Structural and Biochemical Characterization of putative HAD family hydrolase from *Mycobacterium tuberculosis*

THESIS

SUBMITTED TO

JAWAHARLAL NEHRU UNIVERSITY

New Delhi, INDIA

FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE FACULTY OF SCIENCE

IMTECH

LATIKA

CSIR - INSTITUTE OF MICROBIAL TECHNOLOGY
SECTOR 39-A, CHANDIGARH - 160036, INDIA

2022

सीएसआईआर -इमटैक
CSIR-IMTECH

सीएसआईआर -सूक्ष्मजीव प्रौद्योगिकी संस्थान
सैक्टर 39-ए, चण्डीगढ़-160 036 (भारत)
CSIR - INSTITUTE OF MICROBIAL TECHNOLOGY
(A CONSTITUENT ESTABLISHMENT OF CSIR)
Sector 39-A, Chandigarh-160 036 (INDIA)

CERTIFICATE

The research work embodied in this thesis entitled "**Structural and biochemical characterization of putative HAD family hydrolase from *Mycobacterium tuberculosis***" submitted by **Latika** for the degree of Doctor of Philosophy, has been carried out at CSIR-Institute of Microbial Technology, Chandigarh. This research work is original and has not been submitted in part or full for any other degree or diploma to any institute or university.

Dr. Krishan Gopal
Supervisor
Principal Scientist
CSIR-IMTECH
Chandigarh

Latika

दूरभाष : 0172-6665201 } Reception
0172-6665202 }
0172-6665..... (Direct)
फैक्स : 0091-172-2690632 (COA)
Fax : 0091-172-2690585 (Director)
0091-172-2690056 (Purchase)
एस. टी. डी. कोड़ : STD CODE : 172
Web : http://imtech.res.in

Dedicated to My Teachers and Family

Acknowledgement

At the end of this journey, it is a pleasure to express my gratitude to all those who made this thesis possible and a memorable experience for me. I would like to acknowledge everyone who generously contributed directly or indirectly to this journey in the last five years.

First and foremost, I would like to sincerely thank my supervisor, ***Dr. Krishan Gopal,*** *who guided and helped me to complete this milestone. Throughout my tenure, he always encouraged me to pursue my ideas and gave me opportunities to nurture and grow my scientific aptitude. I also thank him for teaching me this interesting field of protein crystallography. I hope for his constant guidance in the future as well.*

My journey would have been incomplete without, ***Dr. Karthikeyan Subramanian,*** *and I thank him for his help and support in efficiently finishing this piece of work. His observations and comments have always helped me to establish the overall direction of the research and to move forward with the investigation in depth. Besides being an excellent supervisor, his simplicity, softness in voice, and kind attitude towards every human being, regardless of their position or power, is always heart-touching.*

I sincerely thank Dr. ***Anil Koul****, former director of the CSIR Institute of Microbial Technology in Chandigarh, for providing excellent infrastructure, facilities, and an ambient atmosphere in which to work efficiently. I thank* ***Dr. Sanjeev Khosla, the*** *present director of our institute, for continued support.*

My sincere thanks to the previous coordinators of the pre-PhD research program, ***Dr. Karthikeyan Subramanian and Dr. Charu Sharma****, and the current coordinator,* ***Dr. Balvinder Singh,*** *for efficiently conducting the program. I also thank* ***Mr. Janki*** *and* ***Mrs Neha*** *for smoothly running all the academic affairs and activities throughout the PhD tenure.*

I also give special thanks to the funding agencies that supported my research work and fellowship. Thanks to DBT for my fellowship.

I extend my gratitude to the late ***Mrs. Sharanjeet Kaur*** *and* ***Mrs. Bhumika*** *for running the MALDI and LC-MS facilities efficiently. I also thank* ***Mr. Deepak Bhatt*** *for help with DNA sequencing.*

I also thank the staff of administration library, finance, and accounts, ESD, instrumentation division, store and purchase, the staff of bioinformatics centre and IT division for their services. I am thankful to the security guards, workers of horticulture and Mess/cafeteria for their services at all times.

My appreciation also extends to my lab members, as I really feel blessed to have such wonderful seniors who helped me in every aspect to complete my work successfully. First and foremost, I thank ***Dr. Suruchi*** *and* ***Dr. Mahesharwaram*** *for teaching me during my initial days in the lab. I would like to acknowledge* ***Dr. Amar*** *for his great help in my work, especially the cloning part. His extraordinary enthusiasm for Science and research and his limitless efforts in this field are commendable. I thank* ***Dr. Dipesh, Dr. Srajan****,* ***Avantika ma'am,*** *and* ***Dr. Satyajeet*** *for their valuable suggestions, motivation, and discussions that helped me a lot. I will always cherish the company of* ***Japleen****, who was always there in need, and together we created memories for a lifetime. I never imagined my stay here without her, and I thoroughly enjoyed working with her .Another member of the JKL group,* ***Komal,*** *whose presence I always enjoyed. She was always available at the time of need. Thank you for the company and the free rides. I extend my thanks to* ***Akriti*** *for helping me whereever and whenever required. She always listened and gave good advice regardless of work or personal issues. I would also like to thank* ***Charan*** *for cloning-related help and discussions, besides the fun and fights. I hope your hard work will pay off soon. I also cherish the company of* ***Rishita,*** *who make sure to keep us up to date on the latest fashion and trends. She had his own vocabulary and pronunciation of every word. I had a great time working with* ***Suraj and Akshay****. No matter how bad my day was, they made sure that I would not remain in that mood for long. They instantly made me smile in any situation. Although i sometimes becomes harsh, but they never complain. Thanks for tolerating me. I always acknowledge the company of* ***Ravneet****, although he is less mischievous than the others but never misses a chance to bother me. Wish him good luck for his future endeavors. I also thank* ***Rishika*** *and cherish the teatime moments spent with her. She is a great artist and a good learner. I'm hoping she can find a good job soon. I thank* ***Vineet*** *for the indents and store-related help. Aside from that, I will always remember the laughter and good times we had in the lab.I also want to thank* ***Dr. Anupama****, who always greets with a smile. She never missed an opportunity to compliment and brighten someone's day. I thank* ***Dr. Neetu*** *for her advice and help whenever required and for sharing her experiences. I also thank* ***Swarnima*** *for her company during the tea times and in the lab. I wish her good luck on her future projects. I will always cherish the wonderful moments with* ***Anshul*** *and all the gossip*

we had regarding anything and everything. I also acknowledge ***Shantanu's*** *company****.*** *He is one of the brilliant students who is well-versed in many fields of science****.*** *I hope he continues to excel in every field****.*** *I also cherished the moments spent with* ***Rithm*** *and wish the best for her. I would like to extend my thanks to* ***Sriraj,*** *who always supported and helped me in my work. He has a good aptitude and attitude towards science and research. I hope he will explore more. I also acknowledge the time spent with* ***Divya*** *and the gossiping we did in our free time. I wish well for her and hope she will do great things in her career. I want to thank* ***Surinder ji,*** *the technical officer in our lab, for providing all the facilities and organising the lab systematically.*

I thank all my batchmates for the wonderful time spent with them during the Pre-PhD. course work, batch trips, and parties in the initial years of a PhD. I especially thank ***Sonal*** *for providing me reagents and other materials whenever required. I thank* ***Shweta, Shivani, Pravin, Parth, Sharmila, Jai, Ankita, Sonal, Radhesham, Madhumita, Dev, Vidhu, Anu, Shalini and Brajender.***

I would like to thank my seniors, ***Dr. Ritesh, Madhuri mam, Yogyita mam, Sapna mam, Narender sir, Pragya mam, Saloni mam, Naincy mam, Lucky mam, Jagrity mam, Kulwinder, and my juniors, Sayil, Nandita, Anu, Sheetal, Chanchal, Khushboo, Yachna Shimona, Sonali and Sapna****.*

Coming to the most important part, or those angels who lift you up when you have forgotten how to fly, we have friends. You can fly high with all the burden on your shoulders if you have good friends and company. I'd like to thank Shweta and Shivani for being the pillars of my strength throughout these five years as we shared good times and memories. I think I had shared more frustration than happiness with you guys, so thanks for bearing with me through all those hard times. I never imagined my life here at IMTECH without ***Shweta.*** *I always went to her when anything good or bad happened. She is loving and most helpful friend I ever find in my life. She pampered me and took care of me in my difficult times.* ***Shivani*** *is my personal teddy bear, whom I want to hug in both good and bad times. She is a source of strength and motivation for me in many situations. She is a good advisor and caring friend. I'd like to thank* ***Pravin, Shiva sir, and Bharti mam*** *in particular for the delicious food made with love on weekends and for sharing all the good times. Pravin is a great cook who makes delicious food with love. He made the best fish I've ever had in my life.* ***Shiva Sir*** *and Bharti Mam are kind and welcoming enough to host a dosa party every other weekend. We had a great time at their*

house and with their parents. Both Uncle and Aunty are so cool and fed us yummy food. ***I*** *thank* ***Parth and Sharmila*** *for all the fun we had together here at IMTECH and also during their wedding in Assam. I extend my gratitude to their families for the welcome, love, warmth, and memories they have given. Besides this, we celebrated our good times at their house, and they were quite welcoming every time. I also cherish the good times I had with* ***Anunay and Manisha*** *during the north-east trip and the fun we did afterwards.* ***Anunay*** *is a genuine and helpful friend to whom I would not hesitate to turn for assistance at any time. He is also a responsible guy. I'd also like to thank* ***Radhika*** *for her company and the memorable time we spent together.*

This chapter of my life was not complete without considering my family in Ambala. Nothing could have made this happen without the love, care, and affection given by them. I thank ***Papa G, Maa, Ruchi,*** *and* ***Jimmy*** *for accepting me and loving me to the extent that it never felt like I was away from home. Thank you for always trusting me and caring about the smallest details. That makes me happy. Besides this, the person who stood like a rock behind me and motivated me to do better in my career and personal life is* ***Jimmy****. He is the one who made me love myself and encouraged me to prioritise my health. Thank you for the constant help, support, and efforts that made me believe in myself and brought back my lost confidence.*

Lastly, I want to thank my parents and family, without whom the completion of this journey would not be possible. I would like to thank my mother for having faith in me and my actions. Thank you for supporting me and my every decision and letting me do whatever I wanted in life. I want to thank my father for allowing me to explore my own world. He never ever raised even a question about my decisions regarding my choices and supported every time. Now is the time to thank my siblings. I want to thank my elder sister, ***Kalpna,*** *for all the madness we did together. She supported me and guided me in every situation. I cherish the company of my brother,* ***Saksham****, especially here in Chandigarh, where he feels at home. I want to thank all of my family members for their concern and for making each of my visits home so special.*

Finally, and most importantly, I thank the almighty God for blessings and strength during the journey of Ph.D. research.

CONTENTS

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Abbreviations

Chemicals

APS	Ammonium persulfate
ATP	Adenine triphosphate
BSA	Bovine serum albumin
BIS-TRIS	2, 2-Bis (hydroxymethyl)-2, 2′, 2″-nitrilotriethanol
CIP	Calf intestinal protease
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Di-thiothretol
EtBr	Ethedium bromide
EDTA	Ethylenediaminetetraacetic acid
ddH_2O	Double distilled water
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MOPS	3-(N-morpholino) propanesulfonic acid
Ni-NTA	Ni- Nitrilotriacetic acid
PEG	Polyethylene glycol
PIPES	Piperazine-N, N′-bis (2-ethanesulfonic acid)
PMSF	Phenylmethylsulfonyl fluoride
SDS	Sodium dodecyl sulfate
α- PGM	Alpha phosphoglucomutase
G1P	Glucose1-phosphate
G1,6BP	Glucose 1,6-bisohosphate
G6P	Glucose6-phosphate

Measurements

Å	Angstrom
cm, mm, μm	Centimetre, millimetre, micrometre
CPS	Counts per second
bp, kbp	Base pair, kilobase pair

Da, kDa	Dalton, kilodalton
℃	Degree Celsius
gm, mg, μg, ng	Gram, milligram, microgram, nanogram
hr, min, sec	Hour, minute, second
L, ml	Litre, millilitre
mA	milliampere
nmoles	nanomoles
O.D.	Optical density
V	Volt

Methods and techniques

CD	Circular dichroism
EMSA	Electrophoretic mobility shift assay
MALDI	Matrix-assisted laser desorption ionisation
MR	Molecular replacement
PCR	Polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SAD	Single wavelength anomalous dispersion

Miscellaneous

I/σI	Signal to noise ratio
IRS	Inverted repeat sequences
LB	Luria bertani
Mtb	*Mycobacterium tuberculosis*
MDR	Multidrug resistant
MRE	Mean residual ellipticity
PDB ID	Protein database identity
PISA server	Protein, interfaces, structures and assemblies server
Rmsd	Root mean square deviation
WHO	World health organization

Publications

- Verma, D. K., Chaudhary, C., **Singh, L.**, Sidhu, C., Siddhardha, B., Prasad, S. E., &Thakur, K. G. (2020). Isolation and Taxonomic Characterization of Novel Haloarchaeal Isolates from Indian Solar Saltern: A Brief Review on Distribution of Bacteriorhodopsins and V-Type ATPases in Haloarchaea. *Frontiers in microbiology*, *11*.554927
- Deep A[#]., Kaur J[#]., **Singh, L**., Gosain T, Singh R, Thakur, K. G. Structural and mechanistic basis of GTP-Rv1045 toxin interactions required for its putative guanylytransferases activity *Mycobacterium tuberculosis.* (Manuscript in communication).
- Deep A[#]., Kaur J[#]., **Singh, L**[#]., Thakur, K. G. Molecular basis of *Mycobacterium tuberculosis* DarT toxin neutralization by DarG antitoxin. (Manuscript in communication.
- **Singh, L**., Karthikeyan, S., Thakur, K. G. Structural and functional studies on *Mycobacterium tuberculosis* putative phosphatase. (Manuscript in preparation).

[#]Equal contribution

Conferences attended

- Participated and volunteered in CCP4 crystallography school and workshop: From data processing to structure refinement and beyond, held at CSIR-IMTECH in October 2018.

- Participated in International Conference on Structure assisted development of Novel Therapeutics, held at RCB, Faridabad in Febraury, 2019.

- Participated in Indo-Italian Elettra Beamline Workshop: Application in Crystallography and Drug Discovery,held at AIIMS, New Delhi in November 2019.

- Participated and poster presentation at 1st Bilateral Meeting of INST Mohali & CSIR-IMTECH held at Mohali in July, 2022

Preface

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB) which is a highly contagious airborne illness spread by aerosols from an infected person to a healthy person. The current therapy to cure the TB infection includes multiple antibiotics that selectively target the essential life processes of the bacteria. However, the ineffectiveness of current therapy regimens against drug resistance strains (MDR and XDR) results in global rise in deaths and infections that are difficult to treat. Therefore, new antimicrobials are urgently required to treat the infections that are caused by drug resistant strains. Among the known approaches for designing new drugs, targeting the enzymes involved in cofactor/coenzyme biosynthesis pathway offers several benefits for the development of new drugs. In most of the bacteria, the biosynthesis of cofactor/coenzyme are endogenous and they are important for their growth. However, the biosynthesis of cofactor/coenzyme are generally absent in humans as they acquire them from the environment through specific transporters. Therefore, it is anticipated that the drug molecules designed against cofactor/coenzyme biosynthesis pathway will be more specific and less toxic to the host. One pathway that is considered to be essential for Mtb but missing in the human is the riboflavin (vitamin B2) biosynthesis pathway. The enzymes involved in riboflavin biosynthesis pathway can be potential drug targets in Mtb. The riboflavin is a sole precursor molecule for the synthesis of cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). These cofactors play a vital role in essential life processes and therefore important for humans, plants, and bacteria, including Mtb. In Mtb, the biosynthesis of riboflavin is catalyzed by five enzymes. Out of these five enzymes, four enzymes are well studied including their biochemical and structural characterization. However, one enzyme, which supposed to do a dephosphorylation of an intermediate in the riboflavin synthesis, is not known yet in Mtb.

In this thesis, we focused on identifying the unknown phosphatase of the riboflavin biosynthesis pathway in Mtb. In the literature, different genes are reported in multiple organisms that encode for the phosphatase involved in the riboflavin biosynthesis pathway. Using bioinformatic approach, we predicted three homologs of the reported phosphatases in Mtb. To validate our prediction and to check and their role in riboflavin biosynthesis pathway, we characterized these potential phosphatases biochemically and structurally. Thus, this thesis consists of five chapters as described below:

- The first chapter provides a brief introduction of background, the research gaps and objectives of the current study.
- The second chapter gives the relevant review of the literature about the work already reported in this field.
- The third chapter provides the detailed information about the material and methods employed in the study.
- The fourth chapter describes the structural and biochemical characterization of Rv3400 from Mtb.
- The fifth chapter involves the characterization of Rv3376 and Rv3813c from Mtb.

Chapter 1
Introduction

1.1. Introduction

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB), which is chronic, highly contagious, and often associated with an asymptomatic latent period following initial infection. TB is one of the world's biggest causes of death and the second leading infectious killer after COVID-19 (statistics from the WHO's 2020 report). Despite the fact that the World Health Organization proclaimed tuberculosis a worldwide health emergency in 1993, the scientific and pharmaceutical communities are still working to find a fast and effective cure. Tuberculosis therapy is a protracted, ongoing process that lasts six to nine months and involves the patient receiving various medications. The current therapy to cure the TB infection consists of multiple antibiotics that selectively target the essential life processes of the bacteria, which includes replication, transcription, translation, etc. Prolonged treatments and poor patient compliance significantly contributed to the emergence of drug resistance strains such as MDR and XDR. Therefore, new therapies are urgently required to combat the drug resistant strains. Such therapies must be inexpensive, accessible, and effective against the resistant strains.

Among the known approaches for designing new drugs, targeting the enzymes involved in cofactor/coenzyme biosynthesis pathway offers several benefits for the development of new drugs. In most cases, the cofactor/coenzyme biosynthesis for bacteria are endogenous and they are important for their growth. However, the biosynthesis of cofactor/coenzyme are generally absent in humans as they acquire them from the environment through specific transporters. Therefore, it is anticipated that the drug molecules designed against cofactor/coenzyme biosynthesis pathway will be more specific and less toxic to the host. One such pathway is the riboflavin (vitamin B2) biosynthesis pathway which is missing in humans but found to be essential for Mtb. Therefore, the enzymes of the riboflavin biosynthesis pathway are contemplated as potential drug targets in Mtb.

Riboflavin, also known as vitamin B2, is an essential metabolite for all life forms and is the sole precursor of flavin mononucleotide (FMN; also known as riboflavin-5'-phosphate) and flavin adenine dinucleotide (FAD). These cofactors play major roles in electron transport, photosynthesis, and fatty acid oxidation in almost all organisms. In the cell, riboflavin is converted to FMN by phosphorylation and FMN is further converted to FAD by adenylation. The complete enzymatic pathway for *de novo* production of riboflavin is available only in plants, fungi, and in certain eubacteria (Bacher, Eberhardt, Fischer, Kis, & Richter, 2000;

Markus Fischer et al., 2004). Mtb and many other deadly pathogens are entirely dependent on the endogenous production of riboflavin for their survival as they lack riboflavin transporters (Dahl, Sylte, & Ravna, 2004). Humans cannot produce or store this vitamin, so they have to obtain it through their diet or through multivitamin supplements. The riboflavin biosynthesis pathway is an ideal target for antimicrobial drugs due to the necessity of enzymes involved for the production of riboflavin in bacteria and their absence in humans (Dahl et al., 2004).

In bacteria, the riboflavin biosynthesis pathway (RBS) consists of five enzymes. In *Escherichia coli* and in several gram-negative bacteria, the genes encoding these enzymes are scattered across the genome, while in Mtb and *Bacillus subtilis* they group together in an operon (Long, Ji, Wang, & Xie, 2010; Vitreschak, Rodionov, Mironov, & Gelfand, 2002). The pathway starts with two branches involving Guanosine-5'-triphosphate (GTP) and ribulose-5-phosphate (Ru5P) as substrates and end up with the production of riboflavin by the action of six enzymes.

One of the intermediates in riboflavin biosynthesis, 5-amino-6-ribitylamino-2,4(1H,3H) pyrimidinedione 5′-phosphate (ARPP), undergoes dephosphorylation to form 5-amino-6-ribitylamino-2,4(1H,3H) pyrimidinedione (ARP). However, the phosphatase that catalyzes the dephosphorylation step in riboflavin biosynthesis pathway remains elusive in Mtb. To understand the riboflavin biosynthesis pathway completely in Mtb and to explore it as a potential drug target we need to address the following questions:

- Which gene is responsible for catalysing the dephosphorylation step of 5-amino-6-ribitylamino-2,4(1H,3H) Pyrimidinedione 5'-phosphate (ARPP) in Mtb?
- What is the mechanism of catalysing this dephosphorylation step?
- Whether the gene encodes for the ARPP phosphatase is part of the rib operon?

Thus, based on the above research gaps we framed following objectives for this thesis.

- Identification of phosphatase(s) potentially involved in riboflavin biosynthesis pathway using bioinformatics analysis.
- Cloning, expression and purification of the putative phosphatase(s), suitable for functional and structural characterization.
- Biochemical and structural characterization of the putative phosphatase(s).
- Investigating the role of identified phosphatases in Riboflavin biosynthesis pathway, if any.

Chapter 2
Review of Literature

2.1 Mycobacterium tuberculosis

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB) which is a highly contagious airborne illness spread by aerosols from an infected person to a healthy person. In 1882, the German scientist Robert Koch was the first microbiologist to report the successful isolation of tuberculosis causative agent, which was termed as *Mycobacterium tuberculosis* a year later. Tuberculosis is the world's one of the biggest causes of death and besides this, the most infectious killer after COVID-19 (statistics from the WHO's 2020 report). In the year 2020, TB killed 1.5 million people out of 10 million people who contracted the disease. (WHO's 2020 report).

2.1.1. Morphology and microbiology

Mtb is a non-sporing, non-motile, and uncapsulated slender rod-shaped bacteria that are either straight or slightly curved and appears either in single, in pairs or in small clumps **(Fig. 2.1).** Within alveolar macrophages, mycobacterial shape ranged from shorter ovals (0.5 to 1 µm in length), to traditional rods (2-4 µm in length), and long filamentous (more than 6-7 µm in length), although width of Mtb remained constant (3µm) (Ufimtseva, Eremeeva, Vakhrusheva, & Skornyakov, 2019). Mtb is neither categorized as Gram-positive or Gram-negative since it lacks the chemical features of either, despite the fact that the bacteria's cell wall contains peptidoglycans (murein). Mtb is an obligate aerobe, as a result, Mtb complexes are invariably identified in the well-aerated upper lobes of the lungs that are infected by tuberculosis. The bacterium is a facultative intracellular parasite that most commonly infects macrophages. It has a long generation rate of 15-20 hours, which may contribute to its virulence. Mtb's cell wall structure and composition deserve special attention because they are unusual among prokaryotes and play a significant role in the bacterium pathogenicity. The cell wall of Mtb is a complex structure composed of peptidoglycans and lipids. Lipids make up more than 60% of the mycobacterial cell wall. Mycolic acids, cord factor, and wax-D are the three primary components of the lipid part of Mtb's cell wall (Alderwick, Harrison, Lloyd, & Birch, 2015).

2.1.2. Infection and transmission

Tuberculosis is transmitted through the air by aerosols or the droplets containing bacteria released from an infected person to a susceptible person (Fennelly et al., 2004). When the people, with pulmonary or laryngeal tuberculosis, cough, sneeze, laugh, or shout, they emit

infectious droplet nuclei, which are tiny water droplets containing bacteria. These little droplet nuclei can float in the air for several hours.

Fig. 2.1 Mycobacterium tuberculosis scanning electron micrograph.

Source- http://textbookofbacteriology.net/tuberculosis.html

Lungs are the primary site of Mtb infection causing the pulmonary tuberculosis (PTB), from where it can further disseminate to other organs like lymph node, gastrointestinal and peritoneal system, central nervous system, urinary, kidney, genital, skin and soft tissue, leading to the extrapulmonary tuberculosis (EPTB) (J. Y. Lee, 2015). The Mtb infection leads to either tuberculosis disease or latent TB infection (LTBI) where, it resides in the dormant state for decades (J. Y. Lee, 2015) (Elder, 1992). The latent Mtb gets activated when the host immune system is compromised due to malnutrition, disease conditions like AIDS or aging, resulting in the tuberculosis disease, which if not treated can be fatal. Mtb is linked to a variety of other human diseases, including pulmonary problems, autoimmune diseases, and metabolic syndromes (Chai, Zhang, & Liu, 2018). Additionally, Mtb infection can affect the human microbiome and endocrine system which is closely related to immune system balance and good health (Vinnard & Blumberg, 2017).

2.1.3. Epidemiology

Mtb is thought to infect about a quarter of the world's population but only 5-15 % developed active tuberculosis while the rest don't fall ill or spread the disease. TB is a centuries old problem and yet quite challenging to overcome (W.H.O, Global tuberculosis report 2021).

Despite being a preventable and treatable illness, tuberculosis claims 1.5 million lives annually, making it one of the biggest infectious killers. According to a WHO report from 2020, India, the world's second populated country, contributed to 25,90,000 people being ill (one in every 12 seconds) and 5,04,000 people dying from tuberculosis (one in every minute), a 7.3 % rise from the year 2015 **(Fig. 2.2).**

Globally, COVID-19 pandemic caused delays in the delivery of medicines and access to TB diagnostic and treatment services, leading to an increase of roughly 1,00,000 fatalities between 2019 and 2020 (a rise in deaths of HIV-negative persons from 1.2 million to 1.3 million, with 5000 more TB deaths among HIV-positive people). A slowdown in the annual fall in the global TB incidence rate has also been brought by the disturbances caused by COVID-19.

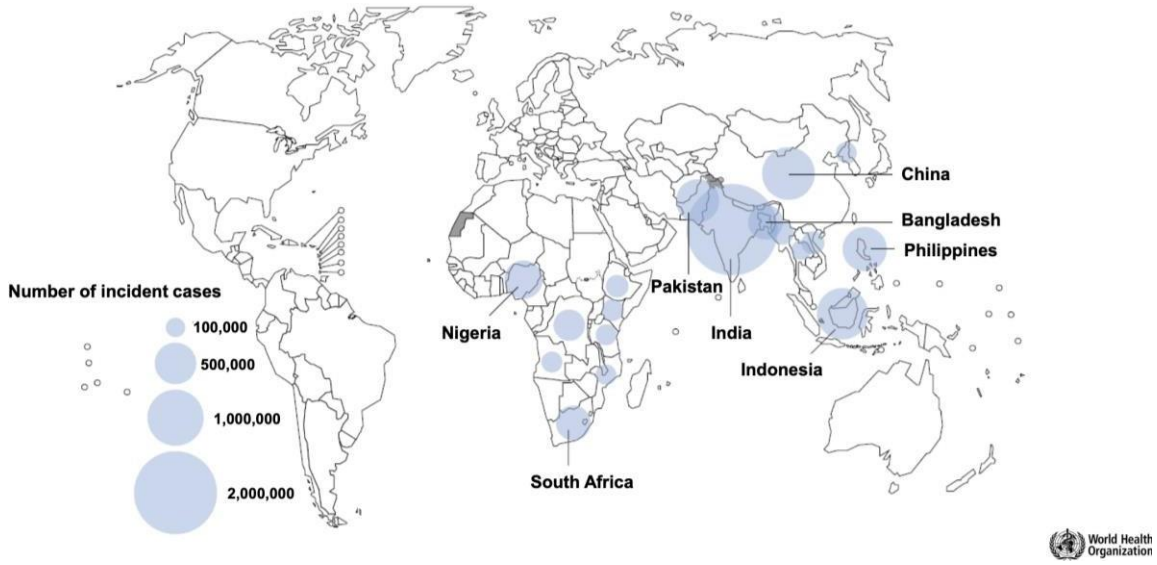


Fig. 2.2 Estimated number of TB incident cases in 2020. The figure is produced from 2020 WHO report.

Source-WHO report 2020 (https://www.who.int/teams/global-tuberculosis).

2.2. Antibiotic resistance and need for new drug molecules for Mtb

Tuberculosis therapy is a protracted, ongoing process that lasts six to nine months and involves the patient receiving various medications. Prolonged treatments, drug resistance strains, HIV pandemic, poor patient compliance, lack of potent vaccine are the challenges to control or eradicate TB. The drug targets against tuberculosis involves inhibition of cell wall components, synthesis of proteins, or nucleic acids, and also targeting the other components of Mtb biology. The mechanisms of action for some medications are yet to be fully identified.

The regimen for the treatment of TB depends on the duration, dose, frequency and the types of anti-TB drugs prescribed (Suarez et al., 2019). The current WHO-recommended TB control strategy, Directly Observed Treatment Short-course (DOTS), is a two-phase oral medication administration treatment. Antibiotics can be used to treat both the active TB infection and the latent TB. The drugs Isoniazid (INH), rifampicin (Rif), ethambutol (EMB), and pyrazinamide (PZA), taken over 6 months, are the current drug-regimen for treating drug-sensitive tuberculosis (Zumla, Nahid, & Cole, 2013). If this treatment fails, a second-line of drugs such as para-aminosalicylate (PAS) and fluoroquinolones are employed, which are probably less efficient or more harmful with substantial side effects. Although the second-line of drugs has a high success rate, it has been hampered by challenges with compliance, leading to the growth of multidrug resistant (MDR), extensively drug resistant (XDR), and totally drug resistant (TDR) strains of the Mtb (Falzon et al., 2015; Udwadia, 2012) **(Fig. 2.3).**

Despite the fact that the World Health Organization proclaimed tuberculosis a worldwide health emergency in 1993, the scientific and pharmaceutical communities are still working to find a fast and effective cure. Exploring new therapeutic targets and developing new treatments will help to combat drug-resistant Mtb. The new medicines must be inexpensive, accessible, and effective against MDR and XDR Mtb strains.

2.2.1. Essential pathway as drug targets

Target-based drug development is universally acknowledged as a successful strategy (Lindsay, 2003) (Knowles & Gromo, 2003). Finding the genes that encode for essential proteins is a preliminary step in determining the drug targets. Like any other organism, Mtb needs a variety of biomolecules to survive. These biomolecules are either synthesized endogenously or taken from the external environment through transporters. Therefore, blocking the proteins that are responsible for the production of these biomolecules or disabling their transport is a powerful means to kill or arrest the growth of the Mtb.

The strategy to identify the drug target(s) in the pathogen must consider that the same targets should be absent or non-essential for the host. The likelihood of host toxicity is decreased by choosing the targets that are exclusive to the pathogen. In order to treat mycobacterial infections, it is therefore important to synthesize the drugs that specifically target the pathways that will be non-essential or absent in the host while being selectively hazardous to the pathogen. In a study by *Sassetti et.al*, a high-throughput method based on a

whole-genome microarray was used to identify essential genes of Mtb by random mutagenesis (C. M. Sassetti, Boyd, & Rubin, 2003).

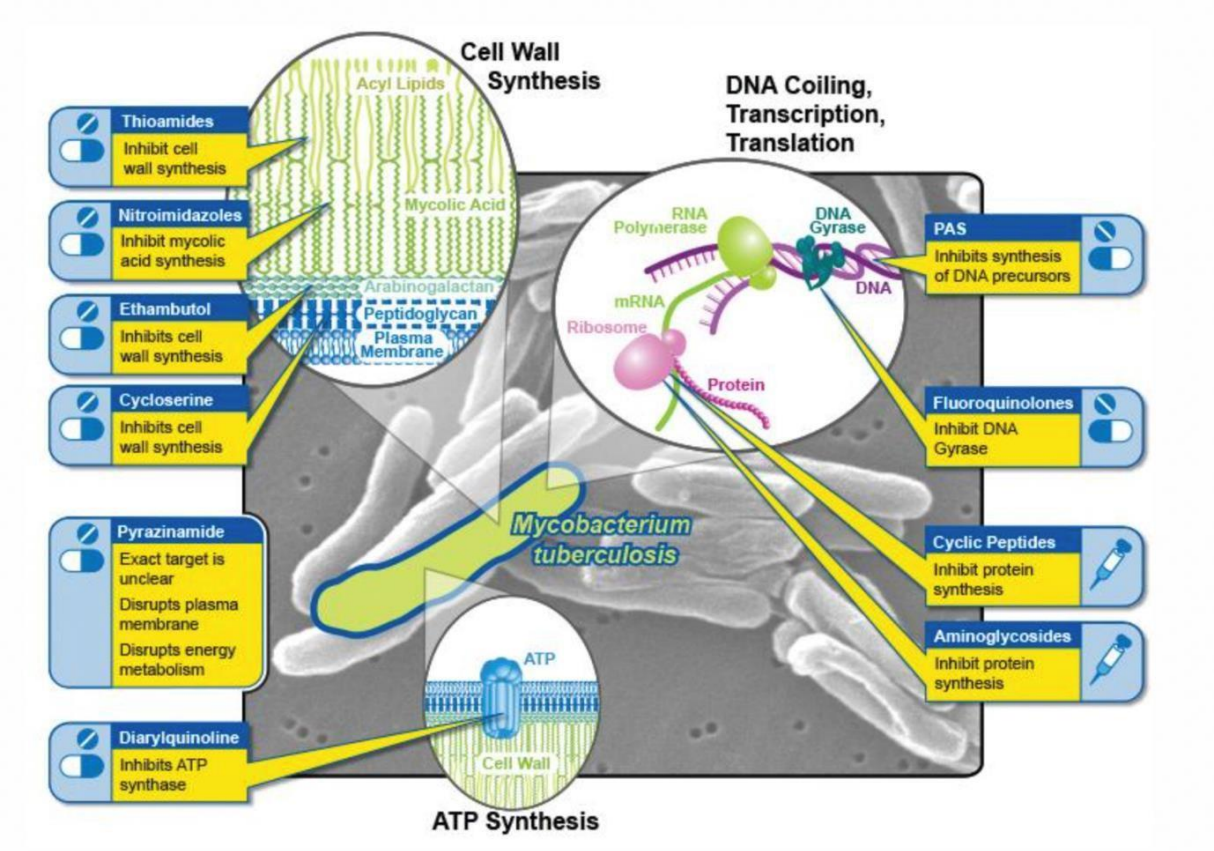


Fig. 2.3 Pharmaceutical TB treatments and their mode of action

Source- https://www.niaid.nih.gov/diseases-conditions/tbdrugs

For instance, many bacteria must produce folic acid, an essential biomolecule, endogenously. Dihydropteroate synthase, an enzyme in the folic acid biosynthesis pathway, converts p-aminobenzoate into dihydrofolate (Lascelles & Woods, 1952). Drugs with sulfonamide bases are structural analogues of p-aminobenzoate and work by preventing dihydropteroate synthase from being produced (Achari et al., 1997). Sulfonamides imitate a necessary substrate and competitively inhibit an important enzyme, making them a useful treatment for a variety of bacterial illnesses. Another important metabolite is D-alanyl-D-alanine which is an important component of bacterial peptidoglycan layer. The β -lactam class of antibiotics is popular class of medication that imitates D-alanyl-D-alanine (Tipper & Strominger, 1965) and disrupting the bacterial cell wall structure which is crucial for the survival of bacteria (Denome, Elf, Henderson, Nelson, & Young, 1999).

2.3. Riboflavin biosynthesis pathway

Vitamin B2, also called as riboflavin (RF), is an important metabolite for almost every organism. The basic structure of riboflavin comprises an isoalloxazine ring with a ribityl side chain at the N10 position (Bartmann et al., 2019). Riboflavin is necessary for protein folding, oxidation-reduction reactions (Tu, Ho-Schleyer, Travers, & Weissman, 2000), and healthy immunological response (Schramm et al., 2014) (Bian et al., 2019; Mazur-Bialy, Buchala, & Plytycz, 2013; B. Zhang et al., 2021). Additionally, it has anti-inflammatory and antioxidant effects (D. Liu & Zempleni, 2014; Mazur-Bialy et al., 2013). One of the most important aspects of riboflavin is that it is the sole precursor of flavin mononucleotide (FMN; also known as riboflavin-5'-phosphate) and flavin adenine dinucleotide (FAD) which are two important coenzymes. These coenzymes play major roles in electron transport, photosynthesis, fatty acid oxidation in almost all organisms. In a cell, riboflavin is modified to FMN by phosphorylation and further converted to FAD by adenylation **(Fig. 2.4).**

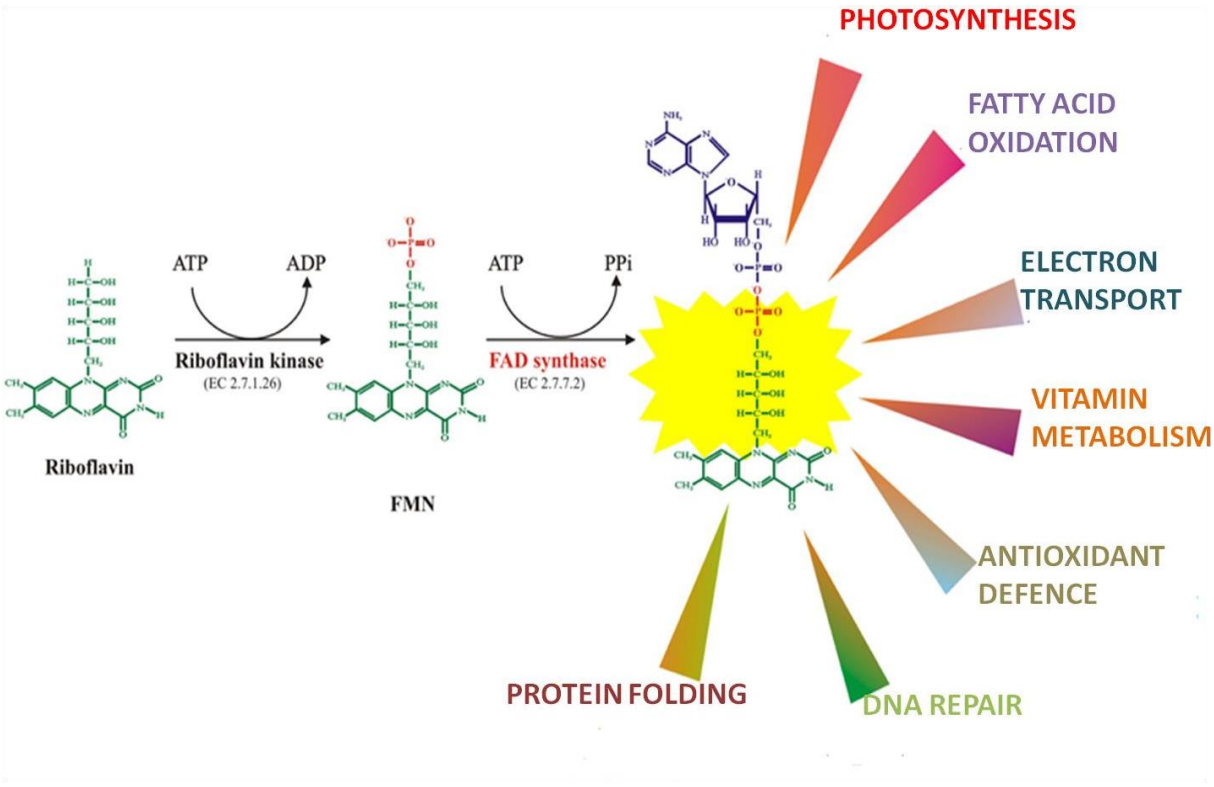


Fig. 2.4. FMN (Flavin mononucleotide) and FAD (Flavin adenine dinucleotide) synthesis from riboflavin and main biological functions in the cell.

Source- https://www.frontiersin.org/articles/10.3389/fchem.2015.00030/full

The whole enzymatic machinery for *de novo* riboflavin production is only encoded by plants, fungi, and certain eubacteria (Bacher et al., 2000). Mtb and many other deadly pathogens are entirely dependent on endogenous riboflavin production as they lack riboflavin transporters (Dahl et al., 2004). Humans cannot produce or store this vitamin and therefore they have to obtain through their diet or by multivitamin supplements. The lack of riboflavin biosynthesis enzymes in humans and the necessity of them in bacteria makes this pathway an excellent target for the designing drugs with minimum host toxicity. (Gerdes et al., 2002) (Long et al., 2010).

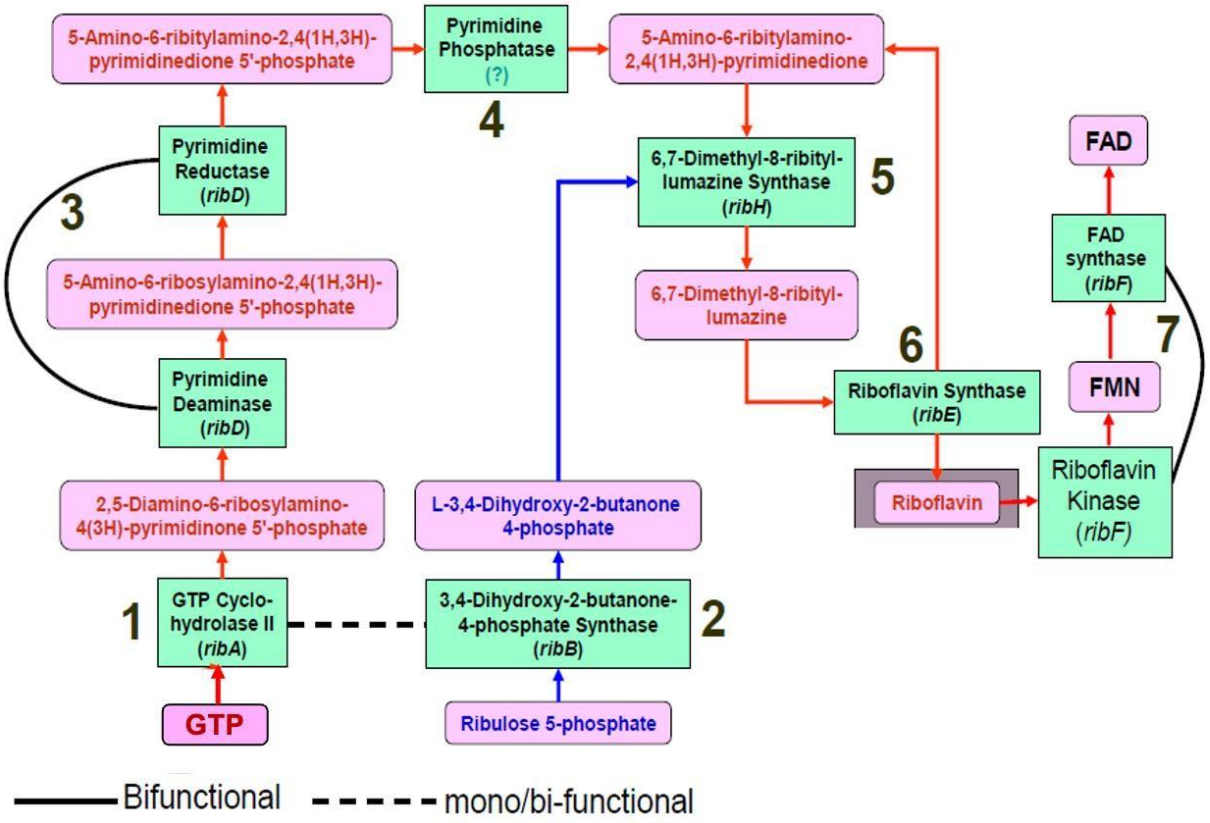


Fig. 2.5 Riboflavin biosynthesis pathway in bacteria. The enzymes involved in the pathway are displayed in green boxes and their respective substrates are shown in pink boxes. Solid line indicates that the enzyme is bifunctional whereas the dotted lines indicate that the enzyme is monofunctional in some organism and bifunctional in others.

In bacteria, the riboflavin biosynthesis pathway (RBS) consists of five enzymes. The RBS starts with two branches involving Guanosine-5'-triphosphate (GTP) as a substrate for one branch and ribulose-5-phosphate (Ru5P) as a substrate for another branch. GTP is derived from the purine biosynthesis pathway and Ru5P from the pentose phosphate pathway. The enzymes involved in the RBS pathway can be part of the same operon and clustered together

or they may be scattered in the genome. In *Escherichia coli* and in several gram-negative bacteria, the genes encoding the enzymes of riboflavin pathway are dispersed across the genome while in Mtb and *Bacillus subtilis* they group together in an operon (Long et al., 2010; Vitreschak et al., 2002). Some enzymes of the RBS pathway encode for two activity and possesses a bifunctional role, while some of them may be monofunctional in one organism and bifunctional in others **(Fig. 2.5).**

The enzymes involved in riboflavin biosynthesis pathway are briefly described as follows:

1). GTP Cyclohydrolase II (GCHII/ribA)

GCHII is the initial enzyme of the RBS pathway that converts the substrate GTP into 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone-5'-phosphate (DARP) with a release of formate (HCOO-) and pyrophosphate (PPi) (**Fig. 2.6)**. In some organisms, GCHII is encoded by a single monofunctional gene ribA (Herz, Eberhardt, & Bacher, 2000), while in some gram-positive bacteria and plants this enzyme exists as bifunctional which is fused with 3,4-Dihrdroxy-2-butanone4-phosphate synthase (DHBPS) (*rib)* and termed as *rib*BA (Singh et al., 2013).

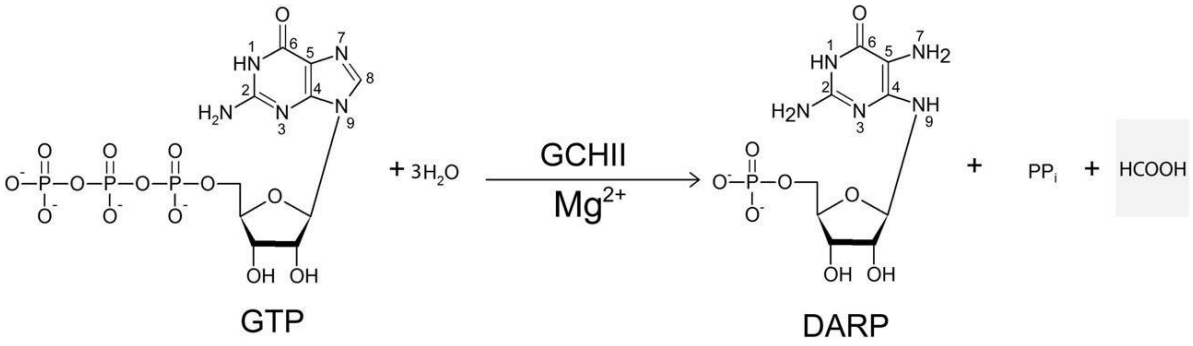


Fig. 2.6 Reaction catalyzed by GCHII. The reaction involves the conversion of GTP into 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone-5'-phosphate (DARP) with a release of formate (HCOO-) and pyrophosphate (PPi).

2). 3,4-Dihrdroxy-2-butanone4-phosphate synthase (DHBPS/ribB)

The DHBPS is another initiating enzyme of the pathway that utilize ribulose-5-phosphate (Ru5P) as substrate and converts into L-3,4-Dihrdroxy-2-butanone4-phosphate (DHBP) and formate (Abbas & Sibirny, 2011) (**Fig. 2.7).** Depending on the organisms, DHBPS and GCHII may present as separate polypeptide or fuse together to form bifunctional enzyme having independent functions (Singh et al., 2013).

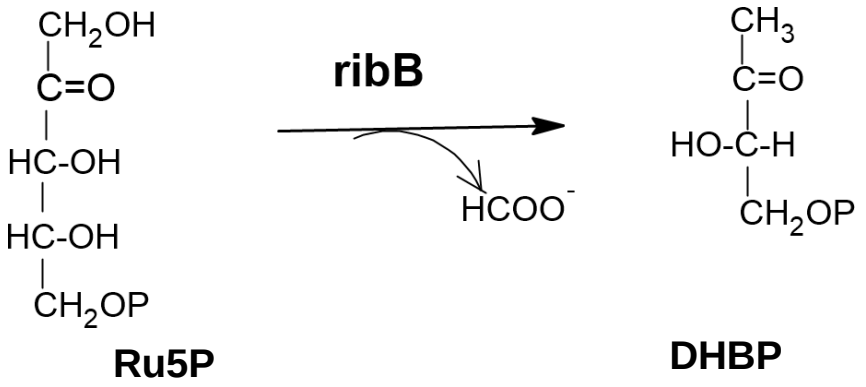


Fig. 2.7 Reaction Catalyzed by DHBPS. The reaction involves the conversion of Ru5P into L-3,4-Dihrdroxy-2-butanone4-phosphate (DHBP).

3). Pyrimidine Deaminase/Reductase (ribD)

Pyrimidine deaminase/reductase is a bifunctional enzyme encoded by single gene *ribD.* The deamination domain is located at the C-terminal and the reductase at the N-terminal that catalyze the deamination and reduction of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone-5'-phosphate (DARP) into 5-Amino-6-ribosylamino-2,4(1H,3H)-pyrimidinone-5'-phosphate (ARPP) **(Fig. 2.8)** (Markus Fischer & Bacher, 2006).

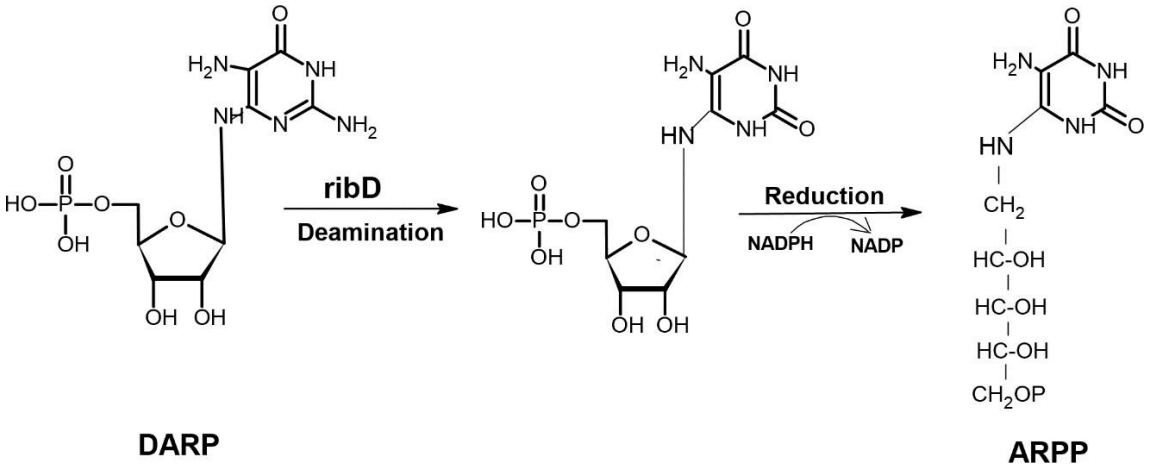


Fig. 2.8 Reaction Catalyzed by deaminase and reductase. The reaction involves the deamination and reduction of DARP into 5-Amino-6-ribosylamino-2,4(1H,3H)-pyrimidinone-5'-phosphate (ARPP).

