

**STRUCTURAL INSIGHTS INTO THE ENZYMES OF
THE SERINE AND BIOTIN BIOSYNTHETIC PATHWAYS IN**

Mycobacterium tuberculosis

A Dissertation

by

SANGHAMITRA DEY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Biochemistry

**STRUCTURAL INSIGHTS INTO THE ENZYMES OF
THE SERINE AND BIOTIN BIOSYNTHETIC PATHWAYS IN**

Mycobacterium tuberculosis

A Dissertation

by

SANGHAMITRA DEY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James C. Sacchettini
Committee Members,	Paul F. Fitzpatrick
	Michael Polymenis
	J. Martin Scholtz
Head of Department,	Gregory D. Reinhart

August 2008

Major Subject: Biochemistry

ABSTRACT

Structural Insights into the Enzymes of the Serine and Biotin Biosynthetic Pathways in
Mycobacterium tuberculosis. (August 2008)

Sanghamitra Dey, B.S.; M.S., Nagpur University, India

Chair of Advisory Committee: Dr. James C. Sacchettini

Mycobacterium tuberculosis (*Mtb*) utilizes different metabolic pathways for its survival during infection. Enzymes of these pathways are often targets for antibiotic development. Genetic studies indicate the importance of the serine and biotin biosynthetic pathways for *Mtb* survival. In this study, enzymes from these pathways were characterized using X-ray crystallographic and biochemical studies.

D-3-phosphoglycerate dehydrogenase (PGDH) catalyzes the first step of phosphorylated serine biosynthesis. In comparison to other forms of PGDH, the *Mtb* enzyme has an insertion near its C-terminus. This insertion results in two different conformations of the subunits in the tetramer, leading to two different environments for cofactor binding. This intervening domain might provide a second binding site for hydroxypyruvic acid phosphate (HPAP) that is responsible for substrate inhibition. Analysis of the HPAP-bound *Mtb* PGDH active site reveals the residues (Arg52, Arg131, and Arg233) involved in substrate interaction and provides insights into a possible enzyme mechanism. *Mtb* PGDH is feedback inhibited by the end product L-serine. Examination of the serine-bound PGDH structure elucidates the key players

(Tyr461, Asp463, and Asn481) involved in this allosteric inhibition, as well as the resultant conformational changes at the regulatory domain interface.

Preliminary biochemical studies of the first enzyme in *Mtb* biotin biosynthesis, 7-keto-8-aminopelargonic acid (KAPA) synthase show that it exists as a dimer in solution and has higher substrate affinity than the *E. coli* enzyme.

The second enzyme, 7, 8-diaminopelargonic acid synthase (DAPAS) uses S-adenosyl methionine and KAPA as substrates in a bi-bi ping-pong mechanism. A comparison of the substrate analog sinefungin-bound *Mtb* DAPAS structure with a KAPA-bound DAPAS model provides a basis for the dual-substrate recognition. Tyr25 is a key player in the substrate specificity in DAPAS; this was confirmed by mutation studies. In certain *Bacillus* species, a Phe replaces this Tyr. The KAPA-bound *B. subtilis* DAPAS structure shows an alteration in the KAPA binding mode.

Substrate and product bound structures of the third enzyme, dethiobiotin synthetase (DTBS) in *Mtb* reveal the important residues involved in its catalysis and provide framework for a possible enzyme mechanism. Comparison to the DTBS structures from *E. coli* and *H. pylori* reveals differences in local conformations.

DEDICATION

To my parents, Samita and Siddheswar Dey, who have always been a great support and source of encouragement for me; to my husband, Sovan for his constant support, patience and love.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. James C. Sacchettini, for his guidance and support throughout my graduate career. I convey my thanks to my committee members, Dr. Paul F. Fitzpatrick, Dr. J. Martin Scholtz, and Dr. Michael Polymenis for their advice throughout the course of this research, and encouragement. I would also like to thank Dr. Gregory A. Grant for a wonderful collaboration in the PGDH project.

I am grateful to Dr. Stanley M. Swanson and Dr. Rosemarie Swanson for critically reviewing my dissertation. Thanks also go to all my lab members and friends for their constant support, help and making my time at Texas A&M University a memorable experience. Lastly, I am also thankful to the Biochemistry department staff for all their administrative help.

NOMENCLATURE

A	Alanine
ACT	Aspartokinase - chorismate mutase – TyrA
ADP	Adenosine-5'-diphosphate
Ala	Alanine
AMP-PCP	Adenylylmethylenediphosphonate
AMPSO	N-(1,1-Dimethyl-2-hydroxyethyl)-3-amino-2-hydroxypropane sulfonic acid
Arg	Arginine
ASU	Asymmetric unit
Asp	Aspartate
Asn	Asparagine
ATP	Adenosine-5'-triphosphate
CNS	Crystallography & NMR system
CoA	Coenzyme A
Cu	Copper
Cys	Cysteine
D	Aspartate
DAPA	7, 8-diaminopelargonic acid
DAPAS	7, 8-diaminopelargonic acid synthase
DNA	Deoxyribonucleic acid
DTB	Dethiobiotin

DTBS	Dethiobiotin synthetase
DTT	Dithiothreitol
F	Phenylalanine
G	Glycine
Gln	Glutamine
Gly	Glycine
Glu	Glutamate
His	Histidine
HPAP	Hydroxypyruvic acid phosphate
HTS	High throughput screening
I	Isoleucine
Ile	Isoleucine
IPTG	Isopropyl-1-thio β -D-galactopyranoside
KAPA	7-keto, 8-aminopelargonic acid
kDa	Kilo Dalton
kbps	kilo base pairs
LLG	Log likelihood gain
M	Methionine
MAD	Multiple anomalous dispersion
MES	2-(N-morpholino)ethanesulfonic acid
Met	Methionine
Mg	Magnesium

MIR	Multiple isomorphous dispersion
MR	Molecular replacement
Mtb	Mycobacterium tuberculosis
N	Asparagine
NAD	Nicotinamide adenine dinucleotide
NBD	Nucleotide-binding domain
NIP	Nitrogenase iron protein
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PEP	Phosphoenol pyruvate
PGDH	D-3-phosphoglycerate dehydrogenase
Phe	Phenylalanine
PIPES	Piperazine-1,4-bis(2-ethanesulfonic acid)
P _i	Inorganic phosphate
PLP	Pyridoxal-5'-phosphate
PMP	Pyridoxamine-5'-phosphate
PMSF	Phenyl methyl sulfonyl fluoride
Pro	Proline
RD	Regulatory domain
R.M.S.D.	Root mean square deviation
SAD	Single anomalous dispersion

SAH	S-adenosyl homocysteine
SAM	S-adenosylmethionine
SBD	Substrate-binding domain
SDS	Sodium dodecyl sulphate
SeMet	Selenomethionine
Ser	Serine
Sm	Samarium
TB	Tuberculosis
TCEP	Tris (2-carboxyethyl) phosphine hydrochloride
TEV	Tobacco etch virus
TLS	Translation libration screw
Trp	Tryptophan
Tyr	Tyrosine
V	Valine
Val	Valine
VS	Virtual screening
W	Tryptophan
Y	Tyrosine

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS	xi
LIST OF FIGURES.....	xv
LIST OF TABLES	xix
 CHAPTER	
I INTRODUCTION.....	1
General description of <i>Mycobacterium tuberculosis</i>	2
Serine biosynthesis.....	4
Phosphorylated pathway of serine biosynthesis.....	5
Non-phosphorylated pathway of serine biosynthesis.....	6
Other possible pathways for serine biosynthesis.....	7
Regulation of serine biosynthesis.....	8
Biotin biosynthesis	8
Biotin biosynthesis in <i>Mycobacterium tuberculosis</i>	9
Need for biotin biosynthetic pathway in <i>Mtb</i>	13
Overview of chapters	15
II STRUCTURAL STUDIES ON D-3-PHOSPHOGLYCERATE DEHYDROGENASE IN <i>Mycobacterium tuberculosis</i>	17
Background	17
Different forms of PGDH.....	18
Kinetic properties of <i>Mtb</i> PGDH	19
Serine inhibition of PGDH.....	24
Methods.....	26
Reagents	26

CHAPTER	Page
Cloning, protein overexpression and purification of <i>Mtb</i> PGDH	27
Cloning, protein overexpression and purification of human PGDH	29
Fluorescence studies of <i>Mtb</i> PGDH	31
Crystallization of <i>Mtb</i> and human PGDH	31
Data collection and processing.....	33
Basics of structure determination.....	34
Results and discussion.....	42
Purification and characterization of <i>Mtb</i> PGDH	42
Structure determination and overall quality of <i>Mtb</i> PGDH apostructure	44
Overall structure of <i>Mtb</i> PGDH	46
Subunit structure in <i>Mtb</i> PGDH	49
Quaternary structure- <i>Mtb</i> PGDH as a tetramer.....	52
Analysis of hinge regions in <i>Mtb</i> PGDH	56
Analysis of the active site of <i>Mtb</i> PGDH	57
Cofactor-binding site.....	58
Mode of substrate recognition and enzyme catalysis.....	63
Role of intervening domain as an anion-binding site.....	67
Mode of serine binding	71
Cloning, purification and crystallization of human PGDH...	76
 III	
CRYSTALLOGRAPHIC STUDIES ON 7, 8-DIAMINOPELARGONIC ACID SYNTHASE IN THE BIOTIN BIOSYNTHETIC PATHWAY	81
Background	81
Methods.....	85
Reagents	85
Cloning, protein overexpression and purification	85
Spectroscopic studies of <i>Mtb</i> DAPAS	86
Crystallization and data collection	88
Structure determination	90
Molecular replacement technique	90
Model building and refinement	94
Results and discussion.....	96
Description of the overall fold in <i>Mtb</i> DAPAS.....	96
Dimer interface.....	96
Comparison with <i>E. coli</i> DAPAS.....	97
Comparison with other PLP-dependent enzymes	99
Analysis of <i>Mtb</i> DAPAS active site	101

CHAPTER	Page
Cofactor-binding site.....	101
Biochemical characterization of <i>Mtb</i> DAPAS	104
Possible substrate recognition in <i>Mtb</i> DAPAS	112
Kinetic properties of Y25A mutant form of <i>Mtb</i> DAPAS	118
Variation in the DAPAS active site.....	120
Variation in KAPA binding in <i>Bacillus subtilis</i> DAPAS.....	121
Search for <i>Mtb</i> DAPAS inhibitors	125
Virtual screening of <i>Mtb</i> DAPAS	130
IV STRUCTURAL INSIGHTS OF DETHIOBIOTIN SYNTHETASE IN <i>Mycobacterium tuberculosis</i>	138
Background	138
Methods.....	140
Reagents	140
Cloning, protein expression and purification of <i>Mtb</i> DTBS .	141
Spectroscopic studies of <i>Mtb</i> DTBS	142
Crystallization of <i>Mtb</i> DTBS	143
Data collection and processing.....	143
Structure determination	145
Results and discussion.....	146
Use of samarium for structure determination of <i>Mtb</i> PGDH	146
Overall structure of <i>Mtb</i> DTBS.....	149
Other structural neighbors of <i>Mtb</i> DTBS	152
Kinetic characterization of <i>Mtb</i> DTBS.....	153
Interaction of <i>Mtb</i> DTBS with substrate DAPA, and its analog KAPA	153
ATP/ADP binding site	159
Interaction of <i>Mtb</i> DTBS with the products DTB and ADP .	161
Comparison with other DTBS structures	167
Possible <i>Mtb</i> DTBS reaction mechanism.....	170
V PURIFICATION AND PRELIMINARY CHARACTERIZATION OF 7-KETO, 8-AMINOPELARGONIC ACID SYNTHASE IN <i>Mycobacterium tuberculosis</i>	173
Background	173
Methods.....	175
Reagents	175
Cloning and protein expression of <i>Mtb</i> KAPAS	175
Purification of His-tagged KAPAS	176
Purification of wild type KAPAS.....	177

CHAPTER	Page
Determination of oligomeric state of <i>Mtb</i> KAPAS	178
Spectroscopic studies of <i>Mtb</i> KAPAS	179
Crystallization trials of <i>Mtb</i> KAPAS	179
Results and discussion.....	180
Purification and crystallization of <i>Mtb</i> KAPAS.....	180
Spectroscopic properties of <i>Mtb</i> KAPAS.....	187
Conclusions	189
VI CONCLUSIONS.....	190
REFERENCES.....	200
VITA	214

LIST OF FIGURES

FIGURE	Page
1.1 Phosphorylated pathway of serine biosynthesis.....	6
1.2 Non-phosphorylated pathway of serine biosynthesis.....	7
1.3 Biotin operon in <i>Mtb</i>	9
1.4 Biotin biosynthetic pathway in <i>Mtb</i>	12
2.1 Reversible enzymatic reaction catalyzed by PGDH.....	18
2.2 Three different forms of PGDH based on sequence alignment	20
2.3 Domain organizations of PGDH in <i>Pyrococcus horikoshii</i>	21
2.4 Domain arrangements of the <i>E. coli</i> PGDH.....	22
2.5 Alignment of PGDH sequences from human, rat liver, <i>Mtb</i> and <i>E. coli</i> ...	23
2.6 Hexagonal rod shaped crystals of <i>Mtb</i> PGDH	33
2.7 Principles of MAD phasing.....	38
2.8 Fluorescence spectra of <i>Mtb</i> PGDH.....	43
2.9 Ramachandran plot generated by PROCHECK for <i>Mtb</i> PGDH apoenzyme (PDBcode-1YGY).....	45
2.10 Domain arrangements of <i>Mtb</i> PGDH.....	47
2.11 Ribbon diagram of superimposition of chain A of <i>Mtb</i> PGDH on chain B	48
2.12 Comparison of the subunit and domain interaction in <i>E. coli</i> and <i>Mtb</i> PGDH	54
2.13 NADH interaction in <i>Mtb</i> PGDH chain B	59
2.14 Differences in cofactor binding site	61

FIGURE	Page
2.15 Catalytic event at the <i>Mtb</i> PGDH active site.....	66
2.16 Overlay of intervening domain of <i>Mtb</i> PGDH on domain II of 2-methylcitrate dehydratase from <i>E. coli</i> (PDB code-1SZQ)	68
2.17 Anion binding site at the intervening domain interface of <i>Mtb</i> PGDH	70
2.18 Changes occurring at the loop region of allosteric site on serine binding ..	74
2.19 Molecular surface representation of the serine binding site.....	75
2.20 Purification of human PGDH.....	79
3.1 Reaction catalyzed by 7, 8-diaminopelargonic acid synthase.....	82
3.2 Alignment of sequences of DAPAS in <i>Mtb</i> , <i>E. coli</i> and <i>B. subtilis</i>	84
3.3 Crystals of DAPAS enzyme.....	89
3.4 Principles of Molecular replacement method.....	93
3.5 Ribbon diagram of <i>Mtb</i> DAPAS structure.....	98
3.6 Superimposition of <i>Mtb</i> DAPAS on <i>Mtb</i> LAT	100
3.7 PLP binding at the active site of <i>Mtb</i> DAPAS.....	103
3.8 PLP reaction intermediates in the first half of the aminotransferase reaction.....	106
3.9 Structure of SAM and its analogs sinefungin and S-adenosyl homocysteine (SAH).....	107
3.10 UV-VIS absorbance scans of <i>Mtb</i> DAPAS with the substrate SAM and the product DAPA.....	109
3.11 Fluorescence spectra of <i>Mtb</i> DAPAS with SAM analog, sinefungin	111
3.12 Model for substrate recognition at the active site of <i>Mtb</i> DAPAS.....	116
3.13 Comparison of absorbance scans of native <i>Mtb</i> DAPAS and Y25A mutant	119

FIGURE	Page
3.14 Parts of the amino acid sequence alignment of <i>Mtb</i> and <i>E. coli</i> DAPAS along with other organisms showing changes at position 25	121
3.15 Binding mode of KAPA at <i>B. subtilis</i> DAPAS active site.....	124
3.16 Changes observed on ampicillin at the <i>Mtb</i> DAPAS active site	127
3.17 Overlay of docked compounds at the <i>Mtb</i> DAPAS active site	132
4.1 Reaction catalyzed by <i>Mtb</i> DTBS	138
4.2 Alignment of <i>E. coli</i> , <i>Mtb</i> and <i>H. pylori</i> DTBS sequences	139
4.3 Crystal of <i>Mtb</i> DTBS	145
4.4 Ramachandran plot for the <i>Mtb</i> DTBS apoenzyme.....	147
4.5 Ribbon diagram representation of <i>Mtb</i> DTBS dimer.....	151
4.6 Coupled assay of <i>Mtb</i> DTBS with pyruvate kinase/lactate dehydrogenase	153
4.7 Electron density for DAPA-carbamate	155
4.8 Mode of DAPA-carbamate binding at the <i>Mtb</i> DTBS active site.....	156
4.9 Schematic diagram of KAPA interactions at the <i>Mtb</i> DTBS active site.....	158
4.10 Interactions of product DTB and phosphate ions in the <i>Mtb</i> DTBS active site.....	162
4.11 Surface representation of the active site of <i>Mtb</i> DTBS.....	165
4.12 Mg ²⁺ ion coordination in the <i>Mtb</i> DTBS active site.....	166
4.13 Superimposition of C _α atoms of DTBS monomer from <i>E. coli</i> and <i>Mtb</i> ...	169
4.14 Schematic representation of the possible mechanism of <i>Mtb</i> DTBS.....	172
5.1 Reaction catalyzed by 7-keto, 8-aminopelargonic acid synthase.....	173

FIGURE	Page
5.2 PCR amplification of Rv1569 gene.....	182
5.3 Elution profile of a Ni affinity column chromatography and SDS-PAGE gel of KAPAS purification.....	183
5.4 Ribbon diagram of <i>E. coli</i> KAPAS (PDB code-1BS0).....	184
5.5 Sequence alignment of <i>E. coli</i> and <i>Mtb</i> KAPAS	185
5.6 Characterization of the KAPA synthase reaction intermediates	188

LIST OF TABLES

TABLE		Page
1	Data collection, refinement and geometry statistics of <i>Mtb</i> PGDH.....	41
2	Intra subunit domain contacts	55
3	Inter subunit domain contacts	55
4	Data collection, refinement and geometry statistics of <i>Mtb</i> DAPAS.....	92
5	Data collection, refinement and geometry statistics of <i>B. subtilis</i> DAPAS	95
6	Kinetic parameters for the single-turnover-half reactions of <i>Mtb</i> and <i>E.coli</i> DAPAS.....	110
7	Compounds having chemical structure similar to the substrate KAPA.....	129
8	Chembridge compounds used for virtual screening	133
9	Data collection, refinement and geometry statistics of <i>Mtb</i> DTBS.....	148

CHAPTER I

INTRODUCTION

The different stages of infection in *Mycobacterium tuberculosis* (*Mtb*) are poorly understood. *Mtb* causes the deadly disease tuberculosis in humans so it is very important to study the various stages of its infection as well as its mode of virulence. With the availability of the genome sequence of *Mtb* and several genetic approaches, we have been able to identify the importance of various genes involved in different metabolic pathways. Detailed knowledge of the enzymes of these pathways that are absent or have alternate mechanisms or regulation in humans will be helpful in the elimination of tuberculosis. Genetic studies show that enzymes of the serine and biotin biosynthetic pathways are crucial for mycobacterial survival. In this dissertation, we have characterized, biochemically as well as structurally, some of the enzymes of these pathways. X-ray crystal structures of these enzymes shed light on the overall folds and on the basis of substrate/inhibitor recognition. The kinetic and structural features of these enzymes can be further used in the development of potential inhibitors.

This dissertation follows the style of *Journal of Biological Chemistry*.

General description of *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (*Mtb*) belongs to the class Actinomycetes and is, thus, closely associated with *Corynebacterium* and *Nocardia* species. Identified in 1882 by Robert Koch, *Mtb* is a rod shaped, aerobic, non-spore forming bacterium that has a complex cell wall containing mycolic acid. This intracellular and intravacuolar pathogen is responsible for causing the deadly disease of tuberculosis in humans. Initially, tubercle bacilli are transmitted via nasal droplets, aerosolized in the form of a sneeze or cough by the infectious person. Most of these bacteria are generally killed by the host immune system. However, some of these mycobacteria can still escape the host's defenses and start replicating inside the alveolar macrophages. During the process of infection, these acid-fast bacilli employ several different strategies for their survival. *Mtb* resides in the macrophages, avoiding phagosome - lysosome fusion through surface sulfatides (1) and impeding acidification or maturation of phagosomes by excluding vesicular proton-ATPase (2). *Mtb* also tolerates the presence of reactive oxygen and nitrogen intermediates in macrophages by producing detoxifying enzymes like catalase and superoxide dismutase (3).

Once adapted to the phagosomal environment, the bacteria continue to multiply exponentially in the vesicles, lysing the macrophage due to high bacterial population and subsequently being engulfed by the neighboring activated macrophages. This process of struggle between the host immune system and the mycobacterium ultimately leads to the formation of granulomas with the recruitment of host CD4⁺ T cells. In this way, the mycobacteria remain contained inside these granulomas and are no longer infectious.

This stage of infection is called the latent state of TB. With a strong host immune system, the granulomas can slowly heal, leaving mild abrasions in their place. However, any weakness in the host immune system due to HIV infection, malnutrition, aging or an immunosuppressant can lead to the breakdown of these granulomas. This breakdown results in the release of infectious bacteria which travels through the blood and lymphatic system to the lungs and other tissues. Under these conditions, a person becomes chronically infected and needs antibiotics to kill these bacteria. Sadly, there are as many as 200 million people with this latent stage infection who are prone to reactivation leading to chronic infection (4).

Even though the morphology of the mycobacterium, its physiology in cell cultures, and the immune response of the host against the mycobacterial infection are well characterized, the lifecycle of *Mtb* during its pathogenesis is still not completely understood, especially not during the latent stage. Most of the currently effective drugs are futile against this stage of bacterial infection. However, a major achievement in TB research in this direction was the annotation of the whole genome of *Mtb* (lab strain H37Rv) in 1998 (5). The circular genome of this strain of *Mtb* has a high GC content of about 65% and encodes 4000 genes (5). Still, 14% of these genes, which might be essential, could not be assigned any function.

In recent years, extensive research using different genetic approaches (high throughput mutagenesis with transposon insertion, microarray-based genomic and transcriptome profiling) has enlightened us, to some extent, about the lifecycle of mycobacterium during different stages of infection and about its diverse virulence

mechanisms (6, 7). Mutation studies have led to the identification of genes involved in various metabolic pathways, which are required at different stages of mycobacterial infection in macrophages (8) and in the mice (9). The microarray analyses allow us to analyze the pattern of gene expression in *Mtb* under different physiological conditions analogous to the host environment during pathogenesis. Genes that are upregulated or downregulated due to different environmental conditions are thought to be important for the survival of this mycobacterium. This valuable information, coupled with the knowledge of the human genome, makes it easier for us to pinpoint the metabolic pathways necessary at different stages of mycobacterial infection that will not be important or present in host cells. Such metabolic pathways of interest are lipid metabolism, amino acid metabolism, carbohydrate metabolism, vitamin, and cofactor metabolism. It is further anticipated that the inhibitors designed against enzymes of these pathways will result in quick bacterial clearance. Thus, it is worthwhile to study the biochemical and the structural features of the enzymes from mycobacterial metabolic pathways in order to possibly use them as templates in the search for better antimicrobials.

To this end, the focus of this dissertation is on the biochemical and three-dimensional structural analyses of enzymes involved in two different metabolic pathways: the serine and biotin biosynthetic pathways.

Serine biosynthesis

From the genome sequence analysis, it appears likely that mycobacteria have all the amino acid metabolic pathways common to other bacteria. However, differences may

lie in the mechanism of regulation of the metabolites either by feedback inhibition or by repression of the enzyme synthesis. L-serine is the predominant source of one-carbon units, in the metabolism of amino acids like methionine and histidine, and in the *de novo* synthesis of purines and pyrimidines (10). This reversible reaction of conversion of serine to one-carbon unit glycine is catalyzed by serine hydroxymethyl transferase (SHMT) (11). L-serine also serves as a precursor for the synthesis of other amino acids like D-serine, cysteine and its derivative taurine, as well as phospholipids.

Though serine is considered to be nonessential due to the presence of all the enzymes needed for its synthesis in the cell, sometimes it can not be produced at an adequate rate to satisfy the needs of the proliferating cells, thereby making it a "conditional essential" (12). Humans with a deficiency in serine biosynthesis show a marked decline in serine and glycine concentrations in plasma and cerebrospinal fluid. This deficiency is the cause of several neurological disorders like congenital microcephaly, dysmyelination, intractable seizures and psychomotor retardation (13-15). There are two common pathways by which serine can be synthesized: the phosphorylated and non-phosphorylated pathways.

Phosphorylated pathway of serine biosynthesis

L-serine can be synthesized from the glycolytic intermediate, D-3-phosphoglycerate (3-PGA) through the phosphorylated pathway of serine biosynthesis. This pathway, first reported by Ichihara and Greenberg (16), is most commonly used by organisms for serine production. D-3-phosphoglycerate dehydrogenase (*serA*, PGDH, EC 1.1.1.95) catalyzes the first step of the oxidation of D-3-phosphoglycerate to

hydroxypyruvate phosphate (HPAP), utilizing NAD^+ as a cofactor. Subsequently, HPAP is converted to phosphoserine by phosphoserine transaminase (*serC*, PSAT, EC 2.6.1.52) requiring pyridoxal-5'-phosphate (PLP) as a cofactor. This phosphoserine is further dephosphorylated into L-serine by phosphoserine phosphatase (*serB*, PSP, EC 3.1.3.3) (Figure 1.1). Annotation of the *Mtb* (H37Rv lab strain) genome shows the presence of all three genes (Rv2996c, Rv0884c, Rv0505c) of this pathway on different operons of this genome. However, very little was known about these enzymes in *Mtb* until we started exploring the first enzyme of this pathway. In early 2007, Baker *et al.* determined the crystal structure of the third enzyme of this pathway, phosphoserine aminotransferase (PDB code-2FYF) in *Mtb* but no further biochemical characterization of this enzyme has been reported.

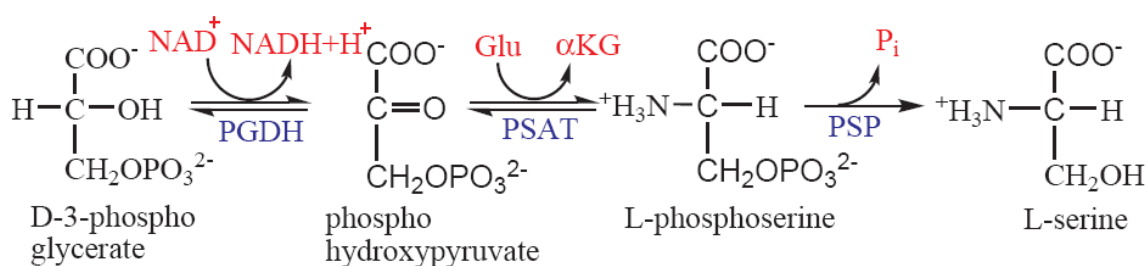


FIGURE 1.1. **Phosphorylated pathway of serine biosynthesis**

Non-phosphorylated pathway of serine biosynthesis

L-serine can also be synthesized in a non-phosphorylated pathway using D-glycerate as the starting material, as studied in detail by H. J. Sallach (17, 18). D-glycerate, acquired from D-2-phosphoglycerate of the glycolytic pathway, is further oxidized to hydroxypyruvate using NAD^+ -dependent D-glycerate dehydrogenase (GDH, EC 1.1.1.29). L-serine is then produced by serine pyruvate aminotransferase (SPT, EC

2.6.1.51) requiring PLP as a cofactor (Figure 1.2). This non-phosphorylated pathway has been found in mammals (19), plants (20, 21) and some microbes, such as the unicellular protozoan *E. histolytica* (22). However, there is no evidence which suggests the existence of these enzymes in *Mtb*. This pathway has been thought to be more involved in serine catabolism than its biosynthesis due to the presence of all the reversible steps, leading to gluconeogenesis (23, 24).

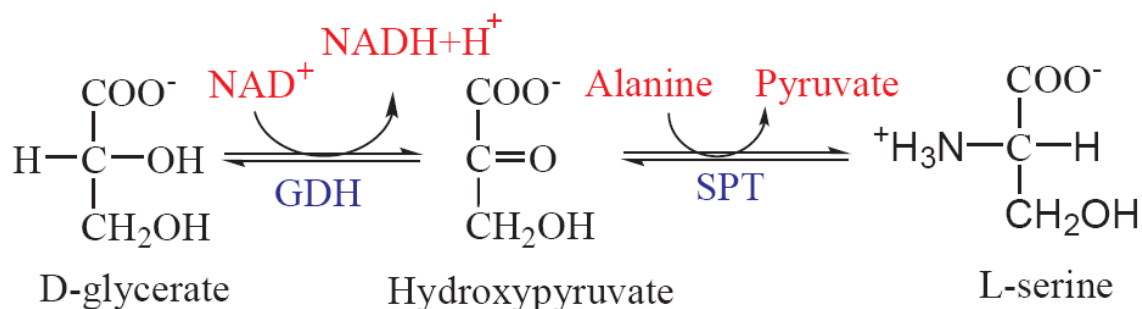


FIGURE 1.2. **Non-phosphorylated pathway of serine biosynthesis**

Other possible pathways for serine biosynthesis

In plants, e.g. *Arabidopsis thaliana*, apart from the above mentioned serine biosynthetic pathways, serine can also be produced from the glycolate pathway (25). This pathway occurs in mitochondria of photosynthetic tissues and only functions in the presence of light. In contrast, the phosphorylated and non-phosphorylated pathways are functional in dark conditions in plastids of non-photosynthetic tissues of the plants. In yeast, serine can also be produced from glyoxylate using TCA cycle intermediates in the absence of glucose (26, 27).

Regulation of serine biosynthesis

In organisms like *Escherchia coli* (*E. coli*), *Bacillus subtilis* (*B. subtilis*), and yeast, the phosphorylated pathway of serine biosynthesis is regulated by feedback inhibition of the first enzyme, PGDH, by the final product, L-serine. Detailed study of the allosteric serine inhibition of PGDH has been carried out in *E. coli*. However, PGDH from plants (20) and mammals (28, 29) does not demonstrate this feedback inhibition by serine. In the case of rat liver tissues, serine biosynthesis is regulated in response to dietary protein and hormonal (insulin, glucocorticoid) intake (30-32). Interestingly, the flux through the serine biosynthetic pathway in rat and rabbit liver is regulated at the level of phosphoserine phosphatase, which is the irreversible and rate limiting step in this case (33). In contrast to the classical feedback inhibition of the serine biosynthetic pathway, it seems that serine biosynthesis in plants and mammals is regulated at the last step by cellular demands for serine. Little is known about the regulation in the non-phosphorylated pathway of serine biosynthesis.

Biotin biosynthesis

The second part of this dissertation focuses on the enzymes of the biotin biosynthetic pathway in *Mtb*. Biotin (vitamin H) is indispensable as a cofactor during carboxylation, transcarboxylation and decarboxylation reactions in certain important pathways like fatty acid biosynthesis, gluconeogenesis, and amino acid metabolism in both prokaryotes and eukaryotes. It functions as a carrier of active CO₂. Like insulin, biotin also plays a role in transcriptional regulation of glucokinase and phosphoenol pyruvate carboxykinase - key enzymes of glucose metabolism in eukaryotes (34, 35).

Utilization of biotin in the food and cosmetics industries is also well known. Since the bioavailability of biotin is very low, it was first chemically synthesized in 1944 by Harris and his colleagues (36). However, this type of industrial scale production is not cost effective and the chemical wastes produced are an environmental hazard. Hence, there is need for more efficient biotechnological methods of biotin synthesis. A better understanding of the enzymes in this pathway, in general, would make it easier to manipulate these enzymes in microbes such as *E. coli*, *B. subtilis*, and *B. sphaericus* to synthesize a large quantity of biotin. Moreover, it is also of particular interest to study this biotin biosynthesis in pathogens like mycobacterium because of its role in various metabolic pathways, especially fatty acid metabolism, which is essential for the bacterial survival during chronic phase of infection (37).

Biotin biosynthesis in Mycobacterium tuberculosis

Bioinformatics studies on the *Mtb* H37Rv genome reveal the presence of all four enzymes in biotin biosynthetic pathway needed for the conversion of pimeloyl-CoA to biotin (38). However, not all the genes are present on the same operon - *bioAFD* are in one operon and *bioB* is in another operon as shown in the figure below (Figure 1.3).

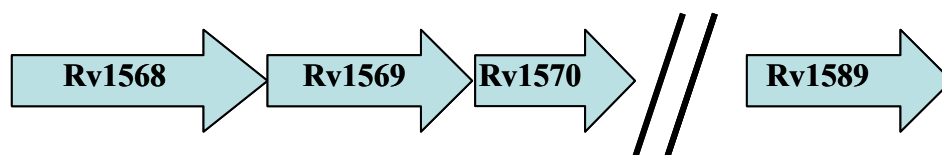


FIGURE 1.3. **Biotin operon in *Mtb***

The first step towards biotin biosynthesis is the decarboxylative condensation of L-alanine and pimeloyl-CoA into 7-keto-8-aminopelargonic acid (KAPA), catalyzed by KAPA synthase (*bioF*: Rv1569), a PLP-dependent enzyme. However, the source of the starting material, pimeloyl-CoA, in *Mtb* is not yet clear (more details in next paragraph). In the second step of biotin synthesis, KAPA is transaminated into 7, 8-diaminopelargonic acid (DAPA) by DAPA synthase (*bioA*: Rv1568), another PLP dependent aminotransferase, requiring S-adenosyl methionine (SAM) as the amino donor. The product of the second step, DAPA is further carboxylated into an ureido ring of dethiobiotin in an ATP and magnesium dependent manner. This penultimate reaction is catalyzed by dethiobiotin synthetase (*bioD*: Rv1570). Finally, dethiobiotin is converted to biotin with the introduction of a sulfur between C6 and C9 carbons of dethiobiotin by biotin synthase (*bioB*: Rv1589), requiring a FeS cluster and generating a free radical by the use of SAM (Figure 1.4).

The source of pimeloyl-CoA has been extensively studied in *Bacillus* and *E. coli*. Each organism follows a different pathway for the synthesis. In *B. subtilis* and *B. sphaericus*, two additional genes, *bioW* and *bioI*, are present in the biotin operon that are involved in this process of pimeloyl-CoA synthesis. Characterization of the *bioI* gene product in *B. subtilis* suggests a function homologous to cytochrome P450 in cleaving carbon-carbon bonds in free fatty acids to produce pimelic acid (39-41). Uptake of this pimelate is thought to be through passive diffusion. Pimeloyl-CoA is synthesized from pimelic acid with the hydrolysis of ATP, catalyzed by pimeloyl-CoA synthetase (*bioW* gene) (42). Other organisms which possess the *bioW* gene and use pimelate as the

precursor are *Staphylococcus aureus*, *Corynebacterium diphtheriae*, *Aquifex aeolicus* and *Methanococcus jannaschii*. In contrast, *E. coli* cannot take up pimelate as confirmed by radioactivity studies (43). Pimeloyl-CoA is synthesized from L-alanine via acetyl-CoA using proteins coded by *bioH* and *bioC* genes. Structural and biochemical studies on the BioH protein in *E. coli* suggest that it may act as an esterase as well as an acyltransferase, in transferring the pimeloyl unit from the bioC protein to coenzyme A to form pimeloyl-CoA (44, 45). The bioC protein, in turn, is thought to be a pimeloyl carrier protein (46). Even so, not all bacteria possessing the *bioC* gene contain the *bioH* gene in their genome but, rather, have various nonorthologous genes (like *bioG* in many proteobacteria or *bioK* in some cyanobacteria or *bioZ* in rhizobia like *Agrobacterium tumefaciens*) which can complement the *E. coli bioH* mutant. In *Mtb*, no homology was found with the genes related to the above two pathways.

In the biotin biosynthetic pathway, preliminary biochemical analysis of different enzymes has been carried out in microbes like *B. sphaericus* (42, 47) and *B. subtilis* (48), lower eukaryotes like *Saccharomyces cerevisiae* (49), and in plants like *Arabidopsis thaliana* (50). However, extensive biochemical and structural studies of all four enzymes of this pathway have been performed in *E. coli*. Several inhibitors are available for these different enzymes in *E. coli*. In plants, numerous inhibitor trials on the enzymes of this pathway were also carried out in search of effective herbicides (51-54).

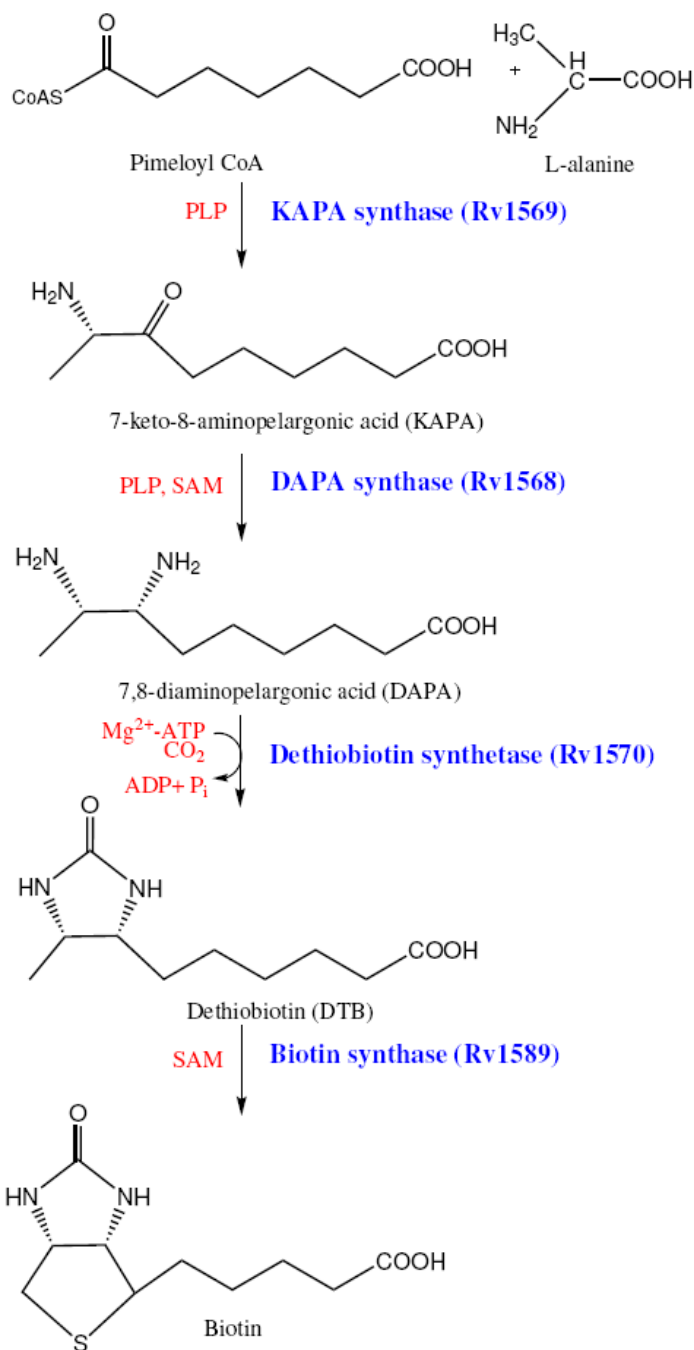


FIGURE 1.4. Biotin biosynthetic pathway in *Mtb*

Need for biotin biosynthetic pathway in Mtb

In recent years, several genetic studies have indicated the importance of the biotin biosynthetic pathway in mycobacteria during *in vitro* growth as well as infection. TraSH (Transposon Site Hybridization) studies on *Mtb* H37Rv strain suggest that genes related to different metabolic pathways are needed at different stages of infection in mice (9). TraSH involves transposon insertion mutation of almost 97% of the 4000 *Mtb* genes, followed by microarray mapping of these mutants. In this study, mutations in most of the genes of the biotin biosynthetic pathway resulted in rapid clearance of the bacterium at the early stages of infection in the mouse model (9). Also, knockout of the *bioF* gene in mycobacterium (H37Rv) has resulted in its inability to grow *in vitro* in culture (55) as well as to establish infection in mice, resulting in rapid clearance during early stages of infection (9). This is not surprising since biotin auxotrophy has also resulted in the attenuation of virulence in other bacterial pathogens like *Vibrio cholerae* (56). A separate study on the *bioA* gene in *Mycobacterium smegmatis* (*M. smegmatis*), an experimental model for mycobacterial physiology, suggests that this gene is needed for the survival in as well as recovery from the latent phase in *M. smegmatis* (57). In this study, the *bioA* mutant strain also survived poorly in rich medium containing biotin, suggesting that *de novo* biosynthesis of biotin is required in this stationary phase. Probably, these cells were not able to take up biotin from the medium. Microarray analysis on mycobacterium gene expression under nutrient-starved conditions shows that the *bioD* gene gets upregulated (2.4 fold) after 96 hours (58). This study was carried out to mimic bacterial persistence in infected hosts (37).

The biotin biosynthetic pathway in *Mtb* can be examined for potential antimicrobial development since only bacteria and plants have the enzymes for the synthesis of this vitamin, but mammals generally get the required biotin from their diet. This pathway came into the spotlight during antibiotic screening in the 1950's and 1970's. Two natural compounds - amikacin (MIC = 3.1 $\mu\text{g/ml}$) and acidomycin (MIC = 0.0625-0.125 $\mu\text{g/ml}$) from *Streptomyces* species were demonstrated to have good whole cell activity against mycobacterium (59, 60). Further studies have indicated that their inhibitory effect is targeted to the biotin biosynthetic pathway (61, 62). Amikacin and acidomycin target the second and fourth enzymes of the pathway. Together, they had a synergistic effect against mycobacterium. However, the interest in this pathway faded due to the ineffectiveness of these antibiotics *in vivo* in animals. This was thought to be due to the ability of *Mtb* to take up sufficient biotin from the host. However, studies conducted at Abbott laboratories, North Chicago, IL, have shown that acidomycin possesses poor bioavailability (63), with 31% of acidomycin recovered from urine within the first hour of administration and the remainder of the dose inactivated by the liver, which might be the reason for its poor performance. Information regarding the bioavailability of amikacin is not available. We still believe that these compounds can provide ideas for developing new inhibitors by modifying their chemical structures, resulting in better biological properties.

In short, this dissertation deals with the characterization of the first enzyme of the serine biosynthesis as well as the first three enzymes of the biotin biosynthesis. These two metabolic pathways are critical in mycobacterium survival during infection.

Elucidation of these enzymes will be helpful in designing inhibitors against these enzymes for the eradication of tuberculosis.

Overview of chapters

Chapter II of this dissertation discusses the structure-function relationship of the first enzyme of the phosphorylated serine biosynthetic pathway, D-3 phosphoglycerate dehydrogenase in *Mtb*. Sequence alignment of PGDH from different organisms suggests that this enzyme in *Mtb* represents a unique form, with a long insertion near the C-terminus, which has not been characterized earlier. X-ray crystallographic methods were used to elucidate the domain organization, role of this extended C-terminal region, mode of substrate binding and catalysis, and serine inhibition in this enzyme. Additionally, biochemical analysis of *Mtb* PGDH has been done in collaboration with Dr. Gregory A. Grant (Washington University School of Medicine, St. Louis, MO). The substrate-bound PGDH structure helps us in mapping the active site. Based on the ligand-bound human truncated PGDH structure, the resulting model of a closed active site provides us with a picture of the catalytic events in this enzyme. Kinetic studies have also indicated that this enzyme is allosterically regulated by L-serine (64), as is *E. coli* PGDH (65). Structural determination of effector-bound *Mtb* PGDH pinpoints the residues involved in the allosteric regulation and the resulting conformational changes. Human PGDH has the same form as *Mtb* PGDH, so chapter II also describes the cloning, protein overexpression and purification procedures for human PGDH, and characterizes and compares this enzyme with its *Mtb* counterpart.

Chapter III deals with the biochemical and three-dimensional structure analyses of 7, 8-diaminopelargonic acid synthase (DAPAS, EC 2.6.1.62), an enzyme involved in the second step of biotin biosynthesis. A unique feature of this aminotransferase enzyme is its usage of SAM as the amino donor. Comparison of the substrate-bound *Mtb* DAPAS structures sheds light on the mode of dual-substrate recognition in this particular transaminase, which was not characterized to date. This chapter also reveals the variation in the mode of substrate binding at the DAPAS active site in certain *Bacillus* species and thermophiles using KAPA-bound *B. subtilis* structure. Following characterization of the *Mtb* DAPAS active site, virtual screening of commercially available drug-like compounds was carried out. In this process, the substrate-bound active site of *Mtb* DAPAS was used as a template and the resultant top hits were tested for inhibitor development using enzyme assays.

Chapter IV elucidates the structural details of the penultimate enzyme in the biotin biosynthesis in *Mtb*, dethiobiotin synthetase (DTBS, EC 6.3.3.3). This study outlines the differences in overall structural conformations in comparison to the enzyme found in *E. coli*, as well as, in *H. pylori*. It further explains the mode of substrate recognition at the active site and the mechanism of this amidoligase reaction, based on the substrate and the product-bound DTBS structures.

Chapter V of this dissertation discusses the cloning, protein overexpression and purification protocols, crystallization trials, and preliminary biochemical characterization of 7-keto, 8-aminopelargonic acid synthase (KAPAS, 2.3.1.47), an enzyme involved in the first step of the biotin biosynthetic pathway in *Mtb*.

CHAPTER II
STRUCTURAL STUDIES ON D-3-PHOSPHOGLYCERATE
DEHYDROGENASE IN *Mycobacterium tuberculosis**

Background

D-3-phosphoglycerate dehydrogenase (PGDH) catalyzes the reversible reaction of oxidation of phosphoglycerate (PGA) to hydroxypyruvic acid phosphate (HPAP) in an NAD^+ dependent manner (Figure 2.1). This is the first step in the serine biosynthetic pathway. This enzyme is a member of the 2-hydroxy acid dehydrogenase family that is specific for substrates with a D-configuration (66). It has sequence/structural homology with other members of this family, e.g. formate dehydrogenase (67), D-glycerate dehydrogenase (68), D-lactate dehydrogenase (69), and erythronate-4-phosphate dehydrogenase (70). PGDH is present in both prokaryotes and eukaryotes.

* Part of this chapter is reproduced with permission from “Crystal Structure of *Mycobacterium tuberculosis* D-3-Phosphoglycerate Dehydrogenase: extreme asymmetry in a tetramer of identical subunits”, Dey, S., Grant, G.A., Sacchettini, J.C. (2005), *The Journal of Biological Chemistry*, 280, 14892-14899, Copyright 2005 by The American Society for Biochemistry and Molecular Biology, Inc.

* Part of this chapter is reproduced with permission from “Structural Analysis of Substrate and Effector Binding in *Mycobacterium tuberculosis* D-3-Phosphoglycerate Dehydrogenase”, Dey, S., Burton, R.L., Grant, G.A., Sacchettini, J.C. (2008), *Biochemistry*, DOI 10.1021/bi800212b, Copyright 2008 by American Chemical Society.

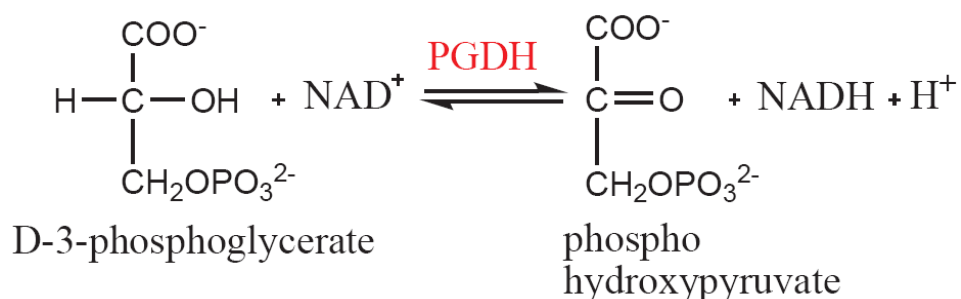


FIGURE 2.1. **Reversible enzymatic reaction catalyzed by PGDH.**

Different forms of PGDH

Sequence alignment of PGDH protein from various species indicates that this enzyme exists in at least three different basic structural forms. None of which appear to be strictly specific for organism type (Figure 2.2). The simplest form, which is found in some bacteria including *Clostridium* and *Pyrococcus*, contains only two domains- the nucleotide-binding domain and the substrate-binding domain (Figure 2.3). The dimeric PGDH from the unicellular parasite *Entamoeba histolytica* (22) differs from other forms of PGDH in having a lysine residue rather than a histidine as one of the members of the catalytic dyad at the active site. The PGDH from some bacteria and lower eukaryotes such as *Saccharomyces cerevisiae* (71) and *Leishmania major* (72) shares similar motif with the *E. coli* enzyme. This enzyme has been extensively studied in *E. coli*, both biochemically (65, 73) as well as structurally (74, 75). In *E. coli* PGDH, in addition to the substrate and nucleotide binding domains, this enzyme possesses a C-terminal domain that is involved in regulation of enzyme activity through binding the effector, L-serine (Figure 2.4). Other microbes such as *Mycobacterium*, *Bacillus subtilis* (76) and

Corynebacterium glutamicum, plants such as *Arabidopsis thaliana* (20), and higher order eukaryotes, including mammals, possess a large polypeptide insertion (~130-200 amino acids) in their C-terminal segment between the substrate-binding and regulatory domains (Figure 2.5).

Kinetic properties of Mtb PGDH

The *serA1* gene that codes for PGDH has been found to be an essential gene by Himar1-based transposon mutagenesis studies in *Mtb* H37Rv strain (77). This gene sequence in *Mtb* is also nearly identical to that in *M. leprae* (89.6%). Generally, only the essential genes are conserved in *M. leprae* because of its long generation time (11-13 days in mice), thus, supporting the essentiality of *serA1* gene in *Mtb* (12, 13). Furthermore, kinetic studies on *Mtb* PGDH provide a K_m value of ~ 85 μM for substrate HPAP and a k_{cat}/K_m of ~ $5.6 \times 10^6 \text{ M}^{-1}\text{sec}^{-1}$ (78) in comparison to the *E. coli* K_m value of ~ 40 μM for substrate HPAP and a k_{cat}/K_m of ~ $7 \times 10^5 \text{ M}^{-1}\text{sec}^{-1}$ (64). In *Mtb* PGDH, the backward reaction from HPAP to PGA is the favored reaction. It has also been demonstrated that the activity of *Mtb* enzyme is greatly affected by the ionic strength of the reaction buffer (64). In addition, *Mtb* PGDH exhibits uncompetitive substrate inhibition at higher concentrations of HPAP (250 μM) (64) analogous to rat liver PGDH (29). This type of substrate inhibition has not been reported in *E. coli* PGDH.

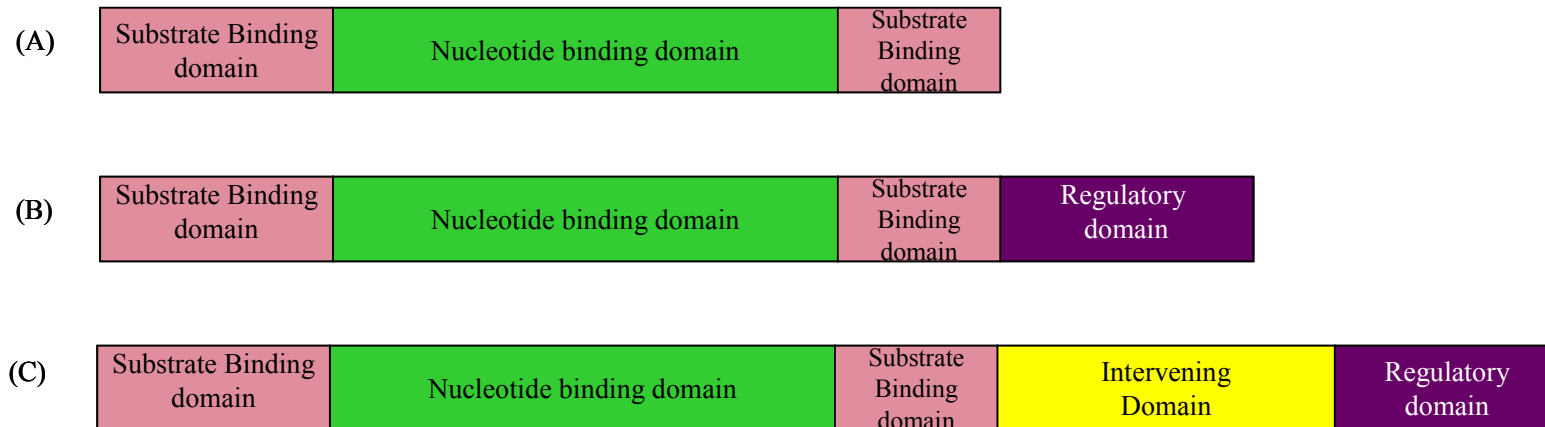


FIGURE 2.2. Three different forms in PGDH based on sequence alignment. The polypeptide arrangement of PGDH falls into three categories. (A) PGDH containing only two domains- substrate-binding and nucleotide-binding as in *P. horikoshii*, *Clostridium*, and *E. histolytica*. (B) PGDH containing a nucleotide-binding domain, a substrate-binding domain and an additional domain- regulatory domain as in *E. coli*, *S. cerevisiae*, and *Leishmania*. (C) PGDH with an insertion of ~130-200 amino acids between the substrate-binding and regulatory domains as in human, rat liver, *Mtb*, and *B. subtilis*.

