Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Tanya Joshi

born in: Agra, Uttar Pradesh, India

Oral examination: 11.09.2024

Structural and functional studies on the 8-demethyl-8-amino-riboflavin-5'-phosphate phosphatase RosC from the roseoflavin producer *Streptomyces davaonensis*

Referees: Professor Dr. Matthias Mayer
Professor Dr. Matthias Mack

ABSTRACT

Roseoflavin is an analogue of riboflavin (Vitamin B_2) and shows antimicrobial activities against various Gram-positive bacteria. Roseoflavin targets flavin mononucleotide (FMN) riboswitches and flavoenzymes that require FMN and flavin adenine dinucleotide (FAD) as cofactors, thus negatively affecting riboflavin uptake, riboflavin biosynthesis and cellular metabolism. The roseoflavin biosynthesis pathway converts FMN to roseoflavin using multiple steps catalysed by the enzymes, RosB, RosC and RosA. This study focuses on the enzyme RosC, which catalyses the dephosphorylation of 8-demethyl-8-amino-riboflavin-5′-phosphate (AFP) to 8-demethyl-8-amino-riboflavin (AF) which both are intermediates of roseoflavin biosynthesis. RosC belongs to the histidine phosphatase superfamily. Dephosphorylation activity was observed for AFP and FMN, which are structurally similar compounds, and dephosphorylation of FMN, in combination with an FMN-producing flavokinase, in principle, would generate a futile cycle. RosC was purified to apparent homogeneity from a recombinant *Escherichia coli* strain. Such a preparation was used to set up crystal screens and four different crystal structures were obtained. Crystal structures of flavin-free RosC, RosC in complex with AF (1.7 Å), RosC in complex with AF (1.25 Å) and inorganic phosphate (1.55 Å, mimicking the AFP bound state) and the RosC variant $RosC^{R33A}$ in complex with riboflavin (1.5 Å) were complemented with comprehensive mutational/kinetic studies. Extensive structural and functional studies on this protein indicated the presence of highly specialized and unique segments and arrangements of residues that assist in binding the substrate. Substrate-induced rigidification of the elongated strand-to-helix insertions and the helical N-terminal extension promote substrate binding and ensure substrate specificity. Aromatic residues in the binding pocket mediate stacking interactions with the isoalloxazine ring of the bound flavin. The amino acid, D166 of RosC, was identified as the critical residue that discriminates between the flavin substrates AFP and FMN. D166 binds the isoalloxazine and ribityl groups of AF, and site-specific exchanges at this position drastically reduce the catalytic activity of RosC. Mutations of D166 to hydrophobic residues (valine, isoleucine and leucine) impart toxic properties to RosC when overproduced in *Escherichia coli* and *Corynebacterium glutamicum* but not when overproduced in *Bacillus subtilis*. The reason for this remains unknown. Two other FMN dephosphorylating enzymes in *S. davaonensis* were identified and characterized by gene expression studies.

ZUSAMMENFASSUNG

Roseoflavin ist ein Analogon von Riboflavin (Vitamin B_2) und zeigt antimikrobielle Aktivitäten gegen verschiedene Gram-positive Bakterien. Roseoflavin zielt auf Flavinmononukleotid (FMN)-RNA Schalter und Flavoenzyme ab, die FMN und Flavin-Adenin-Dinukleotid (FAD) als Kofaktoren benötigen, und wirkt sich damit negativ auf die Riboflavin-Aufnahme, die Riboflavin-Biosynthese und den gesamten Zellstoffwechsel aus. Roseoflavin wird ausgehend von FMN mittels der Enzyme RosB, RosC und RosA in mehreren Schritten in Roseoflavin umgewandelt. Diese Dissertation konzentriert sich auf das Enzym RosC, das die Dephosphorylierung von 8-Demethyl-8-amino-riboflavin-5′-phosphat (AFP) zu 8-Demethyl-8-amino-riboflavin (AF) katalysiert. Beide Flavine sind Intermediate der Roseoflavinbiosynthese. RosC gehört zur Superfamilie der Histidinphosphatasen. RosC dephosphoryliert sowohl AFP als auch FMN, die sich strukturell sehr ähnlich sind. Die Dephosphorylierung von FMN in Kombination mit einer FMN-produzierenden Flavokinase würde im Prinzip einen nutzlosen Zyklus („futile cycle“) erzeugen. RosC wurde bis zur Homogenität aus einem rekombinanten *Escherichia coli*-Stamm gereinigt. Ein solches Präparat wurde für Kristallisationsexperimente verwendet. Es wurden vier verschiedene Kristallstrukturen von RosC gelöst. Die Kristallstrukturen von flavinfreiem RosC (1,7 Å), RosC im Komplex mit AF (1,25 Å), RosC im Komplex mit AF und anorganischem Phosphat (den AFP-gebundenen Zustand nachahmend) (1,55 Å) und der RosC-Variante $RosC^{R33A}$ im Komplex mit Riboflavin (1,5 Å) wurden durch umfassende Mutagenese-Experimente ergänzt. Umfassende strukturelle und funktionelle Studien an diesem Protein wiesen auf das Vorhandensein hochspezialisierter und einzigartiger Segmente und Anordnungen von Aminosäureresten hin, die bei der Bindung des Substrats helfen. Die Substrat-induzierte Versteifung der verlängerten Strang-Helix-Insertionen und die helikale N-terminale Verlängerung von RosC fördern die Substratbindung und gewährleisten die Substratspezifität. Aromatische Reste in der Bindungstasche vermitteln Wechselwirkungen mit dem aromatischen Isoalloxazinring der gebundenen Flavine. Die Aminosäure D166 von RosC wurde als der kritische Rest identifiziert, der zwischen den Flavinsubstraten AFP und FMN unterscheidet. D166 ist an der Isoalloxazin-, Ribityl- und Phosphatbindung von AF beteiligt, und der ortsspezifische Austausch an dieser Position führten zu einer drastischen Verringerung der katalytischen Aktivität von RosC. Der Austausch von D166 gegen hydrophobe Reste (Valin, Isoleucin oder Leucin) verleihen RosC toxische Eigenschaften, wenn es in *E. coli* und *Corynebacterium glutamicum* überproduziert wird, nicht aber, wenn es in *Bacillus subtilis* überproduziert wird. Der Grund dafür ist noch unbekannt. Zwei weitere FMN-dephosphorylierende Enzyme in *S. davaonensis* wurden identifiziert und durch Genexpressionsstudien charakterisiert.

Table of Contents

1. INTRODUCTION

1.1 Flavins and flavin binding proteins

Flavins such as FMN and FAD are derivatives of riboflavin and used as cofactors by a number of enzymes due to their redox properties (Zhuang *et al*., 2022). Most of the flavoproteins utilize FAD as a cofactor while fewer of them utilize FMN. Most of these flavin-binding enzymes are classified as oxidoreductases, and the remaining function as transferases, lyases, isomerases, and ligases (Joosten & van Berkel, 2007 , Pimviriyakul & Chaiyen, 2020). Flavins are useful as cofactors due to their ability to acquire different redox states due to the versatility of the isoalloxazine ring. They can acquire three different forms: oxidized (quinone), one-electron-reduced (semiquinone), and two-electron-reduced (hydroquinone) states, acquiring different protonation states and five physiologically relevant chemical forms (Joosten & van Berkel, 2007, Insińska-Rak *et al*., 2020). Flavins can participate both in one-electron and two-electron transfer processes and, therefore, are unique in their redox capabilities as compared to other coenzymes that can only catalyse either one or two electron transfer processes (Edwards, 2014). Most of these flavin states are found in protein-bound forms, not in free forms. Fluorescence exhibited by flavins has been seen to be completely or mostly quenched in the enzyme-bound state. This quenching can be attributed to the presence of small aromatic hydrocarbons and amino acids like tryptophan, tyrosine, histidine, and methionine (Heelis, 1982). The presence of the hydrophobic residues in the vicinity of the flavin lowers its redox potential (Edwards, 2014). The presence of the tryptophan or tyrosine residues, therefore, plays a photoprotective role by reducing the lifetime of excited state and preventing the formation of reactive oxygen species (Insińska-Rak *et al*., 2020). The flavins, due to their bulky isoalloxazine ring, are usually buried in the hydrophobic core of the enzyme. Flavins bind the enzyme active site through non-covalent interactions, primarily through stacking interactions. The isoalloxazine ring is stacked between two hydrophobic residues (Freigang *et al*., 2002, Knaus *et al*., 2012). Stacking interactions involve aromatic residues like phenylalanine, tyrosine, or tryptophan. The aromatic residues provide substrate specificity as well as allow fine modulation of the redox potential of flavin (Hamdane *et al*., 2015). Most flavodoxins have a Tyrosine present at the *re*-face of the flavin and tryptophan at the *si*-face of the flavin; this feature is relatively conserved in the flavodoxin family. At the *si*- face of the flavin, a Tyr residue has been reported in many flavoenzymes, which mediates stacking interactions (Zhou & Swenson, 1996, Hamdane *et al*., 2015, Arakaki *et al*., 2001).

Flavins are thermostable compounds, but they are photosensitive. Riboflavin undergoes rapid degradation in the presence of UV and visible light to form the degradation products- lumichrome and lumiflavin (Edwards, 2014,Insińska-Rak *et al*., 2020). Upon exposure to light, the isoalloxazine ring system undergoes intramolecular photoreduction wherein the ribityl side chain serves as the electron donor. Apart from lumichrome, the rest of the degradation products retain their absorbance characteristics. The change in the absorbance and emission properties of lumichrome can be attributed to the loss of the characteristic alkyl chain at N10 of the heterocyclic ring (Edwards, 2014). Despite this, lumichrome has been found to play an important role in microbial photoelectric systems (Li *et al*., 2023a), quorum sensing (Rajamani *et al*., 2008), and as a signaling molecule to stimulate symbiotic relationships (Matiru & Dakora, 2005) and plant growth (Dakora, 2015).

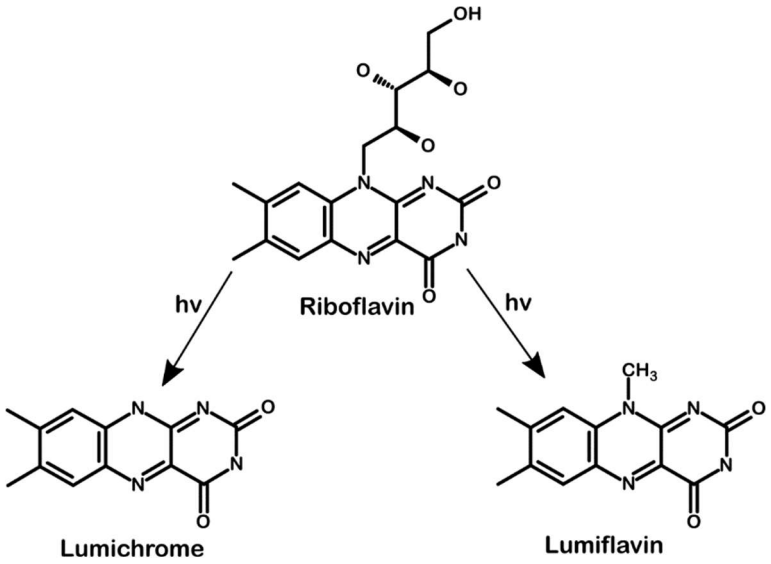


Figure 1: Lumichrome and lumiflavin are photodegradation products of riboflavin.

1.2 Riboflavin biosynthesis pathway

Riboflavin (RF) is a water-soluble yellow pigment that is chemically known as 7, 8-dimethyl-10-ribityl-isoalloxazine. Riboflavin can be synthesized by plants, fungi, and microorganisms, but animals are unable to produce RF due to the absence of the riboflavin biosynthetic pathway and need to acquire it through dietary sources. Riboflavin absorption in humans occurs in the proximal part of the small intestine through a riboflavin transporter (Said *et al*., 2000). Upon uptake, RF is converted to its biologically active forms, FMN and FAD, which act as cofactors in several redox reactions (Mansoorabadi *et al*., 2007). Over the years, metabolic engineering of several microorganisms like *Escherichia coli, Bacillus subtilis*, *Corynebacterium ammoniagenes*, and *Candida* sp. have been done to improve riboflavin biosynthesis (Averianova *et al*., 2020). The

genes involved in the biosynthesis of riboflavin have been widely studied in *E. coli* and *B. subtilis*. *B. subtilis* riboflavin overproducing strains have been engineered and used for large-scale production of riboflavin (Y. Liu *et al*., 2023,C. Liu *et al*., 2024). Due to their enhanced riboflavin production, these strains have found use in increasing the nutritional value of feed, pharmaceutical purposes, and as an additive for food items (Averianova *et al*., 2020). Apart from genes for *de novo* production of riboflavin, many microbes also encode genes for the riboflavin transport system for the uptake of riboflavin from the surroundings (Gutiérrez-Preciado *et al*., 2015). Several bacteria lack a riboflavin uptake system and completely depend upon the riboflavin biosynthetic pathway for riboflavin production. Therefore, the enzymes and regulatory elements of the riboflavin biosynthetic pathway have been regarded as potential targets for antimicrobials (Islam & Kumar, 2023). In most bacteria, the riboflavin biosynthetic genes are clustered together in an operon and are referred to as *rib* genes. The biosynthesis of riboflavin starts with the purine guanosine 5′-triphosphate (GTP)and two molecules of ribulose 5-phosphate (Ru5P). From these precursor molecules, seven enzymatic steps, catalysed by the *rib* genes, lead to the formation of riboflavin. The different *rib* genes and their functions have been summarized in Figure 2. The regulation of expression of the riboflavin biosynthetic genes is under the control of a *cis*-acting regulatory element, the FMN riboswitch. Since riboswitches are genetic elements that regulate important metabolic processes in bacteria but are completely absent in human cells, they are considered promising targets for antibiotics (Fu *et al*., 2022).

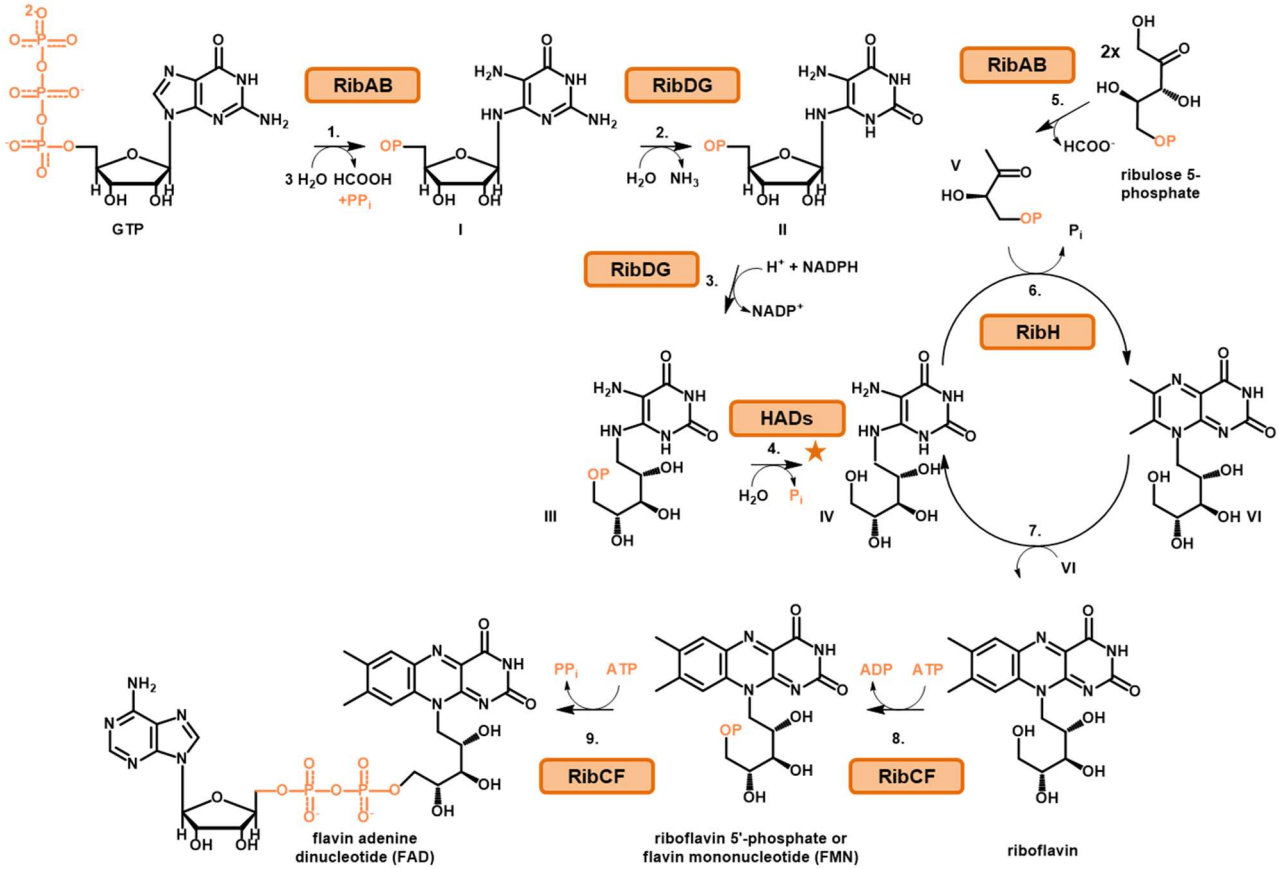


Figure 2: Biosynthesis of riboflavin, FMN and FAD in bacteria.
(Figure adapted from Kißling *et al*., 2020) The compounds are numbered as follows. (I) DARPP, 2,5-diamino-6-ribosyl-amino-4(3*H*)pyrimidinedione-5′-phosphate; (II) ARPP, 5-amino-6-ribosylamino-2,4(1*H*,3*H*)-pyrimidinedione 5'-phosphate; (III) ArPP,5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione-5'-phosphate (IV) ArP, 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione; (V) DHPB, (3*S*)-3,4-dihydroxy-2-butanone 4-phosphate; (VI) 6,7-dimethyl-8-D-ribityllumazine. Riboflavin is synthesized from GTP which comes from the purine biosynthetic pathway. The multi-step reaction is catalyzed by a consortium of enzymes. The first step of the pathway is the conversion of GTP to DARPP (I) *via* bifunctional (as in *Bacillus subtilis*) or monofunctional (as in *Escherichia coli*) RibAB where RibA is GTP cyclohydrolase II. RibB is a 3,4-DHBP synthase that leads to the formation of DHPB (IV) from ribulose-5-phosphate which enters the pathway from the pentose phosphate pathway. The step from conversion of DARPP (I) to ArPP (III) is catalyzed by the bifunctional RibDG. RibD catalyzes the deaminase reaction from DARPP (I) to ARPP(II), and RibG catalyzes the reductase reaction, converting ARPP (II) to ArPP (III). The step includes a deamination of the pyrimidine ring and reduction of the ribosyl side chain (Averianova *et al*., 2020 , Kißling *et al*., 2020). Non-specific phosphatases from the haloacid dehalogenase superfamily catalyze the dephosphosphorylation of ArPP to ArP- marked with an orange star (Sarge *et al*., 2015). RibH is a lumazine synthase that catalyzes the formation of 6,7-dimethyl-8-D-ribityllumazine (VI) by condensation of compounds DHPB (V) and ArP (IV). A dismutation reaction, catalyzed by riboflavin synthase leads to the formation of riboflavin from 2 molecules of compound (VI) (Sebastián *et al*., 2017a). riboflavin kinase (RibC) then converts riboflavin to FMN which is further converted to FAD with the help of FAD synthetase (RibF). Bifunctional (RibCF) FAD synthetases are found in bacteria that catalyze the kinase as well as adenylation steps (Sebastián *et al*., 2017, Bacher *et al*., 2001, Liu *et al*., 2020).

1.3 Roseoflavin and its mechanism of action

Roseoflavin (RoF) is a naturally occurring structural analog of riboflavin (Vitamin B_2) (Otani *et al*., 1974). Roseoflavin production has been studied in *Streptomyces davaonensis*, *Streptomyces cinnabarinus* and *Streptomyces berlinensis* (Matsui *et al*., 1979, Jankowitsch *et al*., 2012 , Liunardo *et al*., 2024). Genomes of two other roseoflavin producers have been published but the isolates themselves have not yet been characterized with regard to roseoflavin biosynthesis. Roseoflavin is taken up by the same transporters that transport riboflavin into the cell. Once taken up roseoflavin is converted to roseoflavin mononucleotide (RoFMN) and roseoflavin adenine dinucleotide (RoFAD) by cytoplasmic flavokinases and FAD synthetases. Due to their structural similarity to FMN and FAD, these modified cofactors bind to the flavoenzymes and reduce their activity (Langer *et al*., 2013a). RoFMN also binds to FMN riboswitches and modulates their activity (Lee *et al*., 2009) (Figure 3). The antimicrobial activity of roseoflavin has been recorded in Gram-positive bacteria like *Bacillus subtilis, Listeria monocytogenes, Staphylococcus aureus and Micrococcus luteus* among others (Otani *et al*., 1974,Otani *et al*., 1980). Overproduction of the riboflavin transporter, RibM, in the *E. coli* strain that was auxotrophic for riboflavin, made it sensitive to roseoflavin, indicating that the antimicrobial action of roseoflavin is dependent on the ability of the bacterial cell to import it into the cell (Langer *et al*., 2013b).

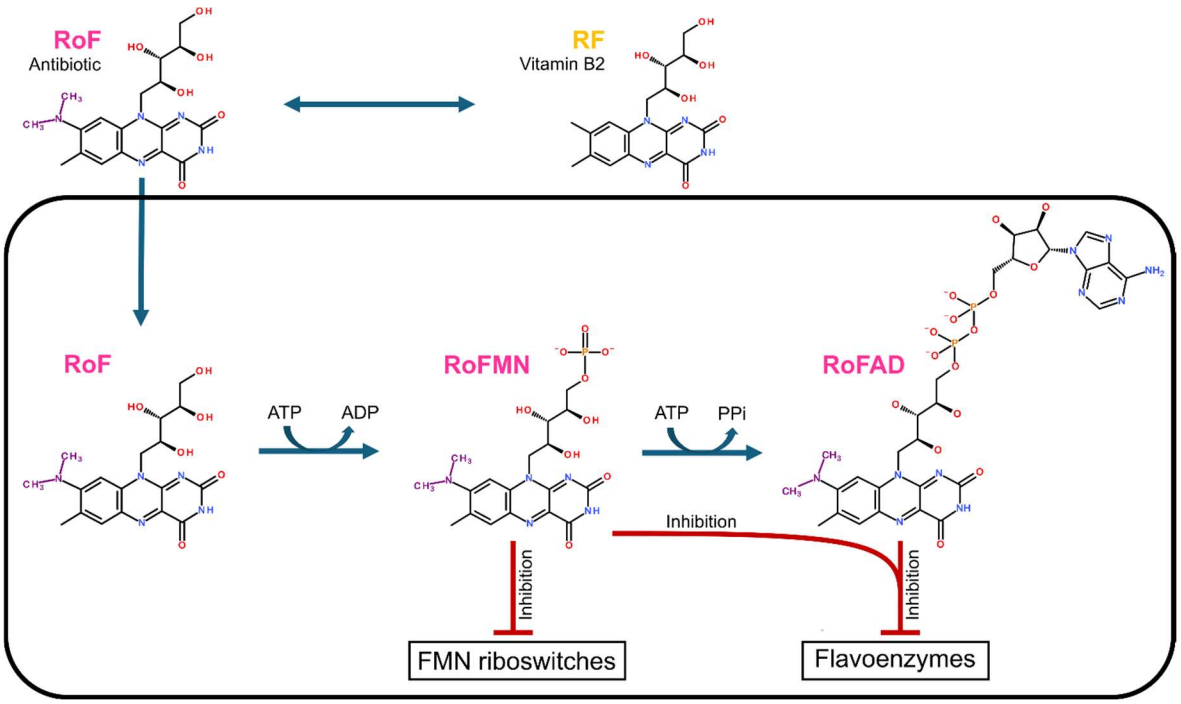


Figure 3: Mechanisms of antimicrobial activity of roseoflavin.
Roseoflavin is taken up by many soil bacteria using riboflavin importers and converted to RoFMN and RoFAD inside the bacterial cells. These cofactors then bind to FMN riboswitches (RoFMN) and flavoenzymes (RoFMN and RoFAD), inhibiting their functions.

1.4 Roseoflavin biosynthetic genes

Roseoflavin producers can be identified by the presence of roseoflavin biosynthetic genes. In *Streptomyces davaonensis*, these genes are located on a 100 kb fragment (Figure 4) and are expressed in the stationary phase.

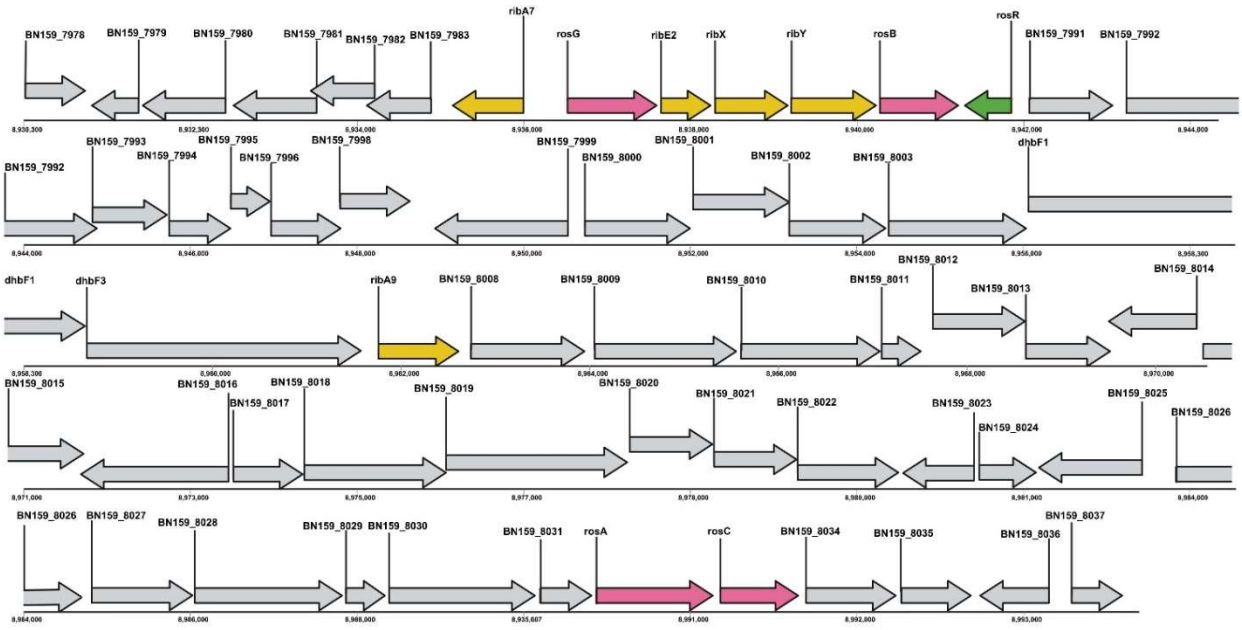


Figure 4: Schematic representation of a subgenomic fragment (nu 8,932,000 to nu 8,994,000) of *Streptomyces davaonensis*.
Roseoflavin metabolic genes (pink) are present in two different clusters. The first cluster (*rosB* cluster) comprises five genes *rosG*, *ribC3*, *ribX*, *ribY* and *rosB*. The gene *rosG* encodes a putative γ-glutamyltransferase/glutathione hydrolase-like enzyme. Possibly, this enzyme generates glutamate which is the amino group donor in the reaction catalyzed by RosB. The riboflavin biosynthetic gene *ribE2* encodes a riboflavin synthase. The genes *ribXY* encode a highly active riboflavin import system, are strongly expressed during the roseoflavin production phase and thus contribute to enhanced cytoplasmic riboflavin levels. The second *ros* gene cluster (*rosA* cluster) comprises 10 genes including the roseoflavin biosynthetic genes *rosA* and *rosC* whereby the functions of the other genes are not known. The riboflavin biosynthetic genes *ribA7* and *ribA9* code for GTP cyclohydrolases II. The gene *rosR* encodes a putative regulator of roseoflavin biosynthesis.

RosA and RosB enzymes have been previously identified and structurally and functionally characterized. *rosA* was the first gene of the RoF biosynthesis to be discovered, which codes for a SAM-dependent dimethyltransferase (EC 2.1.1.B93) (Jankowitsch *et al*., 2011). Gene deletion experiments led to the identification of *rosB* (8-demethyl-8-aminoriboflavin-5′-phosphate synthase), wherein, upon its deletion, *S. davaonensis* was unable to synthesize roseoflavin (Schwarz *et al*., 2016). It was also observed that the last enzyme, RosA was unable to accept AFP as a substrate and could only utilize AF. Therefore, a pathway-specific phosphatase was implicated in the conversion of AFP to AF which would further be used by RosA as a substrate. The phosphatase gene was

identified as RosC with the help of deletion and complementation experiments (Schneider *et al*., 2020). With the discovery and characterization of the main roseoflavin biosynthetic enzymes, the roseoflavin biosynthetic can be broadly accounted for.

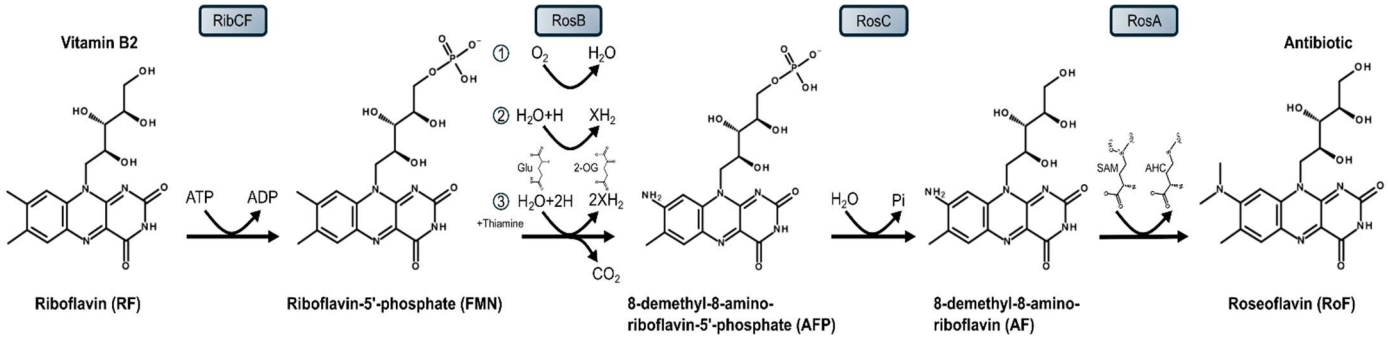


Figure 5:Condensed representation of roseoflavin biosynthetic pathway of *Streptomyces davaonensis*. The roseoflavin biosynthetic pathway leads to the formation of roseoflavin (RoF) from riboflavin (RF). The first step of the pathway is the conversion of riboflavin (vitamin B2) to riboflavin-5'-phosphate (FMN) by the C-terminal flavokinase part of RibCF. RosB (8-demethyl-8-amino-riboflavin-5'-phosphate synthase) catalyzes the first dedicated step of roseoflavin production by *S. davaonensis*. RosB catalyzes the conversion of FMN to AFP in three sub-steps. The first reaction converting FMN to 8-demethyl-8-formyl FMN (FMN-CHO) occurs in the presence of oxygen only. The second step is the conversion of 8-demethyl-8-formyl FMN (FMN-CHO) to 8-demethyl-8-carboxyl FMN (FMN-COOH), the hydrogen acceptor is not known (X). 8-demethyl-8-carboxyl FMN is then decarboxylated and transaminated to 8-demethyl-8-amino-riboflavin-5'-phosphate (AFP) in the presence of thiamine and glutamate with the production of 2-oxoglutarate. The second step of the pathway is the dephosphorylation of AFP to AF, which is carried out by the pathway specific phosphatase, RosC. The S-adenosyl methionine (SAM)-dependent dimethyltransferase, RosA, is responsible for the conversion of 8-demethyl-8-amino-riboflavin to roseoflavin (via 8-demethyl-8-methylamino-riboflavin) (Jankowitsch *et al*., 2011). As a side-product of RosA reaction, S-adenosylhomocysteine (AHC) is formed.

1.5 RosC and the histidine phosphatase superfamily

From the InterPro online tool for the classification of proteins (InterPro - ebi.ac.uk), RosC was seen to belong to the histidine phosphatase superfamily (Carmen Schneider, PhD thesis). The histidine phosphatase superfamily has two branches that share limited sequence similarity. The first branch contains well-studied enzymes like dPGM, PhoE, SixA, TIGAR, Sts-1, and fructose-2,6-bisphosphatase. The second branch comprises acid phosphatases and phytases. Proteins of the two branches share a low level of sequence identity and therefore, structure-function relationships cannot be established in the case of such enzymes. The members of this family catalyse a wide variety of functions e.g. In glycolysis and gluconeogesis (Bond *et al*., 2001, Rigden *et al*., 2003), vitamin synthesis pathways (O'Toole *et al*., 1994), nitrogen uptake systems (Schulte *et al*., 2021), cell signalling and regulatory functions- F26BPase (NIELSEN *et al*., 2004) and TIGAR (Tang *et al*., 2021). They have been implicated in antibiotic production like Neo13, characterized from *Streptomyces fradiae* (Meng *et al*., 2017) which contributes to synthesis of the antibiotic neomycin,

as is the case for RosC. Although the superfamily primarily consists of phosphatases, the earliest known and well-studied members are dPGM (cofactor-dependent phosphoglycerate mutases) which are discussed in Section 1.7.

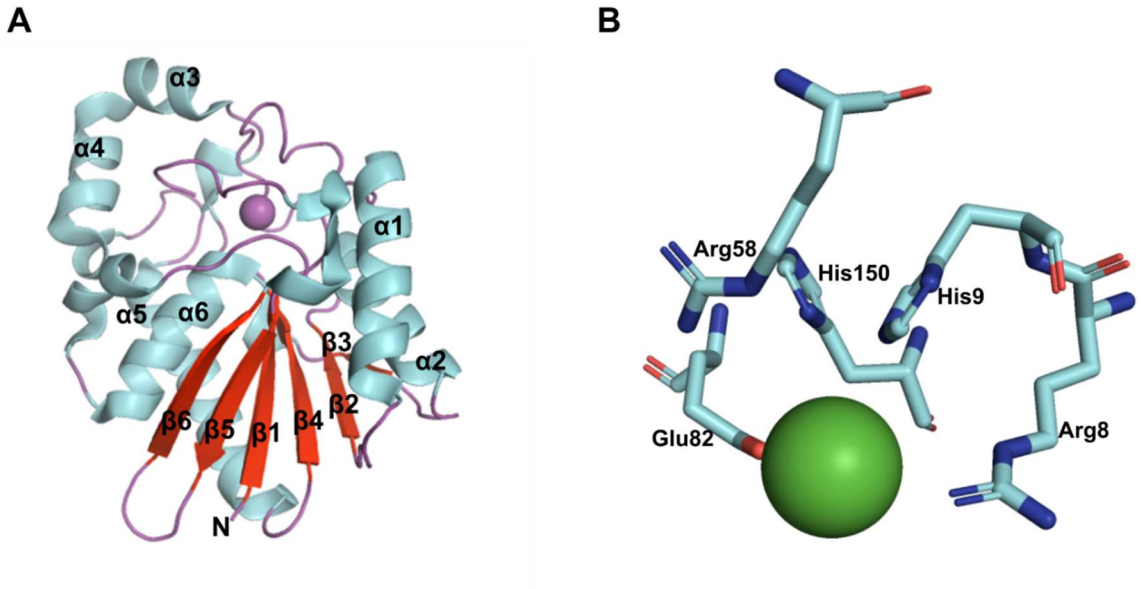


Figure 6: The crystal structure of phosphoserine phosphatase (PDB ID-4IJ5).
A: Ribbon diagram of phosphoserine phosphatase indicating the Rossmannoid fold (α helices – blue, β sheets – red, loop structures- pink). The mixed β sheet is flanked on both sides by α helices. **B:** Catalytic site of phosphoserine phosphatase with bound chloride ion (green). The main catalytic residues, R8, H9, H150, R58, E82 (blue) as seen in the other members of the histidine phosphatase superfamily can be seen surrounding the chloride ion.

1.6 Organization of the catalytic core of histidine phosphatases

A histidine residue is at the catalytic centre of the members of the histidine phosphatase superfamily. This histidine plays a crucial role in the catalytic mechanism, becoming phosphorylated and dephosphorylated during the enzymatic reaction. Two arginines (R33 and R83) and two histidines (H34 and H165) form the major players in the phosphate pocket along with a glutamate or aspartate residue (such as Asp 18 of SixA) that acts as a proton donor for the phosphatase reaction. The RHG motif is conserved throughout the family, where the glycine, in rare cases, is replaced by an alanine residue and even an asparagine. The glycine here forms an important hydrogen bond with the phospho-histidine residue and orients it optimally for catalysis. Another conserved L[S/T] XXG motif is also found where the conserved serine/threonine forms a hydrogen bond with the conserved glycine in the RHG motif. The general catalytic scheme can be summarized in the following steps.The main catalytic histidine residue performs a nucleophilic attack on the phosphorylated substrate, leading to the formation of a phosphohistidine intermediate. This is facilitated by a proton

donor (Asp or Glu) that donates a proton to the leaving group thus allowing the cleavage of the phosphoester bond. The surrounding arginines and histidines stabilize the phosphohistidine intermediate. In the last step, a water molecule, activated by the proton donor, performs a nucleophilic attack on the phosphohistidine intermediate releasing the phosphate group and regenerating the enzyme. The core fold constitutes an α/β/α sandwich motif (Rossmannoid domain fold), comprising of mixed β-sheets flanked by α-helices on both sides (Figure 6A). Different insertions in the core fold may help in carrying out different functions; extensive N and C terminal extensions also assist in these functions. The C-terminal regions of histidine phosphatase superfamily proteins have been of special focus as several proteins show an extended C-terminal domain which extends of the active site. The domain has been shown to form a lid-like structure that covers the active site of the enzyme and is known to contribute to substrate selectivity. These structures often remain disordered until substrate binding occurs, indicated by high B-factor. Salient features of well-characterized histidine phosphatases, their active site residues and unique structural features that are employed in substrate binding have been summarized in Table 1.

