STUDIES ON BIOMEDICALLY IMPORTANT *Mycobacterium tuberculosis* GENES INVOLVED IN NUTRIENT ASSIMILATION AND IMMUNE RESPONSE

Thesis submitted to

Department of Biochemistry
School of Life Sciences
University of Hyderabad
HYDERABAD
INDIA

for the degree of
Doctor of Philosophy

by

PRACHEE

Laboratory of Molecular and Cellular Biology
Centre For DNA Fingerprinting and Diagnostics
HYDERABAD
INDIA

Registration Number: 99LBPH14

2004

DEPARTMENT OF BIOCHEMISTRY
SCHOOL OF LIFE SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD
INDIA

CENTRE FOR DNA FINGERPRINTING
AND DIAGNOSTICS
HYDERABAD
INDIA

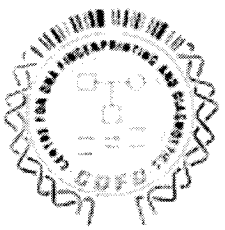
CERTIFICATE

This is to certify that this thesis entitled "Studies on biomedically important *Mycobacterium tuberculosis* genes involved in nutrient assimilation and immune response" comprises the work done by Ms. Prachee under my guidance at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India. This work is original and has not been submitted in part or full for any degree or diploma of any University.

Dr. A S Raghavendra
Dean, School of Life Sciences
University of Hyderabad

Dr. Seyed E Hasnain
Thesis Supervisor
CDFD, Hyderabad

Dr. C K Mitra
Head, Department of Biochemistry
University of Hyderabad



CENTRE FOR DNA FINGERPRINTING AND DIAGNOSTICS
HYDERABAD
INDIA

DECLARATION

I hereby declare that the work presented in this thesis entitled, "Studies on biomedically important *Mycobacterium tuberculosis* genes involved in nutrient assimilation and immune response" has been carried out by me under the supervision of Dr. Seyed E Hasnain at the Centre for DNA FIngerprinting and Diagnostics, Hyderabad, India. This work is original and has not been submitted in part or in full for any degree or diploma of any other University earlier.

Prachee

Prachee

Candidate

To my parents,

for being a constant source of inspiration. Your love and support motivated me to work hard and complete my research objectives.

Acknowledgements

The work presented in this thesis was accomplished with the help of many colleagues and associates. I would like to express my gratitude to each one of them.

I am deeply indebted to my research supervisor Dr Seyed E. Hasnain whose guidance, help, stimulating suggestions and encouragement enriched my graduate experience. I also thank him for his constructive criticism and excellent advice during the preparation of this thesis. Apart from his scientific skills, I have tried to imbibe at least a fraction of his excellent literary and oratory skills, an absolute requirement for sharing one's research with the scientific community. I also appreciate his generosity in sending students to various national and international conferences that are important eye openers in the field of research.

I express an extreme sense of gratitude to the earlier graduate students in the laboratory, Dr Noman Siddiqi, Dr Aruna Ramachandran, Dr Sudip Ghosh and Dr Priya Viswanathan, who introduced me to research in contemporary biology. They also offered an extremely conducive laboratory atmosphere I required to initiate my research work.

I extend my sincere thanks to Dr Bandi Aruna and Niteen Pathak for their contribution towards the completion of many experiments. I am grateful to Y. Sailu, whose computational skills generated several leads for my work. I thank Battu Aruna for her help with biophysical experiments as well as for being an admirable friend. Sincere thanks to Anil, Aisha and G. Savithri for extending their help in various ways. Thanks are due to Dr MD Bashyam whose knowledge and experience in the field of molecular biology was very useful in the formulation of my research objectives. I acknowledge Dr Sangita Mukhopadhyay for her help in the interpretation of the immunological data and Dr Shekhar C. Mande and Dr Akash Ranjan for many helpful discussions.

I am extremely grateful to Dr Abhijit A. Sardesai and Dr. J. Gowrishankar for sharing with me their excellent insights in the field of genetics. Their help was instrumental in the completion of related experiments.

I would like to acknowledge all other members of the Laboratory of Molecular and Cellular Biology for creating a highly inspiring and cheerful atmosphere. The scientific exposure I received at CDFD was a great learning experience.

I thank our research collaborators Dr KJR Murthy and Dr VM Katoch for helpful discussions. I also thank Dr Nasreen Z. Ehtesham for extending her lab facilities at the National Institute of Nutrition, Hyderabad and for providing useful suggestions.

I gratefully acknowledge the financial support of several institutions, the Council of Scientific and Industrial Research for providing a research fellowship and the Department of Biotechnology for funding the projects I worked for. I thank the "Global Alliance for TB Drug Development" for sponsoring my participation in the Gordon Research Conference on TB drug development at Oxford, United Kingdom. This Conference allowed me to interact with leaders in the field of tuberculosis research and gave me an invaluable opportunity to present my research before an international audience.

Heartfelt thanks are reserved for my fiance, Dr Rajat Prakash, who, despite his expertise in mathematical sciences, expressed profound interest in my work, took the effort to go through my thesis drafts and offered useful suggestions for improvement. I look forward to a great life ahead with Rajat.

Finally, I thank the almighty for his benevolence.

Prachee

CONTENTS IN BRIEF

CHAPTER 3

THE HYPOTHETICAL ORF Rv1885c OF *Mycobacterium tuberculosis* ENCODES A MONOFUNCTIONAL AroQ CLASS OF PERIPLASMIC CHORISMATE MUTASE

ABBREVIATIONS

^{0}C	Degree centigrade
aa	amino acid
AP	Alkaline Phosphatase
AS	Anthranilate synthase
AU	Arbitrary Units
BCG	Bacillus of Calmette Guerin
bp	base pair
cfu	colony forming unit
Ci	Curies
CM	Chorismate mutase
COG	Cluster of Orthologous Groups
CPB	Citrate Phosphate Buffer
dATP	2'-deoxyadenosine-5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DOTS	Directly Observed Treatment, short course
DTT	1, 4-Dithiothreitol
DtxR	Diphtheria Toxin Repressor
dTTP	2'-deoxythymidine-5'-triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetra acetic acid (disodium salt)
EtBr	Ethidium Bromide
Fur	Ferric uptake Regulator
Hsp	Heat shock protein
ICS	Isochorismate synthase
IdeR	Iron Dependent Regulator
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase pair
kDa	Kilodaltons
Mtb	*Mycobacterium tuberculosis*
MCS	Multiple cloning site
MeOH	Methanol
MES	Morpholino ethane sulfonic acid
MQ-H_2O	Milli Q water
mg	Milli gram (10^{-3} gram)
min	Minutes
ml	Millilitre
ml	Millilitres (10^{-3} litres)
mM	Millimolar
mmol	Millimoles (10^{-3} moles)
MPTR	Major Polymorphic Tandem Repeat
NCBI	National Centre for Biotechnology Information
ng	Nano gram (10^{-9} gram)
OD	Optical density
PCR	Polymerase chain reaction

PGRS	Polymorphic GC-Rich Sequences
Phe	Phenylalanine
PMSF	Phenyl Methyl Sulfonyl Fluoride
RBS	Ribosome Binding Site
RNA	Ribonucleic acid
rpm	rotations per minute
SDS	Sodium dodecyl sulfate
ss	signal sequence
TB	Tuberculosis
Trp	Tryptophan
Tyr	Tyrosine
μg	Micro gram (10^{-6} grams)
μM	Micromolar (10^{-6} molar)

CHAPTER 1

GENERAL INTRODUCTION

The identification of *Mycobacterium tuberculosis* by Robert Koch in 1882 as the causative agent of tuberculosis, the release of the drug rifampicin in 1970s and the sequencing of the *M. tuberculosis* genome in 1998 are three major events that have revolutionized tuberculosis research. In spite of these breakthroughs, the continued status of TB as the largest killer amongst infectious diseases is an issue of major concern. Although directly observed treatment short course [DOTS] chemotherapy exists to treat the disease, emergence of drug resistant strains has severely threatened the efficacy of treatment [Gleissberg, 1999; Siddiqi *et al*, 2002; 2004]. The recent sequencing of *M. tuberculosis* genome holds promise for the development of new vaccines and design of new drugs [Cole *et al*, 1998]. This is all the more possible when the information from the genome sequence is combined with proteomics and structural and functional genomics. Such an integrated approach has led to the birth of a new field of research christened "post genomics" that holds substantial promise for the identification of novel drug targets and can aid the development of new chemotherapeutic compounds to treat tuberculosis [Chakhaiyar and Hasnain, 2004]. The challenge before the scientific community therefore lies in elucidation of the wealth of information provided by the genome sequence and translation of the same into designing novel therapies for the disease.

1.1 Genome analysis *in silico* : Potentials of bioinformatics

The availability of the genome sequence of various organisms, prokaryotic as well as eukaryotic and even different strains of the same species, as in the case of *M. tuberculosis* and *H. pylori* has provided immense opportunities to the bioinformaticist to analyze the enormous data generated. With the development of efficient bioinformatic tools like PSI BLAST [Position Specific Iterated Basic Local Alignment Search Tool] and COG [Cluster of Orthologous Groups], many genes with unassigned functions are now being annotated quickly and possibly correctly [Muller *et al*, 1999]. The COG database [http://www.ncbi.nlm.nih.gov/COG/], is based upon phylogenetic classification of proteins encoded in complete genomes, has been used to reannotate the genomes of various organisms including *Mycobacterium tuberculosis*. Data mining has also been used for

accurate prediction of protein function [King *et al*, 2000] When the *Mycobacterium tuberculosis* genome was published [Cole *et al*, 1998], about 60% of the genes were labeled as unknowns or hypothetical. Today thanks to new bioinformatic and experimental approaches, this figure has been reduced to about 48%. In a major effort to reannotate the *M. tuberculosis* genome, 22 new protein coding genes most of which are shorter than the initially used 100 codon cutoff have been identified [Camus *et al*, 2002]. However, there are still about 376 proteins that do not have homologues in other bacteria and are possibly unique to *Mycobacterium tuberculosis*.

The bioinformatic approach has also helped in the functional classification of protein families. Computational methods for classification of cyclic nucleotide [cNMP]-binding protein and nucleotide cyclase superfamilies have shown that *Mycobacterium tuberculosis* encodes several more putative cNMP-binding proteins than other prokaryotes; the functions of many of which are unknown [McCue *et al*, 2000].

Identification of secreted proteins of *M. tuberculosis* by *in silico* analysis for the presence of N-terminal signal peptide has also resulted in the identification of about 52 proteins, with experimental proof for about 17 of them [Gomez *et al*, 2000]. Prediction of the existence of these secreted proteins is significant in the sense that it could lead to the identification of potentially antigenic proteins. Associated studies include use of advanced and accurate programmes that can predict T cell epitopes in the context of mycobacterial genome. Development of such programmes alongside *in vitro* methods for screening and confirming epitopes would increase the pace of development of new generation T cell epitope-based vaccines [Sbai *et al*, 2001].

Bioinformatics has also aided our understanding of the possible mechanisms of transcription termination in *Mycobacterium tuberculosis*. Use of specific programmes for deciphering hairpin formation at the 3' end of mRNAs has suggested that the bacterium does not depend much on hairpin formation for transcription termination [Washio *et al*, 1998], hence Rho dependent transcription termination assumes obvious significance. Nevertheless, the probable hairpin forming regions have been evaluated using specific

algorithms and experimental verification of the role of these sequences as transcription terminators suggest that there is a relaxed requirement for the U trail in the mycobacterial genome [Unniraman *et al*, 2001]. While an *in-silico* approach does help in predicting gene function and shortlisting genes for functional analysis, all *in-silico* predictions require a follow up or an *in-vivo* verification and validation.

1.2 Comparative genomics: Evolution of gene function

With the advancement of sequencing technology, the number of organisms being sequenced is increasing at a rapid pace. Amongst the *M tuberculosis* complex, *Mycobacterium tuberculosis* was the first organism to be sequenced followed *by M. leprae* and *M. bovis* [Cole *et al*, 1998, 2001; Garnier *et al*, 2003]. Sequencing *of M bovis BCG*, *M. marinum* and the non-pathogenic bacterium *M. smegmatis* are under various stages of completion and annotation. The availability of the genome sequence of related organisms has dramatically enhanced the study of comparative genomics [Brosch *et al*, 2000]. Comparison at the genetic level provides clues to the possible loss or gain of virulence of one species as compared to the other.

The genome of the leprosy bacillus and the massive deletions and pseudogenes associated with it makes us ponder about the reasons behind its reductive evolution [Cole *et al*, 2001]. The notion that the leprosy bacillus is slowly heading towards is being explored through the genome sequence of the bacterium to understand the complex implications of the loss of several important genes. A speculative article suggests that loss of sigma factors in *M. leprae* could also lead to the accumulation of pseudogenes [Madan Babu, 2003].

Chromosomal polymorphisms between *M. tuberculosis*, *M. bovis*, and *M. bovis* BCG Pasteur have also been studied using the BAC library of all the three [Brosch *et al*, 1998]. The study could detect a novel 12.7-kb segment present in *M. tuberculosis* but absent from *M. bovis* and *M. bovis* BCG Pasteur. Analysis of this segment showed that

this region contains genes whose probable products show low similarity to proteins involved in polysaccharide biosynthesis.

Comparative genomics of various mycobacterial species have revealed several deletions that exist in both virulent and avirulent species. Amongst these deletions, the most widely discussed is the RD1 region which is absent (deleted) from *M. bovis* BCG and *M. microti* [Pym *et al*, 2002]. Restoration of RD1 region in BCG by gene knock-in experiments has shown a clear transformation of the morphology to virulent type. Analysis of the deleted RD1 region by Pym *et al.* [2002] has shown that it encodes few proteins that are localized in the cell wall and others that code for immunodominant T cell antigens, ESAT-6 and CFP-10. Immunogenicity of the RD1 region has also been established using a synthetic peptide approach [Mustafa *et al*, 2002]. While immunocompetent mice could clear wild type BCG, BCG RD1 knock-in showed increased persistence and partial reversal of attenuation. Five other RD loci were also tested for their effect on the virulence of BCG. However, none of them showed a response similar to RD1 knock-ins. A recent report has also shown that recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis in mice [Pym *et al*, 2003]. In a reverse approach, knocking out the RD1 region from *M. tuberculosis* led to reduced ability of the resultant bacterium to infect human and mouse macrophage cell lines [Lewis *et al*, 2003]. These studies point towards the possibility of development of novel recombinant BCG vaccines.

Epidemiological studies for identification of strain specific signatures also form a vital part of comparative genomics as they guide the understanding of evolution and to an extent the pathogenesis of the organism [Tazi *et al*, 2002; Hasnain, 2003; Alland *et al*, 2003; Hasnain and Ahmad, 2004]. Attempts to assign specific signatures to strains from various geographic regions of the world utilizing different techniques like FAFLP [Fluorescent Amplified fragment Length Polymorphism], MIRU VNTR, spoligotyping, IS1081/6110 typing etc have been reported [Siddiqi *et al*, 2001; Ahmed *et al*, 2003, 2004]. The availability of the genome sequence has additionally helped us finemap the unique signatures associated with virulent strains. A noteworthy study suggests that *M.*

tuberculosis strains from HIV seropositive and seronegative patients are very distinct [Ahmed *et al*, 2003]. Availability of a database on strain specific signatures would also help in identification of strains responsible for outbreaks [Raman *et al*, 2000].

1.3 The phenomenon of mycobacterial latency

A major bottleneck in the development of a tuberculosis drug or vaccine has been the ability of mycobacterium to enter a phase of dormancy inside the host cells. Post genomic studies have shown that the glyoxylate shunt enzyme isocitrate lyase is crucial for mycobacterial persistence in macrophages and mice [McKinney *et al*, 2000]. Lipids are the chief carbon source during dormant phase of infection and that perhaps accounts for the existence of a large number of genes coding for proteins of lipid metabolism pathway in the genome of *M. tuberculosis*. The structure of isocitrate lyase has been solved [Sharma *et al*, 2000] but it has not been feasible to design an inhibitor for the enzyme as the amount of inhibitor required to block this actively produced enzyme would be so high that it might even prove toxic to the host.