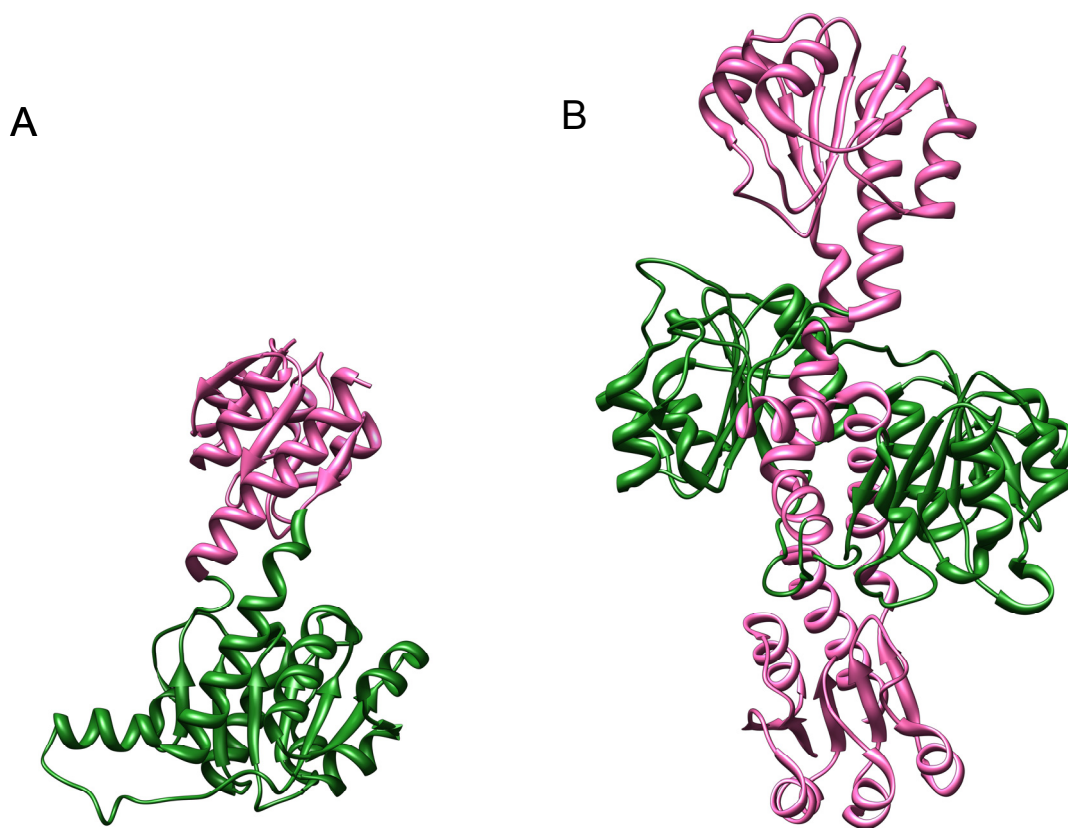


FIGURE 2.3. **Domain organizations of PGDH in *Pyrococcus horikoshii*.** Ribbon diagram representation of (A) monomer (B) dimer of PGDH in *P. horikoshii*. The nucleotide-binding domain is colored green and the substrate-binding domain is colored pink. This PGDH exists as a homodimer as it lacks the regulatory domain.

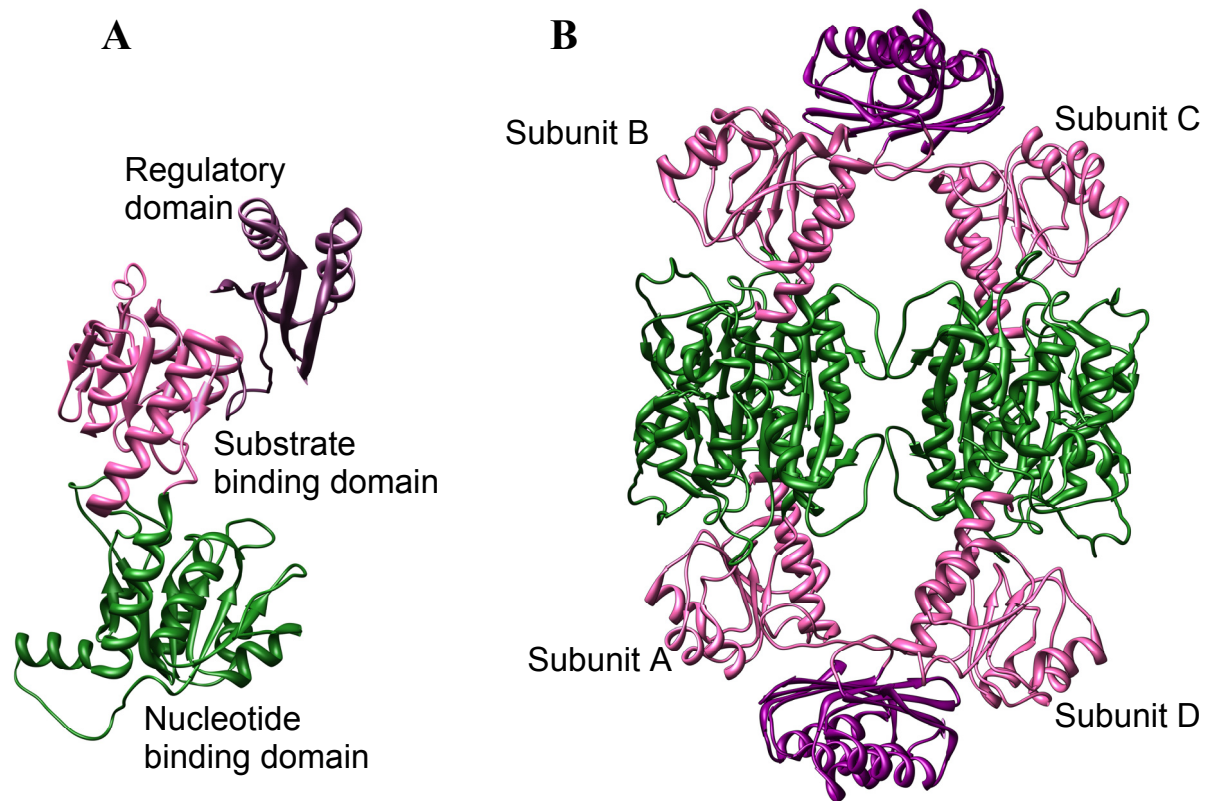


FIGURE 2.4. **Domain arrangements of the *E. coli* PGDH.** Ribbon diagram representation of *E. coli* PGDH (PDB code-1YBA). (A) monomer (B) tetramer. The nucleotide-binding domain is depicted in green, the substrate-binding domain in pink and the regulatory domain in violet.

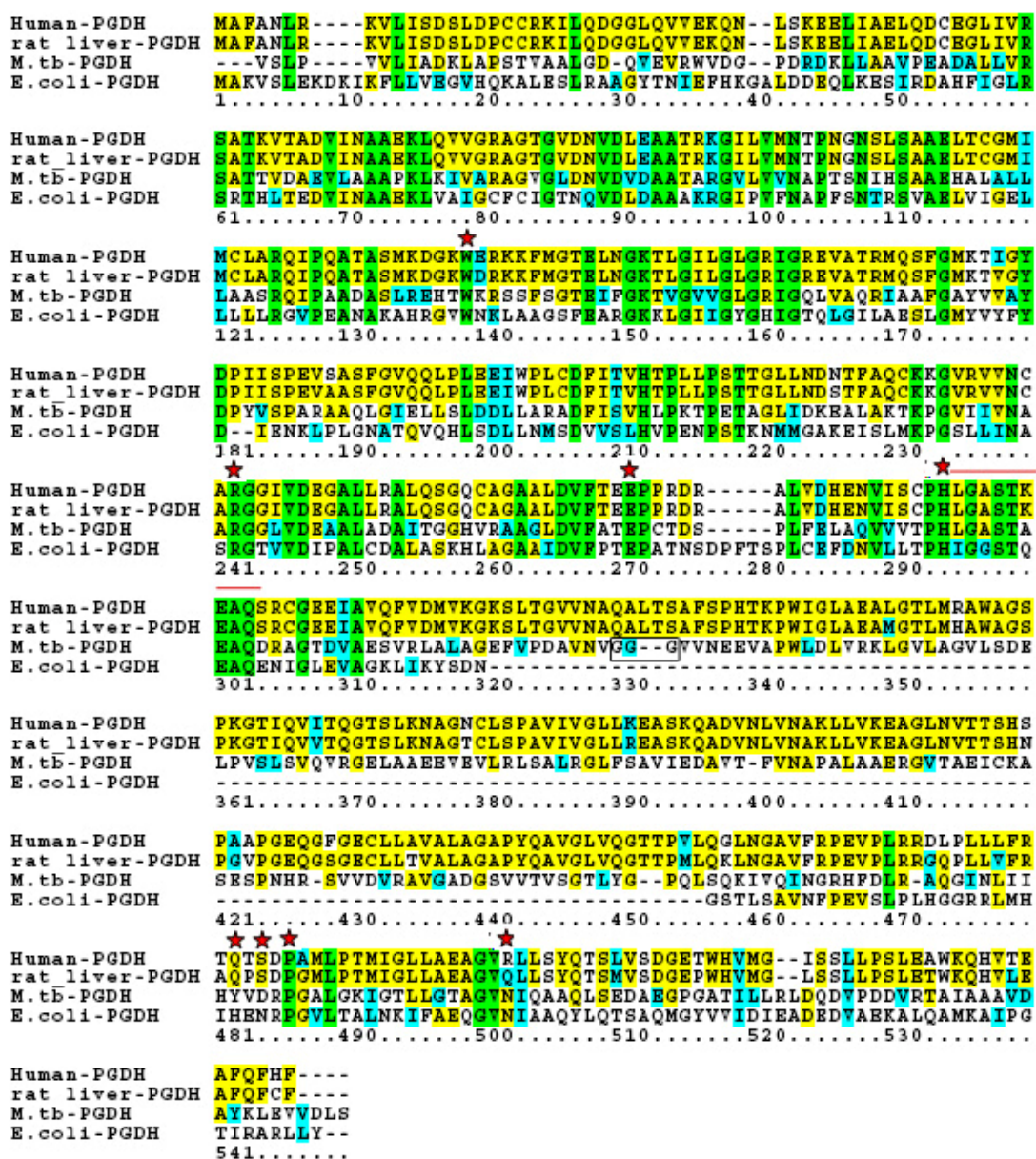


FIGURE 2.5. Alignment of PGDH sequences from human, rat liver, *Mtb* and *E. coli*.

The residues colored in green are the conserved residues in all four PGDH sequence. Identical and similar residues are shown in yellow and cyan. This alignment is prepared using ClustalW (79). All the important residues involved in catalysis and allosteric regulation are marked with a red asterisk. The stretch of three glycine residues (316-318) in *Mtb* is highlighted with a box.

Serine inhibition of PGDH

The phosphorylated pathway of serine biosynthesis is mostly regulated by feedback inhibition of the first enzyme, PGDH, by the end product of the pathway, L-serine. *E. coli* PGDH has been found to be strongly inhibited by L-serine ($IC_{50} \sim 2\text{-}4 \mu\text{M}$) (80). *B. subtilis* PGDH is less sensitive to inhibition by L-serine ($IC_{50} \sim 0.6 \text{ mM}$) and any sensitivity to L-serine appears to be lost under oxidizing conditions (76). The enzymes from *E. coli* and *B. subtilis* are the only homogeneous preparations of PGDH that have been reported to be inhibited by L-serine. While *C. glutamicum* PGDH has also been reported to be inhibited by L-serine, although at very high concentrations ($IC_{50} \sim 10 \text{ mM}$). It has not been studied in homogeneous form (81). In addition, L-serine inhibition of both the *B. subtilis* and *C. glutamicum* PGDH require extensive pre-incubation of the enzyme with the inhibitor before appreciable inhibition can be measured. In the pea (*Pisum sativum*), the sensitivity to L-serine has been reported to be cold labile (82). In crude extracts of wheat germ (83), PGDH activity appears to be inhibited by high concentrations of L-serine, but this sensitivity is lost upon incubation, since the purified wheat germ PGDH is not inhibited by L-serine. Interestingly, L-serine has been reported to have no effect on the activity of PGDH from rat liver (29) and chicken liver (28). *Mtb* PGDH, which belongs to the same form of PGDH as mammals, is inhibited by L-serine with an IC_{50} for L-serine of $\sim 30 \mu\text{M}$, making it the second most sensitive enzyme after *E. coli* PGDH, reported so far. Hence, it appears that there is a lot of variation in the serine inhibition of PGDH with insertion near the C-terminus.

Mtb PGDH is also a V-type enzyme (84) which affects the velocity of the enzyme reaction, similar to *E. coli* PGDH (65). It shows mixed noncompetitive inhibition (64). Some of the other enzymes which also show V-type allosteric regulation are fructose 1, 6 biphosphatase (85), glycerol kinase (86), carbamoyl phosphate synthetase (87), and pyruvate carboxylase (88).

In this chapter, the crystal structures of PGDH from *Mtb* in its apo form, as well as, in complex with its substrate HPAP, and with the inhibitor L-serine will be discussed. We also discuss the cloning and protein purification procedures of different constructs of *Mtb* PGDH. This PGDH structure represents the enzyme with an extended motif near the C-terminus. The apostructure sheds light on how this 130 amino acid insertion affects the domain organization and also provides an image of the effect of this insertion on the overall structure of this PGDH which still maintains the tetrameric form. Remarkably, although the tetramer is composed of symmetry related heterodimers, extreme asymmetry is seen in the tertiary structure of the subunits. The apostructure also hints at a possible role for this insertion as an alternative substrate-binding site, responsible for substrate inhibition of *Mtb* PGDH.

Two other crystal structures of PGDH have been determined since the structure of *Mtb* PGDH apoenzyme was established. One is a truncated form of human PGDH, consisting of the nucleotide-binding and substrate-binding domains (PDB code- 2G76). Native human PGDH is similar to *Mtb* PGDH with the intervening and the regulatory domains. The other structure is from *P. horikoshii* (PDB code- 1WWK), which has the simplest form of PGDH, consisting only of nucleotide-binding and substrate-binding

domains. This PGDH exists as a dimer. Further biochemical characterization of PGDH from either human or *pyrococcus* has not been reported.

Furthermore, the crystal structure of the binary complex of *Mtb* PGDH with the substrate HPAP aids us in identifying the residues involved in substrate binding at the active site. Modeling of the active site closure of *Mtb* PGDH on substrate binding provides us with an insight into the catalytic event occurring in this dehydrogenase. It will also be intriguing to understand the mechanism of serine inhibition in this particular form of PGDH which shows so much variation in serine inhibition. The serine-bound *Mtb* PGDH structure reveals the key players involved in this allosteric regulation as well as the conformational changes occurring on serine binding, both of which are quite different from those of *E. coli* PGDH.

Human PGDH also belongs to the same form of PGDH as *Mtb*. In eukaryotes, the PGDH structure has not been well characterized. Hence, in this study, we will discuss the cloning and protein purification protocols of human PGDH in an effort to characterize and compare it with the *Mtb* PGDH.

Methods

Reagents

Restriction enzymes for *serA* cloning were purchased from New England Biolabs. All the reagents for purification purposes were from Sigma. Selenomethionine was purchased from Acros. HPAP was synthesized from dimethylketal tri-cyclohexylammonium salt (Sigma) using the protocol of Ballou (89). After the

hydrolysis of ketal to HPAP, aliquots of HPAP were stored at -20 °C. NADH, NAD analog (3-acetyl pyridine dinucleotide) and L- serine were also purchased from Sigma.

Cloning, protein overexpression and purification of Mtb PGDH

His-tagged PGDH clone was obtained by amplifying *serA1* gene from *Mtb* H37Rv (lab strain) genomic DNA and further ligating into pET28a vector (Novagen) using *NdeI* and *HindIII* restriction sites. The *NdeI* site contained the ATG start codon and the *HindIII* site was preceded immediately by a stop codon (TCA) following the C-terminal residue of the enzyme, resulting in an N-terminus His-tag protein.

PGDH without a His-tag was obtained by cloning the Rv2996c gene into pET30b vector (Novagen) using *NdeI* and *HindIII* sites. Production of a selenomethionylated (SeMet) protein for MAD (Multiple Anomalous Dispersion) phasing was accomplished by mutation of Ile (ATC) at position 514 to Met (ATG). This mutant was constructed using the QuikChange site-directed mutagenesis kit (Stratagene) with pET30b-*serA1* as a template. The mutated gene was cloned into the pET30b vector in the same way as the wild type. All the constructs were verified by DNA sequencing. The primers used in this study were:

For the His-tagged *Mtb* PGDH

Forward primer

5'AGAGAAGCATATGGTGAGCCTGCCTGTTGTGTTGATCGCGAC3'

Reverse primer

5'CCCAAGCTTTCACGACAGATCGACAACCTCGAGCTTGTAGGCGT3'

For PGDH I514M mutant:

Forward primer: 5' GCGGACGGCGATGGCGGCGGCGG 3'

Reverse primer: 5' CCGCCGCCGCCATCGCCGTCCGC 3'

The pET28a-*serAI* as well as pET30b-*serAI* clones were both transformed into the *E. coli* over-expression strain, BL21 (DE3) (Novagen). The cell culture was grown in LB media with kanamycin (50 µg/ml) at 37 °C until the A₆₀₀ reached 0.6-0.8 and then induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and grown at 18 °C overnight. For SeMet protein, the pET30b-*serAI* (I514M) clone was transformed into *E. coli* B834 (DE3) (Novagen), a methionine auxotroph. Cells were grown in M9 minimal media with all the 19 standard amino acids except Met. Selenomethionine (50 mg/L) and kanamycin (50 µg/ml) were also added to the cell culture. Culture was induced with 1 mM IPTG when A₆₀₀ reached 0.8 and then grown overnight at 18 °C. The cells were harvested by centrifugation at 3000 rpm for 30 min and resuspended in buffer A (depending on the type of purification) supplemented with one tablet of Complete EDTA-free protease inhibitor cocktail (Roche) per cell pellet.

For the His-tagged protein, the cell pellet was resuspended in 20 mM potassium phosphate buffer (pH 7.5), 500 mM NaCl (buffer A). The cells were then lysed using a French press (11,000 psi) and the cell debris was removed by centrifugation at 15,000 rpm for 1 hr. The 0.22 µ filtered supernatant was loaded on a Hitrap Ni column (GE healthcare), pre-equilibrated with buffer A + 5 mM imidazole and then washed with 25 mM imidazole containing buffer A to remove the contaminants. The protein was eluted with imidazole gradient of 25–400 mM. For untagged PGDH protein, the supernatant

was loaded onto a pre-equilibrated (20 mM potassium phosphate buffer (pH 7.5), 5 mM KCl) anion exchange Q sepharose Fast Flow column (GE healthcare) and the contaminants were washed till absorbance reached zero. The protein was eluted using a KCl gradient (50-800 mM). The fractions containing pure PGDH were pooled and concentrated using Centriprep (Millipore). This concentrated protein was applied to a Superdex 200 gel filtration column (GE healthcare) pre-equilibrated with 100 mM potassium phosphate buffer (pH 7.5), 1 mM DTT and 1 mM EDTA. The eluted protein was 98% pure as judged from the 4-12% gradient SDS-PAGE gel. The entire purification procedure was carried out at 4 °C.

Cloning, protein overexpression and purification of human PGDH

The pET30b-*huserA* clone was provided to us by Dr. G. A. Grant (Washington University School of Medicine, St. Louis). *HuserA* gene codes for human PGDH. The pET28b-*huserA* clone for N-terminal His-tagged PGDH protein with TEV cleavage site, as well as, N-terminal His-tagged truncated form of PGDH (Ser322-Phe530) was constructed, using the pET30b-*huserA* as a template with *NdeI* and *HindIII* restriction sites.

The primers used for the N-terminal His-tagged protein,

Forward primer: 5' AGAGAAGCATATGGCTTTTGCAAATCTGCGGAAAGTG 3'

Reverse primer: 5' CCCAAGCTTTTAGAAGTGGAAGTGGAAAGG 3'

For truncated form of human PGDH (323-532) (N-terminal His-tag),

Forward primer: 5' AGAGAAGCATATGACCAAGCCTTGGATTGGTCTGGCA 3'

Reverse primer: 5' CCCAAGCTTTTAGAAGTGGAAAGTGGAAAGG 3'

All the constructs were further verified by DNA sequencing.

All three clones were transformed into the *E. coli* over-expression strain, Rossetta (DE3) (Novagen). The cell culture was grown in LB media with kanamycin (50 µg/ml) and chloramphenicol (34 µg/ml) at 37 °C until the A_{600} reached 0.6-0.8 and then induced with 1 mM IPTG and grown at 18 °C overnight. The cells were harvested by centrifugation at 3000 rpm for 30 min and resuspended in buffer A (depending on the type of purification) supplemented with one tablet of Complete EDTA-free protease inhibitor cocktail (Roche) per cell pellet.

For the His-tagged human PGDH protein, the cell pellet was resuspended in 20 mM potassium phosphate buffer (pH 7.5), 500 mM NaCl (buffer A). The cells were then lysed using a French press (11,000 psi) and the cell debris was removed by centrifugation at 15,000 rpm for 1 hr. The 0.22 µ filtered supernatant was loaded on a Hitrap Ni column (GE healthcare), pre-equilibrated with buffer A + 5 mM imidazole and washed with 25 mM imidazole containing buffer A to remove the contaminants. The protein was eluted with imidazole gradient of 25–400 mM. The His-tag was further cleaved using 1 mg TEV protease per 20 mg purified protein at 4 °C in overnight dialysis against a buffer containing 20 mM phosphate buffer, 200 mM NaCl, 10% glycerol and 2 mM DTT. The cleaved protein was reloaded on Ni column and was eluted as a flow through.

For untagged human PGDH protein, the supernatant was loaded onto a pre-equilibrated (20 mM potassium phosphate buffer (pH 7.5), 5 mM KCl) anion exchange Q sepharose Fast Flow column (GE healthcare) that was washed till absorbance

decreased to zero and, finally, the protein was eluted using a KCl gradient (50-800 mM). The fractions containing pure PGDH were pooled and concentrated by ultrafiltration using Centriprep (Millipore) and applied to a Superdex 200 gel filtration column (XK 16/60, GE healthcare) pre-equilibrated with 100 mM potassium phosphate buffer (pH 7.5), 100 mM NaCl, 10% glycerol, 1 mM DTT and 1 mM EDTA. 2ml fractions were collected at the rate of 0.5 ml/min. The protein was 98% pure as analyzed from a 4-12% gradient SDS-PAGE gel. The entire purification procedure was carried out at 4 °C.

Fluorescence studies of Mtb PGDH

Mtb PGDH (5 µM) was excited at 280 nm and fluorescence emission spectra was scanned in the range of 300-450 nm on a LS 55 luminescence spectrometer (Perkin Elmer Instruments) with constant stirring. Further emission spectra were collected for the enzyme reaction with 200 µM NADH as well as with 0.5 mM HPAP in a 500 µl total reaction volume. All the experiments were carried out in buffer containing 50 mM potassium phosphate (pH 7.5) and 150 mM NaCl at room temperature.

Crystallization of Mtb and human PGDH

His-tagged *Mtb* PGDH protein was initially screened for the purpose of crystallization. Tiny rod shaped crystals in several precipitant conditions were obtained, but further optimization of these conditions did not lead to the improvement of the crystal quality. So crystallization of *Mtb* PGDH without any tag was attempted which eventually provided good quality crystals. Native and SeMet *Mtb* PGDH (13 mg/ml) were both crystallized in 1 M NaK tartrate, 0.1 M MES (pH 6.5) at 18 °C by the vapor diffusion method in hanging drop plates. This precipitant condition was found to

produce crystals of good diffraction quality after setting up a matrix screening with different kits (Hampton Crystal Screen and II, Wizard screen I and II from Emerald BioSystems). Crystals grew within 3-4 days in 4 μ l hanging drops of 2 μ l of PGDH (13 mg/ml) and 2 μ l of precipitant (Figure 2.6). For SeMet protein, crystals took more than a week to grow. The substrate HPAP-bound and inhibitor serine-bound PGDH crystals were obtained by cocrystallising PGDH (13 mg/ml) with 5 mM HPAP + 5 mM NAD analog (3-acetyl pyridine dinucleotide) and 5 mM L-serine + 5 mM NADH, respectively in the same precipitant condition. There were several other precipitant conditions which resulted in crystal formation in the same space group, but they all diffracted poorly. For data collection purposes, crystals were flash frozen in liquid nitrogen using 20% propylene glycol in mother liquor as the cryoprotectant.

Full length human PGDH (10 mg/ml) and truncated human PGDH in complex with 5 mM NADH were also subjected to similar crystallization screening using different matrix screens.

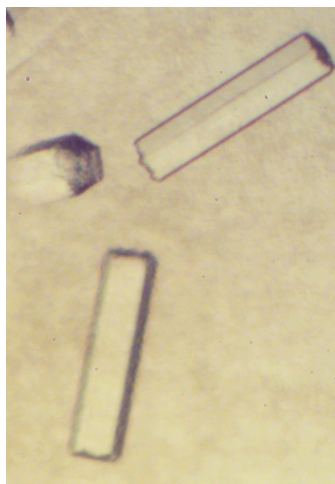


FIGURE 2.6. **Hexagonal rod shaped crystals of *Mtb* PGDH.** Crystals grew to dimensions of $50\ \mu\text{M} \times 50\ \mu\text{M} \times 200\ \mu\text{M}$ in 1 M NaK tartrate, 0.1 M MES (pH 6.5) using hanging drop vapor-diffusion technique.

Data collection and processing

The initial native dataset for *Mtb* PGDH was collected to 2.8 Å resolution at beamline 14-BMC, Advanced Photon Source (APS), Argonne National Laboratory, Chicago. The diffraction dataset was processed and scaled using the HKL2000 suite (90). The space group was found to be $P6_522$. Solvent content calculations using Mathews coefficient (91) indicated either a dimer ($V_m = 3.9$, $V_s = 68.3\%$) or a trimer ($V_m = 2.6$, $V_s = 52.4\%$) in the asymmetric unit. Phasing by molecular replacement was attempted using the *E. coli* PGDH monomer (PDB code-1PSD) or its nucleotide-binding domain (the largest domain) as a search model. No clear solution was obtained using various programs (AMoRe (92), MOLREP (93), PHASER (94)). Subsequently, several different compounds of heavy metals (mercury, platinum, iodine and neodymium) were tried, but they failed to produce an acceptable isomorphous or anomalous signal. SeMet

MAD (Multiple Anomalous Dispersion) phasing of PGDH was originally not possible as this 529 amino acid protein did not contain any methionine residues except the N-terminal methionine which might be cleaved during expression. Based on sequence alignment with *E. coli* PGDH, a single methionine residue was introduced at position 514. This mutation was sufficient to obtain complete and highly redundant Se MAD datasets for PGDH at 3.15 Å, collected at beamline 14-BMD, APS. High resolution datasets of apoenzyme, HPAP, and serine complexed PGDH at 2.3 Å, 2.4 Å, and 2.7 Å were also obtained at beamline 14-BMC, APS. All the datasets were processed and scaled using the HKL2000 package (90). Datasets collected at selenium peak (0.9794 Å) and inflection wavelengths (0.9796 Å), along with native data were enough to provide the phase information for *Mtb* PGDH apoenzyme.

Basics of structure determination

X-rays are electromagnetic waves that interact with the electrons and because of which the electrons oscillate with the same frequency as the incident wave acting as radiation scatters. The range of wavelength particularly useful for protein-crystal structure determination is 0.5-2.5 Å. For in-house source, the X-ray of Cu K α wavelength (1.541 Å) can be generated by rotating anode generators. At synchrotron, particle storage rings are used to generate intense X-ray beam in a whole range of spectrum from 0.5-1.6 Å.

Bragg's law interprets X-ray diffraction by a crystal lattice as reflections from different planes of molecules/lattice points in the crystal.

For constructive interference, this can be written as the equation:

$$2d \sin\theta = n\lambda,$$

where d is the distance between each plane of molecules, θ is the angle of incidence of the X-ray beam, n is an integer and λ is the wavelength of the X-ray beam.

The ultimate goal of protein crystallography is to determine the electron density (ρ) at every position (xyz) of the unit cell. The electron density map ρ_{xyz} can be obtained by Fourier transformation of structure factor F_{hkl} ,

$$\rho_{xyz} = 1/V \sum_{hkl} |F_{hkl}| e^{-2\pi i(hx + ky + lz) + i\alpha_{hkl}}$$

where V is the volume of the unit cell. It is easy to derive the amplitude $|F_{hkl}|$ from the experimental intensities as amplitude is square root of intensity, but the phase angle α_{hkl} cannot be obtained directly from the diffraction pattern. There are three different methods by which this "phase problem" in protein crystallography can be solved. (1) Molecular replacement (MR) (2) Multiple Isomorphous Replacement (MIR), and (3) Multiple Wavelength Anomalous Dispersion (MAD)

The simplest of the three methods for phase determination, molecular replacement (MR) depends on the existence of a solved model that is homologous to the present unknown protein. Further detailed explanation of this method will be provided in chapter III.

Multiple Isomorphous Replacement (MIR) (95) involves either soaking the protein crystal or co-crystallizing with multiple heavy metals. When heavy metals are introduced in the protein, the crystal packing might be affected, resulting in weak

diffraction or loss of isomorphism. The major aim of this method is to get isomorphous data since datasets are collected from different derivative crystals. Isomorphism means almost same unit cell dimensions for both the derivative as well as the native datasets to minimize the error in phase determination. The stable addition of one or more heavy atoms will change the diffraction pattern in comparison to the native crystal as represented by the equation

$$F_{PH} = F_P + F_H,$$

where F_P is the structure factor of the protein, and F_H and F_{PH} are the structure factors for the heavy atom and the derivative protein, respectively. Structure factors can be represented as vectors, with lengths corresponding to their magnitudes and directions corresponding to their phases. There are two different ways to construct vector triangles with two sides whose lengths are $|F_P|$ and $|F_{PH}|$ on the line F_H , resulting in two values for the phase angle α . More than one derivative is required to eliminate this phase ambiguity. The second derivative will also have two possible phase angles for F_P provided the heavy atoms are not at the same positions as in the first derivative. One of these angles will be common between both the derivatives, thus, determining the correct phase.

In situations where the molecular replacement method fails or the structure is totally novel (without any available homologous structure), the method of choice in recent years is single/multiple wavelength anomalous dispersion (SAD/MAD) (96).

MAD (97) involves the integration of heavy metals inside the protein or naturally-occurring atoms in the protein (Fe, Cu or S) and collection of datasets of the

same crystal at different wavelengths to solve the phase problem. Use of single crystal for data collection eliminates any error in phase determination because of non-isomorphism. Since the datasets need to be collected at different wavelengths, a tunable X-ray source such as in synchrotron is required. For *Mtb* PGDH, we have incorporated selenium as the heavy metal, replacing sulfur in methionine, to get the anomalous signal. When X-ray beam hits the PGDH crystal, it sets the electrons to vibrate, generating energy at same frequency (X-ray). This kind of diffraction is “coherent scattering”. If the photon has energy close to transition energy, the electron can go to an excited state and instead of emitting in phase with the incident beam; energy is radiated in a different phase. In this case, the energy of coherent scattering is also reduced as some energy is absorbed to bring the transition. The atomic scattering factor can be written as an equation

$$f(\lambda) = f_0 - \delta f'(\lambda) + i f''(\lambda),$$

where f_0 is the conventional atomic scattering factor,

$\delta f'(\lambda)$ is the reduction of normal scattering at wavelength (λ), and

$i f''(\lambda)$ is the amount of anomalous scattering (out of phase component of the scattering) at this wavelength (λ)

Defining the atomic scattering factor, $f' = f_0 - \delta f'$, the equation becomes

$$f = f' + i f''$$

The anomalous scattering of selenium (f'') will be large at its absorption edge wavelength (0.9794), also called peak wavelength. At inflection (edge) wavelength (0.9796), the normal atomic scattering (f') is reduced to its lowest possible value.

Anomalous scattering disobeys Friedel's law for normal scattering, which states that structure factor for reflections h, k, l and $-h, -k, -l$ have the same amplitude but have opposite signs for phases.

$$|F_{hkl}| = |F_{-hkl}| \quad \alpha_{hkl} = -\alpha_{-hkl}$$

In this case, the amplitude of the Bijvoet pairs $|F_{hkl}|$ and $|F_{-hkl}|$ is different because of the presence of the additional anomalous scattering factor f'' . This difference in diffraction is enough to locate the position of the heavy metal in the unit cell and thereby determine the phase angle (Figure 2.7).

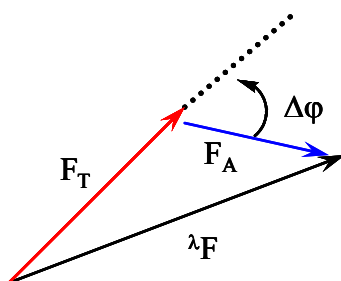


FIGURE 2.7. **Principles of MAD phasing.** F_T is the vector for normal scattering from all atoms, F_A is the vector for anomalous scattering and λF is the vector for total scattering at a particular wavelength (λ) and $\Delta\phi$ is the phase angle difference. During the MAD experiment, the amplitude of total scattering $|\lambda F|$ is measured.

Two selenium sites with occupancies of 1.0 and 0.96 in PGDH were found using direct method SHELXD (98). These sites were further refined and initial phases were obtained using AUTOSHARP (99) with an overall anomalous phasing power of 0.26 and figures of merit of 0.22 and 0.07 for acentric and centric reflections, respectively, for

overall 3.15 Å resolution range. The phases obtained by solvent flattening with DM (100) has an improved figure of merit of ~ 0.72, having two molecules in the asymmetric unit.

Using this unbiased density modified map, chain tracing was done using CAPRA in TEXTAL (101). Several rounds of manual model building was carried out with the Xfit module of the graphics program Xtalview (102) using the Shake&wARP (36) bias minimized electron density maps. The model was further refined against the high resolution (2.3 Å) data using REFMAC (103). The geometry of the model was greatly improved using CNS (104) in the initial stages of model building (rigid body refinement and then simulated annealed refinement at 4000 °K). In rigid body refinement, segments of protein such as domains move as rigid body. Simulated annealed refinement involves heating up of the model to high temperature (4000 °K) and slowly cooling it with the intention to find its preferred geometry at the temperature of data collection i.e. 100 °K. After the model reached a reasonable R factor and R_{free} (below 30%), water molecules were added to the model using automated water picking option in Xtalview and also manually picked using the unbiased map. Reliability factor or R-factor is an indicator of the agreement between the model and the observed structure factor.

It can be computed as:

$$R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

where F_{obs} and F_{cal} are the structure factors of the observed and calculated data.

R_{free} value, which was introduced by Brunger in 1992 (105), assesses the effect of model biasness during refinement. To compute a free R-factor, R_{free} , a set of test reflections

(5%) are randomly chosen and not used during refinement. This value also tells us how well the observed and experimental structure factor of the test set of reflection agrees. The model was subjected to TLS restrained refinement using REFMAC in the final round of refinement. The final structure had an R-factor of 20.4% and an R_{free} of 25% with good stereochemistry as analyzed by PROCHECK (106).

As the datasets for the HPAP-bound and the serine-bound *Mtb* PGDH were isomorphous with the apoenzyme, this structure (PDB code-1YGY) was refined against the ligand-bound datasets using simulated annealing at 4000 °K in CNS to eliminate any biasness as well as improve the geometry of the model. All the steps to yield the final model were similar to the apoenzyme. The overall statistics for data collection, refinement and geometry statistics are provided in Table 1. The coordinates (PDB code-1YGY, 3DDN, 3DC2) and structure factors of the apoenzyme, HPAP-bound, and serine-bound *Mtb* PGDH have been deposited with the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, NJ. The domain movements were analyzed using the program Dyndom (107) which determines the hinge axes and the residues involved in hinge bending. For this analysis, default settings of Dyndom were used. All the structure figures have been prepared using Chimera (108).

TABLE 1
Data Collection, refinement and geometry statistics of *Mtb* PGDH

	Selenium Peak	Selenium Inflection	Native PGDH	PGDH+HPAP	PGDH+serine
Space group	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22
Unit cell dimensions	a=b=165.3 Å c=217.9 Å α=β=90° γ=120°	a=b=165.4 Å c=218.1 Å α=β=90° γ=120°	a=b=165.5 Å c=218.1 Å α=β=90° γ=120°	a=b=165.6 Å c=218.3 Å α=β=90° γ=120°	a=b=165.2 Å c=218.9 Å α=β=90° γ=120°
Molecules per asymmetric unit	2	2	2	2	2
Wavelength (Å)	0.9794	0.9796	0.9	0.9	0.9
Resolution range (Å)	141.42 - 3.15	141.42 - 3.15	48.8-2.3	48-2.4	46.5-2.7
Highest resolution bin (Å)	3.26 - 3.15	3.26 - 3.15	2.38-2.3	2.49-2.4	2.8-2.7
Observed reflections	777664	802306	1117427	1495125	1011203
Unique reflections	57572	57665	78223	69375	48518
Completeness (%) ^a	100 (100)	100 (100)	99.7 (99.9)	100 (100)	99.2 (99.9)
Average redundancy ^a	13.5 (10.2)	13.9 (11.9)	14.3 (9.8)	22.3 (21.6)	20.8 (19.5)
I/σ(I) ^a	19 (3.6)	18.7(4.3)	24.7 (4.8)	26 (3.1)	19.2 (2.8)
R _{sym} ^{a, b}	0.15 (0.68)	0.15 (0.70)	0.095(0.4)	0.038 (0.30)	0.037 (0.42)
<i>Refinement statistics</i> (REFMAC)					
Free R value (5%)			25	24	26
R value (%) ^c			20.4	20	22
No. of protein residues			1054	1052	1049
No. of water molecules			315	165	77
r.m.s.d. bond length, (Å)			0.012	0.011	0.016
r.m.s.d. bond angles, (°)			1.47	1.42	1.75
<i>Ramachandran plot</i> (PROCHECK)					
Most favored region (%)			805 (88.0)	819 (90)	777 (85.4)
Additional allowed regions (%)			107 (11.8)	89 (9.8)	130 (14.3)
Generously allowed regions (%)			2 (0.2)	2 (0.2)	2 (0.2)
Disallowed regions (%)			0	0	0

^a Values in parenthesis for the highest resolution bin

^b $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections

^c $R = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ where F_{obs} and F_{calc} are the observed and calculated structure factors

Results and discussion

Purification and characterization of Mtb PGDH

Initially, the His-tagged *Mtb* PGDH for structural studies was purified using one-step Ni affinity chromatography. However, this protein did not result in good diffraction quality crystals. As a next logical step, untagged *Mtb* PGDH was constructed. This wild type protein was purified using two-step purification protocol of anion exchange and gel filtration chromatography. The protein was found to be 98% pure from the 4-12% SDS-PAGE gel. Crystallization trials with this wild type protein yielded good quality crystals that diffracted to 2.3 Å.

Fluorescence experiments were carried out to characterize *Mtb* PGDH. When excited at 280 nm, *Mtb* PGDH shows a single emission peak at 335 nm which is due to the presence of aromatic residues in the protein. In contrast to *E. coli* PGDH (73, 109), it does not show a second peak at 440 nm which is characteristics of intrinsically bound NADH. This suggests that NADH is not intrinsically present in *Mtb* PGDH. When 200 µM NADH was mixed with the *Mtb* PGDH enzyme, there was the appearance of the second peak at 440 nm with subsequent decrease in the emission peak at 335 nm. This phenomenon is due to the FRET effect of NADH which absorbs the light at 335 nm and emits at 440 nm. On further addition of 0.5 mM substrate HPAP to this enzyme mixture, there is total disappearance of emission peak at 440 nm with subsequent increase in emission peak at 335 nm which indicates the conversion of NADH in the protein to NAD⁺ (Figure 2.8). As the rate of this reaction is very fast, further kinetic studies of this enzyme have been carried out using the stopped flow fluorescence spectrophotometer.

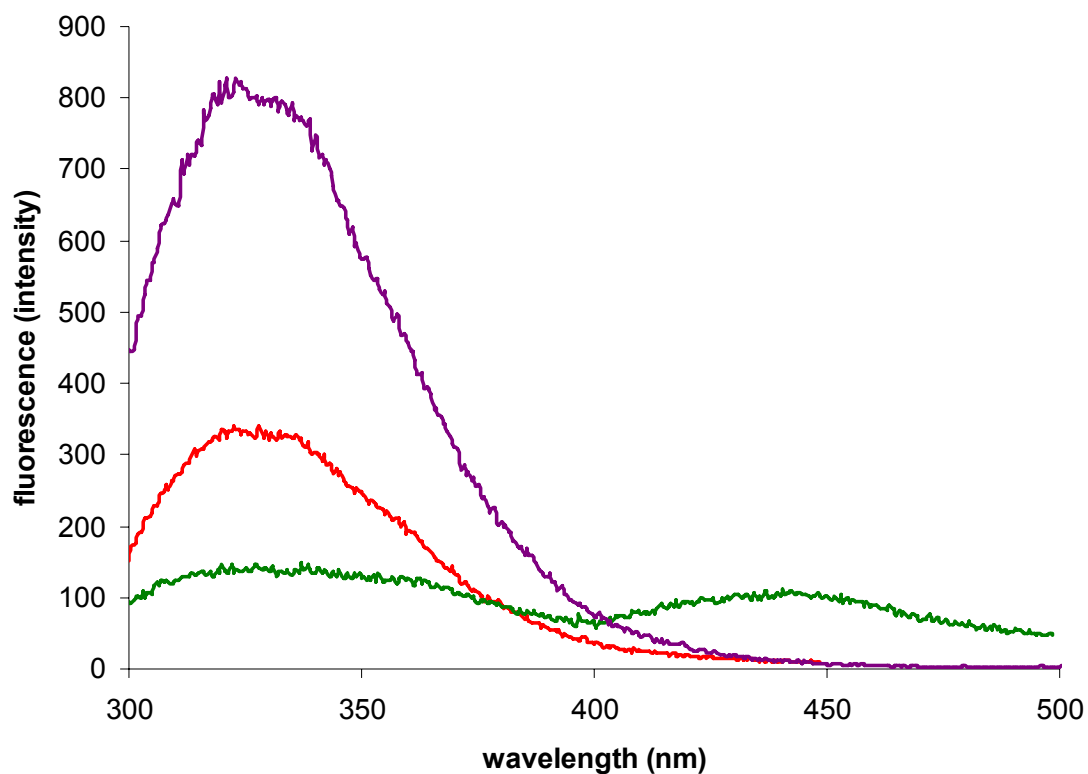


FIGURE 2.8. **Fluorescence spectra of *Mtb* PGDH.** This figure shows the emission spectra of *Mtb* PGDH enzyme alone (violet), after mixing PGDH with 200 μ M NADH (green), and in presence of both 200 μ M NADH and 0.5 mM HPAP (red) in the range of 300-500 nm, when excited at 280 nm.

Structure determination and overall quality of Mtb PGDH apostructure

The 2.3 Å apostructure of *Mtb* PGDH was determined using the selenium MAD phasing (97). To employ this method, the isoleucine residue at position 514 was mutated to methionine since this 529 residue protein lacked any methionine other than the starting Met. This mutation was carried out based on its sequence alignment with *E. coli* PGDH. It is generally recommended that there should be one selenomethionine for approximately 100 amino acids in order to yield sufficient MAD phasing (97). A complete and a highly redundant data set at 3.15 Å resolution provided an interpretable electron density map of high quality from a single selenomethionine in this 529 amino acid protein (2 per asymmetric unit, no NCS averaging). The structure was refined to an Rfactor of 20.4% and R_{free} of 25% with good stereochemistry as analyzed by PROCHECK (Figure 2.9). No outliers were seen in the Ramachandran plot. The refined structure contains all the amino acid residues except the first two N-terminal residues of the protein.

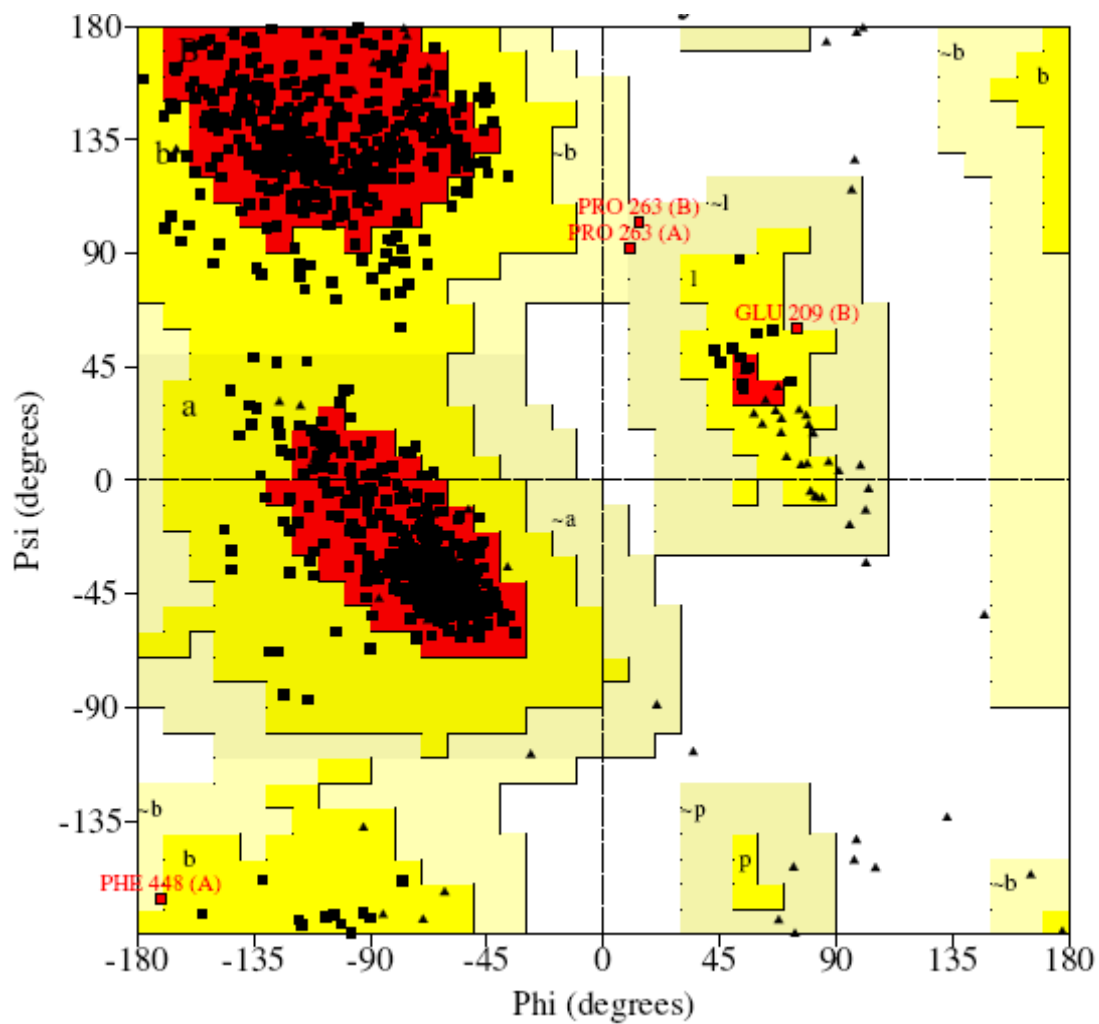


FIGURE 2.9. Ramachandran plot generated by PROCHECK for *Mtb* PGDH apoenzyme (PDB code – 1YGY)

- ▲ Glycine
- Non-glycine and non-proline amino acids

Overall structure of Mtb PGDH

The X-ray crystallographic structure shows that there are two molecules present in the asymmetric unit (chain A and chain B) which form a tetramer with the crystallographic symmetry related molecules (two fold axis of symmetry) (Figure 2.10 B). Superimposition of *Mtb* PGDH chain A on chain B with the nucleotide-binding and substrate-binding domains as reference shows a rotation of the other two domains by approximately 160° as well as flipping of both the same domains. A similar effect is also seen when the other two domains-intervening and regulatory domains are used as a reference. Thus, this demonstrates the presence of two different conformations among the molecules of the asymmetric unit (Figure 2.11).

Each subunit of 55,000 Daltons can be divided into four distinct domains (Figure 2.10 A). These are referred to as the substrate-binding domain (residues 3-98 and 283-319), the nucleotide-binding domain (residues 99-282), the regulatory domain (residues 454-529) and an intervening domain (residues 320-453). The intervening domain between the substrate-binding and regulatory domain is not present in the other two forms of the PGDH, e.g. *E. coli*, *pyrococcus*.

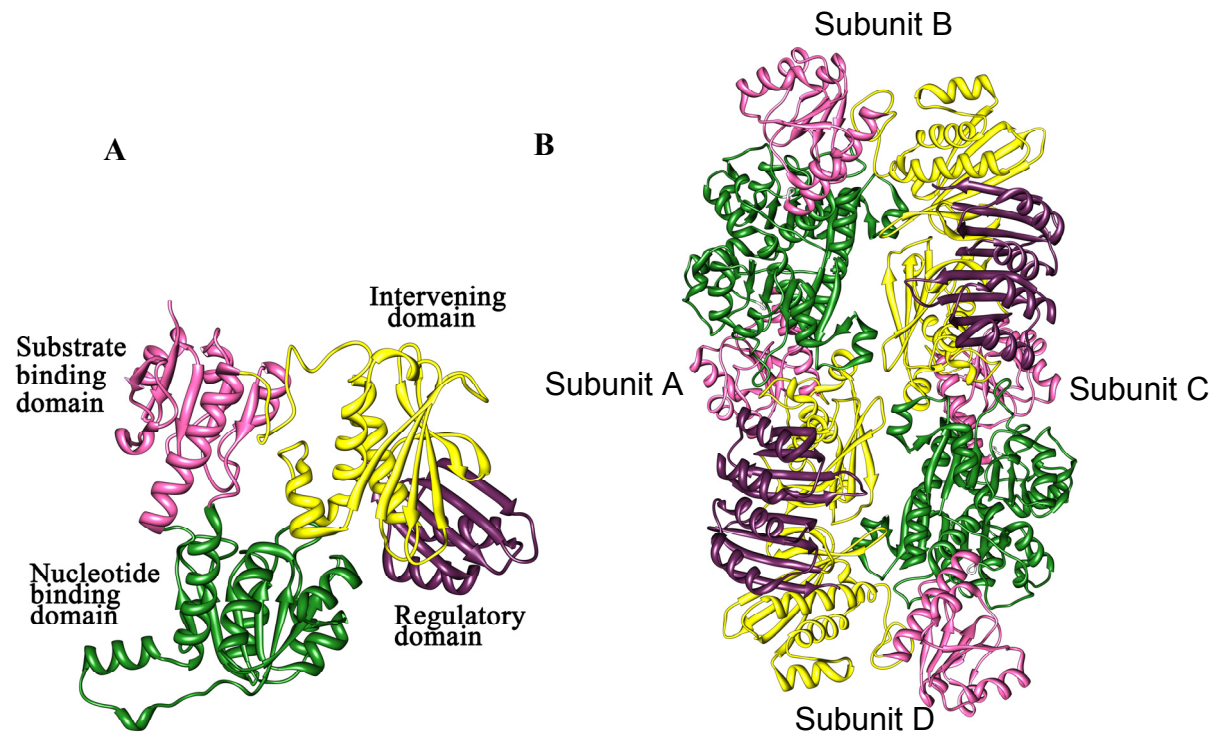


FIGURE 2.10. **Domain arrangements of *Mtb* PGDH.** Ribbon diagram representation of (A) monomeric form (B) tetrameric form of *Mtb* PGDH. The tetramer has three domains in common to *E. coli* PGDH: the nucleotide-binding domain (green), the substrate-binding domain (pink) and the regulatory domain (violet). The fourth additional domain is the intervening domain (yellow).