Table 1:PDB entries of previously characterized histidine phosphatases

PDB ID	Protein	Organism	Active site residues and key features
1H2F	Phosphoglycerate mutase	*Bacillus subtilis*	R9, R59, N16, H151, H10, E83 The C-terminal tail is partially disordered and interacts with R9; participates in substrate binding
1UJC	SixA	*Escherichia coli*	Asp18, Arg21, Ser51, Arg55, H8, R7 and His108 All linker regions are short.
4IJ5	Phosphoserine Phosphatase	*Hydrogenobacter thermophilus*	R8, H9, R58, E82, H150 Dimer like RosC: C-terminal residues form a lid over the active site cleft. No N-terminal extension.
1FBT	Fructose-2,6-bisphosphate	*Rattus norvegicus*	His7, His141, and Glu76, Arg6, His7, Arg56, Two loops interact with each other to bind substrate.
1E58	Cofactor-dependent phosphoglycerate mutase	*Escherichia coli*	His183, Glu88. Arg61, Arg9, Gly184, H10 and N16 C-terminal tail which acquires β-hairpin structure and acts like a lid
2H4X	Bisphosphoglycerate mutase	*Homo sapiens*	Arg-10, Arg-62, His-188, Gly-189, Arg-100, Arg-116, Arg-117, Tyr-92, Asn-190. Residues: 99-122 and C-terminal residues: 236-250 shift towards core domain upon substrate binding
2A6P	Phosphatase	*Mycobacterium tuberculosis*	Arg12, His13, Thr16, Ser19, His24, Thr25, Arg64, Glu85, Tyr88, His147, His149, and Thr173. Loop from one monomer forms a partial lid over the active site of the other monomer; does not affect catalytic activity but affects substrate binding

1.7 The cofactor-dependent phosphoglycerate mutases (dPGMs)

Phosphoglycerate mutases catalyze the reactions in the glycolytic and gluconeogenesis pathways, carrying out interconversion of 2-phosphoglycerate and 3-phosphoglycerate. Different phosphoglycerate mutases exist and catalyze analogous reactions. These enzymes differ in protein size, primary sequence, and structure. The cofactor-dependent phosphoglycerate mutases (dPGMs) catalyze the mutase reactions via a phosphohistidine intermediate and belong to the histidine phosphatase superfamily. The cofactor-independent phosphoglycerate mutases (iPGM) and catalyze the intermolecular transfer of a phosphoryl group between the monophosphoglycerates and the cofactor via a phosphoserine intermediate (Foster *et al*., 2010). While most of the members of the histidine phosphatase superfamily consist of phosphatases, some members are phosphomutases; dPGMs and BPGMs. The mutases simply transfer a phosphate group to a different carbon atom in the compound. The difference between mutases in comparison to the phosphatases is that the mutases can keep the intermediates bound as they reorient within the active site. (Rigden, 2008) . While the core catalytic residues i.e the two histidines, arginines and the proton donor (Asp/Glu) remain conserved in the dPGMs. Certain proteins with homology to dPGMs, but without mutase activity, have been found in a broad range of organisms and have been predicted to act as phosphatases. Some examples are PhoE from *B. Stearothermophilus* and Phosphoserine phosphatase (PspA) homologs from *Thermus thermophilus* (Chiba *et al*., 2013). These dPGM-like phosphatases have been reported to have diverse substrate preferences, as is the case with PspA (Chiba *et al*., 2013).

1.8 The haloacid dehalogenase (HAD) superfamily

Flavin homeostasis is maintained by concerted action of flavokinases and phosphatases. The enzymes that carry out the conversion of riboflavin to FMN (riboflavin kinases) and further to FAD (FMN adenyltransferases) have been extensively studied. While the enzymes that hydrolyse FMN and FAD have been studied in various organisms, their exact roles are unclear (Rawat *et al*., 2011).Two FMN hydrolysing enzymes, YigB and YbjI have been identified in *E. coli.* These enzymes primarily dephosphorylate ARPP, which is an intermediate in the riboflavin biosynthesis pathway. Other similar enzymes have also been identified in *Bacillus* and *Arabidopsis* that catalyse the dephosphorylation of ARPP and FMN. In *Bacillus subtilis*, YcsE, YitU and YwtE have been shown to dephosphorylate FMN (Sarge *et al*., 2015). ARPP and FMN share a high degree of structural similarity which allows these compounds to serve as a substrate for enzymes that do not

have a strict substrate preference. These enzymes belong to the haloacid dehalogenase superfamily. The enzymes also dephosphorylate a variety of other compounds. These enzymes also carry out the dephosphorylation of other compounds such as GMP, ribose-5-phosphate, phosphoenolpyruvate, glucose-6-phosphate, etc (Sarge *et al*., 2015). While most of these FMN hydrolases have broad substrate specificity, a protein from *Arabidopsis thaliana*, localized in the plastids, acts specifically as an FMN hydrolase (Rawat *et al*., 2011).In this plant, a riboflavin kinase homolog is fused to an FMN hydrolase. The N-terminal of the purified recombinant *Arabidopsis* enzyme (AtFMN/FHy) functions as an FMN hydrolase, while the C-terminal domain acts as a riboflavin kinase. (Sandoval & Roje, 2005). Their metabolic regulation and their exact physiological roles are poorly understood. The haloacid dehalogenase (HAD) superfamily represents a large group of enzymes containing phosphoesterases, ATPases, phosphonatases, dehalogenases, and sugar phosphomutases. (Seifried *et al*., 2013a). The enzymes carry out catalysis via a phospho-aspartyl intermediate. The catalytic site has two aspartate residues situated two residues apart. One aspartate residue serves as a general acid or base residue that protonates the leaving sugar group. Another aspartate residue then can deprotonate a water molecule which then performs a nucleophilic attack on the phosphoaspartyl intermediate, leading to the release of a phosphate group. The cofactor used in these reactions is Mg^{2+} and these enzymes are usually highly active at the physiological pH range, i.e 6.0-8-0, indicating general housekeeping roles of the enzymes. Some acid phosphatases are functional are low pH such as 5.0. The HAD superfamily of proteins has distinctive structural features such as a Rossmann-like (Rossmannoid) fold and a cap domain. The fold is different from the Rossmann fold by presence of a squiggle and flap domain and these structures provide the flexibility to the enzyme to attain open and closed conformations.(Burroughs *et al*., 2006) . Even though there is an absence of significant sequence similarity among the members of the HAD superfamily, they have four short conserved catalytic motifs, Motif I contains two conserved aspartate residues in the active that site has a signature motif of DxDx(V/T) (L/V). (Seifried *et al*., 2013b). Motif II contains a highly conserved Thr or Ser residue while Motif III contains are conserved Lys residue. Motif IV consists of conserved $D(X)_x$ D residues.

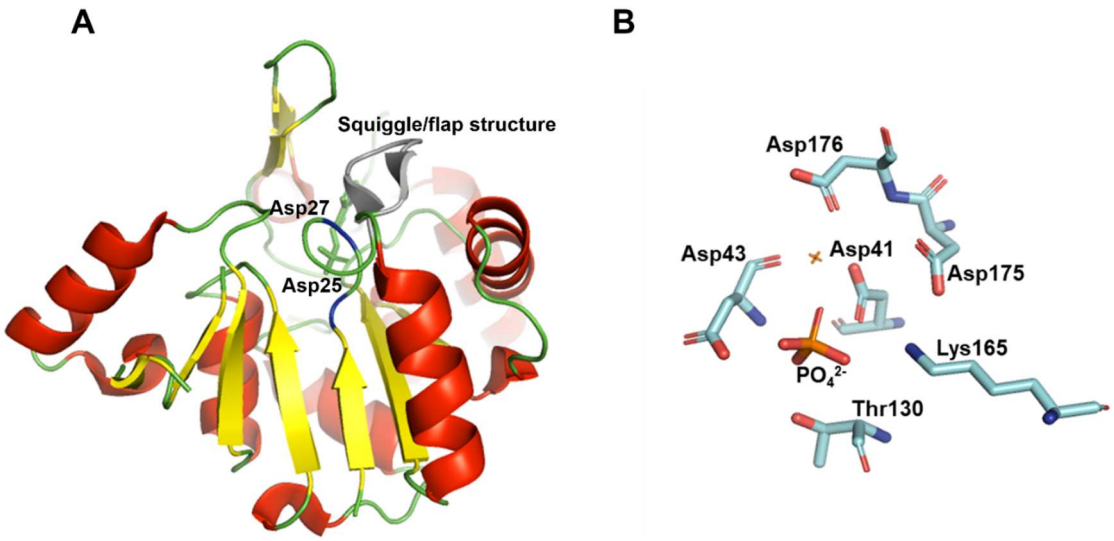


Figure 7: Conserved motifs of haloacid acid dehalogenases (HAD).
(A) The Rossmann-like fold of Chronophin (PDB: 2P69, modified) formed by repeating β-α units. The central sheet consists of five parallel β-strands sequentially arranged (yellow). The key aspartate residues of HAD motif I are depicted in blue, followed by the squiggle and flap domains in the first loop (grey). **(B)** Active site showing the core catalytic residues in the four active site loops of the human mitochondrial deoxyribonucleotidase (PDB: 1MH9, modified). The conserved aspartate (D41 and D43 in Motif I; D175 and D176 in Motif IV), lysine (K165 in Motif III) and threonine (T130 in Motif II) residues are depicted in the active site, surrounded by the phosphate group and Mg^{2+} shown as an orange cross.

Most HAD-like hydrolases contain a highly conserved α/β core domain with a mobile cap domain (Kuznetsova *et al*., 2015). It has been reported, in yeast HADs, that some elements are structurally conserved, such as the active site loops, central β-sheet, and the helical turn, referred to as the squiggle (Kuznetsova *et al*., 2015). Different types of cap domains are present, providing new surfaces for the substrate to interact, thus allowing the diversification of the enzyme function. Based on the topology and locations, cap domains are classified into 3 categories- C0, C1 or C2. C0 caps do not fold into their structural unit while C1 (inserted between motifs I and II) and C2 (inserted between motifs II and III) form independently folded structures that are separated from the catalytic core. The presence of a flexible substrate recognition element in the cap domain which becomes ordered upon substrate binding in the 'closed' cap conformation has been reported. A high B-factor in the unbound state indicates states of high disorder and flexibility which transitions into low B-factor upon the binding of the substrate. This conformation promotes catalysis by favouring interactions between substrate and the substrate binding sites in the cap domain, restricting the influx of solvent and thereby favouring the nucleophilic attack by the aspartate residue. The product is released upon opening of the cap domain, and enzymes are restored to their native state (Park *et al*., 2015).

A common element of loop within the superfamily is the conserved Gly residue located at the N-terminal end of the hairpin turn. Gly residue is essential for conferring flexibility at the hairpin turn position. The Gly residues determine the loop flexibility and conformation (Lahiri *et al*., 2004). Mutation of glycine to proline can affect the ability of the enzyme to attain open and closed conformations. While some members of the HAD superfamily have been implicated for their roles in FMN hydrolysis, they are also known to dephosphorylate other substrates. YbjI and YigB are examples of proteins that prefer compounds in riboflavin biosynthetic pathway (ARPP and FMN) as their substrates. HADs comprise of both high-specificity proteins and proteins with high degree of substrate promiscuity. They are wildly distributed in all life forms and catalyse many biochemical reactions and play important roles in housekeeping roles in a cell.

1.9 Previous studies on RosC

RosC was discovered as the pathway-specific phosphatase involved in the biosynthesis of roseoflavin in 2020. The gene *rosC* was found to be located on the 106 kbp fragment of the *S. davanonesis* chromosome. Gene deletion of a putative AFP phosphatase gene in this fragment, BN159_8033, from *Streptomyces davaonensis* resulted in 70% less RoF production by the resulting strain. When the deletion strain was complemented with an integrative plasmid expressing the BN159_8033 gene under a constitutive promoter, the RoF levels could be restored to the same levels as the wildtype strain. These experiments indicated that the product of the BN159_8033 gene was responsible for the dephosphorylation of AFP to form AF (Schneider *et al*., 2020). Bioinformatic studies indicated that the protein RosC belonged to the histidine phosphatase superfamily. R33, H34, R83, H165 and D166 were identified as key residues (PhD thesis, Carmen Schneider) RosC is capable of AFP, FMN, and RoFMN but is highly specific for AFP. It has a K_M = 34.5 ± 5.6 µM for AFP and K_M = 309 ± 61 µM for FMN (Schneider *et al*., 2020).

1.10 Aim of this study

Histidine phosphatases have been well studied, and many of their members have been structurally characterized. Although the catalytic mechanism for the superfamily is well understood, RosC plays a unique role in dephosphorylating the flavin intermediate, AFP, in the roseoflavin biosynthesis pathway. AFP is structurally similar to the important cellular cofactor, FMN, which is the precursor of roseoflavin. The enzyme prefers AFP as a substrate over FMN, thereby preventing the dephosphorylation of this crucial cofactor. The structural features that allow the accommodation of bulky isoalloxazine ring and the mechanism by which the enzyme distinguishes between AFP and FMN need to be understood. In earlier studies, D166 was annotated as the proton donor of the phosphatase reaction, but its positioning indicates that it is an important substrate-binding residue. In this study, the proton donor has been correctly annotated.

Most of the known FMN hydrolases belong to the haloacid dehalogenase superfamily (HAD), the members of which are known to employ different catalytic mechanisms than the histidine phosphatase superfamily. RosC and many members of the HAD superfamily carry out unspecific dephosphorylation of the cellular cofactor- FMN. Through these studies, the structural and sequence similarities between the RosC and FMN hydrolases from the HAD superfamily have been explored to draw a comparison between the two. The broad aims of these studies are:

- Crystallization of RosC or variants in the presence of substrate, AFP.
- Mutational and kinetic studies on the RosC to understand the roles of residues in the catalytic pocket.
- Understanding the mechanism of discrimination between FMN and AFP by RosC.
- Comparison of RosC with the ARPP/FMN hydrolases from the HAD superfamily.

2. MATERIALS

2.1 Laboratory Equipment

Table 2: List of the equipment used in this study.

Instrument	Specification	Supplier
Autoclave	VarioklavR Dampfsterilisator 175	H+P Labortechnik, Habermoos, Germany
Centrifuge	Multifuge X1R	Thermo Fisher Scientific, Waltham, MA, USA
Electroporator	Gene Pulser II	Bio-Rad Laboratories GmbH, Feldkirchen
FPLC-System	Akta Purifier P-900 with fraction collector	GE Healthcare GmbH, Solingen
Gel-electrophoresis chamber	Mini-Sub Cell GT / Wide Mini-Sub Cell GT	Bio-Rad Laboratories GmbH, Feldkirchen
Gel imaging sytem	Molecular ImagerR GelDoc™ XR, Quantity One 1D-Analysis software	Bio-Rad Laboratories GmbH, Feldkirchen
Heatblock / thermomixer	Thermomixer comfort	Eppendorf, Hamburg
High-speed benchtop homogenizer	MP Biomedicals™ FastPrep™-24	MP Biomedicals Germany GmbH, Eschwege
HPLC-System	1260 Infinity System with 1260 Infinity Diode Array Detector (DAD) und 1260	Agilent Technologies, Waldbronn
Laminar air flow workbench	Variolab Mobilien W 90	Waldner Laboreinrichtungen GmbH & Co KG, Wangen, Germany
Plate-reader	Spark®	Tecan Group, Mannedorf, Switzerland
Powerpacks for gel electrophoresis	Power Pack 200, Power Pack 300	Bio-Rad Laboratories GmbH, Feldkirchen
Sample loop for FPLC	50 ml Superloop	GE Healthcare GmbH, Solingen

SDS running chamber	XCell SureLock Mini-Cell Electrophoresis System	Invitrogen,
Semi-dry blotting system	Perfect Blue Semi Dry Electroblotter	Peqlab Biotechnologie , Clemens
Shaking incubator	Certomat IS	Sartorius Lab Instruments GmbH & Co. KG, Göttingen
Spectrophotometer	Ultrospec 3100 pro	Amersham Biosciences, Freiburg
Table-top centrifuge	Centrifuge 5415 R	Eppendorf, Hamburg
Thermocycler 1	C1000 Touch™ Thermal Cycler	Bio-Rad Laboratories GmbH, Feldkirchen
Thermocycler 2	Vapo.protect Mastercycler® Pro	Eppendorf, Hamburg
Ultracentrifuge	Avanti J-30 I	Beckmann Coulter, Krefeld

2.2 Enzymes

All enzymes were purchased from Thermo Fisher Scientific, Dreieich (Germany). Compatible buffers and additives were provided with each enzyme.

Table 3: List of enzymes used in the study.

Enzyme	Concentration	Incubation
Alkaline phosphatase FastAP	1µl / 20µl reaction volume	37 º C for 10 mins
DreamTaq DNA-Polymerase	0.25 U/µl	See cycler settings
FastDigest *Bam*HI	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 5 mins
Fastdigest *Dpn*I	1µl / 20µl reaction volume	37 º C for 5 mins
FastDigest *Eco*RI	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 20 mins
FastDigest *Hind*III	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 20 mins
FastDigest *Nco*I	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 10 mins

FastDigest *Nde*I	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 60 mins
FastDigest *Sma*I	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 5 mins
FastDigest *Xba*I	1µl / 20µl reaction volume	37 º C, Plasmid DNA- 5 mins, PCR product- 5 mins
Phusion Hot Start DNA-Polymerase	0.02 U/µl	See cycler settings
T4 Ligase	1µl / 20µl reaction volume	22 º C, 60 mins

2.3 Prepacked Columns for HPLC and FPLC

The columns used in this study with FPLC and HPLC systems are listed in the Table 4.

Table 4: List of columns used in this study.

Column	Separation	Manufacturer	Product Specification
Superdex® 200 Increase 10/300 GL	Size-exclusion chromatography	GE Healthcare GmbH, Solingen	28990944
Strep-Tactin® Superflow® high capacity	Strep-tag affinity chromatography	IBA Lifesciences, Göttingen, Germany	2-1240-001
Precolumn: SecurityGuard UHPLC Biphenyl	Column protection and longevity	Phenomenex Ltd. Germany, Aschaffenburg	AJ0-9209
Column: Kinetex 2.6 µm Biphenyl 100 Å LC Column 150 x 2,1 mm	Reverse-phase chromatography	Phenomenex Ltd. Germany, Aschaffenburg	00F-4622-AN
ReproSil-Pur C18, AQ	Reverse-phase chromatography	Dr. Maisch GmbH, Ammerbich-Entringen, Germany	R15.aq.s2546

2.4 General items/ kits

Table 5: List of kits and other items used in this study.

Items/Kits	Manufacturer	Product information
Colorimetric phosphatase assay kit	Merck KGaA, Darmstadt	MAK030
cOmplete EDTA-free Protease Inhibitor Cocktail	Roche Diagnostics GmbH,Mannheim,Germany	04693132001
dNTP-mix (10 mM)	Thermo Fisher Scientific, Dreieich	R0191
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific, Dreieich	SM0311
PageBlue Protein Staining Solution	Thermo Fisher Scientific, Dreieich	24620
PageRuler Plus Prestained Protein Ladder	Thermo Fisher Scientific, Dreieich	26619
Site-directed mutagenesis kits	Agilent GmBH, Darmstadt, Germany	210518
Vivaspin 6 (MWCO 5 kDa) Vivaspin 20 (MWCO 10 kDa)	Sartorius Stedim Biotech GmbH, Göttingen	VS0611, VS2001

2.5 Stock Solutions

Antibiotic and IPTG stock solutions were prepared as mentioned and were sterile-filtered and stored at -20 º C.

Table 6: Stock solutions of antibiotics and IPTG.

Name	Abbreviation	Stock concentration	Selection concentration
Kanamycin	Kan	50 mg/ml in milliQ	50 µg/ml (*E. coli* and *C. glutamicum*)
Chloramphenicol	Cam	50 mg/ml in 80% ethanol	25 µg/ml (*E. coli*) 5 µg/ml (*B.subtilis*)
Ampicillin	Amp	50 mg/ml in milliQ	50 µg/ml (*E. coli*)
Carbenicillin	Carb	50 mg/ml in milliQ	50 µg/ml (*E. coli*)
Isopropyl β-d-1-thiogalactopyranoside	IPTG	1 Molar in milliQ	Various concentrations

2.6. Bacterial strains

Table 7: Bacterial strains used in this study.

Strain		Genotype	Seller
E. coli	Cloning host		
	Top10	mcrA,Δ(mrr-hsdRMSmcrBC), Phi80lacZ(del)M15, ΔlacX74, deoR,recA1, araD139, Δ(ara-leu)7697,galU,galK,rpsL(SmR), endA1, nupG	Invitrogen
	Expression hosts		
	BL21 (DE3) pLysS	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cam^R)	Novagen
	OverExpress™ C43 (DE3)	F – *ompT hsdSB* (r_B- m_B-) *gal dcm* (DE3)	Merck
	Rosetta™ (DE3)	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R)	Novagen
	NEBExpress®	fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10-Tet^S) endA1 Δ(mcrCmrr)114::IS10	New England Biolabs
B. subtilis	168 wildtype	trpC2	Bacillus Genetic Stock Center
C. glutamicum	MB001(DE3)	cg1122-P_{lacI}-*lacI* P_{lacUV5} *lacZ*α-T7 gene *1*-cg1121	(Kortmann *et al*.,2015)

2.7 Growth media

2.7.1 LB (Lysogeny Broth)

LB medium was prepared according to the specification of the premixed medium (Carl *et al*. + Co. KG, Karlsruhe, Germany)

Table 8: General composition of LB medium.

Component/	Concentration (gram/ litre)
NaCl	5
Tryptone	10
Yeast extract	10

2.7.2 BHIS (Brain heart infusion sorbitol medium)

Brain heart infusion (BHI) was available as a premix and was prepared according to the directions of the manufacturer. (Becton Dickinson GmbH, Karlsruhe, Germany).

Table 9: Composition of Brain heart infusion sorbitol (BHIS) medium.

Component / Salt	Concentration (gram/ litre)
Calf brain infusion	7.7 g/l
Beef heart infusion	9.8 g/l
Proteose Peptone	10 g/l
Dextrose	2 g/l
Sodium Chloride	5 g/l
Disodium Phosphate	2.5 g/l

BHI and sorbitol were autoclaved separately in 500 ml distilled water and mixed in a sterile bottle. The final concentration of sorbitol in the medium was 91 g/l.

2.8 Buffers and Solutions

Tables below list the compositions of all buffers and solutions used in this study.

2.8.1 Buffers for preparation of chemically competent *E. coli* cells.

Table 10 (A and B): Preparation of RF1 and RF2

10(A): RF1 (pH 5.8)

Component / Salt	Concentration (gram/ litre)
RbCl	12.08 g/l
$CaCl_2.2H_2O$	1.48 g/l
Glycerol (100 %)	150 ml/l
MnCl x $4H_2O$	9.92 g/l
CH_3COOK	2.96 g/l

10(B): RF2 (pH 6.5)

Component / Salt	Concentration (gram/ litre)
RbCl	1.2 g/l
$CaCl_2.2H_2O$	1.1 g/l
Glycerol (100 %)	150 ml/l
MOPS	2.12 g/l

2.8.2 Buffers and chemicals for preparation of chemically competent *B. subtilis* 168 cells.

Table 11 (A, B and C): Composition of T base, SpC medium and SpII medium

11A: T base

Component / Salt	Concentration (gram/ litre)
$C_6H_5Na_3O_7.2H_2O$	1
$(NH_4)_2SO_4$	2
KH_2PO_4	6
K_2HPO_4 x 3 H_2O	18.3

11B: SpC medium.

Component / Salt	Concentration (gram/ litre)
T base	20
50 % (w/v) glucose	0.2 ml
1.2 % (w/v) $MgSO_4.3H_2O$	0.3 ml
10 % (w/v) Bacto yeast extract	0.4 ml
1% (w/v) casamino acids	0.5 ml

11C: SpII medium

Component / Salt	Concentration (gram/ litre)
T base	200 ml
50 % (w/v) glucose	2 ml
1.2 % (w/v) $MgSO_4.3H_2O$	14 ml
10 % (w/v) Bacto yeast extract	2 ml
1% (w/v) casamino acids	2 ml
0.1 M $CaCl_2$	1 ml

2.8.3 Buffers for agarose gel electrophoresis and SDS PAGE.

2.8.3.1 Tris-acetate EDTA (TAE) buffer for electrophoresis

Table 12: Components of TAE buffer.

Component / Salt	Concentration (gram/ litre)
Tris-HCl	4.85 g/l
Acetic acid	1.14 ml/l
0.5 M EDTA (pH 8.3)	2 ml/l

2.8.3.2 10x Tris-glycine buffer as the SDS running buffer.

Table 13: Components of Tris-glycine buffer.

Component	Molarity
Tris HCl	250 mM
Glycine	192 mM
SDS	1% w/v

1x Tris-glycine buffer was used as the running buffer and was prepared by dissolving 10x Tris-glycine buffer in distilled water.

2.8.3.3 5x protein sample loading dye for SDS PAGE

Table 14: Composition of the 5x protein loading dye.

Component	Concentration	Molarity
1 M Tris-HCl (pH 6.8)	0.25 ml per ml	250 mM
SDS	80 mg/ml	8 %
Bromophenol blue	1 mg/ml	0.1%
Glycerol (80%, v/v)	0.5 ml per ml	10 mM
DTT	0.1 ml per ml	100 mM

2.8.4 Buffers for purification of strep-tagged proteins using Strep-Tactin® column

Table 15 (A, B and C): Buffers for purification of strep-tagged proteins using Strep-Tactin® column

2.8.4.1 Buffer W (pH 8.0) for the equilibration, washing and storage steps

15A: Composition of Buffer W

Component / Salt	Molarity
Tris HCl	100 mM
NaCl	150 mM
EDTA	1 mM

pH adjustment was done using 1M HCl.

2.8.4.2 Buffer E (pH 8.0) for the elution steps

15B: Composition of Buffer E

Component / Salt	Molarity
Tris HCl	100 mM
NaCl	150 mM
EDTA	1 mM
Desthiobiotin	2.5 mM

pH adjustment was done using 1M HCl.

2.8.4.3 Buffer R (pH 8.0) the regeneration of the Strep-Tactin® column.

15C: Composition of Buffer R

Component / Salt	Molarity
Tris HCl	100 mM
NaCl	150 mM
EDTA	1 mM
HABA (2- [4 -hydroxy-benzeneazo] benzoic acid)	1 mM

pH adjustment was done using 1-5M HCl.

2.8.4.4 Buffer W (pH 10.5) for recharging the Strep-Tactin® Column

Buffer W, with the same composition as mentioned above, with a pH of 10.5, was used to remove the bound HABA after the regeneration steps. Further washing and equilibration steps were done using Buffer W.

2.8.5 Buffers for purification of MBP-tagged proteins using amylose column

Table 16 (A and B): Buffers for purification of MBP-tagged proteins using amylose column.

2.8.5.1 Buffer A: Binding buffer (pH 7.2) for the equilibration and washing steps

16A: Components of binding buffer A

Component / Salt	Molarity
Tris HCl	20 mM
NaCl	200 mM
EDTA	1 mM

2.8.5.2 Buffer B: Elution buffer (pH 7.2) for the elution steps.

16B: Components of binding buffer B

Component / Salt	Molarity
Tris HCl	20 mM
NaCl	200 mM
EDTA	1 mM
D (+)-Maltose	10 mM

2.8.6 Buffers for size-exclusion chromatography

Table 17: Components of buffer for size exclusion chromatography (pH-7.6).

Component / Salt	Molarity
Tris HCl	10 mM
NaCl	150 mM

2.8.7 Buffers and solutions for western blot

Table 18 (A, B and C): Buffers and solutions for western blot

2.8.7.1: Towbin buffer (pH- 8.3)

18A: Components of Towbin buffer

Components	Molarity
Tris	25 mM
Glycine	192 mM
Methanol	15 % v/v

10x Tris-glycine buffer (250 mM Tris and 1.92 M) was prepared beforehand, and the 1x dilution was put together freshly with methanol before usage

2.8.7.2 10 x TBS buffer (pH 7.6)

18B: Components of TBS buffer, pH 7.6

Components	Amount	Molarity
Tris base	24 g	198.13 mM
NaCl	88 g	1.5 M

The components mentioned above were dissolved in 900 ml milliQ, and the pH was adjusted to 7.6. The final volume was made up to 1000ml**.**

2.8.7.3 Washing buffer (TBST)

Tris-buffered saline, 0.1 % Tween 20 (TBST), was used during the washing steps. 1x of TBS was prepared by dissolving 10x TBS in milliQ. TBST was prepared by 10µl of Tween 20 was added to 100 ml of 1x TBS.

2.8.7.4 Blocking buffer (5 % BSA/TBST)

1.25 g of Bovine serum albumin (BSA) was dissolved in 25 ml of TBST (Tris-buffered saline, 0.1% Tween 20) was used for blocking. Appropriate dilutions of HRP-conjugated antibody were dissolved in the TBST and incubated overnight. For the western blots in this study, we have used the dilutions 1:500, 1:800 and 1:1000.

2.8.7.5 Staining solution with 3,3'-Diaminobenzidine (DAB)

DAB staining solution was always prepared freshly right before usage and the components were added in the indicated order.

18C: Components of DAB staining solution.

Component	Volume
1 x TBS	10 ml
H_2O_2 (35%)	2 µl
Solution A: DAB in DMF	140 mM (40µl)
Solution B: $CuSO_4$ in milliQ	10 % w/v (10 µl)
Solution C: $NiSO_4$ in milliQ	5 % w/v (1 µl)

2.8.8 Buffer for enzymatic assays

2.8.8.1 Bis-tris propane (BTP) buffer (10x) pH-7.6

Table 19: The composition of the 10x BTP buffer

Component	Molarity
Bis-tris propane	1 M
$CaCl_2$	0.1 M

1x Bis-tris propane buffer was prepared from the 10x stock solution upon dissolving the 10x solution in milliQ solution.

2.8.9 Buffers and Solvents for HPLC

10x Buffer A was prepared with a composition of 400 mM ammonium formate and 400 mM formic acid, pH-3.7. pH adjustment was done using Ammonium hydroxide. 1x Buffer A was prepared from the 10x solution by dissolving it in milliQ, pH was readjusted to 3.7 after dilution. The 1x buffer was filtered using 0.2 µm filter and degassed before use.

Table 20: Buffers used in HPLC

Solution	Components
Buffer/Solution A (1x)	40 mM ammonium formate, 40 mM formic acid, pH-3.7
Solvent B	Methanol
Solvent C	Acetonitrile
Solvent D	HPLC grade milliQ

Buffer A and solvent B were used for the HPLC runs. Solvents C and D were used for column cleaning and storage.

3. METHODS

3.1 Bioinformatic tools

To search for putative FMN hydrolases in *S. davaonensis*, the Uniprot BLASTp tool (Altschul *et al*., 1990) was employed to perform sequence similarity-based searches using *E. coli* protein, YbjI as a query sequence and *S. davaonensis* as a target organism. The query returned a list of sequences, and the best three candidates according to the percentage similarity with the query sequence were checked for ARPP and FMN hydrolase activity. Protein databases like UniProt (https://www.uniprot.org/, Bateman *et al*., 2023) and PDB (https://www.rcsb.org/) were used for mining protein primary structures.

Structures of novel proteins and mutants were determined using Alphafold and Alphafold3 (Abramson *et al*., 2024). Primary structures of $RosC^{D166E}$ and $RosC^{D166L}$ were used for structure predictions.

Structural alignment tools such as I-TASSER (Yang *et al.*, 2015) and DALI (Holm *et al.*, 2010) were used for finding sequences sharing structural similarity to RosC. These sequences were further used to create a sequence alignment using CLC Main workbench (CLC Genomics Workbench 20.0, QIAGEN Aarhus). The active site residues were confirmed using structural as well as sequence alignments and further validated using crystallization and kinetic studies.

Visualization of PDB structures of proteins was done using PyMOL (The PyMOL Molecular Graphics System, Version 2.5.4, Schrödinger, LLC) and WinCOOT (Emsley & Cowtan, 2004).

Figures were drawn using Inkscape (Inkscape Project (2020)), Chimera (Pettersen *et al*., 2004) and GraphPad Prism (GraphPad Prism version 9.5.1 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.)

3.2 Bacterial strains and growth conditions

E. coli and *B. subtilis* bacterial cells were grown on solid media containing LB and agar-agar (15 g/l). The growth of *E. coli* and *B. subtilis* in liquid medium was recorded by cultivating it at 37 °C in 300 ml baffled Erlenmeyer flasks containing 100 ml LB. *C. glutamicum* was aerobically grown in brain heart infusion sorbitol (BHIS) medium. The cultures were grown in baffled Erlenmeyer flasks at 30 °C for 2-18 hours while shaking at 200 rpm.

3.3 Preparation of competent cells

3.3.1 Preparation of *E. coli* competent cells

The following protocol was used for the preparation of the *E. coli* BL21 pLySs cells, C43 DE3, Top10 and Rosetta DE3 competent cells.

E. coli cells were prepared using the rubidium chloride (RbCl) method to make them chemically competent. Cells were streaked on LB plates, and a single colony was picked up and transferred to 5 ml of LB medium, with the addition of appropriate antibiotics. The culture was incubated overnight at 37 °C and 180 rpm in a shaking incubator. Secondary cultures containing 100 ml LB medium in baffled flasks were inoculated from the primary cultures to 1% of the final volume. Appropriate antibiotics were added wherever required. After incubation at 37 °C and 180 rpm to an OD_{600} of 0.6 with continuous shaking, the cells were centrifuged at 5000 rpm for 15 min at 4 °C in 50 ml falcon tubes. The supernatant was decanted, and the cell pellet was resuspended in 16 ml RF1 solution and incubated on ice for 15 minutes. After incubation, the cells were centrifuged, and the supernatant was removed. The cell pellets were pooled and resuspended in 4 ml RF2 solution and stored on ice for 15 minutes. The cells were aliquoted into volumes of 50 µl or 100µl and transferred to pre-cooled tubes. The aliquots were used directly for transformation or stored at -80 °C. All the procedures were performed under the sterile conditions.

3.3.2 Preparation of *B.subtilis* competent cells

Prewarmed SpC medium (5 ml) was inoculated with a single *B. subtilis* colony that was picked from an agar plate. The medium was supplemented with 20 µg/ml tryptophan. The culture was inoculated in a baffled flask at 37°C with constant shaking overnight. In a baffled flask, 1% of the overnight culture was used to inoculate 200 ml of SpII medium supplemented with tryptophan. The cultures were further incubated for 1.5 hours. The cells were harvested by centrifugation at 8000 x g for 5 minutes at room temperature. The supernatant was removed and transferred to a sterile container. The cell pellets were pooled and resuspended in 2 ml of sterile glycerol and 18 ml of the supernatant from the previous step.

3.3.3 Preparation of *C. glutamicum* competent cells

Primary culture of *C. glutamicum* MB001(DE3) was grown overnight at 30°C in BHIS medium in tubes in a volume of 5ml. 100 ml BHIS medium in baffled flasks were inoculated with 1 ml of the

pre-culture. The cultures were grown with constant shaking till an OD_{600} of 2 was attained, and the cells were centrifuged for 20 minutes at 5000 rpm at 4°C. The supernatant was decanted, and the pellets were resuspended in 20 ml each of sterile 10% glycerol. The suspension was centrifuged again at 5000 rpm at 4°C for 5 minutes. The supernatant was discarded, and the pellet resuspended in 40 ml of sterile 10% glycerol as a washing step. The wash step was performed three times to remove the residual salts. After the washing step, the pellet was resuspended in 1 ml of sterile 10% glycerol solution and stored as 100 µl aliquots at -80°C.

3.4 Transformation of various bacteria

3.4.1 Transformation of *E. coli* cells

Chemically competent *E. coli* cells were transformed by adding 50-100 ng of plasmid to competent cells, thawed on ice. The cells were incubated on ice for 30 minutes after addition of plasmid. The cells were subjected to a heat shock at 42°C for 90 seconds, after which, the mixture was incubated on ice for another 5 mins. Subsequently, 900-950 µl sterile LB medium was added to the cells to make up the total volume to 1 ml and incubated 37 °C for 1 h with constant shaking. The transformation mixture was plated on LB agar supplemented with appropriate antibiotics and incubated overnight at 37 °C to select for transformants.

3.4.2 Transformation of *B. subtilis* 168 cells

The transformation of the competent cells was carried out by thawing them at 37 °C. Immediately after thawing, SpII (without $CaCl_2$) and 2% EGTA (0.1 M at pH 8) were mixed gently to the cells. Plasmid preparation (1µg) was taken, to which 500 µl of cell suspension was added. The mixture was incubated at 37 °C with shaking at 180 rpm. The cell mixture was then centrifuged at 500 rpm for 5 minutes, and the cell pellet was resuspended in 200 ml of supernatant. The mixture was then plated on LB plates supplemented with appropriate antibiotics and incubated overnight at 30 °C.

3.4.3 Transformation of *C. glutamicum* MB001

Electrocompetent *C. glutamicum* cells were prepared as described in the preceding section. An aliquot of cells was thawed on ice, and 1µl of isolated plasmid with a concentration of 100-500 ng was added to the tube of competent cells. The mixture was transferred to an electroporation cuvette (Bio-Rad 0.2 cm) previously cooled on ice. Electroporation was done in the electroporator (Bio-

Rad Genepulser TM) at a capacitance of 5 μF, a resistance of 200 Ω and 2500 V for about 4.5 milliseconds. The mixture was then transferred to 5ml of preheated BHIS medium. This was then transferred to a shaking incubator at 30°C for 2 hours with shaking for recovery. The cells were then plated on BHIS agar plates supplemented with kanamycin and kept at 30°C overnight.

3.5 Cloning and Expression of genes in different expression hosts: Overview of general cloning steps

3.5.1 Purification and gel extraction of DNA

Plasmid purification was carried out using GeneJET Plasmid Purification Kit from Thermo Fisher Scientific. Overnight bacterial cultures were used to purify the plasmid. The 1-3 ml of the overnight culture was used according to the instructions provided by the manufacturer. The plasmids were eluted using sterile milliQ in a final volume of 20 μl.

The GeneJET PCR Purification Kit from Thermo Fisher Scientific was used to purify linear DNA fragments up to 20 kb in size. After amplification, the PCR mix was mixed with a binding buffer in a 1:1 (v/v) ratio and processed according to instructions provided by the manufacturer. The elution was performed using milliQ in a 10-30 μl volume.

Plasmid and PCR products were analyzed using agarose gel electrophoresis. Gels were prepared by dissolving suitable amounts of agarose powder in 1x TAE buffer. 0.8% (w/v) agarose gels were prepared to analyze plasmid DNA, while 1% (w/v) agarose gels were used to analyze PCR products.

3.5.2 PCR Cycles

The PCR components were prepared in 0.2 µl PCR tubes and the components added are listed below.

3.5.2.1. PCR amplification of inserts using Phusion polymerase

Table 21: PCR reaction using Phusion polymerase.

21A. Components of PCR reaction for Phusion polymerase

Components	Volume (total reaction volume of 50µl)	Final concentration
Foward Primer (10 µM)	1	0.5 µM
Reverse Primer (10 µM)	1	0.5 µM
5x HF/GC-buffer	4	1x
dNTP-mix (10 mM)	0.4	0.2 mM
DMSO	0.6	3 %
Phusion Hot Start DNA-Polymerase (2 U/ µl)	0.2	0.02 U/ µl
DNA	0.5 – 1.0	50-100 ng
MilliQ	upto. 20µl	

21B. Cycler settings for PCR amplification of inserts using Phusion polymerase

Reaction step	Temperature	Time	Number of cycles
Initial denaturation	98º C	2 min	1
Denaturation	98º C	30 s	20 – 30
Annealing	50 – 60º C	15-20 s	
Elongation	72º C	30 s per kb	
Final Elongation	72º C	5 min	1

3.5.2.2 Confirmation of recombinant plasmids using colony PCR

Colony PCR was performed using primer pairs specific to the PCR reaction to check for the presence of the insert. The reaction mixture was prepared according to the instructions provided by the manufacturer. (Thermo Fisher Scientific) and components are enlisted in Table 22. The thermocycler settings are listed below in Table 22B.

Table 22: PCR reaction using DreamTaq DNA polymerase.

