Characterization of a novel mesophilic CTP-dependent riboflavin kinase and rational engineering to create its thermostable homologs

Yashwant Kumar[1], Reman Kumar Singh[1, 3]*, Amrita B. Hazra[1,2]*

[1]Department of Chemistry, Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pashan, Pune 411008, India

[2]Department of Biology, Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pashan, Pune 411008, India

[3]Current affiliation: Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400 076

*Address correspondence to amrita@iiserpune.ac.in and reman.iitb@iitb.ac.in

ABSTRACT

Flavins play a central role in cellular metabolism as cofactors that are engaged in a wide range of oxidation-reduction reactions in living organisms. Several interesting variations exist in the flavin biosynthesis pathway among the domains of life, and the analysis of enzymes on this pathway have resulted in unique structural and mechanistic insights. The CTP-dependent riboflavin kinase in archaea is one such example - unlike most kinase enzymes that use adenosine triphosphate for conducting phosphorylation reactions, riboflavin kinases from tarchaea utilizes cytidine triphosphate (CTP) to phosphorylate riboflavin to produce flavin mononucleotide (FMN). In this study, we present the characterization of a mesophilic archaeal riboflavin kinase homolog from *Methanococcus maripaludis* (*Mmp*RibK), which is linked closely in sequence to the thermophilic RFK homolog from *Methanocaldococcus jannaschii* (*Mj*RibK). We reconstitute the activity of the CTP-dependent *Mmp*RibK and analyse the molecular factors that contribute to the uncommon properties of this class of enzymes. Furthermore, based on the high degree of sequence similarity between the mesophilic *Mmp*RibK and the thermophilic *Mj*RibK, we establish a set of the residues that are responsible for the thermostability of the enzyme without any loss in activity or substrate specificity. Our work contributes to the molecular understanding of flavin biosynthesis in archaea through the characterization of the first mesophilic CTP-dependent riboflavin kinase. Finally, it validates the role of salt bridges and rigidifying amino acid residues in imparting thermostability to enzymes, with implications in enzyme engineering and biotechnological applications.

INTRODUCTION

Flavin cofactors such as flavin mononucleoide (FMN) and flavin adenine dinucleotide (FAD) are utilized by living organisms in primary metabolism. The biosynthesis of flavins include several steps that are mechanistically intriguing. Especially in archaea, the enzymes involved in the flavin biosynthesis pathway differ significantly in their sequence, structure, and mechanism as compared to their bacterial and eukaryotic counterparts.[1–3] For example, the first enzyme GTP cyclohydrolase in archaea carries out a ring-opening hydrolysis on the

purine imidazole ring of GTP to produce a formlyated intermediate unique to archaea, which is then converted to the diaminopyrimidine product by a formamide hydrolase enzyme.[4,5] Another unusual example is the archaeal riboflavin synthase that catalyzes a dismutation reaction with two molecules of 6,7-dimethyl-8-ribityllumazine in the step leading to the formation of riboflavin. The pentacyclic adduct intermediate formed by the archaeal riboflavin synthase is a diastereomeric flavour of the corresponding bacterial adduct.[6] The example that we analyse in this study is the archaeal riboflavin kinase, which converts riboflavin (Vitamin B_2) into FMN via a phosphorylation reaction. This class of enzymes is unique in several ways: (a) in most known homologs of riboflavin kinase, this conversion is effected by the use of ATP. However, in the case of riboflavin kinase from the thermophilic archaea *Methanocaldococcus janaschii* (*Mj*RibK), CTP is used as a source of phosphate, and is specific for its utilization.[2,7] The only perceptible difference between the various nucleotide triphosphates is the nucleobase, thus this structural fold and class of enzymes appears to have adapted to selecting cytosine over other nucleobases. (b) The characterization of *Mj*RibK showed activity at 70°C without the requirement of a divalent metal ion for its function, which was intriguing. Divalent metal cations such as Mg^{2+}, Mn^{2+} and Zn^{2+} play an important role of stabilizing the negatively charged triphosphate group of the nucleotide and assisting in catalysis, with metal-independent behaviour only being exhibited by a handful of kinases/ pseudokinases.[8–10] (c*) Mj*RibK has a sequence and fold that is unique with no similarity to any other known class of kinase, including a lack of the conventional phosphate binding motif in its active.[7] Such curious features make this class of CTP-dependent riboflavin kinases uniquely distinct from other known kinases and lend itself to mechanistic investigations and enzyme engineering applications. However, as *Mj*RibK shows optimal activity at higher temperatures, it is not well-suited to extensive mechanistic explorations and has limited scope under physiological conditions. As the only reported CTP-dependent

riboflavin kinases in literature are either from hyperthermophiles or thermophiles, we launched a search for a suitable mesophilic candidate that would be amenable to our studies.[2,7,11] Bioinformatic and phylogenetic analysis revealed that *Methanococcus maripaludis*, a closely related mesophilic archaea, contains a homolog of the CTP-dependent riboflavin kinase (*Mmp*RibK). In this study, we report the characterization of *Mmp*RibK. We explore the functional differences between the *Mmp*RibK and *Mj*RibK homologs using biochemical tools and molecular modelling and establish the reason for several of the features that are characteristic of this class of enzymes. Further, we identify key amino acid residues responsible for the difference in activity of *Mj*RibK and *Mmp*RibK at different temperatures and create mutants of *Mmp*RibK which can operate up to 15°C higher than the wild-type enzyme without loss of activity or substrate specificity. Our findings explain some of the unique properties of the archaeal CTP-dependent riboflavin kinases and establish the role of salt bridges made by oppositely charge pairs of residues such as Glu-Lys, Asp-Arg and Asp-Lys and residues that confer rigidity (such as prolines) in creating enzymes that can operate at higher temperatures.