Mycobacterial latency has often been linked to the hypoxic response of the host. However, response of *M. tuberculosis* to hypoxic conditions remained poorly understood for quite sometime. Though microarray analysis has given considerable clues, a recent report is another breakthrough in the understanding of the factors that control hypoxic response in *M. tuberculosis*. The transcription factor that regulates hypoxic response in *M. tuberculosis* has been identified [Sherman *et al*, 2001; Park *et al*, 2003]. Verification of the role of this putative transcription factor [Rv3133c/DosR] was carried out by targeted disruption of the gene followed by a comparative analysis of the transcriptome of the mutant as well as the wild type bacterium. Induction of the hypoxia responsive genes was not observed in the case of *M. tuberculosis* strain with a mutated *dos*R. Conserved DosR binding sites were also observed in the upstream sequences of all the genes under its regulatory control. The host response in the latent phase has also been studied in some detail. It has been shown that the CD8+ subset of T cells is active and produces IFN gamma only in the latent phase of infection [van Pinxteren *et al*, 2000].

Our understanding of the physiology of the bacterium under latent phase has also been increased by a proteomics approach [identification of highly expressed proteins under hypoxic conditions] using techniques like two-dimensional electrophoresis, and protein signature peptide analysis by liquid chromatography-mass spectrometry [Rosenkrands *et al*, 2002]. Targeting latent bacteria is indeed a challenge and using models depicting latency, the genome sequence has indeed helped in the identification of some genes that allow the bacteria to control this phase. Targeting these genes would help in combating the latent phase of infection.

1.4 Whole genome microarray: A clue to gene expression under different conditions

Gene regulation in *Mycobacterium tuberculosis* is tightly controlled by environmental signals and in this context microarray has become a very useful tool for identification of genes expressed under a variety of growth conditions as well as elucidation of the function of regulatory proteins by mutation analysis. Drug induced alterations in gene expression have been studied by DNA microarray technology. Isoniazid induced genes have been studied using microarrays [Wilson *et al*, 1999; Betts *et al*, 2003]. It was observed that isoniazid leads to the expression of proteins physiologically relevant to the drug's mode of action, including an operonic cluster of five genes encoding type II fatty acid synthase enzymes and *fbp*C, which encodes trehalose dimycolyl transferase. Other genes like *efp*A, *fad*E23, *fad*E24, and *ahp*C, which are likely to mediate processes that are linked to the toxic consequences of the drug were also expressed. Identification of these genes may define new drug targets.

Mycobacterium faces a microaerophilic environment within the host. Consequently, the hypoxic response genes of the bacterium assume obvious significance. Whole genome microarray studies identified >100 genes whose expression is rapidly altered by defined hypoxic conditions [Sherman *et al*, 2001]. Many of the genes involved in aerobic metabolism were repressed and a putative two-component sensor response regulator

pair Rv3133c/Rv3132c was induced by hypoxia. Later studies pointed out that Rv3133c is a crucial regulator of hypoxic response of *M. tuberculosis* [Park *et al*, 2003].

Microarray alongwith proteomics has been used to investigate the response *of M. tuberculosis* to nutrient starvation [Betts *et al*, 2002]. This has provided evidence for slowdown of the transcription apparatus, energy metabolism, lipid biosynthesis and cell division in addition to induction of the stringent response and several other genes that may play roles in maintaining long-term survival within the host. This work has therefore generated a model with which one can search for agents active against persistent *M. tuberculosis* and has revealed a number of expressed genes, the products of which could serve as potential targets.

The transcriptional response to low pH which mimics the environment of the phagocytosed mycobacteria has also been studied by microarray technology [Fisher *et al*, 2002]. Eighty-one genes were differentially expressed >1.5-fold, including many involved in fatty acid metabolism. The most highly induced genes showed homology with nonribosomal peptide synthetases/polyketide synthetases.

Microarray based information is available on regulatory proteins as well as several transcription factors of *M. tuberculosis*. The *M. tuberculosis* genome has thirteen annotated sigma factors of which the ECF sigma factor, SigE, has been shown to contribute to pathogen survival following several distinct stresses [Wu *et al*, 1997]. An alternative sigma factor, SigH has also been shown to play a role in the mycobacterial stress response [Raman *et al*, 2001]. Survival of an *M. smegmatis sig*H, *sig*E double mutant was found to be markedly decreased following 53^0C heat shock and exposure to cumene hydroperoxide, an agent generating peroxides *in vivo* [Fernandes *et al*, 1999]. Expression of the second gene in the *sig*H operon is required for complementation of the *sig*H stress phenotypes [Fernandes *et al*, 1999]. Roles of the two sigma factors, SigH and SigE in global gene expression profiles of *Mycobacterium tuberculosis* has been studied using microarray technology. The role of *sig*H in controlling cellular response to the oxidizing agent diamide has been studied using both quantitative reverse

transcription-PCR and microarray [Manganelli *et al*, 2002]. In this study, about 48 genes were identified whose expression increased after exposure of *M. tuberculosis* to diamide. Of these, 39 were not induced in the sig*H* mutant, suggesting their direct or indirect regulation by SigH. Many of these genes include thioredoxin, thioredoxin reductase and enzymes involved in cysteine and molybdopterine biosynthesis which are related to thiol metabolism. SigH was also shown to be the regulator of other sigma factors such as SigB, SigE, and SigH itself. A related study has also shown that the *M. tuberculosis sig*H is dispensable for bacterial growth and survival within the host but is required for the production of immunopathology and lethality [Kaushal *et al*, 2002].

The role of *M. tuberculosis* sigma factors in stress response suggest that they could be involved in pathogenicity of the bacterium [Manganelli *et al*, 2001]. Microarray studies were carried out with *sig*E mutant strains and the SigE regulon was defined after confirmation with RT-PCR analysis. It was shown that a functional SigE is required for full expression of *sig*B and for its induction after SDS exposure but not after heat shock. Genes no longer induced in the absence of SigE were also identified. These genes encode proteins belonging to different classes including transcription regulators, enzymes involved in fatty acid degradation and classical heat shock proteins. Apart from identifying the genes under the regulatory control of SigE, it was also shown that *sig*E mutant is defective in its ability to grow inside both human and murine unactivated macrophages and is more sensitive, than the wild-type strain, to the killing activity of activated murine macrophages [Manganelli, 2001].

Another well-characterized global transcription regulator of *Mycobacterium tuberculosis* is the Iron dependent Repressor [IdeR]. A post genomic approach, in conjunction with microarray technology, has defined the IdeR regulon [Rodriguez *et al*, 2002]. Members of the IdeR regulon encode a variety of proteins, including putative transporters, proteins involved in siderophore synthesis and iron storage, members of the PE/PPE family, enzymes of aromatic amino acid biosynthesis, membrane proteins involved in virulence, transcription regulators, and enzymes involved in lipid metabolism. A PPE member, Rv2430c that is overexpressed in an *M. tuberculosis* IdeR mutant strain has recently

been characterized as an immunodominant antigen [Choudhary *et al*, 2003]. IdeR itself is an attractive drug target as IdeR deletion is lethal for *Mycobacterium tuberculosis*. A high-resolution structure of IdeR has been solved [Feese *et al*, 2001].

1.5 Genomics vs. proteomics: Answers given by proteomics

While the end of the 20th century saw the birth of genomics, the beginning of the 21st century is witnessing the birth and a rapid development of proteomics. Proteomics complements genomics in determining genes that are expressed and translated into a functional protein. Transcriptomics and proteomics are two extremely important tools for the identification of novel drug targets and proteins that could serve as vaccine candidates. These twin technologies are considerably influencing the study of mycobacterial pathogenesis and virulence.

A proteomics approach has suggested that subtle differences in bacterial culture conditions may have important implications for gene expression and protein production in mycobacteria. Study of nutrient starvation in *M. smegmatis* using a proteomics approach has led to the identification of a non specific DNA binding protein Dps that has a role in reducing DNA damage by non specific binding to DNA [Gupta *et al*, 2002]. The *M. tuberculosis* homologue of Dps could thus be performing a similar role. Proteomics study of *M. tuberculosis* using two-dimensional electrophoresis and matrix-assisted laser desorption ionization and nano-electrospray mass spectrometry has also been used to identify 6 new protein of *M. tuberculosis* not predicted by genomics method [Jungblut *et al*, 2001]. A proteomics approach has even been used to identify the T cell antigens of *Mycobacterium tuberculosis* [Covert *et al*, 2001].

Bioinformatics has also aided the field of proteomics in the context of development of methods for clustering protein sequences and building families of potentially orthologous sequences. *In-silico* analyses of protein similarity and genetic neighbourhood has also been used to understand functional relation between proteins [Tekaia *et al*, 1999]. The study observed linkage between genes coding for membrane proteins and those

involved in fatty acid metabolism. ESAT6 like transport proteins were also found to occur with the *mce* operon. The analysis also revealed that very few transport proteins are present in the bacterium, which could be a reflection of its intracellular lifestyle.

1.6 Intracellular survival: Assimilation of nutrients

Despite intensive research efforts focussed on understanding the molecular basis of pathogenicity of *Mycobacterium tuberculosis*, several important issues such as the virulence determinants of the bacterium and the response of the host to the infection etc still remain very poorly understood. An important role of the host defense system is to reduce the amount of available nutrients in the vicinity of the pathogen to a bare minimum, wherein it has to survive for long periods of time [Betts *et al*, 2002; Wayne *et al*, 1994; Primm *et al*, 2000]. Circumvention of the host defense system and acquisition of essential nutrients from such an environment are important aspects of survival of pathogenic organisms.

Amino acid auxotrophs of *M. tuberculosis* are not known to survive or multiply in macrophages [Bange *et al*, 1996; Gordhan *et al*, 2002; Hondalus *et al*, 2000; Smith *et al*, 2001], which suggest that these amino acids are perhaps not available within the compartment of the macrophage in which the bacteria reside. Another important nutrient that is not available to *M. tuberculosis* inside the macrophage is iron [Rodriguez and Smith, 2003]. Iron is required for many of the cellular and biochemical functions of living systems. Ionic forms of iron especially iron [III], its most common state, are very insoluble at neutral or alkaline pH [Baker *et al*, 2003]. The need for iron is reflected by its requirement as a cofactor for several enzymes that facilitate electron transport, oxygen transport and other life-sustaining processes. The mammalian host defense system tries to limit the availability of iron to pathogens by complexing iron to molecules like transferrin, lactoferrin, ferritin and heme [Finkelstein *et al*, 1983]. In the macrophages, the residence of *M. tuberculosis*, the levels of iron are believed to be further brought down by the effects of the immunomodulator IFNγ, which does so by decreasing the levels of expression of transferrin receptor [Byrd and Horwitz, 1993]. Hence, acquisition

of iron from an iron limiting milieu presents a significant challenge to the pathogen and is a very important factor that largely determines the pathogenicity of *Mycobacterium*.

1.7 Iron acquisition by mycobacteria

To circumvent the solubility problem of iron, many microbes, plants and even higher organisms synthesize and utilize very specific low molecular weight iron chelators called siderophores ["iron bearers"]. When grown under iron deficient conditions, mycobacteria synthesize and secrete siderophores in excess of their own dry cell weight to sequester and solubilize iron [Ratledge, 2004]. The genes responsible for siderophore biosynthesis are expressed under low iron conditions [Neilands, 1995]. In fact, synthesis of siderophores is tightly regulated by the external environment i.e. they are produced strictly under iron limiting conditions.

1.7.1 Mycobacterial siderophores: Structure and biosynthesis

Siderophores of varied structures are employed by different organisms to competitively bind iron and gain selective growth advantages [Clarke *et al*, 2001; Raymond *et al*, 2003; Neilands *et al*, 1995]. Mycobacterial siderophores can be classified into various types depending upon the structure of the molecule or its eventual location, that is whether it is membrane bound or secreted into the external medium. According to one form of classification, the secreted forms of mycobacterial siderophores are referred to as exochelins and the cell wall associated forms are termed as mycobactins. However, lately more emphasis has been laid on the structure based classification of siderophores [De Voss *et al*, 1999]. Accordingly, mycobacterial siderophores can be categorized into two kinds, mycobactins or the salicylate derived siderophores and exochelins or the non salicylate containing peptidic siderophores. Exochelins are related to peptido-hydroxamate siderophores and are usually water soluble. Mycobactins are characterized by the presence of a phenyloxazolidine ring and are usually cell wall associated but presence of specific functional groups may render even the mycobactins water soluble. The secreted forms of mycobactins have been referred to as carboxymycobactins [Lane

et al, 1998]. However, this term may not be very appropriate as methyl ester as well as free acid forms of these compounds have been reported [Wong *et al*,1996]. Some species of mycobacteria produce both the types while others produce either of the two. To cite examples, *M. tuberculosis* produces only the mycobactin type of siderophore, *M. vaccae* produces only the exochelin type and *M. smegmatis* produces mycobactinss as well as exochelins [Ratledge, 2004].

1.8 Iron dependent regulatory proteins of *M. tuberculosis*

Iron metabolism is stringently controlled and connected with several regulatory systems. The published genome sequence of *M. tuberculosis* has listed four regulatory proteins possibly involved in the iron metabolism of the organism [Cole *et al*, 1998]. These four proteins are, FurA [Rv1909c], FurB [Rv2359], IdeR [Rv2711] and SirR [Rv2788]. FurA [150aa] has been annotated as a ferric uptake regulatory protein based upon similarity to family of Fur proteins. It shows 32.3% identity in a 133 aa overlap when compared to the *Legionella pneumophila* ferric uptake protein. FurB [130aa] has also been listed as ferric uptake regulatory protein based upon similarity to *E. coli* ferric uptake regulation protein [148aa]. The two share 37.9% identity in a 132 aa overlap. FurB is also similar to *Streptomyces coelicolor* Fur protein [59.4% identity in 133 aa overlap] IdeR [230aa] has been annotated as iron dependent repressor based upon homology to the well characterized DtxR or diphtheria toxin repressor from *Corynebacterium diphtheriae* [57.85% identity in a 230 aa overlap]. SirR [228aa] has been annotated as a probable iron dependent transcriptional repressor based upon similarity to *Staphylococcus epidermidis* putative iron dependant repressor SirR [214 aa]. There is a 33.7% identity in 193 aa overlap between the two. SirR also has some similarity to *M. tuberculosis* iron-dependant regulator IdeR [26.5% identity in 223 aa overlap].

FurA and FurB are well documented in *E. coli* as ferric uptake regulatory proteins [Bagg and Neilands, 1987]. In many gram-negative bacteria, iron regulated virulence determinants are regulated by Fur [Hassan and Sun, 1992; Bijlsma *et al*, 2002; Ochsner and Vasil, 1996]. In spite of the reasonable degree of homology with the *E. coli*

counterparts, in *M. tuberculosis*, FurA and FurB have not been shown to be directly involved in regulation of genes required for iron acquisition. However, it has been shown that *M. tuberculosis* FurA does regulate the expression of *kat*G as well as autoregulates its expression [Pym *et al*, 2001; Zahrt *et al*, 2001]. Even SirR, which shows similarity to the iron dependent regulator from *Staphylococcus epidermidis* has not been shown to be directly involved in a similar phenomenon in *Mycobacterium tuberculosis*. The protein for which maximum evidence exists for the possible role as an iron dependent regulator of *M. tuberculosis* is IdeR.

1.9 Iron-dependent regulation of aromatic amino acid biosynthesis in *M. tuberculosis*

Few enzymes of the aromatic amino acid biosynthesis pathway have also been reported to be under the regulatory control of IdeR [Gold *et al*, 2001; Rodriguez *et al*, 2002]. The genes that appear in published reports as IdeR regulated include prephenate dehydratase [pheA], anthranilate synthase/isochorismate synthase [*trp*E2/*mbt*I], probable phosphoribosyl-AMP pyrophosphatase [h*is*E], Probable imidazole glycerol-phosphate dehydratase [*his*B like] and phosphoribosyl-AMP 1,6 cyclohydrolase [*his*I]. *trp*E2/*mbt*I as well as *phe*A have been detected by microarray experiments as IdeR regulated genes [Rodriguez *et al*, 2002]. The IdeR regulation of genes encoding enzymes involved in aromatic amino acid biosynthesis suggests a possible involvement of their cognate gene products in iron assimilation. It is to be noted that all iron acquisition systems possess aromatic rings.

Rv2386c has been predicted to code for an isochorismate synthase [Quadri *et al*, 1998] that can catalyze the conversion of chorismate to isochorismate, the precursor for salicylate. Since the mycobacterial siderophore mycobactin is a salicylate derived siderophore, the IdeR regulation of Rv2386c may not be simply coincidental [Quadri *et al*, 1998; De Voss *et al*, 2000]. This enzyme was earlier annotated as an anthranilate synthase [*trp*E2] on the basis of observed homology between the two classes of

enzymes [Cole *et al*, 1998]. It is important to note that anthranilate is just an amino analogue of salicylate and hence could also serve as a precursor for salicylate.