22A: Components of PCR reaction for DreamTaq DNA polymerase

Components	Volume (total reaction volume of 50µl)	Final concentration
Foward Primer (10 µM)	1	0.2 µM
Reverse Primer (10 µM)	1	0.2 µM
10x DreamTaq Green buffer	5	1x
dNTP-mix (10 mM)	2	0.4 mM
DreamTaq DNA polymerase	1.25	0.025 U/ µl
DNA	Template	-
MilliQ	upto. 50µl	-

22B: Cycler settings for PCR amplification of inserts using Phusion polymerase

Reaction step	Temperature	Time	Cycles
Initial denaturation	95 º C	3	1
Denaturation	95 º C	30	25-35 cycles
Annealing	Primer specific	15-20 s	
Elongation	72 º C	30-60 s / kb	
Final elongation	72 º C	5 min	1

3.5.2.3 Cloning and expression of *rosC* and variants thereof in *E. coli*

The *S. davaonensis* gene BN159_8033 (Genbank CCK32411.1) encoding for RosC was codon-adapted for expression in *E. coli* using the Java Codon Adaptation Tool (JCAT) (Grote *et al.* 2005). Codons specifying an N-terminal Strep-tag (WSHPQFEK) were attached to the 5'-end of *rosC*. (Thermo Fisher Scientific, Darmstadt, Germany) synthesised the corresponding nucleotide sequence and delivered as a GeneArt cloning vector pMA-RQ <*rosC*$_{opt}$_N>. A PCR-product was generated using ppSG-IBA <*rosC*$_{opt}$_N> as a template (Schneider *et al*., 2020) and the oligonucleotides C2 and C4 listed in Table 24. The expression vector pET24a<*rosC*> was constructed by ligating *Nde*I and *Hind*III treated pET24a (Merck KGaA, Darmstadt, Germany) to this PCR-product. For cloning of the C-terminal his-tagged RosC, primers C1 and C3 were used to amplify the PCR product using the ppSG-IBA <*rosC*$_{opt}$_N> as the template and then used to ligate into the pET24a plasmid resulting in the formation of the construct, pET24a<*rosC*_C>.

E. coli BL21 (DE3) pLySs (Merck KGaA, Darmstadt, Germany) was transformed with pET24a-*rosC* using a standard heat shock protocol for transformation of chemically competent *E. coli* cells

(Sambrook *et al.* 1989). For expression of strep tagged *rosC*, the resulting strain was grown at 37 °C in LB medium supplemented with 50µg/ml Kanamycin and 25µg/ml Chloramphenicol.

Gene expression was induced by adding 0.6 mM isopropylthiogalactopyranoside (IPTG) at OD_{600} of 0.6-0.8. After 3-4 hours of further aerobic incubation with constant shaking, cells were harvested by centrifugation (8,000 *g*, 15 min, 4 °C). The cell pellet was stored at -20 °C until cell disruption. Expression of $rosC^{D166L}$ was done using a similar expression construct (pET24a<$rosC^{D166L}$>) but using the strain *E. coli* C43 DE3 (Merck KGaA, Darmstadt, Germany). The culturing of the cells was the same as described above, but the IPTG induction was done at an OD_{600} of 0.8-1 using 0.1 mM IPTG, and cells were grown for 3 hours at 37°C or overnight at 20 °C after induction.

3.5.2.4 For cloning and expression of *rosC* and variants thereof in *B. subtilis*

For expression of *rosC* and variants thereof in *B. subtilis* the plasmid pHT254 was used *rosC* from *S. davaonensis* was codon adapted using GenSmart™ Codon Optimization tool (available online at https://www.genscript.com/tools/gensmart-codon-optimization) for its expression in the *Bacillus* host cells. The sequence (S11) was obtained as a string sequence from Invitrogen (Thermo Fisher Scientific, Darmstadt) and was used as a template for PCR amplification of the gene using primers C6 and C7 (Table 24). The *E. coli*, *B.subtilis* shuttle vector pHT254 (MoBiTec, Göttingen)was used for cloning and expression. The expression vector pHT254<*rosC*> was constructed by ligating *rosC* with plasmid backbone after digesting them with *Bam*HI and *Sma*I. The cloning was performed in *E. coli* Top10, and afterwards, the plasmid was transferred to *B. subtilis* 168 for expression of *rosC.* 50 µg/ml Carbenicillin and 5 µg/ml Chloramphenicol were used for the plasmid maintenance in *E. coli* and *B. subtilis*, respectively. *B. subtilis* competent cell preparation and transformation of the host cells was carried out as described in Sections 3.3.2 and 3.4.2. For overproduction of RosC, the resulting strain was grown at 37 °C in LB medium. Expression was induced by adding 0.1-1 mM isopropylthiogalactopyranoside (IPTG) at an optical density (OD_{600}) of 0.6-0.8. After 3-4 hours of further aerobic incubation with constant shaking, cells were harvested by centrifugation (8,000 g, 15 min, 4 °C). The cell pellet was stored at -20 °C until cell disruption.

3.5.2.5 Cloning and expression of *rosC* and variants thereof in *C. glutamicum*

For expression of *rosC* and variants in *C. glutamicum*, the plasmid pMKEx2 was used. *rosC* from *S. davaonensis* was codon adapted using GenSmart™ Codon Optimization tool (available online at https://www.genscript.com/tools/gensmart-codon-optimization) for its expression in the *Corynebacterium glutamicum* MB001 (DE3) host cells. The sequence (S12) was obtained as a string

sequence from Invitrogen and was used a template for PCR amplification of the gene using primers C8 and C9. The expression vector pMKEx2 <*rosC*> was constructed by ligating *rosC* and plasmid pMKEx2 after digestion with restriction enzymes, *Nco*I and *Kpn*I. The procedure for *C. glutamicum* competent cell preparation and electro-transformation was done as described in Sections 3.3.3 and 3.4.3. The medium was supplemented with 25µg/ml Kanamycin. Expression was induced by adding 2.5 mM isopropylthiogalactopyranoside (IPTG) at an optical density (OD_{600}) of 0.8-1. The culture was further incubated till an OD of 10 was reached and cells were harvested by centrifugation (8,000 g, 15 min, 4 °C). The cell pellet was stored at -20 °C until cell disruption.

3.5.2.6 Cloning and gene expression of *S. davaonensis* FMN hydrolases

Expression of genes, BN159_ 4587 (GenBank CCK28966.1) and BN159_0855 (GenBank CCK25234.1) was done by amplifying the genes by PCR using *S. davaonensis* genome as a template. Primer pairs C10, C11 and C12, C13 were used for this purpose. The amplified PCR products were each ligated to the pET24a empty plasmid backbone after digestion with *Nde*I and *Hind*III. The *E. coli* expression strain, Rosetta (DE3) was transformed with recombinant plasmids, forming pET24a<BN159_4587> and pET24a< BN159_0855> using the heat shock method. The cells were grown at 37º C in LB, supplemented with 50 µg/ml kanamycin and 25 µg/ml chloramphenicol. The expression of genes was carried by IPTG induction of the cultures using 1mM IPTG after an OD_{600} was 0.6-0.8 was attained by the cells.

3.5.2.7 Cloning and gene expression of Δ22*rosC*_C, Δ22 *rosC* _MBP and Δ22*rosC*_MBP

rosC$_{opt}$, without the first 22 amino acids (Δ22*rosC*$_{opt}$) was amplified using PCR with primers C3 and C5 using pET24a< *rosC*$_{opt}$_N > as a template. The amplified product was digested with *Nde*I and *Hind*III and ligated to pET24a empty plasmid backbone. The STOP codon was removed from the gene to allow in-frame translation of a C-terminal hexahistidine tag. This plasmid is pET24a < Δ22*rosC*_C>.

Δ22*rosC* gene was amplified using primers C2 and C5 and the PCR product was ligated to empty pMAL-c5X plasmid after digestion with *Nde*I and *Hind*III to create pMAL < Δ22*rosC*_MBP >.

Primers C1 and C2 were used to amplify *rosC*$_{opt}$ without the strep-tag, using pET24a< *rosC*$_{opt}$_N > as a template and ligated to pMAL-c5X empty plasmid after digestion with *Nde*I and *Hind*III. This plasmid is pMAL < *rosC*_MBP >.

For expression of pMAL < Δ22*rosC*_MBP > and pMAL < *rosC*_MBP >, the cells were grown in baffled flasks at 37 º C in LB supplemented with 50 µg/ml Carbenicillin. 0.4 mM IPTG was added to induce the expression of genes when the OD_{600} reached 0.6-0.8 and the cells were incubated further at 37 º C at constant shaking for 3-4 hours.

3.6 Plasmids and Constructs

Table 23: List of all the plasmid and constructs used in this study.

Plasmid	Insert/Protein	Description
pET24a (+)	-	Empty plasmid Resistance: Kan
pHT254	-	Empty plasmid Resistance: Carb in *E. coli* and Cam in *B. subtilis*
pMKEx2	-	Empty plasmid Resistance: Kan in *E. coli* and *C. glutamicum*
pET24a<*rosC*_N>/ pET24a< $rosC_{opt}$_N >	Insert: *rosC*_N Translated protein: $RosC^{StrepN}$	*rosC* wt codon adapted for expression in *E. coli.* Sequence S10 was PCR amplified using primer pairs C4 and C2 and cloned in the restriction sites *Nde*I and *Hind*III. Translated protein has an in-frame strep-tag on the N-terminal end. Resistance: Kan (50µg/ml)
pET24a<*rosC*_C>	Insert: *rosC*_C Translated protein: $RosC^{HisC}$	*rosC* wt codon adapted for expression in *E. coli*. Sequence S10 was PCR amplified using primer pairs C1 and C3 and cloned in the restriction sites *Nde*I and *Hind*III. Stop codon was removed from the gene to allow translation of an in-frame his-tag on the C-terminal end of the protein.
pET24a<Δ22*rosC*_C>	Δ22*rosC*_C Translated protein: $\Delta22RosC^{HisC}$	*rosC* wt codon adapted for expression in *E. coli.* Sequence S10 was PCR amplified using primer pairs C5 and C3 with the codons for the first 22 amino acids, resulting in the sequence (Δ22*rosC*) and cloned in the restriction sites NdeI and Hind*III.* The stop codon was removed from the gene to allow translation of an in-frame his-tag on the C-terminal end of the protein. Resistance**:** Kan (50µg/ ml)
pHT254<*rosC*>	Insert: *rosC* wt Translated protein: **RosC** (no tag)	*rosC* wt codon adapted for expression in *B. subtilis.* Sequence S11 was PCR amplified using primer pairs C6 and C7 and cloned in the restriction sites *Bam*HI and *Sma*I Resistance: Carb (50 µg/ml) in *E. coli* and Cam (5µg/ ml) in *B. subtilis*
pMKEx2<*rosC*>	Insert: *rosC* wt Translated protein: **RosC** (no tag)	*rosC* wt codon adapted for expression in *C.glutamicum.* Sequence S12 was PCR amplified using primer pairs C8 and C9 and cloned in *Nco*I and *Kpn*I restriction sites. The second amino acid was changed from G to S using site-directed mutagenesis. Resistance: Kan (50 µg/ml) in *E. coli* and *C. glutamicum*
pMAL<*rosC*_MBP>	Insert: *rosC*_MBP Translated protein**:** $RosC^{MBP}$	*rosC* wt codon adapted for expression in *E. coli.* Sequence S10 was PCR amplified using primer pairs C1 and C2 and cloned in the restriction sites *Nde*I and

		*Hind*III. The translated protein has an in-frame N-terminal MBP tag. Resistance: Carb (50µg/ ml)
pMAL<Δ22*rosC*_MBP >	Insert: Δ22*rosC*_MBP Translated protein: $\Delta22RosC^{MBP}$	*rosC* wt codon adapted for expression in *E. coli.* Sequence S10 was PCR amplified using primer pairs C5 and C2 with the codons for the first 22 amino acids, resulting in the sequence (Δ22*rosC*) and cloned in the restriction sites *Nde*I and *HindI*II. The translated protein has an in-frame N-terminal MBP tag. Resistance: Carb (50µg/ ml)
pET24a<BN159_0855>	BN159_0855	BN159_0855 from *S. davaonensis* amplified by using primer pairs C12 and C13 and cloned into the pET24a plasmid between the restriction sites *Nde*I and *Hind*III Resistance: Kan (50µg/ ml)
pET24a<BN159_4587>	BN159_4587	BN159_4587 from *S. davaonensis* amplified by using primer pairs C10 and C11 and cloned into the pET24a plasmid between the restriction sites *Nde*I and *Hind*III Resistance: Kan (50µg/ ml)
pET24a<*ybjI*>	Insert: *ybjI* Translated protein: YbjI	YbjI from *E. coli* amplified by using primer pairs C14 and C15 and cloned into the pET24a plasmid between the restriction sites *Nde*I and *Hind*III Resistance: Kan (50µg/ ml)

Table 24: List of all the primers used in this study

No.	Description/Mutation	Primer name	Sequence
C1	Amplification of *rosC* wt gene with restriction site *Nde*I at the 5'- end	RosC_fw_NdeI	5'-AAACATATGTCTGACGGTCGTGAATCTTTC-3'
C2	Amplification of *rosC* wt gene with STOP codon, restriction site *Hind*III is at the 3'- end.	RosC_rev_HindIII_STOP	5'-TTTAAGCTTAATTAGATAACGTCAGACG-3'
C3	Amplification of *rosC* wt gene without STOP codon, restriction site *Hind*III is at the 3'- end; in-frame translation of 6x His-tag.	RosC_rev_HindIII	5'-AAAAAGCTTGATAACGTCAGACGGAGCAGC-3'
C4	Amplification of *rosC* wt gene with restriction site *Nde*I at the 5'- end, strep tag is present at N-terminal end of the RosC protein	RosC_fw_NdeI_n-strep	5'-AAACATATGTGGAGCCACCCGCAGTTCG-3'
C5	Amplification of *rosC* gene with 22 amino acids deleted from the N-terminal end of the protein, restriction site *Nde*I is present at the 5'- end. MBP tag is present at the N-terminal end of translated RosC protein.	MBP_Δ22RosC_fw_NdeI	5'-AAACATATGGTTCCGGGTGTTTCTGAAG-3'
C6	Forward primer for amplification of *rosC* wt with 5'-*Bam*HI restriction site.	RosC_BamHI_fw_B.sub	5'-AAAGGATCCATGAGCGACGGCAGAGAATC-3'
C7	Reverse primer for amplification of *rosC* wt with 3'- *Sma*I restriction site.	RosC_SmaI_rev_B.sub	5'-TTTCCCGGGTTAGATCACGTCACTTGGGG-3'
C8	Forward primer for amplification of *rosC* wt with 5'- *Nco*I restriction site.	RosC_NcoI_fw_C.glu	5'-AAACCATGGGTGACGGTCGAGAATCC-3'
C9	Reverse primer for amplification of *rosC* wt with 3'- *Kpn*I restriction site.	RosC_KpnI_rev_C.glu	5'-AAAAGGTACCTTAAATGACATCGCTCGGG-3'
C10	Forward primer for amplification of *BN* 159_4587 gene with 5'-*Nde*I restriction site.	S.dav_Seq1_fw_NdeI	5'-AAACATATGCGGGACAATGGCGCGGTGAC-3'
C11	Reverse primer for amplification of *BN*159_4587 with 3'-*Hind*III restriction site.	S.dav_Seq1_rev_HindIII	5'-AAAAAGCTTTTAACGCTCCGCCAGCAACTGCT CGATCAC-3'
C12	Forward primer for amplification of *BN*159_0855 with 5'-*Nde*I restriction site.	S.dav_Seq2_fw_NdeI	5'-AAACATATGGCCACGACGTCGGCTCTCCCGC-3'
C13	Reverse primer for amplification of *BN*159_0855 with 3'-*Hind*III restriction site	S.dav_Seq2_rev_HindIIII	5'-AAAAAGCTTCTACGGGACGAACGCGTCGATGA ACC-3'
C14	Forward primer for amplification of *ybji* with 5'-*Nde*I restriction site.	Ybji_fw_NdeI	5'-AAACATATGAGCATTAAATTAATTGCG-3'

C15	Reverse primer for amplification of *ybji* with 3'-*Hind*III restriction site.	Ybji_rev_HindIII	5'-AAAAAGCTTTTATTGGTCAAATGGCGCT-3'
S1	Forward primer for sequencing of the multiple cloning sites (MCS) of the pET24a plasmid	pET_T7_fw	5'-TAATACGACTCACTATAGGG-3'
S2	Reverse primer for sequencing of the multiple cloning sites (MCS) of the pET24a plasmid	pET_T7_rev	5'-CTAGTTATTGCTCAGCGGT-3'
S3	Forward primer for sequencing of the multiple cloning sites (MCS) of the pHT254 plasmid	pHT254_fw	5'-TATTATAAGAATTGCGGAATTGTGAG-3'
S4	Reverse primer for sequencing of the multiple cloning sites (MCS) of the pHT254 plasmid	pHT254_rev	5'-CAGTTGCAGACAAAGATCTCC-3'
S5	Forward primer for sequencing of the multiple cloning sites (MCS) of the pMKEx2 plasmid	pMKEx2_fw	5'-ACTCCTGCATTAGGAAGCAGC-3'
S6	Reverse primer for sequencing of the multiple cloning sites (MCS) of the pMKEx2 plasmid	pMKEx2_rev	5'-TTTTAGCTATCTGTCGCAGCG-3'
M1	g496a_a497c	D166T_fw	5'-CAGGATAACCAGAGCAGAGGTGTGACCAACA ACAGCAACA-3'
M2		D166T_rev	5'-TGTTGCTGTTGTTGGTCACACCTCTGCTCTGGT ATCCTG-3'
M3	a497t	D166V_fw	5'-GATAACCAGAGCAGAGACGTGACCAACAACAG C-3'
M4		D166V_rev	5'-GCTGTTGTTGGTCACGTCTCTGCTCTGGTTATC-3'
M5	g496a_a497t	D166I_fw	5'-CAGGATAACCAGAGCAGAGATGTGACCAACAA CAGCAACA-3'
M6		D166I_rev	5'-TGTTGCTGTTGTTGGTCACATCTCTGCTCTGGTT ATCCTG-3'
M7	g508c_t510g	V170L_fw	5'-AGAGAACCCATCAGGATCAGCAGAGCAGAGTC TGAC-3'
M8		V170L_rev	5'-GTCACGACTCTGCTCTGCTGATCCTGATGGGTT CTT-3'
M9	g508a	V170I_fw	5'-AGAACCCATCAGGATAATCAGAGCAGAGTCGT GAC-3'
M10		V170I_rev	5'-GTCACGACTCTGCTCTGATTATCCTGATGGGTT CT-3'
M11	c493g_a494c	H165A_fw	5'-AACCAGAGCAGAGTCGGCACCAACAACAGC AACACG-3'
M12		H165A_rev	5'-CGTGTTGCTGTTGTTGGTGCCGACTCTGCTCTG GTT-3'
M13	t382g_t383c	F128A_fw	5'-TTTTTCCGGTTCCGGGGCCGGGTAAACACCGT GC-3'
M14		F128A_rev	5'-GCACGGTGTTTACCCGGCCCCGGAACCGGAAA AA-3'
M15	t541g_g542c	W181A_fw	5'-CCAGCTGACCCGCACCCAGGTTCATCAGAGAA CC-3'

M16		W181A_rev	5’-GGTTCTCTGATGAACCTGGGTGCGGGTCAGCT GG-3’
M17	t112g_a113c	Y38A_fw	5’-GTCTTCCAGACCGGTGGCAGAGTCAGCGTGAC GG-3’
M18		Y38A_rev	5’-CCGTCACGCTGACTCTGCCACCGGTCTGGAAG AC-3’
M19	a361g_a362c	K121A_fw	5’-TAAACACCGTGCGGTGCCAGTTCAGACGGAC GA-3’
M20		K121A_rev	5’- TCGTCCGTCTGAACTGGCACCGCACGGTGTT TA-3’
M21	t22g_t23c_c24c	F8A_fw	5’-CAGAACGCATAACTTCCAGGGCAGATTCACGA CCGTCAGAC-3’
M22		F8A_rev	5’-GTCTGACGGTCGTGAATCTGCCCTGGAAGTTAT GCGTTCTG-3’
M23	t499g_c500c_t501t	S167A_fw	5’-CAGGATAACCAGAGCAGCGTCGTGACCAACA AC-3’
M24		S167A_rev	5’-GTTGTTGGTCACGACGCTGCTCTGGTTATCCTG-3’
M25	t32c	V11A_fw	5’-TTCGTAAACAGAACGCATAGCTTCCAGGAAAG ATTCACG-3’
M26		V11A_rev	5’-CGTGAATCTTTCCTGGAAGCTATGCGTTCTGTT TACGAA-3’
M27	t46g_a47c	Y16A_fw	5’-CCAACCAGGTAACGTTCGGCAACAGAACGCAT AACTTCCAG-3’
M28		Y16A_rev	5’-CTGGAAGTTATGCGTTCTGTTGCCGAACGTTAC CTGGTTGG-3’
M29	t55g_a56c	Y19A_fw	5’-GAACACCAACCAGGGCACGTTCGTAAACAGA ACGCATAACTT-3’
M30		Y19A_rev	5’-AAGTTATGCGTTCTGTTTACGAACGTGCCCTGG TTGGTGTTC-3’
M31	c379g	P127A_fw	5’-TTCCGGGAACGCGTAAACACCGTGCGG-3’
M32		P127A_rev	5’-CCGCACGGTGTTTACGCGTTCCCGGAA-3’
M33	t44c	V15A_fw	5’-CCAGGTAACGTTCGTAAGCAGAACGCATAACT TCC-3’
M34		V15A_rev	5’-GGAAGTTATGCGTTCTGCTTACGAACGTTACCT GG-3’
M35	c498g	D166E_fw	5’-GCTGTTGTTGGTCACGAGTCTGCTCTGGTTATC-3’
M36		D166E_rev	5’-GATAACCAGAGCAGACTCGTGACCAACAACA GC-3’
M37	a497g	D166G_fw	5’-GCTGTTGTTGGTCACGGCTCTGCTCTGGTTATC-3’
M38		D166G_rev	5’-GATAACCAGAGCAGAGCCGTGACCAACAACAG C-3’
M39	g319c	E107Q_fw	5’-AGTTGGTACGAACTTGACGCAGACGAGCGTC-3’
M40		E107Q_rev	5’-GACGCTCGTCTGCGTCAAGTTCGTACCAACT-3’
M41	a320c_a321c	E107A_fw	5’-CCCAGTTGGTACGAACGGCACGCAGACGAGCG TC-3’
M42		E107A_rev	5’-GACGCTCGTCTGCGTGCCGTTCGTACCAACTG GG-3’
M43	g496a	D166N_fw	5’-TTGCTGTTGTTGGTCACAACTCTGCTCTGGTTA TC-3’

M44		D166N_rev	5’-GATAACCAGAGCAGAGTTGTGACCAACAACA GCAA-3’
M45	g496c_a497t_c498a	D166L_fw	5’-CGTGTTGCTGTTGTTGGTCACCTATCTGCTCTG GTTATCCTGATG-3’
M46		D166L_rev	5’-CATCAGGATAACCAGAGCAGATAGGTGACCAA CAACAGCAACACG-3’
M47	a556g_t557c_g558a	M186A_fw	5’-TCAGCGGCAGGATCGCGTCCAGCTGACCCC-3’
M48		M186A_rev	5’-GGGGTCAGCTGGACGCATCCTGCCGCTGA-3’
M49	a356c	E119A_fw	5’-CGTGCGGTTTCAGTGCAGACGGACGACCG-3’
M50		E119A_rev	5’-CGGTCGTCCGTCTGCACTGAAACCGCACG-3’
M51	c97g_g98c	R33A_fw	5’-TCTGAAGTTTGGCTGATCGCTCACGCTGACTCT TACAC-3’
M52		R33A_rev	5’-GTGTAAGAGTCAGCGTGAGCGATCAGCCAAAC TTCAGA-3’
M53	c247g_g248c	R83A_fw	5’-CGGTCTGCTGAGCAGCGTGAGCACCAGAAG-3’
M54		R83A_rev	5’-CTTCTGGTGCTCACGCTGCTCAGCAGACCG-3’
M55	a497c	D166A_fw	5’-GCTGTTGTTGGTCACGCCTCTGCTCTGGTTATC-3’
M56		D166A_rev	5’-GATAACCAGAGCAGAGGCGTGACCAACAACA GC-3’
M57	c100g_a101c	H34A_fw	5’-CTGAAGTTTGGCTGATCCGTGCCGCTGACTCTT ACAC-3’
M58		H34A_rev	5’-GTGTAAGAGTCAGCGGCACGGATCAGCCAAAC TTCAG-3’
M59	c493g_a494c	H165A_fw	5’-CGTGTTGCTGTTGTTGGTGCCGACTCTGCTCTG TT-3’
M60		H165A_rev	5’-AACCAGAGCAGAGTCGGCACCAACAACAGCA AACG-3’

3.7 Growth curves of *E. coli*, *B. subtilis* and *C. glutamicum*

Growth curves of *E. coli, B. subtilis* and *C. glutamicum* were performed in baffled flasks with constant agitation. Overnight cultures were grown in tubes and were used to inoculate secondary cultures in baffled flasks. The secondary cultures were inoculated from the primary cultures at an OD_{600} of 0.03-0.05.

Cultures of *E. coli* and *B. subtilis* were grown at 37° C, the protein production induction was done at an OD_{600} of 0.5, and the cell growth was measured at regular intervals until the stationary phase was reached.

C. glutamicum cultures were grown at 30° C and the induction of the protein production was done at an OD_{600} of 0.8 using 2.5 mM IPTG, after which, the OD_{600} was measured at regular intervals for ~10 hours., till the cultures attained stationary phase.

3.8 Site-directed mutagenesis of *rosC* from *S. davaonensis*

Site directed mutagenesis of the plasmid pET24a<$rosC_{opt}$_N> was carried out using a QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Waldbronn, Germany). The mutagenic oligonucleotides were designed individually for every mutation using the web-based QuikChange Primer Design Program (www.agilent.com/genomics/qcpd). The PCR cycles for site directed mutagenesis has been described in Table 25. Both mutagenic oligonucleotides contain the desired mutation in the middle of the primer flanked by 10-15 bases of correct sequence. Following reaction mixtures were prepared in a total volume of 50µl.

Table 25: Site-directed mutagenesis using Quikchange lightening kit.

25A: Components of Quikchange reaction

Components	Volume (µl)
Plasmid DNA (100 ng/µl)	0.5
10x Quikchange reaction buffer	5
Forward Primer (10 µM)	0.8
Reverse Primer (10 µM)	0.8
dNTP-Mix (10 mM)	1
QuikSolution	1.5
MilliQ water	Up to 50µl

Afterwards, the total reaction mixture was divided into two separate PCR tubes, containing 25µl reaction mixture each. To one of the tubes, 0.5µl of Quikchange enzyme was added, and the other was treated as a negative control without any addition of the enzyme. The tubes were incubated in the thermocycler with the following settings

25B: Thermocycler settings for site-directed mutagenesis using Quikchange lightening kit

Reaction step	Temperature	Time	Number of cycles
Initial denaturation	95°C	2 mins	1
Denaturation	95°C	20 s	18
Primer-annealing	60°C	10 s	
Elongation	68°C	30s per kb	
Final elongation	68°C	5 mins	1

After the completion of the PCR cycle, 0.5µl of *Dpn*I enzyme (provided with the kit) was added to each tube. The tubes were further incubated at 37° C for 5 minutes, after which the reaction mixtures were directly used to transform competent *E. coli* Top10 cells.

3.9 Molecular and Biochemical techniques

3.9.1 Preparation and clarification of *E. coli* cell-free extracts

Frozen cell pellets of *E. coli* BL21 (DE3) overproducing the N-terminally strep tagged RosC WT were resuspended in 30 ml washing buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA; pH 8.0), containing cOmplete™, EDTA-free, Protease Inhibitor Cocktail tablet (Merck KGaA, Darmstadt, Germany) according to specifications. The pellets of NEB express strains overproducing the N-terminally MBP tagged proteins were resuspended in column buffer (20 mM Tris-HCl, 200 mM NaCl, pH 7.4) containing the Protease Inhibitor Cocktail tablet.

Cells were lysed using the French press with a pressure of 2,000 bar at 10 °C. Six cycles of 40 seconds each at 4.5 m/s were carried out in the FastPrep-24™ cell-homogenizer (MP Biomedicals, Eschwege, Germany)

Cell debris and unbroken cells were removed by two centrifugation steps (8,000 g, 4 °C, 10 min followed by 108000 g, 4 °C, 30 min).

3.9.2 Purification of strep-tagged RosC (RosC^{StrepN}) wild-type and variants

Chromatographic steps were performed using the ÄKTApurifier™ system (GE Healthcare). The cell lysate obtained via above mentioned steps (Section 3.9.1) from the lysis of *E. coli* BL21 pLySs expressing N-strep-tagged RosC was applied to a 5 ml Strep-Tactin® Superflow® high-capacity cartridge column (IBA Lifesciences, Göttingen) equilibrated with washing buffer at a flow rate of 0.5 ml per min. As soon as the UV signal returned to its baseline, elution of Strep-tagged protein was accomplished by applying a gradient of 3.33 % per min (100 % Buffer E in 30 min) elution buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 10 mM desthiobiotin; pH 8.0). Eluted fractions were concentrated using Vivaspin 6 centrifugal concentrators with a molecular weight cut-off of 10 kDa (Sartorius, Göttingen, Germany) at 6,000 *g* and 4 °C. Eluted fractions were concentrated as described above and loaded onto a Superdex® 200 Increase 10/300 GL (GE Healthcare) equilibrated in running buffer (10 mM Tris HCl, 150 mM NaCl; pH 7.6) at a flow rate of 0.5 ml per min. Fractions were collected, glycerol was added to a final concentration of 20 % (v/v) and the enzyme was stored at 4 °C.

3.9.3 Purification of MBP-tagged proteins from *E. coli*

This method has been used to purify wild-type (RosC^{MBP}) and MBP-tagged Δ22RosC (Δ22RosC^{MBP}) proteins. The cell lysate preparation and clarification steps were the same as mentioned in the Section 3.9.1. Gravity flow columns containing 2ml of amylose raisin (New England Biolabs) were used to bind MBP-tagged proteins. The columns were loaded with 10 ml clarified cell-free extracts of the proteins and washed with 5 column volumes of column buffer (20 mM Tris-HCl, 200 mM NaCl, pH 7.4) to wash off unspecifically bound proteins. MBP-tagged proteins were eluted by adding elution buffer (20 mM Tris-HCl, 200 mM NaCl, 10 mM Maltose, pH 7.4) in three steps and collecting the eluted protein after each step. Three fractions of protein of 1ml each were collected during elution and loaded on SDS PAGE for checking their purity.

3.9.4 Determination of native molecular masses of proteins

The native molecular mass of RosC^{StrepN} and Δ22RosC^{MBP} were determined by size exclusion chromatography using a Superdex® 200 Increase 10/300 GL (GE Healthcare). 10 mM Tris HCl, 150 mM NaCl; pH 7.6 was used as the running buffer at a 0.5 ml/min flow rate. Bovine serum albumin (66.5 kDa), ovalbumin (42.7 kDa), chymotrypsinogen (25.6 kDa) and ribonuclease A (13.7 kDa) and IgG (150 kDa) were used as standards.

3.9.5 Determination of protein concentrations

Protein concentration was estimated by the method of Bradford (Bradford, 1976). The samples were prepared by diluting them to concentrations of 1:100 and 1:200 in water. 200µl Bradford reagent (Protein assay dye reagent concentrate (5x), Bio-Rad, Düsseldorf, Germany) was added to make the total volume up to 1000 µl. The total protein concentration was detected by checking absorbance at 595nm (A_{595}). Appropriate dilutions were made according to the protein concentration.

3.9.6 Visualization of proteins using SDS PAGE

The proteins were visualized by SDS-PAGE after denaturation with loading buffer (Section 2.8.3.3) at 90º C for 10 minutes . Novex™ WedgeWell™ 4 to 20 % Tris-glycine precast gels, (Invitrogen, Thermo Fisher Scientific, Germany) were used along a marker PageRuler Plus Prestained Protein Ladder (Thermo Fisher Scientific, Darmstadt, Germany) with a molecular weight range from 10 – 250 kDa was used. The initial run voltage was kept as 90V but subsequently increased to 110V after the samples moved completely from the wells into the gel.

Gels were stained according to specifications using PageBlue™ Protein Staining Solution (Thermo Fisher Scientific, Darmstadt, Germany). The gels were later washed with water to destain them.

3.9.7 Visualization of proteins using western blot

Immunological detection of proteins was done using western blots. The proteins to be detected were C-terminally his-tagged and protein separation was first done using SDS PAGE. $RosC^{HisC}$, $RosC^{D166L\_HisC}$, and $\Delta22RosC^{HisC}$ proteins were detected using this method. After loading the proteins on SDS–PAGE, proteins were transferred to a nitrocellulose membrane (0.45 µm, PALL Corporation, Crailsheim, Germany). 1x Towbin buffer was freshly prepared from the 10x Towbin stock buffer. The membrane was sandwiched between 3 layers of blotting paper on either side, which were presoaked in cooled 1x Towbin buffer. For western blotting of the $\Delta22RosC^{His}$ and $RosC^{HisC}$ proteins, semi-dry protein transfer was done using a blotter (Perfect Blue Semi Dry Electroblotter, Peqlab Biotechnologie , Clemens) at 60 mA for 60 minutes. Membranes were blocked for 60 minutes at 4º C using a blocking solution containing 5% BSA in PBST/TBST buffer (Phosphate/Tris buffer saline with 0.1% Tween 20) and then incubated with the mouse anti-6xHis Tag (Thermo Fischer Scientific, MA1-21315-HRP) primary antibody (1:500-1:1000 times dilution) overnight at 4º C on a rocking platform. After three washes in PBST/TBST, membranes were incubated in a freshly prepared staining solutions containing DAB (3,3'-diaminobenzidine) as mentioned in the

Table 18. The blot was immediately washed with distilled water upon the appearance of protein bands and the incubation times were kept short so as prevent unspecific staining of the blot.

3.9.8 AFP and FMN phosphatase assay and HPLC analysis of flavins

AFP and FMN phosphatase activity of purified RosC and variants thereof was analyzed in 1x Bis-tris propane (BTP) buffer (100 mM BTP, 200 µM $CaCl_2$; pH 7.6) containing 5 – 400 µM AFP. Mixtures were equilibrated at 37 °C for 5 min. The phosphatase assays were started by addition of 0.1 – 1 µM purified RosC. After appropriate time intervals, aliquots were removed, treated with 5 % (w/v) TCA, and placed on ice to stop the reaction. Samples were centrifuged (8,000 *g*, 10 min, 4 °C) and filtered (0.2 µm regenerated cellulose membrane filter) before analysis by HPLC-DAD. HPLC-DAD analysis was performed at a flow rate of 0.2 ml per min at 50 °C using a Phenomenex Biphenyl HPLC column (2.6 µm particle size, 150 mm x 2.1 mm; Aschaffenburg, Germany) in combination with the Agilent 1260 Infinity system. Flavins were separated using a gradient elution programme. The approximate retention times and the HPLC solvent gradients are listed in Table 26 A and B, respectively.

RosC activity is expressed as micromoles of AF formed per minute from AFP. Assays were carried out by varying the concentration of the substrate AFP (ranging from 5 µM to 400 µM) to determine the kinetic parameters. Apparent initial rates were plotted against the substrate concentrations. Kinetic constants K_M and k_{cat}/V_{max} were determined by fitting the data to the Michaelis-Menten equation using a nonlinear regression curve plotted using GraphPad.

Table 26 : Conditions for analysis of flavins using HPLC.

26A: List of the flavins and their retention times and wavelengths.

Flavin	Retention time	Wavelength (DAD)
AF	8,3	480 nm
AFP	10,2	480 nm
FAD	12,6	445 nm
RP	15,3	445 nm
RF	21,3	445 nm
RoFMN	21,2	503 nm
RoF	23,1	503 nm

26B: Method describing steps of the HPLC gradient used to separate and detect flavins.

Gradient step	Time / min	% of Buffer A	% of Buffer B
Separation	0	85	15
	3	77	23
	3.1	73	27
	5	70	30
	6.5	68	32
	13.5	68	32
	20	5	95
Washing	20.1	0	100
	24.1	0	100
	24.5	100	0
	29.5	100	0
Re-equilibration	30	85	15

3.9.9 5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5'-phosphate (ARPP) and FMN hydrolase assays

ARPP was synthesized using RibD and RibG proteins from *E. coli* using GTP as the initial substrate. The procedure was as described by Hasse *et al.*, 2013; supporting information and Sa *et al.,* 2016. The synthesized ARPP was confirmed using MALDI-TOF.

Proteins coded by genes *rosC*, BN159_4587, BN158_0855 were tested for ARPP phosphatase and FMN hydrolase activities. The ARPP phosphatase reaction was carried out in a volume of 500µl using 100 mM Bis-tris propane (BTP) buffer, pH 7.5, 20mM DTT, 8mM $MgCl_2$ (200µm $CaCl_2$ for RosC instead) and 100µl of synthesized ARPP. Cell free extracts of *E. coli* Rosetta strain overproducing the putative FMN/ARPP hydrolases were used for the assay. The reaction mixture was incubated at 37 °C for 1 hour after addition of 0.2 mg/ml of total protein. The dephosphorylated product, ARP was detected by HPLC-FLD (excitation (λ_{ex}) = 410 nm, emission at (λ_{em})= 485 nm) after derivatization with diacetyl as described in Sa *et al.,*2016. The derivatized product was filtered using Vivaspin 6 (MWCO 3 kDa) to remove proteins in the reaction mixture and the sample was analyzed using HPLC using column- ReproSil-Pur C18, AQ (particle size- 5µm; 250 mm x 4.6 mm; Dr. Maisch GmbH, Ammerbich-Entringen, Germany).The compounds were eluted using an isocratic gradient of 5% Acetonitrile and 95% water at 25º C with a flow rate of 0.8 ml/min. Cell-free extracts of *E. coli* BL21 pLySs overproducing the protein YbjI from *E. coli* were used as a

positive control for the reaction. Reaction mixture untreated with protein and treatment with RosC were used as a negative control for ARPP phosphatase activity.

The FMN hydrolase assay was carried out in presence of 100 mM BTP supplemented with 200µM $MgCl_2$ (200µM $CaCl_2$ was used in the case of RosC) pH 7.5. Cell-free extracts of *E. coli* Rosetta overproducing BN159_4587 and BN158_0855 were used for the FMN hydrolase assay. 100µM FMN was used as substrate along with 0.2 mg/ml total protein. The samples were incubated at 37 ° C for 1 hour to observe the dephosphorylation of FMN to riboflavin. The samples were subjected to TCA treatment and analyzed using HPLC.