4). Pyrimidine phosphatase

Pyrimidine phosphatase is the least characterized enzyme in RBS pathway and very limited information is available. It catalyzes the dephosphorylation of ARPP to 5-Amino-6-ribosylamino-2,4(1H,3H)-pyrimidinone (ARP) (**Fig. 2.9**) (Abbas & Sibirny, 2011)

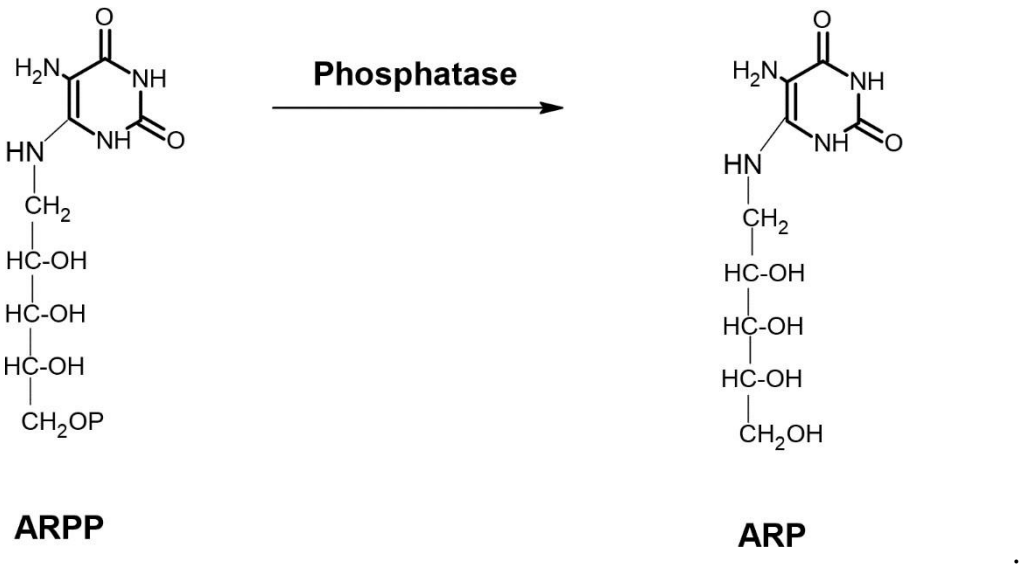


Fig. 2.9 Reaction Catalyzed by phosphatase. The reaction catalyzes the conversion of ARPP into 5-Amino-6-ribosylamino-2,4(1H,3H)-pyrimidinone (ARP) by the action yet unknown phosphatase.

5). Lumazine synthase (ribH)

Lumazine synthase is well studied and characterized enzyme of RBS pathway. It exists in two forms-pentameric and a 60mer form. It catalyse the condensation of ARP and DHBP to form 6,7-Dimethyl-8-ribityl-lumazine (DMRL) (**Fig. 2.10)** (Kis & Bacher, 1995).

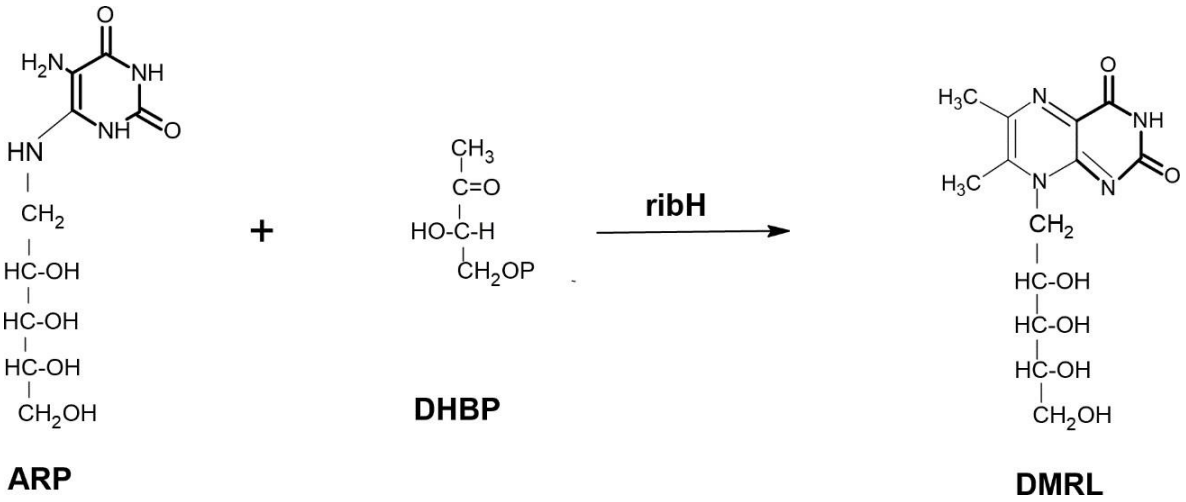


Fig. 2.10 Reaction Catalyzed by Lumazine synthase. The reaction catalyzes the condensation of ARP and DHBP to form 6,7-Dimethyl-8-ribityl-lumazine (DMRL).

6). Riboflavin synthase (ribE)

Riboflavin synthase is encoded by gene *ribE* which catalyzes the dismutation reaction on two molecules of DMRL to form one molecule of riboflavin and one molecule of ARP (Plaut, Smith, & Alworth, 1974) (**Fig. 2.11).** The riboflavin thus formed is utilized by the next enzyme whereas ARP is recycled and again used by ribH.

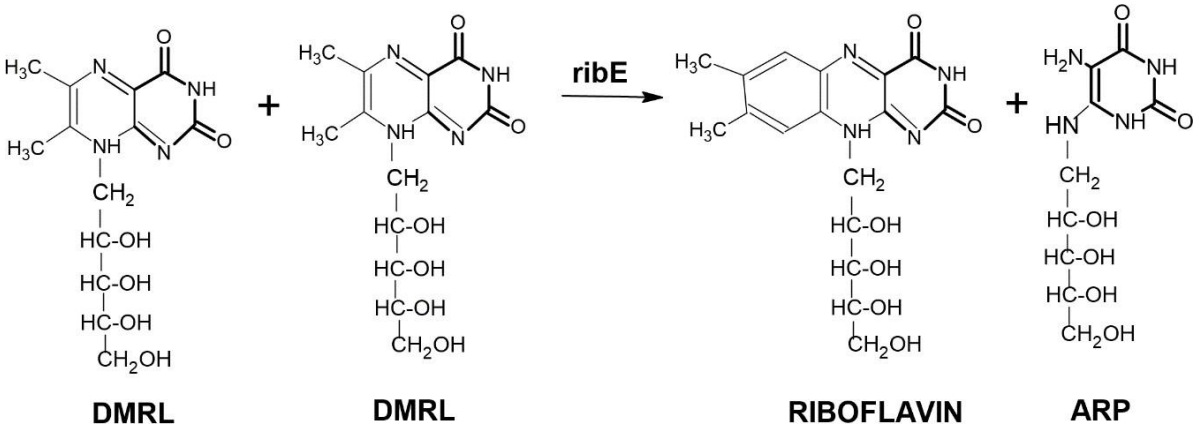


Fig. 2.11 Reaction Catalyzed by Riboflavin synthase. The reaction catalyzes the dismutation reaction on two molecules of DMRL to form one molecule of riboflavin and one molecule of ARP.

7). Riboflavin kinase /FAD synthase (ribF)

The enzyme riboflavin kinase is present in both prokaryotes and eukaryotes. In eukaryotes RF synthase and FAD synthetase is encoded by separate polypeptides whereas in prokaryotes both the functions are carried out by single bifunctional enzyme (Karthikeyan et al., 2003). Riboflavin kinase carries out the phosphorylation of riboflavin, converting it into FMN (Flavin mononucleotide) whereas FAD synthase converts FMN to FAD (Flavin adenine dinucleotide) by adenylation (**Fig. 2.12).**

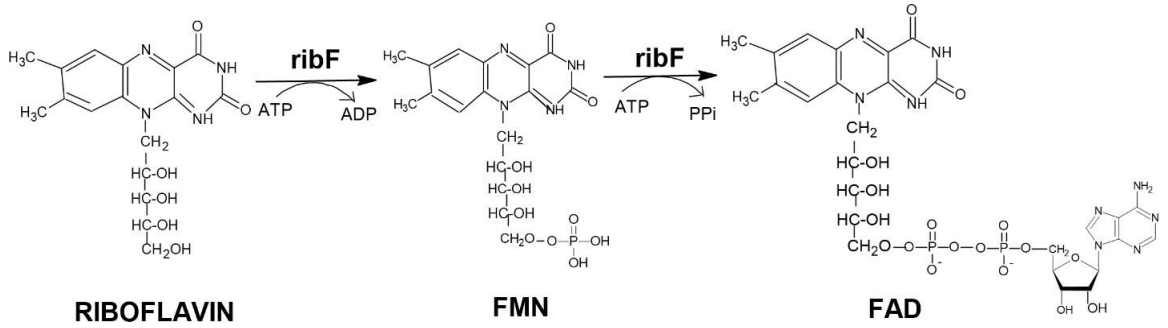


Fig. 2.12 Reaction Catalyzed by Riboflavin kinase /FAD synthase. The reaction involves the phosphorylation of riboflavin to FMN (Flavin mononucleotide) by riboflavin kinase and further adenylation by FAD synthase into to FAD (Flavin adenine dinucleotide).

2.3.1. Pyrimidine phosphatase (5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione)

In riboflavin biosynthesis, 5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate (ARPP) is an intermediate formed by the action of enzyme ribD on 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone-5'-phosphate (DARP). It then goes through additional dephosphorylation step to form 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (ARP). The ARP is required by the lumazine synthase, the next enzyme in RF synthesis (Bacher et al., 2000; M. Fischer & Bacher, 2008) in order to form another intermediate that eventually forms riboflavin. However, the enzyme and the mechanism of dephosphorylation is not yet clear. In addition, nonspecific phosphatases are unlikely to be involved in the biosynthesis of RF since they wouldn't be able to distinguish between the phosphorylated products of GTP cyclohydrolase II, reductase, and deaminase (Thung, Bach, Fasy, Jordon, & Schaffner, 1990).

Therefore, it has been hypothesized that an unidentified phosphatase is involved in the biosynthesis of riboflavin. *B. subtilis* served as one of the microbial producers of riboflavin at the industrial scale. It is recombinantly modified for the overexpression of rib operon (Bretzel et al., 1999). These recombinant strains of *B. subtilis* produce large amount of riboflavin without requiring the overexpression of hypothetical phosphatase. Considering this, it is speculated that the activity of this hypothetical phosphatase must be extremely high (Averianova, Balabanova, Son, Podvolotskaya, & Tekutyeva, 2020).

In vitro hydrolysis of ARPP and FMN by two *E. coli* enzymes - ybjI and yigB from the haloacid dehalogenase (HAD) superfamily was discovered (Haase et al., 2013). When grown without riboflavin, the single yigB or ybjI deletion mutants in *E. coli* show no phenotype, demonstrating the need for a double knockout strain to determine their *in vivo* role in riboflavin production. (Haase et al., 2013). Three HAD phosphatases, YcsE, YwtE, and YitU, acting on ARPP, were discovered in a *B. subtilis* strain that produces riboflavin. Targeted deletion of YcsE in *B. subtilis* did not result in the auxotrophy of riboflavin; hence, double or triple mutants are still needed to validate the *in vivo* functions of these enzymes in the riboflavin synthesis (Sarge et al., 2015). *Arabidopsis* plastidial FMN hydrolase AtcpFHy1 (At1g79790) was identified and its action on ARPP was later discovered *in vitro* (Rawat, Sandoval, Wei, Winkler, & Roje, 2011). Later on, the same group identified AtPyrP2, localized in plastids of *Arabidopsis thaliana*, to be the missing enzyme for catalyzing the dephosphorylation step of riboflavin biosynthesis pathway (Sa, Rawat, Thornburg, Walker, & Roje, 2016). However, such enzyme is not yet been identified in any pathway.

2.4. Haloacid dehalogenase family

Most enzymes are classified as members of families containing sequence-related proteins that catalyze similar function but have diversified their substrate preferences. One of the most prevalent enzyme superfamily in all organisms is haloacid dehalogenase (HAD)-like hydrolases superfamily which has 33 major families and 479,051 sequences in the database (InterPro IPR023214) (Burroughs, Allen, Dunaway-Mariano, & Aravind, 2006; Koonin & Tatusov, 1994). The HAD-like hydrolases in bacteria catalyzes the carbon or phosphoryl group transfer reactions in a range of substrates, and thus termed as Haloacid dehalogenase family and employs an aspartate residue in its active site for the catalysis (Koonin & Tatusov, 1994). These enzymes are present in all three-super kingdom of life and exhibit a wide range of biological activities (Burroughs et al., 2006). Most genomes are predicted to have multiple HAD-like proteins, including 183 in humans, 169 in *Arabidopsis thaliana*, 28 in *Escherichia coli*, 84 in *Caenorhabditis elegans*, and 30 in Mtb. (Allen & Dunaway-Mariano, 2009). Most of the HAD family proteins, including phosphatases, P-type ATPases, phosphonatases, and phosphotransferases, are involved in phosphoryl transfer (Aravind, Galperin, & Koonin, 1998). Originally named after dehalogenases, enzymes that catalyze carbon-halogen bonds, this superfamily is currently dominated by phosphatases (79%) and ATPases (20%), along with phosphonates and phosphomutases (Allen & Dunaway-Mariano, 2009; Aravind et al., 1998; Burroughs et al., 2006). In all kingdoms of life, a wide family of enzymes called HAD-like phosphatases, in particular, is responsible for the majority of metabolic phosphomonoester hydrolysis activities (Pandya, Farelli, Dunaway-Mariano, & Allen, 2014). Different biological functions that HAD-like phosphatases can carry out include primary and secondary metabolism, control of enzyme activity or protein synthesis, cell maintenance, and nutrient uptake (Allen & Dunaway-Mariano, 2009). Numerous HAD-like phosphatases have already been biochemically and structurally characterized from various organisms, such as the phosphoserine phosphatase SerB (Wang, Kim, Jancarik, Yokota, & Kim, 2001), the UMP nucleotidase NagD (Tremblay, Dunaway-Mariano, & Allen, 2006), and the inorganic pyrophosphatase from BT2127 (Huang et al., 2011). Substrate preferences of the 19 soluble *E. coli* HAD-like phosphatases revealed that the majority of these enzymes have extremely broad and overlapping substrate profiles and are active against phosphorylated coenzyme, organic acid, and nucleotide substrates (Kuznetsova et al., 2006).

2.4.1. Conserved motifs in HAD family

HAD family members can be identified by the presence of four short HAD signature motifs that have the conserved catalytic residues based on the amino acid sequence alignments, even though the total sequence identity is often quite low (often <15%) (Aravind et al., 1998) **(Fig. 2.13).** The HAD superfamily members contains four conserved sequence motifs according to extensive sequence comparisons (Burroughs et al., 2006).

The Motif I, present at the N-terminus has the DXD signature residues. It is a metal binding motif where the cofactor Mg^{2+}, coordinates with the carboxylate group of the first Asp and the backbone carbonyl of the second Asp. Furthermore, during catalysis, the initial Asp functions as a nucleophile to create an aspartyl-intermediate (Baker et al., 1998; Collet, Gerin, Rider, Veiga-da-Cunha, & Van Schaftingen, 1997; Morais et al., 2000; Qian, Stanley, Hahn-Hagerdal, & Radstrom, 1994; Seal & Rose, 1987). The second acidic residue functions as a universal acid-base in phosphatase and phosphomutase members of the superfamily, whereas for ATPase and phosphonatase, threonine and alanine were employed respectively. In many cases, the second aspartate protonates the substrate leaving group in the first step and deprotonates the nucleophile of the second step (Lahiri, Zhang, Dunaway-Mariano, & Allen, 2002a). The presence of threonine at this site in the ATPases reduces the rate of hydrolysis of aspartyl phosphate, allowing for the delay needed for the subsequent conformational change. The second aspartate is replaced by an alanine in the phosphonatases, which is consistent with the unique function of the enamine intermediate and is produced when these proteins hydrolyze aspartyl phosphate as a general acid-base catalyst (Burroughs et al., 2006).

The motif II of the HAD family includes the consensus sequence hhhhhh(S /T) (h stands for hydrophobic residues) and contains a conserved Thr (or Ser) that corresponds to S2 strand (**Fig. 2.14)**. Motif III involved a conserved Lys residue that is usually present at the N terminal of the helix located upstream of S4 (Burroughs et al., 2006; Koonin & Tatusov, 1994). The stability of the hydrolysis reaction intermediates is aided by motif II and motif III. According to the known structures of HAD hydrolases, the lysine in motif III can be found in either of two structural contexts. The lysine is embedded into the helix that is present prior the strand S4, in the Cof hydrolases, acid phosphatases, phosphoserine phosphatases, and P-type ATPases. But in other HAD hydrolases, it comes out of the loop right before the helix. Motif III is less conserved than the other motifs in the HAD family because of the variation in the secondary structure context of the lysine.

Motif IV is composed of conserved acidic residues. The three basic signatures for these pattern IV terminal acidic residues are generally DD, GDxxxD, or GDxxxxD or ED (for Beta phopshoglucomutases) (where x is any amino acid). The Mg^{2+} ion must be coordinated in the active site by these acidic residues along with those found in motif I (Lahiri et al., 2002a; Morais et al., 2000). The active site of HAD superfamily is created by the spatial arrangement of motifs I through IV around a single "binding cleft" at the C-terminal end of the strands of the central sheet. (Hisano et al., 1996; Peisach, Selengut, Dunaway-Mariano, & Allen, 2004). The hairpin flap following motif I, partially covers this binding cleft. The catalytic cavity is extensively shielded by additional inserts that appear between the two strands of the flap or just after motif III (Lahiri, Zhang, Dunaway-Mariano, & Allen, 2003a) (Burroughs et al., 2006; Rinaldo-Matthis, Rampazzo, Reichard, Bianchi, & Nordlund, 2002).

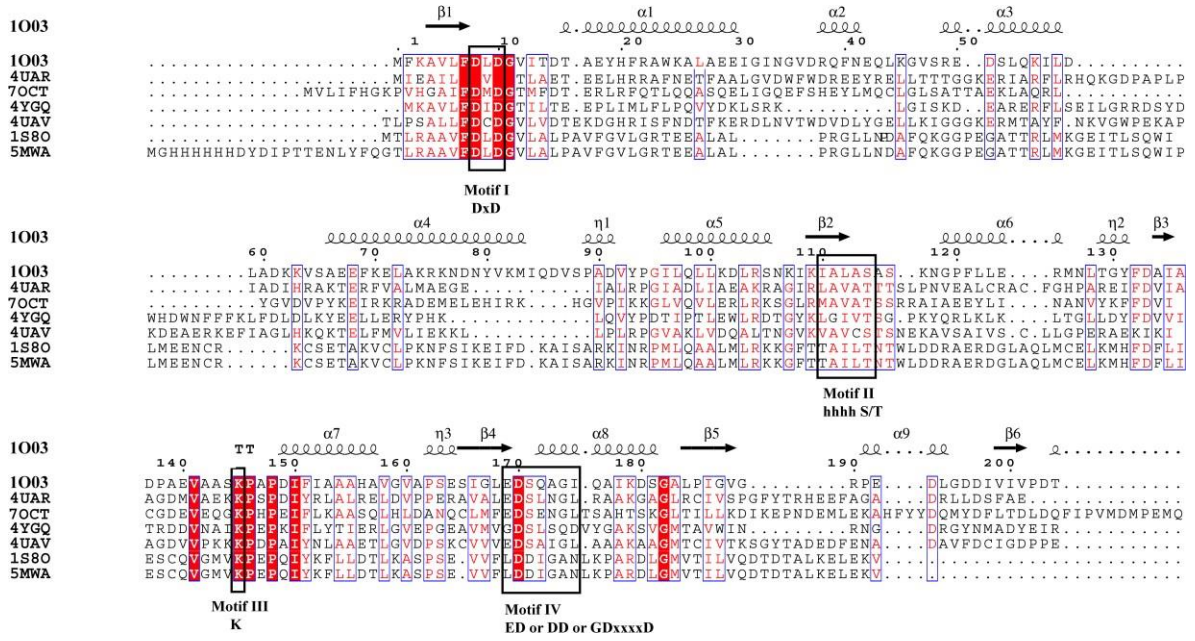


Fig. 2.13 Multiple sequence alignment of HAD-domain containing proteins. The alignment shows the conserved structural motif I, II, III and IV. The numbering and secondary structure elements of alignment corresponds to 1O03, a Beta-phosphoglucomutase from *Lactococcus lactis*. Conserved residues are shown in red on white background and identical residues are displayed in white on red background. The amino acid residues conserved between the groups are boxed. Motifs that are conserved between the HAD members are marked and placed within the box. Sequences are identified by the PDB ID; 4UAR (Sugar phosphatase from *Rhodobacter sphaeroides*), 7OCT (a bifunctional enzyme having mannito-1-phosphate dehydrogenase and phosphatase activity from *Acinetobacter baumannii*), 4UAR (HAD phosphatase from *Thermococcus onnurineus*), 4UAV (phosphatase from *Arabidopsis thaliana*), 1S8O (serine acetyltransferase from *Haemophilis ifluenzae*) and 5MWA (a bifunctional enzyme having lipid epoxide hydrolase and lipid phosphatase activity from *Homo*

sapiens). Multiple sequence alignment was performed with the Multalign server and the figure was generated by ESpript 3.0 on web server (https://espript.ibcp.fr/ESPript/ESPript/) (Robert & Gouet, 2014).

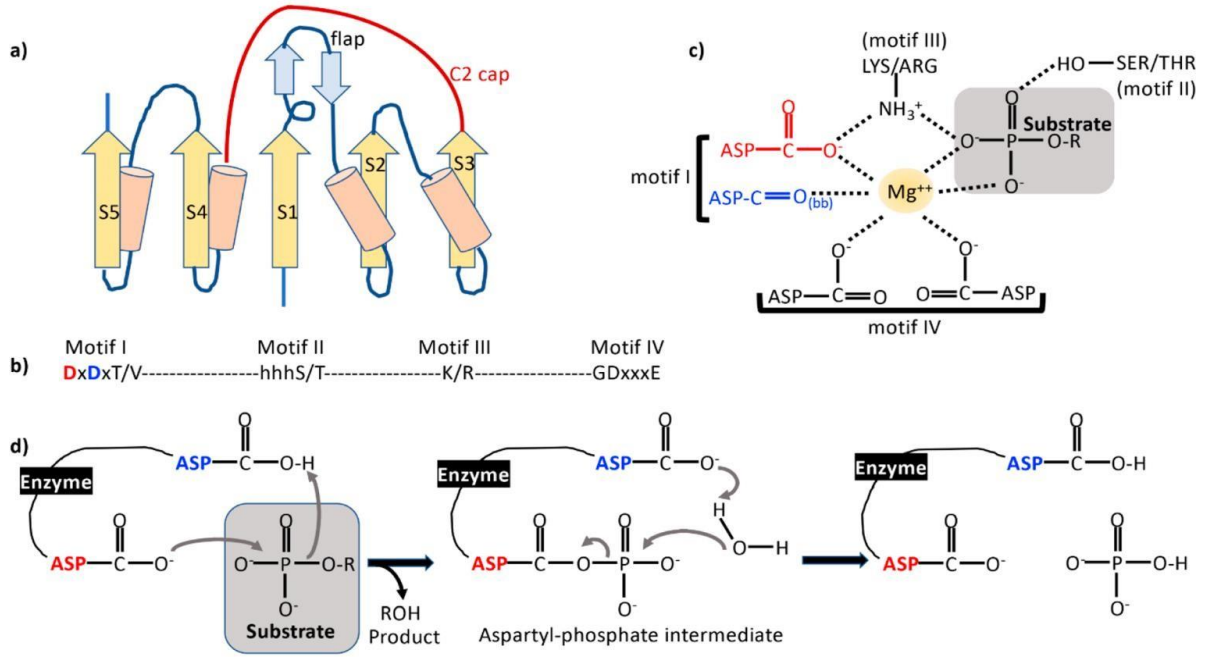


Fig. 2.14 Conserved motifs of HAD-phosphatases (a) Topology of haloacid dehalogenase (HAD) phosphatase domain. α-helixes are shown as cylinders whereas β-strands are shown as yellow arrows and are numbered S1–S5. Sample flap and C2 cap domains are illustrated. (b) The four conserved signature motifs of HAD phosphatases. (c) The conserved amino acids residues in the conserved motifs in co-ordination of the obligate Mg^{2+} ion and substrate interaction. bb denotes a backbone carbonyl group. (d) Schematic diagram of the reaction mechanism of HAD phosphatase. Motif's I, nucleophilic Asp (red) and the general acid-base Asp (blue) from motif I are color coded.

Source-https://www.mdpi.com/1422-0067/22/8/3925/html

2.4.2. Structural features of HAD

The HAD catalytic domain usually spans 200-250 aa residues, which constitutes 13-14 α-helices and 5-6 β-strands. The members of HAD family have two domains- a helical cap domain and the α/β core domain. The α, β-core catalytic domain with a central parallel β-sheet and α-helices on either side is shared by all HAD members. The core catalytic domain of the HAD superfamily is composed of a three-layered α/ β that adopt the architecture typical of the Rossmannoid (Rossmann-like fold) class. In most cases, the centre sheet has at least five strands (S1-S5) in a 54123-strand sequence and is parallel (**Fig. 2.14).** Two crucial structural motifs separate the HAD fold from other Rossmannoid folds: first the "squiggle," a distinctive

six-residue structural motif that assumes a nearly full single helical turn and is located immediately downstream of strand S1. Second, two strands extending from the domain's centre form a β-hairpin turn that is located downstream of the squiggle (Park, Guggisberg, Odom, & Tolia, 2015). This structural pattern forms the "flap." The squiggle can alternately be tightly or loosely twisted since it is close to a helical shape. The flap adjacent to the active site moves as a result, producing alternate closed and open states (Burroughs et al., 2006).

In addition to the squiggle' and 'flap' signature elements that contribute to the open and close conformation of the enzyme, mobile inserts known as cap modules are also present. The cap modules can provide more extensive shielding for the catalytic cavity rather than the flaps and also determine the substrate specificity (Burroughs et al., 2006; Park et al., 2015). The reactions that are catalyzed by the majority of HAD hydrolases are heavily dependent on the these caps, which frequently include residues needed for specificity or auxiliary catalytic functions (Baker et al., 1998; Olsen, Hepburn, Moos, Mariano, & Dunaway-Mariano, 1988) (Kurihara et al., 1995). The HAD caps fall into three general categories: (1) The structurally most basic members of the HAD superfamily are the C0 caps that has small inserts in either of the two points of cap insertion. (2) Inserts in the middle of the β-hairpin of the flap motif are referred to as C1, which usually contain α -helices and they fold into a structural unit different from the core domain. (3) Inserts in the linker that come right after strand S3 are referred to as C2 caps. These C2 caps mostly consists of α/ β fold. Although occasionally proteins may have both caps at once, the majority of members of the HAD superfamily have either a C1 cap or a C2 cap (**Fig. 2.15).**

2.5. Phosphoglucomutase

Phosphoglucomutase (PGM) catalyze the interconversion of D-glucose 1-phosphate (G1P) and D-glucose 6-phosphate (G6P) in carbohydrate metabolism. PGM is the connecting link between the glycolysis and gluconeogenesis pathway. Phosphoglucomutase are further divided into two classes: alpha-phosphoglucomutase (α-PGM, EC 5.4.2.2), which is present in all eukaryotes and prokaryotes, and beta-phosphoglucomutase (β-PGM, EC 5.4.2.6), is present in bacteria and protists. The specificity for α - and β-D-glucose phosphates, as well as the protein fold, separates these two groups. The phosphohexomutase superfamily contains α-PGM, while the haloacid dehalogenase (HAD) superfamily contains β-PGM (Shackelford, Regni, & Beamer, 2004). Compared to α-PGM, which has been extensively studied, β-PGM is not well studied.

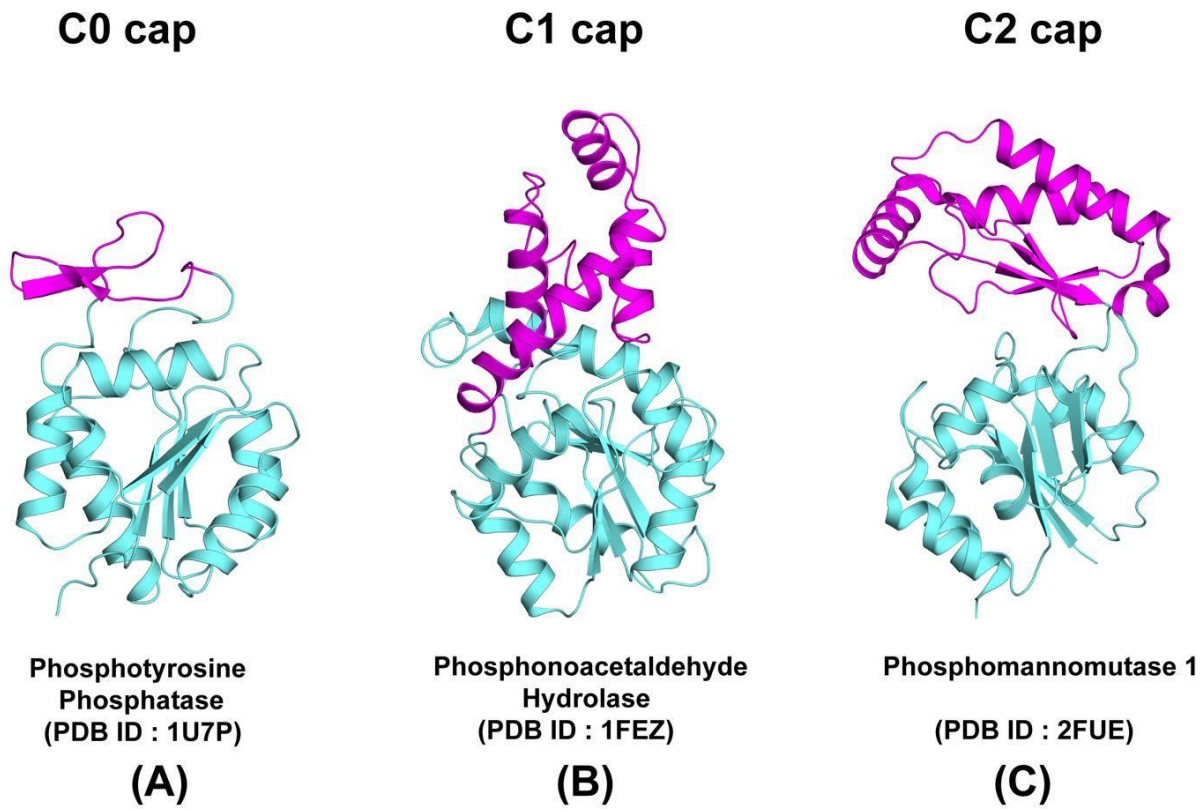


Fig. 2.15 Structural diversity of HAD phosphatase cap domains. (A) X-ray crystal structures of Phosphotyrosine phosphatase MDP-1 (PDB ID: 1U7P) (Peisach et al., 2004) from *Mus musculus*, (C0 cap as an unstructured loop), (B) Phosphonoacetaldehyde hydrolase (PDB ID: 1FEZ) (Morais et al., 2000) (C1 cap consisting of a four-helical bundle), and (C) Phosphomannomutase 1 (2FUE) (Silvaggi et al., 2006) (α /β-fold as a C2 cap). The cap domains are colored in magenta whereas catalytic domain is shown in aquamarine.

2.5.1. Alpha phosphoglucomutase (α-PGM)

The α- PGMs (~50 kDa) belongs to phosphohexomutase super family (**Fig. 2.16)**. The α-PGMs had shown to play an important role in energy metabolism by converting α-Glucose-1-phosphate (G1P) to α-Glucose-6-phosphate (G6P) (J. B. Dai, Liu, Ray, & Konno, 1992; Y. Liu, Ray, & Baranidharan, 1997; Ray, Post, Liu, & Rhyu, 1993). They are also involved in biosynthesis of alginate and lipopolysaccharides (Olvera, Goldberg, Sanchez, & Soberon-Chavez, 1999; Ye, Zielinski, & Chakrabarty, 1994) (Davey, Caiazza, & O'Toole, 2003). The cycle involves a phosphorylated active site serine phosphorylating α-G1P. The intermediate α-G1,6BP formed as a result, undergoes reorientation and binds the enzyme in the opposite direction. The product G6P is eventually released from the active site leaving the enzyme in the phosphorylated state (Ray, Burgner, & Post, 1990). Phosphorylated α-PGM transfers the

phosphoryl group to the next α-G1P, forming an intermediate (α-G16P) that will reorient to finally release the product α-G6P (Ray & Long, 1976).

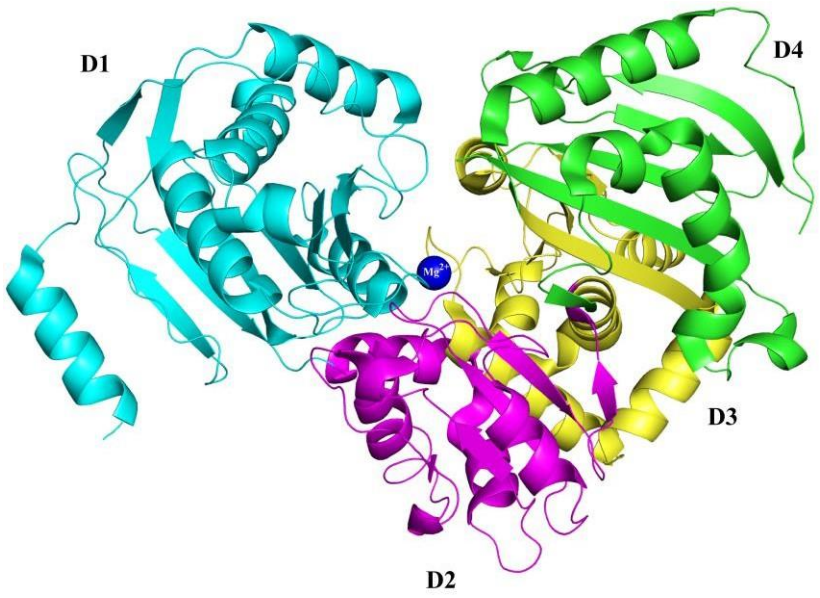


Fig. 2.16 Crystal structure of human α-PGM (PDB ID: 6SNP) (Backe et al., 2020). The monomers are colored by domains. Domain 1 (residues 1-205), marked as D1, is shown in cyan. D2 or Domain2 (residues 206-319) displayed in magenta, D3 or Domain 3 (320-441) shown in yellow, and domain 4 or D4 (residues 442-546) in green colour. The bound Mg^{2+} ion in the active site is shown as a blue sphere.

2.5.2. Beta-phosphoglucomutase (β-PGM)

Beta-phosphoglucomutase (β-PGM) is a member of the HAD superfamily and is the best-characterized till date. The primary cellular function of β-PGM is to sustain growth on maltose or trehalose, which involves the isomerization of β-D-glucose 1-phosphate (G1P) to β-D-glucose 6-phosphate (G6P), via an intermediary D-glucose 1,6-bisphosphate (G1,6BP) (Barrozo et al., 2018a). β-PGM has been characterized from *Lactococcus lactis* (Qian et al., 1994), *Bacillus subtilis* (Mesak & Dahl, 2000), and *Euglena gracilis* (Belocopitow & Marechal, 1974). Similar to α-PGM, β-PGM is known to use Mg^{2+} as a cofactor and G16P as a phosphoryl donor (Belocopitow & Marechal, 1974). The two catalytic scaffolds, however, differ significantly from one another, suggesting that the corresponding phospho enzymes are produced and stabilized using various mechanisms. β-PGM of *L. lactis* has been studied extensively and X-ray structures of phosphorylated, ligand (substrate, intermediate) and Mg^{2+} bound enzyme has been solved (Lahiri et al., 2002a) (Lahiri et al., 2003a; Tremblay, Zhang, Dai, Dunaway-Mariano, & Allen, 2005). The enzyme was depicted in an open active site conformation in the phosphorylated form. In ligand and cofactor bound structure, the active

site was observed in closed state and the Asp8 carboxylate oxygen formed the covalent bond to the phosphate of the β-D-glucose 1-phosphate. In one of the structure it was observed that the substrate analogue had D-galactose 1-phosphate in an active site closed conformation, with its C (6) OH group adjacent to the Asp8 nucleophile and its phosphate group at the distal phosphate binding site. (**Fig. 2.17).** This is an inhibitory complex of substrate attached to the dephosphoenzyme that is a dead end. Also, the enzyme with the bound Mg^{2+} as cofactor has been studied (G. Zhang et al., 2005a) and the Mg^{2+} bound β-PGM interacts with the intermediate G1,6BP during the catalytic cycle for the phosphorylation.

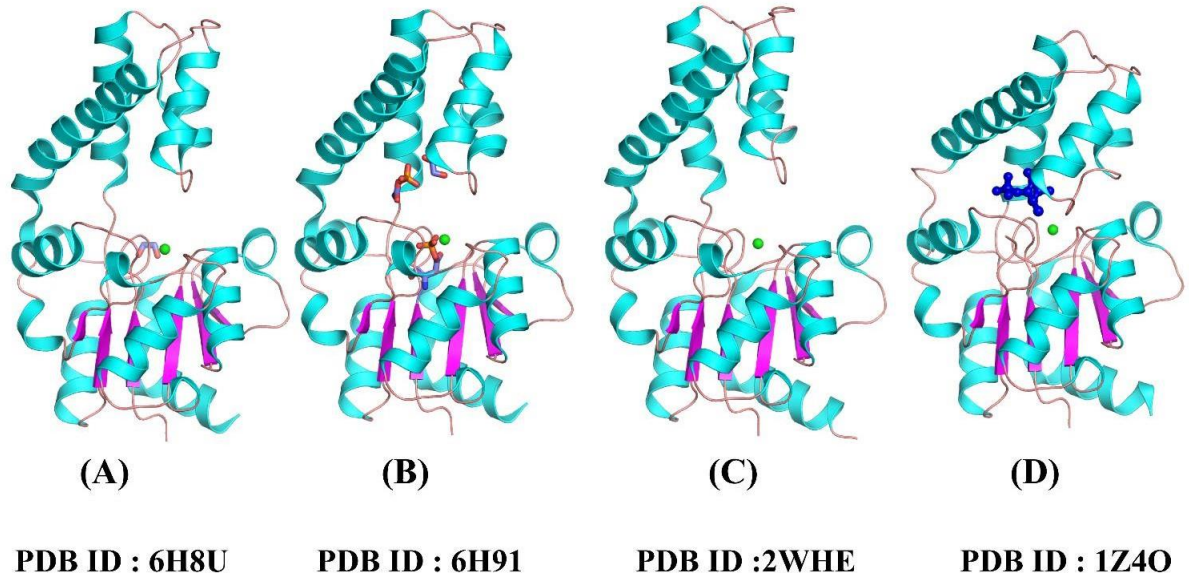


Fig. 2.17 Crystal structures of *Lactococcus lactis* β-PGM depicting open and closed conformation of the enzyme. **(A)** 6H8U, in open conformation with bound $Mg^{2+}$and EDO (1,2-Ethanediol) (deposited in PDB but unpublished), **(B)** 6H91, Phosphorylated β-PGM in open conformation with bound $Mg^{2+}$and EDO (Lahiri, Zhang, Dunaway-Mariano, & Allen, 2002b), **(C)** 2WHE, in open conformation without ligands (Baxter et al., 2010) and **(D)** 1Z4O, with bound inhibitor α-galactose 1-phosphate and in closed conformation (Tremblay et al., 2005).

2.6. Trehalose in Mtb

In *M. tuberculosis*, trehalose is a prevalent non-reducing disaccharide present in free form or as glycoconjugates in the cytosol and the cell wall (Elbein, Pan, Pastuszak, & Carroll, 2003). Trehalose is the non-reducing sugar and found in the cytoplasm of mycobacteria grown in standard laboratory media, and it makes about 1.5–3.0% of the total dry weight of the cells (Woodruff et al., 2004). The trehalose pool is constantly changing according to metabolic studies in the rapid growing *Mycobacterium smegmatis*, indicating it is not just accumulated as an energy reserve but also involved in other functions (Elbein & Mitchell, 1973). Trehalose

is crucial for the composition of mycobacterial cell envelope that mainly involves the trehalose monomycolate (TMM) and trehalose dimycolate (TDM) found on cell surfaces (Jankute, Cox, Harrison, & Besra, 2015). Trehalose has numerous roles in the physiology of Mtb. It serves as a source of carbon and energy, protects Mtb from desiccation, freezing, and osmotic stress, and is essential for Mtb virulence (Shi et al., 2013) **(Fig. 2.18).** Mtb alters its trehalose metabolism as an adaptive metabolic strategy and remodel it to overcome the stresses by stochastic formation of persisters during growth-adverse environments (Fisher, Gollan, & Helaine, 2017; Keren, Minami, Rubin, & Lewis, 2011; T. K. Wood, Knabel, & Kwan, 2013). The stresses that led to the development of Mtb PLBs (persister-like Bacilli) included nutritional starvation-induced systemic metabolic damage, intrabacterial ATP depletion, ETC dysregulation, and a corresponding loss of redox equilibrium (Dutta & Karakousis, 2014). To overcome these difficulties, preexisting trehalose was internally catalyzed through modified metabolic activities, resulting in an increase in the carbon flux toward the biosynthesis of glycolysis (GL) and pentose phosphate pathway intermediates (PPP), while a decrease in the carbon flux toward the biosynthesis of cell surface TMM/TDM. Initial substrates for the GL and PPP intermediates were formed by the trehalose-catalytic shift, which also provided an alternative supply of ATP, NADPH, and antioxidants (J. J. Lee et al., 2019).

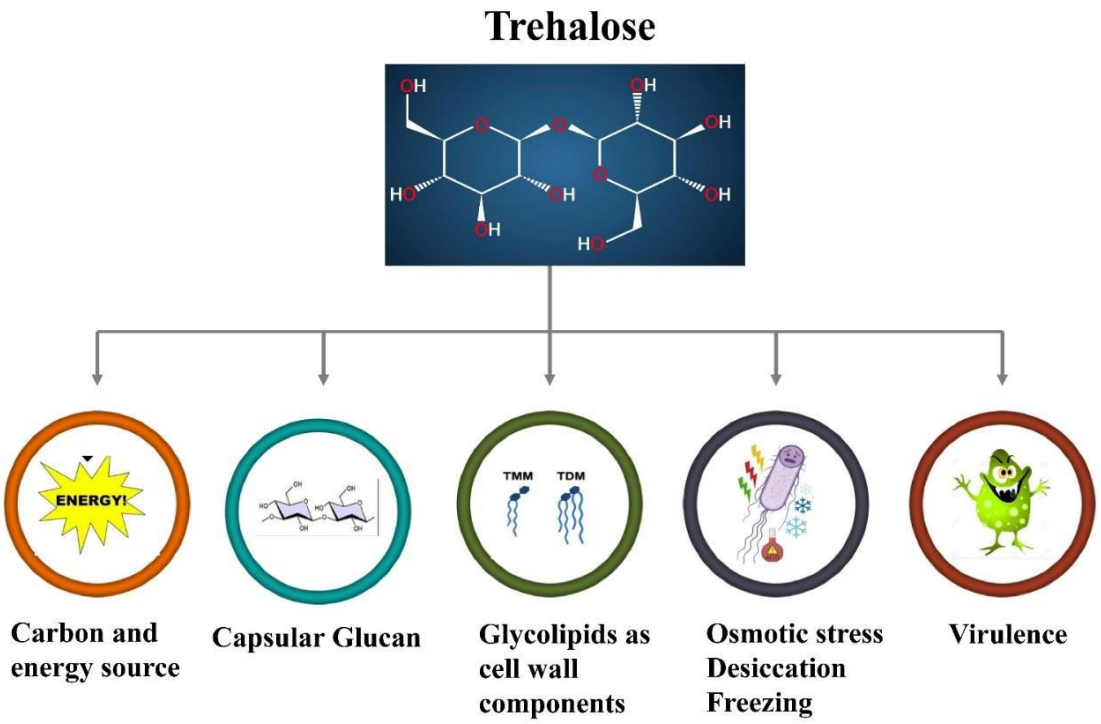


Fig. 2.18 Various functions of Trehalose in Mtb.

2.6.1. Trehalose biosynthesis pathway in Mtb

Mycobacterium tuberculosis can use exogenous trehalose as well as synthesize it *de novo* by three pathways. Endogenously, trehalose can be produced by mycobacteria from glucose 6-phosphate and UDP-glucose (the OtsA-OtsB pathway), from maltose (the TreS pathway) and from glycogen-like α (1-->4)-linked glucose polymers (the TreY-TreZ pathway) (De Smet, Weston, Brown, Young, & Robertson, 2000).

The main pathway (the OtsA/B or TPS-TPP2 pathway) consists of two steps. Trehalose phosphate synthase (TPS, OtsA) catalyses the first step by transferring a glycosyl group from UDP-glucose to glucose-6-phosphate to create trehalose-6-phosphate (T-6-P). The second step, where T-6-P is dephosphorylated to produce free trehalose, is catalyzed by T-6-P phosphatase (TPP, OtsB) (De Smet et al., 2000).

Trehalose synthase (TreS), an enzyme that converts trehalose into maltose and vice-versa, is a component of the TreS pathway. The Pep2 enzyme, a maltokinase that converts maltose into maltose 1-phosphate, connects the TreS and the GlgE pathways. The essential enzyme Glgb adds -1,6-linked branches to linear glucans, while the crucial enzyme maltosyl-transferase GlgE extends glucan chains from maltose 1-phosphate forming capsular glucan (Pan et al., 2004).

Maltooligosaccharides (glycogen) acts as the substrate for the third pathway, TreY/Z. Maltooligosyl trehalose synthase (TreY), the first enzyme, isomerizes the maltosyl at the reducing end of the maltooligosaccharide to a trehalose. The terminal disaccharide is hydrolyzed by the second enzyme, maltooligosyl trehalose trehalohydrolase (TreZ), to produce free trehalose.

Chapter 3
Materials and Methods

This chapter of the thesis encompasses the general material and methods employed during the experiments.

3.1. Chemicals, biological media and enzymes

The chemicals used in this study were procured from various commercial sources. All the chemicals required for various experiments including those required for preparing buffer solutions for agarose gel electrophoresis, SDS PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis), protein purification and crystallization were obtained from Sigma (USA). The nickel-nitrilotriacetic acid (Ni-NTA) agarose was obtained from Qiagen (Germany). Other chemicals used were of analytical grade and are specified otherwise. All the components used to prepare the biological media were obtained from HIMEDIA (India) and MERK (USA). The centrifuge-based concentrators (10 kDa cut off), syringe filters and membrane filters (0.45 and 0.22 µm) were obtained from Millipore (USA). The restriction enzymes, DNA polymerase, DNA ligase, DNA ladder and protein markers were obtained from ThermoFisher Scientific (USA), New England labs (NEB) (USA) and Fermentas (USA). The crystallization screening kits were obtained from Hampton research (USA), Qiagen (Germany), Emerald Biosystems (USA) and Molecular dimensions (UK). The MRC 96 well, 48 well and 24 well sitting drop crystallization plates were obtained from Molecular dimensions (UK). The 24 well hanging drop plates were obtained from Qiagen (Germany).

3.1.1. Bacterial strains and plasmids

3.1.1.1. Bacterial strains

The *E. coli* bacterial strains used in the study were *E. coli* Top10 (Invitrogen), *E. coli* DH5α (Invitrogen), *E. coli* BL21 (DE3), *E. coli* Rosetta (DE3) from Novagen (USA).