MATERIALS AND METHODS

Materials: Genomic DNA of *Methanocaldococcus janaschii* DSM 2661 and *Methanococcus maripaludis* S2 were a gift from Biswarup Mukhopadhyay at Virginia State University. The primary assay reagents riboflavin (RF) 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 and purification and enzyme assays were obtained from Sigma-Aldrich, TCI chemicals, Rankem, SRL chemicals and Himedia, unless otherwise stated. Polymerase chain reaction (PCR) reagents and kits were from Agilent and Himedia, and the high fidelity PrimeSTAR GXL DNA polymerase from

DSS Takara Bio USA was used for cloning. Sypro-orange dye used in the thermal shift assay was obtained from Thermofisher Scientific.

Bioinformatic and phylogenetic analysis for selection of *Mmp*RibK as the candidate: To identify a suitable mesophilic CTP-dependent riboflavin kinase homolog for our studies, the FASTA protein sequence of *Mj*RibK (Swiss-Prot accession number Q60365) was used to perform a Basic Local Alignment Search Tool (BLAST) analysis.[12–14] This yielded 1813 distinct 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 the list of BLAST search candidates. To do so, we obtained 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.[15,16] The transcriptional regulator domains were removed using BioEdit, so as to only consider the riboflavin kinase domain sequence for analysis. Finally the phylip4.0 alignment file was 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 bacterial were also conducted using the method described above.

Molecular cloning, overexpression and purification*: MjribK* and *MmpribK,* the riboflavin kinase genes from *Methanocaldococcus jannaschhi* and *Methanococcus maripaludis* S2, respectively, were amplified using PCR and inserted between the NdeI and BamHI restriction sites of pET21a plasmid vector using restriction-free cloning.[17] The plasmids containing the *MjribK* and *MmpribK* genes were transformed into *E. coli* BL21(DE3) competent cells and an expression check was conducted to verify whether the corresponding proteins were being expressed. Autoclaved Luria Bertani broth supplementated with 50 µg/mL of kanamycin was

inoculated with 1% of overnight grown primary culture and left at 37°C inside the incubator shaker. When the OD600 of culture reached 0.6-0.8, protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18°C and left to shake for 15 h, after which the cells were centrifuged to pellets and stored at -80 degrees. Purification of 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 buffer containing 50 mM tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole 250 mM imidazole and 0.025% β-mercaptoethanol. The concentration of protein was checked using Bradford assay using BSA as standard.[18,19] All mutants of *Mmp*RibK were cloned, expressed and purified in the same manner. The primers used for the cloning of wild-type and mutants and the plasmids created are given in the supplementary information Table S2 and S3, respectively. The purity and molecular weight of the enzymes were analyzed by 15% SDS-PAGE gel electrophoresis (Figure S2 A). 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 and its mass was verified to be 16.37126 kDa (corresponding to the calculated mass of the *Mmp*RibK protein with the hexahistidine tag.(Figure S2B))

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 previous reported conditions for the assay of *Mj*RibK.[2] The reaction mixture was incubated at 37°C overnight and visualized using fluoresce-based thin layer chromatography (TLC) (Merck Silica gel 60 F_{254}) using the long wavelength (365 nm) of handheld UV lamp (Spectroline E-series UV lamp) nm with the solvent system - water: acetic acid: n-butanol in 5:1:4 ratios.[20] 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 5µm NX-C18 110 Å, 250 × 4.6 mm). The HPLC method used for analysis was derived from the one previously used to assay *Mj*RibK.[2] 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 mL/min 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 *Mmp*RibK were assayed and analyzed in a manner identical to the wild-type enzyme, unless otherwise mentioned.

The functional characterization and optimization of *Mmp*RibK using NaCl, alternate nucleotides, different metal ions, varying temperatures pH and guanidinium hydrochloride assisted chemical denaturation were conducted as described in the supplementary information.

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. From the plot of the rate, initial velocities at different concentrations are determined. The Michaelis-Menten and Lineweaver-Burk data were plotted with Origin 9, using nonlinear fitting with the Hills

algorithm, $v = V_{max} \frac{x^n}{k^n + x^n}$ where n is the number of cooperative sites. For single-substrate model, n is fixed to 1.

Structural analysis and ligand docking: The PDB structure of the wild type and mutant *Mmp*RibK enzymes were modelled using Modeller.[21] The stereochemical quality of the modelled structures was estimated using programs Procheck and.[22,23] All structural analysis were done using molecular visualization softwares Visual Molecular Dynamics (VMD) and UCSF Chimera.[24,25] 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 are done using Autodock 4.[26]

Molecular Dynamics simulations: We docked CTP and riboflavin in the MmpRibK structure and then performed a molecular dynamics (MD) simulation. We used AMBER99SB force field for the protein. However, to create the amber force field of CTP and riboflavin we used antechamber module of AmberTool.[27] First, we conducted a quantum mechanical optimization of the CTP using HF/6-31G* basis set in GAUSSIAN03 software and then restricted electrostatic potential (RESP) charges on the atoms of CTP and RF were calculated. The topology and co-ordinates generated from AmberTools were converted into the GROMACS format using a perl program amb2gmx.pl. We put the CTP-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 300K 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 and at 372 K using Nose-Hooverthermostat. All the analysis is done over production run.

Thermal shift assay: The thermostability of mutants has been tested using thermal shift assay.[28] 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 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 enzyme.

