



UNIVERSITY OF  
**LEICESTER**

**Studying the mechanism and action  
of Purine biosynthesis enzymes in  
*Mycobacterium Tuberculosis* H37RV**

Thesis submitted for the degree of

**Doctor of Philosophy**

at the University of Leicester by

**Samah Jameel Almehmadi MSc (New Haven)**

Department of Molecular and Cell Biology

University of Leicester

August 2018

# Abstract

## Studying the mechanism and action of Purine biosynthesis enzymes in *Mycobacterium Tuberculosis* H37RV

*Samah Almehmadi*

MtbSAICAR and MtbpurD are two essential enzymes in the *Mycobacterium Tuberculosis* purine biosynthesis pathway. The aim of this work is to improve our understanding of these enzymes by characterising the kinetics and determining their structures. MtbpurD has been expressed and purified but it failed to crystallise, and so homology modelling was used to compare it to homologues enzymes. MtbSAICAR was successfully crystallised and the X-ray crystal structures determined. The kinetic parameters in solution were also measured. The MtbSAICAR structures were solved in apo form and in complex with substrate and inhibitor ligands, the binding of ligands does not induce major conformation changes. Although all the solved structures are isomorphous, there are some differences in mobility seen upon ligand binding. Comparing MtbSAICAR with homologues enzymes revealed that despite variation in oligomeric organisation, they share similar folds and active sites. The similarities and differences have been evaluated and compared using both global and local alignments. Two models have been proposed for the mechanism and action of SAICARs. The structures of MtbSAICAR presented here are consistent with the mechanism in which the first step is the phosphorylation of the substrate (CAIR) by ATP, rather than the formation of a phosphoenzyme intermediate. The comparisons between MtbSAICAR and the human equivalent enzyme (HsSAICAR) show differences which might have the potential to be exploited as a target for specific inhibitors as new drugs to combat tuberculosis.

# Acknowledgment

I would like to express my special appreciation and thanks to my supervisor Prof. Peter Moody for his support during my PhD studies. I would also like to thank my thesis committee members Prof. Mark Carr and Prof. Russell Wallis for their valuable suggestions and support both for experimental work and thesis writing up.

Words cannot express how grateful I am to my PhD colleague Dr. Adnan Ayna for his unconditional support in experimental studies and advice for a better quality of thesis data. For all of this help thank you so very much Dr. Adnan. I have also much appreciation to Aziz Al-Ethari for his kind support in any difficulties I encountered during my lab works. I would also like to thank Dr. Louise Fairall and Dr. Mikael Feracci for their help in data collection and processing.

I would like to acknowledge Dr. Richard Cowan for his help on using Circular Dichroism (CD) and I Cycler iQ Real-Time PCR. I wish my deep thanks to Dr. Peter Coombs for his useful advice enzyme assay experiment. I would like to express my deep gratitude to Dr Sofia M. Kapetanaki for providing missing urgent materials while is in the way. I will be ever grateful to Dr. Xiaowen Yang for his valuable discussion in Cloning. I am also glad to thank Ghalia Shelmani for her kindness in letting me use a spectrophotometric enzyme assay (plate reader).

My acknowledgments would not be complete without giving my deep thanks to my family especially my father Jameel Almeahmadi, my mother Fatimah Almeahmadi, my husband Khalid Alsaedi and to my little daughter Nisreen Alsaedi for their support, encouragement and much love. I am so lucky to have such a wonderful family. Thank you very much love you all.

Lastly, I would like to thank the Ministry of Higher Education Saudi Arabia for funding this PhD.

*dedicated to*  
*my parents*  
*&*  
*husband*

## Table of Contents

<b>Abstract</b> .....	<b>II</b>
<b>Acknowledgment</b> .....	<b>III</b>
<b>Abbreviations</b> .....	<b>VIII</b>
<b>Chapter1. Introduction</b> .....	<b>1</b>
1.1. <i>Mycobacterium tuberculosis</i> (Mtb) .....	2
1.1.1. Characteristic features of <i>Mycobacterium tuberculosis</i> (Mtb) .....	2
1.1.2. Cell Wall characteristics of <i>M. tuberculosis</i> .....	3
1.2. Tuberculosis (TB).....	4
1.2.1. Pathogenicity of <i>M. tuberculosis</i> (Mtb).....	5
1.2.2. Diagnosis of <i>M. tuberculosis</i> (Mtb) .....	9
1.2.3. <i>M. tuberculosis</i> (Mtb) Dormancy .....	10
1.2.4. Treatment.....	11
1.3. Purine biosynthesis Pathway .....	12
1.3.1. Introduction .....	12
1.3.2. The steps of the purine biosynthesis Pathway in <i>M. tuberculosis</i> (Mtb).....	14
1.3.2.1. Amidophosphoribosyltransferase (purF).....	15
1.3.2.2. Phosphoribosylamine glycine ligase (purDs) .....	16
1.3.2.3. Phosphoribosylglycinamide formyltransferase (purN).....	16
1.3.2.4. Formyl glycinamide ribonucleotide amidotransferase (purL) .....	17
1.3.2.5. 5-aminoimidazole ribonucleotide (AIR) synthetase (purM).....	18
1.3.2.6. N5-carboxyaminoimidazole ribonucleotide synthase (purK)&N5-carboxyaminoimidazole ribonucleotide mutase(purE) .....	19
1.3.2.7. SAICAR Synthetase (SAICARS) .....	20
1.3.2.8. Adenylosuccinate lyase (purB).....	21
1.3.2.9. AICAR transformylase /IMP cyclohydrolase (ATIC) .....	22
1.3.3. SAICAR Synthetase and Phosphoribosylamine glycine ligase of <i>Mycobacterium tuberculosis</i> .....	23
1.3.3.1. SAICAR Synthetase (SAICARS).....	24
1.3.3.2. Phosphoribosylamine glycine ligase (purDs).....	32
1.3.4. Aims of the study.....	36
<b>Chapter2. Materials and Methods</b> .....	<b>37</b>
2.1 Materials.....	37
2.1.1. Chemical and Reagents.....	37
2.1.2. Chromatographic Media and Membranes .....	37
2.1.3. Crystallisation supplementary .....	37
2.2. Molecular biology techniques .....	37
2.2.1. Primer design .....	37
2.2.2. PCR (ORF Amplification).....	38

2.2.3. Gene cloning and sequencing .....	38
2.2.4. Protein expression and purification.....	39
2.2.4.1. Transformation .....	39
2.2.4.2. Overexpression trials .....	39
2.2.4.3. Purification of N-terminal 6xHis tagged MtbSAICAR and MtbpurD.....	40
2.2.4.4. Gel Filtration Chromatography of MtbSAICAR and MtbpurD .....	41
2.3. Bioinformatics.....	42
2.3.1. Homology model alignment of MtbpurD .....	42
2.4. Biophysical characterisation .....	43
2.4.1. Circular Dichroism .....	43
2.4.1.1. Secondary structure determination and thermal stability of MtbSAICAR-MtbpurD.....	44
2.4.2. Thermal shift assay of proteins .....	45
2.4.3. SAICAR Synthetase kinetic properties .....	46
2.4.4. Crystallisation trials .....	48
2.4.4.1. MtbSAICAR Crystallisation Experiments.....	48
2.4.4.2. MtbSAICAR co-crystallisations.....	49
2.4.5. Introduction to X-ray crystallography .....	49
2.4.5.1. Data Collection .....	51
2.4.5.2. Theory of Bragg Law in diffraction pattern.....	52
2.4.5.3. Data Processing .....	53
2.4.5.4. Phase determination and structural analysis.....	53
2.4.5.5. Refinement and structural validation .....	54
2.4.5.6. Optimal structural alignment process.....	55
<b>Chapter3. Characterisation of SAICARs of <i>M. tuberculosis</i> (MtbSAICAR).</b> .....	57
3.1 Introduction .....	57
3.2. Preparation and Purification of MtbSAICAR.....	57
3.3. Secondary Structure of MtbSAICAR.....	61
3.4. Dye Binding Thermal shift assay of SAICARs <i>M. tuberculosis</i> (MtbSAICAR).....	63
3.5. <i>In vitro</i> enzyme activity of MtbSAICAR .....	65
3.6. Summary .....	68
<b>Chapter4. Structural determination of SAICARs in <i>M. tuberculosis</i> and refinement, results .....</b>	69
4.1 Introduction .....	69
4.2. Crystallisation of MtbSAICAR-ATP complex.....	69
4.3. Data Collection and analysis of MtbSAICAR-ATP complex.....	71
4.4. Structure determination of MtbSAICAR-ATP complex.....	72
4.5. Structure of the MtbSAICAR-ATP complex.....	75
4.6. Crystallisation of apo MtbSAICAR.....	80
4.7. Data Collection and analysis of apo MtbSAICAR .....	81
4.8. Structure determination of apo MtbSAICAR .....	82
4.9. Crystallisation of MtbSAICAR-ADP-AMZ complex .....	85
4.10. Data Collection and Analysis of MtbSAICAR-ADP-AMZ complex .....	85
4.11. Structure determination of MtbSAICAR-ADP-AMZ complex.....	86
4.12. Overall Structure of MtbSAICAR-ADP-AMZ complex .....	88
4.13. Crystallisation of MtbSAICAR-ACP-IMP complex.....	93

4.14. Data Collection and analysis of MtbSAICAR -ACP-IMP complex .....	93
4.15. Structure determination of MtbSAICAR in complex with competitive inhibitor ACP and IMP.....	95
4.16. Overall structure of MtbSAICAR in complex with competitive inhibitors ACP and IMP .....	96
4.17. Summary .....	101
<b>Chapter5. Comparison within MtbSAICAR structure and with other homologues structures (ScSAICAR, TmSAICAR, EcSAICAR, SpSAICAR and HsSAICAR) .....</b>	<b>102</b>
5.1 Introduction .....	102
5.2. Comparison MtbSAICAR structures .....	102
5.2.1. Comparison the structure of MtbSAICAR-ACP-IMP to MtbSAICAR-ATP and MtbSAICAR-ADP-AMZ complexes ..	102
5.3. Comparison within MtbSAICAR-ADP-AMZ complex with other homologues structures (ScSAICAR, TmSAICAR, EcSAICAR, SpSAICAR and HsSAICAR).....	105
5.3.1. Structure alignment of MtbSAICAR-ADP-AMZ complex and TmSAICAR .....	106
5.3.2. Structure comparison of MtbSAICAR-ADP-AMZ and ScSAICAR complexes.	111
5.3.3. Structure comparison of MtbSAICAR-ADP-AMZ and EcSAICAR complexes..	117
5.3.4. Structure alignment of MtbSAICAR-ADP-AMZ and SpSAICAR complexes....	122
5.3.5. Structure alignment of MtbSAICAR-ADP-AMZ and HsSAICAR .....	126
5.3.6. Structure alignment of MtbSAICAR-ADP-AMZ and MaSAICAR.....	132
5.4. Mechanism and action of SAICAR synthetase (SAICARs) .....	135
5.5. Summary .....	137
<b>Chapter6. Characterisation studies of purD of <i>M. tuberculosis</i> (MtbpurD) .....</b>	<b>139</b>
6.1 Introduction .....	139
6.2. Preparation of MtbpurD .....	139
6.3. Determination of Secondary structure of MtbpurD.....	144
6.4. Investigating thermal stability of MtbpurD by thermal shift assay .....	146
6.5. Crystallisation trials of MtbpurD.....	148
6.6. Homology model alignment of MtbpurD .....	150
6.7. Summary .....	154
<b>Chapter7. Conclusion .....</b>	<b>155</b>
<b>References .....</b>	<b>160</b>

## Abbreviations

<b>2Fo-Fc</b>	approximating to the electron density of the protein
<b>°</b>	degree
<b>σ</b>	sigma (Standard deviation)
<b> F </b>	structure factor amplitudes
<b>Å</b>	angstrom (1Å=0.1nm)
<b>a,b,c</b>	crystallographic cell vectors
<b>α, β, γ</b>	crystallographic cell angles
<b>ACP</b>	adenylyl-imidodiphosphate
<b>ADP</b>	adenosine di phosphate
<b>AMP</b>	adenosine 5'-monophosphate
<b>AMZ</b>	5-aminoimidazole-4-carboxamide ribonucleotide
<b>ATP</b>	adenosine tri phosphate
<b>bp</b>	base pair
<b>B-factor</b>	temperature factor
<b>BSA</b>	bovine serum albumin
<b>CD</b>	circular dichroism
<b>CCD</b>	charge-coupled device
<b>Completeness</b>	percentage of observed reflections from the all possible reflections
<b>Da</b>	dalton
<b>DLS</b>	diamond light source
<b>DMSO</b>	dimethylsulfoxide
<b>DNA</b>	deoxyribonucleic acid
<b>dNTP</b>	deoxyribonucleotide triphosphate
<b>DTT</b>	dithiothreitol
<b>EDTA</b>	Ethylenediamine-tetraacetic acid
<b>F1</b>	forward primer
<b>Fc</b>	calculated structure factors
<b>Fo</b>	observed structure factors
<b>Fo-Fc</b>	difference electron density map
<b>fTHF</b>	N10-formyl-tetrahydrofolate
<b>GMP</b>	guanosine 5'-monophosphate
<b>HEPES</b>	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
<b>I</b>	intensity of each reflection
<b>Iσ</b>	uncertainties of intensity
<b>I/σI</b>	signal-to-noise
<b>K</b>	kelvin
<b>k<sub>cat</sub></b>	Turnover number
<b>kDa</b>	kilo Dalton
<b>K<sub>M</sub></b>	michaelis constant
<b>IMP</b>	Inosine 5'-monophosphate

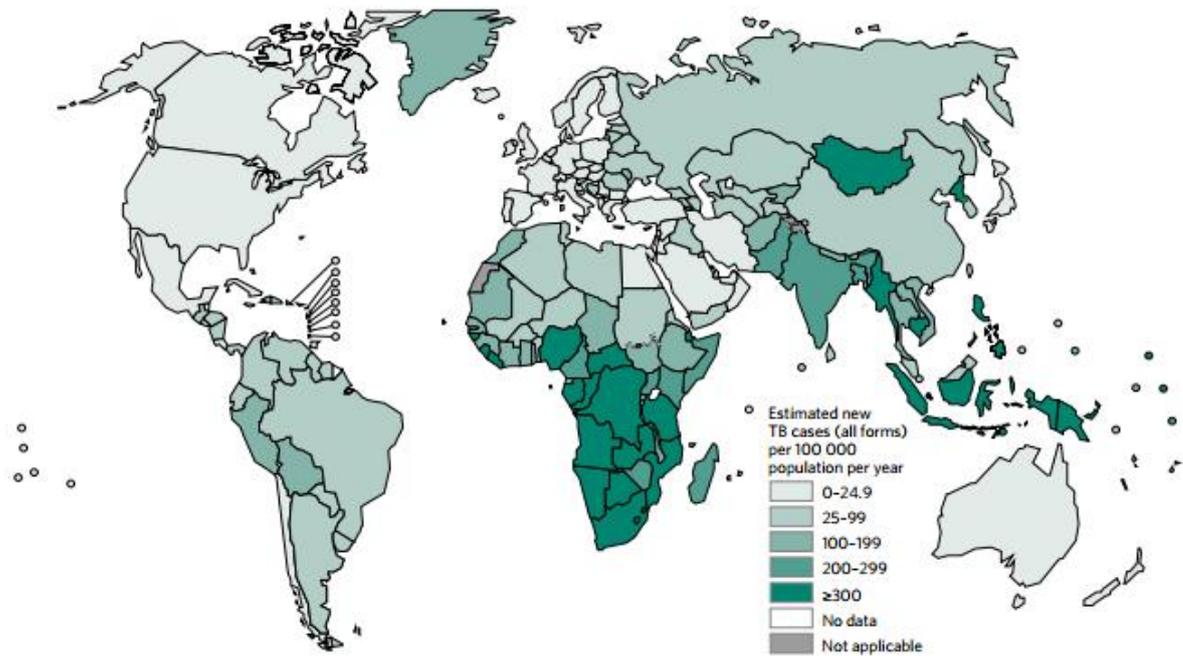
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>mAU</b>	milli absorbance unit
<b>mdeg</b>	millidegrees
<b>MgCl<sub>2</sub></b>	magnesium chloride
<b><math>\mu</math>l</b>	microliter
<b><math>\mu</math>M</b>	micromolar
<b>MDR</b>	multi drug resistance
<b>MR</b>	molecular replacement
<b>M.W</b>	molecular Weight
<b>Ni-NTA</b>	nickel-nitrilotriacetic acid
<b>nm</b>	nanometer
<b>OD</b>	optical density
<b>ORF</b>	open reading frame
<b>PCR</b>	polymerase chain reaction
<b>PDB</b>	protein Data Bank
<b>PKM2</b>	pyruvate kinase protein isoform2
<b>R5P</b>	D-Ribose 5-phosphate disodium salt
<b>R1</b>	reverse primer
<b>Redundancy</b>	number of observations/number of unique
<b>R factor</b>	residual Factor
<b>R free</b>	R factor calculated subset(~5%) of data
<b>RFU</b>	relative fluorescence unit
<b>R merge</b>	the internal agreement between multiple observations of unique intensity
<b>RMSD</b>	root Mean Square Deviations
<b>RNA</b>	ribonucleic acid
<b>rpm</b>	revolutions per minute
<b>S</b>	second
<b>SAICAR</b>	succinyl-5-aminoimidazole-4-carboxamide-1-ribose-5'-phosphate
<b>SDS-PAGE</b>	sodium dodecyl sulfate polyacrylamide gel electrophoresis
<b>SUC</b>	succinic acid
<b>TEV protease</b>	tobacco etch virus (TEV) protease
<b>Tm</b>	melting temperature
<b>Tris-HCl</b>	tRIS hydrochloride
<b>V<sub>0</sub></b>	initial reaction velocities
<b>v/v</b>	volume/volume
<b>V<sub>max</sub></b>	maximum velocity
<b>WHO</b>	world health organisation
<b>XDR TB</b>	extensively drug resistance strain

## Chapter 1. Introduction

Tuberculosis (TB) is one of the world's most lethal diseases (Riley, 1983). Understanding the physiology and metabolism of the pathogen may facilitate understanding of the disease. The formulation by Koch of his famous four postulates to establish the causes of a disease, whereby he hypothesised that a microorganism isolated from similarly infected tissues, grown in freshly prepared culture and inoculated into a new susceptible animal must be able to be regrown in culture, led to his identification of the causative agent of TB. In 1882, Koch confirmed that *Mycobacterium tuberculosis* (Mtb) is the causative agent of TB (Sakula, 1982).

The World Health Organisation (WHO) provides a comprehensive overview of TB occurrence worldwide. Each year, a summary of the prevalence of new TB cases and mortality worldwide is reported. In 2015, 10.4 million patients and 1.4 million deaths were reported due to TB infection (WHO 2016). The infection rate of TB infection in the United States of America, Canada, and New Zealand are considered to be low as seen in Figure (1). TB symptoms have a relationship to human immunodeficiency virus (HIV) infection. It has been estimated that HIV patients who have previously been infected with *M. tuberculosis* have an increased chance of reactivation of the disease (Dierberg and Chaisson, 2013).

Since the discovery of *M. tuberculosis*, extensive studies have been carried out on its cell wall, the characteristic features, pathogenesis, diagnosis, treatment, and dormancy. Researchers conclude that there are many factors that promote TB infection including homelessness, poverty, and AIDS (Waalder, 2002). The infection is commonly transferred from human to human during close contact between individuals such as family members, friends, work colleagues and others (Rose et al., 1979). Although the current TB treatment is mostly effective against replicative *M. tuberculosis* (Zhang and Yew, 2009), drug toxicity (Nguyen and Menzies, 2001) and emerging outbreaks of HIV means that new effective therapeutic agents are required. Defining the essential virulence genes may provide antimicrobial targets by inhibiting the key biochemical processes. This demands a further understanding of the *M. tuberculosis* genome.



**Figure (1)** Estimation of New TB cases in 2015. The rate of TB is indicated by the gradient of color and the estimated number is given below the figure. The lighter the color the less new incident cases will be (WHO 2016). [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/)

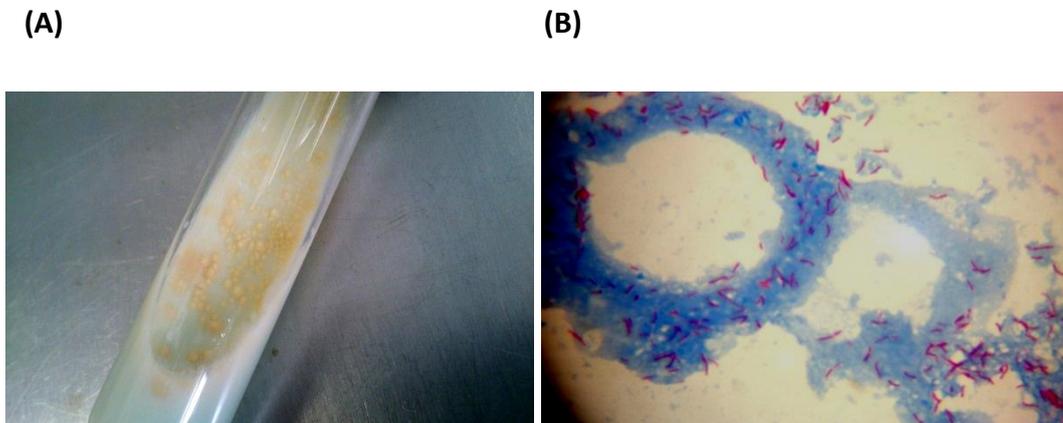
### 1.1. *Mycobacterium tuberculosis* (Mtb)

*M. tuberculosis* is a member of *Actinobacteria* (gram-positive bacteria with a high G+C content ranges which vary between 51% and 70%) (Ventura et al., 2007). *M. tuberculosis* is a species of the *Mycobacterium* genus, which is distinguished from other closely related bacteria by phenotypic and biochemical tests e.g. niacin positive (Kubica et al., 2006). Its special characteristic ability to remain alive inside the host macrophages, where may become dormant provides an additional challenge to treatment. From the dormant state, the pathogen is capable of activating to cause the disease at any time (Lin and Flynn, 2010).

#### 1.1.1. Characteristic features of *Mycobacterium tuberculosis* (Mtb)

*M. tuberculosis* (Mtb) is classified as one of the slow-growing bacteria of *Mycobacterium* genus (Parish and Stoker, 1999), which require 4 to 6 weeks to grow colonies. This slow growth may be due to the hydrophobic and complex cell wall structure that hinders nutrient uptake as described in Section 1.1.2 (Ryan, Ray, and Sherris, 2004; Gebhardt et al., 2007). The morphological characteristics of the non-

motile *M. tuberculosis* is characterised either under the microscope or in culture. Mtb bacilli have a rod shape and stain red against the blue background, staining the *M. tuberculosis* bacilli is described briefly in Section 1.2.2 whereas in culture they are identified as buff-coloured rough dry colonies, as shown in Figure (2) (Essa et al., 2013).

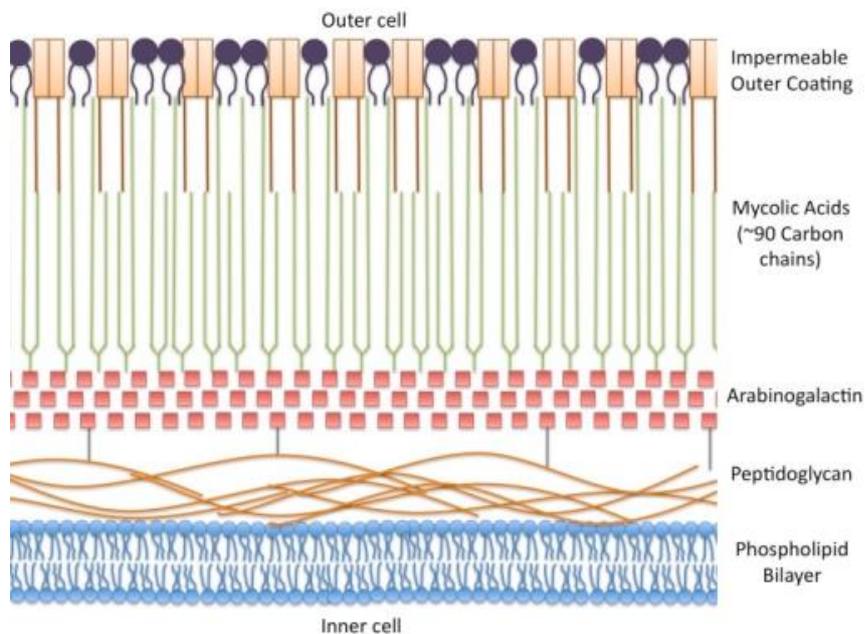


**Figure (2)** Characteristics feature of *Mycobacterium tuberculosis* (Mtb). (A) *M. tuberculosis* colonies (4 to 6 weeks) are grown on medium shown as buff-coloured rough dry. (B) Gram positive *M. tuberculosis* on microscopic slide stained by acid-fast stain and appeared as rod like shape coloured red against a blue background (Orduz and Ribón, 2015).

### 1.1.2. Cell Wall characteristics of *M. tuberculosis*

The survival of *M. tuberculosis* and the process involved in the development of disease are related to the complexity of the cell wall with a high lipid content (about 60%) (Indrigo, Actor, and Hunter, 2002). The cell wall (Figure 3) consists mainly of three layers located outside the phospholipid bilayer. In mycobacterial species, the stability and rigidity of cell wall is provided by the presence of peptidoglycan (PG), which constitutes 30%-70% of Gram-positive cell walls (Schleifer and Kandler, 1972). PG in *M. tuberculosis* is unusual in having *N*-glycolylmuramic acid (MurNGlyc) rather than *N*-acetylmuramic acid (MurNAc), which is shown to have a role in increasing the resistance of mycobacteria to the bactericidal enzyme lysozyme. This was demonstrated with a study of the effect of lysozyme on a mutant mycobacterial culture with a deletion in the *NamH* gene that reveals a decrease in the growth rate because of this vulnerability to lysozyme (Raymond et al., 2004).

The inner PG layer binds to the arabinogalactan (AG) (polymers of galactose and arabinose) layer by phosphodiester bonds. AG connected to PG through linker between the C6 atom of MurNGlyc residues and galactan (Jankute et al., 2015). The next layer of mycolic acids consists of long chain fatty acids (Smith, 2011). Trehalose-free mutant strains with altered mycolic acid layers showed increased susceptibility to antibiotics, indicating the significance of mycolic acid on the permeability of the cell wall (Gebhardt et al., 2007).



**Figure (3)** Components involved in the formation of the cell wall. Inner coats are peptidoglycan (PG), arabinogalactan (AG), while mycolic acids are next to the outer layer (Smith, 2011).

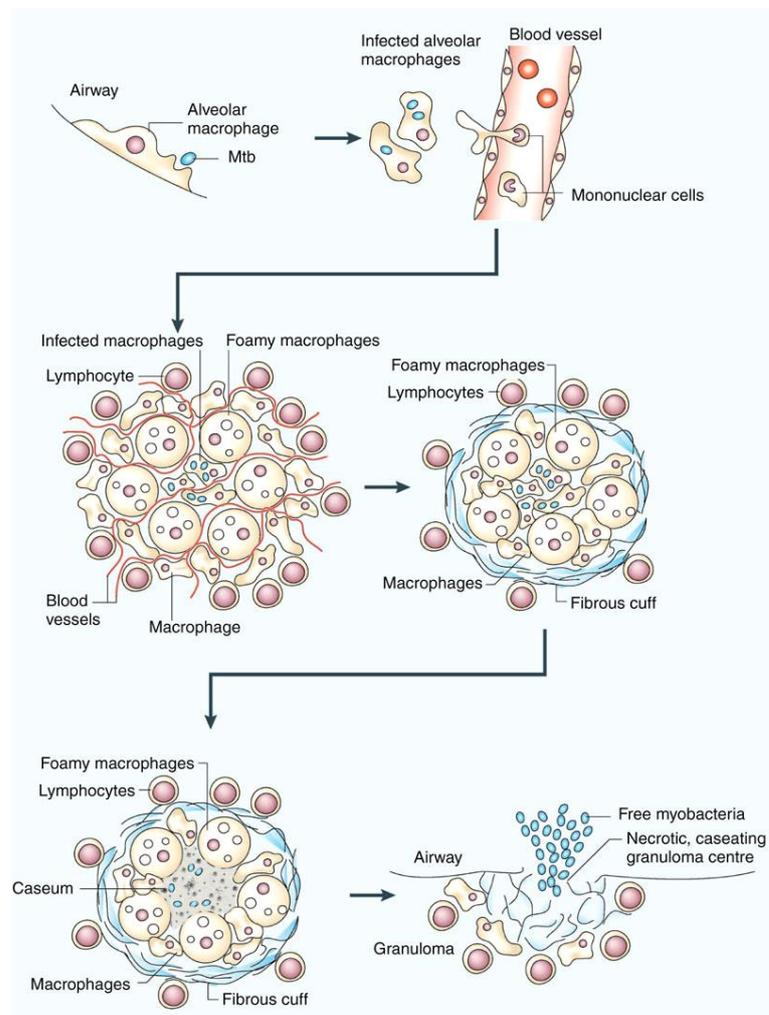
## 1.2. Tuberculosis (TB)

It has been estimated that around 30% of people exposed to *M. tuberculosis* develop the disease soon after infection (Herrera et al., 2011). Approximately 10% of *M. tuberculosis*-infected persons activate the disease during their lifetime and of these 50%-80% develop symptoms in the first 2 years of primary infection (Kaye and Frieden, 1996). Once the disease takes hold, the infection can spread to other individuals easily, this is because the bacilli can stay viable for a long time in the air after being disseminated by speaking, sneezing and coughing. The disease can infect the lungs (pulmonary) or extra pulmonary (where it infects several parts of the human body). The infection can also be active or latent, active TB patients experience clear signs and

symptoms, whereas in latent TB the bacilli remain inactive (not replicating) until the disease reactivates by the trigger reasons as described in Section 1.2.3 (Mahalakshmi.T and Karpagam, 2012).

#### **1.2.1. Pathogenicity of *M. tuberculosis* (Mtb)**

Once *M. tuberculosis* reaches the alveoli, the bacilli may die when the host's alveolar macrophages engulf them with the initiation of phagosome-lysosome fusion formation. If the macrophages are infected, they attempt to control the infection by producing cytokines such as tumour necrosis factor-alpha (TNF- $\alpha$ ) and other chemokines in order to attract other immune cells such as neutrophils and lymphocytes (Russell, 2006), these then form the granuloma simultaneously releasing toxic materials (reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI)). The granuloma or tubercle is established in order to prevent spreading of infection to lungs and other organs as well as to increase the burden of the immune cells around the lesion (Fenton and Vermeulen, 1996; Flynn and Chan, 2001). The granulomas are divided into three types, solid, necrotic and caseous. The solid type consists of different immune cells surrounding a small number of bacilli existing in an inactive state. Individuals with a good immune response will thus contain the infection. The formed granuloma will then diminish in size and become calcified, however should necrotic granuloma form with high amounts of cell debris become included, the bacilli may be reactivated. The caseous granuloma is formed when the bacilli recovers its metabolic activity and becomes highly reproductive (Reece and Kaufmann, 2012; Gengenbacher and Kaufmann, 2012). The infection can then spread out not only in the lung (pulmonary) but also to other organs (extra pulmonary) (Sreeramareddy et al., 2008) such as the pancreas and central nervous system (Kandola and Meena, 2014). Figure (4) shows the formation of the granuloma during Mtb infection.



**Figure (4)** Processes involved in granuloma formation during TB infection. The bacilli are inhaled through the respiratory system and the alveolar macrophages engulf them. Immune cells attracted to the lesion site by means of the proinflammatory response lead to granuloma formation in which the infection is controlled. Where the tubercle diminishes in size and becomes calcified. When the immune system is compromised, the infection cycle restarts and the bacilli are released to become free through the airway (from Russell et al., 2009).

Phagosomal maturation (phagosome-lysosome fusion) is a process in which the bacteria are engulfed in order to eliminate it. The process starts when the *M. tuberculosis* is engulfed by the phagosome, which enters into a maturation cycle that eventually leads to the elimination of the invader. The phagosome goes from its early and matures into the phagolysosome (whereby the pH becomes 4.5). But the bacteria can escape, thereby increasing its chance of survival and being able to replicate (Poirier and Av-Gay, 2012). There are some proteins that have a role in the *M. tuberculosis* pathogenicity and strategies for increasing survival. Protein tyrosine

phosphatase (PtpA) lessens the acidity of the phagosome and its maturation by inhibition of the host's vacuolar H<sup>+</sup>-ATPase (V-ATPase). This ATP-dependent proton pump uses the energy of ATP to transport protons which increases the acidity of the phagosome from pH 6.5 to 5.0 which kills the bacteria (Wong et al., 2011).

PtpA is found to be the substrate of protein tyrosine kinase A (PtkA), which phosphorylates it at residues Y128 and Y129 and so enhance its activity. It has been seen that inhibition of the interaction between PtkA and PtpA affects bacterial growth rate (Zhou et al., 2014; Jaiswal et al., 2018). A recent study on PtkA (Bach, Wong, and Av-Gay, 2009; Niesteruk et al., 2018) revealed that deletion of the gene coding PtkA makes *M. tuberculosis* unable to prevent acidity production in the phagosome during the infection and decreases the growth rate inside the macrophage. But an *in vitro* study revealed that the PtkA mutant strain was found to be able to resist oxidative stress such as H<sub>2</sub>O<sub>2</sub>. Thioredoxin reductase TrxB2 is also a substrate for PtkA, and it is part of the mechanism used to eliminate toxic species such as reactive oxygen intermediates (ROI), it is activated by phosphorylation at Y32 by PtkA. Mutation of Y32 prevents phosphorylation and thus deactivates. However, this role of TrxB2 has been challenged by complementation studies that show the same characteristics as the wild-type. Other proteins that become down-regulated by a lack of PtkA, such as rRNA methyltransferases may provide protection against oxidative stress (Wong et al., 2018).

EsxA (also called ESAT-6) (Brodin et al., 2004) is "virulence factor" protein that is expressed by *M. tuberculosis* and encoded by gene in Region of Difference 1 (RD1) that is found in all pathogenic *Mycobacteria* but is absent in the non-pathogenic *M. bovis* bacillus Calmette-Guérin (BCG) strain (Wang et al., 2009). It is also an *in vitro* T cell antigen that is able to identify TB active patients (Arend et al., 2000). EsxA has a role in membrane permeabilisation during the translocation of *M. tuberculosis* into the macrophage cytosol after the inhibition of the phagosomal maturation and the C-terminal site has a major role in this process (Houben et al., 2012) and a role in the destruction of the macrophage cell membrane by the apoptotic process (Derrick and Morris, 2007). IFN- $\gamma$  a cytokine that has an important role in the immune system and survival of the host (Green, DiFazio, and Flynn, 2012; Kak, Raza and Tiwari, 2018) this

induces the autophagy process whereby the bacteria are eliminated (Mizumura, Choi and Ryter, 2014). EsxA is found to be able to reduce IFN- $\gamma$  gene transcription and expression (Wang et al., 2009)

The essential role of EsxA has also been studied at the molecular level, a single residue mutation at glutamine 5 (Q5) to either Lysine or Valine may affect the translocation of *M. tuberculosis* into the macrophage cytosol. Valine was shown to up-regulate the membrane permeabilisation activity while lysine was seen to down regulate it (Zhang et al., 2016). The importance of EsxA in the virulence and pathogenesis of *M. tuberculosis* has been evaluated using *M. marinum* (*Mm*) as a model; as this wild-type strain shares similar EsxA properties with *M. tuberculosis*. The activity of EsxA on the attachment and survival inside the macrophage has been assessed using *Mm*, *Mm* $\Delta$ EsxA, *Mm* $\Delta$ EsxA(WT), *Mm* $\Delta$ EsxA(Q5K) and *Mm* $\Delta$ EsxA(Q5V) and *M. smegmatis* (*Ms*). It has been seen that wild-type *Mm* has a higher attachment rate and survival than mutant and *Ms* strains. Suggesting the importance of EsxA. Also, *Mm* $\Delta$ EsxA(WT) which restore the EsxA features and *Mm* $\Delta$ EsxA(Q5V) were shown to have higher survival rates than *Mm* $\Delta$ EsxA(Q5K) suggesting higher virulence increasing macrophage cell death (Zhang et al., 2016).

The effect on EsxA activity with phthiocerol dimycocerosates (DIM) (a lipid component in *M. tuberculosis* cell wall) (Camacho et al., 2001), another virulence factor, has also been studied and it has a significant role in *M. tuberculosis* virulence. It was examined using a phagosomal damage marker, such as galectin-3 and ubiquitinated protein, and found that mutated forms are unable to cause damage to the phagosomal membrane, unlike the wild and complementation strains. In the mutant strains, phagosomal damage marker was prevented from accumulation around infected macrophages (Augenreich et al., 2017). The effect of DIM on EsxA has been explored and it was found that DIM induces EsxA activity, this has been confirmed upon a study using BCG wild-type strain (*Mycobacterium bovis* Bacillus Calmette-Guérin vaccine strain) which is deficient in the ESX-1 systems. Comparison between mutant strains (DIM-proficient BCG::ESX-1 and DIM-deficient BCG::ESX-1) found that the presence of DIM in the former strain increases the accumulation of phagosomal markers around the phagosome carrying the strain (Augenreich et al., 2017).

GarA has been determined to be a substrate for Ser/Thr-protein kinase G (PknG) in *M. tuberculosis*. This has been determined using KESTREL (Kinase Tracking and Substrate Elucidation) (Knebel, 2001) as well as other methods (Mueller and Pieters, 2017). The importance of this protein and its potential role in *M. tuberculosis* pathogenesis has been studied. It has been found that mutation in this protein ( $\Delta\text{garA}_{\text{Mt}}$ ) prevents the strain from growing inside the macrophage as well as its propagation inside the lungs of mice. Whereas the wild-type strain (H37RV) showed the opposite and was able to grow in the lungs and disseminate further to the spleen. This information demonstrates the essential role of this protein in the virulence of *M. tuberculosis* (Rieck et al., 2017). *pknG* was found to promote the growth of *Mycobacteria* inside the host and probably prevents phagosome maturation by phosphorylating the host components (Pieters, 2008; Li and Xie, 2011). The important role of *pknG* was assessed *in vivo* and *in vitro*, an early study revealed that *pknG* mutant strains (*Rv* $\Delta$ G) weaken inside the host and unable to resist the oxidative stress produced (Cowley et al., 2004). This was supported by an *in vitro* study, which confirmed this situation as well as sensitivity to an acidic environment (Khan et al., 2017). In the latency condition, *pknG* phosphorylates GarA, which maintains the viability of *mycobacteria*. *pknG* mutant strains prevent phosphorylation of residue T21 of GarA decreasing the survival rate of the *Mycobacteria* (Khan et al., 2017).

### **1.2.2. Diagnosis of *M. tuberculosis* (Mtb)**

Controlling the spread of the disease requires rapid diagnosis following the occurrence of initial symptoms like coughing, bloody sputum, and weight loss. There are a variety of diagnostic techniques for monitoring the TB disease. Microscopic examination is the traditional method of diagnosis and is an efficient, quick and inexpensive method. *M. tuberculosis* is classified as gram-positive because of the high similarity to their genes. Although Mtb has a thick peptidoglycan layer and lack a true outer layer in their cell wall structures, it does not retain the Gram dye as Gram-positive bacteria do thus they are called acid-fast bacilli (Section 1.1.1.) (Hett and Rubin, 2008). Ziehl-Nielsen is traditional staining method used to identify *M. tuberculosis* (Parish and Stoker, 1999; Forrellad et al., 2013). To explain the method, in brief, the sample is applied on a slide

then stained using the acid-fast stain carbol fuchsin, then the decolouriser acid alcohol and counterstain of methylene blue (Dezemon, Muvunyi, and Jacob, 2014).

Tuberculin skin test (TST) is another diagnostic test in which a small amount of *M. tuberculosis* antigens are injected intradermally into the patient. The immune response is assessed within 48 to 72 hours (Nayak and Acharjya, 2012). Although TST test is more convenient, false negatives in patients diagnosed with immune-suppressed patients or false positives in people with vaccinated Bacilli Calmette-Gue´rin (BCG) can be obtained. Immunological test Interferon- $\gamma$  release assays (IGRAs), which measures the IFN- $\gamma$  expression level indicates TB infection (Pai, Riley and Colford, 2004). The technique is expensive and requires a normal blood sample. Like TST, it also unable to discriminate between both TB forms especially in areas of high dissemination of Latent TB (Turetz and Ma, 2016). Active TB has clear signs and symptoms, which contribute to the diagnosis of the disease whereas the latent state does not. Recent studied estimates that there are 169 genes expressed differently, some of these genes are implicated in inflammation, immune system response and apoptosis (Lee et al., 2016).

### **1.2.3. *M. tuberculosis* (Mtb) Dormancy**

The majority of infected individuals possess the infection in the latent state; a poorly understood form (Deb et al., 2009). This describes the situation where bacterial infection is not diagnosed and the infected individual is asymptomatic (McDermott, 1959). The bacteria lose their morphological feature seen in the acid-fast stain. It is characterised by the bacteria's slow growth in a non-replicating state where the metabolic activity is reduced (Kaprelyants, Gottschal, and Kell, 1993; Kell and Young, 2000). Treatment is therefore difficult (McDermott, 1959; Zhang, 2004).

Several factors such as a change in pH, insufficient nutrient and oxygen contribute to dormancy (Shleeva, Salina, and Kaprelyants, 2010). The dormant cell tries to survive by compensating oxygen depletion by using a nitrate respiration strategy that protects them from hypoxia, acidic and reactive nitrogen species toxicity (Weber et al., 2000; Tan et al., 2010). But at any time the cells can be activated by means of some factors such as immunodeficiency virus (HIV) infection, renal failure and diabetes (Horsburgh,

2004). Using animal (rabbit) models, the disease is reactivated by means of an immunosuppression agent such as triamcinolone (Chao and Rubin, 2010; Subbian, Eugenin, and Kaplan, 2014) such that the dormant bacilli growth will be active when exposed to enough oxygen (Wayne and Diaz, 1967).

Research on immune cells and their cytokines revealed that CD4+T cells have a role in the control and eradication of infected macrophages (Flynn and Ernst, 2000) and that up-regulation of IFN- $\gamma$  enables the destruction the bacilli by activation of the macrophage (Flynn, 1993). In addition, the absence of critical cytokines like TNF- $\alpha$  can increase the burden of bacilli by preventing granuloma formation; this study has taken place in mice (Senaldi et al., 1996).

#### **1.2.4. Treatment**

The treatment consists of a course over six months; the first two months include four drugs (isoniazid, rifampicin, pyrazinamide, ethambutol) then another four months of two drugs (isoniazid, rifampicin) (Zhang, 2005). Isoniazid (Isonicotinic acid hydrazide, INH) is an inexpensive and specific drug that targets active *M. tuberculosis* bacilli (Bardou et al., 1998). INH treatment inhibits mycolic acid formation, an important virulence factor and component of antibacterial drug resistance (Quemard, Lacave and Laneelle, 1991). INH is actually a prodrug, enters the cytoplasm of *M. tuberculosis* and is activated by the heme enzyme catalase-peroxidase (KatG) to become toxic to the *M. tuberculosis*. The activated form reacts with NADH and inhibits the enoyl reductase enzyme (Banerjee et al., 1994) involved in mycolic acid synthesis (InhA), an inhibitor ternary complex forms and thus disrupts mycolic acid formation. Unlike INH, rifampicin is also active against dormant *M. tuberculosis* bacilli. It binds to  $\beta$ -subunit of the RNA polymerase and prevents messenger RNA production. Mutations in the gene coding the enzyme may result in rifampicin resistance strains (Nath and Ryoo, 2013; Palomino and Martin, 2014).

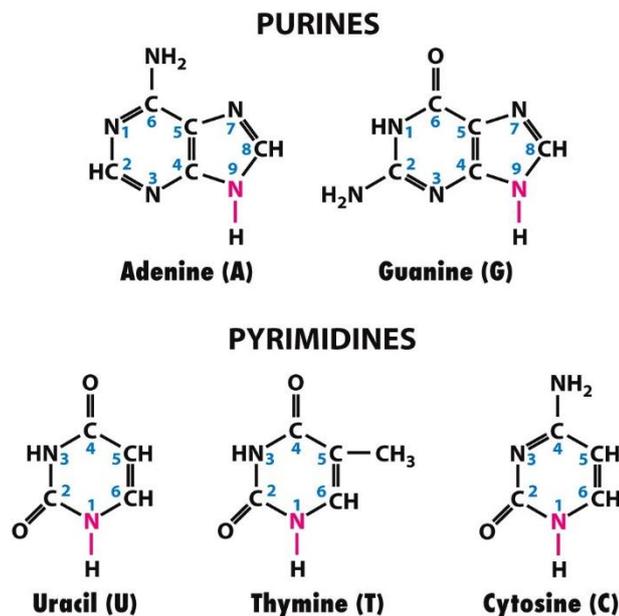
The emergence of resistance and loss of effectiveness of these drugs due to the long period of treatment giving rise to poor patient compliance, made worse by poverty, (Wallis et al., 1999). Resistance to both INH and rifampicin may be shown as multi-drug resistance (MDR). This indicates another drug combination is required, the second line

drugs are capreomycin, kanamycin, and amikacin, but these require prolonged administration (up to 24 months), have high toxicity and are costly (Hum and Sungweon, 2013). Extensive drug resistance strains have been seen to occur (XDR TB) (Andrews et al., 2007). Analysis of this microorganism's genetics has identified factors associated with virulence and survival. The virulent strain of *M. tuberculosis* H37RV is the first *Actinobacteria* species whose genome was sequenced (Cole et al., 1998).

### 1.3. Purine biosynthesis Pathway

#### 1.3.1. Introduction

Nucleotides are important for the essential biological processes in living organisms including DNA and RNA production (Carlucci et al., 1997), as an energy carrier, in the biosynthetic process and for signal transmission (Lane and Fan, 2015). Nucleotides are comprised of phosphate, a ribose sugar and a purine or pyrimidine nitrogenous base (nucleosides) (Zöllner, 1982). The important nitrogenous bases are adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U). Both adenine (A) and guanine (G) are purines whereas cytosine (C), thymine (T) and uracil (U) are pyrimidine bases. The chemical structures of purine and pyrimidine bases are illustrated in Figure (5).

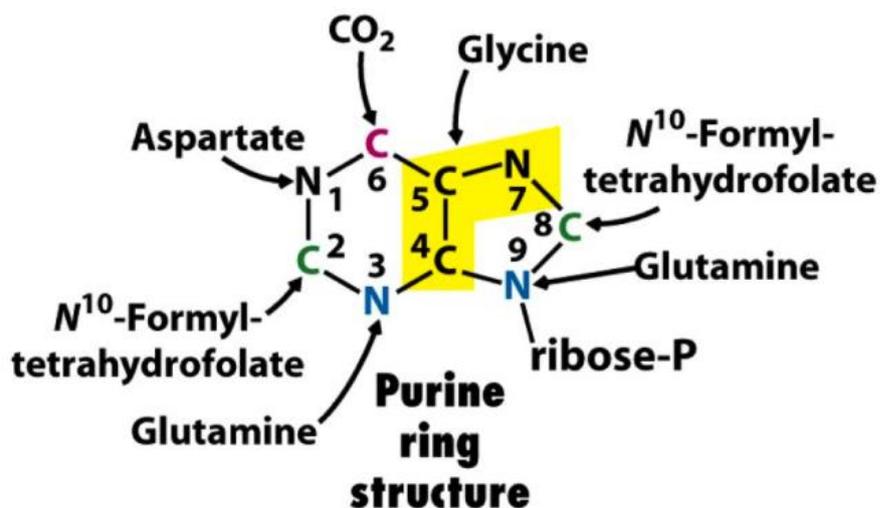


**Figure (5)** Nitrogenous bases involved in the formation of nucleic acids. Adenine and Guanine are classified as purines whereas Cytosine, Thymine, and Uracil are pyrimidines (from Lodish et al., 2016).

The building blocks of purines and pyrimidines can be synthesised from recycled bases (salvage pathway) or from scratch (*de novo* synthesis) (Nelson, Cox and Lehninger, 2009). In purine *de novo* synthesis, bases assembled bound to ribose rings, whereas in the formation of pyrimidine bases, the bases are initially built then attached to the ribose. In the salvage pathway, the resynthesising of the purine and pyrimidine nucleotides is from free bases. The action of this pathway is particularly important in some organisms that rely primarily on the salvage pathways. These include parasites and bacteria such as *Giardia lamblia* (Sarver and Wang, 2002), *Mycoplasmas* (Wang et al., 2001) and *Helicobacter pylori* (Liechti and Goldberg, 2011). Whereas other microorganisms including *M. tuberculosis* rely on *de novo* pathways, the synthesis of the purine ring in *M. tuberculosis* from precursors is described in Section 1.3.2.

The synthesis of purine nucleotides through the salvage pathway is energetically favorable for the cell because it only utilises one ATP molecule to form nucleotides within the pathway (Moffatt and Ashihara, 2002). The enzymes involved in salvaging adenosine and its base adenine have been studied in more detail than the other enzymes of the purine biosynthesis pathway. The adenine-recycling pathway is mediated by adenine phosphoribosyl-transferase enzyme (APT) (EC: 2.4.2.7), which is involved in converting the adenine base and phosphoribosyl-1-pyrophosphate (PRPP) into AMP and pyrophosphate (PPi). This reaction shares the precursors with the salvage pathway and purine biosynthesis pathway (Moffatt and Ashihara, 2002). A similar enzymatic reaction that is catalysed by hypoxanthine-guanine phosphoribosyltransferase (HGPRT) occurs for the free purine base guanine to form guanosine 5'-monophosphate (GMP) named guanylate (Torres and Puig, 2007).

*De novo* synthesis of purines and pyrimidines is achieved through two separate metabolic pathways. Purine nucleotides are generally synthesised following the same mechanism in different organisms. For the synthesis of purine nucleotides from scratch (Bolton, 1954) inosine 5'-monophosphate (IMP) is the primary product which is then converted to either adenosine 5'-monophosphate (AMP) or guanosine 5'-monophosphate (GMP). The contribution of the IMP synthesis precursors shown in Figure (6).



**Figure (6)** Precursors involved in Purine nucleotide structure. Formation of purine bases uses nitrogen from aspartate, glutamine and glycine amino acids labelled as N1, N3-N9, and N7 respectively. Carbon atoms C2 and C8 are provided by a tetrahydrofolate activated derivative, C6, C4, and C5 from glycine are highlighted in yellow. Figure taken from (Berg et al., 2015).

In humans, purine biosynthesis is a target for anti-cancer therapy (Zhao et al., 2013) as cancer cells rely on the *de novo* pathway for purine biosynthesis. Folate contributes two enzymatic reactions of the purine *de novo* pathway and provides carbon 2 and carbon 8 of purine nucleotides (Deacon et al., 1985). Both eukaryotic and prokaryotic organisms require folate for their cell development and reproduction (Rollins and Lindley, 2005). Antifolate drugs such as Pemetrexed have a great effect on these cancer cells as it inhibits folate formation. In addition, succinyl-5-aminoimidazole-4-carboxamide-1-ribose-5'-phosphate (SAICAR), the product of SAICAR synthetase (an enzyme involved in the purine biosynthesis pathway) has been shown to have a role in cancer cells proliferation as the presence of SAICAR in complex with pyruvate kinase protein isoform2 (PKM2) induces its activity which is found to be required for cancer cell growth. It has also been observed that PKM2 has a role in activation some genes that contribute in malignancy; replacement of this isoform with another may reduce the cancer cells viability (Keller et al., 2014).

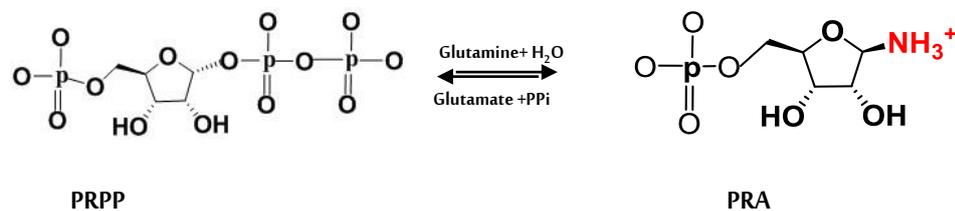
### 1.3.2. The steps of the purine biosynthesis Pathway in *M. tuberculosis* (Mtb)

In the 1950s, Buchanan and his colleagues studied the purine biosynthesis pathways in pigeon and chicken livers (Hartman and Buchanan, 1959). They showed that the pathway consists of ten enzymatic steps and utilises four ATP molecules to fulfill the

synthesis of the inosine 5-monophosphate (IMP). The enzymatic steps of *M. tuberculosis* (Mtb) Inosine-5-phosphate biosynthesis via the *de novo* pathway are described below.

### 1.3.2.1. Amidophosphoribosyltransferase (purF)

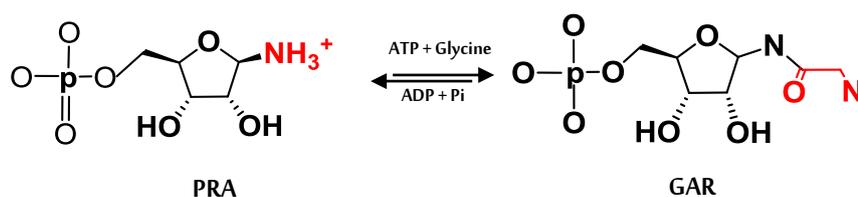
The initial substrate of the pathway is phosphoribosyl-1-pyrophosphate (PRPP), which comes from Ribose-5-phosphate, the product of the oxidative phase of the pentose phosphate pathway. Amidophosphoribosyltransferase (purF) (EC: 2.4.2.14) is the first enzyme, it catalyses the conversion of phosphoribosyl-1-pyrophosphate (PRPP) to phospho- $\beta$ -D-ribosyl-amine (PRA) (Muchmore et al., 1998) whereby the pyrophosphate of PRPP is replaced by the amino group from a glutamine side chain (Smith, 1998) as illustrated in Figure (7). This enzyme is also known as glutamine phosphoribosylpyrophosphate amidotransferase (GPAT) and is an enzyme of the amidotransferases family, whose members also participate in the formation of nucleotide bases, amino acids and antibiotics (Smith et al., 1994). The structures of the enzyme from bacterial microorganisms such as *Escherichia coli* (Muchmore et al., 1998) and *Bacillus subtilis* (Smith et al., 1994) have been solved, although homologous and having 40% sequence identity, these two enzymes are differentiated by a bound metal presence in the *B. subtilis* enzyme which is replaced by water in the *E. coli* enzyme (Zhang, Morar, and Ealick, 2008). It was demonstrated that a single mutation in the Fe<sub>4</sub>S<sub>4</sub> binding site is sufficient to lose the enzyme activity (Makaroff, Paluh, and Zalkin, 1986; Grandoni et al., 1989).



**Figure (7)** The Biochemical reaction of Amidophosphoribosyltransferase (purF)

### 1.3.2.2. Phosphoribosylamine glycine ligase (purDs)

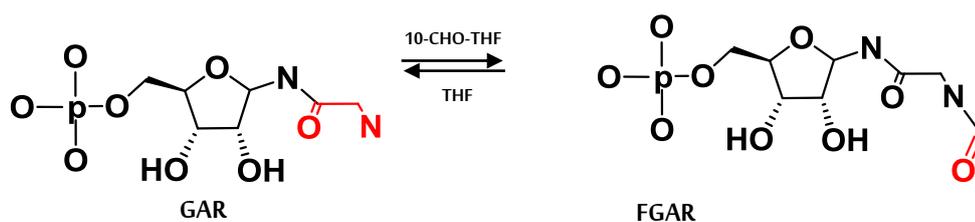
The second step is catalysed by phosphoribosylamine glycine ligase (purDs) (EC: 6.3.4.13). It catalyses the conversion of 5-phospho-D-ribosylamine into N (1)-(5-phospho-D-ribosyl) glycinamide (glycinamide ribonucleotide, GAR) by linking the carboxyl group from glycine to the amino group of PRA to form an amide bond (Zhang et al., 2008). The reaction is illustrated in Figure (8). Phosphoribosylamine glycine ligase will be discussed further in Section 1.3.3.2.



**Figure (8)** The Biochemical reaction of Phosphoribosylamine glycine ligase (purDs).

### 1.3.2.3. Phosphoribosylglycinamide formyltransferase (purN)

The third step is catalysed by phosphoribosylglycinamide formyltransferase (purN) (EC: 2.1.2.2). In its reaction, a formyl functional group, the source of the (C8) atom of the purine nucleotide ring, which is transferred by purN using  $\text{N}^{10}$ -formyl-tetrahydrofolate (fTHF) as the formyl donor (Figure 9). Folate biosynthesis is an anti-cancer target as folate is a critical source for nucleic acid formation by purN (Zhao et al., 2000). This enzyme is monofunctional in the case of *M. tuberculosis*. The crystal structure of purN from *M. tuberculosis* has been solved and it forms a dimer (Zhang et al., 2009), each monomer is composed of two domains; N-terminal (V1–T114) and C-terminal (L115–G215) domains. The active site of the enzyme is located at the interface of these domains. The dimerisation interface mostly consists of non-polar residues (57%) and eight polar amino acids that connect them (Zhang et al., 2009).



**Figure (9)** The biochemical reactions of Phosphoribosylglycinamide formyltransferase (purN).

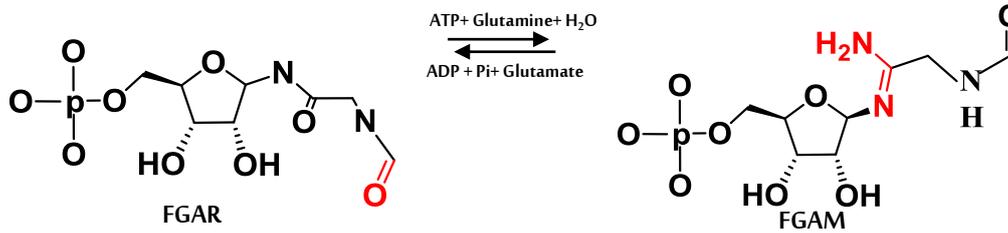
In *E. coli*, an additional enzyme, phosphoribosylglycinamide (GAR) transformylases enzyme (purT), is required, it uses formate as substrate to form formylglycinamide ribonucleotide (FGAR) (Nagy, Mccorkle and Zalkini, 1993). Although two structures have been solved with GAR ATP Mg<sup>2+</sup> complex and ATP Mg<sup>2+</sup> which has an extra Mg<sup>2+</sup>. This Mg<sup>2+</sup> is bound to the  $\gamma$ -phosphate group of ATP, water molecules, and atoms of the enzyme. The enzyme was also solved with Adenylyl-imidodiphosphate (ACP), Mg<sup>2+</sup> and GAR complex and has been compared to ATP Mg<sup>2+</sup> and GAR complex. The conformation of hydroxyl side chain of S160 is oriented differently with a distance 2.6Å and 4.8Å respectively from the  $\gamma$ -phosphate group. It is proposed that the shift of the side chain of S160 toward the solvent allows formate to move into active site (Thoden et al., 2002). Although purN of *M. tuberculosis* shares 36% sequence identity with the equivalent human purN enzyme, the human purN is organisationally different. It is the C-terminal domain of trifunctional human enzyme (HsGART) as described in Section 1.3.3.2.

#### **1.3.2.4. Formyl glycinamide ribonucleotide amidotransferase (purL)**

The fourth step is catalysed by formyl glycinamide ribonucleotide amidotransferase (purL) (EC: 6.3.5.3). It exists as two different types, the large and small purL. The large purL is found in almost all gram-negative bacteria and eukaryotes whereas the small purL is only observed in gram-positive bacteria (Morar et al., 2008). The molecular weight of large purL is about 140 kDa and it is composed of three domains including FGAM synthetase, glutaminase (where the ammonia is formed) and an N-terminal domain, which may have role in ammonia channeling. The small purL has a molecular weight of between 66–80 kDa and is homologous to the FGAM synthetase domain of the large purL (Morar et al., 2006).

In *M. tuberculosis*, the Phosphoribosylformylglycinamide synthase subunit of purL catalyses the reaction using formylglycinamide ribonucleotide (FGAR) and glutamine in order to form formylglycinamide ribonucleotide (FGAM). This domain forms a complex with other, Phosphoribosylformylglycinamide synthase (purQ) and Phosphoribosylformylglycinamide synthase (purS) each with a specific function. Within this complex, purQ produces ammonia, purL transfers ammonia into FGAR to

produce FGAM and purS may act as a supportive molecule in ammonia channeling between purQ and purL (Anand et al., 2004).