3.1.1.2. Plasmids

In this study, the pETDuet-1 vector (a pBR322 derived plasmid) from Novagen, was modified in such a way that the *BamHI* restriction site in the vector was replaced with in-frame *NheI* restriction site, thus creating a pETDuetN vector. The pETDuetN vector is an *E. coli* expression vector containing a constitutive T7 promoter with an ampicillin resistant gene as a selection marker. This vector has a 6xHis-tag at the 5' end and has two multiple cloning sites (MCS) followed by a T7 promoter/lac operator and a ribosome binding site (rbs). Another vector used in the study was pET28c (Novagen) that has a 6xHis-tag at the 5' and 3' end of

multiple cloning sites with a thrombin cleavage site at the N-terminal end of the cloned insert (Dubendorff & Studier, 1991). The kanamycin resistant gene is the selection marker for this vector. **(Fig. 3.1).**

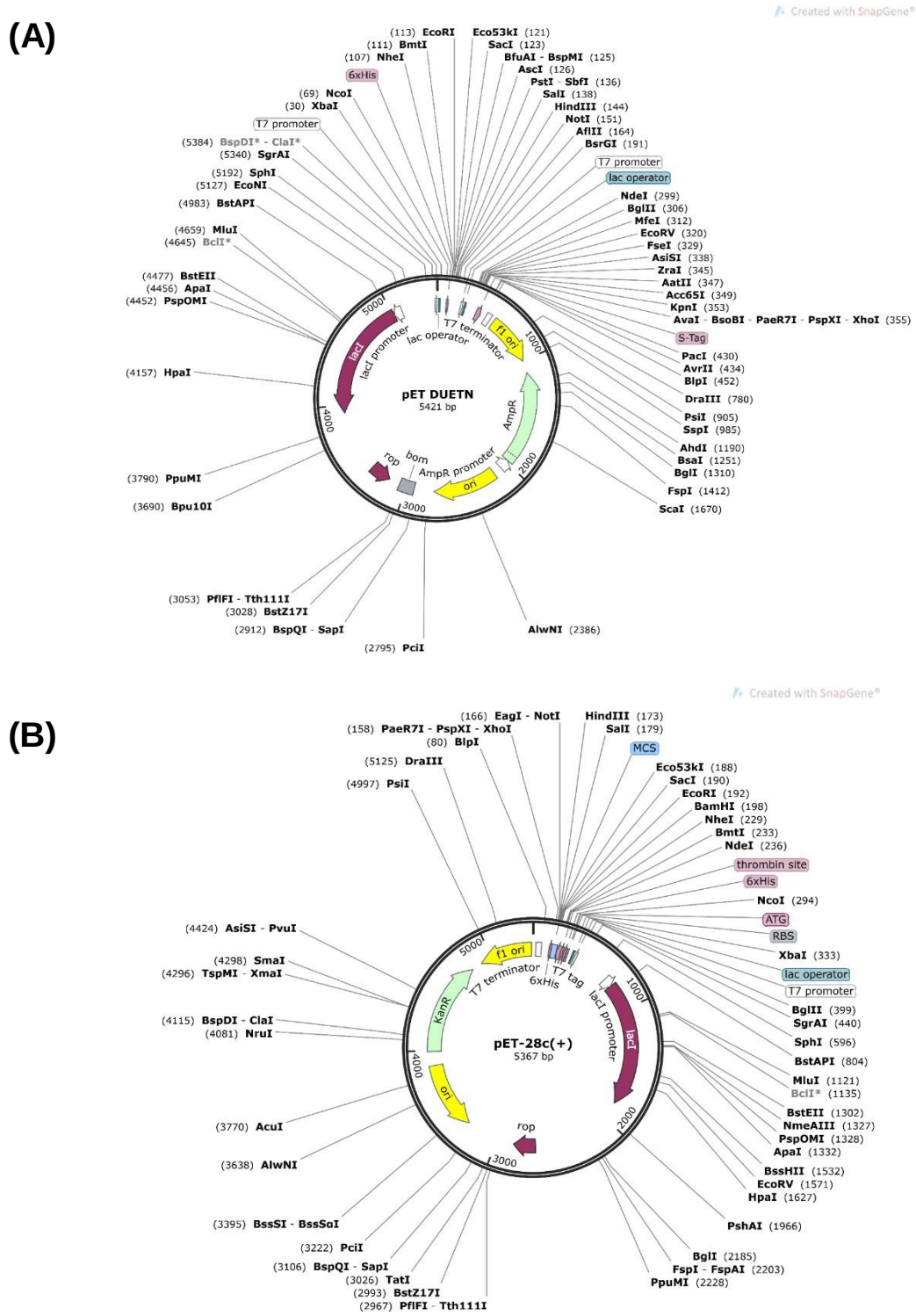


Fig. 3.1 (A) Plasmid map of pETDuetN (modified pETDuet1) having two MCS sites **(B)** Plasmid map of pET28c vector having Thrombin cleavage site. http://www.snapgene.com/resources/plasmid_files/pet_and_duet_vectors_(novagen)/pET-28c(+)/

3.2. Sterilization conditions for Autoclave

Buffer solutions, reaction components, water and media (Broth and Agar) used in the experiments were sterilized by autoclaving at 15 psi for 15 minutes at 121°C.

3.3. Bacterial growth media

3.3.1. Luria-Bertani media

Component	**Amount Required for 1L**
Tryptone	10 g (1%)
Yeast extract	5 g (0.5 %)
Sodium Chloride (NaCl)	10 g (1%)

The Luria-Bertani (LB) broth media was prepared by either mixing 25 g of ready to use media from Himedia (India) or by dissolving the individual components as mentioned above in 1 L of double distilled water. Sodium hydroxide (NaOH) was used to adjust the pH to 7.5 before autoclaving.

3.3.2. LB agar plates

For making LB-agar plates, 1.5% (w/v) of agar powder was added to the desired volume of LB media and autoclaved. Alternatively, 2.8 g of nutrient agar from Himedia (India) was dissolved in 100 mL of double distilled water and subjected to autoclave before use. The autoclaved agar media was cooled down to 45ºC and poured into petriplates about 20-25 mL and further allowed them to cool to become solid under sterile conditions. For making agar plates supplemented with desired antibiotics, the appropriate concentration of antibiotic was added to agar media before pouring into the petriplates.

3.4. Antibiotic stock solution

Various stock solution of antibiotics including ampicillin, kanamycin and chloramphenicol were prepared by dissolving the appropriate amount in the solvent. The stock solution of kanamycin and ampicillin were prepared in water while the chloramphenicol was prepared with isopropanol. The stock solutions were sterilized by passing through a 0.22 µm syringe filter and stored at -20ºC for further use.

Antibiotic	**Stock concentration**	**Working concentration**
Ampicillin	100 mg/ml	100 µg/ml
Kanamycin	50 mg/ml	50 µg/ml
Chloramphenicol	35 mg/ml	35 µg/ml

3.5. Buffer for competent cells preparation ($CaCl_2$ method)

Calcium chloride method was used for the preparation of all the competent cells used in the study. The pH of the buffer was adjusted to 7.0 by NaOH and the solution was filtered through 0.45 µm filter membrane before autoclaving. The composition of the $CaCl_2$ buffer is given below-

Component	**Amount**
$CaCl_2$	60 mM
Glycerol	15 %
1,4-Piperazinediethanesulfonic acid (PIPES) buffer	10 mM
Double Distilled water	For volume makeup

3.6. Competent cell preparation

Each step of the competent cell preparation was carried out in laminar flow hood. The components including LB media, $CaCl_2$ buffer solution and centrifugation tubes were autoclaved before use.

- The glycerol stock of competent cells (Top10, Bl21(DE3), Rosetta (DE3) stored at -80°C) were streaked onto LB agar plates (no antibiotics for Top10 and BL21 (DE3) but chloramphenicol was added for Rosetta (DE3) as it carries the plasmid conferring antibiotic resistance). The streaked plates were incubated overnight at 37°C.
- A single bacterial colony from the plate was used to inoculate in 10 ml of LB media. The media having the colony was incubated in an orbital shaker overnight at 37 °C at 250 rotations per minute (RPM).
- The 1% of fully grown primary culture was used to inoculate 200 ml LB media (secondary culture) and incubated further at 37^{o}C and 250 RPM in an incubator shaker till the optical density at 600 nm (OD_{600nm}) reached a value of 0.4-0.5 (mid log phase).

- The culture was kept on ice for 30 minutes with gentle swirling occasionally, and transferred to a pre-chilled 500 ml centrifugation tube followed by centrifugation at 4°C, 3,500 RPM for 10 minutes.
- The supernatant was discarded and 40 ml of $CaCl_2$ buffer (pre-chilled) was added immediately and the pellet was resuspended by gentle swirling on ice and centrifuged at 4°C, 3,500 RPM for 10 minutes.
- The supernatant was discarded and the pellet was gently resuspended in 40 ml of $CaCl_2$ solution in chilled condition.
- The cell suspension was incubated further for 30 minutes on ice and again centrifuged at 3,500 RPM for 10 minutes at 4°C.
- The pellet was gently resuspended in 2 ml of $CaCl_2$ buffer discarding the supernatant.
- Aliquots of 100 µl cell resuspension were dispensed in cryovials (pre-chilled).
- The cryovials were immediately stored at -80°C for further use.

3.7. Analysis of protein samples by SDS-PAGE

SDS PAGE electrophoresis was used for the detection, quantity and quality of protein of interest. Usually 10-15 % SDS-PAGE was used in this study. The components required for the preparation of stock solutions are described below-

Buffer/ Solutions	**Composition**	**Amount**
Polyacrylamide mix solution (30% w/v)	Acrylamide	29 g
	N,N-methylene-bis-acrylamide	1g
	Distilled water	upto 100 mL
1.5 M Resolving gel Buffer (pH 8.8)	Tris Base	27.25 g
	Distilled water	upto 150 mL
1.0 M Stacking gel Buffer (pH 6.8)	Tris Base	18.17 g
	Distilled water	upto 150 mL
10 % SDS solution	SDS	1 g
	Distilled water	upto 10 mL
10 % Ammonium per sulphate solution	APS	0.1 g
	Distilled water	upto 1mL

10X SDS Running Buffer stock solution (pH 8.3)	Tris Base	30 g
	Glycine	144 g
	SDS	10 g
	Distilled water	upto 1L
5X Sample loading buffer	Tris HCl (pH 6.8)	0.15 M
	SDS	5.0 % (w/v)
	Glycerol	25.0 % (v/v)
	β-mercaptoethanol	12.5 mM
	Bromophenol blue	0.06 % (w/v)
SDS Staining solution	Methanol	40 % (v/v)
	Acetic acid	10 %(v/v)
	Coomassie Brilliant Blue R25	0.1 % (w/v)
	Distilled water	upto 1L
SDS Destaining solution	Methanol	40 % (v/v)
	Acetic acid	10 %(v/v)
	Distilled water	upto 1L

3.7.1. SDS-PAGE Composition

Composition of resolving gel (15mL)

Components	**12 % gel (mL)**	**15 % gel (mL)**
Double distilled water	4.9	3.4
Acrylamide Mix (29.1)	6.0	7.5
Resolving gel buffer (pH 8.8)	3.8	3.8
10 % SDS	0.15	0.15
10 % APS	0.15	0.15
Tetramethylethylenediamine (TEMED)	0.006	0.006

Composition of stacking gel (5mL)

Components	**5 % gel (mL)**
Double distilled water	3.4
Acrylamide Mix (29:1)	0.83
Resolving gel buffer (pH 8.8)	0.63
10 % SDS	0.05
10 % APS	0.05
Tetramethylethylenediamine (TEMED)	0.005

The SDS-PAGE was performed in accordance with the Laemmli procedure to examine the expression and purity of the protein (Laemmli, 1970). The protein samples were prepared by combining 1X final concentration of 5X sample loading buffer. Prior to loading onto the SDS-PAGE, the samples were boiled for 5 minutes and centrifuged at 12,000 x g for 10 minutes. After loading the samples, SDS-PAGE was run using a Biorad electrophoresis system at 150 volts and 30 mA current in the SDS running buffer. In the gel, the protein molecular weight markers were also run simultaneously to compare the molecular weight of the protein of interest. The SDS-gels were stained with coomassie staining for 30-45 minutes, then using destaining solution, the gel was allowed to destain with gentle agitation with the help of rocker for 1-2 hours until the protein bands were clearly visible.

3.8. Polymerase Chain Reaction (PCR)

3.8.1. PCR Reagents

- Phusion High-Fidelity DNA polymerase with HF (High-Fidelity) and GC buffer, ThermoFisher Scientific, USA
- 10 µM Forward primer (100 µM stock)
- 10 µM Reverse primer (100 µM stock)
- dNTP Mix (premixed aqueous solutions of dATP, dCTP, dGTP and dTTP, 10 mM each), Thermo Scientific, USA
- Autoclaved double distilled water.

3.8.2. Oligonucleotide Primers

The oligonucleotide primers used for PCR amplification were synthesized by Sigma-Aldrich, India. The details of primers used in the study are listed in the respective chapters.

The master mix prepared was stored at -20^{o}C and used further to set up a PCR reaction.

To resuspend the oligos/primers received in dry form, autoclaved double distilled water was used. The vials containing oligos were briefly centrifuged to settle the content before opening the tube. The information about the quantity of water was mentioned in the specification sheet or concentration of the oligo (in nM) was multiplied with 10. The resultant product is the amount of water (in µl) that has to be added in the tube to make a stock of 100 µM. Mix the oligos with water by vortexing the vial for 5-10 minutes.

Alternatively, a spectrophotometrically based protocol can also use for the quantitation of oligonucleotides. For this protocol, add an aliquot of the resuspended oligonucleotide to a final volume of 1,000 µl with water (water = 1,000 µl – volume of oligonucleotide added). For mixing the contents, vortex or pipette up and down for 15 seconds. The absorbance of this dilution was measured at 260 nm (A260) in a UV spectrophotometer. The concentration of oligonucleotide in the stock solution can be calculated by the formula given below-

Concentration in µg/ml = A260 × Weight per OD × dilution factor*
(* the dilution factor is determined by 1000/amount of resuspended oligo added for the dilution, ie., if 50 µl of resuspended oligo is added to 950 µl of water to read the absorbance, the dilution factor would be 1,000/50 = 20)

3.8.3. PCR master mix

For amplification of target genes, PCR was carried out in a reaction volume of 50 µl. A master mixture was prepared for the ease and accuracy with the components given below-

Composition	**Amount**
10X HF or GC Buffer	200 µl
10 mM dNTP’s	20 µl (0.2 mM)
DNA polymerase	7 µl
Double distilled water	Add volume upto 1000 µl

The master mix prepared was stored at -20 ^{o}C and used further to set up a PCR reaction.

3.8.4. PCR reaction

The PCR reaction was carried out using the master mix, primers and template DNA as described below-

Composition	**Amount (Final concentration)**
Master mix	49 µl
Forward primer (100 µM)	0.2 µl (0.4 µM)
Reverse primer (100 µM)	0.2 µl (0.4 µM)
Template DNA	0.5 µl (50 ng)

The above-mentioned PCR reaction was carried out in a Master Cycler gradient PCR machine (Eppendorf, Germany) using the standard protocol as described below.

Step1	Initial denaturation	95^{o}C for 5 minutes
Step 2	Denaturation	95^{o}C for 1 minute
Step 3	Annealing	54^{o}C - 60^{o}C for 1 minute (Depending on the target gene)
Step 4	Extension	72^{o}C for 1 minute
Step 5	Final Extension	72^{o}C for 10 minutes
Repeat step 2-4 for 30 cycles		
Step 6	Hold at 4^{o}C	

3.9. DNA Agarose gel electrophoresis

For the detection of DNA, agarose gel electrophoresis was used. The concentration of the agarose was determined based on the expected molecular weight of the DNA fragments. For 1% agarose gel, 0.30 g of agarose was dissolved in 30 mL of 1X TAE (Tris acetate EDTA) buffer followed by boiling for 2-3 minutes. After slight cooling 1 % w/v of ethidium bromide (EtBr) was added. The gel was then poured into a horizontal gel running apparatus containing a comb and allowed to solidify for well formation

The TAE buffer, EtBr solution and loading dye was prepared using the following components:

Buffer/ Solutions	**Stock Solution (50X)**	**Working Solution (1X)**
TAE Buffer (1L)	2M Tris Base (242g)	40 mM
	1M Glacial Acetic Acid (57 mL)	20 mM
	0.5M EDTA pH 8.0 (100 mL)	10 mM
	Composition	**Amount**
Ethidium Bromide stock solution (1% w/v)	Ethidium Bromide	0.10 g
	Distilled water	10mL
6X DNA loading buffer	Bromophenol blue	0.25 % (w/v)
	Glycerol	30 % (v/v)
	Xylene Cyanol	0.25 % (w/v)
	Distilled water	10 mL

The 50x TAE buffer was sterile filtered by passing through a 0.45 µm filter and stored at room temperature for further use.

The 6X loading dye was added to the final concentration of 1X before loading the sample e.g PCR product in the gel. The loading dye contains Bromophenol blue and xylene blue as tracking dyes that moves in the same direction of separating DNA and demarcates their leading edge. In addition to the loaded sample, 1 kb DNA ladder (marker) was run alongside to determine the size of the desired DNA fragment. The horizontal electrophoresis was carried out in the presence of 1X TAE buffer and operated at 90 volts and 300 mA current. The UV transilluminator was used to visualize DNA fragments.

3.10. Purification of PCR products by gel extraction method

With the help of Agarose gel electrophoresis, the DNA fragments were separated on the basis of their size.

The DNA fragments from the agarose gel were extracted and purified using GeneJet gel extraction kit (Thermo Scientific, USA), according to the protocol provided by the manufacturer and it is described briefly below:

- The desired DNA fragment was cut from the agarose gel using a clean and sharp scalpel under UV transilluminator with safety precautions.

- The gel slice was transferred to a clean 1.5 ml micro-centrifuge tube and 1:1 volume of binding buffer from the kit was added.
- The tube was incubated at 50-60°C for 10 minutes (until the gel slice dissolves completely). The process can be fastened by vortexing the tube contents every 2–3 minutes during incubation.
- After the gel slice was completely solubilised, 1 volume of isopropanol was added.
- The solution containing the dissolved DNA sample (upto 800 µl) is transferred to the GeneJET purification column and incubated for 1-2 min at room temperature for efficient binding.
- The tube was then centrifuged at 4000 × g for 2 minute and flow through was discarded.
- 500 µl of wash buffer from the kit was added to the column and centrifuged further at 12,000 × g for 1 minute.
- The flow through was discarded and additional spin was given for 2 minutes at 12,000 g to remove any residual traces of ethanol.
- In the next step, the GeneJET purification column was placed in a clean 1.5 ml microcentrifuge tube and 30 µl of elution buffer (10 mM Tris. HCl, pH 8.5) or autoclaved H_2O was added directly to the column matrix and incubated for 2 minutes. Finally, the DNA was eluted from the column by centrifuging the tube at 12,000 × g for 1 minute.

3.11. Nucleic acid quantification

The concentration of the DNA was estimated by measuring the absorbance at 260 nm using a ND-1000 Nanodrop Spectrophotometer (Thermo Scientific, USA). One unit of the absorbance at 260 nm was equivalent to a concentration of 50 ng/µl of double-stranded DNA (Sambrook et. al., 1989). By measuring the OD_{260}/OD_{280} ratio, the purity of the DNA was evaluated. The ideal value for OD_{260}/OD_{280} ratio for a well purified DNA should be around 1.8. (Wilfinger, Mackey, & Chomczynski, 1997).

3.12. Preparation of vector/Plasmid

The plasmid DNA was isolated from recombinant *E. coli* cells using the GeneJET plasmid isolation kit (Thermo Scientific, USA). Below is a list of the various processes used for plasmid purification as per the manufacturer's methodology.

- The glycerol stock of desired plasmid was streaked onto LB agar plates with the respective antibiotic selection marker. The streaked plates were incubated overnight at 37°C.

- A primary culture of 5 mL of LB medium supplemented with or without antibiotic (as per requirement) was inoculated using a single colony that was taken from a freshly streaked LB agar plate. The primary culture was then incubated overnight at 37°C at 200 RPM in an incubator shaker.
- Next day, harvesting of cells were done by centrifuging the culture at 6000 x g for 15 min at 4°C. After discarding the supernatant, the pellet was resuspended using 250 µl of resuspension buffer from the kit.
- The resuspended cells were lysed by adding 250 µl of lysis buffer and gently mixed by inverting 4–6 times, followed by incubation at room temperature for 2 minutes.
- 300 µl of neutralization buffer was added to the cell lysate, mixed immediately but gently by inverting 4-6 times. The cell debris and chromosomal DNA was pelleted down by centrifuging the lysate for 12-15 min at maximum speed (12,000 RPM).
- The supernatant was immediately transferred to the GeneJET spin column by gentle pipetting or decanting without disturbing the precipitant.
- The spin column was then incubated for 1-2 minutes for its binding with DNA, followed by its centrifugation at 4000 rpm for 2 minutes. The flow through was discarded.
- The spin column was washed two times with 500 µl of wash buffer and centrifuged at 12,000 RPM for 2 minutes.
- Centrifugation for an additional 3 min at 12,000 RPM was done to remove any residual wash solution. The column was incubated for 5 minutes at room temperature for the evaporation of residual ethanol.
- The GeneJET spin column was placed in a clean 1.5 ml microcentrifuge tube and 30 µl of elution buffer (10 mM Tris. HCl, pH 8.5) or autoclaved H_2O was added directly to the column matrix. The plasmid DNA was eluted by incubating the column for 2 minutes followed by centrifugation at 12,000 × g for 1 minute.

3.13. Digestion of plasmid and PCR amplified genes

The isolated plasmid DNA and purified PCR products were digested with the same set of restriction enzymes that creates a either blunt or sticky end required for efficient ligation. The restriction enzymes used in this study were mentioned in their respective chapter.

Generally, the FastDigest restriction enzymes from Thermo Scientific, USA were used for digestion. According to standard protocol, 1µl (10 Unit) of enzyme was used for digestion of 1 µg of plasmid DNA. The digestion was carried out at 37 °C for 2-3 hours. The incubation time and the amount of enzyme depends on the concentration of the DNA used for digestion. The general reaction protocol is described below-

Component	**Amount**
Plasmid DNA/ PCR product	1µg plasmid/ PCR product
Enzyme I	0.1µl (1 Unit)
Enzyme II	0.1µl (1 Unit)
10 x Fast Digest Green buffer	2 µl (1 X)
Autoclaved water	Adjust to make total volume 20 µl

To avoid self-ligation of the linearised plasmid DNA, the digested plasmid was subjected to alkaline phosphatase treatment. 1 unit of fast alkaline phosphatase from Thermo Scientific (USA) was added to the digested plasmid and incubated for 10 minutes at 37°C to dephosphorylate its 5' end. Both the digested plasmid and PCR product were run on a agarose gel and further purified using GeneJET gel extraction kit (Thermo Scientific, USA) according to the protocol provided by the manufacturer.

3.14. Ligation

Following digestion and purification of the plasmid DNA and PCR product, both of them were ligated using T4 DNA ligase (Thermo Scientific, USA) as described by Sambrook et al., 1989.

The amount of digested PCR product required for ligation was calculated by the formula-

$$\text{PCR product required (ng)} = \frac{50 \times \text{PCR product size (bp)} \times \text{molar ratio}}{\text{plasmid size (bp)}}$$

A ratio of 1:3 for the digested plasmid and PCR products, respectively, was used for ligation. The ligation reaction was carried out at 25°C for 2-3 hours.

A 2X ligation master mixture made with the following components-

Component	**Amount**
10 X Ligation buffer	20 µl
15 % PEG 4000	33 µl
T4 DNA ligase	2.5 µl (2.5 U)
Water	Add upto 100 µl

The general ligation reaction consists of following components -

Component	**Amount**
Digested Plasmid	50-60 ng
Digested PCR product	1:3 or 1:6 (Plasmid:PCR product)
2X Ligation master mix	10 µl (1X)
Water	Add upto 20 µl

3.15. Transformation of plasmid into competent cells

The ligated reaction of digested plasmid and PCR product was transformed in the $CaCl_2$ competent cells, as per manufacturer protocol which is described below-

- For transformation, the competent cells that was stored in -80°C was thawed for 3-5 minutes on ice. The ligation mixture or the plasmid DNA (50-100 ng) was added to the competent cells and further incubated on ice for 20-30 minutes.
- The cells were given a heat shock at 42°C for 90 seconds and immediately kept back in ice for 2 minutes.
- To the reaction mixture 900 µl of fresh LB was added and mixed gently. The cells were allowed to grow for 45 minutes by incubating at 37ºC with 200 RPM in an incubator shaker.
- The cells were centrifuged at 3,000 RPM for 5 minutes and the supernatant was discarded except for 50 µl, which was used to resuspend the pellet. The resuspended cells were plated on LB agar plate supplemented with appropriate antibiotics and incubated for 14–16 hours at 37°C.

3.16. Screening of colonies for putative clones

3.16.1. Screening of clones by colony PCR

The colony PCR was used for the detection of gene of insert in the plasmid construct. In order to screen multiple colonies, a PCR reaction mixture was made using recombinant Dream Taq polymerase (Thermo Scientific, USA) and other components. 10 µl of this Dream Taq master mixture is distributed in PCR tube. To each tube a small amount of colony was added and mixed gently with the help of a pipette tip. The PCR reaction is then carried out in a thermocycler using the optimized manufacturers protocol for Dream Taq polymerase. The PCR product was then run on 1% agarose gel to check the positive colonies.

Master mixture composition-

Component	**Amount**
10 X Dream Taq buffer	100 µl
10 mM dNTP's	20 µl (0.2 mM)
Taq polymerase (5U/µl)	7 µl
Autoclaved water	Add up to 1000 µl

PCR reaction for screening of clones by colony PCR-

Component	**Amount**
Dream taq master mix	49.5 µl
Forward Primer (100 µM)	0.2 µl (0.4 µM)
Reverse Primer (100 µM)	0.2 µl (0.4 µM)
Final volume	50 µl

The final mix (50 µl) was distributed in five 0.2 mL PCR tubes (10 µl in each). The PCR reaction was carried out in Master Cycler gradient PCR machine (Eppendorf) with the standard protocol as described below.

Step 1	Initial Denaturation	95^{o}C for 1-3 minutes
Step 2	Denaturation	95^{o}C for 30 seconds
Step 3	Annealing	54^{o}C-60^{o}C for 1 minute
Step 4	Extension	72^{o}C for 1 minute
Step 5	Final extension	72^{o}C for 5-15 minutes
Repeat step 2-4 for 30 cycles		
Step 6	Hold at 4^{o}C	

After the colony PCR, amplified products were load on 1 % agarose gel along with 1Kb DNA ladder. The colonies that had shown the positive amplification or band corresponding to the gene of interest were selected and used to inoculate 5ml of LB broth supplement with appropriate antibiotic. The primary culture was then incubated overnight at 37°C and 200 RPM incubator shaker. To isolate plasmid of these clones, the kit from Thermo Scientific, USA (GeneJET Plasmid isolation kit) was used as per the manufacturer's discussed previously in section 3.9.

3.16.2. Double digestion to confirm clone

To confirm the gene of insert, the isolated plasmid was double digested with appropriate restriction enzymes (Fast Digest, Thermo Scientific, USA) for 20 min at 37°C. The digested products were run on agarose gel to check the fallout of desired size.

3.17. DNA sequencing of gene of interest in the plasmid

The clone confirmed by double digestion was checked for the correct sequence. The plasmid was sequenced using an automated DNA sequencer to confirm the exact sequence and ensure the correct reading frame in the plasmid. The details of sequencing PCR reagents and primers used is as follows:

3.17. DNA sequencing of gene of interest in the plasmid

The clone confirmed by double digestion was checked for the correct sequence. The plasmid was sequenced using an automated DNA sequencer to confirm the exact sequence and ensure the correct reading frame in the plasmid. The details of sequencing PCR reagents and primers used is as follows:

3.17.1. Sequencing PCR Reagents

- 5X Sequencing buffer (Applied Biosystems, USA)
- Terminal-Ready Reaction (TRR) enzyme mix (Applied Biosystems, USA)
- 10 µM Forward primer (100 µM stock) or 10 µM Reverse primer (100 µM stock)
- Autoclaved double distilled water.

To confirm the sequence of clones, the sequencing was done using T7 promoter and T7 terminator primers and their sequences are listed below-

Table T7: Primers used for sequencing

Primer name	**Primer Sequence**
T7 Promoter	5'TAA TAC GAC TCA CTA TAG GG 3'
T7 Terminator	5' GCT AGT TAT TGC TCA GCG G 3'

The composition of the sequencing PCR reaction is as follows:

Composition	**Amount**
DNA (Plasmid)	50-100 ng
5X Sequencing buffer	2 µl
T7 primer (Forward or Reverse)	1 µM
Terminator ready reaction mix (TRR)	1 µl
Double distilled water	Add upto 10 µl

3.17.2. Sequencing PCR protocol

Step 1	Initial Denaturation	96°C for 5 min
Step 2	Final Denaturation	96°C for 30 sec
Step 3	Annealing	50°C for 15 sec
Step 4	Extension	60°C for 4 min
Repeat step 2-4, for 30 cycles		
Step 5	Hold at 4°C	

3.18. Expression of recombinant protein

After the confirmation of the desired gene in the vector, the plasmid was transformed into either BL21 (DE3) or Rosetta (DE3) *E. coli* cells. The expression of the gene was initially checked on a small scale from the freshly formed colonies.

To prepare primary culture, a single colony from the LB-agar plate was inoculated in 5 ml LB media with appropriate antibiotics and allowed to grow at 37°C for 10–12 hours at 200 RPM. A secondary culture was prepared by inoculating 1 % of the primary culture in a 10 ml LB media along with required antibiotics. This secondary culture was kept in an incubator shaker at 37°C and was grown until the absorbance at OD_{600} reached a value of 0.6–0.8. At this point, the culture was separated into two tubes each with 5 ml of culture. In one of the tubes, 0.3 mM isopropylthio-D-galactoside (IPTG) was added and treated as induced sample, while the other was treated as uninduced sample. Both uninduced and induced cell cultures were grown further at 37°C for 4-5 hours before the cells were harvested by centrifugation at 4000 x g for 5 minutes at 4°C. The harvested cells were resuspended in lysis buffer containing 50 mM Tris base pH 8.0, 150 mM NaCl and 1 mM DTT and boiled at 100 °C for 5 min. The resuspended cells were mixed with Sodium dodecyl sulfate (SDS) loading dye and loaded on 15 % SDS-PAGE to check the protein expression.

Once the expression of recombinant protein in *E. coli* was confirmed, the solubility of the recombinant protein was examined. For this, the induced bacterial pellet was dissolved in the lysis buffer and the cells were disrupted by sonication (Sonics, USA) for one minute (30 seconds ON and 30 seconds OFF cycle with 20 % amplitude). By centrifuging the lysate at 12000 x g for 10 min at 4°C, the cell debris containing the insoluble protein was separated from the supernatant. The pellet and supernatant was loaded on a 15 % SDS-PAGE to determine whether the protein is soluble and elutes in the supernatant fraction. If the protein was insoluble at 37°C, expression at lower temperatures with various IPTG concentrations were attempted to get the protein in a soluble fraction. To obtain a good amount of protein for structural studies, the culture volume was ramped up to 2-3 liters after standardizing the expression of protein in soluble fraction. The individual chapters include information on the specific requirements for soluble protein expression.

3.19. Preparation of selenomethionine-labelled protein

The minimal media was prepared by dissolving M9 salts ($Na_2HPO_4.7H_2O$, 33.97 g/l, KH_2PO_4, 15 g/l, NaCl, 2.5 g/l, NH_4Cl, 5 g/l) in 1 L Milli Q water, filtered and then sterilized by autoclaving. 1M $MgSO_4$ and 1M $CaCl_2$ were also prepared separately and autoclaved. 20% glucose, trace elements ($FeCl_3$.$6H_20$, $ZnCl_2$, $CuCl_2$, $CoCl_2.6H_2O$, H_3BO_3, $MnCl_2.6H_2O$), vitamin B-complex and amino acids (threonine 100 mg/l, phenylalanine 100 mg/l, lysine 100 mg/l, isoleucine 50 mg/l, valine 50 mg/l, selenomethionine 60 mg/l) (all the chemicals

including amino acids are purchased from Sigma-Aldrich Co.) were prepared and filter sterilized. The final media used to grow cells was made by mixing 200 ml of M9 salts, 2 ml of 1M $MgSO_4$, 20 ml 20% glucose, 100 ml of 1M $CaCl_2$, 1ml of vitamin B-complex and 1x of trace elements and the volume was made up to 1L using autoclaved water. The primary culture was grown in 10 ml LB media and the cells were harvested in the log phase by centrifugation at 4000 RPM for 5 min. The pellet obtained was resuspended in 5 ml of minimal media and used to inoculate the 1L minimal media as secondary culture. The culture was incubated at 37 °C with shaking at 200 RPM. At $O.D._{600nm}$ of ~0.6, all the amino acids -threonine 100 mg, phenylalanine 100 mg, lysine 100 mg, isoleucine 50 mg, valine 50 mg, including the selenomethionine 60 mg were added and incubated at 37 °C with shaking at 200 RPM. After 15 min, the culture was induced with 0.3 mM IPTG and further incubated at 16 °C for 18 to 20 hrs. Harvesting of cells by centrifugation at 9000xg for 10 min, was done and protein was purified using Ni-NTA chromatography followed by size exclusion chromatography as described in later sections.

3.20. Purification of recombinant proteins

The proteins involved in the study were initially purified by Ni-NTA chromatography followed by gel filtration chromatography. The various steps involved in the protein purification are discussed below-

3.20.1. Affinity chromatography with Ni-Nitrilotriacetic Acid (Ni-NTA chromatography)

The proteins in this study have a 6xHis-tag at either N-terminal or C-terminal end depending on the vector used. This 6xHis-tag was used for purification of over-expressed protein by Ni-NTA chromatography (Qiagen). The 6X His-tag binds efficiently to the Ni^{2+} which are immobilized on agarose beads using nitriloacetic acid (NTA) (Crowe et al., 1994) (PMID: 7921034). This chromatography, which is used to separate proteins from other proteins in solution by taking advantage of 6XHis-tag, is also known as immobilized metal affinity chromatography (IMAC). As shown in a range of proteins, including enzymes, transcription factors, and vaccines, the 6X His tag typically has no interference with the purified protein's structure or function (Carson, Johnson, McDonald, Brouillette, & Delucas, 2007; Chant, Kraemer-Pecore, Watkin, & Kneale, 2005) PMID: 15642465. Typically, the supernatant having the soluble recombinant protein of interest is passed through the Ni-NTA beads (Merk-Sigma Aldrich, USA) packed in a polypropylene column (Qiagen, Germany). The wash buffer

is used to remove the non-specifically bound proteins from the beads, and imidazole-containing elution buffer is used to elute the target protein from the column.

3.20.2. Dialysis

Dialysis is the process of separating molecules in solution through a semi-permeable membrane by taking advantage of difference in their rates of diffusion. It is frequently used for buffer exchange to eliminate excess salts, and reducing agents from protein containing buffer (such NaCl, imidazole, and DTT). In our study, the Small Wonder-Lyzer dialysis bag (Excellion Innovations and Inventions Inc., US patent no. 6368509) was used to dialyze the recombinant protein that was purified by Ni-NTA chromatography against a desired buffer. These dialysis bags have a cellulose membrane with a molecular weight of 10 kDa cutoff.

3.20.3. Gel-filtration chromatography

Gel filtration or size exclusion chromatography is used for additional purification of the protein and also to determine its oligomeric state in solution. In gel filtration chromatography, proteins are separated based on their hydrodynamic radius or its molecular size (Hagel, 1998) and thus also known as size exclusion chromatography. The separation of the proteins is achieved by the porous matrix for which molecules have different degree of access. The macromolecules smaller in size (have access to the matrix) comes later as they get trapped in the matrix while the large molecules are excluded from the matrix and come out quickly from the column. In this study we used Superdex^{TM} 210/300 GL (GE Healthcare) column for additional purification.

The oligomeric state of a protein in solution is determined by comparing the elution volume against a standard curve of proteins of known molecular weight. The proteins standards in calibration curve to estimate the size of unknown proteins are - aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa) and procured from GE Healthcare, USA.

3.21. Protein concentration and estimation

Protein concentrator with a 10 kDa cut off membrane (Millipore, USA) was used to concentrate protein by centrifuging at 2,500 x g at 4 °C. The concentration of the protein was estimated by Bradford method (Bradford, 1976). The absorbance of known concentration of Bovine serum albumin (BSA) was used as the reference standard. Perkin Elmer Lambda

UV/VIS spectrophotometer was used to measure the absorbance at 595 nm. The protein concentration of an unknown sample was determined by plotting the absorbance at 595 nm against several known concentrations of BSA. Briefly, 500 µl of the sample (499 µl water +1µl of protein) and 500 µl of Bradford reagent (Sigma Chemicals, USA) were mix together and allowed to stand for 5 min, followed by recording the absorbance at 595 nm. The absorbance recorded was compared with the standard to calculate the exact concentration of protein.

Alternatively, the concentration of recombinant protein was also determined using nanodrop Denovix nanodrop (DS-11 Series Spectrophotometer). This instrument employs the molecular weight and extinction coefficient of the protein which is related to its tryptophan (W), tyrosine (Y) and cysteine (C) amino acid composition.

The concentration of a purified protein in solution is also measured by using absorbance spectroscopy. The absorbance, *A*, is a linear function of the molar concentration, *C*, according to the Beer-Lambert law: $A = \varepsilon \times l \times c$, where e is the molar absorption coefficient and *l* is the cell path length.

3.22. MALDI-TOF mass spectrometry

Mass spectrometry is an analytical technique in which chemical compounds are ionized into charged molecules and ratio of their mass to charge (m/z) is measured. The introduction of matrix aided laser desorption ionization (MALDI) and electron spray ionization (ESI) improved the applicability of MS to large biological molecules like proteins. In MALDI-TOF mass spectrometry, the ion source is matrix-assisted laser desorption/ionization (MALDI), and the mass analyzer is time-of-flight (TOF) analyzer. In MALDI-TOF, the analyte should be soluble to at least about 0.1 mg/ml in some solvent. And the matrix is dissolved to yield either a saturated solution or a concentration of about 10 mg/ml. The solution of the analyte is then mixed to that of the matrix. For optimized MALDI spectra, the molar matrix-to-analyte ratio is normally adjusted as to fall into the range from 1000: 1 to 100,000: 1, and then the mixture is then spotted onto a metal target plate for analysis. After drying, the mixture of the sample and matrix co-crystallizes and forms a solid deposit of sample embedded into the matrix. The plate is subsequently loaded into the MALDI-TOF instrument and analyzed by software associated with the respective system. The MALDI leads to the sublimation and ionization of both the sample and matrix. These generated ions are separated depending on

m/z through a TOF analyzer, and a spectral representation of these ions is generated and analyzed by the MS software, generating an MS profile (Boesl, 2017; Schiller et al., 2007; Singhal, Kumar, Kanaujia, & Virdi, 2015).

To find the intact mass of the protein, we have used an Applied Biosystems Voyager DESTR mass spectrometer and sinapinic acid as a matrix. Desalting of protein sample was done by buffer exchange using centrifugal ultrafiltration devices Vivaspin (Sigma, USA). The sample was centrifuged at 12000 g for 10 min at 4 °C to remove precipitation and the final concentration of the protein was kept around 0.8-1mg/ml.

3.23. Circular dichroism studies

Circular dichroism (CD) is considered to be one of the excellent spectroscopic methods to find out the secondary structure and folding characteristics of the protein under investigation. The CD is most frequently used to track modifications in secondary structure protein folding due to changes brought by environmental factors, mutations, or interactions with other molecules (Kelly, Jess, & Price, 2005; Kelly & Price, 2000). Therefore, CD spectroscopy can provide structural and thermodynamic data on macromolecules like proteins (Greenfield, 2006).

Electromagnetic waves contain electric and magnetic field components that oscillate perpendicularly in the direction of a light beam's propagation. The directionality of these components defines the waves' polarization. In unpolarized light or white light, the electric and magnetic fields oscillate in many different directions. In linearly polarized light, the electromagnetic wave oscillates along a single plane while in circularly polarized light (CPL) two electromagnetic wave planes are at a 90° phase difference to one another and this plane rotates as the light beam propagates. The circular dichroism is simply the differential absorption of left-handed and right-handed circularly polarized light (CPL). Time dependent electric and magnetic fields are associated with the beam of the light. If the molecule under investigation has chiral (asymmetric) chromophores, either right-handed CPL or left-handed CPL will be absorbed to a greater extent. A CD signal can therefore be either positive (left-handed CPL absorbed stronger) or negative (right-handed CPL absorbed stronger). This causes the plane of the light wave to rotate in an ellipse and the light is referred to be elliptically polarized. CD is expressed as either degrees of ellipticity, which is defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse, or in units of ΔE

which is the difference in absorbance of right CPL and left CPL by an asymmetric molecule. Molar ellipticity ([θ]) is CD corrected for concentration. The units of molar ellipticity are (deg×cm2/dmol) (Greenfield, 2006; Kelly et al., 2005; Kelly & Price, 2000).

CD spectroscopy in the "far-UV" spectral region is used to determine the protein's secondary structure (190-250 nm). At this wavelength range, CD signal arises due to the peptide bond when it is located in a regular folded environment. The positive band at 193 nm and negative bands at 222 nm and 208 nm can be found in the CD spectra of proteins with α-helices. Negative bands at 218 nm and positive bands at 195 nm are present in proteins with clearly defined anti-parallel β-sheets, whereas disordered proteins show very low ellipticity above 210 nm and negative bands close to 195 nm (Greenfield, 2006; Kelly et al., 2005; Kelly & Price, 2000).

For the determination of secondary structure content of the protein used in this study, far-UV CD spectra were collected on a Jasco–810 CD spectropolarimeter that was flushed with nitrogen at the rate of 9–12 L/min. A quartz cuvette with a path length of 1 mm was used to obtain the data, which ranged from 250 nm to 198 nm. Each obtained spectra consisted of an average of five scans and was recorded as raw ellipticity. Data collection for all protein samples utilized a concentration of 0.2 mg/ml. The mean residue ellipticity (MRE) was calculated by raw ellipticity according to the equation,

$$[\theta]_{MRE} = ([\theta] \text{ X } 100 \text{ X } M_r) /(c \text{ x } l \text{ x } N_A)$$

where $[\theta]_{MRE}$ depicts the mean residue ellipticity, [θ] stands for the raw ellipticity in degrees, Mr is molecular weight of the protein (in Da), l is the path length of the cuvette (in cm), c is the concentration of the protein (in mg/ml), and the number of residues of the protein is depicted by $N_{A.}$

3.24. X-ray crystallography

In this study, the three-dimensional protein structure was determined using the X-ray crystallography technique. The crystallization of the protein would be the initial stage in X-ray crystallography. This technique has found to be a major contributor in the Protein Data Bank (RCSB PDB, http://rcsb.org) (Rose et al., 2017). Numerous techniques have been generated and modified over the time to generate protein crystals that are suitable for diffraction. The most widely used methods are vapor diffusion, free interface diffusion, batch,

and dialysis procedures (McPherson & Cudney, 2014) (Mcpherson, 2001; McPherson, 2017) (McPherson, 2004) (Chayen & Saridakis, 2008; Dessau & Modis, 2011) (McPherson & Cudney, 2006). During crystallization process, the protein enters into super-saturated state and precipitates as crystals under regulated conditions. In the vapor-diffusion method (such as sitting-drop and hanging-drop), the supersaturation of the protein is caused by the equilibration of the vapor pressure across the precipitant concentration gradient between the drop containing protein solution and the reservoir. In the current study, the sitting drop vapor diffusion method was initially employed to set up crystallization utilizing the purified and concentrated protein (McPherson, 1997). Different crystallization screens were employed to obtain the crystals.

The details of the commercially available crystallization screens used in this study are given below-

S.No.	**Name of the kit**	**Manufacturer**	**Catalog number**
1.	Crystal screen HT	Hampton Research, USA	HR2-110
2.	PEG Ion2 screen	Hampton Research, USA	HR2-098
3.	PEG Rx2 screen	Hampton Research, USA	HR2-084
4.	Index screen	Hampton Research, USA	HR2-144
5.	SaltRX screen	Hampton Research, USA	HR2-106 & HR2-107
6.	NATRIX screen	Hampton Research, USA	HR2-116 & HR2-117
7.	Proplex screen	Molecular Dimensions, UK	MD1-38
8.	PGA screen	Molecular Dimensions, UK	MD1-50
9.	Structure screen	Molecular Dimensions, UK	MD1-03
10.	Wizard screen	Molecular Dimensions, UK	MD15-W12-T
11.	JCSG screen	Molecular Dimensions, UK	MD1-37
12.	Midas screen	Molecular Dimensions, UK	MD1-106

3.24.1 Crystallization plates and Accessories

For initial crystallization screening, Swissci 96-well sitting-drop vapour diffusion UV compatible crystallization plate and 96-well MRC 3-well sitting drop plates (Molecular Dimensions, UK) were used. The crystallization drop was set up using an NT8 robotics system

(Formulatrix Inc., USA). For the optimization of potential crystallization hits, 48-well sitting drop plates (Molecular Dimensions, UK) and 15-well hanging drop plates (Qiagen, Germany) were used. The crystallization plates were sealed with the help of sealing tape and sealing paddle (Hampton Research, USA). Further optimization was also done using 72-well microbatch plate (Hampton Research, USA) to improve the quality of crystals. Microbatch crystallization is a method where the protein and reagent are combined and sealed in a plate, tube, container, or sealed under a layer of oil. A layer of 1:1 of Silicon oil:Paraffin oil (Hampton Research, USA) was spread over the microbatch plate to control the rate of evaporation. For data collection, crystals were mounted with the help of loops (Molecular Dimensions, UK and Hampton Research, USA) of various sizes and shapes (mesh, round and elliptical) and with different diameters (0.025, 1, 2 mm) depending on the crystal morphology.

3.24.2. X-ray diffraction data collection

The next step towards determination of a structure involves obtaining the diffraction pattern of the crystal. Crystals were mounted using the cryo loops and the data sets were collected under cryogenic conditions (100 K) using a constant flow of nitrogen vapor stream. The single wavelength anomalous dispersion (SAD) data were collected using synchrotron radiation at 11.2 C Elettra, Italy while native data was collected at home source using a Mar345dtb image plate detector mounted on a Rigaku Micromax-007 HF rotating anode X-ray generator operated at 40 KV and 30 mA. Details of data collection for each data set will be discussed in the respective chapter.

3.24.3. Data collection strategy

For optimum X-ray diffraction data collection, consideration of several things like Bravais lattice selection, X-ray intensity, wavelength, crystal-to-detector distance, exposure time, oscillation angle, redundancy, resolution, *etc* (Dauter, 1999) is involved and briefly described below-

- The number of images needed to collect a full set of reflections is constrained by symmetry, so understanding a Bravais lattice is extremely helpful in developing a data collecting strategy.
- The distance between the crystal and the detector was optimized during data collection such that it could collect high resolution data and the spots are well separated.

- An oscillation angle of 1º was found to be a good compromise between speed and the data quality in the predicted reflection profiles due to higher background because bigger oscillation angles tend to reduce the signal to noise ratio and accuracy.
- The exposure time depends on the size of the crystal, oscillation range and diffraction quality (Dauter, 1999). The exposure time was chosen to provide accurate statistics at the maximum resolution while not overexposing the detector with strong low-angle spots.

3.24.4. Data processing, integration and scaling

The data processing of the X-ray diffraction data can be divided into quite three distinct steps (Otwinowski & Minor, 1997). The first step is to determine the crystal cell parameters, specifically the crystal lattice i.e. unit cell, space group, and orientation. This procedure is called **indexing**, and determining these parameters will aid in a preliminary calculation of crystal mosacity.

- The second step is the **integration of the images** (with concurrent refinement of crystal, beam and detector parameters) in segments.
- The third and the final step involved **scaling and merging**. This involves placing all data on a common scale, merging multiple observations to give a unique dataset while rejecting outliers.
- The X-ray diffraction data collected from a single protein crystal can be processed using various software packages like HKL2000 (Otwinowski & Minor, 1997), iMosflm (Battye, Kontogiannis, Johnson, Powell, & Leslie, 2011; Leslie, 2006) and XDS (Kabsch, 2010) and others. In the study we have used XDS for data processing.

X-ray Detector Software (**XDS**) processes single-crystal diffraction data. It also consists of three steps: indexing, integrating and scaling. XDS will index and integrate the reflections. There is one JOB= line in XDS.INP (provided by beamline software) which specifies a list of tasks: JOB= XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT.
Further, XDS comes with additional programs XSCALE and XDSCONV. Scaling of data can be done using XSCALE and XDSCONV can convert the final processed files in to various other file formats used by the programs for structure solution.

3.24.5. Phase problem in X-ray Crystallography

The exposure of an X-ray beam on a single crystal produces a diffraction pattern which usually consists of spots. Each spot may then be represented by a single wave representing the

magnitude and relative phase of the X-rays scattered along that direction. The following data must be known in order to compute the electron density at a position (xyz) in the unit cell of a crystal: (a) the reflection indexes (hkl), (b) the intensity of the reflections, I(hkl), and (c) the phase angles of the reflections (hkl). This information is used to calculate the electron density ρ at a point (xyz) by using the following mathematical equation.

$$\rho(xyz) = 1/V \Sigma| F_{hkl}| \exp (i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

where V is the volume of the unit cell and α_{hkl} is the phase associated with the structure-factor amplitude $|F_{hkl}|$. The amplitude is easily calculated by taking the square root of the intensity, but the phases are lost in the experiment. This is known as phase problem in crystallography. In order to visualize and construct the individual atomic positions, we need to obtain phase information and therefore estimation of phases is an essential step of structure determination (Harrison, 1993).

3.24.6. Recovering the phases

Phases and amplitudes have no direct relationship. The only connection that can be made between these two entities is through electron density or molecular structure. As a result, phases can be identified if one can establish prior knowledge of structure or electron density. Various methods to determine phases are known, viz., direct method (as applied to experimental phasing), molecular replacement and isomorphous replacement.

3.24.7. Experimental phasing

Phases can be determined by experimental phasing, which relies on a few atoms in the macromolecule with known special properties, such as a large number of electrons and/or anomalous scattering. These properties are exploited to determine the location of the special atoms (substructure), and then knowledge of the substructure in one or more crystals is used to deduce phase information for the entire macromolecule. Thus, experimental phasing solves the phase problem by using the differences in amplitudes from different diffraction experiments to determine the structure factor phases. In modern crystallography, it is most common to use single-wavelength or multiple-wavelength anomalous diffraction experiments (SAD/MAD).

In PHENIX, the primary program for performing experimental phasing is *phenix.autosol*. This is a comprehensive, automatic tool that performs experimental phasing

with the MAD, MIR, SIR, or SAD methods. The program locates the substructure, estimates phases, performs density modification, identifies non-crystallographic symmetry, and builds and refines a preliminary model.

To run *phenix.autosol*, diffraction data is provided, typically one or more anomalous data sets, or a native data set plus one or more derivative data sets (in case of performing SIR/MIR phasing). Also the sequence of macromolecule is provided, and information about the anomalous scatterers and/or heavy atoms. The *phenix.autosol* results include a model file containing the anomalous scatterers, an MTZ file containing the experimentally determined phases, the optimized phases from density modification, and, if the maps are of sufficient quality, an atomic model.

3.24.8. Structure Refinement

The model refinement is the last step of the structure determination. The preliminary model is far from ideal. The refinement is done to optimize the agreement between the model and the observed X-ray data in order to improve the phases. Until a final model with accepted statistics is obtained, this interactive model construction was carried out repeatedly.

Several parameters can be modified to obtain the best model

1. PHENIX refinement needs a structure factor file (output.mtz) and a PDB file of the first atomic model created after solving the structure.
2. Crystallographic Information File (CIF) of the ligand is also needed as an input if the structure contains the ligand.
3. The refinement strategy uses rigid-body, real-space, and XYZ refinements. The refinement strategy also incorporates individual or collective B-factors.
4. A variety of restraints, including model restraints, secondary structure restraints, and Ramachandran restrictions, as well as a variety of weighting, may be enforced

$$R_{factor} = \frac{\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||}{\sum_{hkl} |F_o(hkl)|}$$

During refinement, a number of parameters, including atomic coordinates, atomic displacement parameters (ADPs), scale factors, and twinning, are fine-tuned until the best fit between an observed and predicted diffraction pattern is reached. The R_{factor}, which quantifies

the disparity between the observed structure factor amplitudes ($|F_o|$) and calculated structure factor amplitudes $|F_c|$, provides information about the accuracy of a model fit into the diffraction data.