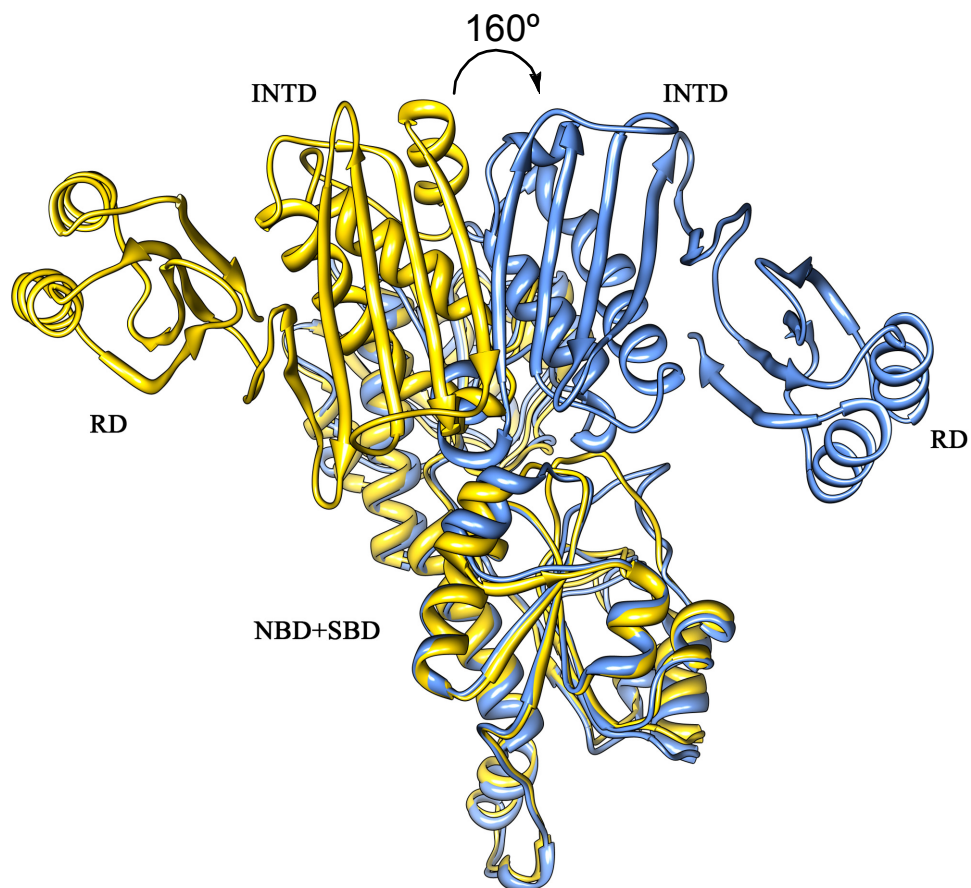


FIGURE 2.11. Ribbon diagram of superimposition of chain A of *Mtb* PGDH on chain B. Figure shows superimposition of nucleotide-binding domain (NBD) and substrate-binding domain (SBD) of chain A (yellow) on chain B (blue) resulting in an flipping and rotation of intervening domain (INTD) and regulatory domain (RD) by 160°, indicating the presence of two different conformations in the asymmetric unit.

Subunit structure in Mtb PGDH

The largest domain, the nucleotide-binding domain, in *Mtb* PGDH consists of residues Asn99 to Gly282 and contains seven parallel β strands (β 8- β 7- β 6- β 9- β 12- β 11- β 10) interconnected with helices (α 6- α 7- α 8- α 9- α 10- α 11- α 12- α 13). This is a variation of the Rossmann fold which is a characteristic of dehydrogenases (110). Sequence alignment of the nucleotide-binding domain in *Mtb* and *E. coli* PGDH shows 33.9% sequence identity and superimposition of these domains in chain A shows a r.m.s.d. of 5.28 Å for 183 C_{α} atoms. This deviation is largely due to the presence of a helix (α 9) between residues Ser176 and Gln182 in *Mtb* PGDH whereas a loop is present in *E. coli* PGDH, as well as, variable loop regions in this domain for both enzymes (Pro173 to Tyr174 in *Mtb* and Ala271 to Asp275 in *E. coli*). However, this helix α 9 in the nucleotide-binding domain is also present in human PGDH (Pro179 - Phe185). Superimposition of the *Mtb* PGDH nucleotide-binding domain of chain B on chain A (r.m.s.d. = 1.5 Å for 183 C_{α} atoms) shows that they have different conformations in the loop region (Leu203 - Ile213) where nucleotide binds.

The substrate-binding domain is geometrically adjacent to both the nucleotide-binding and intervening domains and is composed of residues from two segments of protein sequence (Ser3 - Ser98 and Ala283 - Val319). This amino terminal domain contains five parallel β -strands (β 2- β 1- β 3- β 4- β 5) flanked by five helices (α 1- α 2- α 3- α 4- α 5) like *E. coli* PGDH (r.m.s.d = 3.87 Å for 131 C_{α} atoms). It has only 29.7% sequence identity with *E. coli* PGDH in comparison to 39% sequence identity with human PGDH in both regions 3-98 and 283-319. There is no significant difference in the structure of

the substrate-binding domain in chains A and B of *Mtb* PGDH (r.m.s.d. for 131 C_α atoms is 0.96 Å).

The intervening domain (Val320 - Glu453) is named as such because it intervenes between the substrate-binding and regulatory domains where L-serine binds. This is a new domain not present in the other two forms of PGDH. This two layered sandwich domain consists of six β strands (β14-β13-β15-β16-β17-β18) out of which the first two strands (β14-β13) are flanked by 2 helices (α14-α15). The rest of the four strands form an anti-parallel β sheet. Superimposition of the intervening domains of chain B on chain A in *Mtb* PGDH (r.m.s.d. = 1.8 Å for 133 C_α atoms) does not show any significant secondary structure differences. However, as noted previously, the orientation of the intervening domain in chain B is approximately 160° opposite and flipped in comparison to that of chain A as shown in Figure 2.11. This flipping results in the regulatory domain of chain A occupying a position much closer (9 Å) to its nucleotide-binding domain than in chain B (71 Å). In addition, the reorientation of the domains results in the formation of two different interfaces between the nucleotide-binding and intervening domains in the two unique subunits. Because of this flipping, a flexible loop region (Ala376-Asn386) in the intervening domain of chain B is totally exposed to the solvent, leading to its disordered electron density.

The C-terminal regulatory domain of *Mtb* PGDH consists of residues Gly454 to Ser529 and contains four anti-parallel β strands (β19-β20-β21-β22) interconnected with two helices (α16-α17). This kind of arrangement (βαββαβ) belongs to the ACT (Aspartokinase, Chorismate mutase, and TyrA) domain family (111) which is very

common among enzymes that participate in amino acid metabolism. The interface of these domains also act as a locus for amino acid binding, which is involved in its metabolic regulation (negative feedback regulation) (e.g. ATP phosphoribosyl transferase of *Mtb* (112) involved in histidine biosynthesis, threonine deaminase (113) involved in isoleucine biosynthesis, α -isopropylmalate synthase (114) involved in leucine biosynthesis, and prephenate dehydratase (115) involved in phenylalanine biosynthesis). Thus, it has also been called the "regulatory domain in amino acid metabolism". This domain has also been the site for small molecules in proteins such as thiamine binding proteins involved in thiamine transport (116) and NikR protein involved in regulating the Ni level in the cells (117). Interestingly, this type of secondary structure arrangement is also seen in other proteins such as nucleoside diphosphate kinase and aspartate carbamoyltransferase involved in pyrimidine biosynthesis (118). These enzymes are not directly involved in amino acid metabolism and lack an allosteric site at the ACT domain interface. Superimposition of the regulatory domain of *Mtb* and *E. coli* PGDH (r.m.s.d. = 3.21 Å for 75 C_α atoms) shows a shift (4.8 Å) in one of the helices in *Mtb* PGDH (Pro507 - Val518) due to longer loop regions in *Mtb* PGDH (Glu489 - Ala496 and Leu502 - Val506). Kinetic analysis on serine inhibition of *Mtb* PGDH has shown that serine exhibits a mixed type of noncompetitive inhibition (19), so it is anticipated that it will bind at a site other than the active site. In *E. coli* PGDH, serine-binding site exists at the interface of the two regulatory domains. In *Mtb* PGDH apoenzyme, changes in the orientation of the side chains of residues Asp463 and Arg464 at this site in chains A and B have been noticed when both the adjacent regulatory

domains are superimposed (r.m.s.d. = 0.8 Å for 75 C_α atoms). These changes in orientation of Asp463 and Arg464 side chains at the regulatory domain interface might indicate a path for serine entry. In *E. coli* PGDH tetramer, only two molecules of serine, one at each interface, are required for optimal inhibition (65). If the same is true for *Mtb* PGDH, this orientation change at the interface may delineate two sites that will bind serine easily from the other two that will not. However, an answer to the question of what is triggering this change in orientation in one subunit of an apoenzyme, but not in the other, is not clear.

Quaternary structure-Mtb PGDH as a tetramer

The overall arrangement of subunits in this tetramer is very unusual. The presence of the extra 130 amino acids polypeptide (the intervening domain) causes the movement of the entire regulatory domain towards the nucleotide-binding domain forming a curved structure in each subunit as seen in figure 2.8 B in comparison to the elongated form seen in *E. coli* PGDH (figure 2.4). The asymmetry in the tetramer, due to the shifting of this regulatory domain interface from the extreme ends, as in an elongated *E. coli* PGDH tetramer, to the sides of a more globular *Mtb* PGDH tetramer, is unique.

There are two major points of interaction between the adjacent subunits in the *Mtb* tetramer. One, which is very extensive, is present between the nucleotide-binding domains of adjacent subunits, with a buried surface area of 2402 Å². This interface is also preserved in the homologous enzymes like D-glycerate dehydrogenase (3), formate dehydrogenase (67) as well as *E. coli* (8) and human PGDH (PDB code-2G76). The other interface is between the regulatory domains from adjacent subunits, with buried

surface area of 1294 \AA^2 . This interaction, responsible for tetramer formation, is also found in *E. coli* PGDH (74). In this interface, two adjacent regulatory domains create a sheet of eight antiparallel β strands. L-serine binds at this interface, forming a hydrogen bond network between the two domains. PGDH from *E. histolytica* (22) as well as *P. horikoshii* (PDB code-1WWK) exist as a dimer since they both lack this regulatory domain.

Apart from the above mentioned interactions which are also seen in other forms of PGDH, there are certain unique interactions observed only in *Mtb* PGDH. They consist of: (1) Interaction between the nucleotide-binding domain of subunit B with the intervening domain of subunit C and by symmetry, the nucleotide-binding domain of D and the intervening domain of A. (2) Interaction between the nucleotide-binding domain of subunit A with the intervening domain of subunit C, and, by symmetry the nucleotide-binding domain of C and the intervening domain of A. (3) Interaction between the intervening domains of subunits A and D (and B and C). It is worth mentioning here that the helix $\alpha 9$ (Ser176-Gln182) of the nucleotide-binding domain, which is the unique feature of this form of PGDH is involved in both the above kind of interactions (Table2). Intersubunit interactions were also observed in *E. coli* PGDH tetramer, but between the same domains i.e. between Asn190 residue of opposite nucleotide-binding domains (A&C and B&D).

The flipping of the two unique subunits at the junction between the substrate-binding and intervening domains results in two different and unique sets of intra subunit interactions between the nucleotide-binding and intervening domains. These are listed in

Table 3 and consist of the unique interactions between the nucleotide-binding and intervening domains of subunit B (and D) and the nucleotide-binding and intervening domains of subunit A (and C).

A diagram comparing the subunit and domain arrangement of *E. coli* and *Mtb* PGDH is shown in Figure 2.12.

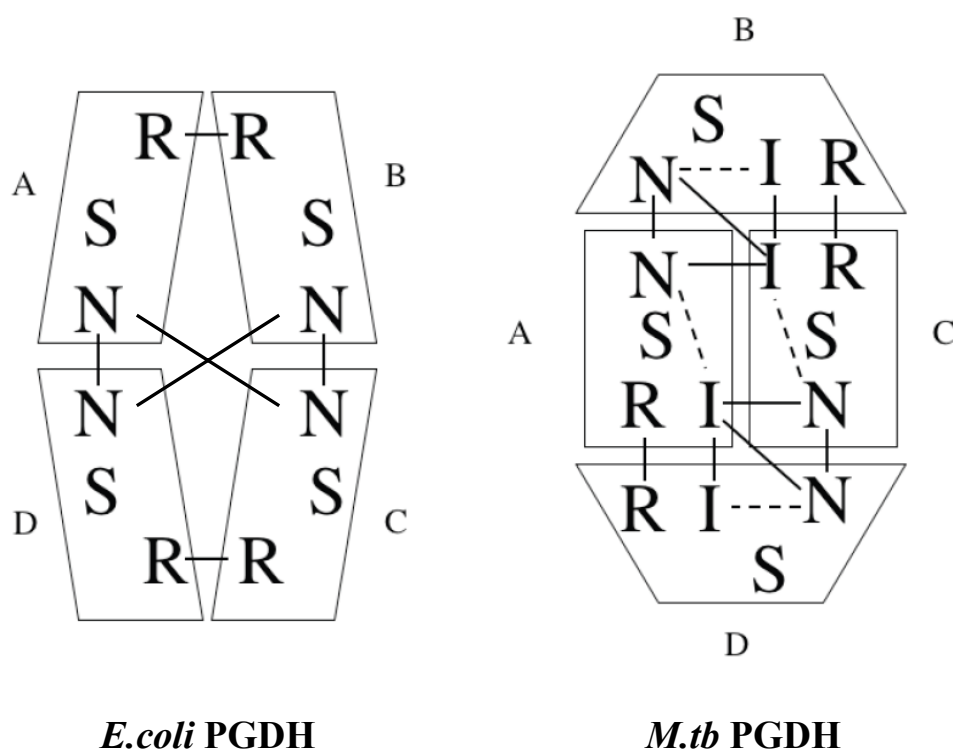


FIGURE 2.12. Comparison of the subunit and domain interaction in *E. coli* and *Mtb* PGDH. Subunits are depicted as geometric representation. Parallelograms are used for the similar conformations in *E. coli* PGDH, and rectangles and trapezoids for the two different conformations in *Mtb* PGDH. The subunits are labeled A-D. N is nucleotide-binding domain, S is substrate-binding domain, I is intervening domain and R is regulatory domain. The solid lines depicts the inter subunit interactions and dotted lines depict the intra subunit interactions that are not continuous in the polypeptide chain.

TABLE 2
Inter subunit domain contacts

Nucleotide-binding domain of B (D)	Intervening domain of C (A)	Distance (Å)
Gln 159 NE2	Glu 405 OE1	2.66
Gln 182 NE2	Arg 410 O	2.55
Gln 182 O	Glu 405 N	3.12
Gln 182 OE1	Ser 406 N	2.99

Nucleotide-binding domain of A (C)	Intervening domain of C (A)	Distance (Å)
Gln 182 NE2	Ser 348 OG	3.05

Intervening domain of A (B)	Intervening domain of D (C)	Distance (Å)
Asn 408 OD1	His 409 ND1	3.22
His 409 ND1	Asn 408 OD1	2.72
Arg 410 N	Asn 408 O	3.33
Arg 410 NH2	Ser 406 O	2.8

TABLE 3
Intra subunit domain contacts

Nucleotide-binding domain of B (D)	Intervening domain of B (D)	Distance (Å)
Arg 179 NH1	Glu 323 OE1	2.95
Arg 179 NH2	Glu 323 OE2	3.25
Arg 179 NE	Glu 356 OE2	2.83
Gln 182 OE1	Arg 410 NE	3.3
Lys 206 NZ	Glu 360 OE2	2.97

Nucleotide-binding domain of A (C)	Intervening domain of A (C)	Distance (Å)
Arg 152 N	Glu 393 OE1	2.87
Gln 155 NE2	Glu 393 O	2.82
Asp 172 OD1	Arg 394 NH1	3.28
Asp 172 OD2	Arg 394 NH1	3.1
Ser 176 N	Arg 394 O	3.1
Arg 179 NH2	Ala 392 O	3.36

Analysis of hinge regions in Mtb PGDH

In *Mtb* PGDH, the nucleotide-binding and substrate-binding domains are linked to the intervening and regulatory domains by a long loop (316-322) which is responsible for the 160° rotation between the domains, resulting in tetramer formation. Interestingly, the point of transition for this rotation is occupied by three consecutive glycine residues (316-318). In *E. coli* PGDH, glycine residues are also found between domains which act as hinge regions and play a role in the cooperativity of serine inhibition. These are Gly294-Gly295, found between the nucleotide-binding and substrate-binding domains and Gly336-Gly337, present between the substrate-binding and regulatory domains. Thus, these hinge regions are present at both ends of a long helix connecting the catalytic site to the regulatory site. Mutation studies on *E. coli* Gly294-Gly295 resulted in reduced enzyme activity and mutation of Gly294 alone was found to be important in the cooperativity of serine inhibition (119). Also, mutations in Gly336-Gly337 (G336, 337V) have resulted in weak serine binding and less sensitivity of the enzyme for serine (120). In comparison, in *Mtb* PGDH, Gly282-Ala283 is present in the loop region between the nucleotide-binding and substrate-binding domains. There is only a short loop of three residues present between the regulatory and intervening domains. This loop is devoid of any Gly or Ala residues. This indicates that there is less flexibility in this region. This arrangement might indicate that the intervening and regulatory domains act together as an ensemble for allosteric inhibition. A long helix (α 14) in *Mtb* PGDH, with Gly282-Ala283 on one end and a long loop containing Gly316-Gly318 on the other end, connecting the catalytic site to this

ensemble, is present. Similar mutant studies have been carried out to investigate the role of these three consecutive glycine residues in *Mtb* PGDH. Single and double mutants of the glycine residues 316-318 had no appreciable effect on serine inhibition. No protein expression was found in triple mutant (G316, 317, 318V) of this region (78). Thus, these findings may indicate that these glycine residues are not involved in the process of serine inhibition but, rather, maintain the overall tetrameric structure of this protein.

Analysis of the active site of Mtb PGDH

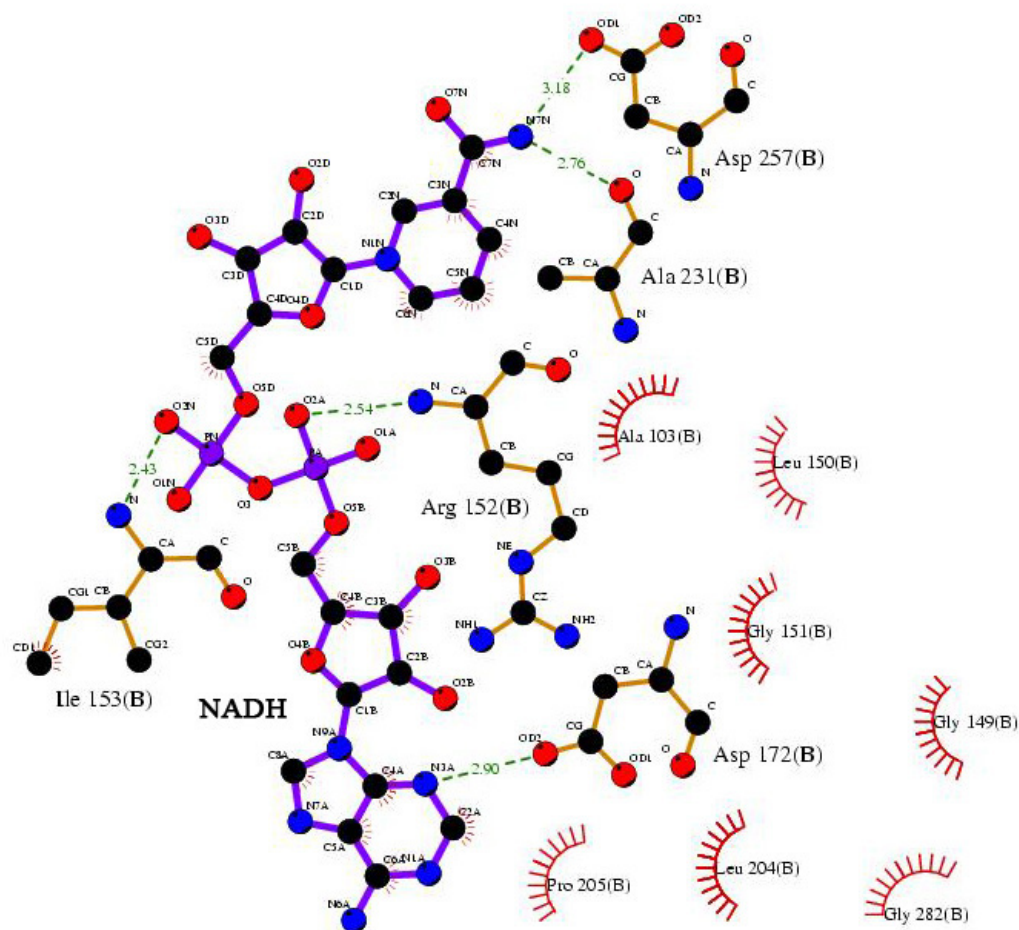
The *Mtb* PGDH catalytic site is present in a cleft between the nucleotide and substrate binding domains. Each PGDH subunit has its own distinct active site and these sites are ~ 32 Å apart at the nucleotide-binding domain interface. The site is mostly lined by the loop regions Arg51-Val56, Ala73-Val80, Lys131-Ser136, Val229-Leu236, Pro279-Gly293 and Ile374-Phe384. The side chains of the basic residues in these loop regions are predominantly oriented in the active site for binding the negatively charged substrates (PGA and HPAP) of this enzyme. A loop region from the adjacent subunit (Lys131'-Ser136') also contributes to its formation. This region is dominated by polar residues like Lys and Ser and contains the Arg132' required for anchoring the substrate in the active site. The loop also contains the conserved Trp residue (Trp130) that maintains the architecture of the active site as well as the nucleotide-binding domain interface of the tetramer. This residue supports the hydrophobic pocket containing the catalytic dyad Glu262-OE1 and His280-ND1, characteristics of the dehydrogenases. In *E. coli* PGDH, the residues in this loop region (Asn140-Gly145) are dominated by Ala and Gly and

presumably provide flexibility to this region. Mutagenesis studies on these residues in *E. coli* suggest that this region is important for the cooperativity of serine inhibition of catalytic activity (121). Interestingly, in human PGDH, which is not inhibited by L-serine, this region is occupied by bulky side chain residues such as Lys, Phe and Met, suggesting less flexibility in this region.

Cofactor-binding site

Since *Mtb* PGDH requires NAD^+/NADH as a cofactor for its catalysis, it contains the Rossmann fold motif consisting of a $\text{Gly}^{149}\text{-X}^{150}\text{-Gly}^{151}\text{-X}^{152}\text{-X}^{153}\text{-Gly}^{154}\text{-X}_{\text{aa}17}\text{-Asp}^{172}$ sequence (110), which delineates the cofactor binding site. Interestingly, the PGDH crystals, which were formed in the presence of NADH/NAD^+ or its analog, 3-acetyl pyridine dinucleotide, do not show the electron density for the cofactor in any of these structures. However, NAD^+/NADH can be modeled into the *Mtb* PGDH active site by superimposing the nucleotide-binding domain of either *E. coli* or human PGDH onto the *Mtb* PGDH structure as the residues interacting with the cofactor are conserved. Based on this model, OD1 and OD2 atoms of Asp172 residue in the Rossmann fold should form hydrogen bonds with the O2 and O3 atoms of the ribose sugar of adenosine moiety. The adenine ring of the cofactor is located in a hydrophobic pocket comprised of Pro173 and 205, Val148, Tyr171, and Leu204 and 213 residues. The two phosphate atoms connecting the nicotinamide part with the adenine part of NADH partly interact with the main chain amides of Arg152 and Ile153. The multiple Gly residues in the Rossmann fold prevent any spatial interference when the cofactor binds. The N7 atom of the nicotinamide ring forms hydrogen bonds with Asp257-OD1 as well as the main

chain oxygen of Ala231 (Figure 2.13).



Key

- | | |
|------------------------------|---|
| Ligand bond | His 53 Non-ligand residues involved in hydrophobic contact(s) |
| Non-ligand bond | Corresponding atoms involved in hydrophobic contact(s) |
| Hydrogen bond and its length | |

FIGURE 2.13. **NADH interaction in *Mtb* PGDH chain B.** The figure shows the possible interactions of protein residues with NADH modeled at the Rossmann fold of chain B. This figure is generated by Ligplot (122).

In *Mtb* PGDH, the cofactor binding site exists in two different environments in contrast to all the known PGDH structures (*E. coli*, human, and *Pyrococcus*). This is possibly due to the presence of two different domain orientations in this enzyme. In chain A, the site is covered by the intervening domain due to its close vicinity to the nucleotide-binding domain, thus forming a channel (Figure 2.14 A). It appears that either the cofactor has to enter through this channel or the intervening domain has to change its conformation to further expose the binding site. In this regard, the channel opening appears to be large enough to allow the cofactor to enter (8 Å x 25 Å). In chain B, due to the change in the orientation of the intervening and regulatory domains, the cofactor-binding site is present in an open cleft. However, in chain B of all *Mtb* PGDH structures, the cofactor-binding site appears to be blocked by the long flexible loop (Leu203 - Ile213) present between strand β 8 and helix α 11 due to the proximity of the neighboring symmetry related molecule. However, movement of this loop to an orientation similar to chain A would allow the cofactor to bind at this site as shown in the figure 2.14 B. Indeed, the presence of these two different environments for NADH binding has also been confirmed by stopped flow fluorescence experiments (78) which shows that there are two sets of NADH binding sites.

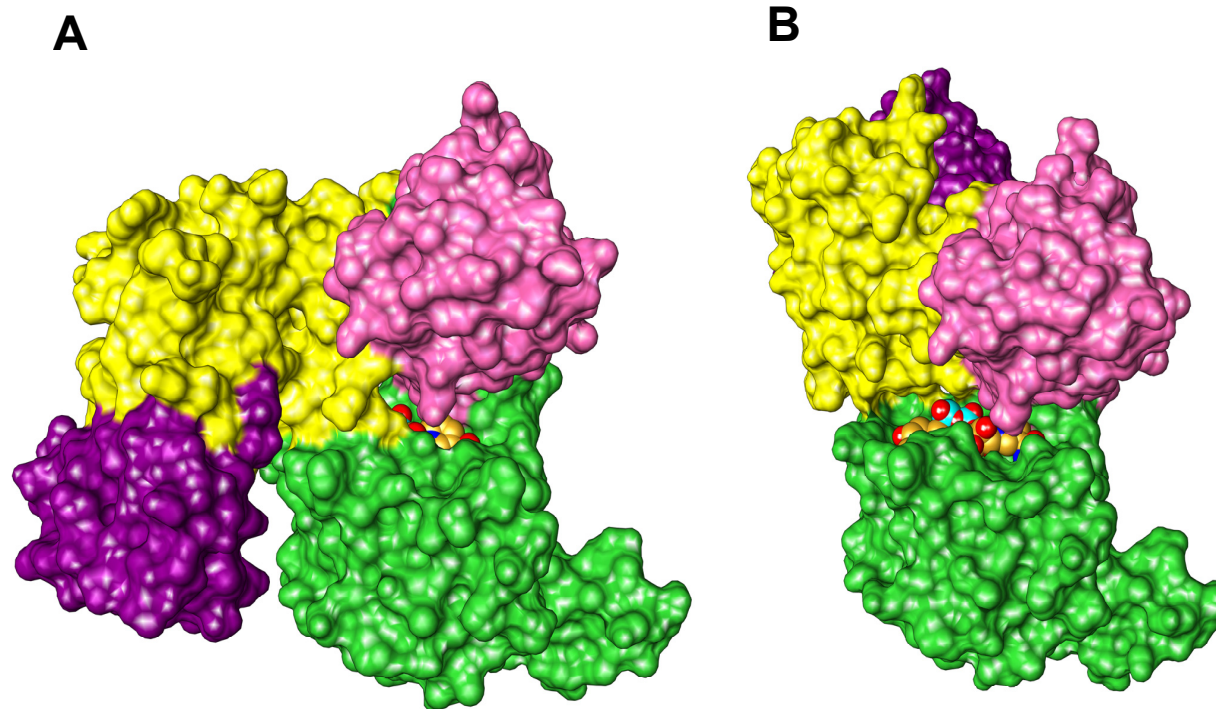


FIGURE 2.14. **Differences in cofactor binding site.** Molecular surface representation of *Mtb* PGDH structure showing two different cofactor-binding sites. In both the chains, the four domains are represented in different colors: the nucleotide-binding domain (green), the substrate-binding domain (pink), the intervening domain (yellow) and the regulatory domain (violet). NADH is represented in orange CPK form. (A) In chain A, the cofactor-binding site resides in a channel which is formed by the intervening (yellow) and nucleotide binding domains (green). (B) In chain B, the loop, which blocks the cofactor-binding site in these structures, has been placed in an orientation similar to chain A. As seen in the figure, this site is now totally exposed to the solvent, as the intervening domain is oriented in the other direction ($\sim 160^\circ$ rotation) in this chain. NADH has been modeled in both the figures based on *E. coli* PGDH structure.

Previously, tartrate present in the crystallization condition was thought to force the structure into a stable form which is not capable of binding cofactor. However, crystals obtained from different crystallization conditions that lacked tartrate had the same space group and cell dimensions as ones that contained tartrate but diffracted weakly. Though these datasets generated poor maps, they were also devoid of the distinct electron density for the cofactor in their structures. Thus, these two different binding environments for NADH do not appear to be a crystallographic artifact due to the presence of tartrate. However, it appears that the high salt concentrations (1 M NaK tartrate) present in the crystallization condition or 100 mM potassium phosphate buffer used as the protein storage buffer might be one of the reasons for the absence of cofactor in these structures. This fact is supported by the analysis of stopped flow fluorescence experiments which showed weak NADH binding in the presence of high ionic strength buffer (78).

Mode of substrate recognition and enzyme catalysis

In *Mtb* PGDH, the reverse reaction of reduction of HPAP to PGA is the favored reaction. The K_m value for the substrate HPAP is $\sim 85 \mu\text{M}$ and a k_{cat}/K_m of $\sim 5.6 \times 10^6 \text{ M}^{-1}\text{sec}^{-1}$ (78). In order to understand the mode of enzyme catalysis in *Mtb*, structure determination of ternary complex of PGDH with HPAP and NAD analog was carried out. The 2.5 Å HPAP bound *Mtb* PGDH structure revealed the residues interacting with the substrate at the active site, although the electron density for HPAP is observed only at the chain B active site. This structure lacks the presence of the cofactor. The active site cleft remains open on substrate binding, thus indicating a nonproductive form of PGDH structure. In this crystal structure, the phosphate group of HPAP forms hydrogen bonds with Arg51-NE (3.5 Å) and Gln289-OE1 (2.6 Å) from one subunit, and Arg132'-NH1 from an adjacent subunit (2.8 Å). The carboxyl group of the substrate is hydrogen bonded to Asn99-ND2 (3.16 Å) and the keto group of HPAP interacts with the main chain amide of Gly76 (3.08 Å) (Figure 2.15A). On comparing this binding of HPAP to the binding of substrate analog malate in the truncated human PGDH structure, HPAP does not seem to be properly oriented, with the keto group pointing away from the catalytic His280 (4 Å). Also, we did not see the expected interaction of the substrate C1-carboxyl group with an arginine residue which is seen in the structure of human PGDH (Arg235) and many other dehydrogenases (67). Arg233 is a conserved residue in *Mtb* PGDH that corresponds to Arg235 in human PGDH, and Arg240 in *E. coli* PGDH. In *Mtb* PGDH, the side chain of Arg233 is facing away from the active site. In fact, in chain A, Arg233 side chain forms a salt bridge with Glu379 whereas in chain B it appears to be flexible, resulting in reduced

density for the side chain. Thus, in the absence of a cofactor, HPAP is not properly oriented for the enzyme to carry out catalysis in this structure. Based on the model of cofactor-bound PGDH, side chain of Asn99 has to move out of the way for the nicotinamide part of the cofactor to bind. It may be possible that the presence of the cofactor plays a key role in the proper positioning of the substrate for hydride transfer from His280 by stabilizing the formation of a salt bridge between Arg233 and the carboxyl group of the HPAP. Indeed, stopped flow fluorescence experiments (78) with *Mtb* PGDH shows that the substrate HPAP binds first and then the cofactor NADH, during the enzyme reaction at the active site. This is in contrast to *E. coli* PGDH, in which NADH is intrinsically bound as evident from the fluorescence spectra of purified *E. coli* PGDH (73).

In the ternary complex of substrate analog malate and NAD⁺-bound truncated human PGDH structure, the active site cleft exists in a closed conformation. Superimposing the *Mtb* PGDH nucleotide-binding and substrate-binding domains on the same domains of malate bound human PGDH reveals a 15° rotation of the substrate-binding domain towards the nucleotide-binding domain upon ligand binding at the active site. This domain motion is analyzed by Dyndom algorithm (107). Even with an open cleft, the HPAP in *Mtb* PGDH structure seems to occupy a site similar to that of malate in the human PGDH structure, suggesting that the binding of both cofactor and substrate is needed for the active site to close and for productive contacts to be made with HPAP. Based on this superimposition, a model for the closed conformation of the active site cleft in *Mtb* PGDH has been prepared to elucidate the mechanism of enzyme action (Figure

2.15B). Interestingly, all three of the catalytically important residues (His280, Glu262 and Arg233) are present on different loop regions of the nucleotide-binding domain. His280 is present on the Thr278-Ser284 loop connecting the nucleotide-binding domain to the substrate-binding domain, Glu262 is present on the loop between β 10 and α 13, and Arg233 is present on the loop between β 9 and α 12, near the nicotinamide ring, at the entrance of the active site cleft. The rest of the cofactor lies at the base of the active site cleft. In the initial step of PGDH catalysis, HPAP will bind at the active site first and then on NADH binding, the side chain of Arg233 will orient toward the HPAP-binding site. As a result, the C1-carboxyl group of HPAP will form a salt bridge with the guanidinium group of Arg233 and the C2-keto group will point towards the His280 on closure of the active site cleft. The interaction of His280-ND1 with Asp262-OE1 properly positions the imidazole ring for proton transfer. The generation of a partial positive charge on C2 of HPAP will allow for the transfer of hydride ion from NADH to the substrate resulting in the formation of NAD^+ and phosphoglycerate.

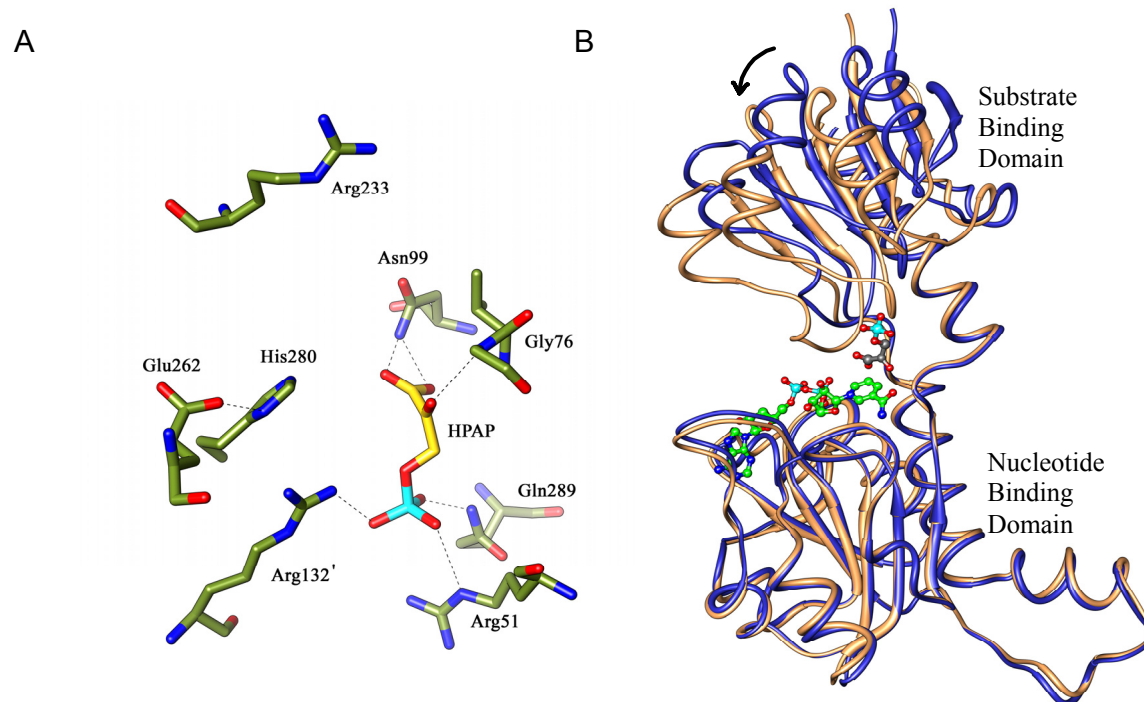


FIGURE 2.15. **Catalytic event at the *Mtb* PGDH active site:** (A) HPAP binding at the PGDH active site. HPAP is shown in stick form (yellow) and shows all the residues interacting with HPAP at the active site. (B) Ribbon diagram representation of the superimposition of the *Mtb* apoenzyme (blue) on ligand bound *Mtb* PGDH model (gold) based on malate and NAD bound human PGDH. NADH (green) and HPAP (gray) are shown in ball and stick form. As seen in the figure, there is rotation of approximately 15° of the substrate-binding domain (arrow) in response to substrate and cofactor binding. The intervening domain and the regulatory domain have not been shown for the sake of clarity.

Role of intervening domain as an anion-binding site

To further explore the role of the intervening domain in *Mtb* PGDH, VAST (123) and DALI (124, 125) structural alignment searches, using the coordinates of this domain, were carried out. Interestingly, these search algorithms found structural similarity within a region of domain II of 2-methylcitrate dehydratase from *E. coli* (VAST score = 8.1, PDB code-1SZQ; r.m.s.d of 2.8 Å for 83 C_α atoms) (Figure 2.16). This enzyme is involved in propionate catabolism and catalyzes the conversion of 2- methylcitrate to 2 methyl cis-aconitate. Unfortunately, similar to the intervening domain, the role of domain II is also unknown. Apart from numerous other candidates with Z score of 2 - 4.5, the DALI search also found a structural similarity with chorismate mutase in *B. subtilis* (Z score = 5, PDB code-2CHS; r.m.s.d. of 3.8 Å for 83 C_α atoms). Chorismate mutase is a homotrimer with the monomer having a single domain that catalyzes the conversion of chorismate to prephenate in aromatic amino acid biosynthesis (126). There was also some structural similarity with urocanate hydratase from *Pseudomonas putida* (VAST score = 7.8, PDB code-1UWK; r.m.s.d. of 2.3 Å for 52 C_α atoms) (127) which is involved in the histidine degradation pathway. Interestingly, the structural similarity is seen in a region that is part of the NAD⁺ binding domain in this enzyme, but no such NAD⁺ binding fingerprint region is present in the intervening domain of the *Mtb* PGDH. This homology search, unfortunately, did not provide us with any clues about the role of the intervening domain.

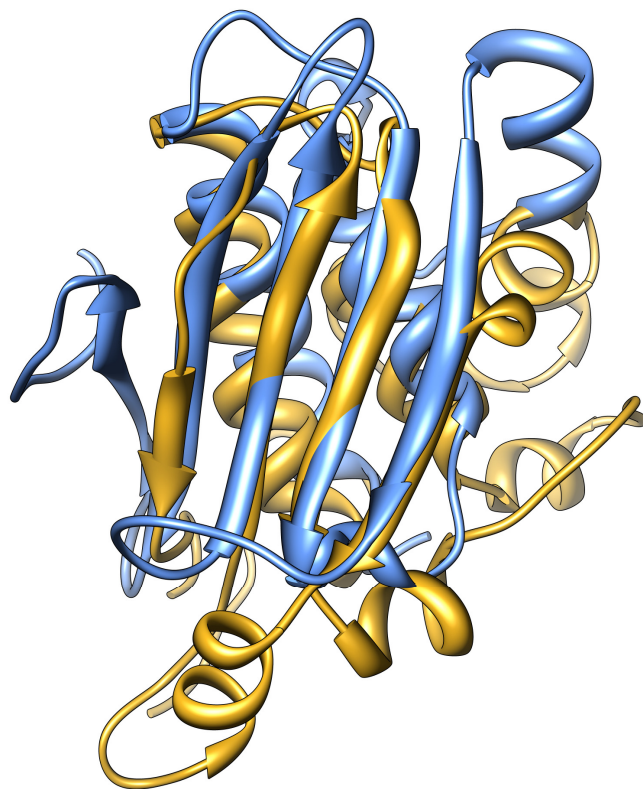


FIGURE 2.16. **Overlay of the intervening domain of *Mtb* PGDH on domain II of 2-methylcitrate dehydratase from *E. coli* (PDB code-1SZQ).** The figure shows a ribbon diagram representation of the superimposition of the intervening domain of *Mtb* PGDH (blue) and the domain II of methylcitrate dehydratase (gold). As seen in the figure, the four antiparallel β strands in the front and the two helices at the back align well in both the structures with an r.m.s.d. of 2.8 Å.

Intriguingly, in the *Mtb* PGDH structures, there is some extra electron density visible in the region where adjacent intervening domains converge. This electron density is surrounded by the positively charged residues His447, Lys439', Arg451', and Arg501' (Figure 2.17). This density can be easily fitted with a tartrate, which is part of the

crystallization solution. The tartrate is within hydrogen bonding distance of His447-ND1 from one subunit (2.5 Å) and Lys439'-NZ (3.5 Å), Arg451'-NH1 & -NH2 (2.98 Å), and Arg501'(3.5 Å) from the adjacent subunit. Arg501' is contributed by the adjacent regulatory domain and forms an ion-pair with the C4 carboxyl group of tartrate. Two molecules of tartrate bind to this interface due to the symmetry of the structure and appear to form an inter-subunit bonding network. In *Mtb*, tartrate does not seem to be involved in any metabolic process so it is not the natural ligand, binding to PGDH. Based on its structural similarity with tartrate, we can infer that this could be another binding site for the substrate HPAP, responsible for uncompetitive inhibition of this PGDH. Indeed, mutagenesis studies on the residues in this region (Arg446Ala, His447Ala, Lys439'Ala, Arg451'Ala, and Arg501'Ala) have resulted in reduced substrate inhibition in *Mtb* PGDH (128). Triple mutation (Lys439', Arg451', and Arg501') of residues from one subunit shows complete elimination of the substrate inhibition. This theory is further supported by the fact that the PGDH in rat liver, which has an intervening domain, also shows substrate inhibition (29), but HPAP does not inhibit *E. coli* PGDH, which lacks this domain. Thus, the intervening domain in *Mtb* PGDH may be required for additional regulation of serine biosynthetic pathway during the chronic infection of the host. This site can also act as a template for the development of inhibitors in this enzyme.

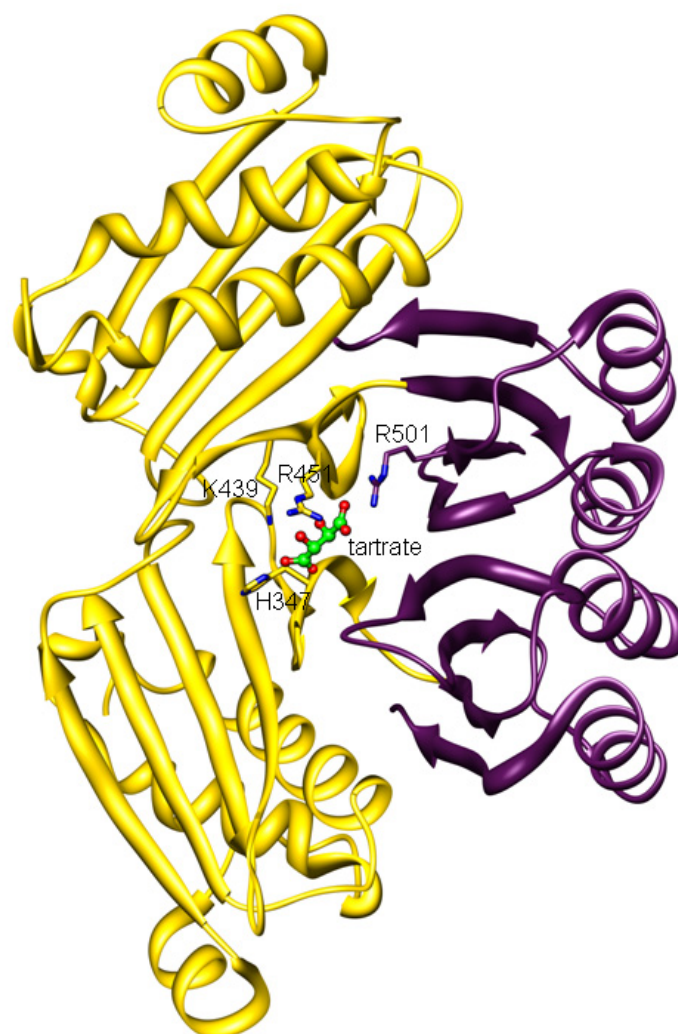


FIGURE 2.17. Anion binding site at the intervening domain interface of *Mtb* PGDH.

The figure shows the interface of the regulatory domain (purple) and the intervening domain (yellow). In this figure, tartrate is shown in ball and stick form (green) interacting with the positively charged residues from both the domains. There are two tartrate bound at the interface due to two fold axis. In this figure, only one tartrate interaction is shown for the clarity of the figure.

Mode of serine binding

L-serine inhibits *Mtb* PGDH with an IC_{50} of $\sim 30 \mu\text{M}$ (64). This allosteric regulation of *Mtb* PGDH shows cooperativity with a Hill coefficient near 2.0 (78). Analysis of the 2.7 Å serine bound *Mtb* PGDH structure shows that L-serine binds at the interface of the regulatory domains, forming hydrogen bonds to both domains in a manner analogous to that seen in *E. coli* PGDH (9). Due to non-crystallographic two-fold axis, there are two serine molecules binding 17 Å apart at the regulatory domain interface. This allosteric site is 35 Å away from the active site. When the serine-bound PGDH structure is compared to the HPAP-bound structure or the apostructure, no major conformational changes or large domain movements are observed.

Even though this domain shares the $\beta\alpha\beta\beta\alpha\beta$ secondary structure with its *E. coli* counterpart, the sequence identity is only 30% and the residues that interact with serine are not well conserved. In the inhibited structure of *E. coli* PGDH, residues His344, Asn346, and Asn364' interact with L-serine at the regulatory domain interface (9). In the serine-bound *Mtb* PGDH structure, the L-serine carboxyl group forms a hydrogen bond with the hydroxy group of Tyr461 (2.7 Å) and Asp463-OD1 (2.5 Å). The amino group of serine hydrogen bonds with Asn481'-OD1 (2.9 Å) from the adjacent subunit. No water molecules are present in the regulatory site for bond formation with the serine side chain hydroxyl group as seen in *E. coli* PGDH, indicating that the site is more compact. Instead, the serine hydroxyl group makes direct hydrogen bond contact with the main-chain amide of Leu468. Mutation of all the residues that interact with L-serine (Y461A, D463A, and N481A) has resulted in reduced serine inhibition with an increase

in IC₅₀ values for serine.

There are significant local changes in the orientations of the side chains of the residues at this allosteric site. The side chains of Asp463 and Arg464, present on the loop region at the allosteric site, flip their positions when serine binds in the regulatory domain (Figure 2.18). This flipping is necessary for hydrogen bond formation between the effector molecule and Asp463 side chain. Even though all the residues in this loop region (Val462-Gly466) are quite similar to *E. coli* PGDH, this flip is not seen in the *E. coli* allosteric site. In *E. coli* PGDH, there is a movement (4 Å) of this loop region near the proline residue (Pro348) to cover the allosteric site in response to serine binding. In *Mtb* PGDH, this change occurs at the flipped region to cover the serine-binding site. Because of this flipping, considerable movement in the neighboring residues (His460-Val462 and Pro465 - Gly466) is also observed. A surface representation of the regulatory domain shows that the path to the binding site, at the regulatory domain interface, is closed upon serine binding (Figure 2.19).

The orientation of the side chain of Asn481' from the adjacent subunit also changes, in addition to all the aforementioned changes. When serine binds, the side chain orients to face Asp463, forming a hydrogen bond with the serine amino group (2.9 Å). These interactions further strengthen the regulatory domain interface. The mutation of Asn481' residue (N481A) from the adjacent subunit has resulted in loss of cooperativity of serine binding with a decrease in the Hill coefficient to 1.1 (78). In serine unbound PGDH, Asn481'-ND2 hydrogen bonds with Asp503-OD2 on the neighboring loop of the same chain (~3.4 Å). There is also a change in the orientation

of the side chain of His460-NE2 which forms hydrogen bonds with Asp490-OD1 (3.1 Å) and OD2 (3.4 Å) when serine binds. In response to this change, the orientation of the phenol side chain of Tyr521 also shifts 90°. There is also variation in the orientation of Glu489 which, in the absence of serine, is hydrogen bonded to Arg464 but faces toward the anion-binding site when serine binds. Interestingly, all these residues, which undergo changes on serine binding, are present on the loop regions adjacent to the allosteric site.

Different members of the ACT domain family show different mechanism for allosteric regulation, e.g. *E. coli* aspartokinaseIII changes its oligomeric state from a dimer to a tetramer in response to the lysine binding at the allosteric site (129). Also, in *E. coli* PGDH, serine binding prevents the movement of the nucleotide-binding domain with respect to the other two domains (74). In comparison, the changes seen in *Mtb* PGDH in response to serine binding seem to be quite different.

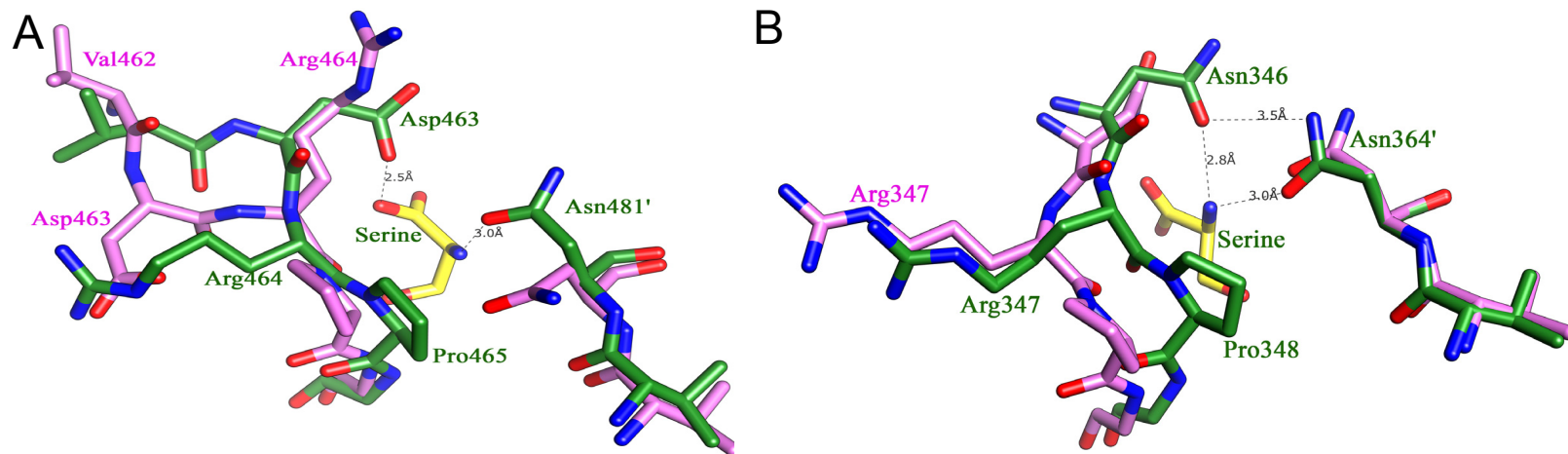


FIGURE 2.18. Changes occurring at the loop region of allosteric site on serine binding. (A) In *Mtb* PGDH, superimposition of the regulatory domain of chain A of the apoenzyme (purple) and the serine bound form (green) shows a flip in the side chains of Asp463 and Arg464, present on the loop region lining the allosteric site. Also visible, is the change in orientation of Asn481' side chain from the adjacent subunit on serine binding, which covers the allosteric site. (B) In *E. coli* PGDH, superimposition of the regulatory domain of chain A of the apoform (purple) with the serine inhibited form (green) shows transition of the loop region at Pro348, with no flipping of side chain on serine binding. In this case, transition of the Pro348 covers the allosteric site on serine binding. In both the figures, serine is shown at this site in the stick form (yellow). Not all the interactions with serine have been shown in this figure to maintain the clarity of the figure.