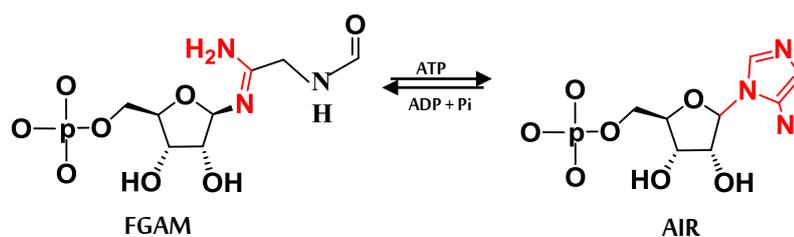


**Figure (10)** The biochemical reactions of Formyl glycinamide ribonucleotide amidotransferase (purL).

The crystal structure of the small purL from *Thermotoga maritima* requires the products of the other subunits to be functionally active. This complex includes the enzymes purQ and purS. Each of these proteins is bound to the other two and then all three proteins combine. purQ of *T. maritima* (TmpurQ) composes of four  $\alpha$ -helices surrounded by nine  $\beta$ -strands. purS of *T. maritima* (TmpurS) is a dimer consisting of two monomers each composed of three  $\beta$ -strands surrounded by two  $\alpha$ -helices. purL of *T. maritima* (TmpurL) is a monomer and is composed of four subdomains classified as A1,A2,B1,B2. The connection between these proteins is mostly mediated by loops (Morar et al., 2008).

#### 1.3.2.5. 5-aminoimidazole ribonucleotide (AIR) synthetase (purM)

The fifth step is catalysed by 5-aminoimidazole ribonucleotide (AIR) synthetase (EC: 6.3.3.1) (purM). The molecular structure organisation of purM differs between organisms. In bacteria and plants, the enzyme is monofunctional but in yeast it is part of a bifunctional enzyme, which includes purD enzyme (Li et al., 1999). In humans, this is part of a trifunctional enzyme (HsGART) as described in Section 1.3.3.2. In *M. tuberculosis*, the enzyme is also called Phosphoribosylformylglycinamide cyclo-ligase, it is a monofunctional enzyme which catalyses the conversion of 2-(formamido)-N (1)-(5-phospho-D-ribosyl) acetamide (FGAM) into 5-amino-1-(5-phospho-D-ribosyl) imidazole (AIR) (Caspi et al., 2015). The first step of the reaction is that the carbonyl group of FGAM is phosphorylated by ATP and then cyclises with the loss of the phosphate to form AIR. This binds the carbonyl group to nitrogen.



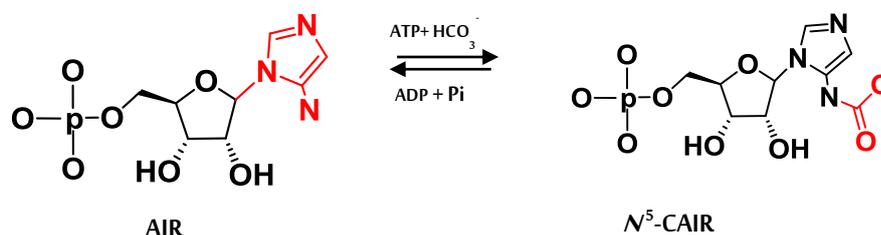
**Figure (11)** The biochemical reactions of 5-aminoimidazole ribonucleotide (AIR) synthetase (purM).

The structure of the enzyme from *E. coli* appears to be homodimer with each monomer composed of two domains, the N-terminal (domain A) and C-terminal (domain B) (Li et al., 1999). The structure of the purM component from humans shows a close similarity to the structure of pur M from *E. coli* with an RMSD of C $\alpha$  positions of 1.3Å.

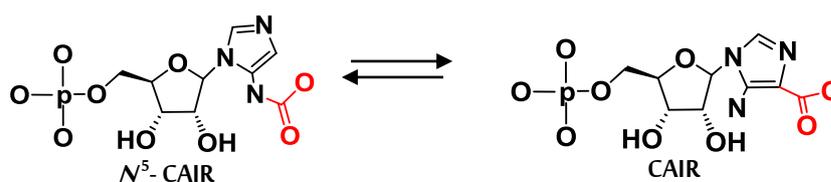
### 1.3.2.6. N5-carboxyaminoimidazole ribonucleotide synthase (purK) & N5-carboxyaminoimidazole ribonucleotide mutase (purE)

The sixth step is catalysed by the bifunctional enzyme that includes N5-carboxyaminoimidazole ribonucleotide synthase activity (purK) (EC: 6.3.4.18) and N5-carboxyaminoimidazole ribonucleotide mutase activity (purE) (EC: 5.4.99.18). 4-carboxy-5-aminoimidazole ribonucleotide (CAIR), the substrate for SAICAR Synthetase (SAICARs) as discussed further in Section 1.3.3.1. is produced by the activity of these two enzymes resulting in the addition of a carboxyl group to AIR as depicted in Figure (12).

(A)



(B)

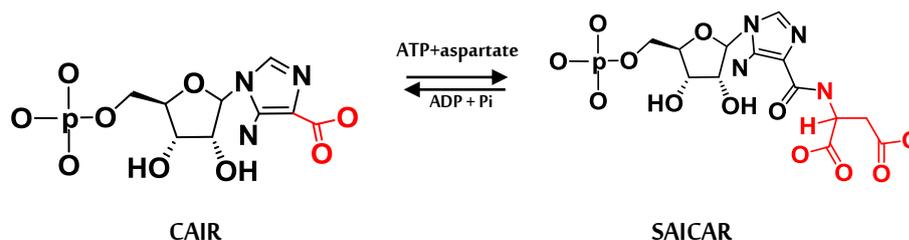


**Figure (12)** The biochemical reactions of N5-carboxyaminoimidazole ribonucleotide synthase (purK) (A). N5-carboxyaminoimidazole ribonucleotide mutase (purE) (B).

As described by Thoden et al., in *E. coli*,  $N^5$ -CAIR synthetase (purK) converts 5-aminoimidazole ribonucleotide (AIR) to  $N^5$ -CAIR then CAIR by  $N^5$ -CAIR mutase (purE class 1) (Thoden, Holden and Firestine, 2008). Although the purK enzyme was crystallised with ATP  $Mg^{2+}$  and AIR, no electron density for AIR was seen whilst the ATP is replaced by ADP+Pi, suggesting hydrolysis of ATP had taken place. The proposed catalytic mechanism was based on a model suggesting the position of AIR using information derived from purT. Because purT (Section 1.3.2.3) shares 25% sequence identity with purK and RMSD  $C\alpha$  value 1.5Å, superimposition suggested the position of AIR is equivalent to that of GAR, the substrate of purT. In the model the phosphate group of ATP bonds to an oxygen atom of bicarbonate to form carboxyphosphate, however, the distance between one of the oxygen atoms of carboxyphosphate and  $N_5$  of AIR is around 4Å suggested decomposition of bicarbonate to carbon dioxide ( $CO_2$ ) and Pi, the  $CO_2$  then attached to AIR (Thoden, Holden and Firestine, 2008). In humans, CAIR is produced directly by the action of AIR carboxylase (AIRc) (purE classII) described in Section 1.3.3.1.

#### **1.3.2.7. SAICAR Synthetase (SAICARs)**

The seventh step is catalysed by SAICARs (EC: 6.3.2.6), this uses 4-carboxy-5-aminoimidazole ribonucleotide (CAIR) to produce succinyl-5-aminoimidazole-4-carboxamide-1-ribose-5'-phosphate (SAICAR) as depicted in Figure (13). In *M. tuberculosis* SAICARs (MtbSAICAR) is monofunctional while in humans, the sixth and seventh steps are combined. The gene coding for single polypeptide chain includes bifunctional enzymes incorporating purE classII (AIRc) and HsSAICAR of aminoimidazole ribonucleotide carboxylase-aminoribosyl aminoimidazole succinocarboxamide ribonucleotide synthetase (PAICS; AIRc-HsSAICAR) (Mathews et al., 1999). Further discussion of MtbSAICAR and other SAICARs enzymes is given in Section 1.3.3.1.

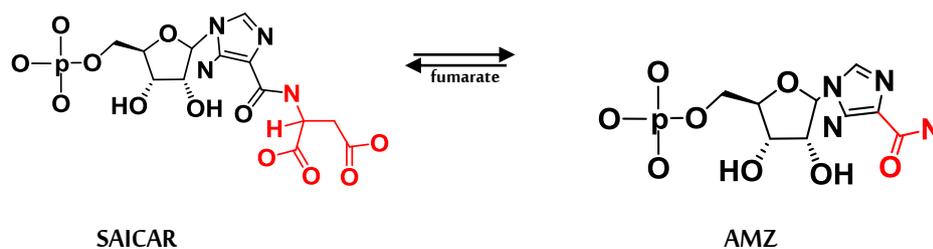


**Figure (13)** The biochemical reactions of SAICAR Synthetase (SAICARs).

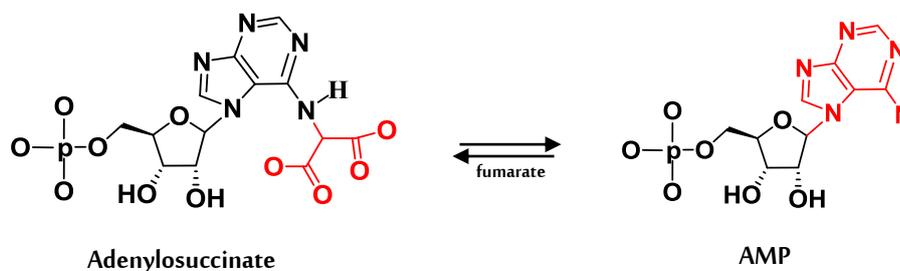
### 1.3.2.8. Adenylosuccinate lyase (purB)

The eighth step is catalysed by adenylosuccinate lyase (purB) (EC: 4.3.2.2) which involves two activities. One activity contributes by producing 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR; AMZ) (IMP biosynthesis pathway) and the other activity is a part of (AMP biosynthesis pathway) (Figure 14A and B) respectively (Ducati et al., 2011).

(A)



(B)



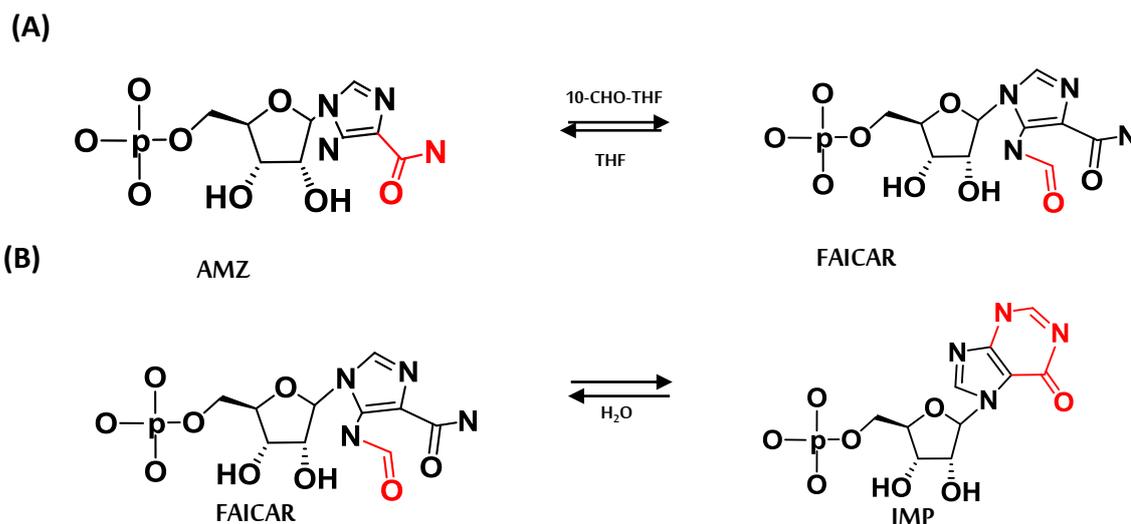
**Figure (14)** The biochemical reactions of Adenylosuccinate lyase (purB), reaction involved in AMZ (A) and AMP biosynthesis (B).

There is a reported attempt in determining the structure of purB of *M. tuberculosis* but the crystals did not diffract. The *M. tuberculosis* (MtbpurB) structure has been modeled using the *Mycobacterium smegmatis* MspurB structure as a template because of the high sequence identity (84.5%). The modelled tetrameric structure shares an identical fold with other Mycobacterial organisms and is highly conserved to the human than other bacterial enzymes (Banerjee et al., 2014). The crystal structure

of adenylosuccinate lyase (purB) from *E. coli* has also been solved and is seen to be a homotetramer. Each subunit is composed of N-terminal, middle and C-terminal domains. The enzyme has been solved in both apo and as an adenylosuccinate (SAMP) complex. Upon ligand binding, the catalytic loop becomes more ordered and closes onto the active site. The catalytic loop poses an open conformation to release fumarate then AMP products (Tsai et al., 2007; Kozlov et al., 2009). The loop in human, *E. coli* and mycobacterial enzymes (MspurB and MtbpurB) are found to have highly-charged, moderately charged and less charged residues respectively. This loop is also seen to be disordered in apo form and becomes ordered upon ligand binding in the human as well. Whereas in the *Mycobacterial* enzymes apo form, it is described as slightly ordered (Banerjee et al., 2014).

#### **1.3.2.9. AICAR transformylase /IMP cyclohydrolase (ATIC)**

AICAR transformylase /IMP cyclohydrolase (ATIC), purH (EC: 2.1.2.3) is a bifunctional enzyme that catalyses the ninth and tenth steps of the Inosine-5'-phosphate biosynthesis pathway. This enzyme is a folate dependent enzyme (like purN, the third enzyme in the pathway). In the first reaction of purH, it uses  $N^{10}$ -formyltetrahydrofolate to transfer a formyl group to 5-aminoimidazole-4-carboxamide ribonucleotide (AMZ). The resulting product, 5-formylaminoimidazole-4-carboxamide ribonucleotide (FAICAR), then forms inosine 5'-monophosphate (IMP) by losing a water molecule. In eukaryotic organisms, the structure of the purH enzyme is seen to be a dimer with two domains per monomer as is the *M. tuberculosis* enzyme. The C-terminal domain is involved in FAICAR formation (Figure 15A), while in the N-terminal domain FAICAR is cyclised to form the purine nucleotide IMP (Figure 15B) (Le Nours et al., 2011).



**Figure (15)** The biochemical reactions of AICAR transformylase /IMP cyclohydrolase (ATIC), FAICAR (A) and IMP formation (B).

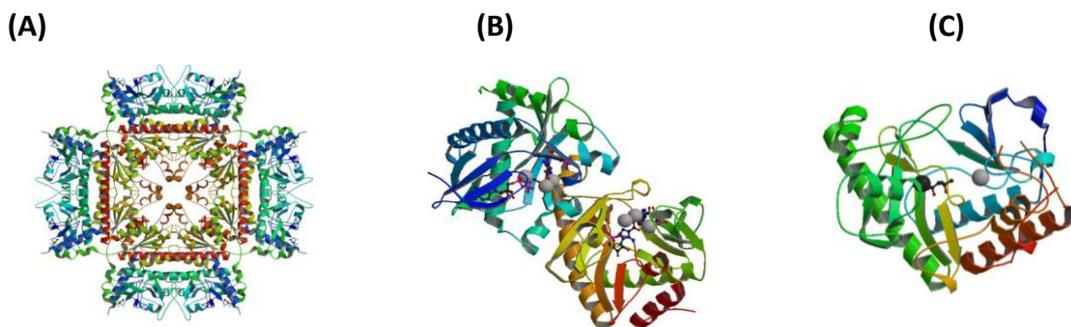
### 1.3.3. SAICAR Synthetase and Phosphoribosylamine glycine ligase of *Mycobacterium tuberculosis*.

The previous sections outlined the *de novo* pathway of *M. tuberculosis* synthesis of the purine IMP and outlines each biochemical reaction briefly. Phosphoribosylamine glycine ligase (purDs) and SAICAR Synthetase (SAICARs) are the second and seventh steps of this pathway respectively. They are two essential enzymes of purine biosynthesis pathway and the genes coding these enzymes have been shown to be fundamental for growth and survival of this microorganism (Sasseti, Boyd and Rubin, 2003).

This is demonstrated in *in vitro* and *in vivo* studies in which mutating the gene encoding MtbSAICAR disrupts the virulent nature of *M. tuberculosis* strain and makes *in vitro* growth dependant on the addition of a purine precursor. In addition, *in vivo* the mutated pathogen is eliminated when injected intravenously into mice, in contrast to the wild type, which maintains its virulence (Jackson et al., 1999). Another study using human serum showed that a strain with a defect in the essential genes *MtbSAICAR* and *MtbpurD* (Samant et al., 2008) would not grow. Both of these are ligases whereby two molecules are joined by carbon-nitrogen bonds where ATP is utilised (Galperin and Koonin, 2008). The following sections will describe the structure and function of both enzymes in detail.

### 1.3.3.1. SAICAR Synthetase (SAICARs)

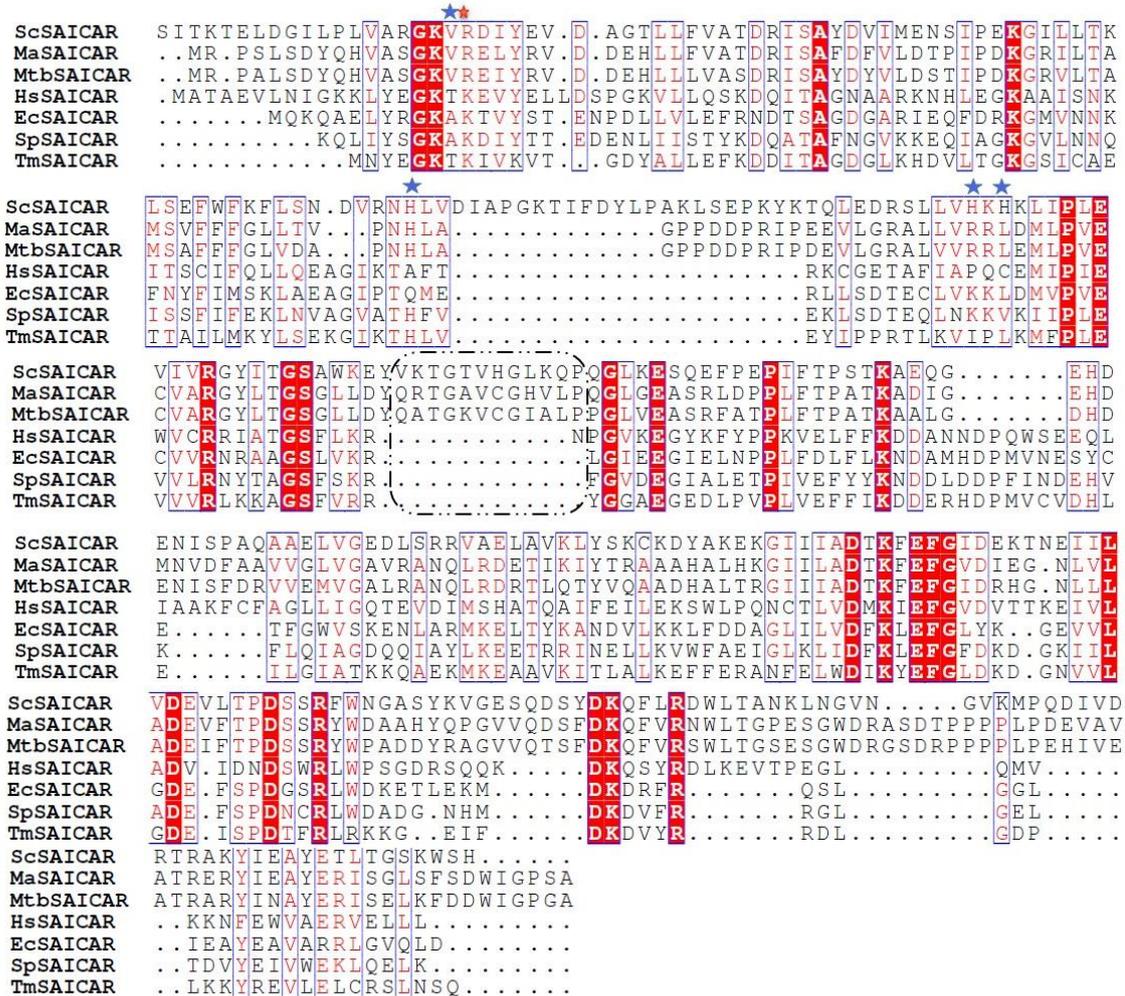
A Basic Local Alignment Search Tool (BLAST) (Altschul, 1990) revealed that SAICARs of *M. tuberculosis* (MtbSAICAR) has 76%-20% sequence identity with the SAICARs of *Mycobacterium abscessus* (MaSAICAR), *Saccharomyces cerevisiae* (ScSAICAR), *Thermotoga maritima* (TmSAICAR), *Streptococcus pneumoniae* (SpSAICAR), *E. coli* (EcSAICAR) and human (HsSAICAR). Only a few crystal structures of SAICARs have been solved to date. The crystal structure and oligomeric organisation of SAICARs varies from organism to organism. For example, the human enzyme (HsSAICAR) is a homooctamer (Li et al., 2007), *E. coli* (EcSAICAR) is a homo dimer (Ginder et al., 2006) and *M. abscessus* (MaSAICAR) is a monomer (Figure 16). The details of each structure will be discussed and compared.



**Figure (16)** Oligomeric organisation of SAICARs from three different organisms. (A) Human homooctamer structure of PAICS (PDB 2H31) coloured as a rainbow. (B) EcSAICAR (PDB 2GQS) a homo dimer coloured as a rainbow from N to C- terminals and assembled with CAIR,  $Mg^{2+}$  and ADP. (C) A monomer of MaSAICAR (PDB 3R9R) all are coloured as N to C-terminals rainbow and shown with  $Mg^{2+}$ ,  $Cl^-$  and (4R)-2-Methylpentane-2, 4-DIOL (MRD).

Alignments of SAICARs amino acid sequences revealed that the residues identified from crystal structures that form the binding sites for the substrate analogues 5-aminoimidazole-4-carboxamide ribonucleotide (AMZ) and inosine 5'-monophosphate (IMP), the competitive inhibitor of CAIR are mostly retained in the SAICARs of all the organisms compared (Figure 17). In the ATP binding site, the hydrophobic V16 of MtbSAICAR replaces the polar T7 and T17 in SAICARs from human (HsSAICAR) and *T. maritima* (TmSAICAR) respectively. H69 replaces Q69 in *E. coli* (EcSAICAR) and the small hydrophobic residue A75 in HsSAICAR. The hydrophobic residue L93 replaces the polar H112 in *S. cerevisiae* (ScSAICAR) and is C90 in HsSAICAR. The basic residue R91 is the most variable as it is changed variously to polar, charged residues, being H110 in

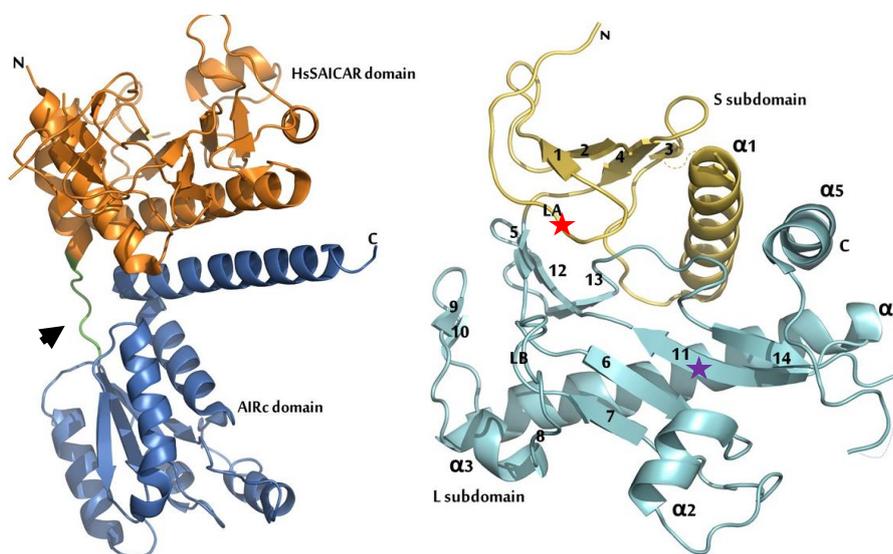
ScSAICAR, K82 in EcSAICAR, and the hydrophobic, I75 in TmSAICAR and P88 in HsSAICAR. The sequence alignments are discussed further in the context of structure alignments in Chapter 5 (Figure 82-85-88-91 and94).



**Figure (17)** Sequence alignment and comparison of MtbSAICAR with homologues, the notable differences in residues in the binding sites (identified in the main text) are labeled with blue stars whilst similar residues are labeled with red stars. The red boxes with white characters indicate strict identity and the blue frames indicate similarity across the group. Residues expected to prevent dimerisation are indicated by a dashed box. The similarity between HsSAICAR and MtbSAICAR is 40%. Sequence alignment is done using MAFFT (Kato and Standley, 2013), which processed using BLOSUM62 matrix then alignment drawn using ESript3 (Robert and Gouet, 2014). All sequences extracted from Protein Data Bank (PDB) and missing residues added using Uniprot (Apweiler, 2004).

In humans, 4-(N-succinylcarboxamide)-5-aminoimidazole ribonucleotide synthetase (HsSAICAR) is a part of the bifunctional enzyme phosphoribosylaminoimidazole carboxylase/phosphoribosylaminoimidazole succinocarboxamide synthetase (PAICS; AIRc, HsSAICAR), i.e. the 6th and 7th steps of the pathway. It requires another enzyme,

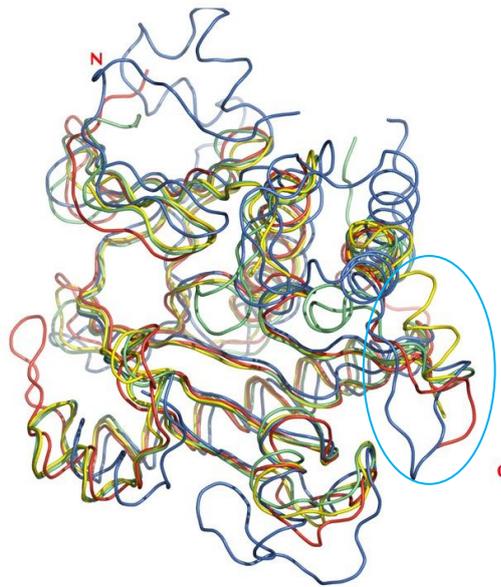
(purE classII), 5-aminoimidazole ribonucleotide carboxylase (AIRc) to complete its action. purE classII (AIRc) produces CAIR, the substrate of HsSAICAR (Li et al., 2007). The structure is octameric with each monomer consisting of two domains. AIRc activity is the C-terminal and located in the core while the HsSAICAR function is the N-terminal domains, forming four dimers surrounded the octameric AIRc. The domains are connected by a short linker of 6 residues (K261-C266), as shown in Figure (18). HsSAICAR is divided into two subdomains named small (S) (L7-Q90) and large (L) (C91-L260) subdomains as shown in Figure (18). PAICS is a multi-catalytic enzyme where the intermediate is channeled between the two active sites. Substrates are efficiently transferred from one active site to the other via a tunnel termed the SAICAR-path. This is made up of the AIRc-alley and SAICAR-hall.



**Figure (18)** Structural of a monomer phosphoribosylaminoimidazole carboxylase/ phosphoribosylaminoimidazole succinocarboxamide synthetase (PAICS). Left the monomer structure of PAICS consists of two domains, AIRc is the C-terminal and located in the core while SAICARs is the N-terminal. Right HsSAICAR (PDB 2H31) is divided into two subdomains named S (L7-Q90) and L (C91-L260) subdomains connected by a linker consisting of 6 residues (K261-C266) coloured light green and indicated by black arrow (Li et al., 2007). The expected ATP and CAIR binding sites are indicated by red and purple stars respectively.

Three subunits of AIRc domain contribute to the formation of the active site where the AIRc-alley is located. Four subunits (two subunits of each AIRc and HsSAICAR) contribute in transferring the intermediate, which is located between AIRc-alley and SAICAR-hall. The third part consists of a dimer of HsSAICAR, which contributes to the

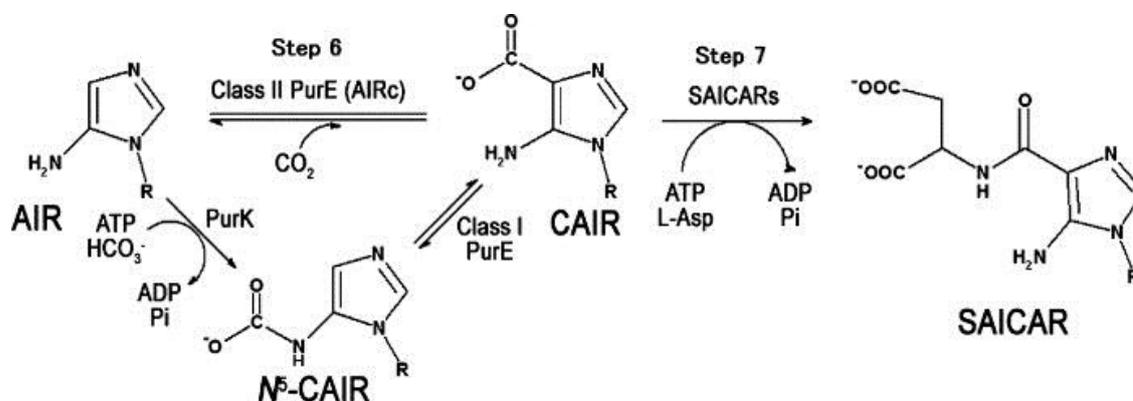
SAICAR-path. The SAICAR-path has several residues that are highly conserved among PAICS enzymes; this demonstrates the importance of the SAICAR-path. Charged residues regulate the transfer of CAIR to the binding site: two groups of negatively charged residues distributed equally in both sides of the SAICAR-path to the CAIR binding site ensure transfer of CAIR. The phosphate group of CAIR is seen to have an important job in facilitating the CAIR channeling. Overall, there are four tunnels connected both enzymes distributed equally and the significance of this tunnel is to transport of CAIR from one active site to the other. Comparison between the structures of (AIRc-HsSAICAR), a multi-functional enzyme, and other SAICARs enzymes show that they are highly similar in their fold but around the C- terminal of the subdomain there is some variation and this may be because of its connection to AIRC domain (Figure 19).



**Figure (19)** Structural comparison of SAICARs among different organisms, HsSAICAR domain (PDB 2H31, red), EcSAICAR (PDB 2GQS, green), TmSAICAR, (PDB 1KUT, yellow), and ScSAICAR (PDB 1A48, blue). Different regions around C-terminal are circled (the area where purE classII of human is bound) (Li et al., 2007).

A comparison of (AIRc-HsSAICAR) with the monofunctional enzymes such as EcSAICAR shows that the latter does not appear to have any sign of tunnel channeling. Although EcSAICAR is homologous to the HsSAICAR (with 32.1% sequence identity), the mechanism of producing CAIR substrate, which is involved in the reaction, is different from that of in EcSAICAR. purE class1 needs the product of another enzyme, N5-

carboxyaminoimidazole ribonucleotide synthase (purK), (*N*<sup>5</sup>-CAIR) to form CAIR (Li et al., 2007) whilst in human (AIRc-HsSAICAR) converts AIR to CAIR using additional CO<sub>2</sub> in the reaction as described in Figure (20).



**Figure (20)** Different reactions involved in synthesis of the substrate for SAICARs enzymes in *E. coli* and human. Humans for step 6 of the pathway. While *E. coli* requires the product of purK (*N*<sup>5</sup>-CAIR), R indicates ribose 5-phosphate. (Li et al., 2007).

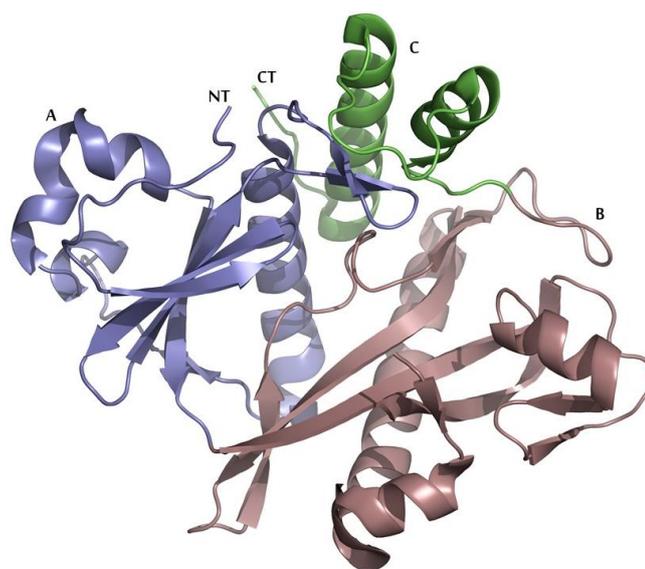
In *E. coli*, EcSAICAR catalyses the reaction of the eighth step of the purine biosynthesis pathway. The crystal structure of the enzyme was solved in both the presence of ADP, and in the ADP+CAIR complex. The enzyme is a dimer similar to TmSAICAR and has a highly similar fold with RMSD C $\alpha$  of around 1.2Å. The structure has two domains in each subunit, domain A fuses to domain B and forms a cleft where ADP Mg<sup>2+</sup> binds in one site while CAIR and two Mg<sup>2+</sup> are found on another cleft site.

ScSAICAR (PDB 1A48) (Levdikov et al., 1998) shares 26% sequence identity with EcSAICAR and RMSD C $\alpha$  value of 2.19Å. The oligomeric state of the protein is monomeric. It may form this monomer because of the presence of an additional 13 residues that are expected to have a role in preventing dimerisation as explained in Chapter5, Section 5.3.1 (Zhang et al., 2006). Comparison of EcSAICAR and ScSAICAR structures shows the presence of a few insertions at residues 1, 77,105 and 221, the third insertion at residue 105 may have an important job in preventing the dimerisation of ScSAICAR (Ginder et al., 2006).

Two different mechanisms have been suggested Ginder et al., 2006 described a mechanism for the enzyme by which ATP directly acts as a phosphorylation agent to CAIR, and then aspartate acts as the amino group donor. By contrast, Wolf et al., 2014

noted the separation of substrates and suggested the involvement of some other amino acid residues contributing to the biochemical reaction. They proposed that the phosphate group is transferred from the side chain E178 onto E89 then it is transferred onto the carboxylate group of CAIR. The amino group of aspartate then becomes free and attaches to CAIR in order to form the product SAICAR.

In the crystal structure of ScSAICAR from *S. cerevisiae*, the enzyme is packed as a monomer, it has been solved in the presence of the ligands ATP, ADP-AMZ and succinate as well as in the apo form (Levdikov et al., 1998; Urusova et al., 2003). The monomer is composed of three domains named A, B, C. Domain A consists of 111 residues (S2-H112), domain B (L113-Q255) and domain C (D256- H306). Both A and B domains are composed of five  $\beta$ -strands with peripheral  $\alpha$ -helices while domain C consists of two  $\alpha$ -helices as shown in Figure (21).



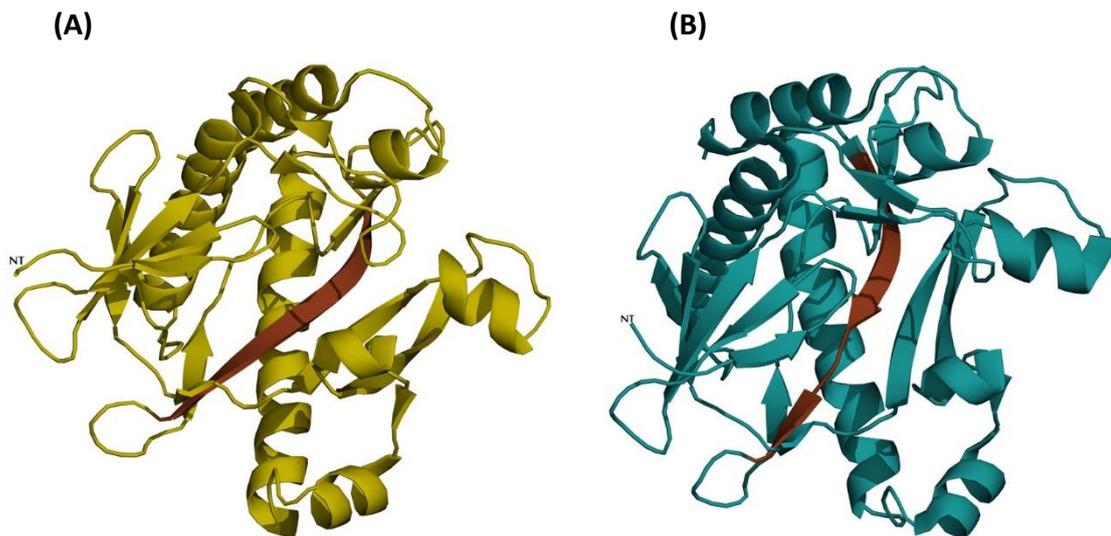
**Figure (21)** Domains classification of ScSAICAR structure (PDB 1A48), domain A (S2-H112), domain B (L113-Q255) and domain C (D256-H306), coloured Violet, light brown and green respectively. NT represents the N-terminal and CT represent the C-terminal.

The active site of ScSAICAR is located in a wide cleft between these domains where the natural substrate is supposed to bind. The comparison between ScSAICAR ATP complex (PDB 1OBD) and ScSAICAR apo (PDB 1A48) revealed that the position of  $\beta$ -phosphate of ATP is located in the same binding site to that of a sulfate ion. The structure also provides some details about the position of ATP and their relative

atoms. ATP lies above phosphate binding loop residues 14-24; the adenine base is in the anti-conformation and it is located in a hydrophobic pocket, but also make H-bonds with surrounding residues.

In addition, the ribose ring and  $\beta$ -phosphate also make some interactions with the enzyme (Levdikov et al., 1998). The crystal structure of ScSAICAR has also been solved in complex with products (ADP), substrate and analogues (AMZ, and succinic acid) (PDB 2CNQ) (Urusova et al., 2003). AMZ is similar to CAIR but has an amino group where CAIR has a carboxyl group. AMZ was observed to have two positions on the structure. One of them is located at the active site where the enzymatic reaction occurs and the other is located in a peripheral site of the protein. Similar to those of succinic acid, an analogue of the substrate aspartate, also has two binding sites in the structure. The presence of ADP was also clearly observed in the crystal structure with two positions, one at the active site while the other is peripheral (Urusova et al., 2003).

In the case of *S. pneumoniae* (SpSAICAR), the enzyme catalysis is the same as in MtbSAICAR in which CAIR and aspartate react to produce SAICAR with additional help from ATP that acts as phosphorylation donor (Wolf et al., 2014). The enzyme is a homodimer. The enzyme was crystallised in the presence of ADP, aspartate,  $Mg^{2+}$  and CAIR (structure1) or ADP and  $Mg^{2+}$  (structure2). The crystal structure of the enzyme was solved in the presence of AIR instead of CAIR this is may be due to crystallisation in an acidic solution. The differences between the two structures (Figure 22), have been described and there is a small difference in secondary structure between residues 171-181 of structure1 solved with ADP, aspartate,  $Mg^{2+}$  and AIR whereby a long  $\beta$ -strand whilst in the presence of ADP and  $Mg^{2+}$  it divides into two  $\beta$ -strands connected by a loop. This is because of the movement of K176 due to the absence of aspartate in this structure. (Wolf et al., 2014).



**Figure (22)** Crystal structures of *S. pneumoniae* (SpSAICAR). (A) SpSAICAR (PDB 4FE2). (B) SpSAICAR (PDB 2CNQ). The secondary structure of segment 171-181 is coloured brown.

The crystal structure of *T. maritima* (TmSAICAR) has been solved to 2.2Å. It shares 26% sequence identity and has the same overall fold as *S. cerevisiae* (ScSAICAR). Unlike ScSAICAR, the crystal structure of TmSAICAR is a dimer both in solution and the crystal where N and C-terminal of the monomers face each other. These monomers are covalently linked to each other through disulfide bond formation between C126 of chain A and C126 of chain B. These two residues are responsible for maintaining the dimeric state (Figure 23). The monomers have the same topology described as three subdomains; the first two are composed of  $\alpha+\beta$ , which poses ATP binding site and third is composed of  $\alpha$ -helix C-terminal. Both monomers are also connected by a salt bridge formed between D126 of chain A and K145 of chain B and vice versa 8-9Å away from the disulfide bridge (Zhang et al., 2006).

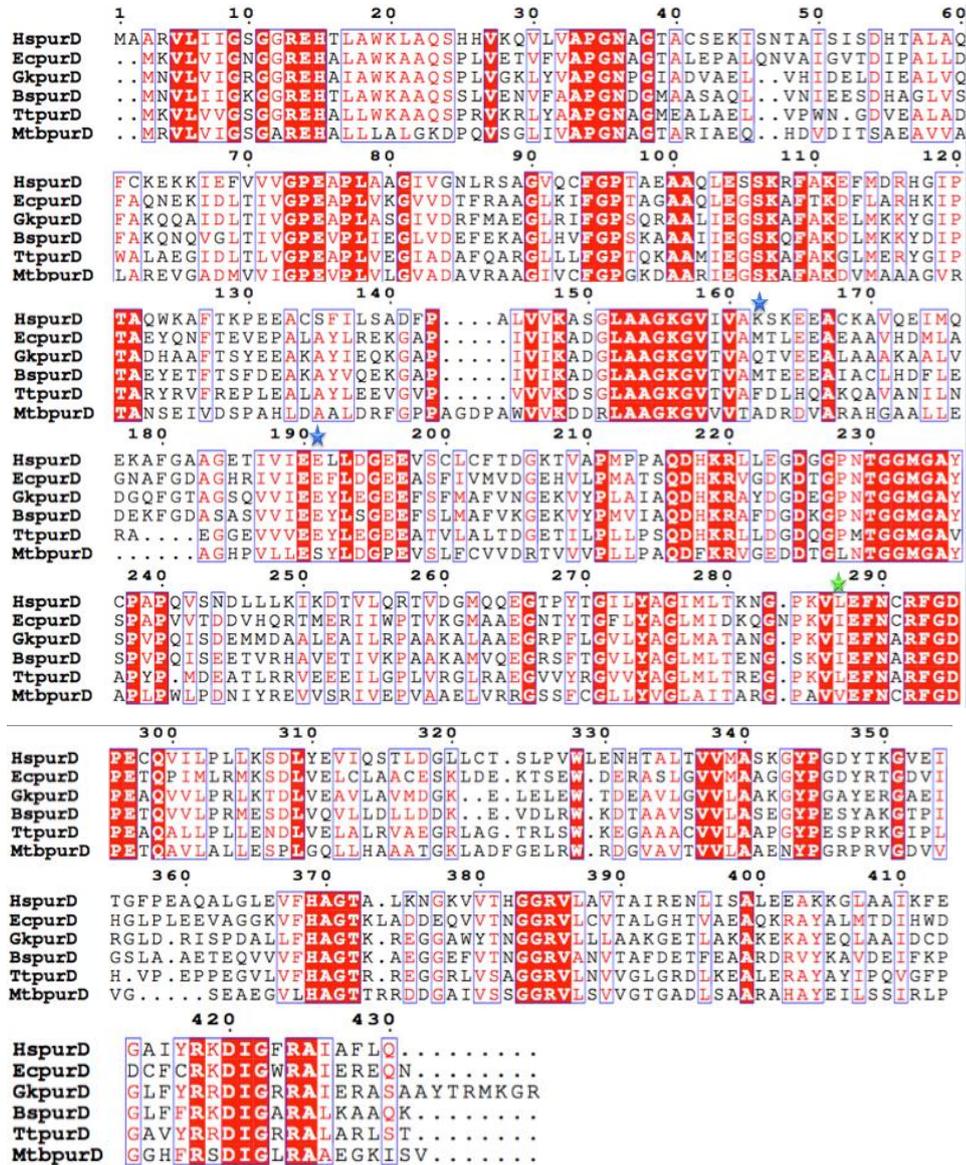


**Figure (23)** Dimerisation scheme of TmSAICAR, the chains coloured red and blue. Both chains are connected by a disulfide bond formed between C126 of chain A and C126 of chain B that is indicated by yellow stick (Zhang et al., 2006).

### 1.3.3.2. Phosphoribosylamine glycine ligase (purDs)

Phosphoribosylamine glycine ligase of *M. tuberculosis* (MtbpurD) (EC: 6.3.4.13) is encoded by an essential gene that is reported to affect the bacterial growth once mutated (Samant et al., 2008). In bacteria, purDs is a monofunctional enzyme whereas, in humans, it is a part of the trifunctional human enzyme (HsGART). The catalytic activity of the enzyme uses ATP as phosphorylation agent (Wang et al., 1998; Fawaz, Topper and Firestine, 2011). The proposed chemical reaction of purDs suggests the carboxyl group of glycine is phosphorylated by the  $\gamma$ -phosphate of ATP, then the amino group of PRA is attached to a glycine carboxyl group (Zhang, Morar and Ealick, 2008). In the absence of a crystal structure of MtbpurD, the multiple sequence alignment approach was utilised using MAFFT (Kato and Standley, 2013). Initially, the alignments were analyzed to compare the residues involved in the ATP and Glycine binding in the purDs of *T. thermophilus* (TtpurD), Human (HspurD), *B. subtilis* (BspurD) and *E. coli* (EcpurD) and *G. kaustophilus* (GkpurD). The structural of HspurD (PDB 2QK4) (Welin et al., 2010), was used as a reference for this comparison. Most of the residues involved in ATP+Glycine binding appear to be conserved, with some minor differences in the ATP

binding sites. E191 of HspurD is replaced by S185 in MtbpurD. K162, which makes an ionic interaction with the phosphate of ATP in HspurD, is replaced by a small hydrophobic residue A162 in MtbpurD. The structure of an homology model of MtbpurD is discussed in Chapter 6, Section 6.6.



**Figure (24)** Sequence alignment of purD from different organisms. purD of each organism is represented as (MtbpurD) for *M. tuberculosis*, *T. thermophilus* (TtpurD) (PDB 2IP4), Human (HspurD) (PDB 2QK4), *B. subtilis* (BspurD) (PDB 2XD4) and *E. coli* (EcpurD) (PDB 1GSO), *G. kaustophilus* (GkpurD) (PDB 2YS6). The sequence identity and similarity between HspurD and MtbpurD are 39% and 56% respectively. Sequence alignment is done using MAFFT (Katoh and Standley, 2013), using the BLOSUM62 matrix and the alignment drawn using ESript3 (Robert and Gouet, 2014). All sequences extracted from the Protein Data Bank (PDB) and missing residues added using Uniport (Apweiler, 2004). The red boxes with white characters indicate strict identity and blue frames indicate similarity. Blue stars identify differences between HspurD and MtbpurD noted in the text, and a green star identifies a conservatively changed residue.

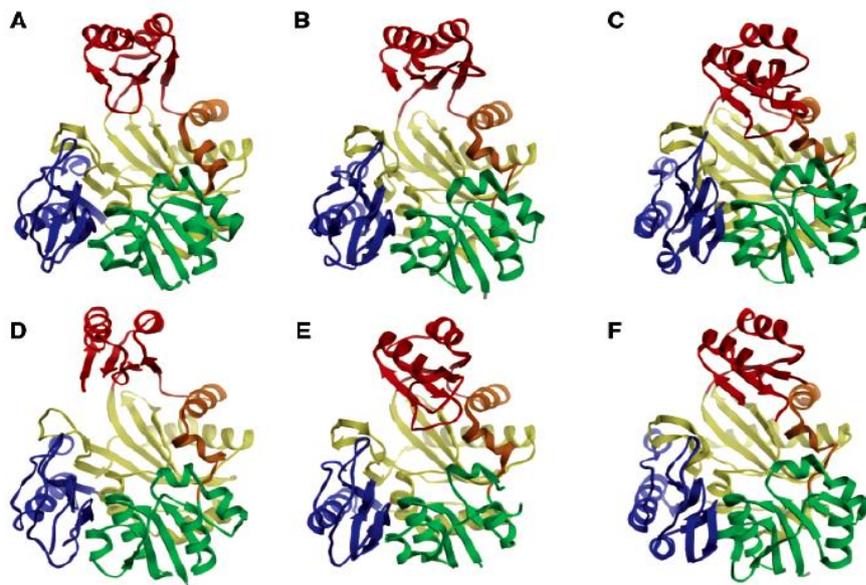
The trifunctional human enzyme (HsGART) consists of a single polypeptide chain with three enzyme activities. Each of these enzymes catalyses a specific step involved in inosine-5-phosphate biosynthesis via a *de novo* pathway. Phosphoribosylamine glycine ligase (purDs) (EC: 6.3.4.13) takes the role of the second step activity which is located at the N-terminal region of HsGART and purN, the C-terminal glycinamide ribonucleotide transformylase (GARTfase; purN) (EC: 2.1.2.2.) which takes the role of the third step, and 5-aminoimidazole ribonucleotide (AIR) synthetase (purM) (EC: 6.3.3.1) which catalyses the fifth step and is placed at the mid of HsGART domains (Welin et al., 2010).

The crystal structure of the human purD (HspurD) (PDB 2QK4) consists of four domains and the ATP binding site (located between A and B domains), it has the same topology as EcpurD (below). The adenine group of ATP forms H-bonds with charged and hydrophobic residues, ribose forms H-bond with charged residue and phosphate group binds to charge polar residues. Unlike EcpurD, which is a monomer, HspurD is a dimer consisting of two monomers that are homologous to EcpurD, sharing a high degree of sequence identity of about 51% and a positional RMSD C $\alpha$  of 1.7Å.

The crystal structure of *E. coli* (EcpurD) (PDB 1GSO) is a monomer and consists of four domains (N, A, C and B). A and B domains contain most of the amino acid residues involved in the ATP binding site while the N and C domain are responsible for the substrate binding site (Wang et al., 1998). In EcpurD, the domains have been described as follows, domain N (1-118), B domain (119-191) is connected to N domain by residues (114-122) and A domain residues (188-192). Both of these segments provide some flexibility of the B domain in relation to other domains. The A domain (192-329) forms a cleft with the B domains, similar to many ATP binding enzymes. The C domain (330-426) is part of the core of the enzyme and is the smallest domain and seems absent in some of ATP dependent enzymes but other domains (N, A and B) are highly conserved (Wang et al., 1998).

The crystal structure of *Thermus thermophilus* (TtPurD) (PDB 2IP4) is composed of two monomers; the structure was solved in the apo form only. The apo form of TtpurD shows a sulfate ion located at the active site center, this is interpreted as the position of a phosphate group when the enzyme is liganded (Sampei et al., 2010). The crystal structure of *Geobacillus kaustophilus* (GkPurD) (PDB 2YS6) (composed of one chain) has been solved in the presence of three ligands,  $\text{PO}_4^{3-}$ , AMP/  $\text{PO}_4^{3-}$  and AMP/glycine. The AMP binding site is positioned an equivalent position to the ADP of EcPurD complex. Although ATP was included in crystallisation conditions, only AMP was seen in both solved structures. This is explained by ATP having disordered  $\beta$  and  $\gamma$ -phosphate groups, rather than by hydrolysis (Sampei et al., 2010). The crystal structure of the *Aquifex Aeolicus* (AaPurD) (PDB 2YW2) has also been solved in two forms, the apo and the ATP-liganded form.

The structures of purDs are well conserved; the sequence identity between TtpurD, GkpurD, AapurD, and EcpurD with HspurD is around 44%-50% (Sampei et al., 2010). In addition, these purDs enzymes share the same four domains as EcpurD identified as N, A, C in the center of the enzyme while domain B locates externally and found to be variable among purDs enzymes. For example, the B domain is seen in the open form in EcpurD (1GSO), TtpurD (2IP4) and GkpurD and in closed form in the structure of *T. maritime* (TmPurD) (PDB 1VKZ), HspurD (2QK4) and AapurD (2YW2) as given in Figure (25) (Sampei et al., 2010).



**Figure (25)** 3D secondary structures of phosphoribosylamine glycine ligase in different organisms. (A) *T. thermophilus* (TtpurD) (PDB 2IP4), (B) *G. kaustophilus* (GkpurD) (PDB 2YS6), (C) *A. Aeolicus* (AapurD) (PDB 2YW2), (D) *E. coli* (EcpurD) (PDB 1GSO), (E) *T. maritime* TmpurD (PDB 1VKZ) and (F) HspurD (PDB 2QK4). purDs in all structure consist of four domains N, B0, B, A and C coloured green, orange, red, yellow and blue respectively (Sampei et al., 2010).

#### 1.3.4. Aims of the study

The previous sections describe the general characteristics and pathogenicity of *M. tuberculosis*. An overview was presented of the purine biosynthesis pathway of this pathogen and compared with other organisms. The purine biosynthesis pathway of *M. tuberculosis* was addressed with the emphasis on data derived from genomic and experimental analysis, which revealed that there are some differences in purine biosynthesis pathways in different organisms.

The main focus of this PhD is the investigation of enzymes in the purine biosynthesis pathway with the aim of increasing the understanding of the pathway. The particular emphasis of the work described in this thesis is given to understanding the structure and function of *M. tuberculosis* SAICAR synthase (MtbSAICAR), this was mainly done by solving the X-ray crystal structure of apo MtbSAICAR and its complexes with various ligands including the substrate analogue AICAR (AMZ) and the competitive inhibitor IMP. A second topic of the work was to investigate the structure of MtbpurD and to characterise it using biophysical methods including circular dichroism, thermal shift assays, crystallisation trials and computational studies.

## Chapter 2. Materials and Methods

### 2.1 Materials

#### 2.1.1. Chemical and Reagents

All reagents used in that study for solution experiments and crystal preparation were purchased from Sigma unless otherwise indicated. All buffers and solutions were prepared using double distilled water. KOD Hot start DNA polymerase and Rosetta (DE3) were purchased from Merck Millipore. Enzyme assays used the Biomol Green kit from Enzo Life Sciences (Enzolifesciences.com, 2016).

#### 2.1.2. Chromatographic Media and Membranes

Ni-NTA Agarose was purchased from Qiagen™ and Superdex 200 10/300 GL column from GE Healthcare. Concentrator tubes of 15ml and 500µl with cutoff 10 kDa was from Millipore. Dialysis membrane was from Scientific Laboratory Suppliers.

#### 2.1.3. Crystallisation supplementary

JCSG+ crystallisation screen, 24 well sitting-drop plates, 96 well plates, Magnetic caps, vials, and Litho™ loops (0.1-0.3mm) were purchased from Molecular Dimensions (UK). Tape used for sealing the crystallisation plates was purchased Hampton Research Ltd, California, USA.

### 2.2. Molecular biology techniques

#### 2.2.1. Primer design

The sequences of the genes of the *M. Tuberculosis* H37RV purine biosynthesis pathway were retrieved from the TB Database (Galagan et al., 2010). The primers were designed with the addition of a homology region of the vector to the 5' end of both forward and reverse primers, based on the instructions given by PROtein EXpression Laboratory (PROTEX) at University of Leicester. Primers were ordered from Eurofins MWG Operon in Germany. The primers are given in Table (1)

**Table (1)** Primer used throughout this study.

Genes Name 5'→ 3'	N-terminal Homology Region of Vector	Primers
<b>purC-F1</b>	TACTTCCAATCCATG	ATGCGCCCCGCATTGTCCGACT
<b>purC-R1</b>	TATCCACCTTTACTGTCA	TCATGCGCCAGGGCCGATCCA
<b>purD-F1</b>	TACTTCCAATCCATG	GTGCGCGTCCTGGTAATCGGT
<b>purD-R1</b>	TATCCACCTTTACTGTCA	CTAGACGCTGATCTTCCCCT

### 2.2.2. PCR (ORF Amplification)

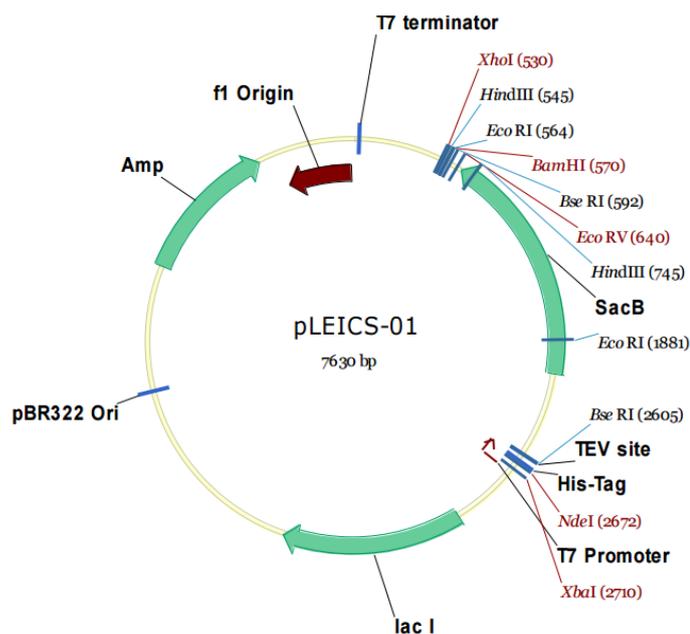
In order to amplify the genes for constructing the expression systems in pLEICS vectors, KOD Hot start DNA polymerase (Novagen) was used. Generally, 20µl reaction volumes were used with a concentration of 1x reaction buffer containing 1.5mM Mg<sub>SO4</sub>; 0.2mM dNTP; 0.3mM primers; DNA template (3-12ng chromosomal DNA). The reaction mixture made up to the required volume with autoclaved ddH<sub>2</sub>O. DNA amplification was carried out in a thermal cycler (Eppendorf Scientific Support Inc., UK). The extension time and annealing temperature were varied depending on the primers used, see (Table 2). The genetic material for *M. tuberculosis* H37RV was supplied by Dr. Galina Mukamolova University of Leicester.

**Table (2)** Cycling parameters in ORF amplification of *MtbpurC* and *MtbpurD*.

Gene	Polymerase activation	Denaturation	Annealing	Elongation
<i>MtbpurC</i>	94°C- 2 minutes	94°C-15 seconds	60°C-60 seconds	68°C -60 seconds
<i>MtbpurD</i>	94°C- 2 minutes	94°C-15 seconds	57°C- 60 seconds	68°C -30 seconds

### 2.2.3. Gene cloning and sequencing

Advances in recombinant DNA allowed the production of large amounts of protein that have in turn been used to drive advances in structural biology, increasing our understanding the molecular structure and function of specific genes and gene products. The general process of cloning consists of several steps, starting with introducing the gene of interest after amplification into a vector of choice, which is then replicated inside a host. To enable this, PCR products were cloned into the pLEICS-01 vector (Figure 26) from the Protein Expression Laboratory (Protex), and then the success of cloning was confirmed by sequencing using the Protein Nucleic Acid Chemistry Laboratory (PNAACL) at University of Leicester.



**Figure (26)** Vector map details of PLEICS-01 used in cloning of PCR products

## 2.2.4. Protein expression and purification

### 2.2.4.1. Transformation

The vector containing the insert was transformed into *E. coli* expression vector Rosetta (DE3) (Invitrogen). This process involved several steps. First, the chemically competent Rosetta (DE3) cells were mixed with 1µl of plasmid DNA on ice. After that, the tube was kept on ice for 30 minutes and heat shocked for 45 seconds at 42°C then kept on ice for 2 minutes. Subsequently, 100µl of 2x Yeast extract Tryptone media (2xTY) (16g tryptone, 5g NaCl and 10g of yeast extract per liter) was added to the microcentrifuge tube and incubated for one hour at 37°C. The incubated culture was spread onto 2xTY agar plates supplemented with 100µg/ml ampicillin and 34µg/ml chloramphenicol and kept at 37°C overnight.