*phe*A [Rv3838c, encoding prephenate dehydratase] is also under the regulatory control of IdeR [Gold *et al*, 2001]. PheA catalyses a committed step in the biosynthesis of the aromatic amino acid phenylalanine.

These findings suggest that either aromatic amino acids are directed towards the biosynthesis of iron acquisition systems or that IdeR, being a global regulator has an independent role in governing the expression of aromatic amino acid biosynthesis enzymes.

1.10 The aromatic amino acid biosynthesis pathway of *M. tuberculosis* and its role in pathogenicity of the bacterium

A large number of secondary metabolites containing aromatic rings are derived from intermediates of aromatic amino acid biosynthesis or the aromatic amino acids themselves [Pittard J, 1987]. Knowledge of the biosynthesis and regulation of aromatic amino acids in actinomycetes, including *Mycobacterium tuberculosis* is limited but will be of importance in understanding the mechanism of pathogenesis of the bacterium.

Parish and Stoker [2002] have reported that the aromatic amino acid biosynthesis pathway is essential for *M. tuberculosis*. In this study, the authors tried to disrupt *aro*K of *M. tuberculosis* by a two-step homologous recombination procedure [aroKDelta::hyg]. The authors however observed that *aro*K could not be deleted without providing a second functional copy of *aro*K in the genome. *aro*K codes for shikimate kinase that catalyses the conversion of shikimate to shikimate-3-phosphate, the fifth step in the formation of chorismate from erythrose-4-phosphate. This study suggested that the chorismate pathway is essential for *M. tuberculosis* and genes of the pathway cannot be disrupted even in the presence of suitable media supplements, which is possible in the case of *E. coli*. It is likely that due to the existence of multiple branchpoints of the

chorismate pathway, unless all the end products are supplied, growth cannot be rescued. The authors also tried to isolate mutants that could grow with chorismate as a media supplement but were unable to do so. This also suggested that there may be other branchpoints of the pathway and media supplementation of all the end products are required. It is also possible that these enzymes are bifunctional and even if one media supplement is provided, the other pathway that is affected is still required for survival. Bifunctionality of *aro*K in *E. coli* has earlier been reported by Vinella *et al* [1996].

Considering the fact that there are no mammalian equivalents of aromatic amino acid biosynthetic pathway enzymes, these enzymes could also serve as novel and highly specific drug targets.

1.11 The annotation status of aromatic amino acid biosynthesis enzymes of *M. tuberculosis*

The initial genome sequence of *M. tuberculosis* listed about 16 genes whose products could be involved in aromatic amino acid biosynthesis [Table 1.1].

This list was later increased by the annotation of a few more genes like a glutamine amidotransferase [*trp*G, Rv0013] and chorismate mutase [Rv1885c, Rv0948c] after more advanced protein family signature determining programmes like COG and Pfam were developed [Tatusov *et al*, 2001; Bateman *et al*, 2000, 2004].

1.12 Novel drugs and vaccine development: Is it a near possibility

The genomics approach has led to the identification of genes/proteins important to the pathogen and is fast progressing towards the process of a viable vaccine development for tuberculosis. However, the challenge lies in selecting the most probable candidate immunogens or virulence genes from amongst the 4000 odd genes of *M. tuberculosis*. The availability of the genome sequence of pathogens in combination with a better

Table 1.1:Annotated aromatic amino acid biosynthesis enzymes of *Mycobacterium tuberculosis*

ORF	gene	Function
Rv3227	*aro*A	3-phosphoshikimate1-carboxyvinyl transferase
Rv2538c	*aro*B	3-dehydroquinate synthase
Rv2537c	*aro*D	3-dehydroquinate dehydratase
Rv2552c	*aro*E	shikimate 5-dehydrogenase
Rv2540c	*aro*F	chorismate synthase
Rv2178c	*aro*G	DAHP synthase
Rv2539c	*aro*K	shikimate kinase I
Rv3838c	*phe*A	prephenate dehydratase
Rv3754	*tyr*A	prephenate dehydrogenase
Rv1613	*trp*A	tryptophan synthase a chain
Rv1612	*trp*B	tryptophan synthase b chain
Rv1611	*trp*C	indole-3-glycerol phosphate synthase
Rv2192c	*trp*D	anthranilate phosphoribosyltransferase
Rv1609	*trp*E	anthranilate synthasecomponent I
Rv2386c	*trp*E2	anthranilate synthase component I
Rv0346c	*aro*P2	probable aromatic amino acid Permease

understanding of protective pathogen specific immune responses has laid the foundations for the development of new generation recombinant vaccines.

1.12.1 Recombinant BCG vaccines

Recombinant BCG vaccines are gaining popularity as a preventive cure for tuberculosis on account of the observed protective efficacy of the same in model infections. A recombinant BCG vaccine [rBCG30] expressing and secreting the 30-kDa major secretory protein of *Mycobacterium tuberculosis,* has been shown to increase survival after challenge, as compared to parental BCG, in the guinea pig model of pulmonary TB [Horwitz *et al*, 2000, 2003]. Recombinant BCG expressing ESAT6 has also been shown to provide mild protection against tuberculosis infection [Bao *et al*, 2003]. In a very promising recent report, recombinant BCG bearing the entire RD1 deletion region along with the flanking region which is predicted to contain the secretory signal has shown an enhanced protection against TB infection in mice [Pym *et al*, 2003]. This suggests that the inability of BCG to elicit a protective immune response is due to the loss of important regions of the genome, restoration of which could make a better vaccine than BCG.

1.12.2 DNA vaccines for tuberculosis

DNA vaccines are a likely possibility considering the report of hsp65 based DNA vaccine providing protection in mice against intravenous infection of tuberculosis [Lowrie *et al*, 1999]. However, in another study where the major antigens, ESAT6 and Hsp60 were injected into mice as DNA vaccines, protection was observed only in the case of intravenous bacterial infection and not in the case of the more general aerosol mode of infection [Taylor *et al*, 2003].

1.12.3 *M. tuberculosis* amino acid auxotrophs as potential vaccine candidates

The recent advancements in generation of *M. tuberculosis* knockouts and mutants have tremendous potential for the development of novel vaccines [Pavelka *et al*, 2003; Hondalus *et al*, 2000; Jackson *et al*, 1999; Sambandamurthy *et al*, 2002]. It is no longer

required to attenuate the bacterium using serial passage as was done in the case of BCG. Smith *et al* [2001] also used this concept to test *M. tuberculosis* auxotrophs as potential vaccine candidates. Briefly, the authors studied the virulence and vaccine potential of *M. tuberculosis* strains containing defined mutations in genes involved in methionine [*metB*], proline [*proC*], or tryptophan [*trpD*] biosynthesis. While the *metB* mutant was a prototroph, the *proC* and *trpD* mutants were auxotrophic for proline and tryptophan, respectively. It was observed that while the *met*B mutant strain of *M. tuberculosis* survived inside murine macrophage for over 10 days, about 90% of the *pro*C and *trp*D mutants were killed. In SCID mice, H37Rv and *met*B mutant strains were equally virulent, the *pro*C mutant was attenuated and the *trp*D mutant was avirulent. The authors also carried out vaccination studies with these strains. They observed that while a *pro*C mutant gave protection equivalent to that imparted by BCG, the protection provided by the *trp*D mutant *M. tuberculosis* strain was stronger that BCG. The authors have concluded from these studies that the *proC* and *trpD* mutants are promising vaccine candidates.

Parish [2003] has also investigated the ability of proline, histidine and tryptophan auxotrophs of *M. tuberculosis* to survive different starvation conditions. Parish and co-workers had earlier reported that a tryptophan auxotroph of *M. tuberculosis* is severely attenuated in both macrophage and mouse models of infection [Smith *et al*, 2001]. However, in that report there was no evidence to ascertain if the bacterial clearance was caused by death due to tryptophan starvation or by active killing of the bacteria by the macrophage. Hence, Parish further studied the survival ability of *M. tuberculosis* tryptophan auxotroph upon withdrawal of tryptophan in axenic cultures. The author observed that while proline auxotrophs could survive single amino acid starvation, tryptophan and histidine auxotrophs could not. However, all these strains were able to survive complete starvation. Additionally, tryptophan and histidine auxotrophs also showed a restricted growth in THP-1 cell lines.

1.12.4 Effective chemotherapy for tuberculosis

While development of a new vaccine for tuberculosis is the prime agenda of tuberculosis research, control of tuberculosis by chemotherapy is also the focus of many laboratories. The increasing information on the biology of *M. tuberculosis* and the availability of the genome sequence has provided an entire list of novel drug targets to kill the pathogen. Identification of new drug targets is important as most of the existing targets have been altered [mutated] in the bacterium [Siddiqi *et al*, 1998, 2002]. Genomics has given us a brief idea about the complete sets of genes, enzymes and proteins possessed by the bacterium and subsequent research has already uncovered several leads for new drug targets.

Genomics has the potential for identifying new drug targets as well as to understand why some drugs are without effect whereas others are fully active. *M. tuberculosis* genome has been known to possess quite a few probable drug efflux pumps, which might have a role in imparting drug resistance to the bacterium. A recent report has also shown that one of the probable drug efflux pump is overexpressed in a drug resistant clinical isolate of *Mycobacterium tuberculosis* [Siddiqi *et al*, 2004]. These pumps could themselves serve as drug targets. Other important potential drug targets in microbial genomes could be outer-membrane proteins, host-interaction factors, permeases, enzymes of intermediary metabolism, systems for DNA replication, transcription, and repair, translation apparatus etc. The probable role and essentiality of all these genes should be studied in detail to develop comprehensive multi-pronged combinations of antibiotics that would effectively fight bacterial pathogens [Galperin and Koonin, 1999; Smith, 2003]. Virulence genes, uncharacterized essential genes, species specific genes, unique enzymes, drug transporters etc are also the current focus of research for new drug target identification.

It is paradoxical that while there has been no dearth in the identification of drug targets, no new drug has been introduced into the market even in the post genomic era. However, the post genomic era has indeed suggested a few promising drugs for the

disease. One of them includes the nitroimidazopyrans, a class of drugs that inhibit the biosynthesis of protein and cell wall lipids [Stover *et al*, 2000]. This drug could target replicating as well as non-dividing bacteria and was shown to have potent bactericidal activity against multidrug resistant bacteria. It is presumed that the application of functional genomics tools, such as microarray and proteomics, in combination with modern approaches, such as structure-based drug design and combinatorial chemistry will lead to the development of new drugs that are not only active against drug-resistant TB but also can shorten the course of TB therapy.

In short, sequencing alone will not bring about a paradigm shift in medical research but a thorough understanding of the implications of the same will definitely ensure one. The initial outcome of any genome sequencing project is just an unpunctuated string of 4 letters [A,T,G,C]. Genomics involves compiling this information and post genomics involves mining useful information from the genome. The genome approach has been applied to tuberculosis and the targets set by commercial companies and research institutions suggest optimistically that a permanent cure for the disease is not a distant vision.

1.13 Ensuring a better diagnosis for tuberculosis

The most popular diagnostic test for tuberculosis, the tuberculin skin test has limitations on account of background obtained in BCG vaccinated individuals. One of the major problem of TB diagnosis is the inability of the present methods to destinguish active infection from relpased, pulmonary and extrapulmonary infection and also distinguish TB cases from BCG vaccinated controls. Development of new PCR based methods for TB diagnosis are gaining more popularity on account of their higher specificity and sensitivity [Su, 2002]. Comparative genome sequencing has given some clues on the regions that are unique to *M. tuberculosis* and are deleted in the case of BCG. These regions, which also encode prominent T antigens like ESAT6, are being exploited to develop a specific antigen based diagnostic assay for tuberculosis [Gennaro, 2000]. Recent studies from our laboratory have shown that specific antigens of *M. tuberculosis*

can be used to identify TB infection [Choudhary *et al*, 2003; Chakhaiyar *et al*, 2004; Banerjee *et al*, 2004].

1.14 Objectives of the present work

The present study was undertaken with the prime objective of identification and characterization of novel genes that could help *M. tuberculosis* to survive inside the nutrient deficient conditions presented by the host. Iron being an essential micronutrient, the focus of the study was on identification of genes of *M. tuberculosis* that are either directly involved in iron acquisition or are responsive to iron stress. It was also considered worthwhile studying the aromatic amino acid biosynthesis enzymes of *M. tuberculosis*, as most of the corresponding genes are not only essential for the bacterium but there is no human or mammalian counterpart of their cognate gene products. This study was specifically focussed on amino acid biosynthesis enzymes that are IdeR regulated or the ones that utilize chorismate as substrate. These enzymes are required not only for the production of aromatic amino acids but are also indirectly involved in the biosynthesis of secondary metabolites with an aromatic ring. Several novel findings with respect to the mechanism of action of these enzymes were made. The present study has a bearing not only on the understanding of the role of these chorismate-utilizing enzymes in the pathogenesis of *Mycobacterium tuberculosis* but also on the study of behaviour and activity profile of divergently as well as convergently evolved group of enzymes. Additionally, to shed light on host pathogen interaction during tuberculosis infection, the study was continued with the identification of novel mycobacterial proteins/antigens that could also help develop new tools for TB diagnosis. The expectation was that, such studies would enhance an understanding of the mechanism of bacterial survival during tuberculosis infection and simultaneously aid in TB diagnosis.

CHAPTER 2

IDENTIFICATION OF PROBABLE IdeR *BINDING SITES IN THE UPSTREAM SEQUENCES OF Mycobacterium tuberculosis ORFs*

2.1 ABSTRACT

In pathogenic bacteria, many virulence factors and iron acquisition systems are regulated by iron dependent transcription regulators [Litwin and Calderwood,1993]. One of the key regulators of such systems in *Mycobacterium tuberculosis* is IdeR [Iron dependent regulator], first identified as a homologue of the DtxR [Diphtheria toxin Repressor] protein *of Corynebacterium diphtheriae* [Schmitt *et al.*,1995]. IdeR has been known to govern the expression of a wide variety of genes ranging from those involved in iron acquisition and oxidative stress response to ones that code for enzymes involved in aromatic amino acid biosynthesis [Gold *et al.*, 2001; Rodriguez and Smith, 2003]. In spite of the wealth of information available on IdeR regulated genes of *M. tuberculosis*, there is still an underlying possibility that there are novel genes/pathways that have gone undetected and identification of which could give new insights into understanding the pathogenesis of *M. tuberculosis*. Therefore, an *in silico* approach employing positional relative entropy method was used to identify potential IdeR binding sites in the upstream sequences of all the 3919 ORFs of *M. tuberculosis*. While many of the predictions made by this approach overlapped with the ones already identified by microarray experiments and binding assays [suggesting the accuracy of the method used], few genes for which there has been no evidence for IdeR regulation were additionally identified. These include *fec*B [a periplasmic lipoprotein] and Rv1404 [a probable transcription regulator]. *In vitro* binding assays with the urease upstream sequence indicated a possible regulation of the urease operon of *M*. *tuberculosis* by IdeR. Since unavailability of iron is a nutritional stress for the bacterium, genes under the control of iron dependent regulatory proteins are of prime importance for the pathogen's survival inside the nutrient deficient environment of the host. Intricacies of the iron dependent regulatory mechanism with respect to the activity of urease operon, transcription regulators and transporters are discussed in this chapter.

2.2 INTRODUCTION

The pathogenecity of disease causing microbes is absolutely dependent on their ability to acquire iron from the iron deficient environment of the host. The success of *M. tuberculosis* in the establishment of an infection is also dependent upon its ability to acquire iron from its neighbouring environment.

While low iron is a limiting factor for the pathogen growth and survival, even high iron is detrimental as it leads to the formation of highly reactive hydroxyl radicals *via* the Fenton reaction. Hence, acquisition of iron by pathogenic bacteria has to be tightly regulated. In mycobacteria, this regulation is brought about by the Iron-dependent Repressor [IdeR] protein, the functional homologue of the Diphtheria toxin Repressor [DtxR] protein of *Corynebacterium diphtheriae* [Schmitt *et al.*,1995].