RESULTS AND DISCUSSION

Bioinformatic and phylogenetic analyses reveal that *Mmp*RibK is one of the closest mesophilic homologues of the thermophilic enzyme *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* (ST 1). Next, we constructed a phylogeny of the curated set of the archaeal riboflavin kinases we obtained from the BLAST search. 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 sequence of *Mmp*RibK and *Mj*RibK are 50.4% identical inspite of *Mmp*RibK belonging to a mesophilic archaea (Figure 1A, Table S1). Based on these results, we selected *Mmp*RibK as our candidate mesophilic CTP-dependent riboflavin kinase for further analysis.

NCBI conserved domain search shows that *Mmp*RibK belongs to the superfamily pfam01982 and 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 regulator/ riboflavin kinases.

A comparison of archaeal riboflavin kinase sequences (including *Mmp*RibK) with their bacteria and eukaryotic homologs revealed that these are two distinct types of enzymes, differing even in the conserved motif responsible for triphosphate binding (GTLN in archaeal versus PTAN in non-archaeal riboflavin kinases), as noted previously (Figure 1B, Figure S1A).[7] However, among the archaeal riboflavin kinases, a large degree of conservation is observed especially within the catalytic domains, as evidenced by the sequence alignment of *MmpRibK* S2 strain with other homologous archaeal enzymes (Figure 1C, Figure S1B). The archaeal riboflavin kinases contain the GTLN domain and the glycine-rich GxGEG loop, responsible for nucleotide recognition and binding, mainly via the phosphate group. Another interesting feature is that the conserved sequence LRxxxL loops around the nucleobase portion of CTP and contains a small α-helix which is not present in the corresponding loop of ATP-dependent riboflavin kinases. The nucleobase cytosine makes backbone interactions with Leu and Arg residues of LRxxxxL domain. Finally, the conserved riboflavin binding domain PxxTxx is found to have mainly backbone interactions with riboflavin.[7]

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

We cloned, expressed and purified the enzymes *Mj*RibK and *Mmp*RibK (Figure S2). 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 $^{\circ}$C. No conversion of riboflavin 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 2A, B).

Next, to compare the activity of the *Mmp*RibK and *Mj*RibK homologs, we carried out reactions using different nucleotide triphosphates under conditions reported previously.[2] 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. On the other hand, *Mj*RibK shows the highest specificity with CTP, but shows 10% and 6% conversion of riboflavin to FMN with UTP and ATP, respectively (Figure S3 A,B)). Next, when boiled at 100°C for 5 min, *Mmp*RibK lost its activity completely while *Mj*RibK has no significant change in activity. Lastly, *Mmp*RibK requires Mg^{2+} ion for its catalytic activity, as compared to *MjRibK* which is unique in its ability to catalyze the phosphate transfer reaction in absence of any divalent metal ion (Figure S4A).

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

The sequence of *Mmp*RibK was subjected to homology modelling, and the resulting structure was compared to the crystal structure *Mj*RibK (PDB: 2VBV). The modelled apo *Mmp*RibK was established to have high similarity to *Mj*RibK. Next, we docked the ligands CTP and riboflavin into the *Mmp*RibK structure and conducted energy minimization to obtain the most favourable configuration. Figure 2C shows the overlap of the docked *Mmp*RibK with *Mj*RibK containing CTP and riboflavin. Both the apo- and docked *Mmp*RibK models show >90% of residues are in the most favoured regions of the Ramachandran plot, which indicates that these are high quality models (Figure S6 A, B). [22] The overall quality factor of the modelled structures were also calculated on the basis of statistics of non-bonded atomic

interactions and distribution of atoms using ERRAT and found to be 68.9% and 82.4% for the apo and docked models of *Mmp*RibK, respectively. (Figure S6 C,D).[23] The overall quality factor for the *Mj*RibK is also shown in for comparison (Figure S6 D). Interestingly, molecular dynamics simulations of the whole protein *Mmp*RibK as well as that of the CTP-binding site within show that the root mean square deviation (RMSD) decreases sharply when CTP and riboflavin are docked in the model, as compared to the apo *Mmp*RibK structure (Figure 2C inset). Theoretically, this indicates that *Mmp*RibK acquires a more favorable conformation in the presence of CTP. Our findings were verified experimentally by a thermal shift assay, in which the broader depression of the apo WT *Mmp*RibK changed into sharp depression on adding CTP (Figure S2C).[28] 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 catalyse the riboflavin kinase reaction with ATP, but also indicates that ATP is unable bind in the enzyme's nucleotide binding site.

Experimental and theoretical analysis of the activity of *Mmp*RibK and *Mj*RibK show that the former exhibits greater molecular flexibility and lower stability under a range of temperatures

We began our functional analysis of the enzyme *Mmp*RibK by testing its ability to show activity at varying temperatures. *Mmp*RibK was found to be most active in the temperature range of 40°C-50°C, after which its activity declined and then was completely inactivate after 70°C (Figure 3A). In a similar experiment with *Mj*RibK, we observed the optimal activity extending from 45°C to 100°C (Figure 3A). Next, the stability of *Mmp*RibK and *Mj*RibK enzymes were probed using chemical denaturation experiments. Both enzymes were subject to varying concentrations of denaturing agent guanidium hydrochloride (Gdn-HCl) and their activities were analyzed using fluorescence-TLC (Figure S7). We observed

that *Mmp*RibK tolerates concentrations of Gdn-HCl only up to 1.2 M, while there is no significant change is activity of *Mj*RibK up to 2.5 M Gdn-HCl. 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 destabilizing its structure (Figure 3B). In sharp contrast, we find that the RMSF of *Mj*RibK remains similar at temperatures of 27°C and 72°C. Taken together, our experimental results and MD simulations strongly suggests 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.