### 2.2.4.2. Overexpression trials

Several colonies were tested for overexpression by inoculation into 10ml of 2xTY media. The media was supplemented with 100µg/ml of ampicillin and 34µg/ml chloramphenicol and incubated in a shaking incubator at 37°C overnight. 150µl of the overnight culture was added into 10ml of 2xTY media supplemented with 100µg/ml of ampicillin and 34µg/ml chloramphenicol. This culture was grown at 37°C to maximise

the cell growth. Once the optical density at 600nm (OD<sub>600</sub>) reached between 0.5-0.7, IPTG was added using the times and temperatures shown in Table (3 and 4). After the incubation periods, all tubes were centrifuged at 4000 rpm for 15 minutes at 4°C. The pellet of each cell was suspended into 750µl of lysis buffer (50mM HEPES pH7.5, 500mM NaCl, 1mM DTT and 1 complete protease inhibitor (Roche-Germany)). The cell suspension was lysed by sonication using a Soniprep 150 for 10 seconds on, 30 seconds off, three times whilst being kept over ice. The soluble fraction was separated by centrifugation at 14000 rpm for 15 minutes at 4°C. The soluble and insoluble fractions were collected for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

**Table (3)** MtbSAICAR IPTG Induction conditions varying incubation time and temperature

<b>IPTG concentrations</b>	<b>Temperatures</b>	<b>Time of incubation</b>
<b>400µM</b>	37°C	2, 3, 4 hours and overnight
<b>150µM</b>	20°C	Overnight
<b>—</b>	37°C	Overnight
<b>50µM</b>	20°C	Overnight
<b>150µM</b>	20°C	Overnight
<b>250µM</b>	20°C	Overnight
<b>—</b>	20°C	Overnight

**Table (4)** MtbpurD IPTG Induction conditions varying incubation time and temperature

<b>IPTG concentrations</b>	<b>Temperatures</b>	<b>Time of incubation</b>
<b>50µM</b>	20°C	Overnight
<b>150µM</b>	20°C	Overnight
<b>250µM</b>	20°C	Overnight
<b>—</b>	20°C	Overnight
<b>400µM</b>	37°C	Overnight
<b>—</b>	37°C	Overnight

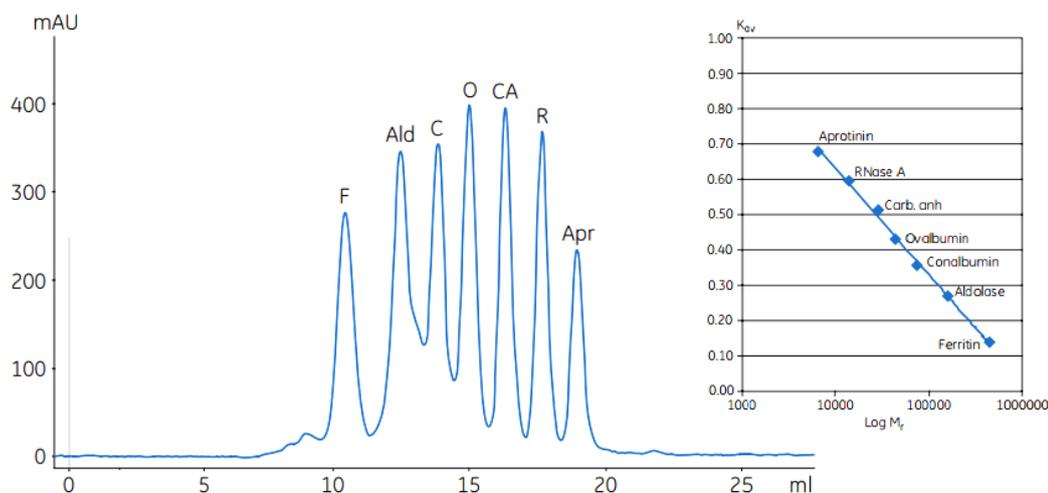
### **2.2.4.3. Purification of N-terminal 6xHis tagged MtbSAICAR and MtbpurD**

After determination of the best overexpression conditions, the protein was prepared on a large scale and purified using standard purification techniques with some modifications (Cheng et al., 2010). A similar protocol was used for both enzymes,

specifically: the cell pellet used to extract MtbSAICAR was resuspended into using 300mM NaCl, 20mM Tris-HCl pH8.0, 1mM DTT, 10mM imidazole pH8.0 and 1 complete protease inhibitor (Roche-Germany). For MtbpurD 500mM NaCl, 20mM Tris-HCl pH8.0, 1mM DTT, 10mM imidazole pH8.0 and 1 complete protease inhibitor (Roche-Germany) was used. This was then sonicated (10 seconds on, 30 seconds off for three times over ice) then centrifuged for 30 minutes at 4°C at 18000 rpm (SS-34 Sorvall). The supernatant was then added into 1ml of Ni-NTA agarose (Qiagen) slurry then agitated for one hour at 4°C on a roller and before being centrifuged. The resin was then washed, for MtbSAICAR with buffer (300mM NaCl, 20mM Tris-HCl pH8.0, 20mM Imidazole, 1mM DTT) and for MtbpurD with 500mM NaCl, 20mM Tris-HCl pH8.0, 20mM Imidazole, 1mM DTT) four times and the fifth wash collected, finally the 6xHis tagged enzyme was eluted with elution buffer for MtbSAICAR of (300mM NaCl, 20mM Tris-HCl pH8.0, 250mM Imidazole, 1mM DTT) and for MtbpurD (500mM NaCl, 20mM Tris-HCl pH8.0, 250mM Imidazole, 1mM DTT) and left for 15 minutes before being centrifuged for 5 minutes at 4°C. TEV protease was then added at a protein ratio of 1:100 for MtbpurD and 1:50 for MtbSAICAR. The sample was dialysed at 4°C overnight with shaking into 150mM NaCl, 20mM Tris-HCl pH8.0, 1mM DTT for MtbSAICAR and, 200mM NaCl, 20mM Tris-HCl pH8.0, 1mM DTT for MtbpurD.

#### **2.2.4.4. Gel Filtration Chromatography of MtbSAICAR and MtbpurD**

To evaluate the molecular weight and oligomeric state in solution of the proteins of interest as well as to maximise purity, the prepared mixtures of proteins were subjected to further gel filtration chromatography steps. The principle is that larger molecules are retarded less in the matrix and elute sooner. The oligomeric states of MtbSAICAR and MtbpurD were determined using Superdex 200 Increase 10/300 GL (GE Healthcare) which is a gel filtration media appropriate for proteins between 10000 Da and 600000 Da to determine approximate molecular weight. 20mM Tris-HCl Buffer pH8.0 and the protein sample were first filtered using a 0.22µm filter (Sigma Aldrich) to remove any contaminants. Calibration Kit proteins were used to provide a calibration curve with a range of known molecular weight proteins. The molecular weights of MtbSAICAR and MtbpurD were determined using the standard calibration curve illustrated in Figure (27).



**Figure (27)** Calibration by separation of standards with Superdex 200 10/300 GL column. Aprotinin (Apr) (6500), Ribonuclease A (Rnase A; R) (13700), Carbonic anhydrase (Carb. anh; CA) (29000), Ovalbumin (O) (44 000), Conalbumin(C) (75000), Aldolase2 (Ald) (158000). The figure is taken from <http://www.gelifesciences.co.jp/catalog/pdf/28407384.pdf>.

## 2.3. Bioinformatics

### 2.3.1. Homology model alignment of MtbpurD

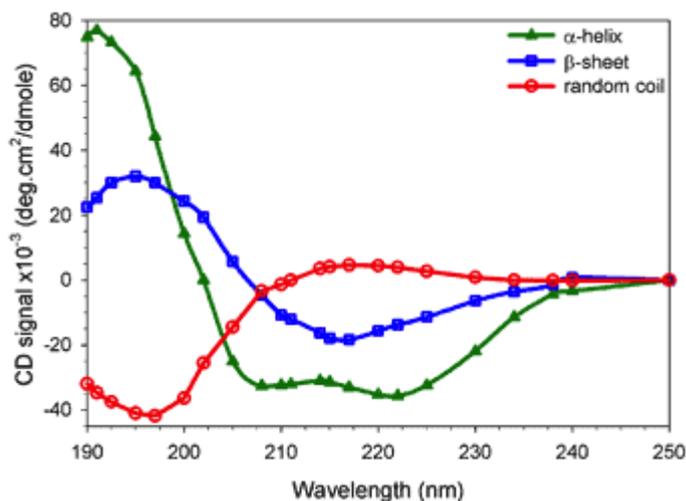
A homology model of MtbpurD was built using the protein fold recognition server PHYRE<sup>2</sup> (Kelley et al., 2015) in order to predict the 3D fold of MtbpurD. In this process a query hidden Markov model (HMM) is built using the information from the sequence alignment (built using the query sequence) by HHblits; sequence database (lightning-fast iterative protein sequence searching by HMM-HMM alignment) database (Zimmermann et al., 2017) and secondary structure features by PSIPRED (a Protein Sequence Analysis). The HMM model is run against HMMs database of proteins (with known structure) and alignments are selected based on coverage and confidence. The backbone of models are built (based on this selected alignment) whilst any insertion or deletion is fixed by a loop modeling process. Because the aim is to generate a perfect query model, the regions with detectable and undetectable homology to template models is refined by Poing and *ab initio* modelled (questionable step) respectively. The final model is achieved by rebuilding the backbone then the side chains are added. 80% accuracy is claimed if the backbone modelled correctly (Kelley et al., 2015).

The MtbpurD homology model was constructed using six templates including *Aquifex Aeolicus* (AaPurD) (PDB 2YYA) *Ehrlichia chaffeensis* (PDB 3LP8), *Neisseria gonorrhoeae* (PDB 5vev), *Geobacillus kaustophilus* (PDB 2ys6), Human (PDB 2qk4), *Bacillus subtilis* (PDB 2xd4) with >90% confidence (which reflects the homology of the templates). This indicates that both template and model are expected to have the same fold and the thus the prediction of the core shape will be reliable. Although a sequence identity >30-40%, is considered the threshold for a reliable model, some useful modelling can be done even when sequence identity <15% if the confidence is high. Here MtbpurD shares 38%, 37%, 41%, 43%, 39% and 42% sequence identity respectively with the homologue template structures. The binding sites for the natural substrates of purD, ATP, and glycine, were compared by superimposing the model structure over to the homologous enzymes from Human (HspurD) (PDB 2QK4), the overall model was also compared to the structure of *E. coli* purD (EcpurD) (PDB 1GSO). The similarities and differences of the whole structure as well as ATP binding site are determined as shown in Chapter 6, Section 6.6.

## **2.4. Biophysical Characterisation**

### **2.4.1. Circular Dichroism**

Circular dichroism (CD) is a technique that is commonly used for the estimation of the secondary content and to indicate whether or not the protein of interest is folded properly in solution and estimate thermal stability. It can also be used to investigate protein-protein and protein-ligand interactions. (Kelly, Jess, and Price, 2005). Circular dichroism spectroscopy estimates the variation between the absorption of left and right-handed circularly polarised light, this arises due to structural asymmetry. Secondary structural elements show characteristic CD spectra, if a protein is  $\alpha$ -helix dominant then negative bands at 222nm, 208nm and a positive band at 193nm are seen. In contrast, predominantly  $\beta$ -sheet proteins show a negative band around 218nm and positive bands at 195 nm. Disordered proteins give a positive band at 210nm, with negative bands near 195 nm (Figure 28).



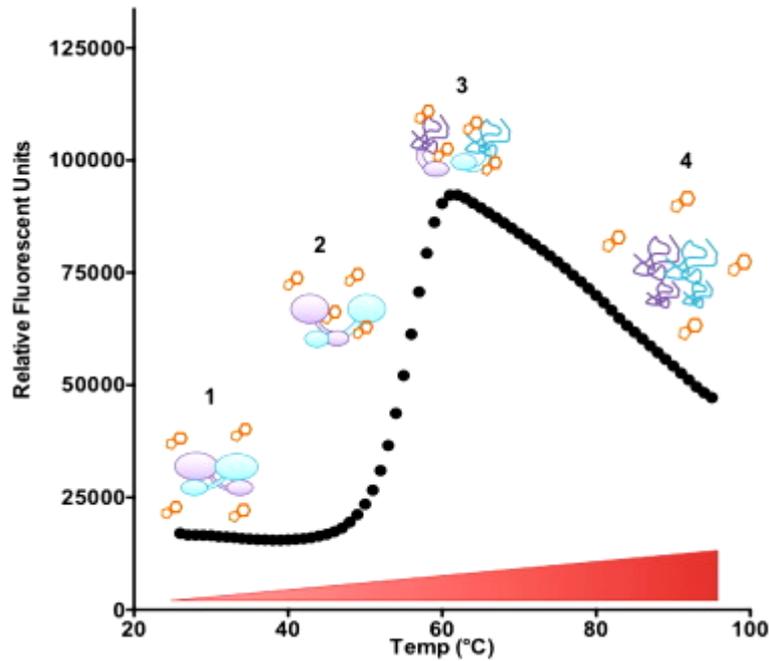
**Figure (28)** Circular Dichroism spectra comparing secondary structure content from <http://www.fbs.leeds.ac.uk/facilities/cd/>

#### 2.4.1.1. Secondary structure determination and thermal stability of MtbSAICAR-MtbpurD

MtbSAICAR was buffer exchanged with 20mM Tris-HCl pH8.0, 50mM NaCl whereas for MtbpurD the buffer was exchanged with 20mM phosphate buffer pH7.4 and 100mM sodium fluoride prior to the experiment. The protein concentration was estimated using Bradford dye-binding protein assay and adjusted to give concentrations of MtbSAICAR of 0.56mg/ml and MtbpurD of 0.15mg/ml respectively. Calibration used the specific buffer in the absence of protein, then the CD spectra of the sample was acquired using Chirascan-Plus CD spectrometer (Applied PhotoPhysics, UK). The signal was subtracted from the blank and the CD curve plotted. The CD spectrum plots the rotation (mdeg) versus wavelength. Using GraphPad Prism (version 6.00 for Windows, GraphPad Software, La Jolla California USA) the folding state of the recombinant MtbSAICAR and MtbpurD was calculated. The thermal stability of MtbSAICAR was determined by measuring the CD spectra over a range of temperatures from 5°C to 90°C; increasing in 0.5°C increments. The change in CD is measured at a single wavelength, 222nm. The secondary structure content of MtbpurD was calculated using online server using the method of Raussens et al., (<http://perry.freeshell.org/raussens.html> (Raussens, Ruyschaert and Goormaghtigh, 2003)).

#### **2.4.2. Thermal shift assay of proteins**

The thermal shift assay is a technique that can be used to test the effect of different ligands on the thermal stability of the protein of interest. As the temperature is increased (20-95°C), the protein undergoes conformational changes as the protein changes from the folding to the unfolded state. The process is monitored by observing the changes in fluorescence of sypro orange dye (excitation 492nm, emission 610nm) (Life Technologies, 5,000x stock). The dye does not bind to the protein when it is folded, as the hydrophobic regions are masked from the solvent. As the protein unfolds with an increase in temperature, the hydrophobic regions of the protein become accessible and the dye binds tightly to these hydrophobic regions (Niesen, Berglund and Vedadi, 2007). When the protein is further denatured and starts to aggregate, the dye is excluded. The process is illustrated in Figure (29). The experiment was carried out using IQ5 (Bio-Rad). 4µM of MtbSAICAR or MtbpurD was mixed with various ligands combined with 2µl of Sypro orange dye (1:1000) of 1:80 stocks into a total volume of 25µl per well (repeated 12times) in 96 wells plate using a buffer of either 20mM Tris-HCl pH8.0 and 50mM NaCl or 20mM HEPES pH8.0 and 50mM NaCl. The temperature raised from (20-95°C) with 0.5°C increments and for each step the temperature was maintained for 20 seconds. 12 observations were used to give the average relative fluorescence unit (RFU) of each temperature. The unfolding curve was plotted temperature on X-axis versus relative fluorescence units (RFU) on the Y-axis. This plot was then used to calculate the melting temperature (T<sub>m</sub>) using GraphPad Prism (version 6.00 for Windows, GraphPad Software, La Jolla California USA).



**Figure (29)** Principle of Thermal shift assay. 1. Folded protein mixed with Sypro orange dye in solution. 2. As the temperature increases, the protein undergoes unfolding causing increases the binding of the dye to hydrophobic regions and in parallel the fluorescence signal increases. 3. The protein reaches maximal unfolded state and at same time maximal fluorescence as there is maximal binding of the dye to hydrophobic regions. 4. The protein starts to aggregate, causing the dye to be dissociated and accordingly the fluorescence signal decreases. Figure is taken from (Abbott et al., 2017).

### 2.4.3. SAICAR Synthetase kinetic properties

The Michaelis-Menten model is a common method for enzyme kinetics. According to this approach the substrate (S) reversibly binds to an enzyme to form an enzyme-substrate complex (ES), which then irreversibly reacts to generate a product (P) and to regenerate the free enzyme E. This system can be represented schematically as follows:



**Figure (30)** The Michaelis-Menten enzymatic reaction scheme, E represents Enzyme, S indicates substrate, ES is the complex of E and S and E+P are the products of breakdown ES complex

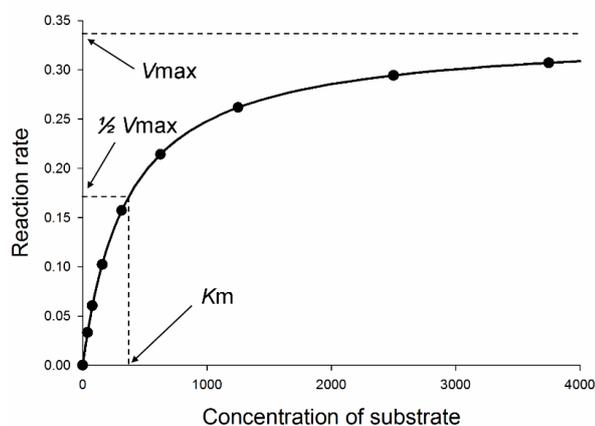
The equations derived from this model contain two important terms. One of them is velocity (V) which represents the reaction rate and the other is Michaelis constant ( $K_M$ ) which is the substrate concentration at which the reaction velocity is 50% of the  $V_{max}$ . At the start of the reaction, the concentration of P will be negligible (and so there will

be no reverse reaction) and so initial rates are observed.  $V_{\max}$  is the maximum velocity achieved by the system, at saturating substrate concentrations. The Michaelis-Menten equation for this system is:

$$v = \frac{V_{\max} [S]}{K_M + [S]}$$

**Figure (31)** The Michaelis-Menten equation

According to the Michaelis Menten scheme, at a fixed enzyme concentration, the proportion of active sites of the enzyme molecule occupied is proportional to the concentration of substrates. When all the active sites have been occupied by the substrate, the enzyme is saturated and is therefore working at its maximum capacity, increasing the concentration of substrate further will not increase the rate of enzymatic catalysis. A plot of the Michaelis-Menten equation's behavior on initial reaction rate ( $V$ ) as a function of substrate concentration is given in Figure (32), showing the significance kinetic parameters  $V_{\max}$  and  $K_M$ .



**Figure (32)** Michaelis Menten ideal curve, plotting  $V$  (the reaction rate) against  $[S]$ ,  $V_{\max}$  is the maximum reaction velocity,  $K_M$  is the Michaelis constant at which the substrate concentration gives a reaction rate of  $\frac{1}{2} V_{\max}$ . The figure is taken from <http://orgchem.tsu.ru/enzyme/Michaelis-Menten%20kinetics.htm>.

The enzymatic activity studies were carried out by measuring the rate of phosphate production using the Biomol Green kit (Enzolifesciences.com, 2016) to measure the phosphate concentration. This assay follows the absorbance change at 600-680 nm (i.e. from yellow to green) for 20-30 min and is conducted in 96 wells plates. In each

case, 100µl of Biomol Green Reagent was added to the assay reaction mixture. The rate of phosphate production was measured following the color change.

The assay was calibrated with a series of dilutions (to 40µM, 20µM, 10µM, 5µM, 2.5µM, 1.25µM, 0.625µM) from phosphate standard (stock concentration 800µM) with 50µl of the assay buffer as blank (50mM Tris-HCl pH8.0, 0.5mM EDTA, 6mM MgCl<sub>2</sub>, 0.01% BSA) plotted against the absorbance ( $A_{620}$ ). Each measurement was repeated twice, the absorbance averaged then subtracted from the blank. This calibration allows the changes in phosphate concentration during the kinetic experiments to be determined by changes in absorbance. These numbers fit along phosphate concentrations of standard values which help to calculate the slope after schemed in x y chart type drawn and the best-fit straight line is placed (Figure 50). The standard curve slope is used to divide the absorbance of each experimental samples measured at a specific time for every parameter (ATP, CAIR, and aspartate). This will estimate the released phosphate at a specific time for each substrate (at a constant concentration of MtbSAICAR at 100nM). The released phosphate (y-axis) was then plotted against time (seconds, x-axis) to determine the rate of the reaction at every kinetic parameter. The rate of reaction produced (y-axis) of each sample for each substrate plot against the substrate concentrations used (x-axis) to understand the relationship between the rate of reaction and substrate concentration as well as calculating the  $V_{max}$ ,  $K_M$ .

#### **2.4.4. Crystallisation trials**

##### **2.4.4.1. MtbSAICAR Crystallisation Experiments**

To investigate the best crystallisation condition for the purified enzymes, commercial JSCG+ crystallisation screen (Molecular Dimensions) was first used. 96-well plates were set up using Mosquito crystallisation robot (TTP labtech). 100nl of mother liquor from 100µl in the wells was mixed with 100nl of protein solution. The plates were then was sealed and kept at room temperature and checked frequently for crystal growth. As a good crystal is not always obtained from an initial crystallisation screen, the conditions need to be optimised around pH, concentration of protein, precipitant concentration, and temperature.

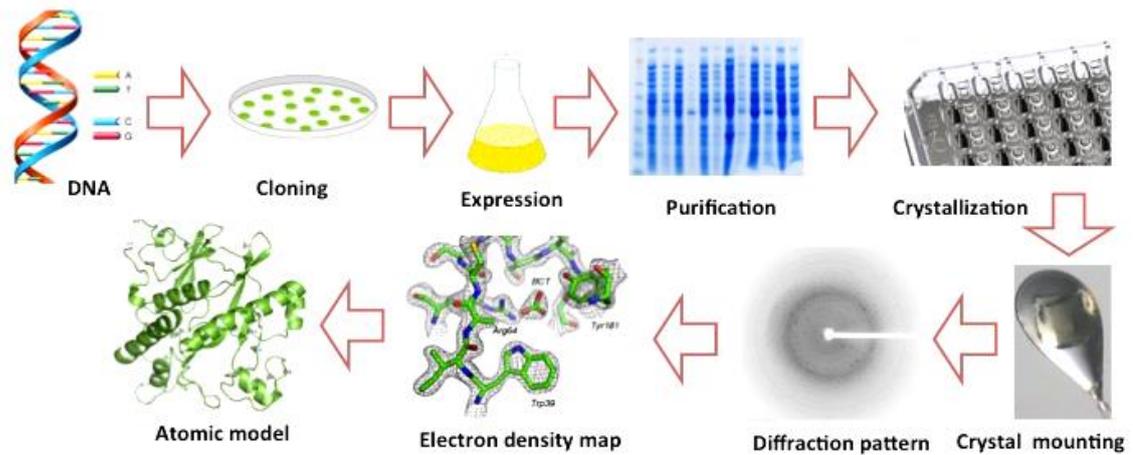
#### **2.4.4.2. MtbSAICAR co-crystallisations**

MtbSAICAR with a concentration of 15.3mg/ml was prepared with 10mM of ATP-Mg<sup>2+</sup> and 4mM aspartate), 10mM of Adenylyl-imidodiphosphate (ACP) and Inosine 5'-monophosphate) (IMP), 10mM of ADP and analogue 5-aminoimidazole-4-carboxamide ribonucleotide (AMZ). 1:1 ratio of protein was mixed with mother liquor containing 0.1M HEPES pH8.0, 0.02M MgCl<sub>2</sub> and 22% Polyacrylic acid sodium salt 5100. The protein was either crystallised in apo form or with ligands. The plates were kept at room temperature and observed within hours. Prior data collection, the crystals produced from co-crystallisation conditions were rested in mother liquor containing 20% glycerol and the specified ligands.

#### **2.4.5. Introduction to X-ray crystallography**

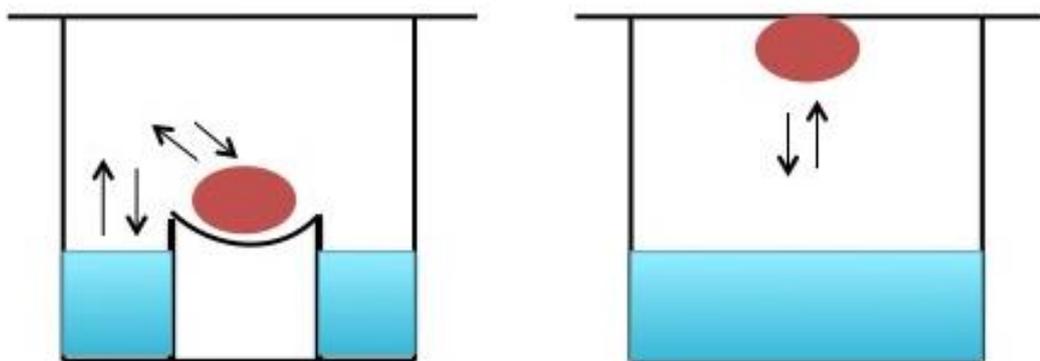
The three-dimensional structure of protein gives information about its properties and function. X-ray crystallography provides accurate and precise structure in atomic detail (Hasegawa, 2012), information that cannot be achieved from light microscopy. The X-ray method provides an important in understanding the details of enzymatic mechanisms and protein-ligand, protein- protein and protein-DNA interactions.

One important step in understanding mechanism and action of protein through X-ray crystallography is obtaining three-dimensional well-ordered crystals. Obtaining such a high quality crystal typically involves several steps starting from ORF amplification of the gene of interest to crystallisation trials and optimisation. These steps are schematically represented in Figure (33). Growing perfect crystals is not easy as the crystal growth is influenced by many factors such as impurities in the protein, the presence of extraneous chemicals, temperature, concentration of precipitant/protein, pH of buffer and so on (Abola et al., 2000). That's why the success ratio in getting crystals is only around 30%-40% (Chayen and Saridakis, 2008).



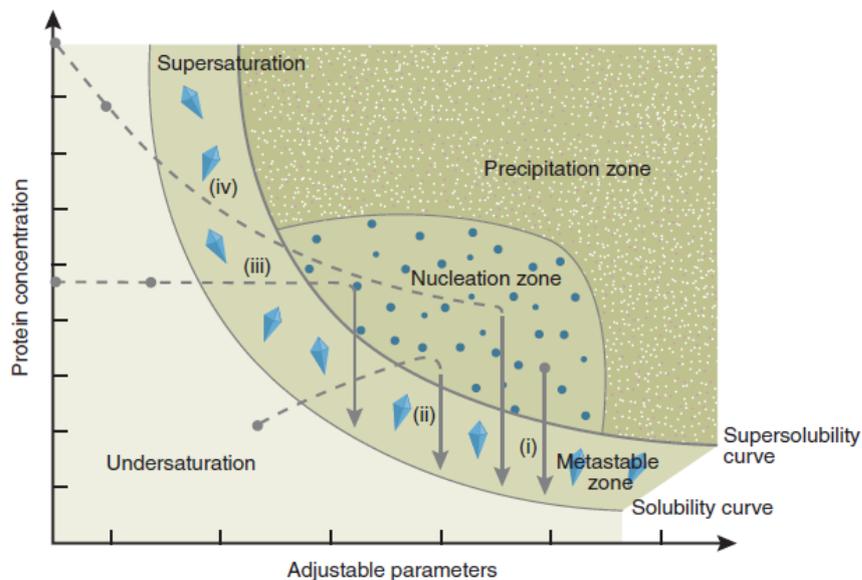
**Figure (33)** Steps involved in structural determination of a protein.

The methods used for crystallisation of proteins include vapor diffusion, free interface diffusion, batch, and dialysis. Among those techniques, vapour diffusion is the most commonly used. There are two types, hanging drop and sitting drop (Figure 34). Both of these use the same principle with the only difference being that in sitting drops the protein is mixed with the precipitant solution and sits on a tiny well. In hanging drops, the protein solution is mixed with a precipitant solution on a cover slip and put upside down over the precipitant reservoir and sealed. In this system, water will evaporate from the drop on the slide into the reservoir because of the difference in the concentration of precipitant between drop and well. This evaporation will continue until the concentration of precipitant in the drop and reservoir reach equilibrium.



**Figure (34)** Vapor diffusion style in protein crystallisation sitting drop left, hanging drop right, 1:1 $\mu$ l of protein solution from 0.5 up to 1ml buffer solution.

Basically, in a crystallisation experiment, the crystal is obtained in a condition where the solubility of the protein is gently decreased. This process is shown in the phase diagram of crystal growth (Figure 35). At the beginning of the crystallisation experiment, the protein is in the soluble phase. As vapour diffusion occurs and the solubility limit is exceeded due to evaporation, the protein solution in tiny wells will become supersaturated and this will cause the protein to be nucleated and protein molecules will be separated from solution and then crystals will grow.



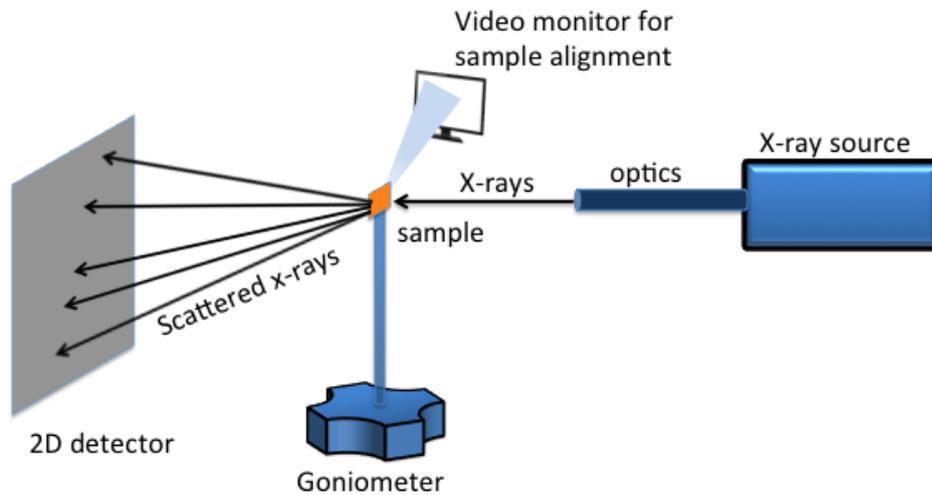
**Figure (35)** Phase diagram of crystal growth. (i) Microbatch, (ii) Vapor diffusion, (iii) Dialysis and (iv) Free interface diffusion (FID) (Chayen and Saridakis, 2008).

#### 2.4.5.1. Data Collection

Once a well-ordered crystal is produced from wet lab experiments, the next step is the collection of data. The data can be collected using both in-house X-ray sources and /or synchrotron radiation. Prior to data collection, the crystal first needs to be mounted using appropriately sized loops and cryo-protected then plunged into liquid nitrogen. This will protect the crystal from drying and lessen radiation damage. After these steps, the crystal is positioned on the goniometer accurately so that the X-rays can hit the crystal properly.

For data collection, several parameters can be varied including the oscillation angle, the wavelength of X-rays, crystal to detector distance, and exposure time so that the best quality of data can be obtained. The diffraction data is collected, as reflections

and is recorded on a charge-coupled device (CCD) or other electronic detector (Figure 36).

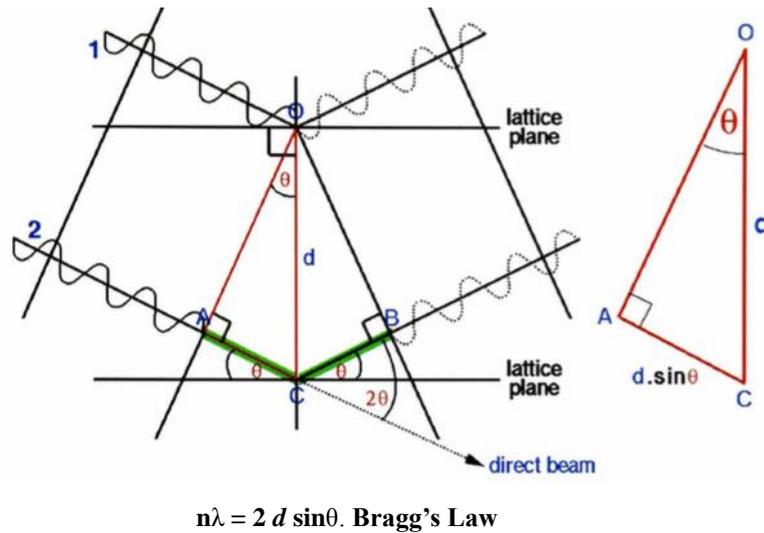


**Figure (36)** General view of Data Collection process on X-ray diffractometer. Picture is taken from (B. B. He, U. Preckwinkel, and K. L. Smith, *Advances in X-ray Analysis*, 2000, **43**, 273).

The choice of exposure time is critical a short exposure time may result in less resolution, as the signal to noise ratio is reduced on the other hand the intensities and hence signal to noise might be improved by increasing the exposure time, however, this may affect the crystal due to possible radiation damage.

#### **2.4.5.2. Theory of Bragg Law in the diffraction pattern**

In order to get a diffraction pattern and spots, the two waves should be in phase, as they should interfere constructively. Destructive waves are not going to produce a diffraction pattern. In constructive interference, the waves are in parallel to each other as explained by Bragg law (Bragg and Bragg, 1913) in Figure (37).



**Figure (37)** Theory of Bragg Law in the diffraction pattern. Taken from (Lovering, 2003)

#### 2.4.5.3. Data processing

Although a few images are enough to get information of crystal symmetry, a full data set will be required to be recorded in order to determine the position of each atom precisely. This is achieved by collecting data as the crystal is rotated through a sufficient angle (depending on the crystal symmetry) to collect the unique data. The first step of data processing is indexing. In indexing, initially two or more images are selected and the positions of the spots are used to calculate the unit cell and suggest a probable Laue symmetry and likely space group and the orientation of the crystal. The indexing allows each spot to be given an index (h, k, l). Following indexing, the positions of all the spots are predicted, and the intensities integrated. In integration, the intensity of each reflection (labeled I) and the associated errors or uncertainties ( $I\sigma$ ) are estimated. Following integration, the data is compared internally using Pointless (Evans, 2011), which uses the possible symmetry operations and systematic absences to suggest a likely space group. After determining the symmetry, the redundant data is scaled together and a unique list of reflections created, the internal agreement is used to assess the quality of the data, this is implemented in the program Aimless (Evans and Murshudov, 2013). Structure factor amplitudes,  $|F|$ , are then calculated from the intensities.

#### **2.4.5.4. Phase determination and structural analysis**

The structure factor amplitudes of each h k l reflection alone are not sufficient to determine the three-dimensional structure of a protein. The Fourier transform also requires a phase for each reflection. Determining the phase can be achieved by several methods including direct methods, isomorphous replacement and Molecular replacement (MR) (Taylor, 2010). Among all of these methods, MR is going to be explained in this paper, as it is the method used to solve the phase problem for the research presented here. MR is a technique used to solve the molecular structure of a protein based on another previously solved structure. The suitability of the homologous structure depends on its sequence identity. The higher sequence identity the better and at least 30% similarity normally required.

Defining the correct orientation and precise position of the molecule is important for estimating the initial phase and uses the Patterson map. The Patterson is calculated from a Fourier transform of the square of the structure factor amplitudes (i.e. the intensities without phases) The Patterson is calculated from the search model and compared to the Patterson from the observed data, by rotating and translating the model and an optimal agreement of the Patterson shows the position and orientation of the protein we are trying to solve. From this model, phases can be calculated to start the process of refinement. This calculation is done in Phaser in CCP4 (McCoy et al., 2007).

#### **2.4.5.5. Refinement and Structural validation**

Initial phasing will provide an electron density map, but the model will require modification and adjustment from the search model (Kleywegt, 2000). The initial model is used as a starting point towards reaching the final model. Then refinement process fixes any remaining errors in the best solution, this is achieved by manual rebuilding in COOT and computational least-squares refinement using Refmac5, CCP4 (Murshudov, Vagin, and Dodson, 1997) or PHENIX (Adams et al., 2011).

COOT (macromolecular model building, model completion, and validation) (Emsley and Cowtan, 2004), allows the interactive assessment of the structure for missing parts, poor fit to electron density and geometric parameters. The improvement of the

structure proceeds iteratively between interactive rebuilding in COOT and refinement in Refmac5. Analysis of the quality of the final model refinement is done using validation processes by means of important elements such resolution and Residual Factor (R-factor), which is the most commonly used monitor of the overall agreement the model to the observed data. The R-factor compares the calculated structure factors (from the model) and observed structure factors. An aim of refinement is to bring the calculated structure factors as close as possible to the observed ones (Figure 38).

$$R = \frac{\sum \left| |F_o| - |F_c| \right|}{\sum |F_o|}$$

**Figure (38)** Residual Factor which calculates the differences between calculated structure factor amplitudes and observed ones.

The molecular model building and refinement are dependent on the crystallographic data that is collected at low temperature from the optimal crystal. Poor quality crystals with more disorder regions may have negative effects on the quality of the data (Spek, 2009). Further parameters are monitored during refinement and validation such as R-free, Root Mean Square Deviations (RMSD) from ideal geometric values of bond lengths and angles, and Ramachandran plots also help in the assessment. Geometry was analysed to compare with ideal values of protein using structural validation servers Procheck (Laskowski et al., 1993).

#### **2.4.5.6. Optimal Structure alignment process**

Comparison and analysis of homologous enzyme structures and discussion of their similarities and differences requires optimal superimposition. In some cases, this will be best achieved by simply matching over the whole chain by minimising the RMS differences between C $\alpha$  positions in, for e.g. a pairwise comparison. However, should there be insertions or deletions, it is better not to include these. The exclusion of residues needs to be considered when quantifying how well structures superimpose. The Q-score value addresses this by using both RMSD and N<sub>align</sub> (the number of

residues aligned in the superposition) (Krissinel, 2012). This metric is considered superior to overall or selective RMSD. In the case of a multi-domain structure, very similar individual domains (or sub-domains) may have different orientations with respect to each other and a global superimposition would give misleading metrics about the similarities between structures, In this study, the overlaid structures were examined over both the entire length (i.e. the whole polypeptide chains) and over particular segments (i.e. per domain). The superimposed alignment for whole chains was calculated by PDBeFold (Krissinel and Henrick, 2004), whereas the optimal superimposition for individual domains used the GESAMT (General Efficient Structural Alignment of Macromolecular Targets) method as this avoid limitation within the SSM algorithm used by PDBeFold (Krissinel, 2012).

The Q-score equation is given by.

$$Q = \frac{N_{\text{algn}}^2}{(1 + (\text{RMSD}/R_0)^2) * N_{\text{resi1}} * N_{\text{resi2}}}$$

**Figure (39)** Q-score equation. It evaluates the relationship between RMSD between  $C\alpha$  atoms of the aligned structures and  $N_{\text{algn}}$  (which is the number of aligned residues).  $N_{\text{resi1}}$  and  $N_{\text{resi2}}$  count for number of residues in the alignment from each structure,  $R_0$  is an experimental normalising factor (standardised at 3Å). For identical structures ( $N_{\text{algn}}=N_{\text{resi1}}=N_{\text{resi2}}$  and  $\text{RMSD}=0$ ) the Q-score would equal 1 whereas for dissimilar structures the Q-score would tend towards 0, as the as the denominator increases with RMSD and the numerators decrease with  $N_{\text{algn}}$ .

## Chapter 3. Characterisation of SAICARs of *M. tuberculosis* (MtbSAICAR)

### 3.1. Introduction

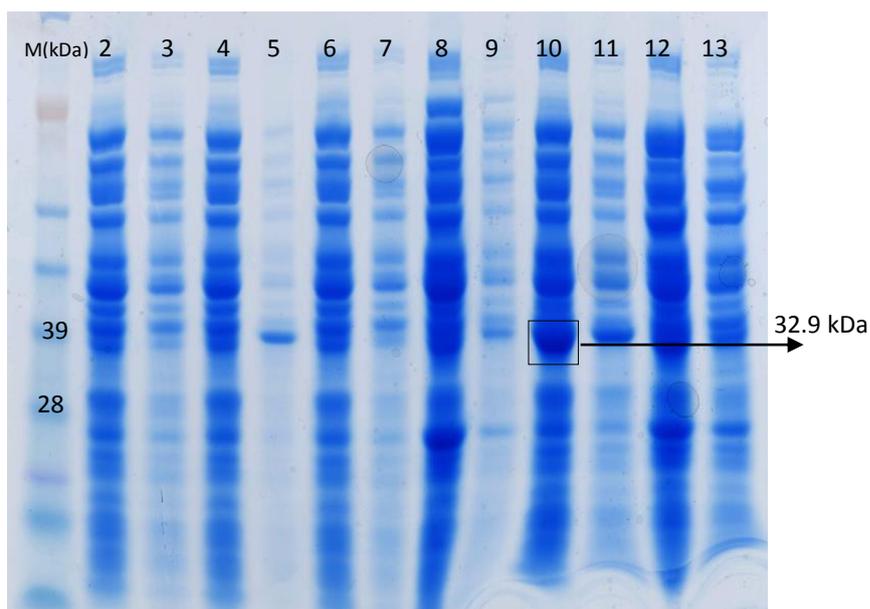
SAICAR synthase of *M. tuberculosis* (MtbSAICAR) catalyses the reaction of 4-carboxy-5-aminoimidazole ribonucleotide (CAIR), ATP and aspartate in order to produce SAICAR, ADP, and Pi in the *de novo* purine pathway. This chapter describes experiments to characterise MtbSAICAR. This was achieved by the cloning, overexpression and purification of MtbSAICAR, followed by the determination of the secondary structure content and stability of the enzyme using circular dichroism and the determination of the kinetic parameters of the enzyme through activity measurements.

### 3.2. Preparation and Purification of MtbSAICAR

The Open Reading Frame of the gene encoding MtbSAICAR was amplified (Figure 40), cloned into the expression vector pLEICS-01, by the protein expression service PROTEX at the University of Leicester and overexpressed in *E. coli* Rosetta (DE3) as described in Chapter 2, Section 2.2.2-2.2.4. The highest expression level of soluble protein was obtained when the expression induced with 150  $\mu$ M IPTG at an OD<sub>600</sub> of 0.50 and grown at 20°C overnight (Figure 41).

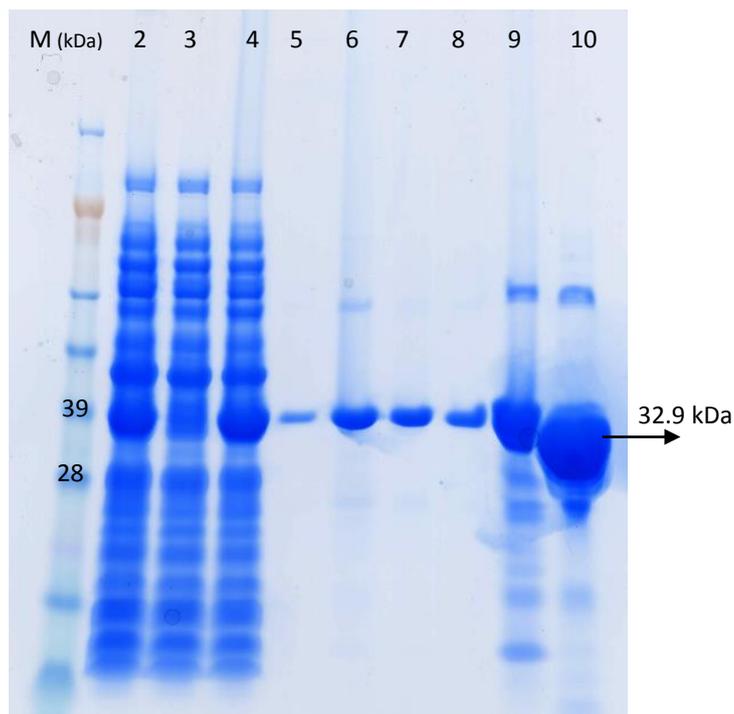


**Figure (40)** ORF amplification of *MtbSAICAR* in 1% (w/v) agarose gel electrophoresis, 1. Molecular weight DNA marker (Bioline 1 kb hyper ladder). 2. PCR product. 3. In the presence of 6% DMSO, the ORF does not amplify.



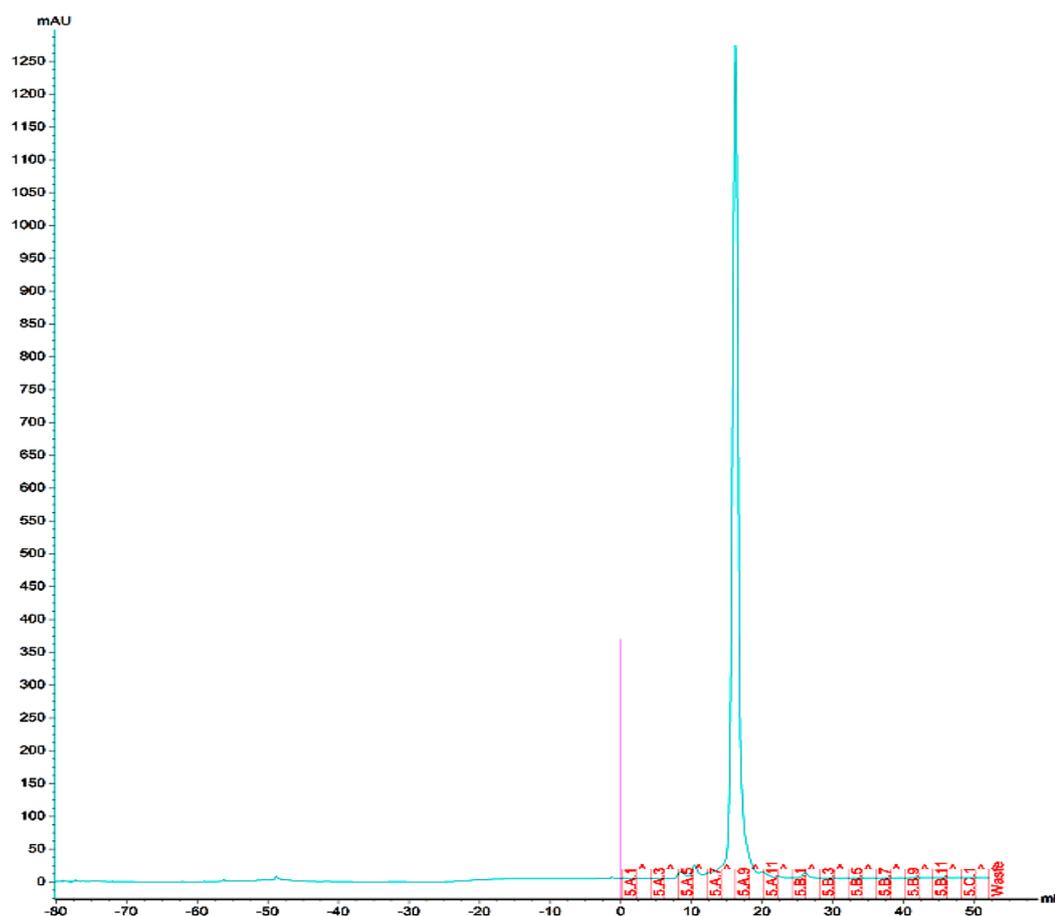
**Figure (41)** Optimisation of overexpression condition of MtbSAICAR in *E. coli* Rosetta (DE3). Lane 1: M. SeeBlue2 (Life Technologies) protein marker. Lane 2: Supernatant from cells induced with 400 $\mu$ M IPTG for 2 hours at 37 $^{\circ}$ C. Lane 3: Pellet from cells induced with 400 $\mu$ M IPTG for 2 hours at 37 $^{\circ}$ C. Lane 4: Supernatant from cells induced with 400 $\mu$ M IPTG for 3 hours at 37 $^{\circ}$ C. Lane 5: Pellet from cells induced with 400 $\mu$ M IPTG for 3 hours at 37 $^{\circ}$ C. Lane 6: Supernatant from cells induced with 400 $\mu$ M IPTG for 4 hours at 37 $^{\circ}$ C. Lane 7: Pellet from cells induced with 400 $\mu$ M IPTG for 4 hours at 37 $^{\circ}$ C. Lane 8: Supernatant from cells induced with 400 $\mu$ M IPTG at 37 $^{\circ}$ C overnight. Lane 9: Pellet from cells induced with 400 $\mu$ M IPTG at 37 $^{\circ}$ C overnight. Lane 10: Supernatant from cells induced with 150 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 11: Pellet from cells induced with 150 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 12: Supernatant from cells uninduced at 37 $^{\circ}$ C overnight. Lane 13: Pellet from cells uninduced at 37 $^{\circ}$ C overnight. Supernatant and pellet mean soluble and insoluble fractions respectively.

Recombinant MtbSAICAR was purified and the purity of the protein was assessed (Figure 42) as described in Chapter2, Section 2.2.4.3. The TEV site was not cleaved after purification. The identity of the protein was confirmed using tryptic digest mass spectrometry (Figure 43). The oligomeric state of the protein was investigated using gel filtration as described in Chapter2, Section 2.2.4.4. and the result suggested that the protein is a monomer in solution (Figure 44). MtbSAICAR elutes in the fraction corresponding to the expected M.W of 32.9kDa, assessed by the calibration supplied by the manufacturers (GE Healthcare) (Chapter2, Section 2.2.4.4).



**Figure (42)** Purification of MtbSAICAR expressed with 6xHis affinity tag. Samples were taken at each stage of the purification (10 $\mu$ l of the sample and 10 $\mu$ l 2XSDS sample buffer) and run on a 4-12% polyacrylamide gel. Lane 1: M. SeeBlue2 (Life Technologies) protein marker. Lane 2: soluble cell extract, Lane 3: unbound protein flow through of Ni-NTA column, Lane 4: Ni-NTA bead, Lane 5: unbound proteins (with 20mM imidazole: fourth wash), Lane 6: MtbSAICAR Ni-NTA bead, Lane 7: MtbSAICAR solution (with 250mM imidazole), Lane 8: Ni-NTA bead, Lane 9: Overnight TEV cleavage at 4°C. Lane 10: Concentrated MtbSAICAR after gel filtration. See Chapter 2 for the experimental protocol.





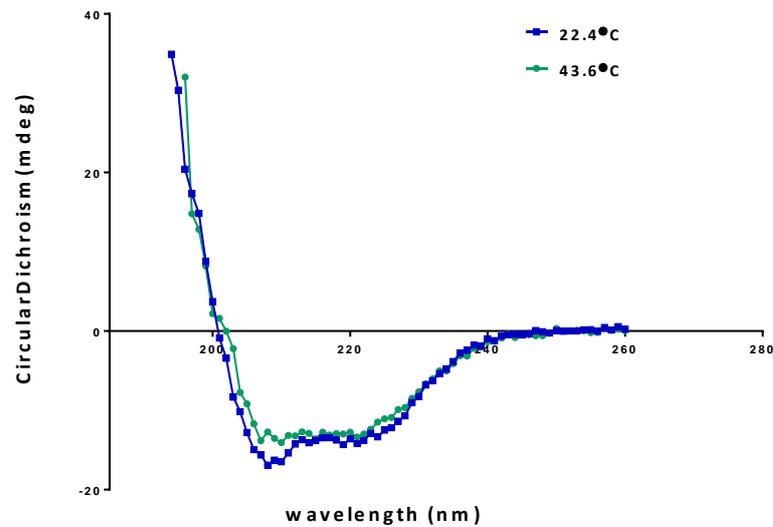
**Figure (44)** Determination of the solution oligomeric state of MtbSAICAR using Superdex 200 Increase 10/300 GL gel filtration chromatography (GE Healthcare). The protein eluted is detected by  $A_{280}$ . The calibration curve is given in Chapter2, Section 2.2.4.4.

### 3.3. Secondary Structure of MtbSAICAR

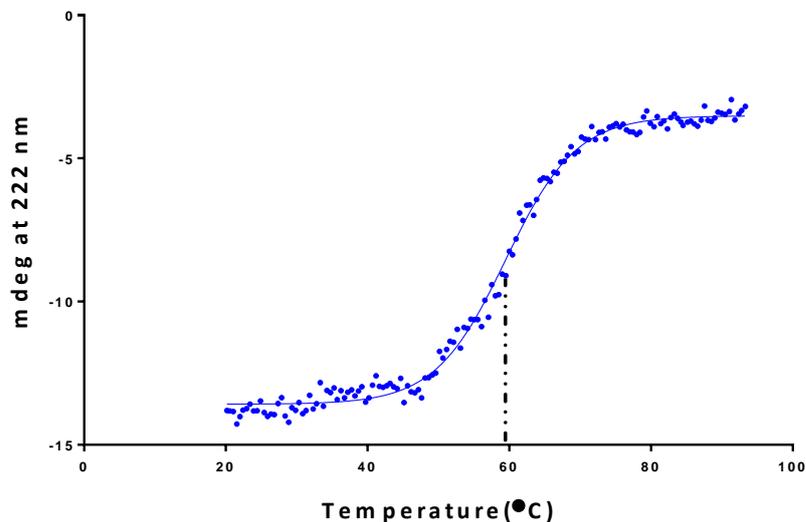
The enzyme was also characterised by far UV Circular Dichroism (CD) initially to determine whether the recombinant protein is folded and subsequently to estimate the secondary structure content and compare this with calculated and predicted values. For that purpose, both computational and experimental techniques were used. PSIPRED Protein Sequence Analysis (McGuffin, Bryson, and Jones, 2000; Buchan et al., 2013) predicted that the protein is predominantly  $\alpha$ -helix, containing 27.2%  $\alpha$ -helix with 9.0%  $\beta$ -strand (Figure 45).



(A)



(B)



**Figure (46)** (A) CD spectrum of apo MtbSAICAR at two temperatures. (B) Thermal unfolding of apo MtbSAICAR. The experiments were carried out as explained in Chapter2, Section 2.4.1.1.

### 3.4. Dye Binding Thermal shift assay of SAICARs *M. tuberculosis* (MtbSAICAR)

The experimental thermal denaturation curves (Figure 47 and 48) are similar to the ideal thermal profile Chapter2, (Figure 29). As the temperature raises gradually, the fluorescence signal initially increases then rapidly declines. Apo MtbSAICAR displays a melting temperature of 43.9°C. The thermal stability of the protein was tested in various conditions (Table 5 and 6), indicating that the enzyme is more stable when it makes a complex with AMZ and ADP as shown in Figure (47). The effect of the following ligands Adenylyl-imidodiphosphate (ACP), 5-aminoimidazole-4-carboxamide

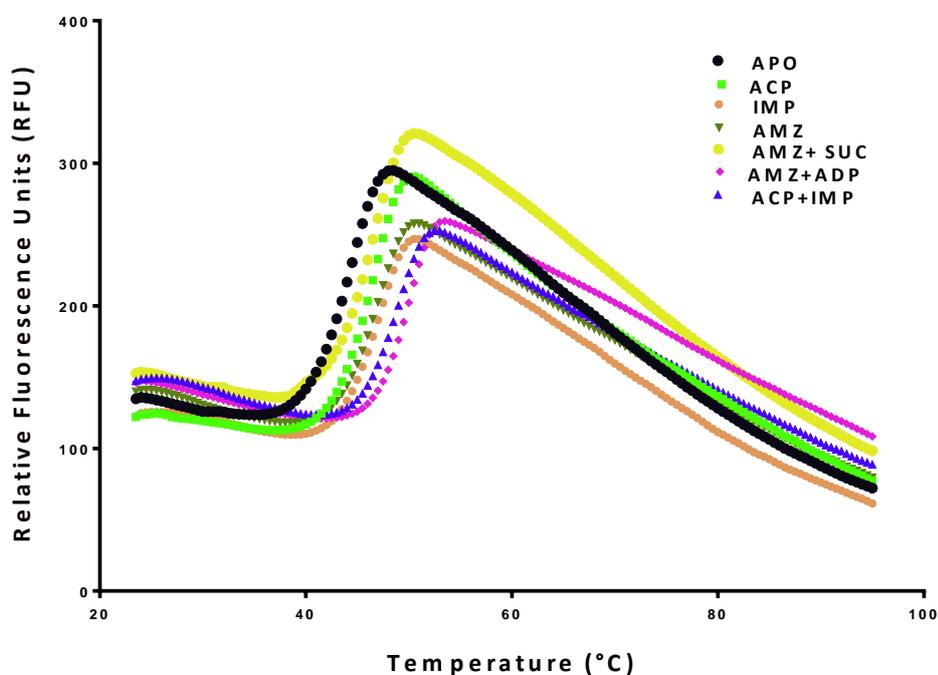
ribonucleotide (AMZ), inosine 5'-monophosphate (IMP) and Succinic acid (SUC) was also tested (Figure 49).

**Table (5)** Assessing melting temperature of MtbSAICAR with different ligands, tested at a concentration of 10mM.

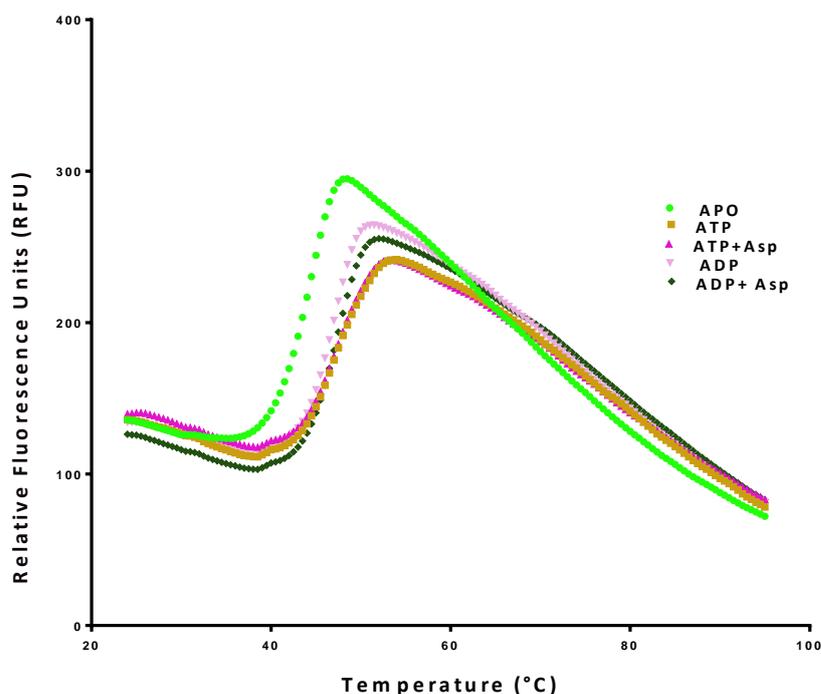
MtbSAICAR Conditions	Temperature (°C)	Change in temperature (°C)
Apo	43.9±0.07	—
ACP	45.9±0.06	2.0
ACP+IMP	48.4±0.04	4.5
AMZ	46.3±0.05	2.4
AMZ+ADP	49.5±0.02	5.6
IMP	46.6±0.06	2.6
AMZ+SUC	46.1±0.05	2.2

**Table (6)** Assessing melting temperature of MtbSAICAR with different tested at concentrations of 5mM of each ATP and ADP, 4mM of aspartate (Asp).

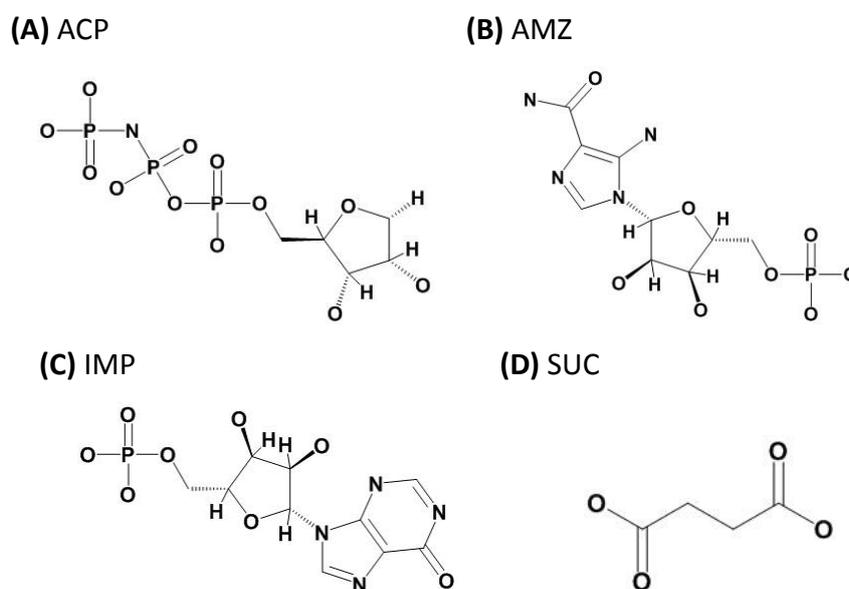
MtbSAICAR Conditions	Temperature (°C)	Change in temperature (°C)
Apo	43.9±0.07	—
ATP	47.7±0.04	3.8
ATP+Asp	47.6±0.04	3.7
ADP	46.9±0.04	3
ADP+Asp	47.2±0.04	3.3



**Figure (47)** Thermal stability of apo MtbSAICAR and in the presence inhibitors and substrate analogues using dye binding thermal shift assay. For methods see (Chapter 2, Section 2.4.2.)