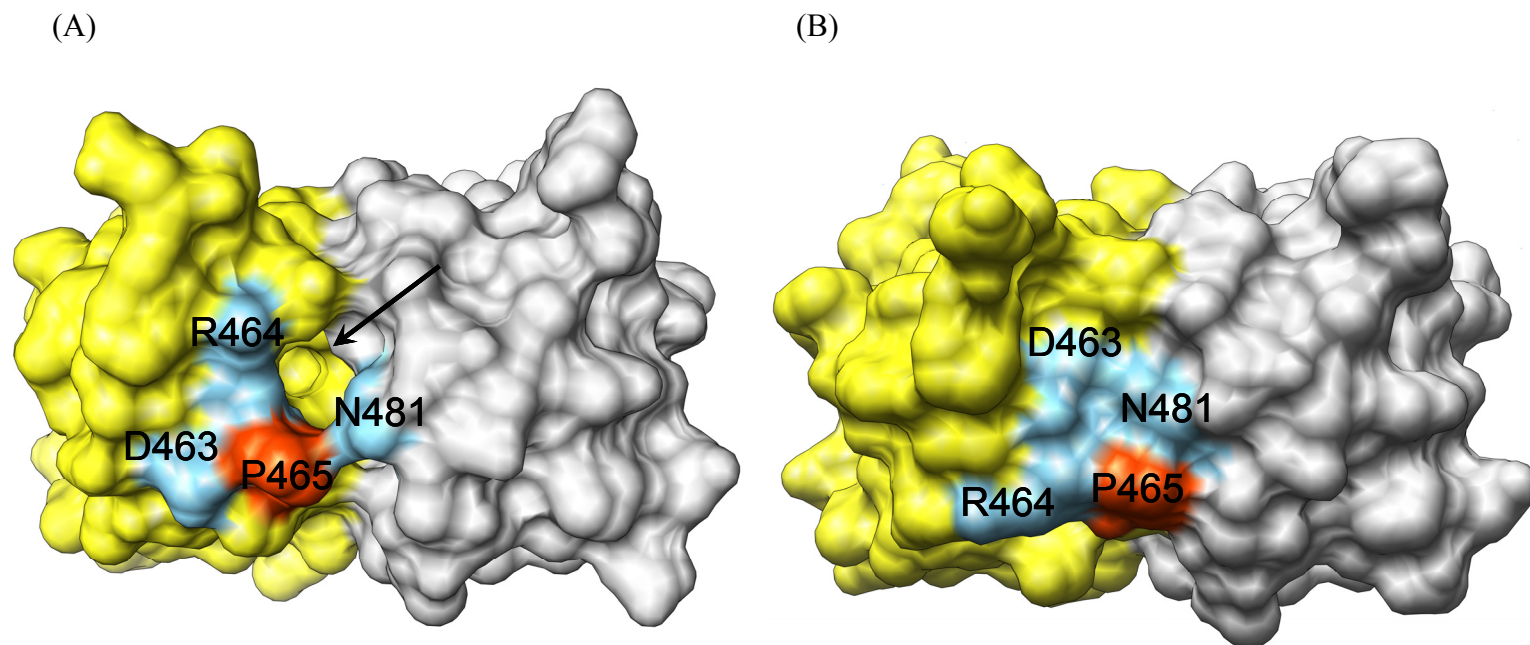


FIGURE 2.19. **Molecular surface representation of the serine binding site.** This figure shows the molecular surface representation of the regulatory domain interface with the changes occurring at the allosteric site on serine binding. Adjacent regulatory domains are colored gray and yellow with the labeled residues lining the allosteric site. (A) The presence of an opening (channel) for entry of serine (shown with an arrow) (B) The closure of this entrance on serine binding, as shown in the figure.

The *Mtb* PGDH structures reported in this chapter and the insights gained from them will aid considerably in understanding the mechanism of action of the ACT domains in regulation not only for this form of PGDH but also for other less closely related members of the ACT domain family. In addition, since the human enzyme is insensitive to serine, this structure may also provide a stepping-stone for the development of new drugs against *Mtb*. As a future work, this allosteric site can be used as a template for virtual screening against the fragment-based library. In addition, as the mechanism of feedback inhibition is not quite clear from this serine-bound *Mtb* structure, a situation, similar to leucine-bound *Mtb* LeuA structure (114); further crystallization trials can be conducted in search for other crystal forms which may lead to a different *Mtb* PGDH structure that can answer this question.

Cloning, purification and crystallization of human PGDH

Sequence alignment of human PGDH with *Mtb* PGDH has shown that humans also possess a similar form of PGDH with insertion near the C-terminus but have a sequence identity of only 33%. Further analysis of the sequences of both human and *Mtb* PGDH showed that all the residues interacting with the cofactor and the substrate at the active site are conserved in these proteins. However, human PGDH lacked sequence identity in the other two domains: the intervening and regulatory domains (20%). The residues that are thought to be responsible for the substrate inhibition in *Mtb* PGDH are not conserved in human PGDH even though it is thought to be also inhibited by HPAP due to its high sequence identity (94%) with rat liver PGDH. The stretch of three glycine residues (G316-G318) which are the locus for the 160° rotation of the nucleotide-binding

and substrate-binding domains in relation to the other two domains during tetramer formation in *Mtb* PGDH, are also not present in human PGDH. This fact raises the question of how the domains as well as subunits are arranged in this form of PGDH in humans. Since, human PGDH has a high sequence identity with rat liver PGDH, which does not show any serine inhibition, we also wanted to know about the inhibitory effect of L-serine on human PGDH.

Although a truncated version of human PGDH structure with nucleotide-binding and substrate-binding domain was recently deposited in the protein data bank (PDB code – 2G76), it does not provide us with any information regarding the domain organization as well as oligomeric formation of the whole mammalian protein. Hence, it would be necessary to determine the three-dimensional structure of the full-length protein to understand the differences in domain organization and regulatory domain interface between human and *Mtb* PGDH.

Toward this end, we have successfully cloned and purified the human PGDH (both His-tagged and untagged) to homogeneity using various purification protocols described in the methods section. His-tagged PGDH protein was purified using a single-step Ni affinity column where as the wild type PGDH was purified using anion exchange and size exclusion chromatography (Figure 2.20). Further kinetic studies have shown that the human protein is insensitive to serine inhibition (78). This fact is quite fascinating since PGDHs having an insertion near the C-terminus show different inhibitory effects of serine, the end product of this biosynthetic pathway and hence prompted us to determine the three-dimensional structure of human PGDH. The

resulting human PGDH structure will explain the role of the C-terminal domain that is present even though the enzyme is not inhibited by L-serine, and also determine the residues involved in substrate inhibition.

We have further tried to crystallize human PGDH complexed with the cofactor NADH. The crystallization procedure always requires a high protein concentration (10 mg/ml or greater), but human PGDH could not be concentrated beyond 4 mg/ml as the protein precipitates at higher concentrations. Several tricks were devised to overcome this problem: the addition of 200 mM LiSO_4 or reducing agents like 10 mM DTT/TCEP to the protein storage buffer or incubation of the dilute protein with the cofactor NADH to stabilize the protein before concentration. The addition of LiSO_4 as well as TCEP did not help in stabilizing the protein; however, the addition of 10 mM DTT as well as incubation of the protein with NADH resulted in protein concentrations of up to 16 mg/ml. It was also observed that human PGDH was mostly stable in potassium phosphate (pH 7.5) or HEPES buffer (pH 7.5) than in Tris buffer (pH 7.5). To further facilitate the process of crystallization in this protein, we have also constructed a truncated version of human PGDH (Ser323-Phe533) containing only the intervening and regulatory domains. Crystallization trials for all the different constructs of human PGDH with different matrix screens from Hampton Research as well as Emerald BioSystems have been carried out and we are awaiting the results from these trials.

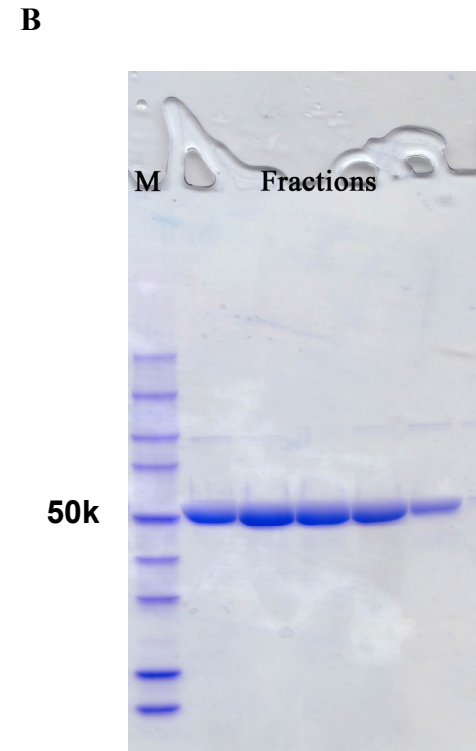
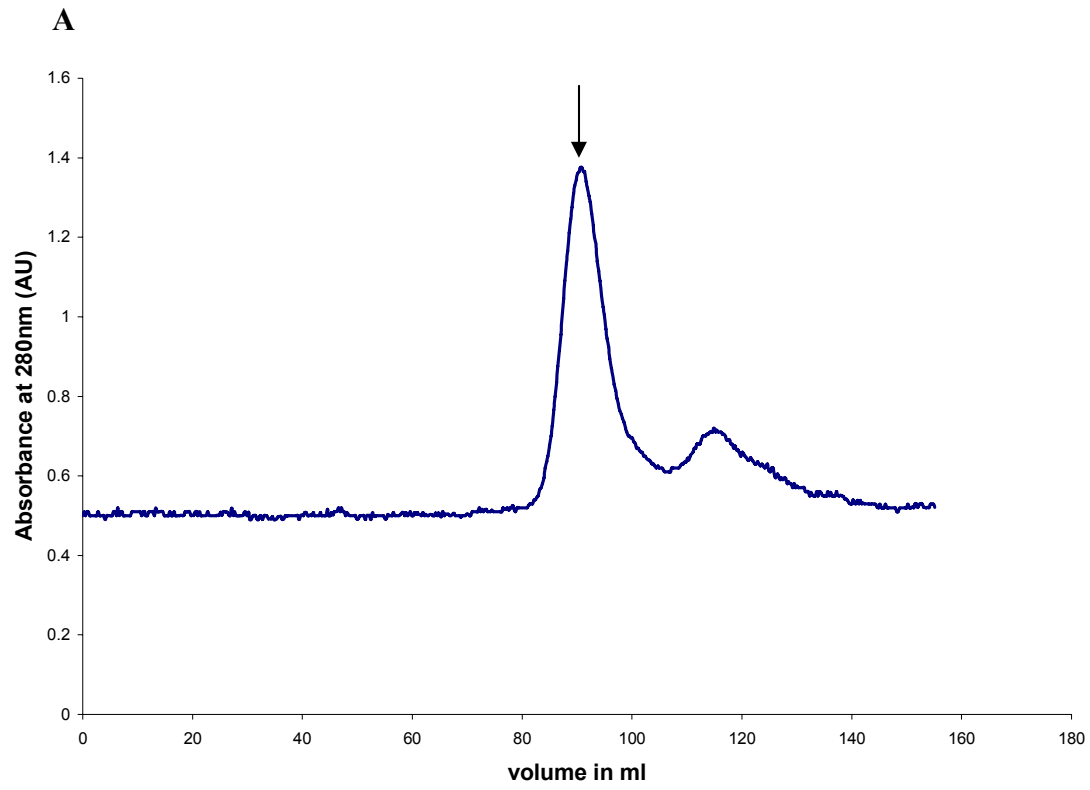


FIGURE 2.20. Purification of human PGDH. This figure shows (A) the chromatogram of a gel filtration column for human PGDH purification (B) the 4-12% SDS-PAGE gel stained with Coomassie Blue dye, containing the fractions of wild type human PGDH purified using gel filtration column. Lane M specifies Perfect protein[™] marker (10-225 kDa) (Novagen). The molecular weight of human PGDH is approximately 57 kDa.

In conclusion to the human PGDH part of the chapter, we have successfully cloned and purified different constructs of human PGDH in preparation for the three-dimensional characterization of this enzyme.

CHAPTER III
CRYSTALLOGRAPHIC STUDIES ON 7, 8-DIAMINOPELARGONIC ACID
SYNTHASE IN THE BIOTIN BIOSYNTHETIC PATHWAY

Background

7, 8-Diaminopelargonic acid synthase (DAPAS), a pyridoxal-5'-phosphate (PLP) dependent enzyme, catalyzes the antepenultimate step of biotin biosynthesis. It catalyzes the transamination reaction of 7-keto, 8-aminopelargonic acid (KAPA) into 7, 8-diaminopelargonic acid (DAPA) (Figure 3.1). This is the only aminotransferase known to utilize S-adenosyl methionine (SAM) as the amino donor for its reaction.

Preliminary biochemical characterization of the DAPAS enzyme has been carried out in *B. sphaericus* (47), *Brevibacterium divaricum* (130), and *B. subtilis* (48). Studies were conducted in *B. subtilis* mainly to increase the production of biotin and dethiobiotin by fermentation process for commercial purposes (48). Although DAPAS has been extensively studied, biochemically (131-133) as well as structurally (134, 135) in *E. coli*, the mechanism of dual substrate recognition in this transaminase has not been resolved to date.

Genetic approaches in the form of transposon mutagenesis site hybridization (TraSH) (9) in *Mtb* H37Rv strain as well as *bioA* mutant studies in *M. smegmatis* (57) have indicated the role of a *bioA* gene, which codes for DAPAS, in persistence and survival during mycobacterial infection. Furthermore, recent kinetic analysis on the *Mtb* DAPAS (136) shows that its kinetic efficiency is quite low compared to the *E. coli* enzyme (133) (K_{cat} for *Mtb* DAPAS ($1 \pm 0.2 \text{ min}^{-1}$) versus *E. coli* DAPAS ($0.013 \pm 0.001 \text{ s}^{-1}$)) even though it has 48% overall sequence identity with the *E. coli* counterpart.

Hence, in order to understand the active site architecture, we have characterized the crystal structure of *Mtb* DAPAS in complex with the cofactor PLP. In addition, characterization of the substrate-bound DAPAS structures elucidates the basis of recognition of the enzyme's two diverse substrates, SAM and KAPA. Utilizing the active site of KAPA-bound *Mtb* DAPAS as a template, we further carried out virtual screening of large compound libraries to search for potent inhibitors.

Furthermore, in this study, in order to understand the variations in the architecture of the DAPAS active site in certain *Bacillus* species and thermophiles, we have also determined the three-dimensional structure of the *B. subtilis* DAPAS enzyme in the apofrom, PLP-bound form as well as in a ternary complex with PLP and substrate KAPA using X-ray crystallography (Figure 3.2).

```

M.tb_DAPAS      MAAATGGLTPEQIIA*VDGAHLWHFYSSIGREAVSFVVAVAAHGAWLTLIRGQPIEVLDA
E.coli_DAPAS    MTTDD-----LAFDQRHILHPYTSMT--SPLPVYPVSAEGCELIISDGR--RLVDG
B.sub_DAPAS     MTHDL-----IEKSKKHLWL*PFTQM*KDYDENPLTIESGTG---IKVKDINGKEYYDG
1.....10.....20.....30.....40.....50.....

M.tb_DAPAS      MSSWNTAIHGHGHPALDQALTTQLRVMNHVMFGGLTHEFARLARLAKLLVDITPAGLDTVFF
E.coli_DAPAS    MSSWNAAIHGYNHPQLNAAMKSQLDAMSHVMFGGIT*HAF*IELCRKLVAMTPQPLECVFL
B.sub_DAPAS     FSSVNLNVHGHRKKE*LDDAIKKQLGKTAHSTLLGMTNVFATQLAETLIDISPKKLTRV*FY
61.....70.....80.....90.....100.....110.....

M.tb_DAPAS      SDSGSVSVEVAAKMALQYWRGRGLPGKRRLMTWRGGYHGDTFLAMSIDCFPHGGMHS----
E.coli_DAPAS    ADSGSVAVEVAMKMALQYWQAKG-EARQRFLTFRNGYHGDTFGAMSVCDP*DNSMHS----
B.sub_DAPAS     SDSGA*EAMEIALKMAFYWKNI*GKPEKQKFIAMKNGYHGDTIGAVSVGSIELFHHVYGPL
121.....130.....140.....150.....160.....170.....

M.tb_DAPAS      LWTDLAAQVFAPQVPR---DYDPAYSAAFEAQLAQHAGELAAVVVEFPVQAGGMRPH
E.coli_DAPAS    LWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAAVIEEPIVQAGGMRMY
B.sub_DAPAS     MFESYKAPIPYVYRSES*GDPDECRDQCLRELAQLLEEHEEIAALSIE*SMVQAGSGMIVM
181.....190.....200.....210.....220.....230.....

M.tb_DAPAS      DPRYHDLR*LDICRRYEVLLIFDEIATGFGRTGALFAADHAGVSPDIMCVGKALTGGYLSL
E.coli_DAPAS    HPEWIKRIRKICDREGILLI*ADEIATGFGRTGKLFACEHAEIAPDILCLGKALTGGTMTL
B.sub_DAPAS     PEGYLAGVRELCTTYD*VLMIVDEVATGFGRTGKMFACEHENVQPDLM*AAGKGITGGYLP
241.....250.....260.....270.....280.....290.....

M.tb_DAPAS      AATLCTADVAHTISAG--AAGALMHGPTFMANPLACAVSVASVELLLGQDWRTRITELAA
E.coli_DAPAS    SATLTREVAETISNG--EAGCFMHGPTFMGNPLACAAANASLAILES*GDWQQQVADIEV
B.sub_DAPAS     AVTFA*EDIYKAFYDDYENLKTFFHGH*SYTGNQLGCAVAL*ENLALFESENIVEQVAEKSK
301.....310.....320.....330.....340.....350.....

M.tb_DAPAS      GLTAGIDTARALPAVTVRVCGAIG-----VIECDRFVDLAVATPAALDR---GVWL
E.coli_DAPAS    QLREQ*LAPARDAEMVADV*RVLGAIG-----VVE*THFVNMAALQKFFVEQ---GVWI
B.sub_DAPAS     KLHFLQDLHALPHVGD*IRQLGFCGAELVRSKETKEPYEADRRIGYKVS*LKMREL*GLT
361.....370.....380.....390.....400.....410.....

M.tb_DAPAS      RPF*RNLVYAMPPYICTPAEITQITSAMVEVARLVGSLP-
E.coli_DAPAS    RPFGLIYLMPPYIILEQQLQRLTAAVNR*AVQDETFFCQ
B.sub_DAPAS     RPLGDVIAFLPPLASTAEELSEMVAIMKQAIHEVTSLED
421.....430.....440.....450.....

```

FIGURE 3.2. Alignment of sequences of DAPAS in *Mtb*, *E. coli* and *B. subtilis*. The residues colored in green are conserved among DAPAS in all three organisms. Residues, which are identical and similar in all the three sequences, are colored yellow and cyan. The residues, which are involved in catalysis, are marked with red asterisks. This alignment is prepared using ClustalW (79)

Methods

Reagents

Restriction enzymes for *bioA* cloning were purchased from New England Biolabs. All the chemicals for purification and assays including SAM, SAH and sinefungin were purchased from Sigma except KAPA which was purchased from Toronto Research Chemicals Inc., Canada. DAPA was chemically synthesized by Dr. Joel Freundlich of Princeton University using the protocol of Gibson *et al.* (137).

Cloning, protein overexpression and purification

The *bioA* gene (Rv1568) from *Mtb* H37Rv genomic DNA was PCR amplified and cloned into pET28b with TEV site using *NdeI* and *HindIII* restriction sites. The *NdeI* site contained the ATG start codon and the stop codon preceded the *HindIII* site. The Y25A mutant form of DAPAS was constructed using QuikChange mutagenesis kit (Stratagene) with pET28b-Rv1568 as the template. The primers used in this study were:

For the native DAPAS:

Forward primer: 5' AGAGAAGCATATGGCTGCGGCGACTGGCGGGTT 3'

Reverse primer: 5' CCCAAGCTTTCATGGCAGTGAGCCTACGA 3'

For the Y25A DAPAS mutant:

Forward primer: 5' CATCTGTGGCACCCCTGCCCAGCTCCATCGGCAG 3'

Reverse primer: 5' CTGCCGATGGAGCTGGCAGGGTGCCACAGATG 3'

Both the clones (pET28b-Rv1568 and Y25A mutant) were then transformed into the *E. coli* expression host BL21 (DE3). The cell culture was grown in LB media with kanamycin (50 µg/ml) at 37 °C till the A_{600} reached 0.8 and then induced with 0.5 mM IPTG at 25 °C and grown overnight. There was no protein expression seen at temperatures lower than 25 °C and protein was mostly insoluble at temperatures above 25 °C. The cells were harvested by centrifuging at 3000 rpm for 30 min and resuspended in 50 mM Tris buffer (pH 7.5), 500 mM NaCl, 1 mM PMSF, 20 µg/ml DNase, and 100 µM PLP. The cells were lysed using French press at 11,000 psi and the cell debris was removed by centrifuging the cell lysate at 15,000 rpm for 1 hour. The supernatant was then loaded on a HisTrapTM FF Ni column, which was pre-equilibrated with 50 mM Tris buffer (pH 7.5), 500 mM NaCl, and the protein, was eluted with a gradient of imidazole (50 mM - 400 mM). The protein is 98% pure as judged from the SDS-PAGE. The fractions containing the pure protein were pooled and dialyzed overnight against a buffer containing 25 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 1 mM DTT.

The *B. subtilis bioA* gene was PCR amplified from the *B. subtilis* 168 genomic DNA (Bacillus Genetic Stock Centre, Ohio State University), further cloned, and the His-tagged protein was purified in the same way as the *Mtb* DAPAS.

Spectroscopic studies of Mtb DAPAS

Absorbance spectra for the *Mtb* DAPAS enzyme (12 µM) alone and in reaction with substrate SAM (1 mM) and its analogs (1 mM)-sinefungin and SAH in the range of 300-600 nm were measured on the Cary 100 spectrophotometer (Varian). Similar spectra

for the Y25A mutant form of DAPAS with 1 mM SAM and DAPA were also scanned. On excitation at 280 nm, fluorescence emission scans for the *Mtb* DAPAS (5 μ M) with 1 mM sinefungin, and S-adenosyl homocysteine (SAH) were followed in the range of 300-450 nm on the LS 55 luminescence spectrometer (Perkin Elmer Instruments) with constant stirring. All the experiments were performed at room temperature. The buffer used was 50 mM AMPPO (pH 8.5), 150 mM NaCl, 10% glycerol, 4 mM β -mercaptoethanol and 100 μ M PLP. Concentrations of DAPAS and its Y25A mutant were calculated from their absorbance at 280 nm using extinction coefficient 64490 /mM/cm and 63210 /mM/cm, respectively (138). Further kinetic parameters for *Mtb* DAPAS with SAM and sinefungin have been determined by performing the first half of the enzyme reaction at 335 nm to measure the formation of PMP (ϵ of PMP = 8300 /M/cm) (139). The Michaelis-Menten equation was used to calculate the K_m^{app} and K_{max} values with the curvefit option in the Kaleidagraph software.

For the inhibition assays of the top ranked compounds from virtual screening, we have tried two different concentrations of the compounds, 10 μ M and 50 μ M. All the compounds were dissolved in 100% DMSO to prepare stock solutions (10-100 mM). In a typical 100 μ l enzyme reaction, 5 μ M of *Mtb* DAPAS was incubated with the compound for 1 min. The buffer used was 50 mM AMPPO (pH 8.5), 150 mM NaCl, 10% glycerol, 4 mM β -mercaptoethanol and 100 μ M PLP. The reaction was started by the addition of 800 μ M of SAM and change in absorbance of the enzyme reaction was monitored at 335 nm. We also confirmed that DMSO did not affect the enzyme reaction

by replacing the DMSO dissolved compounds with an equal volume of DMSO in the reaction mixture and monitoring the absorbance at 335 nm.

Crystallization and data collection

The crystal trials for *Mtb* and *B. subtilis* DAPAS were carried out at 18 °C using the crystal screen I and II, and Index screen of Hampton Research and wizard I, II and III of Emerald Biosystems. Crystals were initially obtained using robotic screening with a 1:1 ratio on 96 well intelliplates (Hampton Research). Yellow crystals for *Mtb* DAPAS would appear after heavy precipitation in conditions having PEG 8000 as the common precipitant (Figure 3.3A). These crystals have a single random mutation of H315R in the protein. The quality of the crystals could not be improved by any optimization method (different types and concentrations of PEG, additive screens, detergent screens etc.). A 2.2 Å intensity dataset for *Mtb* DAPAS crystal in PEG 8000, 0.1 M Tris buffer (pH 7.0), 0.1 M MgCl₂ (WIZARD II # 43) was collected after a lot of screening.

To determine the crystal structure of SAM-bound DAPAS, we soaked the crystals with different concentrations (10-40 mM) of SAM, but the crystals resulted in no diffraction to X-rays. For cocrystallization, mixing high concentrations of SAM with the protein, led to precipitation of the protein. The dataset for the sinefungin-bound DAPAS was obtained by soaking the crystal in 10 mM sinefungin for 45 min. This condition was obtained after the DAPAS crystals were soaked with different concentrations of sinefungin at different time points. This dataset was also collected in-house at the resolution of 2.2 Å. This DAPAS crystal was obtained in 20% PEGMME 5000, bis-tris buffer (pH 7.5) (# 46 of Hampton Research's Index Screen). However,

soaking the DAPAS crystals for a longer time in sinefungin or using higher concentration of sinefungin also resulted in deteriorated or no diffraction of the crystal.

Crystals for *B. subtilis* DAPAS were obtained in 4 μ l drop of protein (10 mg/ml) and precipitant (20% PEG 3350, 0.2 M sodium thiocyanate, #13 PEG/Ion screen of Hampton Research) in a 1:1 ratio (Figure 3.3B). The quality of the crystal greatly improved with the addition of 5% xylitol from the additive screen (#56 of Hampton Research) to the crystallization drops. The PLP-bound and PLP + KAPA-bound DAPAS datasets were obtained by cocrystallization of the DAPAS protein with 1 mM PLP and 1 mM PLP + 5 mM KAPA respectively, in hanging drop plates. Simply soaking the DAPAS crystals with PLP (5 mM) or PLP (5 mM) + KAPA (5 mM) would result in no diffraction of the crystals.

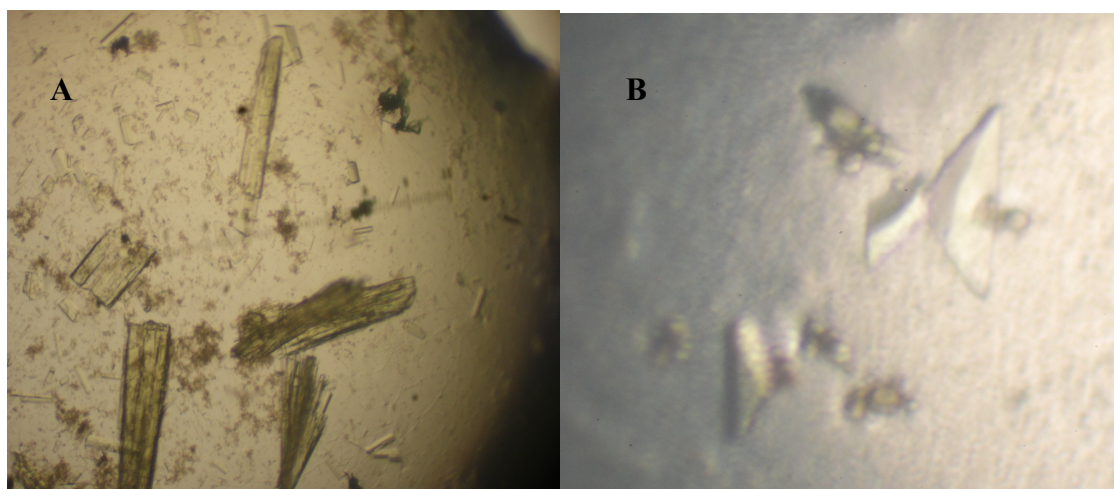


FIGURE 3.3. **Crystals of DAPAS enzyme.** (A) *Mtb* and (B) *B. subtilis* using hanging drop-vapor diffusion method in PEG conditions from Hampton research. DAPAS crystals are in the dimension of 50 μ M x 50 μ M x 50 μ M.

All the DAPAS crystals were cryoprotected with paratone oil and the data were collected at the home-source on Rigaku Raxis IV ++ image plate detector at 120 °K. All the datasets for *Mtb* and *B. subtilis* DAPAS were processed and scaled using the HKL2000 suite (90).

Structure determination

The structure of PLP-bound *Mtb* DAPAS was determined using molecular replacement (MR) with Phaser (94). The *E. coli* PLP-bound DAPAS structure (PDB code - 1QJ5) was used as the search model (LLG = 1437.6) for this MR. Even though, both the *Mtb* DAPAS structures possessed the same orthorhombic space group ($P2_12_12_1$) they had quite a different unit-cell dimensions as provided in the Table 4. So the structure of sinefungin-bound DAPAS was solved by MR using the PLP-bound *Mtb* DAPAS as the model (LLG = 4189.75). Phasing of the *B. subtilis* DAPAS dataset was also achieved by MR with Phaser using the coordinates of chain A of *Mtb* DAPAS (PDB code - 3BV0; LLG = 279). Using chain A of *E. coli* DAPAS (PDB code - 1QJ5) as a model, the LLG was 243.

Molecular replacement technique

Molecular replacement method that was used to solve the crystal structures of DAPAS in this chapter requires the presence of an already solved homologous model with at least 25% sequence identity between the two protein sequences. In this case, we used the *E. coli* DAPAS structure (PDB code-1QJ5) as the search model (sequence identity of 48%). The aim is to properly place the known *E. coli* DAPAS structure in the

unit cell which was determined from the diffraction data of the *Mtb* DAPAS. This idea of molecular replacement was first initiated by Rossmann and Blow in 1962 (140). The structure factors obtained from the search model and the diffraction data were used to calculate the corresponding Patterson maps. This map is a Fourier map calculated with the square of the structure factor amplitude and phase angle of zero. The idea is to best overlap the Patterson map generated from the *E. coli* DAPAS onto the Patterson map of the *Mtb* DAPAS.

Molecular replacement involves two steps: calculation of the rotation and translation functions. In the rotation function search, the search molecule is systematically rotated through all the possible orientations (α , β , γ). The characteristic pattern for each orientation is compared with the Patterson map from the experimental data, and close correspondences suggest that the orientation of the experimental molecule and the search molecule are the same. After the orientation of the experimental molecule has been determined by this rotation function search procedure, a related procedure can be carried out using the properly oriented search molecule, this time systematically translating the search molecule through the asymmetric unit (x , y , z) and calculating the corresponding Harker vectors, again looking for a match to the experimental Patterson map (Figure 3.4).

Over the years, many mathematical transformations have been developed to make the MR search computationally more efficient. In recent years, program such as Phaser (94) uses maximum-likelihood-based algorithms for the calculation of fast rotation and translation functions for MR solutions.

TABLE 4
Data collection, refinement and geometry statistics of *Mtb* DAPAS

	<i>Mtb</i> -DAPAS	<i>Mtb</i> -DAPAS + sinefungin
<i>Data collection</i>		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	a= 63.1 Å b=66.5 Å c=203.4 Å α=β=γ=90°	a=62.6 Å b=83.4 Å c=157.9 Å α=β=γ=90°
Molecules per asymmetric unit	2	2
Wavelength (Å)	1.54	1.54
Resolution range (Å)	35-2.2	30-2.2
Highest resolution bin (Å)	2.3-2.2	2.3-2.2
Observed reflections	1149410	618111
Unique reflections	121230	30270
Completeness (%) ^a	99.5 (96.1)	99.9 (99.8)
Average redundancy ^a	18.7 (10.8)	20.4 (16.4)
I/σ(I) ^a	19.3 (1.6)	39.8 (5.2)
R _{sym} ^{a,b}	0.16 (0.85)	0.081 (0.59)
<i>Refinement statistics</i> (REFMAC)		
Free R value (5%)	24	23
R value (%) ^c	19	18.7
No. of protein residues	860	841
No. of water molecules	200	280
r.m.s.d. bond length, (Å)	0.009	0.016
r.m.s.d. bond angles, (°)	1.2	1.81
<i>Ramachandran plot</i> (PROCHECK)		
Most favored region (%)	629 (90.4)	631 (89.1)
Additional allowed regions (%)	60 (8.6)	65 (9.2)
Generously allowed regions (%)	3 (0.4)	5 (0.7)
Disallowed regions (%)	4 (0.6)	4 (0.6)

^a Values in parenthesis for the highest resolution bin

^b $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

^c $R = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ where F_{obs} and F_{calc} are the observed and calculated structure factors

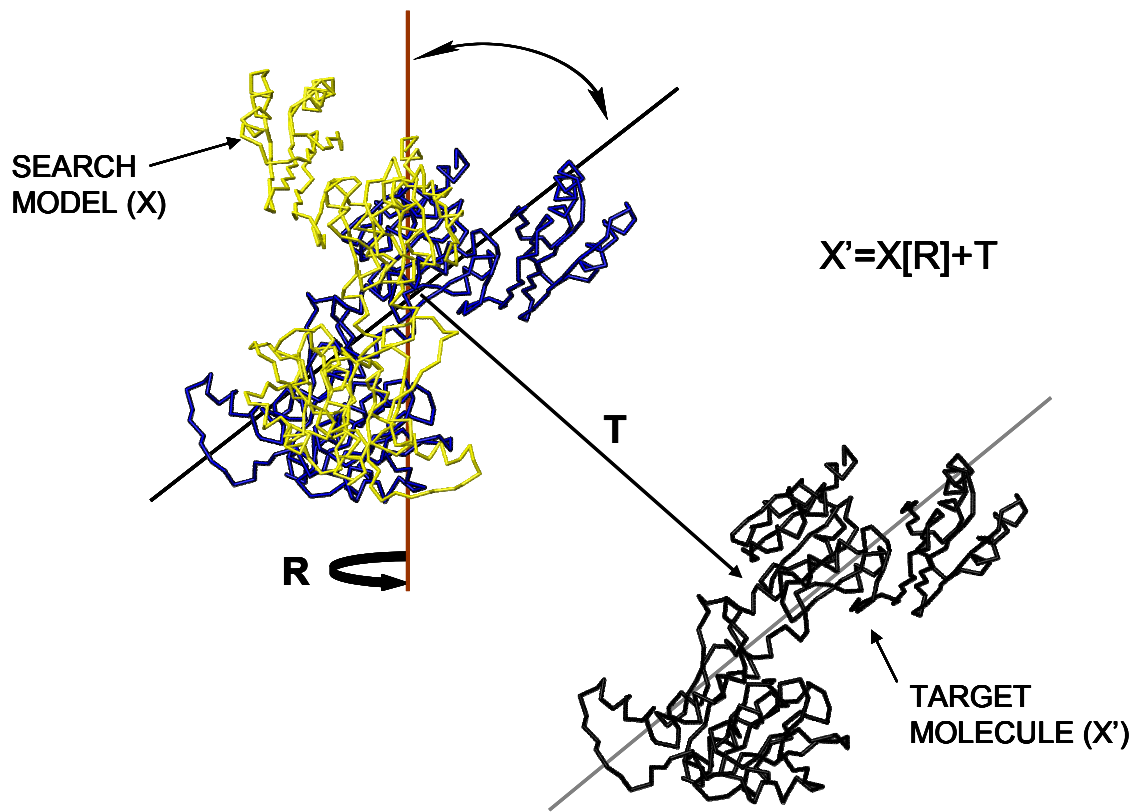


FIGURE 3.4. **Principle of Molecular replacement method.** This figure represents the principle of molecular replacement strategy to obtain phases for structure determination. The correct position, X' in the unit cell of search model with coordinates X is obtained by applying the rotation function R and translation function T to X .

Model building and refinement

Iterative rounds of model building were performed using the Xfit module of Xtalview (102) and Coot (141) with the Shake&wARP (142) bias-minimized electron density maps. These models were refined (rigid body and restrained refinement) using REFMAC of the CCP4 suite (103). When the models have reached decent Rfactor and R_{free} values below 30%, water molecules were added to the model, and inspected manually during refinement. Criteria for adding water molecules to the model are as follows: 1) a minimum of 3σ peak present in the $F_{\text{obs}} - F_{\text{calc}}$ difference map 2) water molecule should make sensible hydrogen bonds with protein atoms, and 3) the B factor values of the water molecules should not be unusually high compared to the protein residues during refinement. The quality of the models was validated by generating the Ramachandran plots using PROCHECK (106). Data collection, refinement and geometry statistics of *Mtb* DAPAS and *B. subtilis* DAPAS are provided in the Table 4 and 5, respectively. All the structure figures in this chapter have been prepared using Chimera (108).

TABLE 5
Data collection, refinement and geometry statistics of *B. subtilis* DAPAS

	<i>Bs</i> -DAPAS	<i>Bs</i> -DAPAS + PLP	<i>Bs</i> -DAPAS +PLP+ KAPA
<i>Data collection</i>			
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell dimensions	a=58.1 Å b=102.8 Å c=74.53 Å α=γ=90° β=105.1°	a=58.1 Å b=102.8 Å c=74.53 Å α=γ=90° β=105.1°	a=57.72 Å b=104.9 Å c=74.96 Å α=γ=90° β=105.3°
Molecules per asymmetric unit	2	2	2
Wavelength (Å)	1.54	1.54	1.54
Resolution range (Å)	35-2.2	35-1.9	30-2.26
Highest resolution bin (Å)	2.3-2.2	2.0-1.9	2.34-2.26
Observed reflections	1149410	372563	157974
Unique reflections	121230	65085	39726
Completeness (%) ^a	99.5 (96.1)	98.5 (96.8)	98 (82.2)
Average redundancy ^a	18.7 (10.8)	5.7 (5.6)	4.0 (2.9)
I/σ(I) ^a	19.3 (1.6)	22.1 (2.0)	10.5 (1.2)
R _{sym} ^{a,b}	0.16 (0.85)	0.06 (0.72)	0.12 (0.73)
<i>Refinement statistics</i> (REFMAC)			
Free R value (5%)	26	24	27
R value (%) ^c	23	20	21
No. of protein residues	830	833	894
No. of water molecules	250	385	261
r.m.s.d. bond length, (Å)	0.012	0.009	0.014
r.m.s.d. bond angles, (°)	1.7	1.32	1.74
<i>Ramachandran plot</i> (PROCHECK)			
Most favored region (%)	657 (89.5)	658 (90.0)	681 (87.3)
Additional allowed regions (%)	67 (9.1)	68 (9.3)	94 (12.1)
Generously allowed regions (%)	6 (0.8)	1 (0.1)	1 (0.1)
Disallowed regions (%)	4 (0.5)	4 (0.5)	4 (0.5)
PDB code	3DRD	3DOD	3DU4

^a Values in parenthesis for the highest resolution bin

^b $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

^c $R = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ where F_{obs} and F_{calc} are the observed and calculated structure factors

Results and discussion

Description of the overall fold in Mtb DAPAS

Mtb DAPAS has two distinct domains in its monomer, a small domain (residues 1- 60 and 339 – 437) and a large domain (residues 61 – 338), similar to *E. coli* DAPAS (134) (Figure 3.5A). The N-terminal part of the small domain consists of one helix (α 1) and three antiparallel β strands (β 1- β 3) connected by a long loop, and the C-terminal part starts with a helix-loop-helix having three antiparallel β sheets (β 11- β 13) and three helices (α 10- α 12) laid against these strands. The large domain consists of seven β strands (β 4- β 10- β 9- β 8- β 7- β 5- β 6), six of which are parallel with the exception of β 10 which is placed antiparallel in the sheet. These β strands are interconnected by helices (α 2- α 9), which also bear the cofactor-binding site. Superposition of the chain A on chain B of *Mtb* DAPAS using 435 C_{α} atoms results in an r.m.s.d. of 0.38 Å, indicating similar structure between the two chains as well as less flexibility between the large and small domains in contrast to aminotransferases of subclass I (143).

Dimer interface

The asymmetric unit of the crystal structure consists of a homodimer which is mostly formed by the extensive contact between the large domains of both subunits (Figure 3.5B). The buried surface area between the two monomers forming the dimer is approximately 4021 Å², 13.6% of the total surface area of the protein. In this dimer formation, the helix α 2 (Pro74-Arg85) of the large domain of subunit A is placed antiparallel to the helix α 2' of the large domain of subunit B at the non-crystallographic

2-fold axis, having only one hydrogen bond contact between the main-chain oxygen of Leu84 with Asp77-OD1 through a water molecule, at both ends of the helix. There are several other hydrophobic interactions throughout the interface. Monomer contacts are also present in the form of hydrogen bond interactions between His71-ND1 and Asp88'-OD1 (2.74 Å) at the loop regions, connected with this helix α_2 . There are also hydrogen bond interactions between the N-terminal part of the small domain and the large domain of the two subunits at His20-ND1:Asp116' main-chain oxygen (2.5 Å) and His23-ND1:Phe93' main-chain oxygen (2.82 Å). The helix α_1 of the small domain from one subunit crosses over the helix α_3' of the large domain of the other subunit. This mostly involves hydrophobic interactions with Pro, Val and Ile residues from both the subunits.

Comparison with E. coli DAPAS

The 437 amino acid *Mtb* DAPAS, which is eight residues longer than *E. coli* DAPAS, has a very similar overall structure with the exception of the loop regions (r.m.s.d. for 420 C_α atoms = 1.07 Å). This is also the case with other members of this class of aminotransferases, as discussed later. The orientation and length of the loop regions (Asp51 - Gly52, Thr111- Asp116, Pro189 - Asp196 and Ala307 - Asn322) of the *Mtb* enzyme differ from their *E. coli* counterpart. The position of helix α_6 (Pro197 - Glu213) differs with a movement of 2 Å. Certain interactions unique to *Mtb* DAPAS, not visible in *E. coli* enzyme, are a salt bridge between Arg246:Glu204 (3.0 Å) and several hydrogen bond networks (Tyr195-OH: Asp239- OD2: Ser200- γ O, Lys105-NZ: Asp109 (2.9 Å and 2.7 Å), Trp45-NE1: Glu418-OE2 (3 Å)).

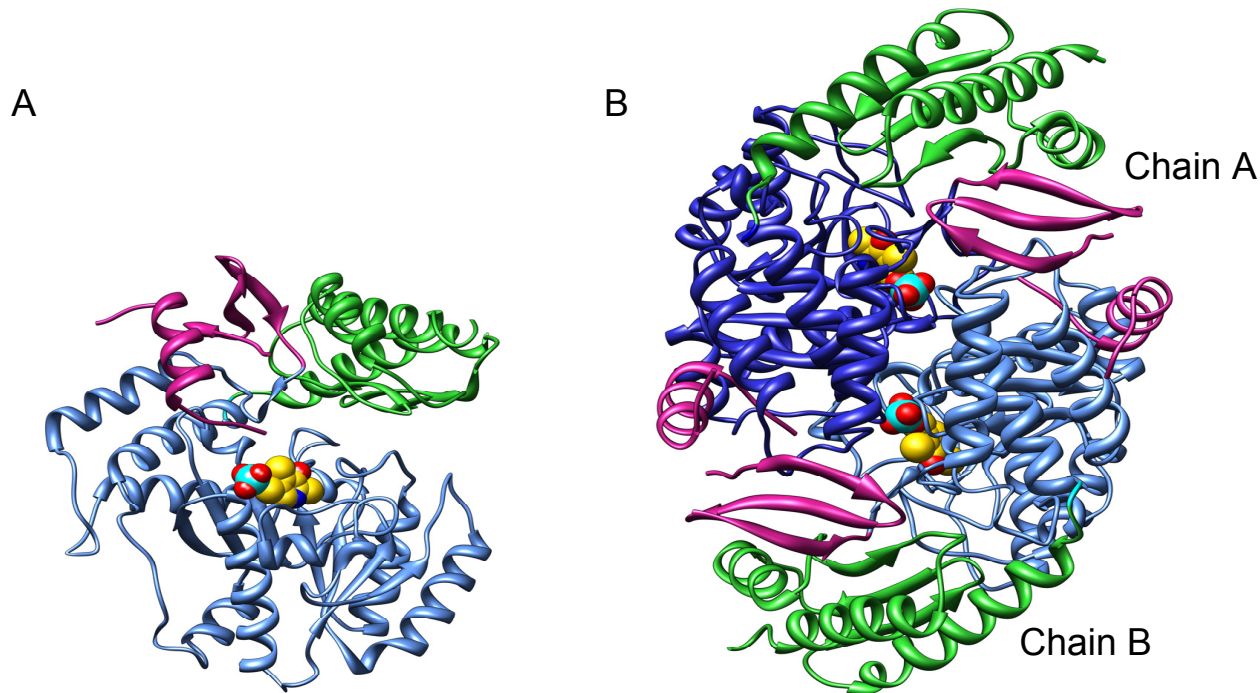


FIGURE 3.5. **Ribbon diagram of *Mtb* DAPAS structure.** (A) Monomer (B) Dimer of *Mtb* DAPAS. The monomer is divided into small and large domains. The small domain consists of residues from N- (pink) and C- (green) termini. The large domain of chain A and chain B are colored dark and light blue, respectively. As shown in the figure, the active sites are present at the interface of the dimer, 18 Å apart. The cofactor, PLP is shown as a sphere (yellow) to indicate the position of the active site.

Comparison with other PLP-dependent enzymes

A VAST (123) search for other structural homologs, using the coordinates of *Mtb* DAPAS, found an interesting similarity to lysine aminotransferase (Rv3290c) in *Mtb* (PDB code-2CJG, VAST score-34.6 with r.m.s.d. for 403 C_α atoms of 2.9 Å). Other homologs with high VAST scores were *E. coli* 2, 2 dialkylglycine decarboxylase (PDB code-2DKB: VAST score - 41.9 with r.m.s.d. for 402 C_α atoms of 2.3 Å) and GABA aminotransferase (PDB code-1OHW: VAST score - 33.9 with r.m.s.d. for 398 C_α atoms of 2.7 Å). *E. coli* KAPA synthase (PDB code -1BS0) (144), also related to the biotin biosynthetic pathway, which catalyzes the first reaction of KAPA formation, has a VAST score of 31.1 with r.m.s.d. for 350 C_α atoms of 3.5 Å. It has a similar secondary structure with *Mtb* DAPAS, mostly at the N-terminus with change in the orientation of the helix α1. In *Mtb* lysine aminotransferase (LAT), which is also highly upregulated in a nutrient starved TB model and a member of the subclass II of aminotransferases, L-lysine is the amino donor (145). Both of these enzymes have similar structural folds with the exception of variable loops and a few helices (α2, α4, α6, α7, α14) having different length and orientations (Figure 3.6). *Mtb* LAT uses a glutamate switch to distinguish between the substrates at the active site. The residues Glu243 and Arg422, required for substrate recognition in LAT, are not conserved in *Mtb* DAPAS. Dialkylglycine decarboxylase (146) belongs to the aminotransferase subclass II and PLP fold type I family, which catalyzes both decarboxylation and transamination reactions and does not undergo conformational changes on ligand binding at the active site. GABA aminotransferase (147) which also belongs to subclass II of aminotransferases catalyzes

the degradation of GABA into succinic semialdehyde. When all of these above mentioned enzyme structures were compared with *Mtb* DAPAS, despite having less sequence identity (21-27%), the overall fold is quite similar to *Mtb* DAPAS with the exception of the loop regions. These structural similarities strongly suggest that the *Mtb* DAPAS also belongs to subclass II of the aminotransferases (148).

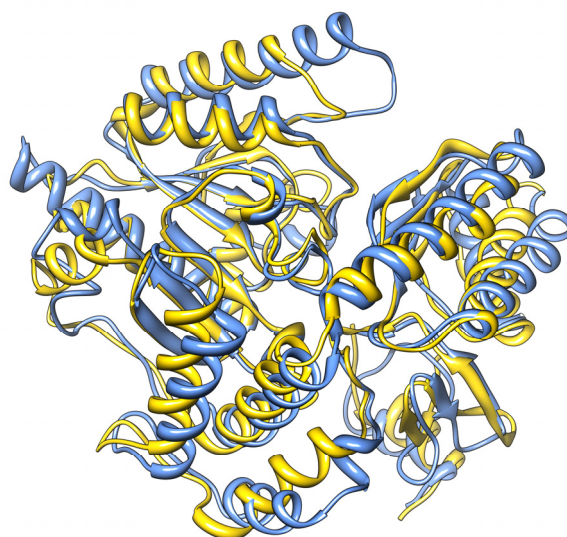


FIGURE 3.6. Superimposition of *Mtb* DAPAS on *Mtb* LAT. This figure is a ribbon diagram representation of an overlay of chain A of *Mtb* DAPAS (gold) on *Mtb* LAT monomer (blue; PDB code-2CJG) with r.m.s.d. of 2.9 Å for 403 C_α atoms.

Analysis of Mtb DAPAS active site

The active site of DAPAS is approximately 24 Å deep with a wide entrance of dimension 22 Å x 23 Å which narrows to a dimension of 14 Å x 10 Å at the bottom. This cavity is mainly formed by the loop regions Pro24 - Ser34, Arg156 - Asp160, Gln224 - Gly228, and Arg400-Arg403 from one subunit and Ala307'-Asn322' and Met87'-His97' from the other subunit. Thus, the active site cavity consists of residues from both the domains of one subunit but only the large domain of the adjacent subunit. Even though residues from both chains form the catalytic site, each subunit has its own distinct site, approximately 18 Å apart. The entrances of the cavities are present on opposite sides of the non-crystallographic two-fold dimer axis. In this structure, the loop at the entrance of the active site (Tyr25-Val33), as well as, the long loop that lines the active site (Ala307'-Pro317') are disordered. The latter loop has four alanine and three glycine residues in the sequence (307'-AGAAGALMHGP-317') at the entrance of the cavity which makes it quite flexible on solvent exposure. This stretch of residues in the active site is not conserved in *E. coli* as well as *B. subtilis* enzyme. In addition, 13 other residues, which are lining the *Mtb* active site, are also not conserved in *E. coli* DAPAS.

Cofactor-binding site

The inherent feature of any aminotransferase enzyme is the presence of cofactor PLP (vitamin B₆) at the active site. The electron density for PLP is clearly visible at the bottom of the active site cavity in both subunits of the DAPAS structure. This cofactor forms a covalent link with the N_ε group of conserved Lys283 as evident from the continuous electron density between PLP and Lys283 (Figure 3.7). In *Mtb* DAPAS, the

phosphate groups of the PLP are hydrogen bonded to the main-chain amides of Gly124 (3.5 Å), Ser125 (3.2 Å), and Thr318' from a loop in the adjacent subunit (3.2 Å). The pyridine nitrogen forms a hydrogen bond with the carboxylate-OD2 atom of the conserved Asp254 (2.8 Å), which is further stabilized by a hydrogen bond interaction with His158-ND1 (2.7 Å). The pyridine ring is also sandwiched between the perpendicularly stacked phenol ring of Tyr157 from the front (*re* side) and Ile256 from the back (*si* side). The hydroxyl group of PLP does not directly form any hydrogen bonds with the protein atoms, but interactions are seen with the carboxyl group of Glu220 and the main-chain amide of Ala226 via a water molecule (2.8 Å). There are several other ordered water molecules present around PLP in the active site. Thus, all these protein residues interacting with the cofactor PLP are contributed by both the subunits, characteristics of PLP fold type I family (149). These interactions with PLP are also conserved in *E. coli* DAPAS.