The diffraction data is divided into two sets for cross validation: 1) a large working set which comprises approximately 90-95% of the data 2) a small test set, comprising the remaining of 10-5% of the data. The diffraction data present in the working set is used for refinement. It provides a more objective guide during model building and refinement process. The cross validation of the structural refinement is monitored by R_{free} which is calculated (in the same way as of R_{factor}) for the remaining 5-10% complementary test set. If the model is correct and errors are statistical, R_{free} is expected to be close to R_{factor}.

The most popular software applications for structural refinement are **PHENIX** (both maximum likelihood and simulated annealing) and **REFMAC** (maximum likelihood refinement) (Murshudov et al., 2011) (Vagin et al., 2004). In our study we have employed **PHENIX** to refine the structure. The PHENIX software suite refinement module is called phenix.refine (Adams et al., 2010). It offers a comprehensive set of tools for refinement across a wide range of resolution limits (subatomic to low) using X-ray, neutron or both types of data simultaneously. (https://www.osti.gov/pages/biblio/1625674).

3.24.9. Model building with coot (Crystallographic Object-Oriented Toolkit)

Crystallographic Object-Oriented Toolkit (COOT) was used for model building, model completion and validation. After every cycle of refinement, the model and the map are displayed in COOT (Emsley & Cowtan, 2004) for manual model building by manipulating the model using real space refinement, mutation, manual rotation and translation, rotamer fit, and many more. In COOT, the coordinate file and the structure factor file are read and the model building is done by inspecting two different types of maps. One map is the difference Fourier map (Fo-Fc) and other is Fourier (2Fo-Fc) map. These maps provide a comparison between the measured electron density of the crystal and the electron density explained by the current model, indicates the error in model and assist in improving the model. The difference Fourier is displayed for the regions where atoms need to be added (represented in green by default showing positive density) and for the regions where modelled atoms are fitted wrongly (represented in red by default showing negative density). In COOT, the model is altered by mutating or reorientating side chains, realigning backbone to fit it better in the electron density

map followed by refinement. The ligand atoms, water molecules and ions are added during the final stages of refinement.

3.25. Structure validation

The final model was validated using the program PROCHECK (Laskowski, MacArthur, Moss, & Thornton, 1993) which examines the overall and residue-by-residue geometry of a protein structure to determine its stereochemical quality. This program was used to assess the quality of the model after every refinement cycle. It validates the structure by analyzing the following parameters:

- Ramachandran plot
- Ramachandran plots for Glycine and Proline
- Main-chain parameters
- Side-chain parameters
- Main-chain bond length distributions
- Main-chain bond angle distributions

3.26. Deposition of the structure

The final structure of the protein is deposited in the Protein Data Bank (PDB). PDB is the large repository of structural data of biological macromolecules. It offers information on the 3D forms of proteins, nucleic acids and complex components. It includes data obtained by X-ray crystallography, nuclear magnetic resonance (NMR) spectrometry and cryo electron microscopy (cryo EM) submitted by biologists and biochemists from all over the world. The final structure was also validated using PDB validation tool (https://validate-rcsb-2.wwpdb.org/) of Research Collaboratory for Structural Bioinformatics (RCSB). The structure was submitted at RCSB PDB using the deposition server (https://deposit-1.wwpdb.org/).

3.27. Graphical representation and structure superimposition

PYMOL (2.5.2)- Schrodinger, 2022) was used for visualization of molecular structures and making publication quality images. It also helps to perform superimposition of two or more structures. The structures can be aligned by all atoms, c-alpha atoms or by the backbone atoms. It also calculates the root mean square deviation RMSD of the aligned structure.

3.28. Comparison of structure

To find structural homologs for a given protein, a search can be performed using PDBeFOLD (https://www.ebi.ac.uk/msd-srv/ssm/ssmcite.html). PDBeFOLD compares the already submitted structure in the PDB database with the query structure. The results provide PDB Ids and sequence alignments of the results along with Z-scores, root mean square deviations and number of Cα aligned.

3.29. Sequence Alignment

The sequence alignment was done using multalign web server ("Multiple sequence alignment with hierarchical clustering" (Corpet, 1988). The alignment figures with superimposed structure features were generated using ESPRIPT (Robert & Gouet, 2014).

Chapter4
Characterization of Rv3400

4.1. Introduction

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis which includes pulmonary tuberculosis (TB), gastrointestinal TB, liver TB, TB peritonitis, etc. It accounts for the death of 1-1.5 million people every year and thus known to be the second infectious leading cause of death worldwide (WHO report, 2020). The current therapy to cure the TB infection consists of multiple antibiotics that selectively targets the essential life processes of the bacteria that includes translation, transcription, replication etc. However, the fundamental cause of the ineffectiveness of current therapy regimens results in the global rising of resistance leading to MDR and XDR TB strains. Finding new antimicrobial therapeutic targets against Mtb with great potency and specificity is therefore urgently required. Identifying the genes that allow Mtb to thrive and persist in the host will aid in finding new drug targets. These genes should be exclusive (unique/essential/present) to pathogens and not to the host. For example, the enzymes of the riboflavin biosynthesis (RBS) pathway which are present in Mtb and not in humans, can be targeted against the pathogen for drug designing (Long et al., 2010). The biosynthesis pathway which synthesizes riboflavin, a sole precursor molecule for the synthesis of flavin cofactors Flavin mononucleotide (FMN) and Flavin adenine dinucleotide (FAD) is shown to be essential for the survival of Mtb (Fassbinder, Kist, & Bereswill, 2000). The pathway involved four enzymes for the formation of riboflavin in Mtb. The pathway begins with two enzymes GTP cyclohydrolase II (GCHII) and 3, 4-dihydroxy 2-butanone 4-phosphate synthase (DHBPS) utilizing GTP and ribulose-5-phosphate (Ru5P) as substrates, respectively. One of the intermediate steps in the riboflavin biosynthesis pathway is the dephosphorylation of 5-amino-6-ribitylamino-2,4(1*H*,3*H*) pyrimidinedione 5′-phosphate (ARPP) to 5-amino-6-ribitylamino-2,4(1*H*,3*H*) pyrimidinedione (ARP) by a putative phosphatase. However, the phosphatase that catalyzes the dephosphorylation of ARPP to ARP remain unknown in Mtb. Thus the aim of the current study is to identify the unknown phosphatase, which catalyse the conversion of ARPP to ARP of riboflavin biosynthesis in Mtb, and validate it both enzymatically and structurally.

Recently it has been reported that the dephosphorylation step in riboflavin biosynthesis is catalysed by the broad spectrum of hydrolases that belongs to the haloacid dehalogenase (HAD) family. Infact, the genes *yigB* and *ybjl* from *E.coli* (Haase et al., 2013), *ycsE, ywtE* and *yitU* from *Bacillus subtilis* (Sarge et al., 2015) and *AtPyrP2* and *AtGpp1/PyrP3* from *Arabidopsis thaliana* (Sa et al., 2016) are reported to catalyze the dephosphorylation of ARPP

to ARP of riboflavin biosynthesis pathway in their respective organisms. Utilizing these known phosphatases sequences as a query sequence in PSI-BLAST (Position specific iterated-BLAST), we predicted that Rv3400 or Rv3376 or Rv3813c could be a potential phosphatase that may catalyse the conversion of ARPP to ARP of riboflavin biosynthesis pathway in Mtb. In this chapter we have focused on the characterization of Rv3400.

In Mtb, Rv3400 is annotated as a hypothetical protein and is shown to be non-essential for *in vitro* growth of H37Rv (Minato et al., 2019) (DeJesus et al., 2017) (C. M. Sassetti et al., 2003) but required for growth in C57BL/6J mouse spleen, by transposon site hybridization (TraSH) (Christopher M. Sassetti & Rubin, 2003). Rv3400 also showed a positive interaction with the *mce*4 and thus possibly involved in the virulence. (Joshi et al., 2006). The positive interaction primarily occurs between genes of the same pathway and dependent on each other for their functions. *mce loci* of Mtb have been implicated as important determinants of virulence as it promotes the uptake of bacteria into nonphagocytic cells and thus designated as mycobacterial cell entry (Joshi et al., 2006). In the Tuberculist database, Rv3400 is annotated as hypothetical protein as this protein is not yet characterized experimentally, however, it has been predicted as a beta phosphoglucomutase based on its amino acid sequence. The Rv3400 also shares a weak sequence similarity with the phosphatases that are shown to be involved in riboflavin biosynthesis in other organisms. Notably, the phosphatase and mutase shares a common structural fold, similar catalytic sites and therefore sometimes it is difficult to distinguish between a phosphatase and mutase just based on their amino acid sequence only (Mendes, Maranha, Alarico, da Costa, & Empadinhas, 2011). Initial BLAST results suggested that Rv3400 show sequence similarity with beta phosphoglucomutase. However, we found that Rv3400 is also one of the homologs of the phosphatase involved in the riboflavin biosynthesis pathway in Mtb. Since Rv3400 is an uncharacterized protein and to rule out the ambiguity of either mutase or phosphatase, we have carried out the characterization of Rv3400 from Mtb. The biochemical and structural characterization indeed confirms that the Rv3400 encodes for a beta-phosphoglucomutase of Haloacid dehalogenase (HAD) family.

4.2. Experimental procedure

4.2.1. Bioinformatics analysis

To find the potential phosphatase that catalyze the conversion of ARPP to ARP in Mtb, the known phosphatase sequences (Haase et al., 2013), (Sarge et al., 2015) and default parameters against Mtb-H37Rv database. Similarly, the PSI-BLAST (Position Specific Iterative BLAST) with default parameters was also employed to identify the distantly related homologs of the potential phosphatase in Mtb. Based on sequence similarity and query coverage results from PSI-BLAST we predicted that the three genes Rv3400, Rv3376 and Rv3813c in Mtb may encode for potential phosphatase.

The Pfam database was also used for the identification of the protein family of Rv3400. A search was performed using Rv3400 protein sequence against the Pfam database. Multiple sequence alignment and Protein family analysis was done to find the family of proteins which belong to the Rv3400. For the prediction of the secondary structure of Rv3400, PSIPRED (https://bio.tools/psipred) server was used.

4.2.2. Cloning of Rv3400 constructs

The DNA encoding Rv3400 (789 bp) was PCR amplified using Mtb H37Rv genomic DNA as a template and designed primers (**Table1**). The primers were synthesized by Sigma-Aldrich, India. The amplified PCR product was purified using gel extraction kit (Thermo Fisher scientific Cat no-K0702) and digested with the *NheI* and *XhoI* restriction enzymes for 1 hr at 37°C. The digested PCR fragment was ligated into pETDuet-N vector (a modified pETDuet-1 vector where *BamHI* restriction site was mutated to in-frame *NheI* restriction site with N-terminal His_6-site), which was also digested with the same set of restriction enzymes. The ligation was carried out with T4 DNA ligase and the reaction was incubated at 22°C for 2-4 hrs. The ligated product was transformed in *E. coli* TOP10 cells (Novagen) followed by plating on nutrient agar supplemented with ampicillin. The plates were incubated for 16-18 hrs at 37°C or until colonies appeared. The partial colonies from the plates were carefully marked and checked for the plasmid harboring Rv3400 gene by colony PCR. The colonies showing the positive amplification in the PCR were inoculated in 5ml of LB broth supplemented with ampicillin as a selection marker. The primary culture was then kept in an incubator shaker overnight at 37°C at 200 RPM. Next day the cultures were harvested and plasmids were isolated using a commercial miniprep kit (GeneJet kit, Thermo Scientific, USA). The integration of the desired gene into the plasmid was further confirmed by double

digestion using *Nhe1* and *Xho1* restriction enzymes for 20 min at 37°C. The digested products were run on 1% agarose gel to check the fallout of desired size. The correctness of the sequences were verified by automated DNA sequencing. The resulting clone was named as pDuetN-Rv3400-FL.

In addition, several gene constructs were created to aid the purification and crystallization of Rv3400. The constructs Rv3400-Δ16 and Rv3400-Δ29 were made by deleting the disordered amino acids as predicted by PSIPRED server. For the construct Rv3400-Δ16, the first 16 amino acid residues and for Rv3400-Δ29, the first 29 residues were deleted. The Rv3400-Δ16 and Rv3400-Δ29 genes were cloned in pETDuet-N vector using the same protocol as that of full-length Rv3400 gene. The primers used for these constructs are listed in Table 1.

Table1-Table showing the list of primers and their nucleotide sequence used in the study.

S.No.	PRIMER NAME	NUCLEOTIDE SEQUENCE
1	Rv3400FL Nhe1_F	GATATAGCTAGCATGGCGAACTGGTATCGCCCGAAC
2	L17Rv3400Nhe_F	ATATAGCTAGCCTGGGTCTGCCCGAGAAGGTGCGTGC
3	L30Rv3400-Nhe_F	ATATATGCTAGCCTCGACGGTGTGCTCACCGATACCGC
4	Rv3400Xho1_R	ATATCTCGAGCTACAGCAGCTCGGCGAGATCG
5	Rv3400 L126M_F	GTTGCACAAGCTGATGCGCGACGATGGGG
6	Rv3400L197&202M_R	GTAACCCCCAACATTTCTGCCGCGCGCATGAAGGAGTCGGG
7	Rv3400 D29N_R	TCGGTGAGCACACCGTCGAGATTGAACAGGCAAGCACGCACCT
8	Rv3400 D31N_R	CGGTATCGGTGAGCACACCGTTGAGGTCGAACAGGCAAGCACGC
9	Rv3400_D29E_F	AGGTGCGTGCTTGCCTGTTCGAGCTCGACGGTGTGCTCACCGA
10	RV3400 D29E_R	TCGGTGAGCACACCGTCGAGCTCGAACAGGCAAGCACGCACCT
11	Rv3400_D31E_F	GCGTGCTTGCCTGTTCGACCTCGAGGGTGTGCTCACCGATACCG
12	Rv3400 D31E_ R	CGGTATCGGTGAGCACACCCTCGAGGTCGAACAGGCAAGCACGC
13	Rv3400_D29A_F	AGGTGCGTGCTTGCCTGTTCGCGCTCGACGGTGTGCTCACCGA
14	Rv3400 D29A_R	TCGGTGAGCACACCGTCGAGCGCGAACAGGCAAGCACGCACCT
15	Rv3400_D31A_F	GCGTGCTTGCCTGTTCGACCTCGCGGGTGTGCTCACCGATACCG
16	Rv3400 D31A_ R	CGGTATCGGTGAGCACACCCGCGAGGTCGAACAGGCAAGCACGC
17	MP_Fwd	GGTGCCGCGCGGCAGCCATATGGCTAGCATGATAAATCAGCGGTT
18	MP_Rev	CTCAGTGGTGGTGGTGGTGCTCGAGTCACTCATTAGGCGTTC

4.2.3. Site directed mutagenesis

To assist the crystal structure determination of Rv3400 by single wavelength anomalous diffraction (SAD) method, additional methionine residues were introduced in Rv3400-Δ16

construct. Two methionine mutant constructs Rv3400-Δ16-T1 and Rv3400-Δ16-T2 were generated based on the combined results of Multiple sequence alignment (MSA) (**Fig. 4.1**) and the PSI-blast based secondary structure PREDiction (PSIPRED) analysis of Rv3400. The amino acid Leucine was selected as it is the `safe' residue substitution of methionine due to similar non-polar side chain and similar hydrophobicity to methionine [difference ~0.5 kcal/mol (Finney, Gellatly, Golton, & Goodfellow, 1980; Guy, 1985). According to MSA analysis, the leucine residues where methionine was present in the Rv3400 homologs, were selected as site of mutation. From the secondary-structure prediction, mutated residues were selected in the helices as they can accumulate mutation without changing the structural stability of protein due to the higher numbers of inter-residue contacts (Abrusan & Marsh, 2016). For the construct Rv3400-Δ16-T1, three methionine residues were incorporated by replacing the leucine residue at 90, 197 and 202 positions whereas for Rv3400-Δ16-T2, the leucine residue at 126, 197 and 202 positions were replaced by methionine. The methionine mutants were made using site-directed mutagenesis method (Zheng, Baumann, & Reymond, 2004)The amplification was done by PCR using Phusion polymerase (Thermo Fischer Scientific, USA), a set of forward and reverse primers (**Table1)** and pETDuetN-Rv3400-Δ16 DNA as template. The PCR product was purified by commercially available PCR cleanup kit (GeneJet kit, Thermo Scientific, USA). The amplified and purified PCR product was incubated with *DpnI* at 37ºC for 3-4 hours to destroy the methylated template plasmid. The unmethylated DNA was transformed into *E. coli* Top10 competent cells and plated on nutrient agar supplemented with ampicillin. The plate was incubated for 14-16 hours or till the colonies appeared. A single colony was inoculated in 5 ml LB media and allowed to grow overnight 37 ºC for the isolation of the plasmid. Mutations at the required positions were confirmed by automated DNA sequencing.

In addition, the catalytic site mutants D29A, D29E, D29N, D31A, D31E and D31N of Rv3400 were also generated by using the same protocol as mentioned above. Mutations at the required positions were confirmed by DNA sequencing.

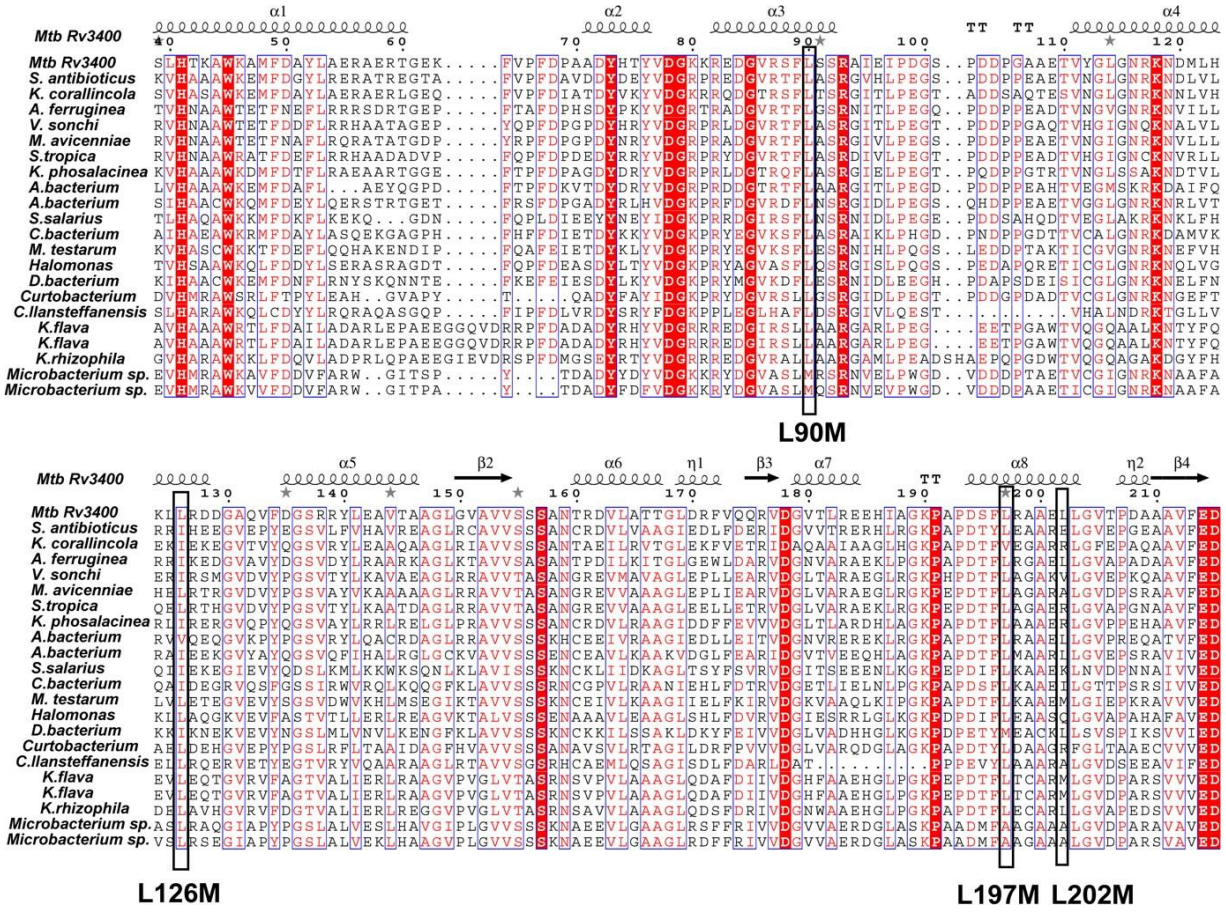


Fig. 4.1 Multiple sequence alignment of homologs of Rv3400. The alignment shows the residues selected for leucine to methionine point mutation at position 90, 126,197 and 202 and are boxed. Multiple sequence alignment was performed with the Multalign server and the figure was generated by ESpript 3.0 on web server (https://espript.ibcp.fr/ESPript/ESPript/) (Robert & Gouet, 2014).

4.2.4. Protein expression and purification

The pDuetN-Rv3400-FL plasmid containing *Rv3400* gene was transformed in *E. coli* Rosetta DE3 (Novagen) for its expression. The transformed *E. coli* cells were plated on the nutrient agar plates supplemented with ampicillin and chloramphenicol as selection markers and were incubated overnight at 37°C. The primary culture was obtained by innoculating a single colony in 10 ml of LB media containing ampicillin and chloramphenicol and was grown overnight at 37°C in an incubator shaker at 200 RPM. The 1 % of overnight grown culture was used to inoculate 1L of fresh LB media containing ampicillin and chloramphenicol as secondary culture in a 2L flask. The secondary culture was allowed to grow at 37°C in an incubator shaker with constant shaking at 200 RPM till the absorbance at 600 nm reached a value ~ 0.6-0.8. The expression of Rv3400 was induced by adding 0.3 mM isopropyl β-D-1-

thiogalactopyranoside (IPTG) and incubated further for 3 hrs at 37°C. The cells were harvested by centrifugation at 8,000xg for 10 min at 4°C. The cell pellet was resuspended in a lysis buffer containing 50 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol and 1mM Dithiothreitol (DTT). Also, 1mM phenylmethylsulfonyl fluoride (PMSF) was added along with 1 tablet of cocktail protease inhibitor (Roche, Applied Science, Mannheim, Germany). The lysis of the cells was done by sonication (Sonics, NewTown, CT USA) for 30-45 minutes at pulse rate of 8 seconds ON and 12 seconds OFF cycle with 25% amplitude. By centrifuging the lysate at 12,000 g for 45 min at 4°C, the cell debris were separated from the supernatant containing the soluble Rv3400 protein. The N-terminal His-tagged Rv3400 protein was purified using His-select Ni-NTA beads (Merk-Sigma Aldrich, USA). The supernatant was passed through the Ni-NTA column which was pre-equilibrated with the lysis buffer followed by washing the beads with the same buffer. The Rv3400 protein was eluted from the column with a lysis buffer containing 20-500 mM imidazole and each fraction of eluted protein was loaded on 15% SDS-PAGE to check the expression of Rv3400 and its purity. An Amicon concentrator, (Millipore, USA) with a 10 kDa cut off membrane was used to concentrate the Rv3400 protein. The concentration of protein was estimated by Denovix nanodrop spectrophotometer (DS-11 Series). Further, purification was done by using the gel filtration chromatography using Superdex^{TM} 210/300 GL (GE Healthcare) column that was equilibrated with the buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, 10% glycerol and 1mM DTT before injecting the protein. The expression and purification of the deletion constructs Rv3400-Δ16 and Rv3400-Δ23 were done by using the same protocol as that of full-length Rv3400 protein.

4.2.5. Preparation of selenomethionine-labelled protein

The plasmids containing genes *Rv3400-Δ16-T1* and *Rv3400-Δ16-T2* were transformed in Rosetta (DE3) cells (Novagen). The transformed cells were plated on nutrient agar supplemented with ampicillin and chloramphenicol and incubated at 37°C for 16 hours. A single colony was used to inoculate 10 ml LB media as primary culture and the cells were harvested in the log phase by centrifugation at 4000 RPM, for 5 min. The pellet obtained was resuspended in 5 ml of minimal media. The resuspended cells were used to inoculate the 1 L minimal media (secondary culture). The culture was incubated at 37 °C with shaking at 200 RPM. When absorbance at 600 nm reached ~ 0.6, all the amino acids -threonine 100 mg, phenylalanine 100 mg, lysine 100 mg, isoleucine 50 mg, valine 50 mg, and selenomethionine 60 mg were added. The culture was further incubated at 37°C with shaking at 200 RPM. After

15 min, the culture was induced with 0.3 mM IPTG and incubated at 16°C for 18 to 20 hrs. Harvesting of cells by centrifugation at 9000xg for 10 min was done and protein was purified using Ni-NTA chromatography followed by size exclusion chromatography similar to that of wild type Rv3400 protein.

4.2.6. Crystallization, structure determination and refinement

The purified Rv3400-FL (10 mg/ml) protein was used for crystallization trials in presence of commercially available screens from Hampton Research, USA and Molecular Dimensions, UK. 96-well sitting-drop high-throughput crystallization trays were set up by using a automated NT8 robotics system (Formulatrix Inc., USA). Each drop contained 150 nl of protein sample (10 mg/ml) and 150 nl of reservoir solution. The total reservoir volume was 40µl in the Swissci 96-well plate. The incubation of the crystallization trays was done at 20°C in a Rock Imager 1000 (Formulatrix) for storage with automatic pre-scheduled imaging. In the initial crystal screening, hits were obtained within a day in conditions containing 0.2 M Ammonium Sulphate, 0.05 M sodium acetate pH 4.5 and 30% (w/v) Polyethylene glycol (PEG) and 0.2 M Sodium acetate trihydrate, 0.1 M sodium cacodylate pH 6.5 and 30% (w/v) Polyethylene glycol (PEG) 8000. These initial conditions were further optimized by varying the pH, precipitant and molarity of the buffer used. Also, microbatch under oil plates (Hampton Research, USA) were used to improve the crystal quality.

The selenomethionine (selmet) labelled crystals of Rv3400-Δ16-T2 were obtained using the same protocol as for Rv3400-FL. The crystals appeared in almost all the conditions as mentioned above but the good diffraction quality crystals were obtained in a condition having 0.2 M Ammonium Sulphate, 0.05 M sodium acetate pH 4.5 and 30% PEG 8000 after 2 months. The selmet derivative crystals were diffracted and data was collected at 11.2 C beamline at ELETTRA synchrotron radiation source, Trieste Italy. The complete data set for the selmet labelled Rv3400-Δ16-T2 was collected at wavelength (0.97910Å) using DECTRIS PILATUS 6M detector. The data was collected under cryogenic conditions at 100 K. The data set of 360 images was collected by keeping the crystal to detector distance at 392.7 mm. The data set was collected with 0.5° oscillation and an exposure time of 0.5 sec per image. The X-ray diffraction data set was processed using XDS software package (Kabsch,2010). Using AutoSol module in the Phenix software suite, the crystal structure of the protein was determined by SAD method. The refinement of the model was done by several iterative cycles of model building by Coot and refinement using PHENIX.

4.2.7. Other computational tools

The amino acid sequence of Rv3400 (Accession number: NC_000962.3) was retrieved from GenBank (Altschul et al., 1997). Basic Local Alignment Search Tool (BLAST) programs was used to obtain homologs of Rv3400 from different databases. The sequence alignment was done using multalign web server (Corpet, 1988). The alignment figures with superimposed structure features were generated using ESPRIPT(Robert & Gouet, 2014). PSI-blast based secondary structure PREDiction (PSIPRED) was used to predict the secondary structure of proteins http://bioinf.cs.ucl.ac.uk/psipred/ (McGuffin, Bryson, & Jones, 2000). To find structural homologs for the query protein, a search was performed using PDBeFOLD (https://www.ebi.ac.uk/msd-srv/ssm/ssmcite.html). To generate the molecular graphic figures and to calculate r.m.s.d, we used PyMOL (The PyMOL) Molecular Graphics System, Version 2.5, Schrödinger, LLC). For modelling of conserved residues on the Rv3400 structure, the ConSurf server was used (Armon, Graur, & Ben-Tal, 2001)

4.2.8. ARPP Phosphatase assay

In riboflavin biosynthesis, 5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate (ARPP) is formed after the sequential action of GTP cyclohydrolaseII and Pyrimidine Deaminase/Reductase on GTP. It then goes through additional dephosphorylation step to form 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (ARP). The ARP compound is required by the lumazine synthase, the next enzyme in riboflavin synthesis for the formation of riboflavin. A recombinant GTP cyclohydrolase II (ribA) and pyrimidine deaminase/reductase (ribD) from *Helicobacter pylori* were used to produce ARPP via an enzymatic conversion from GTP in a three-step process (Richter et al., 1997). A 500 µl reaction mixture containing 10 mM Tris-HCl, pH 8.0, 8 mM $MgCl_2$, 3 mM dithiothreitol (DTT), 2 mM NADPH, 5 mM GTP, 100 µg *H. pylori* ribA, and 50 µg *V. cholera* ribD were used for the synthesis of ARPP. The reaction mixture was incubated at 37°C for 30-60 minutes. The proteins were removed by passing the reaction mixture through a centrifugal filter with a 3000 Da molecular weight cut-off. The ARPP thus formed was incubated with the Rv3400 for 1 hour and the ARP formed was detected through liquid chromatography-mass spectrometry (LC-MS).

Another strategy was also attempted where all the enzymes involved in riboflavin biosynthesis pathway were expressed and purified and used for the formation of riboflavin along with the potential phosphatase. Briefly, the GTP cyclohydrolase II (ribA) converts GTP

into 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone (DARP). The 3,4-Dihrdroxy-2-butanone4-phosphate synthase (ribB) is another starting enzyme of the pathway that utilize ribulose-5-phosphate (RUBP) as substrate and converts into L-3,4-Dihrdroxy-2-butanone4-phosphate (DHBP). The DARP is transformed into 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate (ARPP) by hydrolytic deamination of the pyrimidine ring and reduction of the ribose side chain by enzyme pyrimidine Deaminase/Reductase (ribD). The 5′-phosphate group of ARPP must be removed by a phosphatase in order to obtain 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (ARP) which is the committed substrate of lumazine synthase (ribH). It condenses ARP with Ru5P to form 6,7-Dimethyl-8-ribityl-lumazine (DMRL). Riboflavin synthase (ribE) catalyzes a dismutation reaction on two molecules of DMRL to form one molecule of riboflavin and ARP. 6,7-Dimethyl-8-ribityl-lumazine (DMRL) was formed which was converted to riboflavin by riboflavin synthase (ribE). For this strategy we expressed and purified all the enzymes involved in the pathway that has been previously cloned in the lab, except pyrimidine phosphatase. The purified enzymes ribA from *H. pylori* (Yadav & Karthikeyan, 2015) , ribB (Islam, Kumar, Singh, Salmon, & Karthikeyan, 2015), ribD, ribH and ribE (unpublished data from the thesis of Zeyaul Islam entitled Structure-Functionl analysis of enzymes involved in riboflavin biosynthesis pathway of *V. cholerae*) were incubated with the substrate GTP and Ribulose-5-phosphate along with the potential phosphatase. The reaction mixture contained 1mM GTP, 2mM Ru5P, 50mM Tris HCl, 10mM $MgCl_2$, 3mM DTT, 50µg each of ribA, ribB, ribD, ribH and ribE and 50 µg of Rv3400. The reaction was incubated at 37°C for 1-2 hours for the formation of riboflavin, a yellow-coloured compound that could further be verified spectrophotometrically at 440 nm (MacKintosh, C. (1993).

4.2.9. General phosphatase assay

To check the general phosphatase activity of Rv3400, para-Nitrophenyl Phosphate (pNPP) assay was. pNPP is a non-proteinaceous chromogenic substrate for alkaline and acid phosphatases. The hydrolysis of pNPP by phosphatase releases an inorganic phosphate and the conjugate base of para-nitrophenol (pNP) (J. Zhang, Lu, Lei, Hou, & Wu, 2017). The pNP is a yellow-coloured compound that has a maximum absorbance at 405 nm. pNPP (Thermofischer) was dissolved in a buffer containing 100mM Tris (pH 7-9), 100mM NaCl and 5mM $MgCl_2$. The Rv3400 protein at different concentrations (1-50 µM) was incubated with pNPP. The formation of yellow colour was monitored spectrophotometrically at 405 nm.

4.2.10. Pyrophosphatase assay

Pyrophosphatase activity of the proteins were determined using a sigma pyrophosphate assay kit (CAT-MAK168). For the assay, pyrophosphate concentration in the sample is determined by the use of fluorogenic pyrophosphate sensor (PPi sensor). The PPi sensor in presence of pyrophosphate results in the production of fluorescent product (λ_{ex}=316/λ_{em}=456 nm) proportional to the pyrophosphate present. The reaction was carried out in 96 well flat bottom black ELISA plates. Sodium pyrophosphate was used as a substrate and different concentrations of Rv3400 (1µM-5µM) was used. The reaction mixture consists of 100 mM Tris pH 9.5, 100 mM NaCl and 5mM $MgCl_2$ and Rv3400. The reaction was incubated for 30 minutes at 37°C. After incubation, 50 µl of reaction mixture was mixed with 50 µl master mix (assay buffer + pyrophosphate sensor stock solution, provided in the kit) and incubated further at room temperature for 10-30 minutes as per the manufacture's protocol. The pyrophosphate release in each reaction was measured at 316 nm and 456nm to check the pyrophosphatase activity of protein.

4.2.11. Beta-phosphoglucomutase (β-PGM) assay

The Beta-phosphoglucomutase (β-PGM) catalyzes the conversion of β-D-Glucose 1-phosphate (β-G1P) to β-D-Glucose 6-phosphate (βG6P). The substrate (β-G1P) was synthesized enzymatically by the action of Maltose phosphorylase (MP) (EC.2.4.1.8) on maltose (Nilsson & Radstrom, 2001). The gene encoding maltose phosphorylase from *Bacillus subtilis* was cloned in pET 28c vector and expressed in BL21(DE3). The MP protein was purified by Ni-NTA followed by gel filtration chromatography. To synthesize β-G1P, the purified MP was incubated with maltose (0.2 M) overnight at 30°C with 200 µg of the enzyme. (Johnson et al., 2018). The formation of β-G1P was confirmed by Liquid chromatography–mass spectrometry (LC–MS) where peak corresponding to the β-D-Glucose 1-phosphate was observed at 259 Da. The concentration of β-D-Glucose 1-phosphate formation is equivalent to the β-D-Glucose in the reaction. To estimate β-D-Glucose in the reaction, the YSI 2900 analyzer was used to measure the concentration of glucose.

To check the β-PGM activity, the protein Rv3400 was incubated with the reaction mixture containing substrate. The reaction mixture contained 50 mM HEPES (pH 7.5), 450 µM β-D-Glucose 1-phosphate, 2 mM $MgCl_2$, 30 µM α-glucose 1,6-bisphosphate, 0.4 mM NADP, 15U of glucose 6-phosphate dehydrogenase (G6PD) (HIMEDIA, India). The reaction was started by adding 50 µg of the protein. It was a coupled reaction in which product β-G6P

was catalyzed by the enzyme G6PD and led to the reduction of NADP. Thus, the increase in the absorbance of NADPH which is directly proportional to the formation of product β6P was monitored at 340 nm for 15-30 minutes using UV–Visual spectrophotometer (Perkin Elmer Lambda 25, USA).

The kinetic constants V_{max} and K_m were determined by monitoring the β-PGM activity of Rv3400 at varying concentration of substrate β-D-Glucose 1-phosphate (0-750 μM). The reaction rate was calculated using the extinction coefficient for NADPH (76.22 l/mmol/cm). All the kinetic parameters were determined by non-linear regression method to fit the data to Michaelis-Menton equation using program ORIGIN 8.5. All the experiments were done in triplicate for each concentration of substrate.

4.2.12. Effect of temperature and pH on catalytic activity of Rv3400

To study the influence of pH, temperature and different metal ions on β-PGM activity of Rv3400, enzymatic reactions were carried out using the same components as well as the same method as described above. The relative activity of the enzyme for β-D-Glucose 1-phosphate was examined at various pH values ranging between 5.0 to 9.0 using suitable buffers. Citrate buffer of pH 5 and 5.5, sodium phosphate buffer of pH 6 and 6.5, HEPES buffer of pH 7 and 7.5, and TRIS buffer of pH 8, 8.5 and 9 were used. The effect of temperature on the enzyme activity was determined at various temperatures ranging from 20 to 65°C. Similarly, the effects of different divalent metal ions such as Mn^{2+}, Fe^{2+}, Co^{2+}, Zn^{2+}, Ni^{2+} Ca^{2+}, Ba^{2+} and Cd^{2+} on the Rv3400 activity were studied.

4.2.13. Circular dichroism studies

The circular dichroism (CD) studies of Rv3400 protein were done to check the folding and secondary structure content of the proteins. The far-UV CD spectra were collected on a Jasco–810 CD spectropolarimeter that was flushed with nitrogen at the rate of 9–12 L/min. A quartz cuvette with a path length of 1 mm was used to obtain the data, which ranged from 250 nm to 198 nm. Each obtained spectra consisted of an average of five scans and was recorded as raw ellipticity. Data collection for all protein samples utilised a concentration of 0.2 mg/ml. The mean residue ellipticity (MRE) was calculated by raw ellipticity according to the equation,

$$[\theta]_{MRE} = ([\theta] \text{ X } 100 \text{ X } M_r) / (c \text{ x } l \text{ x } N_A)$$

where $[\theta]_{MRE}$ depicts the mean residue ellipticity, [θ] stands for the raw ellipticity in degrees, Mr is molecular weight of the protein (in Da), l is the path length of the cuvette (in cm), c is the concentration of the protein (in mg/ml), and the number of residues of the protein is depicted by $N_{A.}$

4.3. Results

4.3.1. Bioinformatics analysis of Rv3400

In order to find unknown phosphatase of riboflavin biosynthesis pathway in Mtb, the phosphatases that are known to involve in riboflavin biosynthesis pathway of other organisms were employed as query sequence in Basic Local Alignment Search Tool (BLAST). The BLAST results depicted very less similarity (6-10%) or no similarity between the reported phosphatase and the corresponding homologs in Mtb. So, PSI-BLAST (Position specific Iterative BLAST) was used to pick any distant related homologs in Mtb (**Table 2**). Based on the analysis of PSI-BLAST results, three genes Rv3400, Rv3376 and Rv3813c, which showed same sequence similarity and query coverage were selected for further studies. The analysis also found a homolog- Rv1225c, but it has very less sequence similarity and query coverage in comparison to the other three genes, rendering it less significant for further studies (**Table 3)**. The BLAST search using the amino acid sequence of Rv3400 revealed that it is conserved throughout the *Mycobacterium* species and shares about 98% sequence identity among them. The BLAST results also suggested that it shows high similarity with Haloacid dehalogenase family (HAD) and more specifically with beta-phosphoglucomutase. Similarly, multiple sequence alignment (MSA) (**Fig. 4.2**) and Pfam (protein family also predicted Rv3400 to be the member of haloacid dehalogenase family proteins as they contain the conserved HAD signature motifs. Since Rv3400 has shown similarity with both phosphatase and beta-phosphoglucomutase, the biochemical and structural characterization will help to find the actual function of the protein.

Table 2- Table showing the reported phosphatases of RBS pathway and their homologs in the Mtb with percent identity and query coverage.

S. No.	Gene (Reported phosphatases)	Potential Homologs in *Mycobacterium tuberculosis*	Sequence Similarity (%)	Query coverage (%)
1.	*YigB (E. coli)*	*Rv1225c*	29	32
2.	*Ybjl (E. coli)*	*Rv3813c*	27	94
3.	*YcsE (B. subtilis)*	*Rv3813c*	27	94
4.	*YwtE (B. subtilis)*	*Rv3813c*	27	98
5.	*YitU (B. subtilis)*	*Rv3813c*	29	98
6.	*AtPyrP1 (A. thaliana)*	*Rv3376*	33	27
7.	*AtPyrP2(A. thaliana)*	*Rv3376*	27	39
8.	*AtGpp1/PyrP3(A. thaliana)*	*Rv3400*	24	63
6.	*AtPyrP1 (A. thaliana)*	*Rv3376*	33	27
7.	*AtPyrP2(A. thaliana)*	*Rv3376*	27	39
8.	*AtGpp1/PyrP3(A. thaliana)*	*Rv3400*	24	63

Table 3- Table showing the selected potential phosphatases in Mtb based on high sequence similarity and query coverage.

S. No.	Homologs in *Mycobacterium tuberculosis*	Predicted function	Sequence similarity (%)	Query coverage (%)
1.	Rv3813c	Cof-type HAD-IIB family hydrolase	27	94
2.	Rv3376	HAD family phosphatase	27	39
3.	Rv3400	Beta-phosphoglucomutase family hydrolase	24	63

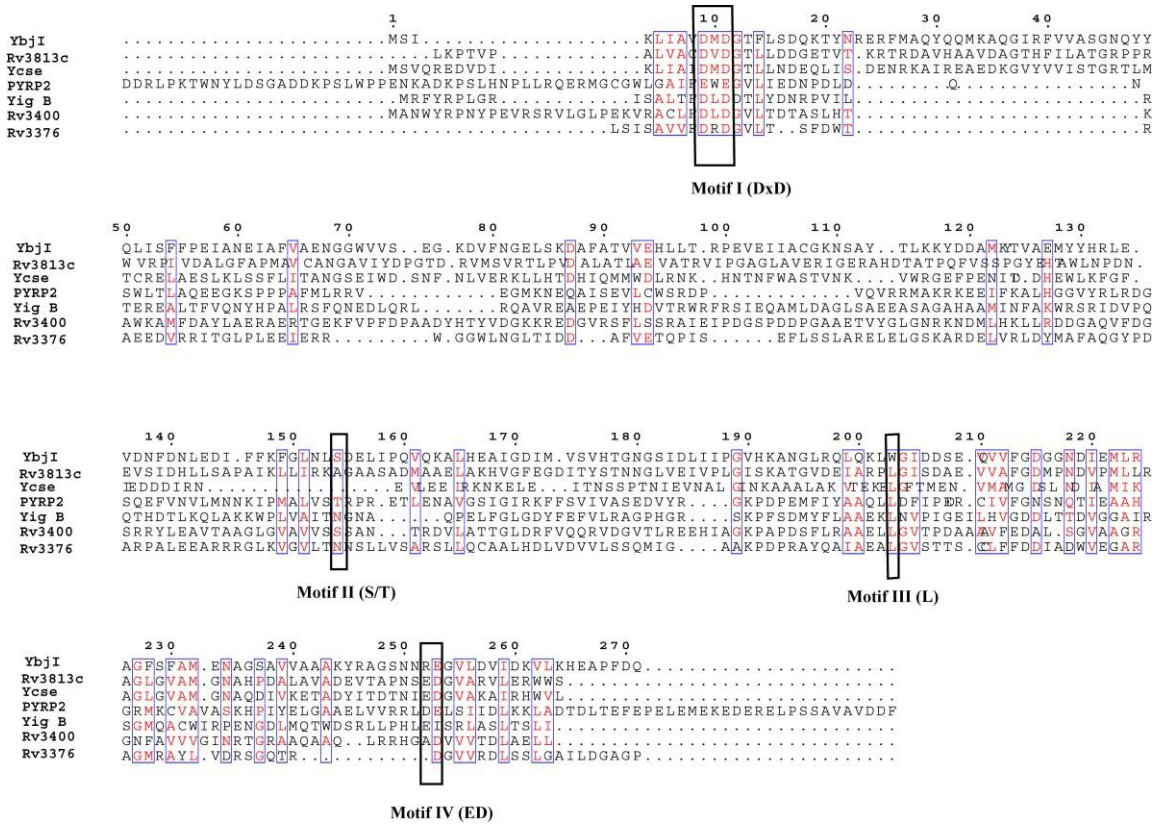


Fig. 4.2 Multiple sequence alignment of known and potential phosphatases of riboflavin biosynthesis. The alignment shows the conserved structural motif I, II, III and IV of HAD. Conserved residues are shown in red on white background and identical residues are displayed in white on red background. The amino acid residues conserved between the groups are boxed. Multiple sequence alignment was performed with the Multalign server and the figure was generated by ESpript 3.0 on web server (https://espript.ibcp.fr/ESPript/ESPript/) (Robert & Gouet, 2014).

4.3.2. Cloning of Rv3400 gene

The Rv3400 gene (789 base pairs) encoding for probable hydrolase (Mycobrowser, (Kapopoulou, Lew, & Cole, 2011) of *Mycobacterium tuberculosis* was amplified using the genomic DNA of Mtb H37Rv with standard polymerase chain reaction (PCR) protocol. The amplified gene was cloned in the pETDuetN vector with 6x His Tag at the N-terminal end of Rv3400. The integration of Rv3400 was confirmed by double digestion with *NheI* and *XhoI* restriction enzymes and confirmed the correct sequence by automated DNA sequencing (**Fig. 4.3.**). As described in the material and methods of this chapter, in addition to full length constructs several deletion constructs were engineered (**Table 4**) to aid in getting quality crystals in terms of diffraction. Since, there are only two methionines in the native Rv3400 protein, additional methionine residues were incorporated in the protein by mutating three

leucine residues in the construct pET DuetN-Rv3400- Δ16-T1 (L90,197 and 202 to M) and pET DuetN-Rv3400-Δ16-T2 (L126,197 and 202 to M) (**Table 4**). All these constructs were confirmed by automated DNA sequencing.

Table 4- Table showing the list of constructs and their description.

S. No.	Construct	Description
1.	pET DuetN-Rv3400-FL	N-term His-tag Full length Protein (262 aa)
2.	pET DuetN-Rv3400-Δ16	N-term His-tag protein where first 16 amino acids were deleted (17-262 aa)
3.	pET DuetN-Rv3400-Δ29	N-term His-tag protein where first 29 amino acids were deleted (30-262 aa)
4.	pET DuetN-Rv3400-Δ16-T1	N-term His-tag protein where first 16 amino acids were deleted and L90,197 and 202 replaced with M (17-262 aa)
5.	pET DuetN-Rv3400-Δ16-T2	N-term-His-tag protein where first 16 amino acids were deleted and L126,197 and 202 replaced with M (17-262 aa)

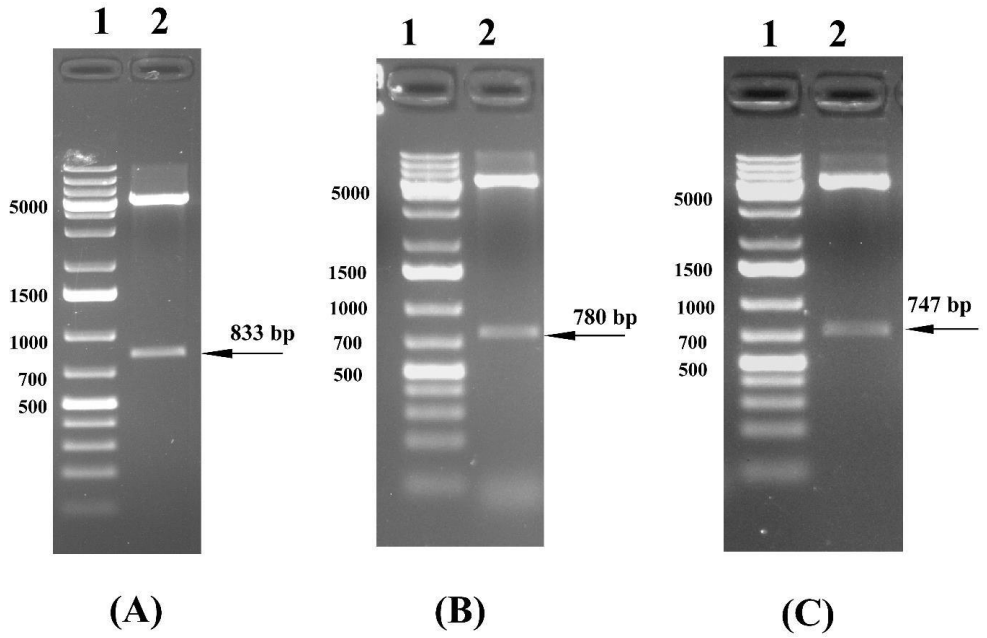


Fig. 4.3. Cloning of Rv3400-FL, Rv3400-Δ16 and Rv3400-Δ29. **(A)** 1% Agarose gel showing the *NheI* and *XhoI* digested pETDuetN-Rv3400-FL and expected release of 833 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv3400-FL from pETDuetN vector. (B) 1% Agarose gel showing the *NheI* and *XhoI* digested pETDuetN-Rv3400-Δ16 and expected release of 780 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv3400-Δ16 from pETDuetN

vector. **(C)** 1% Agarose gel showing the *NheI* and *XhoI* digested pETDuetN-Rv3400-Δ29 and expected release of 747 bp Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv3400-Δ29 from pETDuetN vector.