3.9.10 Detection of phosphatase activity using phosphate release assay

Phosphatase activity against other potential substrates was measured by using a phosphate colorimetric assay kit (MAK030, Merck KGaA, Darmstadt). The assay was performed in 69 well plates to monitor the release of free phosphate as a result of enzymatic activity. Stock solutions of substrates to be tested were made in HPLC-grade milliQ water. 100µM substrate and 1 µM purified enzyme were added to a well. The final volume was made up to 300 µl using milliQ. The plates were then incubated at 37 ° C for 30 minutes. After incubation, 60µl of the substrate was added to each well and the mixed 2-3 times using pipetting. The plates were further incubated for 45 minutes away from light at room temperature. The assay was then carried out by recording the absorbance at 650 nm (A_{650}) using Spark plate reader. (Tecan Group, Mannedorf, Switzerland)

3.9.11 Crystallization of purified RosC

For crystallization purified RosC was stored in 10 mM TRIS/HCl pH 7.5 and 150 mM NaCl at a concentration of 15 mg/ml. Sitting drop experiments were performed with the Rigaku CrystalMation robot system and commercial screens of JBS HT I, II, pentaerythritol, PACT++ (Jena Bioscience), PGA (Molecular Dimension) and JCSG Core Suite I-IV (Qiagen). Crystallization experiments were performed with the flavin-free wild-type RosC, RosC incubated with 0.5 mM AF and with 0.5 mM AF and 5 mM phosphate as well as $RosC^{R33A}$ supplemented with 0.5 - 1.0 mM riboflavin. The corresponding crystallization conditions and cryoprotectant solutions are listed in Table 29. Data of the obtained crystals were collected at the PXII beamline at the Swiss-Light-Source, Switzerland and processed with XDS and XSCALE (Kabsch, 2010) (Table 29)

. The structure of RosC was determined by the molecular replacement method using PHASER (McCoy *et al*., 2007) with a combined model of Robetta (Kim *et al*., 2004) and phosphoserine

phosphatase (4IJ5,(Chiba *et al*., 2013) as initial coordinates. The correction of the models and the incorporation of ligands and solvent were performed with COOT (Emsley *et al*., 2010). For refinement, PHENIX (Afonine *et al*., 2012) was applied. Original cif parameter files for the substrates AF and riboflavin was produced by PHENIX-elBOW and slightly modified according to the electron density. The structure was analyzed by COOT and MOLPROBITY(Davis *et al*., 2007). figures were produced with Chimera (Pettersen *et al*., 2004).

4. RESULTS

4.1 Bioinformatic studies

4.1.1 Comparison of the primary structure of RosC with 8-demethyl-8-amino-riboflavin-5'-phosphate phosphatases from other roseoflavin producers

When the sequence of RosC from *S. davaonensis* was used as a query sequence to search for similar sequences, the highest degree of similarity was found in the case of 8-demethyl-8-amino-riboflavin-5'-phosphate phosphatase (AFP phosphatases) from *S. davaonensis, S. cinnabarinus, S. berlinensis*, which have been shown to produce roseoflavin. A similar protein was also found in *Streptomyces* NBRC 14336 and *Streptomyces* HUAS 2-6 which are putative roseoflavin producers (Liunardo *et al*. 2024) . RosC from *S. davaonensis* shares 96.8% sequence identity with RosC from *S. cinnabarinus*, 93.7% with that of *S. berlinensis,* 93.7 % with *Streptomyces* sp. NBRC 14336 and 84.3% with *Streptomyces* sp. HUAS 2-6 (Figure 8). Sequences in each of the organisms are highly similar with all the residues in the catalytic pocket being conserved (Figure 8).

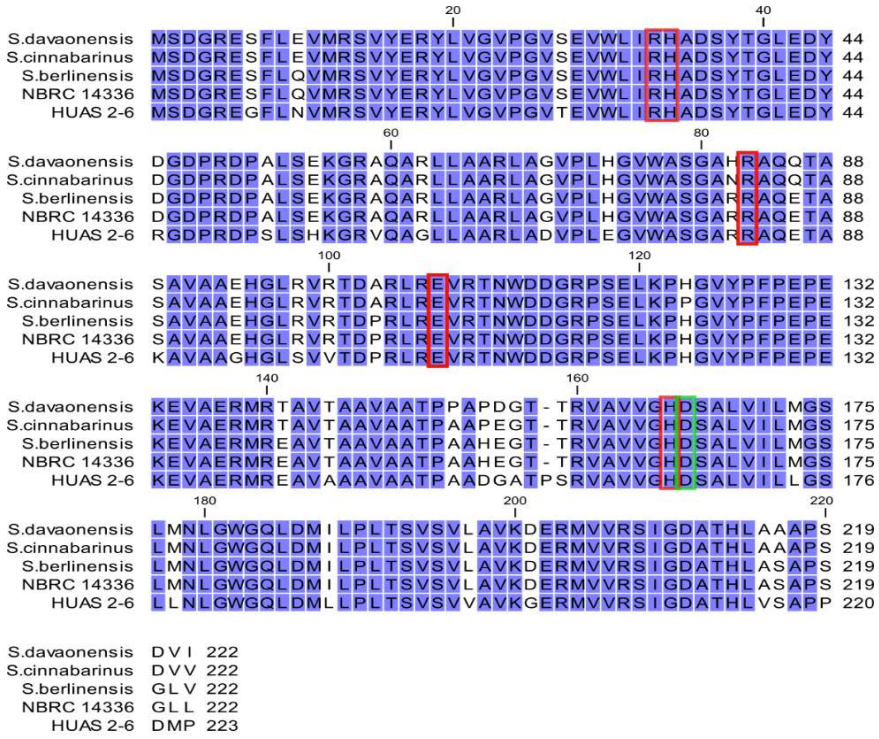


Figure 8: Sequence alignment of the protein sequence of RosC from *Streptomyces davaonensis* with 8-demethyl-8-amino-riboflavin-5'-phosphate phosphatases from other known and predicted roseoflavin producers.
The primary structure of RosC from *S. davaonensis* was used to carry out a BLASTp search highly similar proteins were found in the Roseoflavin producers *Streptomyces cinnabarinus, Streptomyces berlinensis* and the putative roseoflavin producers *Streptomyces* sp. NBRC 14336 and *Streptomyces* sp. HUAS 2-6. Amino acids that play an important role in catalysis and are conserved in all structures are marked (red box). D166 is also marked (green box) and is indicative of AFP phosphatases.

4.1.2 Sequence and structural alignment with previously characterized histidine phosphatases

Sequence alignment of RosC with well-characterized histidine phosphatases gave initial insights into the important residues in RosC activity. Most of these proteins belong to the family of enzymes that function as cofactor-dependent phosphoglycerate mutases (dPGMs). Residues R33, H34, R83, and H165 have been identified as the key catalytic residues. Sequence alignment indicates that the residue that acts as the proton donor is a glutamate residue at position 107 (E107), which is conserved in the sequences that share a high degree of similarity with RosC (Figure 10). H34 is the main catalytic residue that carries out the nucleophilic attack on the substrate leading to the formation of the phosphohistidine intermediate. The region with the first 27 residues does not align with other histidine phosphatases and seems to be unique to the AFP phosphatases. RosC, upon crystallization, was also used for structural alignment with previously characterized histidine phosphatases that share a high structural similarity with it. RosC shares the highest similarity with Phosphoserine phosphatase and cofactor-dependent phosphoglycerate mutase from *E. coli*, PDB

IDs – 4IJ5 and 1E58, respectively. The catalytic pocket of the two enzymes were superimposed on RosC to confirm all the active site residues. Upon the structural alignment of RosC with PSP (4IJ6/4IJ5), it could be observed that even though the core is relatively conserved, RosC has the presence of unique regions/insertions that do not align with the crystal structures of the two proteins. The unique regions span from residues 1-27, 33-53, 103-131, and 178-192 (Figure 9.) Residues 1-22 from the N-terminal of RosC, fold into an alpha helix which is connected to the enzymatic core by a four amino acid linker region. N-terminal helix here on has been referred to as the cap structure or the cap domain. The cap domain along with Insertion I (33-53), Insertion II (103-131), and Insertion III (178-192) form the unique segments of RosC and other 8-demethyl-8-amino-riboflavin-5'-phosphatase phosphatases.

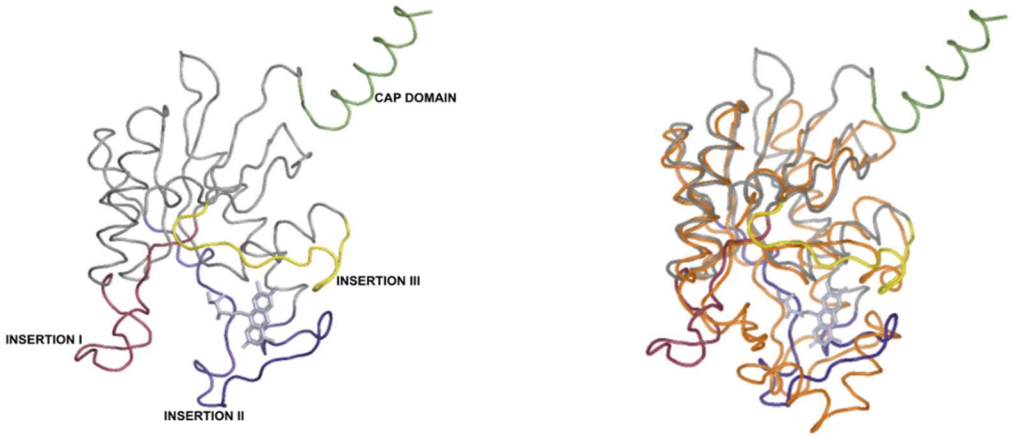


Figure 9: Ray trace diagram of RosC highlighting the unique segments and its comparison with phosphoserine phosphatase (4IJ6).
The figure on the left shows RosC (gray) highlighting the cap domain (green) and the Insertions I (red), II (purple) and III (yellow). These distinct insertions between secondary structures and fusions at the N- and C-terminus to the core fold determine the individuality among family members. Insertion II (103-131) is highly diverse with respect to length and conformation. Insertions I (33-53) and III (178-192), although less variable, substantially contribute to the size and shape of the substrate binding site. Solely, RosC is endowed with a long N-terminal extension protruding to the partner subunit, while its C-terminal arm is short in contrast to other family members. The figure on the right shows the superposition of RosC with Phosphoserine phosphatase - PDB ID- 4IJ6 (orange). The root mean square deviation is 2.8 Å (178 of 211 residues). The insertions and as well the cap domain does not align well with the phosphoserine phosphatase backbone.

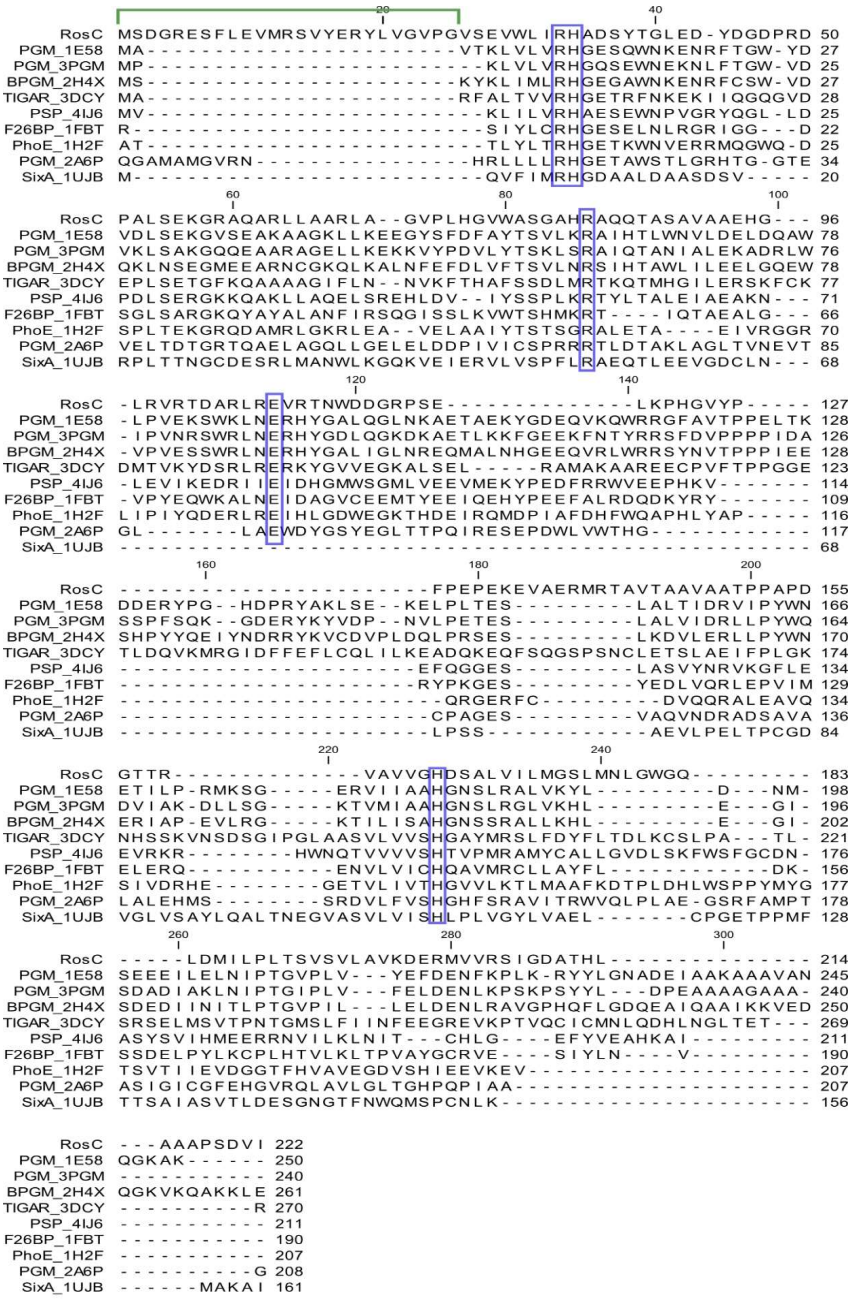


Figure 10: Sequence alignment of RosC with previously characterized histidine phosphatases
Sequence alignment was done using CLC Main Workbench and the following PDB entries were used for sequence alignment ; 1E58-cofactor-dependent phosphoglycerate mutase from *Escherichia coli*, 3PGM - Phosphoglycerate mutase from *Saccharomyces cerevisiae*, 2H4X- Bisphosphoglycerate mutase from *Homo sapiens,* 3DCY- TIGAR (apoptosis regulator) from *Homo sapiens,* 4IJ6- phosphoserine phosphatase from *Hydrogenobacter thermophilus*, 1H2F- cofactor-dependent phosphoglycerate mutase PhoE from *Bacillus stearothermophilus*, 2A6P - cofactor-dependent phosphoglycerate mutase from *Mycobacterium tuberculosis*, 1UJB - histidine phosphatase SixA from *Escherichia coli.* The conserved residues in RosC R33, H34, R83, E107, and H165 are indicated by the blue boxes. The 22 amino acids from the N-terminal cap domain are unique to RosC and do not align with any other protein from the histidine phosphatase super-family (green).

4.2 Expression of *rosC* and purification of RosC variants

4.2.1 Overproduction of RosC in *E. coli* BL21 pLySs host cells

For obtaining proteins to set up crystal structures, the expression of *rosC* was optimized under the inducible T7 promoter using *E. coli* BL21 pLySs as a host. Different IPTG concentrations were used to determine the optimal IPTG concentration for the gene expression. Protein production was determined using SDS PAGE. Protein was observed upon IPTG induction with 0.1-0.6 mM IPTG Figure 11. This expression system was used for overproducing $RosC^{StrepN}$ and mutant proteins.

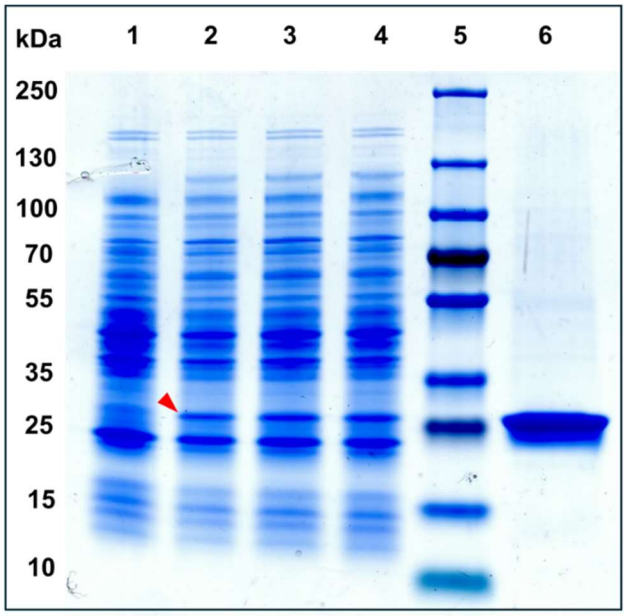


Figure 11: Overproduction of N-terminally strep-tagged RosC ($RosC^{StrepN}$) in *E. coli* BL21 pLySs. Lanes from left to right- **Lane 1:** Cell-free extract (CFE) of *E. coli* BL21 pLySs with pET24a containing *rosC* without IPTG induction. **Lane 2:** Cell-free extract (CFE) of *E. coli* BL21 pLySs with pET24a containing *rosC* with IPTG induction using 0.1 mM IPTG **Lane 3:** CFE of *E. coli* BL21 pLySs expressing *rosC* using 0.4 mM IPTG **Lane 4:** CFE of *E. coli* BL21 pLySs expressing *rosC* using 0.6 mM IPTG **Lane 5:** Protein molecular weight marker. **Lane 6:** Purified N-terminally strep tagged RosC, $RosC^{StrepN}$. The red arrow indicates the position of RosC. *rosC* was expressed in all the cases after addition of 0.1- 0.6 mM IPTG and the molecular weight of the protein on the gel is approximately 25 kDa which is very close to the calculated molecular weight of the protein i.e 25.58 kDa.

4.2.2 Purification of RosC to attain highly pure protein

The pET24a<*rosC*> expression construct was used for the over-production of strep-tagged RosC and mutant proteins in *E. coli* host cells. The cell-free extracts overproducing $RosC^{StrepN}$ were first purified using affinity chromatography (Figure 12A) after loading it on the Strep-Tactin® Superflow® prepacked column as described in Section 3.9.2.Elution of bound protein was done using a linear gradient of Buffer E containing desthiobiotin using a linear gradient of 3.33 % Buffer E per min, and a flow rate of 1ml /min (Figure 12A). RosC eluted at a concentration between 35 - 40% Buffer E corresponding to a concentration of 0.88-1 mM desthiobiotin. The protein was collected in fractions of 1 ml each. The purity of proteins was confirmed using SDS PAGE which showed relatively pure protein levels (Figure 13). This eliminated the need for an intermediate chromatographic step. For purifying proteins further for crystallization, size-exclusion chromatography was performed as the polishing step. Fractions containing purified RosC were collected and pooled. The protein was concentrated and applied to a Superdex™ 200 Increase 10/300 GL prepacked column for size exclusion chromatography (Figure 12B). The eluted protein was used for structural and biochemical analysis. Samples from each stage were analyzed on SDS PAGE for purity check. Analysis using SDS PAGE revealed the presence of highly pure protein which was concentrated and used for setting up crystal screens.

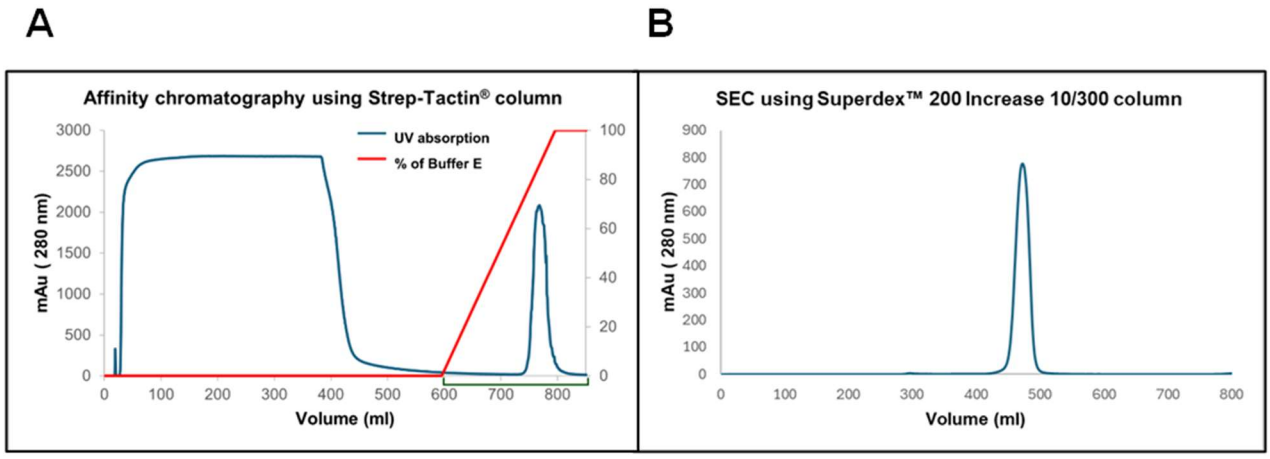


Figure 12 (A and B): Elution profile of RosC after (A) affinity chromatography and (B) size-exclusion chromatography
For setting up crystal screens, 50-100 ml of cleared cell-free extract (CFE) was applied to the 5 ml Strep-Tactin® Superflow column with Buffer E containing desthiobiotin using a gradient of 3.33 % Buffer E / min (denoted by the red line, Figure 12A). The protein was eluted in fractions of 1 ml each and pooled together. The green line marks the region of fractionation. The pooled proteins were concentrated to a volume of 500µl and then applied to the Superdex™ 200 Increase 10/300 GL at a flow rate of 0.5 ml/min (Figure 12B).

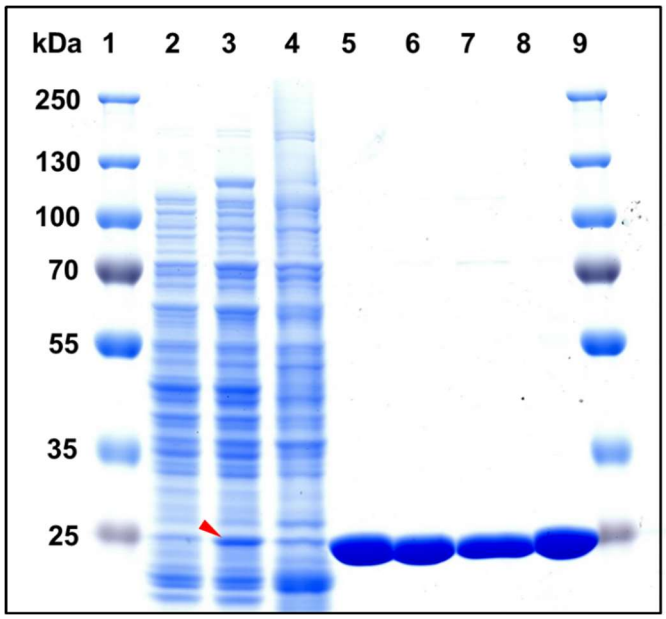


Figure 13: Purification of RosC^{StrepN} to apparent homogeneity from *E. coli* host cells
SDS PAGE showing stepwise purification of strep-tagged RosC and variants from *E. coli* BL21 pLySs host cells by affinity chromatography and size-exclusion chromatography. **Lane 1:** Protein molecular weight marker, **Lane 2:** Cell-free extract (CFE) of *E. coli* BL21 pLySs containing pET24a empty plasmid (control) **Lane 3:** CFE of BL21 pLySs containing pET24a containing *rosC* gene **Lane 4:** Flow through after passing CFE expressing *rosC* through Strep-Tactin® column. **Lanes 5-7:** Different fractions of eluted RosC protein after affinity chromatography using Strep-Tactin® column **Lane 8:** RosC protein after size exclusion chromatography using Superdex® 200 Increase 10/300 GL **Lane 9:** Protein molecular weight marker. Lanes 2-4 contain 20 µg total protein, lanes 5-8 contain 8 µg total protein.

4.2.3 Purification of RosC^{StrepN} mutants

One-step purification was used for the purification of RosC wildtype and RosC mutants for biochemical studies. They were purified using affinity chromatography with the ÄKTApurifier FPLC system using the Strep-Tactin® Superflow® prepacked columns or gravity flow columns packed with Strep-Tactin® Sepharose® resin. The purity of the proteins was confirmed using SDS PAGE and RosC could be visualized at 25 kDa. The proteins that were purified using the FPLC system had fewer impurities as compared to those purified using the gravity flow columns. RosC proteins that were purified using the gravity flow columns were seen to be co-purified with high molecular weight impurities, The most prominent impurity, has a molecular weight of ~55 kDa (Figure 14).

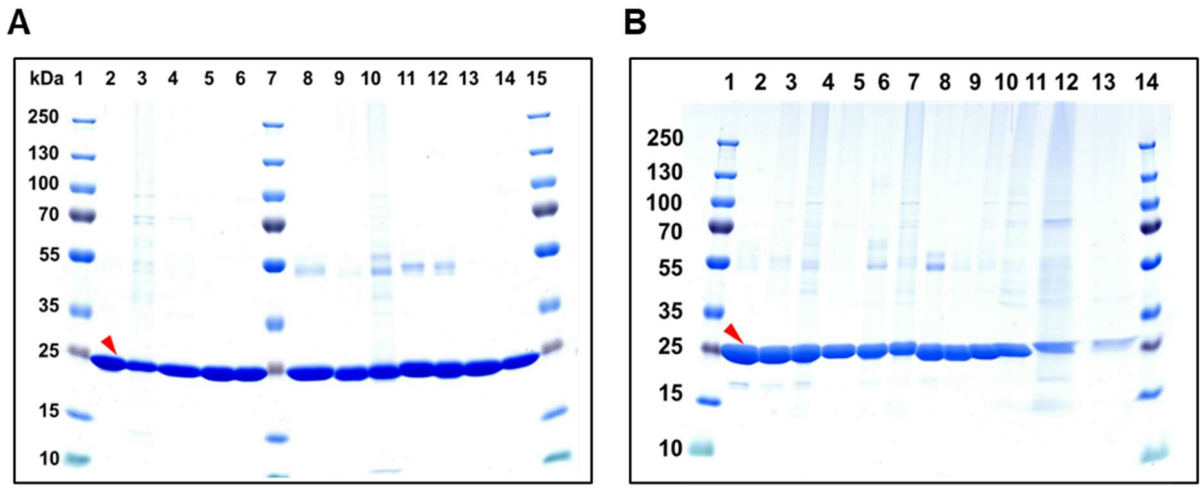


Figure 14:Purification of RosC^{StrepN} and variants from *E. coli* host cells using affinity chromatography. SDS PAGE was performed to analyze the N-terminally strep-tagged RosC variants purified for biochemical analysis using either ÄKTApurifier FPLC system or gravity flow columns packed with Strep-Tactin® Sepharose® resin. **A. Lane 1**: Protein molecular weight marker **Lane 2:** RosC^{Y19A} **Lane 3**: RosC^{Y16A} **Lane 4**: RosC^{V15A} **Lane 5**: RosC^{V11A} **Lane 6**: RosC^{F8A} **Lane 7:** Protein molecular weight marker **Lane 8:** RosC^{D166E} **Lane 9:** RosC^{D166N} **Lane 10: RosC^{D166T} Lane 11:** RosC^{D166G} **Lane 12**: RosC^{D166L} **Lane 13:** RosC^{D166A} **Lane 14:** RosC^{StrepN} **Lane 15:** Protein molecular weight marker**. B**. **Lane 1**: Protein molecular weight marker **Lane 2:** RosC^{E107Q} **Lane 3:** RosC^{E107A} **Lane 4:** RosC^{K121A} **Lane 5:** RosC^{E119A} **Lane 6:** RosC^{M186A} **Lane 7:** RosC^{Y38A} **Lane 8:** RosC^{V170L} **Lane 9:** RosC^{V170I} **Lane 10**: RosC^{D47A} **Lane 11:** RosC^{W181A} **Lane 12:** RosC^{P127A} **Lane 13:** RosC^{F128A} **Lane 14:** Protein molecular weight marker. Each well was loaded with 8µg of protein in Gel A. In each lane in Gel B, 5-8µg protein was loaded.

4.3 RosC is not a broad specificity phosphatase

Colorimetric assay was to measure free phosphate and was used to determine the phosphatase activity of RosC on various substrates. The phosphatase activity of FMN hydrolases from the HAD superfamily has been observed on many compounds. Some important metabolites that commonly dephosphorylated substrates were tested as substrates for RosC. RosC does not accept any of the listed compounds as substrates apart from AFP and FMN (Table 27).

Table 27 : Phosphatase activity of RosC on different phosphorylated compounds

Compound	RosC
AFP	+
FMN	+
ARPP	-
NADP	-
FAD	-
Acetyl pyruvate	-
PLP	-
TPP	-
dNTP mix	-
Fructose-6-phosphate	-
Fructose-1,6-phosphate	-

4.4 RosC prefers AFP as a substrate at physiological pH

pH studies were performed on RosC to determine the optimal pH for dephosphorylation of AFP and FMN. For the dephosphorylation of 200 μM AFP, it was observed that the specific activity increased with the decrease in pH, showing the highest value at a pH of 7 and decreasing sharply as the pH values decreased. The specific activity also decreased as the values increased up to a pH of about 9. For FMN, the maximum activity was observed at a pH of 5 and decreased as the pH values were increased or decreased. Maximum specific activity for AFP is at a pH-7 while a maximum activity for FMN is observed at a pH of about 5. At the physiological pH of 6-7, RosC prefers AFP as a substrate (Figure 15)

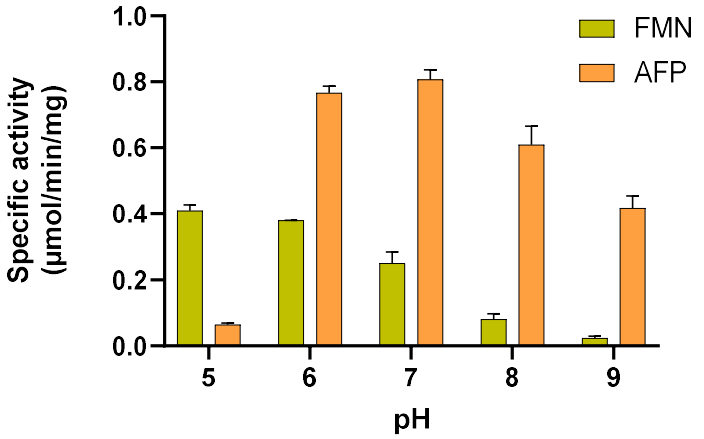


Figure 15: Determination of optimal pH for dephosphorylation of AFP and FMN by RosC
pH studies were done in 50 mM Tris HCl and 10 mM $CaCl_2$ buffer and various pH levels were tested for calculating the specific activity of $RosC^{StrepN}$ using 200µM AFP and 200µM FMN as substrates. Buffers with different pH values ranging from 5-9 were tested at 37 º C.

4.5 Several RosC mutants were purified for crystallization of protein with bound AFP

Based on kinetic data and that from structural data of RosC, several mutants were selected as potential candidates for setting up crystal screens to obtain RosC bound with the substrate, AFP. As a result, four mutants were selected for crystallization studies- $RosC^{E107A,}$ $RosC^{E107Q,}$ $RosC^{H34A+H165A}$, $RosC^{H34C+E107A}$. These mutants were selected due to their diminished phosphatase activity as they carried a mutation in the key catalytic residues. The mutants were purified as mentioned in Section 3.9.2 and crystal screens were set up with each mutant with AFP. Upon setting up the crystal screen, E107Q appeared as crystals but no AFP was bound to the crystals (Table 28). A new strategy was adopted for crystallization which included the crystallization of protein in the presence of AF (product) and inorganic phosphate to closely mimic the conditions for crystallization of the wild-type protein in presence of AFP. The crystal structure was solved in the conditions mentioned in Table 29.

Table 28: RosC mutants purified for obtaining crystals of the protein in presence of substrates.

Mutant	Total protein	Results of crystallization
$RosC^{H34A+H165A}$	Low purification yield	-
$RosC^{H34C+E107A}$	24 mg	Precipitation of sample
$RosC^{E107A}$	17.7 mg	Precipitation of sample
$RosC^{E107Q}$	8 mg	Crystals observed at pH 7.7 and 8.5

4.6 RosC crystals were obtained under four different conditions

Even though it was not possible to obtain a crystal structure of RosC in the presence of substrates AFP and FMN, 4 high-resolution crystal structures were solved in the presence of products, RF and AF. The setting of the crystal screens and determination of the electron density of the proteins was done in collaboration with the **Max Planck Institute of Biophysics, Frankfurt.** Four high-resolution crystal structures available for RosC under different conditions have been summarized in Table 29.

Table 29: Conditions for crystallization of RosC and variant RosC^{R33A} in different conditions.

Data set	RosC in the absence of flavins	RosC in the presence of 8-demethyl-8-amino-riboflavin (AF)	RosC in the presence of 8-demethyl-8-amino-riboflavin and inorganic phosphate (AF+ PPi)	RosC in the presence of riboflavin (RF)
Protein solution	10 mM Tris pH 7.5, 150 mM NaCl	10 mM Tris pH 7.5, 10% glycerol, 0.5 M 8-demethyl-8-amino-riboflavin	10 mM Tris pH 7.5, 10% glycerol 0.5 M 8-demethyl-8-amino-riboflavin	10 mM Tris pH 7.5, 10% glycerol, 0.5 M riboflavin
Precipitant solution	22% PEG 3350, 0.1 M Bis-Tris-propane pH 5.7, 0.2 M NaJ	27% PEG1500, 0.1 M sodium tartrate dibasic dihydrate, Bis-Tris, Gly-Gly (TBG), pH 5.0	29% PEG 1500, 0.1 M TBG, pH 5.0, 5 mM KH_2PO_4, pH 7.4	25% PEG 1500, 0.1 M TGB, pH 5.0
Data collection				
Wavelength [Å]	1.0	1.0	1.0	1.0
Space group	$P4_32_12$	$P2_12_12_1$	P1	P1
Resolution [Å]	50.0-1.7 (1.7-1.8)	50.0-1.25 (1.25-1.35)	50.0-1.4 (1.4-1.5)	50.0-1.5 (1.5-1.6)
Cell axes [Å]	75.9, 162.0	58.8, 81.4, 89.4	51.9, 51.9, 81.2, 73.5, 81.7, 70.5	51.9, 52.0, 81.4, 74.2, 81.8,70.8
Mol. asym. unit	αβ	αβ	2αβ	2αβ
Completeness [%]	99.9 (87.0)	99.1 (92.4)	91.4 (72.7)	87.7 (52.9)
R_{sym} [%]	6.1 (139.0)	5.2 (54.1)	5.3 (76.0)	4.7 (57.1)
I/σ_I	13.0 (1.7)	19.7 (1.1)	15.7 (4.1)	16.3 (3.3)
Redundancy	6.9 (7.1)	11.2 (4.2)	3.55 (3.4)	3.5 (3.4)
CC	99.9 (87.0)	100.0 (47.4)	99.8 (71.6)	99.8 (80.1)
Refinement				
No. of atoms polypeptide, ligand, solvent	3053, 12, 116	3464, 60, 313	6942, 238,393	3880, 227, 337
Resolution [Å]	50.0 – 1.7	1.25 - 50.0	1.4 -50.0	1.75 - 50.0
Reflections	52743	117822	138577	92346
$R_{working}$, R_{free} [%]	20.3, 22.1	19.1, 20.9	16.5, 20.5	21.4, 22.3
$B_{average}$ [A^2], polypeptide, ligand, solvent	55.8, 65.8, 45.2	29.2, 24.3, 32.9	21.4, 36.9, 27.5	31.0, 201.5, 39.5

Rmsd: **bond length [Å]** **bond angles [°]**	 0.016 1.5	 0.011 1.2	 0.006 0.91	 0.006 0.93
Clash score	6.4	2.0	4.5	2.2
Ramachandran **favored [%]** **outliers [%]**	 94.7 0.5	 97.5 0.2	 98.4 0.0	 97.7 0.0
PDB code	**8S6Q**	**8S6R**	**9EMU**	**8S7M**

4.7 Elucidation of the crystal structure of RosC

The crystal structure of RosC was determined at 1.7 Å resolution and revealed a compact homodimeric enzyme. Each monomer is built up of a mixed six-stranded β-sheet sandwiched between two helices on both sides. This core fold is expanded by specifically designed supplementary segments inserted between strands 26:32 and helix 54:69, strand 99:102 and helix 132:149 as well as helix 166:177 and strand 193-198 termed insertions I, II and III. In addition, a helical segment is fused to the N-terminus of the core fold, which is segregated and attached to the partner monomer of the dimer (Figure 25)10 residues from the N-terminal end remain unfolded. These supplementary segments are mostly flexible and partly even completely disordered as are indicated by the B-factor. The active site is in front of its four parallel strands of the central β-sheet involving the supplementary segments for substrate binding.

RosC belongs to a widespread histidine phosphatase superfamily; the structurally related representatives - phosphoserine phosphatase (4IJ6) and SixA (1UJC) have been used to draw comparisons. Distinct insertions between secondary structures and fusions at the N- and C-terminus to the core fold determine the individuality among family members. Insertion II (103-131) is highly diverse with respect to length and conformation. It varies between a short loop (SixA) via expanded segments of approximately 30-40 amino acids without significant secondary structures (as in RosC, phosphoserine phosphatase, fructose-2,6-bisphosphatase) to an entire helical domain as found in glucose-1-phosphatase and phytase. Insertions I (33-53) and III (178-192), although less variable, substantially contribute to the size and shape of the substrate binding site. Solely, RosC is endowed with the long N-terminal extension protruding to the partner monomer while its C-terminal arm is short in contrast to other family members as PhoE, for which a prolongation to the substrate binding site has been reported.

4.7.1 Binding of AF and phosphate

The crystal structure of the RosC-AF complex, determined at 1.25 Å resolution, revealed a fully occupied AF (Figure 16 A). The riboflavin derivative is arranged nearly perpendicular to the plane of the β-sheet with the tip of the ribityl (C5O) being directed towards the central β-sheet. The ribityl and isoalloxazine are surrounded and thereby fixed by insertion II (103-131) assisted by insertions I (33-53) and III (178-192) and the N-terminal helix of the partner monomer (Figure 17A). Upon AF binding, these supplementary segments become rigidified and simultaneously change their positions and conformations in the range of 2-3 Å to optimally interact with AF and with each other (Figure 17C and D). After the induced-fit process, insertion II (101-132) packs against insertions I and III via a salt bridge between Asp113-OD1 and Arg49-NH1 and a hydrophobic patch between Tyr126, Pro127, Phe128 and Trp181, respectively, as prominent interface features. In addition, insertion II (112-121) forms an extended contact area with the N-terminal partner helix (segment 7-18), which is also, linked with insertions I and III by a hydrophobic patch between Phe8', Leu9', Met12', Tyr38 and Leu41, and a hydrogen-bond between Tyr19-OH and Asp183OD1, respectively.

AF forms multiple contacts with the polypeptide (Figure 17C and D). The rather planar amino-isoalloxazine ring is sandwiched between Val11, Val15 and Tyr19 at its *re*-side and Pro127, Phe128 and Trp181 at its *si*-side (Section 4.7.3). A large number of hydrogen bonds are formed between the polypeptide and, in particular, polar atoms of the pyrimidine ring of AF frequently mediated by H_2O molecules (Figure 16 B). Of crucial relevance is the amino group of AF (replaced by a methyl group in RBF), which forms a hydrogen bond with the Met186 carbonyl oxygen and a salt bridge with Asp166 positioned at the N-terminal end of helix 166:177 (Figure 16 B). The Asp166-OD1/OD2 groups are firmly held in their positions by interactions with the Leu188-NH and ribityl hydroxyl groups, respectively.