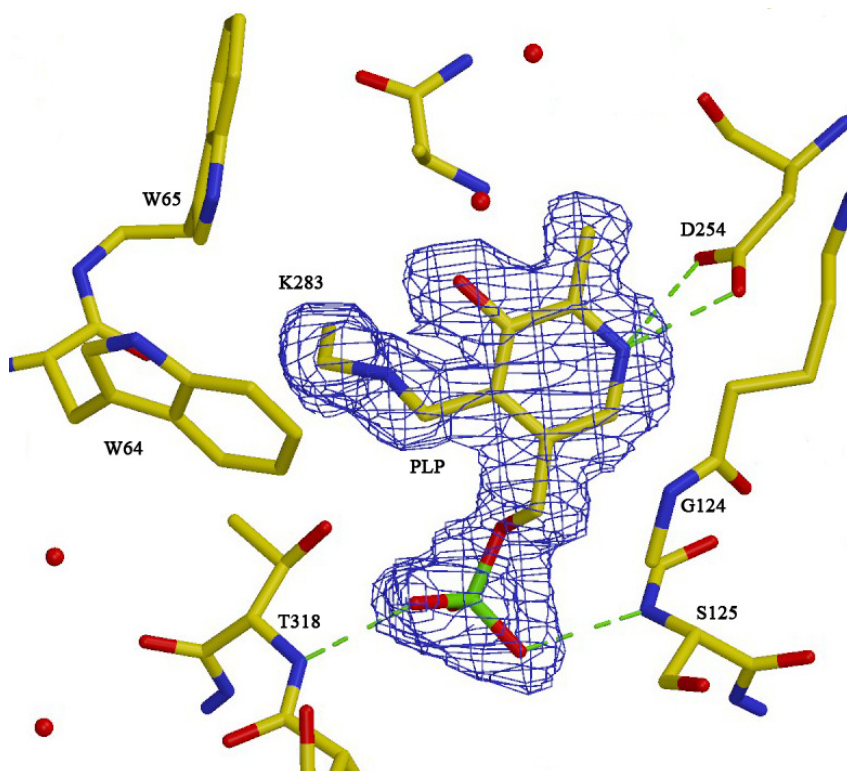


FIGURE 3.7. PLP binding at the active site of *Mtb* DAPAS. This figure shows the internal-aldimine form of the enzyme with the $2F_o - F_c$ electron density map at 1σ level for PLP-Lys283 covalent linkage. All the hydrogen bond interactions with PLP are shown as green dotted lines and at a distance below 3.5 \AA . The surrounding protein residues as well as cofactor PLP are shown in yellow stick form. Atoms are colored based on their type (N as blue, O as red, P as green)

Biochemical characterization of Mtb DAPAS

DAPAS is an aminotransferase that is responsible for the transfer of the amino group from SAM to KAPA to form DAPA, using PLP as a cofactor. The unique aspect of this enzyme is its usage of SAM as the amino donor for the transamination reaction. In *E. coli* DAPAS, other analogs of SAM such as SAH (4 mM), S-methyl-L-methionine (4 mM), adenosine (4 mM), and amino acids such as L-methionine (4 mM), L-leucine (8 mM) were not found to be the amino donor, thus, *E. coli* DAPAS was thought to be very specific for SAM (131). Recent studies of *Mtb* DAPAS have also shown that amino acids such as L-Asp, L-Glu, L-Met, L-Lys do not act as the amino donor (136), even at a concentration of 5 mM. This is in contrast to other aminotransferases such as glutamine transaminase K (150) and glutamine:phenylpyruvate aminotransferase (151) which can use a variety of amino acids like Asp, Glu, Met, and aromatic amino acids like Phe, Tyr, Trp as substrates. In this chapter, we have studied the *Mtb* DAPAS reaction with SAM and its analogs, and product DAPA, discussed later in this section.

All aminotransferases contain the cofactor PLP as an internal-aldimine, having an imine linkage with the ϵ -amino group of lysine at the active site. After the noncovalent binding of the first substrate at the active site, the amino group of the invariant lysine residue is replaced by the amino group from the substrate, thus forming an external-aldimine complex with the enzyme. PLP abstracts electrons from this substrate in the external-aldimine complex, acting as an electron sink. The release of any one of the C_{α} substituents leads to the formation of a resonance stabilized quinonoid intermediate in the PLP-dependent reactions (152) (Figure 3.8).

This cofactor PLP acts as a chromophore having absorption maxima around 340-550 nm due to the presence of delocalized electrons on the pyridine ring. For this reason, the PLP form of the protein is yellow in color. Around 320-340 nm range, the predominant species is the PMP form of the enzyme which is the result of sp^3 hybridization of the pyridine C4 atom and reduced π -electron delocalization. Sometimes a quinonoid intermediate, which absorbs at 490 nm, is formed during the PLP-dependent reaction. This shift in absorbance to a longer wavelength is also due to π -electron delocalization.

In this study, we utilized these chromophore properties of the cofactor (PLP or PMP) in studying the single-turnover half-reactions of *Mtb* DAPAS with substrate SAM and its analogs sinefungin and S-adenosyl homocysteine (SAH), as well as, with the product DAPA using spectroscopic methods. SAM and SAH differ from sinefungin by the presence of a sulfonium ion at the δ position. The sulfonium ion in SAM is methylated; sinefungin has an amine on its δ carbon (Figure 3.9).

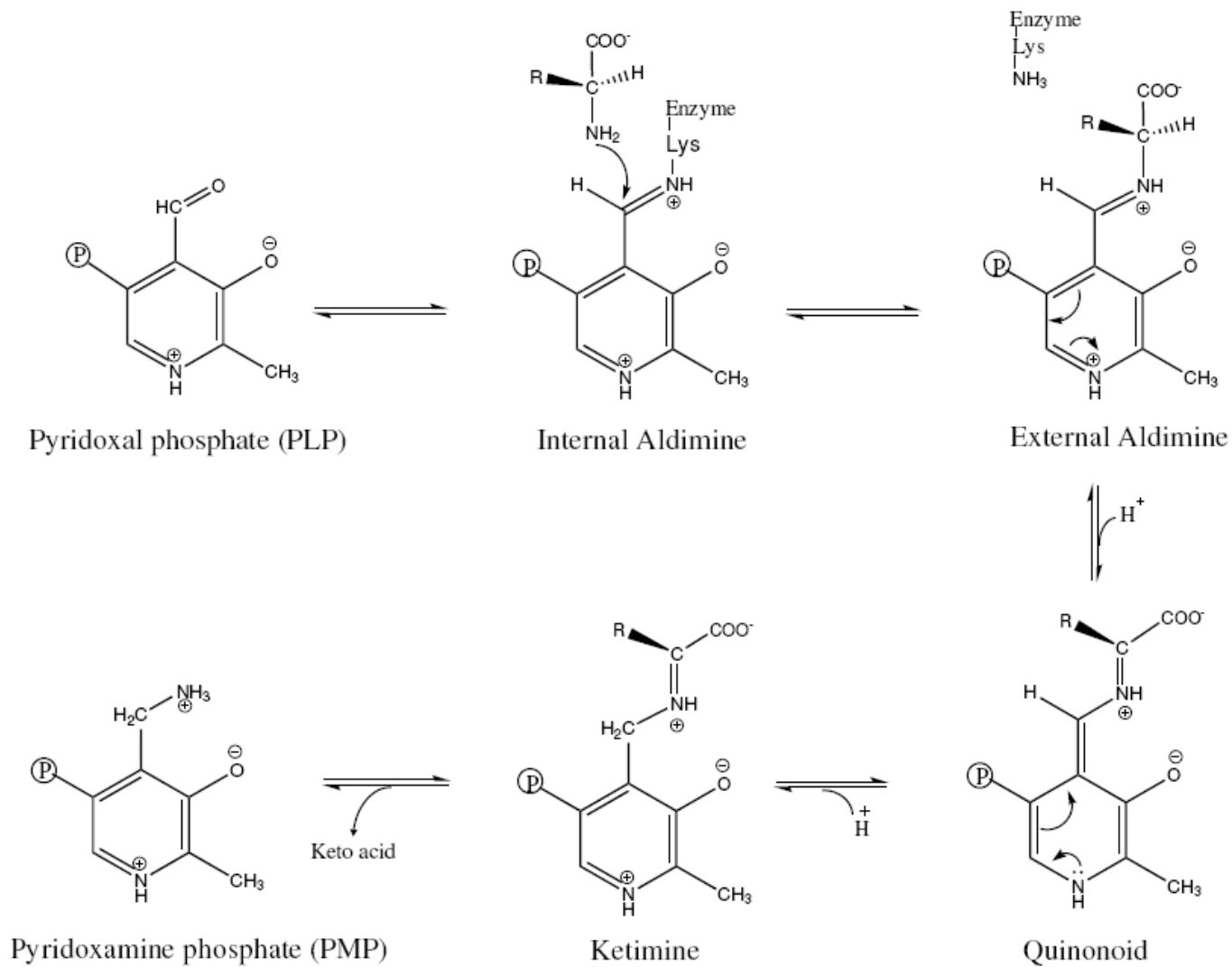
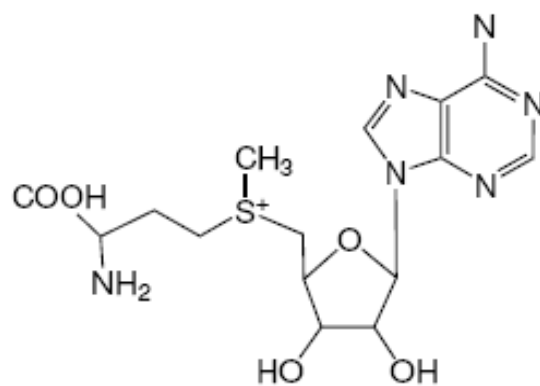
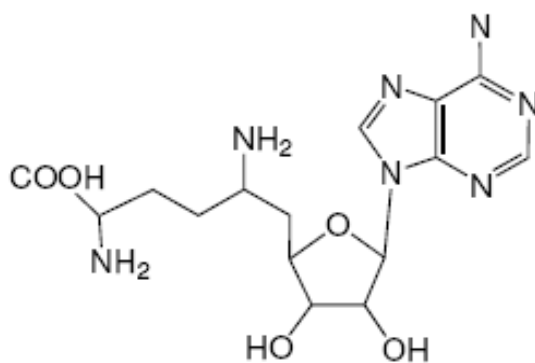


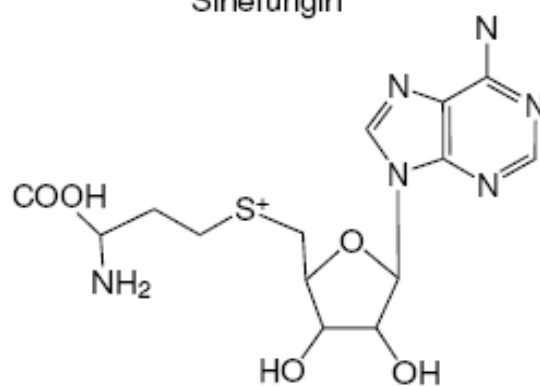
FIGURE 3.8. **PLP reaction intermediates in the first half of the aminotransferase reaction.**



S-adenosyl methionine (SAM)



Sinefungin



S-adenosyl homocysteine (SAH)

FIGURE 3.9. Structure of SAM and its analogs sinefungin and S-adenosyl homocysteine (SAH).

An absorbance scan of *Mtb* DAPAS showed absorption maxima at 420 nm due to the presence of PLP. This peak at 420 nm decreases with a corresponding increase in absorbance peak at 335 nm, when *Mtb* DAPAS reacts with 1 mM SAM or sinefungin. This peak at 335 nm is due to the formation of PMP. In contrast to the *E. coli* DAPAS reaction with SAM, the presence of the quinonoid intermediate at 487 nm was not detected in these reactions. This may indicate that this intermediate does not accumulate during the transaminase reaction and it is rapidly converted into the product. Similar absorption spectra were also observed when PLP-DAPAS was mixed with product DAPA (reverse of the second half-reaction) and scanned in the range from 300 to 500 nm. The reaction also shows a decrease in the absorption peak at 420 nm, with a subsequent increase in absorbance at 335 nm. This change in the absorption peaks results from the formation of PMP in the single-turnover transamination reaction with DAPA. From the absorbance spectra of *Mtb* DAPAS with SAM and with DAPA, it appears that the PMP formation occurs faster with DAPA compared to SAM (Figure 3.10). This is similar to the *E. coli* DAPAS half-reactions with SAM and with DAPA where the K_{\max} value for the DAPA reaction is 49 times greater than that of SAM (133). Interestingly, similar kinetic studies with the other SAM analog SAH (concentrations up to 3 mM) showed no absorption peak at 335 nm. This indicates that SAH may not be capable of donating its amino group to PLP. This finding is analogous to that reported for the *E. coli* (153) and *B. sphaericus* DAPAS (47) where SAH is not the amino donor.

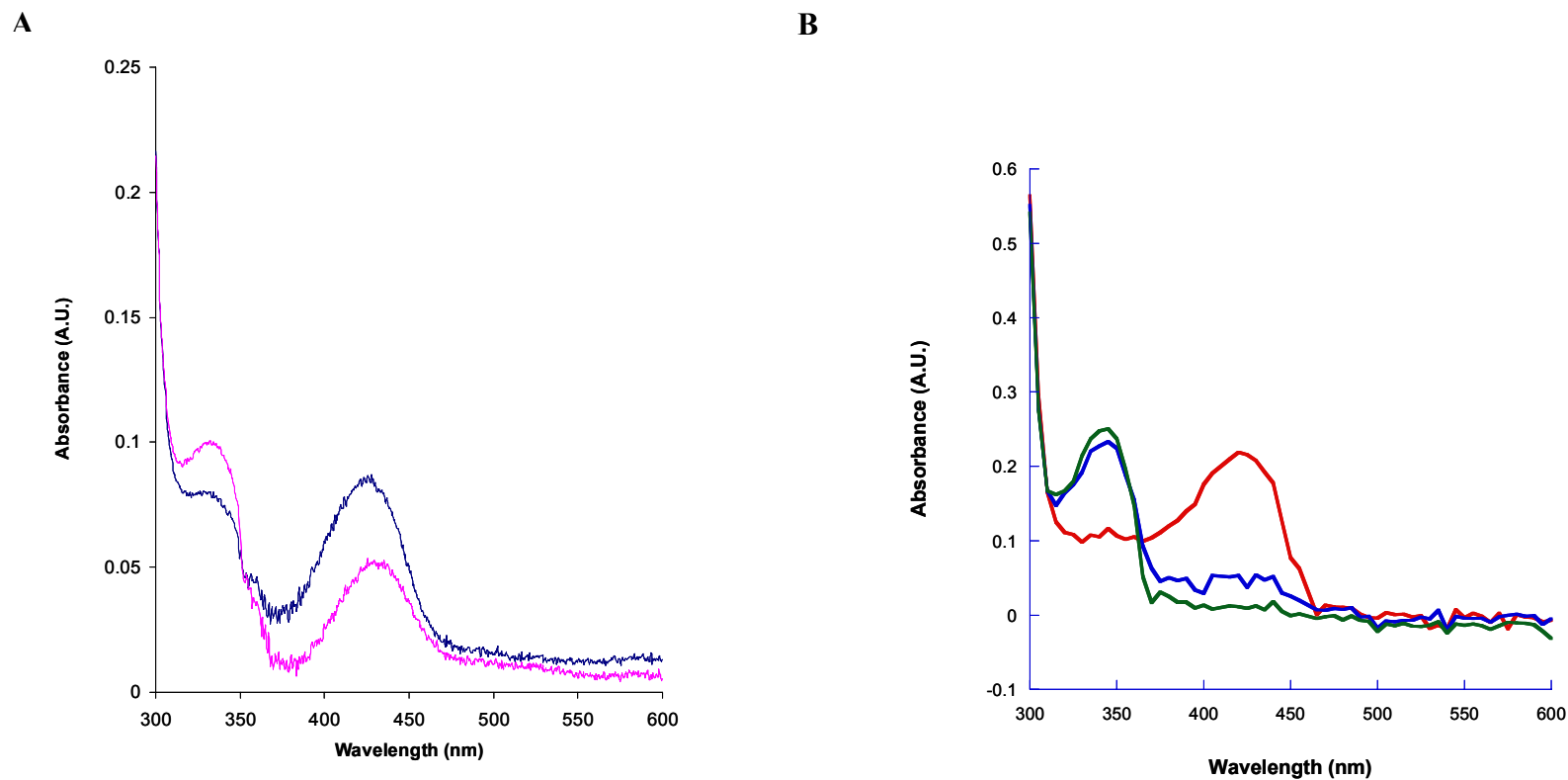


FIGURE 3.10. UV-VIS absorbance scans of *Mtb* DAPAS with the substrate SAM and the product DAPA. (A) The absorption spectra of *Mtb* DAPAS alone (blue) and after mixing DAPAS with 1 mM SAM (purple), after 30 sec. These spectra are very similar to the one reported by Bhor *et al.* (154) (B) The absorption spectra of *Mtb* DAPAS alone (red), immediately after the addition of 0.5 mM DAPA (blue), and 30 seconds (green) after addition of DAPA. The buffer used in these scans is 50 mM AMPISO (pH 8.5), 150 mM NaCl, 10% glycerol, 5 mM β mercaptoethanol at room temperature.

Further kinetic parameters for the single-turnover half-reactions of *Mtb* DAPAS with SAM and sinefungin have been calculated by monitoring the increase in absorbance at 335 nm with varying concentrations of substrate. The K_m^{app} values for sinefungin and SAM in *Mtb* DAPAS is nearly similar to the K_m^{app} value for SAM in *E. coli* DAPAS (132). However, the enzyme efficiency for the first-half reaction in *E. coli* DAPAS is much higher compared to its *Mtb* counterpart (Table 6).

TABLE 6
Kinetic parameters for the single-turnover-half reactions of *Mtb* and *E.coli* DAPAS

	K_m^{app} (SAM)	K_{max} (SAM)	K_m^{app} (sinefungin)	K_{max} (sinefungin)
<i>Mtb</i> DAPAS	0.45±0.2 mM	3.6×10^{-6} /sec	0.7±0.2 mM	5.2×10^{-7} /sec
<i>E.coli</i> DAPAS	0.3 mM	0.016 /sec	Not determined	Not determined

Furthermore, fluorescence emission spectra of *Mtb* DAPAS on excitation at 280 nm results in a peak at 335 nm. The *Mtb* DAPAS reaction with 1 mM sinefungin, when excited at 280 nm, demonstrated a decrease in the emission peak at 335 nm with a subsequent increase in the peak at 385 nm (characteristic of PMP form of the enzyme). The initial emission peak of the enzyme alone at 335 nm, upon excitation at 280 nm, resulted from the presence of aromatic residues in the protein. This phenomenon of an increase in emission at 385 nm is due to the FRET (Förster resonance energy transfer) effect on chromophore PMP, which absorbs the light at 335 nm and emits at 385 nm, thus, confirming the formation of PMP on DAPAS reaction with sinefungin. Hence, this

study indicates that sinefungin can also bind at the active site and donate its amino group to PLP. Thus, it acts as a substrate for *Mtb* DAPAS (Figure 3.11).

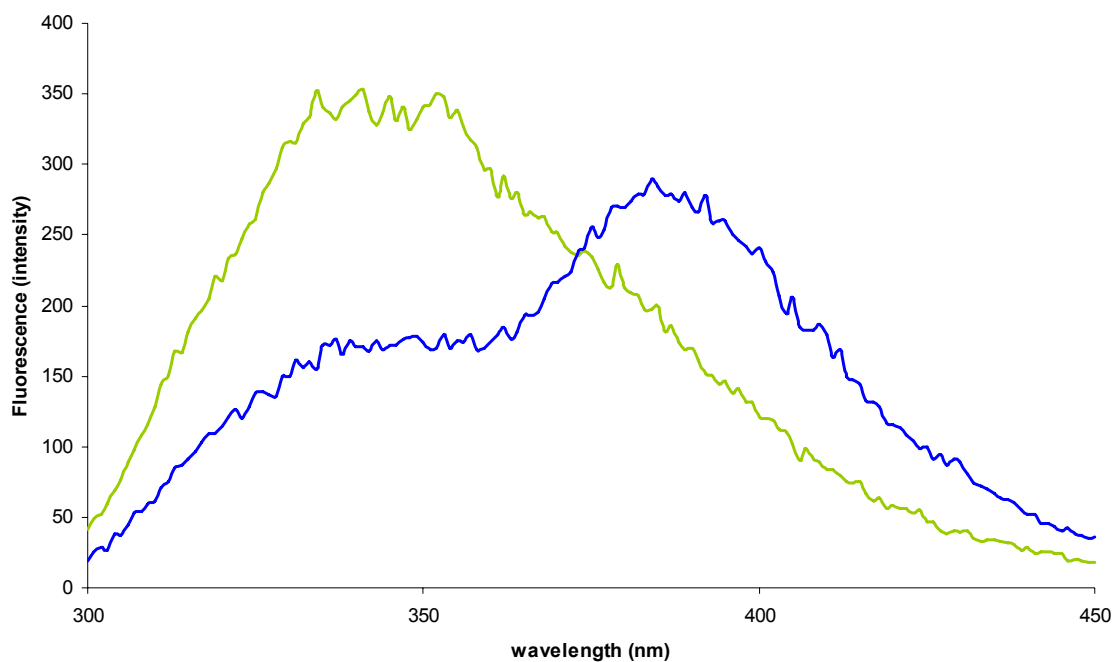


FIGURE 3.11. Fluorescence spectra of *Mtb* DAPAS with SAM analog, sinefungin.

Upon excitation at 280 nm, fluorescence emission spectra were scanned in the range between 300–450 nm for *Mtb* DAPAS (12 μ M) alone (green), and after mixing DAPAS with 1 mM sinefungin (blue) after 30 sec at room temperature.

Possible substrate recognition in Mtb DAPAS

In order to understand the mode of the SAM binding at the active site, attempts were made to determine the crystal structure of *Mtb* DAPAS complexed with the physiological substrate SAM, which were unsuccessful. Therefore, the crystal structure determination of DAPAS complexed with the SAM analog, sinefungin was tried and phases were obtained at a resolution of 2.18 Å. In this structure, sinefungin can be fitted to the extra electron density, which is present adjacent to the C4' atom of PLP at a distance of 3.56 Å, in the active site of the DAPAS chain A. On further refinement of this enzyme structure, the occupancy of sinefungin is found to be lower than one (0.8) due to high B-factor. The electron density for sinefungin is mostly disordered in chain B also due to its low occupancy. Surprisingly, in this structure, PLP is still covalently attached to the conserved Lys283 in both the chains; no PMP form is visible. This may be due to a short period (45 minutes) of 10 mM sinefungin soaking to avoid damage to the crystal and the smeary diffraction pattern. We believe that this structure depicts the precatalytic state of sinefungin binding. In this structure, the amino group of sinefungin points towards the internal-aldimine complex (3.56 Å) thus, oriented properly to form an external-aldimine complex with the enzyme. The carboxy oxygen atoms of sinefungin form hydrogen bonds with the main-chain amide of Gly316' (3.4 Å) and the hydroxyl group of Tyr157 via a water molecule in the active site. The aliphatic chain of sinefungin spans the entire length of the active site cavity making van der Waals interactions with residues Trp64, Trp65, and Phe402. At the entrance of the active site cavity, the electron density for the adenine ring of sinefungin is mostly disordered due to its complete

exposure to the solvent. The loop near the entrance of the cavity is also disordered (Ser26-Ala32), similar to the native structure. However, the loop region (Ala307'-Pro317') lining the active site at the interface of the dimer in chain A is ordered due to the presence of the ligand. Superimposition of the sinefungin-bound structure on the PLP-bound *Mtb* DAPAS structure shows that the enzyme does not undergo any major conformational changes (r.m.s.d. for 429 C α atoms is 0.3 Å) except for the movement of the side chains of Trp64, Trp65, Met409, and Phe402 to accommodate sinefungin in the active site. The changes in the active site are in the form of 90° rotation of the indole side chain of both Trp64 and Trp65, causing the side chain of Met409 to move by 2 Å. There is also a 180° rotation of Tyr25 side chain (based on *E. coli* DAPAS apostructure). Thus, the change in the unit-cell dimension of sinefungin-bound DAPAS structure (a= 62.6 Å, b=83.4 Å, c=157.9 Å, $\alpha=\beta=\gamma=90^\circ$) in comparison to PLP-bound structure (a= 63 Å, b=66.5 Å, c=203.4 Å, $\alpha=\beta=\gamma=90^\circ$) might be due to the crystal packing.

Based on the precatalytic form of the sinefungin-bound DAPAS structure, SAM was modeled in the DAPAS active site. Comparing this model with the sinefungin-bound structure, the amino group of SAM is positioned 2 Å further inside the active site cavity. This positioning of the amino group of SAM inside the active site would allow the formation of the Schiff's base with the C4' group of PLP. The carboxy oxygen atoms of SAM would interact with the hydrogen bond network of Tyr157-OH: Asp160-OD2. The aliphatic chain in SAM would span the hydrophobic active site pocket, and the adenine ring of SAM would have stacking interactions with the phenol ring of Tyr25. This stacking interaction of Tyr25 with the adenine ring is analogous to Tyr85 in the structure

of 1-aminocyclopropane-1-carboxylate (ACC) synthase-AMA enzyme-adduct complex (PDB code-1M4N). ACC synthase also uses SAM as the substrate; AMA is an amino-oxy analog of SAM (155). These interactions are in contrast to other members of subclass II aminotransferase in which the carboxy groups of the amino donor are mostly hydrogen bonded to the Arg residues in the active site. The sulfonium ion of SAM might form a hydrogen bond with the hydroxyl group of Tyr407, which is also conserved in all DAPAS, and the oxygen atom of the ribose sugar might form hydrogen bonds with Arg400 and the N-atom of Trp65 imidazole ring (Figure 3.12A).

To locate the binding site of the amino group acceptor KAPA in the active site, we have superimposed the KAPA-bound *E. coli* DAPAS structure on the *Mtb* enzyme structure. In the *E. coli* DAPAS structure, KAPA is bound to the PLP form of the enzyme, thus, it provides an unproductive form of the enzyme complex. Based on this superimposition, in *Mtb* DAPAS, the 7-keto group of KAPA would point towards the internal-aldimine at a distance of approximately 2.6 Å in the model. The 8-amino group of KAPA would form a hydrogen bond with the hydroxyl group of Tyr25 (2.5 Å) as well as with the main-chain oxygen of Gly316' (2.5 Å). Tyr25 is involved in stabilizing KAPA binding in the active site as well as the proper positioning of KAPA for amino group acceptance by forming a hydrogen bond with the N8-amino group. The carboxyl group of KAPA has a specificity for conserved Arg400, whose side chain would move 3 Å towards KAPA to form the salt bridge (3.15 Å), resulting in a hydrogen bond breakage between Arg400 and Thr389 observed in the PLP-bound structure. It is worth mentioning here that this movement of the guanidinium group of Arg400 in response to

the KAPA binding in the active site is not observed with sinefungin binding. The rest of the aliphatic chain in KAPA is within van der Waals distance to Tyr25, Trp64, Tyr157, and Phe402 of the active site (Figure 3.12B). Analogous to sinefungin binding, KAPA binding would also result in the movement of the side chains of a few residues in the active site (Trp64, Trp65, and Met409). Thus, no major conformational changes in terms of large domain movement are observed (r.m.s.d.= 0.3 Å for 429 C_α atoms). This is in contrast to some aminotransferases, like aspartate aminotransferase, in which the active site closes on ligand binding with a 13° rotation of the small domain as a rigid body toward the large domain (156). Thus, from the sinefungin-bound DAPAS structure and the KAPA-bound DAPAS model, we can derive that both SAM and KAPA will bind at the same place in the active site with differences in the movement of side-chains of the active site residues. Kinetic studies have shown that at high concentrations (14±2 μM), KAPA has also been found to be an inhibitor of this enzyme, forming a dead end complex in the active site (136). This also indicates that both KAPA and SAM occupy a similar position in the active site, even though they have very different size and characteristics.

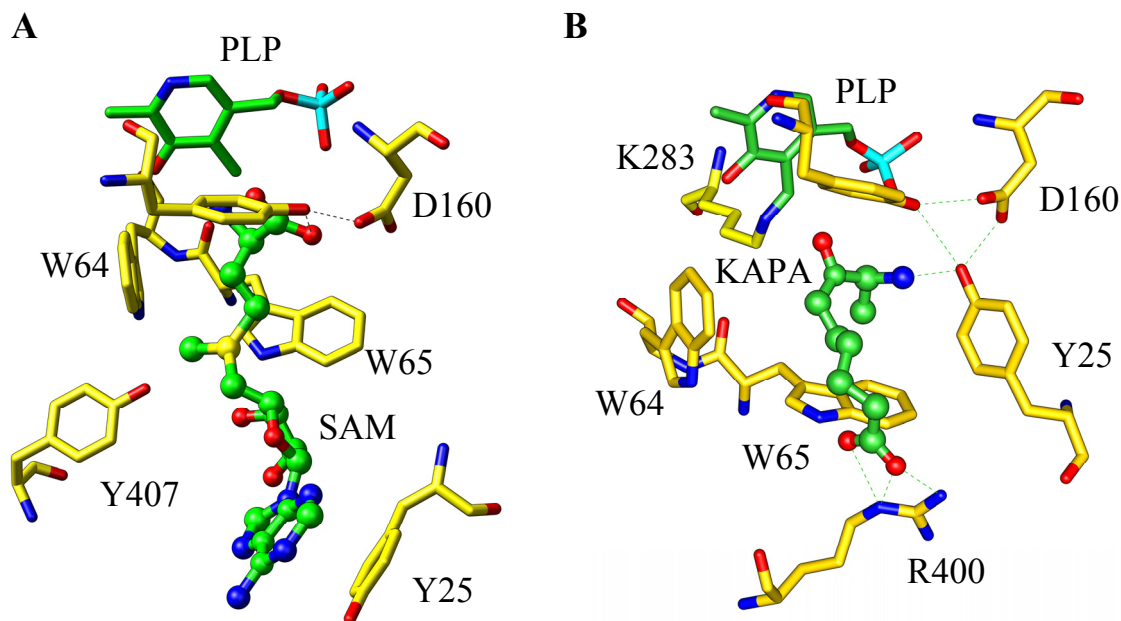


FIGURE 3.12. **Model for substrate recognition at the active site of *Mtb* DAPAS.**

(A) Interaction of amino group donor SAM with DAPAS: SAM is shown in green ball and stick form in the figure. The active site residues of DAPAS are shown in yellow color. The amino group of SAM is pointing towards PLP and the carboxy group would hydrogen bond with the network of Tyr157-OH and Asp160-OD2, and the Tyr25 has stacking interaction with the adenine ring of SAM. The side chains of Trp64 and Trp65 have changed their orientation to accommodate SAM in the active site. (B) Interaction of amino group acceptor KAPA with DAPAS: KAPA is shown in green ball and stick form in the figure. KAPA is binding at a similar position as SAM in the active site. In this case, PLP is covalently linked to Lys283. Tyr25 side chain has changed its orientation to form hydrogen bond with the amino group of KAPA. The carboxy group of KAPA forms salt bridge with Arg400.

Transaminases utilize various modes to identify the dual substrates in the same active site since they follow the Bi Bi ping-pong mechanism. These enzymes employ two different mechanisms or their combination to recognize the substrates at the active site. The induced fit model proposed by Koshland (157, 158) involves the conformational change of the active site on substrate binding and the Lock and Key model proposed by Fischer (159) suggests the complementarity of the active site and the substrate to each other (active site is a lock that all the substrates can fit in as a key). In the case of subclass I enzymes like aspartate aminotransferases (156, 160) and glutamine: phenylpyruvate aminotransferase (151), they follow the induced fit model, since there is a large movement of the small domain (13° rotation) involved in substrate recognition. Other transaminases such as lysine aminotransferase utilizes the glutamate switch (145) or tyrosine aminotransferase (161) and GABA aminotransferase (162) utilize the arginine switch. Acetylornithine aminotransferase, which uses acetylornithine and alpha-ketoglutarate as substrates, employs a combination of a lock and key and the induced fit mechanism for substrate recognition (161). In this regard, DAPAS appears to follow the lock and key mechanism in which certain residues in the active site change their side chain orientations to accommodate SAM and KAPA. There is no large conformational change in the form of domain movement as mentioned above. When SAM binds in the DAPAS active site, its amino group forms a Schiff's base with the PLP and the carboxy group of SAM hydrogen bonds to the Tyr157-OH: Asp160-OD2 hydrogen bond network. The conserved Tyr25 changes its orientation of the phenol group (approximately 180°) to provide a platform for stacking interactions with the

adenine ring. Following the release of SAM when KAPA binds in the same active site region to the PMP form of the enzyme, the side chain of Tyr25 will change its orientation (90°) to interact with the amino group of KAPA. This bond formation with Tyr25 helps in positioning KAPA in the active site for amino group acceptance from PMP. In response to this interaction, the guanidinium group of Arg400 also moves towards KAPA to form a salt bridge with the carboxy group of KAPA.

Kinetic properties of Y25A mutant form of Mtb DAPAS

Based on the analysis of aforementioned structural studies, we recognized a key role played by Tyr25 in substrate recognition in the *Mtb* DAPAS active site. In order to confirm the role of Tyr25 in substrate recognition, we mutated this residue at position 25 to Ala. Single-turnover half-reactions of the mutant *Mtb* DAPAS with the substrate SAM and the product DAPA were monitored spectrophotometrically. The Y25A mutant form of the DAPAS did not show any detectable enzyme activity with SAM at 335 nm, in the form of increase in absorbance, even at concentrations of up to 3 mM. There was also no change in the absorbance peak at 420 nm on SAM addition (Figure 3.13A). The mutant showed approximately 70% enzyme activity with product DAPA (0.5 mM) in comparison to the native protein (Figure 3.13B). Thus, this mutation study confirms the role of Tyr25 in substrate recognition.

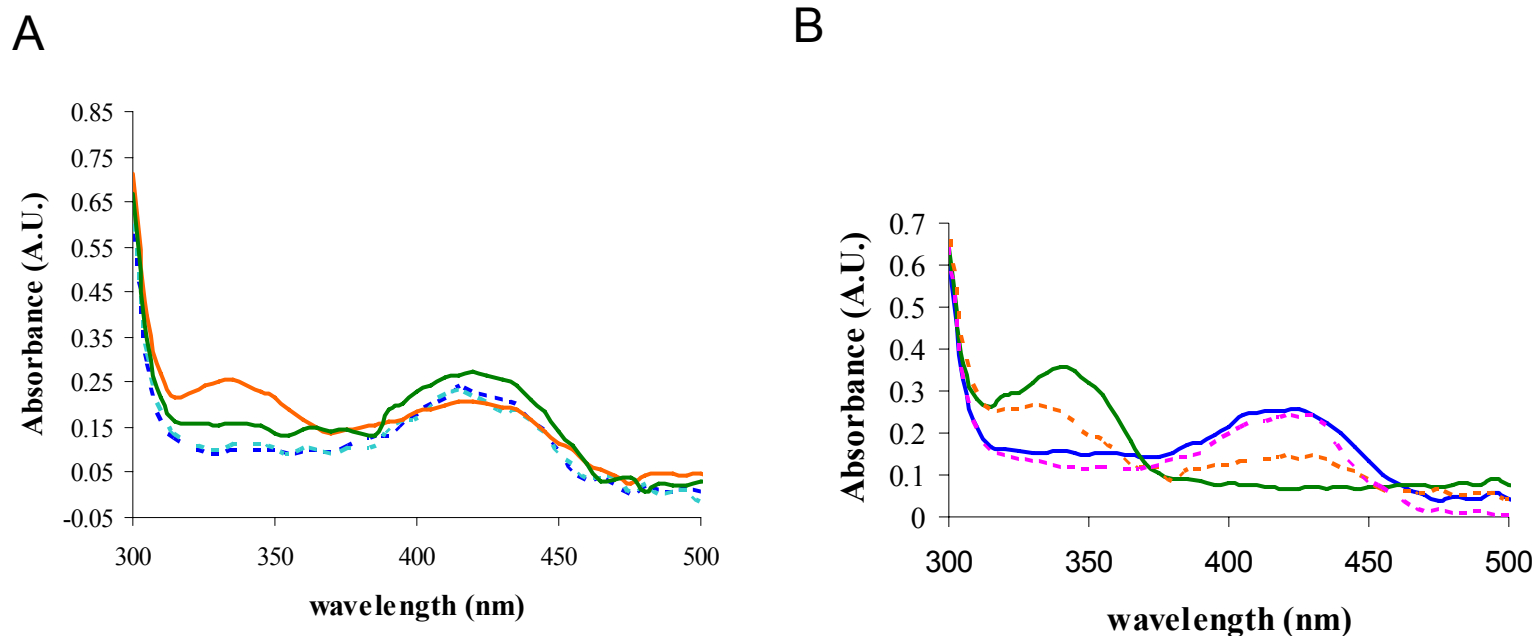


FIGURE 3.13. **Comparison of absorbance scans of native *Mtb* DAPAS and Y25A mutant.** The figure shows absorption spectra of DAPAS reaction with (A) SAM and (B) DAPA. In both the plots, the absorption spectra for native DAPAS is shown in solid line and for mutant DAPAS is shown as dotted line. As shown in figure A, there is no change in the absorption spectra of mutant form on reaction with SAM (blue) where as for native DAPAS, there is a decrease in absorption spectra at 420 nm with an increase in peak at 335 nm (orange). In figure B, in comparison to the native DAPAS reaction with DAPA (green), Y25A mutant shows approximately 70% enzyme activity with product DAPA (orange). The absorbance scans of native DAPAS and mutant enzymes are shown in solid blue and dotted magenta lines.

Variation in the DAPAS active site

The *Mtb* DAPAS structures indicated the crucial role played by Tyr25 in substrate binding at the active site, both in SAM binding with recognition of the adenine ring as well as in proper positioning of the KAPA for amino group acceptance. Sequence alignment of DAPAS from different organisms show that Tyr25 is a conserved residue in most organisms except in certain species of Bacillus such as *B. subtilis* and *B. cereus* as well as *Brucella melitensis*, *Leptospira interrogans*, and *Aquifex aquenox* where Phe replaces this Tyr residue in the protein sequence (Figure 3.14). Mutation studies on the replacement of this Tyr residue with Phe in *E. coli* DAPAS (132) have resulted in 60-fold less enzyme activity for the reverse second half-reaction with DAPA, but with no effect on the first half-reaction with SAM, since Phe can still stack against the adenine ring of SAM. We also noticed that on mutating Tyr25 to Ala, the *Mtb* enzyme totally loses its activity against SAM and retains only 70% activity with DAPA in the reversed second half-reaction. Thus, this raises the question of how substrate binding especially KAPA binding, takes place in these organisms to maintain the enzyme activity for their survival.

	25		157	160																	
<i>B. subtilis</i> DAPAS	L	F	T	Q	M	K	D	-----	K	E	A	Y	H	G	D	T	I	G	A	V	
<i>B. cereus</i> DAPAS	H	P	F	T	Q	M	K	D	-----	K	N	G	Y	H	G	D	T	I	G	A	V
<i>A. aeolicus</i> DAPAS	H	P	F	T	Q	M	K	V	-----	S	E	A	Y	H	G	D	T	V	G	A	V
<i>B. melitensis</i> DAPAS	H	P	F	T	Q	H	R	L	-----	H	S	-	Y	H	G	D	T	I	G	T	M
<i>E. coli</i> DAPAS	H	P	Y	T	S	M	T	S	-----	R	G	G	Y	H	G	D	T	F	L	A	M
<i>Mtb</i> DAPAS	H	P	Y	S	S	I	G	R	-----	R	N	G	Y	H	G	D	T	F	G	A	M

FIGURE 3.14. **Parts of the amino acid sequence alignment of *Mtb* and *E. coli* DAPAS along with other organisms showing changes at position 25.** (Numbers are based on *Mtb* DAPAS). All the DAPAS sequences are from Swiss-Prot and TrEMBL databases: *B. subtilis* (P53555), *B. cereus* (Q818W8), *A. aeolicus* (O66557), *B. melitensis* (Q8YBV8), *E. coli* (P12995), and *Mtb* (O06622).

Variation in KAPA binding in Bacillus subtilis DAPAS

To understand the mode of KAPA binding in the DAPAS active site of the aforementioned organisms, we have determined the crystal structures of DAPAS in *B. subtilis* as an apoprotein, PLP-bound complex, and PLP and KAPA bound ternary complex at resolutions of 2.2 Å, 1.9 Å and 2.3 Å respectively. Molecular replacement technique was used to determine the phase information for these structures. Sequence alignment of DAPAS from different organisms shows that *B. subtilis* DAPAS possess 32% sequence identity with *Mtb* DAPAS and 33% with *E. coli* DAPAS. The overall structure of *B. subtilis* DAPAS is quite similar to both *Mtb* as well as *E. coli* DAPAS with pair wise r.m.s.d. of ~1 Å for 325 C_α atoms. Similarly, all three DAPAS structures of *B. subtilis* superimpose very well with a pair wise r.m.s.d. of 0.41 Å for 414 C_α atoms, indicating similar structures. However, in the apo form as well as in the PLP bound form of this homodimeric enzyme, large parts of the active site (Lys143-Glu172) are found to be totally disordered in both the chains. Even binding of cofactor PLP did not help in

ordering this region of the active site. On the other hand, on KAPA binding, the electron density for the above-mentioned 29 residues is ordered in both the chains of this structure. This suggests that the active site in *B. subtilis* stabilizes only on substrate binding.

In contrast to the KAPA-bound *E. coli* DAPAS structure, there are very few local conformational changes visible in this nonproductive KAPA-bound *Bacillus* structure. This change is in the form of side chain movement of Arg403 to form a salt bridge with the carboxy group of KAPA (2.8 Å). However, in this case, the side chain of Arg residue moves backward by 2.2 Å in contrast to that observed in *E. coli* DAPAS. In *E. coli*, the side chain of Arg residue moves forward towards KAPA by 3 Å. Interestingly, the amino group of KAPA is hydrogen bonded to the OH group of Tyr146 (2.8 Å) and the main-chain oxygen of Gly315' (2.65 Å). The keto group of KAPA points towards the internal-aldimine. Thus, this kind of interaction still maintains the hydrogen bond network of Tyr146-OH: Asp149-OD2 in absence of a Tyr residue at position 17 (*B. subtilis* numbering). It is worth mentioning here that Tyr146 and Asp149 are part of the loop region which is disordered in the apoform and PLP-bound form of the enzyme. In comparison to the KAPA-bound *E. coli* structure, there is a shift of 1 Å in the position of KAPA binding in the active site towards Tyr146 for hydrogen bond formation (Figure 3.15). In this way, the substrate KAPA stabilizes the loop region of the active site.

In the *E. coli* Y17F mutant, an additional water molecule at the active site fulfils the role of the hydroxyl group of Tyr17. These findings may also indicate the importance of this hydrogen bond network of Tyr25:Tyr157:Asp160 (*Mtb* numbering) in the maintenance of the active site structure. It may also be noted here that there are two consecutive Trp residues in the *Mtb* (W64 and W65) and the *E. coli* (W52 and W53) DAPAS active site, which change their conformation on substrate binding, in place of Val53, Trp54 in *B. subtilis*. The empty space created by this substitution of Trp to Val53 in *Bacillus* active site is filled by Leu82' (Gly in *Mtb* and *E. coli*) from the neighboring subunit. Thus, this KAPA-bound structure answers the question of how the enzyme maintains its activity in the absence of Tyr25 but leaves us with another question of why DAPAS from *Bacillus* and other above-mentioned organisms possess Phe rather than Tyr at the active site. This question can be answered by interpreting the phylogenetic tree analysis of DAPAS from different organisms which reveals that DAPAS from *B. subtilis* evolved long before DAPAS from *E. coli* and *Mtb* DAPAS.

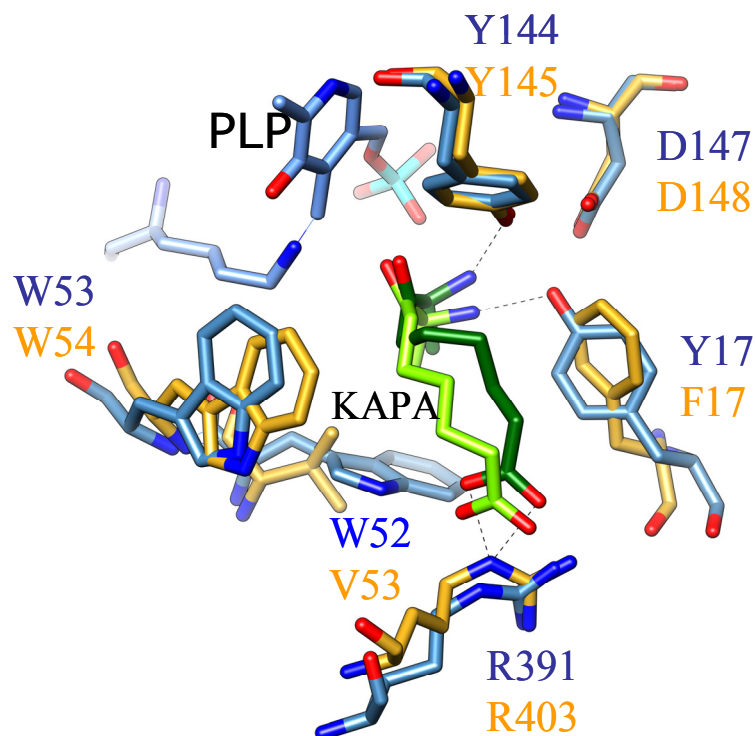


FIGURE 3.15. Binding mode of KAPA at *B. subtilis* DAPAS active site.

Superimposition of KAPA-bound *E. coli* DAPAS (blue) upon the respective *B. subtilis* DAPAS (orange) shows the binding pattern of KAPA in absence of Tyr at position 17 of *Bacillus* DAPAS. As shown in this figure, PLP is covalently linked to Lys, thus, providing an unproductive form of the enzyme. KAPA in this figure is shown in stick form, which is colored light green in *E. coli* and dark green in *B. subtilis* DAPAS. Arg403 forms salt bridge with the carboxy group of KAPA. Val53 and Trp54 have hydrophobic interactions with KAPA.

Search for Mtb DAPAS inhibitors

To search for potent inhibitors, there are mainly three different ways: using substrate mimics/suicide substrates/transition state mimics, high throughput screening or virtual screening.

DAPAS is a PLP-dependent enzyme and PLP acts as a cofactor for many other enzymes, which catalyze diverse reactions. Even though the reactions are different, they share certain similarities in the enzyme mechanism i.e. aldimine formation with PLP. Hence, most of the inhibitors of these enzymes utilize a mechanism-based inhibition such as substrate mimics, activated nucleophiles or electrophiles as well as reversible competitive inhibitors.

The known natural antibiotic, ampicillin (ACM) irreversibly inhibits *Mtb* DAPAS with $K_i = 12 \pm 2 \mu\text{M}$ (136). It has been found to be most effective against *Mtb* in cell culture (60) in comparison to other bacteria and yeast. Ampicillin inhibition of any other PLP-dependent enzyme has not been reported so far. A model of ACM binding at the *Mtb* DAPAS active site, which was generated based on the cis-ampicillin-bound structure of *E. coli* DAPAS (PDB code-1MLY) (135), suggests a mechanism of suicide substrate inhibition. In this structure, ACM occupies the same place in the active site as the substrates (KAPA or SAM). In this process of ACM binding, PLP would be tilted at an angle of 17° in comparison to the PLP complexed DAPAS structure to form a covalent bond with the 4-amino group of ACM and results in an irreversible aromatic adduct.

This aromatic adduct would be further stabilized by hydrophobic interactions with Trp64, Trp65, and Phe402 at the active site. All the other local changes in the form of side chain movements observed on KAPA binding at the active site are also seen with ACM binding. The 1-amino group of the ACM does not seem to interact with any part of the protein and thus, may not be essential for its inhibitory effect. Similar to KAPA binding, the side chain of Arg400 would move closer (3 Å) to hydrogen bond with the carboxy group of the antibiotic (2.8 Å), thus, providing specificity (Figure 3.16).

Although the natural antibiotic, ACM was found to be ineffective *in vivo* (60), we can still use this compound as a template to design new inhibitors by modifying its chemical characteristics to improve its biological properties. Infact, attempts have been made recently to synthesize variations of ACM like 4-(4c-aminocyclohexa-2,5-dien-1-yl) propanol which also irreversibly inhibited *Mtb* DAPAS ($K_i = 20 \pm 2 \mu\text{M}$) and had a good whole cell activity in *M. smegmatis* (MIC = 10 $\mu\text{g/ml}$) (136).

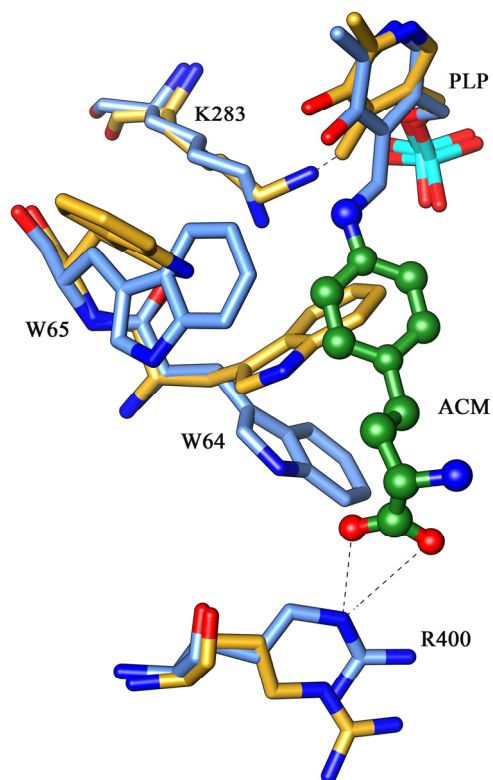
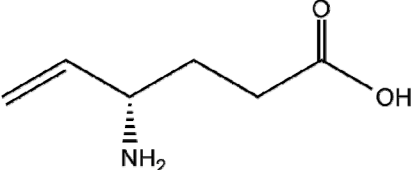
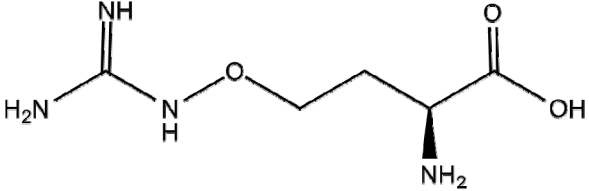
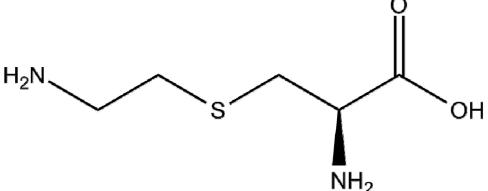
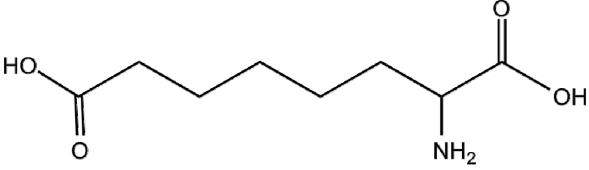
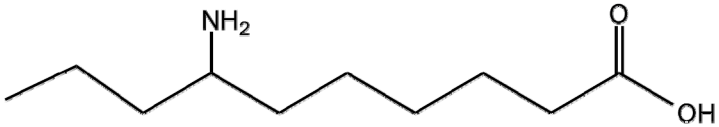


FIGURE 3.16. Changes observed on ampicillin binding at the *Mtb* DAPAS active site. This figure shows a superimposition of *Mtb* DAPAS structure upon ACM-bound *Mtb* DAPAS model. Irreversible inhibitor ampicillin is modeled in the active site of *Mtb* DAPAS based on the ACM-bound *E. coli* DAPAS structure (PDB code-1MLY). ACM is shown in ball and stick form (green) having covalent linkage with C4' atom of PLP shown in stick form. The *Mtb* DAPAS and the ACM-bound structure are colored gold and blue, respectively. This structure shows the conformational changes at the active site on ACM binding. Also seen is Lys283, which is present next to the ACM aromatic adduct; its side chain is oriented away from the PLP. PLP is also tilted at certain angle to covalently link with ACM.

Cycloserine and gabaculine effectively inhibit other PLP enzymes alanine racemase (163), and GABA aminotransferase (164) and ornithine aminotransferase (165), respectively. Even though these inhibitors follow a same inhibition mechanism as of ACM, they were reported not to inhibit *Mtb* DAPAS (136). In this study, we also found that vigabatrin, which is another potent inhibitor of GABA aminotransferase (147) and forms irreversible aromatic adduct analogous to ACM, does not inhibit the *Mtb* DAPAS even at concentrations of 1 mM. This insensitivity of *Mtb* DAPAS may be due to these compounds not being long enough to form both, covalent bond with PLP and ionic bond with Arg400, at the active site. Also, there are no other Arg residues present in the active site lining that can provide hydrogen bond partners for the carboxy group of these inhibitors or provide any specificity to these compounds in the *Mtb* DAPAS active site. Interestingly, several other compounds, which possess a chemical structure analogous to substrate KAPA, also did not show any inhibition (Table 7). This may suggest that *Mtb* DAPAS enzyme is very specific for its own substrates and possesses scaffolds at the active site which can distinguish between different substrates. For the inhibitor studies of *Mtb* DAPAS, the single-turnover half-reaction of PLP form of DAPAS (5 μ M) with SAM was monitored at 335 nm.