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

We began by conducting a kinase activity assay with *Mmp*RibK without any added metal ion, based on the reported observation that *Mj*RibK functions in the absence of any metal cofactor.[2] *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 S4A). 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+}, we found that *Mmp*RibK is able to phosphorylate riboflavin most effectively in the presence of Mg^{2+} and Zn^{2+} (Figure S4B). 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 S4B). This preference

is reminiscent of the rat liver riboflavin kinase which is reported to exhibit a preference of Zn^{2+} ion followed by Mg^{2+} ion.[29]

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 enzyme up to pH 11 (Figure S4C). A similar trend where increase in pH leads to higher activity has been previously noted for the *Neurospora crassa* riboflavin kinase.[30]

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 in comparison to *Mj*RibK. As *Methanococcus 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 S4D). To our knowledge, this is the first riboflavin kinase which has been shown to require a high concentration of NaCl.

Determination of kinetic parameters

Next, we conducted the steady state kinetic analysis of *Mmp*RibK. In the presence of saturating concentration of CTP, riboflavin displays the follows kinetic parameters: V_{max} 5.92 µM min^{-1}, K_m 185.8 µM and k_{cat} 0.197 s^{-1} (Figure 4A, Figure S5A). Under saturating concentration of riboflavin, the kinetic parameters for CTP are: V_{max} =3.85 µM min^{-1}, K_m= 235.96 µM and k_{cat} = 0.141 s^{-1} (Figure 4B, Figure S5B). We summarized and compared these values to the parameters reported for *Mj*RibK and established that they were quite similar (Table 1).[2]

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.

Molecular basis of thermostability of *Mj*RibK and rational engineering of *Mmp*RibK to make it thermostable

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 excellent candidates for analysing 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 *MmpRibK* and *MjRibK* structures. The *Mj*RibK had more salt bridge interactions than *Mmp*RibK, including one within the nucleobase loop region in its active site (Table 2, Figure 5 A-C). This is expected, as salt bridges are noted to confer rigidity and subsequently thermostability to proteins.[31] In order to test these predicted salt bridges, we introduced the salt bridge forming residues from *Mj*RibK into the *Mmp*RibK sequence, and created three mutants (Mut 1-3) (Table3, Figure 5 A-C). Additionally, insertion of proline residues are known to confer thermostability in a mesophilic enzyme.[32] Based on the sequence alignment of *Mj*RibK and *Mmp*RibK, we replaced residues T21 and H22 in *Mmp*RFK by two consecutive prolines to create Mut4 (Table3, Figure 5D). Mut 1-4 were tested for activity at 60 °C, where the wild-type enzyme drastically loses activity. Except Mut 1, the remaining three mutants showed similar or better turnover as compared to the wild-type enzyme. Hence, we created combinations of Mut 2-4 involving salt bridges and proline residues to generate Mut5-8. (Table 3).

Thermal shift analysis and specific activity of the *Mmp*RibK mutants validate the role of a subset of amino acids within the enzyme:

We performed a thermal shift analysis with wild-type *Mmp*RibK and its mutants in order to establish their melting temperatures. 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 6A). Our results show that Mut1 (with a 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) and 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 6A, Figure S8, Table 3). The remaining combination mutants Mut 5-8 all displayed melting temperatures higher than the wild-type enzyme, but surprisingly, they were lower than Mut 2 and Mut 3 (Figure 6A).

Since the thermostabilization of an enzyme may occur at the cost of its activity, we next 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° and 70°C (Figure 6B). WT *Mmp*RibK has reduced activity at 60°C and 70°C in comparison to that at 45°C (also seen in Figure 3A). At 45°C, the activities of WT *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 an elevated temperature of 60°C, all mutants except Mut1, have higher activity when compared to WT *Mmp*RibK. The highest activity is displayed by Mut5 which contains two salt bridges (Table 3). The WT and mutants have significantly diminished activity at 70°C. 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.

Discussion

The cofactor biosynthesis pathways in archaea have unravelled many interesting enzymes and reaction mechanisms to date, and the CTP-dependent riboflavin kinase is one such enzyme. Until now, only one thermophilic homolog *Mj*RibK had been characterized. In this study, we report the reconstitution of the activity of the first mesophilic CTP-dependent riboflavin kinase homolog *Mmp*RibK, and establish its biochemical properties.

On testing for various reaction parameters, we found that *Mmp*RibK varies in several ways from *Mj*RibK. A feature of *Mj*RibK that stands out is that it does not require a metal ion for its function.[2] In contrast, we found that *Mmp*RibK requires a metal ion for activity, and gives the best activity with Mg^{2+} and Zn^{2+}. As most kinases use a metal ion to stabilize the nucleotide triphosphate within its active site and to catalyse the phosphate transfer reaction, we tried to understand the role of the metal ion in *Mmp*RibK using a molecular dynamics simulation with the enzymes, substrates, and without and with $MgCl_2$. Our results show that for *Mmp*RibK, the absence of Mg^{2+} results in weak binding of CTP in the active site (Figure S10 A and B), and a larger number of hydrogen bonds stabilizing the CTP in the active site in the presence of Mg^{2+} (Figure S10 C and D). Additionally, a constant distance is being maintained between the Mg^{2+} and the phosphate residues of CTP indicating that the metal ion is important for its stabilization (Figure S10 E). In contrast, for the simulation with *Mj*RibK, the Mg^{2+} is unable to stabilize in the active site in the absence or presence of CTP and riboflavin (data not shown)