2.2.1 IdeR: The iron-dependent Regulator of *M. tuberculosis*

IdeR was first identified as the mycobacterial equivalent of DtxR of *Corynebacterium diphtheriae* [Schmitt *et al.*,1995]. DtxR serves as a repressor of *tox* gene, the structural gene for diphtheria toxin. Apart from this function, DtxR also behaves as a regulatory protein involved in the iron metabolism of the bacterium [Boyd *et al.*,1990; Schmitt *et al.*,1991]. In fact, low iron concentration has often been found to be associated with bacterial toxin production. This includes exotoxin A from *Pseudomonas aeruginosa*, Shiga like toxin from *E. coli* and diphtheria toxin from *C. diphtheriae* [Frank *et al.*, 1989; Calderwood and Mekalanos, 1987; Schmitt & Holmes, 1991]. The function of DtxR was found to be similar to the very well known Fur protein of gram-negative bacteria. Under iron sufficient conditions, DtxR causes repression of genes involved in iron metabolism by binding to their operator sequences, with a high specificity, thereby blocking transcription [Tao *et al.*, 1992]. Since DtxR has been proved to be an iron dependent repressor, homology search was carried out within the *M. tuberculosis* genome to identify a DtxR equivalent [Schmitt *et al.*, 1995]. A protein with 59% overall amino acid identity [in a 230 aa stretch] to DtxR was identified in the *M. tuberculosis* proteome. This protein was named IdeR or 'Iron dependent Repressor'. Initial experiments were done to determine if IdeR represses transcription of DtxR regulated promoters. DNA mobility

shift assays and DNA footprinting analysis showed that IdeR binds to the same promoter sequences to which the Corynebacterial DtxR protein binds. Binding was observed to be metal dependent [Schmitt *et al.*, 1995].

In *M. tuberculosis*, under iron sufficient conditions, IdeR binds to the upstream sequences of genes required for growth in low iron conditions. Under iron limiting conditions, IdeR no longer binds to these promoter regions which are free to allow the binding of RNA polymerase and subsequent transcription of the downstream gene. As IdeR binds to the same promoter regions as DtxR [Schmitt *et al.*,1995], the virulence of a wild type *M. tuberculosis* strain harbouring a constitutively active dominant mutant *dtx*R was assessed [Manabe *et al.*, 1999]. Mice were infected with mutant *M. tuberculosis* strain constitutively expressing mutant DtxR. A significant decrease in bacterial load in both the spleen and the lungs was observed at 16 weeks.

Later, it was shown that *ide*R is an essential gene of *M. tuberculosis* and the encoded protein regulates the expression of genes involved in the metabolism of iron and oxidative stress response [Rodriguez *et al.*, 2002]. In this study, the authors have shown that *ide*R null mutant of *M. tuberculosis* can not be generated without the incorporation of a second copy of the gene. However, they were able to obtain a rare mutant of *ide*R, in which the lethal effects of *ide*R inactivation were alleviated by a suppressor mutation. This mutant showed a restricted iron assimilation capacity and helped the authors in the identification of phenotypic effects resulting from *ide*R inactivation. In this study, the transcription profiles of wild-type, *ide*R mutant, and *ide*R-complemented mutant strains were analyzed using DNA microarray. This resulted in the identification of genes regulated by iron and IdeR. These genes encode proteins involved in siderophore biosynthesis and iron storage, enzymes of aromatic amino acid biosynthesis, putative transporters, members of the PE/PPE family, transcriptional regulators, and enzymes involved in lipid metabolism.

An *ide*R mutant of *M. smegmatis* showed a defective regulation of siderophore biosynthesis [Dussurget *et al.*, 1996]. Later, it was shown that *fxb*A of *M. smegmatis* which codes for a putative formyltransferase for siderophore biosynthesis is under the transcriptional control of the IdeR [Dussurget *et al.*, 1999].

Considering the overall importance of this iron dependent regulatory protein in transcriptional regulation of virulence factors and iron scavenging systems, efforts were initiated to crystallize the protein in order to gain a better understanding of the mode of action of iron dependent regulators [Feese *et al.*, 2001]. Wild type IdeR protein from *M. tuberculosis* has been crystallized in its complete holorepressor form at 2^0A resolution. Structurally, IdeR shares a lot of homology with DtxR [Pohl *et al.*, 1999]. The high-resolution data by Feese *et al.* revealed the first detailed picture of the metal binding sites and N-terminal residues in the fully activated conformation. In addition, the SH3 like fold and metal binding function of domain 3 were confirmed by unbiased electron density.

Until very recently, there was knowledge about only the repressor activity of these iron dependent regulatory proteins. This led to the assumption that high iron conditions are more inhibitory for bacterial growth than low iron. However, recent observation of the presence of IdeR binding sites in the upstream region of the bacterioferritin genes of *M tuberculosis* has led to the elucidation of other functions of the iron dependent regulatory proteins. For the first time it was shown that IdeR also acts as an activator [Gold *et al.*, 2001]. Under high iron conditions, the iron storage protein, bacterioferritin of bacterial systems is upregulated. This leads to enhanced storage of excess of iron for future use.

2.2.2 Mycobacterial promoters regulated by iron dependent regulatory proteins

IdeR of *M. tuberculosis* in association with ferrous ions binds to a 19 bp inverted repeat consensus sequence or iron box [TTAGGTTAGGCTAACCTAA] in the upstream sequences of several genes [Schmitt *et al.*,1995]. Gel mobility shift assays, DNA footprinting, Reporter gene assays and DNA microarray are four techniques that have been exploited by a multitude of workers to determine the genes expressed/repressed in *M. tuberculosis* as a function of iron availability [Gold *et al.*, 2001; Camacho *et al.*, 1999]. Genes involved in iron acquisition and storage have been shown to be IdeR regulated [Dussurget *et al.*, 1996; Gold *et al.*, 2001]. Several other genes not directly involved in siderophore biosynthesis have also been shown to be expressed or repressed as a function of iron stress [Rodriguez *et al.*, 2003; Dussurget *et al.*, 1999]. These reports

suggest that IdeR is a global regulator that controls several genes involved in iron metabolism and processes related to iron metabolism. Experimental evidence for iron mediated regulation of a large number of mycobacterial genes exists. Two divergently transcribed genes, *his*E [a part of the histidine operon] and a PPE gene [Rv2123] have been shown to be IdeR regulated [Rodriguez *et al.*, 1999]. Gel shifts and footprinting assays have revealed that IdeR regulates the expression of these genes by binding to the iron boxes in the regulatory region and binding was dependent upon the concentration of iron in the reaction mix. Few other genes involved in the biosynthesis of siderophores [*mbt*A, *mbt*B, *mbt*I], biosynthesis of aromatic amino acids [*trp*E2, *phe*A, *his*E, *his*B] and others like iron storage proteins [bacterioferritins, *bfr*A, *bfr*B] have also been experimentally shown to be part of the IdeR regulon [Gold *et al.*, 2001]. Functional characterization of genes not apparently involved in iron metabolism would lead to further insights into the relation between iron metabolism and various aspects of mycobacterial physiology.

This chapter describes the use of a computational approach to identify novel genes under the regulatory control of IdeR followed by its experimental verification.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Computational prediction of IdeR binding sites

The complete genome sequence of *Mycobacterium tuberculosis* H37Rv was downloaded from NCBI ftp site *[ftp://ftp.ncbi.nlm.gov*] and IdeR binding sites were collected from literature [Table 2.1]. Profiles for recognition of IdeR binding sites were calculated by positional relative entropy method assuming that each position is independent. A matrix was developed for the purpose, which was used to scan the upstream sequences of all the genes from -400 to +20 of the translation start site to identify potential IdeR binding sites [Yellaboina *et al.*, 2004a, 2004b]. Consensus Iron dependent repressor protein [deR] binding sites were collected from literature and probability distributions of four different nucleotides within the binding sites of known sequences as well as throughout the genome were computed. The probability distributions of nucleotides within and outside regulatory region were used to compute

Table 2.1: Known IdeR binding sites in the upstream sequences of *M.tuberculosis* ORFs

IdeR binding sequence	Downstream ORF
CAAGGTAAGGCTAGCCTTA	Rv1519
TTATGTTAGCCTTCCCTTA	Rv3403c
TTAACTTAGGCTTACCTAA	Rv3839
TTAGGCAAGGCTAGCCTTG	Rv1343c
CAAGGCTAGCCTTGCCTAA	Rv1344
TATGGCATGCCTAACCTAA	Rv1347c
TTCGGTAAGGCAACCCTTA	Rv1348
ATAGGTTAGGCTACCCTAG	Rv2122c
CTAGGGTACCCTAACCTAT	Rv2123
AGAGGTAAGGCTAACCTCA	Rv3402c*
TTAGTGGAGTCTAACCTAA	Rv1876
GTAGGTTAGGCTACATTTA	Rv2386c
CTAGGAAAGCCTTTCCTGA	Rv3841
TTAGCTTATGCAATGCTAA	Rv0282
TTAGGCTAGGCTTAGTTGC	RV0451c
TTAGCACAGGCTGCCCTAA	Rv2383c
TTAGGGCAGCCTGTGCTAA	Rv2384

the relative entropy of segments [length 19bp] along the +20 to -400 regions of all the genes *M. tuberculosis*. Finally segments were sorted according to the relative entropy and segments with high relative entropy were considered as probable Iron dependent repressor binding sites. A complete list of these IdeR binding sites as calculated by the relative entropy method is provided in Table 2.2.

2.3.2 Cloning, expression and purification of *M. tuberculosis* IdeR

pRSETa expression vector [Promega] with an N-terminal 6X His tag was used to clone the ORF Rv2711 of *M. tuberculosis* that encodes IdeR. Briefly, Rv2711 was amplified from *M. tuberculosis* H37Rv DNA using primers with specific restriction enzyme sites. [Forward primer: ATT***GGATCC***ATGAACGAGTTGGTTGATA and Reverse primer: TGT***AAGCTT***GACTTTCTCGACCTTGACC] and the amplicon was cloned into the corresponding sites of pRSETa expression vector. *E coli* BL21DE3 cells transformed with the 6xHis tagged chimeric construct were grown in 1L of LB medium supplemented with 100μg/ml of ampicillin and 10%glycerol. IPTG [0.1mM] was added to a mid log phase culture. The cells were kept in an incubator shaker for another five hours at 27^0C and 150rpm to allow protein expression. After induction, cells were harvested by centrifugation and resuspended in 20ml of lysis buffer [10mM Tris HCl, 100mM NaCl and 10% glycerol, pH7.5] with 0.1mM PMSF and disrupted using a sonicator. After another round of centrifugation for 10 minutes at 10,000g, the supernatant was applied to a Ni-NTA affinity column [Qiagen, USA].

Affinity chromatography: The supernatant was allowed to bind to Ni-NTA column packed in a polypropylene column. The recombinant protein was purified after washing the column with 5 bed volumes of lysis buffer containing 30mM imidazole and eluting with 250mM imidazole. The eluates were analyzed by SDS PAGE and dialyzed against Tris buffer to remove salts and imidazole. The purity of the eluted protein was checked on SDS PAGE with Coomassie Blue staining

Table 2.2: Candidate IdeR boxes in the genome of *Mycobacterium tuberculosis*

Predicted IdeR box	Position of binding site relative to Translation starty site	Score	Gene annotation	Rv number	Predicted Function
TTAGTGGAGTCTAACCTAA	-226	5.2563	*bfr*A	Rv1876	Bacterioferritin
ATAGGCAAGGCTGCCCTAA	**-151**	**.19346**	_	**Rv1846c**	**Predicted transcriptional regulator**
TTAGCACAGGCTGCCCTAA	-86	5.16997	*mbt*A	Rv2384	Peptide arylation enzymes
TTAGGGCAGCCTGTGCTAA*	-32	5.15772	*mbt*B	Rv2383c	Peptide arylation enzymes
TTATGTTAGCCTTCCCTTA*	-2	5.14546	_	Rv3403c	hypothetical protein
CTAGGAAAGCCTTTCCTGA*	-73	5.12055	*bfr*B	Rv3841	Ferritin-like protein
TTAGGCAAGGCTAGCCTTG*	-85	5.09743	_	Rv1343c	hypothetical protein
TTAACTTAGGCTTACCTAA	-36	5.09181	_	Rv3839	hypothetical protein
ATAGGTTAGGCTACCCTAG*	-51	5.07767	PPE	Rv2123	PPE
TTAGGTAAGCCTAAGTTAA	-79	5.04482	*phe*A	Rv3838c	Prephenate dehydratase
CTAGGGTAGCCTAACCTAT*	-95	5.04385	*his*I	Rv2122c	Phosphoribosyl-ATP pyrophosphohydrolase
TTAGGGCAGCCTTGCCTAT	**-146**	**5.0165**	_	**Rv1847**	**Hypothetical protein**
CAAGGCTAGCCTTGCCTAA	-292	4.97724	*fad*D33	Rv1345	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II
CAAGGTAAGGCTAGCCTTA*	-50	4.97077	_	Rv1519	hypothetical protein
CAAGGTAAGGCTAGCCTTA	-345	4.97077	_	Rv1520*	Glycosyltransferases involved in cell wall biogenesis
GTAGGTTAGGCTACATTTA*	-25	4.8669	*trp*E2	Rv2386c	Anthranilate/para-aminobenzoate synthases component I
GCAGGTCAGGCTACCCTTA	-26	4.82224	*mur*B	Rv0482	UDP-N-acetylmuramate dehydrogenase
ATAGGAAAGCCGATCCTTA	-36	4.64865	_	Rv0114	Histidinol phosphatase and related phosphatases
GTAGACCAGGCTCCCCTTG	**-302**	**4.62592**	***fec*B**	**Rv3044**	**ABC-type cobalamin/Fe3+-siderophores transport systems**
TAAGGGTAGCCTGACCTGC	-20	4.61752	_	Rv0481c	hypothetical protein
TTAGGCTAGGCTTAGTTGC*	-112	4.59032	*mmp*54	Rv0451c	mmpS4
GCAACTAAGCCTAGCCTAA	-139	4.54925	_	Rv0452	Transcriptional regulator
CTATGTGATACTGACCTGA	-42	4.5466	*glp*Q2	Rv0317c	Glycerophosphoryl diester phosphodiesterase

AGATGCTAGACTTTCCTGA	**-77**	**4.54327**	_	**Rv1404**	**Transcriptional regulator**
TTACGGCAGCCTGTTGTAA	**-35**	**4.53876**	_	**Rv2663**	**hypothetical protein**
TTAGCTTATGCAATGCTAA*	-50	4.49914	_	Rv0282	hypothetical protein
TTCGGTAAGGCAACCCTTA*	-213	4.41965	_	Rv1348	hypothetical protein
TCACTGTAGTCTTAGCTGA	-179	4.39591	_	Rv0698	hypothetical protein
ATCCGTAAGTCTAAACTTA	-26	4.35929	_	Rv2034	Predicted transcriptional regulators
TTACTGCAATCTCCACTGA	-149	4.33623	*fad*A5	Rv3546	Acetyl-CoA acetyltransferases
TATGGCATGCCTAACCTAA	-31	4.02212	_	Rv1347c	Acetyltransferases, including N-acetylases of ribosomal proteins
TTACCGCGCACTGCTCTAT	-17	3.51297	_	Rv1344	Acyl carrier protein
TATGGCATGCCTAACCTAA	-50	4.02212	_	Rv1347c	Acetyltransferases
GTAGGTTAGGACAGCCTTT	-102	3.92933	_	Rv0338c	Fe-S oxidoreductases
TAATGGCAGACTGTGATTC	-3	3.89219	*ppi*A	Rv0009	Peptidyl-prolyl cis trans isomerase

Sequences with asterisk (*) represent the experimentally confirmed IdeR binding sites. Sequences in bold represent the experimentally unverified novel IdeR binding sites predicted by the positional relative entropy method used in this study.

2.3.3 Gel retardation and south western assays

Gel retardation assay: Binding of IdeR to the predicted iron box was carried out in a 20µl reaction consisting of 1X buffer [10mM Tris HCl, 50mMNaCl, 10% glycerol, 5µg/ml acetylated BSA, 1mM DTT, 1mM PMSF and 50mM $MgCl_2$], 1% NP40, 1µg/ml poly dIdC, purified IdeR [1µM] and 3-5fmol of ^{32}P labeled probe. The probe consisted of the annealed 19bp oligo corresponding to the predicted IdeR box end labeled with ^{32}P using T4 polynucleotide kinase enzyme. Reaction was performed in the presence and absence of $CoCl_2$ (200µM). Unlabeled oligo was used for specific competition. After the addition of labeled probe, the reaction mixture was incubated for 15 min at 25^0C followed by loading on a 4% polyacrylamide gel in 1XTBE buffer. Electrophoresis was carried out at 200V for 30 minutes at 4^0C. After electrophoresis, the gel was dried and analyzed by autoradiography.