TABLE 7
 Compounds having chemical structure similar to the substrate KAPA

COMPOUND	STRUCTURE	Mol. Wt. (g/mol)
Vigabatrin		129.16
Canavanine		176.2
Aminoethylcysteine		200.7
DL-alpha-aminosuberic acid		189.21
Chembridge ID 5190771		187

Virtual screening of Mtb DAPAS

In this study, after mapping the active site of *Mtb* DAPAS, we proceeded to conduct a structure-based virtual screening of this enzyme to search for potent inhibitors. Virtual screening (VS) or *in silico* screening is the selection of the compounds based on their predicted binding, using different computer algorithms in contrast to High throughput screening. The aim here is to identify the compounds that have high likelihood of binding at the active site, thus, minimizing the need for screening large number of compounds experimentally by performing assays. VS has been successfully carried out against other *Mtb* enzymes such as dihydrofolate reductase (DHFR), enoyl-acyl carrier protein reductase (InhA), isocitrate lyase (ICL) etc in search of potential inhibitors. VS mostly involves four steps: 1) Preparation of the compound library and protein template 2) docking of the compounds 3) selection of compounds from the list of top hits, and 4) Validation of these compounds.

In this study, a docking algorithm, DOCK 6.0 (UCSF) was used for virtual screening in collaboration with Dr. T. Ioerger (Dept. of Computer Science, Texas A&M University). “Hits” are defined as the compounds having high docking scores from the VS runs. DOCK scoring function calculates two components of interaction energy: electrostatic and van der Waals energy. This computationally intensive task was performed utilizing the computer resources on the TAMU campus. These docking jobs were run on a Grid cluster (connected to over 1000 computers) using MP Agent software from United Devices, Austin TX.

We utilized the active site of the KAPA-bound DAPAS model as a template for docking. All the non-protein molecules, e.g. water or substrate were removed. Ideal situation is to have both the template and the ligands flexible. However, due to the limitations in computing power, the templates generally act as a rigid body and the ligands (compounds) are flexible entities. In this case, the cofactor PLP was retained in the model to prevent any compound from occupying its position as well as to improve the number of hits due to affinity with PLP. Virtual screening was carried out against 2 million drug-like compounds contained in the commercially available ZINC database (166). These compounds in the database are mostly present in multiple conformations and are based on Lipinski's rule of 5 (167) which states that poor absorption or permeation of compounds is more likely to happen when

- 1) There are more than 5 hydrogen bond donors,
- 2) There are more than 10 hydrogen bond acceptors,
- 3) The molecular weight is greater than 500,
- 4) The Log P (octanol-water partition coefficient) is over 5 and
- 5) Classes of compounds that are substrates for biological transporters

Compounds from the list of top hits of virtual screening were selected for inhibitor studies based on (1) Visual inspection (using ViewDock option in Chimera) to evaluate the mode of binding of the compounds and prevent redundancy (2) The DOCK scores (3) Easy availability, and (4) Economical pricing (Figure 3.17).

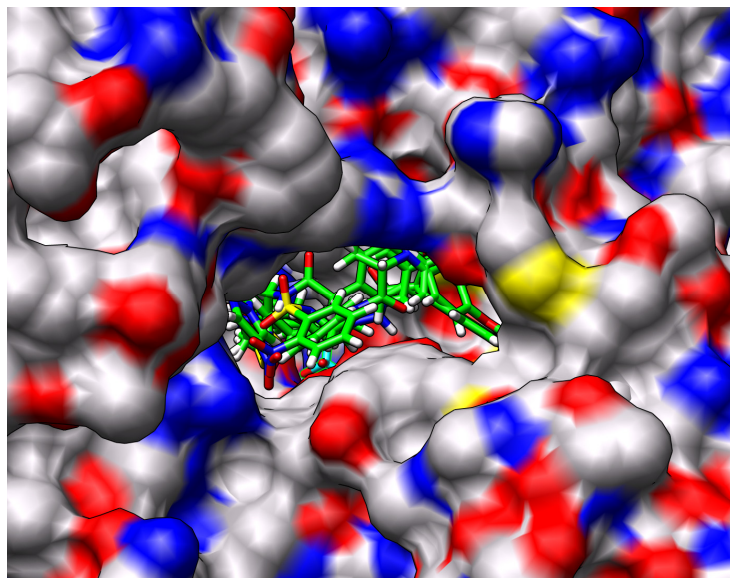


FIGURE 3.17. **Overlay of docked compounds at the *Mtb* DAPAS active site.** This figure shows the active site cavity of *Mtb* DAPAS with the overlay of docked compounds from the virtual screening.

The compounds chosen for inhibitor studies were purchased from Chembridge as presented in Table 8. In this table, molecular weight, calculated logP values and energy scores obtained from the virtual screening using DOCK algorithm is provided.

TABLE 8
Chembridge compounds used for virtual screening

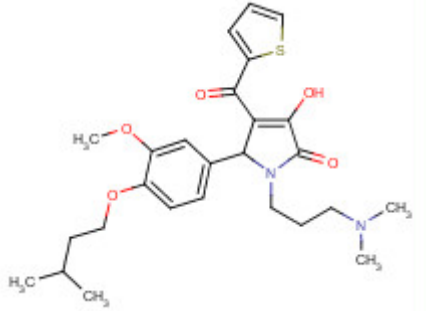
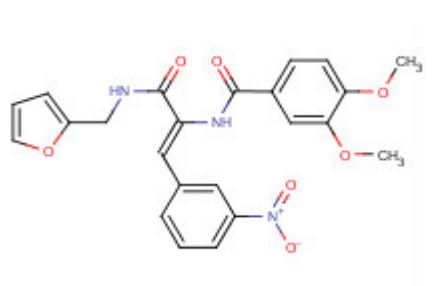
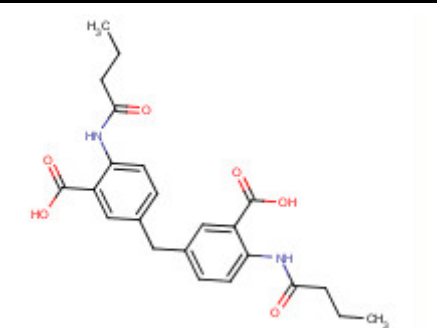
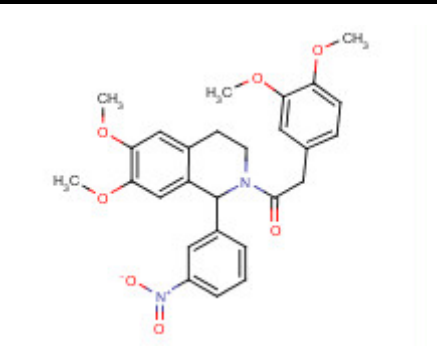
Compound ID	Structure	Mol. Wt.	clogP	Energy Score
5881639		469	2.90	-74.39
5904890		451	3.07	-73.16
6094188		426	5.69	-76.59
6384549		493	3.02	-73.63

TABLE 8 (Continued)

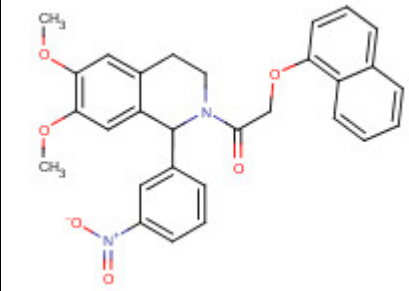
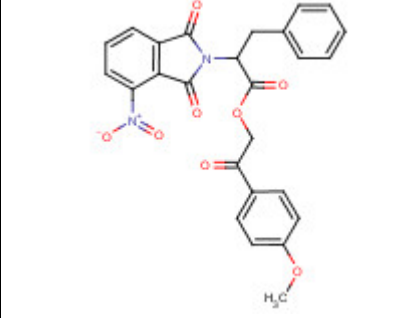
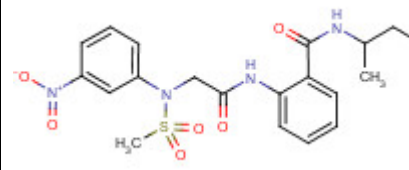
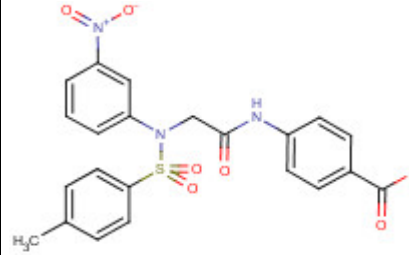
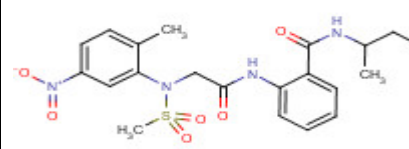
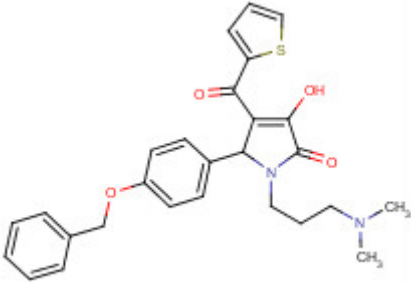
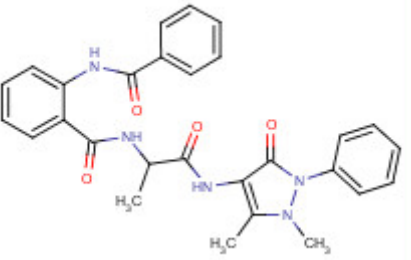
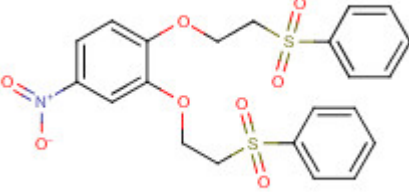
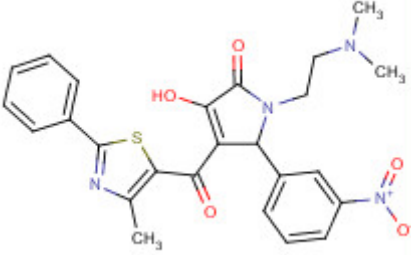
Compound ID	Structure	Mol. Wt.	clogP	Energy Score
6410766		499	4.95	-74.96
6531042		488	4.29	-74.74
7586634		448	3.00	-73.79
7713211		483	4.80	-73.58
7714429		463	3.16	-74.0

TABLE 8 (Continued)

Compound ID	Structure	Mol. Wt.	clogP	Energy Score
7876301		474	4.32	-72.92
7908679		486	2.77	-73.16
6939388		483	4.99	-72.81
6656927		466	3.92	-73.19
7102000		487	5.16	-75.52

TABLE 8 (Continued)

Compound ID	Structure	Mol. Wt.	clogP	Energy Score
7103108		477	5.20	-74.68
6498331		498	4.56	-73.16
6238769		492	2.87	-76.45
7746089		493	4.71	-74.43

Validation of the top candidates is generally carried out using enzyme or binding assays. Due to the presence of unique cell wall structure in *Mtb*, cell permeability tests in the form of whole cell assays in either *M. smegmatis* (mc²155 strain) or other mycobacterium strains are also performed. This combination of virtual screening and enzyme assays is recognized to be an effective way to find inhibitors in the low micromolar range (168). In this study, quick screening of the top hits for an IC₅₀ near 10 μM was carried out by inhibition studies of the single-turnover half-reaction of the *Mtb* DAPAS enzyme with SAM at 335 nm using spectrophotometric methods. Unfortunately, we did not yet find any useful inhibitor in this study. In this regard, more screening of the compounds from the VS results needs to be done.

Future work will involve docking the commercially available compound library against the sinefungin-bound DAPAS active site as a VS template and varying different parameters of the docking algorithm to get better hits. Once we get some compounds with good IC₅₀ values from enzyme assays, their whole cell activity will be tested against different mycobacterium strains. Further, crystallization of the *Mtb* DAPAS with these potent inhibitors, which are active in both assays, will be carried out. Either soaking or cocrystallization will be attempted to understand the mode of inhibitor binding and its mechanism of inhibition. Any successful candidate can be further improved upon by synthetic chemistry and QSAR to develop a more potent analog.

CHAPTER IV

STRUCTURAL INSIGHTS OF DETHIOBIOTIN SYNTHETASE IN

Mycobacterium tuberculosis

Background

Dethiobiotin synthetase (DTBS; EC 6.3.3.3) catalyzes the penultimate step in the biotin biosynthesis: carboxylation of 7, 8-diaminopelargonic acid (DAPA), leading to the formation of the ureido ring of dethiobiotin (DTB). This amidoligase reaction requires Mg^{2+} -ATP (Figure 4.1).

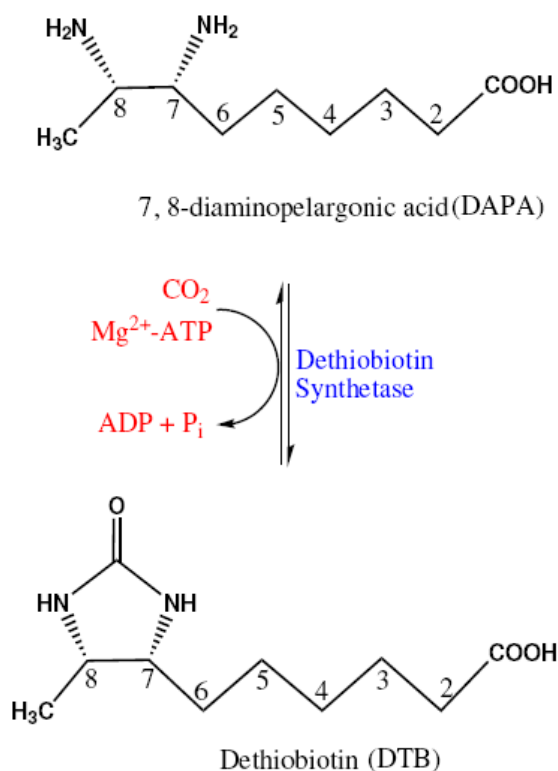


FIGURE 4.1. Reaction catalyzed by *Mtb* DTBS

This enzyme has been extensively studied in *E. coli* (169-173). The *Mtb* DTBS has a poor sequence identity with the *E. coli* DTBS (28%) (Figure 4.2). In 1970, Krell and Eisenberg reported that CO₂ and not HCO₃⁻ is the substrate for this enzyme in *E. coli*, based on their experiments with carbonic anhydrase (169). Structurally and mechanistically, DTBS also differs from other carboxylases like biotin dependent carboxylase (174) and ribulose-1, 5 biphosphate carboxylase (175). The information gained from the *E. coli* DTBS structures were used to develop herbicides in plants such as *A. thaliana* (176). In 2007, another DTBS structure from *H. pylori* (PDB code-2QMO, sequence identity with *Mtb* DTBS – 18%) has been solved by Minor *et al.* However, no further biochemical characterization of the *H. pylori* enzyme is available.

```

E.coli DTBS      SKRYFVTGTDTEVGKTVASCALLQAAKAAAGYRTAGYKPVASGSEKTP*EGLRNSDALALQ-
M.tb DTBS       LTILVVTGTGTGVGKTVVCAALASAAARQAGIDVAVCKPVQTGTARGDDDLAEVGRLAGV-
H.pylori_DTBS   --MLFISATNENAGKTCARLLAQYCNA*CGVKTILLKPIETGVNDAINHSSDAHLFLQDN
1.....10.....20.....30.....40.....50.....

E.coli_DTBS     RNSLQLDYATVNPYTFAEPTS*PHIISAQEGRP*IES--LVMSAGLRAL*EQQADWV*LV*EGA
M.tb DTBS       TQLAGLARYP-----QPMAPAAAAEHAGMALPAR-DQIVRLIADLDRPGRITLVEGA
H.pylori_DTBS   RLLDRSLTLKDI*SFYRYHKVSA*PLIAQQEEDPNAPIDT*DLTQRLHNFTKTYDLVIVEGA
61.....70.....80.....90.....100.....110.....

E.coli DTBS     GCWFTPLSDT-FTFADWVTQEQLPVILVVGVKLG*GINHAMLTAQVIQHAGLTLAGWVAND
M.tb DTBS       GGLLVELAEPGVTLR*VAVDVAAAAALVVVTADL*GTLNHTKLTLEALAAQ*QVSCAGLVIGS
H.pylori_DTBS   GGLCV*PITL*EEN-MLL*FALK*KLKAK*MLLI*SHDN*LGL*INDCLL*ND*FLLKSH*QLDYKIAIN--
121.....130.....140.....150.....160.....170.....

E.coli DTBS     VT-PPGKRHAEYMTTL*TR--MIPAPLLGEIPWLAENPENAA*TKYINLALL-----
M.tb DTBS       WPD*PPGLVAASNRSA*ARIAMVRAAL*PAGAASLDAGDFAAMS*AAAFDRN*WVAGLVG
H.pylori_DTBS   ---LKG*NNTAFHSISL*PYIELFNTRSN*NP*IVIFQQSLKVLMS*FALK-----
181.....190.....200.....210.....220.....230.....

```

FIGURE 4.2. Alignment of *E. coli*, *Mtb* and *H. pylori* DTBS sequences. The conserved, identical, and similar residues between the different DTBS sequences are colored green, yellow, and cyan, respectively. The important residues at the active site are marked by black lines and red asterisks. This alignment is prepared using ClustalW (79).

As mentioned in previous chapters, biotin acts as a cofactor in several carboxylases, including acetyl-CoA carboxylase, which is involved in the initial stages of fatty acid synthesis. This is important for *Mtb* survival during the chronic phase of infection (37). Genetic studies such as microarray analysis conducted on *Mtb* H37Rv strain under nutrient starved conditions report that the *bioD* gene, which codes for DTBS, gets upregulated in 96 hours of starvation (58). This study was performed to simulate the host environment during infection. Also, due to the absence of a homolog of this enzyme in humans, *Mtb* DTBS would be an excellent choice for inhibitor studies.

In order to better understand the DTBS enzyme in *Mtb*, we have carried out the preliminary biochemical characterization of this enzyme using a coupled assay. We further determined the three-dimensional structures of DTBS apoenzyme as well as enzyme complexes with substrate and products using X-ray crystallography. With the help of these structures, we have compared the overall structure of *Mtb* DTBS with the known *E. coli* and *H. pylori* DTBS structures. We have further determined the residues involved in substrate/product binding at the active site. In addition, we have been able to derive a possible mechanism for the enzyme action, which will be further helpful in inhibitor studies.

Methods

Reagents

The restriction enzymes for Rv1570 cloning were purchased from New England Biolabs. All the reagents for purification procedures were from Sigma. KAPA and DTB used in the structural studies were purchased from Toronto Research Chemicals Inc.,

Canada and Sigma, USA, respectively. DAPA was chemically synthesized by Dr. Joel Freundlich of Princeton University using the protocol of Gibson *et al.* (137).

Cloning, protein expression and purification of Mtb DTBS

The *bioD* gene (Rv1570) from the *Mtb* H37Rv genome was PCR amplified and ligated into the pET28b vector with the TEV cleavage site using *NdeI* and *HindIII* restriction sites. The *NdeI* site contained the ATG start codon and the stop codon (TCA) preceded the *HindIII* site. The DTBS (D47A) mutant clone was constructed by using the QuikChange site-directed mutagenesis kit (Stratagene) with pET28b-Rv1570 as the template. The primers used in this study were:

For the native DTBS protein:

Forward primer: 5' AGAGAAGCATATGACGATCCTGGTCGTCA 3'

Reverse primer: 5' CCCAAGCTTTCAGCCGACCAGCCCGGCAACCCA 3'

For the D47A mutant:

Forward Primer: 5' CGCCCGCGGTGCCGACGACCTCG 3'

Reverse Primer: 5' CGAGGTCGTCGGCACCGCGGGCG 3'

The plasmids so obtained were checked for the presence of the insert using restriction digestion as well as sequencing (Gene technologies Lab, TAMU). The pET28b-Rv1570 and the D47A mutant clones were then transformed into the *E. coli* overexpression host BL21 (DE3). The transformed cell culture was grown in LB media with kanamycin (50 µg/ml) at 37 °C till the A_{600} reached 0.8. It was then induced by adding IPTG at a final concentration of 0.5 mM and grown overnight at 25 °C. There was no protein overexpression seen in the cultures grown at temperatures lower than 25 °C. The cells

were harvested by centrifuging at 3000 rpm for 30 min and the cell pellet was resuspended in a buffer containing 50 mM Tris buffer (pH 7.5), 500 mM NaCl, 1 mM PMSF and 20 μ g/ml DNase. The cells were lysed using a French press at 11,000 psi and the cell debris was removed by centrifuging the cell lysate at 15,000 rpm for 1 hour. The supernatant was then loaded on a HisTrapTM Ni column (GE Healthcare) which was pre-equilibrated with 50 mM Tris buffer (pH 7.5), 500 mM NaCl and washed with a buffer containing 25 mM imidazole till the absorbance was zero. The protein was then eluted using an imidazole gradient (50–400 mM). The protein is 98% pure as judged by a 4–12% gradient SDS-PAGE gel. However, the yield of the protein was very low (0.5 mg/L). Hence, the presence of DTBS protein in the fractions as well as in the crystals was confirmed by western blot technique. The expression of the protein could not be improved by trying different cell strains, fusion tags or growth media. The fractions containing the pure protein were pooled and dialyzed overnight against a buffer containing 25 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA and 1 mM DTT.

Spectroscopic studies of Mtb DTBS

Coupled assays for native *Mtb* DTBS and its D47A mutant form with pyruvate kinase/lactate dehydrogenase (40 units) (177) were carried out to determine the kinetic parameters of both the substrates ATP and DAPA while keeping the other substrate at a saturating concentration. The reaction buffer contained 50 mM Tris (pH 7.5), 150 mM NaCl, 15 mM NaHCO₃, 5 mM MgCl₂, 250 μ M phosphoenol pyruvate, 80 μ M NADH. In this 500 μ l coupled reaction, ADP formation was measured by monitoring the decrease in NADH absorbance at 340 nm ($\epsilon = 6.22$ /mM/cm). These assays were carried

out on a Cary 100 spectrophotometer (Varian) at room temperature. The DTBS and the D47A mutant concentrations were calculated from their absorbance at 280 nm using extinction coefficients 12840 /M/cm and 14120 /M/cm, respectively (138).

Crystallization of Mtb DTBS

Initially, microcrystals of *Mtb* DTBS were found in sitting drop vapor-diffusion plates (96 well intelliplates from Hampton Research) at a protein concentration of 2 mg/ml. Crystallization screens with 1:1 ratio of protein to precipitant were setup using a Hydra Plus One dispenser (Art Robbins Instruments). In this process, different precipitant conditions from Hampton Research (Crystal Screen, Crystal Screen II, PEG/Ion and SaltRx screen) and Emerald Biosystems (WIZARD I, II) were used. There were several different conditions in which *Mtb* DTBS crystals were obtained. These conditions were further optimized with protein concentrated to 10 mg/ml, at a ratio of 2:2 in a hanging drop plate at 18 °C. To obtain the protein at concentration of 10 mg/ml, iterative rounds of protein purification was carried out. The dataset for the apoenzyme was collected using the crystal from 1 M sodium citrate, 0.1 M imidazole (pH 7.5) (WIZARD I # 36) (Figure 4.3).

Data collection and processing

A 1.85 Å high resolution dataset of the DTBS crystal was collected in-house using a Rigaku Raxis IV++ image plate detector coupled to a Rigaku (MM007HF) X-ray generator with a copper rotating anode. Using the *E. coli* DTBS structure (PDB code-1BYI, sequence identity-28%) as the search model, structure determination by molecular replacement was not successful. Expression of selenomethionine-substituted protein was

tried in an attempt to solve the structure using MAD phasing. However, we did not get any DTBS protein expression with selenomethionine. As a next logical step, in order to obtain phase information using the anomalous or the isomorphous signal, native DTBS crystals were soaked with different heavy metals (mercury, platinum, iridium and iodine) and datasets were collected at the home source and the synchrotron (mercury dataset). However, all these attempts were also unsuccessful. Ultimately, we were able to obtain the phase information from samarium-derivatized crystals. In this regard, native DTBS crystals were soaked in 50 mM samarium chloride for 1 hour and a dataset at 2.6 Å resolution with 1° oscillation per image was collected using the home source at Cu K α wavelength (1.54 Å).

DTBS complexed with DAPA, DAPA analog-KAPA and dethiobiotin + ADP datasets were obtained by soaking DTBS apocrystals in 10 mM DAPA, 10 mM KAPA and 10 mM dethiobiotin + 10 mM ADP + 10 mM MgCl₂ overnight, respectively.

All the crystals were cryoprotected with paratone oil and datasets were collected at 120 °K. All the datasets were indexed, integrated and scaled using the HKL2000 suite (90).

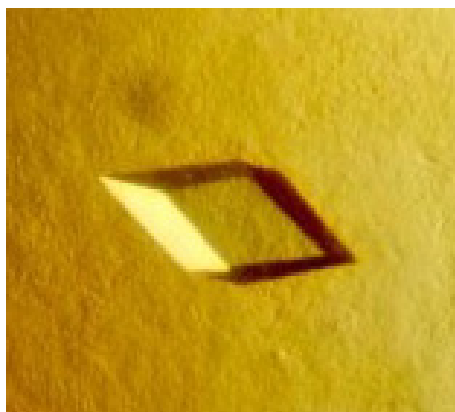


FIGURE 4.3. **Crystal of *Mtb* DTBS.** These rhombohedral shaped crystals of *Mtb* DTBS (50 μM x 75 μM x 100 μM) were obtained in 1 M sodium citrate and 0.1 M imidazole (pH 7.5) using hanging drop vapor-diffusion technique.

Structure determination

For the *Mtb* DTBS structure determination, four sites for samarium were found using direct method- SHELXD (98). After refinement of these samarium sites and initial phasing with AutoSHARP (99), an overall anomalous phasing power of 1.4 and figures of merit of 0.4 and 0.14 for acentric and centric reflections, respectively, were obtained over the full 2.6 Å resolution range. Phases were further improved with a figure of merit of ~0.85 by solvent flattening with DM (100). This phase was extended to a higher resolution using the 1.85 Å dataset of the DTBS apostructure. DAPA, KAPA and dethiobiotin-ADP complexed *Mtb* DTBS structures were further obtained by molecular replacement of the DTBS apostructure as a search model against these datasets.

Iterative rounds of model building were performed with the Shake&wARP (178) unbiased electron density maps using graphics software such as the Xfit module of Xtalview (102) and Coot (141). One round of simulated annealed refinement at 4000 °K

using CNS (104) was performed to improve the geometry and remove bias from the model. Further refinements were carried out using REFMAC from the CCP4 suite (103). Water molecules were added to the models after Rfactor and R_{free} values below 30% were achieved and inspected manually during refinement. All the structure figures have been prepared using Chimera (108).

Results and discussion

Use of samarium for structure determination of Mtb DTBS

Simple molecular replacement using *E. coli* DTBS structure (PDB code-1BYI) as a search model was not successful. MIR/SIR/SAD using heavy metals such as Hg, Pt, Ir, I at concentrations of 5-10 mM was also unsuccessful. Protein structures have been successfully solved using the anomalous behavior of Se or S at their respective absorption wavelength. Se-SAD was not possible as there was no expression of *Mtb* DTBS with selenomethionine. Finally, the crystal structure of *Mtb* DTBS was solved using the single anomalous dispersion (SAD) method with the Lanthanide metal-samarium (Sm) at home source. Since Sm can easily diffuse in the protein and occupy the sites of the Mg ion in DTBS, it can be used at a higher concentration (100 mM) than the abovementioned heavy metals without causing any damage to the crystal. In this study, we obtained the best data at the concentration of 50 mM. At Cu $K\alpha$ wavelength, samarium has large anomalous signal ($f''=12.1 e^-$). Taking advantage of this large anomalous signal, we were able to obtain an excellent electron density map.

The initial phase at 2.6 Å, obtained using the SAD method, was extended to a higher resolution of 1.85 Å and the final DTBS structure was refined to an Rfactor of 22% and an R_{free} of 26%. The stereochemistry of the final model was excellent as no outliers were seen in the disallowed region in the Ramachandran plot as per PROCHECK evaluation (106) (Figure 4.4). The data collection, refinement and geometry statistics are provided in Table 9.

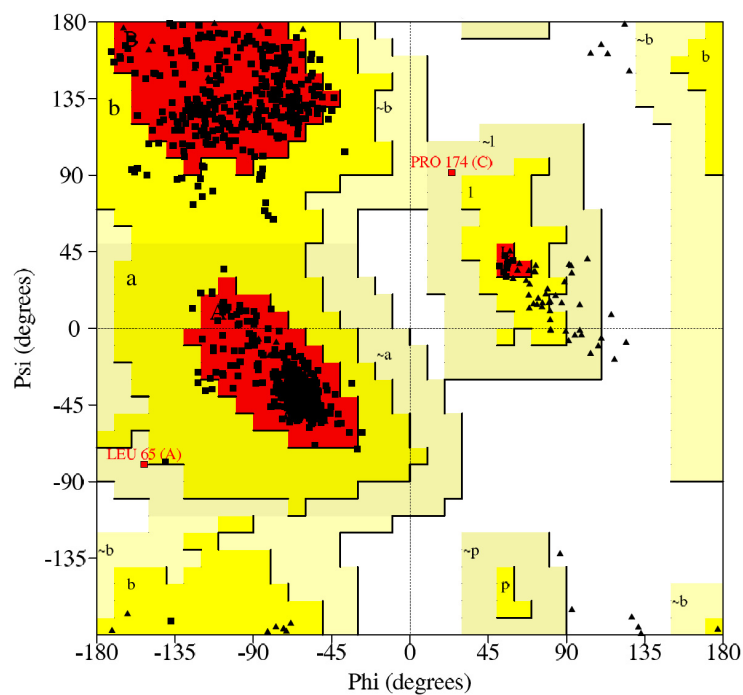


FIGURE 4.4. **Ramachandran plot for the *Mtb* DTBS apoenzyme.** Torsion angles (ϕ and ψ) of the amino acid residues of DTBS are plotted against each other. This plot is mostly used to evaluate the quality of the model.

▲ Glycine ■ Non-glycine and non-proline amino acids

TABLE 9
Data collection, refinement and geometry statistics of *Mtb* DTBS

	DTBS-Sm	Native DTBS	DTBS + DAPA	DTBS + KAPA	DTBS+DTB+ADP
<i>Data collection</i>					
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	a=55.13Å b=104.45Å c=151.13Å α=β=γ=90°	a=55.13Å b=104.45Å c=151.13Å α=β=γ=90°	a=55.13Å b=104.45Å c=151.13Å α=β=γ=90°	a=55.13Å b=104.45Å c=151.13Å α=β=γ=90°	a=55.13Å, b=104.45Å c=151.13Å α=β=γ=90°
Molecules per asymmetric unit	4	4	4	4	4
Wavelength (Å)	1.54	1.54	1.54	1.54	1.54
Resolution range (Å)	35-2.6	35-1.85	31.5-2.11	30-2.14	35- 2.3
Highest resolution bin (Å)	2.7-2.6	1.95-1.85	2.22-2.11	2.24-2.14	2.3-2.2
Observed reflections	392314	1149410	249621	618111	84334
Unique reflections	52020	121230	47727	30270	25289
Completeness (%) ^a	100 (99.9)	99.5 (96.1)	93.9 (86.6)	99.9 (99.8)	99.5 (100)
Average redundancy ^a	7.5 (6.9)	18.7 (10.8)	5.2 (4.6)	20.4 (16.4)	3.3 (3.3)
I/σ(I) ^a	18.7 (2.5)	19.3 (1.6)	31.7 (11.8)	39.8 (5.2)	20.1 (4.8)
R _{sym} ^{a,b}	0.15 (0.90)	0.16 (0.85)	0.04 (0.12)	0.08 (0.60)	0.05 (0.21)
<i>Refinement statistics (REFMAC)</i>					
Free R value (5%)		25	24	27	26
R value (%) ^c		21	20	22	21
No. of protein residues		904	904	904	904
No. of water molecules		321	208	195	180
r.m.s.d. bond length, (Å)		0.011	0.010	0.016	0.012
r.m.s.d. bond angles, (°)		1.3	1.3	1.8	1.6
<i>Ramachandran plot (PROCHECK)</i>					
Most favored region (%)		725(94.9)	708(94.3)	710(94.5)	696 (92.4)
Additional allowed regions (%)		37(4.8)	38(5.1)	38(5.1)	50 (6.6)
Generously allowed regions (%)		1 (0.1)	2 (0.3)	2 (0.3)	5 (0.7)
Disallowed regions (%)		0	0	0	0

^a Values in parenthesis for the highest resolution bin

^b $R_{sym} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

^c $R = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factors

Overall structure of Mtb DTBS

In the *Mtb* DTBS structure, there are four molecules in the asymmetric unit. Each pair of molecules (chains A&B and C&D) forms a dimer. The electron density for all four of the polypeptide chains is well defined in the apostructure.

Each subunit of DTBS (226 amino acids) consists of one globular domain that has seven parallel β strands (β 3- β 2- β 4- β 1- β 5- β 6- β 7) interconnected with α helices. Being an ATP binding protein, the structure contains the classical P loop motif (also called Walker A motif), Gly8-X-Gly10-X-Gly12-Val13-Gly14-Lys15-Thr16, which binds the phosphate group of the nucleotide (179, 180). This motif is in a loop region between the strand β 1 and helix α 1 at the N-terminus of the protein. There is also a Walker B motif at the strand β 4 (Leu104-Thr105-Leu106-Val107-Glu108) of this structure. The N- and C-termini of this structure are in proximity to each other (9 Å between the C $_{\alpha}$ atoms of Met1 and Gly226) on the same side of the protein but do not participate in dimer formation. This dimer interface occupies a buried surface area of approximately 1321 Å² (7.5% of total surface area of the dimer). The 37.2 Å elongated dimer interface mainly involves the helices α 3 (Pro74-Ala81), α 6 (Gly144'-Gln159') and α 7 (Leu177'-Ile189'), the loop (Gly109-Leu124) between the strand β 5 and helix α 5, and the P loop (Thr9-Gly14) between the strand β 1 and helix α 1 of both the subunits (' indicates the residues in the adjacent subunit).

There are six helices present at the dimer interface, the inner helices, $\alpha 5$ and $\alpha 5'$, of both subunits and the outer helices, $\alpha 6$ & $\alpha 3'$ and $\alpha 3$ & $\alpha 6'$, which are arranged perpendicular to each other (Figure 4.5). As the dimer forms at a non-crystallographic two-fold axis, all the contacts between the residues at the interface occur twice. The interface of the helices $\alpha 5$ and $\alpha 5'$ involves a stacking interaction of the imidazole rings of His148 and His148'. The interface of helices $\alpha 3$ and $\alpha 6'$ makes hydrogen bond contacts between Arg188-NH2 and Glu79'-OE2 (3.5 Å) and hydrophobic interactions with Met72, Leu177 and Val178. Other interactions at the dimer interface include hydrogen bonding with the Thr9 main chain oxygen of the P loop with His148-NE2 (2.7 Å), and the Asn147-OD1 of one monomer with the main chain amide nitrogen of Leu114' of the other (2.9 Å) as well as a salt bridge between Arg125 and Glu154' (2.85 Å and 3.0 Å). Superimposition of chain A on chain B of the dimer does not show much variation in the overall structure (r.m.s.d. = 0.62 Å).

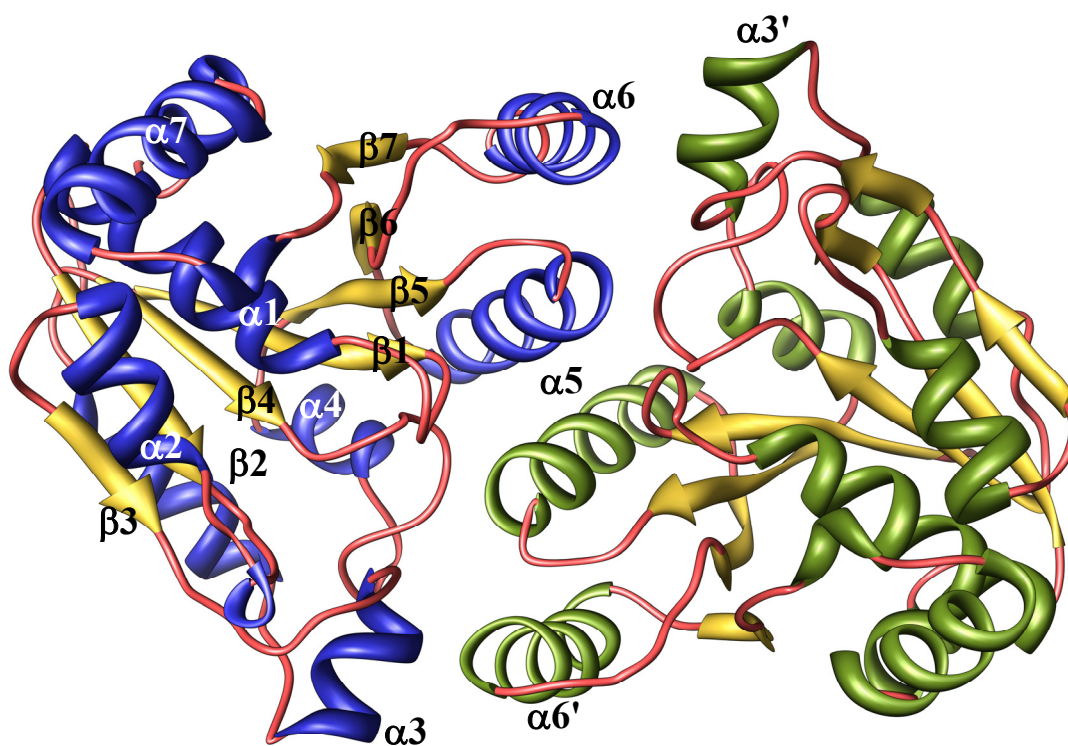


Figure 4.5. **Ribbon diagram representation of *Mtb* DTBS dimer.** This figure shows the dimer interface of *Mtb* DTBS. The β -strands and the loops are colored yellow and pink in both the subunits. The helices are colored green in one subunit and blue in the other.

Other structural neighbors of Mtb DTBS

A search for other structural homologs of *Mtb* DTBS using VAST (123) and DALI (25, 26) algorithms resulted in numerous hits of nucleotide-binding proteins; prominent among them was *E. coli* CTP synthetase (PDB code-1S1M) with a score of 17.8 as analyzed by VAST. CTP synthetase regulates the intracellular CTP levels in bacteria and produces CTP from UTP and glutamine (181). The N-terminus of this protein consists of an amidoligase domain which is similar in structure to DTBS (the r.m.s.d. is 1.13 Å for 74 C_α atoms) with differences in the residues of the P loop (GGVVSSLGKG). Another protein showing a high DALI score of 16.2 is the bacterial cell division regulator MinD from *A. fulgidus* (PDB code-1HYQ) with an r.m.s.d. of 1.1 Å for 77 C_α atoms (182). MinD activity also requires ATP hydrolysis. Unlike DTBS, MinD is a monomer with a slight variation in the P-loop (GKGGTGKT). Furthermore, *Mtb* DTBS also shows structural homology with the nitrogenase iron protein (NIP) from *C. pasteurianum* (PDB code-1CP2) with a DALI score of 14.7 (r.m.s.d. for 75 C_α atoms is 1.14 Å). All seven parallel β strands of both the proteins superimpose quite well at the core of the structure. However, except for the helix α1, all other NIP helices differ in their orientation. NIP, which is involved in nitrogen fixation, couples ATP hydrolysis with electron transfer (183). It requires a FeS cluster for its activity which is coordinated by Cys residues (97 and 132), not present in DTBS.

Kinetic characterization of *Mtb* DTBS

DTBS catalyzes an amidoligase reaction to form dethiobiotin by using DAPA and Mg^{2+} -ATP. We carried out a coupled reaction of *Mtb* DTBS with pyruvate kinase/lactate dehydrogenase (177) (Figure 4.6) to study the formation of ADP by monitoring the decrease in the absorbance of NADH at 340 nm. The kinetic parameters, K_m and K_{cat} values for ATP were found to be $29 \mu\text{M}$ and $2.9 \times 10^{-4}/\text{s}$ and for DAPA were $2 \mu\text{M}$ and $1.5 \times 10^{-2}/\text{s}$, keeping the other substrates at saturation. This study shows that the K_m for ATP in *Mtb* DTBS ($29 \mu\text{M}$) is quite high compared to that of the *E. coli* DTBS ($\sim 7 \mu\text{M}$) (173). This might explain the low occupancy of ATP analogs in the active site of *Mtb* DTBS structures as discussed later in this chapter.

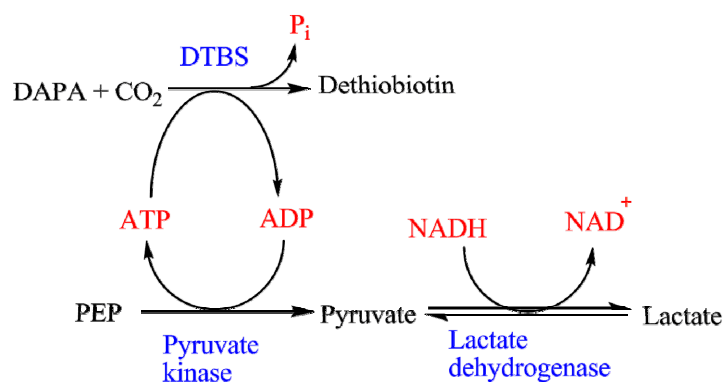


FIGURE 4.6. **Coupled assay of *Mtb* DTBS with pyruvate kinase/lactate dehydrogenase.**

Interaction of *Mtb* DTBS with substrate DAPA, and its analog KAPA

In order to determine the binding mode of the substrate at the active site, we have soaked the *Mtb* DTBS crystals with substrate DAPA. The 2.2 \AA crystal structure of *Mtb*

DTBS complexed with DAPA reveals a long distinct electron density at the dimer interface, which fits DAPA very well. Interestingly, there is some additional electron density attached to the N7 atom of DAPA which fits the carbon dioxide molecule very well (Figure 4.7). The carboxy group of the carbamate occupies the position of the water molecule (W11 in chain A) at the active site of the apostructure and its carboxy oxygen atoms are within hydrogen bond distance to the catalytic Lys37-NZ (3 Å and 3.4 Å) and the Thr41 main-chain amide nitrogen (2.8 Å). The N7 atom of DAPA also has hydrogen bond links with Thr41-OG1 (2.9 Å) and the N8 atom with Thr11-OG1 through an ordered water molecule (W221 in chain A). The long aliphatic chain of this nonstandard amino acid spans the active site pocket at the dimer interface. This pocket formation is contributed by a stretch of residues from the helix α_6 (Leu143'-Asn147') from the adjacent subunit and loops Pro71-Pro74 and Ala110-Gly112. One of the C1-carboxy oxygen atoms of DAPA coordinates with the main-chain amide of Leu146' (2.8 Å) and an ordered water molecule (2.6 Å). The other C1-carboxy oxygen atom is within hydrogen bonding distance of the main-chain amide of Asn147' (2.6 Å) in the helix α_6 and of the α_7 helix's Asn182'-OD1 through a neighboring water molecule (3 Å). This water molecule is not present in the active site of the *Mtb* DTBS apostructure. Due to the presence of this water molecule at the active site, there is a peptide flip at Gly111 of the loop region 110-112. This phenomenon of peptide flipping is not observed in DAPA-bound *E. coli* DTBS, even though both enzymes share the same sequence at this region 106-112 (LVEGAGG). Furthermore, in *E. coli* DTBS, the Tyr187' on the helix α_6 changes its side chain orientation (approximately 90°) upon DAPA binding to hydrogen

bond with the carboxy oxygen of DAPA. This Tyr187' is replaced by the Asn182' in *Mtb* DTBS (Figure 4.8).

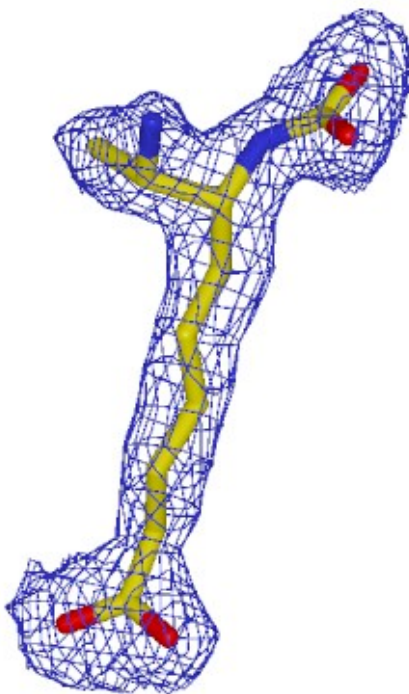


FIGURE 4.7. **Electron density for DAPA-carbamate.** This figure shows the electron density of Shake-warp omit map for DAPA-carbamate at 3σ level, visible in the *Mtb* DTBS active site. This also confirms that DTBS in *Mtb* is specific for carbamate at N7 position of DAPA.

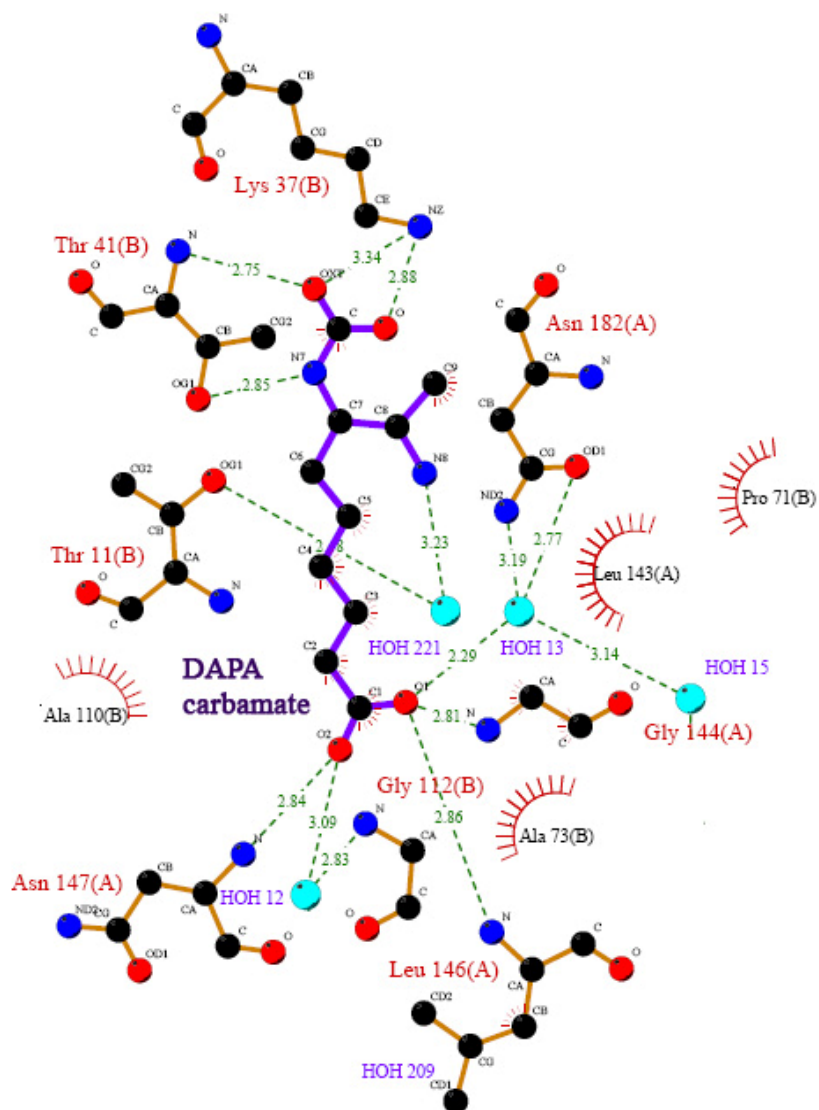


FIGURE 4.8. Mode of DAPA-carbamate binding in the *Mtb* DTBS active site.

Residues interacting with the substrate DAPA-carbamate in the DTBS active site are shown in this figure. This plot has been prepared using Ligplot (122). The protein residues are colored orange and the ligand DAPA-carbamate is colored purple.

Hydrogen bonds are shown in green dotted lines. Hydrophobic interactions are shown in brick-red sunbursts. Water molecules are shown in cyan. Atoms are colored by type (C- black, O- red, N- blue).

Similar interactions are also observed for KAPA at the DTBS active site (Figure 4.9). However, as evident from the weak electron density for KAPA and its high B factor, the occupancy of KAPA (0.5) is much lower than DAPA (1.0) at the active site of all four chains of the asymmetric unit. This low occupancy of KAPA might suggest its low affinity as a substrate for this enzyme. In this KAPA-bound DTBS structure, there is also a presence of sulfate ion from the crystallization solution, bound at the P loop (Gly8-Thr16). This sulfate ion mimics the β phosphate of the nucleotide. The 7-keto-oxygen atom of KAPA is oriented towards this sulfate ion with an ordered water molecule (W24 in chain A) occupying a similar position to the γ -phosphate of the nucleotide at the active site (3.2 Å). The N8 atom of KAPA forms a hydrogen bond with Asp47-OD2 (2.9 Å) at the active site. Thus, KAPA also occupies a similar position to that of DAPA in the hydrophobic pocket of the enzyme, away from the P loop.

The DAPA-bound DTBS structure illustrates that the first substrate DAPA can bind to the enzyme in the absence of the second substrate ATP. It also demonstrates the enzyme's specificity for the N7-DAPA carbamates as carbamylation of DAPA might have occurred at either N7 or N8 position in crystallization drops from atmospheric CO₂. Furthermore, the diamino group of the substrate is bound to one subunit and the carboxy group is in contact with the other adjacent subunit. Thus, the substrate is anchored to both the subunits at the active site.

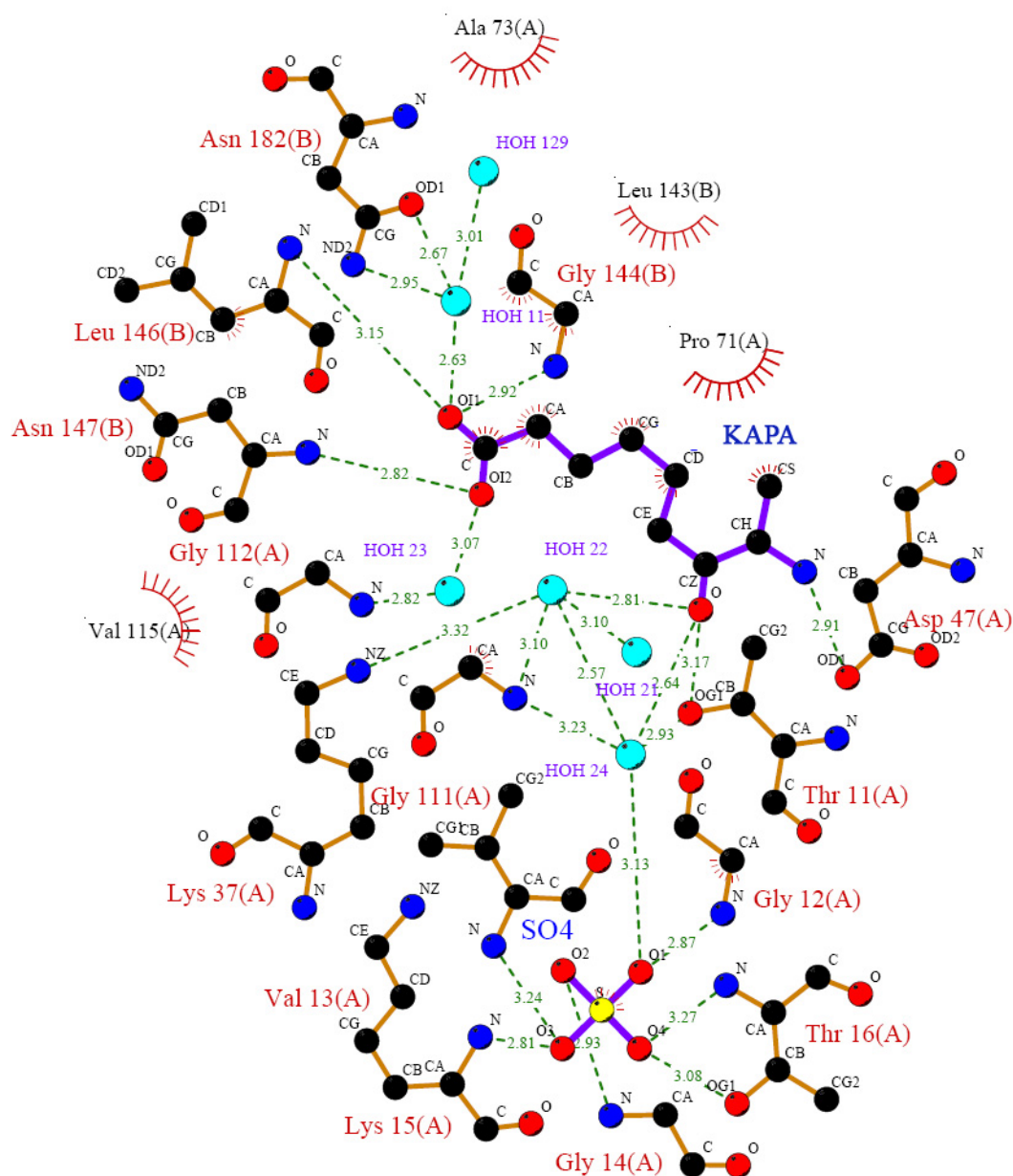


FIGURE 4.9. Schematic diagram of KAPA interactions in the *Mtb* DTBS active site. The protein residues are colored in orange. KAPA and the sulfate ion are colored purple. Water molecules are shown in cyan. Hydrogen bonds are shown in green dotted lines. Hydrophobic interactions are shown in brick-red sunbursts. Atoms are colored by type (C- black, O- red, N- blue, S- yellow). This plot has been prepared using Ligplot (122).