**Figure (48)** Thermal stability of MtbSAICAR compared apo enzyme and in the presence of natural ligands using the thermal shift dye binding assay. ADP is a substrate analogue of ATP.

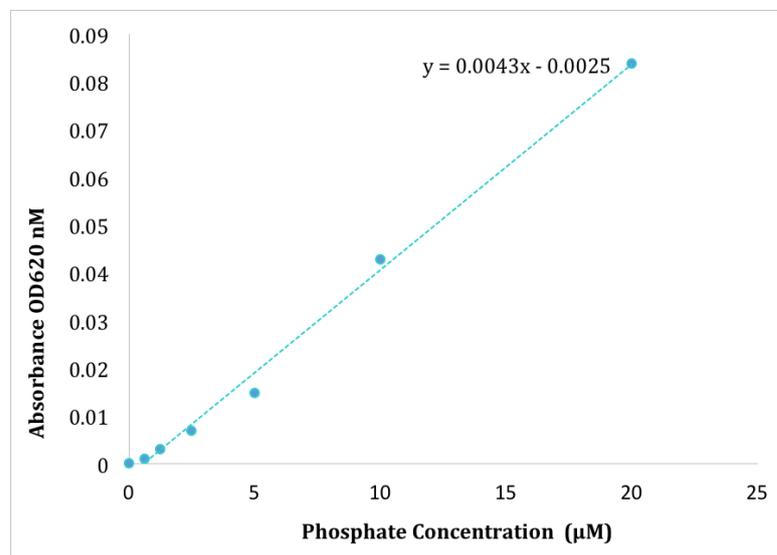


**Figure (49)** Ligands tested with the thermal shift assay. Succinic acid (SUC) is a competitive inhibitor of Aspartate. The structural formulae are drawn using <https://pubchem.ncbi.nlm.nih.gov/edit2/index.html>

### 3.5. *In vitro* enzyme activity of MtbSAICAR

A spectrophotometric enzyme assay (in a plate reader) was used to measure the activity of MtbSAICAR *in vitro* for each substrate used (ATP, CAIR, and aspartate). The measurement of the rate of phosphate production is used to calculate the rate of the

enzymatic reaction. This is achieved using a calibration of phosphate concentration to  $A_{620}$ , and thus the change in colour is used to calculate the change in phosphate concentration, and from this, the slope is calculated. Figure (50) shows the standard curve and the complete methods are described in Chapter2, Section 2.4.3.



**Figure (50)** Standard curve plot of serial phosphate standard concentration (40µM, 20µM, 10µM, 5µM, 2.5µM, 1.25µM, 0.625µM and blank of assay buffer) was plotted on X-axis and Absorbance OD620nm on Y-axis. The assay buffer includes (50mM Tris-HCl pH8.0, 0.5mM EDTA, 6mM MgCl<sub>2</sub>, 0.01% BSA)

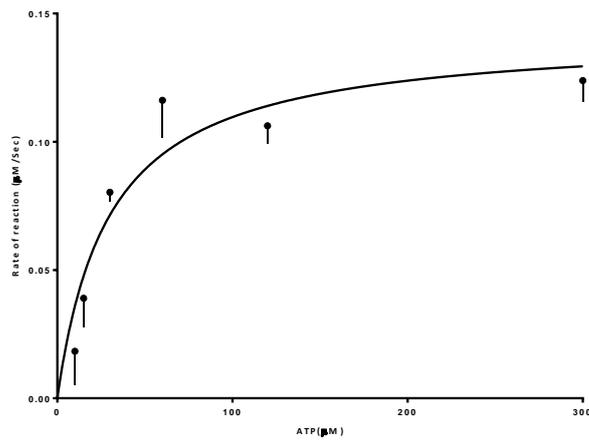
The activity of MtbSAICAR follows hyperbolic Michaelis-Menten kinetics as shown in Figure (51). The reaction rate increases with the substrate concentration and becomes constant where the enzyme is saturated with its substrate. The parameters  $K_M$ , (the substrate concentration at which half the enzyme's active sites are occupied by the substrate), and  $V_{max}$ , the saturated rate, were calculated from the plots of substrate concentration versus velocity. The  $K_M$ ,  $V_{max}$  and  $k_{cat}$  ( $s^{-1}$ ) values for ATP, CAIR, and aspartate are shown in Table (7).

**Table (7)** Kinetic parameters of MtbSAICAR for ATP, CAIR, and aspartate (µM). (Kinetics parameters for were calculated with GraphPad Prism (version 6.00 for Windows, GraphPad Software, La Jolla California USA).

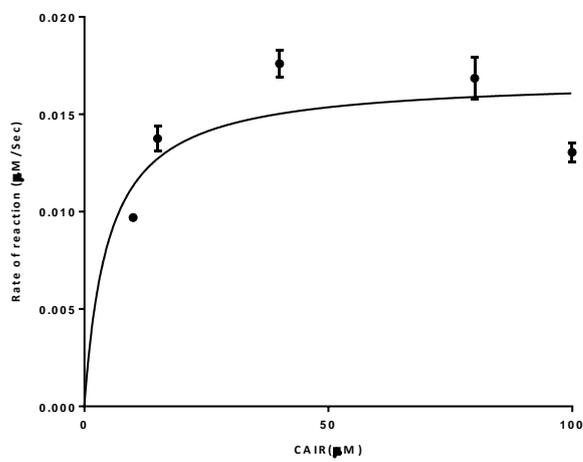
Substrate	$K_M$ (µM)	$K_M$ Range (µM)	$V_{max}$ (µM/s)	$k_{cat}$ ( $s^{-1}$ )
ATP	29.14	30 – 60*	0.14	1.41
CAIR	6.43	~10*	0.017	0.18
Asp	142.3	200-800*	0.03	0.29

\*Expected  $K_M$  range provided by Dr. Peter Coombs MRC technology, UK (unpublished)

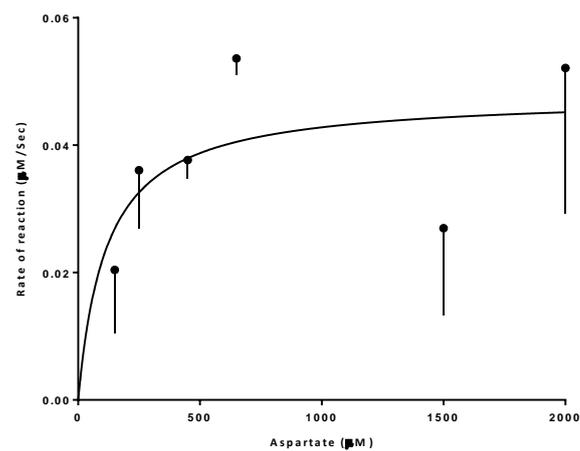
(A)



(B)



(C)



**Figure (51)** Rate dependence on concentration for MtbSAICAR. (A) ATP ( $\mu\text{M}$ ), (B) CAIR ( $\mu\text{M}$ ), and (C) aspartate ( $\mu\text{M}$ ). The reaction rate unit ( $\mu\text{Mol/sec}$ ) is on the Y-axis and substrates concentration unit ( $\mu\text{M}$ ) is on the X-axis. Data was average of the twice-repeated experiment.

### 3.6. Summary

The aim of the work described in this chapter was to characterise MtbSAICAR. This was achieved by ORF amplification followed by cloning into an overexpression vector, overexpression, and purification of the enzyme. The oligomeric state of the enzyme was investigated via gel filtration chromatography, suggesting the protein is in monomeric form in solution as shown in Figure (44). Following purification, the thermal stability of the enzyme was studied by thermal shifting assay and CD experiments. CD results are consistent with the predicted mainly  $\alpha$ -helix content and show it is stable with an unfolding mid-point of 59.47°C. Stability studies by thermal shift assay show that the enzyme is slightly more stable when it is in the presence of ligands such as IMP, aspartate, and AMZ, presumably because it is in complex, as the unfolding temperatures are lower in the presence of dye, suggesting this has a destabilising effect. Activity assays and preliminary kinetic measurements show the enzyme preparation is active and appears to follow Michaelis-Menten kinetics, estimates of kinetic parameters are obtained.

## **Chapter4. Structural determination of SAICARs in *M. tuberculosis* and refinement, results**

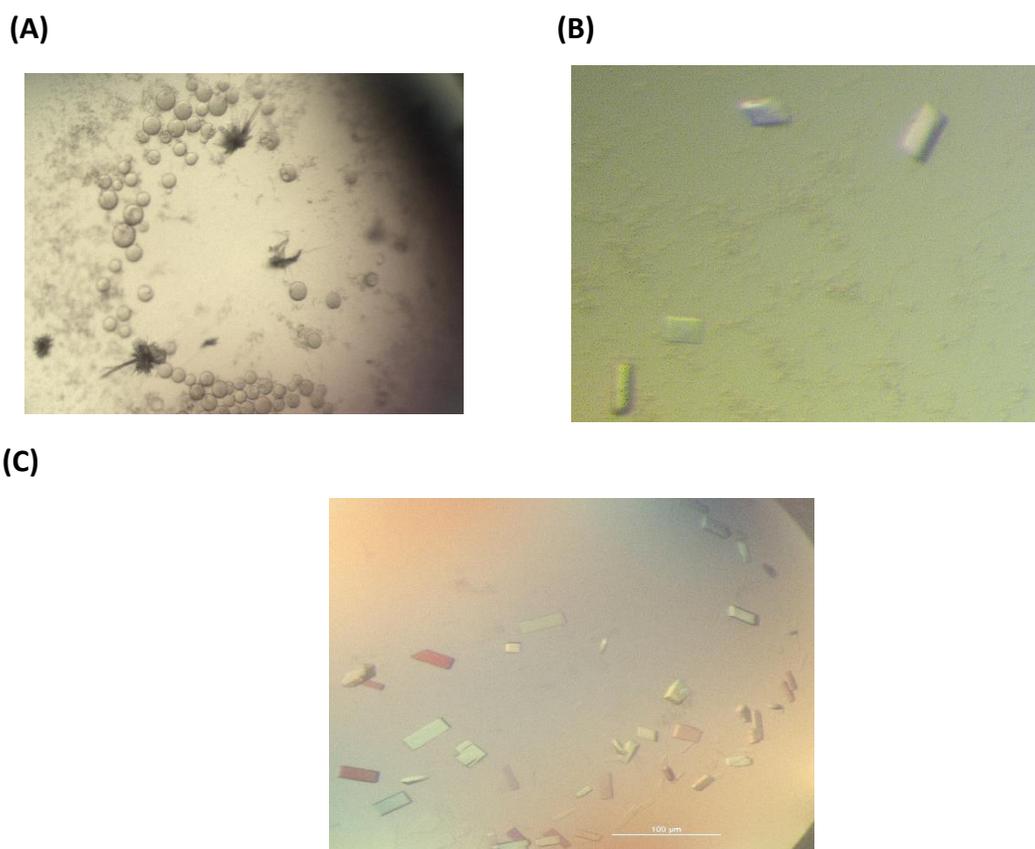
### **4.1. Introduction**

The crystal structure of SAICARs from various organisms including human (HsSAICAR) (Li et al., 2007), *E. coli* (EcSAICAR) (Ginder et al., 2006), *S. cerevisiae* (ScSAICAR) (Levdikov et al., 1998), *S. pneumoniae* (SpSAICAR) (Wolf et al., 2014), *T. maritima* (TmSAICAR) (Zhang et al., 2006) have previously been solved and the mechanism and action of the enzyme has been studied in some detail. However, previous attempts at solving the crystal structure of SAICARs from *M. tuberculosis* (MtbSAICAR) have failed (Baugh et al., 2015). This chapter describes the determination of the crystal structure of apo MtbSAICAR along with its complexes with ATP, ADP- {5-aminoimidazole-4-carboxamide ribonucleotide (AMZ)}, and {Adenylyl-imidodiphosphate (ACP)- inosine 5'-monophosphate (IMP)}. The overall structure and the details of the protein-inhibitor and protein-substrate interactions are described in this chapter.

### **4.2. Crystallisation of MtbSAICAR-ATP complex.**

The initial crystallisation conditions for MtbSAICAR-ATP (10mM ATPMg<sup>2+</sup>) complex were investigated using the vapour diffusion method with sitting drops using the commercially available JSCG+ crystallisation screen (Molecular Dimensions) and the Mosquito crystallisation robot (TTP labtech). Crystals were found in 0.1M HEPES pH7.5, 0.02M MgCl<sub>2</sub>, and 22% poly (acrylic acid) 5100. However, these crystals were not of diffraction quality, therefore these initial crystallisation conditions were optimised with a systematic exploration of pH gradients, different concentrations of precipitants and salts. The highest quality of crystals was found when the pH of the buffer was changed to pH8.0. This precipitant/buffer mix is insufficient for cryoprotection of the crystals, but the inclusion of 20% (v/v) glycerol was found to be sufficient for this. The rationale behind adding a cryoprotectant agent is that it allows vitrification of the aqueous phase around the crystal rather than forming ice crystals that will damage the crystals, and give ice diffraction, thus affecting the quality of the diffraction data. The first crystals of MtbSAICAR were grown in presence of 10mM ATP-4mM aspartate were prepared for data collection by transferring them into a cryo buffer containing 0.1M HEPES pH8.0, 0.02M MgCl<sub>2</sub>, 22% poly (acrylic acid) 5100 and

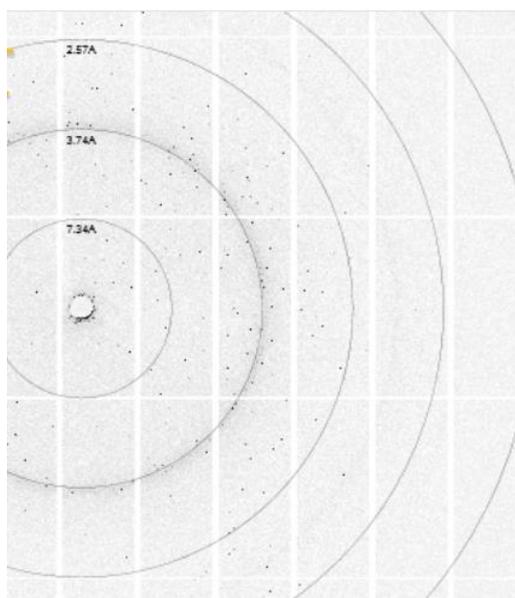
20% (v/v) glycerol. The protein crystals were harvested using appropriately sized cryo loops (Hampton research), then flash frozen under a stream of boiled off liquid nitrogen at 100 Kelvin prior to data collection. X-ray data from the first crystal was collected at Diamond Light Source using the I04 beamline. A wavelength of 0.97Å was used. In total, 180 images were collected with a 0.5° oscillation width. The crystal produced a 4.2Å diffraction data set. However, no ligand was seen bound to the enzyme. This might have been because the crystal was harvested into a cryo-buffer without the ligands, allowing dissociation. Accordingly, subsequent experiments incorporated ligands into the harvesting cryo-buffer. In order to maximise the probability of binding ligands, to limit osmotic shock and to “relax” them after being moved into cryo-buffer, the next crystals first had 4µl of cryo-buffer added into crystal drop, and then they were incubated in the presence 10mM ATP Mg<sup>2+</sup>-4mM aspartate for 1 hour. Representative crystals of MtbSAICAR along with its ligand complexes are shown in Figure (52).



**Figure (52)** Typical crystals of MtbSAICAR. Initial crystals were produced using JCSG+ commercial screen (A) From 0.1M HEPES pH7.5, 0.02M MgCl<sub>2</sub>, and 22% poly (acrylic acid) sodium salt 5100 (B) The crystals grown in 0.1M HEPES pH8.0, 0.02M MgCl<sub>2</sub>, and 22% polyacrylic acid sodium salt 5100. (C) Co-crystallised and rested in 10mM ATP Mg<sup>2+</sup>- 4mM aspartate, crystal size of 0.1mm.

### 4.3. Data Collection and analysis of MtbSAICAR-ATP complex.

A complete crystallographic data set was collected from crystals prepared with ATP and Asp using the I24 microfocus beamline macromolecular X-ray crystallography (MX) facility at Diamond light source (DLS). An example of the diffraction is given in Figure (53). A wavelength of 0.96Å was used with an oscillation width of 0.10°. The crystal diffracted to 1.69Å with a space group of P1 21 1. All data processing steps were performed by Xia2 (Winter, 2009), (an automated data reduction pipeline used during data collection at DLS), the data collection statistics are given in Table (8).



**Figure (53)** A diffraction pattern image of MtbSAICAR-ATP complex.

**Table (8)** Data collection and reduction parameters for MtbSAICAR-ATP complex

Data Collection	
<b>Cell Dimensions</b>	
<i>a, b, c</i> (Å)	51.32 99.89 74.09
<i>α, β, γ</i> (°)	90 110.20 90
<b>Space group</b>	P 1 21 1
<b>Resolution</b>	1.69 (1.73-1.69)*
<b>Completeness</b>	98.0 % (89.2)*
<b>Rmerge (%)</b>	7.3 (68.7)*
<b>I/σI</b>	10.70 (1.60)*
<b>Redundancy</b>	3.4 (3.3)*

\*The outer shell values are shown in parenthesis

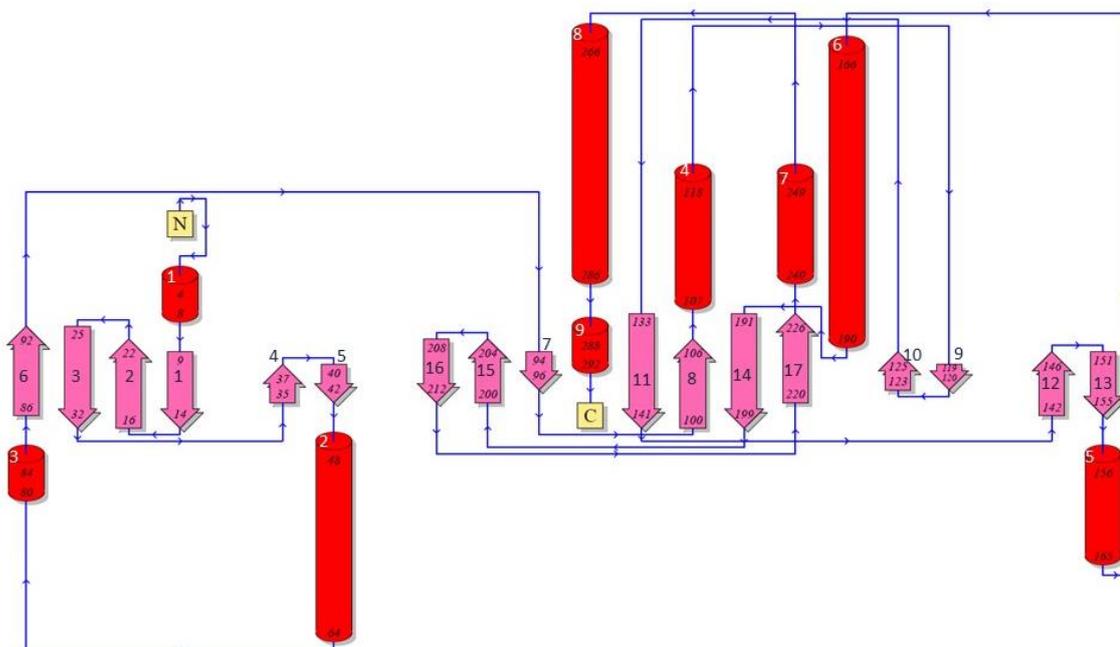
#### 4.4. Structure determination of MtbSAICAR-ATP complex

The crystal structure of the previously solved SAICARs of *M. abscessus* (MaSAICAR) PDB entry 3R9R (Baugh et al., 2015) was used as the basis for the search model for molecular replacement. Using Chainsaw (Stein, 2008) a modified model was built using the coordinates and sequence alignment file prepared using Clustal Omega (Sievers et al., 2014). The output model was then used as a search model for molecular replacement. PHASER (McCoy et al., 2007) requires for molecular replacement the coordinate file (the search model) and reflection file (MTZ) from data collection as well as an estimate of solvent content. The solvent content was calculated by Matthew's method (Matthews, 1968), using the molecular weight of the protein and the cell parameters from data collection, to be approximately 47%. The solution from PHASER, (the phased reflection file (mtz) and coordinate files) was opened with the molecular structure visualisation program COOT (Emsley and Cowtan, 2004). The examination of these preliminary electron density maps showed clear positive difference (Fo-Fc) electron density consistent with ATP being bound (although the phosphates appeared to be in more than one conformation) but no density was seen for aspartate. The ATP molecule was then built into that electron density, and the rest of the model rebuilt using the 2Fo-Fc and difference electron densities as a guide. The model was then subjected to restrained refinement using Refmac5 (Murshudov, Vagin and Dodson, 1997). After refinement, the coordinate and reflection file (phased on the refined model) were reopened with COOT and although the ATP had refined well into density there was a positive peak close to the phosphate group of ATP. Mg<sup>2+</sup> ions were fitted into that density and further refinement was performed. After refinement, the absence of difference density confirmed the presence of ATPMg<sup>2+</sup> in the structure. At the final cycle of the refinement, water molecules were added automatically (Emsley and Cowtan, 2004) and were analysed interactively. A water molecule was rejected if the B-factor was higher than 60Å<sup>2</sup> and the distance between hydrogen bond (H-bond) donor and acceptor atoms was greater than 3.6Å, or if two water molecules were less than 2.5Å away from each other. The stereochemical quality of the refined MtbSAICAR-ATP complex was validated using Procheck (Laskowski et al., 1993). Most of the residues (97.9%) were found to be the favored region of the Ramachandran plot. The statistics for refinement are given in Table (9). Briefly, the structure of the

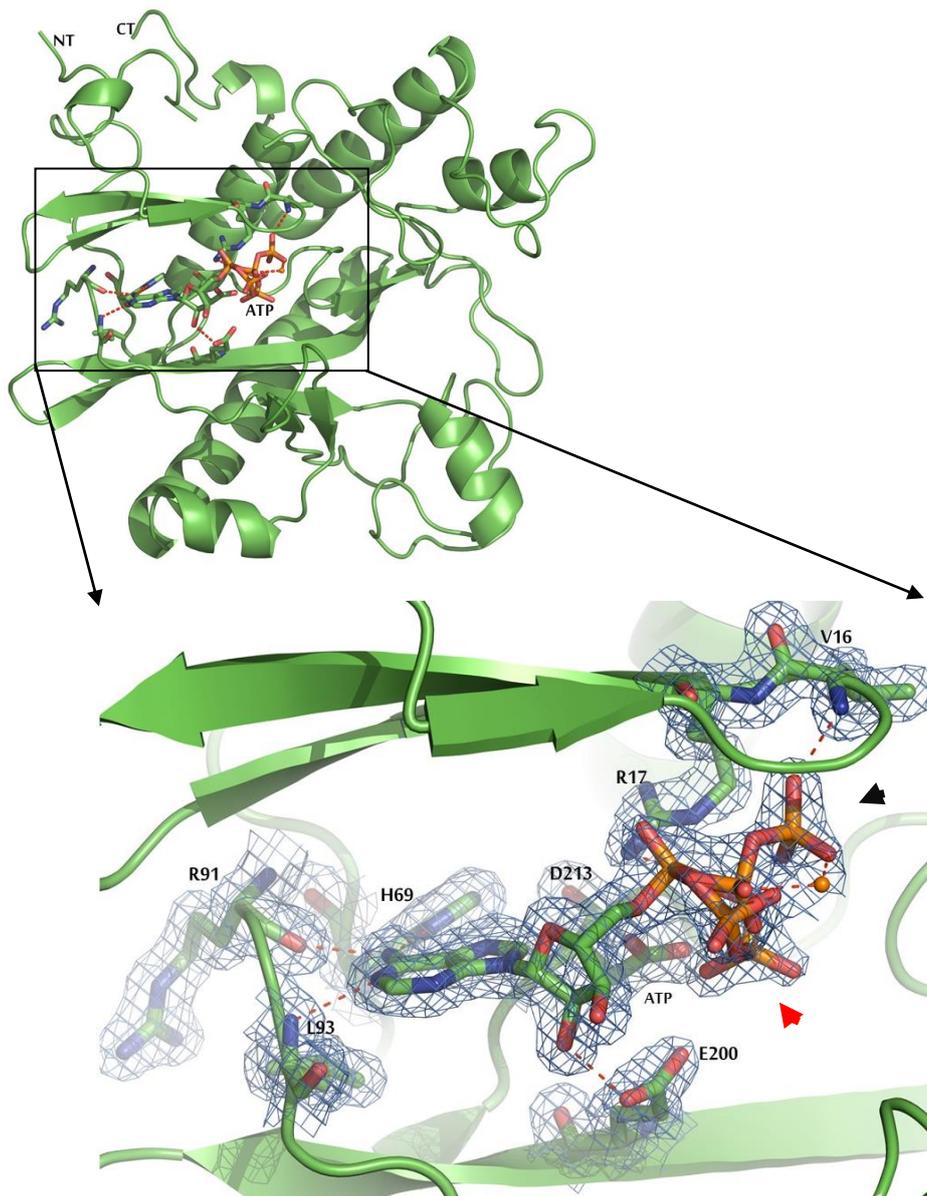
MtbSAICAR-ATP complex contains 9  $\alpha$ -helices and 17  $\beta$ -strands as shown in Figure (54). The crystal structure of MtbSAICAR consists of one molecule per asymmetric unit. One molecule of ATP and  $Mg^{2+}$  was bound to the monomer in the active site cleft, which is located between  $\beta$ -strands surrounded by  $\alpha$ -helices (Figure 55).

**Table (9)** Final refinement parameters for MtbSAICAR-ATP complex

Refinement	
Resolution (Å)	1.69
Number of reflections	73122
R factor	0.17
R free	0.21
R.M.S	
Bond Length (Å)	0.02
Bond Angle (°)	1.99
No. atoms	
Protein	2325
Water	906
Overall B-factors (Å <sup>2</sup> )	23.26



**Figure (54)** Topology diagram of the secondary structure of MtbSAICAR-ATP complex. The structure is made up of 9  $\alpha$ -helices (in red), and 17  $\beta$ -strands (in pink). The identity (number) of residues is labelled at each end of the secondary structure elements. The figure is drawn with PDBsum (Laskowski et al., 1997).



**Figure (55)** Structure of MtbSAICAR-ATP complex showing the residues involved in binding ATP. The overall fold and secondary structure are shown as a green cartoon. Selected residues and ATP are shown in stick form along with electron density (in blue  $2F_o-F_c$  map contoured at  $1\sigma$ ). The phosphate groups of ATP is shown in both linear and bent positions (black and red arrows respectively). NT and CT show the N and C-terminal. The figure was drawn with Pymol (DeLano, 2002).

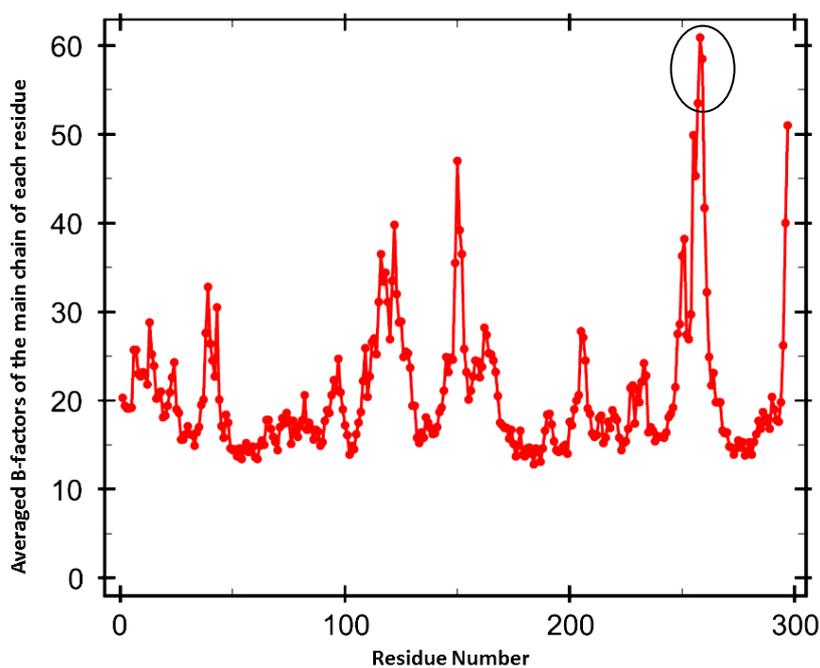
#### 4.5. Structure of the MtbSAICAR-ATP complex

For any interpretation based on structure, the accuracy of the atomic positions must be considered. The average accuracy of atomic coordinates in a refined X-ray crystal structure is reflected in the quality and quality of the X-ray data, and the agreement between the data and the model. That is, it depends on resolution and R-factors. The relationship between these was first described by Luzzati (Luzzati, 1952). However, the average is of little value when considering specific parts of a model. When data of sufficient resolution is available, it is possible to estimate individual atomic positional uncertainties using programs such as SHELX (Sheldrick, 2015). The scattering of X-rays from an individual atom depends on the apparent (or real) thermal motion of that atom. The averaging over  $10^{12}$  to  $10^{15}$  protein molecules in the crystal means that the variations in position of an atom are observed as the same as motion, that can be considered as thermal motion.

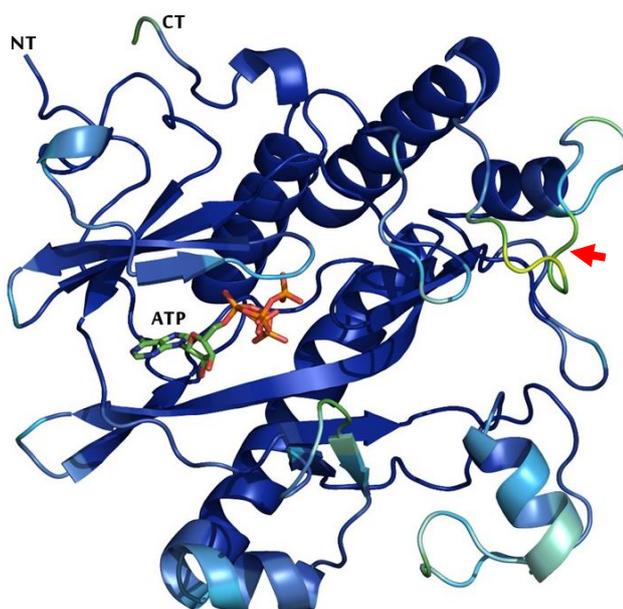
The term “temperature factor” is used because the vibration of an atom (or molecule) is dependent on its temperature, the displacement seen is a combination of the actual thermal displacement of any given atom and the differences in position between the equivalent atoms throughout the crystal, the observed “temperature factor” therefore reflects both static and dynamic disorder. The observed displacement depends on various factors and is reflected in the resolution of the data that can be collected from the crystal (increased B-factors means less order and poor resolution) and is affected by packing. The movement about the average position is evaluated by calculation of “temperature factors” (B-factor) describing the displacement of atoms using this equation  $B_j = 8\pi^2 \langle r^2_j \rangle$ , where  $r^2$  is the displacement (of atom j) from its mean position (Yang, Wang and Zhang, 2016). At worse than atomic resolution the B-factors may be under-determined and influenced by the process of refinement (Yang, Wang and Zhang, 2016). Expected B-factors are expected to range from 15-50Å<sup>2</sup> for main chain atoms. Atomic B-factors are a useful indicator of how well ordered individual atoms are, and therefore reflect confidence in their positions

Prior to detailed analysis of the structures, the B-factor profiles were calculated. The B-factor reflecting the uncertainty of an atom's position about its average. Strictly, it shows the magnitude of vibration averaged over the copies within the crystal structure; as such it may be correlated with the functions of regions of a protein. Regions of higher B-factors may have greater flexibility, and this may be required for catalytic function (for example). In the case of the MtbSAICAR-ATP complex, averaged B-factors of the main chain atoms of almost all residues are well ordered and within the standard range ( $15-50\text{\AA}^2$ ) (Figure 56A). A few residues near the C-terminal site, as indicated in the graph below, have higher B-values than the others, this suggests flexibility in the loop they form. The residues that form the ligand binding regions are well ordered and their residues have a low B-factor, indicated by blue colour in (Figure 56B).

(A)



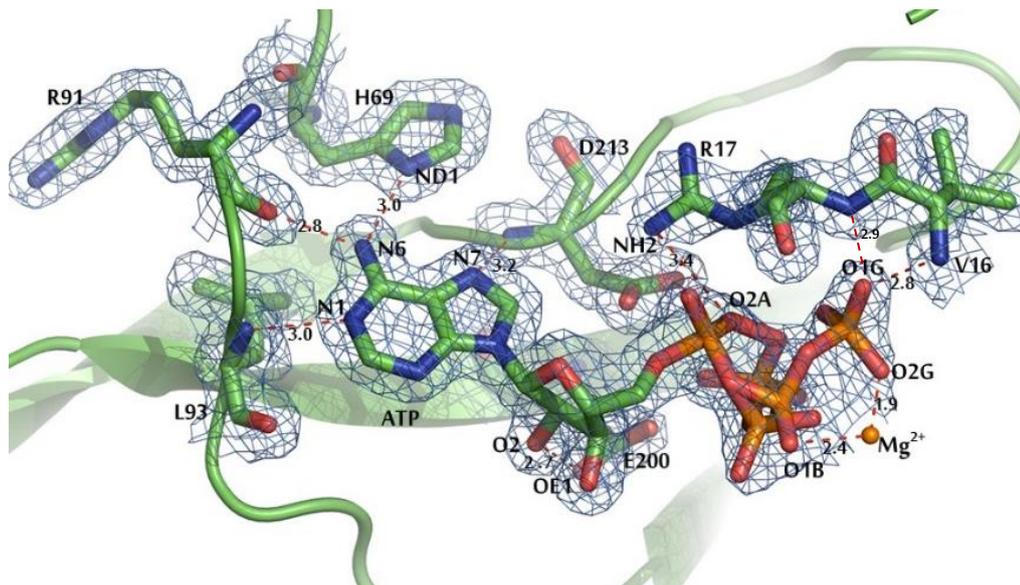
(B)



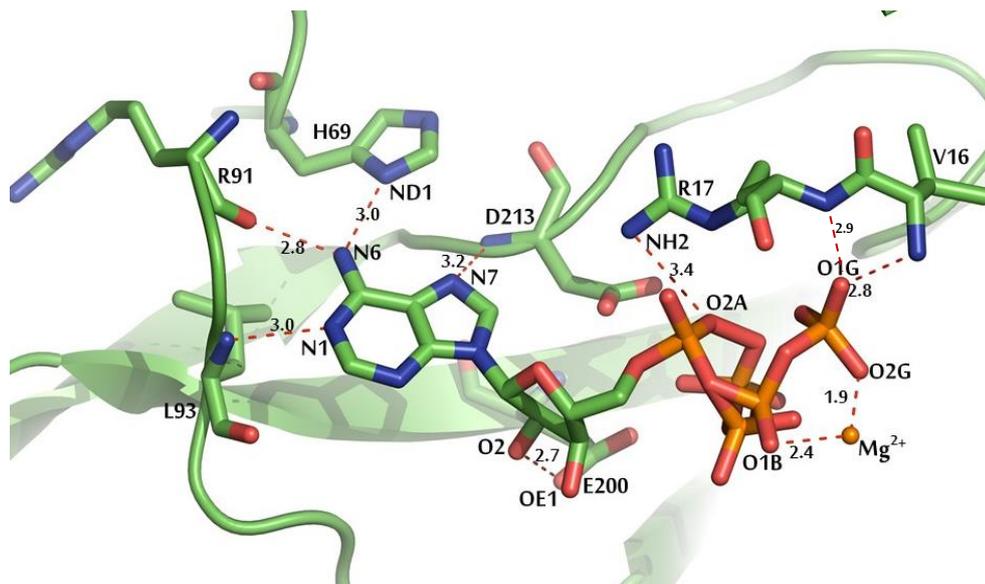
**Figure (56)** Refined B-factors for main chain atoms of the MtbSAICAR-ATP complex. (A) B-factors plotted by sequence of the MtbSAICAR-ATP complex. Almost all of the main chain atoms are within standard range ( $15\text{-}50\text{\AA}^2$ ) and consider rigid, this may be caused by the presence of the ligand. A few residues (S258 and D259) near the C-terminal have higher B-values as indicated by a circle. (B) Three-dimensional structure showing the B-factors, coloured from dark blue (low) to lemon (moderate) respectively. The residues with higher B-values are indicted by the red arrow. The region where the ligand is bound is well ordered as indicated by blue colour. The ligand (ATP) position is indicated as a stick representation. Graph is drawn using Baverage utility in CCP4.

The ATP molecule is located in the active site cleft and appears to be in the anti-conformation. The interactions of ATP with the enzyme include hydrophobic interactions, H-bonds, and electrostatic interactions. These are shown in Figure (57A and B). The adenine group is located in the hydrophobic pocket, surrounded by G14, I19, L27, V90, M95, A212, and L148. The adenine ring also makes specific interactions with the main and side chain atoms of the enzyme; N1 of the adenine ring makes an H-bond with main chain N of L93, N7 of adenine makes H-bond with the main chain N of D213, and N6 of adenine makes an H-bond with main chain O of R91 and the side chain ND1 of H69. The C2'OH (labelled O2) of the ribose makes an H-bond with the carboxylic acid group of E200. O2A of the  $\alpha$ - $\beta$  phosphate link forms an ionic interaction with the guanidinium R17, O1G makes H-bond with main chain N of R17 and the main chain N of V16.  $Mg^{2+}$  interact with water as well as the  $\beta$  and  $\gamma$ -phosphate groups of ATP. A LIGPLOT diagram (Laskowski et al., 1997) of the ATP binding site is given in Figure (58).

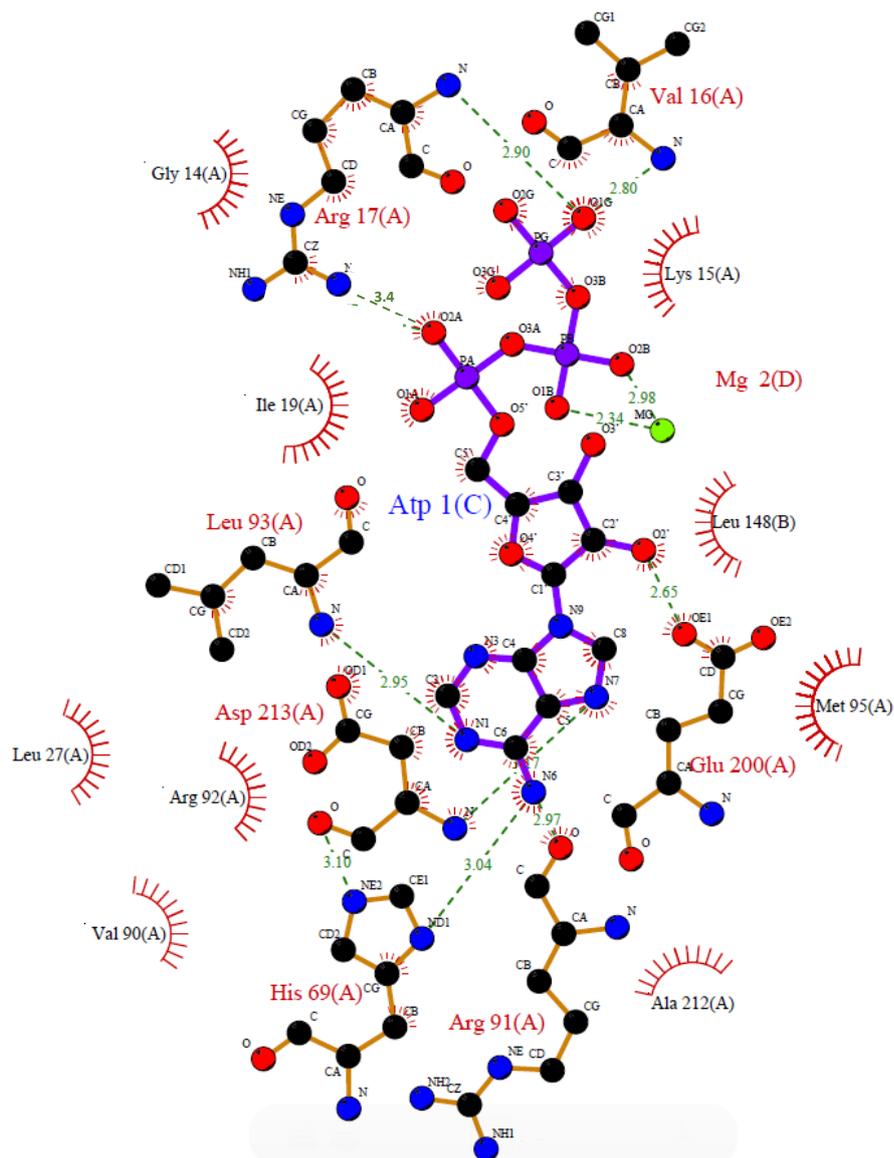
(A)



(B)



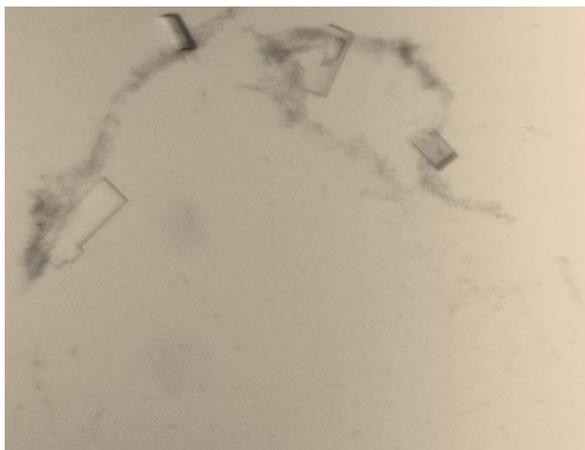
**Figure (57)** Key residues binding to ATP in MtbSAICAR (A) The ATP binding site is shown inside the electron density map. (B) without electron density and showing measured distances in Å.  $Mg^{2+}$  is shown as an orange sphere. Residues interacting with the ligand are shown as sticks. The distances are given in Å. The map in (A) is contoured in blue at  $1\sigma$  for 2Fo-Fc map.



**Figure (58)** ATP binding site of MtbSAICAR drawn using LIGPLOT, PDBsum (Laskowski et al., 1997). Distances are given in Å.

#### 4.6. Crystallisation of apo MtbSAICAR

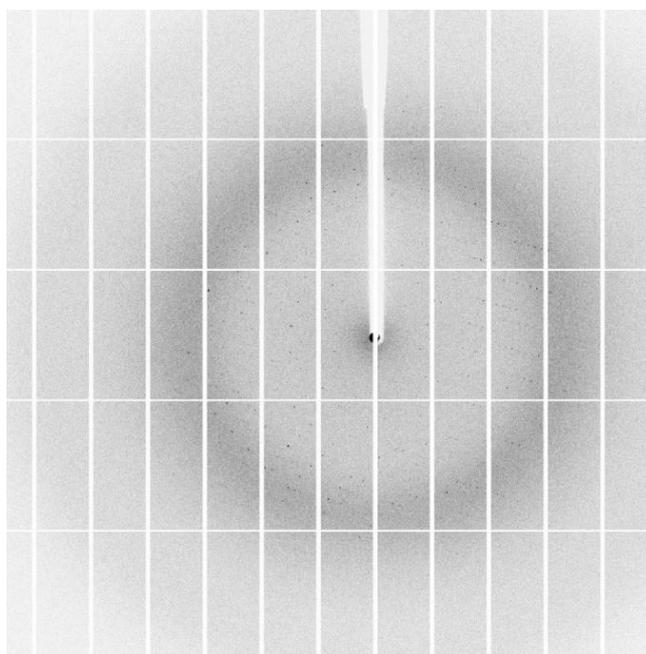
The crystallisation of apo MtbSAICAR was achieved using a mother liquor consisting of 0.1M HEPES pH8.0, 0.02M MgCl<sub>2</sub>, 22% poly (acrylic acid) 5100 as before. Plates were monitored for crystal growth at room temperature. The first crystal growth was observed the next day with a crystal size of 0.1mm (Figure 59). Crystals were prepared for cryocrystallography by transferring them to a cryo mixture of the mother liquor incorporating 20% (v/v) glycerol. The crystals were harvested using appropriately sized cryo loops in the cryoprotectant and flash frozen prior to data collection.



**Figure (59)** Typical crystals of apo MtbSAICAR.

#### **4.7. Data Collection and analysis of apo MtbSAICAR**

A complete crystallographic data set was collected using the macromolecular X-ray crystallography (MX) I04-1 beamline at Diamond light source (DLS). A diffraction pattern is shown in Figure (60). A wavelength of  $0.92\text{\AA}$  was used and data collected with an oscillation width of  $1.0^\circ$ . The crystal diffracted to  $2.29\text{\AA}$  with space group P1 21 1. All data processing steps were performed by Xia2 (Winter, 2009). The data collection and reduction statistics are given in Table (10).



**Figure (60)** Diffraction pattern of apo MtbSAICAR. 360 images of  $1^\circ$  were collected from a single crystal at  $0.92\text{\AA}$ .

**Table (10)** Data collection parameters of apo MtbSAICAR

Data Collection	
<b>Cell Dimension</b>	
<i>a,b,c</i> (Å)	52.11 100.32 72.59
<i>α, β, γ</i> (°)	90 109.79 90
<b>Space group</b>	P 1 21 1
<b>Resolution</b>	2.29 (2.33-2.29)*
<b>Completeness</b>	99.11 (99.49)*
<b>Rmerge (%)</b>	16.3 (162.8)*
<b>I/σI</b>	11.1 (1.2)*
<b>Redundancy</b>	6.9 (7.0)*

\*Outer shell values are shown in parenthesis

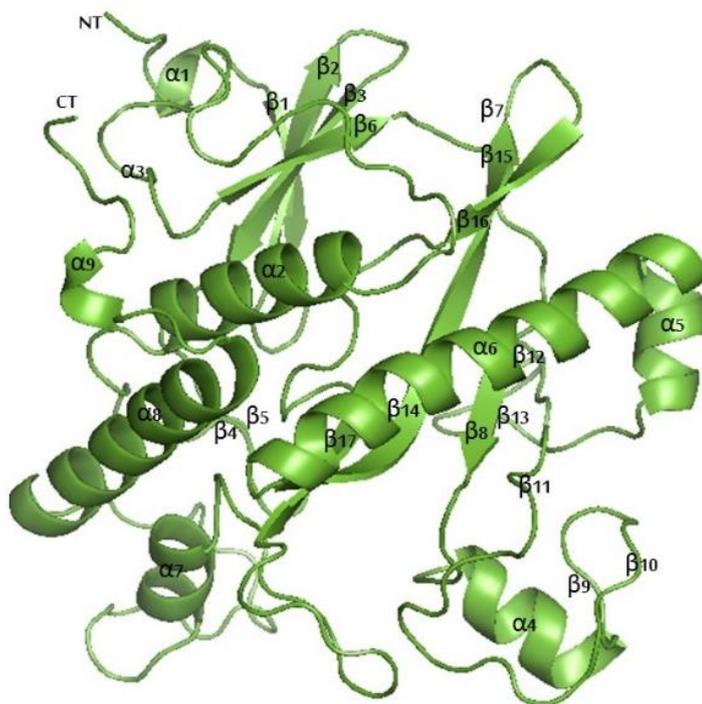
#### 4.8. Structure determination of apo MtbSAICAR

The crystal structure of the previously refined MtbSAICAR-ATP complex was used as a search model (without ATP) for molecular replacement in PHASER (McCoy et al., 2007). Although the space group is the same and the cell dimensions are similar to the ATP complex, PHASER was used to correctly orient the model in the crystal. After PHASER, the output maps and coordinate files were opened with the structure visualisation program COOT (Emsley and Cowtan, 2004). The coordinates were adjusted manually to correct the position of atoms where necessary. The structure was subjected to further refinement (restrained refinement) using Refmac5 (Murshudov, Vagin and Dodson, 1997) with the addition of water molecules. The final refinement parameters are given in Table (11).

**Table (11)** Refinement parameters of apo MtbSAICAR

Refinement	
<b>Resolution (Å)</b>	2.29
<b>Number of reflections</b>	29819
<b>R factor</b>	0.20
<b>R free</b>	0.26
<b>R.M.S</b>	
<b>Bond Length (Å)</b>	0.02
<b>Bond Angle (°)</b>	1.97
<b>No. atoms</b>	
<b>Protein</b>	2325
<b>Water</b>	136
<b>Overall B-factors (Å<sup>2</sup>)</b>	48.71

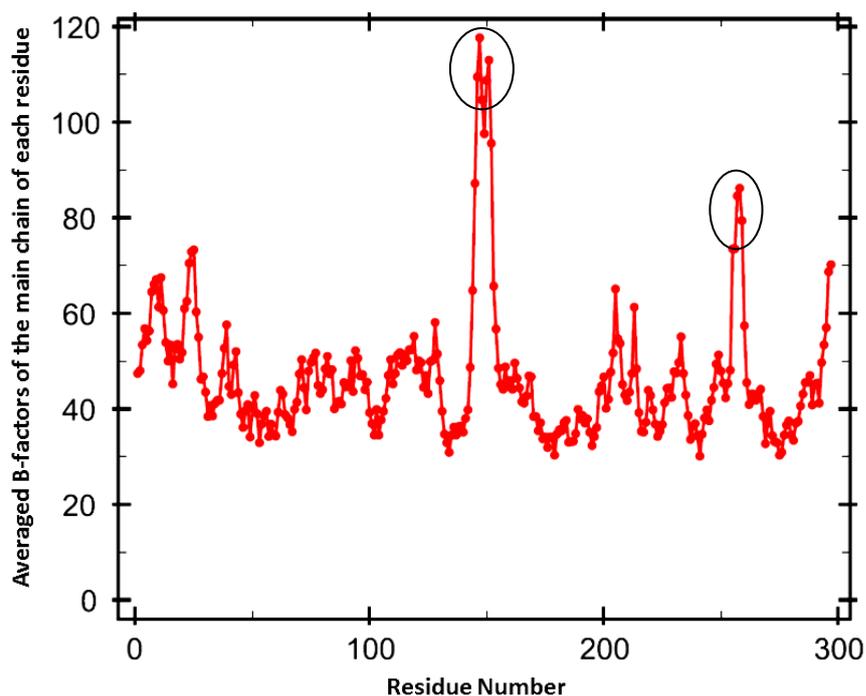
The overall structure of MtbSAICARs apo form is very similar to previously solved MtbSAICAR-ATP complex (Figure 61). The stereochemical quality of the apo structure was validated using Procheck (Laskowski et al., 1993), most of the residues (90.1%) are in the favored regions of the Ramachandran plot.



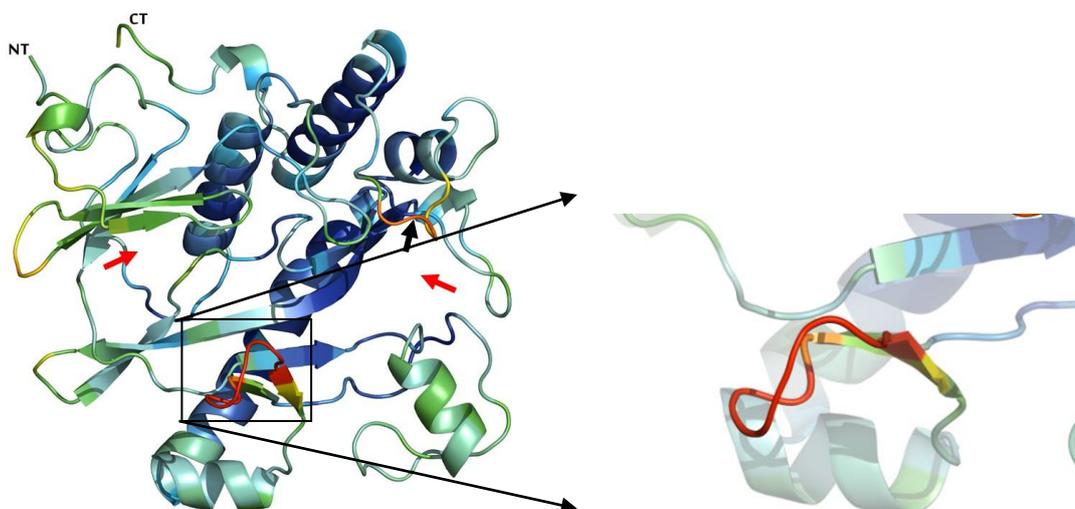
**Figure (61)** Unliganded (Apo) structure of MtbSAICAR. The structure consists of 9  $\alpha$ -helices, 17  $\beta$ -strands, the secondary structure is labelled according to topology generated by PDBsum (Figure 53) (Laskowski et al., 1997).

The B-factor profile has also been calculated for apo structure and reveals that many of the residues exceed the standard B-factor range ( $15-50\text{\AA}^2$ ). The average B-factor is higher than the average B-factor values of the ATP complex, reflecting the lower resolution data, with some main chain atoms having very high B-factors (exceeding  $100\text{\AA}^2$ ). The overall profile is similar to that of the ATP complex, except for a relatively higher peak around residue 150, Figure (62A-62B). Figure (58) shows that in the complex L148 forms part of the ribose-binding region and therefore might be less constrained in the apo structure. Unlike the MtbSAICAR-ATP complex, the apo structure is less ordered, and the binding of the ligand reduces mobility. The B-factors are shown by colour in the structure cartoon of Figure (62B).

(A)



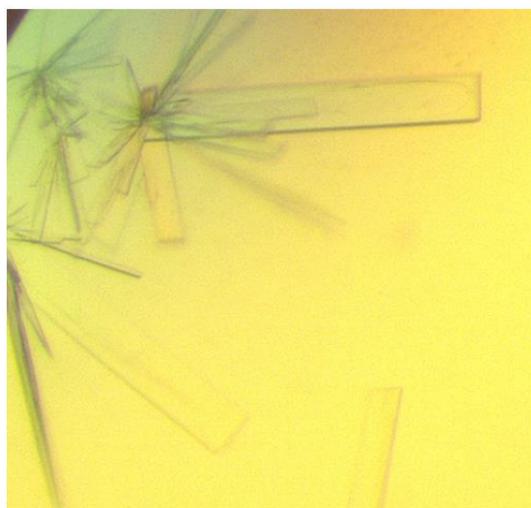
(B)



**Figure (62)** B-factors of the main chain of apo MtbSAICAR. (A) Averaged B-factor the main chain atoms plotted by residue. Residues with B-factor exceeding  $60\text{\AA}^2$  (K145-E153) and (D255-D259) form part of loops and are indicated by circles. (B) Three-dimensional structure cartoon showing B-factors as a colour spectrum, with low values in blue and high values in red. The segment (145-153) is shown enlarged and the region (D255-D259) is indicated by black arrow. The absence of bound ligand might be the reason behind the increased disorder of this form. Expected ligand positions are indicated by red arrows. Graph is drawn using Baverage utility in CCP4.

#### 4.9. Crystallisation of MtbSAICAR-ADP-AMZ complex

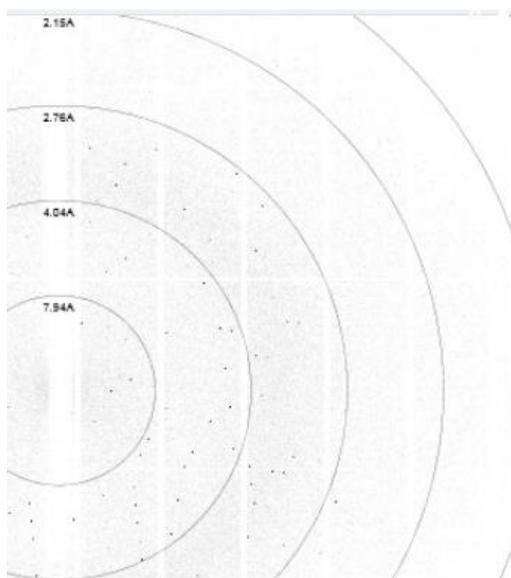
The crystallization of MtbSAICAR – ADP - {5-aminoimidazole -4 – carboxamide ribonucleotide (AMZ)} complex was set up using 0.1M HEPES pH8.0, 0.02M MgCl<sub>2</sub>, 22% poly (acrylic acid) 5100 as before, but also included 10mM ADP and 10mM AMZ. Plates were monitored for crystal growth at room temperature. The first crystal growth was observed the next day with a crystal length of 0.3mm (Figure 63). Crystals were prepared for cryocrystallography by transferring them into 4μl of cryo mixture (mother liquor with 20% glycerol and 10mM ADP and 10mM AMZ) and left for 1 hour. The crystals were then harvested using appropriately sized cryo loops and plunged into liquid nitrogen prior to data collection.



**Figure (63)** A crystal of MtbSAICAR-ADP-AMZ complex, grown and soaked in 10mM ADP and 10mM AMZ for 1 hour.

#### 4.10. Data Collection and analysis of MtbSAICAR-ADP-AMZ complex

A complete crystallographic data set was collected using the macromolecular X-ray crystallography (MX) facility at Diamond light source (DLS). The I04 beamline was used and a diffraction image is shown in Figure (64). Diffraction data were collected to 1.77Å. A wavelength of 0.97Å was used with an oscillation width of 0.1° over 180°. All data processing steps were performed by Xia2 (Winter, 2009), the data collection statistics are given in Table (12).



**Figure (64)** 0.1° Diffraction pattern from MtbSAICAR-ADP-AMZ complex crystal.

**Table (12)** Data collection and reduction parameters for MtbSAICAR-ADP-AMZ complex

Data Collection	
<b>Cell Dimension</b>	
<i>a, b, c</i> (Å)	50.88 100.44 73.69
<i>α, β, γ</i> (°)	90 110.14 90
<b>Space group</b>	P 1 2 1
<b>Resolution</b>	1.77 (1.82-1.77)*
<b>Completeness</b>	99.8 (98.7)*
<b>Rmerge (%)</b>	4.7 (62)*
<b>I/σI</b>	17.10 (1.80)*
<b>Redundancy</b>	3.4 (3.3)*

\*Outer shell values are shown in parenthesis

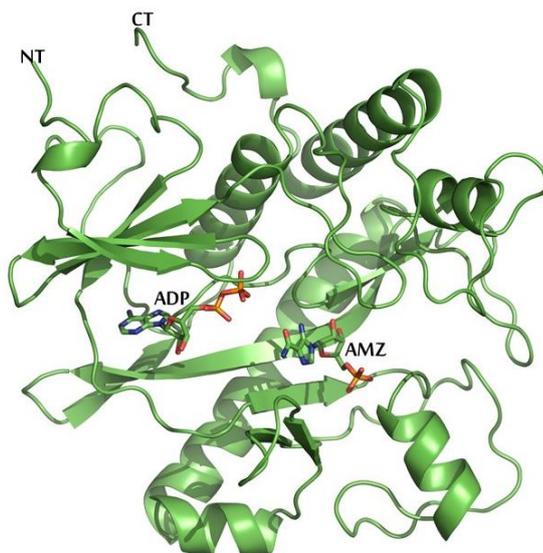
#### 4.11. Structure determination of MtbSAICAR-ADP-AMZ complex

As before, the crystal structure of the previously refined MtbSAICAR-ATP complex (without ATP) was used as a search model for molecular replacement in PHASER (McCoy et al., 2007). After PHASER, the output maps and coordinate files were opened with the structure visualisation program COOT (Emsley and Cowtan, 2004). The presence of a bound ADP molecule at active site cleft as well as a molecule of CAIR (AMZ) was clearly seen with positive difference electron density corresponding to the shape of these ligands. ADP and AMZ were fitted manually into the appropriate

electron densities. The enzyme structure with ligands was then refined (restrained refinement) using Refmac5 (Murshudov, Vagin and Dodson, 1997). The coordinates were adjusted manually in COOT to correct the position of ligands and atoms of residues when necessary. A positive electron density peak was still apparent around the phosphate of ADP and this was surrounded by three water molecules. The density analysis and bond distance suggested it could be Mg<sup>2+</sup> and hence Mg<sup>2+</sup> was placed into this density. After the model was subjected to another round of refinement with the addition of waters, all positive difference densities around the binding site had disappeared. The refinement statistics are shown in Table (13). The stereochemical quality of the structure was validated using Procheck (Laskowski et al., 1993), most of the residues (92.1%) are in the favored regions of the Ramachandran plot. The structure is similar to the other MtbSAICAR structures. The positions of the ligands in the active site cleft shown in Figure (65).