South western assay: The bacterial extract overexpressing *M tuberculosis* IdeR was separated on an SDS/PAGE gel and the proteins were electrophoretically transferred to a nitrocellulose membrane in a buffer containing 25mM Tris, 190mM glycine and 20% methanol for 16hrs at 30mA. The protein on the membrane was renatured by incubating in blocking buffer [2% non fat dry milk, 1% BSA, 10mM Hepes NaOH, 0.1mM EDTA, 200mM NaCl, 50mM $MgCl_2$ and 16µg/ml sonicated sperm DNA]. After renaturation, the membrane was placed in a hybridization bag with binding buffer [blocking buffer with 0.2% non-fat dry milk and 10^6 cpm/ml labeled oligo. Hybridization was performed with constant shaking for 16 hours. The membrane was briefly rinsed in blocking buffer without skimmed milk or BSA, dried, covered with plastic wrap and subjected to autoradiography.

2.4 RESULTS

2.4.1 Novel IdeR binding sites are present upstream of *fec*B, a periplasmic lipoprotein coding gene and Rv1404, a putative transcriptional regulator

The consensus IdeR binding site collected from published literature was used to identify similar IdeR binding regions in the –400 to +20 regions of all the 3919 ORFs of *Mycobacterium tuberculosis*. A complete list of IdeR binding sites with the highest scores as calculated by the positional relative entropy method is shown in Table 2.2.

To date, the most detailed study on prediction of IdeR binding sites alongwith experimental verification has been done by Gold *et al.* [2001]. There had been an earlier report on IdeR regulation of two divergently transcribed genes of *Mycobacterium tuberculosis* as well [Rodriguez *et al.*, 1999]. Additionally, microarray analysis of genes induced by low iron and in an IdeR mutant strain have also shed light on the iron-dependent regulation of mycobacterial genes [Rodriguez *et al.*, 2002]. As the first step towards the analysis of our predictions, the results were compared with the available information on IdeR regulated genes. Though most of these genes were earlier known to be IdeR regulated, the present study suggest for the first time that a part of the ferric dicitrate type transporter complex, FecB, a periplasmic lipoprotein and Rv1404, a putative transcriptional regulator are possibly regulated by IdeR [Table 2.2].

The upstream sequence of *fec*B shows the presence of an IdeR box at –302 position. On account of absence of reports on the details of the iron transport system of *M. tuberculosis*, the ferric dicitrate transporter system does seem to be an important candidate. A new transcription regulator [Rv1404] and a hypothetical protein Rv2663 that were not earlier predicted to be part of the IdeR regulon could be identified in this study.

2.4.2 IdeR binds to the IdeR box present in the intergenic region between the ORFs Rv1846c and Rv1847c

While many of the IdeR binding sites predicted by this study overlapped with ones predicted by earlier workers, experimental evidence demonstrated by *in vitro* binding experiments and reporter gene assays is available for only a few. These include *his*E, Rv2123 [Rodriguez *et al.*, 1999], Rv3402, *mbt*I *his*G, *mbt*A, *mbt*B, *mbt*I, Rv3402 and *bfr*A [Gold *et al.*, 2001] etc. As per our prediction, the IdeR box upstream of the ORF Rv1846c shows one of the highest similarity score to the IdeR consensus sequence. However, experimental evidence for the same does not exist. Moreover, Rv1846c does not figure in the list of genes induced in an IdeR mutant strain [Rodriguez *et al.*, 2003]. Hence, it was decided to determine if IdeR binds to this predicted iron box. The ORF, Rv2711 that encodes IdeR was cloned in the *Bam*HI and *Hind*III sites of pRSETa vector with an N-terminal Histidine tag and expressed and purified as a recombinant protein in *E. coli* BL21 cells [Figure 1]. Purified IdeR was used in gel retardation and south-western assays to test if it binds to the predicted IdeR box in the intergenic region between Rv1846c and Rv1847 [Figure 2A]. As evident from the gel shift assay results shown in Figure 2B, IdeR does bind to the 19bp IdeR binding sequence present in the intergenic region between Rv1846c and Rv1847. The binding could be competed out using cold oligos indicating the specificity of binding.

To convincingly demonstrate binding of IdeR to the above mentioned probable iron box, south-western analysis was carried out. *E. coli* BL21 cells transformed with recombinant plasmids carrying *M. tuberculosis ide*R were grown to mid log phase and fractionated by electrophoresis on polyacrylamide gel. The gel was probed with radiolabeled oligonucleotide corresponding to the predicted iron box. While the vector control lysate lane [Figure 2.2C, lane 1] did not show any binding, the induced cultures showed a positive binding. This experiment generated evidence for the involvement of IdeR in the regulation of the divergently transcribed ORFs, Rv1846c and Rv1847 of *M. tuberculosis*.

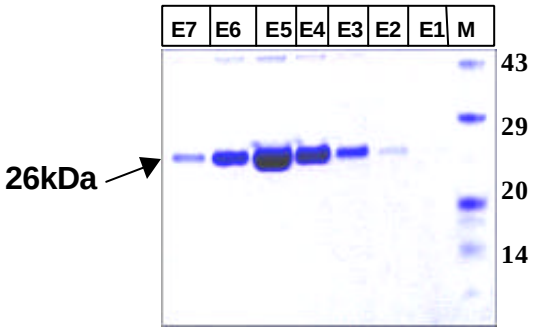


Figure 2.1: Purification of the Iron-dependent Regulator (IdeR) of *Mycobacterium tuberculosis* as a recombinant protein in *E. coli*. *M. tuberculosis* IdeR (coded by ORF Rv2711) was cloned in the *Bam*HI/*Hind*III sites of pRSETa expression vector and purified as a 6X Histidine tagged recombinant protein in *E coli* BL21 cells using affinity chromatography procedures. Purified protein was fractionated on a 10% polyacrylamide gel and stained with Coomassie Brilliant Blue dye. M represents the protein molecular size marker (Broad range, Genei), E1-E7 show the successive eluted fractions of the recombinant protein. Arrowhead indicates the position of the pure eluted protein.

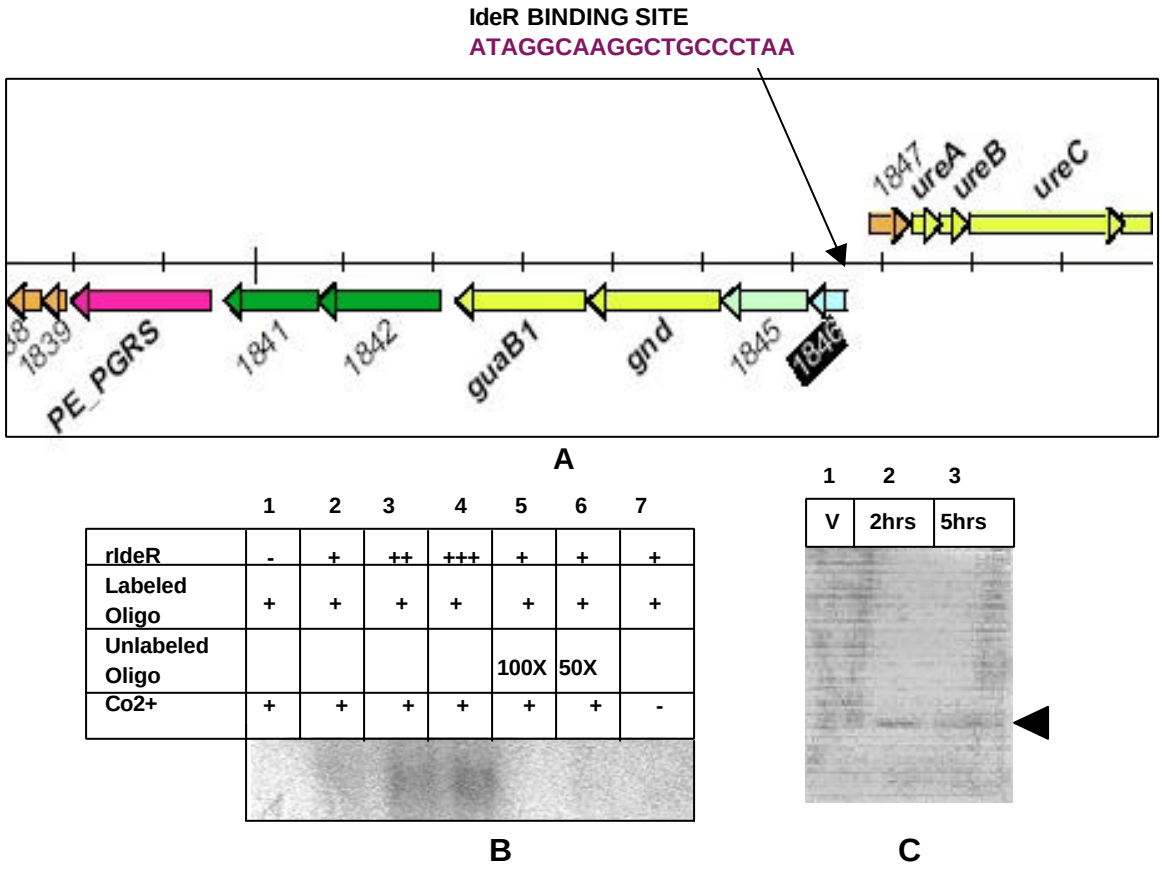


Figure 2.2: A: Schematic representation of the divergently transcribed ORFs, Rv1846c and Rv1847 with an IdeR binding site in the intergenic region. *ure* A, B and C are genes of the urease operon. Rv1841 and Rv1842 are putative hemolysin coding genes. B: Autoradiogram of the gel retardation assay demonstrating the binding of IdeR to the predicted iron box shown in A. Binding was specific as indicated by the disappearance of the band upon addition of cold oligo (Lanes 5 and 6). C: Autoradiogram of the south western assay demonstrating the binding of *Mtb* IdeR in *E. coli* BL21 cell lysates (indi\uced for 2hrs and 5hrs) to the predicted iron box shown in A. The cultures were induced for 2hrs and 5hrs, fractionated on a 10% polyacrylamide gel, transferred to a nitrocellulose membrane, renatured and hybridized with labeled oligo. Arrowhead indicates the position of the band. Specificity of binding was confirmed by the absence of the corresponding band in vector control lane (lane1).

2.5 DISCUSSION

Mycobacterium tuberculosis shows remarkable ability to survive various kinds of stresses from the initiation to the establishment of a full fledged infection in the host. To gain an insight into the molecular basis of this adaptive nature of the organism, it is important to identify genes specifically expressed under different stress conditions. Since non-availability of soluble form of iron is an important form of nutritional stress presented by the host to the bacterium, it is assumed that genes responsible for the acquisition of iron are essential for full virulence and establishment of a successful infection. *M. tuberculosis* has an elaborate network of genes for the biosynthesis of siderophores, the iron acquisition systems [Qadri *et al.*, 1998]. Recent experiments have shown that these genes are regulated by iron-dependent regulatory proteins [Gold *et al.*, 2001]. Transcriptional control plays a key role in regulating gene expression in response to various environmental conditions. Apart from the production of siderophores as a function of low iron availability, *M. tuberculosis* also produces many other iron regulated proteins, which are the probable virulence factors of the bacterium [Rodriguez *et al.*, 2003].

2.5.1 The ferric dicitrate type transporter complex of *M. tuberculosis* as a probable IdeR regulated system

While a number of transporter proteins like Rv1463 (an ABC transporter), Rv2459 (a probable drug efflux pump), Rv1348 (membrane protein similar to Yersiniabactin uptake system) etc have been earlier predicted to be IdeR regulated, the present work suggests for the first time IdeR dependent regulation of *fec*B of *M. tuberculosis*. FecB has been annotated as a probable Fe[III] dicitrate binding periplasmic lipoprotein. As *E. coli fec* operon is also under the regulatory control of Fur like protein, it is likely that the ferric dicitrate system could be an integral part of membrane iron transport in *M. tuberculosis*.

The *fec* operon is very well characterized in *E. coli* and a dyad repeat sequence GAAAATAATTCTTATTTCG present upstream to *fec*A has been proposed to serve as the binding site of the Fur iron repressor protein in *E. coli* [Zimmermann *et al.*, 1984, Pressler *et al.*, 1988]. It is thus likely that FecB of *M. tuberculosis* could also be part of

the iron transport complex of the bacterium and the regulation of the gene is brought about by IdeR, the Fur equivalent of *M. tuberculosis*. Additionally, as predicted by the method described above as well as the NCBI pattern search by Gold *et al.* [2001], another membrane protein coded by Rv1348c [similar to the yersiniabactin uptake system] shows an IdeR box in its upstream sequence. This protein also appears to be an important candidate in the uptake of siderophore like compounds.

2.5.2 Regulation of a probable MarR equivalent transcriptional regulator, Rv1404 by IdeR

Quite a few transcription regulators are known to be under the regulatory control of IdeR [Rodriguez *et al.*, 2002]. Results presented above could also identify Rv1404, a novel transcriptional regulator that shares some similarity to the Multiple antibiotic resistance regulator [MarR] protein from *E. coli*, as a probable IdeR regulated gene. If the antibiotic resistance regulator function of Rv1404 is proven, this could provide a clue to iron dependent regulation of antibiotic resistance in *M. tuberculosis*. Here, it would also be important to mention that the ORF Rv1846c that is also predicted to have an IdeR box in its upstream sequence also shows some similarity to the penicillase repressor protein of *Bacillus licheniformis*. These findings suggest that IdeR could be a global regulator that activates even other regulatory proteins that take care of the iron dependent regulation of a broader network of *M. tuberculosis* genes.

2.5.3 Regulation of the urease operon and hemolysins by IdeR

The ORFs Rv1846 and Rv1847 have interesting predicted functions that are important in the context of the pathology of *M. tuberculosis*. While Rv1847 is a hypothetical protein probably a thioesterase involved in the biosynthesis of aromatic compounds, Rv1846c codes for a transcription regulator with some similarity to the penicillase repressor protein of *Bacillus licheniformis*. Interestingly, Rv1847 is also present in the same operon that codes for genes involved in the biosynthesis of the urease enzyme. It is known that *Mycobacterium tuberculosis* survives in the acidic, toxic and hostile environment of the macrophage phagolysosome. One mechanism of survival is to somehow increase the pH of the phagolysosome. In this respect, the activity of the urease operon assumes

importance as it could possibly help in neutralization of the acidic pH [Clemens *et al.*, 1995]. However, the mechanism of regulation of *M. tuberculosis* urease operon has not yet been described anywhere. As an iron box exists upstream to the urease operon (directly upstream of ORF, Rv1847), it was tempting to speculate that urease could also be regulated by IdeR. Additional evidence springs up from the fact that in many pathogenic bacteria like *H. pylori*, the urease operon is regulated by ferric uptake regulatory [Fur] proteins [Bijlsma *et al.*, 2002]

Along with the prediction of a high score, experimental evidence for IdeR dependent regulation of *M. tuberculosis* urease operon has been provided in this work. Urease has been implicated in the virulence of several other pathogenic micro orgamisms. In *H.pylori*, *Salmonlla typhimurium* and *Escherichia coli*, urease is regulated by Ferric uptake regulator in response to pH [Bijlsma *et al.*, 2002; Heimer *et al.*, 2002]. In case of *M. tuberculosis*, ammonia generated by the action of urease may be of considerable importance in alkalinizing the microenvironment of the organism and preventing phagosome-lysosome fusion and phagosome acidification. In addition ammonia generated by the action of urease should be available to the organism for assimilation of nitrogen into biomolecules.

The genes downstream to Rv1846c [Rv1841c, Rv1842c]. are predicted to code for hemolysin like proteins In *E. coli*, hemolysins are regulated by H-NS [Wyborn *et al.*, 2004]. Additionally, even the global regulators FNR and CRP are known to bind to the upstream of the hemolysin like genes [Wyborn *et al.*, 2004]. In *Staphylococcus aureus*, ClpX and ClpP regulate the expression of the hemolysin virulence factors [Frees *et al.*, 2003]. This study gives the first evidence of a probable IdeR regulation of hemolysin coding genes of *M. tuberculosis* which might aid the bacterium to acquire iron from heme.