ATP/ADP binding site

To understand the interaction of nucleotides with the enzyme and the resulting conformational changes on its binding, the crystal structures of *Mtb* DTBS complexed with ATP analogs (AMP-PCP and ATP γ -S) as well as ADP were determined. However, only partial electron density for these entities was visible at the active site. Phosphate ions were clearly visible at the P loop but electron density for the rest of the molecules was less well defined. Hence, to pinpoint the residues involved in the interaction with the nucleotide at the active site, the *E. coli* DTBS structure complexed with ADP (PDB code-1DAM) has been overlaid on this *Mtb* structure. In a surface representation of the *Mtb* DTBS active site, it appears that this 17 Å long nucleotide binding region is completely exposed to solvent, resulting in disordered electron density for ATP/ADP. This might be due to a variation in the conformation of the loop regions Gln168-Val176 and Leu196-Asp204 near the adenine ring-binding pocket in comparison to the *E. coli* enzyme. A unique feature of this loop region is the presence of multiple proline residues (Pro172, Pro174, and Pro175) which would make the loop less flexible. It also appears that the only interaction of the protein with the nucleotide's ribose sugar would occur between its O2' atom and the carboxy group of Glu52 probably via a water molecule. This Glu residue is present on the helix α 2 of *Mtb* DTBS which is positioned 5.3 Å closer to the active site than in its *E. coli* counterpart. The corresponding residue in *E. coli* DTBS, which makes direct hydrogen bond contact with the O2' atom of the nucleotide's ribose sugar is Glu211 present in the loop region (Trp209-Glu212) between the strand β 7 and the helix α 9. This nucleotide interaction in *E. coli* DTBS orders the

loop region which is disordered in its apostructure (PDB code-1DTS). This loop stabilization upon nucleotide binding has also been reported in other enzymes such as GTPases (184) and adenylosuccinate synthetase (185). Intriguingly, clear electron density for this loop (Leu196-Asp204) in *Mtb* DTBS can be seen in the unliganded structure as well. This might be due to the presence of the long helix $\alpha 7$ just behind the loop region, not present in *E. coli* DTBS. Furthermore, the nitrogen atoms of the adenine ring have hydrogen bond contacts with this loop. The N1 atom would hydrogen bond with the main-chain amide of Gly199 from the loop region (2.7 Å) and the N7 atom would bond with the main-chain oxygen of Gly169 (2.7 Å). It is evident from this study that the solvent exposed nucleotide-binding site is located adjacent to the hydrophobic DAPA binding pocket.

DAPA and ATP occupy distinct binding sites in the protein similar to the *E. coli* DTBS enzyme. It appears that if the γ phosphate of ATP is bound to one monomer it would be too far (23.8 Å) from the carboxy group of the DAPA-carbamate bound to the same monomer to perform the catalytic reaction. Therefore, it is anticipated that the enzyme reaction will involve the ATP bound to one monomer with the DAPA from the other, with both subunits forming the active site. The DTBS dimer still possesses two active sites, which are approximately 13 Å apart at the interface, with the entrances facing the same side of the dimer.

In addition, based on the superimposition of the apostructure on the substrate-bound structures (r.m.s.d. of 0.3 Å for 226 C_{α} atoms), it appears that there are no significant changes in the conformation of the enzyme upon substrate binding. This is

also evident from the soaking experiments, since the crystals did not crack in the presence of substrate. There are local conformational changes in the form of movement of the loop regions Thr9-Gly14 (2.4 Å), Gly109-Leu113 (1 Å) and Thr140'- Gly144' (0.8 Å) in response to substrate binding. Most of these changes result in increased compactness of the active site for better substrate interaction.

Interaction of Mtb DTBS with the products DTB and ADP

To obtain a clear picture of the mechanism of the enzyme reaction and understand the binding mode of the products at the active site, we have tried to capture the ternary complex by soaking the DTBS apoenzyme crystal with dethiobiotin, Mg^{2+} -ADP and phosphate. The 2.3 Å electron density map of this DTBS complex clearly reveals the position of the product DTB just opposite the P loop in the active site. The C1-carboxyl group of DTB buried inside the active site pocket forms hydrogen bond contacts with the main-chain amide nitrogen atoms of Leu146' (2.8 Å) and Asn147' (2.6 Å), and with Asn182' via a water molecule, in the same manner as with the substrate DAPA. There is a peptide flip at Gly111 due to the presence of this water molecule, not seen in the *E. coli* DTB complexed structure. As a result, the main-chain oxygen of Gly111 forms a hydrogen bond with Thr7-OG1 via a water molecule. The aliphatic chain of DTB spans the space between the dimer interface, making hydrophobic contacts with Pro74, Ala73, Gly111, and Val115. The N7 atom of the ureido ring of DTB is hydrogen bonded to the Asp47-OD1 (3.3 Å) and in this process, the salt bridge between Arg45 and Asp47 is broken at the active site. The N8 atom of DTB is hydrogen bonded to the Thr11-OG1 of the P loop (3.4 Å) (Figure 4.10).

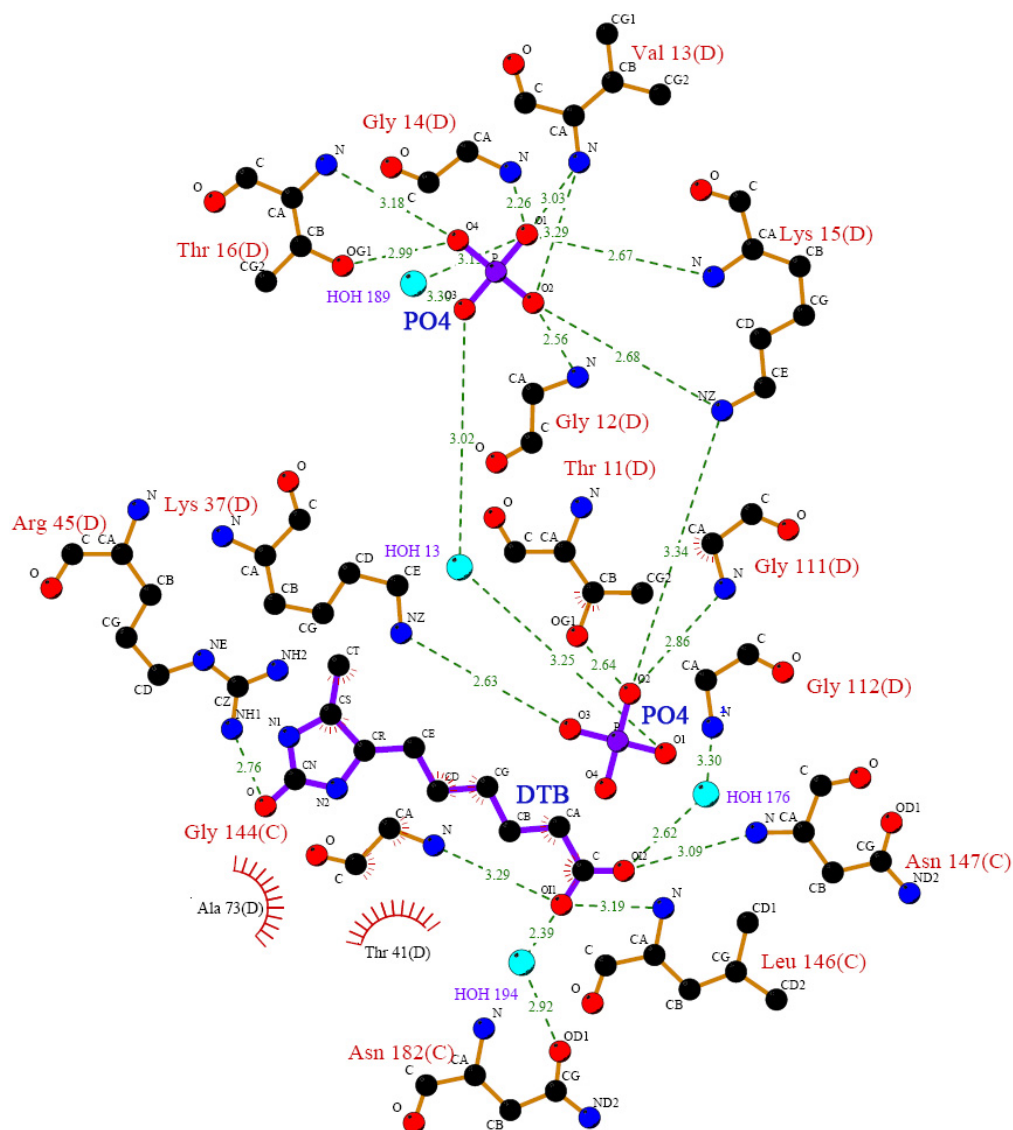


FIGURE 4.10. Interactions of product DTB and phosphate ions in the *Mtb* DTBS active site. The protein residues are colored orange, and DTB and phosphate ions are colored purple. Water molecules are in cyan. Hydrogen bonds are shown in green dotted lines. Hydrophobic interactions are shown in brick-red sunbursts. Atoms are colored by type (C- black, O- red, N- blue, P- purple). This plot has been prepared using Ligplot (122).

Comparison of the DTB-bound *Mtb* structure to the DAPA-bound structure, it is clear that the substrate DAPA and the product dethiobiotin both occupy the same position in the active site. There is no major change in the conformation of residues lining the active site during the substrate or the product binding except in the orientation of the ligand to fit in this hydrophobic pocket. The C1-carboxy end of the ligand anchors to the other subunit of the dimer and the diamino end of the substrate participates in catalysis. In the DTB-bound *E. coli* DTBS structure, the aliphatic chain of the product was curved compared to DAPA in order to fit into the active site. It was thought to be due to the formation of the ureido ring in DTB (172). Contrary to this, a similar curving of the aliphatic chain of DTB is not observed in the *Mtb* structure, probably due to the hydrogen bond formation of the N7-ureido ring of DTB with Asp47 rather than Thr41. It is worth mentioning here that at this position in *Mtb* DTBS, the active site is lined by a short loop (Gly42-Asp47), whereas in *E. coli* DTBS there is an extension of two antiparallel β strands (Gly42-Ser53). Interestingly, there are three contiguous Asp residues (47-49) in this region of the *Mtb* active site, Asp47 hydrogen bonds with Arg45 of the active site, as well as, with the substrate/product, Asp48 hydrogen bonds with Arg67 (3.3 Å) and Asp49-OD1 forms hydrogen bond with Lys37-NZ (2.9 Å) or a Mg^{2+} ion.

In this ternary complex of DTBS, there is presence of two phosphate ions at the nucleotide-binding site in place of the product ADP. One phosphate ion is bound at the P loop (Gly8-Thr16) which is similar to the β phosphate of ADP and a second one is present between the P loop bound phosphate and dethiobiotin, occupying the γ

phosphate's position. This mimics the position of the intermediate state DTB-Pi-ADP complex at the active site, just before the release of the products (Figure 4.11). In response to the phosphate bound at the P loop, all the main-chain amide N atoms in this region (Thr11-Lys15) are oriented inward, creating a positively charged environment that causes the phosphate group of the nucleotide to bind tightly. Also, the Thr11 side chain changes its rotamer 90° in response to hydrogen bond formation with the γ phosphate-O4, with a movement of 1 Å towards the active site. These phosphate ions (β and γ) are kept in position by conserved lysine residues (15 and 37) and the magnesium ion.

Thus, these DTBS structures also indicate that the P loop is involved in binding both substrates at the active site, providing a reason for the longer than usual peptide sequence in this region, as compared to other enzymes with a P loop.

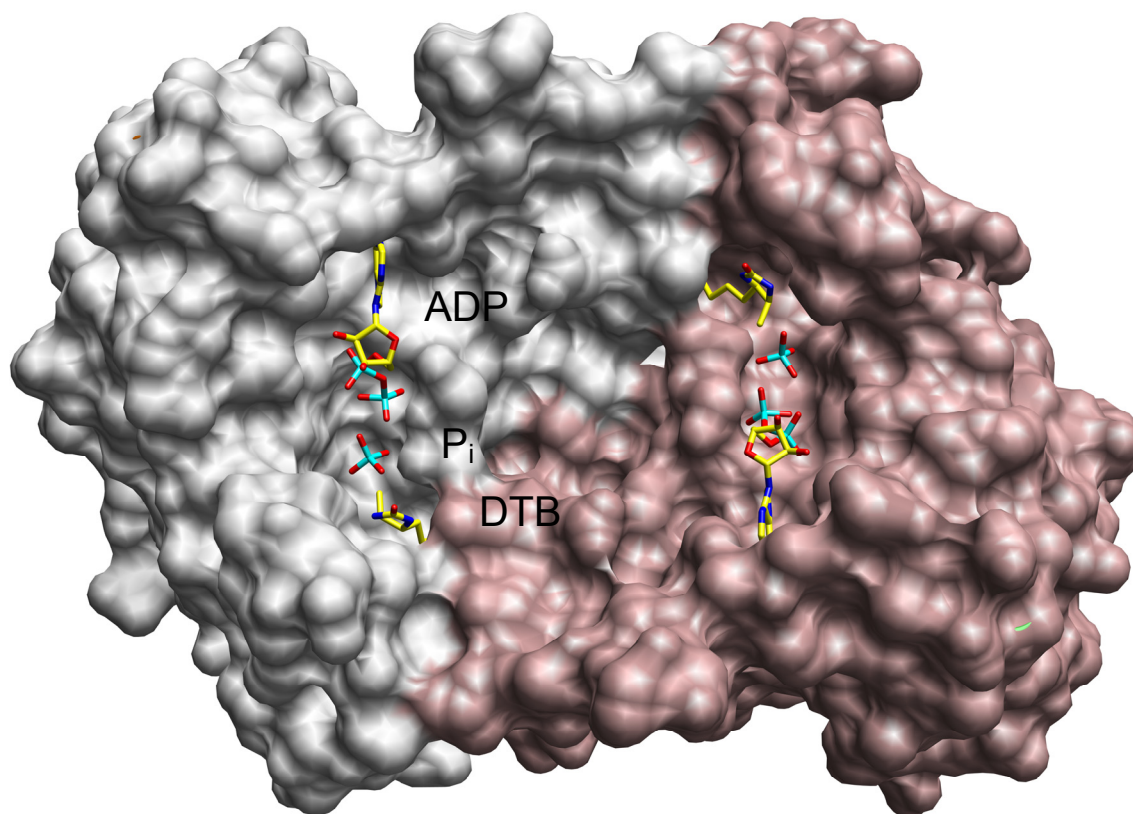


FIGURE 4.11. **Surface representation of the active sites of *Mtb* DTBS.** This figure represents the dimeric form of dethiobiotin synthase in *Mtb*, showing the products bound at both the active sites, 13 Å apart. Subunit A and B are colored grey and rose, respectively. At the active site, DTB is positioned inside the hydrophobic pocket, anchored to the other subunit and ADP is modeled in the open cleft, completely exposed to the solvent. The product inorganic phosphate is positioned between DTB and ADP.

Like other nucleotide-binding proteins, *Mtb* DTBS uses divalent metal ion such as Mg^{2+} for stabilizing the phosphate groups (mostly β and γ) of the nucleotides. This structure displays an octahedral arrangement of the Mg^{2+} ion with six ligands. These ligands are the two oxygen atoms- one from each of the phosphate groups β and γ , the side chains of residues Thr16, Asp49, and Glu108 and a water molecule (figure 4.12). These interacting amino acid residues are also conserved in *E. coli* and almost conserved in *H. pylori* DTBS with Asp49 being replaced with Asn46. However, *Mtb* DTBS lacks the second Mg^{2+} binding site as reported in *E. coli* DTBS structure (186).

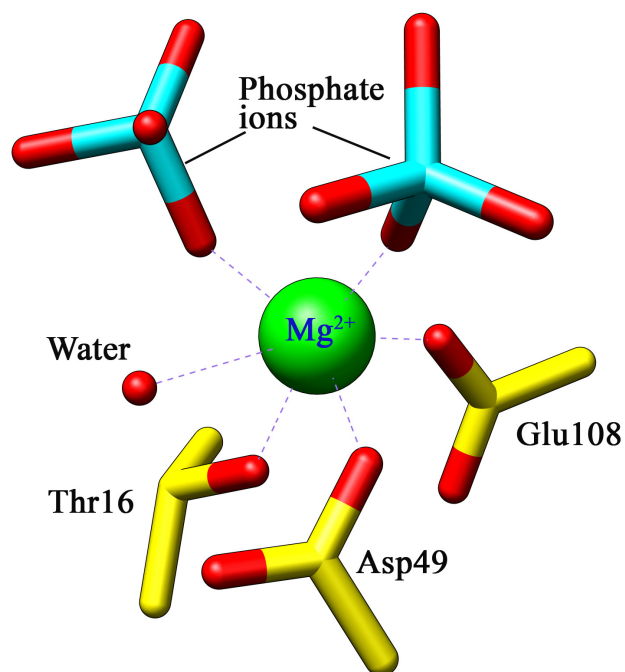


FIGURE 4.12. Mg^{2+} ion coordination in the *Mtb* DTBS active site. The figure shows the octahedral arrangement of Mg^{2+} ion (green) with the active site residues Thr16, Asp49 and Glu108, and the phosphate groups (β and γ) of the nucleotide, and a water molecule.

Comparison with other DTBS structures

Superimposition of the monomer of *E. coli* DTBS (PDB code–1BYI) on *Mtb* DTBS results in an r.m.s.d. of approximately 1.79 Å which might be due to the low sequence identity of *Mtb* DTBS (28%) with *E. coli* DTBS. Comparison of both structures reveals significant structural differences, which explains the failure of molecular replacement for the *Mtb* DTBS structure determination using the *E. coli* DTBS structure as the search model (Figure 4.13). Interestingly, when the *Mtb* DTBS structure was overlaid with a recently solved DTBS structure from *H. pylori* (PDB code-2QMO, sequence identity-18%, r.m.s.d. = 1.05 Å for 111 C_α atoms), similar differences in the overall structure were found. Firstly, in *E. coli* DTBS active site, two extended antiparallel β–strands of unknown function are present between the residues Gly42-Ser53 in place of a short loop (Gly42-Asp47) in *Mtb*. There is a long loop (Lys35-Ser49) present in this region in *H. pylori* DTBS. Secondly, a long loop followed by a short one-turn helix is seen in both the *E. coli* enzyme (Arg62-Pro72) as well as in *H. pylori* (Leu61-Ile70) in place of a single β–strand (β3) in *Mtb* DTBS (Gln61-Leu65). Thirdly, there is a stretch of three short one-turn helices connected by loops in the region Pro209-Asn224 in *E. coli* DTBS instead of a long helix (α7, Ala205-Ala215) followed by a short one-turn helix at the C-terminus in *Mtb*. Some parts of this C-terminal region have disordered electron density in the *E. coli* apoenzyme as well as in the substrate complexed structures (PDB codes – 1DTS, 1DAI and 1DAE). This might be due to the presence of these short stretches of helix at this terminus, responsible for its flexibility upon solvent exposure. In this case, the *H. pylori* DTBS differs in having a very different

structure at the C-terminus; the strand $\beta 7$ and the helices $\alpha 6$ and $\alpha 7$, which is the binding site for the adenine ring of the nucleotide from *Mtb*, are missing in *H. pylori*. Therefore, the exact mode of nucleotide binding in the *H. pylori* DTBS remains unknown until a structure of DTBS complexed with nucleotide can be determined. Even the residues in the P loop motif differ in all three DTBS enzymes. The *Mtb* enzyme has glycine residues in place of the charged residues in the loop region of its *E. coli* (Glu and Asp) and *H. pylori* (Asn) counterparts. Furthermore, at the P loop of the *E. coli* enzyme, Glu12 forms a salt bridge with Lys148 of the active site. As the nucleotide binds the Walker A motif, this bond is broken and Glu12 orients toward the β -phosphate of ATP and DAPA-carbamate at the active site. Initially, Glu12 was thought to participate in *E. coli* enzyme catalysis as a general base/acid, however, mutation studies on Glu12 (E12A) suggest that this residue has synergistic effects on substrate binding and is not involved in catalysis (187). Lys148 stabilizes the P loop by supporting the pocket formed at the back of the loop through an interaction with Glu12. In *Mtb* DTBS, Gly12 and Asp142 are present in place of Glu12 and Lys148, respectively. Thus, this ionic interaction is missing in the *Mtb* DTBS active site. These comparisons also indicate that the *Mtb* DTBS possesses some structural features which are not present in the DTBS of the other two organisms.

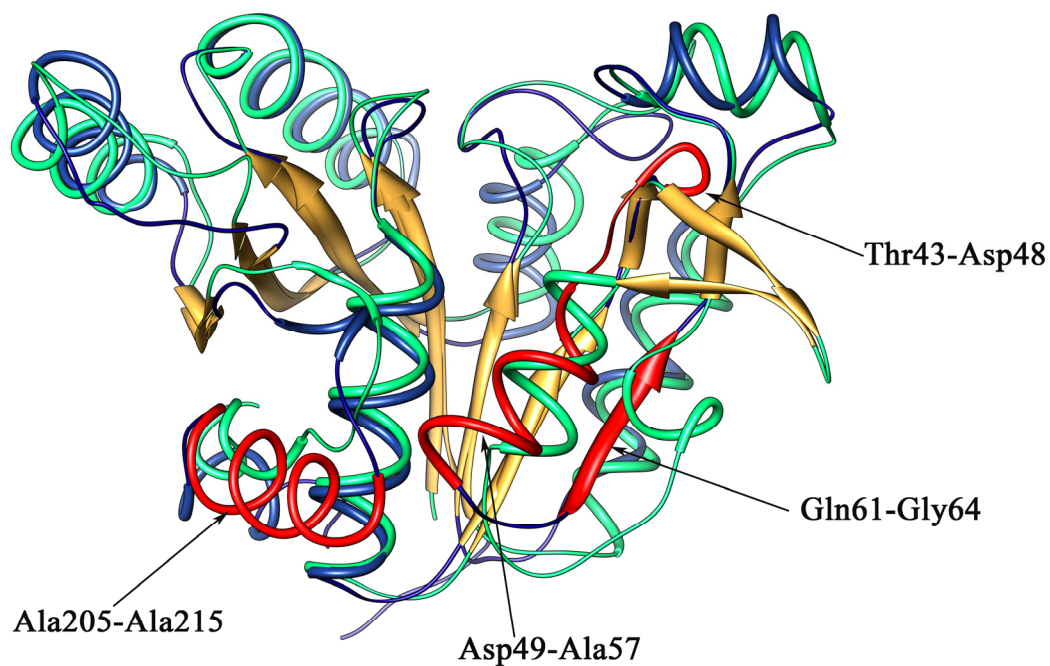


FIGURE 4.13. **Superimposition of C_{α} atoms of DTBS monomer from *E. coli* and *Mtb*.** Ribbon diagram of an overlay of *E. coli* DTBS (PDB code-1BYI) structure (cyan) on *Mtb* DTBS (blue). All the core β strands (gold) from both the structures overlay very well. The arrows point toward the major differences in the two structures (red). The r.m.s.d. for 224 C_{α} atoms for both chains is 1.79 Å.

Possible Mtb DTBS reaction mechanism

Based on the structures of *Mtb* DTBS complexed with substrates and products, we can outline a possible reaction mechanism for this amidoligase. Like *E. coli* DTBS (170), the *Mtb* DTBS reaction involves three steps, the carbamylation of the N7 position of DAPA, the hydrolysis of ATP and the formation of a carbamate–phosphoric anhydride intermediate, and ultimately, the closure of the ureido ring to form DTB with the release of inorganic phosphate.

The first step of the enzyme reaction is the carbamylation of the substrate DAPA. A carbamylation reaction is possible either at the N7 or the N8 position of DAPA in solution (188). The DAPA-bound *Mtb* DTBS structure suggests that this enzyme has specificity for the N7-carbamate. However, it is not clear, whether this enzyme catalyzes the carbamylation at the N7 position of DAPA or if it happens freely in solution. In the next step of the reaction, one of the carbamate oxygen atoms attacks the γ phosphate of ATP, based on the positions of the carbamate-carboxyl oxygen atom and P_i in DAPA and DTB complexed DTBS structures, respectively. As a result, there is cleavage of the P-O bond between the β and γ phosphate of ATP. One of the protons from the amino group of the N8 atom is also abstracted by the water molecule (W221 in chain A) present next to it. Furthermore, the Lys37 residue, which is in the close vicinity of the phosphate-carbamate anhydride intermediate, stabilizes this complex. In the final step of the formation of the ureido ring of DTB, a second deprotonation of the N8 atom of the carbamate-phosphoric anhydride must occur so that it can attack the carbamyl carbon to close the ring. Initially, we thought that Asp47 could act as a base to abstract a proton

from the N8 atom due to its presence in the loop region of the active site, a unique feature of this enzyme. However, our kinetic analysis on the D47A mutant suggests that this Asp residue might not be involved in proton abstraction from N8 atom since the mutant enzyme still retains the normal activity. In this situation, it is quite difficult to further pinpoint which residue might act as a base due to the lack of a DTBS crystal structure complexed with carbamate phosphate anhydride. However, based on the analogy with the p21^{ras} protein mechanism (189, 190), it can be speculated that the oxygen atom on the phosphate of the intermediate complex could act as a base. After the ureido ring closure, there is transfer of electrons, resulting in the breakage of the C-O bond and the release of the inorganic phosphate to form DTB (Figure 4.14).

Thus, using the knowledge gained from the different *Mtb* DTBS complexed structures as well as the preliminary biochemical analysis of this enzyme, our future goal will be to search for potential inhibitors against this enzyme. In this regard, the analogs of DAPA carbamate and variable length analogs of substrate DAPA and KAPA seem to be good candidates. The natural antibiotic acidomycin has a chemical structure similar to DTB/biotin. It also had a good whole cell activity against *Mtb* (MIC = 0.0625-0.125 µg/ml) (59). However, our studies show that *Mtb* DTBS is not the target of acidomycin since it did not inhibit the DTBS reaction even at a concentration of 200 µM. Besides targeting the DAPA-binding site, the mimics of the carbamate-phosphoric anhydride complex can also act as good inhibitors for this enzyme as they will irreversibly bind at the active site, not resulting in product formation.

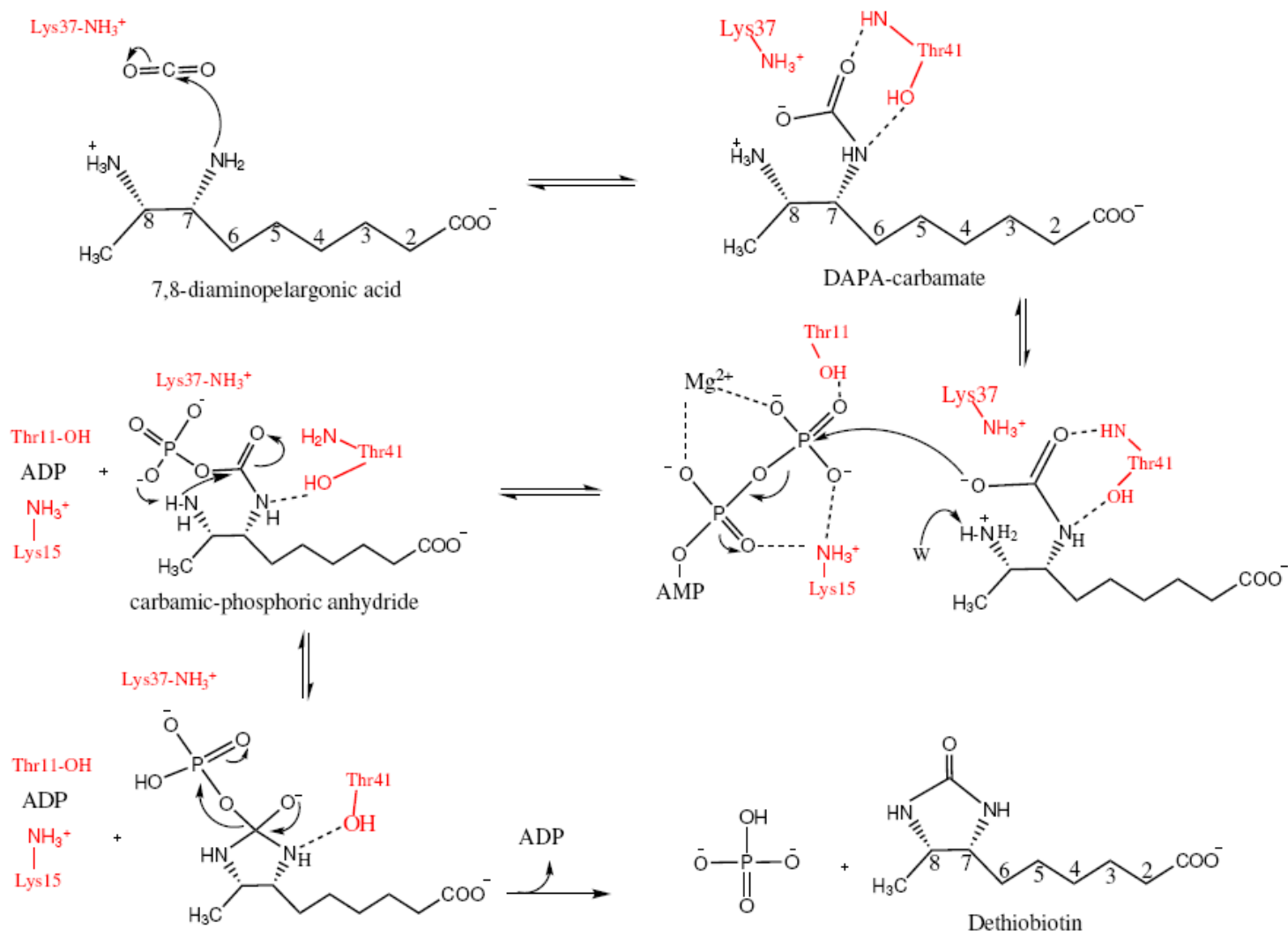


FIGURE 4.14. Schematic representation of the possible mechanism of *Mtb* DTBS.

CHAPTER V

PURIFICATION AND PRELIMINARY CHARACTERIZATION OF

7-KETO, 8-AMINOPELARGONIC ACID SYNTHASE

IN *Mycobacterium tuberculosis*

Background

7-keto, 8-aminopelargonic acid synthase (KAPAS; EC 2.3.1.47) catalyzes the first step in the biotin biosynthetic pathway. During this step, the condensation reaction of L-alanine with pimeloyl-CoA and the simultaneous decarboxylation of L-alanine occur. This process requires pyridoxal 5'-phosphate (PLP) as a cofactor and results in the formation of 7-keto, 8-aminopelargonic acid (KAPA) (Figure 5.1). In *Mtb*, the source of the starting material for biotin production, pimeloyl-CoA, is not clear.

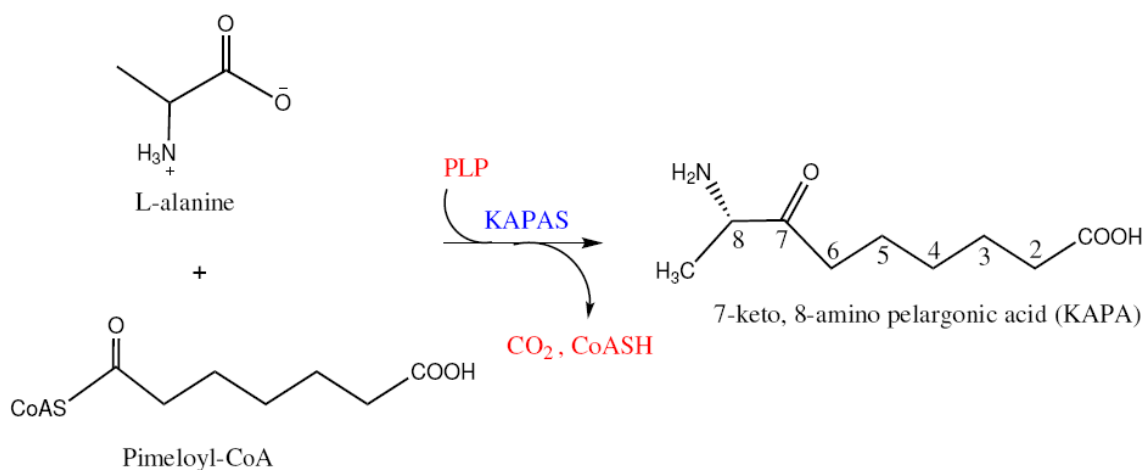


FIGURE 5.1. Reaction catalyzed by 7-keto, 8-aminopelargonic acid synthase.

KAPAS has been studied in both *E. coli* (191, 192) and *B. sphaericus* (42, 193). The only known crystal structure of KAPAS is determined in *E. coli* and, likewise, inhibition studies of this enzyme have also been carried out only in this organism. Biochemical and crystallization trials of KAPAS in *B. sphaericus* have been performed for increasing the biotin production in the food industry. Preliminary biochemical studies have also been performed in plants such as *Arabidopsis thaliana* (50) and pathogens like *Mtb* (194) with the intent to develop herbicides and antibiotics.

Knockout studies on the *Mtb* H37Rv *bioF1* gene, which codes for KAPAS, have suggested that this gene is required during the early stages of mycobacterial infection in mice (9). Further *in vitro* studies of the *bioF1* mutant strain in culture results in cell death instead of growth arrest suggesting that it cannot survive without biotin (55). This indicates the requirement for a *de novo* biotin biosynthetic pathway for mycobacterial survival. Hence, KAPAS in *Mtb* is a good candidate for inhibitor design against this pathogen.

To this end, we will discuss the cloning, protein purification and preliminary biochemical characterization of the KAPAS in *Mtb* in order to further facilitate the process of structure determination and inhibitor studies in this enzyme.

Methods

Reagents

The restriction enzymes for Rv1569 cloning were purchased from New England Biolabs. All the reagents have been purchased from Sigma except KAPA was purchased from Toronto research chemicals Inc., Canada and pimeloyl-CoA was provided to us by Dr. Richard Lee, University of Tennessee.

Cloning and protein expression of Mtb KAPAS

The 1161 base pair *bioF1* gene (Rv1569) from *Mtb* H37Rv genome was PCR amplified and cloned into pET28b with the TEV site using *NdeI* and *HindIII* restriction enzymes. This clone generates the His-tagged KAPAS. The *NdeI* site contained the ATG start codon and the stop codon preceded the *HindIII* site. The different primers used in this study were:

For N-terminal His-tagged or wild type KAPAS,

Forward Primer: 5' AGAGAAGCATATGAAAGCCGCCACGCAGGCACG 3'

Reverse Primer: 5' CCCAAGCTTTCAACGGCGCGCCACGGCGAGAACATC 3'

For C-terminal His-tagged KAPAS,

Forward Primer: 5' AGAGAAGCATATGAAAGCCGCCACGCAGGCACG 3'

Reverse Primer: 5' CCCAAGCTTACGGCGCGCCACGGCGAGAACATC 3'

For truncated N-terminus KAPAS (40-386),

Forward Primer: 5' AGAGAAGCATCTGGTCGCCACCGAGCTGGACTTGG 3'

Reverse Primer: 5' CCCAAGCTTTCAACGGCGCGCCACGGCGAGAACATC 3'

For untagged KAPAS, the Rv1569 gene was inserted in the pET30b vector with *NdeI* and *HindIII* sites. The truncated Rv1569 gene was also cloned into pET28b with TEV sites to generate the N-terminal His-tagged truncated protein. All the constructs were checked for the presence of the insert by DNA sequencing.

The clones (pET28b-Rv1569, pET30b-Rv1569 and the truncated version of Rv1569) were transformed into the *E. coli* expression host BL21 (DE3). The cell culture was grown in LB media with kanamycin (50 µg/ml) at 37 °C till the A_{600} reached 0.8. The culture was then induced with 0.5 mM IPTG at 25 °C and grown overnight. There was no expression of KAPAS at temperatures lower than 25 °C.

Purification of His-tagged KAPAS

Cells were harvested by centrifuging at 3000 rpm for 30 min and resuspended in 50 mM Tris buffer (pH 7.0), 500 mM NaCl, 10% glycerol, 20 µg/ml DNase, 100 µM PLP and one tablet of Complete EDTA-free protease inhibitor cocktail (Roche) per cell pellet (6 grams). The cells were lysed using a French press at 11,000 psi and the cell debris was removed by centrifuging the cell lysate at 15,000 rpm for 1 hour. The supernatant was then loaded on a HisTrapTM FF Ni column (GE healthcare) which was pre-equilibrated with buffer A (50 mM Tris buffer (pH 7.0), 500 mM NaCl, 10% glycerol, 4 mM β-mercaptoethanol). The column was then washed with buffer A containing 50 mM imidazole to remove all the contaminants. The protein was eluted using an imidazole gradient (100-500 mM) in the buffer. KAPAS was 99% pure as judged by the 4-12% gradient SDS-PAGE gel. The fractions containing the protein were

pooled and dialyzed overnight against a buffer containing 25 mM Tris (pH 7.0), 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol and 100 μ M PLP. The protein was further concentrated by ultrafiltration using 30 kDa molecular weight cutoff Centriprep (Millipore) filters before being used for crystallization experiments.

Purification of wild type KAPAS

Purification of wild type KAPAS (untagged) began by resuspending cell pellet in buffer containing 50 mM potassium phosphate (pH 7.0), 10% glycerol, 2 mM DTT, 20 μ g/ml DNase and one tablet of Complete EDTA-free protease inhibitor cocktail (Roche) per cell pellet. The cells were lysed using a French press at 11,000 psi and the cell debris was removed by centrifuging the cell lysate at 15,000 rpm for 1 hour. After centrifugation, the supernatant was saturated with 30% ammonium sulfate. The solution was again centrifuged at 10,000 rpm for 15 min.

Phenyl sepharose HP column: The resultant pellet was further resuspended in buffer containing 50 mM potassium phosphate (pH 7.0), 30% $(\text{NH}_4)_2\text{SO}_4$, 10% glycerol, 1 mM DTT and loaded on phenyl sepharose HP column (GE Healthcare) equilibrated with buffer A (50 mM potassium phosphate (pH 7.0), 1 M $(\text{NH}_4)_2\text{SO}_4$, 1 mM DTT and 10% glycerol). After the column was washed thoroughly with 100 ml of buffer A to remove the contaminants, the protein was eluted with 400 ml of linear gradient of $(\text{NH}_4)_2\text{SO}_4$ in buffer A (1-0 M).

SP sepharose FF column: The fractions containing KAPAS were pooled and loaded on a SP sepharose column pre-equilibrated with buffer B (50 mM PIPES (pH 6.5), 1 M NaCl,

10% glycerol, 1 mM DTT). After the column was washed with 100 ml of buffer B containing 25 mM NaCl, the protein was eluted with 400 ml of linear gradient of NaCl in buffer B (50 mM - 800 mM). The fractions containing KAPAS were pooled and concentrated, using 30 kDa molecular weight cutoff Centriprep (Millipore) filters.

Gel filtration: In the last step, the concentrated KAPAS is injected into a gel filtration column (Superdex 200 XK16/60, GE healthcare) which was pre-equilibrated with 25 mM PIPES (pH 6.5), 100 mM NaCl, 10% glycerol, 2 mM DTT and 1 mM EDTA and eluted with the same buffer at a rate of 0.5 ml/min.

The fractions containing KAPAS were pooled and concentrated as before and used for crystallization experiments. All the purification experiments have been carried out at 4 °C.

Determination of oligomeric state of Mtb KAPAS

KAPAS with concentration of 10 mg/ml was injected into the gel filtration column (Superdex 200 XK16/60, GE Healthcare) which was pre-equilibrated with buffer containing 25 mM bis-tris propane (pH 7.0), 50 mM NaCl, 2 mM DTT and 1 mM EDTA at 4 °C. The protein was eluted at the rate of 0.5 ml/min with 2 ml fractions being collected. BioRad standards (thyroglobulin (670kDa), BSA (158kDa), chicken ovalbumin (44kDa), equine myoglobin (17kDa) and vitamin B₁₂ (13.5kDa)) were used for the calibration purpose.

Spectroscopic studies of Mtb KAPAS

Absorbance spectra for the *Mtb* KAPAS (10 μ M) reaction with 10 mM alanine and 100 μ M pimeloyl-CoA was measured in the range of 300-600 nm using a Cary 100 spectrophotometer (Varian). These experiments were conducted at room temperature, using a buffer containing 50 mM AMPSO (pH 9.0), 150 mM NaCl, 10% glycerol, 4 mM β -mercaptoethanol and 100 μ M PLP. KAPAS concentration was calculated from its absorbance at 280 nm using its extinction coefficient ($\epsilon = 21270$ /mM/cm) (138).

For the measurement of the kinetic parameters of *Mtb* KAPAS, we performed the enzyme (2.5 μ M) reaction with Ellman's reagent (DTNB- 5, 5'-dithio-bis (2-nitobenzoic acid)) at 412 nm. 5mM stock solution of Ellman's reagent was prepared in 50mM phosphate buffer (pH7.5) and stored at -80 °C. The buffer used for the reaction was 20 mM potassium phosphate buffer (pH 7.5), 150 mM NaCl, 10% glycerol and 2 mM EDTA. The reaction was monitored by measuring the absorbance of thiophenolate formation at 412 nm ($\epsilon = 13600$ /M/cm) at room temperature using a Cary 100 spectrophotometer. Michaelis-Menten equation was used to calculate the K_m and K_{cat} values for L-alanine and pimeloyl-CoA, with the curvefit option in the Kaleidagraph software.

Crystallization trials of Mtb KAPAS

Crystallization trials for *Mtb* KAPAS in complexes with 100 μ M PLP, substrates 10 mM L-alanine and 100 μ M pimeloyl-CoA as well as product KAPA (5 mM) were carried out at 18 °C using the crystal screen I and II, SaltRx, PEG/Ion, Membfac, Index

screen of Hampton Research and Wizard I, II and III of Emerald BioSystems. Initial crystal screening was performed using Hydra Plus One dispenser (Art Robbins instruments) with a 1:1 ratio of KAPAS protein (10 mg/ml) and crystallization condition, on 96 well intelliplates (Hampton Research).

Results and discussion

Purification and crystallization of Mtb KAPAS

We have successfully cloned the Rv1569 gene from *Mtb* H37Rv genomic DNA and overexpressed KAPAS protein in *E. coli* BL21 (DE3) host strain. In figure 5.2, the DNA agarose gel shows the amplified 1.2 kbps Rv1569 gene from *Mtb* H37Rv genome. This gene was inserted in the modified pET28b vector. The pET28b-Rv1569 clone was transformed into an overexpression host *E. coli* BL21 (DE3). Both His-tagged and wild type *Mtb* KAPAS (untagged) proteins were purified to homogeneity using different purification protocols outlined in the methods section. Figure 5.3 A represents a typical Ni column purification profile which provided one-step purification of KAPAS. SDS-PAGE analysis of *Mtb* KAPAS from this purification experiment revealed a single band at 40 kDa which is in agreement with the molecular mass determination of the Rv1569 gene sequence (Figure 5.3 B). This His-tagged protein was further used to setup crystallization trials.

There are instances in crystallization trials when an N-terminal His-tag may hinder crystal lattice formation due to its flexible nature. Cleavage of this His-tag from the protein or no tag can readily lead to crystal formation. Hence, to cleave the N-

terminal His-tag of KAPAS, we used TEV protease. The reaction was carried out at a ratio of 20 mg of KAPAS to 1 mg of TEV protease at 4 °C under overnight dialysis. The buffer containing 20 mM potassium phosphate buffer (pH 7.0), 150 mM NaCl, 10% glycerol and 5 mM β -mercaptoethanol was used. However, the protein became cloudy under overnight dialysis and TEV cleavage was also not successful, resulting in only 20% cleavage of KAPAS. This might indicate that the protein is not stable during TEV cleavage. To circumvent this problem, C-terminal His-tagged KAPAS was purified. However, this construct was not very pure and had low yield.

As a next logical step towards crystallization of *Mtb* KAPAS, we constructed the wild type KAPAS protein. This protein was purified in four steps: ammonium sulfate precipitation (30% saturation), phenyl sepharose (hydrophobic) column, SP sepharose (anionic) column and gel filtration chromatography. The protein from this protocol was judged 99% pure from the 4-12% SDS –PAGE gel. This protein was further concentrated up to 10 mg/ml and crystallization screens were set up using over 1000 different precipitant conditions.

Different buffers (Tris, potassium phosphate, Hepes, PIPES, bis-tris propane) with pH conditions in the range 6.0-9.5 have been screened to determine the stability of the protein as well as to setup crystallization trials. This protein is stable in potassium phosphate and Hepes buffer (pH 7.0). Although *Mtb* KAPAS seemed to be stable in phosphate buffer, crystallization trials in this buffer mostly resulted in the formation of salt crystals.

Based on the crystal structure of *E. coli* KAPAS (PDB code-1BS0), the N-terminal region contains a helix followed by a long loop (Figure 5.4). It was believed that this floppy region would prevent crystal formation. In order to facilitate the process of crystallization in KAPAS, 40 residues at the N-terminus have been truncated. Further crystallization trials of this truncated protein in complex with ligands such as substrate/product as well as inhibitor are in progress.

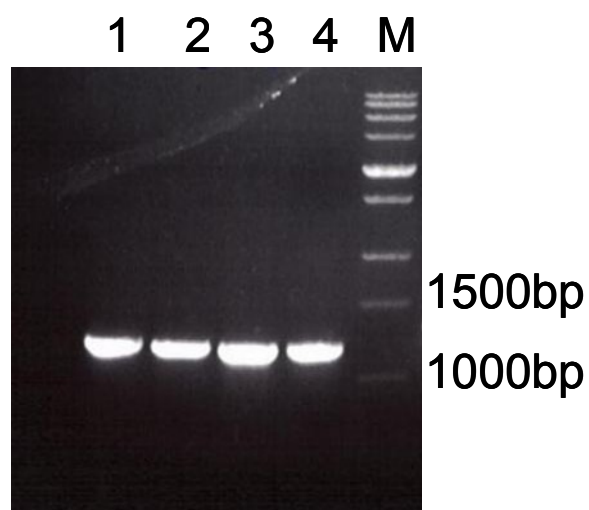


FIGURE 5.2. **PCR amplification of Rv1569 gene.** 1% agarose gel shows the PCR amplified Rv1569 gene product from *Mtb* H37Rv genomic DNA. 1161 bps Rv1569 gene product is shown in lanes 1-4. The extreme right lane, M is the Perfect DNATM marker (0.5–12 kbp) (Novagen).

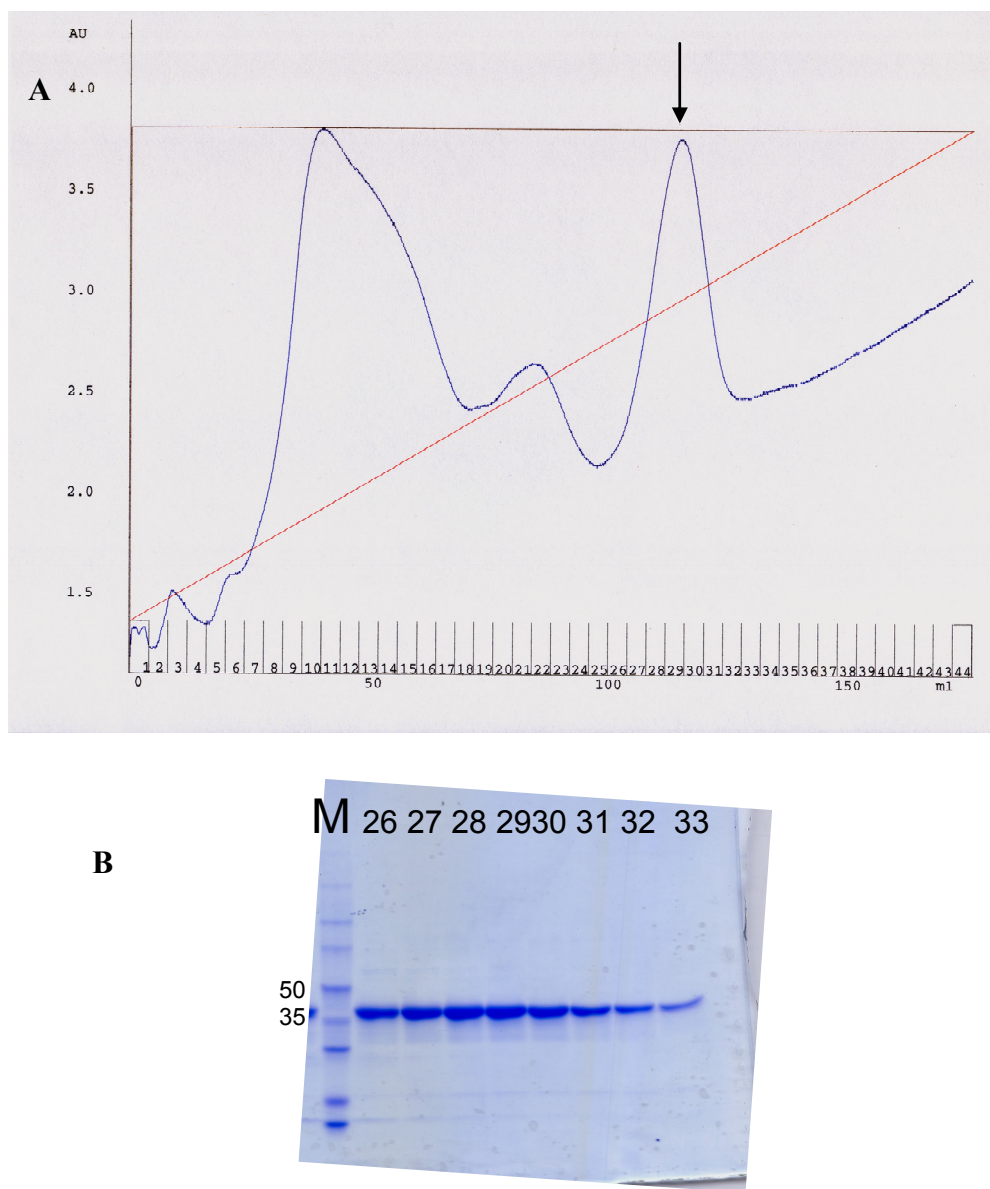


FIGURE 5.3. Elution profile of a Ni affinity column chromatography and SDS-PAGE gel of KAPAS purification. (A) The elution profile of the Ni affinity chromatography showing the absorbance peak at 280 nm (marked by an arrow) which contains the fractions with *Mtb* KAPAS. (B) 4-12% gradient SDS-PAGE gel stained with Coomassie Blue showing lanes with *Mtb* KAPAS in the fractions eluted from the Ni column and Perfect protein™ marker (10-225 kDa) (M). As evident in the figure, KAPAS runs at 40 kDa on PAGE gel.



FIGURE 5.4. **Ribbon diagram of *E. coli* KAPAS (PDB code-1BS0).** The helices, β strands, and the loops are colored blue, yellow and green, respectively. In order to yield better crystals, the N-terminal helix with the loop (pink) was truncated in *Mtb* KAPAS.