*Mmp*RibK and *Mj*RibK share a high degree of sequence identity as evidenced by their proximity on the phylogenetic tree and from the similar proportions of individual amino acids of which they are made (Figure 1A, Table S5). Interestingly, the significant differences in the mol% of amino acid residues are observed in the ones that are reportedly involved in salt

bridges (Arg, Lys, Asp, Glu, His) and in providing rigidity to the protein (Pro)(Figure S9).[33] As expected, the thermophilic *Mj*RibK has more salt bridge-forming and rigidity-conferring residues as compared to the mesophilic *Mmp*RibK with the exception of His.[31] We also experimentally observed that *Mmp*RibK functions optimally at a temperature much lower than *Mj*RibK. So, we probed the biochemical basis for thermal stability of *Mj*RibK using bioinformatics and computational tools and introduced specific thermostability elements i.e. salt bridge residues and Pro to confer rigidity at the corresponding positions in the *Mmp*RibK enzyme. By this process, we successfully created mutant variants of *Mmp*RibK that show optimal activity at 60 °C i.e. 15 °C higher than the wild-type enzyme. Lastly, as *Mmp*RibK is a riboflavin kinase that is specific for its use of CTP as its phosphate donor, we investigated whether the thermostable mutants we had created had either lost activity or had lost their specificity for CTP at 60 °C. We find that with the exception of one mutant, all the other mutants we created had similar or higher activity than the wild-type *Mmp*RibK and continue to show a high specificity for CTP even at 60 °C (Figure 6B). Taken together, our results establish that our method of rational engineering the amino acids of a mesophilic protein homolog by comparing it with the corresponding thermophilic homolog using theoretical tools, followed by systematically introducing them in the mesophilic enzyme is a reliable general strategy to produce thermostable enzymes for various applications.

Conclusion

In this study, we report the 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, and the only homolog characterized so far in this class of enzymes is from the thermophilic archaea *M. janaschii*. Our studies on the characterization of *Mmp*RibK and creation of its thermostable mutants not only adds

important information to this class of enzymes, but also puts forth an important candidate for mechanistic and structural investigations and bioengineering applications in the future.

Acknowledgements

We are grateful to Dr. Biswarup Mukhopadhyay for his gift of the genomic DNA of *M. janaschii* and *M. maripaludis* used in this study. We thank Yamini Mathur, Rupali Sathe, Ateek Shah, Bappa Ghosh and Nishant Singh for their critical inputs on experiments and their execution. We thank Harinath Chakrapani, Santosh Kumar Jha, Arnab Mukharjee, and Gayathri Pananghat for insightful discussions. We thank Mohan Reddy at IISER Pune for his help with MALDI-TOF analysis and the help received from the LC-MS and MALDI-TOF facilities and their managers at IISER Pune. Contributions by summer student Lavisha Parab and semester project student Saswata Nayak are also acknowledged. Y.K. acknowledges IISER Pune and University Grant Commission, India, and R.S acknowledges CSIR and IIT Bombay postdoctoral fellowship for funding. .

Funding:

The work was supported by grants from the Ministry of Science and Technology, Government of India - Department of Science and Technology SERB Core Research Grant **CRG/2019/003270** to A.B.H. and the Department of Biotechnology (DBT)-Ramalingaswami Re-entry fellowship to A.B.H.

References

1. Sato, T. & Atomi, H. Novel metabolic pathways in Archaea. *Curr. Opin. Microbiol.* **14**, 307–14 (2011).

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

3. 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.* **75**, 321–360 (2011).

4. Graham, D. E., Xu, H. & White, R. H. A member of a new class of GTP cyclohydrolases produces formylaminopyrimidine nucleotide monophosphates. *Biochemistry* **41**, 15074–84 (2002).

5. Grochowski, L. L., Xu, H. & White, R. H. An iron(II) dependent formamide hydrolase catalyzes the second step in the archaeal biosynthetic pathway to riboflavin and 7,8-didemethyl-8-hydroxy-5-deazariboflavin. *Biochemistry* **48**, 4181–8 (2009).

6. 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.* **280**, 28541–6 (2005).

7. Ammelburg, M. *et al.* A CTP-dependent archaeal riboflavin kinase forms a bridge in the evolution of cradle-loop barrels. *Structure* **15**, 1577–90 (2007).

8. Mukherjee, K. *et al.* CASK Functions as a Mg2+-independent neurexin kinase. *Cell* **133**, 328–39 (2008).

9. Baas, A. F. *et al.* Activation of the tumour suppressor kinase LKB1 by the STE20-like pseudokinase STRAD. *EMBO J.* **22**, 3062–72 (2003).

10. Gerlits, O. *et al.* Metal-free cAMP-dependent protein kinase can catalyze phosphoryl transfer. *Biochemistry* **53**, 3179–86 (2014).

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

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

(2008).

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

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

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

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

17. 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).

18. 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).

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

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

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

22. 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).

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

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

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

26. No Title.

27. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T, S. F. No Title.

28. 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).

29. 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).

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

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

32. 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**, 277–83 (1994).

33. Lecocq, M., Groussin, M., Gouy, M. & Brochier-Armanet, C. The Molecular Determinants of Thermoadaptation: Methanococcales as a Case Study. *Mol. Biol. Evol.* (2020) doi:10.1093/molbev/msaa312.