In summary, this study enhances the current understanding of the complex network of *M. tuberculosis* genes expressed/repressed as a consequence of iron stress. The study also adds considerably to the understanding of the various mechanism of survival adopted by the bacterium to survive inside in the iron deficient environment presented by the host.

CHAPTER 3

THE HYPOTHETICAL ORF Rv1885c OF Mycobacterium tuberculosis ENCODES A MONOFUNCTIONAL AroQ CLASS OF PERIPLASMIC CHORISMATE MUTASE

3.1 ABSTRACT

Chorismate mutase [CM] is a branchpoint enzyme of the aromatic amino acid biosynthesis pathway and catalyzes a committed step in the biosynthesis of tyrosine and phenylalanine. Naturally occurring variants of the enzyme chorismate mutase are known to exist that exhibit diversity in enzyme structure, regulatory properties, association with other proteins etc. These features have been the apparent reasons behind extensive studies that have been carried to date on the enzyme. Chorismate mutase was not annotated in the initial sequenced genome of *Mycobacterium tuberculosis* on account of low sequence similarity that exists between known CMs. In the present work, CM activity was assigned to the hypothetical protein corresponding to ORF Rv1885c of *M. tuberculosis*. Extensive biochemical and biophysical characterization of the recombinant protein suggests its semblance to the aroQ class of CMs, prototype examples of which include the *E. coli* and yeast CMs. These results also demonstrate that unlike the P and the T proteins of *E. coli*, *M. tuberculosis* CM does not have any associated prephenate dehydratase or dehydrogenase activity, indicating its monofunctional nature. Further, the ability of the 33 amino acids at the N-terminus of *M. tuberculosis* CM to act as a signal sequence to export *E. coli* alkaline phosphatase into the periplasmic space suggests that *M. tuberculosis* CM belongs to the periplasmic subclass of aroQ class of CMs. A redundancy of CM in *M. tuberculosis* genome was evident from the observation that the hypothetical protein coded by ORF Rv0948c also shows *in vitro* CM activity. The present study is the first report wherein biochemical evidence has been provided for the existence of two monofunctional CMs in any pathogenic bacteria. Considering the fact that the enzyme does not have a human homologue, understanding the mechanism of its action would open up avenues for the development of novel therapeutic interventions for tuberculosis.

3.2 INTRODUCTION

Mycobacterium tuberculosis has developed ingenious mechanisms to survive inside the hostile environment presented by the host and to acquire essential nutrients from this adverse environment [Wayne *et al.*, 1994; Betts *et al.*, 2002]. These attributes enable the bacterium to establish a successful infection inside the host. For the development of new therapeutic strategies to check tuberculosis, there is a need for the identification of novel targets that are not only unique to *M. tuberculosis* but blocking of which would either prove lethal to the bacterium or render it extremely susceptible to the host immune response. In this context, understanding the mechanism of action of the aromatic amino acid pathway enzymes of *M. tuberculosis* assumes utmost importance as most of the corresponding genes have been reported to be essential for the bacterium and there is no human or mammalian counterpart of the same. [Parish and Stoker, 2002; Sassetti *et al.*, 2003]. Moreover, amino acid auxotrophs of *M. tuberculosis* do not survive or multiply in macrophages [Bange *et al.*, 1996; Gordhan *et al.*, 2002], suggesting that these nutrients are not available within the compartment of the macrophage in which the bacteria reside.

Chorismic acid is the last common precursor in the aromatic amino acid biosynthesis pathway and is a substrate for multiple enzymes [Gibson and Jackman, 1963; Dosselaere *et al.*, 2001] [Figure 3.1]. Hence it has always been of interest to explore how a microbe partitions chorismate into diverse metabolic pathways. Various enzymes that utilize chorismate as a substrate include chorismate mutase [CM], anthranilate synthase [AS], isochorismate synthase [ICS] and p amino benzoate synthase [Haslam, 1993; Weiss and Edwards, 1980].

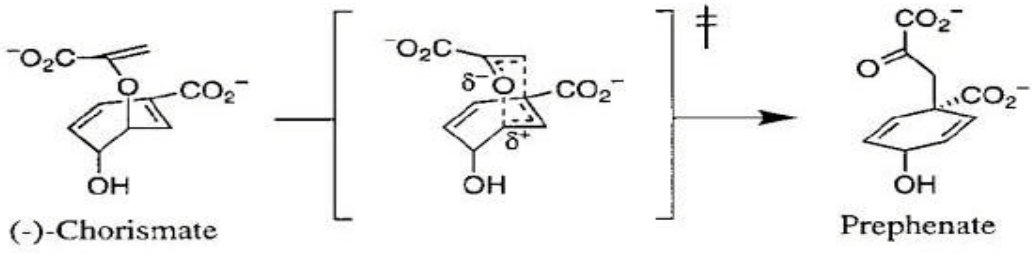
3.2.1 Chorismate mutase

Chorismate mutase [EC 5.4.99.5] is a branchpoint enzyme of the shikimate pathway and plays a crucial role in the biosynthesis of aromatic amino acids. The enzyme channels chorismate towards the biosynthesis of phenylalanine and tyrosine away from that of tryptophan. Chorismate mutase is thus important for regulating the balance of aromatic

amino acids in the cell. Since the shikimate pathway and hence chorismate mutase only exists in bacteria, fungi and higher plants, the enzyme may be extremely useful in developing not only herbicides but also anti- bacterial and anti-fungal therapeutics [Kishore and Shah, 1988].

Chorismate mutase [CM; EC 5.4.99.5] catalyzes the first committed step in the biosynthesis of the aromatic amino acids phenylalanine and tyrosine in bacteria, fungi and higher plants [Haslam, 1993]. CM catalyzes the conversion of chorismate to prephenate, which is then converted to the precursors of tyrosine or phenylalanine by the subsequent action of prephenate dehydrogenase or prephenate dehydratase respectively.

The conversion of chorismate to prephenate involves breaking an ether bond between the ring moiety and a carboxylic side-chain of the substrate, and reconnecting the two components at another spot by alkylation [Hur and Bruice, 2002]. It is generally agreed that the conversion proceeds *via* a pericyclic [Claisen type] rearrangement, involving a chair-like transition state as shown below.



This reaction is a rare example of an enzyme catalyzed pericyclic reaction and is one of the reasons behind extensive studies that have been carried to date on the enzyme [Guilford *et al.*, 1987; Kast *et al.*, 2000; Hur and Bruice, 2003; Guimaraes *et al.*, 2003; Kienhofer *et al.*, 2003; Heather *et al.*, 1996]. Despite extensive studies, however, the enzyme mechanism and in particular the 10^6 to 10^7 fold rate acceleration by chorismate

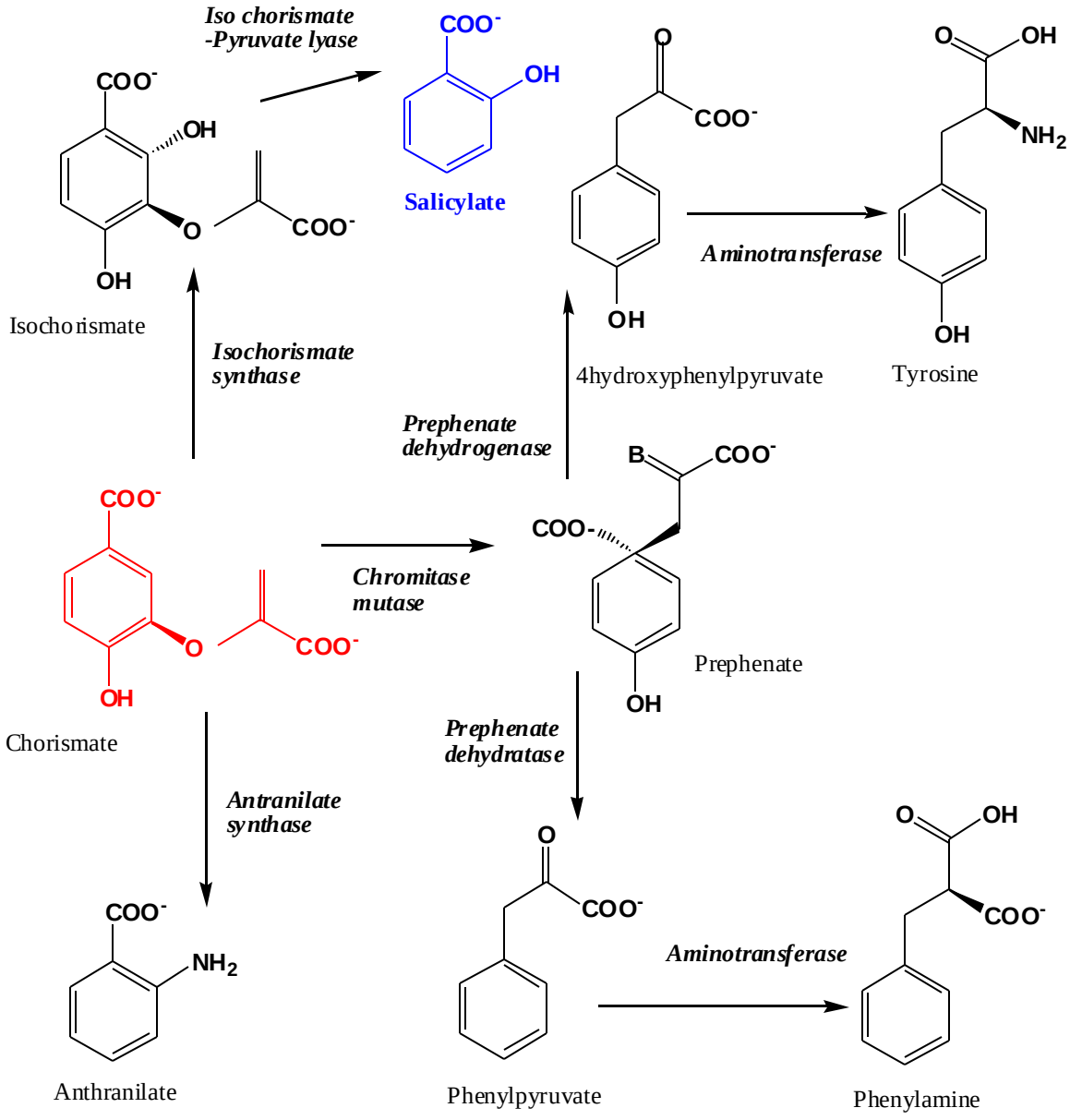
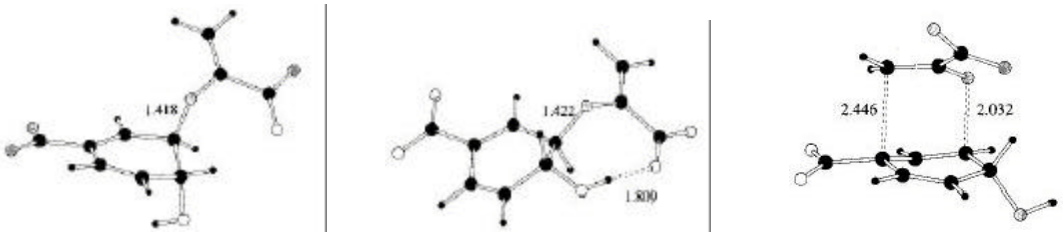


Figure 3.1: Chorismate is a substrate for multiple enzymes

mutase is poorly understood. The proposed mechanism by which CM catalyzes the conversion of chorismate to prephenate is by preferential binding of the enzyme to the less stable pseudo-diaxial form of the enolpyruvate side chain of chorismate [Sogo *et al.*, 1984]. It thus positions chorismate for a rearrangement in the chair-like transition state [Copley *et al.*, 1985; Guo *et al.*, 2001].The pseudo-diaxial and pseudo-diequitorial conformers of chorismate as proposed by Wiest and Houk [1994] are shown below.



pseudo- diequitorial **pseudo-diaxial** **Transition state**

Chorismate to prephenate reaction is also known to occur spontaneously. Knowles *et al.* have worked out the detailed mechanism of catalyzed and uncatalyzed conversion of chorismate to prephenate [Addadi *et al.*, 1983]. In the uncatalyzed reaction, the pseudodiequitorial form of chorismate is more populated as its stability is increased by the formation of an internal hydrogen bond between the hydroxyl group and the side-chain carboxylate. However, the pseudodiaxial conformer is the more reactive conformer that is capable of progressing towards the transition state [Carlson *et al.*, 1996]. As stated in the previous paragraph, chorismate mutases catalyzes the conversion of chorismate to prephenate by stabilizing the less stable but more reactive pseudo-diaxial form of chorismate.

Apart from a unique reaction mechanism, the CM enzyme has also gathered attention due to the presence of completely different protein fold and ligand binding pockets in different organisms, which suggests separate evolutionary origins of the enzyme. A brief description of the same is provided in the proceeding sections.

3.2.1.1 Classification of chorismate mutase

Chorismate mutase enzyme can be classified into three groups; AroQ class, prokaryotic type; AroQ class, eukaryotic type; and AroH class. These three classes fall into two structural folds [AroQ and AroH class], which are completely unrelated. Additionally, there are subsets of the AroQ class that are periplasmic or the ones that have large regulatory domains. The former are often referred to as periplasmic CMs and the latter are referred to as AroR CMs.

The two types of the AroQ structural class [the *Escherichia coli* CM dimer and the yeast CM monomer] can be structurally superimposed, and the topology of the four-helix bundle forming the active site is conserved. However, the AroH class of CMs have an entirely different structure. While the AroQ proteins are generally all-helix-bundle proteins [Lee *et al.*, 1995], while the AroH proteins possess a trimeric pseudo α/β barrel structure [Chook *et al.*, 1993]. A subset of the AroQ CMs, commonly referred to as AroR are structurally similar to AroQ with a four-helix-bundle structure around the active site. AroR CMs are seen only in eukaryotes and can be viewed as a version of AroQ that has inserted loops required for allosteric regulation [Strater *et al.*, 1997; Goers *et al.*, 1984; MacBeath *et al.*, 1998]. The periplasmic chorismate mutases have an N-terminal signal sequence, that defines their periplasmic location [Calhoun *et al.*, 2001; Xia *et al.*, 1993]. The extent of dissimilarity, yet a common reaction mechanism of the two classes of chorismate mutase enzymes suggests an independent evolutionary origin of analogous enzymes [Galperin *et al.*, 1998].

Chorismate mutases encoded by both *aroQ* and *aroH* are widely distributed among prokaryotes. AroH is the more rare group of CMs and most familiar examples of the same include those of *Bacillus* species [*B. subtilis, B. halodurans etc*], cyanobacteria [*Synechocystis* sp., *Synechococcus* sp], *Streptomyces coelicolor,* and *Thermus thermophilus* etc. In contrast, AroQ class of CMs are more widely distributed [*Escherichia coli*, *Saccharomyces cerevisisae*]. Even multiple paralogs of *aroQ* are commonly present in a single organism. In *Bacillus subtilis,* genes encoding both AroQ and AroH co-exist [Jensen and Nester, 1965].

Chorismate mutase frequently exists as a fusion protein and the most common associations are seen with other enzymes of the aromatic amino acid biosynthesis pathway like prephenate dehydrogenase, prephenate dehydratase, DAHP synthase etc. *E. coli* has two CMs, one in association with prephenate dehydrogenase [P protein] and the other with prephenate dehydratase [T protein]. In other organisms like *Bacillus* sp, *Methanococcus*, Yeast etc, CM exists in two copies, one as a monofunctional protein and the other as a fusion protein with other aromatic amino acid biosynthesis enzymes [Llewellyn *et al.*, 1980].

The crystal structures of CMs from *Escherichia coli* [EcCM] [Lee *et al.*, 1995], *Bacillus subtilis* [BsCM] [Chook *et al.*, 1994], and *Saccharomyces cerevisiae* [Xue *et al.*, 1994] have been solved and their significant features are described in the following sections.

3.2.1.1.1 Chorismate Mutase of *Escherichia coli* [AroQ; prokaryotic]

The Chorismate mutase of *E. coli* consists of residues l-l09 of the P-protein. *E. coli* CM is the the prototypical AroQ chorismate mutase and consists of homodimers of intimately entwined α-helices. The individual subunit has three helical segments [residues 6-42, 49-65, 70-l00]. Coiled-coil and helix-helix interactions between the two longest segments that create a catalytically functional, elongated homo-dimer with two equivalent, elbow-shaped active sites that are highly charged and completely enclosed [Lee *et al.*, 1995] [Figure 3.2].