4.3.3. Expression and purification of Rv3400

The plasmid containing the gene *Rv3400* (pET DuetN-Rv3400-FL) was transformed in *E. coli* Rosetta DE3 (Novagen) expression cells and induced with 0.3 mM IPTG. The overexpressed protein was purified using Ni-NTA affinity chromatography followed by size exclusion chromatography. The purity was assessed by loading the fractions on 15% SDS-PAGE. The size of the purified proteins in SDS-PAGE matched well with the estimated molecular weight. The deletion constructs Rv3400-Δ16 and Rv3400-Δ29 were also expressed and purified using the same protocol followed by the wild type (**Fig.4.4).** Se-Met labelling of constructs Rv3400-Δ16-T1 and Rv3400-Δ16-T2 was done (detailed in materials and methods) and purified using Ni-NTA and size exclusion chromatography (**Fig. 4.5).**

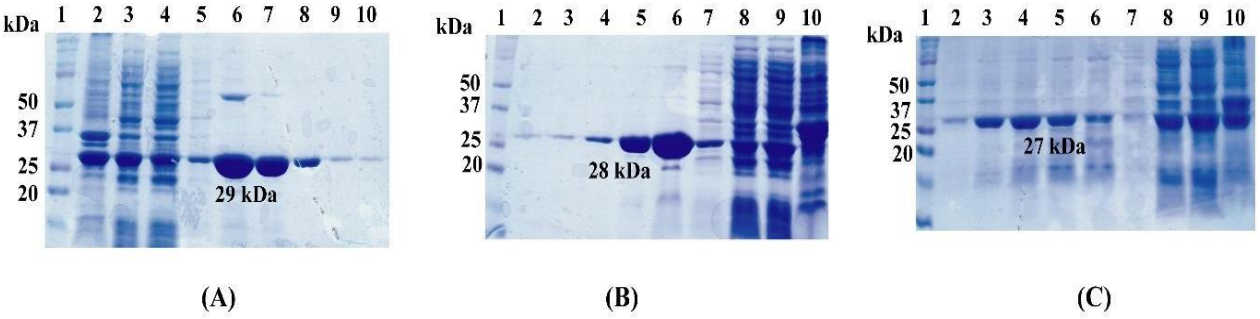
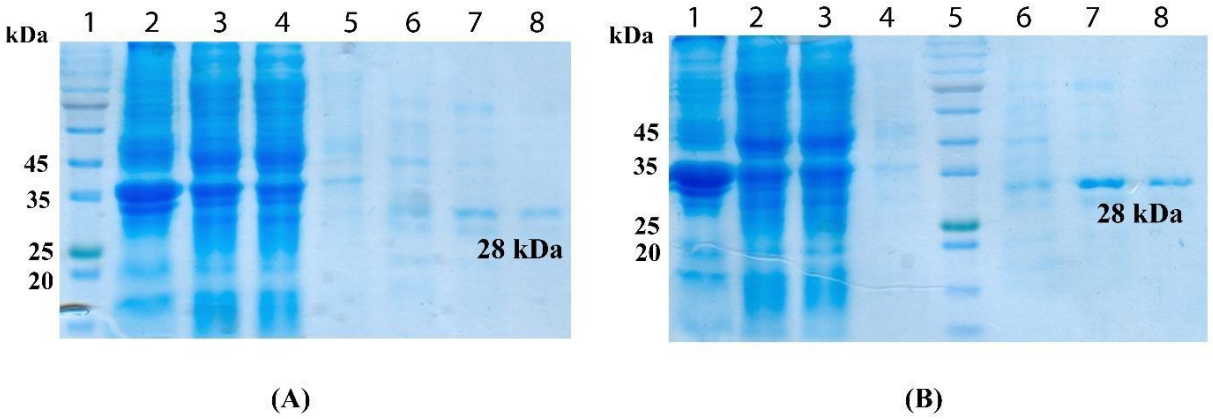


Fig. 4.4 Ni-NTA purification elution profile of Rv3400-FL, Rv3400-Δ16 and Rv3400-Δ29. (A)15% SDS-PAGE showing purification profile of Rv3400-FL using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Pellet, Lane 3: Supernatant, Lane 4: Flowthrough, Lane 5: Wash, Lane 6: Elution 1, Lane 7: Elution 2, Lane 8: Elution 3, Lane 9: Elution 4 and Lane 10: Elution 5. (B) 15% SDS-PAGE showing purification profile of Rv3400-Δ16 using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Elution 5, Lane 3: Elution 4, Lane 4: Elution 3, Lane 5: Elution 2 and Lane 6: Elution 1, Lane 7: Wash, Lane 8: Flowthrough, Lane 9: Supernatant, and Lane 10: Pellet. (C) 15% SDS-PAGE showing purification profile of Rv3400-Δ29 using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Elution 5, Lane 3: Elution 4, Lane 4: Elution 3, Lane 5: Elution 2 and Lane 6: Elution 1, Lane 7: Wash, Lane 8: Flowthrough, Lane 9: Supernatant, and Lane 10: Pellet.



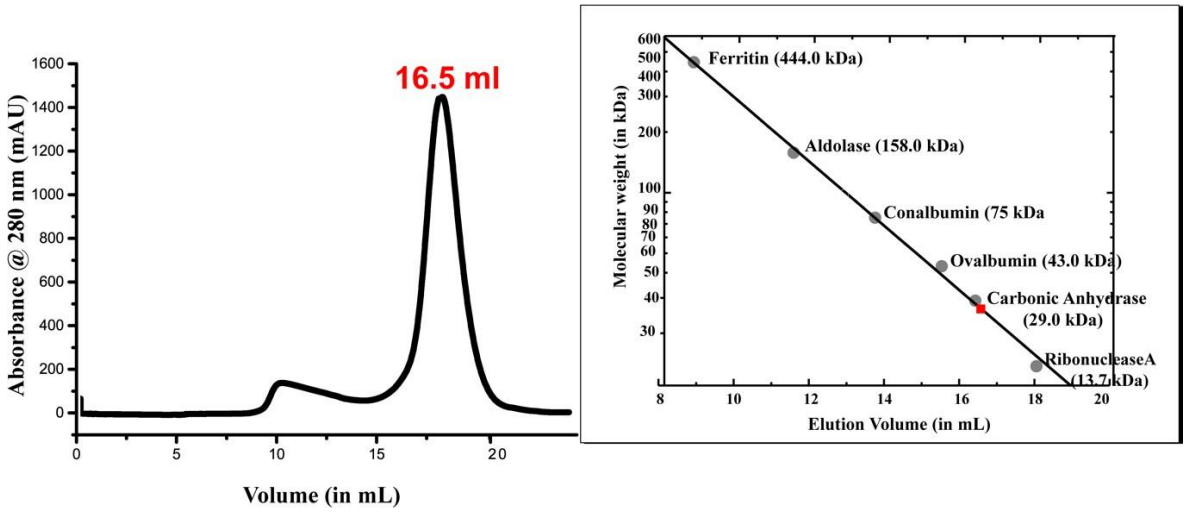
(A) (B)

Fig. 4.5 Ni-NTA purification elution profile of Selenomethionine labelled proteins Rv3400-Δ16-T1 and Rv3400-Δ16-T2.

(A) 15% SDS-PAGE showing purification profile of Rv3400-Δ16-T1 using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Pellet, Lane 3: Supernatant, Lane 4: Flowthrough, Lane 5: Wash, Lane 6: Elution 1, Lane 7: Elution 2, and Lane 8: Elution 3. (B) 15% SDS-PAGE showing purification profile of Rv3400-Δ16-T2 using Ni-NTA chromatography. Lane1: Pellet, Lane 2: Supernatant, Lane 3: Flowthrough, Lane 4: Wash, Lane 5: Molecular weight marker (Biorad), Lane 6: Elution 1, Lane 7: Elution 2, and Lane 8: Elution 3.

4.3.4. Determination of the oligomeric state of Rv3400

The further purification and oligomeric state of Rv3400 was determined using the prepacked analytical size-exclusion chromatography Superdex-200 column (GE Healthcare, USA).



(A) (B)

Fig. 4.6 Oligomeric state of Rv3400: Analytical size exclusion chromatography profile of Rv3400 (ASEC). The peak eluted at ~16.5 ml corresponds to 29 kDa and matches well with the calculated molecular weight of Rv3400 (29 kDa). The ASEC experiments suggested that the Rv3400 exists as monomer in solution. The inset shows the standard calibration curve for Superdex-200 column (GE Healthcare, USA).

The purified protein from Ni-NTA chromatography was injected into the superdex-200 column and the elution profile was compared with the elution profile of standard proteins (**Fig. 4.6).** The results after comparison with standard indicates that the Rv3400 exists as a monomer in solution. Similarly all the variants of Rv3400 were subjected to the gel filtration chromatography for the additional purification and also eluted as monomer as that of full-length.

4.3.5. MALDI-TOF studies of Rv3400

The intact molecular mass of Rv3400 was determined by MALDI-TOF (Matrix Assisted Laser Desorption Ionization-Time of Flight) mass spectrometry. The concentration of protein sample was 0.2 mg/ml and sinapinic acid was used as a matrix. The calculated mass of Rv3400-FL expressed with 14 extra amino acids (MGSSHHHHHHSQAS) is 29511 Da, after removal of formyl methionine. The observed molecular weight of Rv3400 in MALDI experiment was comparable to the calculated molecular weight (**Fig. 4.7).**

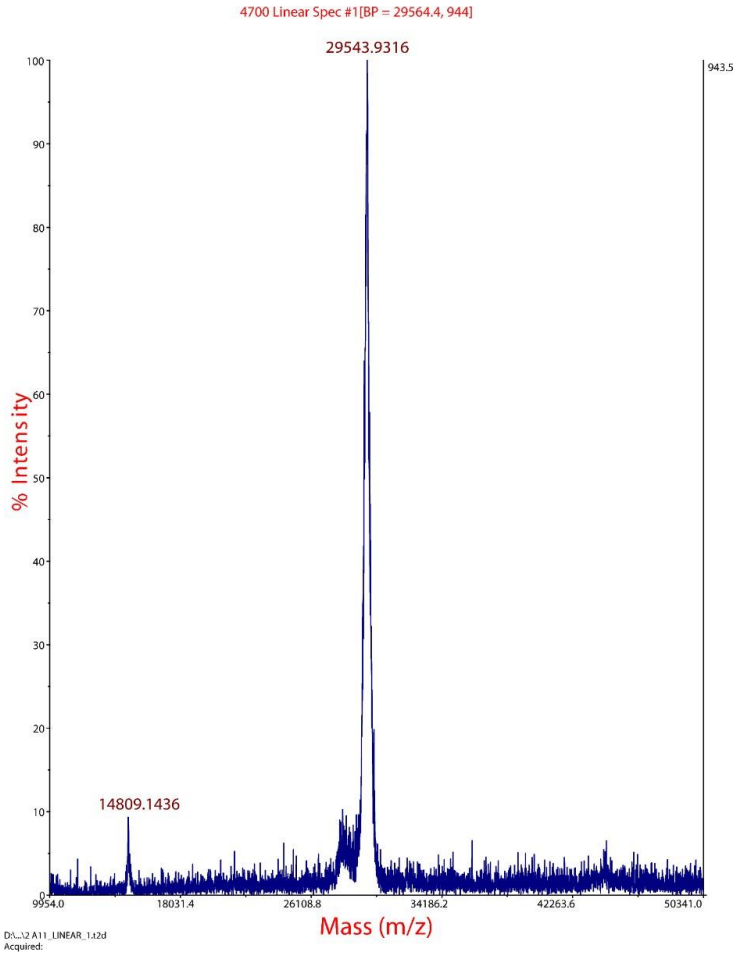


Fig. 4.7 MALDI-TOF analysis of Rv3400. The major peak observed at ~29543.9 Da corresponds to Rv3400.

4.3.6. Rv3400 lacks phosphatase activity

Para-Nitrophenyl Phosphate (pNPP), a common substrate, was utilised to test the phosphatase activity of Rv3400. Dephosphorylation of the pNPP yields pNP, a yellow-coloured compound that can be measured spectrophotometrically at 405 nm. Alkaline phosphatase was used as the positive control in the assay which shows an increase in the absorbance at 405 nm due to the formation of pNP (**Fig. 4.8).** However, for Rv3400, no absorbance was observed at405 nm. The results indicated that Rv3400 was unable to catalyze the dephosphorylation of pNPP and thus lacks the phosphatase activity.

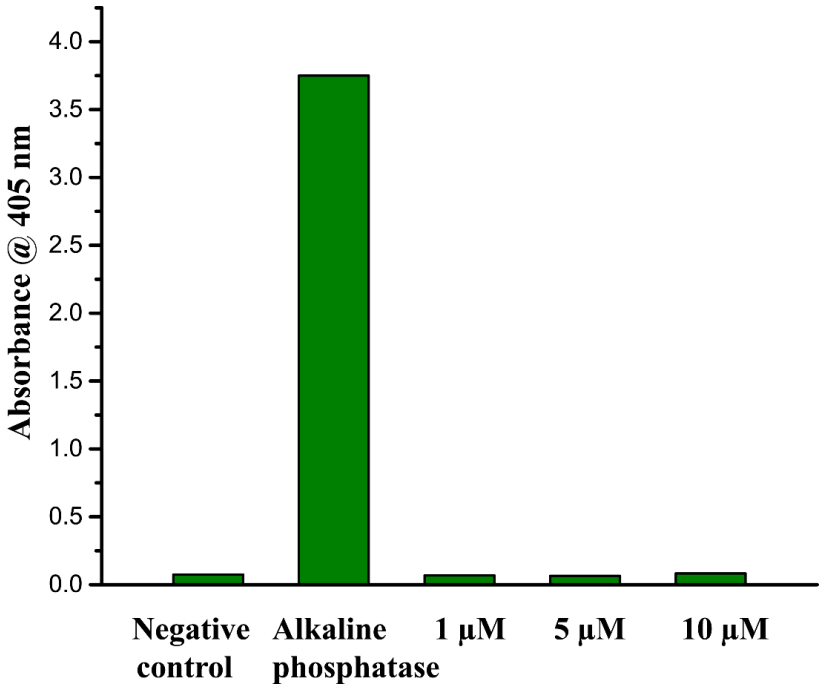
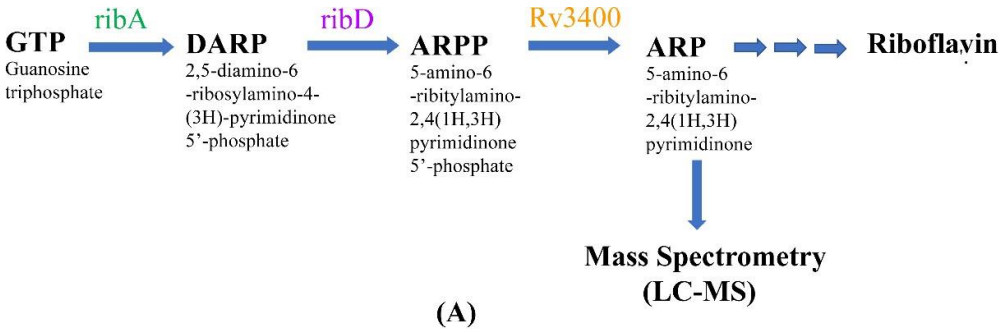


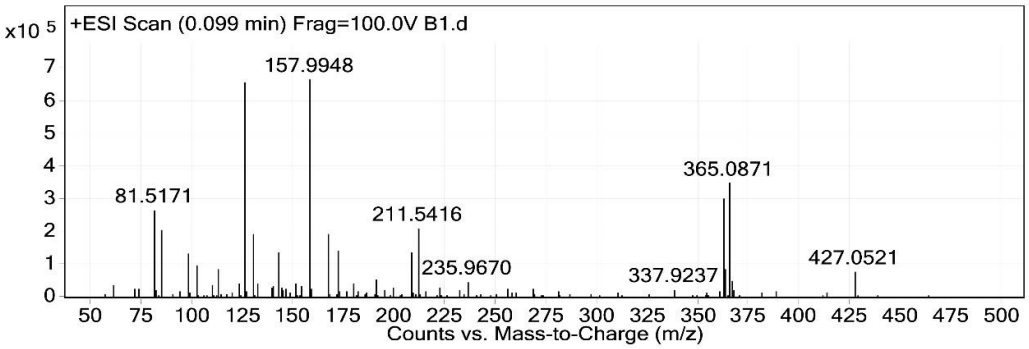
Fig. 4.8 Testing phosphatase activity in Rv3400. Phosphatase activity of Rv3400 using general substrate para-Nitrophenyl Phosphate (pNPP). Negative control is the reaction mixture containing buffer and substrates except protein, AP-Alkaline phosphatase, a known phosphatase used as positive control, and different concentrations of Rv3400 ranges from 1µM, 5 µM and 10 µM were used. The assay was done only once to check the phosphatase activity of Rv3400. (The experiment was done only once to check the phosphatase activity of Rv3400).

As discussed previously, the catalysis of ARPP to ARP in riboflavin biosynthesis is due to the action of a yet unknown phosphatase in Mtb. To find whether Rv3400 can play a role in the conversion of ARPP to ARP, two strategies were employed. In the first strategy, the required ARPP was produced using the substrate GTP and enzymes ribA and ribD. The formation of ARPP was confirmed by spectrophotometric assay (Sa et al., 2016). The enzymatically produced ARPP was then incubated with Rv3400 and the peak corresponding ARP (277Da) was monitored using liquid chromatography-mass spectrometry (LC-MS). The analysis of

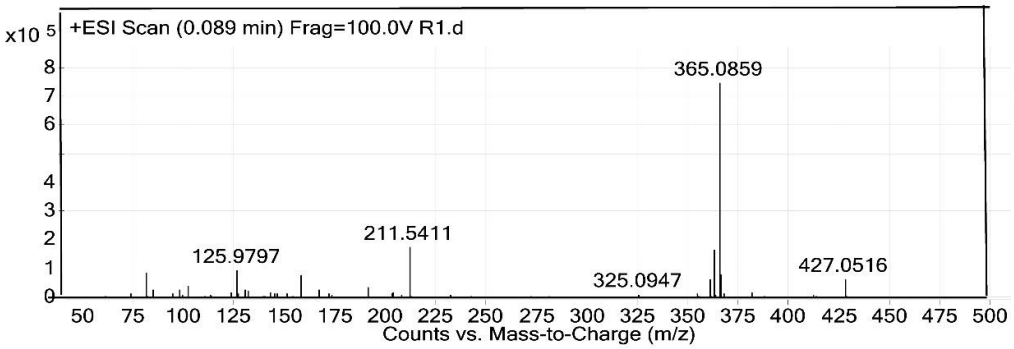
LC-MS data revealed that there was no peak observed that corresponds the molecular weight of ARP (**Fig. 4.9).**



(A)



(B)



(C)

Fig. 4.9 Role of Rv3400 in riboflavin biosynthesis pathway. (A) Schematic diagram for Strategy I depicting the workflow for the formation of ARPP from GTP and dephosphorylation by the potential phosphatase to form ARP. **(B)** LC-MS analysis revealed that no peak corresponding to molecular weight of ARP (~277 Da) was observed when ARPP was incubated with protein Rv3400.

In strategy II, all the enzymes of riboflavin biosynthesis pathway (ribA, ribB, ribD, ribH and ribE) were purified and incubated with the substrate GTP and Ru5P along with Rv3400. The

reaction was monitored for the formation of the end product i.e., riboflavin, a yellow-coloured compound which can be monitored spectrophotometrically at 440nm. However, no riboflavin formation was observed suggesting that Rv3400 is unable to catalyze the conversion of ARPP to ARP and thus lacks the dephosphorylation activity (**Fig. 4.10).** These results indicated that Rv3400 is not a phosphatase and probably have some other function.

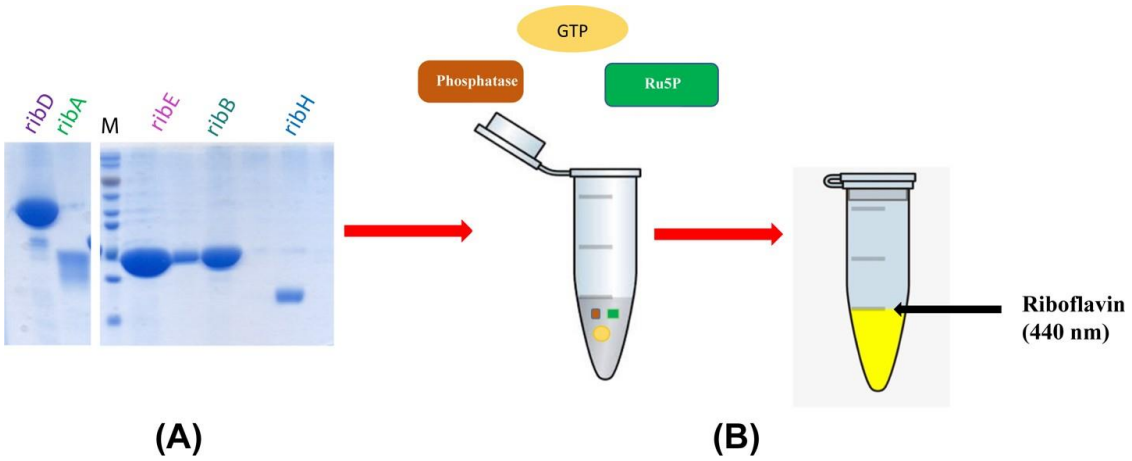


Fig. 4.10 Role of Rv3400 in riboflavin biosynthesis pathway. Strategy II depicting the workflow. **(A)** All the enzymes of RBS pathway were purified by Ni-NTA chromatography and GFC. **(B)** The enzymes were incubated with the substrates GTP and Ru5P and Rv3400 enzyme.

4.3.7. Crystallization, data collection and refinement of Rv3400

Initially, crystallization attempts for Rv3400-FL protein was performed using various commercial screens available from Molecular Dimensions and Hampton research. Very fine needle shaped crystals (**Fig. 4.11, A)** appeared by next day in the condition containing 0.2 M Ammonium Sulphate, 0.1 M sodium acetate pH 4.6 and 30% PEG 4000, however it did not diffract well in the X-ray beam. The optimization to improve quality of the crystals by varying concentration of protein, incubation temperature, different crystallization conditions and construct designing was done. One construct Rv3400-Δ16 was made which yielded rectangular shaped crystals that appeared after two months in the condition containing 0.2 M Ammonium Sulphate, 0.05 M sodium acetate pH 4.5 and 30% PEG 8000. (**Fig. 4.11, B**) and also diffracted well (**Fig. 4.11, D**). The sequence similarity search using BLAST (Altschul et al., 1997) with Rv3400 amino acid sequence against the Protein Data Bank (PDB) revealed no significant hit (found only hits with maximum 31% identity). So it was uncertain to solve the structure of Rv3400 by molecular replacement method using the existing models that are available in the Protein Data Bank (PDB) due to poor sequence similarity. Therefore, we chose to determine the crystal structure of Rv3400 by SAD phasing using selenium-methionine

(Semet) labelled protein. We incorporated additional 3 methionine residues replacing Leucine at 97, 197 and 202 position making Rv3400-Δ16-T1 construct and Leucine at 126, 197 and 202 positions for Rv3400-Δ16-T2 construct (detailed in materials and method of this chapter) as the native sequence had only 2 methionines.

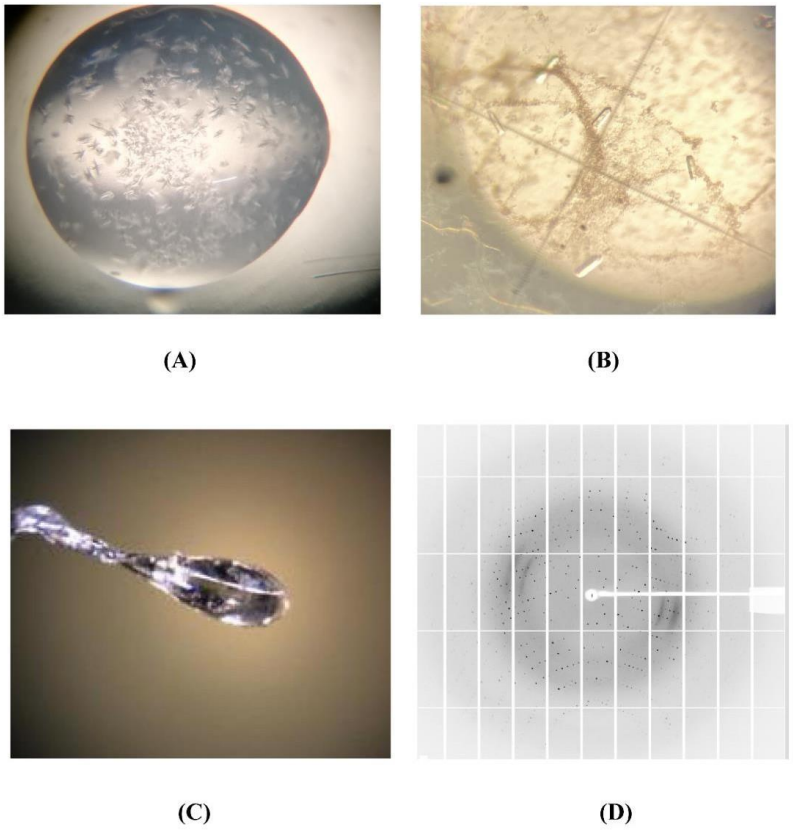


Fig. 4.11. Crystals and diffraction pattern of Rv3400 (A) Crystals of Rv3400-FL, **(B)** Crystals of Rv3400- Δ16-T2 **(C)** Mounted crystal of Rv3400- Δ16-T2 **(D)** Diffraction pattern of Rv3400-Δ16-T2**.**

The Rv3400-Δ16-T1 yielded rectangular shaped crystals that appeared after two months in the condition containing 0.2 M Ammonium Sulphate, 0.05 M sodium acetate pH 4.5 and 30% PEG 8000. (**Fig. 4.11, B**). The X-ray diffraction dataset for Rv3400-Δ16-T2 construct was collected at 11.2C beamline at ELETTRA synchrotron radiation source, Trieste Italy. The data obtained was further processed that involves indexing, integration and scaling using XDS software package (Kabsch, 2010). The structure of Rv3400-Δ16-T2 was solved by SAD method using AUTOSOL module as implemented in PHENIX suite of programs (Adams et al., 2010). The model thus obtained for Rv3400-Δ16-T2 was further refined using PHENIX. Notably, an unintentional mutation of a histidine residue at position 123 was mutated to leucine during the cloning of Rv3400-Δ16-T2 construct. Therefore, the residue at 123 position in the crystal structure was modelled as leucine.

Table 5: Data collection and refinement statistics

	Se-Rv3400
Data Collection	
Wavelength (Å)	0.97910
No. of crystals used	1
Resolution range (Å)	41.92 – 1.70 (1.73-1.70)
Space Group	$C222_1$
Unit cell parameters	
a, b, c (Å)	97.90, 106.38, 81.20
α, β, γ (°)	90°, 90°, 90°
Total number of reflections	498700
No. of unique reflections	85784
No. of unique reflections (non-anomalous)	44752
Average Mosaicity (°)	0.09
Redundancy	11.1 (3.5)[a]
Average $I/\sigma I$	30.3 (1.6)
Completeness (%)	95.4 (66.2)
R_{merge} (%)[b]	5.7 (67.6)
CC1/2	1.00 (0.65)
Refinement statistics	
Resolution range (Å)	33.34 – 1.70
No. of reflections used in the refinement	85775
R_{cryst} (%)[c]	15.4
R_{free} (%)[d]	17.3
RMSD[e]	
Bond lengths (Å)	0.006
Bond angles (°)	0.758
Ramachandran plot statistics	
Most favoured (%)	98.78
Allowed regions (%)	1.22
No. of protein atoms	1922
No. of solvent atoms	310
Others	12
Cl atoms	2
Glycerol atoms	6
1,2-Ethanediol atoms	4
Wilson B-factor (Å)	19.5
Average B-factor ($Å^2$)	
Protein atoms	26.22
Solvent atoms	38.02
Others	35.46
PDB CODE	8H5S

[a]Values for the last shell are in parentheses
[b]$R_{merge}=\sum_{hkl}\sum_i | I_i(hkl) - I(hkl) | / \sum_{hkl}\sum_i I_i(hkl)$ where $I(hkl)$ is the intensity of reflection hkl
[c]$R_{cryst} = \sum_{hkl} \| F_{obs} | - | F_{calc} \| / \sum | F_{obs} |$
[d]$R_{free}$is the cross-validated R-factor computed for the test set of 5% of unique reflections
[e]Root Mean Square Deviation

The crystals of Rv3400-Δ16-T2 diffracted upto 1.7 Å and belong to orthorhombic space group $C222_1$ with unit cell parameters of a=97.903 Å, b=106.389 Å, c=81.204 Å, α=β=γ=90°. The model building and refinement were carried out iteratively till no electron density was left to model. All the residues were traced in the electron density map except for the residues Asp 127, Asp 128, Asp 129 and His 186. These residues showed clear electron density for the back bone atoms, however, the electron density was absent for the side chains. Therefore, the side chains for these residues were modelled based on the rotamers.

The R_{cryst} and R_{free} converged to 15.4 % and 17.3 % respectively. The Ramachandran plot analysis of the final model using PROCHECK showed 98.78 % of the residues are in the most favoured region while 1.22 % of the residues are in the allowed region. The final model of Rv3400 consists of 273 amino acid residues, two Cl, one glycerol (GOL) and one EDO (1,2-Ethanediol) molecule. The data collection and refinement statistics for Rv3400 are shown in **Table 5.**

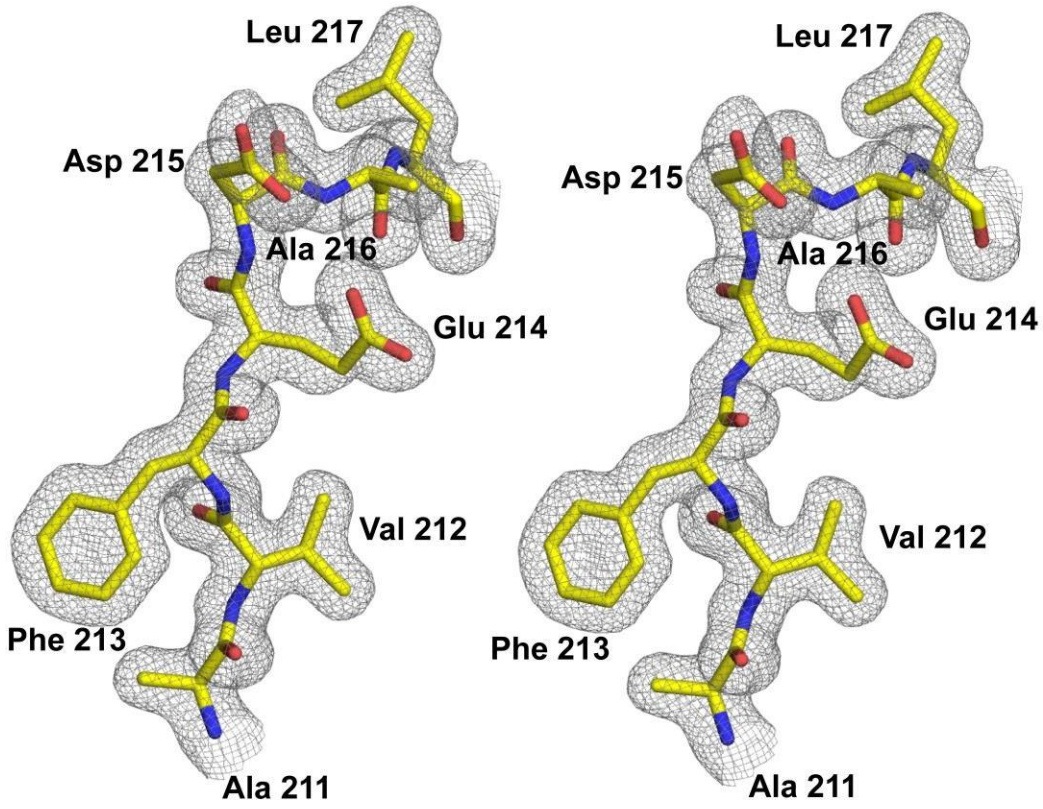


Fig. 4.12 Stereoview showing the 2 Fo-Fc fourier map for Rv3400.

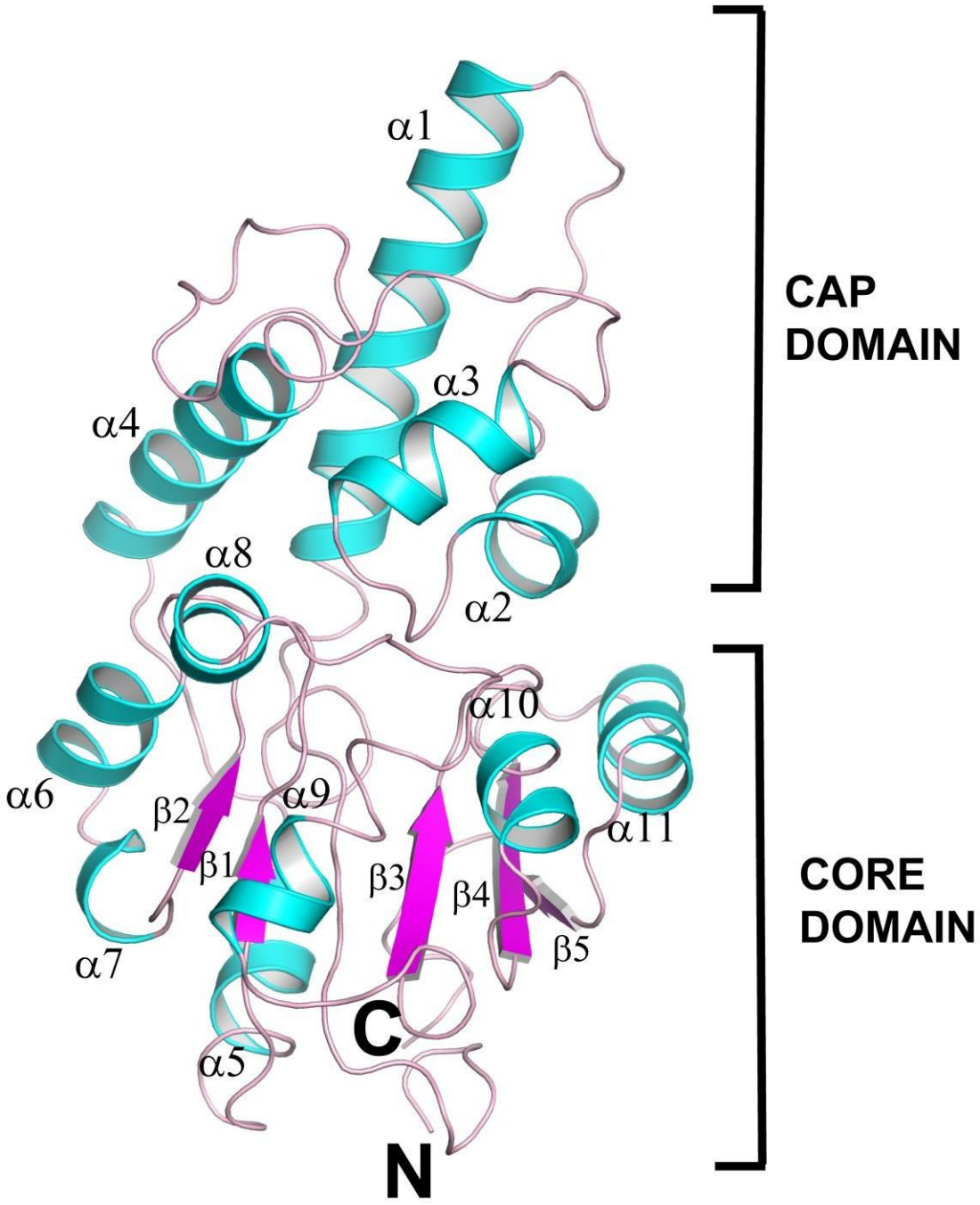


Fig. 4.13 Overall structure of Rv3400 showing α-helices and 6 β-stands. Secondary structures are labeled and represented in cyan for α-helices, magenta for β-stands and light pink for loops. The cap domain is represented by the 4 α-helices while the core domain or the catalytic domain contains the α/ β fold.

4.3.8. Overall structure of Rv3400

The overall structure of Rv3400 is composed of 262 aa which constitute 13 α-helices and 6 β-strands. The Rv3400 structure have two domains - a helical cap domain (residues 37-129) and the α/β core domain (residues 17-36 and 129-262) **(Fig. 4.13).** Rv3400 exhibits four widely spaced motifs. Motif I, present at the N-terminus has the DxD signature residues which is usually a metal binding motif. Motifs II and III including a highly conserved Thr (or Ser) and Lys, respectively. Motif IV is composed of conserved acidic residues E and D. Rv3400 consists of a Rossmanoid fold, a three-layered α/β sandwich with five parallel strands composing the central β-sheet. In addition to the four conserved motifs, Rv3400 exhibit C1 type of cap module as it contains four α-helices in the middle of the flap motif (2 β-sheets) which is important to determine substrate specificity.

4.3.9. Comparison of Rv3400 with other structural homologs

We searched for structural homologs of Rv3400 using PDBefold server (Velankar et al., 2010) (Krissinel & Henrick, 2004) and the top hits include Beta phosphoglucomutase from *Lactococcus lactis* (PDB ID: 5OLY) with rmsd of 1.2 Å and Beta phosphoglucomutase from *Clostridioides difficile* (PDB ID 4GIB) with rmsd of 1.3 Å (**Fig. 4.14, A).** Other homologs include pyrophosphatases from *Bacteroides thetaiotaomicron* (PDB ID 3Qu5) with rmsd of 1.4 Å (**Fig. 4.14, B)**, hydrolases from *Bacillus cereus* (PDB ID 1SWV with rmsd of 1.5, phosphatases from *Escherichia coli K-12* (PDB ID-1TE2) with rmsd of 1.8. The structure of a Beta phosphoglucomutase from *Lactococcus lactis* (PDB ID: 5OLY) shares 30% sequence identity with Rv3400. Similarly, a a pyrophophatase from *Bacteroides thetaiotaomicron* (PDB ID 3Qu5) also shares 23 % sequence identity with Rv3400. Based on the information obtained through this analysis, we tried to find the activity of the Rv3400. So pyrophosphatase and Beta phosphoglucomutase activity assays were employed to find the activity of Rv3400.

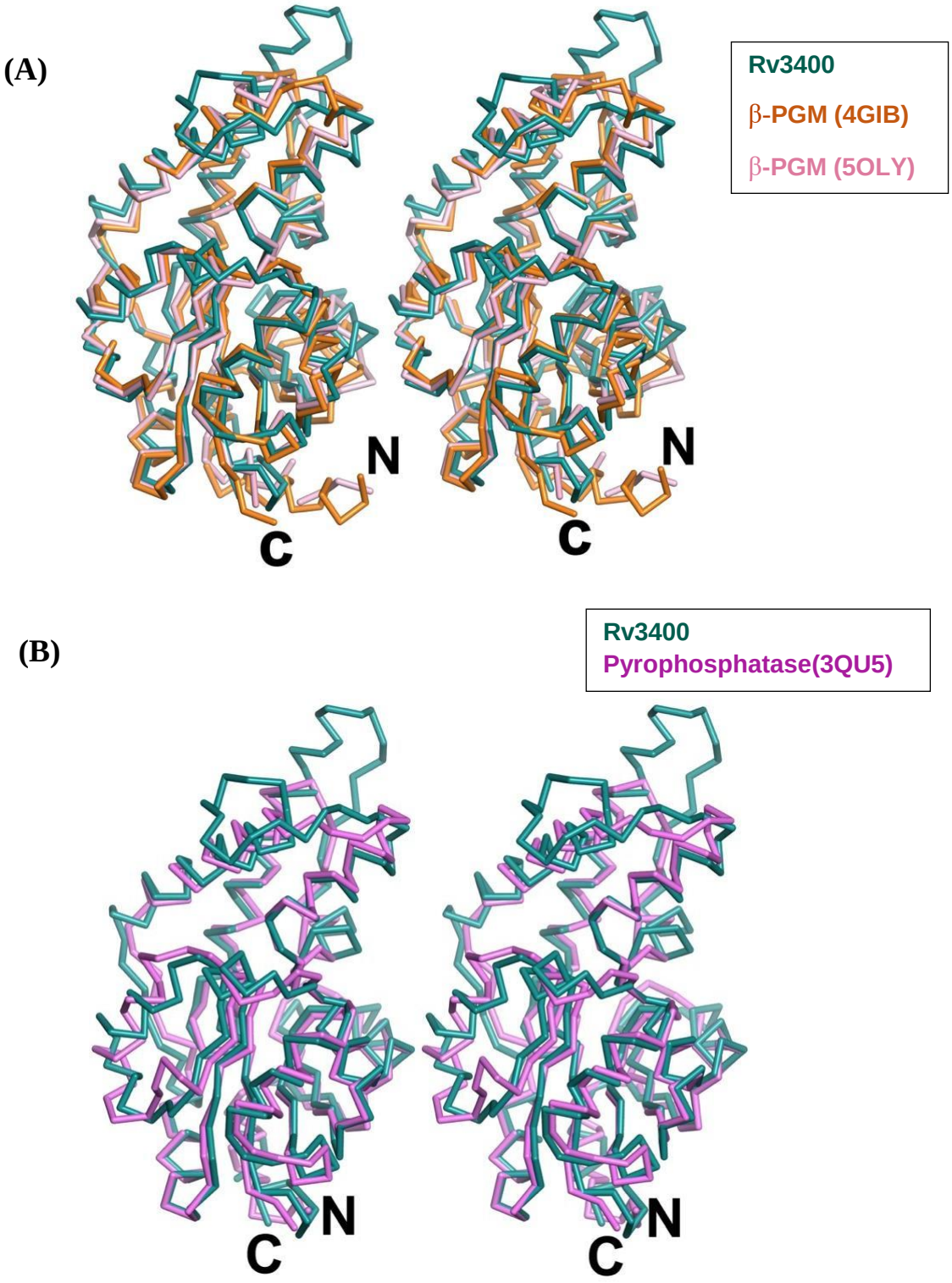


Fig. 4.14. (A) Stereoview of the **s**tructural superposition of Rv3400 (Deep teal) superimposed with proteins of Beta phosphoglucomutase family from different organisms *Lactococcus lactis***, PDB ID:5OLY** (light pink), and *Clostridioides difficile***, PDB ID:4GIB** (TV orange) **and (B)** Stereoview of the **s**tructural superposition of Rv3400 (Deep teal) with other proteins of Pyrophosphatase family (*Bacteroides thetaiotaomicron***, PDB ID: 3QU5** (Voilet).

4.3.10. Pyrophosphatase activity of Rv3400

To check if Rv3400 possess any pyrophosphatase activity, kit-based detection (Sigma, CAT-MAK168) of free pyrophosphate in the reaction was determined. The PPi sensor (provided in the kit) in presence of free pyrophosphate results in the production of fluorescent product (λ_{ex}=316/λ_{em}=456 nm) proportional to the pyrophosphate present in the reaction. The substrate sodium pyrophosphate was incubated with the Rv3400 followed by the addition of pyrophosphate sensor stock solution as per the manufacture's protocol. The results indicated that Rv3400 was unable to convert pyrophosphate to orthophosphate suggesting that it lacks any pyrophosphatase activity (**Fig. 4.15).** However, an increase in fluorescence intensity was observed in Rv3400 when compared to the blank.

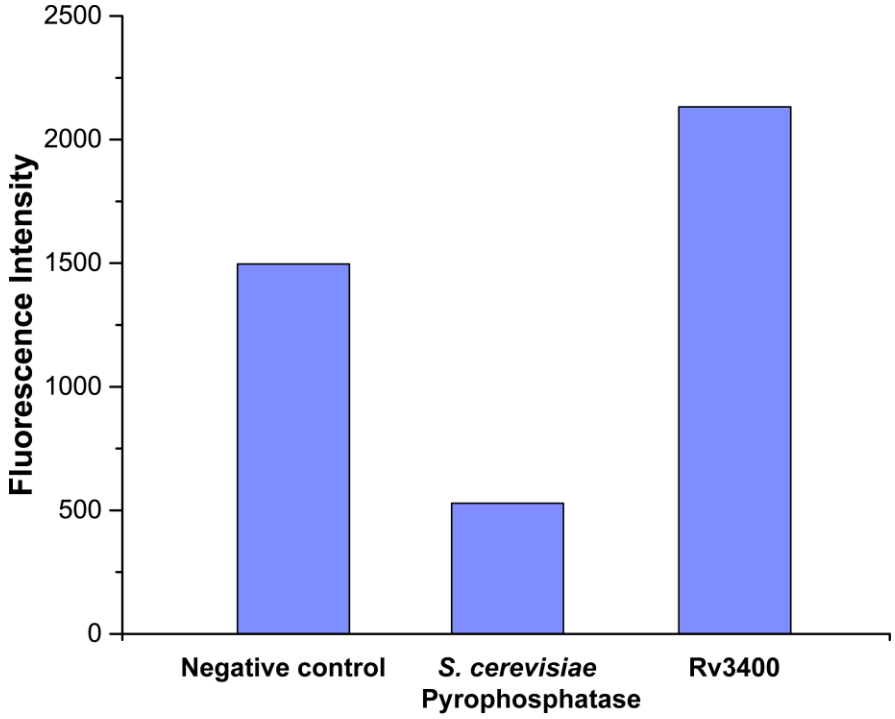


Fig. 4.15 Pyrophosphatase activity of Rv3400. Pyrophosphatase activity of protein was determined by using substrate Sodium pyrophosphate. The negative control contained the buffers and substrate except the protein, Pyrophosphatase from *S. cerevisiae* was used as positive control and different concentrations of Rv3400 (1µM-5µM) were used in the assay (data here shown only for 5 µM). (The experiment was done only once to check the activity of the protein).

4.3.11. Beta-phosphoglucomutase activity (β-PGM) of Rv3400

β-PGM activity of Rv3400 was measured spectrophotometrically by employing the method reported previously (G. Zhang et al., 2005a). It was a coupled reaction in which the substrate

β-D-Glucose 1-phosphate (β-G1P) was converted into the β-D-Glucose 6-phosphate (β-G6P) by the enzyme Rv3400 (β-PGM). The product β-G6P was further catalyzed by the enzyme glucose 6-phosphate dehydrogenase (G-6PD) and led to the reduction of NADP. Thus, the increase in the absorbance of NADPH which is directly proportional to the formation of product β-G6P was monitored at 340 nm for 30 minutes using UV-visual spectrophotometer (Perkin Elmer lambda 25). An increase in the absorbance at 340 nm indicated that the enzyme was functionally active and catalyze the β-PGM activity (**Fig. 4.16).**

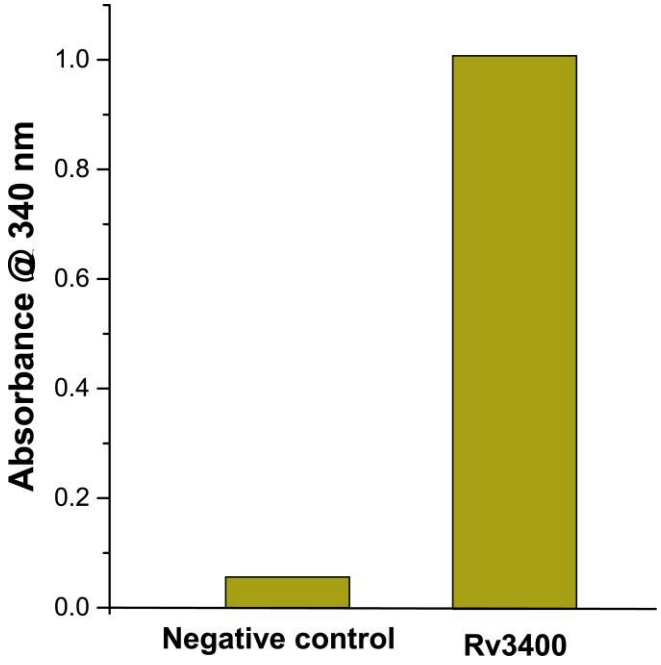
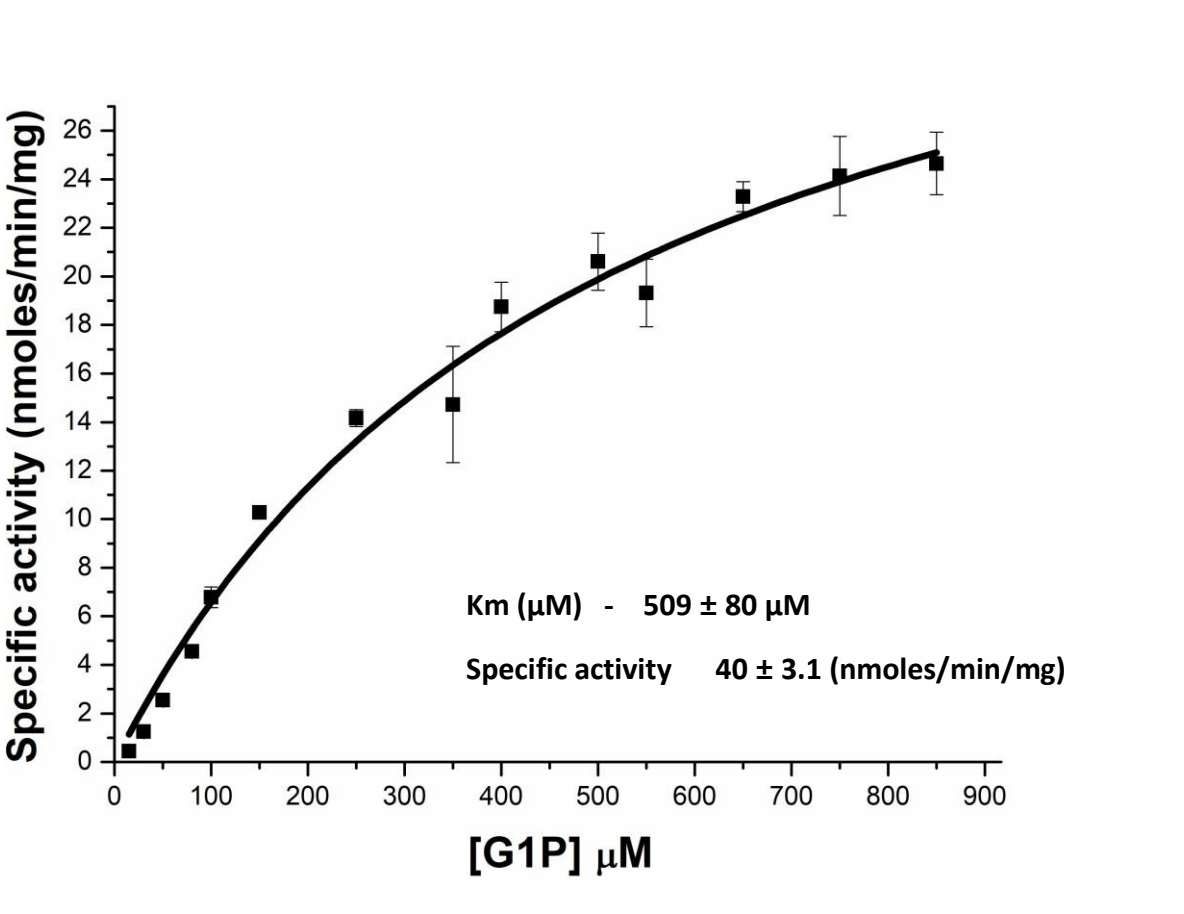


Fig. 4.16 β-PGM activity of Rv3400. β-PGM activity of proteins using substrate β-G1P which was synthesized from maltose. The negative control is reaction mixture without protein. Different concentrations of Rv3400 protein (1µM-5µM) were used (data here shown only for 5 µM).