**Table (13)** Refinement parameters of MtbSAICAR-ADP-AMZ complex

<b>Refinement</b>	
<b>Resolution (Å)</b>	1.77
<b>Number of reflections</b>	64072
<b>R factor</b>	0.16
<b>R free</b>	0.19
<b>R.M.S</b>	
<b>Bond Length (Å)</b>	0.02
<b>Bond Angle (°)</b>	2.13
<b>No. atoms</b>	
<b>Protein</b>	2325
<b>Water</b>	706
<b>Overall B-factors (Å<sup>2</sup>)</b>	25.65



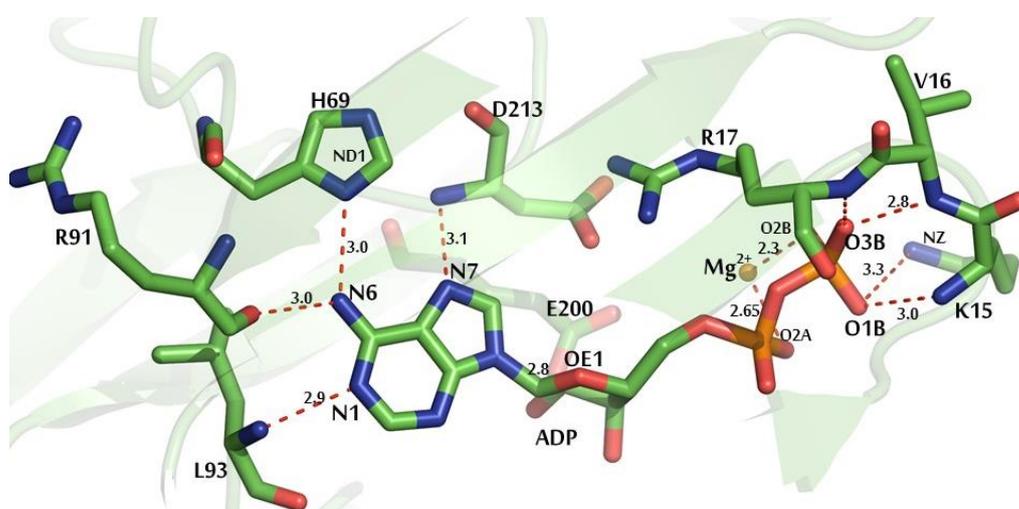
**Figure (65)** Cartoon representation of MtbSAICAR-ADP-AMZ complex showing the ADP and AMZ binding sites. N and C-terminal are labelled NT and CT. The active site is located within  $\beta$ -strands are surrounded by  $\alpha$ -helices.

#### 4.12. Overall Structure of MtbSAICAR-ADP-AMZ complex

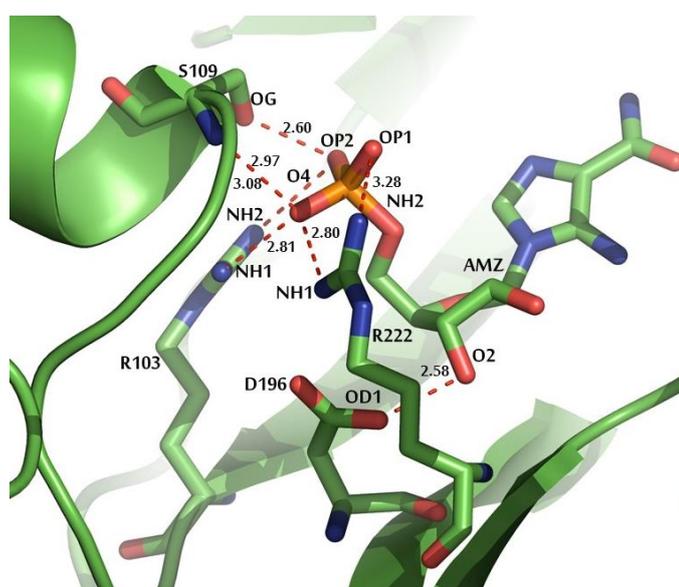
The structures described here only show one AMZ and one ADP bound to the protein, unlike the crystal structure of SAICARs of *S. cerevisiae* (ScSAICAR) complex which shows additional binding at what is presumed to be a peripheral non-enzymatic site (Chapter5, Section5.3.2.). ADP binding in the active site cleft is stabilised by extensive H-bonds, electrostatic, and hydrophobic interactions. Overall the ADP binding is very similar to that of ATP, the adenine base is surrounded by hydrophobic residues including G14, I19, L27, V90, M95, and A212. The adenine base forms specific H-bonds, the N1 atom with the main chain N of L93, the N6 atom with the side chain of H69 at ND1 and the main chain O of R91 and the N7 atom with main chain N of D213. The C2'OH (O2) of ribose makes H-bonds with the side chain of carboxylic acid E200, O3B of the  $\beta$ -phosphate group forms H-bonds with main chain V16N and R17N, O2B of  $\beta$ -phosphate makes ionic interactions with  $Mg^{2+}$ . O1B of the  $\beta$ -phosphate also makes ionic interaction with side chain NZ of K15 and H-bonds with main chain N of K15.  $Mg^{2+}$  makes interactions with the atoms of ADP and water forming H-bonds with four water molecules, O2A and O2B of the  $\beta$ -phosphate group of ADP (Figure 66A). The substrate analogue AMZ binds in close proximity to ADP. AMZ (Figure 66B) makes H-bonds with several atoms. The O4 atom of the phosphate group forms ionic interactions with guanidinium of R222, H-bond to main chain N of S109 and ionic interactions with

guanidinium of R103. The phosphate group also makes an H-bond with side chain OG of S109 and an ionic interaction with the guanidinium of R103. The phosphate group also makes an ionic interaction with the guanidinium of R222. The C2'OH (O2) of the ribose ring forms H-bonds with the side chain OD1 of D196. The adenine base doesn't interact directly with residues of the enzyme, instead, it makes an H-bond to a water molecule. A schematic representation of ADP, AMZ and Mg<sup>2+</sup> binding site is shown using PDBsum (Laskowski et al., 1997), and illustrated in Figure (67).

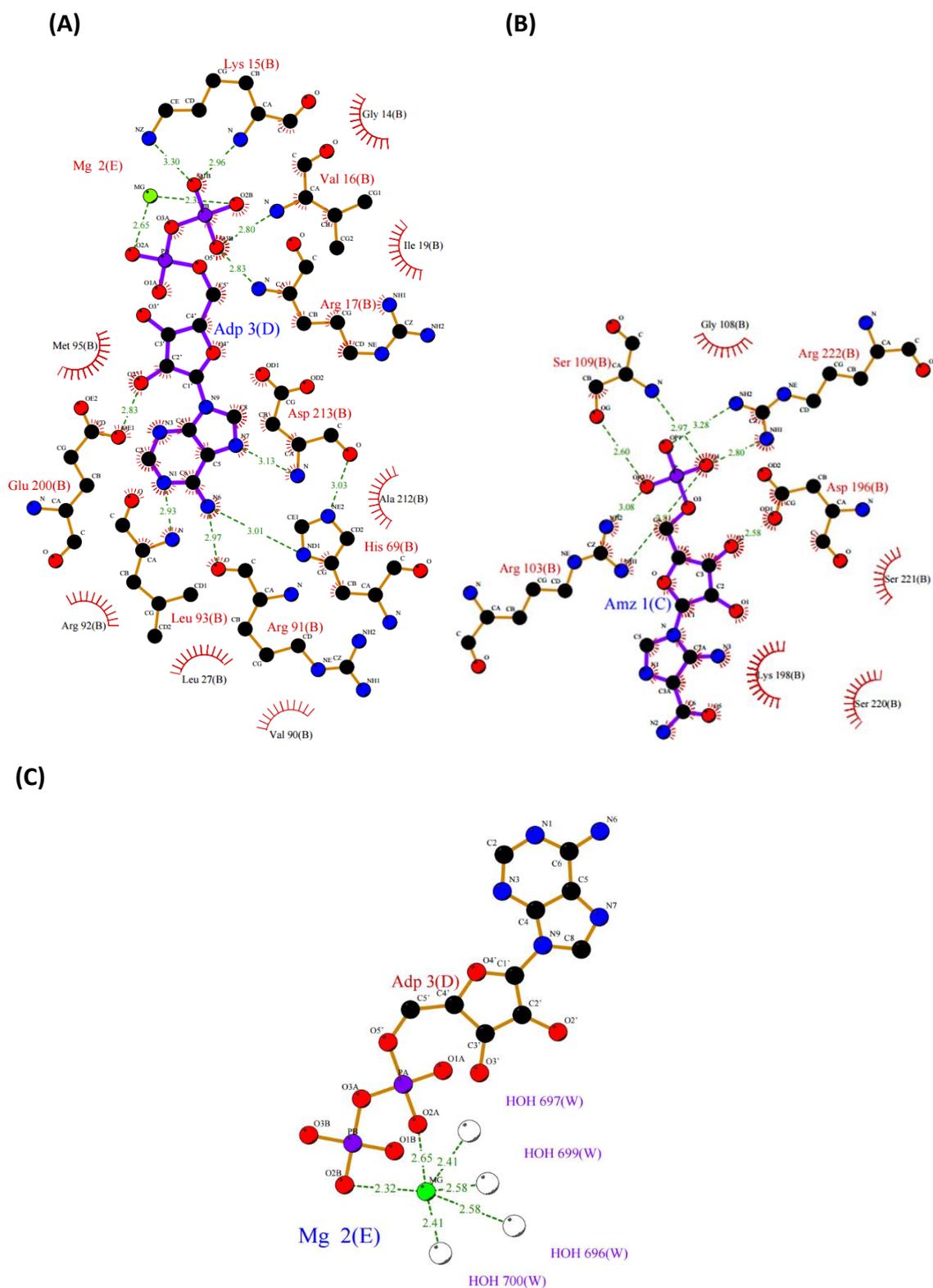
(A)



(B)



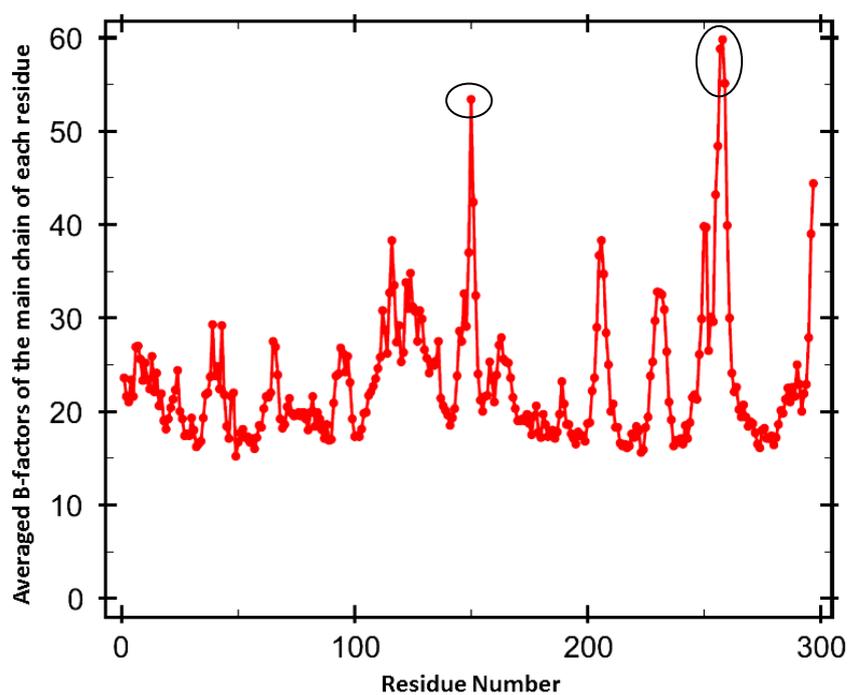
**Figure (66)** Key residues showing the mode of binding of ADP and AMZ in MtbSAICAR. (A) ADP binding site showing measured distances (B) AMZ binding sites with measured distances. The 2Fo-Fc electron density maps are coloured blue contoured at 1 $\sigma$ . Distances are given in Å.



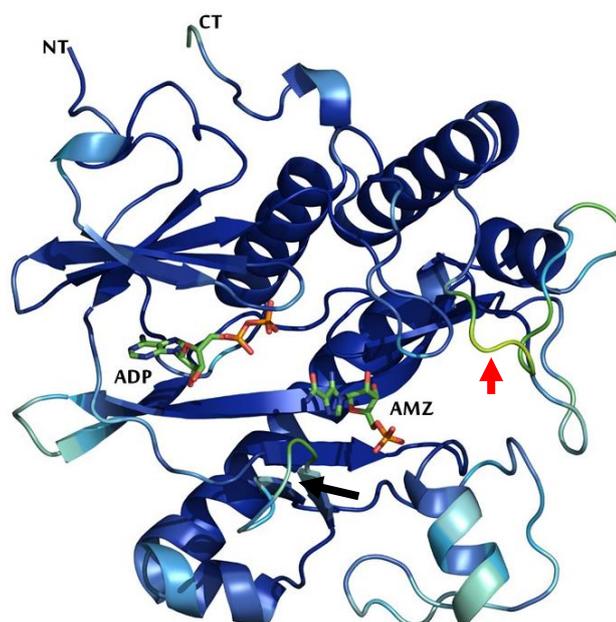
**Figure (67)** ADP, AMZ and Mg<sup>2+</sup> binding sites in MtbSAICAR-ADP-AMZ complex, (A) ADP and (B) AMZ binding site interactions. (C) Interactions of Mg<sup>2+</sup> coloured green, distance to interacting residues are labeled. The distances are given in Å. The figures were drawn using LIGPLOT.

The B-factor profile has been calculated for MtbSAICAR-ADP-AMZ complex. The averaged B-factor values of all main chain in each residue shown is within the standard range ( $15-50\text{\AA}^2$ ) with few residues (around H151) and(G257-D259) forming loops exceeding  $50\text{\AA}^2$  as indicated (Figure 68A). This peak is, however much less than the corresponding one in the apo structure, L148 forms a hydrophobic interaction in the ATP complex (Figure 58) and this may reduce the motion of the nearby H151. Suggesting the presence of ligand increases the rigidity of the area around it. The profile has peaks in similar positions to that of the ATP complex and the comparison with the apo form again shows that once liganded, the stability of enzyme increases. The regions around the ligand binding site are well ordered (low B-factors) and are shown in blue colour in Figure (68B).

(A)



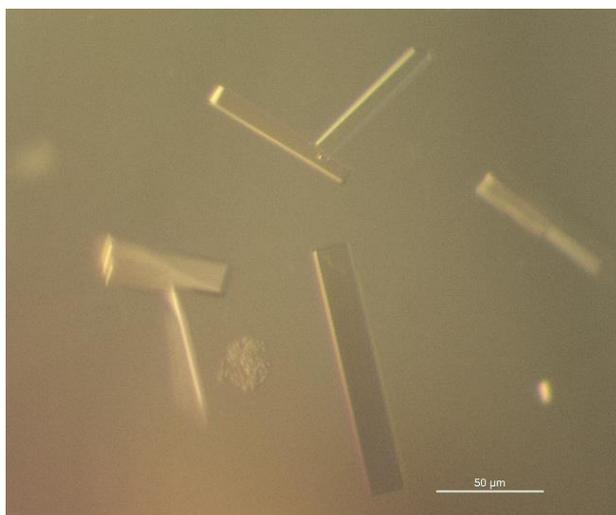
(B)



**Figure (68)** B-factors for MtbSAICAR-ADP-AMZ complex. (A) B-factor plot of main chain atoms. Residues with high B-factors are found to be parts of loops. (B) Three-dimensional ribbon cartoon showing B-factors coloured from blue (low) to red (high). Residues (H151) and (G257-D259) are part of a loop as indicated in graph (A) by a black circle, in the 3D structure (B) by black and red arrows respectively. The stability of the enzyme is increased by becoming liganded, the ligand positions are indicated. Graph is drawn using Baverage utility in CCP4.

#### 4.13. Crystallisation of MtbSAICAR-ACP-IMP complex

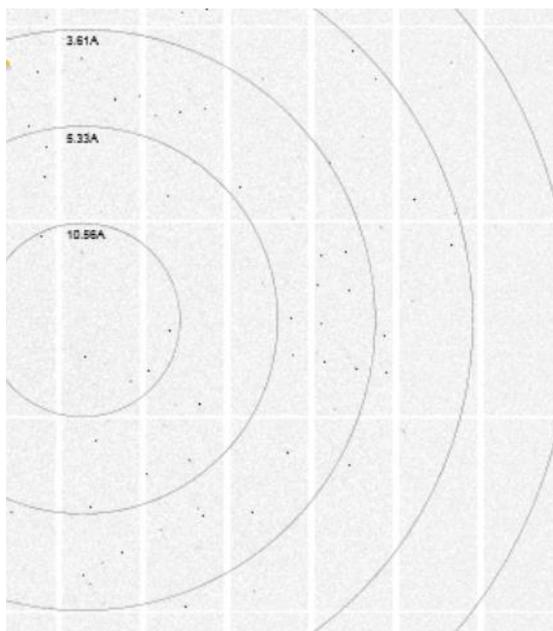
MtbSAICAR-{Adenylyl-imidodiphosphate (ACP)-inosine 5'-monophosphate (IMP)} complex crystals were grown in mother liquor consisting of 0.1M HEPES pH8.0, 0.02M MgCl<sub>2</sub>, and 22% poly (acrylic acid) 5100, 10mM of Adenylyl-imidodiphosphate (ACP) and inosine 5'-monophosphate (IMP) in crystallisation well. Plates were monitored for crystal growth at room temperature. The first crystal growth was observed the next day and crystals grew to a length of 0.4mm (Figure 69). Crystals were harvested for cryocrystallography by transferring them into 4µl cryo mixture of the mother liquor, 20% glycerol including 10mM of each of ACP and IMP and rested in for 1 hour. The crystals were then mounted using appropriately sized cryo loops and plunged into liquid nitrogen for storage prior to data collection.



**Figure (69)** A typical crystal of MtbSAICAR-ACP-IMP complex, prepared with 10mM ACP, 10mM IMP and 20% glycerol.

#### 4.14. Data Collection and analysis of MtbSAICAR-ACP-IMP complex

A complete crystallographic data set was collected using the macromolecular X-ray crystallography (MX) facility at Diamond light source (DLS). The microfocus I24 beamline was used for data collection and a diffraction image is shown in Figure (70). The crystal diffracted to 1.99Å with space group P1 21 1. A wavelength of 0.96Å was used to collect 180° data with an oscillation width of 0.1°. All data processing steps were performed in Xia2 (Winter, 2009). The data collection statistics are given in Table (14).



**Figure (70)** Diffraction pattern of MtbSAICAR-ACP-IMP complex. 1800 images were collected from a single crystal.

**Table (14)** Data collection and reduction parameters of MtbSAICAR-ACP-IMP complex

Data Collection	
<b>Cell Dimension</b>	
<i>a, b, c</i> (Å)	51.17 99.23 73.65
$\alpha, \beta, \gamma$ (°)	90.00 110.17 90.00
<b>Space group</b>	P 1 21 1
<b>Resolution</b>	1.99 (2.04-1.99)*
<b>Completeness</b>	94(68.3)*
<b>Rmerge (%)</b>	3.7 (29.6)*
<b>I/<math>\sigma</math>I</b>	13.3(1.4)*
<b>Redundancy</b>	3.1 (2.0)*

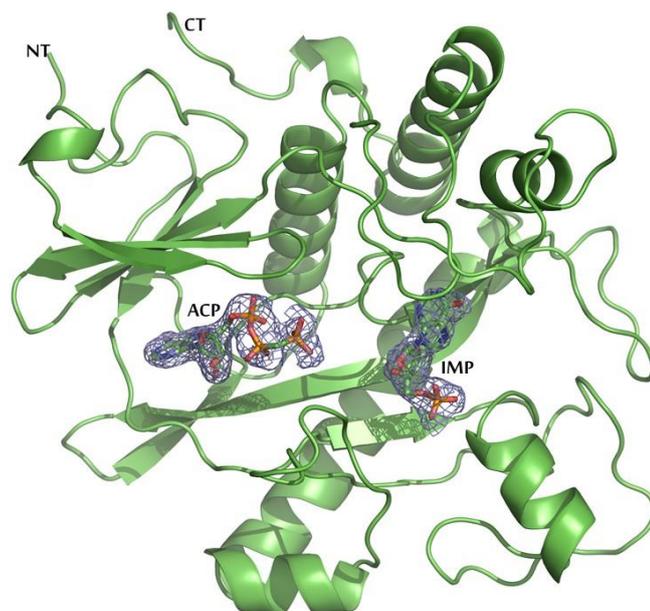
\* Outer shell values are shown in parenthesis

#### 4.15. Structure determination of MtbSAICAR in complex with competitive inhibitor ACP and IMP

As before, the crystal structure of the previously refined MtbSAICAR-ATP complex (without ATP) was used as a search model for molecular replacement in PHASER (McCoy et al., 2007). After PHASER, the output maps and coordinate files were opened with the structure visualisation program COOT (Emsley and Cowtan, 2004). The presence of ACP and IMP were clear with positive difference electron density corresponding to the shape of these ligands being visible. ACP and IMP were fitted manually into the electron densities. The structure was refined automatically (restrained refinement) using Refmac5 (Murshudov, Vagin and Dodson, 1997) after adding these ligands. The coordinates were adjusted manually in COOT to correct the position of ligands and protein atoms where necessary. The files were subjected to another refinement with the addition of waters, at the end of refinement all positive densities near the binding sites had disappeared. The final refinement parameters are given in Table (15). The overall structure is very similar to other MtbSAICAR structures (Figure 71). The stereochemical quality of the structures solved with ACP-IMP complex was validated using Procheck (Laskowski et al., 1993), most of the residues (92%) were found to be in the favored regions of the Ramachandran plot.

**Table (15)** Refinement parameters of MtbSAICAR-ACP-IMP complex

Refinement	
Resolution (Å)	1.99
Number of reflections	42528
R factor	0.17
R free	0.22
R.M.S	
Bond Length (Å)	0.02
Bond Angle (°)	2.19
No. atoms	
Protein	2325
Water	390
Overall B-factors (Å <sup>2</sup> )	44.87



**Figure (71)** Cartoon representation of MtbSAICAR-ACP-IMP complex showing ACP and IMP binding. N and C-terminal are labelled NT and CT. The active site is located within  $\beta$ -strands are surrounded by  $\alpha$ -helices.

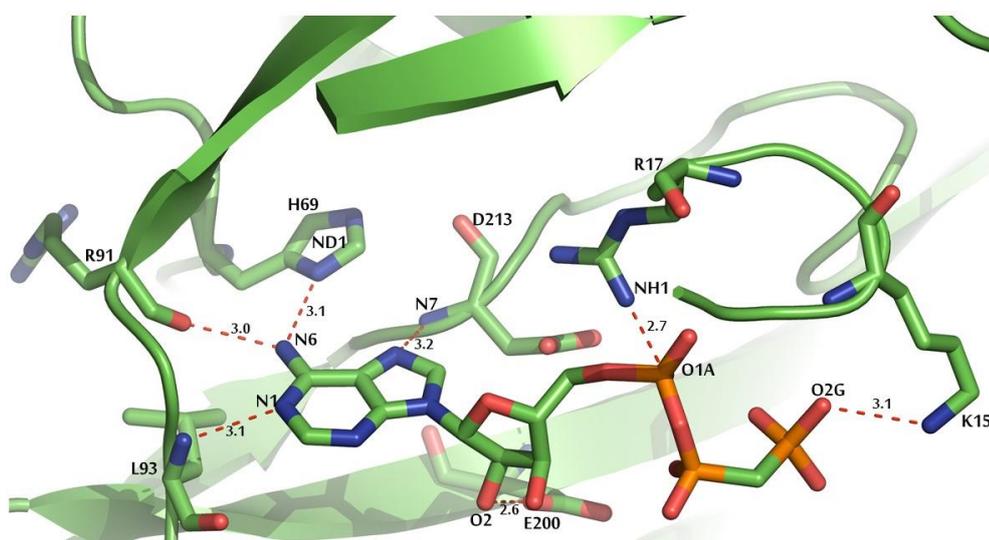
#### **4.16. The overall structure of MtbSAICAR in complex with competitive inhibitors ACP and IMP**

ACP is a competitive inhibitor of ATP (Nelson et al., 2005) and this suggests it should bind at the same site as ATP. IMP is a competitive inhibitor to CAIR (Nelson et al., 2005) and therefore competes to bind in the same place as CAIR. Figure (72 A and B) shows the interactions with ACP and IMP, both of which are located in close vicinity to each other. ACP has a similar binding site as ATP, the adenine ring of ACP is located in the hydrophobic pocket similarly to ATP and makes ionic interactions and H-bonds to several residues. The O2G  $\gamma$ -phosphate of ACP makes ionic interaction with side chain NZ of K15. The phosphate group of ACP makes an ionic interaction to guanidinium of R17, N6 of the inosine ring makes H-bond to the main chain O of R91 and side chain ND1 of H69, the N7 of the adenine ring makes an H-bond to main chain N of D213, N1 of adenine ring makes an H-bond to the main chain N of L93 and C2'OH (O2) of ribose ring make H-bond to side chain carboxylic acid of E200 as shown in Figure (73A).

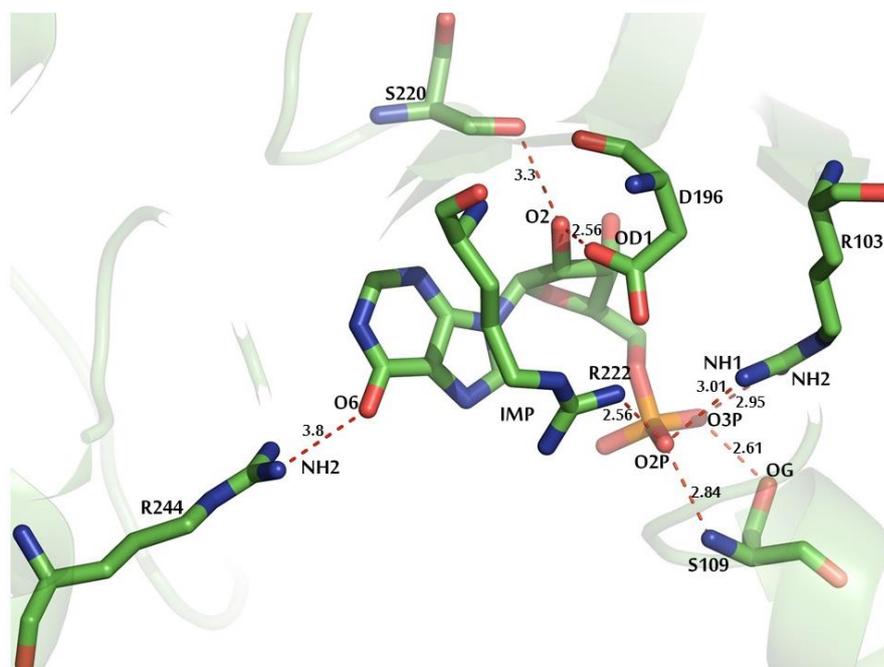
IMP is mostly located within a region formed by hydrophilic residues making H-bonds and ionic interactions with the local residues. The phosphate group makes an H-bond to the main chain N of S109 and an ionic interaction with the guanidiniums of R103

and R222. The O3P of the phosphate group makes H-bond to side chain OG of S109 and ionic interaction to guanidinium of R103. O6 of the inosine base points toward the guanidinium of R244 but is 3.8Å away (Figure 73B), the C2'OH (O2) of ribose make H-bonds to the main chain of S220 and side chain OD1 of D196.

(A)

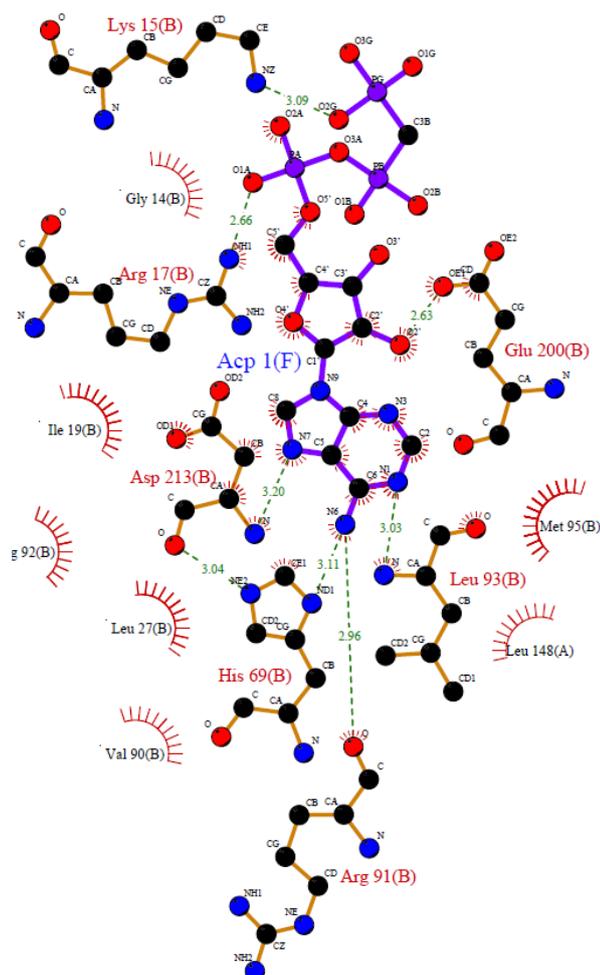


(B)

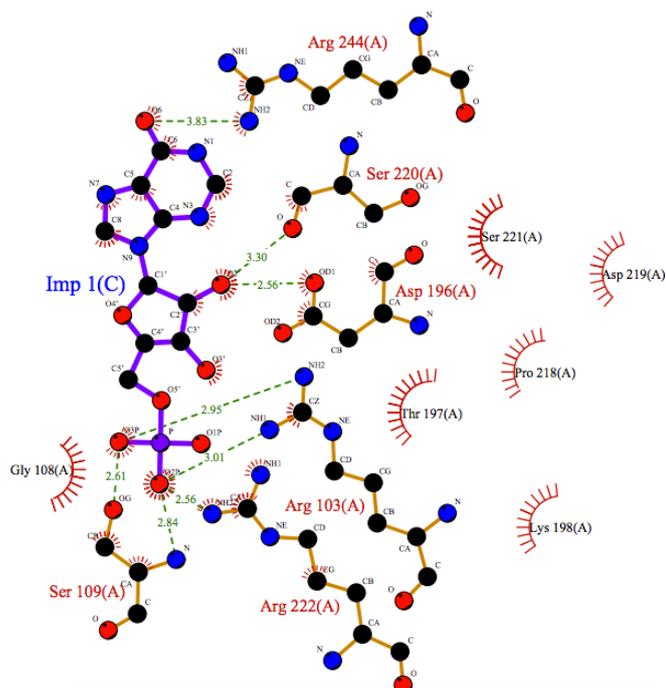


**Figure (72)** The interactions within the MtbSAICAR-ACP-IMP complex (A) ACP binding site showing distances. (B) IMP binding site showing distances, residues involved in the interactions are shown up as sticks, the distance between the NH2 atom of R244 and O6 of IMP is beyond the H-bond range. The 2Fo-Fc maps are coloured blue and contoured at 1 $\sigma$ .

(A)



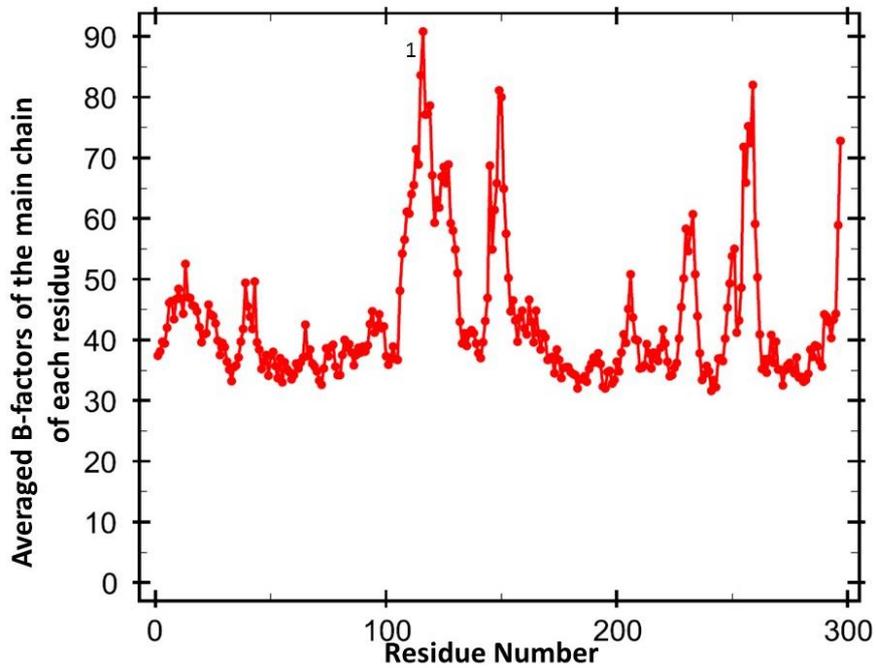
(B)



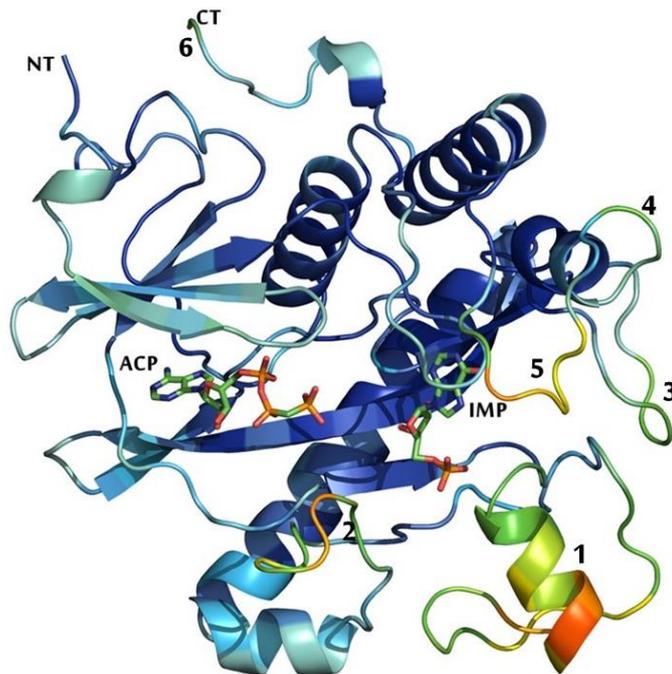
**Figure (73)** LIGPLOT representation of MtbSAICAR-ACP-IMP complex binding sites. (A) Residues involved in ACP binding site. (B) Residues involved in IMP binding.

The B-factor profile of MtbSAICAR in complex with competitive inhibitors ACP and IMP shows that it is the least ordered among liganded MtbSAICARs structures (ATP, ADP-AMZ complexes). The averaged B-factor is similar to the apo structure with many residues exceeding the standard B-factor range (15-50Å<sup>2</sup>) (Figure 74A). Residues (K145-E153), (Y229-V234), (S250-E251), (D255-P261) and (G296-A297) with high B-factors are found in loop regions. An  $\alpha$ -helix (S109-T117) which is part of the segment (T107-E131) is found to be bounded by loops on both sides is relatively less ordered than in the other structures. The highest B-values regions are indicated in order with numbers from 1-6 in (Figure 74B).

(A)



(B)



**Figure (74)** B-factors of the main chain atoms of the MtbSAICAR-ACP-IMP complex. (A) B-factors plotted by residue. (B) Three-dimensional cartoon showing B-factors coloured from blue (low) to red (high). The segments with high B-factors ( $>50\text{\AA}^2$ ) are (T107-E131), (K145-E153), (Y229-V234), (S250-E251), (D255-P261) and (G296-A297) and are numbered accordingly (1-6). (T107-E131) forms an  $\alpha$ -helix (S109-T117) bounded by loops on both sides. (K145-E153), (Y229-V234), (S250-E251), (D255-P261) and (G296-A297) are parts of loop suggesting flexibility. The area where the ligands bound are ordered (coloured blue). The graph is drawn using Baverage utility in CCP4.

#### 4.17. Summary

MtbSAICAR catalyses the seventh step of purine *de novo* pathway converting 4-carboxy-5-aminoimidazole ribonucleotide (CAIR) into succinyl-5-aminoimidazole-4-carboxamide-1-ribose-5'-phosphate (SAICAR) using ATP as energy source and aspartate. This chapter describes the solution of the crystal structure MtbSAICAR in apo form as well and in complex with ligands including ATP (activator), ACP and IMP (inhibitors) and substrate analogues (ADP, AMZ).

The crystal structures of MtbSAICAR in apo and liganded forms (ATP, ACP-IMP, and AMZ-ADP) are described. Co-crystallisation experiments showed ligands bound. In summary, the enzyme structure consists of one molecule per asymmetric unit and composed of 9  $\alpha$ -helices and 17  $\beta$ -strands (Figure 54). Analysis of the structure suggested that atoms, where the ligands are bound, are highly ordered, as indicated in Figure (56-62-68 and 74). The MtbSAICAR ligand complexes, ATP, ACP, and ADP are found in equivalent positions. Each structure contains only one molecule of the corresponding ligand, but the positions of phosphate groups in each complex are located differently (ATP, ADP, and ACP). In the MtbSAICAR-ATP complex, the electron density of the phosphate group shows it occupies two positions (bent and linear conformations). In the MtbSAICAR-ACP complex, the phosphate is in the bent conformation whereas in MtbSAICAR-ADP complex it is in the linear form. The adenine bases of AMZ and IMP are oriented differently and both are made H-bonds to the water molecule and the guanidinium of R244 respectively as discussed in the next chapter (Chapter5, Section 5.2.1.).

## **Chapter 5. Comparison within MtbSAICAR structures and with other homologues structures (ScSAICAR, TmSAICAR, EcSAICAR, SpSAICAR and HsSAICAR)**

### **5.1. Introduction**

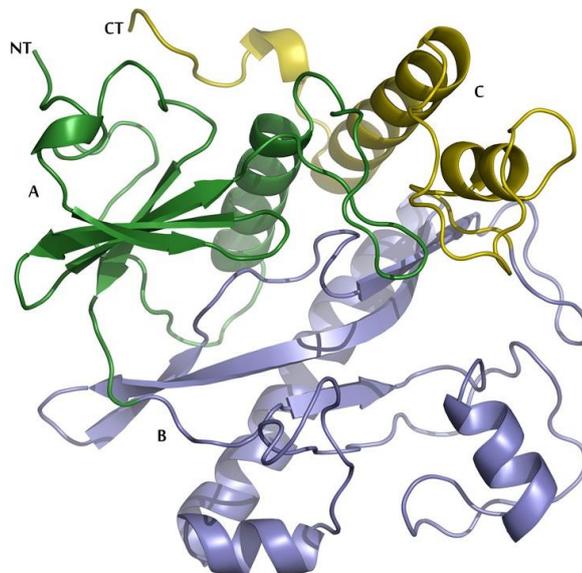
The crystal structure of MtbSAICAR has been solved both in the apo form and in complex with various ligands such as activator (ATP), substrate analogues (ADP and AMZ) and inhibitors (ACP and IMP). This chapter compares the apo structure of MtbSAICAR with its liganded complexes (ATP, ADP, AMZ, ACP and IMP). The MtbSAICAR-ADP-AMZ complex is also compared to homologous structures in the PDB i.e. the SAICARs from *S. cerevisiae* (ScSAICAR), *T. maritima* (TmSAICAR), *E. coli* (EcSAICAR), *S. pneumoniae* (SpSAICAR) and human (HsSAICAR). A detailed description of the ligand binding sites and differences between them will be presented in this chapter.

### **5.2. Comparison MtbSAICAR structures**

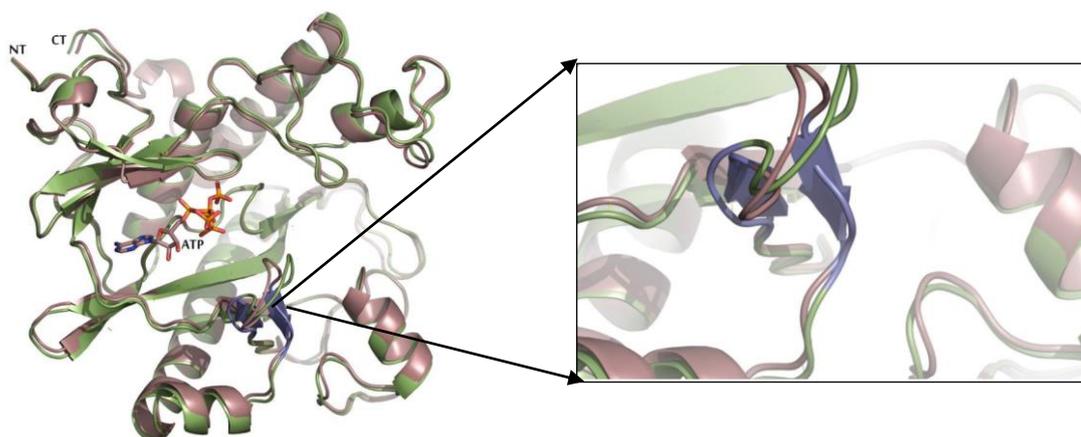
All of the MtbSAICAR crystal structures reported here have been determined from isomorphous crystals, and so the structures can be directly compared.

#### **5.2.1. Comparison of the structure of MtbSAICAR-ACP-IMP to MtbSAICAR-ATP and MtbSAICAR-ADP-AMZ complexes**

MtbSAICAR consists of three domains (based on the ScSAICAR classification (Levdikov et al., 1998)). The domains are shown in Figure (75). Superimposition of the structures of apo and the ATP complex shows that they are very similar; the only differences being a small secondary structure feature in regions (A143, T144, and K145) and (D152, E153, and N154) that are part of a  $\beta$ -strand in the complex whereas in the apo structure they form a less regular loop as shown in Figure (76).

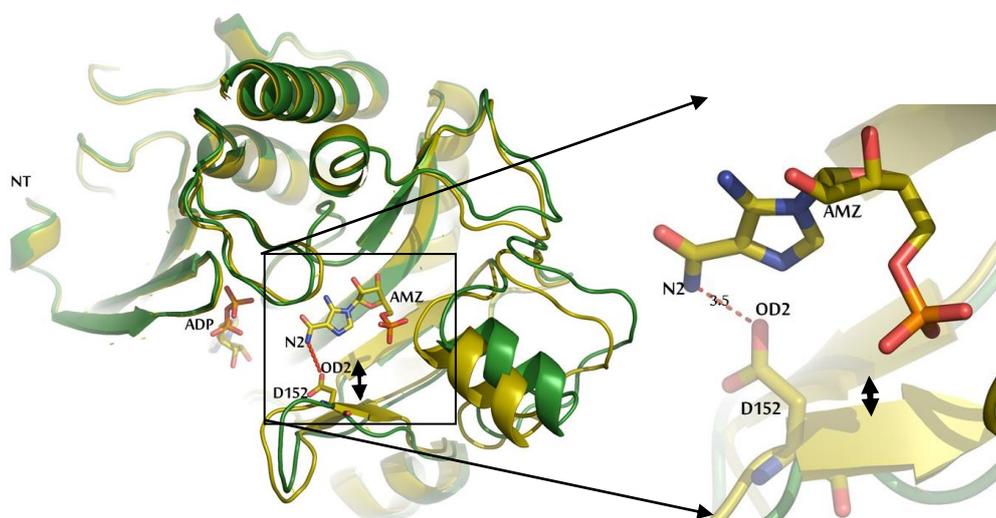


**Figure (75)** Domains representation of MtbSAICAR. Domain A is coloured green, Domain B is coloured slate blue, and domain C is coloured yellow. The domains classification is based on ScSAICAR (Levdikov et al., 1998). NT and CT represent the N-terminal and C-terminal respectively.



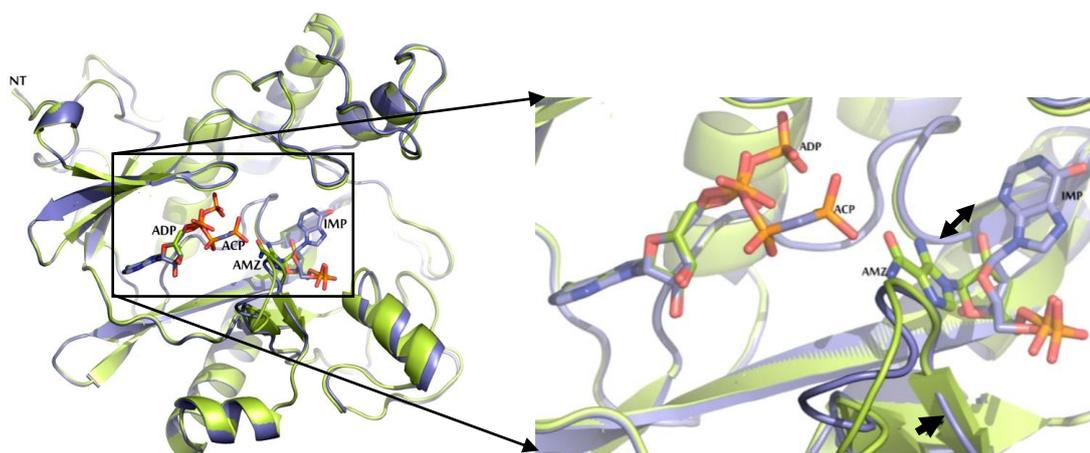
**Figure (76)** Superimposition of MtbSAICAR apo and its ATP complex. Both structures are very similar in their overall fold, the region of secondary structure difference is shown magnified, the ATP complex is coloured violet and the apo structure is coloured light green respectively.

Superimposition of the structures of apo and the ADP-AMZ complex reveal that they are also very similar. Figure (77) D152 shows the small differences in the secondary structure near the area where the AMZ is bound. This residue has high observed temperature factors and can be considered disordered.



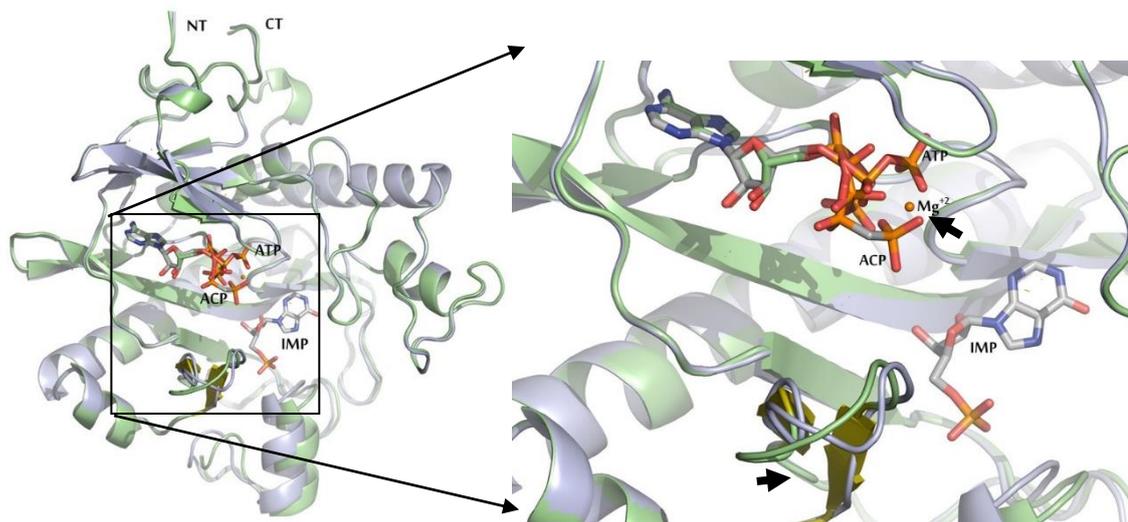
**Figure (77)** Comparison of the structures of SAICARs of *M. tuberculosis* (MtbSAICAR) in apo (olive) and in complex with ADP-AMZ (forest green). Differences in secondary structure are indicated by the double-ended black arrow. NT represents the N-terminal.

Comparison between the structures of SAICARs *M. tuberculosis* (MtbSAICAR) ADP-AMZ and ACP-IMP complexes gives similar observations to the comparison between apo and ADP-AMZ structures. Both of them have the small differences in secondary structure as shown in Figure (78). Although IMP is a competitive inhibitor with respect to CAIR and might be expected to bind in a very similar way, the adenine ring is oriented differently when compared to the adenine ring of AMZ, as shown below.



**Figure (78)** Superimposition of ACP-IMP and ADP-AMZ complexes revealed some differences between the positions of the phosphate groups of ACP and ADP. The phosphate groups are in bent and linear conformations respectively. The adenine ring of IMP orients differently compared to the adenine ring of AMZ (indicated by double-headed black arrow). The single black arrow shows the loop region (in the case of ACP-IMP) that forms a  $\beta$ -strand in the ADP-AMZ complexes. ACP-IMP and ADP-AMZ are coloured slate and lemon green respectively.

The structures of MtbSAICAR-ATP and ACP-IMP complexes are also very similar. ACP is bound in the same site as ATP. The ATP phosphates are seen in both linear and bent conformations, whereas ACP is solely in the bent conformation. A small secondary structure segment is different between ATP and ACP-IMP complexes, which are part of a  $\beta$ -strand in the ATP complex and loop in ACP-IMP as shown in Figure (79).



**Figure (79)** Superimposition of ATP and ACP-IMP complexes coloured light green and light blue respectively shows some differences between the positions of the phosphate group of ATP and ACP as well as small secondary structure changes. The ATP complex is coloured olive and the ACP-IMP slate, the  $\beta$ -strand/loop change is indicated by the black arrow.

### 5.3. Comparison of MtbSAICAR-ADP-AMZ complex with homologous structures (ScSAICAR, TmSAICAR, EcSAICAR, SpSAICAR and HsSAICAR)

**Table (16)** Deposited SAICAR Structures in the PDB

Organism	PDB ID	Resolution ( $\text{\AA}$ )
MtbSAICAR-ADP-AMZ complex	-	1.77 $\text{\AA}$
ScSAICAR-ADP-AMZ complex	(2CNQ)	1.00 $\text{\AA}$
TmSAICAR apo	(1KUT)	2.20 $\text{\AA}$
EcSAICAR-ADP-CAIR complex	(2GQS)	2.05 $\text{\AA}$
SpSAICAR-ADP-AIR-Asp complex	(4FE2)	2.28 $\text{\AA}$
HsSAICAR apo	(2H31)	2.80 $\text{\AA}$

In order to discuss the structures determined here and to compare them with the published structures of homologous enzymes (Table 16); the structures need to be first aligned (superimposed). The overall (global) secondary structures were aligned using Secondary Structure Matching (SSM) (Krissinel and Henrick, 2004) and locally using General Efficient Structural Alignment of Macromolecular Targets (GESAMT) in CCP4 (Krissinel, 2012), and root-mean-square deviation (RMSD) calculated. Because the domains move with respect to each other, appropriate comparison sometimes requires local rather than global matching. In a pairwise comparison, how well the C $\alpha$  backbones, when aligned, match each other is reflected in the calculated Q-score, the values for most of the comparisons are seen to be moderate, indicating structural similarity. The tables below give the values from the 3D alignments generated using PDBeFold (SSM) (Krissinel and Henrick, 2004).

### **5.3.1. Structure alignment of MtbSAICAR-ADP-AMZ complex and TmSAICAR structures**

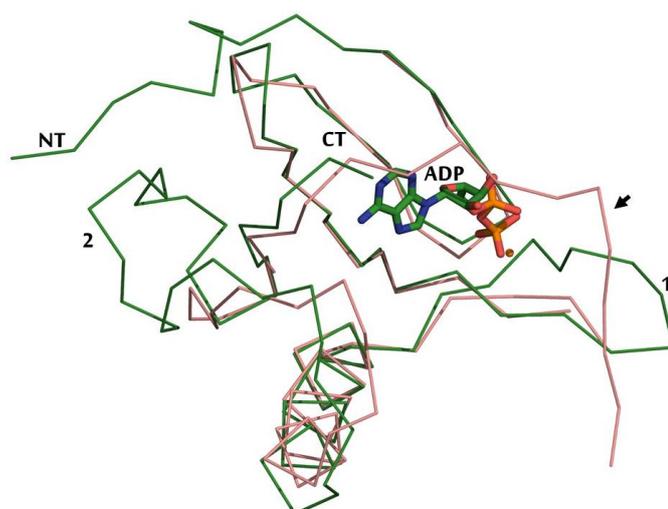
Comparison between molecular structures of the enzymes of different organisms requires optimal superimposition in order to assess similarities and differences between them precisely and eliminate possible errors from inappropriate comparisons. As described in the methods section (Chapter 2, Section 2.4.5.6), the structural comparison between MtbSAICAR-ADP-AMZ complex and TmSAICAR have been done both globally and locally. For the global pairwise comparisons, the whole chain (C $\alpha$  atoms) of the enzymes are superimposed and processed through PDBeFold (Krissinel and Henrick, 2004). The Q-Score value of the global comparison is 0.42 (Table 17). Whereas for local comparisons, the chains of the enzymes are divided into two domains (domain A; ADP binding site) and (domain B; AMZ binding site). In both methods, the alignment of the superimposition is assessed using the Q-Score value (Krissinel, 2012). The Q-Score value for global comparisons (Table 17) (0.42) is seen to be less than the superimposed domain A (0.51) alone, whereas it is nearly equal to the overlay domain B (0.41). This illustrates the advantages of local comparison of superimposed structures in domains to highlight regions of difference. Examination of C $\alpha$  traces (the main chain) of the superimposed A domains revealed that structure of the region around the ADP binding site is well conserved as indicated in (Figure 80A).

The graph of differences of superimposed domain A showed that residues which have a greater distance score ( $\geq 3 \text{ \AA}$ ) are part of a loop, as shown in Figure (80B), but these are not part of the active site. Overlaying domain B of the enzymes revealed that they are both very similar in terms of the fold and the region around the AMZ binding site is well conserved. The residues that have a larger distance value ( $>3 \text{ \AA}$ ) either form part of a loop that may indicate some flexibility or are part of the  $\alpha$ -helix found at the C-terminal end in both enzymes. The differences between the  $\alpha$ -helices of both organisms may be explained by the differing two loops labelled as 2 and 3 in the Figure (81A).

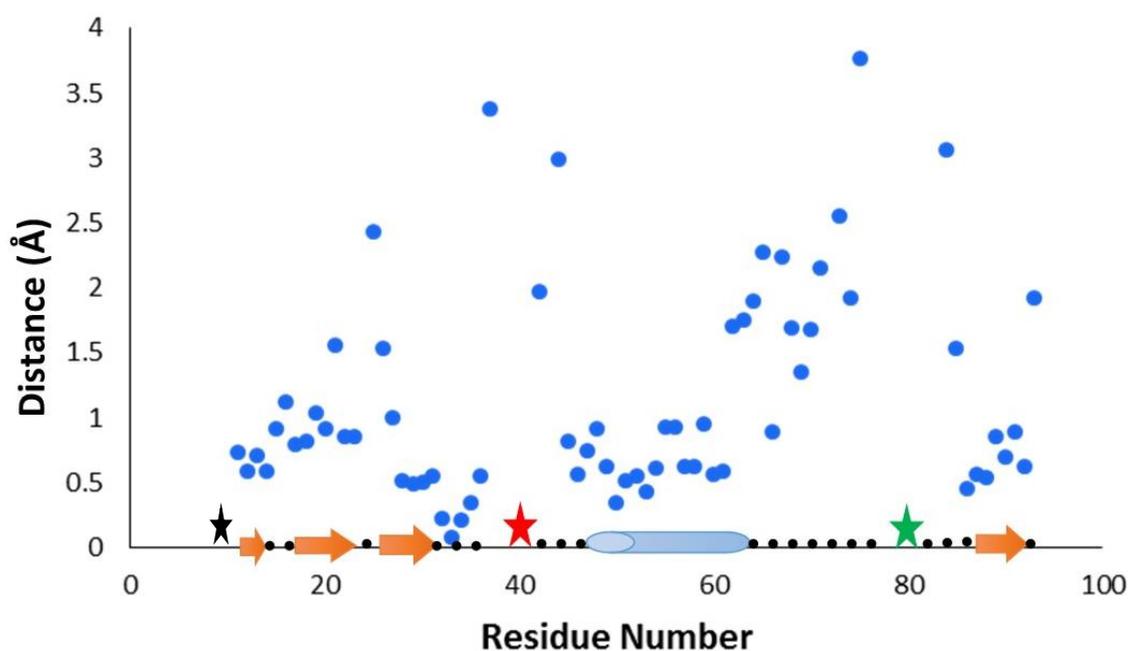
**Table (17)** 3D alignment and comparison between whole chain TmSAICAR (PDB 1KUT) and target MtbSAICAR-ADP-AMZ complex. Sequence identity (%seq) is the percentage of identical residues between the aligned sequences.  $N_{\text{resi}}$  represents the number of the residues in the whole polypeptide chain. The Q-score assesses the alignment based on both RMSD, and  $N_{\text{algn}}$  information where RMSD is averaged for  $C\alpha$  atoms of the alignment and  $N_{\text{algn}}$  defines the number of aligned residues between the polypeptide chains, a Q value of 1 indicate similar structure. A segment found to be in the query structure but not in the target is given by  $N_g$ , the number of gaps (of one residue or more) among the sequence. A zero-gap count is for perfect  $C\alpha$  alignment. %sse represents the similarity in secondary structure features that are found in both structures. Calculation of the superimposed alignment was performed using PDBeFold (Krissinel and Henrick, 2004).