AroQ CMs have composite active sites. Each of the two active sites of *Ec*CM dimer is formed by 3 helices from one monomer and a helix from the other subunit. It has been proposed that such domain-swapped dimers might have evolved from active, monomeric precursors [Bennet *et al.*, 1995]. By inserting a 180^0 turn into the middle of helix 1 of the thermostable *Methanococcus jannaschii*.CM, active site was generated with residues from a single polypeptide chain [MacBeath *et al.*, 1998].

Aromatic amino acid biosynthesis is generally very well regulated because of the high energy costs to synthesize these amino acids. In fact, amongst all amino acids, aromatic

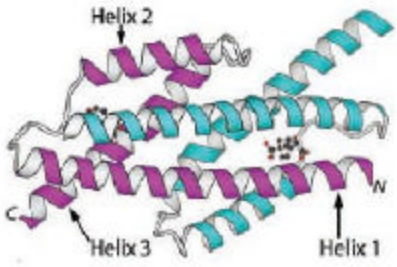


Figure 3.2: Structure of *Escherichia coli* chorismate mutase shown in a ribbon diagram representation. Each subunit has three helices. A transition state analog inhibitor is bound in each active site and is represented in ball-and-stick form [Lee *et al*, 1995].

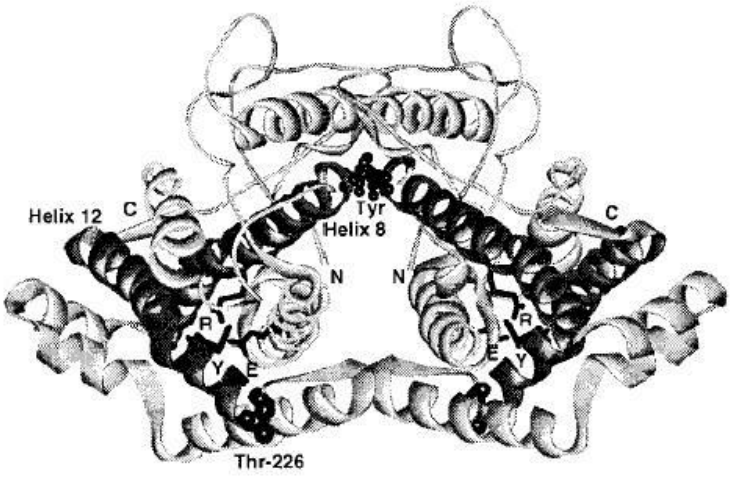


Figure 3.3: Ribbon diagram of the structure of the T state of yeast chorismate mutase. The N and C termini are labeled as N and respectively. Tyr represents the two tyrosine molecules located at the interface of dimer. Helix 8, which connects the allosteric active sites, and helix 12 are shown in dark color. The side chains Glu-23 [E] and Tyr-234 [Y], which are involved in the hydrogen bonding interaction with the active residue Arg-157 [R], are drawn stick model. Thr-226, which connects the 220s loop and helix 12 is indicated. [Strater *et al* 1997]

amino acids are the most expensive to synthesize [L-Phe: 65, L-Tyr: 62, and L-Trp: 78 ATP metabolic equivalents, respectively] [Walker-Simmons and Atkinson, 1977]. In *Escherichia coli*, the biosynthesis of phenylalanine and tyrosine is subjected to feedback regulation. The catalytic and regulatory domains of *E. coli* P-protein have been studied by Zhang *et al.* [1998] using genetically engineered proteins. The authors have shown that the mutase, dehydrogenase and regulatory activities of P-protein reside within specific domains. Residues 286-386 were found to be crucial for phenylalanine binding.

3.2.1.1.2 Chorismate mutase of *Saccharomyces cerevisiae* [Aro Q; eukaryotic]

The yeast CM is a homodimer consisting of two 30-kDa polypeptide chains coded by the ARO7 gene and shows homotropic activation by the substrate chorismate. The crystal structures of wild-type yeast CM co-crystallized with tryptophan and an *endo*-oxabicyclic transition state analogue inhibitor [Bartlett and Johnson, 1985] has been determined [Strater *et al.* 1997]. YCM has essentially an all-helix structure, which has similarities to the CM domain of the P protein from *E. coli*. Both proteins contain a four-helix bundle that surrounds the active site cavity. Using 94 residues of 22% sequence identity, these helix bundles can be superimposed with a root-mean square deviation [rmsd] of 1.06 Å [Xue and Lipscomb, 1995] [Figure 3.3].

Yeast CM shows approximately 10-fold allosteric activation by tryptophan and 10-fold inhibition by tyrosine [Schmidheini *et al.*, 1990]. This antagonism plays an important role in maintaining the correct balance of aromatic amino acids in the cell. The structural basis for these regulatory processes has also been described [Strater *et al.*, 1996, 1997].

Phenylalanine is not an inhibitor of yeast chorismate mutase. This is in contrast to the homologous chorismate mutase from *Arabidopsis thaliana*, which is inhibited by both tyrosine and phenylalanine [Eberhard *et al.*, 1993].

The kinetic properties of the enzyme can be viewed in terms of a transition from a conformation having a low affinity for chorismate, the T [tense] state, to a conformation with high affinity for the substrate, the R [relaxed] state [Monod *et al.*, 1965]. The influence of the allosteric regulators is best explained by local conformational changes induced by their competitive binding to the regulatory sites, thereby stabilizing either of the two states or inducing a conformational change in the overall enzyme structure.

3.2.1.1.3 Chorismate mutase of *Bacillus subtilis* [AroH class]

BsCM is a member of the AroH class of CMs. This class adopts a trimeric, pseudo-α/β barrel fold [Chook and Lipscomb, 1993]. Chook *et al.* [1993] were the first to solve the crystal structure of the monofunctional chorismate mutase [EC 5.4.99.5] of *Bacillus subtilis* at 2.2^0A resolution. The enzyme was found to be a homotrimer, with beta-sheets from each monomer packing to form the core of a pseudo-alpha beta-barrel with helices on the outside of the trimer [Figure 3.4]. In addition, the active sites were located by using data from the enzyme structure complex with an endo-oxabicyclic inhibitor that mimics the transition state of the reaction. The structure of this complex was refined to 2.2^0A with a current R value of 0.182 for a model that included 1388 residues, 12 inhibitor molecules, and 530 water molecules in the asymmetric unit. Three active sites were seen in each trimer at the interface of two adjacent subunits.

The structure was refined further when it was solved at 1.3^0A resolution. In this structure, the C-terminal tails of the subunits of the trimer were refined further and were shown to be hydrogen bonded to residues of the active site of neighboring trimers in the crystal. This bonding could therefore cross-link the molecules in the crystal lattice.

AroH and AroQ CMs share some common active site features. For example, in the crystal structures with an oxabicyclic transition state analog [TSA], both enzymes show multiple cationic groups [Arg and Lys] interacting with the carboxylates and the ether oxygen [Lee *et al.*, 1995; Chook *et al.*, 1993]. Additionally, both possess a Glu residue that hydrogen bonds to the hydroxyl group of the TSA. Despite their different folds, both enzymes are likely to utilize similar catalytic mechanisms.

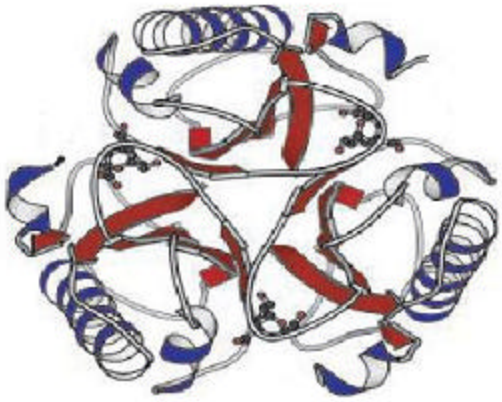


Figure 3.4: Structure of *Bacillus subtilis* chorismate mutase. The monofunctional chorismate mutase from *B. subtilis* is a homotrimer and adopts a pseudo-α/β barrel fold. A transition state analog, shown in a ball-and-stick representation, is bound in each of the active sites, which are located at the trimer interfaces [Chook and Lipscomb, 1993].

It was earlier believed that there were apparently no catalytic residues for CM and catalysis was due to the geometry and polarity of the active site of CM which stabilizes the transition state [Chook *et al.* 1993]. Later however, a different mechanism was proposed for EcCM as well as BsCM, that involves a catalytic group [Xue and Lipscomb 1995 Strater *et al.* 1996].

Chorismate mutase was not annotated in the initial genome sequence *of M. tuberculosis*. However, the recent COG [Cluster of Orthologous groups, NCBI] annotation of the *M. tuberculosis* genome suggests that CM activity can be attributed to two ORFs, Rv1885c and Rv0948c [http://www.ncbi.nlm.nih.gov/COG/]. The ORF Rv1885c has been annotated in the TubercuList web-server [http://genolist.pasteur.fr/TubercuList/], Institute Pasteur, as a conserved hypothetical protein with some similarity to the monofunctional chorismate mutases of *Erwinia herbicola* [28.6% identity in a 133aa overlap]. Rv0948c has also been annotated as a conserved hypothetical protein, equivalent to a conserved hypothetical protein [105 aa] from *Mycobacterium leprae* [NP_301237.1|NC_002677] which is also similar to N-terminus of some chorismate mutase/prephenate dehydratase enzymes.

This chapter addresses the question whether these two hypothetical proteins indeed show *in vitro* chorismate mutase activity and if they do, what is their proximity or distance with respect to the known two classes of CMs [AroQ and AroH]. The approach involved expressing the *M. tuberculosis* ORFs, Rv1885c and Rv0948c in *E. coli* and determining the biochemical and biophysical parameters of the encoded proteins. While more extensive studies were carried out with the protein coded by ORF Rv1885c, it was also demonstrated that Rv0948c also possesses CM activity, though with a reduced turnover. Kinetic and regulatory studies of rRv1885c indicate several unique properties of the enzyme which include feedback regulation by pathway specific as well as cross pathway specific ligands in the same manner. A genetic approach was used for functional characterization of the predicted N-terminal signal sequence of *M. tuberculosis* CM. These studies provide sufficient evidence to conclusively place the protein coded by ORF Rv1885c of *M. tuberculosis* in the aroQ class of periplasmic CMs.

3.3 EXPERIMENTAL PROCEDURES

3.3.1 Bacterial strains and plasmids

All *E. coli* and *M. tuberculosis* strains used in this part of the work are briefly described in Table 3.1. The plasmid vectors [with their sources] and the recombinant plasmids constructed are also described in the same table. The details of primers used in this part of the work are given in Table 3.2

3.3.2 Media, chemicals, buffers, and enzymes

Luria Broth (LB) supplemented with 200 μg/ml of ampicillin was used for selection and amplification of plasmids in *E. coli* hosts. LB with 4% glycerol was used for growth of bacterial cells for protein purification purposes.
XP medium was used for direct AP [Alkaline phosphatase] assay on bacterial plates. The composition of the medium is as follows.

10X XP medium [1Litre]

15g of Agar-agar was dissolved in 900ml of water and autoclaved. To this, 100ml of 10X XP medium [sterile] was added. The XP medium consisted of

1M CaCl2	2ml
10mM FeCl3	2ml
10mM ZnCl2	2ml
20% Glucose	10ml
Vitamin B1	500μl
20mg/ml XP [fresh]	2ml

3.3.3 Cloning, overexpression and purification of recombinant proteins in *E. coli*

E. coli BL21DE3 cells transformed with the 6xHis tagged chimeric construct [Rv1885c / Rv0948c cloned in the *Nde*I/*Xho*I sites and *Nde*I/*Hind*III sites respectively of pET23a expression vector [Novagen]] were grown in 1L of LB medium supplemented with

Table 3.1: Bacterial Strains and Plasmids

Strains/plasmids	Description	Source / Reference
Mycobacterium tuberculosis		
H37Rv	Laboratory Strain [Virulent]	Cole ST *et al*., 1998
E. coli		
DH5α		
BL21 DE3		
MG1655	Wild Type	
MG1655*pho*A+*pro*C-	*E. coli* proline auxotroph	This work
MG1655*pho*A-*pro*C+	*E. coli* Alkaline phosphatase negative prototrophic strain	This work
Plasmids		
pET23a	Expression vector	Novagen
pET23aCM1	pET23a derivative with *M. tuberculosis* CM1 [Rv1885c] cloned in *Nde*I/*Xho*I sites	This work
pET23aCM2	pET23a derivative with *M. tuberculosis* CM2 [Rv0948c] cloned in *Nde*I/*Xho*I sites	This work
pGEMTEasy	Cloning Vector	Promega
pGEMT AP	pGEMT Easy derivative with ss less *E.coli pho*A cloned in vector MCS	This work
pETCMssAP	pET23aCM1 derivative with ss less *E.coli pho*A cloned in *Sal*I/*Xho*I sites downstream of *M. tuberculosis* CM ss	This work
pBAD18 MCS2	Expression vector with the arabinose inducible promoter, P_{BAD}	Guzman *et al.*
pBADCMssAP	pBAD MCS2 derivative with ss less *E. coli* phoA cloned in *Xba*I/*Hind*III sites	This work

ss: signal sequence

Table 3.2: Sequences of primers used to amplify *M. tuberculosis* chorismate mutase (Rv1885c and Rv0948c) and *E. coli* Alkaline phosphatase (AP)

ORF/ Gene	FORWARD PRIMER	REVERSE PRIMER
	*Nde*I	*Xho*I
Rv1885c	AT***CATATG***TTGCTTACCCGTCCACGTGA	AT***CTCGAG***GGCCGGCGGTAGG
	*Nde*I	*Hind*III
Rv0948c	AAT***CATATG***AGACCAGAACCCCCACATCACGA	ATA***AAGCTT***GTGACCGAGGCGGCCCCT
	*Sal*I	*Xho*I *Hind*III
E coli AP	AT***GTCGAC***ACACCAGAAATGCCTGTTCTG	AT***CTCGAGAAGCTT***TTATTTCAGCCCCAGAGC

100μg/ml of ampicillin and 10% glycerol. IPTG was added to the mid log phase culture. The cells were kept in an incubator shaker for another five hours at 27^0C/150rpm to allow protein expression. A low temperature was used to allow the protein to enter the soluble phase. After induction, purification of the protein was carried out in a manner similar to that described for IdeR in Chapter 2 [Section 2.3.2]

3.3.4 Enzyme assays and kinetic studies

Chorismate mutase activity assays were carried out as described elsewhere with a few modifications [Davidson and Hudson, 1987]. This assay is based on enzymatic conversion of chorismate to prephenate, which is terminated by the addition of HCl. Prephenate can be further converted to phenylpyruvate under alkaline conditions which can be measured spectrophotometrically [Krappmann *et al.*, 1999]. Reaction volumes of 100 μl contained 10mM Tris HCl, pH7.5, chorismic acid 0.5 to 4mM, optionally tyrosine, phenylalanine or tryptophan [100 μM to 500 μM]. The reaction was started by the addition of 100 to 500 pico moles of the purified protein to the prewarmed chorismate solution [at 30^0C] and was stopped by the addition of 10μl of 1.6NHCl. After a second incubation of the mix at 37^0C for about 15 minutes, 890μl of 1.58N NaOH was added and the absorbance was recorded at 320nm. To exclude the absorbance caused by the uncatalyzed arrangement of chorismate, blanks of increasing chorismate concentrations without enzyme were prepared and their absorbance was also recorded. These readings were then subtracted from the absorbance measured for enzyme activities at different concentrations of the substrate.

For determination of the optimum pH for CM activity, buffers of different pKa [CPB, pH4, CPB pH4.5, MES pH6, HEPES pH7, Tris HCl pH7.5, Tris HCl pH8] were used CM activity was also assessed at different temperatures [15^0C to 80^0C]. One unit of enzyme was defined as the amount of enzyme required to form 1 μmol of product/min at 37 °C.
Prephenate dehydratase [PDT] and prephenate dehydrogenase [PDH] assays were carried out as described elsewhere [Gething *et al.*, 1976; Davidson and Hudson, 1987].

The PDT assay is described in detail in Chapter 5, section 5.3.3.

Phe and Tyr Feedback inhibition assays

Allosteric regulation of CM activity by L-Phe and L-Tyr was measured at 100-1000μM concentrations of the effectors.