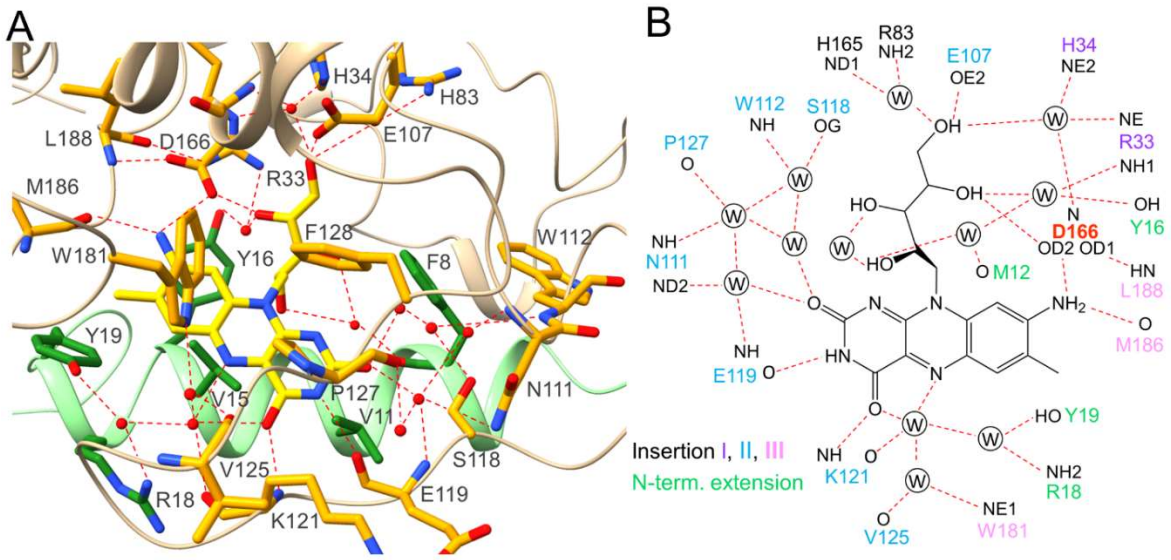


Figure 16: Interactions between RosC and the reaction product (AF).
A: Detailed RosC-AF interactions depicting the. amino acid residues of the two RosC monomers (tan and light green) contacting AF (yellow carbons) are drawn as sticks (carbons in orange or green depending on the subunit origin) and water molecules as red balls. The key amino acid D166 and its contacts to the C8 amine of AF and the hydroxyl of its ribityl are shown in the upper part. **B:** Schematic representation of the hydrogen-bond network between selected amino acids of RosC and AF. The ribityl and isoalloxazine of AF are characteristically surrounded by the four specifically designed supplementary segments (insertions I-III and an N-terminal extension) thereby constituting a novel flavin binding pocket. Amino acids of linker II (blue) directly or indirectly (*via* water molecules, "circled W") interact with heteroatoms of the polar, pyrimidine of isoalloxazine derivative.

The phosphate binding mode was explored based on the structure of the tertiary RosC-AF-phosphate complex determined at 1.55 Å resolution (Figure 18A). Phosphate directly sits in front of the four parallel strands of the central β-sheet thereby surrounded by the loops following them, helix 166:177 and AF. As strictly conserved in other histidine phosphatases the phosphate oxygens are hydrogen-bonded with two histidine residues (His34 and His165), two arginine residues (Arg33 and Arg83) and one glutamate (Glu107). Further hydrogen bonds are formed to Tyr38-OH, Asp166N and the AF-5-OH group. His34, forming the histidine-phosphate intermediate, is kept in its position by a hydrogen bond with the Ala35 carbonyl oxygen, which is in most family members a glycine (RHG motif) due to the limited space. Ala35 also forms a hydrogen bond to Ser87, a residue of a conserved region. Binding of AF and phosphate create a strained conformation of the active site, such that the ribityl and neighbouring Tyr38 and Met12 are not well defined and present in several weakly occupied conformations. Asp166-OD2 interacts with the C4 hydroxyl group in the RosC-AF structure (Figure 16 A) but predominantly with the C3 hydroxyl group in the RosC-AF-phosphate structure (Figure 18 A).

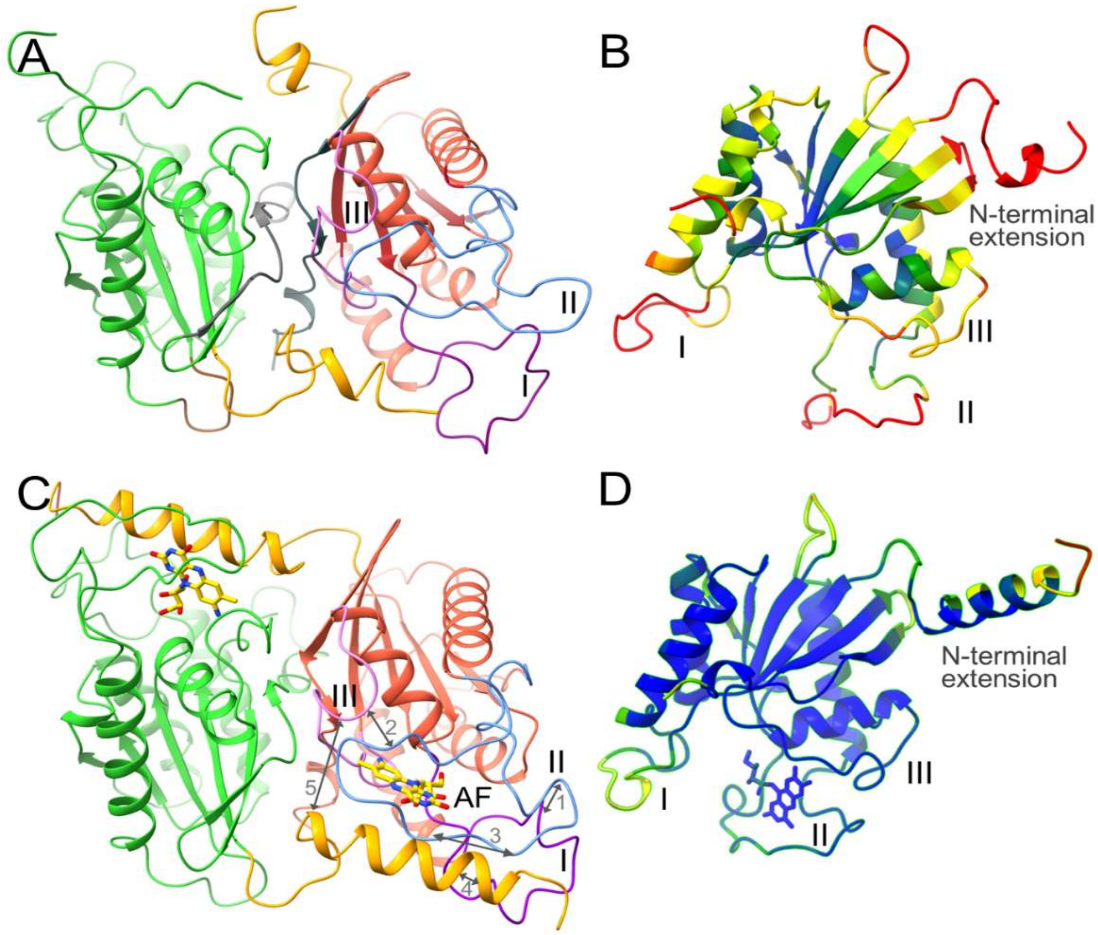


Figure 17 Crystal structures of the 8-demethyl-8-amino-riboflavin-5'-phosphate (AFP) phosphatase RosC from *S. davaonensis*.
A: The wild-type RosC dimer in the absence of flavins. Each subunit (green and red) is made up of an α/β-fold. The extended homodimer interface is mainly constructed by the N (yellow)- and C (dark gray)-terminal segments of the enzyme. The C-terminal β-strands of both subunits (dark gray) connect the two six-stranded β-sheets of the two subunits. Insertions I (32-54), II (101-132) and III (178-192) are highlighted in purple, blue and pink. Segment 153-158 (brown), only found in RosC is involved in stabilizing the linker (yellow) to the N-terminal helix (yellow) projecting into the partner subunit. Notably, in the absence of flavins the N-terminal helix is largely disordered and is better visible in C. **B:** B-factor coloured RosC monomer (B=0-25 (blue), 25-50 (green), 50-75 (yellow), >75 (red) Å^2). Insertions I-III and the N-terminal helix of the flavin-free RosC structure are highly flexible reflected in a high B-factor that describes the attenuation of X-ray scattering caused by thermal motion. **C:** The RosC-8-demethyl-8-amino-riboflavin (AF) complex. AF is flanked by the three RosC-specific insertions I (purple), II (blue) and III (pink) and the N-terminal helix (yellow) of the partner subunit (green). These segments mutually stabilize each other or themselves (marked by gray arrows 1-5). Insertion II packs against insertions I and III by a salt bridge (1) (OD1 of D113 and NH1 of R49) and a hydrophobic patch (Y126, P127, F128 and W181) (2), respectively and multiply against the N-terminal partner helix (segment 7'-18'; partner subunit residues are labelled with an apostrophe) (3). The latter is also linked with insertions I and III by a hydrophobic patch (F8', L9', M12', Y38 and L41) (4) and noticeably by a hydrogen-bond (OH of Y19' and OD1 of D185) (5), respectively. **D:** B-factor coloured monomer of the RosC-AF complex (B=0-25 (blue), 25-50 (green), 50-75 (yellow), >75 (red) Å^2). When compared with the flavin-free structure (B) the B-factors of insertions I, II and III and the N-terminal helix of the RosC-AF complex are strongly decreased indicating rigorous rigidification of the four supplementary segments.

4.7.2 The key residues of the RosC catalytic pocket

Based on the information from the structural and sequence alignments, a new catalytic scheme for RosC has been presented in this study. According to what has been studied for the histidine phosphatases, His34 is the main histidine residue that acts as a nucleophile and attacks the substrate, forming a phosphohistidine intermediate (Figure 17B). H165, R33 and R83 stabilize the negatively charged phosphate group by electrostatic interactions (Figure 17A and B). When the substrate transfers its phosphate group to the enzyme, the proton donor donates a proton to the leaving group. The proton donor also activates the water molecule to attack the phosphohistidine intermediate for the release of the phosphate group and the regeneration of the enzyme. The proton donor in RosC is E107 (Figure 17 B).

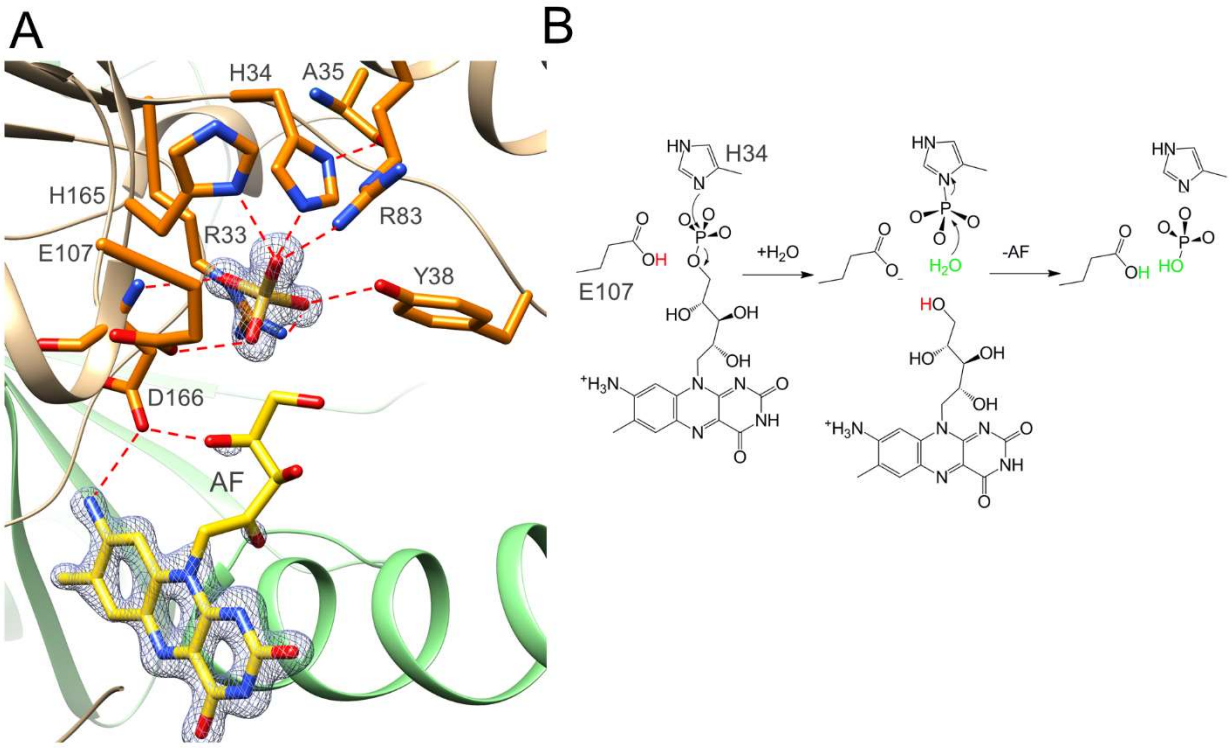


Figure 18 The phosphate-binding pocket of the active site of RosC and the catalytic mechanism.
A: Structure of the RosC-AF-phosphate structure at 1.55 Å resolution. The catalytic apparatus of RosC is primarily built of segments that follow the two β-strands 28:32 and 158:163 hosting the phosphate-contacting residues R33, H34, A35, Y38, H165, D166, and R83. **B:** Reaction mechanism of the RosC-. AFP binds to the softened state of the supplementary segments (Figure 18A and 18B) that allows the bulky substrate to reach its binding site. After a pronounced induced-fit process AFP is almost completely enveloped by the rigidified protein matrix. The imidazole of H34, ideally positioned between the three free oxygens of the AFP phosphate, attacks the phosphorous by an in-line nucleophilic reaction. R33, R83 and H165 fix the phosphate oxygens and stabilize the transition state towards the histidine phosphate intermediate. E107-OD2 with a distance of 2.6 Å to the oxygen bridging the phosphate phosphorous and the ribityl 5'-OH in the substrate is properly placed to serve as proton donor for the released product AF. The nucleophilic attack of OH^- (E107 accepts a proton from the water molecule) onto the H34-phosphate and the subsequent cleavage of the phosphate histidine bond completes the reaction.

4.7.3 Flavins are positioned within the active site by hydrophobic residues

The hydrophobic flavin sandwich, also referred to as the aromatic sandwich is an arrangement wherein the flavin moiety is sandwiched between two aromatic amino acid residues. This arrangement creates a hydrophobic environment around the flavin ring through π-stacking interactions with the aromatic side chains. In RosC, this positioning is seen to be facilitated by Y16, F128, and W181 residues (Figure 19). The involvement of Y16 (towards the *re*-side of the flavin) and F128 and W181 (towards the *si*-side) in substrate binding is evident from the kinetic data. (Table 30). Additional hydrophobic contacts are provided by P127, Y126, Y19, V11 and V15.

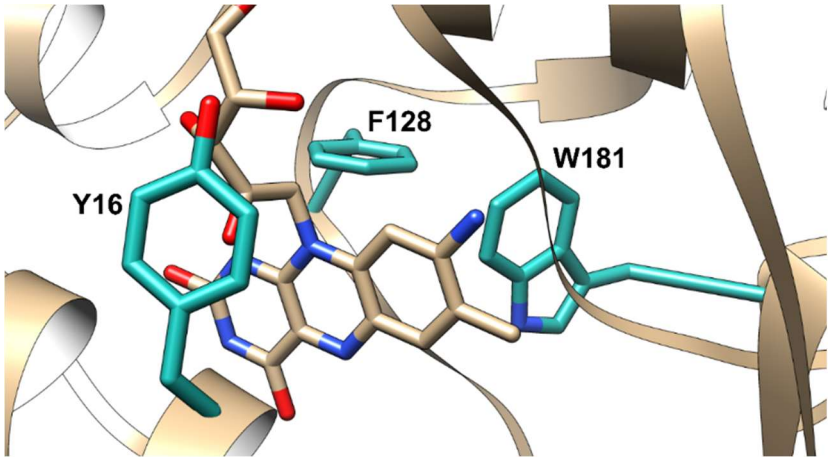


Figure 19: RosC backbone with bound AF highlighting the main substrate binding residues.
Y16, F128, and W181 (green) are the substrate binding residues forming the hydrophobic sandwich around the isoalloxazine ring. The flavin hydrophobic sandwich is a property of flavin binding proteins, used to stabilize the heterocyclic ring of flavins. Mutations in these residues affect the ability of the protein to bind AFP as well as FMN.

4.7.4 D166 is the key residue that mediates discrimination between AFP and FMN

RosC has evolved to dephosphorylate AFP as a substrate instead of FMN for the successful overproduction of roseoflavin. The Aspartate at position 166 could potentially act as the main discriminating residue according to the data obtained from the crystal structure of RosC with AF. According to the crystal structure, other residues –V170, I187, S167, and M168 in the proximity of C8 atom could also be responsible for providing substrate specificity to RosC (Figure 20). Increased hydrophobicity around the C8 could allow better binding of FMN to the protein as compared to AFP. Mutational studies were carried out to confirm the role of surrounding residues and to increase the hydrophobicity around the C8 atom for better binding of the methyl group of riboflavin. The proteins carrying mutations D166L, D166I, D166V, D166T, D166N, V170L, V170I, M186A, I187L

were checked to assess their ability to dephosphorylate FMN. None of the residues showed higher dephosphorylation of FMN compared to AFP. RosC^{D166L} shows a higher specific activity on 200µM FMN with respect to AFP, but the activities are greatly reduced as compared to the RosC wildtype protein. Mutations D166T, D166N and M186A showed a significant decrease in specific activity. According to the structural data, the specificity of AFP against FMN is mainly provided by the salt bridge between AF-N8M and Asp166-OD2; the latter is strongly fixed with the protein matrix (Figure 23). If FMN would bind, the short distance between charged Asp166 and the non-polar C8M methyl group is energetically unfavourable, however, a significant evasion movement is not easily possible. Simultaneous interaction with the ribityl side chain is equally important. All types of rearrangements, perhaps a reorientation of the Asp166 side chain would alter the position/conformation of the ribityl and phosphate group of FMN and directly affect the catalytic activity. The tight binding of AF provided by rigidification, involving D166 and AFP, directly correlates with the specificity against FMN. In addition, the dual binding of D166 to N8M and ribityl/phospho can be considered as an instrument for promoting the specificity of AF against RBF. A mutation that affects the interaction of the C8 carbon without affecting the salt bridges with the ribityl group could not be isolated.

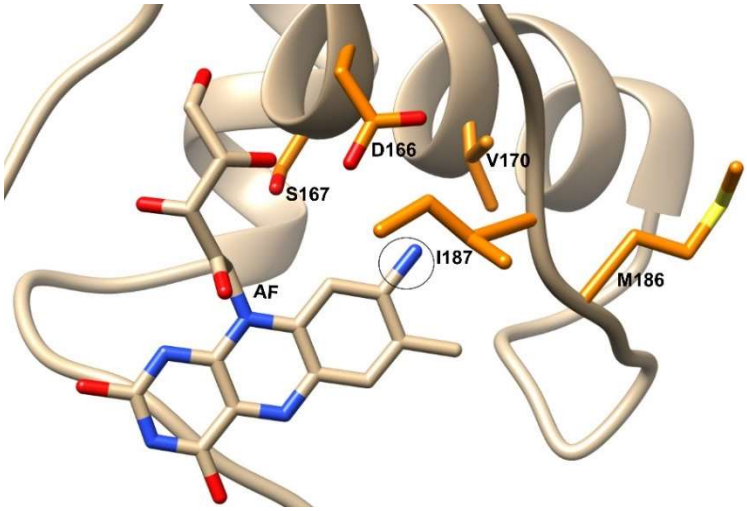


Figure 20: Residues surrounding the amino group at C8 (N8M) of the isoalloxazine ring in AFP.
Residues D166, V170, I187, S167, M186 (orange) were investigated for their involvement in providing substrate specificity to RosC as they closely surround the methyl group in case of FMN or the amino group in case of AFP (circled).

4.7.5 The specificity of AFP binding can be accounted for by its resonance structures

Isoalloxazine rings acquire different resonance states or mesomeric forms depending on the redox state of the molecule. The ability of flavins to acquire different mesomeric forms is due to the conjugated system of double bonds in their isoalloxazine ring, which allows for the delocalization of π electrons The relative stability of the mesomeric states relies heavily on the surrounding residues in the protein-bound state. Possible mesomeric forms of AFP can explain the specificity of hydrogen bond formation between the amino group of AFP and D166. In one of the forms, there is an iminium ion at the C8 position, which can form a salt bridge with the aspartate side chain. No mesomeric forms in FMN allow this interaction due to the presence of a methyl group instead of an amino group in the case of FMN, thus providing a possible reason for stronger AFP binding. (Figure 21).

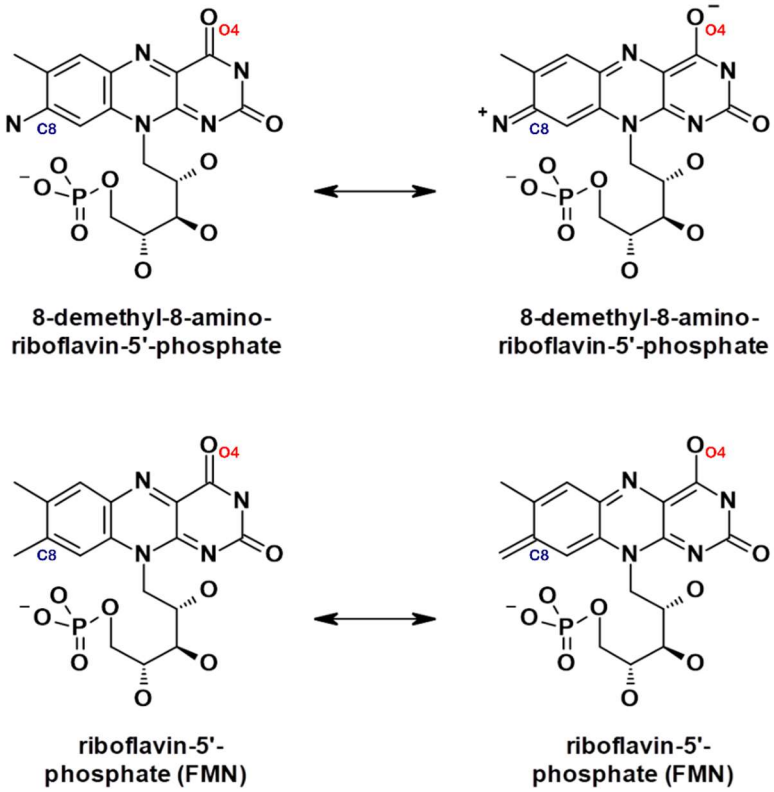


Figure 21: The isoalloxazine ring of both FMN and AFP can attain different mesomeric forms or resonance structures.
In one of the forms, AFP has a positive charge at the N8M group (iminium) which enables the formation of a salt bridge with the Asp166-OD2. The resonance structure of the quinonoid form of AF would allow the formation of a salt bridge between the positively charged iminium ion at C8 and the negatively charged D166 carboxylate of RosC. The negative charge at O4A of the quinonoid form of AF could possibly be stabilized by interaction with the main chain amine of K121 and the proximity of the positively charged R18 and K121 side chains, which polarize a nearby water molecule. Notably, such an interaction would not be possible for FMN.

4.8 Characterization of high-affinity RosC mutants

4.8.1 RosC^{D166E} and RosC^{R33A} were purified with bound riboflavin

Upon purification using affinity chromatography, it was observed that mutants RosC^{D166E} and RosC^{R33A} were purified with a yellow compound bound to them. HPLC analysis of this compound indicated that in both cases, the yellow compound was riboflavin (Figure 22).

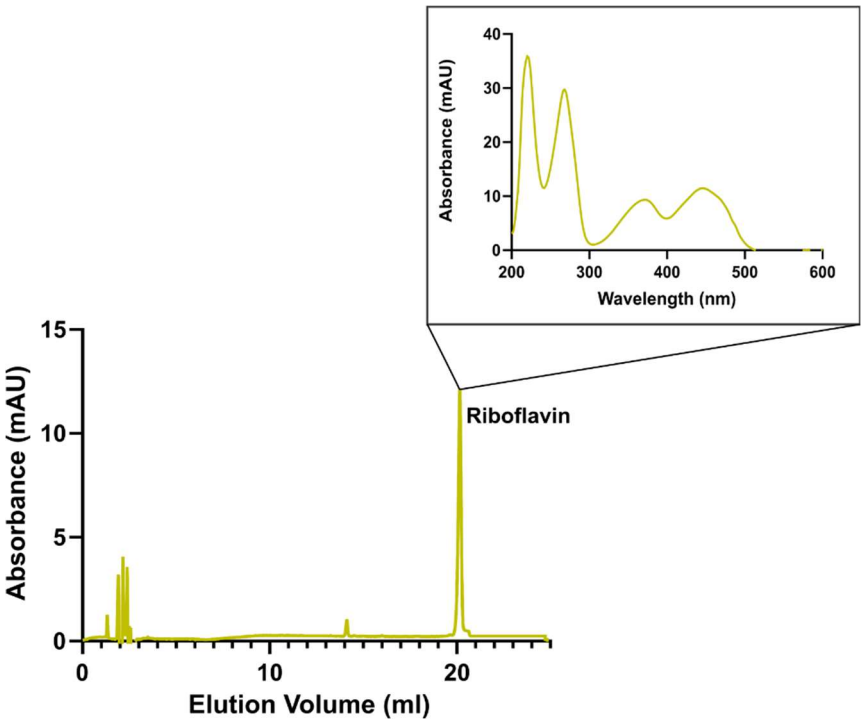


Figure 22: Mutants RosC^{R33A} and RosC^{D166E} were purified with bound riboflavin.
RosC variants RosC^{R33A} and RosC^{D166E} were purified with a bound flavin that imparted them a yellow colour. The flavin was determined to be riboflavin upon HPLC-DAD analysis of the supernatant of denatured protein preparation. The absorbance spectrum and the retention time of the bound compound corresponded to the riboflavin standard, showing an absorbance maximum at $\lambda = 445$ nm.

4.8.2 Crystal structure of R33A mutant with bound riboflavin

The structure of RosC^{R33A} with bound riboflavin gives insights into the mechanism of riboflavin binding by the variant. The structure of the RosC^{R33A}-RBF complex determined at 1.5 Å resolution revealed a highly similar flavin binding mode as described for the RosC-AF complex, including the polypeptide rigidification event (Figure 23). Eliminating the long arginine side chain in the RosC^{R33A} variant results in a shift of Leu190 and Asp166 towards the vacant space occupied in the centre by a solvent molecule that bridges Asp166 and His34 (Figure 23A). Most interestingly, the displaced Asp166 reorients its side chain from the isoalloxazine in the RosC-AF complex (Figure

23B) to the ribityl in the RosC^{R33A}-RBF complex (Figure 23A). Enabled by a small rearrangement of the ribityl, a hydrogen bond is formed between Asp166 and the C4 and C5 hydroxyl groups. Consequently, the distance is increased from 3.0 Å between Asp166-OD2 and AF-N8M to 4.2 Å between the Asp166-OD2 and RBF-C8M. The polypeptide further reacts on the introduced C8M methyl by small side chain rearrangements of Ile187 and Val170 towards the non-polar group and by the displacement of the isoalloxazine ring by 0.5 Å. Thus, the distance of 3.0 Å between Met185-CO and AF-N8M is increased to 3.8 Å between Met185-CO and RBF-C8M. Altogether, the capability of RosC^{R33A} to bind riboflavin with a high affinity can be plausibly rationalized based on the structural data. The mutant R33A is therefore able to bind FMN/riboflavin better due to structural rearrangements and binding is feasible when D166 has space to evade the C8 methyl group as seen in the RosC^{R33A}-RBF structure (Figure 23A). Binding, however, is only feasible at the expense of catalytic inactivity as R33 provides important electrostatic interactions for the phosphatase activity.

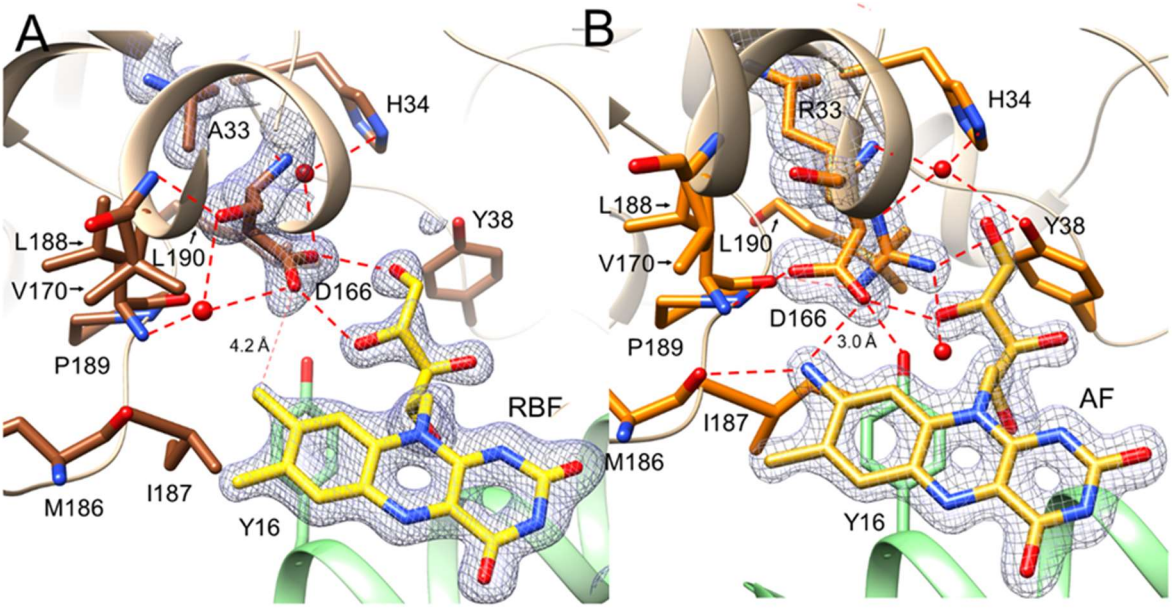


Figure 23: Binding of riboflavin to the variant RosC^{R33A}
A: Structure of the RosC^{R33A} riboflavin complex at 1.5 Å resolution. D166 evades the C8 methyl of riboflavin and interacts with hydroxyl groups at C4 and C5 of the ribityl part (red dashed line). This conformational change is enabled by the space created by replacing the bulky arginine at position 33 (R33) with alanine (A33). **B:** The RosC-AF structure for comparison. D166 forms a salt bridge with the amino group at C8 of the AF isoalloxazine. Without considerable structural arrangements, a flavin with a methyl group at C8, such as riboflavin or FMN, would not bind to the flavin binding site.

4.8.3 Structural characteristics that afford better binding of riboflavin to RosC^{D166E}

Structural prediction of RosC^{D166E} using Alphafold hinted how this exchange could lead to stronger binding of both AFP and FMN by the variant. This was also corroborated by the kinetic data (Figure 24, Table 30). D166 interacts with the amino group of the isoalloxazine ring as well as the hydroxyl group at C4 of the ribityl part. In the case of E166, these interactions are preserved but strengthened due to shortened distances between the interacting groups. The interaction between the Glu-OD2 group and the hydroxyl group at C4 of the ribityl chain is more affected as the distances are significantly reduced in the case of E166 (1.8 Å) as compared to D166 (2.3 Å). This explains the strong and indiscriminate binding of the ribityl part of the flavin to RosC^{D166E}. The interaction between E166 side chain and the amino group is not as strongly affected (bond length reducing from 3.1 Å in case of D166 to 2.9 Å in case of E166). Since the ribityl part is identical in both AFP and FMN, E166 shows a higher specificity towards both compounds.

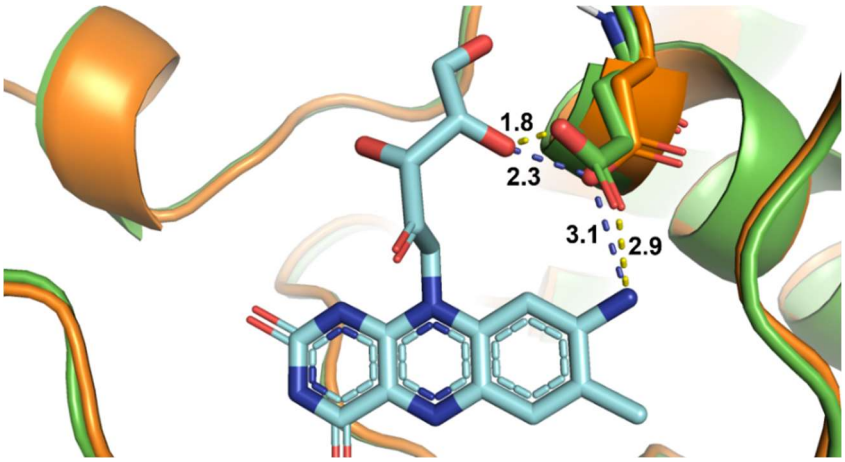


Figure 24: Structural superposition of RosC with bound AFP and structure of RosC^{D166E} predicted using Alphafold.
We postulate that the distance between the oxygen atom (OD2) of the residue at position 166 and the hydroxyl group at C4 of ribityl is shorter in the case when glutamate occupies this position, (E166 - green) in RosC^{D166E} as compared to when aspartate occupies this position in RosC (D166 - orange). The shortening of distance from 2.3 Å (blue dashed lines) in the case of D166 to 1.8 Å (yellow dashed lines) in the case of E166 could be affording a better binding of AFP and FMN to RosC^{D166E}.

4.9 Kinetic studies on RosC and mutants

Steady-state kinetics were determined for RosC and mutants to determine the role of various active site residues. Kinetic constants were determined by using initial reaction rates µmol min^{-1} mg^{-1} protein against the substrate concentrations used. Kinetic constants were calculated from a non-linear regression curve based on the Michaelis Menten equation, and data was plotted and calculated using Graphpad Prism software. For the high-affinity mutants, $RosC^{D166E}$ and $RosC^{V170L}$, kinetic constants were calculated for both FMN and AFP as substrates. The range of substrate concentration was 0.5 – 300 µM for AFP and 2-500 µM for FMN. Purified protein concentration was used in the range of 0.1µM- 1 µM. The kinetic constants for various RosC variants have been summarized in Table 30. RosC has a V_{max} of 1.0 µmol min^{-1} mg $protein^{-1}$ (k_{cat} = 25 min^{-1}) and a K_M of 35 µM (AFP). For dephosphorylation of FMN, the kinetic parameters are V_{max}= 0.2 µmol min^{-1} mg^{-1} (k_{cat} = 5 min^{-1}) and K_M = 309 µM (Schneider *et al*., 2020).

Table 30: Kinetic parameters for $RosC^{StrepN}$ catalyzed dephosphorylation of substrates AFP and FMN

Variant	Substrate	V_{max} (µmol/min/mg)	K_M (µM)	k_{cat} (min^{-1})
V170L	AFP	1.03 ± 0.80	1.33 ± 0.13	26.41
	FMN	0.356 ± 0.028	109.6 ± 17.55	9.09
D166E	AFP	1.57 ± 0.063	11.10 ± 1.96	40.17
	FMN	0.594 ± 0.045	24.45 ± 3.48	15.1
F128A	AFP	0.0249 ± 0.001	22.98 ± 8.79	0.64
W181A	AFP	0.0379 ± 0.005	69.19 ± 27.6	0.97
Y38A	AFP	0.163 ± 0.059	5.69 ± 0.98	4.17
Y16A	AFP	0.403 ± 0.016	1.78 ± 0.356	10.35
P127A	AFP	0.054 ± 0.001	12.91 ± 1.95	1.38
D166A	AFP	0.11 ± 0.014	39.93 ± 15.26	2.81

Several RosC variants were purified and tested for their ability to dephosphorylate 200µM of flavin substrate. Different concentrations of proteins ranging from 0.2-1µM were used for each assay. $RosC^{StrepN}$ was always assayed with the tested mutant to determine the relative activity of the mutant with respect to the wild-type protein. The results have been summarized in Table 31

Table 31: Specific activities of different mutants in the presence of 200µm AFP and FMN.

Mutation	Specific activity (AFP) (µmol/min/mg)	Specific activity (FMN) (µmol/min/mg)	Relative activity with respect to $RosC^{WT}$ for AFP (%)	Relative activity with respect to $RosC^{WT}$ for FMN (%)
D166E	1.5 ± 0.045	0.567 ± 0.026	156.25	283.5
D166A	0.11 ± 0.014	-	12.9	-
D166N	0.0318 ± 0.003	0.0291 ± 0.0037	3.32	14.55
D166L	0.006 ± 0.0004	0.0073 ± 0.0004	0.62	3.6
D166G	0.427 ± 0.057	0.088 ± 0.029	44.48	44
D166V	-	-	-	-
D166I	-	-	-	-
D166T	-	-	-	-
V170L	1.00 ± 0.074	0.234 ± 0.008	104.17	117
V170I	0.793 ± 0.095	0.1956 ± 0.008	82.61	97.5
E119A	0.58 ± 0.016	0.0515 ± 0.005	60	25.75
K121A	0.55 ± 0.06	0.091 ± 0.006	56.88	45.5
M186A	0.0017 ± 0.0001	-	0.17	-
E107Q	0.0016 ± 0.0002	0.0016 ± 0.0002	-	-
H34A+H165A	-	-	-	-
D47A	1.711 ± 0.078	0.007 ± 0.0004	178.23	3.45
Y38A	0.167 ± 0.009	0.0046 ± 0.0014	17.4	2.25
P127A	0.051 ± 0.00245	-	-	-
F128A	0.024 ± 0.002	0.0032 ± 0.004	2.45	1.5
W181A	0.029 ± 0.0024	-	3.02	-
Y19A	0.608 ± 0.03	0.026 ± 0.014	73.4	12.3
Y16A	0.405± 0.005	0.044 ± 0.022	48.7	20
V11A	0.48 ± 0.021	0.021 ± 0.02	57.8	10
V15A	0.890 ± 0.010	0.0142 ± 0.005	107.2	6.7
F8A	0.591 ± 0.049	0.016 ± 0.0002	71.2	7.6
G22P	-	-	-	-
G25P	-	-	-	-
$RosC^{MBP}$	0.319 ± 0.009	0.0976 ± 0.007	40	48.5
$\Delta22RosC^{MBP}$	-	-	-	-

4.10 The function of the N-terminal cap domain of RosC

4.10.1 General Overview

The cap domain of RosC is a unique structure that is absent in the rest of the characterized members of the histidine phosphatases. This alpha helix structure extends into the active site of the adjacent monomeric unit and participates in substrate binding as mentioned in Section 4.7. The inner residues, facing the flavin are hydrophobic/apolar and participate in shielding the core of the enzyme from the surrounding solvent, while the residues that are in contact with the surrounding solvent are polar and responsible in mediating solvent contacts, ensuring solubility of the protein. The substitution of any of these amino acids with alanine does not significantly affect the activity of the enzyme for AFP as well as FMN (Table 31), indicating that the alpha helix serves as a lid, the primary role of which is to provide a hydrophobic cover to facilitate the binding of the flavin.