Based on the elution profile of size exclusion chromatography, which contained a single peak, the molecular weight of KAPAS in solution was calculated to be approximately 80 kDa (data not shown). This indicates the oligomeric state of KAPAS as a dimer. This is analogous to *E. coli* KAPAS protein. Furthermore, on sequence alignment of KAPAS from both the organisms, we found that *Mtb* KAPAS has 36% sequence identity with *E. coli* KAPAS (Figure 5.5).

```

E.coli_KAPAS MSWQEKINAALDARRAADALRR-RYPVAQAGRWLVADDRQYLNFSNDYLGLSHHPQII
M.tb_KAPAS   MKAATQARIDDSPLAWLDAVQRQRHEAGLRRCLRFRPAVATELDLANSYDLGLSRHPAVI

E.coli_KAPAS RAWQCGAEQFGIGSGSGHVSQYSVVHQALFEELAEWLGYSRALLFISGFANQAVIAAM
M.tb_KAPAS   DGGVQALRIWGCATGSRRLVTGDTKLHQQFHAELAEFVGAAGLLFSSQYTNLGAUVGL

E.coli_KAPAS MAKEDRIAADRLSHASILEAASLSPSQLRRFAENDVTHLARLLASPCPGQQMVVTEGVFS
M.tb_KAPAS   SGPSSLLVSDARSHASLVDA CRLSRARVVVTPHRDVAVDAAALRSRDEQRAVVVTDVFS

E.coli_KAPAS MDGDSAPLAIEIQQTQQHNGWLMVDDAHGTGVIGEQGRGSCWLQKVK--PELLVVTFGK
M.tb_KAPAS   ADGSLAPVRELLVCRRHGALLLVDEAHGLGVRG-GGRLLYELGLAGAPDVVMTTTLK

E.coli_KAPAS GFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARREKLA
M.tb_KAPAS   ALGSGGGVVLGPTPVRAHLIDAARPFIFDTGLAPAVGAARAALRVLQAEFWRPQAVLNH

E.coli_KAPAS ALITRFRAGVQDLFFTLADSCSAIQPLIVGDNRAQLQAEKLRQQCWVTAIRPPTVPAG
M.tb_KAPAS   AGELARMCVAAVPD-----SAMVSVILGEPESA VAAA AACLDA CVKVCFRPPTVPAG

E.coli_KAPAS TARLRLTLTA AHMQDIDRLLLEVLIHGNG-----
M.tb_KAPAS   TSRLRLTARA SLNAGELELARRVLTVDLAVARR

```

FIGURE 5.5. Sequence alignment of *E. coli* and *Mtb* KAPAS. The residues colored in green and cyan depict conserved and similar residues in both organisms, respectively.

To gain some insights into the residues, which might constitute the active site, structural alignment using the ligand bound complexes of *E. coli* KAPAS structure with *Mtb* KAPAS homology model (generated by Swiss model) was carried out. Key residues from the *E. coli* AONS-PLP structure (PDB code-1BS0) include Lys236 (*Mtb* Lys239) which binds PLP to form the internal-aldimine and His133 which forms a stacking interaction with the PLP ring, also conserved in *Mtb* (His134). Other residues in the

active site of the *E. coli* enzyme that contribute to PLP binding are Asp204, Ala206, His207, Arg361 which are also conserved in *Mtb* (Asp205, Ala207, His208, Arg358). Based on the ternary complex of KAPA-PLP bound AONS structure (PDB code-1DJ9), the O7-keto group of KAPA is hydrogen bonded with His133 (*Mtb* His134) and the carboxylate group of KAPA is hydrogen bonded with Arg21 and Arg349 which corresponds to Gln21 and Arg346 in *Mtb*. However, there is an Arg present next to Gln21 in *Mtb*, which might be involved in the hydrogen bond interaction with the carboxylate of KAPA, or the amino group of Gln21 can also serve as the hydrogen bond partner. The methylene long chain of KAPA is in van der Waals contact with Val79 (*Mtb* Val80), Met180 (*Mtb* Ala181), Ile263' (*Mtb* Ile266) and Tyr264' (*Mtb* Phe267). Other residues involved in catalysis of KAPAS are Ser179 (*Mtb* Ser180) and Glu175 (*Mtb* Asp176). Although there are some differences in KAPA interaction between the two enzymes, we expect the active site of the *E. coli* enzyme to be similar to the *Mtb* KAPA synthase. However, there is no three-dimensional structure of KAPAS complexed with substrates L-alanine or pimeloyl-coA available. Hence, there is lack of information regarding the residues involved in substrate interaction as well as conformational changes accompanying the substrate binding at the active site. Since these substrates have very different structural characteristics compared to the product KAPA, three-dimensional structure of substrate-bound *Mtb* KAPAS still needs to be determined to pinpoint the residues involved in its binding.

Spectroscopic properties of Mtb KAPAS

Taking advantage of the chromophore properties of cofactor PLP and PMP, we studied the absorption spectra of wild type *Mtb* KAPAS reaction with the substrates, L-alanine and pimeloyl-CoA in the range of 300-600 nm. As shown in the figure 5.6 A, the absorption spectra of KAPAS shows two peaks at 345 nm and 425 nm, which are characteristic of the PLP form of the KAPAS (internal-aldimine) in non-coplanar and coplanar states, respectively. Addition of the first substrate, L-alanine shows an increase in the absorbance peak at 425 nm indicating the formation of external-aldimine complex of the enzyme with subsequent decrease in absorption at 345 nm (coplanar form of PLP). Further addition of pimeloyl-CoA to this enzyme-external-aldimine complex leads to the formation of the carbanion reaction-intermediate characterized by a third peak at 532 nm (Figure 5.6 B).

To calculate the kinetic parameters of *Mtb* KAPAS for substrates L-alanine and pimeloyl-CoA, we performed the enzyme reaction with Ellman's reagent (DTNB). Breakdown of pimeloyl-CoA to CoA-SH was monitored by measuring the increase in absorbance of resultant thiophenolate formation at 412 nm. The K_m and K_{cat} values calculated for L-alanine were 400 μM and $7 \times 10^{-3}/\text{s}$ and for pimeloyl-CoA were 4 μM and $9.1 \times 10^{-3}/\text{s}$, keeping the other substrate concentration constant, at room temperature. These values are quite similar to the values reported for *Mtb* enzyme by another group (194) but different from the values obtained for *E. coli* KAPAS (K_m for L-alanine is 900 μM and K_m for pimeloyl-CoA is 70 μM).

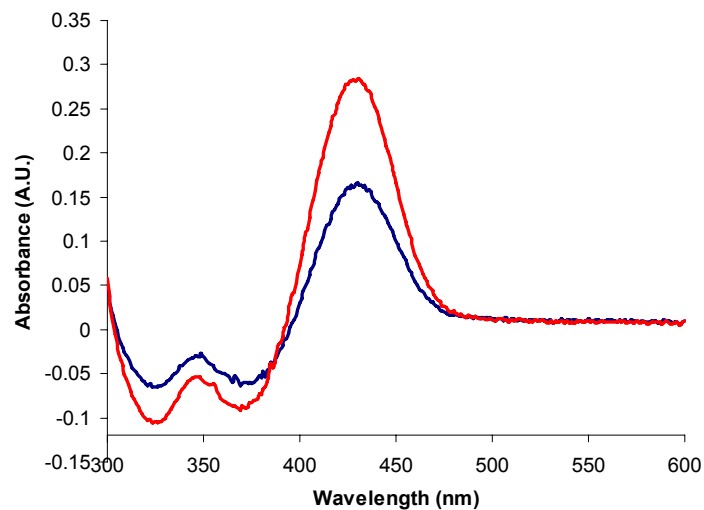
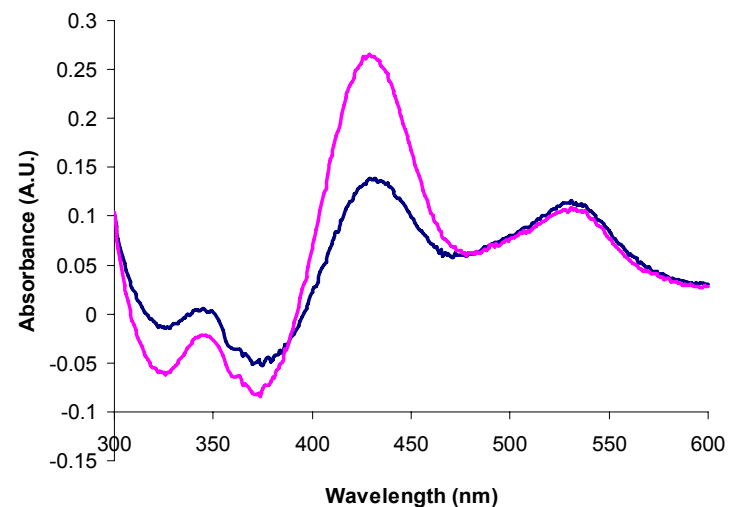
A**B**

FIGURE 5.6. **Characterization of the KAPA synthase reaction intermediates.** Absorbance spectra in the range of 300-600 nm for (A) 10 μM KAPAS enzyme alone (blue) and a reaction mixture with 10 mM L-alanine after 30 sec (red) (B) 10 μM KAPAS + 10 mM L-alanine + 100 μM pimeloyl-CoA (0 sec-blue and 30 sec-pink). All the spectrophotometric experiments were conducted at room temperature.

Conclusions

In this chapter, we have described, in detail, the purification protocols for both His-tagged and wild type *Mtb* KAPAS, one of the interesting drug targets in the *Mtb* biotin biosynthetic pathway. We also tried to trouble-shoot some of the problems faced during the purification of this protein. To understand the catalytic behavior of this enzyme, preliminary biochemical characterization of KAPAS was carried out using spectroscopic methods. This analysis can be further utilized as an enzyme assay for inhibitor studies. Kinetic studies on *Mtb* KAPAS show that the K_m values for the substrates L-alanine and pimeloyl-CoA are much lower than those of the *E. coli* enzyme. In addition, different crystallization trials of the apoenzyme and enzyme - substrate/product complexes were carried out for the determination of the three-dimensional structures of this protein and we are awaiting the results. The active site in these KAPAS structures will further act as a template in virtual screening against the large compound libraries in search for a potent inhibitor.

CHAPTER VI

CONCLUSIONS

A better understanding of the physiology of *Mycobacterium tuberculosis* (*Mtb*) during its pathogenesis requires knowledge of the different metabolic pathways responsible for its survival. Research with *Mtb* is difficult because of its slow growth, clump formation in cultures due to its lipid-rich cell wall, biosafety level 3 work conditions, and the high cost of maintaining the animal models. However, the availability of the genome sequence of *Mtb* and the different genetic approaches to study them has paved our way towards the elimination of this organism. Different genetic methods have indicated the importance of various metabolic pathways, including the serine and biotin biosynthetic pathways, in the survival of *Mtb* during different stages of its pathogenesis. In this regard, we have studied the first enzyme of the serine and first three enzymes of the biotin biosynthetic pathways in *Mtb*. We had the opportunity to analyze three different classes of enzymes: oxido-reductase, transferase, and ligase, using different X-ray crystallographic techniques.

In the first step of phosphorylated serine biosynthesis, D-3-phosphoglycerate dehydrogenase (PGDH) catalyzes the reversible oxidation of D-3-phosphoglycerate to hydroxypyruvic acid phosphate (HPAP), using NAD^+ as a cofactor. Although this enzyme has been extensively studied in *E. coli*, *Mtb* PGDH has not been characterized to date. *Mtb* PGDH differs from *E. coli* PGDH in having an insertion of 130 amino acids near its C-terminus. In Chapter II of this dissertation, studies of the structure–function

relationship of PGDH in *Mtb* have been carried out in collaboration with Dr. Gregory A. Grant (Washington University School of Medicine, St. Louis, MO). *Mtb* PGDH was cloned, purified and further characterized, both structurally and biochemically. The three-dimensional structure of this 529-residue protein was determined in a nonstandard way. Since this protein lacked any methionine other than the starting one, this residue was introduced in the protein by mutating Ile514, located in the regulatory domain. This mutant was selected based on its sequence alignment with *E. coli* PGDH. We were able to obtain the phase information with only one selenomethionine residue in this 529 amino acid protein, well below the theoretically estimated value of one methionine residue per 100 amino acids for phase determination (97). The reason we could get phase information with one selenomethionine might be a result of a highly redundant data (14-fold) and a high symmetry space group (P6₅22) for this PGDH crystal. The resulting 2.3 Å *Mtb* PGDH apostructure shows somewhat different domain organization in its monomer due to the presence of the insertion in comparison to *E. coli* PGDH. Superimposition of chain A on chain B in the asymmetric unit shows two different conformers. This conformational change between the adjacent subunits in the asymmetric unit is in the form of a 160° rotation and flipping of the nucleotide binding and substrate binding domains in relation to the intervening and regulatory domains. The pivot point for this change is located in a stretch of three glycine residues (316-318) on the long loop region, which connects both the nucleotide- and the substrate-binding domain to the intervening and the regulatory domain. Thus, the PGDH tetramer is

arranged in an asymmetric manner, generated by the heterodimer with a noncrystallographic two-fold axis.

Because of the two different structural conformers, the cofactor-binding site in this dehydrogenase exists in two different environments: one inside a channel and the other as an open cleft. The presence of two different NADH-binding sites was further verified by stopped flow fluorescence spectroscopic experiments (78). Failure to trap the cofactor at the active site of *Mtb* PGDH structures was a result of the presence of the loop (203-213) blocking the cofactor-binding site in chain B due to the symmetry related molecule. Also, the causes for the weak cofactor binding were the high salt concentration (1 M NaK tartrate) present in the crystallization solution and the high ionic strength of the protein storage buffer (100 mM potassium phosphate). This fact is also supported by the stopped flow fluorescence analysis of NADH-binding in high ionic strength buffer (78).

A 2.4 Å crystal structure of a binary complex of *Mtb* PGDH with the substrate HPAP pinpointed certain residues in the active site responsible for the enzyme activity. These residues (Arg 51, Arg132, and Arg233), which interact with HPAP, are mostly conserved in both *E. coli* and human PGDH. All the positively charged residues at the active site are oriented inwards to facilitate in the binding of the negatively charged substrates (HPAP and PGA). This structure also identified the His280 and Glu262 residues at the active site which act as a catalytic dyad in dehydrogenases. In this complexed structure, we did not see the productive form of the enzyme due to the absence of a cofactor. However, comparison of this structure with the human truncated

PGDH (with the bound substrate analog malate) provided us with a model for the resulting conformational changes at the active site cleft during hydride transfer. This change is in the form of a 15° rotation of the substrate-binding domain with respect to the nucleotide-binding domain which prevents solvent exposure during hydride transfer. Fluorimetric studies also indicate that the binding of NADH is facilitated by the presence of substrate in the active site, resulting in the conformational changes needed for catalysis. Thus, the order of substrate binding in *Mtb* PGDH is HPAP first and then NADH, which is the reverse of *E. coli* PGDH.

Presence of tartrate bound to the intervening domain of *Mtb* PGDH provided us insights into a possible role for this domain (64). The role of tartrate as a physiological ligand is ruled out, as *Mtb* does not require tartrate in any metabolic reaction. However, this binding provided us a hint about the role of the intervening domain as a site for HPAP-binding due to its structural resemblance to tartrate. *Mtb* PGDH is uncompetitively inhibited at higher HPAP concentrations (250 μM) (128). This type of substrate inhibition has also been reported in rat liver PGDH which contains an intervening domain (29) but not in *E. coli* PGDH which lacks this domain. The HPAP-binding site is termed an anion-binding site since all the positively charged residues from the intervening and regulatory domain interface converge to interact with HPAP. Mutational studies on these residues involved in HPAP binding (R446A, H447A, R451'A, R501'A, and L439'A) demonstrate reduced substrate inhibition (' indicates the residue is from adjacent subunit). Triple mutation of the residues from one subunit (R451A, R501A, L439A) resulted in complete elimination of the substrate inhibition

(128). This kind of second substrate-binding site may be needed for additional regulation of this enzyme during *Mtb* infection of macrophages and can be used as a site for inhibitor design.

Kinetic studies have also shown that *Mtb* PGDH is subject to feedback inhibition by the end product, L-serine, with an $IC_{0.5}$ of $\sim 30 \mu\text{M}$. It exerts V-type inhibition and exhibits a mixed noncompetitive type of inhibition (64). This allosteric regulation shows cooperativity with a Hill coefficient of 2. In agreement with the kinetic data, the serine-bound structure of *Mtb* PGDH revealed the presence of two serine molecules at the interface of the regulatory domains due to a non-crystallographic two-fold axis. This allosteric site is approximately 35 \AA away from the active site. This structure also pinpointed the residues involved in the interaction with serine (Tyr461, Asp463, and Asn481'), which are not conserved in *E. coli* except for Asn481'. Mutation of any of these residues resulted in greatly reduced serine inhibition. Mutation of Asn481' from the adjacent subunit (N481'A) also resulted in a loss of cooperativity for serine binding (78) which suggests that the hydrogen bonding of Asn481' with serine strengthens the regulatory domain interface. There are no major conformational changes in the form of domain movements on serine binding in comparison to the apostructure. However, local changes at the allosteric site in the form of flipping of the side chains of Asp463 and Arg464 as well as change in side chain orientation of Asn481' occurs with serine binding in contrast to the movement of a loop in *E. coli* PGDH. This flipping of the side chain results in the closure of the allosteric site channel present at the regulatory domain interface, resulting in hiding serine from solvent exposure. Finally, this structural

analysis also suggested that *Mtb* PGDH may not follow the same mechanism for serine inhibition as its *E. coli* counterpart. Thus, all this information gained from structural studies on serine feedback inhibition in *Mtb* PGDH will further assist in understanding the mechanism of allosteric regulation in other members of this extended form of PGDH as well as in other proteins which possess the ACT domain. Also, due to the human PGDH insensitivity to serine, this allosteric site in *Mtb* PGDH can act as a template for inhibitor design using fragment based library.

Furthermore, these structural and kinetic studies of *Mtb* PGDH indicate that it is closely related to the mammalian PGDH (presence of intervening domain, substrate inhibition, etc). Nonetheless, *Mtb* PGDH is inhibited by L-serine which suggests that it also possesses metabolic properties similar to *E. coli* PGDH. These observations indicate that the *Mtb* PGDH is a link between the *E. coli* and the mammalian enzymes (64). The phylogenetic tree of PGDH from various organisms, generated by Ali *et al.*, also supports this idea (22).

Human PGDH is not inhibited by L-serine. Sequence alignment of human PGDH with *Mtb* PGDH shows a lot of variation in the two enzymes (sequence identity 33%) even though both belong to the same form of PGDH containing an intervening domain. Determination of the three-dimensional structure of human PGDH will elucidate the differences in domain and subunit organization in this enzyme in comparison to *Mtb*. This structure will also provide us a hint on the role of the C-terminal region, as this enzyme does not show regulatory properties. In this regard, we have discussed the

cloned, purified and set up crystallization trials in preparation for the structure determination.

The focus of the second part of this dissertation was on the characterization of the enzymes participating in different steps of the biotin biosynthetic pathway using biochemical and X-ray crystallographic techniques. Biotin acts as a cofactor in various important metabolic pathways in *Mtb*. Various genetic studies have also indicated the importance of the enzymes of the biotin biosynthetic pathway in mycobacterium survival during infection in macrophages (8) as well as in mice (9). The fact that humans lack homologs of these enzymes makes them excellent targets for antibiotic development.

Knockout studies on the *bioF* gene in the *Mtb* H37Rv strain show that absence of this gene, which codes for the first enzyme 7-keto, 8-aminopelargonic acid (KAPA) synthase (KAPAS) in the biotin biosynthetic pathway, results in complete clearance of mycobacterium in mice (9). This enzyme catalyzes the simultaneous decarboxylation and transamination of pimeloyl-CoA to KAPA, requiring pyridoxal 5'-phosphate (PLP) as a cofactor. We have discussed cloning and protein purification procedures as well as biochemical characterization of *Mtb* KAPAS in chapter V. This will further help in three-dimensional structure determination as well as in designing enzyme assays for inhibitor studies. Our studies show that this enzyme exists as a dimer of 80 kDa in solution and has a much higher substrate affinity than its *E. coli* counterpart.

In chapter III, examination of the 2.2 Å crystal structure of the second enzyme in biotin biosynthesis, 7, 8-diaminopelargonic acid (DAPA) synthase (DAPAS) in *Mtb*, revealed folds similar to those in *E. coli* DAPAS. This DAPAS structure was determined

using the molecular replacement technique. This enzyme catalyzes the transamination reaction of KAPA to DAPA. In this process, it requires PLP as a cofactor and utilizes a unique amino donor, SAM. Structural homology with other members of subclass II aminotransferases confirms its membership in this subclass. Despite the fact that the structure of *E. coli* DAPAS was known, its mode of dual-substrate (SAM and KAPA) recognition was a mystery to date. This question was answered by determining the structure of the ternary complex of *Mtb* DAPAS with PLP and the SAM analog sinefungin. Residues from both the subunits of the dimer contribute to the active site formation. Comparison of this structure with a model of KAPA complexed *Mtb* DAPAS provided us insights into the mechanism of substrate recognition in this particular aminotransferase enzyme, which is quite different from other enzymes of the subclass II. The substrate-bound structures identified the important active site residues involved in catalysis, also present in *E. coli*. Besides the residues Tyr157, Asp160, Arg400 involved in substrate binding at the active site, we recognized Tyr25 to be a key player in substrate recognition, both in positioning KAPA for amino group acceptance as well as in stabilizing the adenine ring of SAM by providing platform for stacking interactions. This fact was further confirmed by constructing the Y25A mutant, which did not show any DAPAS activity with the amino donor SAM. However, the reverse of the second half reaction of DAPAS with product DAPA showed only 70% activity. Furthermore, in an effort to identify potent inhibitors for *Mtb* DAPAS, virtual screening of large compound library has been attempted against the active site of a KAPA bound DAPAS model.

Certain *Bacillus* species as well as *Brucella melitensis*, *Leptospira interrogans*, and *Aquifex aquenox* lack this Tyr residue in the 25th position at the active site; instead, they have a Phe at a corresponding position in the protein sequence. This raises questions about the maintenance of the enzyme activity needed for the survival of these organisms, as it is evident from the previous paragraph that Tyr25 plays an important role in substrate recognition. The amino donor SAM can still bind at the active site since Phe can have stacking interactions with the adenine ring. However, an alternate binding mode for the substrate KAPA at the active site in the absence of the Tyr residue was unraveled on determination of the KAPA-bound *B. subtilis* DAPAS structure. The KAPA-bound form of DAPAS structure, as well as, the apoform and the PLP-enzyme complex structure in *B. subtilis* were solved using molecular replacement with *Mtb* DAPAS as the search model (sequence identity 32%). Interestingly, the active site in the apostructure as well as the PLP-bound structure contains a large peptide segment (Lys143-Glu172) which is disordered. The active site in *Bacillus* gets ordered only on the binding of the substrate KAPA, emphasizing the need for the hydrogen bond network of Tyr25:Tyr157:Asp160 (*Mtb* numbering) for the maintenance of the DAPAS active site.

Chapter IV described the three-dimensional structure of the 1.85 Å *Mtb* dethiobiotin synthetase (DTBS) apoenzyme, the third enzyme in the biotin biosynthetic pathway. This enzyme catalyzes the formation of dethiobiotin from DAPA. This apoenzyme structure was determined using a single anomalous dispersion method with samarium at wavelength of 1.54 Å (Cu K α). We utilized the strong anomalous behavior

of the Lanthanide metals at this wavelength. The differences in the local structural conformations of *Mtb* DTBS in comparison to *E. coli* and *H. pylori* DTBS were discussed in this chapter. The substrate and the product bound *Mtb* DTBS structures identified the key residues involved in catalysis, some of which are not conserved in the *E. coli* counterpart. These structures also illustrated that one of the substrates DAPA, as well as, the product DTB both bind in a hydrophobic pocket formed by residues from both the subunits of the dimer whereas the other substrate ATP binds in an open cleft made up of residues from a single subunit. The DAPA-bound *Mtb* DTBS structure further indicated the preference for carbamate formation at the N7 position of DAPA. Analyses of the substrate and the product complexed DTBS structures, and of the site directed mutagenesis of active site residue (D47A) have further helped us to elucidate a possible mechanism of enzyme action in a rather solvent exposed active site.

With this knowledge of the active site architecture and biochemical characterization in these enzymes (KAPAS, DAPAS, and DTBS) in biotin biosynthesis and the mode of binding of substrates and natural inhibitors, we now have a starting point for designing new inhibitors against these enzymes, with the goal of eradication of *Mycobacterium tuberculosis*.

REFERENCES

1. Sturgill-Koszycki, S., Schaible, U. E., and Russell, D. G. (1996) *EMBO J.* **15**, 6960-6968
2. Sturgill-Koszycki, S., Schlesinger, P. H., Chakraborty, P., Haddix, P. L., Collins, H. L., Fok, A. K., Allen, R. D., Gluck, S. L., Heuser, J., and Russell, D. G. (1994) *Science* **263**, 678-681
3. Escuyer, V., Haddad, N., Frehel, C., and Berche, P. (1996) *Microb. Pathog.* **20**, 41-55
4. McKinney, J. D. (2000) *Nat. Med.* **6**, 1330-1333
5. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaiia, 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. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) *Nature* **393**, 537-544
6. Clark-Curtiss, J. E., and Haydel, S. E. (2003) *Annu. Rev. Microbiol.* **57**, 517-549
7. Murry, J. P., and Rubin, E. J. (2005) *Trends Microbiol.* **13**, 366-372
8. Rengarajan, J., Bloom, B. R., and Rubin, E. J. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8327-8332
9. Sasseti, C. M., and Rubin, E. J. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12989-12994
10. Snell, K., Natsumeda, Y., and Weber, G. (1987) *Biochem. J.* **245**, 609-612
11. Cybulski, R. L., and Fisher, R. R. (1976) *Biochemistry (Mosc.)* **15**, 3183-3187
12. Eagle, H. (1959) *Science* **130**, 432-437
13. de Koning, T. J., and Klomp, L. W. (2004) *Curr. Opin. Neurol.* **17**, 197-204

14. Jaeken, J., Detheux, M., Van Maldergem, L., Foulon, M., Carchon, H., and Van Schaftingen, E. (1996) *Arch. Dis. Child.* **74**, 542-545
15. Jaeken, J., Detheux, M., Van Maldergem, L., Frijns, J. P., Alliet, P., Foulon, M., Carchon, H., and Van Schaftingen, E. (1996) *J. Inherit. Metab. Dis.* **19**, 223-226
16. Greenberg, D. M., and Ichihara, A. (1957) *J. Biol. Chem.* **224**, 331-340
17. Sallach, H. J. (1956) *J. Biol. Chem.* **223**, 1101-1108
18. Willis, J. E., and Sallach, H. J. (1962) *J. Biol. Chem.* **237**, 910-915
19. Walsh, D. A., and Sallach, H. J. (1966) *J. Biol. Chem.* **241**, 4068-4076
20. Ho, C. L., Noji, M., Saito, M., and Saito, K. (1999) *J. Biol. Chem.* **274**, 397-402
21. Ho, C. L., and Saito, K. (2001) *Amino Acids* **20**, 243-259
22. Ali, V., Hashimoto, T., Shigeta, Y., and Nozaki, T. (2004) *Eur. J. Biochem.* **271**, 2670-2681
23. Cheung, G. P., Cotropia, J. P., and Sallach, H. J. (1969) *Arch. Biochem. Biophys.* **129**, 672-682
24. Rowsell, E. V., Snell, K., Carnie, J. A., and Al-Tai, A. H. (1969) *Biochem. J.* **115**, 1071-1073
25. Tolbert, N. E. (1980) *The biochemistry of plants*, Academic Press, New York
26. Takada, Y., and Noguchi, T. (1985) *Biochem. J.* **231**, 157-163
27. Melcher, K., and Entian, K. D. (1992) *Curr. Genet.* **21**, 295-300
28. Walsh, D. A., and Sallach, H. J. (1965) *Biochemistry (Mosc.)* **4**, 1076-1085
29. Achouri, Y., Rider, M. H., Schaftingen, E. V., and Robbi, M. (1997) *Biochem. J.* **323 (Pt 2)**, 365-370

30. Achouri, Y., Robbi, M., and Van Schaftingen, E. (1999) *Biochem. J.* **344** (Pt 1), 15-21
31. Fallon, H. J., Hackney, E. J., and Byrne, W. L. (1966) *J. Biol. Chem.* **241**, 4157-4167
32. Hayashi, S., Tanaka, T., Naito, J., and Suda, M. (1975) *J Biochem* **77**, 207-219
33. Fell, D. A., and Snell, K. (1988) *Biochem. J.* **256**, 97-101
34. Dakshinamurti, K. (2005) *J. Nutr. Biochem.* **16**, 419-423
35. Romero-Navarro, G., Cabrera-Valladares, G., German, M. S., Matschinsky, F. M., Velazquez, A., Wang, J., and Fernandez-Mejia, C. (1999) *Endocrinology* **140**, 4595-4600
36. Goldberg, M. W., Sternbach, L. (1949). US Patent 2,489,237
37. Gomez, J. E., and McKinney, J. D. (2004) *Tuberculosis (Edinb)* **84**, 29-44
38. Camus, J. C., Pryor, M. J., Medigue, C., and Cole, S. T. (2002) *Microbiology* **148**, 2967-2973
39. Green, A. J., Rivers, S. L., Cheeseman, M., Reid, G. A., Quaroni, L. G., Macdonald, I. D., Chapman, S. K., and Munro, A. W. (2001) *J. Biol. Inorg. Chem.* **6**, 523-533
40. Cryle, M. J., and De Voss, J. J. (2004) *Chemical communications (Cambridge, England)*, 86-87
41. Stok, J. E., and De Voss, J. (2000) *Arch. Biochem. Biophys.* **384**, 351-360
42. Ploux, O., Soularue, P., Marquet, A., Gloeckler, R., and Lemoine, Y. (1992) *Biochem. J.* **287** (Pt 3), 685-690
43. Pai, C. H., and McLaughlin, G. E. (1969) *Can. J. Microbiol.* **15**, 809-810
44. Tomczyk, N. H., Nettleship, J. E., Baxter, R. L., Crichton, H. J., Webster, S. P., and Campopiano, D. J. (2002) *FEBS Lett.* **513**, 299-304

45. Sanishvili, R., Yakunin, A. F., Laskowski, R. A., Skarina, T., Evdokimova, E., Doherty-Kirby, A., Lajoie, G. A., Thornton, J. M., Arrowsmith, C. H., Savchenko, A., Joachimiak, A., and Edwards, A. M. (2003) *J. Biol. Chem.* **278**, 26039-26045
46. Lemoine, Y., Wach, A., and Jeltsch, J. M. (1996) *Mol. Microbiol.* **19**, 645-647
47. Izumi, Y., Kano, Y., Inagaski, K., Yamada, H. (1981) *Agr. Biol. Chem* **45**, 1983-1989
48. Van Arsdell, S. W., Perkins, J. B., Yocum, R. R., Luan, L., Howitt, C. L., Chatterjee, N. P., and Pero, J. G. (2005) *Biotechnol. Bioeng.* **91**, 75-83
49. Phalip, V., Kuhn, I., Lemoine, Y., and Jeltsch, J. M. (1999) *Gene* **232**, 43-51
50. Pinon, V., Ravanel, S., Douce, R., and Alban, C. (2005) *Plant Physiol.* **139**, 1666-1676
51. Nudelman, A., Dana, M.-M., Abraham, N., Dennis, F., and Vernon, W. (2004) *Tetrahedron* **60**, 1731-1748
52. Alan R Rendina, Taylor, W. S., Gibson, K., Lorimer, G., Rayner, D., Lockett, B., Kranis, K., Wexler, B., Marcovici-Mizrahi, D., Nudelman, A., Nudelman, A., Marsilii, E., Chi, H., Wawrzak, Z., Calabrese, J., Jia, W. H. J., Schneider, G., Lindqvist, Y., and Yang, G. (1999) *Pesticide Science* **55**, 236-247
53. Ashkenazi, T., Pinkert, D., Nudelman, A., Widberg, A., Wexler, B., Wittenbach, V., Flint, D., and Nudelman, A. (2007) *Pest Manag. Sci.* **63**, 974-1001
54. Ashkenazi, T., Widberg, A., Nudelman, A., Wittenbach, V., and Flint, D. (2005) *Pest Manag. Sci.* **61**, 1024-1033
55. Dey, S., Rubin, E. J., and Sacchettini, J. C. (2008) Structural characterization of 7, 8-diaminopelargonic acid synthase and dethiobiotin synthetase, two important enzymes in the biotin biosynthetic pathway in *Mycobacterium tuberculosis*, manuscript in preparation.
56. Chiang, S. L., and Mekalanos, J. J. (1998) *Mol. Microbiol.* **27**, 797-805

57. Keer, J., Smeulders, M. J., Gray, K. M., and Williams, H. D. (2000) *Microbiology* **146 (Pt 9)**, 2209-2217
58. Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A., and Duncan, K. (2002) *Mol. Microbiol.* **43**, 717-731
59. Grundy, W. E., Whitman, A.O., Rdzok, E.G., Hanes, M.E., Sylvester, J. C. (1952) *Antibiot. Chemother.* **II**, 399-408
60. Okami, Y., Kitahara, T., Hamada, M., Naganawa, H., and Kondo, S. (1974) *J. Antibiot. (Tokyo)*. **27**, 656-664
61. Hotta, K., Kitahara, T., and Okami, Y. (1975) *J. Antibiot. (Tokyo)*. **28**, 222-228
62. Kitahara, T., Hotta, K., Yoshida, M., and Okami, Y. (1975) *J. Antibiot. (Tokyo)*. **28**, 215-221
63. Hwang, K. (1952) *Antibiot. Chemother.* **II**, 453-459
64. Dey, S., Hu, Z., Xu, X. L., Sacchettini, J. C., and Grant, G. A. (2005) *J. Biol. Chem.* **280**, 14884-14891
65. Sugimoto, E., and Pizer, L. I. (1968) *J. Biol. Chem.* **243**, 2081-2089
66. Grant, G. (1989) *Biochem. Biophys. Res. Commun.* **165**, 1371-1374
67. Lamzin, V. S., Dauter, Z., Popov, V. O., Harutyunyan, E. H., and Wilson, K. S. (1994) *J Mol Biol* **236**, 759-785
68. Goldberg, J. D., Yoshida, T., and Brick, P. (1994) *J Mol Biol* **236**, 1123-1140
69. Razeto, A., Kochhar, S., Hottinger, H., Dauter, M., Wilson, K. S., and Lamzin, V. S. (2002) *J Mol Biol* **318**, 109-119
70. Ha, J. Y., Lee, J. H., Kim, K. H., Kim do, J., Lee, H. H., Kim, H. K., Yoon, H. J., and Suh, S. W. (2007) *J Mol Biol* **366**, 1294-1304
71. Dietrich, F. S., Mulligan, J., Hennessy, K., Yelton, M. A., Allen, E., Araujo, R., Aviles, E., Berno, A., Brennan, T., Carpenter, J., Chen, E., Cherry, J. M., Chung, E., Duncan, M., Guzman, E., Hartzell, G., Hunicke-Smith, S., Hyman, R. W.,

- Kayser, A., Komp, C., Lashkari, D., Lew, H., Lin, D., Mosedale, D., Davis, R. W., and et al. (1997) *Nature* **387**, 78-81
72. Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., and Myler, P. J. (2005) *Science* **309**, 436-442
73. Sugimoto, E., and Pizer, L. I. (1968) *J. Biol. Chem.* **243**, 2090-2098
74. Schuller, D. J., Grant, G. A., and Banaszak, L. J. (1995) *Nat. Struct. Biol.* **2**, 69-76
75. Thompson, J. R., Bell, J. K., Bratt, J., Grant, G. A., and Banaszak, L. J. (2005) *Biochemistry (Mosc.)* **44**, 5763-5773
76. Saski, R., and Pizer, L. I. (1975) *Eur. J. Biochem.* **51**, 415-427
77. Sasseti, C. M., Boyd, D. H., and Rubin, E. J. (2003) *Mol. Microbiol.* **48**, 77-84
78. Dey, S., Burton, R. L., Grant, G. A., and Sacchettini, J. C. (2008) *Biochemistry*, DOI 10.1021/bi800212b.
79. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res* **22**, 4673-4680
80. Pizer, L. I. (1963) *J. Biol. Chem.* **238**, 3934-3944

81. Peters-Wendisch, P., Netzer, R., Eggeling, L., and Sahm, H. (2002) *Appl. Microbiol. Biotechnol.* **60**, 437-441
82. Slaughter, J. C., and Davies, D. D. (1975) *Methods Enzymol.* **41**, 278-281
83. Rosenblum, I. Y., and Sallach, H. J. (1970) *Arch. Biochem. Biophys.* **137**, 91-101
84. Monod, J., Wyman, J., and Changeux, J. P. (1965) *J Mol Biol* **12**, 88-118
85. Becker, A., Liewald, J. F., Stypa, H., and Wegener, G. (2001) *Insect Biochem. Mol. Biol.* **31**, 381-392
86. Hurley, J. H., Faber, H. R., Worthylake, D., Meadow, N. D., Roseman, S., Pettigrew, D. W., and Remington, S. J. (1993) *Science* **259**, 673-677
87. Rodriguez-Aparicio, L. B., Guadalajara, A. M., and Rubio, V. (1989) *Biochemistry (Mosc).* **28**, 3070-3074
88. Phillips, N. F., Snoswell, M. A., Chapman-Smith, A., Keech, D. B., and Wallace, J. C. (1992) *Biochemistry (Mosc).* **31**, 9445-9450
89. Ballou, C. E. (1960) *Biochemical Preparations* **7**, 66-68
90. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
91. Matthews, B. W. (1974) *J Mol Biol* **82**, 513-526
92. Navaja, J. (1994) *Acta Crystallogr. A.* **50**, 157-163
93. Vagin, A. A., and Teplyakov, A. (1997) *J Appl Cryst* **30**, 1022-1025
94. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 432-438
95. Ke, H. (1997) *Methods Enzymol.* **276**, 448-461
96. Hendrickson, W. A., Smith, J. L., and Sheriff, S. (1985) *Methods Enzymol.* **115**, 41-55

97. Hendrickson, W., and Ogata, C. (1997) *Methods Enzymol.* **276**, 494-523
98. Sheldrick, G., and Gould, R. (1995) *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst Chem* **51**, 423-431
99. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr Sect. D Biol Crystallogr* **52**, 43-48
100. Cowtan, K. D., and Zhang, K. Y. (1999) *Prog. Biophys. Mol. Biol.* **72**, 245-270
101. Ioerger, T. R., and Sacchettini, J. C. (2002) *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 2043-2054
102. McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156-165
103. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240-255
104. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D. Biol. Crystallogr.* **54**, 905-921
105. Brunger, A. T. (1992) *Nature* **355**, 472-475
106. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) *J Mol Biol* **231**, 1049-1067
107. Hayward, S., and Berendsen, H. J. (1998) *Proteins* **30**, 144-154
108. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J Comput Chem* **25**, 1605-1612
109. Grant, G. A., Hu, Z., and Xu, X. L. (2002) *J. Biol. Chem.* **277**, 39548-39553
110. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature* **250**, 194-199
111. Aravind, L., and Koonin, E. V. (1999) *J Mol Biol* **287**, 1023-1040

112. Cho, Y., Sharma, V., and Sacchettini, J. C. (2003) *J. Biol. Chem.* **278**, 8333-8339
113. Gallagher, D. T., Gilliland, G. L., Xiao, G., Zondlo, J., Fisher, K. E., Chinchilla, D., and Eisenstein, E. (1998) *Structure* **6**, 465-475
114. Koon, N., Squire, C. J., and Baker, E. N. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8295-8300
115. Tan, K., Li, H., Zhang, R., Gu, M., Clancy, S. T., and Joachimiak, A. (2008) *J. Struct. Biol.* **162**, 94-107
116. Devedjiev, Y., Surendranath, Y., Derewenda, U., Gabrys, A., Cooper, D. R., Zhang, R. G., Lezondra, L., Joachimiak, A., and Derewenda, Z. S. (2004) *J Mol Biol* **343**, 395-406
117. Schreiter, E. R., Sintchak, M. D., Guo, Y., Chivers, P. T., Sauer, R. T., and Drennan, C. L. (2003) *Nat. Struct. Biol.* **10**, 794-799
118. Gouaux, J. E., Stevens, R. C., and Lipscomb, W. N. (1990) *Biochemistry (Mosc)*. **29**, 7702-7715
119. Grant, G. A., Hu, Z., and Xu, X. L. (2001) *J. Biol. Chem.* **276**, 17844-17850
120. Grant, G. A., Xu, X. L., and Hu, Z. (2000) *Biochemistry (Mosc)*. **39**, 7316-7319
121. Grant, G. A., Hu, Z., and Xu, X. L. (2005) *Biochemistry* **44**, 16844-16852
122. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng.* **8**, 127-134
123. Madej, T., Gibrat, J. F., and Bryant, S. H. (1995) *Proteins* **23**, 356-369
124. Holm, L., and Sander, C. (1996) *Sci. Justice* **273**, 595-602
125. Holm, L., and Park, J. (2000) *Bioinformatics* **16**, 566-567
126. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8600-8603

127. Kessler, D., Reteý, J., and Schulz, G. E. (2004) *J Mol Biol* **342**, 183-194
128. Burton, R. L., Chen, S., Xu, X. L., and Grant, G. A. (2007) *J. Biol. Chem.* **282**, 31517-31524
129. Kotaka, M., Ren, J., Lockyer, M., Hawkins, A. R., and Stammers, D. K. (2006) *J. Biol. Chem.* **281**, 31544-31552
130. Izumi, Y., Sato, K., Tani, Y., Ogata, K. (1975) *Agr. Biol. Chem.* **39**, 175-181
131. Stoner, G. L., and Eisenberg, M. A. (1975) *J. Biol. Chem.* **250**, 4029-4036
132. Sandmark, J., Eliot, A. C., Famm, K., Schneider, G., and Kirsch, J. F. (2004) *Biochemistry (Mosc)*. **43**, 1213-1222
133. Eliot, A. C., Sandmark, J., Schneider, G., and Kirsch, J. F. (2002) *Biochemistry (Mosc)*. **41**, 12582-12589
134. Kack, H., Sandmark, J., Gibson, K., Schneider, G., and Lindqvist, Y. (1999) *J Mol Biol* **291**, 857-876
135. Sandmark, J., Mann, S., Marquet, A., and Schneider, G. (2002) *J. Biol. Chem.* **277**, 43352-43358
136. Mann, S., and Ploux, O. (2006) *FEBS J* **273**, 4778-4789
137. Gibson, K. J., Lorimer, G. H., Rendina, A. R., Taylor, W. S., Cohen, G., Gatenby, A. A., Payne, W. G., Roe, D. C., Lockett, B. A., and Nudelman, A. (1995) *Biochemistry (Mosc)*. **34**, 10976-10984
138. Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) *Anal. Biochem.* **200**, 74-80
139. Peterson, E. A., and Sober, H. A. (1954) *Journal of American Chemical Society* **76**, 169
140. Rossmann, M. G., and Blow, D. M. (1962) *Acta Crystallogr.* **15**, 24-31
141. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126-2132

142. Reddy, V., Swanson, S. M., Segelke, B., Kantardjieff, K. A., Sacchettini, J. C., and Rupp, B. (2003) *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 2200-2210
143. Goldberg, J. M., and Kirsch, J. F. (1996) *Biochemistry (Mosc)*. **35**, 5280-5291
144. Alexeev, D., Alexeeva, M., Baxter, R. L., Campopiano, D. J., Webster, S. P., and Sawyer, L. (1998) *J Mol Biol* **284**, 401-419
145. Mani Tripathi, S., and Ramachandran, R. (2006) *J Mol Biol* **362**, 877-886
146. Toney, M. D., Hohenester, E., Keller, J. W., and Jansonius, J. N. (1995) *J Mol Biol* **245**, 151-179
147. Storici, P., De Biase, D., Bossa, F., Bruno, S., Mozzarelli, A., Peneff, C., Silverman, R. B., and Schirmer, T. (2004) *J. Biol. Chem.* **279**, 363-373
148. Mehta, P. K., Hale, T. I., and Christen, P. (1993) *Eur. J. Biochem. / FEBS* **214**, 549-561
149. Schneider, G., Kack, H., and Lindqvist, Y. (2000) *Structure* **8**, R1-6
150. Han, Q., and Li, J. (2004) *Eur. J. Biochem.* **271**, 4804-4814
151. Goto, M., Omi, R., Miyahara, I., Hosono, A., Mizuguchi, H., Hayashi, H., Kagamiyama, H., and Hirotsu, K. (2004) *J. Biol. Chem.* **279**, 16518-16525
152. Metzler, D. E., Ikawa, M., and Snell, E. E. (1954) *J. Am. Chem. Soc.* **76**, 648-652
153. Stoner, G. L., and Eisenberg, M. A. (1975) *J. Biol. Chem.* **250**, 4037-4043
154. Bhor, V. M., Dev, S., Vasanthakumar, G. R., and Surolia, A. (2006) *IUBMB Life* **58**, 225-233
155. Capitani, G., Eliot, A. C., Gut, H., Khomutov, R. M., Kirsch, J. F., and Grutter, M. G. (2003) *Biochim. Biophys. Acta* **1647**, 55-60
156. Hohenester, E., and Jansonius, J. N. (1994) *J Mol Biol* **236**, 963-968
157. Koshland, D. E. (1994) *Journal Angrew Chem Int Ed Eng* **33**, 2375

158. Koshland, D. E. (1958) *Proc. Natl. Acad. Sci. U. S. A.* **44**, 98-104
159. Fischer, E. (1894) *Ber. Dt. Chem. Ges.* **27**, 2985-2993
160. Jager, J., Moser, M., Sauder, U., and Jansonius, J. N. (1994) *J Mol Biol* **239**, 285-305
161. Hirotsu, K., Goto, M., Okamoto, A., and Miyahara, I. (2005) *The Chemical Record* **5**, 160-172
162. Toney, M. D., Pascarella, S., and De Biase, D. (1995) *Protein Sci.* **4**, 2366-2374
163. Fenn, T. D., Stamper, G. F., Morollo, A. A., and Ringe, D. (2003) *Biochemistry (Mosc)*. **42**, 5775-5783
164. Rando, R. R., and Bangerter, F. W. (1976) *J. Am. Chem. Soc.* **98**, 6762-6764
165. Shah, S. A., Shen, B. W., and Brunger, A. T. (1997) *Structure* **5**, 1067-1075
166. Irwin, J. J., and Shoichet, B. K. (2005) *J. Chem. Inf. Model.* **45**, 177-182
167. Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) *Adv Drug Deliv Rev* **46**, 3-26
168. Klebe, G. (2006) *Drug Discov Today* **11**, 580-594
169. Krell, K., and Eisenberg, M. A. (1970) *J. Biol. Chem.* **245**, 6558-6566
170. Huang, W., Jia, J., Gibson, K. J., Taylor, W. S., Rendina, A. R., Schneider, G., and Lindqvist, Y. (1995) *Biochemistry (Mosc)*. **34**, 10985-10995
171. Huang, W., Lindqvist, Y., Schneider, G., Gibson, K. J., Flint, D., and Lorimer, G. (1994) *Structure* **2**, 407-414
172. Kack, H., Sandmark, J., Gibson, K. J., Schneider, G., and Lindqvist, Y. (1998) *Protein Sci.* **7**, 2560-2566
173. Eisenberg, M. A., and Krell, K. (1979) *Methods Enzymol.* **62**, 348-352

174. Waldrop, G. L., Rayment, I., and Holden, H. M. (1994) *Biochemistry (Mosc)*. **33**, 10249-10256
175. Taylor, T. C., Backlund, A., Bjorhall, K., Spreitzer, R. J., and Andersson, I. (2001) *J. Biol. Chem.* **276**, 48159-48164
176. Rendina, A. R., Taylor, W. S., Gibson, K., Lorimer, G., Rayner, D., Lockett, B., Kranis, K., Wexler, B., Marcovici-Mizrahi, D., Nudelman, A., Nudelman, A., Marsilii, E., Chi, H., Wawrzak, Z., Calabrese, J., Huang, W., Jia, J., Schneider, G., Lindqvist, Y., and Yang, G. (1999) *Pesticide Science* **55**, 236-247
177. Cleland, W. W. (1979) *Anal. Biochem.* **99**, 142-145
178. Kantardjieff, K. A., Hocht, P., Segelke, B. W., Tao, F. M., and Rupp, B. (2002) *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 735-743
179. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) *EMBO J.* **1**, 945-951
180. Ramakrishnan, C., Dani, V. S., and Ramasarma, T. (2002) *Protein Eng.* **15**, 783-798
181. Endrizzi, J. A., Kim, H., Anderson, P. M., and Baldwin, E. P. (2004) *Biochemistry* **43**, 6447-6463
182. Cordell, S. C., and Lowe, J. (2001) *FEBS Lett* **492**, 160-165
183. Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B., and Rees, D. C. (1998) *J Mol Biol* **280**, 669-685
184. Pai, E. F., Krenzel, U., Petsko, G. A., Goody, R. S., Kabsch, W., and Wittinghofer, A. (1990) *EMBO J.* **9**, 2351-2359
185. Silva, M. M., Poland, B. W., Hoffman, C. R., Fromm, H. J., and Honzatko, R. B. (1995) *J Mol Biol* **254**, 431-446
186. Kack, H., Gibson, K. J., Lindqvist, Y., and Schneider, G. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5495-5500

187. Yang, G., Sandalova, T., Lohman, K., Lindqvist, Y., and Rendina, A. R. (1997) *Biochemistry* **36**, 4751-4760
188. Gibson, K. J., Lorimer, G. H., Rendina, A. R., Taylor, W. S., Cohen, G., Gatenby, A. A., Payne, W. G., Roe, D. C., Lockett, B. A., and Nudelman, A. (1995) *Biochemistry* **34**, 10976-10984
189. Schweins, T., Geyer, M., Scheffzek, K., Warshel, A., Kalbitzer, H. R., and Wittinghofer, A. (1995) *Nat Struct Biol* **2**, 36-44
190. Schweins, T., Geyer, M., Kalbitzer, H. R., Wittinghofer, A., and Warshel, A. (1996) *Biochemistry* **35**, 14225-14231
191. Alexeev, D., Alexeeva, M., Baxter, R. L., Campopiano, D. J., Webster, S. P., and Sawyer, L. (1998) *J Mol Biol* **284**, 401-419
192. Ploux, O., and Marquet, A. (1996) *Eur. J. Biochem.* **236**, 301-308
193. Ploux, O., and Marquet, A. (1992) *Biochem. J.* **283 (Pt 2)**, 327-331
194. Bhor, V. M., Dev, S., Vasanthakumar, G. R., Kumar, P., Sinha, S., and Surolia, A. (2006) *J. Biol. Chem.* **281**, 25076-25088

VITA

SANGHAMITRA DEY
(sanghamitradey@tamu.edu)

MAILING ADDRESS

Department of Biochemistry and Biophysics
Texas A&M University, College Station, 77843-2128
Phone # (979)-862-7639

EDUCATION

- Ph.D., Biochemistry, Texas A&M University, College Station, TX, 2008
- M.S., Biochemistry, Nagpur University, Maharashtra-India, 1997
- B.S., Biochemistry, Chemistry and Zoology, Nagpur University, Maharashtra-India, 1995

MAJOR PUBLICATIONS

- “D-3-phosphoglycerate dehydrogenase from *Mycobacterium tuberculosis* is a link between the Escherichia coli and mammalian enzymes.” **Dey, S.**, Hu, Z., Xu, X.L., Sacchettini, J.C., and Grant, G. A., *J. Biol. Chem.* 2005; 280(15):14884-91.
- “*Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase: extreme asymmetry in a tetramer of identical subunits.” **Dey, S.**, Grant, G. A. and Sacchettini, J. C., *J. Biol. Chem.* 2005; 280(15):14892-99.
- “The effect of hinge mutations on effector binding and domain rotation in *E.coli* D-3-phosphoglycerate dehydrogenase.” **Dey, S.**, Hu, Z., Xu, X.L., Sacchettini, J.C., and Grant, G. A., *J. Biol. Chem.* 2007; 282(25):18418-26
- “Structural analysis of substrate and effector binding in *Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase.” **Dey, S.**, Burton, R. L., Grant, G. A., and Sacchettini, J. C., *Biochemistry* (2008): DOI 10.1021/bi800212b)
- “Structural characterization of 7, 8-diaminopelargonic acid synthase and dethiobiotin synthetase, two important enzymes in the biotin biosynthetic pathway in *Mycobacterium tuberculosis*.” **Dey, S.**, Rubin, E., and Sacchettini J. C. (2008) manuscript in preparation.