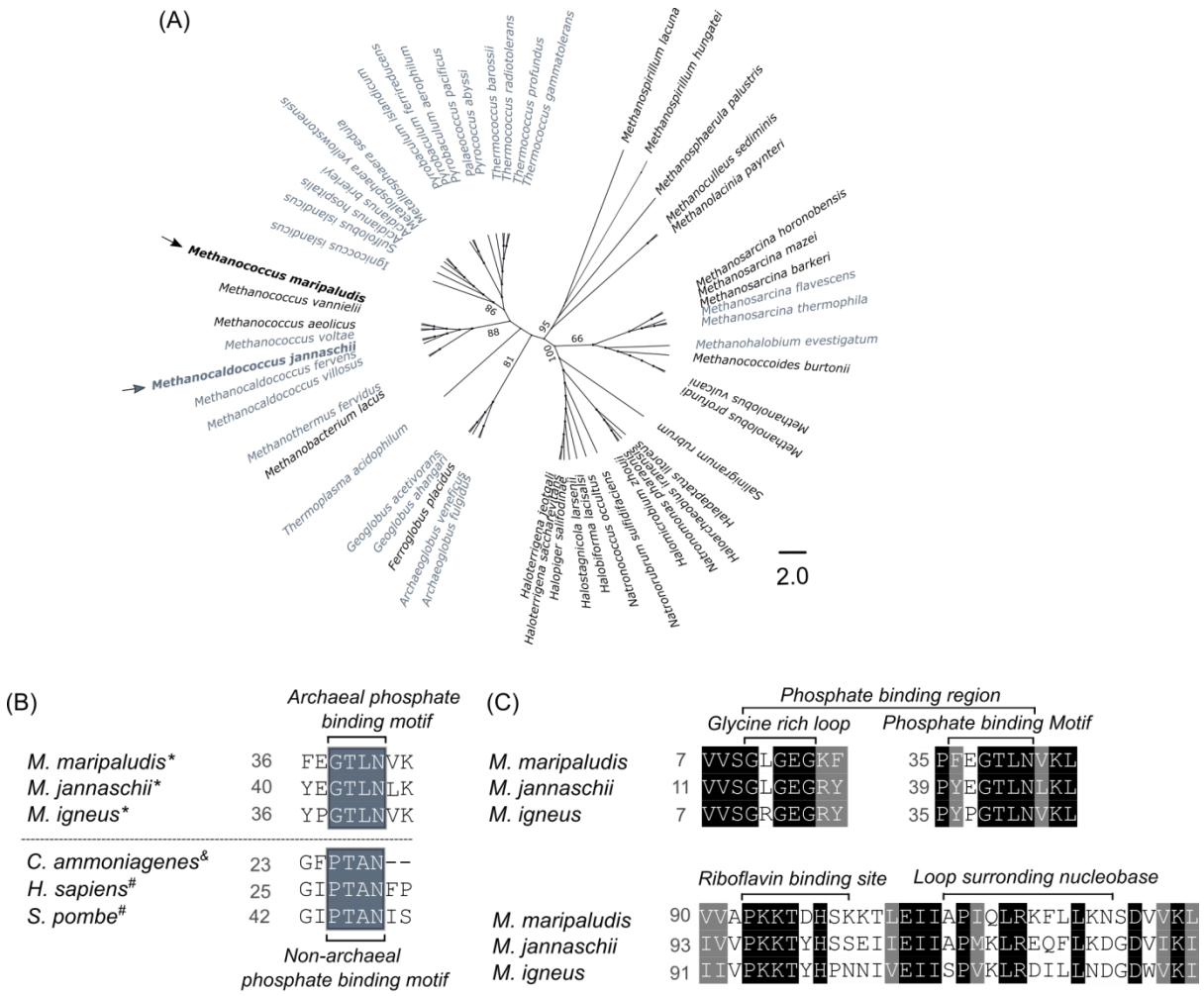

Figure 1: Phylogeny and multiple sequence alignment of riboflavin kinases. **(A)** We constructed a phylogenetic tree of archaeal riboflavin kinase homologs. The grey entries and black entries represent riboflavin kinase sequences derived from thermophilic/hyperthermophilic and mesophilic organisms, respectively. Arrows indicate the thermophilic *Methanocaldococcus janaschii* RibK (*Mj*RibK) and the uncharacterized mesophilic *Methanococcus maripaludis* RibK (*Mmp*RibK). **(B)** Multiple sequence alignment of archaeal and non-archaeal riboflavin kinases indicate distinct phosphate binding motifs – GTLN in the case of archaea and PTAN in the case of bacteria. **(C)** The phosphate-binding domain, riboflavin binding site, and the region surrounding the nucleobase are well conserved across archaeal sequences.