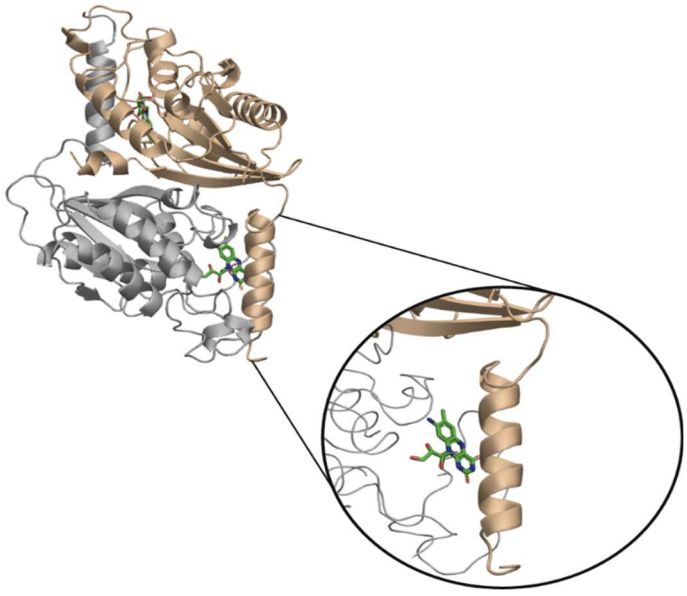


Figure 25: RosC dimer highlighting the cap domain of RosC with bound flavin.
The X-ray structure of RosC in the presence of 8-demethyl-8-amino-riboflavin (AF) revealed a compact homodimeric enzyme (one monomer is shown in gold, the other in gray). The core fold of the enzyme is supplemented by strand-helix insertions and by a unique 22 amino acid N-terminal helical extension. The latter is separated from the core subunit and associates with the partner subunit.

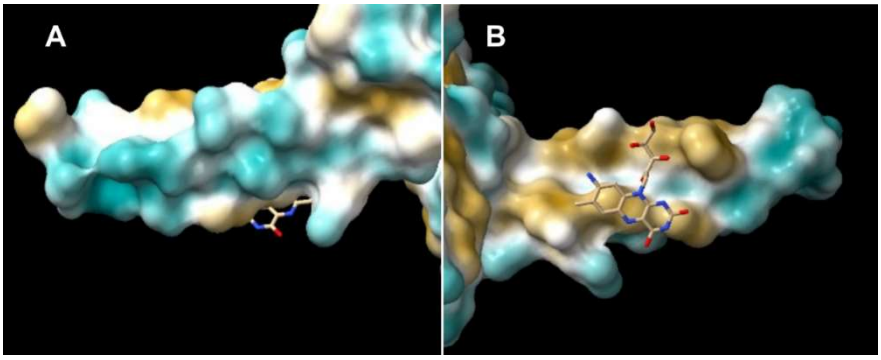


Figure 26: Surface diagram of the cap domain of RosC with bound AF.
Cap domain with hydrophobic residues - F8, V11, M12, V15, Y16, and Y19 (yellow) towards the inner side for the binding and positioning of the flavin. Polar residues- G6, S7, G10, R13, S14, G17, are present towards outer side of the cap domain (blue) for mediating polar contacts with the surrounding solvent.

4.10.2 Deletion of the cap domain affects solubility of RosC

The deletion of the *rosC* cap domain affects the solubility as well as the activity of the protein. There was a loss of solubility upon the deletion of 22 amino acids (that fold into the helix) from the N-terminal region of RosC (Δ22RosC). This protein was C-terminally His-tagged (Δ22RosC^{HisC}). The proteins could not be visualized using SDS PAGE and were detected by blotting them on a nitrocellulose membrane (Figure 27). The protein was probed with an Anti-6x-His Tag Mouse Monoclonal Antibody after DAB staining and was visualized in the pellet fraction of the cell-free extract.

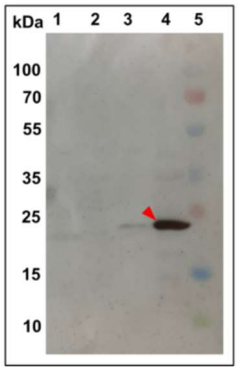


Figure 27: Detection of C-terminally his-tagged Δ22RosC (Δ22RosC^{HisC})
Soluble and pellet fractions of cell-free extracts overproducing the Δ22RosC^{HisC} were blotted on a nitrocellulose membrane and detected by western blot using Anti-6x-His tag mouse monoclonal antibody (1:800 dilution) upon staining with diaminobenzidine (DAB). **Lane 1:** Supernatant of cell-free extract (CFE) of *E. coli* BL21 pLySs containing pET24a with Δ22RosC^{HisC}, uninduced. **Lane 2:** Pellet of CFE of *E. coli* BL21 pLySs containing pET24a with Δ22RosC^{HisC}, uninduced. **Lane 3:** Supernatant of CFE of *E. coli* BL21 pLySs containing pET24a with Δ22RosC^{HisC}, induced using 1mM IPTG. **Lane 4:** Pellet fraction of cell-free extract of *E. coli* BL21 pLySs containing pET24a with Δ22RosC^{HisC}, induced using 1 mM IPTG. **Lane 5.** Protein molecular weight marker. Δ22RosC^{HisC} was detected in the pellet fraction of the CFE. A total of 20µg protein was loaded in each lane. Red arrow indicates the position of the protein, Δ22RosC.

4.10.3 Solubilization of Δ22RosC using a maltose binding protein (MBP) tag

The protein could be solubilized again using the addition of a maltose-binding protein (MBP) tag. The MBP tagged Δ22RosC mutant could be detected at around 65 kDa which corresponds to the calculated molecular weight (64.3 kDa) of the protein. This protein was then purified from the cell-free extract of E. *coli* NEB express strain using an amylose column. The protein was however inactive and was unable to dephosphorylate both AFP and FMN. RosC WT with an N-terminal MBP tag ($RosC^{MBP}$) was used as a control. Both FMN and AFP could be successfully dephosphorylated by $RosC^{MBP}$ indicating that the presence of the bulky MBP tag does not completely hinder the active site of the protein. The addition of the MBP tag, however, affects the specific activity of RosC as the specific activity of the protein is 0.32 µmol/min/mg, only 40% of $RosC^{StrepN}$ (Table 31)

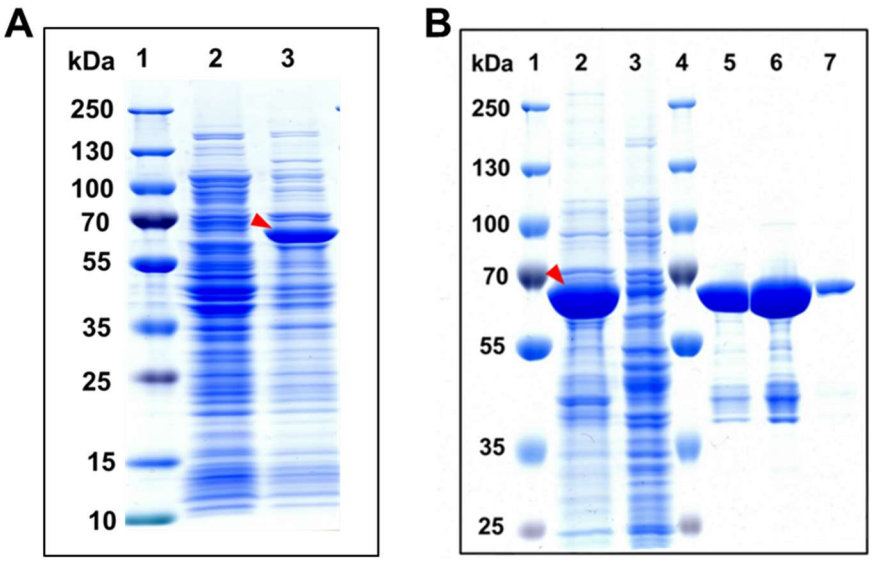
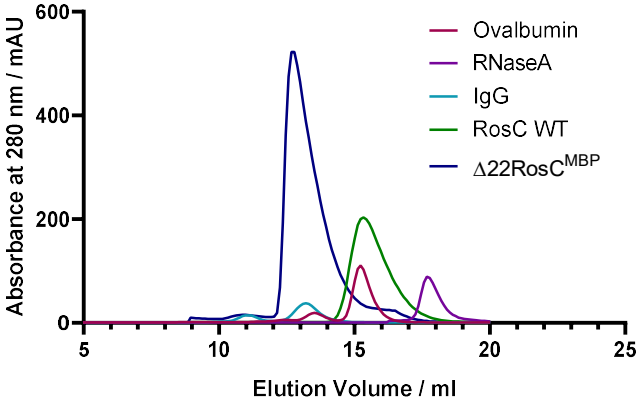


Figure 28: Overproduction and purification of $\Delta22RosC^{MBP}$
SDS-PAGE gels depicting expression of Δ22*rosC*_MBP in *E. coli* NEB express strain and purification of the resulting fusion protein, $\Delta22RosC^{MBP}$ using gravity flow columns containing amylose resin. **A. Lane 1**: Protein molecular weight marker **Lane 2:** Cell-free extract (CFE) of *E. coli* NEB express cells containing pMAL plasmid with *rosC* gene having 22 amino acids deleted from N-terminal end and containing the N-terminal MBP tag ($\Delta22RosC^{MBP}$) without IPTG induction. **Lane 3:** CFE of *E. coli* NEB express cells overproducing the $\Delta22RosC^{MBP}$ protein. Cultures were induced using 1mM IPTG. Red arrow marks the position of $\Delta22RosC^{MBP}$ protein **B. Lane 1:** Protein molecular weight marker **Lane 2:** Cell-free extract (CFE) of *E. coli* from NEB express cells containing pMAL plasmid overproducing Δ22RosC with the N-terminal MBP tag ($\Delta22RosC^{MBP}$) **Lane 3:** Flow-through obtained after wash steps. **Lanes 4:** Protein molecular weight marker. **Lanes 5-7:** Different fractions of eluted $\Delta22RosC^{MBP}$ protein after affinity chromatography using an amylose column. Lanes 3 and 4 contain 20µg protein each. In lanes 5-7, simply 20µl of eluted protein was loaded to check for purity.

4.10.4 Deletion of cap domain does not affect the dimerization of RosC

Size exclusion chromatography was used to determine the role of the cap domain in the dimerization of the protein. Standard proteins with known molecular masses were used and RosC wildtype ($RosC^{StrepN}$) was used as a control. Different proteins eluted differently from the column based on their molecular mass. RosC is known to be a dimeric protein and has a molecular mass of 25.58 kDa. The molecular mass of the MBP tagged Δ22RosC ($\Delta22RosC^{MBP}$) is ~64 kDa and that of the dimeric protein is ~128 kDa. According to the data from size exclusion chromatography, $\Delta22RosC^{MBP}$ protein eluted at an elution volume that is close to that of IgG protein, which has a molecular mass of 150 kDa, indicating that $\Delta22RosC^{MBP}$ elutes as a dimeric protein. The N-terminal cap domain is not involved in the formation of the RosC dimer and other interactions mediate the dimerization of the two monomeric subunits.



Protein	Molecular mass (kDa)	Elution volume (ml)
RNase A	13.7	17.88
Ovalbumin	42.7	15.22
IgG	150	12.74
RosC WT	25.58 (monomeric mass)	15.39
$\Delta22RosC^{MBP}$	Unknown	12.75

Figure 29: Size exclusion chromatography to determine to determine oligomeric states of $RosC^{StrepN}$ and $\Delta22RosC^{MBP}$
Native molecular masses of $RosC^{StrepN}$ and $\Delta22RosC^{MBP}$ were determined by size exclusion chromatography using a Superdex® 200 Increase 10/300 GL column in 10 mM Tris HCl, 150 mM NaCl, pH 7.6 at a flow rate of 1 ml per min. ribonuclease A (1; 13.7 kDa; V = 17.88 ml) ovalbumin (2; 42.7 kDa; V = 15.22 ml), Immunoglobin G (3; 150 kDa; V = 12.74 ml) were used as standards. RosC (green line, V = 15.39 ml) eluted similarly to ovalbumin (2), which has twice the molecular mass of RosC. This suggests that native RosC is a dimer with an approximate molecular mass of 51.16 kDa. N-terminally MBP tagged RosC ($\Delta22RosC^{MBP}$) (64 kDa as a monomer; 128 kDa as a dimer) eluted similarly as IgG with a molecular weight of 150 kDa. This suggests that $\Delta22RosC^{MBP}$ is a dimeric enzyme as well.

4.11 FMN phosphatases from the HAD superfamily from *S. davaonensis*

4.11.1 Bioinformatic studies

BLASTp tool was used to mine proteins in *Streptomyces davaonensis* sharing sequence similarity to known FMN hydrolase from *E. coli*- YbjI. According to the BLASTp results, two different proteins with significant sequence similarity to YbjI were identified to be putative FMN hydrolases in *S. davaonensis*. The genes BN159_0855 and BN159_4587 were successfully expressed in the *E. coli* Rosetta DE3 strain (Figure 30).

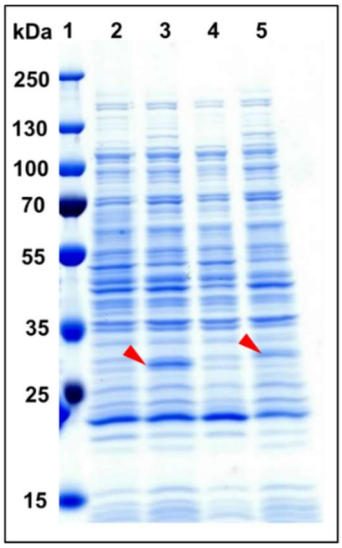


Figure 30: Overproduction of gene products of BN159_0855 and BN159_4587 from *S. davaonensis* in *E. coli*
SDS PAGE to detect the expression of putative FMN hydrolases from *S. davaonensis* in *E. coli* Rosetta DE3 cells. Soluble fractions of cell-free extracts of *E. coli* cells were loaded in each well. Lanes from left to right- **Lane 1:** Protein molecular weight marker. **Lane 2:** Cell-free extract (CFE) of *E. coli* Rosetta strain with pET24a containing gene, BN159_0855, without IPTG induction. **Lane 3:** CFE of *E. coli* Rosetta strain with pET24a containing BN159_0855, induced using 1mM IPTG. **Lane 4:** CFE of *E. coli* Rosetta strain with pET24a containing BN159_4587, without IPTG induction. **Lane 5:** CFE of *E. coli* Rosetta strain with pET24a containing BN159_4587, induced using 1mM IPTG. The calculated molecular masses of BN159_0855 and BN159_4587 are 29.41 kDa and 31.18 kDa, respectively. The corresponding bands can be visualized on the gel when the cultures are induced using IPTG and are marked using red arrow. The total protein loaded in each lane was 20 µg.

4.11.2 Gene products of BN159_0855 and BN159_4587 show FMN hydrolase activity

The cell-free extracts of *E. coli* Rosetta DE3 overexpressing putative FMN hydrolase genes from *S. davaonensis* were tested to determine FMN phosphatase activity. Formation of the product, riboflavin, was observed only in the cases where the protein production was induced by the addition of IPTG, indicating that the protein products of the two genes can dephosphorylate FMN (Figure 31).

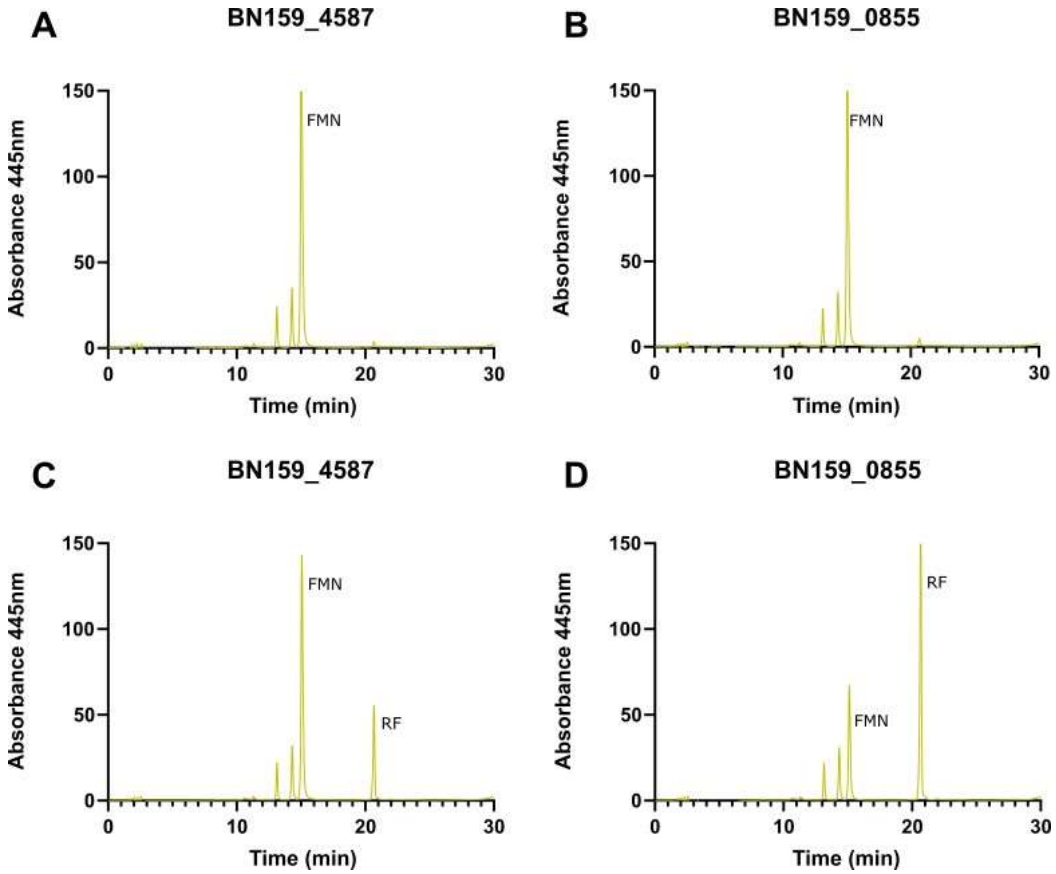


Figure 31: Dephosphorylation of FMN by gene products of BN159_4587 and BN159_0855
Cell-free extracts of *E. coli* Rosetta DE3 strains overexpressing genes BN159_4587 and BN159_0855 were checked for their ability to dephosphorylate FMN. Reactions were carried out at 37 ºC using 100 µM FMN, 100 mM BTP buffer with 200 mM $MgCl_2$ at pH 7.6. **A and B:** Reactions when the enzyme production was not induced (Uninduced). **C and D:** Reactions when the enzyme production was induced upon the addition of 1mM IPTG (Induced). All the reactions were stopped after 60 mins of incubation and samples were treated with 5 % w/v TCA and analyzed using HPLC. Dephosphorylated product- Riboflavin (RF) was observed at the end of 60 mins in samples where the protein production was induced (C and D).

4.11.3 Gene products of BN159_0855 and BN159_4587 show ARPP hydrolase activity

The cell-free extracts of *E. coli* Rosetta DE3, overexpressing genes, BN159_4587 and BN159_0855 from *S. davaonensis* were tested for their ability to dephosphorylate ARPP. ARPP dephosphorylation was monitored using HPLC after derivatization of the reaction components using the fluorescence detector (HPLC-FLD) at λ_{ex}= 410 nm and λ_{em}= 485 nm. It was observed that both BN159_0855 and BN159_4587 were able to dephosphorylate ARPP. RosC is unable to dephosphorylate ARPP and was used as a negative control (Figure 32).

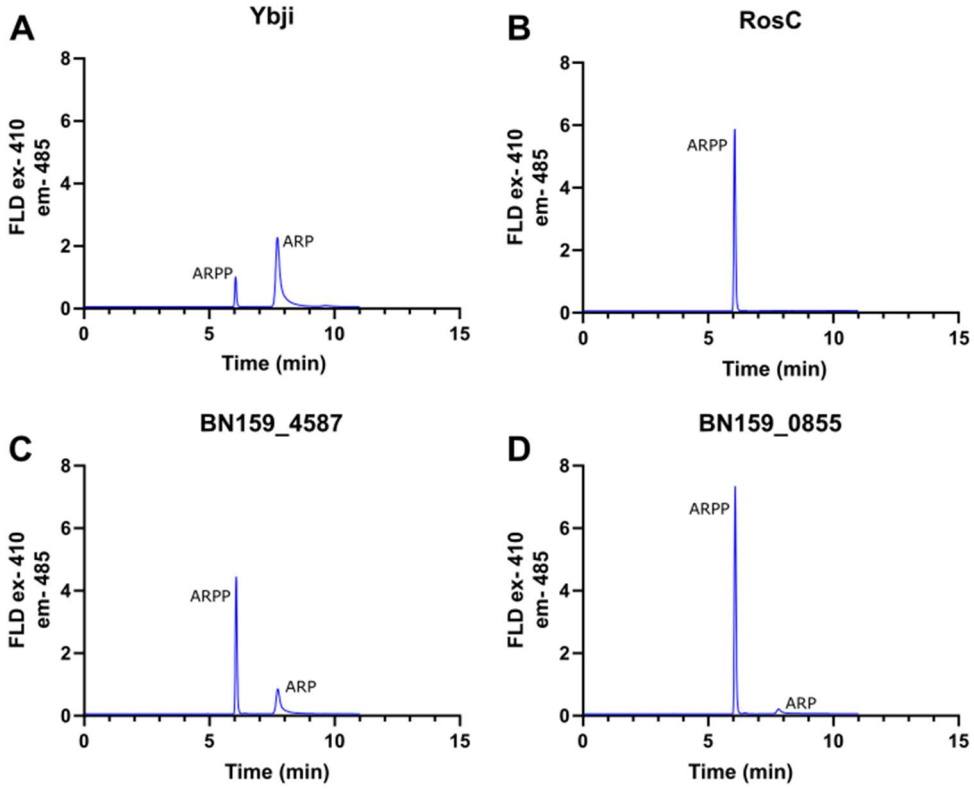


Figure 32: Dephosphorylation of ARPP by gene products of BN159_4587 and BN159_0855
The substrate, ARPP was produced from GTP using *ribA* and *ribD* from *E. coli* as described by (Haase *et al*., 2013); supporting information. The synthesized ARPP was used as a substrate for phosphatase reactions with cell-free extracts of *E. coli* overproducing the gene products from *ybjI*, BN159_0855, and BN159_4587. Purified n-strep tagged RosC ($RosC^{StrepN}$) was used for the assay. Reaction mixtures contained freshly synthesized ARPP, 100 mM BTP buffer with 200 mM $MgCl_2$ (200 mM $CaCl_2$ was used for RosC) at pH 7.6 along with cell-free extracts/purified protein. Reaction tubes were incubated at 37 ° C for 60 mins. The products were derivatized with diacetyl and analysed using HPLC. **A:** YbjI from *E. coli* is known to dephosphorylate ARPP and FMN and was used as a positive control. **B:** RosC cannot dephosphorylate ARPP and, therefore, was used as a negative control. **C and D:** BN149_4587 and BN159_0855 from *S. davaonensis* were also tested for ARPP phosphatase activity. A small amount of the dephosphorylated product, ARP was observed 60 mins after the reaction with the enzymes (C and D).

4.11.4 Gene products of BN159_0855 and BN159_4587 show AFP hydrolase activity

Since FMN and AFP are structurally similar, in order to confirm the substrate specificity of the FMN hydrolases from *S. davaonensis*, their ability to dephosphorylate AFP was tested using an in-vitro assay using $MgCl_2$ as a cofactor. After 60 minutes, the product, AF was observed in reaction mixtures expressing both BN159_0855 and BN159_4587, indicating that both can dephosphorylate AFP (Figure 33).

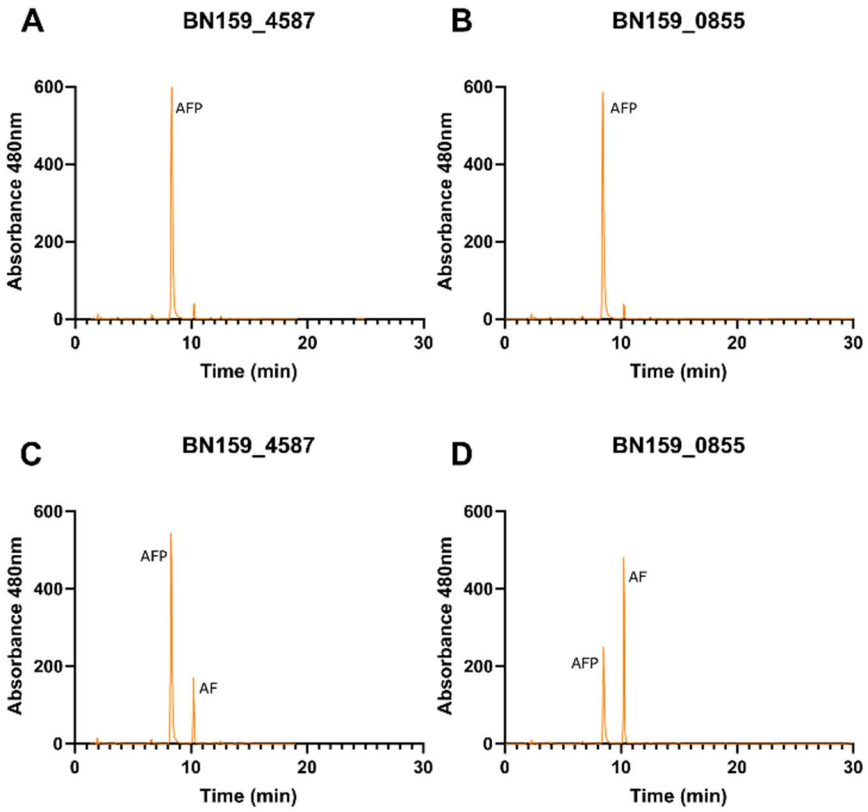


Figure 33: Dephosphorylation of AFP by gene products of BN159_4587 and BN159_0855
Cell-free extracts of cultures overproducing the products of genes, BN159_4587 and BN159_0855 were checked for their ability to dephosphorylate AFP. Reactions were carried out at 37 ⁰ C using 200 µM AFP, and 100 mM BTP buffer with 200 mM $MgCl_2$ at pH 7.6. **A and B :** Reaction products when the enzyme production was not induced (Uninduced). **C and D:** Reaction products when the enzyme production is induced upon the addition of 1mM IPTG (Induced). All the reactions were stopped after 60 mins of incubation and samples were treated with 5% w/v TCA and analysed using HPLC. Dephosphorylated product- AF was observed at the end of 60 mins in samples where the protein production was induced (C and D).

4.11.5 FMN/ARPP hydrolases from *S. davaonensis* and RosC share no sequence or structural similarity

Sequence alignment of protein sequences of known FMN hydrolases from *E. coli, B.subtilis,* and *S. davaonensis* were compared with RosC. There was an overall lack of sequence similarity and the motifs conserved in the HAD superfamily were missing in RosC. Similarly, the key residues that make up the catalytic machinery of RosC were missing in the sequences of HAD superfamily. Upon the structural alignment of RosC with the known FMN hydrolases from *S. davaonensis* from this study, it was observed that RosC that there is very little structural similarity between RosC and the two proteins. Even though both proteins have evolved to catalyze similar reactions, they belong to different superfamilies and have a differently organized catalytic core.

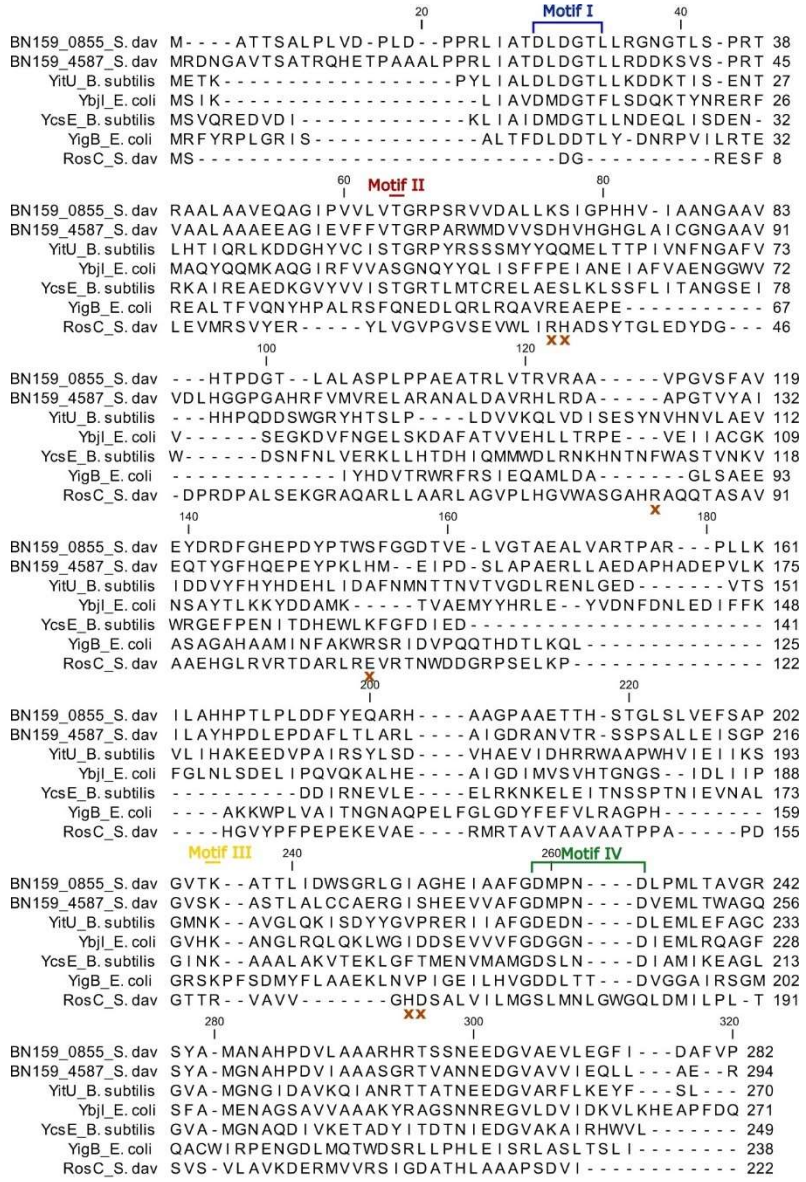


Figure 34: Primary structure alignment of FMN hydrolases from different organisms with 8-demethyl-8-amino-riboflavin-5'-phosphate phosphatase (RosC) from *S. davaonensis.*
Sequence of FMN hydrolases such as BN159_0855 (UniProt ID- K4QXY1), BN159_4587 (UniProt ID - K4QY35) from *S. davaonensis*, YitU (UniProt ID - P70947) and YcsE (UniProt ID - P42962) from *B. subtilis*, YbjI (UniProt ID - P75809) and YigB (UniProt ID - P0ADP0) from *E. coli* were compared with RosC from *S. davaonensis.* The members of the HAD superfamily show a poor sequence similarity within the superfamily but show the presence of certain conserved motifs. Conserved motifs from Haloacid dehalogenases are indicated in different colours (Motif I- Blue, Motif II- Red, Motif III- Yellow and Motif IV- Green). The key residues responsible for RosC activity have been indicated by "x" marking (orange)

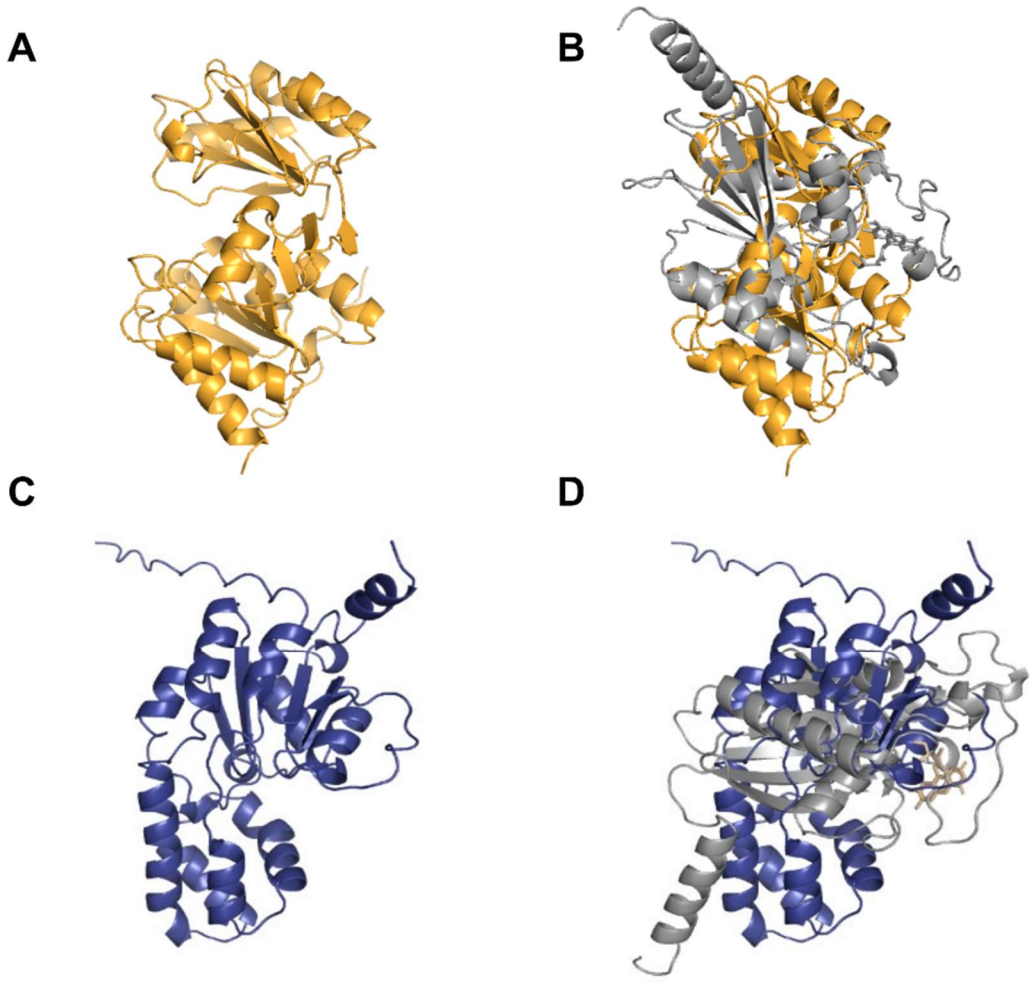


Figure 35: Structural comparison of RosC with FMN hydrolases from *S. davaonensis*.
The structures (A and B) of the gene products of BN159_0855 (UniProt ID - K4QXY1) and BN159_4587 (UniProt ID - K4QY35) were predicted using AlphaFold (Jumper *et al*. 2021) do not show any structural similarity to the FMN phosphatase, RosC, as is visible upon their structural alignments (C and D).

4.12 *rosC*^{D166L}, *rosC*^{D166V} and *rosC*^{D166I} produce toxic gene products in *E. coli*

Mutant RosC^{D166L} could not be optimally produced using the standard expression system used for overexpressing RosC wildtype and other mutants. It was observed that the gene expression caused a decline in cell density with time, showing a growth arrest approximately 1-1.5 hours after IPTG induction of the gene. Hydrophobic exchanges of the D166 led to the production of variants that showed the same physiological effect in *E. coli,* wherein the resulting proteins were toxic to the *E. coli* host cells.

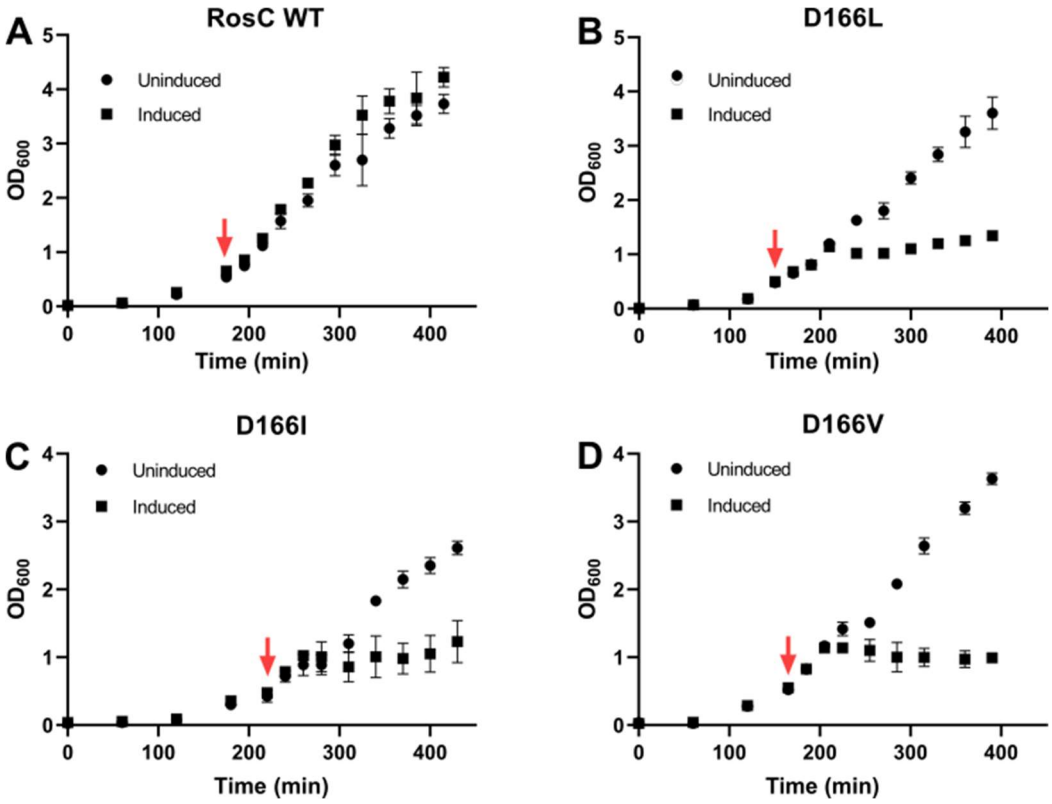


Figure 36: Growth was greatly reduced in recombinant *Escherichia coli* strains that overproduced RosC^{D166L}, RosC^{D166I} and RosC^{D166V}.
The growth of *E. coli* strains containing plasmids that enable overexpression of the *rosC* wild-type gene and variants was monitored. The red arrow shows the timepoint of induction of (A) RosC wild-type (WT, control) (B) RosC^{D166L}, (C) RosC^{D166I} and (D) RosC^{D166V}.