The kinetics parameters such as K_m, specific activity and K_{cat} were calculated for Rv3400 protein. The kinetic parameters for Rv3400 were determined by using the non- linear regression method to fit the data to the Michael Menten equation using the program ORIGIN 8.5 (**Fig. 4.17).** All readings were taken in triplicate for each concentration of substrate. The calculated K_m and specific activity values were 509 ± 80 µM and 40 ± 3.1 nmoles/min/mg respectively. The K_m obtained was comparable to *Lactococcus lactis* K_m 400 ± 40 µM where the reaction was catalyzed without β-G1,6bisP (G. Zhang et al., 2005a) which was otherwise 4± 5 µM.



(A)

Kinetic parameters	Rv3400 (in absence of intermediate)	*L.lactis* (in absence of intermediate)	*L.lactis* (in presence of intermediate)
Km (µM)	509 ± 80	400 ± 40	4 ± 5

(B)

Fig. 4.17 Steady state kinetics of Rv3400. (A) Michaelis-Menten plot showing the kinetic parameters for Rv3400. All readings were taken in triplicate for each concentration of substrate. (B) Comparison of kinetic parameters of Rv3400 with β-PGM of *Lactococcus lactis* in presence and absence of intermediate β-glucose 1,6-bisphosphate (βG1,6bisP).

4.3.12. Biochemical characterization of Rv3400

The β-PGM activity of Rv3400 requires Mg^{2+} for its catalytic activity. The catalytic activity of Rv3400 was checked in the presence of other divalent metal ions. The results suggested that besides Mg^{2+}, Rv3400 possess catalytic activity in the presence of Mn^{2+}(105%), Co^{2+}(108%), Ni^{2+}(105%). However, in case of metal ions Ca^{2+}, Ba^{2+}, Zn^{2+} and Cd^{2+}, no activity was observed. The buffers of different pH from 5.0-9.0 were used to check the β-PGM activity of Rv3400. The optimum pH was found to be 8.0 for Rv3400 catalysis. The activity of Rv3400 was also carried out at various temperature from 20-65℃. The results suggested that the β-PGM activity of Rv3400 was maximum at 30℃ and activity decreased in the temperature range of 50-65℃ (**Fig. 4.18).**

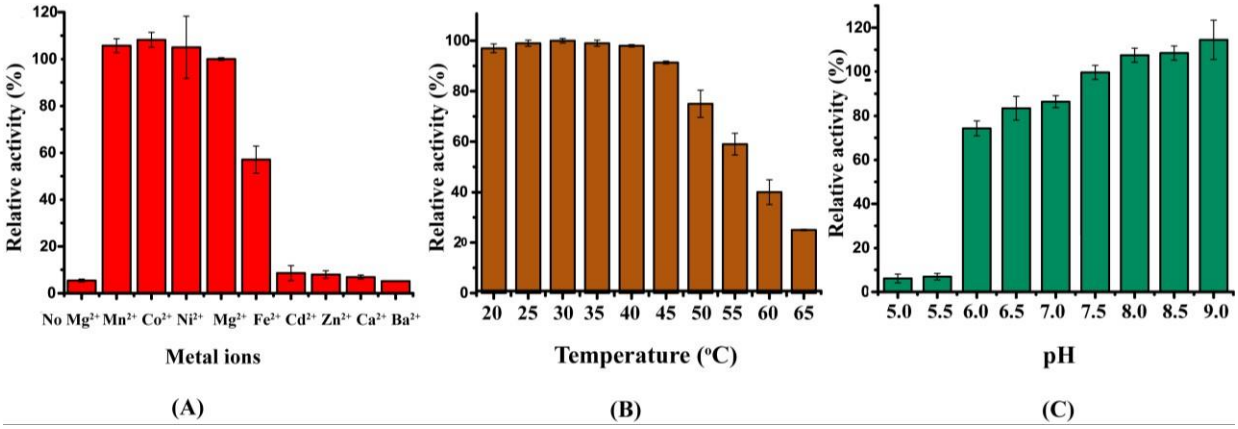
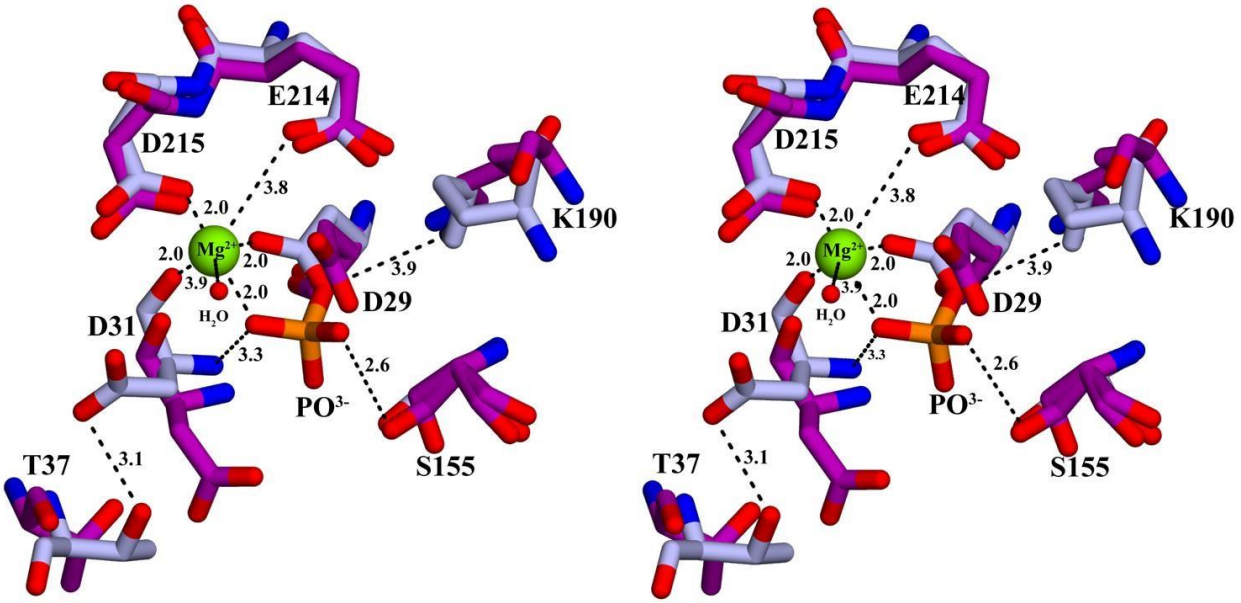


Fig. 4.18 Biochemical characterization of Rv3400 (A) β-GM activity of Rv3400 in presence of different divalent metal ions **(B)** β-PGM activity of Rv3400 in presence of different temperatures, and **(C)** pH profile of β-PGM activity of Rv3400.

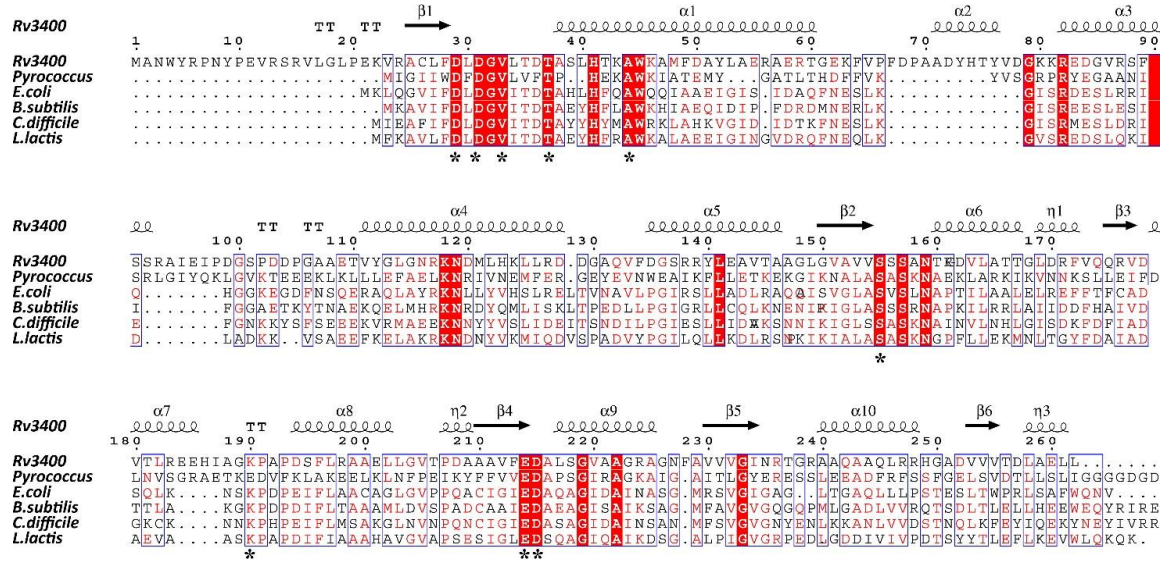
4.3.13. Catalytic site of Rv3400

The active site of Rv3400 is located at the core domain that comprised of β-sheets sandwiched between α-helices and one residue from the cap domain involved in the catalysis. The active site comprised of the side chain Thr37 contributed by the cap domain and side chains of Asp29, Asp31, Val33, Thr37, Ser155, Lys190, Glu214, and Asp215 contributed by the core domain (**Fig. 4.19**). Asp 29, Asp31, Lys190, Glu214, and Asp215 are conserved among all HAD phosphotransferases. When compared, the structural homologs Rv3400 and *L. lactis* (1LVH) shows that Asp 8 (Asp 29 in Rv3400) is phosphorylated. Also, a Mg^{2+} is coordinated with octahedral geometry by the three carboxylate side chains of Asp8 (Asp 29), Glu169 (Glu214), and Asp 170 (Asp 215), the backbone carbonyl oxygen of Asp 10 (Asp 31) and a

water molecule. However, for Rv3400, we did not observe any electron density for phosphate or Mg^{2+} suggesting that it exists in apo form.



(A)



(B)

Fig. 4.19 (A) Stereoview of active site superimposition of Rv3400 (Purple) and *Lactococcus lactis* (light blue) showing the residues involved in the catalysis and in the coordination of Mg^{2+}. (**B)** Multiple sequence alignment of β-PGM from different bacteria including Rv3400 from Mtb showing the conserved active site residues. The numbering and secondary structure elements of alignment corresponds to Rv3400. Conserved residues are shown in red on white

background and identical residues are displayed in white on red background. The amino acid residues that are conserved between the groups are boxed. Amino acid residues that are involved in the formation of active site are shown in star mark. Multiple sequence alignment was performed with the Multalign server and the figure was generated by ESpript 3.0 on web server (https://espript.ibcp.fr/ESPript/ESPript/) (Robert & Gouet, 2014).

4.3.14. Mutations of DxD motif

To catalyze the β-PGM reaction by Rv3400, it has to be phosphorylated at Asp 29 residue. Like other β-PGMs, it is a divalent metal dependent (Mg^{2+}) protein which is crucial for the activity as reported earlier and further verified by our activity assays. The DxD, a conserved motif of HAD family and also a metal binding domain has crucial role to play in the catalysis. So to understand the role of DxD motif in the catalysis of Rv3400, Asp 29 was mutated to Ala, Glu, Asn and similarly Asp 31 to Ala, Glu, Asn were individually mutated by site directed mutagenesis. The Rv3400 active site mutants (*Rv3400D29A, Rv3400D29E, Rv3400D29N, Rv3400D31A, Rv3400D31E and Rv3400D31N*) were cloned in pETDuetN vector. The mutants were confirmed by automated DNA sequencing.

The plasmid having the point mutations in Rv3400 were transformed in Rosetta (DE3) for the expression of the corresponding proteins (Rv3400D29A, Rv3400D29E, Rv3400D29N, Rv3400D31A, Rv3400D31E and Rv3400D31N). The expression and purification of mutant proteins of Rv3400 were done by the same protocol as used in wild type (**Fig. 4.20).**

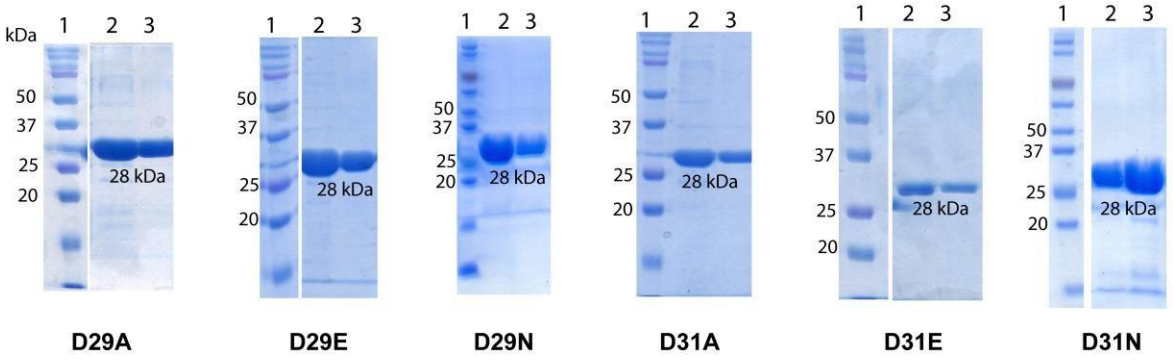


Fig. 4.20 Purification of Rv3400 active site mutants using Ni-NTA chromatography. 15% SDS-PAGE showing Rv3400 active site mutant purified protein using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2 and 3: Elution fractions of mutant proteins of Rv3400.

To know whether the mutations has affected the folding of Rv3400, Circular dichroism (CD) was done for all mutant proteins (Rv3400-D29A, Rv3400-D29E, Rv3400-D29N, Rv3400-D31A, Rv3400-D31E and Rv3400-D31N). The CD studies indicated that there were no gross change in the secondary structure content and protein folding of the mutant proteins when compared to wild type protein. Also, the CD analysis of Rv3400-Δ16 was done to know whether the deletion of first 16 aa residue had affected the protein folding (**Fig. 4.21)**.

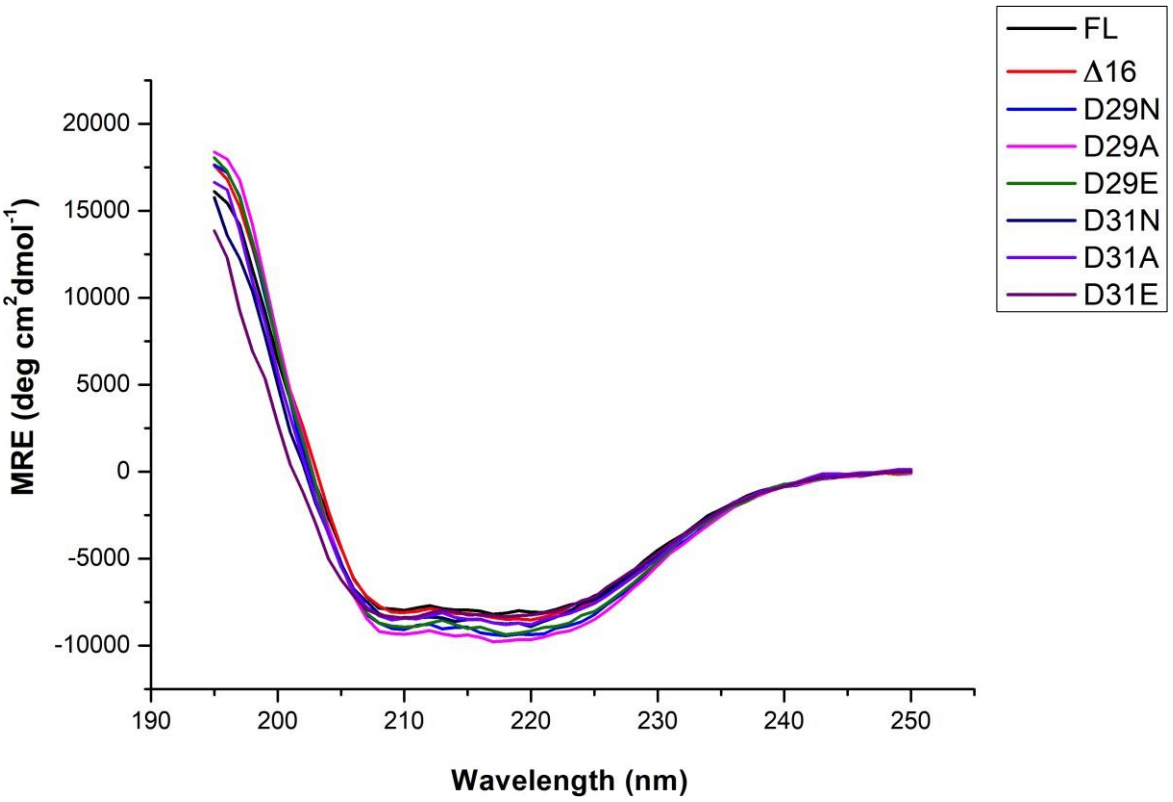


Fig. 4.21 CD spectra of Rv3400-FL, Rv3400-Δ16 and active site mutant proteins. CD spectra show the well folded Rv3400 active site mutants comparable to that of wild type Rv3400.

The catalytic activity of all the mutants was monitored in the manner similar to the wild type protein Rv3400. The β-PGM activity assay performed with the point mutants of D29 and D31 (D29A, D29E, D29N, D31A, D31E and D31N) revealed that they were catalytically inactive, however D31N retained ~30 % activity (**Fig. 4.22**).

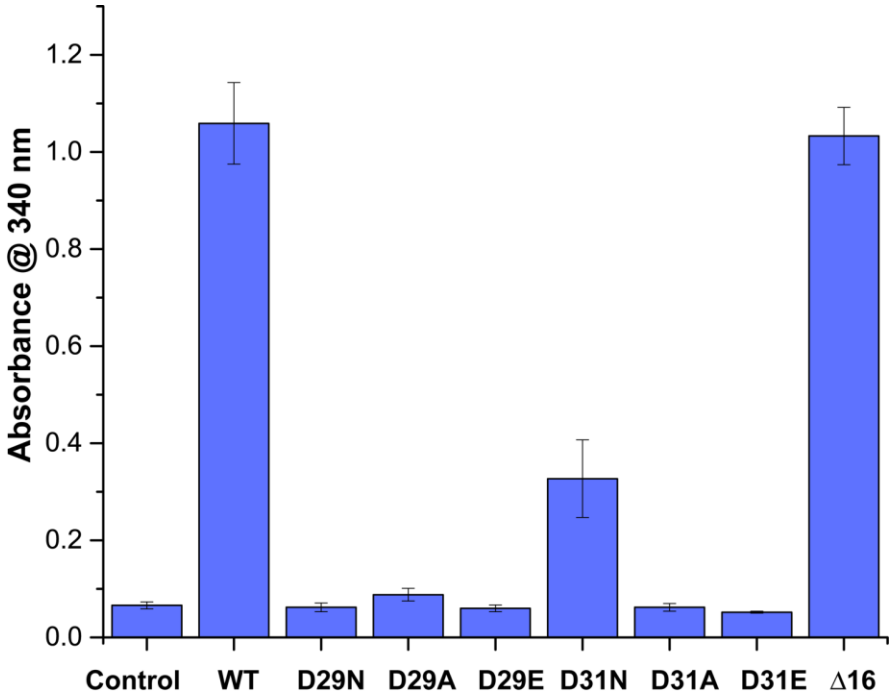


Fig. 4.22 Catalytic activity of Rv3400 mutants. Effect of active site mutants on the β-PGM activity of Rv3400 which has shown loss in the activity except for D29A. The crystallized construct Rv3400-Δ16 was also employed to check its β-PGM.

4.3.15. Activation of the enzyme Rv3400 by substrate induced phosphorylation

In order to carry out the catalysis, β-PGM needs to be phosphorylated at one of the Asp residue. Mutational studies also suggested that if the phosphorylation site (Asp29 in Rv3400) is mutated, there is no phosphorylation and thus become inactive. The literature suggests that the phosphorylation of the β-PGM is done by one of the three mechanisms (**Fig. 4.23).** First mechanism includes autophosphorylation of the enzyme where β-PGM is found to be phosphorylated during the expression and purification of the enzyme by yet unknown molecular events (Lahiri et al., 2002b). Second mechanism involves phosphorylation of the enzyme by intermediate β-G1,6BP (Lahiri, Zhang, Dunaway-Mariano, & Allen, 2003b). The third mechanism involves phosphorylation of the enzyme by the substrate β-G1P (substrate induced phosphorylation) (Mesak & Dahl, 2000). However, the β-PGM activity was also observed in the absence of the intermediate but its presence in the reaction helps to achieve the maximum catalytic efficiency (Kvam, Olsvik, McKinley-McKee, & Saether, 1997). In *Bacillus subtilis* and *Lactococcus lactis*, the β-PGM shows that in absence of an intermediate, β-G1P acts as phosphoryl donor and the K_m observed was 400 μM. But when the same reaction was catalyzed in the presence of intermediate β-G1,6BP, K_m was found to be 4 μM

(Mesak & Dahl, 2000; G. Zhang et al., 2005b), indicating that the enzyme has a strong affinity with the intermediate rather than the substrate. We did β-PGM activity assay for Rv3400 in the absence of intermediate β-G1,6 BP and the K_m observed was 500 μM. Since we observed a higher K_m comparable to that has been reported earlier where no intermediate was used, we are proposing that the enzyme is phosphorylated by the substrate β-G1P. To rule out the possibility of autophosphorylation of Rv3400, we checked the intact mass of the protein through LC-MS. The results shown no additional molecular weight corresponding to the phosphate, thus no autophosphorylation of the protein.

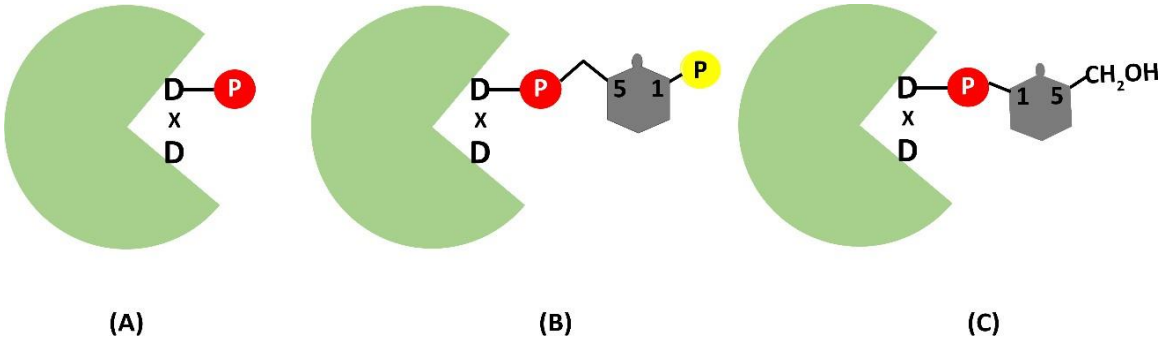


Fig. 4.23. Reported mechanism of phosphorylation of Beta phosphoglucomutase (β-PGM) (A) Autophosphorylation (Lahiri et al., 2002b), (B) Phosphorylation by the intermediate (Lahiri et al., 2003a) and (C) Phosphorylation by the substrate (Mesak & Dahl, 2000).

4.3.16. Proposed Catalytic mechanism of Rv3400

The reaction catalyzed by β-PGM comprises of two-step process, first the conversion of β-G1P → β-G1,6P, followed by the conversion of β-G1,6BP → β-G6P. Two mechanisms for these processes was proposed earlier: a general-base catalyzed mechanism and a substrate assisted mechanism. In general-base mechanism, the nucleophilic hydroxyl group of β-G1P (position 6) attacks the phosphorylated D8 (residues for *Lactococcus lactis,* D29 for Rv3400*),* with its proton transferred to the D10 ("D10-assisted mechanism", D31 for Rv3400). The second is the substrate-assisted mechanism, which also involves the nucleophilic attack of β-G1P , but in this case the hydroxyl group is deprotonated by one of the non-bridging oxygens of the phosphate group ("substrate-assisted mechanism") (**Fig.4.24)** (Barrozo et al., 2018b). Based on our active site mutations results, we propose that Rv3400 follows the general-base mechanism for its catalysis. Initially deprotonated D31 acts as a general base, activating the nucleophile attack on the phosphorylated D29 to yield β-G1,6BP, followed by a second

phosphoryl transfer reaction from the 1-position of β-G1,6BP back to D29, followed by protonation of the leaving group by D31 (G. Zhang et al., 2005a) (J. Dai et al., 2009). The reaction starts with the phosphorylation at D29 residue of Rv3400 by the substrate β-G1P leading to the formation of phosphorylated enzyme and glucose. The phosphorylated enzyme in this conformation then binds to another molecule of the β-G1P from its 6 carbon position that involves the nucleophilic attack of β-G1P on phosphorylated D29 and forms the intermediate glucose-1,6-bisphosphate (β-G1,6BP). The intermediate dissociates from the enzyme leaving an unphosphorylated enzyme. The unphosphorylated enzyme had a tendency to bind to either β-G1P or β-G16BP at their first carbon position. As discussed in the previous section, the affinity of the intermediate is more than that of substrate, Rv3400 binds β-G16BP preferably and facilitates the conversion of β-G1,6BP to β-G6P. The final output is the product β-G6P and a phosphorylated Rv3400 which is ready to start a new cycle. Together with the previous reports (J. Dai et al., 2009), our mutational studies has highlighted the significance of both D29 and D31 in the catalytic cycle. Importantly, the mutation of D29 completely abolish the activity of Rv3400. However, while the mutation of D31 to either Ala or Glu completely abolish the activity, the D31N mutation retains about 30 % catalytic activity indicating the length of the side chain of D31 may play a role in catalytic activity.

However, based on structural information, it has also been suggested that D31 keeps β-PGM in active and closed conformation (G. Zhang et al., 2005b). It has also be observed from the previous structural studies (PDB ID: 6H8U, 6H91, 2WHE,) and the from Rv3400, the enzyme will remain in open conformation during the phosphorylated bound form, slightly closed during the apo-form (Rv3400) and closed after binding of the substrate or ligand (PDB ID:1Z40) and during the catalysis.

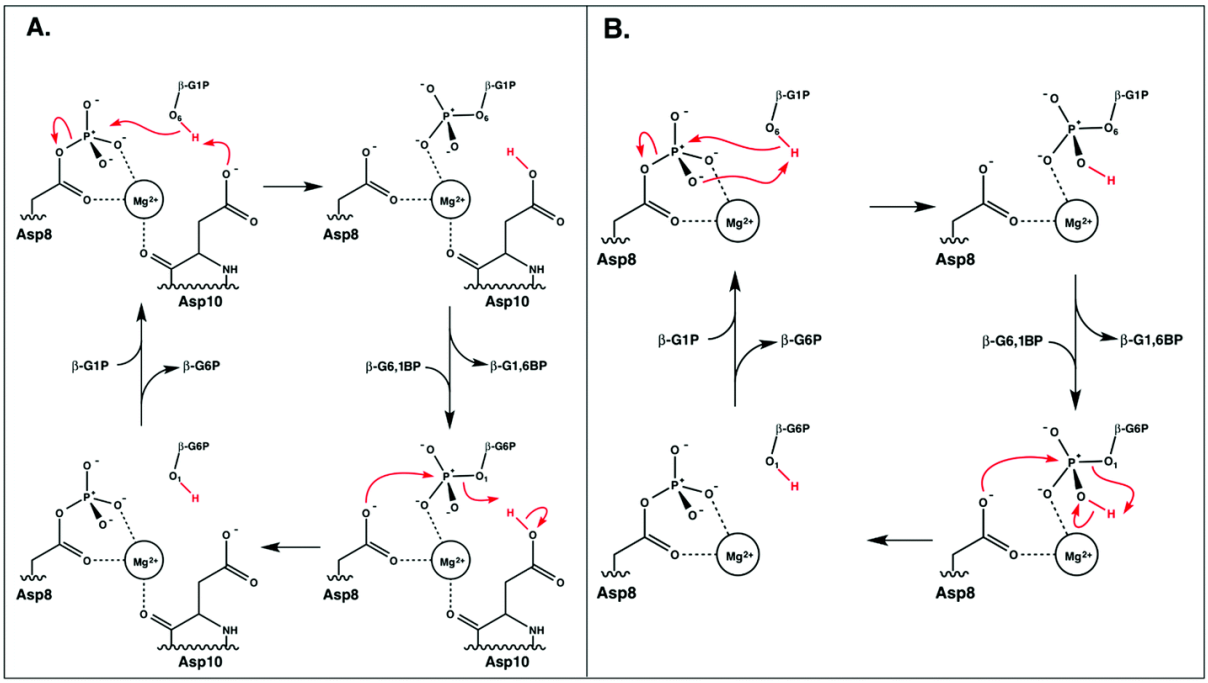


Fig.4.24 Reported mechanisms of catalysis of Beta phosphoglucomutase (β-PGM) (A) The general-base catalyzed mechanism (D10-assisted mechanism) originally postulated by (J. Dai et al., 2009) , proposes that D10 acts as a general acid/base and the proton, highlighted in red, is transferred to and from the nucleophilic hydroxyl of the glucose. (B) The substrate-assisted mechanism has the phosphoryl group acting as the acid–base catalyst.

Source- https://pubs.rsc.org/en/content/articlehtml/2018/ob/c8ob00312b

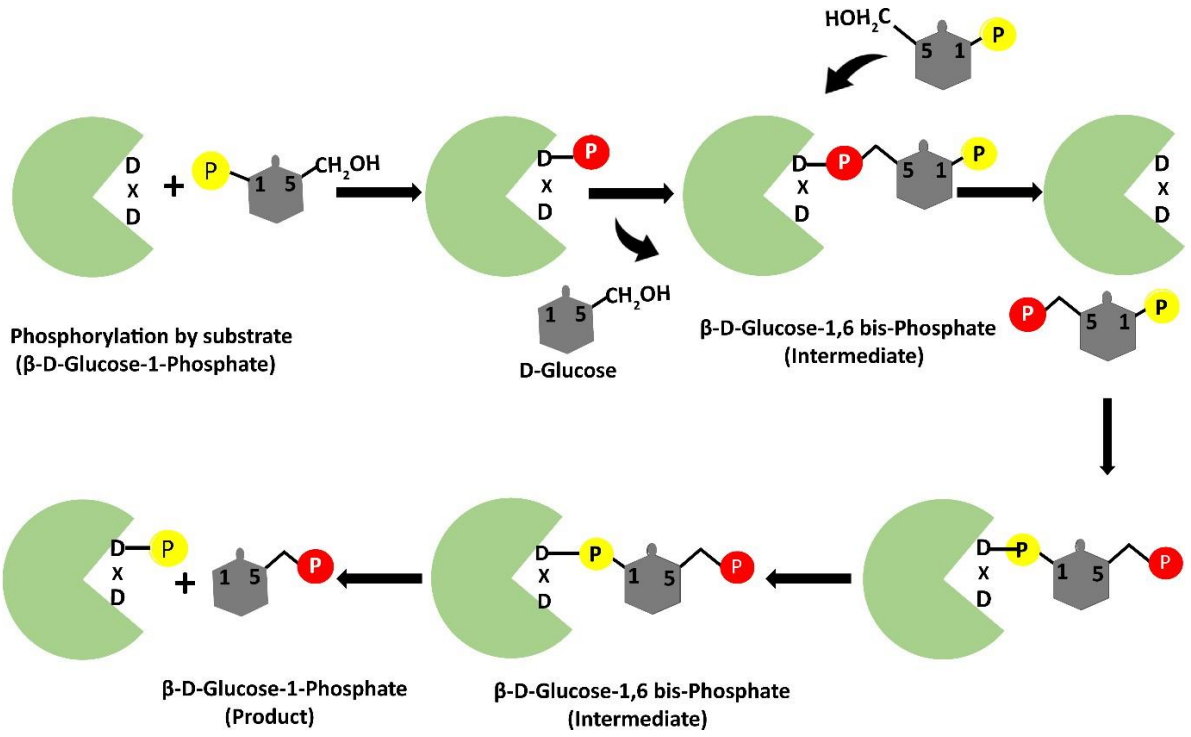


Fig. 4.25 Proposed mechanism for Rv3400 catalysis.

4.4. Discussion

Riboflavin, commonly known as vitamin B2, is the precursor molecule for FMN and FAD and plays an important role in production of energy, cellular activity, growth and development, as well as the metabolism of fat, drugs and steroid. The riboflavin biosynthesis pathway is essential for the survival of Mtb and the enzymes of the pathway can be the potential drug targets for designing new drugs. The pathway involved five enzymes and out of them, four has been characterized structurally and biochemically. However, a phosphatase which catalyze an essential step of the RBS pathway has not been well studied yet. Thus, we started this project to find the unknown phosphatase in the riboflavin biosynthesis pathway. So, we were interested to characterize this phosphatase of Mtb. Literature survey suggests that the members of large haloacid dehalogenase superfamily were shown to perform the dephosphorylation step in different organisms. Two genes yigB and ybjl from *E.coli* (Haase et al., 2013),three genes ycsE, ywtE and yitU from *Bacillus subtilis* (Sarge et al., 2015) and AtPyrP2 and AtGpp1/PyrP3 from *Arabidopsis thaliana* (Sa et al., 2016) are reported to catalyze the dephosphorylation step in riboflavin biosynthesis pathway. To find the homologs of the potential phosphatase involved in RBS pathway the known phosphatase sequences were used as a query sequence in PSI-BLAST (Position specific iterated-BLAST) using default parameters. We predicted three homologs Rv3400, Rv3376 and Rv3813c as potential phosphatase of riboflavin biosynthesis pathway in Mtb. Structural studies were done to characterize these enzymes. We were successful in characterizing Rv3400 which encodes for the beta phosphoglucomutase (β-PGM). We have reported the first high resolution β-PGM (1.7Å) from Mtb. Rv3400 is a monomeric protein of 29 kDa that has magnesium-dependent phosphoryl transfer activity. The active site is located in the cleft between the α/βcore domain (17-36 and 129-262) and the α-helical cap domain (37-129) and during catalysis, the cleft is closed by domain reorientation. Based on the previous studies and our structural and mutational studies for Rv3400, we found that two aspartate residues are important for catalysis of Rv3400. The enzyme has to phosphorylate at Asp29 while Asp 31 helps in the nucleophilic attack by the substrate β-G1P for catalysis and uses Mg^{2+} ion as cofactor. During steady-state catalysis, β-G1P binds to phosphorylated Rv3400 (phosphorylated on D29) and forms an intermediate β-glucose 1,6-bisphosphate (β-G1,6BP). The formation of product, G-6P and the regeneration of Rv3400 were caused by the release of G1,6BP into solution and subsequent rebinding in the alternative orientation based on the high affinity of Rv3400 with the intermediate (J. Dai, Wang, Allen, Radstrom, & Dunaway-Mariano, 2006). A phosphorylating or priming agent is required to initiate the reaction *in vitro* but *in vivo*, this

can carried out by F16BP, β-G1P, G6P, α-glucose 1,6-bisphosphate (α- G1,6BP), acetyl phosphate (AcP), and also with the reaction intermediate, βG1,6BP (H. P. Wood et al., 2020). Since the reaction catalyzed by Rv3400 lacked intermediate β-G1,6BP, we assumed the phosphorylation was done by the substrate β-G1P. However, only the intermediate β-G16BP allows the enzyme β-PGM to reach its maximum catalytic rate, and also an increase in the K_m and V_{max} is observed in its absence.

The role of β- phosphoglucomutase in the physiology of *Bacillus subtilis* and *Lactococcus lactis* is well studied where they are known to play a role in carbohydrate metabolism especially the disaccharide maltose. Insertional mutagenesis of β-phosphoglucomutase enzyme resulted in the growth deficiency of *B. subtilis* on minimal medium containing starch or maltodextrins (maltose to maltoheptaose) (Mesak & Dahl, 2000). In *Lactococcus* both maltose and trehalose were catabolized by the sequential action of phosphorylase and β-phosphoglucomutase (Andersson & Radstrom, 2002) (Looijesteijn, Trapet, de Vries, Abee, & Hugenholtz, 2001). *L. lactis* β-PGM-deficient strain is unable to grow or has impaired growth, when trehalose or maltose is the sole carbon source (Levander, Andersson, & Radstrom, 2001).

In Mtb, trehalose is a prevalent non-reducing disaccharide present in free or in glycoconjugates form in the cytosol as well as in the cell wall (Elbein et al., 2003). Trehalose is crucial for the composition of mycobacterial cell envelope that mainly involves the trehalose monomycolate (TMM) and trehalose dimycolate (TDM) found on cell surfaces (Jankute et al., 2015). Trehalose has various roles in Mtb physiology. It is used as carbon and energy source, protects Mtb against desiccation, freezing, osmotic stress, and also plays an important role in Mtb virulence (Shi et al., 2013). Mtb can use exogenous trehalose as well as synthesize it *de-novo* by three pathways. Endogenously, Mtb can produce trehalose from glucose 6-phosphate and UDP-glucose, from glycogen-like alpha(1-->4)-linked glucose and from maltose (De Smet et al., 2000).

Mtb alters its trehalose metabolism as an adaptive metabolic strategy and remodel it to overcome the stress by stochastic formation of persisters during growth-adverse environments (Fisher et al., 2017; Keren et al., 2011; T. K. Wood et al., 2013). The stress that led to the development of Mtb PLBs (persister-like Bacilli) included nutritional starvation-induced systemic metabolic damage, ETC dysregulation intrabacterial ATP depletion, and accompanied loss of redox homeostasis (Dutta & Karakousis, 2014). In order to resolve these

challenges, pre-existing trehalose is internally catalyzed via modified metabolic activities to shift its carbon flux more toward the biosynthesis of glycolysis (GL) and pentose phosphate pathway intermediates (PPP). This shift also led to the decrease in the carbon flux for the biosynthesis of cell surface TMM/TDM. This catalytic shift produced initial substrates for the glycolysis and pentose phosphate pathway and thus served as alternate biosynthetic sources of ATP, NADPH, and antioxidants (J. J. Lee et al., 2019).

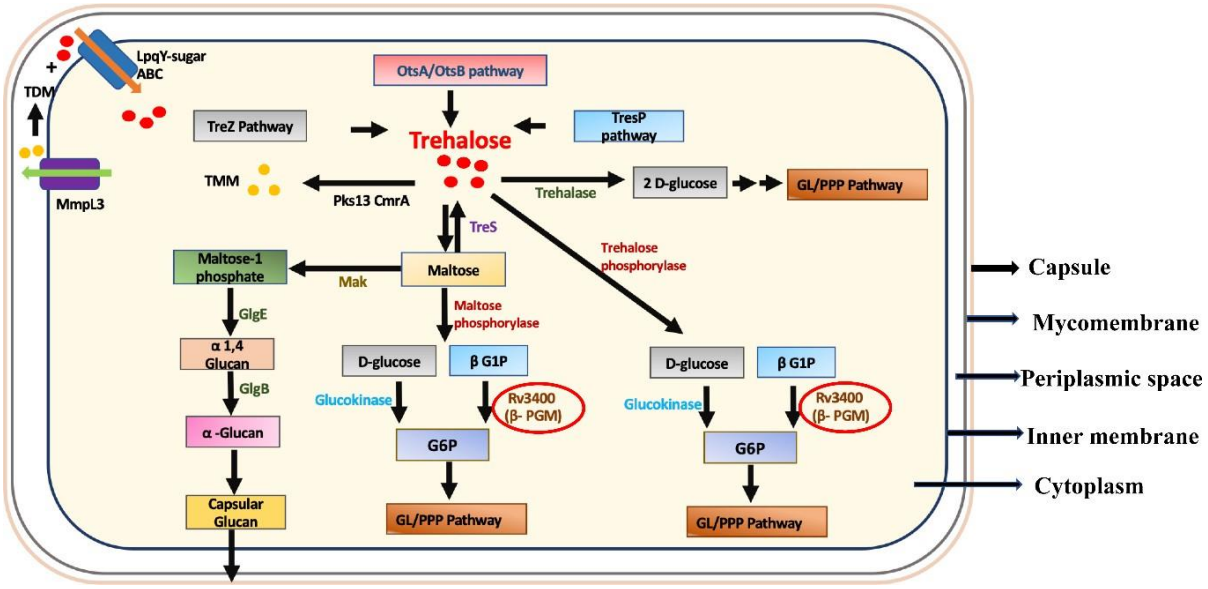


Fig. 4.26. Proposed role of Rv3400 in the physiology of Mtb.

From our present knowledge we propose the following model for trehalose /maltose metabolism in Mtb -

Trehalose acts as energy storage and synthesis of cell wall components trehalose monomycolate (TMM) and trehalose dimycolate (TDM) and alpha glucan for capsule formation, thus involves its catabolism and biosynthesis through various pathways. Biosynthesis generally occurs through three pathways- OtsAB or TPS/Tpp, TreS pathway and TreYZ pathway but only OtsAB pathway is essential for trehalose biosynthesis (De Smet et al., 2000). Trehalose is utilized by the enzyme Pks13 and CmrA to produce TMM which is a mycolyl carrier. TMM is transported from the cytoplasm to the periplasm for biosynthesis of TDM and free trehalose releases in pseudoperiplasmic space (Gavalda et al., 2014; Li et al., 2014).

Trehalose synthase (TreS) is the main enzyme of the TreS pathway and converts trehalose into maltose and vice-versa. Both TresS and GleE pathway are linked by the enzyme Pep2 which is a maltokinase and converts maltose into maltose-1-phosphate. The enzyme maltosyl-transferase GlgE extends glucan chains from maltose-1-phosphate whereas the enzyme GlgB introduces α-1,6-linked branches to linear glucans which then forms capsular glucan.

During starvation conditions or persister formation, free trehalose is metabolized by different enzymes to yield glucose 6-phosphate (β-G6P) that generate ATP and NADPH through glycolysis or Pentose phosphate pathway for the survival of the cell. The reversible enzyme TreS converts trehalose to maltose which can further converted to β-G1P and glucose by the action of Pi-dependent maltose phosphorylase Rv3401 (gene adjacent to Rv3400). Another pathway involves the enzyme trehalose phosphorylase (Trep) that catalyze the reversible hydrolysis of trehalose in presence of inorganic phosphate which can convert trehalose to glucose and β-G1P. The β-G1P formed is catalyzed by the β-phosphoglucomutase to β-G6P whereas the glucose is phosphorylated by glucokinase. In both the pathways, β-G1P serves as a substrate for Rv3400 and is converted to β-G6P. Glucose-6-phosphate thus formed can be used for the synthesis of ATP and NADPH via glycolytic pathway or the pentose phosphate pathway. The enzyme trehalase is known to degrade trehalose into two molecules of glucose in other organisms but its homologs are not present in Mtb.

Chapter 5
Characterization of Rv3376 and Rv3813c

5.1. Introduction

As described in the earlier chapter, we were targeting an essential pathway of Mtb that leads to the formation of riboflavin, a sole precursor molecule for the synthesis of cofactors FMN and FAD (Fassbinder et al., 2000) for drug designing. Riboflavin biosynthesis pathway involves five enzymes utilizing the substrates - GTP and Ribulose-5 phosphate. All the enzyme of this pathway are characterized structurally and functionally except for a phosphatase that dephosphorylates one of the intermediate of the pathway 5-amino-6-ribitylamino-2,4(1*H*,3*H*) pyrimidinedione 5′-phosphate (ARPP) to 5-amino-6-ribitylamino-2,4(1*H*,3*H*) pyrimidinedione (ARP). In the previous chapter we have discussed the bioinformatic methods to identify the homologs using known phosphatases from *E. coli, B. subtilis* and *A. thaliana* in Mtb. Utilizing the known phosphatases as a query sequence, we were able to find three homologs Rv3400, Rv3376 and Rv3813c as the potential phosphatases of riboflavin biosynthesis pathway in Mtb. In the previous chapter we characterized one of the potential phosphatase Rv3400. However, our results indicates that Rv3400 does not play a role in catalyzing the dephosphorylation of ARPP in the riboflavin biosynthesis pathway and instead found to be a β- phosphoglucomutase. To explore whether Rv3376 and Rv3813c could be a potential phosphatase, we focused on the characterization of these enzymes.

In Mtb, Rv3376 is annotated as a conserved hypothetical protein which is shown to be non-essential for *in vitro* growth of H37Rv (DeJesus et al., 2017; Minato et al., 2019) (C. M. Sassetti et al., 2003) but required for survival in primary murine macrophages, by transposon site hybridization (TraSH) in H37Rv (Christopher M. Sassetti & Rubin, 2003). The Rv3376 has been shown to hydrolyze geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) to produce geraniol, farnesol, and geranylgeraniol, respectively (Nakano et al., 2011).

Similarly, another potential phosphatase Rv3813c is annotated as a conserved hypothetical protein and is non-essential for *in vitro* growth of H37Rv in Mtb (DeJesus et al., 2017; Minato et al., 2019) (C. M. Sassetti et al., 2003). The BLAST and Pfam analysis has shown that the Rv3813c belongs to the haloacid dehalogenase family.

5.2. Methodology and results

5.2.1. Bioinformatics analysis

To find the homologs of the potential phosphatase in riboflavin biosynthesis pathway, the known phosphatase sequences were used as a query sequence in Basic Local Alignment Search Tool (BLAST) using default parameters. Similarly, PSI-BLAST (Position- Specific Iterative BLAST) was also employed which helped in finding the remote homologs of the potential phosphatase in Mtb. Finally, genes encoded for potential phosphatases were selected based on sequence similarity and query coverage.

The Pfam database was also used for the identification of the protein family of Rv3376 and Rv3813c. A search was performed using the Rv3376 and Rv3813c protein sequence against the Pfam database. Multiple sequence alignment and protein family analysis was done to find the family of these proteins For the prediction of the secondary structure PSIPRED (https://bio.tools/psipred) server was used.

Table1- Table showing the reported ARPP phosphatases of RBS and their homologs in Mtb with percent identity and query coverage.

S.No.	Gene (Reported phosphatases)	Potential Homologs in *Mycobacterium tuberculosis*	Sequence Similarity (%)	Query coverage (%)
1	*YigB (E. coli)*	*Rv1225c*	29	32
2	*Ybjl (E. coli)*	*Rv3813c*	27	94
3	*YcsE (B. subtilis)*	*Rv3813c*	27	94
4	*YwtE (B. subtilis)*	*Rv3813c*	27	98
5	*YitU (B. subtilis)*	*Rv3813c*	29	98
6	*AtPyrP1 (A. thaliana)*	*Rv3376*	33	27
7	*AtPyrP2(A. thaliana)*	*Rv3376*	27	39
8	*AtGpp1/PyrP3(A. thaliana)*	*Rv3400*	24	63

In order to find the homologs of potential phosphatase of riboflavin biosynthesis pathway in Mtb, the earlier reported phosphatases in *E. coli, B. subtilis, A. thaliana* that are involved in riboflavin biosynthesis pathway was employed as query sequence in BLAST. The results depicted very less similarity (6-10%) between the reported phosphatase and the homologs. So, PSI-BLAST (Position specific Iterative BLAST) was done to pick remote homologs in Mtb. Based on high percent sequence identity and query coverage three genes were selected for our studies- Rv3400, Rv3376 and Rv3813c for further studies (**Table1)**. We

already discussed the characterization of Rv3400 in chapter 4. The BLAST search using the amino acid sequences of Rv3376 and Rv3813c, suggested that these proteins share high similarity with the Haloacid dehalogenase (HAD) family (**Table 2)**. Multiple sequence alignment (MSA) and Pfam (Protein family) also predicted Rv3376 and Rv3813c to be members of Haloacid dehalogenase family proteins.

Table 2- Table showing the selected potential phosphatases in Mtb based on high percent identity and query coverage

S. No.	Homologs in *Mycobacterium tuberculosis*	Predicted function	Percent Identity (%)	Query coverage (%)
1.	Rv3813c	Cof-type HAD-IIB family hydrolase	27	94
2.	Rv3376	HAD family phosphatase	27	39
3.	Rv3400	Beta-phosphoglucomutase family hydrolase	24	63

5.2.2. Cloning of Rv3376 and Rv3813c constructs

The DNA encoding Rv3376 (654 bp) and Rv3813c (822 bp) genes were PCR amplified using Mtb H37Rv genomic DNA as a template and designed primers (detailed in **Table 3**) synthesized by Sigma-Aldrich, India. Both the PCR amplified genes encoding full length protein were cloned in pET 28c vector with N-terminal His_6-site. The Rv3376 gene was cloned in pET 28c between *NdeI* and *HindIII.* Similarly, Rv3813c gene was cloned between *NdeI* and *XhoI* restriction site in pET 28c vector. The digested and purified cloning vector and the specific gene was ligated with T4 DNA ligase, incubated at 22 °C for 2-4 hrs. The ligated product was transformed in *E. coli* TOP10 cells (Novagen) followed by plating on nutrient agar supplemented with Kanamycin. The plates were incubated for 16-18 hrs at 37°C or untill the colonies appeared. The colonies from the plates were carefully marked and checked for plasmid harbouring the gene of interest by colony PCR. The partial colonies that showed the positive amplification were inoculated in 5ml of LB broth supplement with kanamycin. The primary culture was then kept in shaker and incubate overnight at 37°C at 200 RPM. Next day, plasmid isolation was done using plasmid isolation kit from Thermo Scientific, USA (GeneJet kit). The integration of the desired gene into the plasmid was further confirmed by double digestion using restriction enzymes *NdeI* and *HindIII* for Rv3376 and *Nde1* and *XhoI* for Rv3813c for 20 min and incubated at 37°C. The digested products were run on 1 % agarose

gel to check the fallout of desired size. The sequence of the clones was verified by automated DNA sequencing. The resulting clones were named as pET 28c-Rv3376-FL and pET 28c-Rv3813c-FL.