Query		Q-Score	RMSD ( $\text{\AA}$ )	$N_{\text{algn}}$	$N_g$	%seq	Target	
%sse	$N_{\text{resi}}$						%sse	$N_{\text{resi}}$
87	222	0.42	1.91	197	12	23	65	297

(A)

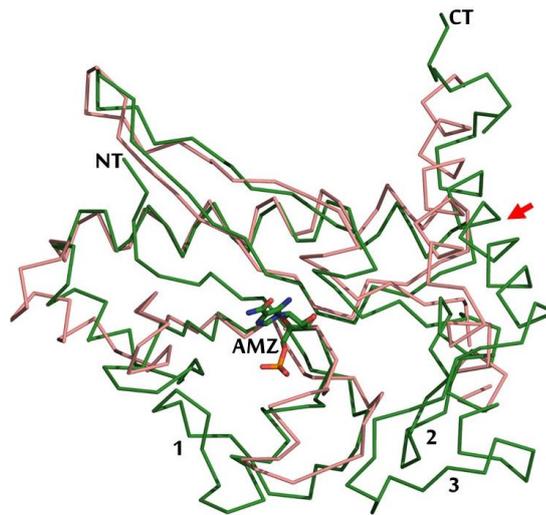


(B)

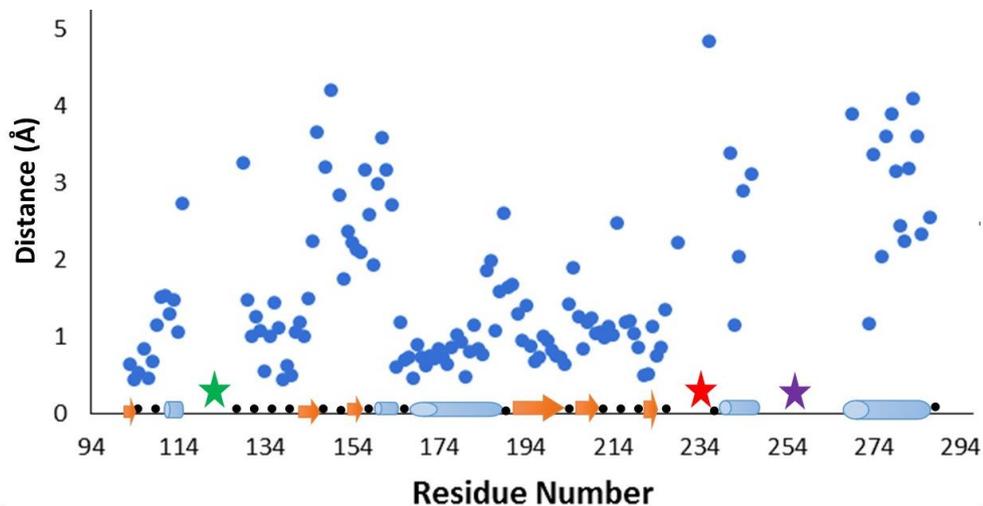


**Figure (80)** Comparison of Domain A of MtbSAICAR ADP complex and TmSAICAR using the GESAMT method (Krissinel, 2012). (A)  $C\alpha$  overlay traces of TmSAICAR (domain A coloured salmon) and MtbSAICAR-ADP-AMZ complex (ADP binding site; domain A coloured forest green). The insertion (M1-H10), is at the N-terminal (loop- $\alpha$ -helix- $\beta$ -strands). Region 1 (Y38-V41) is a gap in TmSAICAR and region 2 is a loop insertion (D76-V83). In TmSAICAR, the C-terminal end of domain A is part of a  $\beta$ -strand (as indicated by a black arrow). The Q-Score value and Global RMSD for 86 aligned residues are 0.51 and 1.382 $\text{\AA}$  from a total number of 80 and 93 residues respectively. (B) Local RMSD versus sequence position of the superimposition, showing that residues with large distance values ( $>3\text{\AA}$ ) are part of a loop. Region 1 and 2 above are indicated by red and green stars. Only the MtbSAICAR-ADP-AMZ complex residues of domain A (ADP domain) (M1-L93) are referenced as well as secondary structure feature of the superimposed residues as shown: loop ( $\bullet\bullet$ ),  $\alpha$ -helix ( $\text{blue oval}$ ) and  $\beta$ -strand ( $\text{orange arrow}$ ). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

(A)



(B)



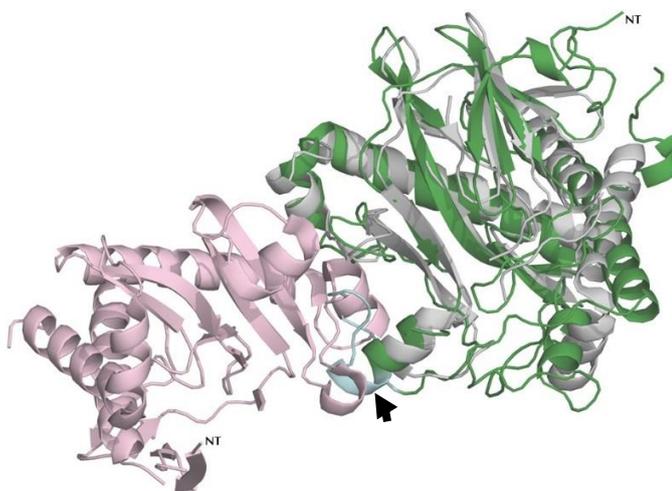
**Figure (81)** Comparison of Domain B of MtbSAICAR-ADP-AMZ complex and TmSAICAR structures using GESAMT method. (A) C $\alpha$  overlay traces of domain B of TmSAICAR (coloured salmon) and MtbSAICAR-AMZ complex (coloured forest green). The Q-Score value and Global RMSD for 129 aligned residues are 0.41 and 1.895Å from a total number of 142 and 204 residues respectively. Of the residues which have distance values  $>3\text{\AA}$  some are found be part of a loop while others are found to be in an  $\alpha$ -helix (F157,V160-E162), this  $\alpha$ -helix is seen to be near area of the loop labelled 1.  $\alpha$ -helix (E267-S285) is near the loops labelled as 2 and 3, this may explain the difference in conformation (indicated by red arrow found near the C-terminal site). (B) Local RMSD value versus sequence position of superimposition. There is loop insertion, 1 (A116-G128), in region 2(D227-K240) there is a gap in TmSAICAR and loop insertion 3 (G249-P266) indicated with green, red, purple stars respectively. The differences in the  $\alpha$ -helices conformation between enzymes at the C-terminal end is indicated by a red arrow and may be caused by the flexibility of the loops labelled as 2 and 3. Only the MtbSAICAR-AMZ complex residues of domain B (E94-A297) are referenced as well as secondary structure features of the superimposed domains shown as loop (••),  $\alpha$ -helix (blue oval) and  $\beta$ -strand (orange arrow). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

The crystal structure of MtbSAICAR is a monomer whereas, in *T. maritime* (TmSAICAR) (Zhang et al., 2006), EcSAICAR (Ginder et al., 2006) and SpSAICAR (Wolf et al., 2014) the structures are dimers. TmSAICAR consists of 8  $\alpha$ -helices and 11  $\beta$ -strands in each subunit whereas the monomeric form of MtbSAICAR is made up of 9  $\alpha$ -helices and 17  $\beta$ -strands (Laskowski et al., 1997). In TmSAICAR, the subunits are connected by a disulfide bond formed between C126 of subunit A and C126 of subunit B as shown in Chapter 1, Section 1.3.3.1. The amino and carboxyl termini of each subunit are located opposite to each other in a head to tail arrangement. Each subunit consists of three domains following the *S. cerevisiae* (ScSAICAR) classification.

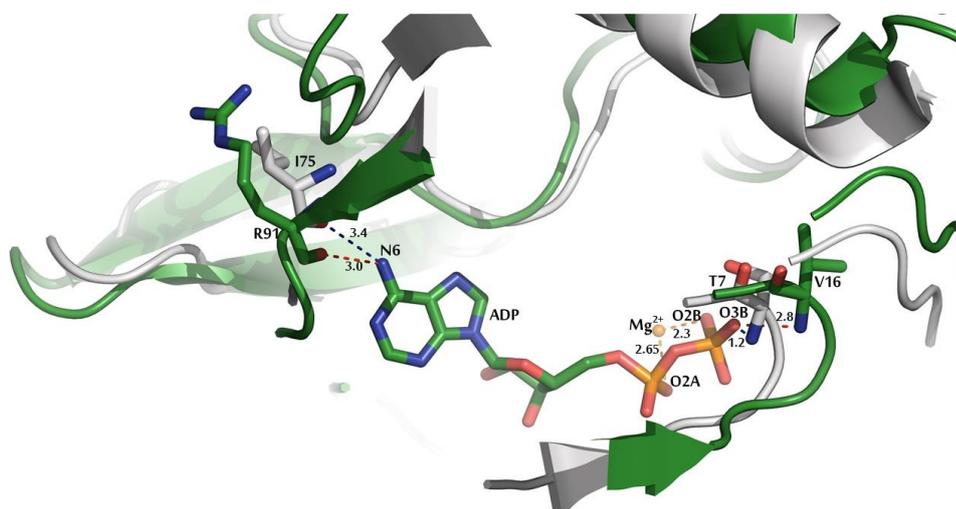
The structure of the apo TmSAICAR compared to ScSAICAR-ATP complex (PDB 1OBD) shows that their overall folds are conserved. However, there are some insertions and one of these (V134-Q146) is expected to have a role in preventing the dimerisation of ScSAICAR (Zhang et al., 2006). This insertion was seen to be absent in TmSAICAR whereas it is seen in the MtbSAICAR-ADP-AMZ complex at (Q115-P127) (indicated by the black arrow in Figure (82A)). The superimposition of TmSAICAR with MtbSAICAR-ADP-AMZ shows the close similarity in structure (Figure 82A).

The structure of apo TmSAICAR shows that it forms the sites corresponding to the ADP and AMZ binding sites. Both sites are located around a region of  $\beta$ -strands as in the MtbSAICAR-ADP-AMZ structure. The only differences seen in the residues forming the ADP binding site are where V16 and R91 of the MtbSAICAR-ADP-AMZ complex correspond to T7 and I75 of TmSAICAR. However, one could argue that these are not significant differences because the ligand interactions are with the main chain, not the side chains (Figure 82B).

(A)



(B)



**Figure (82)** Comparison of MtbSAICAR and TmSAICAR (A) Superimposed MtbSAICAR (coloured forest green) and TmSAICAR chains A and B (coloured pink and grey respectively) (B) The residues around the ADP binding site determined by the superimposition of MtbSAICAR-ADP-AMZ complex and TmSAICAR.

### 5.3.2. Structure comparison of MtbSAICAR-ADP-AMZ and ScSAICAR complexes

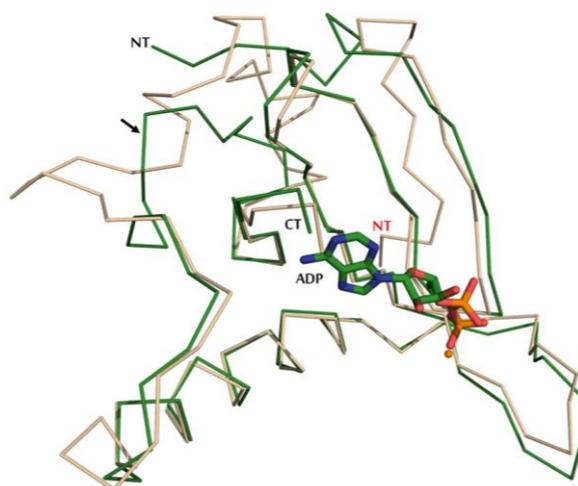
The structural comparison between MtbSAICAR-ADP-AMZ complex and ScSAICAR-ADP-AMZ complex domains was conducted as described in Chapter 2, Section 2.4.5.6. The overlaid structures of MtbSAICAR-ADP-AMZ complex and ScSAICAR-ADP-AMZ complex have been investigated both globally and locally in a similar manner to that described

in Section 5.3.1. For the global superposition, the Q-Score of the whole superimposed chains was been calculated as 0.51 (Table 18). In the comparison of the Q-Score values calculated from both domains separately (Domain A; 0.60) and (Domain B; 0.58), it is seen that Q-score values of both domains are higher than whole chains Q-score value. This indicates that optimal superimposition achieved once the whole chain is divided into domains. The overlaying of Domain A of enzymes from both organisms demonstrates that both share a similar fold. The region around the ADP binding site is well conserved as shown in Figure (83A). It has a few residues (P73-V83) that form part of loop which may allow some flexibility in the MtbSAICAR-ADP-AMZ complex over a segment (I79-L102) composed of few residues of loop then many residues form part of an  $\alpha$ -helix in the ScSAICAR-ADP-AMZ complex, this difference explains why some residues have a large distances ( $>3\text{\AA}$ ) (indicated by black arrow in Figure (83A)). Also, in domain A of MtbSAICAR-ADP-AMZ complex, a few residues with distance values exceeding  $3\text{\AA}$  are found as part of the  $\alpha$ -helix near the N-terminal end, these residues are adjacent to residues which are part of a loop (Figure 83B). The N-terminal site of domain A of ScSAICAR-ADP-AMZ complex found to be part of a loop which has a different conformation as indicated in Figure (83A). In domain B, both domains superimpose very well and the area around the AMZ binding site corresponds closely. In the MtbSAICAR-ADP-AMZ complex, there are two loops (labelled as 1 and 2 in Figure (84A)). The loop labelled 2 may affect the position of  $\alpha$ -helices at the C-terminal site. In MtbSAICAR-ADP-AMZ complex it forms a complete loop but in the corresponding segment of ScSAICAR-ADP-AMZ complex part of this loop forms an  $\alpha$ -helix. This is shown by the few residues involved in this piece (Figure 84B) having larger distance values ( $>3\text{\AA}$ ) than the rest of the structure.

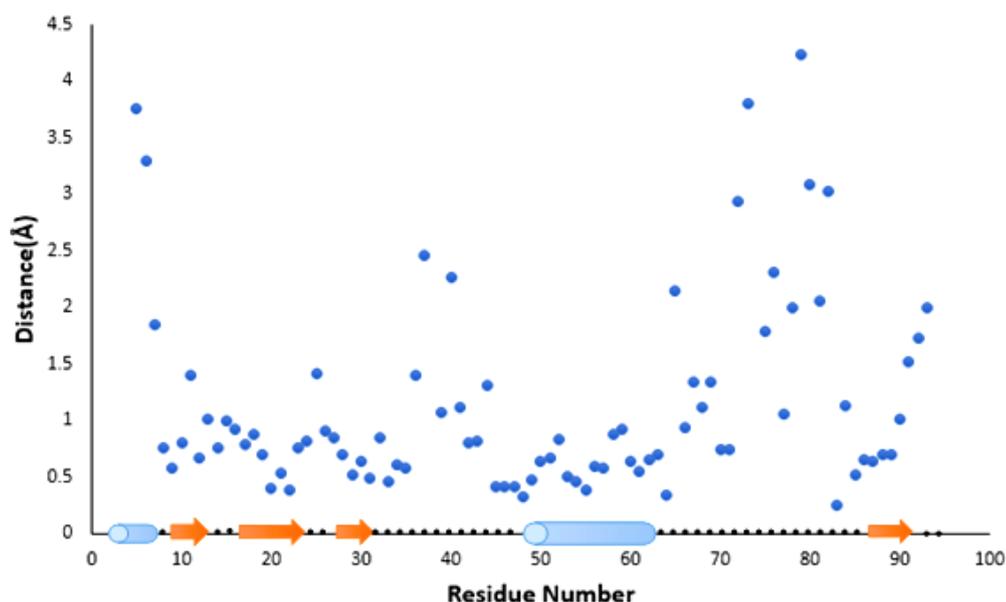
**Table (18)** 3D alignment comparison between query structure ScSAICAR-ADP-AMZ complex (PDB 2CNQ) and target structure (MtbSAICAR-ADP-AMZ complex). The definition of each is given in Table (17). Calculation of the superimposed alignment has been performed using PDBeFold (Krissinel and Henrick, 2004).

Query		Q-Score	RMSD ( $\text{\AA}$ )	$N_{\text{algn}}$	$N_{\text{g}}$	%seq	Target	
%sse	$N_{\text{resi}}$						%sse	$N_{\text{resi}}$
75	301	0.51	1.99	256	8	45	75	297

(A)

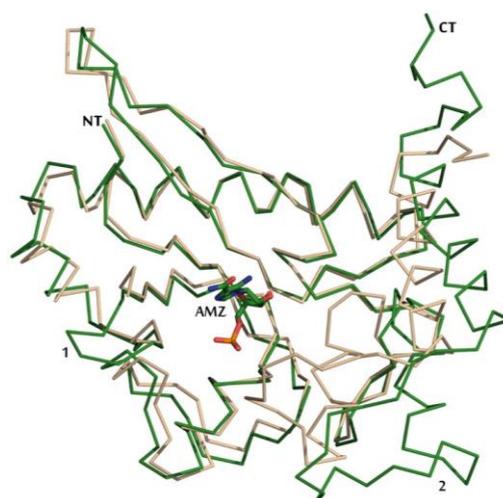


(B)

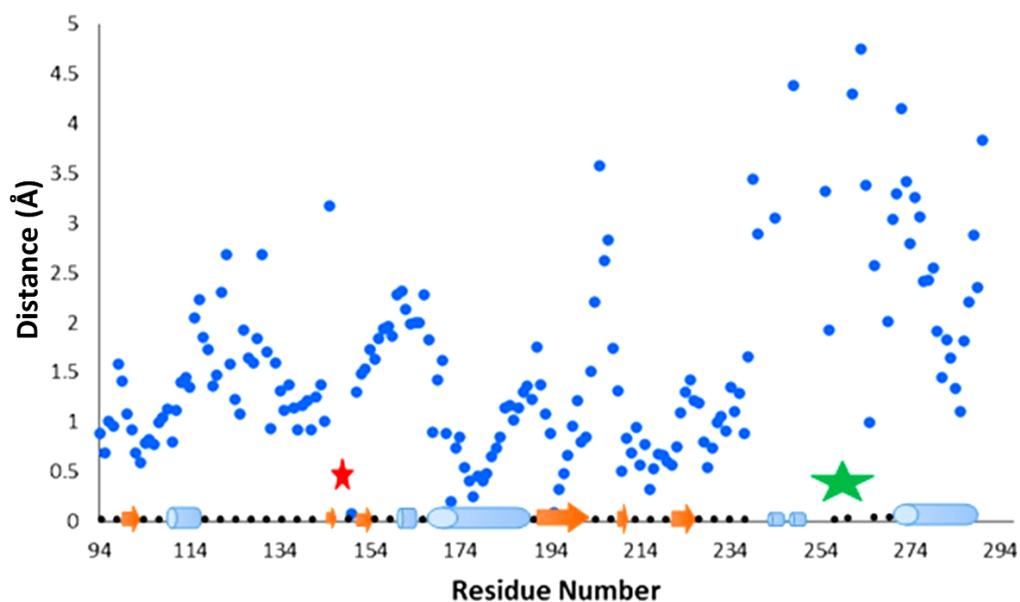


**Figure (83)** Comparison of Domain A of MtbSAICAR-ADP-AMZ complex with ScSAICAR-ADP-AMZ complex using the GESAMT method. (A) C $\alpha$  Overlay traces of domainA of ScSAICAR-ADP-AMZ complex (coloured wheat) and MtbSAICAR-ADP-AMZ complex (coloured forest green). The Q-Score value, Global RMSD for 87 aligned residues are 0.60 and, 1.414Å from the total number of 111 and 93 residues respectively. In ScSAICAR-ADP-AMZ complex, the equivalent residues to (P73-V83) (loop) are (I79-L102) (loop- $\alpha$ -helix) (which has some insertions (A80-K83, P90-Y98), this area is indicated by a black arrow in A and shown by high difference values in the graph (B). N-terminal (NT; coloured red) of domain A of ScSAICAR-ADP-AMZ complex (B) Local RMSD versus sequence position of the superimposition. In MtbSAICAR-ADP-AMZ complex, residues with large distance value ( $\geq 3\text{\AA}$ ) are part of loop whereas two residues ( $\alpha$ -helix) near the N-terminal end have a greater distance value near a loop. Only the MtbSAICAR-ADP-AMZ complex residues of domainA (M1-L93) are referenced and the related secondary structure for the superimposition is shown as loop (••),  $\alpha$ -helix (●) and  $\beta$ -strand (→). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

(A)



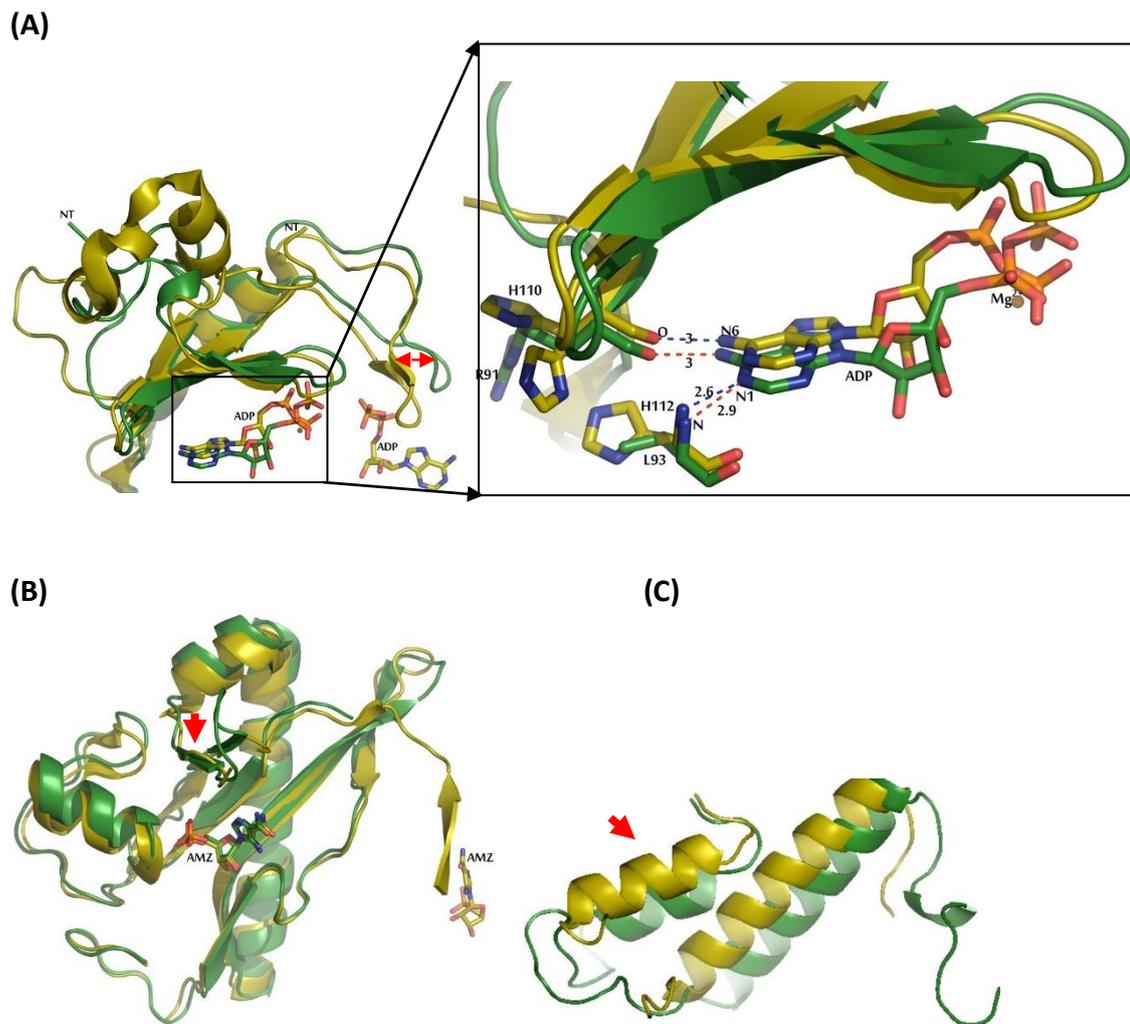
(B)



**Figure (84)** Comparison of Domain B of MtbSAICAR-ADP-AMZ complex and ScSAICAR-ADP-AMZ complex using the GESAMT method. (A) C $\alpha$  Overlay traces of domain B of ScSAICAR-ADP-AMZ complex (coloured wheat) and MtbSAICAR-ADP-AMZ complex (coloured forest green). The Q-Score value and global RMSD for 174 aligned residues are 0.58 and 1.772 Å from a total number of 190 and 204 residues respectively. The structure around the AMZ binding site is well conserved. In MtbSAICAR-ADP-AMZ complex, two loop insertions are indicated by 1 and 2. In Loop 1(A146-G149), there is a gap in ScSAICAR and loop 2 (G249-P266), is part of (Q241-H268;  $\alpha$ -helix-loop). (B) Local RMSD versus sequence position. Many residues near the C-terminal site have larger distance values than the rest and form part of an  $\alpha$ -helix. Loop regions labelled as 1(A146-G149) and 2(G249-P266) are indicated by red and green stars. Only the numbers of MtbSAICAR-ADP-AMZ complex residues of domain B (E94-A297) are referenced and the related superimposition secondary structure is indicated as loop(••),  $\alpha$ -helix (○) and  $\beta$ -strand (➡). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

The overall structure of the domains of ScSAICAR-ADP-AMZ (PDB 2CNQ) (Urusova et al., 2003) and MtbSAICAR-ADP-AMZ domain A (M1-L93) are very similar. However, some secondary structure features are seen to be different. It was observed that the structure of ScSAICAR-ADP-AMZ has a  $\beta$ -strand, S40-A41 and V44-I45, the S40 residue is the closest to  $\beta$ -phosphate of ADP, the corresponding residues S36-A37 and Y40-V41 of MtbSAICAR are part of a loop. The phosphate groups of ADP are in different conformations in ScSAICAR, being bent, whilst in MtbSAICAR the linear conformation is seen, as shown in Figure (85A). In addition to that, there is a difference in the ADP binding site residues, whereby the polar side chains of H112 and H110 are replaced by a hydrophobic and a charged residue (L93 and R91) respectively in MtbSAICAR-ADP-AMZ. However, these interactions (made to N1 and N6) are from the main chain, and so may not affect the binding site.

In comparison, the structure of ScSAICAR-ADP-AMZ Domain B (M95-Q235), also shows some small differences in secondary structure. In MtbSAICAR-ADP-AMZ, residues (143-145) and (152-154) are parts of  $\beta$ -strands. These segments are part of a loop in ScSAICAR-ADP-AMZ. Apart from that, the AMZ binding site is conserved (Figure 85B). Domain C (T236-A297) is well conserved to the structure of ScSAICAR (T256- S305) although it does not make a connection to any ligands here, the  $\alpha$ -helix shows some movements in comparison to ScSAICAR-ADP-AMZ as it is shifted outward while in ScSAICAR-ADP-AMZ it is shifted inward (Figure 85C).



**Figure (85)** The structure of MtbSAICAR-ADP-AMZ (green) and ScSAICAR-ADP-AMZ (olive). (A) Domain A of MtbSAICAR and ScSAICAR, ADP is shown as sticks. NT identifies the N-terminal. (B) Domain B of MtbSAICAR and ScSAICAR. The difference in secondary structure is indicated by a red arrow. The peripheral site in ScSAICAR is also shown. (C) The  $\alpha$ -helices of Domain C, the relative movement between the  $\alpha$ -helix in the structures is pointed out by the red arrow.

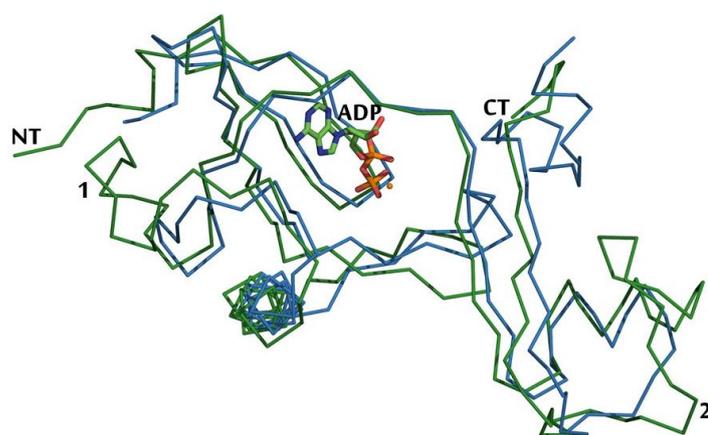
### 5.3.3. Structure comparison of MtbSAICAR-ADP-AMZ and EcSAICAR-ADP-CAIR complexes

The comparison of EcSAICAR-ADP-CAIR and MtbSAICAR-ADP-AMZ complexes (Domain A) were conducted in the same manner as discussed in Chapter 2, Section 2.4.5.6. For the entire length superimposed polypeptide chains, the Q-score is 0.45 which is similar to the Q-score of the divided domains A and B of 0.47 and 0.47 respectively. Although evaluation of the aligned enzymes is easier on a local (per domain) basis, this is not reflected in the Q-values as these are similar to those for global analysis. The superimposed domains are found to be highly conserved in terms of the fold, the structure around the ligand binding site is also well conserved (Figure 86A). Loop insertions in domain A of MtbSAICAR-ADP-AMZ complex are labelled as 1 and 2 in Figure (86A) and as red and green stars in Figure (86B). Comparison of domain B of enzymes of both organisms shows that both are conserved in terms of the fold but the MtbSAICAR-ADP-AMZ complex has loop insertions labelled as 1 (G118-G128), 2(A146-G149), 3(R230-Q235), and 5(G249-P266) (Figure 87A). The residues with greater distance values ( $>3\text{\AA}$ ) are part of loop and  $\alpha$ -helix but those residues which are part of the  $\alpha$ -helix are located near the loop insertion 5 that may affect the displacement of these residues.

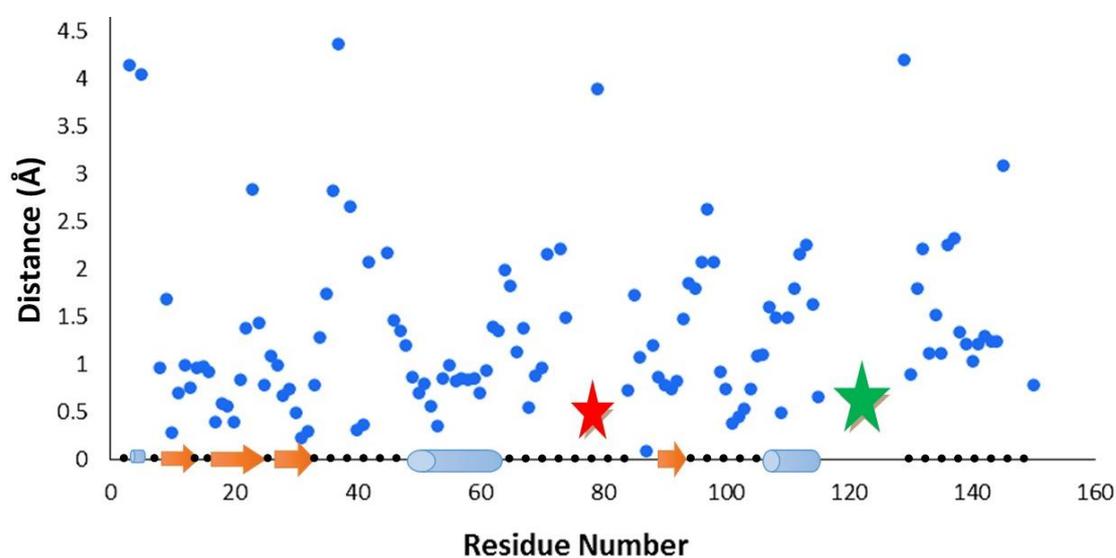
**Table (19)** 3D alignment comparison between query EcSAICAR-ADP-CAIR complex (PDB 2GQS) and target MtbSAICAR-ADP-AMZ complex. The definition of each is given in Table (17). Calculation of the superimposed alignment has been performed using PDBeFold (Krissinel and Henrick, 2004).

Query		Q-Score	RMSD ( $\text{\AA}$ )	N <sub>algn</sub>	Ng	%seq	Target	
%sse	N <sub>resi</sub>						%sse	N <sub>resi</sub>
70	237	0.45	2.02	215	12	24	70	297

(A)

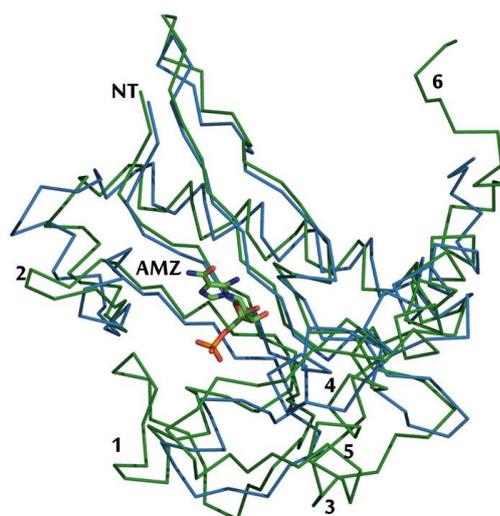


(B)

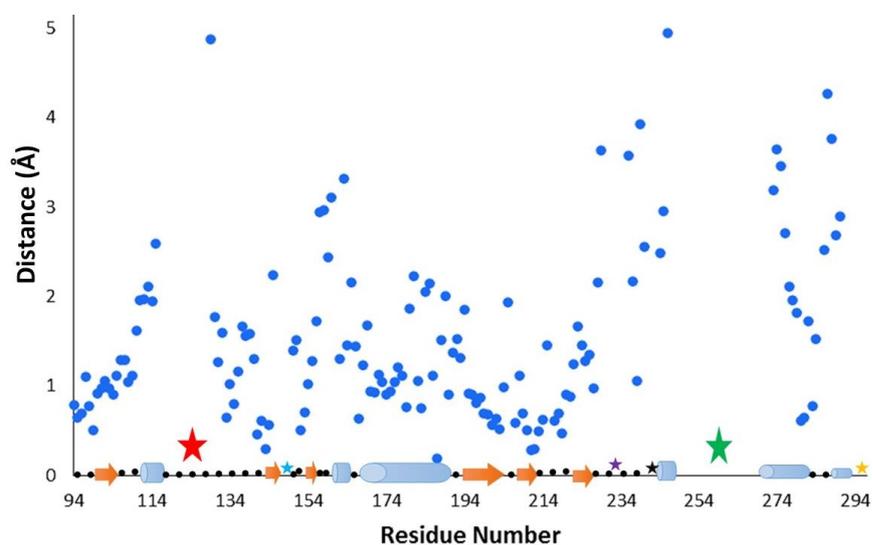


**Figure (86)** Comparison of Domain A of MtbSAICAR-ADP-AMZ and EcSAICAR-ADP-CAIR complexes structures using the GESAMT method (Krissinel, 2012). (A) Ca overlay traces of domain A of EcSAICAR (coloured blue) and MtbSAICAR-ADP-AMZ complex (coloured forest green). The Q-Score value and global RMSD for 116 aligned residues are 0.47 and 1.607 Å from a total number of 150 and 150 residues respectively. Loop insertions 1 and 2 are indicated as stars in the graph. (B) Local RMSD versus sequence position of the superimposition. Residues which have large distance  $>3\text{Å}$  are found to be part of the loop. Loop insertions (D75-V83) and (A116-G128) (labelled as 1 and 2 are shown above) are indicated by red and green stars respectively. Only the MtbSAICAR-ADP-AMZ complex residues of domain A (M1-D150) are referenced and the related secondary structure for the superimposition is shown loop (••),  $\alpha$ -helix (○) and  $\beta$ -strands (➔). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

(A)



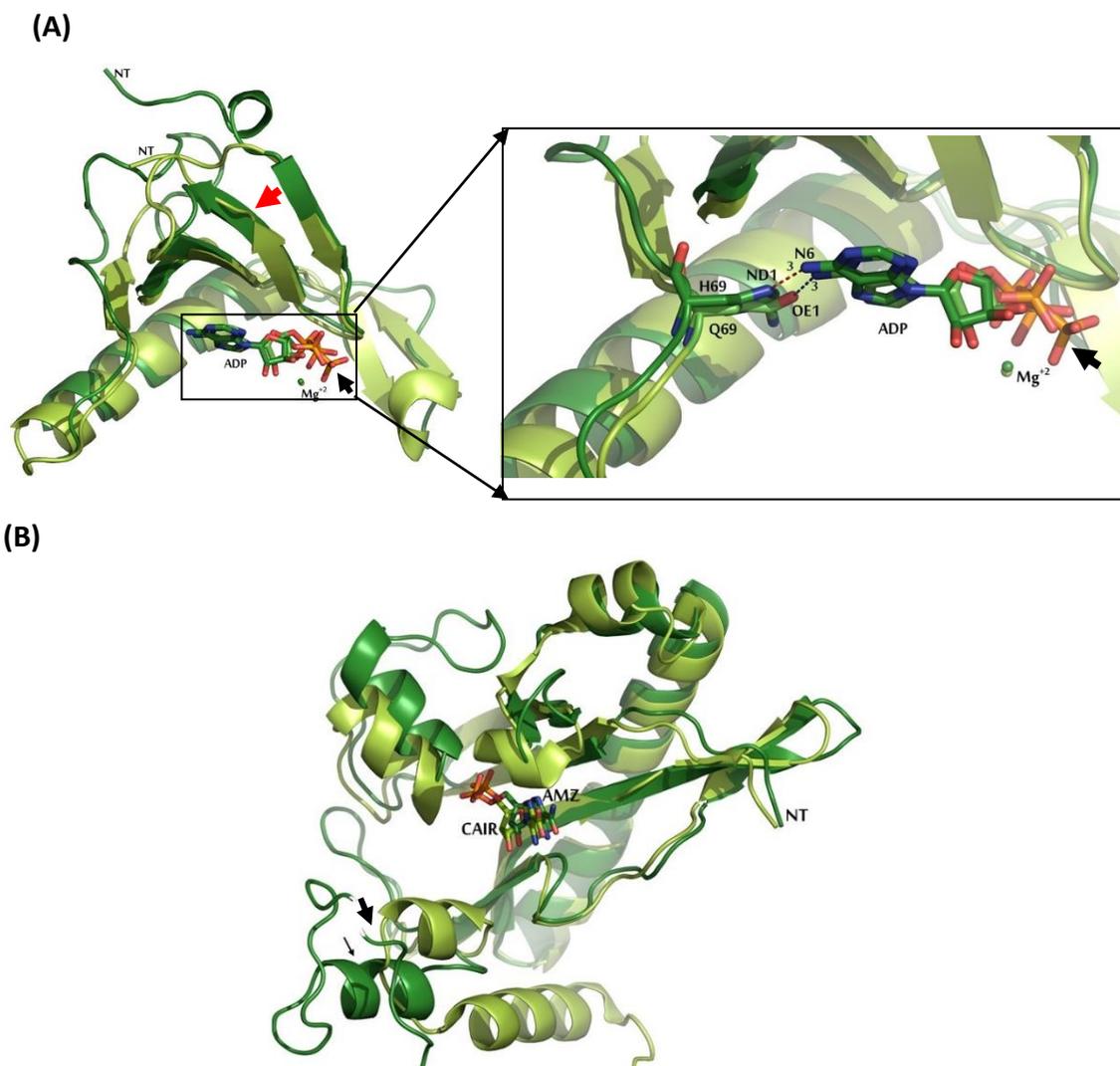
(B)



**Figure (87)** Comparison of Domain B of MtbSAICAR-ADP-AMZ and EcSAICAR-ADP-CAIR complexes structures using the GESAMT method. (A) C $\alpha$  Overlay traces of domain B of EcSAICAR (coloured blue) and MtbSAICAR-ADP-AMZ complex (coloured forest green). The Q-Score value and global RMSD for 141 aligned residues are 0.47 and 1.784 Å from a total number of 153 and 204 residues respectively. Loop insertions 1(G118-G128), 2(A146-G149), 3(R230-Q235), 5(G249-P266) and 6(D291-A297) are indicated, Loop number 6 is located at the C-terminal. The segment indicated by 4(Q241-V243) is part of an  $\alpha$ -helix and found not to superimpose well to the corresponding  $\alpha$ -helix of domain B of the EcSAICAR-ADP-CAIR complex. Residues with a large distance ( $>3\text{\AA}$ ) are part of loop and  $\alpha$ -helix but insertions may result in the relative movement of these residues that are part of the  $\alpha$ -helix. (B) Local RMSD versus sequence position of the superimposition. Segments 1,2,3,4,5 and 6 are indicated by red, blue, violet, black, green and yellow stars. Only the number of MtbSAICAR-ADP-AMZ complex residues of domain B (E94-A297) are referenced and the secondary structure content of the superimposed segments is shown loop (••),  $\alpha$ -helix (○) and  $\beta$ -strands (→). This calculation was carried out using the superpose utility in CCP4 (Krissinel, 2012).

The SAICARs of *E. coli* (EcSAICAR) is a dimer with each subunit consisting of two domains. These are the ADP binding domain A (M1-K82) and the CAIR binding domain B (K83-D237). The crystal structure of the ADP and AMZ (the substrate analogue of CAIR) complex of MtbSAICAR was compared to the structure of EcSAICAR-ADP-CAIR complex. Superimposition of one subunit of EcSAICAR to the monomer structure of MtbSAICAR molecule revealed that the overall fold of both structures is similar. The ADP binding domain A (M1-R91) (based on structural alignment with EcSAICAR-ADP-CAIR) of MtbSAICAR-ADP-AMZ shows the structure of domain A is highly conserved with minor differences in some secondary structure elements such as in MtbSAICAR-ADP-AMZ complex that has segments (R21-D23) which are part of a  $\beta$ -strand. However, in the EcSAICAR-ADP-CAIR complex, the corresponding region (S17-E19) is part of a loop shown by the red arrow in Figure (88A). Both of these segments are located in domain A where the ADP binding site is found but neither of them has direct interactions with ligands. The phosphates of ADP are in a linear conformation in the MtbSAICAR-ADP-AMZ complex but in the EcSAICAR-ADP-CAIR complex ADP they are bent, indicated by the black arrow in Figure (88A). In the ADP binding site side chain atom, ND1 of H69 of MtbSAICAR-ADP-AMZ complex is replaced by OE1 of Q69. Both of these residues are bound to ADP by their side chains as shown in Figure (88A).

In the CAIR binding domain (Domain B) of the EcSAICAR-ADP complex (Ginder et al., 2006) CAIR is replaced by AMZ, the substrate analogue, in the structure of the MtbSAICAR-ADP-AMZ complex (R92-P266). Overall, both domains have similar structures but there are movements of some  $\alpha$ -helices, shown by the black arrow in Figure (88B). The position of AMZ and CAIR, and the residues forming the binding sites are conserved between both organisms.



**Figure (88)** Domains of MtbSAICAR in complex with ADP and AMZ. The domains were determined based on the EcSAICAR description. (A) The ADP binding domain A (M1-R91) of MtbSAICAR. Domain A is coloured forest green and EcSAICAR is coloured lemon. The ADP phosphates are bound in linear and bent conformations in MtbSAICAR and EcSAICAR respectively. NT represents the N-terminal of both structures. The different secondary structures are shown by a red arrow. Residues differences between the proteins are also labeled. The distances between atoms of MtbSAICAR-ADP complex and EcSAICAR-ADP complexes is indicated by red and blue respectively. (B) AMZ binding domain B (R92-P266) of MtbSAICAR. Domain B is coloured forest (green), the CAIR site of domain B of EcSAICAR is coloured lemon; the difference in the secondary structure elements of both proteins is shown with a black arrow. Both AMZ and CAIR sites are shown and coloured corresponding to their cartoon representation.

### 5.3.4. Structure alignment of MtbSAICAR-ADP-AMZ and *S. pneumoniae* SAICAR (SpSAICAR-ADP-AIR-Asp)

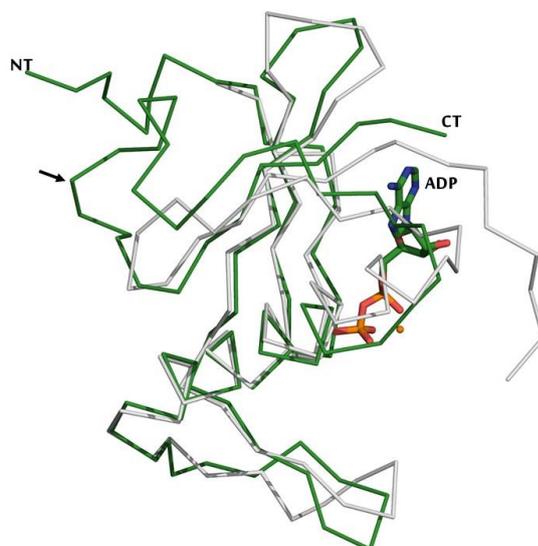
The alignment between MtbSAICAR-ADP-AMZ complex and SpSAICAR-ADP-AIR-Asp is assessed using Q-score value as above. The alignment has been done for the entire length of chain and for the separated domains as is described in Section 5.3.1. The Q-score value of the entire enzymes is 0.44 as shown in Table (20). The Q-score value of the divided domains (A; 0.57) is higher while (B; 0.44) is similar to the whole chain. The aligned domain A of MtbSAICAR-ADP-AMZ complex and SpSAICAR-ADP-AIR-Asp complex revealed that both domains are highly conserved in terms of the fold. The area around ADP binding site correlates is well. The residues that have relatively larger distance ( $>3\text{\AA}$ ) are found to be part of a loop adjacent to a loop insertion as indicated by the black arrow and a red star in Figure (89A-89B).

Comparison of domain B of both enzymes showed that region around the AMZ binding site is also highly conserved. Most of the residues (F157, R158, M163, S245, T273-A275, S285) with large distance values of  $>3\text{\AA}$  are part of  $\alpha$ -helices). Almost all residues of the segment (A116-L129) are part of the loop as labelled 1 and with a red star in Figure (90A-90B). Segments (A147-D150) and (R230-Q235) are parts of loops (labelled 2 and 3 indicated by a green and a grey star respectively). This segment (Q241-V243) is composed of an  $\alpha$ -helix and does not superimpose well, this is may be due to the loop precedes it, this segment is indicated below by number 4 and a black star. A loop insertion (G249-P266) may cause some movement to the  $\alpha$ -helix part of (W246-A272) and prevent it from being closely superimposable. These segments are labelled as 1,2,3,4,5 and indicated by red, green, grey, black and blue stars respectively in Figure(90A-90B).

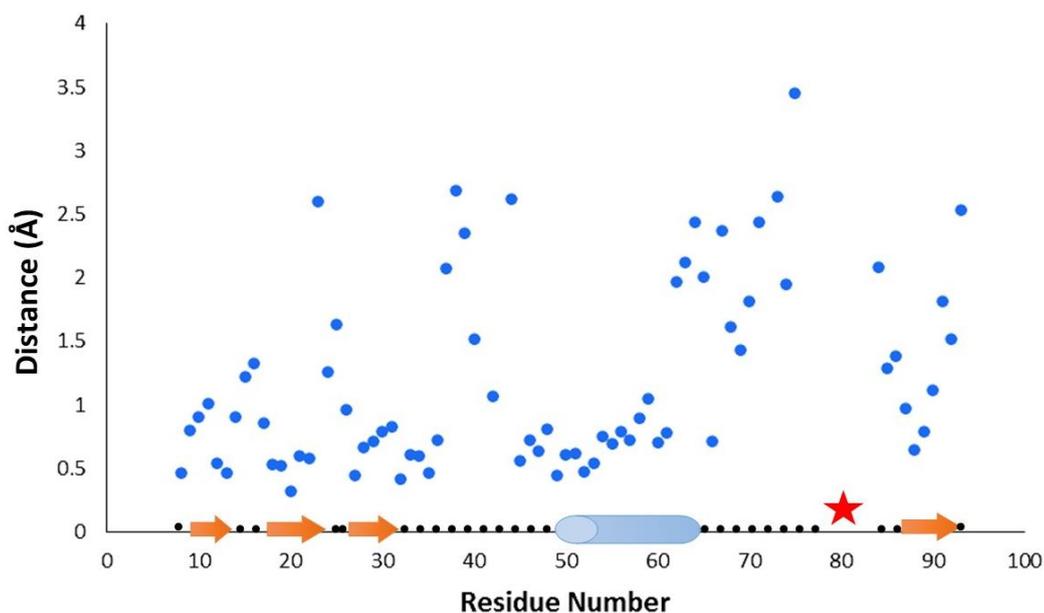
**Table (20)** 3D alignment comparison between whole chain SpSAICAR-ADP-AIR-Asp (PDB 4FE2) and the MtbSAICAR-ADP-AMZ complex. The definition of each is given in Table (17). Calculation of the superimposed alignment has been performed using PDBeFold (Krissinel and Henrick, 2004).

Query		Q-Score	RMSD ( $\text{\AA}$ )	N <sub>algn</sub>	N <sub>g</sub>	%seq	Target	
%sse	N <sub>resi</sub>						%sse	N <sub>resi</sub>
79	234	0.44	2.07	213	11	24	75	297

(A)

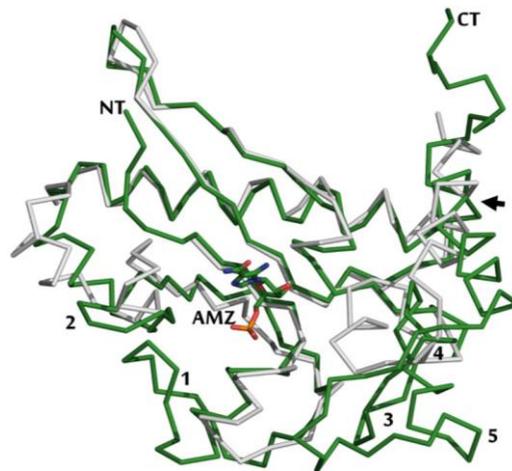


(B)

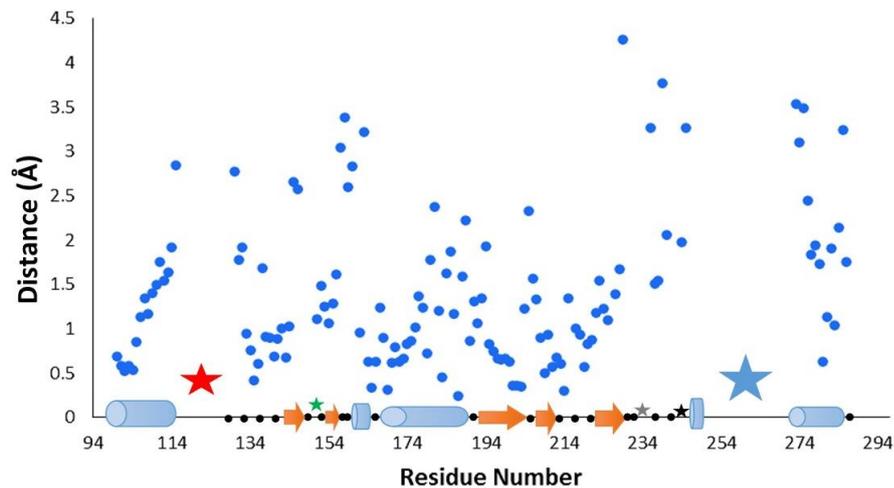


**Figure (89)** Comparison of Domain A of MtbSAICAR-ADP-AMZ complex and SpSAICAR-ADP-AIR-Asp-Mg<sup>2+</sup> complex using the GESAMT method (A) C $\alpha$  Overlay traces of domain A SpSAICAR-ADP-AIR-Asp-Mg<sup>2+</sup> complex (grey) and MtbSAICAR-ADP-AMZ complex (forest green). The Q-Score value and global RMSD for 75 aligned residues are 0.57 and 1.393Å from a total number of 87 and 93 residues respectively. (B) Local RMSD versus sequence position. Residues which have large distance values ( $>3\text{Å}$ ) are part of a loop and their movement may be due to the loop insertion adjacent (indicated by a red star). In domain A of MtbSAICAR-ADP-AMZ complex, loop insertion (D76-V83) is indicated by the black arrow above and as a red star in the graph. Only the MtbSAICAR-ADP-AMZ complex residues of domain A (M1-L93) are referenced and the secondary structure content of the superimposed segments are shown loop (•),  $\alpha$ -helix (blue oval) and  $\beta$ -strands (orange arrow). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

(A)



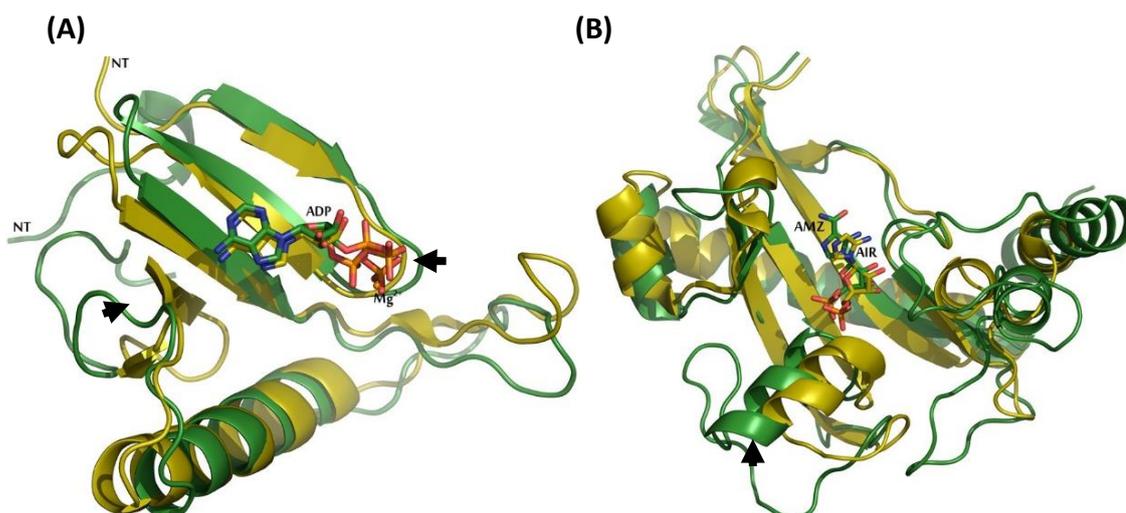
(B)



**Figure (90)** Comparison of Domain B of MtbSAICAR-ADP-AMZ complex and SpSAICAR-ADP-AIR-Asp-Mg<sup>2+</sup> complex structures using the GESAMT method (Krissinel, 2012). (A) Ca Overlay traces of domain B of SpSAICAR-ADP-AIR-Asp-Mg<sup>2+</sup> complex (grey) and MtbSAICAR-ADP-AMZ complex (forest green). The Q-Score value and global RMSD for 130 aligned residues are 0.44 and 1.640Å from a total of 146 and 204 residues respectively. Insertions and unaligned segments have been labelled as 1,2,3,4 and 5 and are described below in order (B) Local RMSD versus sequence position. Most of the residues with a larger distance value of >3Å (F157-R158, M163, S245,T273-A275, S285) are part of α-helices. An insertion (A116-L129) (α-helix-loop) is indicated by a red star. Segments (A147-D150), (R230-Q235) are part of loop indicated by green and grey stars respectively. Segment (Q241-V243) (indicated by black stars) is α-helix which does not superimpose well to the corresponding α-helix, this may be because of the movement caused by the preceding loop. Insertions (W246-A272)(α-helix-loop-α-helix), a loop insertion (G249-P266) are indicated by a blue star. Only the MtbSAICAR-ADP complex residues of domain B (E94-L297) are referenced and the secondary structure content of the superimposed segments is shown loop (•), α-helix (○) and β-strands (➤). This calculation was carried out using superpose in CCP4 (Krissinel, 2012).

The crystal structure of SpSAICAR has been solved in two forms. One structure was solved in complex with ADP, CAIR, aspartate,  $Mg^{2+}$  (PDB 4FE2), and the other structure was solved in the presence of acetate, ADP, AIR and  $Mg^{2+}$  (PDB 4NYE) (Wolf et al., 2014). SpSAICAR is a homodimer with two domains consisting of domain 1 (M1–K81) and domain2 (K82–K235). In SpSAICAR the ADP binding domain (domain A) (M1–K81) is equivalent to (M1-E82) of the MtbSAICAR-ADP-AMZ complex. Superimposition of these domains revealed that they are very similar, as shown in Figure (91A).

The ADP binding site is retained between the enzymes from both organisms but the pyrophosphate of ADP in MtbSAICAR-ADP-AMZ complex is in a linear conformation whereas in SpSAICAR as well as EcSAICAR it is bent. The ADP conformations in SpSAICAR and MtbSAICAR are pointed out by a black arrow in Figure (91A). The residues around the ADP binding site is either conserved or from similar amino acids types. The 5-amino-1-(5-phospho-D-ribosyl) imidazole (AIR) binding site (Domain B) of SpSAICAR (K82-K235) is equivalent to AMZ binding site (Domain B) of MtbSAICAR-ADP-AMZ complex (R92-K288). Both domains and their ligand binding sites are conserved. But some secondary structure region showed slight relative movements as indicated by black arrow (Figure 91B).



**Figure (91)** Comparison of MtbSAICAR-ADP in complex with AMZ; domains were assigned based on SpSAICAR. (A) ADP binding domain A (M1-E82) of MtbSAICAR, domain A is coloured green while in SpSAICAR it is coloured as olive. ADP in MtbSAICAR and SpSAICAR are shown in linear and bent conformations respectively. NT signifies the N-terminal of both structures. The difference in secondary structure is pointed by black bold arrow (B) AMZ binding domain B (R92-K288) of MtbSAICAR. The movement of the secondary structure between both structures is pointed it out with a black arrow.

### 5.3.5. Structure alignment of MtbSAICAR-ADP-AMZ and HsSAICAR

The Optimal alignment of these enzymes has been calculated using the methods described in Chapter 2, Section 2.4.5.6. The comparison between MtbSAICAR-ADP-AMZ complex and HsSAICAR (unliganded apo form) follows the same manner as discussed in Section 5.3.1. The optimal superimposition (for entire length/per domain) is assessed using Q-score value. The Q-score value of the whole chain alignment is 0.37 as shown in Table (21). Comparing the Q-score value of the entire chain alignment to the divided domains (domain A; 0.35, domain B; 0.44) revealed that they are nearly similar but the best alignment is achieved once the enzymes are divided into domains.

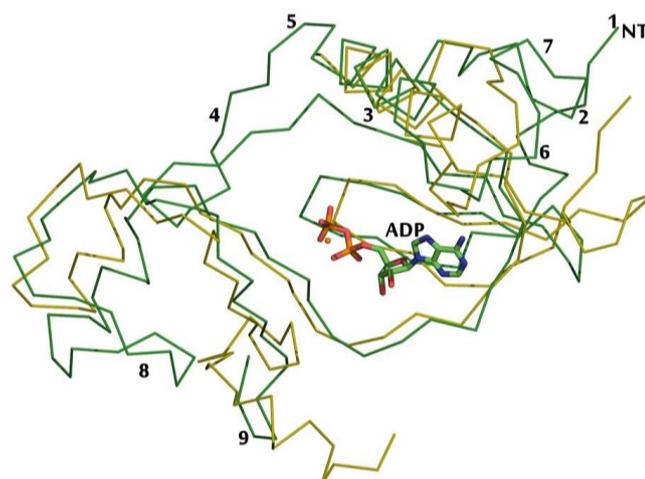
The structural comparison between domain A of MtbSAICAR-ADP-AMZ complex and apo HsSAICAR reveals there are some differences as shown in Figure (92A). Residues with a greater distance ( $>3 \text{ \AA}$ ) are parts of  $\beta$ -strand (H10-S13 and D23), loops (G14, D24, S32, D48, V64, V83, R86, G108, G128, A132, A146 and G149) or  $\alpha$ -helix (G50, R51, T54, and F59). In domain A of MtbSAICAR-ADP-AMZ complex, some loop labelled as 4, 7 and 8 (D33-S44), (D75-E82) and (G118-P127) in Figure (92B) by red, violet, and light green colours respectively. Regions around the ADP binding site are well conserved.

In domain B, many differences are found between MtbSAICAR-ADP-AMZ complex and HsSAICAR, specifically (A116-G128), (A146-G149), (F216), (D228-Y229), (A231-V234), (K240-R276), (I278) and (L287-A297) numbered from 1-8 and shown by stars in Figure (93A-93B). In region 1 (A116-G128), ( $\alpha$ -helix-loop), the loop insertion (G118-G128) indicated by a red star. Region 2, (A146-G149) (Indicated by an orange star) (loop) does not superimpose well to the corresponding region in HsSAICAR, it connects to a regular structure ( $\beta$ -strand) whereas in domain B of HsSAICAR it is a loop, suggesting flexibility. Region 4 and 5 (D228-Y229) and (A231-V234) are parts of loops. A gap is seen in the region near (D228-Y229) in MtbSAICAR. Region 6 (K240-R276) is composed of a loop between two  $\alpha$ -helices [(Q241-T248) and (E267-R276)] respectively, this loop (G249-P266) that may allow some flexibility to the  $\alpha$ -helices near the C-terminus.

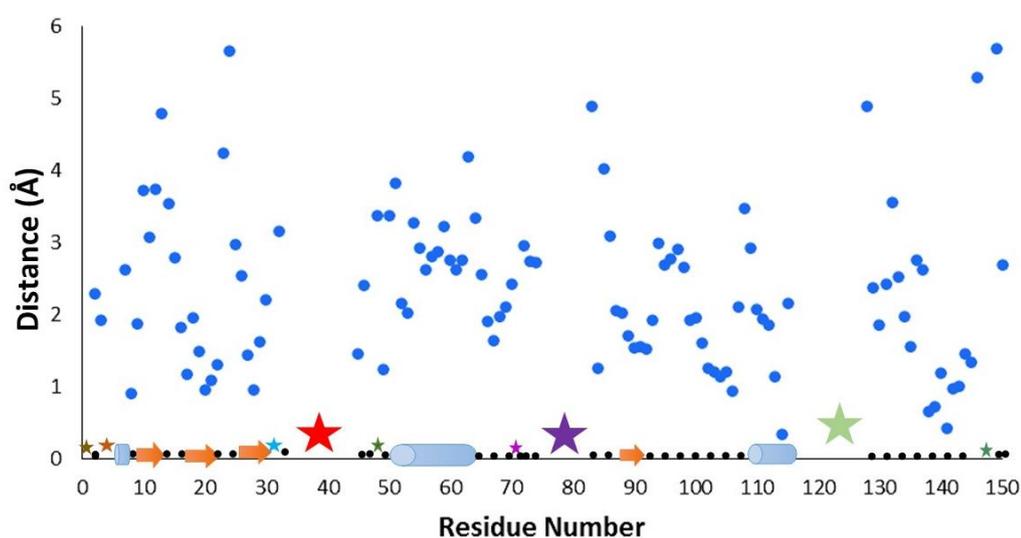
**Table (21)** 3D alignment comparison between whole chain HsSAICAR (PDB 2H31) and target MtbSAICAR-ADP-AMZ complex. The definition of each is given in Table (17). Calculation of the superimposed alignment has been performed using PDBeFold (Krissinel and Henrick, 2004).