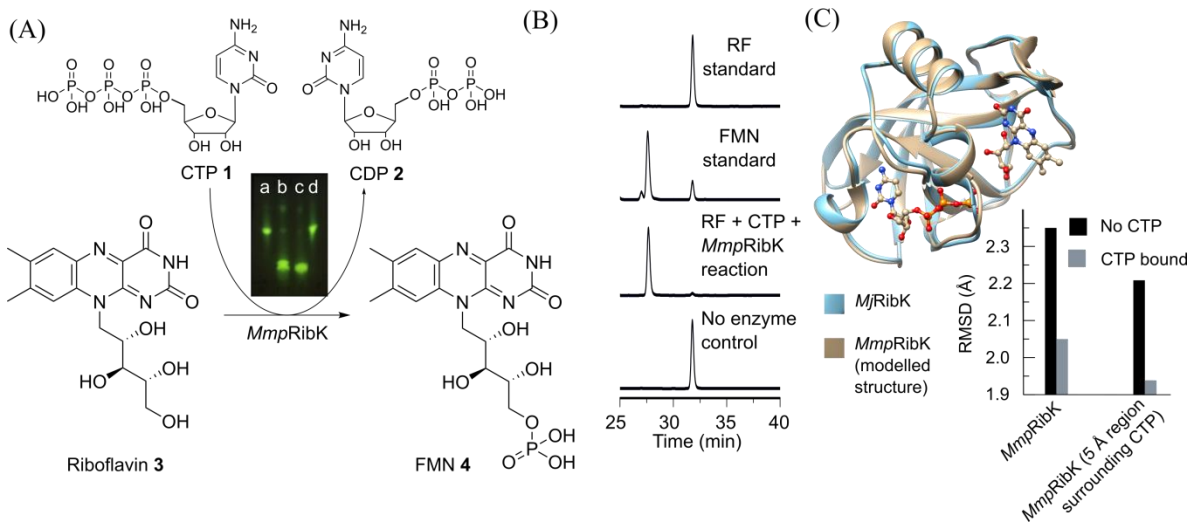

Figure 2: Characterization of the riboflavin kinase activity of *Mmp*RibK, modelling the structure of *Mmp*RibK , and substrate docking **(A)** Reaction catalyzed my *Mmp*RibK to convert riboflavin into FMN. TLC shows the fluorescent spots for (a) riboflavin standard, (b) FMN standard, (c) full reaction, and (d) No enzyme control (d). **(B)** HPLC chromatogram indicating the formation of FMN by *Mmp*RibK. **(C)** The *Mj*RibK (cyan) and modelled *Mmp*RibK (tan) structures superimposed over one another show a high degree of similarity in the tertiary stucture.The modeled structure of *Mmp*RibK (tan) with docked substrates riboflavin and CTP.

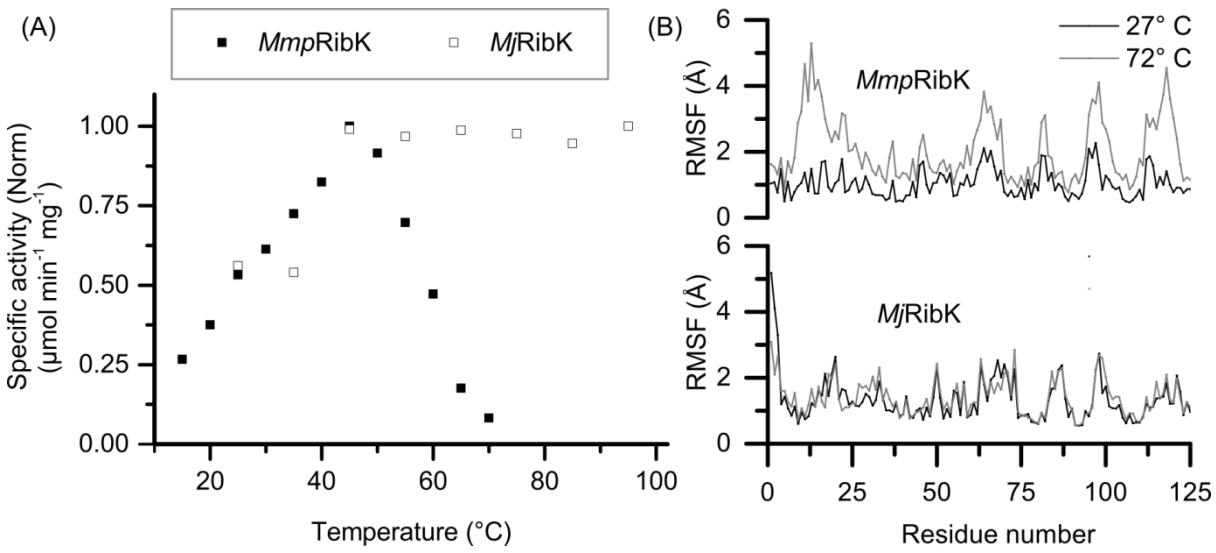

Figure 3: Testing the temperature dependence of *Mmp*RibK and *Mj*RibK: **(A)** The activities of *Mmp*RibK and *Mj*RibK tested from 15°C to 70°C show that the optimum temperature for *Mmp*RibK is 45°C. *Mj*RibK shows good stable activity at temperatures ranging from 45°C to 95°C. The activity of *Mmp*RibK decreases drastically after 65°C and vanishes completely after 70°C. **(B)** 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 70°C indicating that it is , while *Mj*RibK is stable at both temperatures.