Table 3- Table showing the list of primers and their nucleotide sequence used in the study-

S. No.	Primer Name	Nucleotide Sequence
1	Rv3376-FL-F	TATACATATGTTGAGCATTAGCGCGGTTGTTTTCG
2	Rv3376-T17-F	TATACATATGGACTGGACACGTGCCGAGGAGGATG
3	Rv3376-R	TATAAAGCTTTCATGGTCCCGCGCCGTCC
4	Rv3813c-FL-F	TATACATATGTTGAAACCGACTGTGCCGGCG
5	Rv3813C-T6-F	GTATATACATATGGCGCTCGTCGCGTGTGACGTCGAC
6	Rv3813C-T23-F	CGGATATACATATGACCAAACGCACTCGCGACGCGG
7	Rv3813c-R	TTACTCGAGCTAGGACCACCAGCGCTCCAGCA

The deletion mutation constructs of Rv3376 and Rv3813c were created to aid the purification and crystallization of the proteins. For Rv3376, a construct was made where first 17 aa were deleted (found to be disordered region as per PSIPRED analysis) and named as Rv3376-Δ17. Similarly, for Rv3813c, two constructs were made where first 6 and 24 aa were deleted and named Rv3813c-Δ6 and Rv3813c-Δ24 respectively (**Table 4)**. The protocol for cloning of these constructs in pET 28c was similar to that of full-length genes**.**

Table 4- Table showing the list of constructs and their description-

S.No.	Construct	Description
1.	pET 28c-Rv3376-FL	N-term His Tag Full Length Protein (217 aa)
2.	pET 28c-Rv3376-Δ17	N-term His Tag protein where first 17 amino acids were deleted (18-217 aa)
3.	pET 28c-Rv3813c-FL	N-term His Tag Full Length Protein (273 aa)
4.	pET 28c-Rv3813c-Δ6	N-term His Tag protein where first 6 amino acids were deleted (7-273 aa)
5.	pET28c-Rv3813c-Δ23	N-term His Tag protein where first 23 amino acids were deleted (24-273 aa)

The expected release after the double digestion of the plasmid containing desired gene and the sequencing results confirmed the successful cloning of full length and deletion constructs of Rv3376 and Rv3813c (**Fig. 5.1 and Fig. 5.2).**

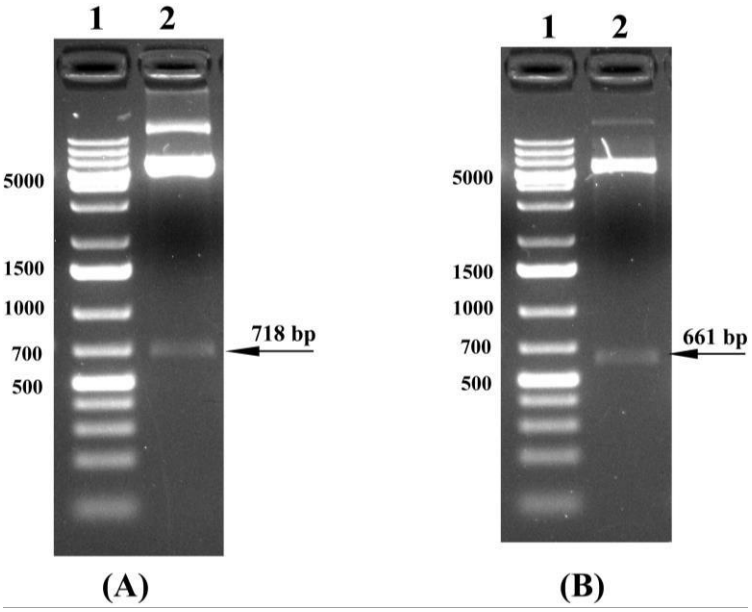


Fig. 5.1 Cloning of Rv3376-FL and Rv3376-Δ17. **(A)** 1% Agarose gel showing the *NdeI* and *HindIII* digested pET28c-Rv3376FL and expected release of 718 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv3376-FL from pET28c vector. **(B)** 1% Agarose gel showing the *NdeI* and *HindIII* digested pET28c-Rv3376-Δ17 with expected release of 661 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv3376-Δ17 from pET28c vector.

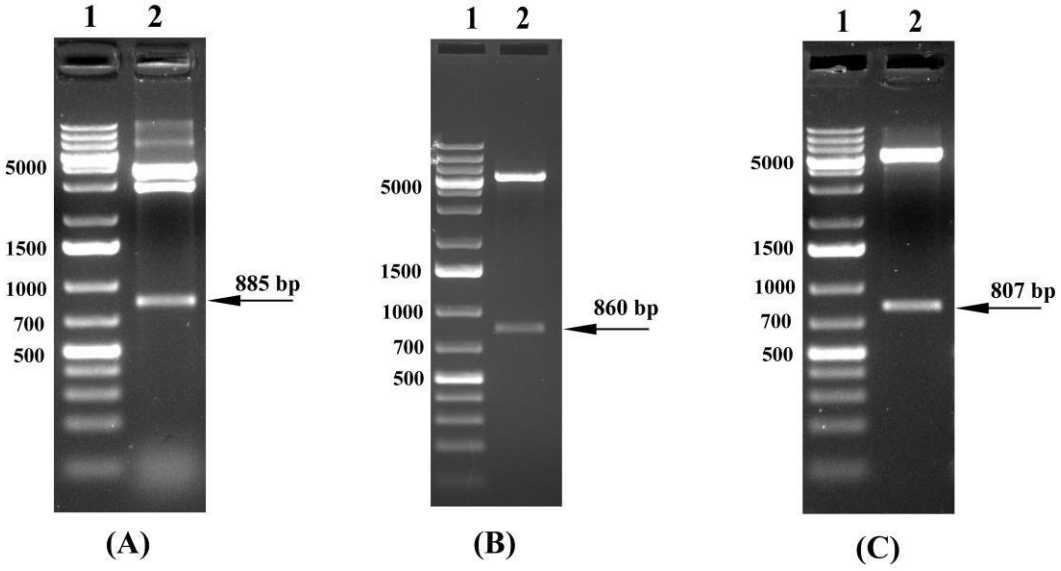


Fig. 5.2. Cloning of Rv3813c-FL, Rv3813c-Δ6 and Rv3813c-Δ23. **(A)** 1% Agarose gel showing the *NdeI* and *XhoI* digested pET28c-Rv3813c-FL and expected release of 885 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv3813c-FL from pET28c vector. **(B)** 1% Agarose gel showing the *NdeI* and *XhoI* digested pET28c-Rv33813c-Δ6 and expected release of 860 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv33813c-Δ6 from pET28c vector. **(C)** 1% Agarose gel showing the *NdeI* and *XhoI* digested pET28c-Rv33813c-Δ23 and expected release of 807 bp. Lane 1: 1 kb DNA ladder, Lane 2: Digested product of Rv33813c-Δ23 from pET28c vector.

5.2.3. Protein expression and purification of Rv3376 and Rv3813c

The plasmid containing Rv3376 gene and the plasmid containing Rv3813c gene were transformed in *E. coli* Rosetta DE3 (Novagen) separately for the expression of the recombinant proteins. The transformed *E. coli* cells were plated on the nutrient agar plates supplemented with kanamycin and chloramphenicol as selection markers and were incubated overnight at 37°C. The primary culture was obtained by inoculating a single colony in 10 ml of LB media containing kanamycin and chloramphenicol and was grown overnight at 37°C with constant shaking at 200 RPM. The 1 % of overnight grown culture was used to inoculate 1L of fresh LB media containing kanamycin and chloramphenicol as secondary culture in a 2L flask. The secondary culture was allowed to incubate at 37 °C in incubator shaker with constant shaking at 200 RPM till the absorbance at 600 nm reached a value of ~ 0.6-0.8. The protein expression was induced by adding 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated further for 3 hrs at 37°C with continuous shaking. The cells were harvested by centrifugation at 8000 g for 10 min at 4°C. The cell pellet was resuspended in a lysis buffer containing 50 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol and 1mM Dithiothreitol (DTT). Also, 1mM phenylmethylsulfonyl fluoride (PMSF) was added along with 1 tablet of cocktail protease inhibitor (Roche, Applied Science, Mannheim, Germany). The lysis of the cells was done by sonication (Sonics, NewTown, CT USA) for 30-45 minutes at pulse rate of 8 seconds ON and 12 seconds OFF cycle with 25 % amplitude. By centrifuging at 12,000 g for 45 min at 4°C, the cell debris and the lysed cells were pelleted down. The supernatant thus obtained contained the soluble protein. The His-tagged protein was purified using His-select Ni-NTA beads (Merk-Sigma Aldrich, USA). The purification of the protein involved the passing of supernatant through the Ni-NTA column which was pre-equilibrated with the lysis buffer followed by washing the beads with the same buffer. The protein of interest was eluted from the column with a lysis buffer containing 20- 500 mM imidazole and each fraction was loaded on 15% SDS-PAGE to check the presence and purity of protein. The size of the purified proteins in SDS-PAGE matched well with the estimated molecular weight of full length and deletion constructs of Rv3376 and Rv3813c **(Fig. 5.3 and Fig. 5.4)**. An Amicon concentrator, (Millipore, USA) with a 10 kDa cut off membrane was used to concentrate the proteins. The concentration of protein was estimated by Denovix nanodrop (DS-11 Series Spectrophotometer). Further, the fractions containing protein were pooled and further purified using gel filtration chromatography using the pre-equilibrated Superdex^{TM} 210/300 GL (GE Healthcare) column in a buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, 10% glycerol and 1mM DTT.

The deletion constructs of Rv3376 named as Rv3376-Δ17 and the deletion constructs of Rv3813c named as Rv3813c-Δ6 and Rv3813c-Δ24 were expressed and purified using the same protocol as that of full-length proteins of Rv3376 and Rv3813c.

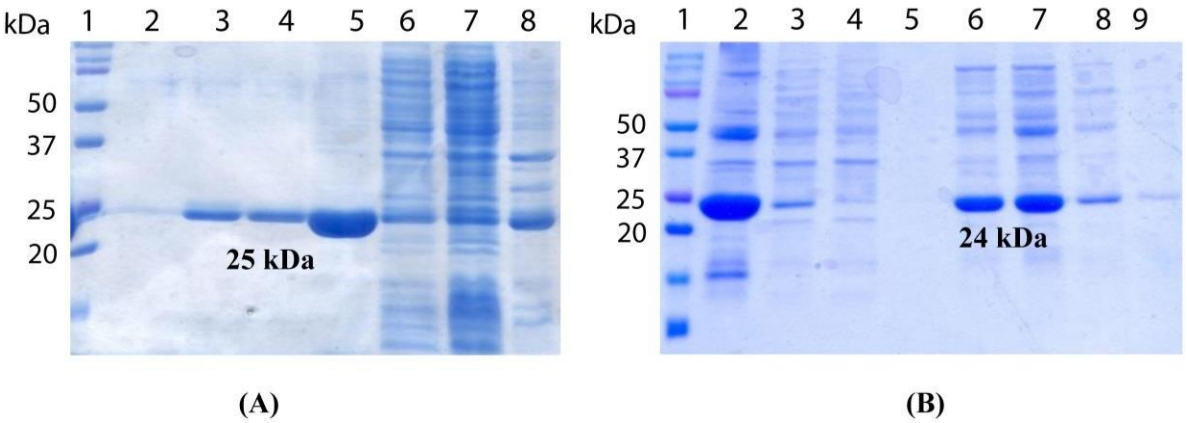


Fig. 5.3 Ni-NTA purification elution profile of Rv3376-FL and Rv3400-Δ17.
(A) 15% SDS-PAGE showing purification profile of Rv3376-FL using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Elution 4, Lane 3: Elution 3, Lane 4: Elution 2, Lane 5: Elution 1, Lane 6: Wash, Lane 7: Flowthrough and Lane 8: Pellet. (B) 15% SDS-PAGE showing purification profile of Rv3376-Δ17 using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Pellet, Lane 3: Supernatant, Lane 4: Flowthrough, Lane 5: Wash, Lane 6: Elution1, Lane 7: Elution 2, Lane 8: Elution 3 and Lane 9: Elution 4.

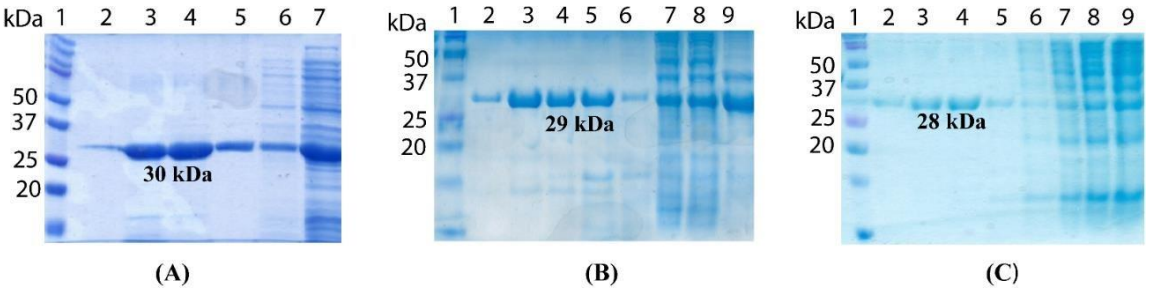


Fig. 5.4 Ni-NTA purification elution profile of Rv3813c-FL, Rv3813c-Δ6 and Rv3813c-Δ24.
(A) 15% SDS-PAGE showing purification profile of Rv3813c-FL using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Elution 4, Lane 3: Elution 3, Lane 4: Elution 2, Lane 5: Elution 1, Lane 6: Flowthrough and Lane 7: Pellet. (B) 15% SDS-PAGE showing purification profile of Rv3813c-Δ6 using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Elution 5, Lane 3: Elution 4, Lane 4: Elution 3, Lane 5: Elution 2, Lane 6: Elution 1, Lane 7: Flowthrough and Lane 8: Supernatant, Lane 9: Pellet. (C) 15% SDS-PAGE showing purification profile of Rv3813c-Δ24 using Ni-NTA chromatography. Lane1: Molecular weight marker (Biorad), Lane 2: Elution 5, Lane 3: Elution 4, Lane 4: Elution 3, Lane 5: Elution 2, Lane 6: Elution 1, Lane 7: Flowthrough and Lane 8: Supernatant, Lane 9: Pellet.

5.2.4. Determination of the oligomeric state of Rv3376 and Rv3813c

The oligomeric state of Rv3376 and Rv3813c were determined using the prepacked analytical size-exclusion chromatography Superdex-200 column (GE Healthcare, USA). The purified protein from Ni-NTA chromatography was injected into the superdex-200 column and the elution profile was compared with that of standard proteins. The results indicates that the Rv3376 and Rv3813c proteins exists as monomer in solution (**Fig. 5.5).** Similarly, all the variants of Rv3376 and Rv3813c proteins were subjected to the gel filtration chromatography using the Superdex-200 column (GE Healthcare, USA), which showed same profile as that of the wild type Rv3376.

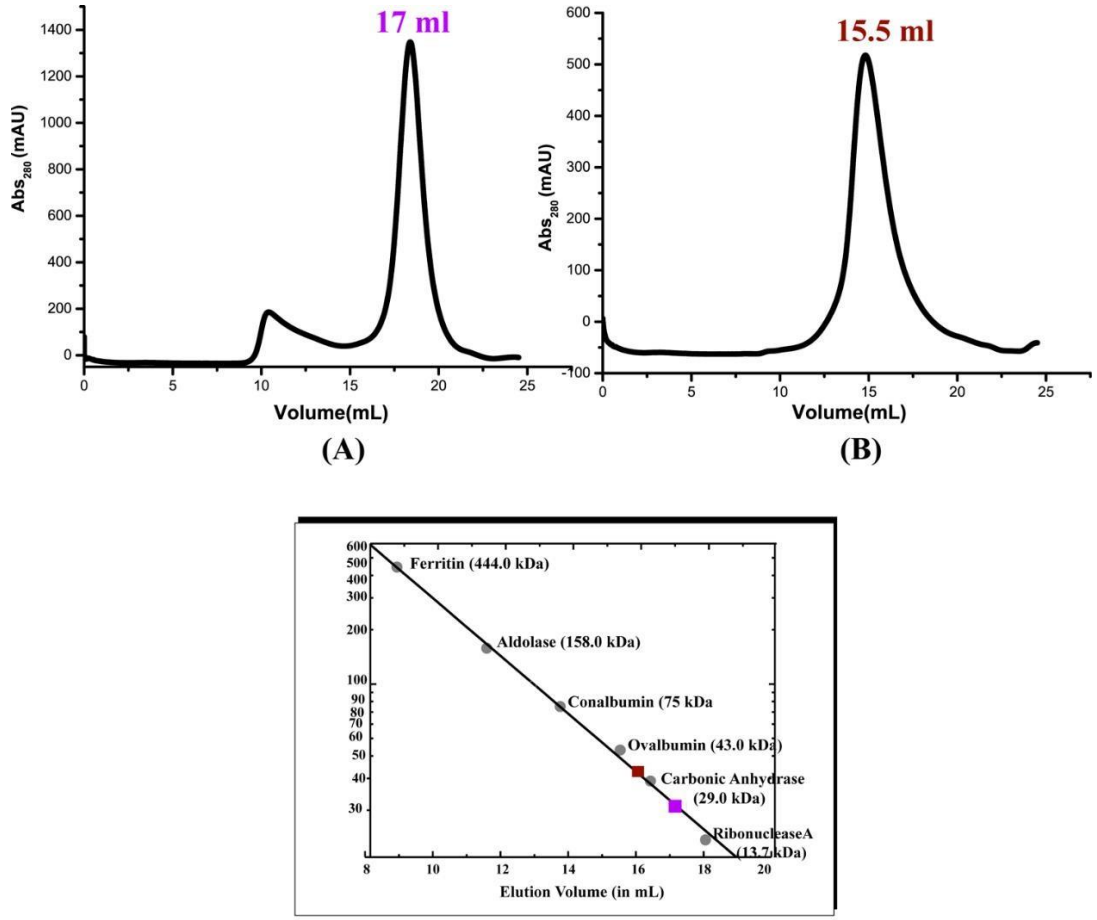


Fig. 5.5 Oligomeric status of Rv3376 and Rv3813c determined using analytical size exclusion chromatography (A) Analytical size exclusion chromatography (ASEC) profile of Rv3376. The peak eluted at ~17 ml corresponds to 27 kDa and matches well with the calculated molecular weight of Rv3376 (27 kDa) when compared to the standard curve. The ASEC experiments suggested that the Rv3376 exists as monomer in solution **(B)** Analytical size exclusion chromatography profile of Rv3813c. The peak eluted at ~15.5 ml corresponds to the 30 kDa and matches with the calculated molecular weight of Rv3813c (30 kDa) when

compared to the standard curve. The ASEC experiments suggested that the Rv3813c exists as monomer in solution **(C)** The inset shows the standard calibration curve for Superdex-200 column (GE Healthcare, USA).

5.2.5. MALDI-TOF studies

The intact molecular mass of Rv3376 and Rv3813c were determined by MALDI-TOF (Matrix Assisted Laser Desorption Ionization-Time of Flight) mass spectrometry. The concentration of protein sample was about 0.2 mg/ml and sinapinic acid was used as a matrix. The calculated mass of Rv3376 protein expressed with 19 extra amino acids (MGSSHHHHHHSSGLVPRGSH) is 25809.08 Da, after removal of formyl methionine. The observed molecular weight of Rv3376 in MALDI experiment was comparable to the calculated molecular weight (**Fig.5.6.).**

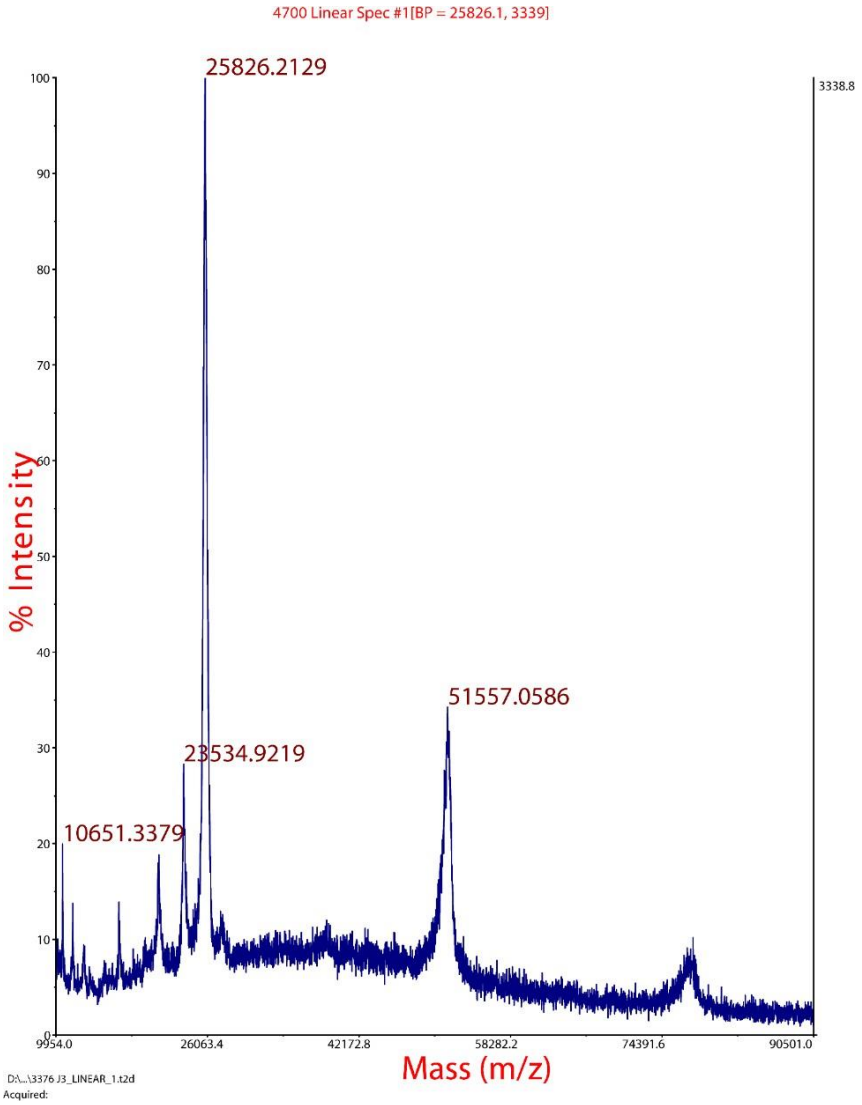


Fig. 5.6 MALDI-TOF analysis of Rv3376 - The major peak observed at ~25826.2 Da corresponds to Rv3376.

For Rv3813c, the calculated molecular mass of Rv3813c with 19 extra amino acids (MGSSHHHHHHSSGLVPRGSH) is 30634 Da, after removal of formyl methionine. The observed molecular weight of Rv3376 in MALDI experiment was comparable to the calculated molecular weight (**Fig. 5.7).**

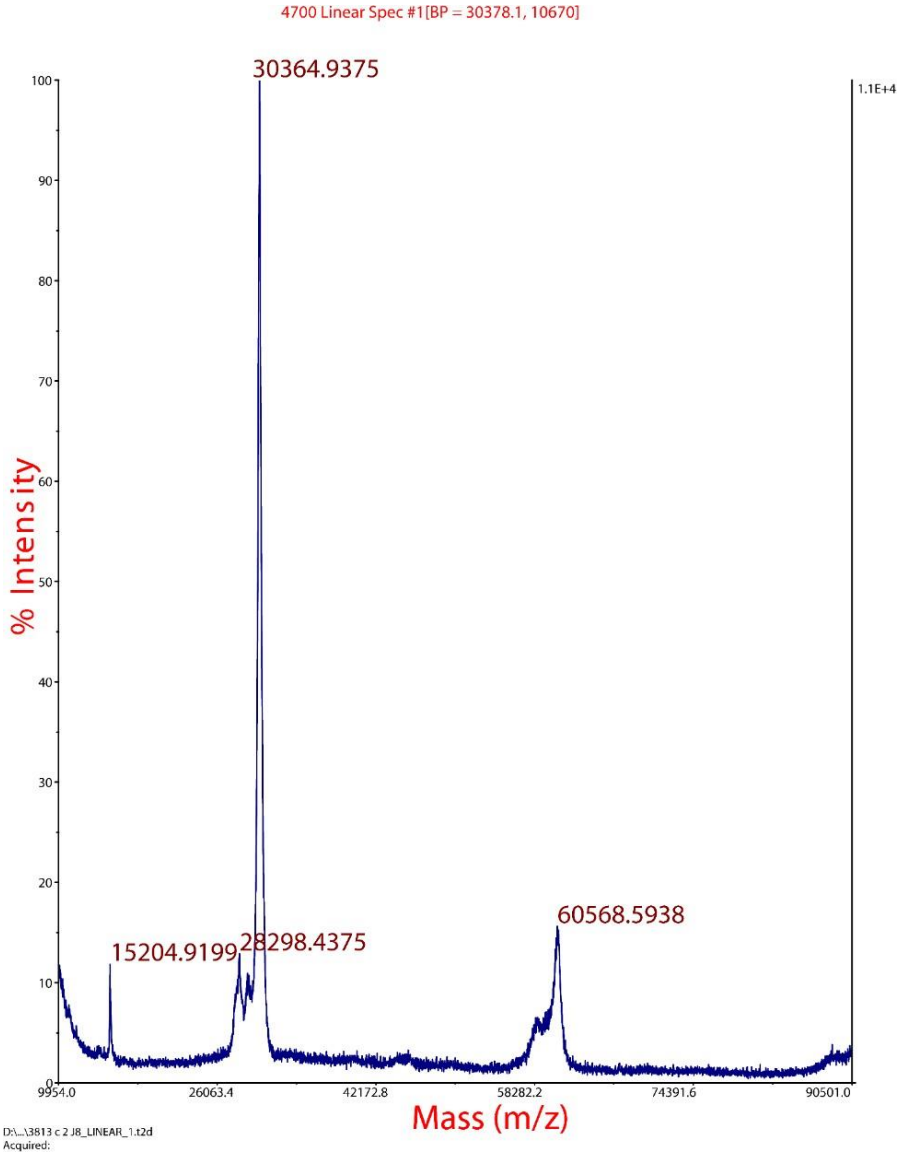


Fig. 5.7 MALDI-TOF analysis of Rv3813c - The major peak observed at ~30364.9 Da corresponds to Rv3813c.

5.2.6. Circular dichroism studies

The circular dichroism (CD) studies of Rv3376 and Rv3813c proteins were done to check the folding and secondary structure content of the proteins. The far-UV CD spectra were collected

on a Jasco–810 CD spectropolarimeter that was flushed with nitrogen at the rate of 9–12 L/min. A quartz cuvette with a path length of 1 mm was used to obtain the data, which ranged from 250 nm to 198 nm. Each obtained spectra consisted of an average of five scans and was recorded as raw ellipticity. Data collection for all protein samples utilised a concentration of 0.2 mg/ml. The mean residue ellipticity (MRE) was calculated by raw ellipticity according to the equation,

$$[\theta]_{MRE} = ([\theta] \text{ X } 100 \text{ X } M_r) /(c \text{ x } l \text{ x } N_A)$$

where $[\theta]_{MRE}$ depicts the mean residue ellipticity, [θ] stands for the raw ellipticity in degrees, Mr is molecular weight of the protein (in Da), l is the path length of the cuvette (in cm), c is the concentration of the protein (in mg/ml), and the number of residues of the protein is depicted by N_A.

To investigate the secondary structure content and protein folding of Rv3376 and Rv3813c, far-UV CD studies was done. The secondary structure content of Rv3376 was determined as 96.7% α- helices. Rv3813c had an α- helical content of 47.5% and β-sheet content of 52.5% as predicted by the BeStSel web server (**Fig. 5.8)** (Micsonai et al., 2018).

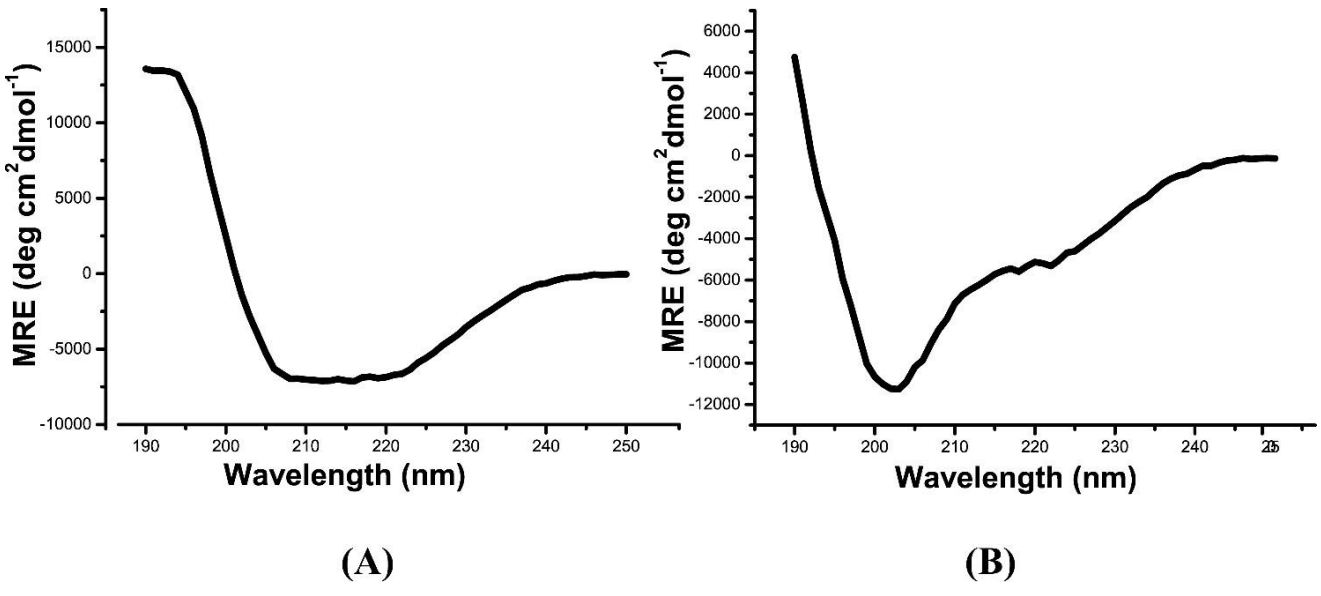


Fig. 5.8 CD analysis of Rv3376-FL and Rv3813c-FL. The far-UV CD spectra showing the folding and secondary structure content of the Rv3376 (A) and Rv3813c (B).

5.2.7. Crystallization trials for Rv3376 and Rv3813c

The purified and concentrated proteins of Rv3376, Rv3813c and their deletion variants were subjected to crystallization trials using commercially available screens from Hampton Research, USA and Molecular Dimensions, UK. A high-throughput 96-well sitting-drop crystallization trays were set up using an NT8 robotics system (Formulatrix Inc., USA). The drops contained 150 nl of protein sample and 150 nl of reservoir solution. The reservoir volume in Swiscci plate was 40µl. The crystallization trays were incubated at 20°C in a Rock Imager 1000 (Formulatrix) for storage and automatic pre-scheduled imaging. Despite trying several attempts, we could not crystallize Rv3376 and Rv3813c proteins. The deletion constructs were made to aid in crystallization but no success in getting the crystals.

5.2.8. ARPP Phosphatase assay

In riboflavin biosynthesis, 5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate (ARPP) is formed after the sequential action of GTP cyclohydrolaseII and Pyrimidine Deaminase/Reductase on GTP. It then goes through additional dephosphorylation by an unknown phosphatase to form 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (ARP). To find whether Rv3376 and Rv3813c can play a role in the conversion of ARPP to ARP, two strategies were employed. In strategy I, the required ARPP was synthesized by using recombinant GTP cyclohydrolase II (ribA) and pyrimidine deaminase/reductase (ribD) proteins via an enzymatic conversion from GTP in a three-step process (Richter et al., 1997). A 500 µl reaction mixture containing 10 mM Tris-HCl, pH 8.0, 8 mM $MgCl_2$, 3 mM dithiothreitol (DTT), 2 mM NADPH, 5 mM GTP, 100 µg *H. pylori* ribA, and 50 µg *V. cholera* ribD were used for the synthesis of ARPP. The reaction mixture was incubated at 37°C for 30-60 minutes. The proteins were removed by passing the reaction mixture through a centrifugal filter with a 3000 Da molecular weight cut-off. The formation of ARPP was confirmed by spectrophotometric assay (Sa et al., 2016). The enzymatically produced ARPP was then incubated with Rv3376 or Rv3813c and the peak corresponding ARP (277Da) was monitored using liquid chromatography-mass spectrometry (LC-MS). The analysis of LC-MS data revealed that no peak was observed that corresponds the molecular weight of ARP (**Fig. 5.9).**

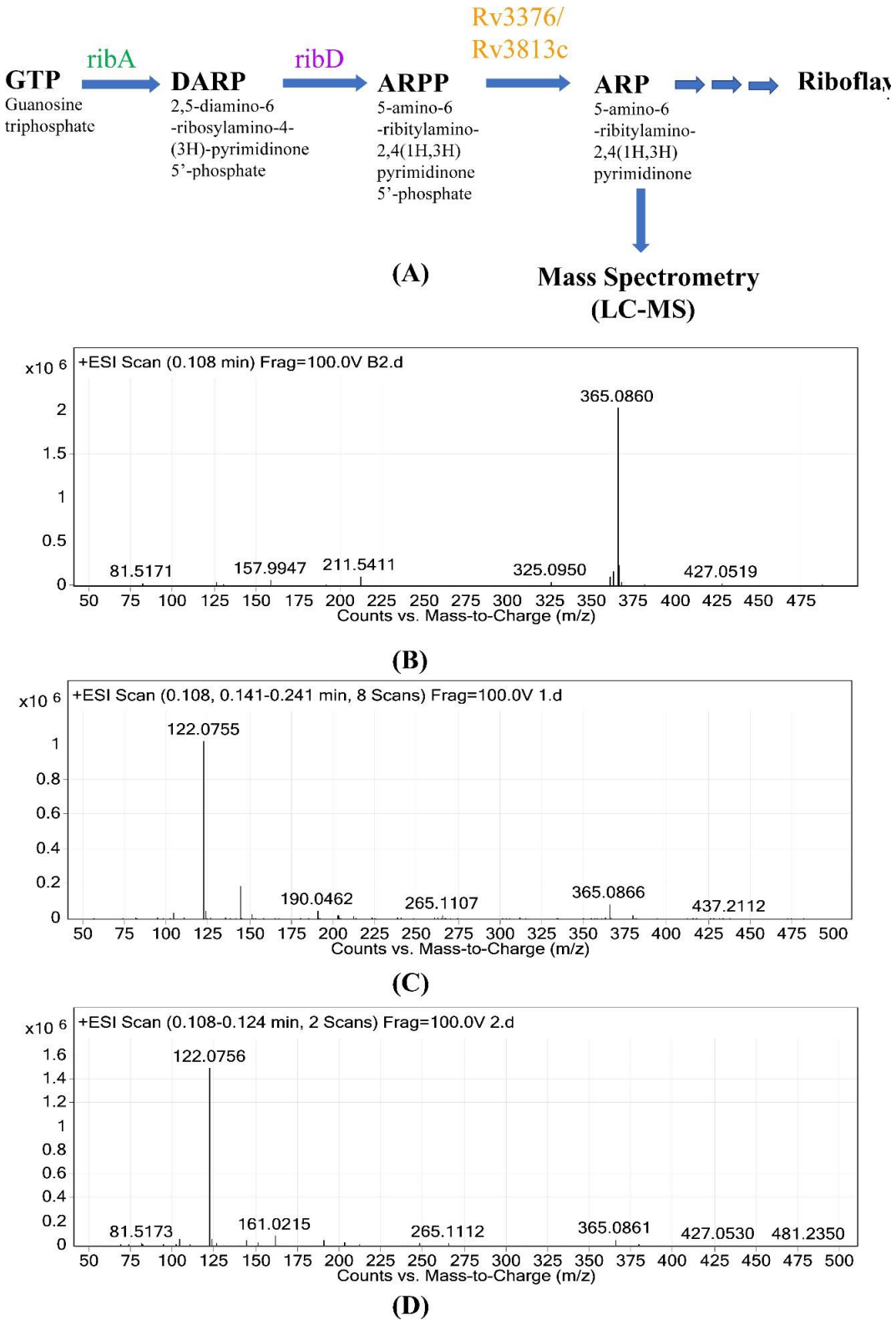


Fig. 5.9 Role of Rv3376 and Rv3813c in riboflavin biosynthesis pathway. (A) Schematic diagram showing the Strategy I and depicting the workflow for the formation of ARPP from GTP and dephosphorylation by the potential phosphatase to form ARP. LC-MS analysis revealed that no peak corresponding to molecular weight of ARP (~277 Da) was observed when ARPP was incubated with either Rv3376 **(C)** or Rv3813c **(D).**

Strategy (II) was attempted where all the enzymes involved in riboflavin biosynthesis pathway were used for the formation of riboflavin along with the potential phosphatase. For this strategy we expressed and purified all the enzymes involved in the pathway that has been previously cloned in our lab, except pyrimidine phosphatase. The purified enzymes ribA from *H. pylori* (Yadav & Karthikeyan, 2015) , ribB (Islam et al., 2015), ribD, ribH and ribE (unpublished data from the thesis of Zeyaul Islam entitled Structure-Functionl analysis of enzymes involved in riboflavin biosynthesis pathway of *V. cholerae*) were incubated with the substrate GTP and Ribulose-5-phosphate along with the potential phosphatase. The reaction mixture contained 1mM GTP, 2mM Ru5P, 50mM Tris HCl, 10mM $MgCl_2$, 3mM DTT, 50µg each of ribA, ribB, ribD, ribH and ribE and 50 µg of Rv3376 or Rv3813c. The reaction was incubated at 37°C for 1-2 hours for the formation of riboflavin, a yellow-coloured compound that could further be verified spectrophotometrically at 440 nm. The reaction containing the potential phosphatases Rv3376 and Rv3813c along with other components was unable to produce yellow colour or riboflavin as the end product. The results indicated that Rv3376 and Rv3813c were not the showing any phosphatase activity in the riboflavin biosynthesis pathway (**Fig. 5.10)**.

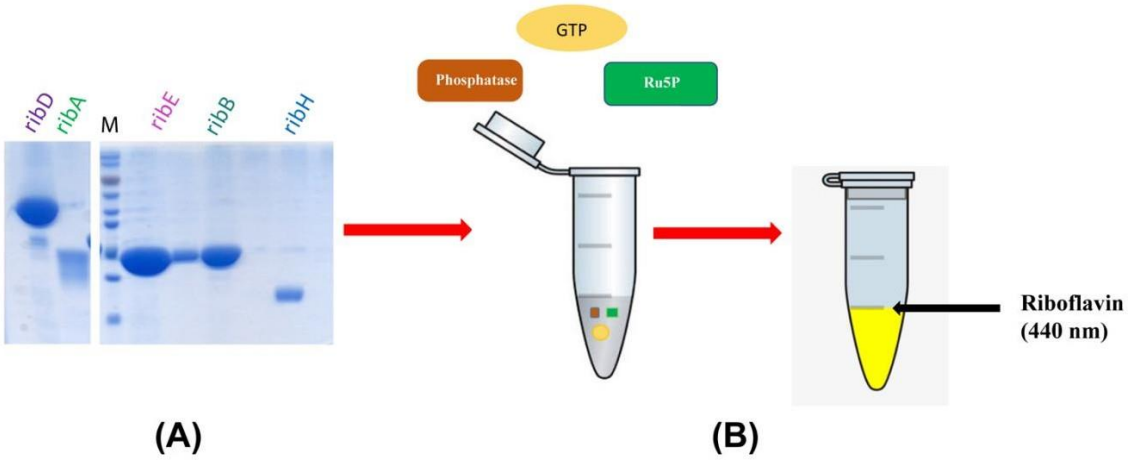


Fig. 5.10 Role of Rv3376 and Rv3813c in riboflavin biosynthesis pathway. Strategy II depicting the workflow. **(A)** All the enzymes of RBS pathway were purified by Ni-NTA chromatography and GFC. **(B)** The enzymes were incubated with the substrates GTP and Ru5P and with the potential phosphatases Rv3376 or Rv3813c enzymes.

5.2.9. Phosphatase activity assay

To check the general phosphatase activity of Rv3376 and Rv3813c, para-Nitrophenyl Phosphate (pNPP) assay was used. pNPP is a non-proteinaceous chromogenic substrate for alkaline and acid phosphatases. The hydrolysis of pNPP by phosphatases releases an inorganic phosphate and the conjugate base of para-nitrophenol (pNP). The pNP is a yellow-coloured compound that has a maximum absorbance at 405 nm. pNPP (Thermofischer) was dissolved in a buffer containing 100mM Tris (pH 7-9), 100mM NaCl and 5mM $MgCl_2$. The Rv3376 and Rv3813c proteins at different concentrations (1-50 µM) were incubated with pNPP separately. The formation of yellow colour was monitored spectrophotometrically at 405 nm. Alkaline phosphatase was used as the positive control to check the formation of pNP which shows an increase in the absorbance at 405 nm (**Fig. 5.11).** However, for both the proteins Rv3376 and Rv3813c, no increase in absorbance was observed that corresponds to pNP. The results indicated that these enzymes lack the phosphatase activity and thus unable to catalyze the dephosphorylation of pNPP.

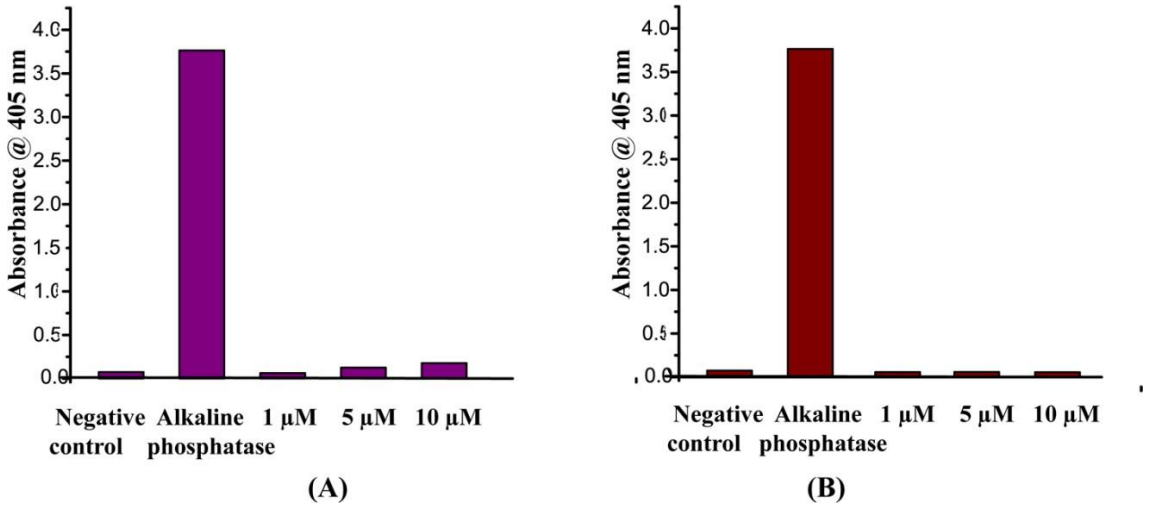


Fig. 5.11 Phosphatase activity of Rv3376 (A) and Rv3813c (B). Phosphatase activity of Rv3376 and Rv3813c using general substrate para-Nitrophenyl Phosphate (pNPP). Negative control is the reaction mixture containing buffer and substrates except protein, AP-Alkaline phosphatase, a known phosphatase used as positive control, and different concentrations of Rv3376 and Rv3813c ranges from 1µM, 5 µM and 10 µM were used. (The experiments were done only once to check the activity of the proteins).

5.2.10. Pyrophosphatase activity assay

Pyrophosphatase activity of the proteins were determined using a sigma pyrophosphate assay kit (CAT-MAK168). For the assay, pyrophosphate concentration of sample is determined by the use of unique fluorogenic pyrophosphate sensor (PPi sensor). The PPi sensor in presence of pyrophosphate results in the production of fluorescent product (λ_{ex}=316/λ_{em}=456 nm) proportional to the pyrophosphate present in the solution. The reaction was carried out in 96 well flat bottom black ELISA plates. Sodium pyrophosphate was used as a substrate and different concentrations of Rv3376 and Rv3813c proteins (1µM-5µM) were used. The reaction mixture consists of 100 mM Tris pH 9.5, 100 mM NaCl and 5mM $MgCl_2$ and Rv3376 or Rv3813c. The reaction was incubated for 30 minutes at 37°C. After incubation, 50 µl of reaction mixture was mixed with 50 µl master mix (assay buffer + pyrophosphate sensor stock solution, provided in the kit) and incubated further at room temperature for 10-30 minutes as per the manufacture's protocol. The pyrophosphate release in each reaction was measured at 316 nm and 456nm to check the pyrophosphatase activity of protein. The results suggested that Rv3376 and Rv3813c were unable to convert pyrophosphate to orthophosphate indicating devoid of any pyrophosphatase activity (**Fig. 5.12).**

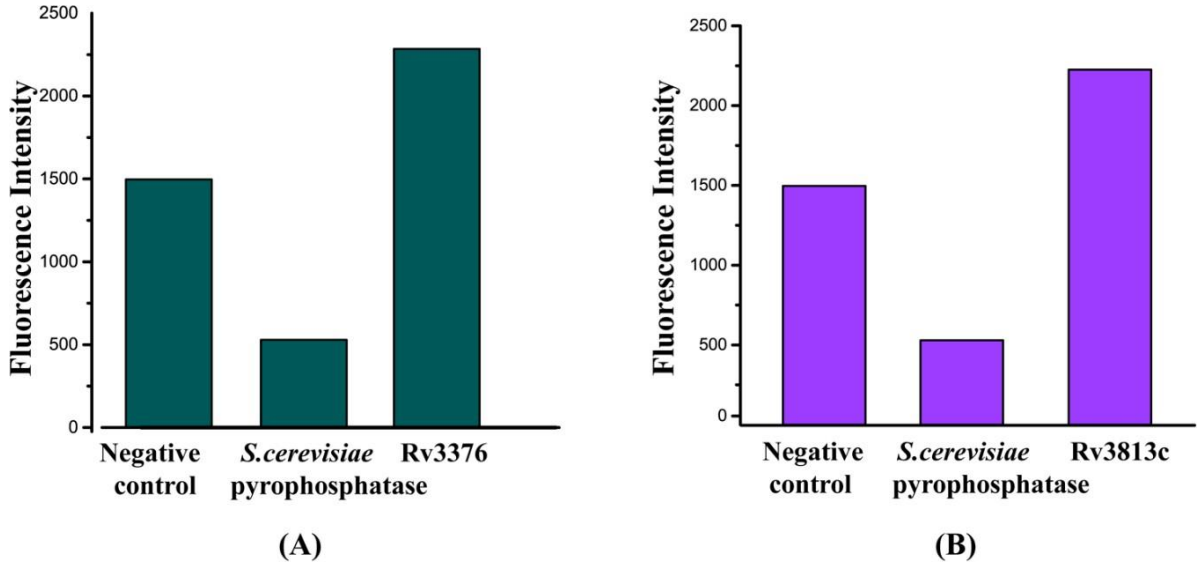


Fig. 5.12 Pyrophosphatase activity of Rv3376 (A) and Rv3813c (B). Pyrophosphatase activity of protein was determined by using substrate Sodium pyrophosphate. The negative control contained the buffers and substrate except the protein, Pyrophosphatase from *S. cerevisiae* was used as positive control and different concentrations of Rv3376 and Rv3813c (1µM-5µM) were used in the assay (data here shown only for 5 µM). (The experiments were done only once to check the activity of the proteins).

5.2.11. Beta-phosphoglucomutase activity (β-PGM) of Rv3376 and Rv3813c

The Beta-phosphoglucomutase (β-PGM) catalyzes the conversion of β-D-Glucose 1-phosphate (β-G1P) to β-D-Glucose 6-phosphate (β-G6P). The substrate (β-G1P) was synthesized enzymatically by the action of Maltose phosphorylase (MP) (EC.2.4.1.8) on maltose (Nilsson & Radstrom, 2001) as described by in the previous chapter. To check the β-PGM activity, the proteins Rv3376 and Rv3813c were incubated with the reaction mixture containing substrate (β-G1P). The reaction mixture contained 50 mM HEPES (pH 7.5), 450 µM β-D-Glucose 1-phosphate, 2 mM $MgCl_2$, 30 µM α –glucose 1,6-bisphosphate, 0.4 mM NADP, 15U of glucose 6-phosphate dehydrogenase (G6PD) (HIMEDIA, India). The reaction was started by adding 50 µg of the protein. It was a coupled reaction in which product β-G6P was catalyzed by the enzyme G6PD and led to the reduction of NADP. Thus, the increase in the absorbance of NADPH which is directly proportional to the formation of product β-6P was monitored at 340 nm for 15-30 minutes using UV–Visual spectrophotometer (Perkin Elmer Lambda 25, USA). Both the proteins do not show any increase in the absorbance at 340 nm corresponded to the formation of NADPH, indicating that the Rv3376 and Rv3813c lack β-PGM activity (**Fig. 5.13).**

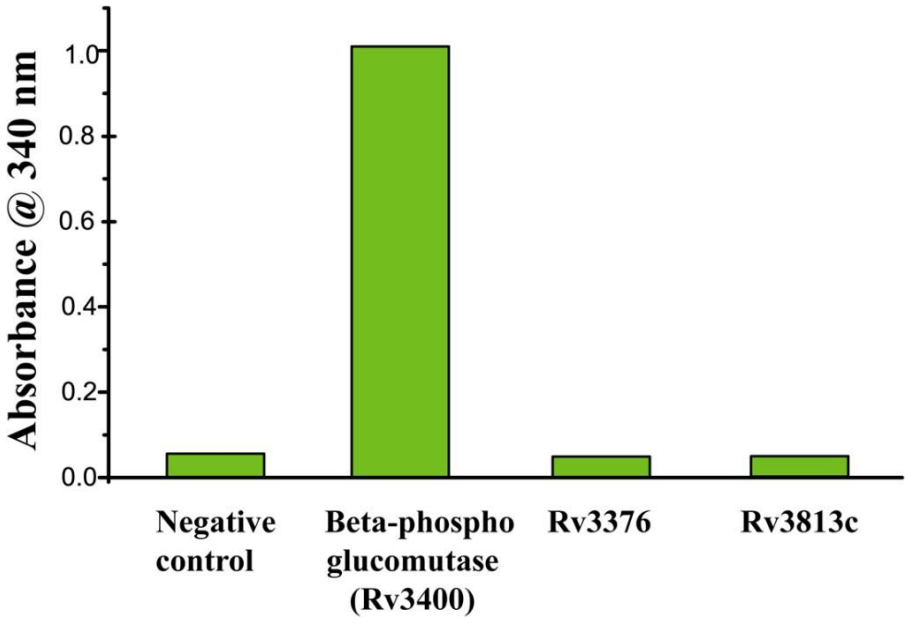


Fig. 5.13 β-PGM activity of Rv3376 and Rv3813c. β-PGM activity of proteins using substrate β-G1P which was synthesized from maltose. The negative control is reaction mixture without protein and Rv3400, a β-PGM was used as positive control. Different concentrations of Rv3376 and Rv3813c proteins (1µM-5µM) were used (data here shown only for 5 µM). (The experiments were done only once to check the activity of the proteins).