Query		Q-Score	RMSD (Å)	N <sub>algn</sub>	N <sub>g</sub>	%seq	Target	
%sse	N <sub>resi</sub>						%sse	N <sub>resi</sub>
73	221	0.37	2.13	192	12	24	55	297

(A)

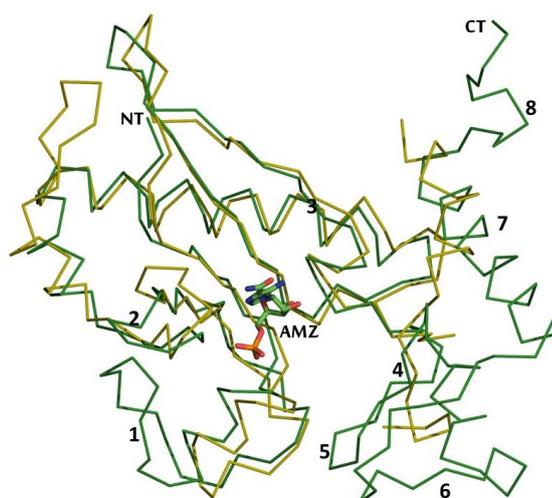


(B)

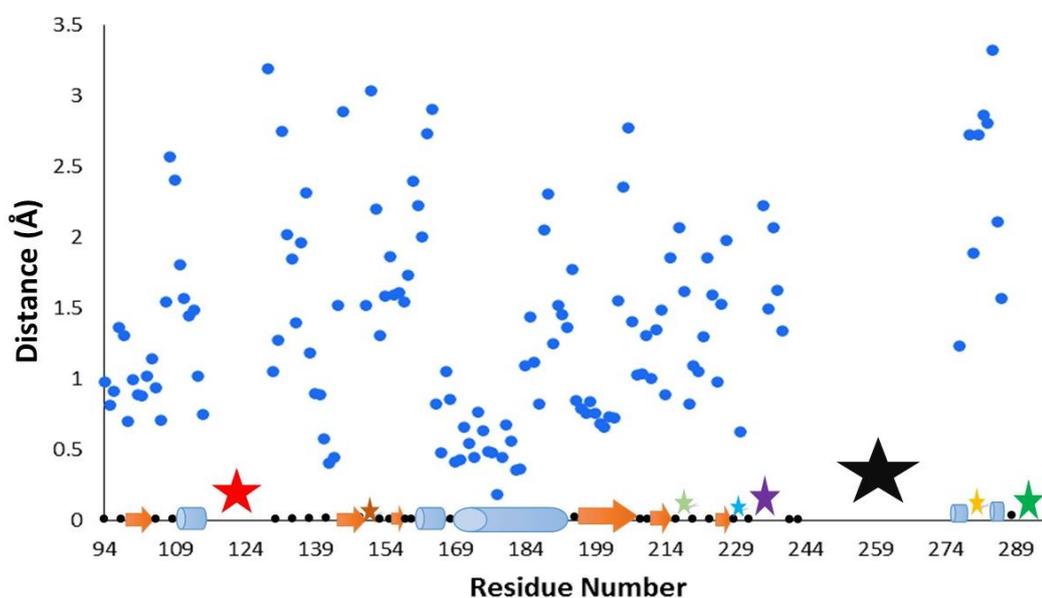


**Figure (92)** Comparison of Domain A of MtbSAICAR-ADP-AMZ complex and HsSAICAR apo structures using GESAMT method (Krissinel, 2012). (A) Overlaid  $C\alpha$  traces of domain A of HsSAICAR (coloured olive) and MtbSAICAR-ADP-AMZ complex (coloured forest green). The Q-Score value and global RMSD for 109 aligned residues are 0.35 and  $2.60\text{\AA}$  from total number 129 and 150 residues respectively. Differences between enzymes of both organisms are labelled from 1-9 and discussed below. (B) Local RMSD versus sequence position of the superimposition. M1(loop) at the N-terminal is indicated by a brownish coloured star. (A4-S6) (loop- $\alpha$ -helix) (orange star). (A31) is part of  $\beta$ -strand followed by loop suggesting flexibility (blue star). A gap in HsSAICAR exists near (D33-S44) (loop) in MtbSAICAR-ADP-AMZ complex, indicated by a red star. (P47) (loop) indicated by a dark green star. (A71) (loop) (magenta star) is located near H69 (part of the active site). Insertion loop (D75-E82) is indicated by a violet star. (A116-P127) ( $\alpha$ -helix-loop), a loop insertion (G118-P127) may induce some flexibility (Light green star). (A147-L148) (loop) at C-terminal. Only the MtbSAICAR-ADP complex residues of domain A (M1-D150) are referenced and the secondary structure content of the superimposed segments is shown loop (••),  $\alpha$ -helix (○) and  $\beta$ -strands (➡). This calculation was carried out the Superpose in CCP4 (Krissinel, 2012).

(A)



(B)

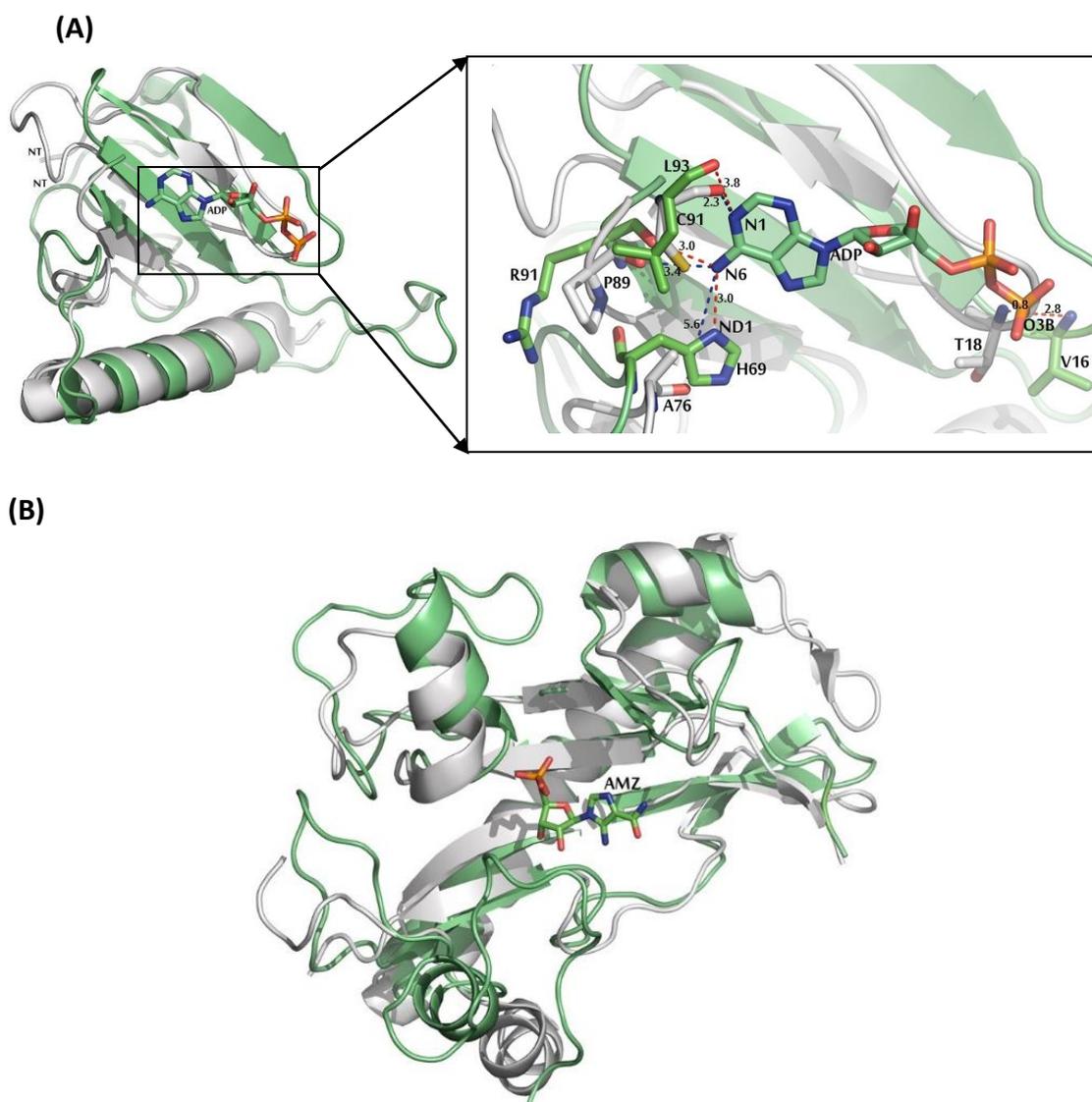


**Figure (93)** Comparison of Domain B of MtbSAICAR-ADP-AMZ complex and HsSAICAR using the GESAMT method (Krissinel, 2012). (A) *C $\alpha$*  Overlay traces of domain B of HsSAICAR (olive) and MtbSAICAR-ADP-AMZ complex (forest green). The Q-Score value and global RMSD for 131 aligned residues are 0.44 and 1.57Å from a total number of 151 and 204 residues respectively. (A116-G128), (A146-G149), (F216), (D228-Y229), (A231-V234), (K240-R276), (I278) and (L287-A297) are numbered from 1-8 and indicated by stars below. (B) Local RMSD versus sequence position of superimposition. (A116-G128), (A146-G149), (F216), (D228-Y229) and (A231-V234) are parts of a loop. A gap exists in HsSAICAR near area of (D228-Y229) of the MtbSAICAR-ADP-AMZ complex. Insertions (K240-R276) is (loop- $\alpha$ -helix-loop- $\alpha$ -helix), (I278) ( $\alpha$ -helix) and (L287-A297) (loop- $\alpha$ -helix-loop). The segments are indicated by red, orange, green, blue, violet, black, yellow and green stars respectively. Only the MtbSAICAR-ADP-AMZ complex residues of domain B (E94-A297) are referenced and the secondary structure content of the superimposed segments is shown loop (••),  $\alpha$ -helix (○) and  $\beta$ -strands(→). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

The crystal structure of HsSAICAR (Li et al., 2007) has only been solved in apo form and therefore the assignment of the binding sites are based on the superimposition of the structures. Overall, the topology of HsSAICAR enzyme consists of 14  $\beta$ -strands and 5  $\alpha$ -helices while MtbSAICAR consists of 9  $\alpha$ -helices and 17  $\beta$ -strands. The topology of MtbSAICAR is shown in Figure (54). The ADP binding site revealed some differences upon superimposing domain A of HsSAICAR with MtbSAICAR (Figure 94A).

Domain A of MtbSAICAR ADP-AMZ complex (M1-R91) is equivalent to (L7-P89) of HsSAICAR. Most of the binding site interactions are formed by the main chain atoms of the enzyme, but it is notable that the imidazole side chain of H69 makes an H-bond to N6 of ADP in MtbSAICAR, but this residue is replaced by A76 in HsSAICAR (the distance between side chain CB of A76 and N6 of the modelled ADP is 4.2Å. The main chain N of T18 makes an H-bond interaction with the phosphate group of ADP and is replaced by main the chain N of V18 in MtbSAICAR. In addition, the main chain N of L93 in MtbSAICAR interacts with N1 of ADP, which is replaced by the main chain O of C91 in HsSAICAR. In HsSAICAR, the side chain thiol group (-SH) of C91 is seen to be close to N1 of the modelled ADP and N6 of ADP at distances of 2.3Å and 2.5Å respectively.

The AMZ binding site of MtbSAICAR-ADP-AMZ (R92-P266) and HsSAICAR (Q90-L260) residues are conserved (R103, S109 and R222 and R101, S107 and R215). These residues are also conserved in the CAIR binding site of EcSAICAR (R94, S100, and R199). The AMZ binding site is shown in Figure (94B).



**Figure (94)** Structure alignment of MtbSAICAR-ADP-AMZ complex and HsSAICAR (coloured green and grey respectively). (A) ADP binding site. (B) AMZ binding site. The ADP binding sites of MtbSAICAR and HsSAICAR are shown in red and blue respectively,  $Mg^{2+}$  has been deleted from ATP for clarification. Distances are given in Å.

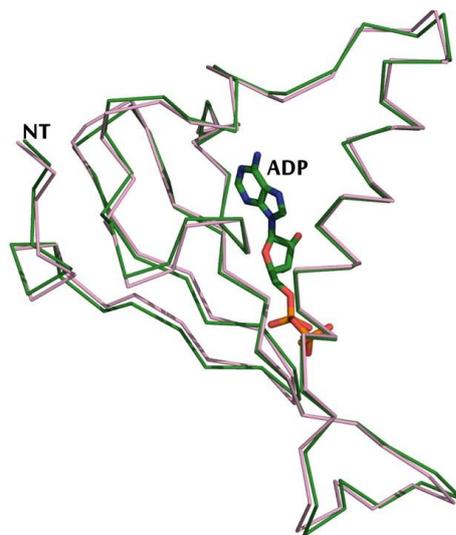
### 5.3.6. Structure alignment of MtbSAICAR-ADP-AMZ and *M. abscessus* SAICAR (MaSAICAR)

*Mycobacterium abscessus* (MaSAICAR) is closely similar to MtbSAICAR with a high sequence identity of 76%. The structure of the enzyme was solved in the apo form, but no more information on complexes has been published. Baugh et al. suggested that the structure of the MaSAICAR would be a good model for MtbSAICAR, as their attempts to form suitable MtbSAICAR crystals failed (Baugh et al., 2015). The absence of MaSAICAR complex structures encouraged us to solve the crystal structure of MtbSAICAR and complexes, and a comparison of MtbSAICAR and MaSAICAR allows us to see if MaSAICAR is indeed a good model for the pathogen's enzyme and to define any further differences. The alignment between MtbSAICAR-ADP-AMZ complex and MaSAICAR is assessed using Q-score values as above. The alignment has been done for the entire length of chain and for the separated domains as is described in Section 5.3.1. The Q-score value of the entire enzymes is 0.91 as shown in Table (22). The Q-score value of the divided domains (A; 0.97) is higher while (B; 0.90) is nearly similar to the whole chain. The aligned domain A of MtbSAICAR-ADP-AMZ complex and MaSAICAR revealed that both domains are highly conserved in terms of the fold. The area around ADP binding site correlates is well and all residues have distance value less than 3Å showing a conserved fold. There are few residues with different secondary structure but they are not part of the active site. In domain B some differences are found between MtbSAICAR-ADP-AMZ complex and MaSAICAR specifically an insertion and gap (A147-G149) and (D259-R260) which is part of loop as indicated in Figure (96A-96B). these difference may indicate that MaSAICAR is not a perfect model for the MtbSAICAR enzyme.

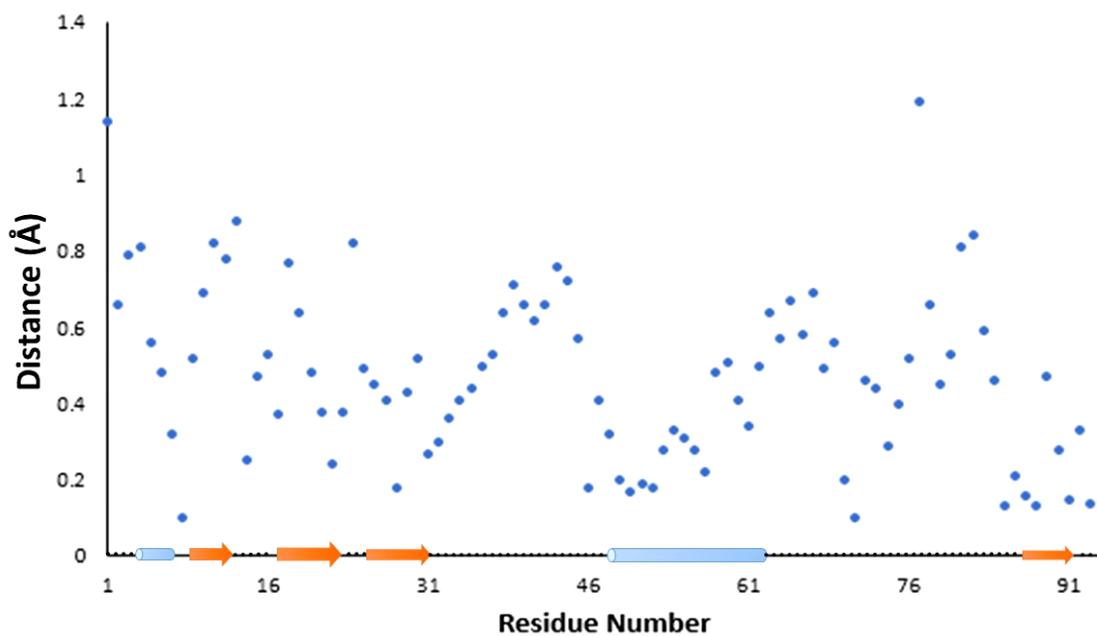
**Table (22)** 3D alignment comparison between whole chain MaSAICAR (PDB 3R9R) and target MtbSAICAR-ADP-AMZ complex. The definition of each is given in Table (17). Calculation of the superimposed alignment has been performed using PDBeFold (Krissinel and Henrick, 2004).

Query		Q-Score	RMSD (Å)	N <sub>algn</sub>	Ng	%seq	Target	
%sse	N <sub>resi</sub>						%sse	N <sub>resi</sub>
95	295	0.91	0.72	290	2	76	90	297

(A)

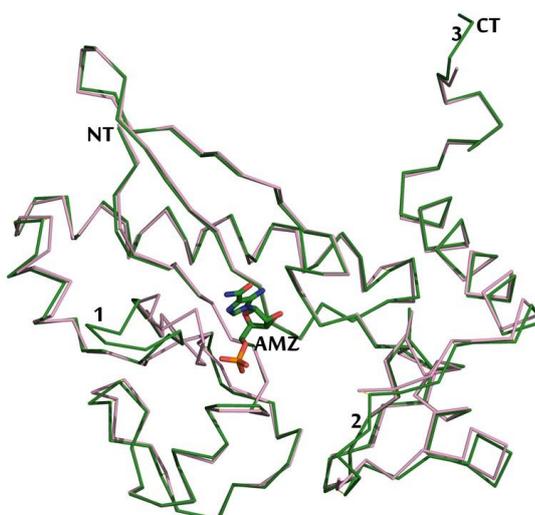


(B)

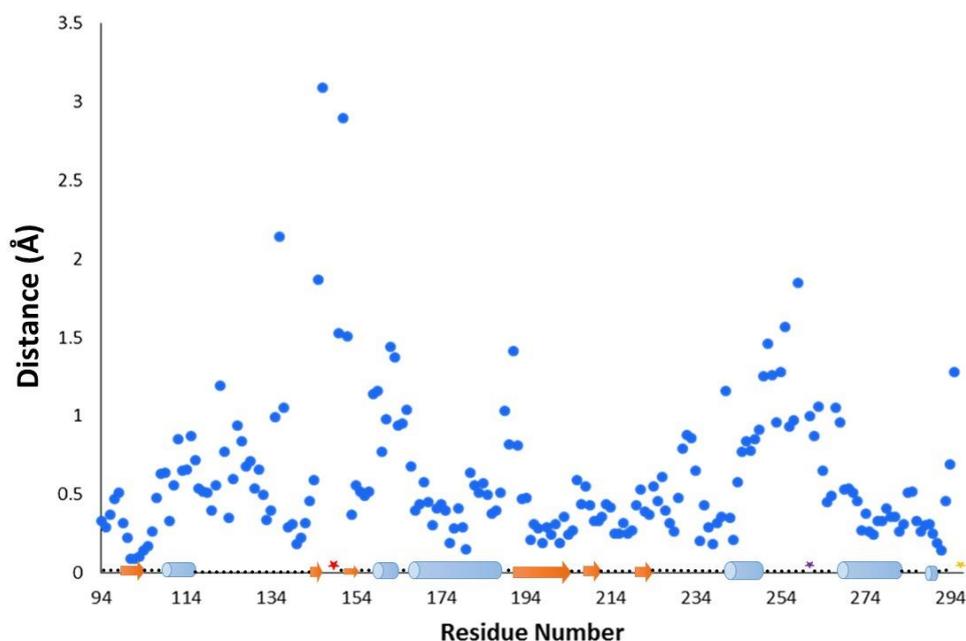


**Figure (95)** Comparison of Domain A of MtbSAICAR-ADP-AMZ complex and MaSAICAR using the GESAMT method (A) C $\alpha$  Overlay traces of domain A MaSAICAR (pink) and MtbSAICAR-ADP-AMZ complex (forest green). The Q-Score value and global RMSD for 93 aligned residues are 0.97 and 0.52Å from a total number of 93 and 93 residues respectively. (B) Local RMSD versus sequence position. All residues have distance values ( $>3\text{\AA}$ ). Only the MtbSAICAR-ADP-AMZ complex residues of domain A (M1-L93) are referenced and the secondary structure content of the superimposed segments are shown loop (••),  $\alpha$ -helix (○) and  $\beta$ -strands (➔). This calculation was carried out using superpose utility in CCP4 (Krissinel, 2012).

(A)



(B)

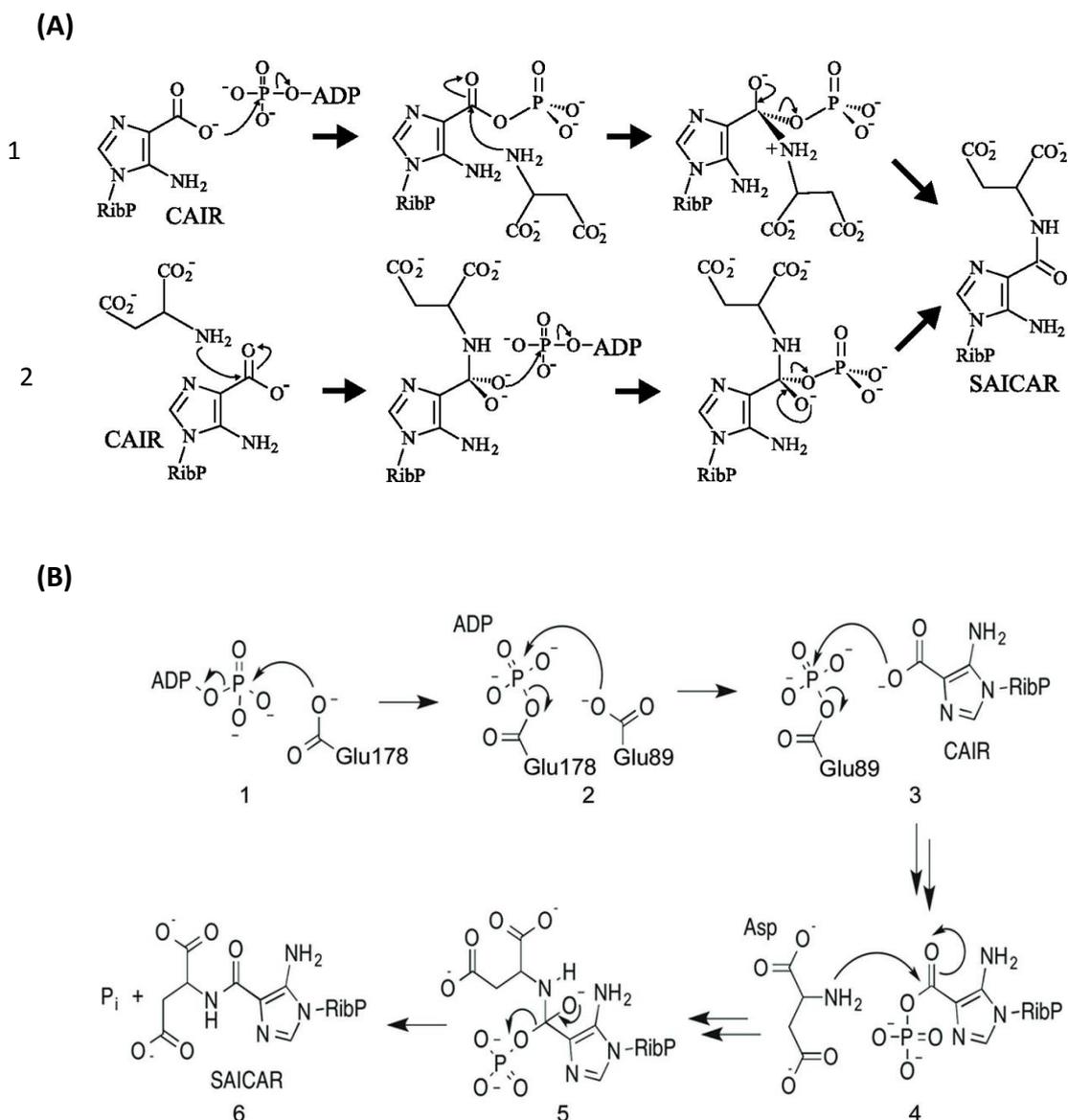


**Figure (96)** Comparison of Domain B of MtbSAICAR-ADP-AMZ complex and MaSAICAR structures using the GESAMT method (Krissinel, 2012). (A) C $\alpha$  Overlay traces of domain B of MaSAICAR (pink) and MtbSAICAR-ADP-AMZ complex (forest green). The Q-Score value and global RMSD for 197 aligned residues are 0.90 and 0.74Å from a total of 200 and 204 residues respectively. Insertion segment (A147-G149) and a gap (D259-R260) in MaSAICAR, both have been labelled as 1 and 2 and shown below as red and violet stars. (B) Local RMSD versus sequence position. The residue (A146) with a high distance value >3Å is part of a loop. Only the MtbSAICAR-ADP complex residues of domain B (E94-L297) are referenced and the secondary structure content of the superimposed segments is shown loop (••),  $\alpha$ -helix (blue oval) and  $\beta$ -strands (orange arrow). This calculation was carried out using superpose in CCP4 (Krissinel, 2012).

#### **5.4. Mechanism and action of SAICAR Synthetase (SAICARs)**

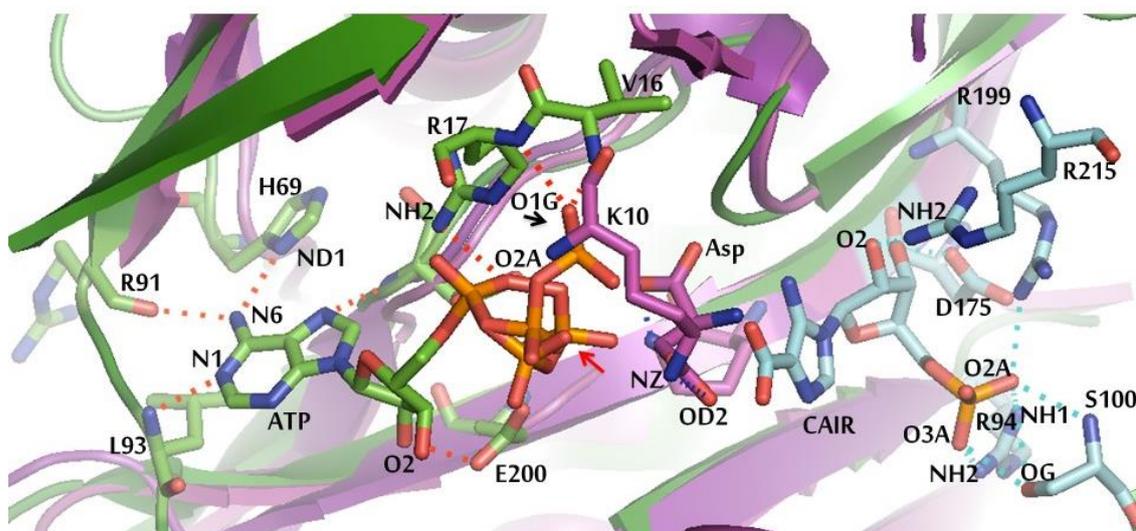
Fully understanding the mechanism and action of the enzyme requires the detailed understanding of the three-dimensional structure at the atomic level. SAICAR Synthetase utilises ATP, aspartate and CAIR to produce SAICAR. Two competing models have been outlined for the catalysis mechanism. Ginder et al., proposes the reaction of EcSAICAR as being either that the CAIR is first phosphorylated by ATP and then the CAIR phosphoester is attacked by the amino group of aspartate, forming a tetrahedral intermediate that collapses to release phosphate and leave the amide of the product SAICAR, or that initially the amino group of aspartate is attached to the carboxyl group of CAIR, forming a tetrahedral intermediate, then ATP phosphorylates this, releasing ADP, afterwards this phosphorylated intermediate collapses to release phosphate and form SAICAR. As no electron density appeared for aspartate in the crystal structure, it was suggested that the former (first) phosphorylation mechanism is the most likely to happen (Ginder et al., 2006) (Figure 97A).

When reporting the structures of SpSAICAR and complexes, Wolf et al., propose the involvement of phospho-enzyme steps, whereby E178 is first phosphorylated by ATP, and the phosphate is then transferred to E89 and then on to CAIR, when it is displaced by aspartate with the subsequent step being similar (Wolf et al., 2014) (Figure 97B).



**Figure 97** Possible models of formation of SAICAR by SAICARs. (A) Two schemes from (Ginder et al., 2006). (B) The contribution of the side chains of E178, E89 in  $\gamma$ -phosphate transition (Wolf et al., 2014). ATP,  $Mg^{2+}$ , Aspartate, CAIR, and residues involved in are determined.

In a hybrid model (Figure 98) using the structure of the MtbSAICAR-ATP complex, where the phosphate groups of ATP are disordered and are seen in dual conformations. The  $\gamma$ -phosphate of ATP is seen to be distant (around 5Å) from the carboxyl group of CAIR. The amino group of aspartate and the carboxyl group of CAIR are close to each other, indicating the carboxyl group of CAIR undergoes a possible conformation change to prevent clashes.



**Figure (98)** Proposed model by superimposition of MtbSAICAR, SpSAICAR, and EcSAICAR. ATP (MtbSAICAR) is shown in linear and bent conformations (labeled by black and red arrows respectively). Aspartate (Asp) (SpSAICAR) and CAIR (EcSAICAR). Residues colored based on their associated ligands.

## 5.5. Summary

The superimposition of the structures of apo and MtbSAICAR-ligand complexes shows them to be very similar except for minor differences in secondary structure between MtbSAICAR (apo and ACP-IMP) and (ATP and ADP-AMZ) given in Figures (76-79). AMZ and IMP complexes have the same binding sites as each other but the adenine ring is oriented differently (Figure 78). Although SAICARs is in a monomeric form in *M. tuberculosis*, *M. abscessus* and *Saccharomyces cerevisiae*, the oligomeric state appears to be different in other organisms, it is a dimer in EcSAICAR, SpSAICAR, and TmSAICAR, and octamer in HsSAICAR (Li et al., 2007). Structures were compared by superimposition both globally and locally by domain.

Examination of the interface of TmSAICAR revealed the absence of segments, which are expected to have a role in preventing the dimerisation of MtbSAICAR and ScSAICAR (Zhang et al., 2006) (Figure 82A). Superimposition of MtbSAICAR with ScSAICAR (solved with ADP and AMZ) showed half the number of ADP and AMZ molecules in the structure, shown in Figure (85A and B). The phosphate groups of ADP appeared to be bent in ScSAICAR and EcSAICAR but linear in MtbSAICAR. In the MtbSAICAR-ADP binding site, two residues are bound differently compared to the ADP binding site of

TmSAICAR and ScSAICAR but this is a main chain interaction and may not affect the behaviour of the binding site (Figure 82B and 85A). In MtbSAICAR ADP binding site, the side chain of a polar residue is replaced by the side chain of another polar residue compared to EcSAICAR-ADP complex (Figure 88A). In HsSAICAR, the ADP modelled binding site revealed two changed residues when compared to the MtbSAICAR-ADP complex (Figure 94A). A proposed model of superimposed structures of EcSAICAR, MtbSAICAR, and SpSAICAR is shown in Figure (98).

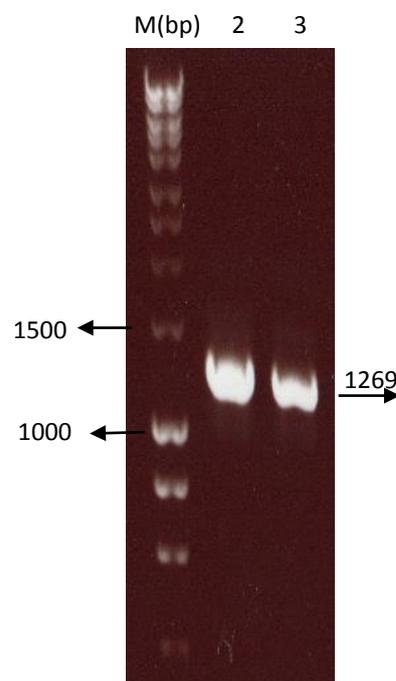
## Chapter 6. Characterisation studies of purD of *M. tuberculosis* (MtbpurD)

### 6.1. Introduction

Phosphoribosylamine glycine ligase (purD) is the enzyme that catalyses the second step of the *de novo* purine pathway, it is a mono functional enzyme in bacteria. In humans, in contrast, it is part of the trifunctional enzyme (HsGART) (Chapter1, Section1.3.3.2) (Adam, 2005). This chapter describes the work done to characterise the MtbpurD enzyme using biophysical methods.

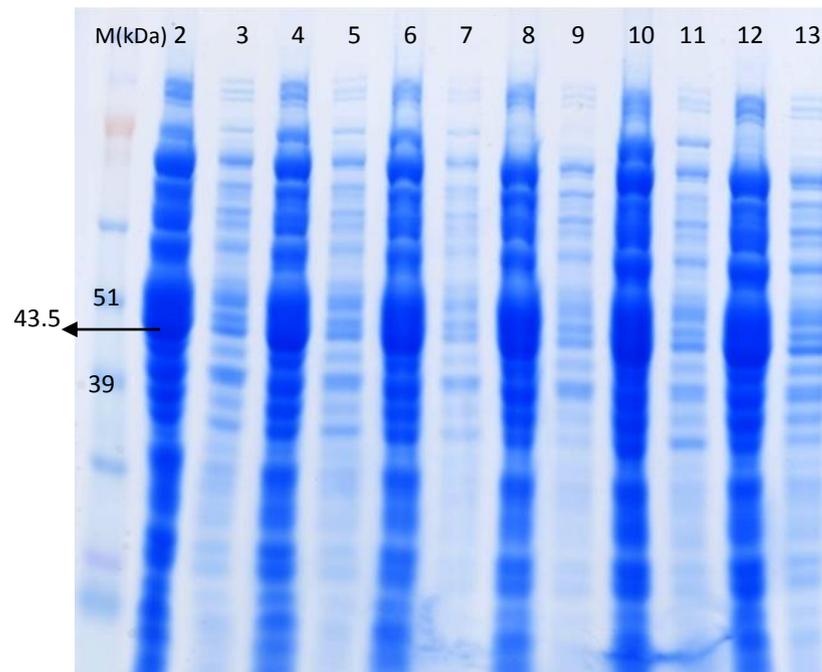
### 6.2. Preparation of MtbpurD

The Open Reading Frame of the gene encoding purD *M. tuberculosis* (MtbpurD) was amplified (Figure 99), and cloned into the expression vector pLEICS-01 by the protein expression facility, PROTEX, at the University of Leicester and overexpressed in *E. coli* strain Rosetta (DE3) as described in Chapter2, Section 2.2.2-2.2.4.

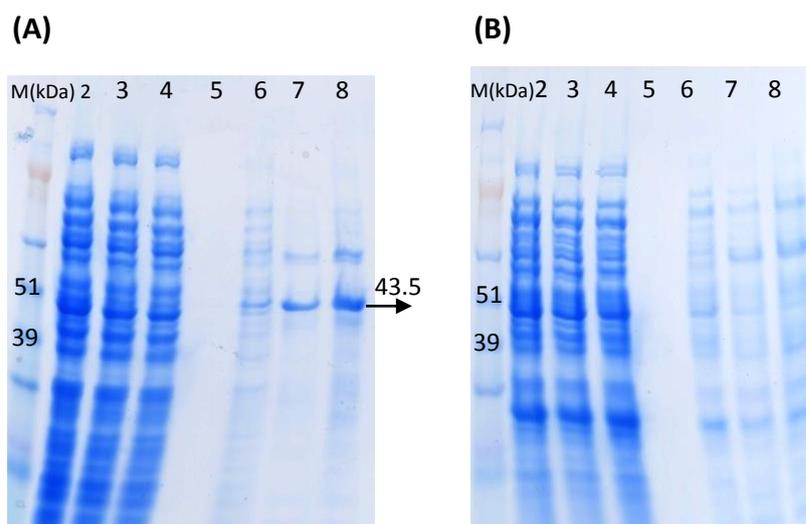


**Figure (99)** ORF amplification of the *MtbpurD* gene, Lane1. M. Molecular weight DNA marker (Bioline 1 kb hyper ladder). Lane 2. PCR product obtained using 1.5mM MgSO<sub>4</sub>. 3. PCR product obtained in the presence of 6% DMSO and 2.5mM MgSO<sub>4</sub>.

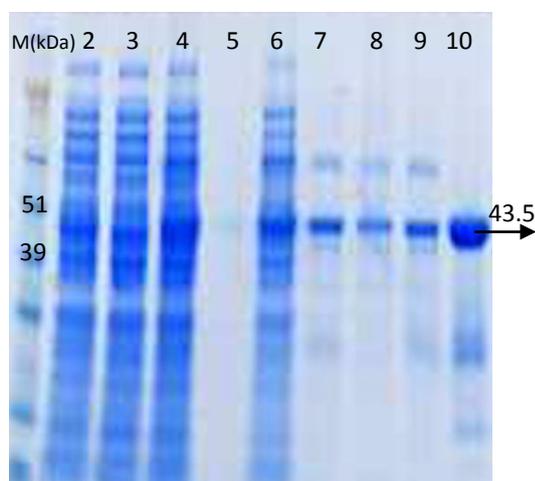
Initial expression trials (Figure 100) did not provide unambiguous optimal conditions for protein expression, therefore further pilot trials were undertaken (Figure 101). These trials explored concentrations for the inducer IPTG 50 $\mu$ M and 400 $\mu$ M post-induction temperatures between 20 $^{\circ}$ C and 37 $^{\circ}$ C. A soluble protein band was obtained when protein expression was induced by 50 $\mu$ M IPTG and grown at 20 $^{\circ}$ C overnight. Purification by Nickel affinity chromatography using Ni-NTA resin and assessed by SDS-PAGE is shown in (Figure 102). The 6xHis N-terminal tag was not cleaved after purification.



**Figure (100)** Optimisation of overexpression conditions for MtbpurD in *E. coli* Rosetta (DE3). Lane 1: M. SeeBlue2 (Life Technologies) protein marker. Lane 2: Supernatant from cells induced with 50 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 3: Pellet from cells induced with 50 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 4: Supernatant from cells induced with 150 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 5: Pellet from cells induced with 150 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 6: Supernatant from cells induced with 250 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 7: Pellet from cells induced with 250 $\mu$ M IPTG at 20 $^{\circ}$ C overnight. Lane 8: Supernatant from cells induced with IPTG at 20 $^{\circ}$ C overnight. Lane 9: Pellet from cells induced with IPTG at 20 $^{\circ}$ C overnight. Lane 10: Supernatant from cells induced with 400 $\mu$ M IPTG at 37 $^{\circ}$ C overnight. Lane 11: Pellet from cells induced with 400 $\mu$ M IPTG at 37 $^{\circ}$ C overnight. Lane 12: Supernatant from cells uninduced at 37 $^{\circ}$ C overnight. Lane 13: Pellet from cells uninduced at 37 $^{\circ}$ C overnight.

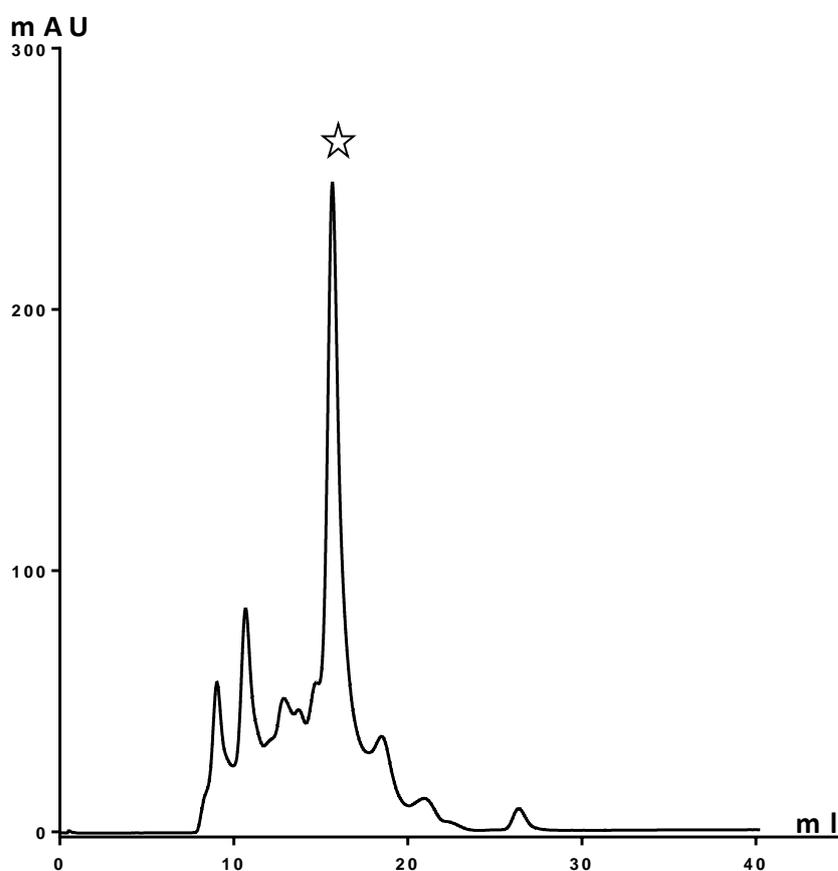


**Figure (101)** Small-scale purification of MtbpurD using *E. coli* Rosetta (DE3). (A) Induced with 50 $\mu$ M IPTG at 20°C overnight. Lane 1: M. SeeBlue2 (Life Technologies) protein marker. Lane 2: soluble cell extract. Lane 3: unbound protein flow through of Ni-NTA column. Lane 4: Ni-NTA bead. Lane 5: unbound proteins (with 20mM imidazole: fourth wash). Lane 6: Ni-NTA bead showing MtbpurD bound. Lane 7: MtbpurD solution (eluted with 250mM imidazole). Lane 8: Ni-NTA bead. (B) Induced with 400 $\mu$ M at 37°C overnight, Lane 1: M. SeeBlue2 (Life Technologies) protein marker. Lane 2: soluble cell extract. Lane 3: unbound protein flow through of Ni-NTA column. Lane 4: Ni-NTA bead. Lane 5: unbound proteins (with 20mM imidazole: fourth wash). Lane 6: MtbpurD Ni-NTA bead. Lane 7: MtbpurD solution (with 250 mM imidazole). Lane 8: Ni-NTA bead.

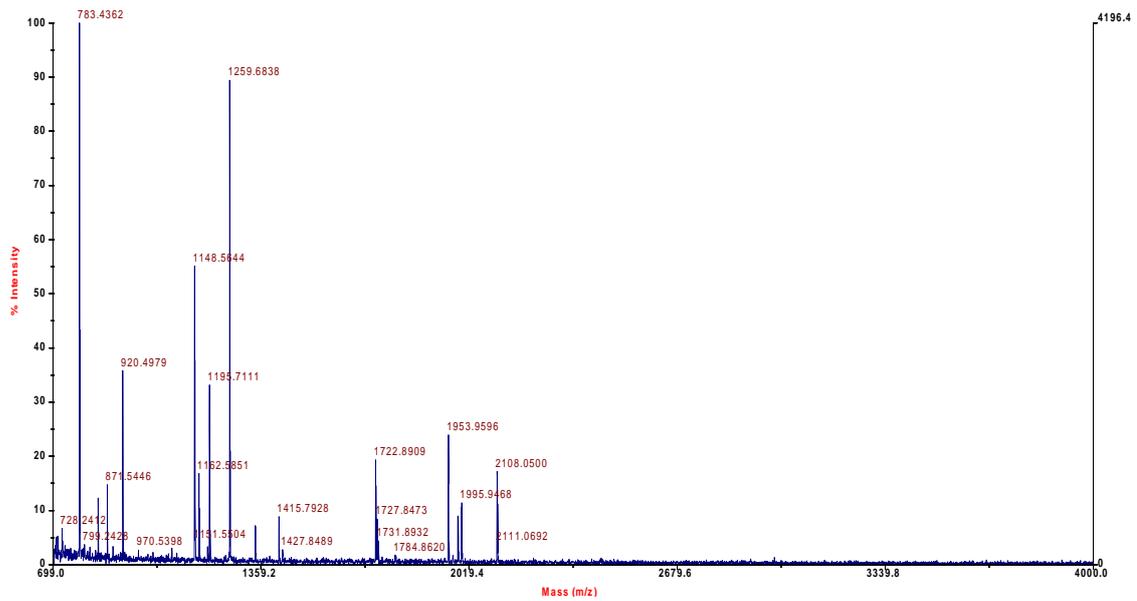


**Figure (102)** Purification of MtbpurD expressed with 6xHis affinity tag. Samples were taken at each stage of the purification (10  $\mu$ l of sample and 10 $\mu$ l 2xSDS sample buffer) and run on a 4-12% polyacrylamide gel. Lane 1: M. SeeBlue2 (Life Technologies) protein marker. Lane 2: soluble cell extract. Lane 3: unbound protein flow through of Ni-NTA column, lane 4: Ni-NTA bead. Lane 5: unbound proteins (with 20mM imidazole: fourth wash). Lane 6: MtbpurD Ni-NTA bead. Lane 7: MtbpurD solution (with 250mM imidazole). Lane 8: Ni-NTA bead. Lane 9: Overnight TEV cleavage at 4°C. Lane 10: Concentrated MtbpurD after gel filtration.

The oligomeric state of the protein was investigated by further purification by gel filtration. These experiments suggested that the protein is a monomer in solution as seen in Figure (103). MtbpurD elutes in the fraction that standard protein calibrations (supplied by the manufacturers (GE Healthcare)) show correspond to the expected monomeric M.W. of 43.5kDa. The calibration curve is given in Chapter2, Section 2.2.4.4. The identity of the protein was confirmed by tryptic digest mass spectrometry (Figure 104) in the Protein Nucleic Acid Chemistry Laboratory (PNAACL), University of Leicester.



**Figure (103)** Gel filtration of MtbpurD preparation using Superdex 200 Increase 10/300 GL (GE Healthcare). Absorbance units are represented as (mAU). The peaks from earlier fractions were evaluated by SDS-PAGE and did not show proteins of MW 43.5kDa. The peak (indicated by a star) corresponds to MtbpurD with expected MW 43.5kDa (from the calibration).

**(A)****(B)**

Range	Observed masses	Calculated masses	Sequence
24-32	871.5480	870.5287	R.VLVIGSGAR.E
43-60	1722.8950	1721.90.61	K.DPQVSGLIVAPGNAGTAR.I
61-80	2108.0540	2107.0909	R.IAEQHDVDITSAEAVVALAR.E
138-156	1994.9490	1993.9705	R.TANSEIVDSPAHLDAALDR.F
157-169	1340.6990	1339.6925	R.FGPPAGDPAWVVK.D
157-172	1726.8390	1725.8475	R.FGPPAGDPAWVVKDDR.L
222-234	1426.8380	1425.8232	R.TVVVPLLPQDFK.R
269-279	1195.7120	1194.6972	R.IVEPVAELVR.R
281-297	1784.8820	1783.9291	R.GSSFCGLLYVGLAITAR.G
298-307	1148.5640	1147.5444	R.GPAVVEFNCR.F
336-343	920.4960	919.4763	K.LADFGELR.W
346-364	1984.0220	1983.0538	R.DGVAVTVVLAENYPGRPR.V
365-384	1952.9730	1951.9964	R.VGDVVVGSEAEGVLHAGTTR.R
397-411	1415.7890	1414.7780	V.LSVVGTGADLSAAR.A
412-422	1259.6830	1258.6669	R.AHAYEILSSIR.L
423-429	783.4360	782.4188	R.LPGGHFR.S

**Figure (104)** Matrix-assisted laser desorption ionisation time of flight (MALDI-TOF) Mass Spectrometry identification of MtbpurD (Susnea et al., 2012). Peptides from a tryptic digest of the MtbpurD preparation (trypsin cuts at the C end of KR residues in the protein, except if the next residue is proline) were analysed using MALDI-TOF Mass Spectrometry. (A) Intensity vs Mass (m/z) spectrum analysis of MtbpurD tryptic digested. (B) Comparison of matched peptide masses (observed and calculated) calculated peptide masses are derived from the SwissProt database (Bairoch and Boeckmann, 1991) by the MASCOT Peptide Mass Fingerprint search engine (<http://www.matrixscience.com>). Figure supplied by PNAFL, University of Leicester.



The secondary structure prediction was compared to the measurements from CD experiments. Prior to taking the CD measurements, the protein buffer was exchanged with 20mM sodium phosphate buffer pH7.4 and 100mM sodium fluoride. The secondary structure content of the MtbpurD in the presence and absence of ligands was calculated from the CD data using the method of Raussens et al., 2003 (Raussens, Ruyschaert, and Goormaghtigh, 2003) and are consistent with the predicted secondary structure content as shown in Table (23).

**Table (23)** Percentages of secondary structure for MtbpurD apo and liganded forms from CD measurements. (A) Apo MtbpurD (B) MtbpurD-Glycine (Gly) (C) MtbpurD-Glycine-ATP (D) MtbpurD-Glycine-ATP-D-Ribose 5-phosphate disodium salt (R5P). The proportions shown are calculated using the online server of Raussens et al., (<http://perry.freeshell.org/raussens.html>) (Raussens, Ruyschaert, and Goormaghtigh, 2003).

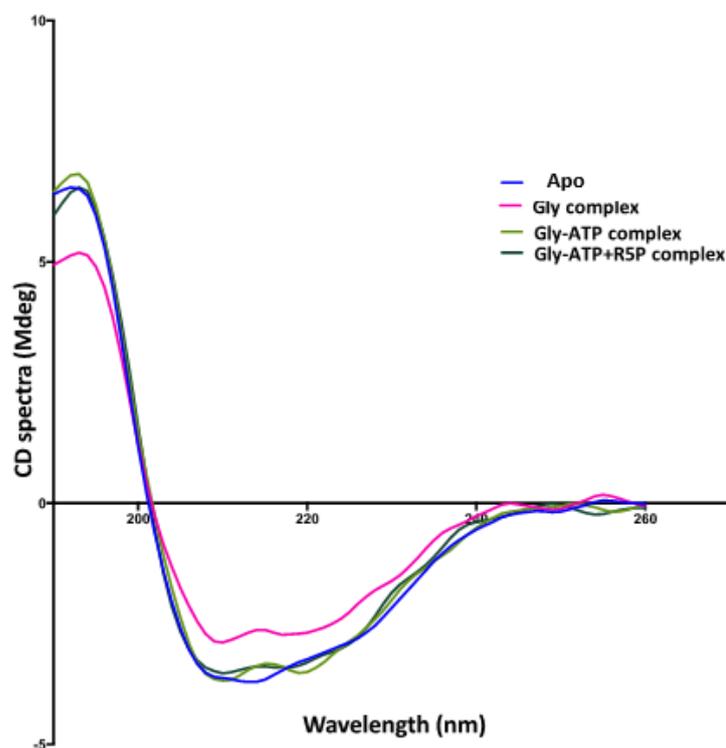
<b>(A)</b>		<b>(B)</b>	
<b>Apo MtbpurD</b>		<b>MtbpurD-Gly</b>	
Helix	45.7%	Helix	47.2%
Beta	7.2 %	Beta	8.2%
Turn	12.5%	Turn	12.5%
Random	30.4%	Random	29.0%
Sum	95.8%	Sum	96.9%

<b>(C)</b>		<b>(D)</b>	
<b>MtbpurD-Gly-ATP</b>		<b>MtbpurD-Gly-ATP-R5P</b>	
Helix	35.1%	Helix	40.3%
Beta	14.7%	Beta	13.2%
Turn	12.5%	Turn	12.5%
Random	31.6%	Random	30.3%
Sum	93.9%	Sum	96.3%

Figure (106) shows the CD spectra of these preparations, the spectra demonstrate the characteristics of the  $\alpha$ -helix dominant protein. The presence of (and presumed binding to) a ligand does not significantly change the secondary structure content of the protein. In addition, the measured secondary structure of MtbpurD was compared with secondary structure content of homologous enzymes from other organisms

whose crystal structure are deposited in the PDB. It was seen that structure of Phosphoribosylamine glycine Ligase from *T. thermophilus* (TtpurD) (PDB 2IP4), *E. coli* purD (EcpurD) (PDB 1GSO), *B. subtilis* purD (BspurD) (PDB 2XD4) are composed of 31%  $\alpha$ -helix+26%  $\beta$ -strand, 28%  $\alpha$ -helix+30%  $\beta$ -strand, 31%  $\alpha$ -helix+29%  $\beta$ -strand respectively (Sampei et al., 2010; Wang et al., 1998; Bertrand et al., 2011). Based on this computational and experimental work, it is likely that the structure of MtbpurD is predominantly  $\alpha$ -helix.



**Figure (106)** Circular Dichroism (CD) of MtbpurD. (A) CD spectra data of the MtbpurD in buffer containing 20mM Phosphate pH7.4 and 100mM Sodium fluoride, concentration of 0.15mg/ml. The blue color shows the CD spectra obtained at 20°C of apo MtbpurD, the pink color shows the CD spectra obtained at 20°C of MtbpurD-Glycine complex, the light green color shows the CD spectra obtained at 20°C of MtbpurD-Gly-ATP complex and the dark green color shows the CD spectra obtained at 20°C of MtbpurD-Glycine- ATP+R5P complex. Gly represents Glycine and R5P represents D-Ribose 5-phosphate disodium salt.

#### 6.4. Investigating thermal stability of MtbpurD by thermal shift assay

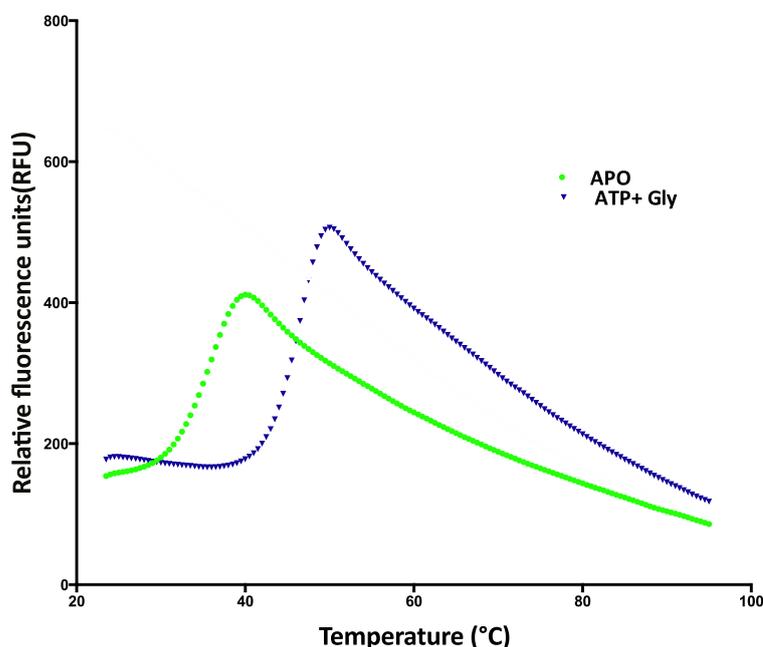
Prior to crystallisation trials, an investigation into the stability of the enzyme in apo and liganded forms (including natural and analogues compounds) was undertaken. The effect of Tris-HCl buffer (the purification buffer) was also explored by comparison with MtbpurD purified in presence of HEPES. The thermal profile of apo and liganded forms

of MtbpurD enzyme were determined between (20-95°C) at 0.5°C increment using a custom designed thermal shift assay in 96 well plates This was also used to compare the effect on the protein's stability of different buffer conditions and components (Chapter2, Section 2.4.2).

The thermal shift dye binding assay used to determine the temperature at which 50% of the protein becomes unfolded. Apo MtbpurD displays a melting temperature of 35.0°C (Figure 107). The thermal stability of the protein is increased in the presence of 10mM ATP+10mM Glycine (Table 24) indicating that the enzyme is more stable when in complex with ATP and Glycine.

**Table (24)** Assessing melting temperature of different condition from MtbpurD in Tris-HCl pH8.0, ligands used in a concentration of 10mM.

MtbpurD Conditions	Tm (°C)	Change in Tm(°C)
Apo	35.0	—
ATP+ Glycine	45.0	10°C

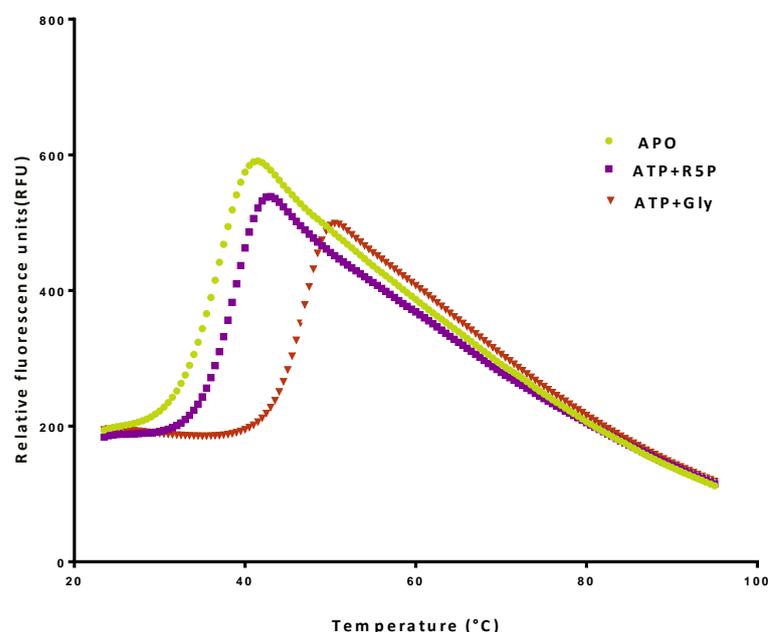


**Figure (107)** Thermal stability of purD *M. tuberculosis* (MtbpurD) enzyme in apo and in the presence of ligands. Tm is estimated from the graph profile. The ligands used throughout this experiment are prepared using 20mM Tris-HCl pH8.0, 50mM NaCl, Each sample was then mixed with 1/1000 Sypro orange dye. The concentration of ATP+Glycine (Gly) is 10mM.

In HEPES buffer, the melting temperatures of MtbpurD in the apo form and with ligand are given in Table (25). The results suggest the enzyme is stable in the presence of ATP+ Glycine both in Tris-HCl buffer and HEPES.

**Table (25)** Assessing melting temperature of different condition from MtbpurD in HEPES pH8.0, 10mM of each ADP, ATP and Glycine and 5mM R5P.

MtbpurD Condition	T <sub>m</sub> (°C)	Change in T <sub>m</sub> (°C)
Apo	36.83±0.13	—
ATP+R5P	38.43±0.10	1.6
ATP+Glycine	46.44±0.07	9.6

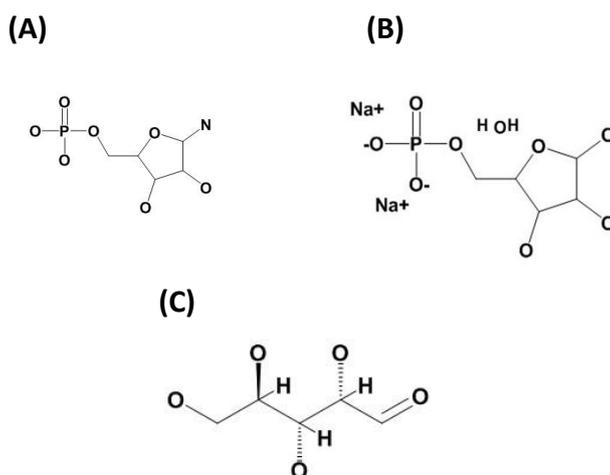


**Figure (108)** Thermal stability of apo and liganded form of purD of *M. tuberculosis* (MtbpurD). Each form is prepared in 20mM HEPES pH8.0, 50mM NaCl. Each sample was then mixed with 1/1000 Sypro orange dye.