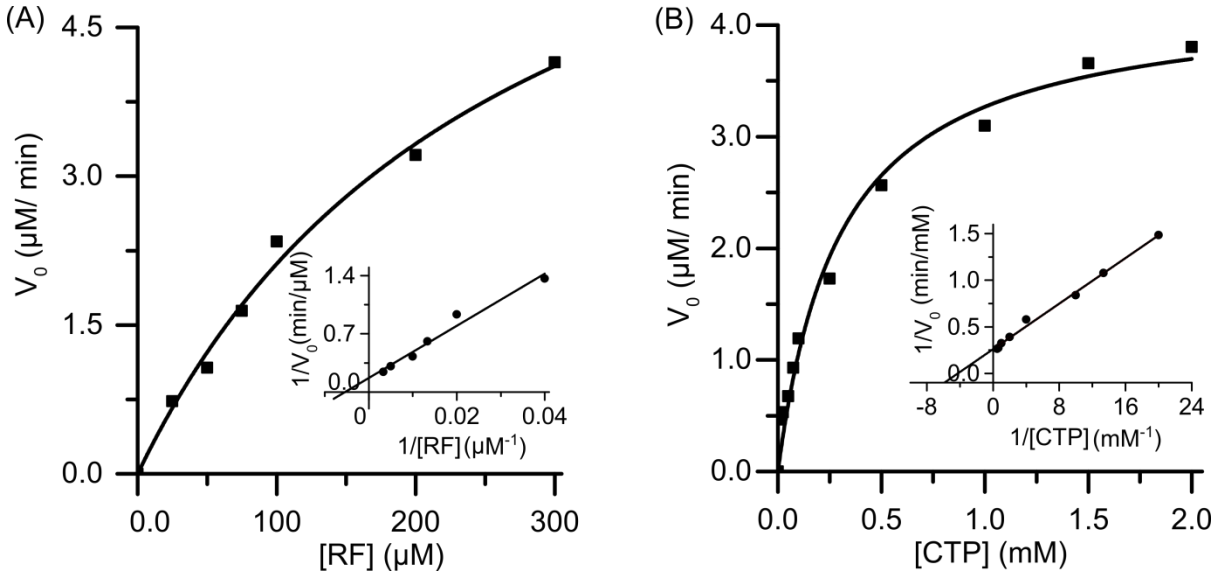

Figure 4: Steady-state kinetics of *Mmp*RibK with substrates riboflavin and CTP: **(A)** Michaelis-Menten plot of *Mmp*RibK for riboflavin under saturating conditions of CTP. We also represent our data as a Lineweaver-Burk plot, shown in the inset. **(B)** Michaelis-Menten kinetics of *Mmp*RibK for CTP under saturating conditions of riboflavin.Lineweaver-Burk plot for the same is given in the inset.

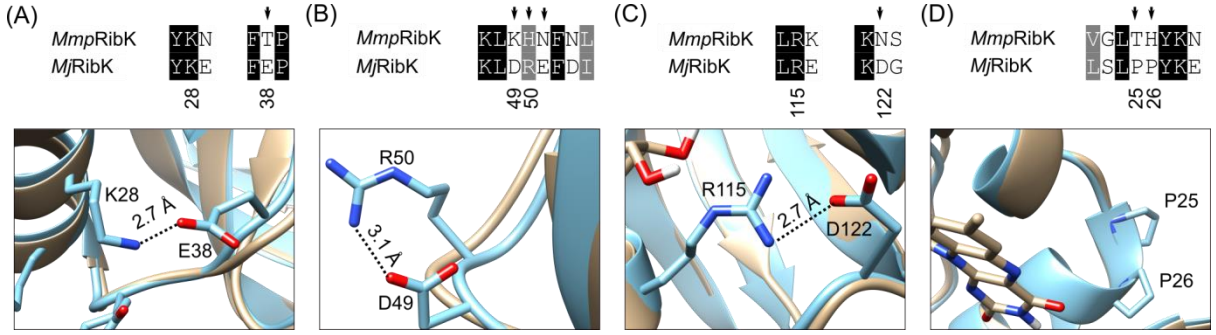

Figure 5: Overlap of modeled structure of *Mmp*RibK and crystal structure of *Mj*RibK (sky blue) depicting the salt bridges and non-conserved proline residues considered for mutation. **(A)** K28-E38 salt bridge in *Mj*RibK corresponding to Mut1 (T34E) **(B)** D49-R50 salt bridge in *MjRibK* corresponding to Mut2 (KHN45-47DRE) **(C)** R115-D122 salt bridge corresponding to Mut3 (N119D) **(D)** Non-conserved proline residues in *Mj*RFK corresponding to Mut4 (TH21-22PP)

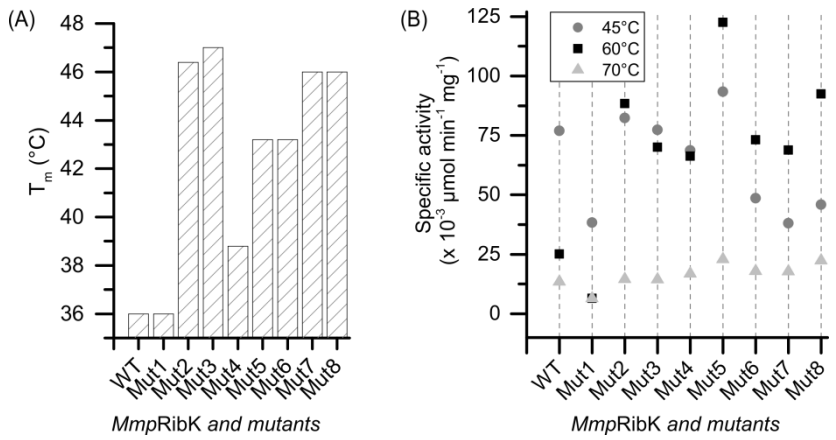

Figure 6: Melting temperature and specific activities of the *Mmp*RibK mutants **(A)** Melting temperatures of the wild-type *Mmp*RibK and its mutants determined by the thermal shift assay **(B)** The specific activities of *Mmp*RibK and mutants measured at 45°C, 60°C, 70°C

Table1: Kinetics parameters for WT*MmpRibK* and WT*MjRibK*

	*Mmp*RibK (this paper)		*Mj*RibK (values obtained from Mashhadi et al., JBact, 2008)	
	Riboflavin	CTP	Riboflavin	CTP
V_{max} (µM min^{-1})	5.92	3.85	Data not available	Data not available
K_m(µM)	185.8	235.96	159	180
k_{cat}(s^{-1})	0.197	0.141	0.334	0.342

Table2: Salt bridges in the *MjRibK* (PDB: 2VBU)

PDB	***MjRibK*: 2VBU**	**Corresponding residues in *MmpRibK***
Ligands	**CDP, MG,**	
# Salt bridges	**Total 8**	

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

Table 3: List of mutants and corresponding mutations

Mutant	**Residue/s of *MmpRibK* replaced with that in *MjRibK***	**Putative salt bridge introduced**	**# of salt bridges introduced**	**# of residues replaced by proline**
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*	NA	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