4.12.1 RosC^{D166L} is a toxic protein in *E. coli* and *C. glutamicum* but not in *B. subtilis*

When RosC^{D166L} was overproduced in *E. coli,* the growth of the cultures declined with time upon induction of protein production. A similar decline of cell growth was observed in the case of *C. glumaticum* but not *B. subtilis* host cells. The overproduced protein could be detected upon overexpression in *B. subtilis* but not in *C. glutamicum.*

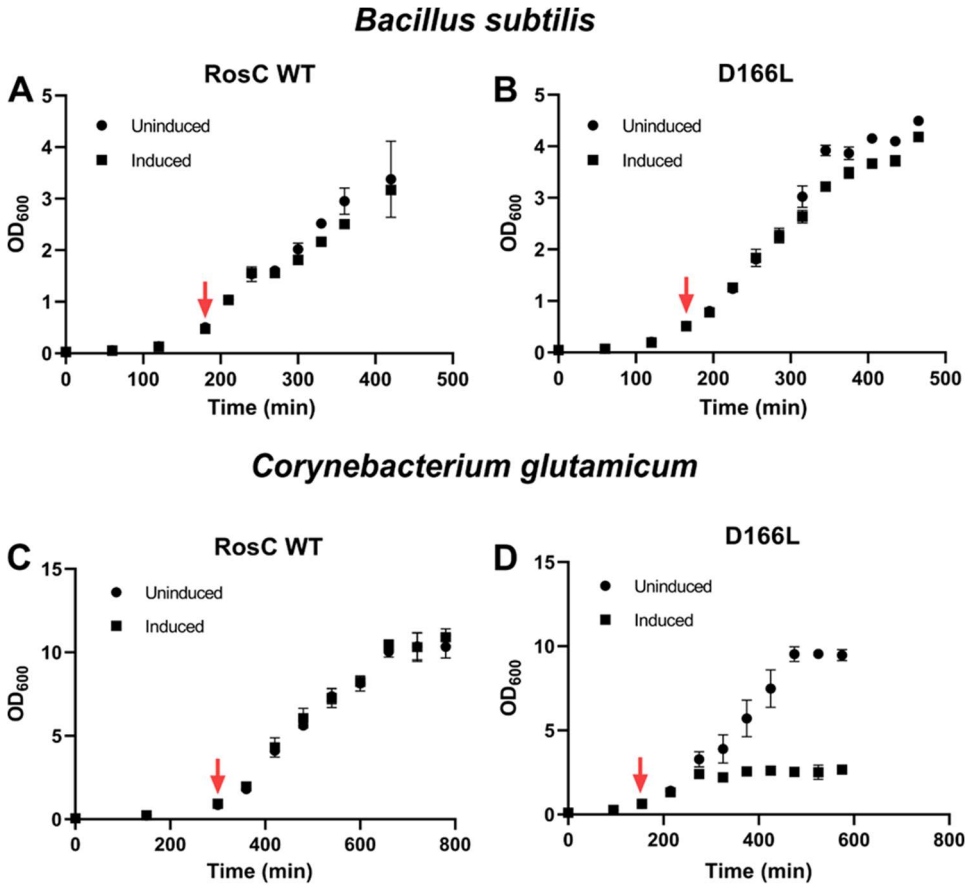


Figure 37: Overproduction of RosC^{D166L} is toxic for *Corynebacterium glutamicum* but not for *Bacillus subtilis*.
Growth of recombinant *B. subtilis* (A and B) and *C. glutamicum* (C and D) strains that overexpress the *rosC* wildtype gene and variants were monitored. The red arrow shows the time point of induction. Reduced growth upon overproduction of RosC^{D166L} was only observed in *C. glutamicum* but not in *B. subtilis*. The successful overproduction of RosC^{D166L} in *B. subtilis* was detected using SDS PAGE. It appears that the target metabolite for the phosphatase RosC^{D166L} is present in *E. coli* and *C. glutamicum* but not in *B. subtilis*. The red arrow shows the timepoint of induction.

4.12.2 Overproduction of RosC^{D166L} variant is suppressed in *E. coli* and *C. glutamicum*

When *E. coli* and *C. glutamicum* were used as expression hosts for the overproduction of RosC^{D166L}, it was observed that while the wildtype protein could be sufficiently overproduced the overproduction of RosC^{D166L} was severely suppressed, and the protein could not be visualized using SDS PAGE (Figure 38). When *B. subtilis* was used as an expression host, both RosC and RosC^{D166L} proteins were visible using SDS PAGE. The protein production levels were comparable for the wildtype and the mutant protein (Figure 38).

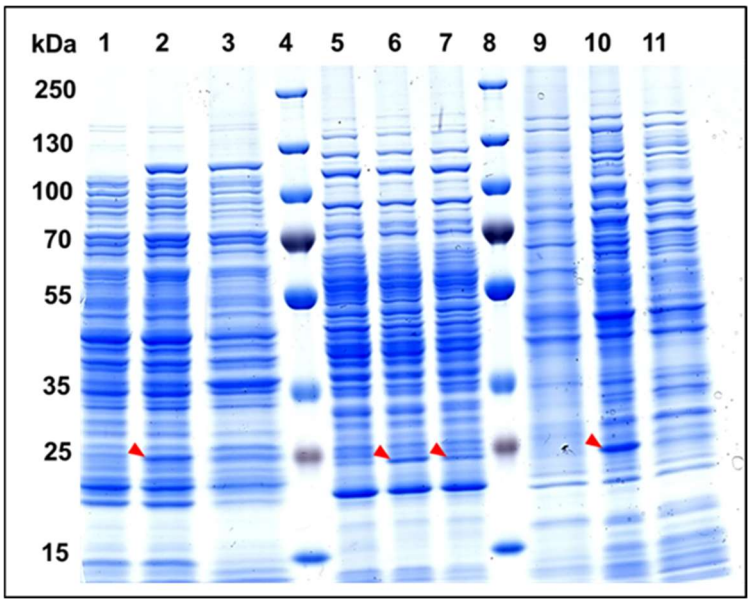


Figure 38: Overproduction of RosC and RosC^{D166L} in *E. coli*, *B. subtilis* and *C. glutamicum*
SDS PAGE depicting the results of expression check of *rosC* and *rosC*^{D166L} in different hosts. Lanes from left to right- **Lane 1:** Cell-free extract (CFE) of *E. coli* BL21 pLySs with pET24a containing *rosC* without IPTG induction**. Lane 2:** CFE of *E. coli* BL21 pLySs with pET24a containing *rosC* with IPTG induction**. Lane 3:** CFE of *E. coli* BL21 pLySs with pET24a containing *rosC*^{D166L} with IPTG induction. **Lane 4:** Protein molecular weight marker **Lane 5:** CFE of *B. subtilis* 168 with pET24a containing *rosC* without IPTG induction. **Lane 6:** CFE of *B. subtilis* 168 with pHT254 containing *rosC* with IPTG induction. **Lane 7:** CFE of *B.subtilis* 168 with pET24a containing *rosC*^{D166L} with IPTG induction. **Lane 8:** Protein molecular weight marker. **Lane 9:** CFE of *C. glutamicum* with pMKEx2 containing *rosC* without IPTG induction. **Lane 10:** CFE of *C. glutamicum* with pMKEx2 containing *rosC* with IPTG induction**. Lane 11:** CFE of *C. glutamicum* with pMKEx2 containing *rosC*^{D166L} with IPTG induction. The *E. coli* and *B. subtilis* expression strains were induced at an OD_{600} of 0.5 and *C. glutamicum* expression strains were induced at an OD_{600} of 0.8. IPTG (1 mM) was used for induction in each case. Overproduction of RosC^{D166L} led to a growth arrest in a recombinant *E. coli* and *C. glutamicum* strain but not in a recombinant *B. subtilis* strain. SDS-PAGE/Coomassie Brilliant Blue R-250 staining of cell-free extracts of the recombinant bacteria showed that RosC ^{D166L} is produced in both bacteria, however, in very small amounts and cannot be detected using SDS PAGE. Production of wild-type RosC was possible in all three hosts and in these cases larger amounts of the protein were found which are visually detectable. CFE containing 20µg total protein were loaded in each lane. The position of the red arrow indicates the position of RosC and RosC^{D166L} proteins.

4.12.3 Visualization of his-tagged RosC and variant RosC^{D166L} using western blot

Since the overproduction of toxic protein, RosC^{D166L} in *E. coli* BL21 pLySs could not be detected using SDS PAGE, both RosC and the variant RosC^{D166L} were C-terminally his-tagged to be detected immunologically using western blot. RosC^{D166L} could be observed but only as a very faint band (Figure 39), in comparison to RosC. The toxic protein is produced in small amounts produced and its production seems to be suppressed. The minute amounts of protein that are produced are toxic to *E. coli* host cells.

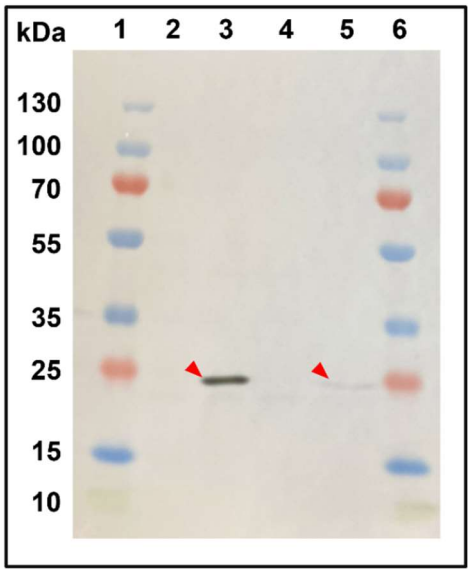


Figure 39: Overproduction of C-terminally his-tagged RosC and RosC^{D166L} proteins
Soluble fractions of cell-free extracts of *E. coli* were blotted on nitrocellulose membrane and visualized using western blot using anti-6x-His tag mouse monoclonal antibody (1:1000 dilution) and stained with diaminobenzidine (DAB). Lanes from left to right. **Lane 1:** Protein molecular weight marker. **Lane 2:** Cell-free extract (CFE) of *E. coli* BL21 pLySs containing pET24a with RosC wildtype protein, without IPTG induction. **Lane 3:** CFE of *E. coli* BL21 pLySs containing pET24a with *rosC wildtype* gene, induced using 1mM IPTG. **Lane 4:** CFE of *E. coli* BL21 pLySs containing pET24a with *rosC*^{D166L} gene, without IPTG induction. **Lane 5:** CFE of *E. coli* BL21 pLySs containing pET24a with *rosC*^{D166L} gene, induced using 1mM IPTG. **Lane 6:** Protein molecular weight marker. Red arrows indicate the position of the RosC and RosC^{D166L} proteins, around 25 kDa. Cells were induced at an OD_{600} of 0.8 and grown for 3 hours after induction. A total of 10µg protein is loaded in each lane. RosC protein (Lane 3) appears as a prominent band indicating that the protein is optimally produced in the expression host while the RosC^{D166L} (Lane 5) is visible as a faint band indicating low levels of production in the host cell.

4.12.4 Several compounds were tested to determine the substrate of RosC^{D166L}

A phosphate colorimetric kit was used to determine the phosphatase activity of RosC^{D166L} towards a variety of compounds that could be potential targets of the enzyme. FMN was initially considered as a potential target but was ruled out as the RosC^{D166L} could not dephosphorylate FMN as effectively as the RosC wildtype protein (Figure 40). Since the overproduction of RosC wildtype is not toxic in *E. coli* host cells, FMN was ruled out as the potential target of RosC^{D166L}. ATP, ADP, AMP, GTP, GMP, GDP, TTP, CTP, ribose-5-phosphate, fructose-6-phosphate, fructose-1,6-phosphate, and ARPP were tested as potential substrates, but no phosphatase activity was detected for any of the compounds.

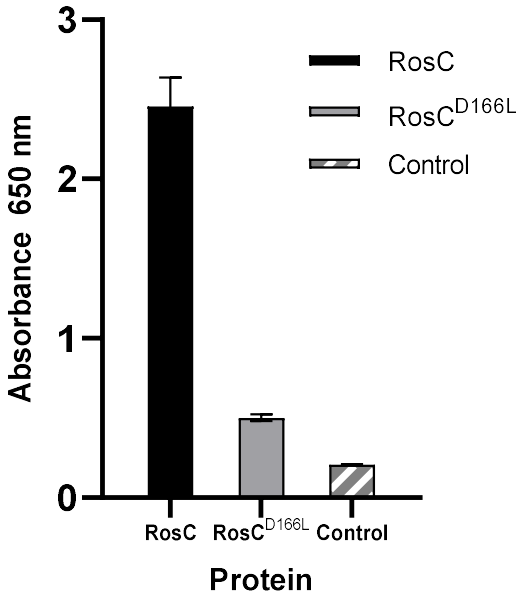


Figure 40: Colorimetric phosphate release assay for detection of phosphatase activity.
As a standard, phosphatase activity of the proteins towards FMN was compared and is depicted here as an example. 100µM FMN was used as a substrate for RosC and RosC^{D166L} to determine the phosphatase activity of the enzymes towards FMN. Purified proteins (0.8 µM) were used. Samples were incubated in the dark for 30 mins after the addition of reagent, and absorbance was measured at 650 nm (A_{650}). No enzyme was added in the case of the control reaction. Other compounds, listed above, were assayed similarly. Each assay was performed in triplicates.

4.12.5 Cell death due to overproduction of RosC^{D166L} is not due to a phosphatase activity

When a double mutant of RosC carrying the mutations E107A and D166L (RosC$^{D166L+E107A}$) was overproduced in *E. coli* BL21 pLySs host cells, the effect was similar to what was observed upon the overproduction of RosC^{D166L}. RosC$^{D166L+E107A}$ lacks the ability of the enzyme to dephosphorylate substrates as effectively as the wildtype enzyme due to mutation of the proton donor, a key residue in the dephosphorylation process. The single mutant, RosC^{E107Q} showed a loss of more than 99% of the phosphatase activity as compared to the wild-type protein. The toxic effect of the double mutation (D166L+E107A) has a similar physiological effect as that of single mutation (D166L), indicating that toxic effect is not due to a phosphatase reaction. The toxic effect could be attributed to the sequestering of an important cofactor or metabolite by the enzyme due to change in its binding preference.

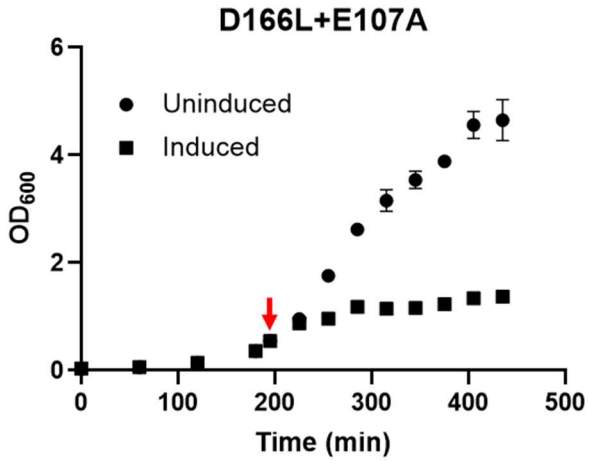


Figure 41: Overproduction of RosC$^{D166L+E107A}$ negatively affects growth of *E. coli* BL21 pLySs host cells. Cultures overproducing the RosC variant, RosC$^{D166L+E107A}$ were grown on LB medium in baffled flasks with constant shaking at 37 ° C. OD_{600} was monitored continuously, and cultures were induced using 1mM IPTG when an OD_{600} of 0.5 was reached. OD_{600} was measured for the cultures after every 30 mins, until stationary phase was reached. The red arrow shows the time point of induction. Both *rosC*$^{D166L+E107A}$ and *rosC*^{D166L} (Figure 36) produce toxic gene products *in E. coli* BL21 pLySs.

4.12.6 RosC^{D166L} binds lumichrome

RosC^{D166L} was purified with a lightly yellow coloured compound bound to it which was detected using fluorescence with an excitation wavelength (λ_{ex}) 260 nm and emission wavelength (λ_{em}) 460 nm. HPLC analysis indicated that the retention time of this unknown compound was the same as that of the lumichrome standard, with an absorption maximum at a wavelength, λ= 350 nm (Figure 42). Analysis of the bound compound using mass spectrometry indicated that compound had an M+H mass of 243.085, corresponding to the lumichrome standard. Since lumichrome is a

degradation product of riboflavin, it is plausible that the source of lumichrome is riboflavin present in the cytoplasm of the *E. coli* cell. Neither RosC wildtype nor other RosC mutants with a mutation at a different residue, could be purified with lumichrome. Mutants RosC^{D166T} and RosC^{D166G} could bind lumichrome but to a much smaller degree than RosC^{D166L} (Table 32).

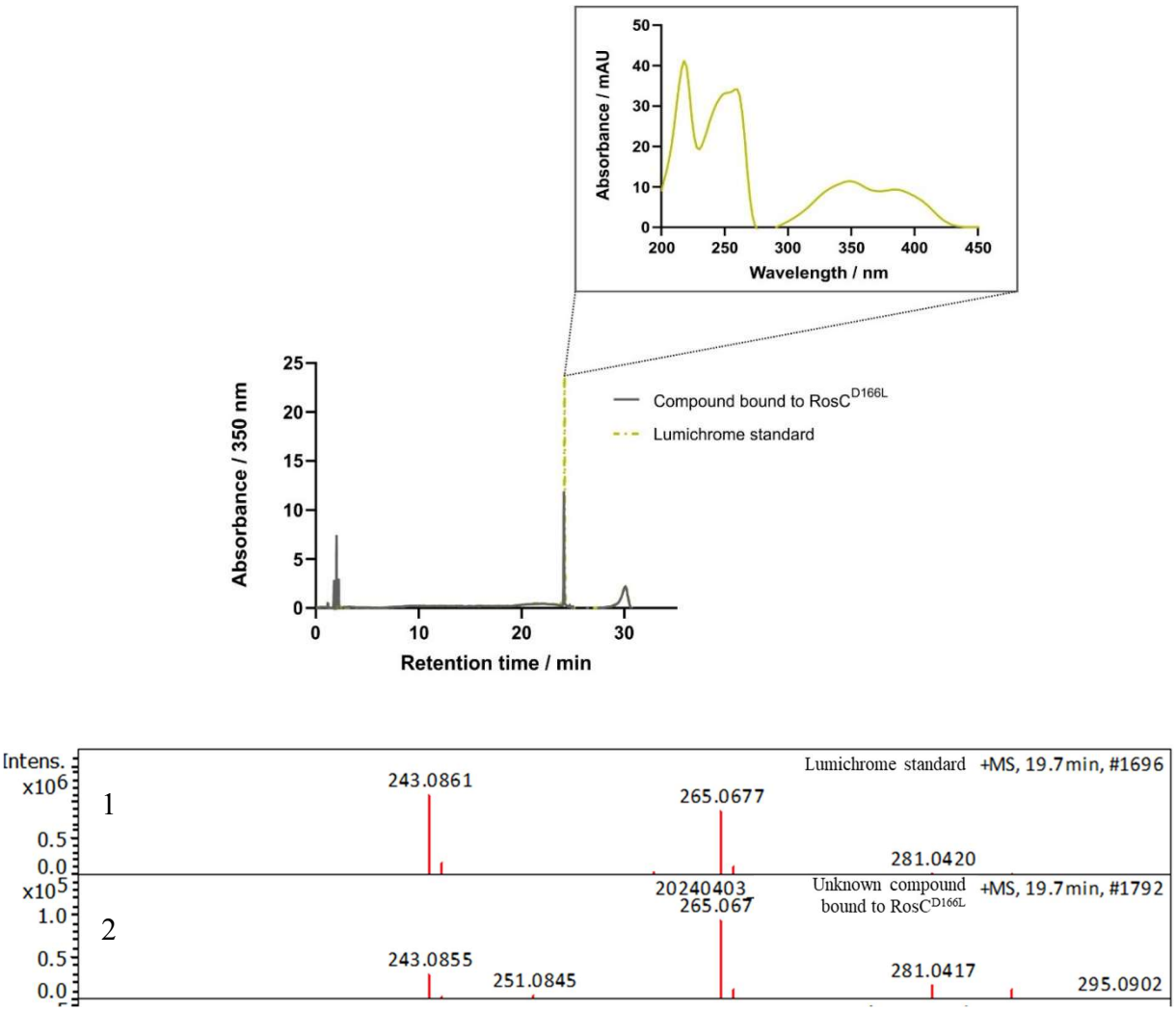


Figure 42:The compound bound to RosC^{D166L} that is responsible for its yellow colour is lumichrome
HPLC analysis of the compound bound to RosC^{D166L} was performed after TCA digestion of the purified enzyme. The compound (gray solid) has the same retention time as the lumichrome standard (yellow dashed), with absorbance maximum at λ= 350 nm. **B.** MS analysis of the unknown compound revealed a mass (M+H) of 243.085 for Lumichrome (1) as well as for the unknown compound (2). Sodium adducts of lumichrome (M+Na) are also visible as prominent peaks with masses 265.067.

Table 32: Amount of lumichrome bound to various RosC mutants carrying a mutation in the D166 residue.

Protein	Lumichrome (µM)	Protein in 10µl (µg)	Bound lumichrome (µM/µg)
D166A	0.22	6.0	0.04
D166L	3.54	11.3	0.32
D166G	1.02	9.3	0.11
D166T	1.53	13.6	0.12
D166N	-	5.8	-
RosC WT	-	19.4	-

4.12.7 Mechanism of lumichrome binding with RosC^{D166L}

D166 is an important substrate binding residue as it mediates two important interactions with AFP that are crucial for binding. The Asp166-OD2 interacts with the iminium group (mesomeric state) via a salt bridge and interacts with the C4 hydroxyl group of the ribityl chain. The first interaction is affected due lack of a salt bridge in the case of FMN/RF binding. However, the second interaction is preserved as the ribityl side chain is identical in both AFP and FMN.

It was hypothesized that D166L mutation would favour the binding of the methyl group at C8 of FMN, due a probable hydrophobic interaction. However, this mutation abolishes the hydrogen bond between Asp-OD2 and the C4-OH of the ribityl chain in the wild-type protein. Therefore, RosC^{D166L} shows very little activity towards FMN (3.6% of WT protein) and AFP (0.6 % of WT protein) as a hydrogen bond with the ribityl chain is essential for substrate binding. However, in the case of lumichrome, the ribityl chain is absent and the L166 thereby affords a better binding to the molecule due to enhanced interaction of the leucine side chain with the C8 methyl group, and this provides an explanation as to why lumichrome was bound to the purified enzyme and not riboflavin. Lumichrome supplementation did not rescue the growth in *C. glutamicum* overproducing RosC^{D166L}. It could not be confirmed whether the toxic effect of the overproduction of RosC^{D166L} was due to the sequestering of any other metabolite. The cause of toxicity remains unknown and is still being investigated.

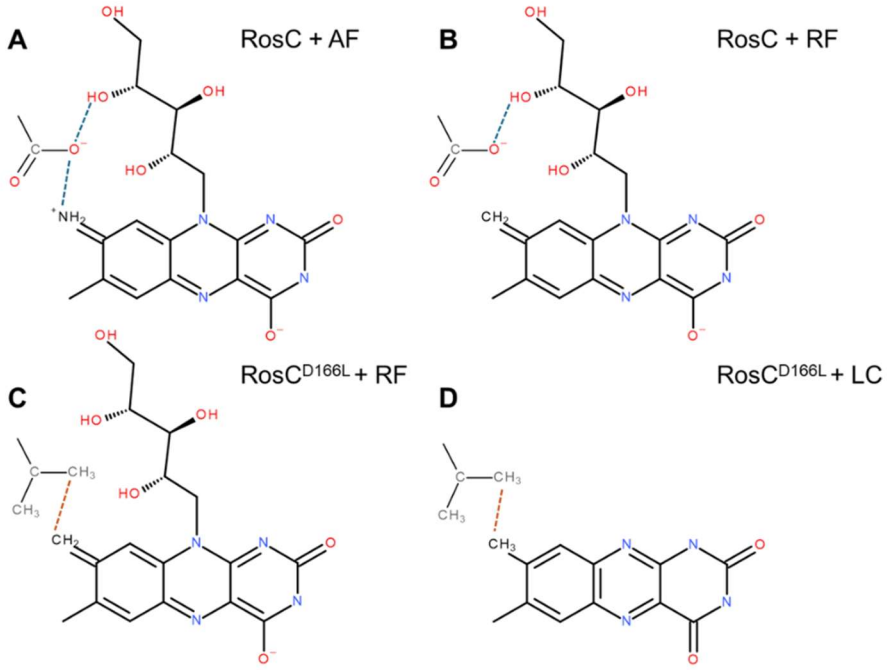


Figure 43:Interaction of RosC residues D166 and L166 with flavins
Aspartate at 166 position mediates two important interactions according to the RosC-AF structure (Figure 18A and 43A). It is the discriminating residue between FMN and AFP and strong binding to AFP is attributed to the salt bridge between the probable iminium intermediate and the side chain of D166, which is abolished when a methyl group replaces the amino group in the case of RF (Figure 43B). This is apparent in the lower K_M value of RosC for AFP as compared to FMN. Mutation D166L hampers the ability of RosC to form a hydrogen bond with ribityl chain, affecting binding of the flavins AF/AFP and RF/FMN (Figure 43C). Lumichrome lacks the ribityl chain and therefore mutation D166L affords better binding of RosC^{D166L} to lumichrome than RosC (Figure 43D).

5. DISCUSSION

In this study, the structural and functional properties of RosC were elucidated. RosC is a unique enzyme that dephosphorylates AFP, catalyzing the penultimate step of the roseoflavin metabolic pathway. The enzyme also dephosphorylates FMN and RoFMN. All three compounds show a high structural similarity (Schneider *et al*., 2020). From these studies, I could show that RosC is highly specific to the compounds in the roseoflavin biosynthesis pathway. The active site of the enzyme was analyzed using bioinformatic studies and structural studies, followed by mutational studies on the key residues to better understand their functions.

5.1 RosC possesses characteristics that distinguish it from other members of the histidine phosphatase superfamily.

RosC belongs to the histidine phosphatase superfamily and closely resembles the phosphoglycerate mutase (PGM) family of proteins (Rigden, 2008). RosC functions only as a phosphatase, and the absence of mutase activity can be attributed to the lack of specific catalytic residues that are conserved in phosphoglycerate mutases (Chiba *et al*., 2013; Johnsen & Schönheit, 2007) family. In this study, we have experimentally identified the catalytic residues and updated the proposed catalytic scheme of RosC. RosC additionally possesses characteristic properties that distinguish it from the other members of the histidine phosphatase superfamily. First is the presence of specific supplementary segments that assist in positioning and binding the flavin isoalloxazine ring in the enzyme active site. RosC has four supplementary segments or insertions that show little or no similarity to the well-characterized phosphoglycerate mutases. Insertions I (32-54), II (101-132), and III (178-192), along with the 22 amino acid N-terminal domain, contain specific residues for flavin binding. Segments I-III and the N-terminal helix of the flavin-free RosC structure have a high B-factor, indicating a state of disorder in the unbound state. The segment comprising residues 153-158 is involved in stabilizing the linker to the N-terminal helix projecting into the partner subunit. These segments and their substrate-induced rigidification comprise the novel flavin binding pocket of RosC. Although an elaborate N-terminal helix has only been found in RosC, the predicted structure of histidine phosphatase from *Chloroflexota* sp. shows a similar N-terminal extension. The enzyme has not yet been characterized, and its physiological function remains unknown. The other genes involved in roseoflavin biosynthesis are absent in the organism, ruling it out as a roseoflavin producer.

The N-terminal extension forms a lid-like structure on the adjacent monomeric unit and is referred to as the N-terminal 'cap domain'. The total length of this extension is 27 amino acids, with the first 22 folding into an alpha-helix followed by a linker comprising five amino acids. The alpha-helix serves as a lid to cover the substrate binding site of the enzyme. The removal of the N-terminal cap domain does not affect dimerization of the protein, indicating that other interactions are responsible for the dimerization process (Watkins & Baker, 2006). The N-terminal amino acids G22 and G25 are located on the linker connecting the cap domain to the enzyme core. Mutation of these residues to proline very likely affects the mobility of the cap domain during substrate-induced rigidification. This negatively affects the activity of the enzyme, as evident from the mutational studies i.e. loss of enzyme activities of $RosC^{G22P}$ and $RosC^{G25P}$. These experiments highlighted the importance of the cap domain for RosC activity. An elaborate cap domain structure is observed in many enzymes that belong to the haloacid dehalogenase superfamily (HAD), some of the members of which carry out non-specific dephosphorylation of FMN (Pandya *et al*., 2014; Sa *et al*., 2016). The primary function of the cap domain is to exclude the surrounding solvent from the active site and to provide substrate specificity to the enzymes (Park *et al*., 2015; Peisach *et al*., 2004, Seifried *et al*., 2013). C-terminal extensions of the histidine phosphatase superfamily members have been studied to protrude over the active side residues to affect the substrate specificity and facilitate solvent exclusion (Chiba *et al*., 2013; Walter *et al*., 1999). The cap domain of RosC carries out a similar role; the inner face facing the flavin is primarily hydrophobic to facilitate the substrate binding, while the outer face is primarily hydrophilic. The polar residues on the outer face of the cap domain facilitate polar interactions with the surrounding solvent, ensuring the solubility of the protein. The removal of the cap domain affects the solubility and stability of the enzyme due to the exposure of the hydrophobic core of the enzyme, thus highlighting the importance of this structure. A RosC mutant with 22 amino acids deleted from the N-terminus (Δ22RosC) was insoluble. It was only observed in the pellet fraction but was soluble as an MBP-fusion ($\Delta22RosC^{MBP}$). The resulting protein was, however, inactive, highlighting the importance of this part of the enzyme. Kinetic and mutational studies on the amino acids of the cap domain indicated that they have little impact on the catalytic activity of RosC, indicating that the role of the cap domain is to function as a facile lid for optimal positioning of the substrate.

The amino acids W181, F128, Y16, and P127 stabilise the isoalloxazine ring by stacking interactions, sandwiching the flavin (hydrophobic sandwich) (Borshchevskiy *et al*., 2015). Their mutation to an alanine residue results in an enzyme with low turnover numbers, thus highlighting their importance (Table 30). The presence of aromatic residues is common in the active site of the

flavin binding protein family of flavodoxins (Sancho, 2006). Most of the crystallized flavodoxins have been reported to possess a tryptophan residue at the *si*-face of the enzyme and a tyrosine residue at the *re*-face of the flavin (Freigang *et al*., 2002). Such an organization has also been reported in the active site of RosB, which possesses Y240 at the *re*-face of the isoalloxazine ring of FMN (Konjik *et al*., 2017). In RosC, the planar amino-isoalloxazine ring is also sandwiched between Val11, Val1 and Tyr19 at its *re*-side and Pro127, Phe128 and Trp181 at its *si*-side. A large number of hydrogen bonds are formed between the polypeptide and, in particular, polar atoms of the pyrimidine ring of AF frequently mediated by water molecules.

RosC shares a striking similarity with the phosphoglycerate mutases (dPGM) family of proteins and the highest structural similarity to phosphoserine phosphatase (PSP) (Chiba *et al*., 2013) and phosphoglycerate mutase from *E. coli* (Bond *et al*., 2001). The regions of high dissimilarity are mainly in the regions with the supplementary sections from RosC, while the catalytic core with the conserved histidine and arginine residues remains relatively conserved. Sequence and structural alignments along with kinetic studies were used to confirm the active site residues, which were H34, H165, R33, R83 and E107 (Figure 44).

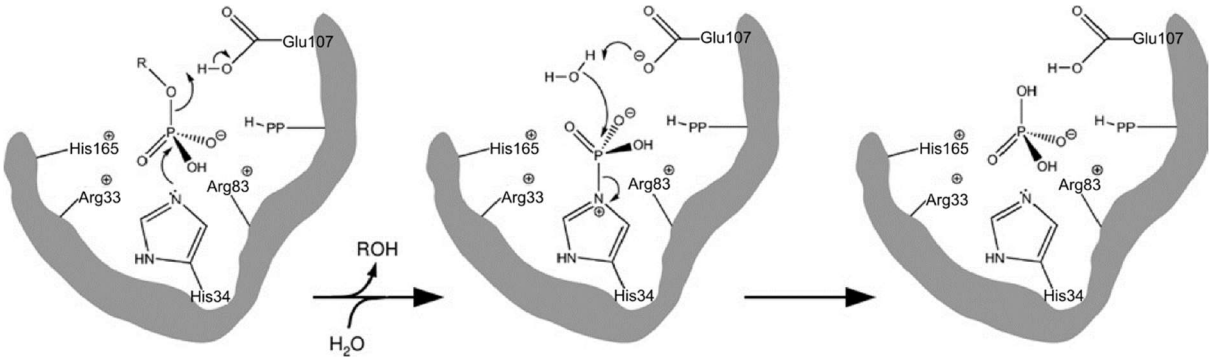


Figure 44: A new catalytic scheme from RosC adapted to the proposed catalytic mechanism for the histidine phosphatase superfamily according to the work of Rigden *et al*.

5.2 Important RosC residues

Several amino acid residues were identified in the vicinity of the amino group at C8 of AFP. Mutational studies were performed to determine their roles and to determine which residue(s) discriminate(s) between AFP and FMN. Mutation of M186 to an alanine residue reduced RosC activity, which was surprising as the carbonyl of the amino acid backbone is the participating atom (and not a side chain) that forms a hydrogen bond with the ammonium group of AFP. S167 also plays an important role in forming hydrogen bonds with the ribityl group; therefore, its mutation to a more hydrophobic residue, such as leucine, negatively impacted this interaction. V170 is also an

important residue in the vicinity of the flavin group, but its mutation to leucine and isoleucine afforded a more hydrophobic environment in the enzymatic core, but it could not be established as the main residue distinguishing between FMN and AFP. Using these studies and the information from the crystal structure, D166 was determined to be the residue discriminating between AFP and FMN.

The residue D166 is an important residue that mediates multiple interactions with the bound flavins. The residue interacts with C8 of the isoalloxazine ring and the 4' hydroxyl of the ribityl part of the flavin. The mutation of D166 to any residue that abolishes the interaction affects binding of the flavins, as in the cases of RosC^{D166A} and RosC^{D166G}. Primary sequence alignments indicated that this residue is usually an aliphatic residue like glycine or alanine in the other phosphatases with high similarity to RosC. The PSP of *Hydrogenobacter thermophilus* has a threonine residue at the position where an aspartate residue (D166) is found in RosC. The threonine in PSP essentially mediates the same dual interactions as D166 in RosC by forming a salt bridge and hydrogen bond with the substrate (Chiba *et al*., 2013). The mutant RosC^{D166T}, however, is inactive as the side chain is shorter than that of aspartic acid. The mutant RosC^{D166L}, RosC^{D166V} and RosC^{D166I} were shown to be toxic to the *E. coli* host cells. D166 mediates the substrate specificity in RosC as the aspartic acid side chain interacts favourably with the amine group at C8 of AFP but not with the methyl group of FMN. The formation of a (putative) salt bridge with the amino group may explain why AFP is a better substrate when compared to FMN. Changing D166 to more hydrophobic residues come at the expense of the loss of interactions with the ribityl side chain; therefore, we could not find a RosC mutant that preferred FMN as a substrate over AFP. A few mutations like D166E and V170L lead to a decrease in the K_M for FMN binding, indicating a higher affinity when compared to the wild-type enzyme. The yellow mutant D166E was purified from *E. coli* host cells with bound riboflavin. It has a lower K_M and higher V_{max} for both FMN and AFP (Table 30), indicating increased affinity, which can be attributed to a stronger interaction with the ribityl chain of the bound flavins than the wild-type enzyme. These experiments again show that D166 is an essential residue concerning substrate binding and can help discriminate between different substrates. The mutation V170L increases the overall hydrophobic environment in the enzyme core, thus leading to better binding of the isoalloxazine ring of both AFP and FMN, which is reflected in the kinetic parameters of the enzyme (Table 30). The R33 mutation shows a particular rearrangement, leading to better riboflavin binding. The mutant R33A is inactive on both FMN and AFP as R33 is one of the key arginine residues and constitutes important catalytic components in the dephosphorylation reaction (Schneider *et al*., 2020). E107 is the proton donor, which is in line with loss of more than 99% of

phosphatase activity in mutant RosC^{E107Q}. Therefore, the inactive RosC^{E107A} and RosC^{E107Q} mutants were the top candidates for setting up crystals in the presence of substrates AFP and FMN. Unfortunately, neither crystallisation of RosC wild-type nor of any mutant in the presence of AFP could be accomplished.

5.3 FMN hydrolases from *Streptomyces davaonensis* and their comparison with RosC

Flavin turnover inside a cell is carried out by many enzymes that work reversibly or irreversibly to maintain flavin levels in a cell. The flavokinases convert riboflavin into FMN and FAD synthetases use FMN as a substrate to synthesize FAD. The interplay between the flavokinases and the phosphatases is essential for flavin homeostasis. Two of these enzymes have been identified in *E. coli* and characterized primarily as ARPP hydrolases, whereby ARPP is an intermediate of the riboflavin pathway (Haase *et al*., 2013). However, they also dephosphorylate FMN with comparatively lower turnover numbers (Kuznetsova *et al*., 2006). Two such FMN/ARPP phosphatases belonging to the HAD family have been identified in *Streptomyces davaonensis* in this work.Gene products of BN159_0855 and BN159_4587 have been identified to catalyze the previously unknown step of the riboflavin biosynthetic pathway, as shown in Figure 2. They dephosphorylated ARPP, FMN and AFP in vitro. The mechanism of action is different from that of RosC, and the reaction occurs via the two-step phosphoaspartyl-transferase mechanism (Seifried *et al*., 2013a). The cap domain structure is present in RosC, and the FMN dephosphorylating enzymes are from the haloacid dehalogenase superfamily. The cap domain attains an open and closed conformation depending on the binding of the substrate. Although the cap domain of RosC does not show the opened and closed conformations, substrate-induced rigidification has been observed for RosC. RosC, BN159_0855 and BN159_4587 are all enzymes that have evolved to catalyse dephosphorylation of flavin phosphates and precursors thereof. However, FMN is the primary substrate of neither RosC nor FMN hydrolases from the HAD superfamily.

In *S. davaonensis*, the HAD superfamily phosphatases play virtually no role in AFP dephosphorylation and roseoflavin production, as they are not expressed in the stationary growth phase (Kissling et al 2020). However, the dephosphorylation of FMN to riboflavin has been attributed to these non-specific phosphatases in the cell. They dephosphorylate a number of compounds. Many members of the HAD superfamily have also been identified to dephosphorylate FAD. FAD pyrophosphatases act on a broad array of nucleotide diphosphate motifs, but none are completely specific to FAD (Pallotta, 2011, McLennan, 2006). FAD pyrophosphatase activities that have been identified in certain micro-organisms are attributed to the pyrophosphatase domain that

is present in the FAD synthetases wherein the C-terminus of the protein is responsible for the riboflavin kinase activity and the N-terminus responsible for FMN adenyltransferase and FAD pyrophosphatase activities (Sebastián *et al*., 2017b). These enzymes modulate their function depending on the concentration of FAD, ATP and inorganic phosphate. There are no known specific enzymes which solely dephosphorylate FAD and FMN. However, the HAD superfamily members perform non-specific dephosphorylation of these cofactors. The absence of enzymes that exclusively dephosphorylate FMN and FAD helps to keep the concentrations of these phosphorylated compounds high within a bacterial cell. RosC prefers AFP as a substrate at the physiological pH over FMN. The substrate preference of RosC, indicated by a low K_M for AFP, ensures that sufficiently high FMN levels are maintained in the cell (Schneider *et al*., 2020). A phosphatase enzyme that actively dephosphorylates FMN would not benefit the roseoflavin producers since it would generate a futile cycle since FMN is the substrate for RosB. Therefore, RosC has evolved into an enzyme that carefully discriminates between the two structurally similar substrates. As is evident from the kinetic data, the mutants RosC^{D166E} and RosC^{V170L} preferred FMN more than the wild-type enzyme. The overproduction of these proteins and FMN hydrolases in *E. coli* was not toxic to the host cells, indicating that the highly specific flavokinases in bacterial cells can significantly counteract the dephosphorylation of FMN and FAD.

5.4 Toxic effects of RosC^{D166L}

Mutational studies were performed on RosC to analyze its important residues. D166 has been determined as a key residue that mediates substrate binding and provides substrate specificity to the enzyme. RosC^{D166L} was hypothesized to show better binding to FMN than RosC as the methyl group of leucine would more favourably interact with the methyl group at the C8 position of FMN when compared to RosC wild-type. Overproduction of RosC^{D166L} leads to decreased growth of the *E. coli* host cells. The recombinant protein could not be detected by SDS PAGE and Coomassie Brilliant Blue R250 staining. Only small amounts of protein could be detected upon western blot analysis (and immunological detection), indicating that protein production is toxic to the producer. The "toxic" effect of the overproduction of the protein could also be observed in *C. glutamicum* host cells but not in *B. subtilis* host cells. Other replacements of D166 by hydrophobic amino acids led to the same effect, and RosC^{D166V} and RosC^{D166I} were also toxic to the *E. coli* host cells. The reason for the toxicity and the genetic elements in *B. subtilis* that allow it to circumvent the toxic effects are unknown. This was not observed in the case of any other RosC mutant.