5.3. Discussion

The riboflavin biosynthesis pathway is well studied in various microorganisms like *Bacillus, E. coli,* Mtb etc. Usually, the genes involved in the riboflavin biosynthesis are clustered in an operon known as *rib* operon. The genes belong to the operon maybe monofunctional or bifunctional. The genes involved in the riboflavin pathway are- *ribA, ribB, ribD, ribH* and *ribE*. The functional and structural characterization of these genes has been reported. However, an unknown phosphatase is also involved in the pathway which has not yet been characterized. So, we were interested in the identification of phosphatase from Mtb and aim to target the riboflavin biosynthesis pathway for drug discovery. In the literature, members of haloacid dehalogenase family were found to catalyze the dephosphorylation step of RBS pathway. Two genes yigB and ybjl from *E. coli* (Haase et al., 2013), three genes ycsE, ywtE and yitU from *Bacillus subtilis* (Sarge et al., 2015) and AtPyrP2 and AtGpp1/PyrP3 from *Arabidopsis thaliana* (Sa et al., 2016) are reported to catalyze the dephosphorylation step converting (ARPP to ARP) in riboflavin biosynthesis pathway. However, BLAST search to identify a homolog of these genes in Mtb database did not yield any result. Therefore, a PSI-BLAST search using the reported phosphatase of riboflavin biosynthesis was done to find the homologs in Mtb. The results revealed at least three genes namely - *Rv3400, Rv3376* and *Rv3813c* that showed a weak sequence similarity range from 27-30 %. We have discussed in detail about Rv3400 in the previous chapter (chapter 4). The amino acid sequence analysis of Rv3376 and Rv3813c using bioinformatics tools predicted that these proteins belong to the haloacid dehalogenase family (HAD). These proteins showed the presence of four short signature motifs that contain the conserved catalytic residues exclusively present in the HAD family.

To characterize the potential phosphatases Rv3376 and Rv3813c and to understand their role in riboflavin biosynthesis pathway, we initiated the studies on these two proteins. We cloned, expressed and purified the potential phosphatase Rv3376 and Rv3813c from Mtb. Different strategies were employed to check the role of the Rv3376 and Rv3813c in RBS pathway. Rv3376 and Rv3813c were failed to dephosphorylate the substrate ARPP to ARP and therefore we concluded that they did not have any role in RBS pathway. We also checked the phosphatase activity of Rv3376 and Rv3813c using a general substrate of phosphatases i.e., pNPP but these proteins were unable to catalyze its dephosphorylation. Since Rv3376 and Rv3813c belongs to HAD family which involves pyrophosphatases and β-phosphoglucomutases, we assessed their role in catalyzing these reactions. The results from

the activity assay suggests that the proteins Rv3376 and Rv3813c lacks both pyrophosphatases and β- phosphoglucomutases activity.

Structural characterization was also initiated to assign a structure-based function determination for both Rv3376 and Rv3813c. Despite trying several attempts, we could not crystallize these proteins. The deletion constructs were made to aid in crystallization but no success in getting the crystals.

Thus, based on the results obtained, we conclude that both Rv3376 and Rv3813c does not have phosphatase activity to catalyze substrates ARPP (riboflavin biosynthesis pathway) and pNPP (general substrate). Moreover, these proteins also lack pyrophosphatases and β-phosphoglucomutases activity. Further studies were required to characterize these proteins structurally and biochemically.

Summary and Future Directions

Summary

Despite being one of the oldest diseases, tuberculosis (TB) is still wreaking havoc in the 21st century. TB, caused by *Mycobacterium tuberculosis* (Mtb), is an infectious disease that can severely affect the respiratory system of humans and also other parts of the body, e.g., the brain, the kidneys, and the spine. TB is a life-threatening disease if not treated well or on time. The prevalence of antibiotic resistance developed by Mtb against the current available treatment poses a threat to mankind. The emergence of multidrug resistant (MDR) and extensively drug resistant (XDR) strains of Mtb has increased the number of deaths worldwide. So, there is an urgent need to find new drugs that are more potent and efficient than existing ones to cope up with the resistant strains of Mtb. The libraries of random transposon mutant and completely saturated transposon library experiments have identified the proteins and genes that are essential for the survival of Mtb. In both the experiments, the enzymes involved in riboflavin biosynthesis (vitamin B2) pathway are shown to be essential for the growth of Mtb. Notably, the Mtb lacks transporters for the riboflavin uptake and required to produce it endogenously, making the enzymes of the riboflavin biosynthesis pathway as potential drug targets. Importantly, this pathway is absent in humans as they uptake riboflavin from their food. The endogenous biosynthesis of riboflavin in Mtb is coordinated by five enzymes using GTP and Ribulose-5-phosphate as substrates. Except for an unknown phosphatase, all four enzymes (ribA2, ribD, ribH and ribE) of riboflavin biosynthesis pathway are well studied. However, the phosphatase that catalyzes the dephosphorylation of an intermediate ARPP to ARP, is still elusive and remains to be identified in Mtb.

Thus, we initiated our study to identify the unknown phosphatase that is involved in riboflavin biosynthesis pathway of Mtb. The literature survey suggested that the genes, *yigB* and *ybjl* from *E. coli* (Haase et al., 2013), *ycsE, ywtE, and yitU* from *Bacillus subtilis* (Sarge et al., 2015) , and *AtPyrP2* and *AtGpp1/PyrP3* from *Arabidopsis thaliana* (Sa et al., 2016), are reported to catalyze the dephosphorylation step in the riboflavin biosynthesis pathway in different organisms. Using the amino acid sequences from the above species and through bioinformatics tools we predicted that three proteins Rv3400, Rv3376, and Rv3813c may play a role as phosphatases in riboflavin biosynthesis pathway in Mtb. Although, these three proteins are not essential for the growth of Mtb, they are shown to be important for its virulence. Importantly, these three proteins are not yet characterized in Mtb. Therefore, to

characterize these proteins and to check whether Rv3400, Rv3376, and Rv3813c possess any phosphatase activity as we predicted, we cloned these three genes and expressed them in *E. coli* system. All three proteins were purified by affinity chromatography followed by gel filtration chromatography and checked for their phosphatase activity. In our experimental setup none of these proteins showed any phosphatase activity, however, Rv3400 showed β-PGM activity. Since Rv3400 has not been characterized yet, we carried out the characterization of this enzyme, further. Our biochemical studies confirmed that Rv3400 indeed encodes for β-PGM and catalyzes the conversion of β-D-glucose 1-phosphate (G1P) to β-D-glucose 6-phosphate (G6P), via β-D-glucose 1,6-bisphosphate (G16BP) as an intermediate. In addition, the crystal structure of Rv3400 determined at 1.7 Å by SAD method revealed the overall fold and active site residues involved in the catalysis. Based on the biochemical and structural characterization we have also proposed a catalytic mechanism for Rv3400.

As noted earlier, Rv3400 is not essential for the growth but shown to be essential for the infection. To understand the role of Rv3400 during infection we further explored different metabolic pathways of Mtb. Rv3400 catalyzes the formation of G6P, which is further required for the generation of ATP and NADPH through glycolysis and pentose phosphate pathway, respectively. Hence, Rv3400 may play a crucial role in utilizing alternate sugars, i.e., trehalose as a carbon/energy source during starvation.

To find the role of other two potential phosphatases Rv3376 and Rv3813c in riboflavin biosynthesis pathway, they were cloned, expressed and purified. However, in our experimental setup Rv3376 and Rv3813c proteins were not able to catalyze the dephosphorylation of ARPP to ARP, indicating that they are not involved in the riboflavin biosynthesis pathway. In fact, different activity assays were performed to determine their functions, e.g., beta-phosphoglucomutase and pyrophosphates, but both proteins were unable to catalyze these activities. Furthermore, attempts were made to crystallize Rv3376 and Rv3813c proteins to structurally characterize them. However, despite different construct designs (mutation, deletion etc.) and rigorous screening, no crystals were obtained for Rv3376 and Rv3813c proteins. Therefore, further studies are required to know the function of Rv3376 and Rv3813c and their role in Mtb physiology.

Future directions

- In the present study, we attempted to find the missing enzyme of the riboflavin biosynthesis pathway (RBS) in *M. tuberculosis*, but despite of several attempts and strategies, we did not find the role of shortlisted genes Rv3400, Rv3376 and Rv3813c in dephosphorylation. So, we may explore other genes that will catalyse the dephosphorylation step of the RBS pathway.

- So far, this is the first study where the crystal structure of a beta-phosphoglucomutase (β-PGM) from Mtb has been solved. We are able to solve the crystal structure of Rv3400 at 1.7 Å resolution using Se-SAD method in apo form. However, for more mechanistic insights, the crystal structure of Rv3400 in presence of ligands, substrate and intermediate are required. Such structures will also help in understanding the conformational differences during the catalysis.

- Based on the available structural and functional information about the β-PGM, we propose its role in trehalose metabolism of Mtb. However further *in vivo* studies are required to confirm the role of Rv3400 and trehalose metabolism in the physiology of Mtb.

- In the present study, efforts were made to characterize the Rv3376 and Rv3813c proteins. We were successful in cloning, expression and purification of these recombinant proteins. Our initial studies suggests that these predicted potential phosphatases do not appear to be involved in dephosphorylation of ARPP in riboflavin biosynthesis pathway. Therefore, further studies are required to identify the function of Rv3376 and Rv3813c and their role in Mtb.

Bibliography

Abbas, C. A., & Sibirny, A. A. (2011). Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. *Microbiol Mol Biol Rev, 75*(2), 321-360. doi:10.1128/MMBR.00030-10

Abrusan, G., & Marsh, J. A. (2016). Alpha Helices Are More Robust to Mutations than Beta Strands. *PLoS Comput Biol, 12*(12), e1005242. doi:10.1371/journal.pcbi.1005242

Achari, A., Somers, D. O., Champness, J. N., Bryant, P. K., Rosemond, J., & Stammers, D. K. (1997). Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nat Struct Biol, 4*(6), 490-497. doi:10.1038/nsb0697-490

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221. doi:10.1107/S0907444909052925

Alderwick, L. J., Harrison, J., Lloyd, G. S., & Birch, H. L. (2015). The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. *Cold Spring Harb Perspect Med, 5*(8), a021113. doi:10.1101/cshperspect.a021113

Allen, K. N., & Dunaway-Mariano, D. (2009). Markers of fitness in a successful enzyme superfamily. *Curr Opin Struct Biol, 19*(6), 658-665. doi:10.1016/j.sbi.2009.09.008

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402. doi:10.1093/nar/25.17.3389

Andersson, U., & Radstrom, P. (2002). Beta-glucose 1-phosphate-interconverting enzymes in maltose- and trehalose-fermenting lactic acid bacteria. *Environ Microbiol, 4*(2), 81-88. doi:10.1046/j.1462-2920.2002.00268.x

Aravind, L., Galperin, M. Y., & Koonin, E. V. (1998). The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem Sci, 23*(4), 127-129. doi:10.1016/s0968-0004(98)01189-x

Armon, A., Graur, D., & Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information11Edited by F. Cohen. *Journal of Molecular Biology, 307*(1), 447-463. doi:https://doi.org/10.1006/jmbi.2000.4474

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. *Front Bioeng Biotechnol, 8*, 570828. doi:10.3389/fbioe.2020.570828

Bacher, A., Eberhardt, S., Fischer, M., Kis, K., & Richter, G. (2000). Biosynthesis of vitamin b2 (riboflavin). *Annu Rev Nutr, 20*, 153-167. doi:10.1146/annurev.nutr.20.1.153

Backe, P. H., Laerdahl, J. K., Kittelsen, L. S., Dalhus, B., Morkrid, L., & Bjoras, M. (2020). Structural basis for substrate and product recognition in human phosphoglucomutase-1 (PGM1) isoform 2, a member of the alpha-D-phosphohexomutase superfamily. *Sci Rep, 10*(1), 5656. doi:10.1038/s41598-020-62548-0

Baker, A. S., Ciocci, M. J., Metcalf, W. W., Kim, J., Babbitt, P. C., Wanner, B. L., . . . Dunaway-Mariano, D. (1998). Insights into the mechanism of catalysis by the P-C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry, 37*(26), 9305-9315. doi:10.1021/bi972677d

Barrozo, A., Liao, Q., Esguerra, M., Marloie, G., Florian, J., Williams, N. H., & Kamerlin, S. C. L. (2018a). Computer simulations of the catalytic mechanism of wild-type and mutant beta-phosphoglucomutase. *Org Biomol Chem, 16*(12), 2060-2073. doi:10.1039/c8ob00312b

Barrozo, A., Liao, Q., Esguerra, M., Marloie, G., Florian, J., Williams, N. H., & Kamerlin, S. C. L. (2018b). Computer Simulations of the Catalytic Mechanism of Wild-Type and Mutant β-Phosphoglucomutase. *Org. Biomol. Chem., 16*, 2060.

Bartmann, L., Schumacher, D., von Stillfried, S., Sternkopf, M., Alampour-Rajabi, S., van Zandvoort, M., . . . Wu, Z. (2019). Evaluation of Riboflavin Transporters as Targets for Drug Delivery and Theranostics. *Front Pharmacol, 10*, 79. doi:10.3389/fphar.2019.00079

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 271-281. doi:10.1107/S0907444910048675

Baxter, N. J., Bowler, M. W., Alizadeh, T., Cliff, M. J., Hounslow, A. M., Wu, B., . . . Waltho, J. P. (2010). Atomic Details of Near-Transition State Conformers for Enzyme Phosphoryl Transfer Revealed by MgF3– Rather Than by Phosphoranes. *Proc. Natl. Acad. Sci. U. S. A., 107*, 4555.

Belocopitow, E., & Marechal, L. R. (1974). Metabolism of trehalose in Euglena gracilis. Partial purification and some properties of phosphoglucomutase acting on beta-glucose 1-phosphate. *Eur J Biochem, 46*(3), 631-637. doi:10.1111/j.1432-1033.1974.tb03659.x

Bian, X., Gao, W., Wang, Y., Yao, Z., Xu, Q., Guo, C., & Li, B. (2019). Riboflavin deficiency affects lipid metabolism partly by reducing apolipoprotein B100 synthesis in rats. *J Nutr Biochem, 70*, 75-81. doi:10.1016/j.jnutbio.2019.04.011

Boesl, U. (2017). Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrom Rev, 36*(1), 86-109. doi:10.1002/mas.21520

Bretzel, W., Schurter, W., Ludwig, B., Kupfer, E., Doswald, S., Pfister, M., & van Loon, A. P. G. M. (1999). Commercial riboflavin production by recombinant Bacillus subtilis: down-stream processing and comparison of the composition of riboflavin produced by fermentation or chemical synthesis. *Journal of Industrial Microbiology and Biotechnology, 22*(1), 19-26. doi:10.1038/sj.jim.2900604

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. *J Mol Biol, 361*(5), 1003-1034. doi:10.1016/j.jmb.2006.06.049

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., & Delucas, L. J. (2007). His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr, 63*(Pt 3), 295-301. doi:10.1107/S0907444906052024

Chai, Q., Zhang, Y., & Liu, C. H. (2018). Mycobacterium tuberculosis: An Adaptable Pathogen Associated With Multiple Human Diseases. *Front Cell Infect Microbiol, 8*, 158. doi:10.3389/fcimb.2018.00158

Chant, A., Kraemer-Pecore, C. M., Watkin, R., & Kneale, G. G. (2005). Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene

regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expr Purif, 39*(2), 152-159. doi:10.1016/j.pep.2004.10.017

Chayen, N. E., & Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods, 5*(2), 147-153. doi:10.1038/nmeth.f.203

Collet, J. F., Gerin, I., Rider, M. H., Veiga-da-Cunha, M., & Van Schaftingen, E. (1997). Human L-3-phosphoserine phosphatase: sequence, expression and evidence for a phosphoenzyme intermediate. *FEBS Lett, 408*(3), 281-284. doi:10.1016/s0014-5793(97)00438-9

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res, 16*(22), 10881-10890. doi:10.1093/nar/16.22.10881

Crowe, J., Dobeli, H., Gentz, R., Hochuli, E., Stuber, D., & Henco, K. (1994). 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol Biol, 31*, 371-387. doi:10.1385/0-89603-258-2:371

Dahl, S. G., Sylte, I., & Ravna, A. W. (2004). Structures and models of transporter proteins. *J Pharmacol Exp Ther, 309*(3), 853-860. doi:10.1124/jpet.103.059972

Dai, J., Finci, L., Zhang, C., Lahiri, S., Zhang, G., Peisach, E., . . . Dunaway-Mariano, D. (2009). Analysis of the Structural Determinants Underlying Discrimination Between Substrate and Solvent in β-Phosphoglucomutase Catalysis. *Biochemistry, 48*, 1984.

Dai, J., Wang, L., Allen, K. N., Radstrom, P., & Dunaway-Mariano, D. (2006). Conformational Cycling in β-Phosphoglucomutase Catalysis: Reorientation of the β-D-Glucose 1,6-(Bis)phosphate Intermediate. *Biochemistry, 45*, 7818.

Dai, J. B., Liu, Y., Ray, W. J., Jr., & Konno, M. (1992). The crystal structure of muscle phosphoglucomutase refined at 2.7-angstrom resolution. *J Biol Chem, 267*(9), 6322-6337. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1532581

Dauter, Z. (1999). Data-collection strategies. *Acta Crystallogr D Biol Crystallogr, 55*(Pt 10), 1703-1717. doi:10.1107/s0907444999008367

Davey, M. E., Caiazza, N. C., & O'Toole, G. A. (2003). Rhamnolipid surfactant production affects biofilm architecture in Pseudomonas aeruginosa PAO1. *J Bacteriol, 185*(3), 1027-1036. doi:10.1128/JB.185.3.1027-1036.2003

De Smet, K. A. L., Weston, A., Brown, I. N., Young, D. B., & Robertson, B. D. (2000). Three pathways for trehalose biosynthesis in mycobacteria. *Microbiology (Reading), 146 (Pt 1)*, 199-208. doi:10.1099/00221287-146-1-199

DeJesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., . . . Ioerger, T. R. (2017). Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *mBio, 8*(1). doi:10.1128/mBio.02133-16

Denome, S. A., Elf, P. K., Henderson, T. A., Nelson, D. E., & Young, K. D. (1999). Escherichia coli mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. *J Bacteriol, 181*(13), 3981-3993. doi:10.1128/JB.181.13.3981-3993.1999

Dessau, M. A., & Modis, Y. (2011). Protein crystallization for X-ray crystallography. *Journal of visualized experiments : JoVE*(47), 2285. doi:10.3791/2285

Dubendorff, J. W., & Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol, 219*(1), 45-59. doi:10.1016/0022-2836(91)90856-2

Dutta, N. K., & Karakousis, P. C. (2014). Latent tuberculosis infection: myths, models, and molecular mechanisms. *Microbiol Mol Biol Rev, 78*(3), 343-371. doi:10.1128/MMBR.00010-14

Elbein, A. D., & Mitchell, M. (1973). Levels of glycogen and trehalose in Mycobacterium smegmatis and the purification and properties of the glycogen synthetase. *J Bacteriol, 113*(2), 863-873. doi:10.1128/jb.113.2.863-873.1973

Elbein, A. D., Pan, Y. T., Pastuszak, I., & Carroll, D. (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology, 13*(4), 17R-27R. doi:10.1093/glycob/cwg047

Elder, N. C. (1992). Extrapulmonary tuberculosis. A review. *Arch Fam Med, 1*(1), 91-98. doi:10.1001/archfami.1.1.91

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2126-2132. doi:10.1107/S0907444904019158

Falzon, D., Mirzayev, F., Wares, F., Baena, I. G., Zignol, M., Linh, N., . . . Raviglione, M. (2015). Multidrug-resistant tuberculosis around the world: what progress has been made? *Eur Respir J, 45*(1), 150-160. doi:10.1183/09031936.00101814

Fassbinder, F., Kist, M., & Bereswill, S. (2000). Structural and functional analysis of the riboflavin synthesis genes encoding GTP cyclohydrolase II (ribA), DHBP synthase (ribBA), riboflavin synthase (ribC), and riboflavin deaminase/reductase (ribD) from Helicobacter pylori strain P1. *FEMS Microbiol Lett, 191*(2), 191-197. doi:10.1111/j.1574-6968.2000.tb09339.x

Fennelly, K. P., Martyny, J. W., Fulton, K. E., Orme, I. M., Cave, D. M., & Heifets, L. B. (2004). Cough-generated aerosols of Mycobacterium tuberculosis: a new method to study infectiousness. *Am J Respir Crit Care Med, 169*(5), 604-609. doi:10.1164/rccm.200308-1101OC

Finney, J. L., Gellatly, B. J., Golton, I. C., & Goodfellow, J. (1980). Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. *Biophys J, 32*(1), 17-33. doi:10.1016/S0006-3495(80)84913-7

Fischer, M., & Bacher, A. (2006). Biosynthesis of vitamin B2 in plants. *Physiologia Plantarum, 126*(3), 304-318. doi:https://doi.org/10.1111/j.1399-3054.2006.00607.x

Fischer, M., & Bacher, A. (2008). Biosynthesis of vitamin B2: Structure and mechanism of riboflavin synthase. *Arch Biochem Biophys, 474*(2), 252-265. doi:10.1016/j.abb.2008.02.008

Fischer, M., Schott, A.-K., Römisch, W., Ramsperger, A., Augustin, M., Fidler, A., . . . Eisenreich, W. (2004). Evolution of Vitamin B2 Biosynthesis. A Novel Class of Riboflavin Synthase in Archaea††This paper is dedicated to the memory of Dr Justus Henatsch, Europa Fachhochschule Fresenius Idstein. *Journal of Molecular Biology, 343*(1), 267-278. doi:https://doi.org/10.1016/j.jmb.2004.08.016

Fisher, R. A., Gollan, B., & Helaine, S. (2017). Persistent bacterial infections and persister cells. *Nat Rev Microbiol, 15*(8), 453-464. doi:10.1038/nrmicro.2017.42

Gavalda, S., Bardou, F., Laval, F., Bon, C., Malaga, W., Chalut, C., . . . Quemard, A. (2014). The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chem Biol, 21*(12), 1660-1669. doi:10.1016/j.chembiol.2014.10.011

Gerdes, S. Y., Scholle, M. D., D'Souza, M., Bernal, A., Baev, M. V., Farrell, M., . . . Osterman, A. L. (2002). From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J Bacteriol, 184*(16), 4555-4572. doi:10.1128/JB.184.16.4555-4572.2002

Global tuberculosis report 2020. Geneva: World Health Organization; 2020. Licence: CC BY-N

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc, 1*(6), 2876-2890. doi:10.1038/nprot.2006.202

Guy, H. R. (1985). Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys J, 47*(1), 61-70. doi:10.1016/S0006-3495(85)83877-7

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, 14*(17), 2272-2275. doi:10.1002/cbic.201300544

Hagel, L. (1998). Gel-Filtration Chromatography. *Current Protocols in Molecular Biology, 44*(1), 10.19.11-10.19.12. doi:https://doi.org/10.1002/0471142727.mb1009s44

Harrison, R. W. (1993). Phase problem in crystallography. *Journal of the Optical Society of America Part A, Optics and Image Science, 10*(5), 1046-1055. Retrieved from http://inis.iaea.org/search/search.aspx?orig_q=RN:24073150

Herz, S., Eberhardt, S., & Bacher, A. (2000). Biosynthesis of riboflavin in plants. The ribA gene of Arabidopsis thaliana specifies a bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase. *Phytochemistry, 53*(7), 723-731. doi:10.1016/s0031-9422(00)00013-3

Hisano, T., Hata, Y., Fujii, T., Liu, J. Q., Kurihara, T., Esaki, N., & Soda, K. (1996). Crystal structure of L-2-haloacid dehalogenase from Pseudomonas sp. YL. An alpha/beta hydrolase structure that is different from the alpha/beta hydrolase fold. *J Biol Chem, 271*(34), 20322-20330. doi:10.1074/jbc.271.34.20322

Huang, H., Patskovsky, Y., Toro, R., Farelli, J. D., Pandya, C., Almo, S. C., . . . Dunaway-Mariano, D. (2011). Divergence of structure and function in the haloacid dehalogenase enzyme superfamily: Bacteroides thetaiotaomicron BT2127 is an inorganic pyrophosphatase. *Biochemistry, 50*(41), 8937-8949. doi:10.1021/bi201181q

Islam, Z., Kumar, A., Singh, S., Salmon, L., & Karthikeyan, S. (2015). Structural basis for competitive inhibition of 3,4-dihydroxy-2-butanone-4-phosphate synthase from Vibrio cholerae. *J Biol Chem, 290*(18), 11293-11308. doi:10.1074/jbc.M114.611830

Jankute, M., Cox, J. A., Harrison, J., & Besra, G. S. (2015). Assembly of the Mycobacterial Cell Wall. *Annu Rev Microbiol, 69*, 405-423. doi:10.1146/annurev-micro-091014-104121

Johnson, L. A., Robertson, A. J., Baxter, N. J., Trevitt, C. R., Bisson, C., Jin, Y., . . . Waltho, J. P. (2018). van der Waals Contact between Nucleophile and Transferring Phosphorus

Is Insufficient To Achieve Enzyme Transition-State Architecture. *ACS Catalysis, 8*(9), 8140-8153. doi:10.1021/acscatal.8b01612

Joshi, S. M., Pandey, A. K., Capite, N., Fortune, S. M., Rubin, E. J., & Sassetti, C. M. (2006). Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc Natl Acad Sci U S A, 103*(31), 11760-11765. doi:10.1073/pnas.0603179103

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta crystallographica. Section D, Biological crystallography, 66*(Pt 2), 133-144. doi:10.1107/S0907444909047374

Kapopoulou, A., Lew, J. M., & Cole, S. T. (2011). The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb), 91*(1), 8-13. doi:10.1016/j.tube.2010.09.006

Karthikeyan, S., Zhou, Q., Mseeh, F., Grishin, N. V., Osterman, A. L., & Zhang, H. (2003). Crystal structure of human riboflavin kinase reveals a beta barrel fold and a novel active site arch. *Structure, 11*(3), 265-273. doi:10.1016/s0969-2126(03)00024-8

Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. *Biochim Biophys Acta, 1751*(2), 119-139. doi:10.1016/j.bbapap.2005.06.005

Kelly, S. M., & Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci, 1*(4), 349-384. doi:10.2174/1389203003381315

Keren, I., Minami, S., Rubin, E., & Lewis, K. (2011). Characterization and transcriptome analysis of Mycobacterium tuberculosis persisters. *mBio, 2*(3), e00100-00111. doi:10.1128/mBio.00100-11

Kis, K., & Bacher, A. (1995). Substrate channeling in the lumazine synthase/riboflavin synthase complex of Bacillus subtilis. *J Biol Chem, 270*(28), 16788-16795. doi:10.1074/jbc.270.28.16788

Knowles, J., & Gromo, G. (2003). A guide to drug discovery: Target selection in drug discovery. *Nat Rev Drug Discov, 2*(1), 63-69. doi:10.1038/nrd986

Koonin, E. V., & Tatusov, R. L. (1994). Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J Mol Biol, 244*(1), 125-132. doi:10.1006/jmbi.1994.1711

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2256-2268. doi:10.1107/S0907444904026460

Kurihara, T., Liu, J. Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., & Soda, K. (1995). Comprehensive site-directed mutagenesis of L-2-halo acid dehalogenase to probe catalytic amino acid residues. *J Biochem, 117*(6), 1317-1322. doi:10.1093/oxfordjournals.jbchem.a124861

Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., . . . Yakunin, A. F. (2006). Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *J Biol Chem, 281*(47), 36149-36161. doi:10.1074/jbc.M605449200

Kvam, C., Olsvik, E. S., McKinley-McKee, J., & Saether, O. (1997). Studies on recombinant Acetobacter xylinum alpha-phosphoglucomutase. *Biochem J, 326 (Pt 1)*, 197-203. doi:10.1042/bj3260197

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. (2002a). Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis. *Biochemistry, 41*(26), 8351-8359. doi:10.1021/bi0202373

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. (2002b). Caught in the Act: The Structure of Phosphorylated β-Phosphoglucomutase From Lactococcus lactis. *Biochemistry, 41*, 8351.

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. (2003a). The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science, 299*(5615), 2067-2071. doi:10.1126/science.1082710

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. (2003b). The Pentacovalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction. *Science, 299*, 2067.

Lascelles, J., & Woods, D. D. (1952). The synthesis of folic acid by Bacterium coll and Staphylococcus aureus and its inhibition by sulphonamides. *Br J Exp Pathol, 33*(3), 288-303. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14935094

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography, 26*(2), 283-291. doi:https://doi.org/10.1107/S0021889892009944

Lee, J. J., Lee, S. K., Song, N., Nathan, T. O., Swarts, B. M., Eum, S. Y., . . . Eoh, H. (2019). Transient drug-tolerance and permanent drug-resistance rely on the trehalose-catalytic shift in Mycobacterium tuberculosis. *Nat Commun, 10*(1), 2928. doi:10.1038/s41467-019-10975-7

Lee, J. Y. (2015). Diagnosis and treatment of extrapulmonary tuberculosis. *Tuberc Respir Dis (Seoul), 78*(2), 47-55. doi:10.4046/trd.2015.78.2.47

Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta crystallographica. Section D, Biological crystallography, 62 Pt 1*, 48-57.

Levander, F., Andersson, U., & Radstrom, P. (2001). Physiological role of beta-phosphoglucomutase in Lactococcus lactis. *Appl Environ Microbiol, 67*(10), 4546-4553. doi:10.1128/AEM.67.10.4546-4553.2001

Li, W., Upadhyay, A., Fontes, F. L., North, E. J., Wang, Y., Crans, D. C., . . . Jackson, M. (2014). Novel insights into the mechanism of inhibition of MmpL3, a target of multiple pharmacophores in Mycobacterium tuberculosis. *Antimicrob Agents Chemother, 58*(11), 6413-6423. doi:10.1128/AAC.03229-14

Lindsay, M. A. (2003). Target discovery. *Nat Rev Drug Discov, 2*(10), 831-838. doi:10.1038/nrd1202

Liu, D., & Zempleni, J. (2014). Low activity of LSD1 elicits a pro-inflammatory gene expression profile in riboflavin-deficient human T Lymphoma Jurkat cells. *Genes Nutr, 9*(5), 422. doi:10.1007/s12263-014-0422-6

Liu, Y., Ray, W. J., Jr., & Baranidharan, S. (1997). Structure of rabbit muscle phosphoglucomutase refined at 2.4 A resolution. *Acta Crystallogr D Biol Crystallogr, 53*(Pt 4), 392-405. doi:10.1107/S0907444997000875

Long, Q., Ji, L., Wang, H., & Xie, J. (2010). Riboflavin biosynthetic and regulatory factors as potential novel anti-infective drug targets. *Chem Biol Drug Des, 75*(4), 339-347. doi:10.1111/j.1747-0285.2010.00946.x

Looijesteijn, P. J., Trapet, L., de Vries, E., Abee, T., & Hugenholtz, J. (2001). Physiological function of exopolysaccharides produced by Lactococcus lactis. *Int J Food Microbiol, 64*(1-2), 71-80. doi:10.1016/s0168-1605(00)00437-2

MacKintosh, C. (1993). In D.G. Hardie (Ed.). Protein Phosphorylation: A Practical Approach. 221. New York: IRL Press).C-SA 3.0 IGO.

Mazur-Bialy, A. I., Buchala, B., & Plytycz, B. (2013). Riboflavin deprivation inhibits macrophage viability and activity - a study on the RAW 264.7 cell line. *Br J Nutr, 110*(3), 509-514. doi:10.1017/S0007114512005351

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics, 16*(4), 404-405. doi:10.1093/bioinformatics/16.4.404

McPherson, A. (1997). Recent advances in the microgravity crystallization of biological macromolecules. *Trends Biotechnol, 15*(6), 197-200. doi:10.1016/S0167-7799(97)01041-X

Mcpherson, A. (2001). A comparison of salts for the crystallization of macromolecules. *Protein Science, 10*(2), 418-422. doi:https://doi.org/10.1110/ps.32001

McPherson, A. (2004). Protein crystallization in the structural genomics era. *Journal of Structural and Functional Genomics, 5*(1), 3-12. doi:10.1023/B:JSFG.0000029199.43875.92

McPherson, A. (2017). Protein Crystallization. *Methods Mol Biol, 1607*, 17-50. doi:10.1007/978-1-4939-7000-1_2

McPherson, A., & Cudney, B. (2006). Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol, 156*(3), 387-406. doi:10.1016/j.jsb.2006.09.006

McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta crystallographica. Section F, Structural biology communications, 70 Pt 11*, 1445-1467.

Mendes, V., Maranha, A., Alarico, S., da Costa, M. S., & Empadinhas, N. (2011). Mycobacterium tuberculosis Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis. *Sci Rep, 1*, 177. doi:10.1038/srep00177

Mesak, L. R., & Dahl, M. K. (2000). Purification and enzymatic characterization of PgcM: a beta-phosphoglucomutase and glucose-1-phosphate phosphodismutase of Bacillus subtilis. *Arch Microbiol, 174*(4), 256-264. doi:10.1007/s002030000200

Micsonai, A., Wien, F., Bulyaki, E., Kun, J., Moussong, E., Lee, Y. H., Kardos, J. (2018). BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res, 46*(W1), W315-W322. doi:10.1093/nar/gky497

Minato, Y., Gohl, D. M., Thiede, J. M., Chacon, J. M., Harcombe, W. R., Maruyama, F., & Baughn, A. D. (2019). Genomewide Assessment of Mycobacterium tuberculosis

Conditionally Essential Metabolic Pathways. *mSystems, 4*(4). doi:10.1128/mSystems.00070-19

Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. (2000). The crystal (MacKintosh, C. (1993). In D.G. Hardie (Ed.). Protein Phosphorylation: A Practical Approach. 221. New York: IRL Press).structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily. *Biochemistry, 39*(34), 10385-10396. doi:10.1021/bi001171j

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., . . . Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 355-367. doi:10.1107/S0907444911001314

Nakano, C., Ootsuka, T., Takayama, K., Mitsui, T., Sato, T., & Hoshino, T. (2011). Characterization of the Rv3378c gene product, a new diterpene synthase for producing tuberculosinol and (13R, S)-isotuberculosinol (nosyberkol), from the Mycobacterium tuberculosis H37Rv genome. *Biosci Biotechnol Biochem, 75*(1), 75-81. doi:10.1271/bbb.100570

Nilsson, U., & Radstrom, P. (2001). Genetic localization and regulation of the maltose phosphorylase gene, malP, in Lactococcus lactis. *Microbiology (Reading), 147*(Pt 6), 1565-1573. doi:10.1099/00221287-147-6-1565

Olsen, D. B., Hepburn, T. W., Moos, M., Mariano, P. S., & Dunaway-Mariano, D. (1988). Investigation of the Bacillus cereus phosphonoacetaldehyde hydrolase. Evidence for a Schiff base mechanism and sequence analysis of an active-site peptide containing the catalytic lysine residue. *Biochemistry, 27*(6), 2229-2234. doi:10.1021/bi00406a063

Olvera, C., Goldberg, J. B., Sanchez, R., & Soberon-Chavez, G. (1999). The Pseudomonas aeruginosa algC gene product participates in rhamnolipid biosynthesis. *FEMS Microbiol Lett, 179*(1), 85-90. doi:10.1111/j.1574-6968.1999.tb08712.x

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol, 276*, 307-326. doi:10.1016/S0076-6879(97)76066-X

Pan, Y. T., Koroth Edavana, V., Jourdian, W. J., Edmondson, R., Carroll, J. D., Pastuszak, I., & Elbein, A. D. (2004). Trehalose synthase of Mycobacterium smegmatis: purification, cloning, expression, and properties of the enzyme. *Eur J Biochem, 271*(21), 4259-4269. doi:10.1111/j.1432-1033.2004.04365.x

Pandya, C., Farelli, J. D., Dunaway-Mariano, D., & Allen, K. N. (2014). Enzyme promiscuity: engine of evolutionary innovation. *J Biol Chem, 289*(44), 30229-30236. doi: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 Crystallogr D Biol Crystallogr, 71*(Pt 9), 1824-1834. doi: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. doi:10.1021/bi0490688

Plaut, G. W., Smith, C. M., & Alworth, W. L. (1974). Biosynthesis of water-soluble vitamins. *Annu Rev Biochem, 43*(0), 899-922. doi:10.1146/annurev.bi.43.070174.004343

Qian, N., Stanley, G. A., Hahn-Hagerdal, B., & Radstrom, P. (1994). Purification and characterization of two phosphoglucomutases from Lactococcus lactis subsp. lactis and their regulation in maltose- and glucose-utilizing cells. *J Bacteriol, 176*(17), 5304-5311. doi:10.1128/jb.176.17.5304-5311.1994

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. *J Biol Chem, 286*(49), 42091-42098. doi:10.1074/jbc.M111.260885

Ray, W. J., Jr., Burgner, J. W., 2nd, & Post, C. B. (1990). Characterization of vanadate-based transition-state-analogue complexes of phosphoglucomutase by spectral and NMR techniques. *Biochemistry, 29*(11), 2770-2778. doi:10.1021/bi00463a021

Ray, W. J., Jr., & Long, J. W. (1976). Thermodynamics and mechanism of the PO3 transfer process in the phosphoglucomutase reaction. *Biochemistry, 15*(18), 3993-4006. doi:10.1021/bi00663a014

Ray, W. J., Jr., Post, C. B., Liu, Y., & Rhyu, G. I. (1993). Structural changes at the metal ion binding site during the phosphoglucomutase reaction. *Biochemistry, 32*(1), 48-57. doi:10.1021/bi00052a008

Rinaldo-Matthis, A., Rampazzo, C., Reichard, P., Bianchi, V., & Nordlund, P. (2002). Crystal structure of a human mitochondrial deoxyribonucleotidase. *Nat Struct Biol, 9*(10), 779-787. doi:10.1038/nsb846

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res, 42*(Web Server issue), W320-324. doi:10.1093/nar/gku316

Rose, P. W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., Burley, S. K. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res, 45*(D1), D271-D281. doi:10.1093/nar/gkw1000

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. *Plant J, 88*(5), 705-716. doi:10.1111/tpj.13291

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, 16*(17), 2466-2469. doi:10.1002/cbic.201500352

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol, 48*(1), 77-84. doi:10.1046/j.1365-2958.2003.03425.x

Sassetti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences, 100*(22), 12989-12994. doi:doi:10.1073/pnas.2134250100

Schiller, J., Suss, R., Fuchs, B., Muller, M., Zschornig, O., & Arnold, K. (2007). MALDI-TOF MS in lipidomics. *Front Biosci, 12*, 2568-2579. doi:10.2741/2255

Schramm, M., Wiegmann, K., Schramm, S., Gluschko, A., Herb, M., Utermohlen, O., & Kronke, M. (2014). Riboflavin (vitamin B2) deficiency impairs NADPH oxidase 2 (Nox2) priming and defense against Listeria monocytogenes. *Eur J Immunol, 44*(3), 728-741. doi:10.1002/eji.201343940

Seal, S. N., & Rose, Z. B. (1987). Characterization of a phosphoenzyme intermediate in the reaction of phosphoglycolate phosphatase. *J Biol Chem, 262*(28), 13496-13500. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2820973

Shackelford, G. S., Regni, C. A., & Beamer, L. J. (2004). Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily. *Protein Sci, 13*(8), 2130-2138. doi:10.1110/ps.04801104

Shi, L., Zhang, H., Qiu, Y., Wang, Q., Wu, X., Wang, H., . . . Lin, D. (2013). Biochemical characterization and ligand-binding properties of trehalose-6-phosphate phosphatase from Mycobacterium tuberculosis. *Acta Biochim Biophys Sin (Shanghai), 45*(10), 837-844. doi:10.1093/abbs/gmt084

Silvaggi, N. R., Zhang, C., Lu, Z., Dai, J., Dunaway-Mariano, D., & Allen, K. N. (2006). The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. *J Biol Chem, 281*(21), 14918-14926. doi:10.1074/jbc.M601505200

Singh, M., Kumar, P., Yadav, S., Gautam, R., Sharma, N., & Karthikeyan, S. (2013). The crystal structure reveals the molecular mechanism of bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II (Rv1415) from Mycobacterium tuberculosis. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 9), 1633-1644. doi:10.1107/S0907444913011402

Singhal, N., Kumar, M., Kanaujia, P. K., & Virdi, J. S. (2015). MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Front Microbiol, 6*, 791. doi:10.3389/fmicb.2015.00791

Suarez, I., Funger, S. M., Kroger, S., Rademacher, J., Fatkenheuer, G., & Rybniker, J. (2019). The Diagnosis and Treatment of Tuberculosis. *Dtsch Arztebl Int, 116*(43), 729-735. doi:10.3238/arztebl.2019.0729

Thung, S. N., Bach, N., Fasy, T. M., Jordon, D., & Schaffner, F. (1990). Hepatocellular carcinoma associated with autoimmune chronic active hepatitis. *Mt Sinai J Med, 57*(3), 165-168. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2164643

Tipper, D. J., & Strominger, J. L. (1965). Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci U S A, 54*(4), 1133-1141. doi:10.1073/pnas.54.4.1133

Tremblay, L. W., Dunaway-Mariano, D., & Allen, K. N. (2006). Structure and activity analyses of Escherichia coli K-12 NagD provide insight into the evolution of biochemical function in the haloalkanoic acid dehalogenase superfamily. *Biochemistry, 45*(4), 1183-1193. doi:10.1021/bi051842j

Tremblay, L. W., Zhang, G., Dai, J., Dunaway-Mariano, D., & Allen, K. N. (2005). Chemical confirmation of a pentavalent phosphorane in complex with beta-phosphoglucomutase. *J Am Chem Soc, 127*(15), 5298-5299. doi:10.1021/ja0509073

Tu, B. P., Ho-Schleyer, S. C., Travers, K. J., & Weissman, J. S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science, 290*(5496), 1571-1574. doi:10.1126/science.290.5496.1571

Udwadia, Z. F. (2012). MDR, XDR, TDR tuberculosis: ominous progression. *Thorax, 67*(4), 286-288. doi:10.1136/thoraxjnl-2012-201663

Ufimtseva, E., Eremeeva, N., Vakhrusheva, D., & Skornyakov, S. (2019). Mycobacterium tuberculosis shape and size variations in alveolar macrophages of tuberculosis patients. *European Respiratory Journal, 54*(suppl 63), PA4605. doi:10.1183/13993003.congress-2019.PA4605

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., & Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2184-2195. doi:10.1107/S0907444904023510

Velankar, S., Best, C., Beuth, B., Boutselakis, C. H., Cobley, N., Sousa Da Silva, A. W., . . . Kleywegt, G. J. (2010). PDBe: Protein Data Bank in Europe. *Nucleic Acids Res, 38*(Database issue), D308-317. doi:10.1093/nar/gkp916

Vinnard, C., & Blumberg, E. A. (2017). Endocrine and Metabolic Aspects of Tuberculosis. *Microbiol Spectr, 5*(1). doi:10.1128/microbiolspec.TNMI7-0035-2016

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., & Gelfand, M. S. (2002). Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res, 30*(14), 3141-3151. doi:10.1093/nar/gkf433

Wang, W., Kim, R., Jancarik, J., Yokota, H., & Kim, S. H. (2001). Crystal structure of phosphoserine phosphatase from Methanococcus jannaschii, a hyperthermophile, at 1.8 A resolution. *Structure, 9*(1), 65-71. doi:10.1016/s0969-2126(00)00558-x

Wilfinger, W. W., Mackey, K., & Chomczynski, P. (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques, 22*(3), 474-476, 478-481. doi:10.2144/97223st01

Wood, H. P., Cruz-Navarrete, F. A., Baxter, N. J., Trevitt, C. R., Robertson, A. J., Dix, S. R., . . . Waltho, J. P. (2020). Allomorphy as a mechanism of post-translational control of enzyme activity. *Nat Commun, 11*(1), 5538. doi:10.1038/s41467-020-19215-9

Wood, T. K., Knabel, S. J., & Kwan, B. W. (2013). Bacterial persister cell formation and dormancy. *Appl Environ Microbiol, 79*(23), 7116-7121. doi:10.1128/AEM.02636-13

Woodruff, P. J., Carlson, B. L., Siridechadilok, B., Pratt, M. R., Senaratne, R. H., Mougous, J. D., . . . Bertozzi, C. R. (2004). Trehalose is required for growth of Mycobacterium smegmatis. *J Biol Chem, 279*(28), 28835-28843. doi:10.1074/jbc.M313103200

Yadav, S., & Karthikeyan, S. (2015). Structural and biochemical characterization of GTP cyclohydrolase II from Helicobacter pylori reveals its redox dependent catalytic activity. *J Struct Biol, 192*(1), 100-115. doi:10.1016/j.jsb.2015.08.004

Ye, R. W., Zielinski, N. A., & Chakrabarty, A. M. (1994). Purification and characterization of phosphomannomutase/phosphoglucomutase from Pseudomonas aeruginosa involved in biosynthesis of both alginate and lipopolysaccharide. *J Bacteriol, 176*(16), 4851-4857. doi:10.1128/jb.176.16.4851-4857.1994

Zhang, B., Tang, J., Wu, Y. B., Cao, J. T., Xing, G. N., Sun, P. X., . . . Hou, S. S. (2021). Effects of riboflavin deficiency on the lipid metabolism of duck breeders and duck embryos. *Poult Sci, 100*(10), 101342. doi:10.1016/j.psj.2021.101342

Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W., & Allen, K. N. (2005a). Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis. *Biochemistry, 44*(27), 9404-9416. doi:10.1021/bi050558p

Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W., & Allen, K. N. (2005b). Catalytic Cycling in β-Phosphoglucomutase: A Kinetic and Structural Analysis. *Biochemistry, 44*, 9404.

Zhang, J., Lu, X., Lei, Y., Hou, X., & Wu, P. (2017). Exploring Tunable Excitation of QDs to Maximize the Overlap with Absorber for Inner Filter Effect-based Phosphorescent Sensing of Alkaline Phosphatase. *Nanoscale, 9*. doi:10.1039/C7NR03673F

Zheng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res, 32*(14), e115. doi:10.1093/nar/gnh110

Zumla, A., Nahid, P., & Cole, S. T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov, 12*(5), 388-404. doi:10.1038/nrd4001

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 December 2020
doi: 10.3389/fmicb.2020.554927

Isolation and Taxonomic Characterization of Novel Haloarchaeal Isolates From Indian Solar Saltern: A Brief Review on Distribution of Bacteriorhodopsins and V-Type ATPases in Haloarchaea

Dipesh Kumar Verma[1], Chetna Chaudhary[1], Latika Singh[1], Chandni Sidhu[2], Busi Siddhardha[3], Senthil E. Prasad[4] and Krishan Gopal Thakur[1]**

[1] Structural Biology Laboratory, G. N. Ramachandran Protein Centre, Council of Scientific and Industrial Research-Institute of Microbial Technology (CSIR-IMTECH), Chandigarh, India, [2] MTCC-Microbial Type Culture Collection & Gene Bank, Council of Scientific and Industrial Research Institute of Microbial Technology (CSIR-IMTECH), Chandigarh, India, [3] Department of Microbiology, School of Life Sciences, Pondicherry University, Puducherry, India, [4] Biochemical Engineering Research and Process Development Centre, Council of Scientific and Industrial Research-Institute of Microbial Technology (CSIR-IMTECH), Chandigarh, India

OPEN ACCESS

Edited by:
Anna-Louise Reysenbach, Portland State University, United States

Reviewed by:
Mohea Couturier, University of Illinois at Urbana–Champaign, United States
Likui Zhang, Yangzhou University, China

**Correspondence:*
Krishan Gopal Thakur
krishang@imtech.res.in
Senthil E. Prasad
esprasad@imtech.res.in

Specialty section:
This article was submitted to Biology of Archaea, a section of the journal Frontiers in Microbiology

Received: 23 April 2020
Accepted: 17 September 2020
Published: 09 December 2020

Citation:
Verma DK, Chaudhary C, Singh L, Sidhu C, Siddhardha B, Prasad SE and Thakur KG (2020) Isolation and Taxonomic Characterization of Novel Haloarchaeal Isolates From Indian Solar Saltern: A Brief Review on Distribution of Bacteriorhodopsins and V-Type ATPases in Haloarchaea. Front. Microbiol. 11:554927. doi: 10.3389/fmicb.2020.554927

Haloarchaea inhabit high salinity environments worldwide. They are a potentially rich source of crucial biomolecules like carotenoids and industrially useful proteins. However, diversity in haloarchaea present in Indian high salinity environments is poorly studied. In the present study, we isolated 12 haloarchaeal strains from hypersaline Kottakuppam, Tamil Nadu solar saltern in India. 16S rRNA based taxonomic characterization of these isolates suggested that nine of them are novel strains that belong to genera *Haloarcula*, *Halomicrobium,* and *Haloferax*. Transmission electron microscopy suggests the polymorphic nature of these haloarchaeal isolates. Most of the haloarchaeal species are known to be high producers of carotenoids. We were able to isolate carotenoids from all these 12 isolates. The UV-Vis spectroscopy-based analysis suggests that bacterioruberin and lycopene are the major carotenoids produced by these isolates. Based on the visual inspection of the purified carotenoids, the isolates were classified into two broad categories i.e., yellow and orange, attributed to the differences in the ratio of bacterioruberin and lycopene as confirmed by the UV-Vis spectral analysis. Using a PCR-based screening assay, we were able to detect the presence of the bacteriorhodopsin gene (*bop*) in 11 isolates. We performed whole-genome sequencing for three *bop* positive and one *bop* negative haloarchaeal isolates. Whole-genome sequencing, followed by pan-genome analysis identified multiple unique genes involved in various biological functions. We also successfully cloned, expressed, and purified functional recombinant bacteriorhodopsin (BR) from one of the isolates using *Escherichia coli* as an expression host. BR has light-driven proton pumping activity resulting in the proton gradient across the membrane, which is utilized by V-Type ATPases to produce ATP. We analyzed the distribution of *bop* and other accessory genes involved in functional BR expression and ATP synthesis in all the representative haloarchaeal

This page is available in the following languages:

Creative Commons License Deed

Attribution 4.0 International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.