### 6.5. Crystallisation trials of MtbpurD

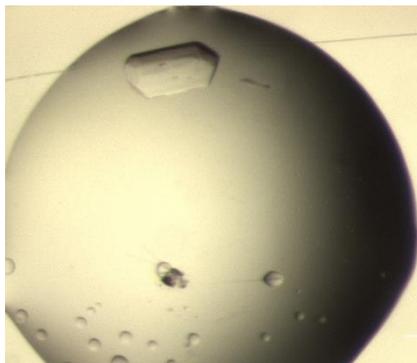
Crystallisation trials of the protein were attempted to further understand the enzyme by seeing details of the structure and the protein-ligand interactions that would become apparent from the three-dimensional structure. The characterisation by CD and thermal shift assays shows the protein to be folded, stable and therefore ready for crystallisation trials. However, the attempts at crystallisation did not yield diffraction quality crystals. Trials were carried using three different approaches and using the

various commercially available screens including JCSG+, PACT, ProPlex, Midas, Morpheus and Nuclear Receptor-Ligand Binding Domain™ (NR-LBD). Batch crystallogenesis was attempted with the Oryx4 crystallisation robot (Douglas Instruments), sitting drop vapor diffusion trials used the Mosquito crystallisation robot (TTP Labtech). In addition to this, microdialysis was also tried however, none of these methods produced a protein crystal. Many parameters that can have been reported to improve crystal formation were also investigated. The included checking the purity of the protein preparation, testing different incubation temperatures for the crystal growth plate (4°C and room temperature), varying the pH and the buffer used for purification (HEPES and Tris-HCl), varying protein concentration (5.3 mg/ml, 15 mg/ml and 30 mg/ml). As ligand binding increases the stability of the protein, the crystallisation conditions were screened in the presence and absence of ligands, including the substrate analogue D-Ribose 5-phosphate disodium salt (R5P), arabinose (Figure 109) and natural substrates of the enzyme ATP and Glycine. Unfortunately, none of these trials produced protein crystals. In addition to these trials, macro and micro seeding was used to attempt to improve the quality of putative initial hits. Similar to all the previous crystallisation trials, no protein crystals were formed. In most of the wells light to heavy precipitates were formed. Sometimes needle-shaped and larger crystals were formed, but these turned out to be salt.

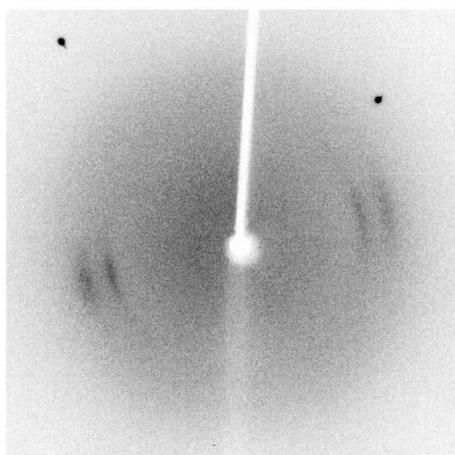


**Figure (109)** Substrate analogues of natural substrate 5-P-β-D-ribosylamine (PRA). (A) 5-P-β-D-ribosylamine (PRA). (B) D-Ribose 5-phosphate disodium salt (R5P). (C) D-arabinose. These compounds were drawn using <https://pubchem.ncbi.nlm.nih.gov/edit2/index.html>.

An example of such a salt crystal is shown in Figure (110), this was formed in the presence of 0.2M Lithium sulfate 0.1M Sodium acetate pH4.5, 50% polyethylene glycol 400. The crystal was mounted on a loop with a size of around 0.1mm and was prechecked using the home X-ray source (University of Leicester). The diffraction pattern suggested that the crystal is a salt crystal as intense spots were seen at high resolution, shown in Figure (111).



**Figure (110)** A crystal grown from a solution of MtbpurD. The crystal was grown in the presence of 0.2M Lithium sulfate 0.1M Sodium acetate pH4.5, 50% polyethylene glycol 400.



**Figure (111)** Representative diffraction pattern of the crystal. The crystal was grown in the presence of 0.2M Lithium sulfate 0.1M Sodium acetate pH4.5, 50% polyethylene glycol.

## 6.6. Homology model alignment of MtbpurD

Attempts to crystallise MtbpurD failed and experimental structure determination was therefore not possible, as sequence alignment analysis indicated that there were some differences in the ATP binding site between HspurD and MtbpurD, a homology model of MtbpurD was built to enable further analysis. A good homology model is

expected to provide a more reliable adjustment for insertions and deletions than simple sequence alignment. A good model should allow a meaningful comparison of the residues involved in the interactions between protein and ligands and therefore make it possible to see whether the ligand and substrate binding sites are conserved among purDs from different organisms.

A homology model of MtbpurD was generated by the protein fold recognition server, PHYRE<sup>2</sup> (Kelley et al., 2015) using six templates *Aquifex Aeolicus* (AaPurD) (PDB 2YYA) *Ehrlichia chaffeensis* (PDB 3LP8), *Neisseria gonorrhoeae* (PDB 5vev), *Geobacillus kaustophilus* (PDB 2ys6), Human (PDB 2qk4), *Bacillus subtilis* (PDB 2xd4) (Chapter 2, Section 2.3.1) as shown in Figure (112(1)), with the aim of showing the sequence differences between homologous structures in a structural context. Overall, the model is superimposed onto EcpurD (PDB 1GSO) as a reference structure. The model is composed of four domains N (M1-R116), B (T117-D188), A (G189- G326), C (V327-A416) coloured light pink, green, violet, and olive respectively in Figure (112(1)). Superimposition of the N domains of these enzymes suggests that there is little evidence for structural divergence; the only difference is that in the MtbpurD model, the region Q43-D45 (which is predicted to be composed of a loop) corresponds to a region that is composed of  $\beta$ -strands. In addition to this, the models also show that most of the residues are conserved with only minor differences such as a positively charged residue R39 of MtbpurD being replaced by the hydrophobic residue L39 in EcpurD.

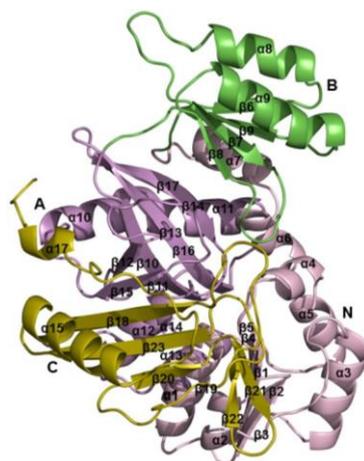
Domain B of both structures is highly conserved, two segments in EcpurD (G174-A180) and (V128-E129) are part of loop whereas the corresponding segments in MtbpurD are part of an  $\alpha$ -helix. ATP binding site is located between A and B domains. In MtbpurD, domain A is composed of a core of  $\beta$ -strands flanked by  $\alpha$ -helices. Superimposition of the A domains of both structures revealed that they are conserved. In this region, there is (M209-A210) and (G192-E193) composed of  $\beta$ -strand whereas corresponding one in MtbpurD is part of a loop. Domain C of MtbpurD appears conserved with an (R364-D365), (I369-S371) and (G381-T382) are parts of  $\beta$ -strands where the corresponding one of EcpurD are part of a loop as shown in Figure (112(2)). As the

crystal structure of EcpurD was solved in the apo form, the putative ATP binding site of MtbpurD model is compared to the HspurD-ATP complex.

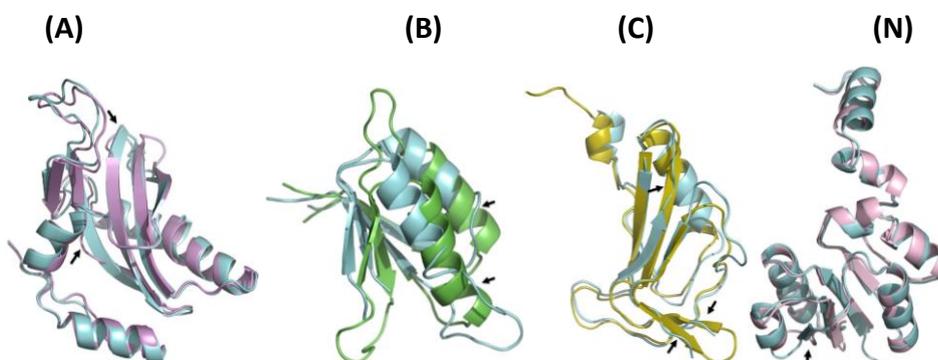
The structure of the model was also compared with the crystal structures solved in the presence of ADP and ATP. The binding site in the homology model was assigned based on superimposition with the Glycinamide ribonucleotide synthetase of *Bacillus subtilis* (BspurD) (PDB 2XD4) in complex with ADP. The residues involved in ADP binding are conserved with only one difference being the charged residue E186 in the BspurD is replaced by the polar amino acid S185 and both main chains make H-bonds to N6 of ADP. The ATP binding site was also examined.

The MtbpurD model was superimposed onto Human purD (HspurD). It can be seen that almost all of the residues are conserved with the only differences observed being in two positions where the negatively charged E191 of HspurD replace the polar residues of MtbpurD model S185 respectively as shown in Figure (112(3)). In both organisms, both residues are the main chain and make H-bonds with N6 of ATP as shown in Figure (112(3)), mutating these residues may not effect ligand-protein stability. ATP makes three ionic interactions so it bound more tightly than in the MtbpurD model but still an experimental structure is required for verification.

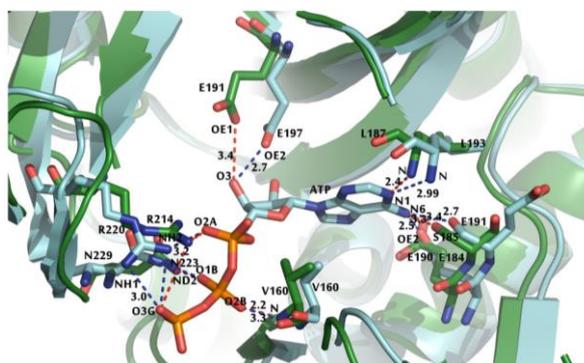
(1)



(2)



(3)



**Figure (112)** Structure analysis of MtbpurD model (1). Cartoon representation of modelled MtbpurD structure; the domains are coloured light pink for N, green B, violet A and olive for C respectively. The structure is labeled based on its secondary structure elements identified by PDBsum (Laskowski et al., 1997). (2). Superimposition of MtbpurD model and over EcpurD (PDB 1GSO) apo form coloured light pink for N, green B, violet A and olive C domains respectively and cyan for EcpurD. The domain is classified as N, B, A, C domains. Differences in secondary structure features are indicated by black arrows. (3) Putative ATP binding site, superimposition of MtbpurD and HspurD (PDB 2QK4), different residues in MtbpurD and HspurD are coloured green and blue respectively. The distances of residues involved in the ATP binding site of MtbpurD and HspurD coloured red and blue. The ATP binding site is located between the A and B domains.

## 6.7. Summary

Prior to characterisation studies, the gene encoding MtbpurD was cloned into an expression vector, overexpressed and purified. The oligomeric state of the protein was determined by gel filtration chromatography seen to be a monomer. The secondary structure of the enzyme was measured in apo and complexed forms by CD, suggesting that the enzyme is  $\alpha$ -helix dominant, which agrees with, theoretical data and data derived from homologous structures (Bertrand et al., 2011). Furthermore, the thermal stability of the enzyme was investigated by using a thermal shift assay, revealing the enzyme is more stable when it makes complexes with ATP+Glycine and in both Tris-HCl buffer and HEPES. Crystallisation trials were carried out to solve the structure of the enzyme to understand protein-ligand interactions. However, the trials did not yield any protein crystals in both whatever purification or crystallisation condition was used. Therefore, the structure of the enzyme was created through homology modeling and putative binding sites of the ligand in MtbpurD were compared to those of other organisms whose structures are deposited in PDB (Sampei et al., 2010).

## Chapter 7. Conclusion

Characterisation of an enzyme in solution in parallel with structure determination facilitates the understanding of the solution properties at the molecular level. Verification of a prepared enzyme's identity, the assessment of purity, and confirmation of the properly folded state should be determined before structural studies are undertaken. Ideally, the enzyme should also have been shown to be in an active state. The establishment of the conditions of optimal stability and catalytic activity are also important information to help with the crystallisation process. MtbSAICAR and MtbpurD have been successfully characterised, the purity, folding, and stability was assessed and the identities of both enzymes confirmed. The kinetic properties of MtbSAICAR were also measured. The enzyme is found to be catalytically active and follows hyperbolic Michaelis-Menten kinetics. The kinetic parameters are found to be similar to those published for the enzyme from another organism (ScSAICAR).

The crystal structures of MtbSAICARs were solved in apo and liganded forms, including ATP, ADP-AMZ, and ACP-IMP complexes. The formation of complexes was achieved using co-crystallisation and a soaking strategy in order to ensure no loss of ligands on cryoprotection. The crystal structures are isomorphous with each other, in the ADP-ACP and AICAR-IMP complexes, these ligands compete for binding at the same sites as ATP and CAIR respectively. The only difference in binding modes, when compared to homologous enzymes, is that the adenine ring was seen to orient differently in IMP in comparison to the natural substrate CAIR in the *E. coli* enzyme. Although all the solved MtbSAICARs structures are quite similar, the average B-value of the main chain atoms of each complex shows that the liganded structures with ATP and ADP-AMZ are more rigid than ACP-IMP complex. These average main chain B values differences are reflected in the overall B-factors values shown in Tables (9, 11, 13 and 15). The increased B-factors may reflect differences in crystalline order resulting in poor resolution of diffraction data, this is probably due to intrinsically less well-ordered enzyme molecules, but other effects on crystal packing cannot be excluded.

The similarities and differences segments between MtbSAICAR (exemplified by MtbSAICAR-ADP-AMZ complex) and homologous enzymes of different organisms have been determined based on global and local comparisons using Q-score value as comparison parameter. From comparisons over the entire length, MaSAICAR showed the highest Q-score as shown in Table (22), it also has the highest sequence identity and similarity (76% and 85% respectively) of the compared enzyme structures. Local alignment using sequence fragments (e.g. domains) is the best way for precise and sensitive comparison of local structures, especially the binding sites of ligands. The aligned domain A (ADP binding site) of available homologous enzyme structures from different organisms revealed that the Q-score of local alignment increases over that of the whole chain. Based on the Q-score values, the most similar to domain A of MtbSAICAR-ADP-AMZ complex are the structures of MaSAICAR (PDB 3R9R) (0.97), ScSAICAR-ADP-AMZ complex (PDB 2CNQ) (0.60) then SpSAICAR-ADP-AIR-Asp (PDB 4FE2) (0.57). The least similar is HsSAICAR with a Q-score 0.35, although the resolution of the HsSAICAR structure is lower and this may affect the result. Based on the local comparison graphs, the structures of all the enzymes were compared and it was found that MtbSAICAR has a loop (P73-V83), this becomes in MaSAICAR (P73-V83) (loop- $\alpha$ -helix), and is deleted in TmSAICAR, EcSAICAR, SpSAICAR and HsSAICAR whereas ScSAICAR has some insertions (A80-K83, P90-K92, S94-Y98) in this area.

For domain B, local alignment confirmed that domain B of MaSAICAR is most similar to MtbSAICAR with Q-score (0.90). Whereas for the other enzymes the values are close to each other with moderate similarities such as for ScSAICAR (0.58), SpSAICAR (0.57), EcSAICAR (0.47), HsSAICAR (0.44), and the least TmSAICAR (0.41). Overall the Q-score calculated using local alignment is better than Q-score of Global values suggesting there are relative domain movements between enzymes, and the local comparisons may be more informative. Comparison of MtbSAICAR with the others revealed that MtbSAICAR has a loop insertion (G118-P127) which is found in MaSAICAR (G118-G128) ( $\alpha$ -helix- $\beta$ -strand-loop-B-strand-loop), in ScSAICAR (G137-G147) is ( $\alpha$ -helix- $\beta$ -strand-loop-B-strand-loop). Whereas it is missing in TmSAICAR, EcSAICAR complex, SpSAICAR, and HsSAICAR.

In MtbSAICAR, a loop insertion segment (A146-G149) is missing in MaSAICAR, EcSAICAR, SpSAICAR, and in HsSAICAR. In ScSAICAR, there is a gap around this area. In TmSAICAR, this has a greater distance difference ( $>3\text{\AA}$ ). (G249-P266) is another insertion in MtbSAICAR that found to be in common with MaSAICAR (G249-P266) (loop- $\alpha$ -helix) but there is a gap in MaSAICAR near the (D259-R260) in MtbSAICAR. There is a deletion and insertion loop in ScSAICAR (N273-V278). This segment is missing in TmSAICAR, EcSAICAR, SpSAICAR, and HsSAICAR.

The segment D228-P266 (loop- $\alpha$ -helix-loop) in MtbSAICAR is compared to corresponding one (H228-P266) ( $\alpha$ -helix-loop- $\alpha$ -helix-loop- $\alpha$ -helix) in MaSAICAR, which found to be similar but there is a gap in MaSAICAR around (D259-260) in MtbSAICAR. In ScSAICAR-ADP-AMZ complex, the corresponding segment is (S248-Q282) (loop- $\alpha$ -helix-loop- $\alpha$ -helix) is similar but there is an insertion (K260-Q261, G274-V278) in ScSAICAR and insertion (G249-W254, G257-R260, P262) in MtbSAICAR. Comparing MtbSAICAR to TmSAICAR (G198-G211) (loop- $\alpha$ -helix) found that there is a gap in TmSAICAR around (D228-K240) in MtbSAICAR, also MtbSAICAR has an insertion (L247-H268) between (G211-D212) in TmSAICAR. In SpSAICAR (D204-E219) (loop- $\alpha$ -helix-loop- $\alpha$ -helix), MtbSAICAR has two insertions (R230-Q235; loop) and (L247-A272;  $\alpha$ -helix-loop- $\alpha$ -helix). In HsSAICAR (G220-E250) (loop- $\alpha$ -helix), there is a gap in HsSAICAR in the area (D228-V234) of MtbSAICAR. There is loop insertion (G249-P266) in MtbSAICAR between (F249-E250) in HsSAICAR. Comparing MtbSAICAR to EcSAICAR (T205-G220) (loop- $\alpha$ -helix-loop- $\alpha$ -helix), there is a loop insertion (R230-Q235) in MtbSAICAR located between (L206-E207) in EcSAICAR. In the EcSAICAR-ADP-CAIR complex R215 interacts with the ribose of 4-carboxy-5-aminoimidazole ribonucleotide (CAIR), the equivalent R244 side chain is beyond H-bond range with the inosine base of IMP in the MtbSAICAR complex. It was suggested that the differences in this region might be exploitable for targeting the enzyme with a selective inhibitor, even though most of the residues are conserved (Ginder et al., 2006).

Details extrapolated from the structure provide information in understanding the mechanism of the enzyme by determining the possible order of the substrates to bind. Ginder et al., propose the reaction of EcSAICAR as being either that the CAIR is first phosphorylated by ATP and then the CAIR phosphoester is attacked by the amino

group of aspartate, forming a tetrahedral intermediate that collapses to release phosphate and leave the amide of the product SAICAR, or that firstly the amino group of aspartate is attached by the carboxyl group of CAIR, forming a tetrahedral intermediate, then ATP phosphorylates this, releasing ADP, then this phosphorylated intermediate collapses to release phosphate and form SAICAR. As no electron density appeared for aspartate in the crystal structure, it was suggested that the phosphorylation first mechanism is the most likely (Ginder et al., 2006). When reporting the structures of SpSAICAR and complexes Wolf et al., proposed the involvement of phospho-enzyme steps, whereby E178 is first phosphorylated by ATP, and the phosphate is then transferred to E89 and then on to CAIR, when it is displaced by aspartate with the subsequent step being similar (Wolf et al., 2014). As seen in EcSAICAR, MtbSAICAR did not show electron density for aspartate even though it was introduced into the crystallisation conditions. In the dimeric SpSAICAR, density for aspartate is only seen in one subunit. In both MtbSAICAR crystal structures, ATP was seen with the disordered  $\gamma$ -phosphate group, perhaps indicating that ATP could be partially hydrolysed. This may suggest that enzyme phosphorylation is the initial step. Crystal structures that include all three substrates (ATP+aspartate), (CAIR+aspartate) and (ATP+CAIR) respectively are required to structurally verify whether there is conformational change upon energy transfer from ATP and to see whether any residues contribute in phosphate transfer (from ATP to CAIR). The structures of MtbSAICAR presented here are consistent with the mechanism in which the first step is the phosphorylation of the substrate (CAIR) by ATP, rather than the formation of a phosphoenzyme intermediate. The involvement of E89 and E178 in phosphate transfer could be tested through site-directed mutagenesis of these residues to glutamine which is isostructural but would not be phosphorylated.

MtbSAICAR is an essential enzyme for *Mycobacterium tuberculosis* and optimistically remains a candidate drug target. Effective inhibitory drugs will either compete for the active site or cause allosteric changes. Selectivity for the pathogen's enzyme over the host is also required. The obvious candidate region to exploit is the one different residue bound to N6 of ATP binding site where the imidazole ring of H69 of MtbSAICAR-ADP-AMZ complex is replaced by the side chain of A76 of HsSAICAR. The

histidine is seen to be important for ATP binding and therefore effective catalysis. The unique pKa of the imidazole might also be taken advantage of in this context. Mutational studies coupled with kinetics could determine the effects of changes in this residue, and also to see if it has a role in the enzyme's stability. Having determined the structure of MtbSAICAR, it becomes possible to use this information to design small molecules that can fit into the active site pocket, target non-conserved structure to create an inhibitor that affects MtbSAICAR but not HsSAICAR, and to bind tighter than the natural substrate. Although MaSAICAR has high sequence identity to MtbSAICAR the MaSAICAR structure has been solved only in apo form, this encouraged us to solve the pathogenic species enzyme in complex with ligands and compare with the non-pathogenic enzyme to see if there is any difference. Based on the comparison, there are some differences between the enzymes suggesting that MaSAICAR is not a perfect model for the MtbSAICAR pathogenic enzyme. The differences in structure would be very important if a PROTAC-type (Neklesa, Winkler and Crews, 2017) approach using Pup/Mpa (Pearce et al., 2008) was invoked as a means of neutralising this enzyme.

## References

- Abbott, J., Livingston, N., Egri, S., Guth, E. and Francklyn, C. (2017). Characterization of aminoacyl-tRNA synthetase stability and substrate interaction by differential scanning fluorimetry. *Methods*, 113, pp.64-71.
- Abola, E., Kuhn, P., Earnest, T. and Stevens, R. (2000). *Nature Structural Biology*, 7, pp.973-977.
- Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Echols, N., Headd, J., Hung, L., Jain, S., Kapral, G., Grosse Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T. and Zwart, P. (2011). The Phenix software for automated determination of macromolecular structures. *Methods*, 55(1), pp.94-106.
- Adam, T. (2005). Purine de novo synthesis-mechanisms and clinical implications. *Klin. Biochem. Metab*, 13(34), pp.177-181.
- Altschul, S. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3), pp.403-410.
- Anand, R., Hoskins, A., Stubbe, J. and Ealick, S. (2004). Domain Organization of *Salmonella typhimurium* Formylglycinamide Ribonucleotide Amidotransferase Revealed by X-ray Crystallography. *Biochemistry*, 43(32), pp.10328-10342.
- Andrews, J., Shah, N., Gandhi, N., Moll, T. and Friedland, G. (2007). Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis: Implications for the HIV Epidemic and Antiretroviral Therapy Rollout in South Africa. *The Journal of Infectious Diseases*, 196(s3), pp.S482-S490.
- Arend, S., Andersen, P., van Meijgaarden, K., Skjøt, R., Subronto, Y., van Dissel, J. and Ottenhoff, T. (2000). Detection of Active Tuberculosis Infection by T Cell Responses to Early-Secreted Antigenic Target 6-kDa Protein and Culture Filtrate Protein 10. *The Journal of Infectious Diseases*, 181(5), pp.1850-1854.
- Augenreich, J., Arbues, A., Simeone, R., Haanappel, E., Wegener, A., Sayes, F., Le Chevalier, F., Chalut, C., Malaga, W., Guilhot, C., Brosch, R. and Astarie-Dequeker, C. (2017). ESX-1 and phthiocerol dimycocerosates of *Mycobacterium tuberculosis* act in

concert to cause phagosomal rupture and host cell apoptosis. *Cellular Microbiology*, 19(7), p.e12726.

Bach, H., Wong, D. and Av-Gay, Y. (2009). Mycobacterium tuberculosis PtkA is a novel protein tyrosine kinase whose substrate is PtpA. *Biochemical Journal*, 420(2), pp.155-162.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K., Wilson, T., Collins, D., de Lisle, G. and Jacobs, W. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science*, 263(5144), pp.227-230.

Banerjee, S., Agrawal, M., Mishra, D., Sharan, S., Balaram, H., Savithri, H. and Murthy, M. (2014). Structural and kinetic studies on adenylosuccinate lyase from Mycobacterium smegmatis and Mycobacterium tuberculosis provide new insights on the catalytic residues of the enzyme. *FEBS Journal*, 281(6), pp.1642-1658.

Bardou, F., Raynaud, C., Ramos, C., Laneelle, M. and Lanrelle, G. (1998). Mechanism of isoniazid uptake in Mycobacterium tuberculosis. *Microbiology*, 144(9), pp.2539-2544.

Baugh, L., Phan, I., Begley, D., Clifton, M., Armour, B., Dranow, D., Taylor, B., Muruthi, M., Abendroth, J., Fairman, J., Fox, D., Dieterich, S., Staker, B., Gardberg, A., Choi, R., Hewitt, S., Napuli, A., Myers, J., Barrett, L., Zhang, Y., Ferrell, M., Mundt, E., Thompkins, K., Tran, N., Lyons-Abbott, S., Abramov, A., Sekar, A., Serbzhinskiy, D., Lorimer, D., Buchko, G., Stacy, R., Stewart, L., Edwards, T., Van Voorhis, W. and Myler, P. (2015). Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis*, 95(2), pp.142-148.

Berg, J., Tymoczko, J., Gatto, G. and Stryer, L. (2015). *Biochemistry*. 1st ed. New York, NY: Freeman.

Bertrand, J., Chen, S., Zalkin, H. and Smith, J. (2011). Nucleotide-bound Structures of Bacillus subtilis Glycinamide Ribonucleotide Synthetase.

Bolton, E. (1954). BIOSYNTHESIS OF NUCLEIC ACID IN ESCHERICHIA COLI. *Proceedings of the National Academy of Sciences*, 40(8), pp.764-772.

Bragg, W. and Bragg, W. (1913). The Reflection of X-rays by Crystals. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 88(605), pp.428-438.

Brodin, P., Rosenkrands, I., Andersen, P., Cole, S. and Brosch, R. (2004). ESAT-6 proteins: protective antigens and virulence factors?. Trends in Microbiology, 12(11), pp.500-508.

Buchan, D., Minneci, F., Nugent, T., Bryson, K. and Jones, D. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research, 41(W1), pp.W349-W357.

Camacho, L., Constant, P., Raynaud, C., Lan elle, M., Triccas, J., Gicquel, B., Daff , M. and Guilhot, C. (2001). Analysis of the Phthiocerol Dimycocerosate Locus of *Mycobacterium tuberculosis*. Journal of Biological Chemistry, 276(23), pp.19845-19854.

Carlucci, F., Rosi, F., Di Pietro, C., Marinello, E., Pizzichini, M. and Tabucchi, A. (1997). Purine nucleotide metabolism: specific aspects in chronic lymphocytic leukemia lymphocytes. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1360(3), pp.203-210.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C., Keseler, I., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. and Karp, P. (2015). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research, 44(D1), pp.D471-D480.

Chao, M. and Rubin, E. (2010). Letting Sleeping dogs Lie: Does Dormancy Play a Role in Tuberculosis?. Annual Review of Microbiology, 64(1), pp.293-311.

Chayen, N. and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. Nature Methods, 5(2), pp.147-153.

Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S., Eiglmeier, K., Gas, S., Barry, C., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N.,

Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M., Rajandream, M., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J., Taylor, K., Whitehead, S. and Barrell, B. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393(6685), pp.537-544.

Cowley, S., Ko, M., Pick, N., Chow, R., Downing, K., Gordhan, B., Betts, J., Mizrahi, V., Smith, D., Stokes, R. and Av-Gay, Y. (2004). The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Molecular Microbiology*, 52(6), pp.1691-1702.

Deacon, R., Chanarin, I., Lumb, M. and Perry, J. (1985). Role of folate dependent transformylases in synthesis of purine in bone marrow of man and in bone marrow and liver of rats. *Journal of Clinical Pathology*, 38(12), pp.1349-1352.

Deb, C., Lee, C., Dubey, V., Daniel, J., Abomoelak, B., Sirakova, T., Pawar, S., Rogers, L. and Kolattukudy, P. (2009). A Novel In Vitro Multiple-Stress Dormancy Model for *Mycobacterium tuberculosis* Generates a Lipid-Loaded, Drug-Tolerant, Dormant Pathogen. *PLoS ONE*, 4(6), p.e6077.

DeLano, W.L., 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, pp.82-92.

Dezemon, Z., Muvunyi, C. and Jacob, O. (2014). Staining techniques for detection of acid fast bacilli: what hope does fluorescein-diacetate (FDA) vitality staining technique represent for the monitoring of tuberculosis treatment in resource limited settings. *International Research Journal of Bacteriology*, 1(1), p.1.

Derrick, S. and Morris, S. (2007). The ESAT6 protein of *Mycobacterium tuberculosis* induces apoptosis of macrophages by activating caspase expression. *Cellular Microbiology*, 9(6), pp.1547-1555.

Dierberg, K. and Chaisson, R. (2013). Human Immunodeficiency Virus–Associated Tuberculosis. *Clinics in Chest Medicine*, 34(2), pp.217-228.

Ducati, R., Breda, A., A. Basso, L. and S. Santos, D. (2011). Purine Salvage Pathway in *Mycobacterium tuberculosis*. *Current Medicinal Chemistry*, 18(9), pp.1258-1275.

- Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp.2126-2132.
- Essa, S., Abdel-Samea, S., Ismaeil, Y. and Mohammad, A. (2013). Comparative study between using Lowenstein Jensen and Bio-FM media in identification of *Mycobacterium tuberculosis*. *Egyptian Journal of Chest Diseases and Tuberculosis*, 62(2), pp.249-255.
- Evans, P. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp.282-292.
- Evans, P. and Murshudov, G. (2013). How good are my data and what is the resolution?. *Acta Crystallographica Section D Biological Crystallography*, 69(7), pp.1204-1214.
- Fawaz, M., Topper, M. and Firestine, S. (2011). The ATP-grasp enzymes. *Bioorganic Chemistry*, 39(5-6), pp.185-191.
- Fenton, M. and Vermeulen, M. (1996). Immunopathology of Tuberculosis: Roles of Macrophages and Monocytes. *Infection and immunity*, 64(3), pp.683-90.
- Flynn, J. (1993). An essential role for interferon gamma in resistance to *Mycobacterium tuberculosis* infection. *Journal of Experimental Medicine*, 178(6), pp.2249-2254.
- Flynn, J. and Chan, J. (2001). Tuberculosis: Latency and Reactivation. *Infection and Immunity*, 69(7), pp.4195-4201.
- Flynn, J. and Ernst, J. (2000). Immune responses in tuberculosis. *Current Opinion in Immunology*, 12(4), pp.432-436.
- Forrellad, M., Klepp, L., Gioffré, A., Sabio y García, J., Morbidoni, H., Santangelo, M., Cataldi, A. and Bigi, F. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*, 4(1), pp.3-66.

Galperin, M. and Koonin, E. (2008). A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Science*, 6(12), pp.2639-2643.

Gebhardt, H., Meniche, X., Tropis, M., Kramer, R., Daffe, M. and Morbach, S. (2007). The key role of the mycolic acid content in the functionality of the cell wall permeability barrier in *Corynebacterineae*. *Microbiology*, 153(5), pp.1424-1434.

Gengenbacher, M. and Kaufmann, S. (2012). *Mycobacterium tuberculosis*: success through dormancy. *FEMS Microbiology Reviews*, 36(3), pp.514-532.

Ginder, N., Binkowski, D., Fromm, H. and Honzatko, R. (2006). Nucleotide Complexes of *Escherichia coli* Phosphoribosylaminoimidazole Succinocarboxamide Synthetase. *Journal of Biological Chemistry*, 281(30), pp.20680-20688.

Grandoni, J., Switzer, R., Makaroff, C. and Zalkin, H. (1989). Evidence that the iron-sulfur cluster of *Bacillus subtilis* glutamine phosphoribosylpyrophosphate amidotransferase determines stability of the enzyme to degradation in vivo. *The American Society for Biochemistry and Molecular Biology*, 264(11), pp.6058-6064.

Green, A., DiFazio, R. and Flynn, J. (2012). IFN- $\gamma$  from CD4 T Cells Is Essential for Host Survival and Enhances CD8 T Cell Function during *Mycobacterium tuberculosis* Infection. *The Journal of Immunology*, 190(1), pp.270-277.

Hartman, S. and Buchanan, J. (1959). Nucleic Acids, Purines, Pyrimidines (Nucleotide Synthesis). *Annual Review of Biochemistry*, 28(1), pp.365-410.

Hasegawa, K. (2012). Introduction to single crystal X-ray analysis. *The Rigaku Journal*, 28(1).

Herrera, V., Perry, S., Parsonnet, J. and Banaei, N. (2011). Clinical Application and Limitations of Interferon-Release Assays for the Diagnosis of Latent Tuberculosis Infection. *Clinical Infectious Diseases*, 52(8), pp.1031-1037.

Hett, E. and Rubin, E. (2008). Bacterial Growth and Cell Division: a *Mycobacterial* Perspective. *Microbiology and Molecular Biology Reviews*, 72(1), pp.126-156.

Horsburgh, C. (2004). Priorities for the Treatment of Latent Tuberculosis Infection in the United States. *New England Journal of Medicine*, 350(20), pp.2060-2067.

Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeón, L., Abdallah, A., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K., van der Laan, T., Kant, A., Bossers-de Vries, R., Willemsen, P., Bitter, W., van Soolingen, D., Brosch, R., van der Wel, N. and Peters, P. (2012). ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cellular Microbiology*, 14(8), pp.1287-1298.

Indrigo, J., Actor, J. and Hunter, R. (2002). Influence of trehalose 6,6'-dimycolate (TDM) during mycobacterial infection of bone marrow macrophages. *Microbiology*, 148(7), pp.1991-1998.

Jackson, M., Phalen, S., Lagranderie, M., Ensergueix, D., Chavarot, P., Marchal, G., McMurray, D., Gicquel, B. and Guilhot, C. (1999). Persistence and Protective Efficacy of a Mycobacterium tuberculosis Auxotroph Vaccine. *Infect Immun*, 67(6), pp.2867–2873.

Jaiswal, S., Chatterjee, A., Pandey, S., Lata, K., Gadi, R., Manda, R., Kumar, S., Reddy, M., Ramachandran, R. and Srivastava, K. (2018). Mycobacterial protein tyrosine kinase, PtkA phosphorylates PtpA at tyrosine residues and the mechanism is stalled by the novel series of inhibitors. *Journal of Drug Targeting*, pp.1-9.

Jankute, M., Cox, J., Harrison, J. and Besra, G. (2015). Assembly of the Mycobacterial Cell Wall. *Annual Review of Microbiology*, 69(1), pp.405-423.

Kak, G., Raza, M. and Tiwari, B. (2018). Interferon-gamma (IFN- $\gamma$ ): Exploring its implications in infectious diseases. *Biomolecular Concepts*, 9(1), pp.64-79.

Kandola, P. and Meena, L. (2014). Extra pulmonary tuberculosis: Overview, manifestations, diagnostic and treatment techniques. *Advanced Materials Reviews*, 1(1), pp.13-19.

Kaye, K. and Frieden, T. (1996). Tuberculosis Control: the Relevance of Classic Principles in an Era of Acquired Immunodeficiency Syndrome and Multidrug Resistance. *Epidemiologic Reviews*, 18(1), pp.52-63.

- Kell, D. and Young, M. (2000). Bacterial dormancy and culturability: the role of autocrine growth factors Commentary. *Current Opinion in Microbiology*, 3(3), pp.238-243.
- Keller, K., Doctor, Z., Dwyer, Z. and Lee, Y. (2014). SAICAR Induces Protein Kinase Activity of PKM2 that Is Necessary for Sustained Proliferative Signaling of Cancer Cells. *Molecular Cell*, 53(5), pp.700-709.
- Kelley, L., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), pp.845-858.
- Kelly, S., Jess, T. and Price, N. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1751(2), pp.119-139.
- Khan, M., Bhaskar, A., Upadhyay, S., Kumari, P., Rajmani, R., Jain, P., Singh, A., Kumar, D., Bhavesh, N. and Nandicoori, V. (2017). Protein kinase G confers survival advantage to *Mycobacterium tuberculosis* during latency-like conditions. *Journal of Biological Chemistry*, 292(39), pp.16093-16108.
- Kleywegt, G. (2000). Validation of protein crystal structures. *Acta Crystallographica Section D Biological Crystallography*, 56(3), pp.249-265.
- Knebel, A. (2001). A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *The EMBO Journal*, 20(16), pp.4360-4369.
- Kozlov, G., Nguyen, L., Pearsall, J. and Gehring, K. (2009). The structure of phosphate-bound *Escherichia coli* adenylosuccinate lyase identifies His171 as a catalytic acid. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 65(9), pp.857-861.
- Krissinel, E. (2012). Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem*, 1(2), pp.76-85.

- Krissinel, E. and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp.2256-2268.
- Kubica, T., Agzamova, R., Wright, A., Rakishev, G., Rüscher-Gerdes, S. and Niemann, S. (2006). *Mycobacterium bovis* Isolates with *M. tuberculosis* Specific Characteristics. *Emerging Infectious Diseases*, 12(5), pp.763-765.
- Kufareva, I. and Abagyan, R. (2011). Methods of Protein Structure Comparison. *Methods in Molecular Biology*, pp.231-257.
- Lane, A. and Fan, T. (2015). Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research*, 43(4), pp.2466-2485.
- Laskowski, R., Hutchinson, E., Michie, A., Wallace, A., Jones, M. and Thornton, J. (1997). PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences*, 22(12), pp.488-490.
- Laskowski, R., MacArthur, M., Moss, D. and Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), pp.283-291.
- Lee, S., Wu, L., Huang, G., Huang, K., Lee, T. and Weng, J. (2016). Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. *BMC Bioinformatics*, 17(S1).
- Le Nours, J., Bulloch, E., Zhang, Z., Greenwood, D., Middleditch, M., Dickson, J. and Baker, E. (2011). Structural Analyses of a Purine Biosynthetic Enzyme from *Mycobacterium tuberculosis* Reveal a Novel Bound Nucleotide. *Journal of Biological Chemistry*, 286(47), pp.40706-40716.
- Levdikov, V., Barynin, V., Grebenko, A., Melik-Adamyan, W., Lamzin, V. and Wilson, K. (1998). The structure of SAICAR synthase: an enzyme in the de novo pathway of purine nucleotide biosynthesis. *Structure*, 6(3), pp.363-376.
- Li, C., Kappock, T., Stubbe, J., Weaver, T. and Ealick, S. (1999). X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution. *Structure*, 7(9), pp.1155-1166.

Lin, P. and Flynn, J. (2010). Understanding Latent Tuberculosis: A Moving Target. *The Journal of Immunology*, 185(1), pp.15-22.

Li, S., Tong, Y., Xie, X., Wang, Q., Zhou, H., Han, Y., Zhang, Z., Gao, W., Li, S., Zhang, X. and Bi, R. (2007). Octameric Structure of the Human Bifunctional Enzyme PAICS in Purine Biosynthesis. *Journal of Molecular Biology*, 366(5), pp.1603-1614.

Li, W. and Xie, J. (2011). Role of mycobacteria effectors in phagosome maturation blockage and new drug targets discovery. *Journal of Cellular Biochemistry*, 112(10), pp.2688-2693.

Liechti, G. and Goldberg, J. (2011). *Helicobacter pylori* Relies Primarily on the Purine Salvage Pathway for Purine Nucleotide Biosynthesis. *Journal of Bacteriology*, 194(4), pp.839-854.

Lodish, H., Berk, A., Kaiser, C., Krieger, M., Bretscher, A., Ploegh, H., Amon, A. and Martin, K. (2016). *Molecular Cell Biology*. W. H. Freeman.

Lovering, A. (2003). X-ray crystallographic studies of therapeutic enzyme; Nitroreductase and AKR1C3. Doctor of philosophy. University of Birmingham.

Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallographica*, 5(6), pp.802-810.

Mahalakshmi, T. M. and Karpagam, K. (2012). The Ongoing Challenge of Latent Tuberculosis- A Over View. *International Journal of Scientific Research*, 3(6), pp.35-37.

Makaroff, C., Paluh, J. and Zalkin, H. (1986). Mutagenesis of Ligands to the [4Fe-4S] Center of *Bacillus subtilis* Glutamine Phosphoribosylpyrophosphate Amidotransferase. *The Journal of biological chemistry*, 261(24), p.1141611423.

Mathews, I., Kappock, T., Stubbe, J. and Ealick, S. (1999). Crystal structure of *Escherichia coli* PurE, an unusual mutase in the purine biosynthetic pathway. *Structure*, 7(11), pp.1395-1406.

Matthews, B. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, 33(2), pp.491-497.

- McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L. and Read, R. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), pp.658-674.
- McDermott, W. (1959). Inapparent Infection: Relation of Latent and Dormant Infections to Microbial Persistence. The R. E. Dyer Lecture. *Public Health Reports (1896-1970)*, 74(6), p.485.
- McGuffin, L., Bryson, K. and Jones, D. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), pp.404-405.
- Moffatt, B. and Ashihara, H. (2002). Purine and Pyrimidine Nucleotide Synthesis and Metabolism. *The Arabidopsis Book*, 1, p.e0018.
- Morar, M., Anand, R., Hoskins, A., Stubbe, J. and Ealick, S. (2006). Complexed Structures of Formylglycinamide Ribonucleotide Amidotransferase from *Thermotoga maritima* Describe a Novel ATP Binding Protein Superfamily. *Biochemistry*, 45(50), pp.14880-14895.
- Morar, M., Hoskins, A., Stubbe, J. and Ealick, S. (2008). Formylglycinamide Ribonucleotide Amidotransferase from *Thermotoga maritima*: Structural Insights into Complex Formation. *Biochemistry*, 47(30), pp.7816-7830.
- Mueller, P. and Pieters, J. (2017). Identification of mycobacterial GarA as a substrate of protein kinase G from *M. tuberculosis* using a KESTREL-based proteome wide approach. *Journal of Microbiological Methods*, 136, pp.34-39.
- Murshudov, G., Vagin, A. and Dodson, E. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D Biological Crystallography*, 53(3), pp.240-255.
- Muchmore, C., Krahn, J., Smith, J., Kim, J. and Zalkin, H. (1998). Crystal structure of glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. *Protein Science*, 7(1), pp.39-51.
- Nagy, P., McCorkle, G. and Zalkin, H. (1993). purU, a source of formate for purT-dependent phosphoribosyl-N-formylglycinamide synthesis. *Journal of Bacteriology*, 175(21), pp.7066-7073.

- Nath, H. and Ryoo, S. (2013). First- and Second-Line Drugs and Drug Resistance. Tuberculosis - Current Issues in Diagnosis and Management.
- Nayak, S. and Acharjya, B. (2012). Mantoux test and its interpretation. Indian Dermatology Online Journal, 3(1), p.2.
- Neklesa, T., Winkler, J. and Crews, C. (2017). Targeted protein degradation by PROTACs. Pharmacology & Therapeutics, 174, pp.138-144.
- Nelson, D., Cox, M. and Lehninger, A. (2009). Lehninger principles of biochemistry absolute ultimate. New York: W.H. Freeman.
- Nelson, S., Binkowski, D., Honzatko, R. and Fromm, H. (2005). Mechanism of Action of EscherichiacoliPhosphoribosylaminoimidazolesuccinocarboxamideSynthetase. Biochemistry, 44(2), pp.766-774.
- Niesen, F., Berglund, H. and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature Protocols, 2(9), pp.2212-2221.
- Niesteruk, A., Jonker, H., Richter, C., Linhard, V., Sreeramulu, S. and Schwalbe, H. (2018). The domain architecture of the PtkA, the first tyrosine kinase from Mycobacterium tuberculosis differs from the conventional kinase architecture. Journal of Biological Chemistry, pp.jbc.RA117.000120.
- Orduz, M. and Ribón, W. (2015). Molecular Epidemiology of Tuberculosis. Tuberculosis - Expanding Knowledge.
- Pai, M., Riley, L. and Colford, J. (2004). Interferon- $\gamma$  assays in the immunodiagnosis of tuberculosis: a systematic review. The Lancet Infectious Diseases, 4(12), pp.761-776.
- Palomino, J. and Martin, A. (2014). Drug Resistance Mechanisms in Mycobacterium tuberculosis. Antibiotics, 3(3), pp.317-340.
- Parish, T. and Stoker, N. (1999). Mycobacteria: Bugs and Bugbears (Two steps forward and one step back). Molecular Biotechnology, 13(3), pp.191-200.

- Pearce, M., Mintseris, J., Ferreyra, J., Gygi, S. and Darwin, K. (2008). Ubiquitin-Like Protein Involved in the Proteasome Pathway of *Mycobacterium tuberculosis*. *Science*, 322(5904), pp.1104-1107.
- Pieters, J. (2008). *Mycobacterium tuberculosis* and the Macrophage: Maintaining a Balance. *Cell Host & Microbe*, 3(6), pp.399-407.
- Poirier, V. and Av-Gay, Y. (2012). *Mycobacterium tuberculosis* modulators of the macrophage's cellular events. *Microbes and Infection*, 14(13), pp.1211-1219.
- Quemard, A., Lacave, C. and Laneelle, G. (1991). Isoniazid inhibition of mycolic acid synthesis by cell extracts of sensitive and resistant strains of *Mycobacterium aurum*. *Antimicrobial Agents and Chemotherapy*, 35(6), pp.1035-1039.
- Raussens, V., Ruyschaert, J. and Goormaghtigh, E. (2003). Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method. *Analytical Biochemistry*, 319(1), pp.114-121.
- Raymond, J., Mahapatra, S., Crick, D. and Pavelka, M. (2004). Identification of the *namH* Gene, Encoding the Hydroxylase Responsible for the N-Glycosylation of the *Mycobacterial* Peptidoglycan. *Journal of Biological Chemistry*, 280(1), pp.326-333.
- Reece, S. and Kaufmann, S. (2012). Floating between the poles of pathology and protection: can we pin down the granuloma in tuberculosis. *Current Opinion in Microbiology*, 15(1), pp.63-70.
- Rieck, B., Degiacomi, G., Zimmermann, M., Cascioferro, A., Boldrin, F., Lazar-Adler, N., Bottrill, A., le Chevalier, F., Frigui, W., Bellinzoni, M., Lisa, M., Alzari, P., Nguyen, L., Brosch, R., Sauer, U., Manganeli, R. and O'Hare, H. (2017). PknG senses amino acid availability to control metabolism and virulence of *Mycobacterium tuberculosis*. *PLOS Pathogens*, 13(5), p.e1006399.
- Riley, R. (1983). The contagiousness of tuberculosis. *Schweiz Med Wochenschr*, 113(3), pp.75-79.
- Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), pp.W320-W324.

Rollins, K. and Lindley, C. (2005). Pemetrexed: A multitargeted antifolate. *Clinical Therapeutics*, 27(9), pp.1343-1382.

Rose, c., Zerbe, J., Lantz, S. and Baily, W. (1979). Establishing Priority during Investigation of Tuberculosis Contacts. *American Review Of Respiratory Disease*, 119(4), pp.603-608.

Russell, D. (2006). Who puts the tubercle in tuberculosis?. *Nature Reviews Microbiology*, 5(1), pp.39-47.

Russell, D., Cardona, P., Kim, M., Allain, S. and Altare, F. (2009). Foamy macrophages and the progression of the human tuberculosis granuloma. *Nature Immunology*, 10(9), pp.943-948.

Ryan, K. and Ray, C and Sherris, J. (2004). *Sherris Medical Microbiology an introduction to infectious diseases*. New York: McGraw-Hill Education/Medical.

Sakula, A. (1982). Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax*, 37(4), pp.246-251.

Samant, S., Lee, H., Ghassemi, M., Chen, J., Cook, J., Mankin, A. and Neyfakh, A. (2008). Nucleotide Biosynthesis Is Critical for Growth of Bacteria in Human Blood. *PLoS Pathogens*, 4(2), p.e37.

Sampei, G., Baba, S., Kanagawa, M., Yanai, H., Ishii, T., Kawai, H., Fukai, Y., Ebihara, A., Nakagawa, N. and Kawai, G. (2010). Crystal structures of glycinamide ribonucleotide synthetase, PurD, from thermophilic eubacteria. *The Journal of Biochemistry*, 148(4), pp.429-438.

Sarver, A. and Wang, C. (2002). The Adenine Phosphoribosyltransferase from *Giardia lamblia* Has a Unique Reaction Mechanism and Unusual Substrate Binding Properties. *Journal of Biological Chemistry*, 277(42), pp.39973-39980.

Sassetti, C., Boyd, D. and Rubin, E. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1), pp.77-84.

Schleifer, K. and Kandler, O. (1972). Peptidoglycan .Types of Bacterial Cell Walls and their Taxonomic Implications. *American Society for Microbiology*, 36(4), pp.407-477.

Senaldi, C., Yin, S., Shaklee, C., Pipet, P., Mak, T. and Ulich, T. (1996). *Corynebacterium parvum*- and *Mycobacterium bovis* bacillus Calmette-Guerin-induced granuloma formation is inhibited in TNF receptor I (TNF-RI) knockout mice and by treatment with soluble TNF-RI. *The Journal of Immunology*, 157(11), pp.5022-5026.

Sievers, F., Wilm, A., Dineen, D., Gibson, T., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. and Higgins, D. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), pp.539-539.

Sheldrick, G. (2015). Crystal structure refinement with SHELXL. *Acta Crystallographica Section C Structural Chemistry*, 71(1), pp.3-8.

Shleeva, M., Salina, E. and Kaprelyants, A. (2010). Dormant forms of mycobacteria. *Microbiology*, 79(1), pp.1-12.

Smith, J. (1998). Glutamine PRPP amidotransferase: snapshots of an enzyme in action. *Current Opinion in Structural Biology*, 8(6), pp.686-694.

Smith, J. (2011). Nanoparticle Delivery of Anti-Tuberculosis Chemotherapy as a Potential Mediator Against Drug-Resistant Tuberculosis. *Yale Journal of biology and Medicine*, 84(4), pp.361-369.

Smith, J., Zaluzec, E., Wery, J., Niu, L., Switzer, R., Zalkin, H. and Satow, Y. (1994). Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science*, 264(5164), pp.1427-1433.

Spek, A. (2009). Structure validation in chemical crystallography. *Acta Crystallographica Section D Biological Crystallography*, 65(2), pp.148-155.

Sreeramareddy, C., Panduru, K., Verma, S., Joshi, H. and Bates, M. (2008). Comparison of pulmonary and extrapulmonary tuberculosis in Nepal- a hospital-based retrospective study. *BMC Infectious Diseases*, 8(1).

Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, 41(3), pp.641-643.

- Susnea, I., Bernevic, B., Wicke, M., Ma, L., Liu, S., Schellander, K. and Przybylski, M. (2012). Application of MALDI-TOF-Mass Spectrometry to Proteome Analysis Using Stain-Free Gel Electrophoresis. *Topics in Current Chemistry*, pp.37-54.
- Tan, M., Sequeira, P., Lin, W., Phong, W., Cliff, P., Ng, S., Lee, B., Camacho, L., Schnappinger, D., Ehrhart, S., Dick, T., Pethe, K. and Alonso, S. (2010). Nitrate Respiration Protects Hypoxic *Mycobacterium tuberculosis* Against Acid- and Reactive Nitrogen Species Stresses. *PLoS ONE*, 5(10), p.e13356.
- Taylor, G. (2010). Introduction to phasing. *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp.325-338
- Thoden, J., Holden, H. and Firestine, S. (2008). Structural Analysis of the Active Site Geometry of N<sup>5</sup>-Carboxyaminoimidazole Ribonucleotide Synthetase from *Escherichia coli*. *Biochemistry*, 47(50), pp.13346-13353.
- Thoden, J., Firestine, S., Benkovic, S. and Holden, H. (2002). PurT-encoded Glycinamide Ribonucleotide Transformylase. *Journal of Biological Chemistry*, 277(26), pp.23898-23908.
- Torres, R. and Puig, J. (2007). Hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency: Lesch-Nyhan syndrome. *Orphanet Journal of Rare Diseases*, 2(1), p.48.
- Tsai, M., Koo, J., Yip, P., Colman, R., Segall, M. and Howell, P. (2007). Substrate and Product Complexes of *Escherichia coli* Adenylosuccinate Lyase Provide New Insights into the Enzymatic Mechanism. *Journal of Molecular Biology*, 370(3), pp.541-554.
- Turetz, M. and Ma, K. (2016). Diagnosis and management of latent tuberculosis. *Current Opinion in Infectious Diseases*, 29(2), pp.205-211.
- Urusova, D., Antonyuk, S., Grebenko, A., Lamzin, V. and Melik-Adamyan, V. (2003). X-ray diffraction study of the complex of the enzyme SAICAR synthase with substrate analogues. *Crystallography Reports*, 48(5), pp.763-767.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G., Chater, K. and van Sinderen, D. (2007). Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiology and Molecular Biology Reviews*, 71(3), pp.495-548.

Waalder, H. (2002). Tuberculosis and poverty. *The International Journal of Tuberculosis and Lung Disease*, 6(9), pp.745-746.

Wallis, R., Patil, S., Cheon, S., Edmonds, K., Phillips, M., Perkins, M., Joloba, M., Namale, A., Johnson, J., Teixeira, L., Dietze, R., Siddiqi, S., Mugerwa, R., Eisenach, K. and Ellner, J. (1999). Drug Tolerance in *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, 43(11), pp.2600–2606.

Wang, L., Westberg, J., Bolske, G. and Eriksson, S. (2001). Novel deoxynucleoside-phosphorylating enzymes in mycoplasmas: evidence for efficient utilization of deoxynucleosides. *Molecular Microbiology*, 42(4), pp.1065-1073.

Wang, X., Barnes, P., Dobos-Elder, K., Townsend, J., Chung, Y., Shams, H., Weis, S. and Samten, B. (2009). ESAT-6 Inhibits Production of IFN- $\gamma$  by *Mycobacterium tuberculosis*-Responsive Human T Cells. *The Journal of Immunology*, 182(6), pp.3668-3677.

Wang, W., Kappock, T., Stubbe, J. and Ealick, S. (1998). X-ray Crystal Structure of Glycinamide Ribonucleotide Synthetase from *Escherichia coli*. *Biochemistry*, 37(45), pp.15647-15662.

Wayne, L. and Diaz, G. (1967). Autolysis and Secondary Growth of *Mycobacterium tuberculosis* in Submerged Culture. *Journal of bacteriology*, 93(4), pp.1374-1381.

Weber, I., Fritz, C., Ruttkowski, S., Kreft, A. and Bange, F. (2000). Anaerobic nitrate reductase (narGHJI) activity of *Mycobacterium bovis* BCG in vitro and its contribution to virulence in immunodeficient mice. *Molecular Microbiology*, 35(5), pp.1017-1025.

Welin, M., Grossmann, J., Flodin, S., Nyman, T., Stenmark, P., Trésaugues, L., Kotenyova, T., Johansson, I., Nordlund, P. and Lehtiö, L. (2010). Structural studies of tri-functional human GART. *Nucleic Acids Research*, 38(20), pp.7308-7319.

Winter, G. (2009). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 43(1), pp.186-190.

- Wolf, N., Abad-Zapatero, C., Johnson, M. and Fung, L. (2014). Structures of SAICAR synthetase (PurC) from *Streptococcus pneumoniae* with ADP, Mg<sup>2+</sup>, AIR and Asp. *Acta Crystallographica Section D Biological Crystallography*, 70(3), pp.841-850.
- Wong, D., Bach, H., Sun, J., Hmama, Z. and Av-Gay, Y. (2011). Mycobacterium tuberculosis protein tyrosine phosphatase (PtpA) excludes host vacuolar-H<sup>+</sup>-ATPase to inhibit phagosome acidification. *Proceedings of the National Academy of Sciences*, 108(48), pp.19371-19376.
- Wong, D., Li, W., Chao, J., Zhou, P., Narula, G., Tsui, C., Ko, M., Xie, J., Martinez-Frailes, C. and Av-Gay, Y. (2018). Protein tyrosine kinase, PtkA, is required for Mycobacterium tuberculosis growth in macrophages. *Scientific Reports*, 8(1).
- Yang, J., Wang, Y. and Zhang, Y. (2016). ResQ: An Approach to Unified Estimation of B - Factor and Residue-Specific Error in Protein Structure Prediction. *Journal of Molecular Biology*, 428(4), pp.693-701.
- Zhang, Q., Wang, D., Jiang, G., Liu, W., Deng, Q., Li, X., Qian, W., Ouellet, H. and Sun, J. (2016). EsxA membrane-permeabilizing activity plays a key role in mycobacterial cytosolic translocation and virulence: effects of single-residue mutations at glutamine 5. *Scientific Reports*, 6(1).
- Zhang, R., Skarina, T., Evdokimova, E., Edwards, A., Savchenko, A., Laskowski, R., Cuff, M. and Joachimiak, A. (2006). Structure of SAICAR synthase from *Thermotoga maritima* at 2.2 Å reveals an unusual covalent dimer. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 62(4), pp.335-339.
- Zhang, Y. (2004). Persistent and dormant tubercle bacilli and latent tuberculosis. *Frontiers in Bioscience*, 9(1-3), p.1136.
- Zhang, Y. (2005). THE MAGIC BULLETS AND TUBERCULOSIS DRUG TARGETS. *Annual Review of Pharmacology and Toxicology*, 45(1), pp.529-564.
- Zhang, Y., Morar, M. and Ealick, S. (2008). Structural biology of the purine biosynthetic pathway. *Cellular and Molecular Life Sciences*, 65(23), pp.3699-3724.

Zhang, Y. and Yew, W. (2009). Mechanisms of drug resistance in *Mycobacterium tuberculosis*. *The International Journal of Tuberculosis and Lung Disease*, 13(11), pp.1320-1330.

Zhang, Z., Caradoc-Davies, T., Dickson, J., Baker, E. and Squire, C. (2009). Structures of Glycinamide Ribonucleotide Transformylase (PurN) from *Mycobacterium tuberculosis* Reveal a Novel Dimer with Relevance to Drug Discovery. *Journal of Molecular Biology*, 389(4), pp.722-733.

Zhao, H., French, J., Fang, Y. and Benkovic, S. (2013). The purinosome, a multi-protein complex involved in the de novo biosynthesis of purines in humans. *Chemical Communications*, 49(40), p.4444.

Zhou, P., Li, W., Wong, D., Xie, J. and Av-Gay, Y. (2014). Phosphorylation control of protein tyrosine phosphatase A activity in *Mycobacterium tuberculosis*. *FEBS Letters*, 589(3), pp.326-331.

Zhao, R., Gao, F., Babani, S. and Goldman, I. (2000). Sensitivity to 5,10-Dideazatetrahydrofolate Is Fully Conserved in a Murine Leukemia Cell Line Highly Resistant to Methotrexate Due to Impaired Transport Mediated by the Reduced Folate Carrier. *Clinical Cancer Research*, 6, pp.3304–3311.

Zimmermann, L., Stephens, A., Nam, S., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. and Alva, V. (2017). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*.

Zöllner, N. (1982). Purine and pyrimidine metabolism. *Proceedings of the Nutrition Society*, 41(03), pp.329-342.