RosC purified from *E. coli* C43 host cells showed that the protein was bound to a riboflavin degradation product-lumichrome. Lumichrome has the same isoalloxazine ring as riboflavin but lacks the ribityl chain. Since L166 cannot interact with the ribityl group, RosC^{D166L} does not bind riboflavin but prefers lumichrome. RosC mutants RosC^{D166T} and RosC^{D166G} also were found to bind lumichrome but not to the extent of RosC^{D166L}. The cells were grown in the presence of visible light which could be the reason for formation of lumichrome, as the latter is produced from riboflavin by photodegradation. Although it has been implicated in its roles in microbial photoelectric systems and as a signaling molecule, formation of lumichrome in liquid cultures has not been extensively studied (Li *et al*., 2023).

The phosphatase activity of RosC^{D166L} was fully abolished by introducing another mutation in the proton donor, E107, resulting in the double mutant, RosC$^{D166L+E107A}$. The double mutant was still toxic to host cells when overproduced in *E. coli.* I, therefore, conclude that the toxic effect of the RosC variants is due to the sequestering of a vital metabolite and not due to a dephosphorylating activity and thus to the degradation of an essential phosphometabolite. Untargeted metabolomic studies to find unknown metabolites are underway to identify and confirm the cause of this effect. These studies are essential as they could help identify novel target metabolites and pathways that could help discover new antimicrobial compounds.

5.5 Conclusion and Outlook

In this work I have successfully carried out the expression of *rosC* from *Streptomyces davaonensis* in various hosts like *E. coli*, *B. subtilis* and *C. glutamicum.* RosC was purified from *E. coli* cells to obtain high-purity protein for crystallization and biochemical studies. In collaboration with scientists from Max Planck Institute of Biophysics, I was able to elucidate the structural characteristics of the protein. I also generated, purified, and kinetically characterized various RosC variants to understand the roles of different amino acids in the catalytic pocket.

RosC also dephosphorylates RoFMN, which interacts with the FMN riboswitches and flavoproteins and negatively affects the bacterial cells. A more detailed investigation is required to determine if RosC contributes to the resistance of *S. davaonensis* to roseoflavin. Since the overproduction of RosC WT protein is not toxic in the *E. coli* cells, we still do not understand how mutation of a single amino acid imparts toxic properties to the variant towards certain host cells. Therefore, the underlying mechanism of toxicity of the RosC^{D166L} protein in *E. coli* and *C. glutamicum* still needs to be investigated.

REFERENCES

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Afonine, P. V, Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica. Section D, Biological Crystallography*, *68*(Pt 4), 352–367. https://doi.org/10.1107/S0907444912001308

Arakaki, A. K., Orellano, E. G., Calcaterra, N. B., Ottado, J., & Ceccarelli, E. A. (2001). Involvement of the flavin si-face tyrosine on the structure and function of ferredoxin-NADP+ reductases. *The Journal of Biological Chemistry*, *276*(48), 44419–44426. https://doi.org/10.1074/jbc.M107568200

Averianova, L. A., Balabanova, L. A., Son, O. M., Podvolotskaya, A. B., & Tekutyeva, L. A. (2020). Production of Vitamin B2 (Riboflavin) by Microorganisms: An Overview. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.570828

Bacher, A., Eberhardt, S., Eisenreich, W., Fischer, M., Herz, S., Illarionov, B., Kis, K., & Richter, G. (2001). *Biosynthesis of riboflavin* (pp. 1–49). https://doi.org/10.1016/S0083-6729(01)61001-X

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Bond, C. S., White, M. F., & Hunter, W. N. (2001). High resolution structure of the phosphohistidine-activated form of Escherichia coli cofactor-dependent phosphoglycerate mutase. *The Journal of Biological Chemistry*, *276*(5), 3247–3253. https://doi.org/10.1074/jbc.M007318200

Borshchevskiy, V., Round, E., Bertsova, Y., Polovinkin, V., Gushchin, I., Ishchenko, A., Kovalev, K., Mishin, A., Kachalova, G., Popov, A., Bogachev, A., & Gordeliy, V. (2015). Structural and functional investigation of flavin binding center of the NqrC subunit of sodium-translocating NADH:quinone oxidoreductase from Vibrio harveyi. *PloS One*, *10*(3), e0118548. https://doi.org/10.1371/journal.pone.0118548

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254. https://doi.org/10.1006/abio.1976.9999

Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., & Aravind, L. (2006). Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *Journal of Molecular Biology*, *361*(5), 1003–1034. https://doi.org/10.1016/j.jmb.2006.06.049

Chiba, Y., Horita, S., Ohtsuka, J., Arai, H., Nagata, K., Igarashi, Y., Tanokura, M., & Ishii, M. (2013). Structural units important for activity of a novel-type phosphoserine phosphatase from Hydrogenobacter thermophilus TK-6 revealed by crystal structure analysis. *The Journal of Biological Chemistry*, *288*(16), 11448–11458. https://doi.org/10.1074/jbc.M112.449561

Dakora, F. D. (2015). Lumichrome: A Bacterial Signal Molecule Influencing Plant Growth. In *Biological Nitrogen Fixation* (pp. 389–396). Wiley. https://doi.org/10.1002/9781119053095.ch38

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, *35*(Web Server issue), W375-83. https://doi.org/10.1093/nar/gkm216

Edwards, A. M. (2014). Structure and general properties of flavins. *Methods in Molecular Biology*, *1146*, 3–13. https://doi.org/10.1007/978-1-4939-0452-5_1

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 12 Pt 1), 2126–2132. https://doi.org/10.1107/S0907444904019158

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 4), 486–501. https://doi.org/10.1107/S0907444910007493

Foster, J. M., Davis, P. J., Raverdy, S., Sibley, M. H., Raleigh, E. A., Kumar, S., & Carlow, C. K. S. (2010). Evolution of bacterial phosphoglycerate mutases: non-homologous isofunctional enzymes undergoing gene losses, gains and lateral transfers. *PloS One*, *5*(10), e13576. https://doi.org/10.1371/journal.pone.0013576

Freigang, J., Diederichs, K., Schäfer, K. P., Welte, W., & Paul, R. (2002). Crystal structure of oxidized flavodoxin, an essential protein in Helicobacter pylori. *Protein Science : A Publication of the Protein Society*, *11*(2), 253–261. https://doi.org/10.1110/ps.28602

Fu, B., Ying, J., Chen, Q., Zhang, Q., Lu, J., Zhu, Z., & Yu, P. (2022). Enhancing the biosynthesis of riboflavin in the recombinant Escherichia coli BL21 strain by metabolic engineering. *Frontiers in Microbiology*, *13*, 1111790. https://doi.org/10.3389/fmicb.2022.1111790

Gutiérrez-Preciado, A., Torres, A. G., Merino, E., Bonomi, H. R., Goldbaum, F. A., & García-Angulo, V. A. (2015). Extensive Identification of Bacterial Riboflavin Transporters and Their Distribution across Bacterial Species. *PloS One*, *10*(5), e0126124. https://doi.org/10.1371/journal.pone.0126124

Haase, I., Sarge, S., Illarionov, B., Laudert, D., Hohmann, H.-P., Bacher, A., & Fischer, M. (2013). Enzymes from the haloacid dehalogenase (HAD) superfamily catalyse the elusive dephosphorylation step of riboflavin biosynthesis. *Chembiochem : A European Journal of Chemical Biology*, *14*(17), 2272–2275. https://doi.org/10.1002/cbic.201300544

Hamdane, D., Bou-Nader, C., Cornu, D., Hui-Bon-Hoa, G., & Fontecave, M. (2015). Flavin–Protein Complexes: Aromatic Stacking Assisted by a Hydrogen Bond. *Biochemistry*, *54*(28), 4354–4364. https://doi.org/10.1021/acs.biochem.5b00501

Heelis, P. F. (1982). The photophysical and photochemical properties of flavins (isoalloxazines). *Chemical Society Reviews*, *11*(1), 15. https://doi.org/10.1039/cs9821100015

Insińska-Rak, M., Prukała, D., Golczak, A., Fornal, E., & Sikorski, M. (2020). Riboflavin degradation products; combined photochemical and mass spectrometry approach. *Journal of Photochemistry and Photobiology A: Chemistry*, *403*, 112837. https://doi.org/10.1016/j.jphotochem.2020.112837

Islam, Z., & Kumar, P. (2023). Inhibitors of riboflavin biosynthetic pathway enzymes as potential antibacterial drugs. *Frontiers in Molecular Biosciences*, *10*, 1228763. https://doi.org/10.3389/fmolb.2023.1228763

Jankowitsch, F., Kühm, C., Kellner, R., Kalinowski, J., Pelzer, S., Macheroux, P., & Mack, M. (2011). A novel N,N-8-amino-8-demethyl-D-riboflavin Dimethyltransferase (RosA) catalyzing the two terminal steps of roseoflavin biosynthesis in Streptomyces davawensis. *The Journal of Biological Chemistry*, *286*(44), 38275–38285. https://doi.org/10.1074/jbc.M111.292300

Jankowitsch, F., Schwarz, J., Rückert, C., Gust, B., Szczepanowski, R., Blom, J., Pelzer, S., Kalinowski, J., & Mack, M. (2012). Genome Sequence of the Bacterium Streptomyces davawensis JCM 4913 and Heterologous Production of the Unique Antibiotic Roseoflavin. *Journal of Bacteriology*, *194*(24), 6818–6827. https://doi.org/10.1128/JB.01592-12

Johnsen, U., & Schönheit, P. (2007). Characterization of cofactor-dependent and cofactor-independent phosphoglycerate mutases from Archaea. *Extremophiles : Life under Extreme Conditions*, *11*(5), 647–657. https://doi.org/10.1007/s00792-007-0094-x

Joosten, V., & van Berkel, W. J. H. (2007). Flavoenzymes. *Current Opinion in Chemical Biology*, *11*(2), 195–202. https://doi.org/10.1016/j.cbpa.2007.01.010

Kabsch, W. (2010). XDS. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 125–132. https://doi.org/10.1107/S0907444909047337

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, *32*(Web Server issue), W526-31. https://doi.org/10.1093/nar/gkh468

Kißling, L., Schneider, C., Seibel, K., Dorjjugder, N., Busche, T., Kalinowski, J., & Mack, M. (2020). The roseoflavin producer Streptomyces davaonensis has a high catalytic capacity and specific genetic adaptations with regard to the biosynthesis of riboflavin. *Environmental Microbiology*, *22*(8), 3248–3265. https://doi.org/10.1111/1462-2920.15066

Konjik, V., Brünle, S., Demmer, U., Vanselow, A., Sandhoff, R., Ermler, U., & Mack, M. (2017). The Crystal Structure of RosB: Insights into the Reaction Mechanism of the First Member of a Family of Flavodoxin-like Enzymes. *Angewandte Chemie (International Ed. in English)*, *56*(4), 1146–1151. https://doi.org/10.1002/anie.201610292

Kortmann, M., Kuhl, V., Klaffl, S., & Bott, M. (2015). A chromosomally encoded T7 RNA polymerase-dependent gene expression system for Corynebacterium glutamicum: construction and comparative evaluation at the single-cell level. *Microbial Biotechnology*, *8*(2), 253–265. https://doi.org/10.1111/1751-7915.12236

Kuznetsova, E., Nocek, B., Brown, G., Makarova, K. S., Flick, R., Wolf, Y. I., Khusnutdinova, A., Evdokimova, E., Jin, K., Tan, K., Hanson, A. D., Hasnain, G., Zallot, R., de Crécy-Lagard, V., Babu, M., Savchenko, A., Joachimiak, A., Edwards, A. M., Koonin, E. V, & Yakunin, A. F. (2015). Functional Diversity of Haloacid Dehalogenase Superfamily Phosphatases from Saccharomyces cerevisiae: BIOCHEMICAL, STRUCTURAL, AND EVOLUTIONARY INSIGHTS. *The Journal of Biological Chemistry*, *290*(30), 18678–18698. https://doi.org/10.1074/jbc.M115.657916

Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V, Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A. V, Arrowsmith, C. H., Koonin, E. V, Edwards, A. M., & Yakunin, A. F. (2006). Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *The Journal of Biological Chemistry*, *281*(47), 36149–36161. https://doi.org/10.1074/jbc.M605449200

Lahiri, S. D., Zhang, G., Dai, J., Dunaway-Mariano, D., & Allen, K. N. (2004). Analysis of the substrate specificity loop of the HAD superfamily cap domain. *Biochemistry*, *43*(10), 2812–2820. https://doi.org/10.1021/bi0356810

Langer, S., Hashimoto, M., Hobl, B., Mathes, T., & Mack, M. (2013a). Flavoproteins Are Potential Targets for the Antibiotic Roseoflavin in Escherichia coli. *Journal of Bacteriology*, *195*(18), 4037–4045. https://doi.org/10.1128/JB.00646-13

Langer, S., Hashimoto, M., Hobl, B., Mathes, T., & Mack, M. (2013b). Flavoproteins are potential targets for the antibiotic roseoflavin in Escherichia coli. *Journal of Bacteriology*, *195*(18), 4037–4045. https://doi.org/10.1128/JB.00646-13

Lee, E. R., Blount, K. F., & Breaker, R. R. (2009). Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biology*, *6*(2), 187–194. https://doi.org/10.4161/rna.6.2.7727

Li, X., Tian, X., Yan, X., Huo, N., Wu, X., & Zhao, F. (2023a). Lumichrome from the photolytic riboflavin acts as an electron shuttle in microbial photoelectrochemical systems. *Bioelectrochemistry*, *152*, 108439. https://doi.org/10.1016/j.bioelechem.2023.108439

Li, X., Tian, X., Yan, X., Huo, N., Wu, X., & Zhao, F. (2023b). Lumichrome from the photolytic riboflavin acts as an electron shuttle in microbial photoelectrochemical systems. *Bioelectrochemistry (Amsterdam, Netherlands)*, *152*, 108439. https://doi.org/10.1016/j.bioelechem.2023.108439

Liu, C., Xia, M., Fang, H., Xu, F., Wang, S., & Zhang, D. (2024). De novo engineering riboflavin production Bacillus subtilis by overexpressing the downstream genes in the purine biosynthesis pathway. *Microbial Cell Factories*, *23*(1), 159. https://doi.org/10.1186/s12934-024-02426-w

Liu, S., Hu, W., Wang, Z., & Chen, T. (2020). Production of riboflavin and related cofactors by biotechnological processes. *Microbial Cell Factories*, *19*(1), 31. https://doi.org/10.1186/s12934-020-01302-7

Liu, Y., Zhang, Q., Qi, X., Gao, H., Wang, M., Guan, H., & Yu, B. (2023). Metabolic Engineering of Bacillus subtilis for Riboflavin Production: A Review. *Microorganisms*, *11*(1). https://doi.org/10.3390/microorganisms11010164

Liunardo, J. J., Messerli, S., Gregotsch, A., Lang, S., Schlosser, K., Rückert-Reed, C., Busche, T., Kalinowski, J., Zischka, M., Weller, P., Nouioui, I., Neumann-Schaal, M., Risdian, C., Wink, J., & Mack, M. (2024). Isolation, characterisation and description of the roseoflavin producer *Streptomyces berlinensis* sp. nov. *Environmental Microbiology Reports*, *16*(2). https://doi.org/10.1111/1758-2229.13266

Mansoorabadi, S. O., Thibodeaux, C. J., & Liu, H. (2007). The diverse roles of flavin coenzymes--nature's most versatile thespians. *The Journal of Organic Chemistry*, *72*(17), 6329–6342. https://doi.org/10.1021/jo0703092

Matiru, V. N., & Dakora, F. D. (2005). The rhizosphere signal molecule lumichrome alters seedling development in both legumes and cereals. *New Phytologist*, *166*(2), 439–444. https://doi.org/10.1111/j.1469-8137.2005.01344.x

Matsui, K., Juri, N., Kubo, Y., & Kasai, S. (1979). Formation of roseoflavin from guanine through riboflavin. *Journal of Biochemistry*, *86*(1), 167–175.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(Pt 4), 658–674. https://doi.org/10.1107/S0021889807021206

McLennan, A. G. (2006). The Nudix hydrolase superfamily. *Cellular and Molecular Life Sciences : CMLS*, *63*(2), 123–143. https://doi.org/10.1007/s00018-005-5386-7

Meng, X., Wang, W., Xie, Z., Li, P., Li, Y., Guo, Z., Lu, Y., Yang, J., Guan, K., Lu, Z., Tan, H., & Chen, Y. (2017). Neomycin biosynthesis is regulated positively by AfsA-g and NeoR in Streptomyces fradiae CGMCC 4.7387. *Science China Life Sciences*, *60*(9), 980–991. https://doi.org/10.1007/s11427-017-9120-8

NIELSEN, T., RUNG, J., & VILLADSEN, D. (2004). Fructose-2,6-bisphosphate: a traffic signal in plant metabolism. *Trends in Plant Science*, *9*(11), 556–563. https://doi.org/10.1016/j.tplants.2004.09.004

Otani, S., Kasai, S., & Matsui, K. (1980). *[34] Isolation, chemical synthesis, and properties of roseoflavin* (pp. 235–241). https://doi.org/10.1016/0076-6879(80)66464-7

Otani, S., Takatsu, M., Nakano, M., Kasai, S., & Miura, R. (1974). Letter: Roseoflavin, a new antimicrobial pigment from Streptomyces. *The Journal of Antibiotics*, *27*(1), 86–87.

O'Toole, G. A., Trzebiatowski, J. R., & Escalante-Semerena, J. C. (1994). The cobC gene of Salmonella typhimurium codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. *The Journal of Biological Chemistry*, *269*(42), 26503–26511.

Pallotta, M. L. (2011). Evidence for the presence of a FAD pyrophosphatase and a FMN phosphohydrolase in yeast mitochondria: a possible role in flavin homeostasis. *Yeast (Chichester, England)*, *28*(10), 693–705. https://doi.org/10.1002/yea.1897

Pandya, C., Farelli, J. D., Dunaway-Mariano, D., & Allen, K. N. (2014). Enzyme promiscuity: engine of evolutionary innovation. *The Journal of Biological Chemistry*, *289*(44), 30229–30236. https://doi.org/10.1074/jbc.R114.572990

Park, J., Guggisberg, A. M., Odom, A. R., & Tolia, N. H. (2015). Cap-domain closure enables diverse substrate recognition by the C2-type haloacid dehalogenase-like sugar phosphatase Plasmodium falciparum HAD1. *Acta Crystallographica. Section D, Biological Crystallography*, *71*(Pt 9), 1824–1834. https://doi.org/10.1107/S1399004715012067

Peisach, E., Selengut, J. D., Dunaway-Mariano, D., & Allen, K. N. (2004). X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily. *Biochemistry*, *43*(40), 12770–12779. https://doi.org/10.1021/bi0490688

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

Pimviriyakul, P., & Chaiyen, P. (2020). Overview of flavin-dependent enzymes. *The Enzymes*, *47*, 1–36. https://doi.org/10.1016/bs.enz.2020.06.006

Rajamani, S., Bauer, W. D., Robinson, J. B., Farrow, J. M., Pesci, E. C., Teplitski, M., Gao, M., Sayre, R. T., & Phillips, D. A. (2008). The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Molecular Plant-Microbe Interactions : MPMI*, *21*(9), 1184–1192. https://doi.org/10.1094/MPMI-21-9-1184

Rawat, R., Sandoval, F. J., Wei, Z., Winkler, R., & Roje, S. (2011). An FMN hydrolase of the haloacid dehalogenase superfamily is active in plant chloroplasts. *The Journal of Biological Chemistry*, *286*(49), 42091–42098. https://doi.org/10.1074/jbc.M111.260885

Rigden, D. J. (2008). The histidine phosphatase superfamily: Structure and function. In *Biochemical Journal* (Vol. 409, Issue 2). https://doi.org/10.1042/BJ20071097

Rigden, D. J., Littlejohn, J. E., Henderson, K., & Jedrzejas, M. J. (2003). Structures of phosphate and trivanadate complexes of Bacillus stearothermophilus phosphatase PhoE: structural and functional analysis in the cofactor-dependent phosphoglycerate mutase superfamily. *Journal of Molecular Biology*, *325*(3), 411–420. https://doi.org/10.1016/s0022-2836(02)01229-9

Sa, N., Rawat, R., Thornburg, C., Walker, K. D., & Roje, S. (2016). Identification and characterization of the missing phosphatase on the riboflavin biosynthesis pathway in Arabidopsis thaliana. *The Plant Journal : For Cell and Molecular Biology*, *88*(5), 705–716. https://doi.org/10.1111/tpj.13291

Said, H. M., Ortiz, A., Moyer, M. P., & Yanagawa, N. (2000). Riboflavin uptake by human-derived colonic epithelial NCM460 cells. *American Journal of Physiology. Cell Physiology*, *278*(2), C270-6. https://doi.org/10.1152/ajpcell.2000.278.2.C270

Sancho, J. (2006). Flavodoxins: sequence, folding, binding, function and beyond. *Cellular and Molecular Life Sciences : CMLS*, *63*(7–8), 855–864. https://doi.org/10.1007/s00018-005-5514-4

Sandoval, F. J., & Roje, S. (2005). An FMN Hydrolase Is Fused to a Riboflavin Kinase Homolog in Plants. *Journal of Biological Chemistry*, *280*(46), 38337–38345. https://doi.org/10.1074/jbc.M500350200

Sarge, S., Haase, I., Illarionov, B., Laudert, D., Hohmann, H.-P., Bacher, A., & Fischer, M. (2015). Catalysis of an Essential Step in Vitamin B2 Biosynthesis by a Consortium of Broad Spectrum Hydrolases. *Chembiochem : A European Journal of Chemical Biology*, *16*(17), 2466–2469. https://doi.org/10.1002/cbic.201500352

Schneider, C., Konjik, V., Kißling, L., & Mack, M. (2020). The novel phosphatase RosC catalyzes the last unknown step of roseoflavin biosynthesis in Streptomyces davaonensis. *Molecular Microbiology*, *114*(4), 609–625. https://doi.org/10.1111/mmi.14567

Schulte, J. E., Roggiani, M., Shi, H., Zhu, J., & Goulian, M. (2021). The phosphohistidine phosphatase SixA dephosphorylates the phosphocarrier NPr. *The Journal of Biological Chemistry*, *296*, 100090. https://doi.org/10.1074/jbc.RA120.015121

Schwarz, J., Konjik, V., Jankowitsch, F., Sandhoff, R., & Mack, M. (2016). Identification of the Key Enzyme of Roseoflavin Biosynthesis. *Angewandte Chemie International Edition*, *55*(20), 6103–6106. https://doi.org/10.1002/anie.201600581

Sebastián, M., Lira-Navarrete, E., Serrano, A., Marcuello, C., Velázquez-Campoy, A., Lostao, A., Hurtado-Guerrero, R., Medina, M., & Martínez-Júlvez, M. (2017a). The FAD synthetase from the human pathogen Streptococcus pneumoniae: a bifunctional enzyme exhibiting activity-dependent redox requirements. *Scientific Reports*, *7*(1), 7609. https://doi.org/10.1038/s41598-017-07716-5

Sebastián, M., Lira-Navarrete, E., Serrano, A., Marcuello, C., Velázquez-Campoy, A., Lostao, A., Hurtado-Guerrero, R., Medina, M., & Martínez-Júlvez, M. (2017b). The FAD synthetase from the human pathogen Streptococcus pneumoniae: a bifunctional enzyme exhibiting activity-dependent redox requirements. *Scientific Reports*, *7*(1), 7609. https://doi.org/10.1038/s41598-017-07716-5

Seifried, A., Schultz, J., & Gohla, A. (2013a). Human HAD phosphatases: structure, mechanism, and roles in health and disease. *The FEBS Journal*, *280*(2), 549–571. https://doi.org/10.1111/j.1742-4658.2012.08633.x

Seifried, A., Schultz, J., & Gohla, A. (2013b). Human HAD phosphatases: structure, mechanism, and roles in health and disease. *The FEBS Journal*, *280*(2), 549–571. https://doi.org/10.1111/j.1742-4658.2012.08633.x

Tang, J., Chen, L., Qin, Z.-H., & Sheng, R. (2021). Structure, regulation, and biological functions of TIGAR and its role in diseases. *Acta Pharmacologica Sinica*, *42*(10), 1547–1555. https://doi.org/10.1038/s41401-020-00588-y

Walter, R. A., Nairn, J., Duncan, D., Price, N. C., Kelly, S. M., Rigden, D. J., & Fothergill-Gilmore, L. A. (1999). The role of the C-terminal region in phosphoglycerate mutase. *The Biochemical Journal*, *337 (Pt 1)*(Pt 1), 89–95.

Watkins, H. A., & Baker, E. N. (2006). Structural and functional analysis of Rv3214 from Mycobacterium tuberculosis, a protein with conflicting functional annotations, leads to its characterization as a phosphatase. *Journal of Bacteriology*, *188*(10), 3589–3599. https://doi.org/10.1128/JB.188.10.3589-3599.2006

Zhou, Z., & Swenson, R. P. (1996). The Cumulative Electrostatic Effect of Aromatic Stacking Interactions and the Negative Electrostatic Environment of the Flavin Mononucleotide Binding Site Is a Major Determinant of the Reduction Potential for the Flavodoxin from *Desulfovibrio vulgaris* [Hildenborough]. *Biochemistry*, *35*(50), 15980–15988. https://doi.org/10.1021/bi962124n

Zhuang, B., Liebl, U., & Vos, M. H. (2022). Flavoprotein Photochemistry: Fundamental Processes and Photocatalytic Perspectives. *The Journal of Physical Chemistry. B*, *126*(17), 3199–3207. https://doi.org/10.1021/acs.jpcb.2c00969

LIST OF ABBREVIATIONS

AF	8-demethyl-8-amino-riboflavin
AFP	8-demethyl-8-amino-riboflavin-5'-phosphate
ARP	5-amino-6-ribitylamino-2,4(1H,3H) pyrimidinedione
ARPP	5-amino-6-ribitylamino-2,4(1*H*,3*H*) pyrimidinedione 5′-phosphate
DAB	3,3'-diaminobenzidine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide, Riboflavin-5'-phosphate
FPLC	Fast protein liquid chromatography
HAD	Haloacid dehalogenase
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
PAGE	Polyacrylamide gel electrophoresis
RF	Riboflavin
RoFAD	Roseoflavin adenine dinucleotide
RoFMN	Roseoflavin mononucleotide
SDS	Sodium dodecyl sulfate
TCA	Trichloroacetic acid
WT/wt	Wild type

TABLE OF FIGURES

TABLE OF TABLES

SUPPLEMENT

List of sequences

(Red- Protein tags, Green- linker region)

S1: Amino acid sequence of RosC from *S. davaonensis*

RosC

MSDGRESFLEVMRSVYERYLVGVPGVSEVWLIRHADSYTGLEDYDGDPRDPALSEKGRAQ
ARLLAARLAGVPLHGVWASGAHRAQQTASAVAAEHGLRVRTDARLREVRTNWDDGRPS
ELKPHGVYPFPEPEKEVAERMRTAVTAAVAATPPAPDGTTRVAVVGHDSALVILMGSLM
NLGWGQLDMILPLTSVSVLAVKDERMVVRSIGDATHLAAAPSDVI

S2: **RosC from *S. davaonensis* with N-terminal strep-tag**

RosC^{StrepN}

MWSHPQFEKIEGRMSDGRESFLEVMRSVYERYLVGVPGVSEVWLIRHADSYTGLEDYDGD
PRDPALSEKGRAQARLLAARLAGVPLHGVWASGAHRAQQTASAVAAEHGLRVRTDARLR
EVRTNWDDGRPSELKPHGVYPFPEPEKEVAERMRTAVTAAVAATPPAPDGTTRVAVVG
HDSALVILMGSLMNLGWGQLDMILPLTSVSVLAVKDERMVVRSIGDATHLAAAPSDVI

S3: **RosC from *S. davaonensis* with C-terminal his-tag**

RosC^{HisC}

MSDGRESFLEVMRSVYERYLVGVPGVSEVWLIRHADSYTGLEDYDGDPRDPALSEKGRAQ
ARLLAARLAGVPLHGVWASGAHRAQQTASAVAAEHGLRVRTDARLREVRTNWDDGRPS
ELKPHGVYPFPEPEKEVAERMRTAVTAAVAATPPAPDGTTRVAVVGHDSALVILMGSLM
NLGWGQLDMILPLTSVSVLAVKDERMVVRSIGDATHLAAAPSDVIKLAAALEHHHHHH

S4: RosC from *S. davaonensis* with the first 22 amino acids deleted, containing an N-terminal maltose binding protein (MBP) tag.

Δ22RosC^{MBP}

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDII
FWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDL
LPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDV
GVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKV
NYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGA
VALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEAL
KDAQTNSSSNNNNNNNNNLGIEGRISHMVPGVSEVWLIRHADSYTGLEDYDGDPRDPALS
EKGRAQARLLAARLAGVPLHGVWASGAHRAQQTASAVAAEHGLRVRTDARLREVRTNW
DDGRPSELKPHGVYPFPEPEKEVAERMRTAVTAAVAATPPAPDGTTRVAVVGHDSALVIL
MGSLMNLGWGQLDMILPLTSVSVLAVKDERMVVRSIGDATHLAAAPSDVI

S5: RosC from *S. davaonensis* with the first 22 amino acids deleted, containing a C-terminal hexahistidine tag.

Δ22RosC^{HisC}

MVPGVSEVWLIRHADSYTGLEDYDGDPRDPALSEKGRAQARLLAARLAGVPLHGVWASG
AHRAQQTASAVAAEHGLRVRTDARLREVRTNWDDGRPSELKPHGVYPFPEPEKEVAERM
RTAVTAAVAATPPAPDGTTRVAVVGHDSALVILMGSLMNLGWGQLDMILPLTSVSVLAVK
DERMVVRSIGDATHLAAAPSDVIKLAAALEHHHHHH

S6: RosC from *S. davaonensis*, containing an N-terminal maltose binding protein (MBP) tag.

RosC^{MBP}

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDII
FWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDL
LPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDV
GVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKV
NYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGA
VALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEAL
KDAQTNSSSNNNNNNNNNNLGIEGRISHMSDGRESFLEVMRSVYERYLVGVPGVSEVWLIR
HADSYTGLEDYDGDPRDPALSEKGRAQARLLAARLAGVPLHGVWASGAHRAQQTASAVA
AEHGLRVRTDARLREVRTNWDDGRPSELKPHGVYPFPEPEKEVAERMRTAVTAAVAATP
PAPDGTTRVAVVGHDSALVILMGSLMNLGWGQLDMILPLTSVSVLAVKDERMVVRSIGDA
THLAAAPSDVI

S7: Amino acid sequence of gene product of BN159_4587 from *S. davaonensis*

MRDNGAVTSATRQHETPAAALPPRLIATDLDGTLLRDDKSVSPRTVAALAAAEEAGIEVFF
VTGRPARWMDVVSDHVHGHGLAICGNGAAVVDLHGGPGAHRFVMVRELARANALDAVR
HLRDAAPGTVYAIEQTYGFHQEPEYPKLHMEIPDSLAPAERLLAEDAPHADEPVLKILAYH
PDLEPDAFLTLARLAIGDRANVTRSSPSALLEISGPGVSKASTLALCCAERGISHEEVVAFGD
MPNDVEMLTWAGQSYAMGNAHPDVIAAASGRTVANNEDGVAVVIEQLLAER

S8: Amino acid sequence of gene product of BN159_0855 from *S. davaonensis*

MATTSALPLVDPLDPPRLIATDLDGTLLRGNGTLSPRTRAALAAVEQAGIPVVLVTGRPSR
VVDALLKSIGPHHVIAANGAAVHTPDGTLALASPLPPAEATRLVTRVRAAVPGVSFAVEYD
RDFGHEPDYPTWSFGGDTVELVGTAEALVARTPARPLLKILAHHPTLPLDDFYEQARHAA
GPAAETTHSTGLSLVEFSAPGVTKATTLIDWSGRLGIAGHEIAAFGDMPNDLPMLTAVGRS
YAMANAHPDVLAAARHRTSSNEEDGVAEVLEGFIDAFVP

S9: Amino acid sequence of protein YbjI from *E. coli*

MSIKLIAVDMDGTFLSDQKTYNRERFMAQYQQMKAQGIRFVVASGNQYYQLISFFPEIANE
IAFVAENGGWVVSEGKDVFNGELSKDAFATVVEHLLTRPEVEIIACGKNSAYTLKKYDDA
MKTVAEMYYHRLEYVDNFDNLEDIFFKFGLNLSDELIPQVQKALHEAIGDIMVSVHTGNGSI
DLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSAVVAAA
KYRAGSNNREGVLDVIDKVLKHEAPFDQ

RosC sequences used in this study

S10: Sequence of *rosC* from *S. davaonensis* codon adapted for its expression in *E. coli*

ATGTCTGACGGTCGTGAATCTTTCCTGGAAGTTATGCGTTCTGTTTACGAACGTTACC
TGGTTGGTGTTCCGGGTGTTTCTGAAGTTTGGCTGATCCGTCACGCTGACTCTTACAC
CGGTCTGGAAGACTACGACGGTGACCCGCGTGACCCGGCTCTGTCTGAAAAAGGTCG
TGCTCAGGCTCGTCTGCTGGCTGCTCGTCTGGCTGGTGTTCCGCTGCACGGTGTTTG
GGCTTCTGGTGCTCACCGTGCTCAGCAGACCGCTTCTGCTGTTGCTGCTGAACACGGT
CTGCGTGTTCGTACCGACGCTCGTCTGCGTGAAGTTCGTACCAACTGGGACGACGGT
CGTCCGTCTGAACTGAAACCGCACGGTGTTTACCCGTTCCCGGAACCGGAAAAAGAA
GTTGCTGAACGTATGCGTACCGCTGTTACCGCTGCTGTTGCTGCTACCCCGCCGGCTC
CGGACGGTACCACCCGTGTTGCTGTTGTTGGTCACGACTCTGCTCTGGTTATCCTGAT
GGGTTCTCTGATGAACCTGGGTTGGGGTCAGCTGGACATGATCCTGCCGCTGACCTC
TGTTTCTGTTCTGGCTGTTAAAGACGAACGTATGGTTGTTCGTTCTATCGGTGACGCT
ACCCACCTGGCTGCTGCTCCGTCTGACGTTATCTAA

S11: Sequence of *rosC* from *S. davaonensis* codon adapted for its expression in *B.subtilis*

ATGAGCGACGGCAGAGAATCGTTCCTTGAAGTCATGCGAAGCGTTTATGAACGCTAT
TTAGTCGGTGTTCCTGGTGTTTCTGAGGTGTGGCTGATACGCCATGCAGATTCATAC
ACAGGGCTGGAAGATTATGATGGAGACCCAAGAGACCCCGCTCTATCAGAGAAGGGC
CGTGCTCAAGCAAGGCTCCTGGCAGCCCGCCTGGCCGGTGTGCCGCTTCACGGAGTA
TGGGCGTCCGGGGCTCACCGGGCACAGCAAACGGCCTCGGCGGTTGCGGCCGAGCAT
GGGTTGCGTGTCAGAACAGATGCCCGACTGCGTGAAGTACGCACAAACTGGGATGAC
GGAAGGCCTTCTGAATTAAAACCGCATGGCGTGTACCCATTTCCGGAGCCGGAAAAA
GAAGTGGCAGAGCGGATGAGAACGGCGGTAACCGCTGCCGTAGCGGCAACACCGCCT
GCGCCTGATGGCACGACAAGAGTCGCTGTTGTCGGCCATGATTCCGCGCTTGTCATT
TTAATGGGAAGCTTGATGAATCTCGGCTGGGGACAGCTTGATATGATTTTGCCGTTA
ACTTCTGTTTCAGTGCTTGCTGTCAAAGATGAAAGAATGGTAGTTCGGAGCATCGGC
GATGCAACTCATTTGGCAGCTGCCCCAAGTGACGTGATCTAA

S12: Sequence of *rosC* from *S. davaonensis* codon adapted for its expression in *C. glutamicum*

ATGAGTGACGGTCGAGAATCCTTTCTTGAAGTCATGCGCTCCGTCTATGAGCGCTACC
TCGTGGGTGTTCCTGGTGTGAGTGAGGTGTGGCTGATCCGCCATGCGGATTCCTACA
CCGGCTTGGAAGACTACGACGGCGATCCCCGCGATCCAGCATTATCAGAAAAGGGCC
GCGCCCAAGCACGGTTGCTTGCTGCGCGACTTGCTGGAGTTCCGCTGCACGGGGTCT
GGGCTTCGGGGGCACACCGCGCACAGCAGACTGCATCCGCGGTTGCCGCAGAGCACG
GCCTGCGTGTGCGGACGGACGCCCGTCTACGTGAAGTGCGTACCAACTGGGATGATG
GACGCCCCTCAGAGCTGAAGCCACATGGAGTGTATCCGTTCCCAGAGCCTGAGAAAG
AAGTTGCGGAACGCATGCGTACTGCTGTCACCGCGGCGGTGGCCGCCACACCACCAG
CTCCTGATGGCACCACCCGCGTCGCAGTTGTCGGTCACGACTCTGCACTGGTGATTCT
CATGGGCAGCTTGATGAATCTCGGCTGGGGCCAACTCGACATGATCCTGCCCCTCACG

TCGGTTTCTGTTCTTGCCGTAAAAGACGAAAGAATGGTGGTAAGGTCCATCGGTGAT
GCTACCCACCTGGCGGCCGCCCCGAGCGATGTCATTTAA

ACKNOWLEDGEMENTS

I would like to extend my deepest and sincerest gratitude to Professor Matthias Mack for allowing me to work in his lab. His guidance and support have been instrumental in my professional and personal growth. For his belief in my scientific capabilities and for giving me the freedom to cultivate them; I can never thank him enough.

I am extremely grateful to our collaborators from the Max Planck Institute of Biophysics, Dr Ulrich Ermler and Ms Ulrike Demmer, for their indispensable contributions to the project by providing me with the crystallographic data and scientific inputs on the project.

I am extremely grateful to all my colleagues for maintaining a positive and invigorating environment in the lab. I would like to thank Dr. Carmen Schneider for her help in the formative years of my PhD. Also, deeply grateful for my labmates - Guoxia, Johannes, Lars, Ben, Ahmed, Alina, Julia, Chantel, and Jan. Many thanks to my colleagues from the biochemistry lab as well, especially my friend Caro. Also thankful to all the students who have worked with me. It was nice to know and work with everyone.

I would also like to extend gratitude to Kerstin, Sonja, and Mischl for their help and all the technical support they have provided. I would especially like to thank Frau Christiane Koch for all her scientific and technical help, patience, and kindness.

I would also like to express my gratitude to the graduate school- HBIGS for the scientific support and learning opportunities provided by them. Many thanks to Professor Rebecca Wade and Professor Matthias Mayer for their invaluable scientific inputs on my research. I would once again like to thank Professor Prasenjit Bhaumik for the scientific background and the encouragement he has provided me.

I got a lot of inspiration from my friends Prakriti and Ashita, who have been there for me consistently throughout all these years. I have been incredibly lucky to have met Dr. Anna Hübenthal and Duc Do, who have great colleagues and friends. Anna has been a constant support and inspiration and there is no one else I would want to have shared this journey with. I am grateful to my friend, Karan who always made me feel at home, away from home.

I would like to thank Jai, for inspiring me, encouraging me, and always believing in me, even when I did not believe in myself.

Most importantly, I would like to thank my parents and my brother, for their unconditional support and love. I am grateful to my family for everything they have done for me. This work is dedicated to them.