3.3.5 Isolation of the periplasmic fraction of *E. coli* BL21 cells

The periplasmic fraction of *E. coli* was isolated following an osmotic shock procedure [Qiagen, USA]. About 1l of induced culture was harvested by centrifugation at 4000g for 20 minutes and the pellet was resuspended in 30mM Tris HCl, pH 8.0 at 80ml per gram wet weight. The cells were kept on ice and to it 500mM EDTA was added dropwise to a final concentration of 1mM. This was followed by another incubation on ice for 5-10 minutes with gentle agitation. The cell suspension was then centrifuged at 8000g for 20 minutes at 4^0C. The supernatant contained the osmotic shock fluid consisting of periplasmic proteins. The periplasmic and the cytoplasmic fractions were quantified and used in western blots for a comparative estimation of protein expression.

3.3.6 Western Blot

Western blot with the cytosolic and periplasmic fractions of *E. coli* BL21 cells expressing *M. tuberculosis* CM [Rv1885c] was carried out with monoclonal anti Histidine antibody using the manufacturer's instructions [Santa Cruz Biotech, USA].

3.3.7 Limited Proteolysis

To study the effects of various ligands on enzyme stability, integrity and accessibility of active sites, limited proteolysis of *M. tuberculosis* CM was carried out. Trypsin was taken as the protease of choice. 10μg recombinant protein [1μg/μl concentration] was taken to which different concentrations [100μM to 5mM] of the ligand [Tyrosine, Phenylalanine, tryptophan, Salicylate] was added. The reaction buffer contained 50 mM Tris pH 7.5 and 100mM NaCl. Trypsin protein molar ratio was kept at 1:1000 and the reaction was

incubated at 25^0C for 30 minutes, following which trypsin was inactivated by the addition of 0.5mM PMSF. The reaction was mixed with equal volume of 2X Tris-Tricine loading dye and loaded on a 10% Tris Tricine gel. The gel was silver stained and scanned.

3.3.8 Silver staining of protein gels:

3.3.8.1 Gel Preparation

After the completion of electrophoresis, the gel was soaked in SOLUTION 1 [50ml MeOH + 10ml Acetic Acid + 40ml MQ-H_2O]. Following this, the gel was transferred to SOLUTION 2 [5ml MeOH + 7ml Acetic Acid + 88ml MQ-H_2O] for 30 minutes**.** Solutions 1 and 2 were prepared in advance and stored at room temperature. The gel was further transferred to SOLUTION 3 [20ml 25% glutaraldyhyde [Sigma G6257] + 30ml MQ-H2O] for 30 min. Solution 3 was made fresh. The gel was washed with MQ- H_2O either 4 times for 30 minutes [> 100 ml/gel/wash] or 1 time overnight [> 500 ml/gel/wash]. The gel was then soaked in DTT solution [0.5 ml 100x DTT + 50ml MQ-H_2O] [100X DTT = 0.5 g/ml, at -20 C]

3.3.8.2 Gel Staining

The gel was kept for 30 min in SILVER SOLUTION [0.05g Silver Nitrate + 50ml MQ-H_2O]. STOP solution [2.42 g Citric Acid monohydrate + 5 ml MQ-H2O] was prepared in advance. DEVELOPER solution [30 ml 37% Formaldehyde + 3g Sodium Carbonate (anhydrous) + 100 ml MQ-H_2O] was also prepared before hand.

The gel was dumped in Silver Solution and rinsed briefly with MQ-H_2O. Next, the gel was rinsed with ~20ml Developer solution for 5 min then the remaining developer solution was added. Development of bands was closely monitored. Just before bands were dark enough, the entire 5ml of STOP solution was added. When bubbling stopped [~5min], the gel was rinsed with MQ-H_2O. Gel was stored in MQ-H_2O until it was scanned.

3.3.9 CD Spectrometry

The CD spectrum of the recombinant protein [Rv1885c] was recorded in a wavelength range of 190-250nm in steps of 1nm with four accumulations at each step. The spectral baseline was corrected by subtracting the respective blanks. Ellipticity, represented in milli degrees was plotted as a function of wavelength [nm]. Percentage helicity for secondary structure determination was calculated using the online available K2D software [http://www.embl-heidelberg.de/~andrade/k2d/]

3.3.10 Analytical size exclusion chromatography

The oligomeric state of the recombinant protein corresponding to ORF Rv1885 of *M. tuberculosis* was determined using analytical size exclusion chromatography on a Superdex 200 [HP 10/30] FPLC column from Pharmacia. Chromatography was performed at room temperature with 10mM Tris and 100mM NaCl as the running buffer, the same buffer in which the recombinant protein was eluted. A standard curve was prepared according to the instruction manual using the LMW Gel Filtration calibration kit from Pharmacia Biotech. The void volume was determined using Blue dextran 2000. The elution parameter Kav was calculated as Kav= Ve-Vo/ Vt-Vo, where Ve= elution volume for the protein, Vo= column void volume, Vt= total bed volume. Kav was plotted against log MW. The protein sample was loaded on the gel filtration column at a concentration of 4mg/ml in the presence of DTT. The elution volume of the protein was used to calculate its molecular weight using the standard curve.

3.3.11 Construction of an *E. coli pho*A negative strain and N-terminal signal sequence characterization of *M. tuberculosis* chorismate mutase

Wild type *E. coli* MG1655 strain was transduced with P1 phage carrying a *pro*C mutation linked to a kanamycin resistance marker [15% linkage]. The resultant transductants were selected for proline auxotrophy. A second round of transduction was carried out in this strain using a Δ*pho*A P1 phage where *phoA*- character was linked to *pro*C+ character [90% linkage]. 90% of these transductants were able to grow on minimal medium but

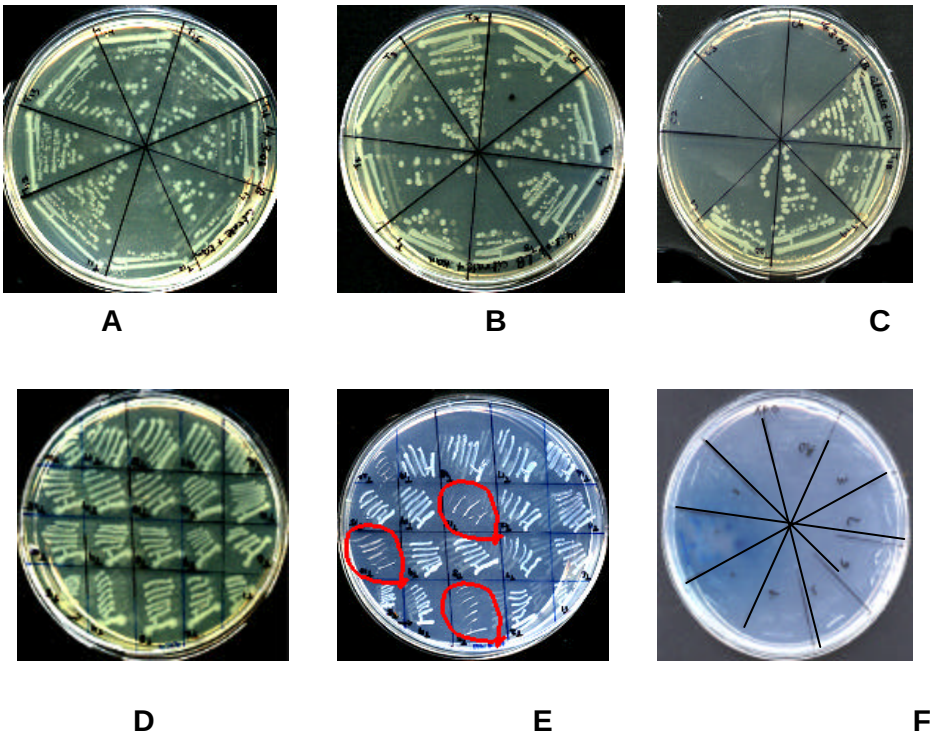


Figure 3.5: Construction of an *E. coli* strain lacking intrinsic alkaline phosphatase activity (*E. coli pho*A- strain)

E. coli strain MG1655 (wild type) was transduced with P1 phage carrying a mutation in *pro*C, where the *pro*C- character (for proline auxotrophy) was linked to a kanamycin resistance marker. The transductants were selected on LB kanamycin plates.

A,B,C: Purified kanamycin resistant colonies obtained after first transduction. Few colonies which grew in the control plates (without P1 transduction) did not grow upon restreaking (empty sectors in plate C)

D, E: The *pro*C- character of transductants was tested on LB (D) and Minimal medium (E). Colonies unable to grow on minimal medium (encircled red) were designated as MG1655 *pho*A+*pro*C- and selected for second round of transduction with P1 phage carrying the *pro*C+ *pho*A- segment.

F: After the second P1 transduction, the transductants were selected on Minimal XP plates. The sector with blue colonies is a positive control (*pho*A+) *E. coli* strain. 90% of the transductants were *pho*A negative (white on Minimal XP plates). This strain was referred to as a *pho*A negative *E. coli* strain (MG1655*pho*A-*pro*C+).

failed to develop blue colour on XP medium. This strain was designated as an *E. coli phoA* negative strain [Figure 3.5] and was used for functional characterization of N-terminal signal sequence of *M. tuberculosis* chorismate mutase. *E. coli pho*A [alkaline phosphatase] lacking the signal sequence was amplified from wild type *E. coli* genomic DNA and cloned in the MCS of pGEMTEasy vector followed by subcloning in *Sal*I/*Xho*I sites downstream to the 33amino acid signal sequence of the pET23aCM1 construct. This construct was named pETCMssAP. The *Xba*I/*Hind*III cassette from pETCMssAP alongwith the optimized RBS was transferred to pBAD18MCS2 vector [Guzman *et al.*, 1995] [Figure 3.6]. Expression from pBAD is driven by the arabinose inducible promoter P_{BAD}. Subsequently, the *pho*A negative *E. coli* strain was transformed with the pBADCMssAP chimeric construct and plated on LB-XP plates. As *E. coli* AP is known to be active only in the periplasmic space [Brickman and Beckwith, 1975], blue colour was expected only if *M. tuberculosis* CM signal sequence could export *E. coli* AP to the periplasmic space.

3.3.11.1 Preparation of P1 phage lysates

The donor strain was grown overnight in 0.3ml Z broth and mixed with 10^7 pfu [plaque forming unit] of P1 phage. Adsorption was allowed to occur at 37^0C for 20 minutes followed by the addition of 10ml of Z broth and incubation at 37^0C in a 100ml flask with slow shaking until growth. Lysis of the culture followed [after approximately 4 hours]. The lysate was treated with chloroform, centrifuged and the clear supernatant was stored at 4^oC.

Z broth:

LB medium	100 ml
$CaCl_2$ [0.5 M]	0.5 ml

3.3.11.2 P1 Transduction

The recipient cells were grown to $1x10^9$ cfu/ml in rich broth. $CaCl_2$ was added to a concentration of 1mM. 0.1 to 0.2 ml cells were mixed with P1 lysate to give a MOI (multiplicity of infection) of 0.01 to 0.1 and allowed adsorption at 37^oC for 15 min.. About 1/4 to 1/2 of the mixture were plated on selective plates containing 0.01M sodium citrate and appropriate antibiotic and incubated overnight at 37^oC.

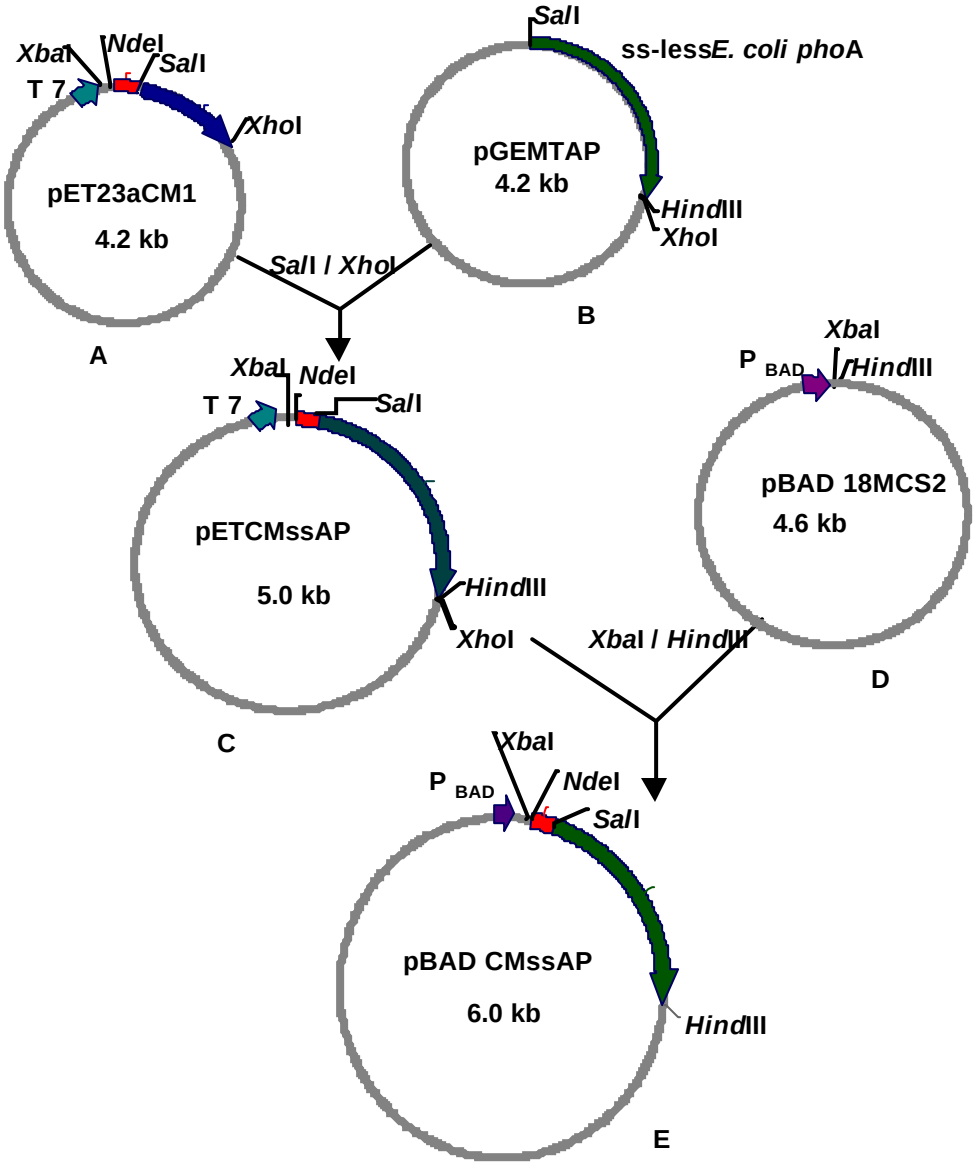


Figure 3.6: Construction of a recombinant plasmid with the signal sequence (ss) of *M. tuberculosis* chorismate mutase (CM) cloned upstream of ss less *E. coli pho*A. A: ORF Rv1885c of *M. tuberculosis* cloned in the *Nde*I/*Xho*I sites of pET23a vector. The region between *Nde*I and *Sal*I is the putative signal sequece of *M. tuberculosis* CM B: *E. coli pho*A (gene coding for alkaline phosphatase) lacking the ss cloned in the MCS of pGEMTEasy vector. C: A recombinant plasmid generated by transferring the *Sal*I/*Xho*I cassette from pGEMTAP to the *Sal*I/*Xho*I site of pET23aCM1 E: The *Xba*I/*Hind*III fragment from pETCMssAP was transferred to pBAD 18 (D) to generate pBADCMssAP (E). This was the final construct used for functional characterization of the of ss of *M. tuberculosis* CM.

3.4 RESULTS

3.4.1 The hypothetical ORF, Rv1885c of *M. tuberculosis* encodes a functional chorismate mutase enzyme

The gene corresponding to the ORF Rv1885c of *M. tuberculosis* was PCR amplified, cloned and expressed in *E. coli* BL21 cells and recombinant protein was purified to near homogeneity [Figure 3.7]. The kinetic properties of the purified protein were studied using CM activity assays. The substrate saturation curves were hyperbolic for the enzyme i.e. the enzyme followed Michealis Menton's kinetics [Figure 3.8]. There was no indication of positive homotropic co-operativity of *M. tuberculosis* CM towards chorismate. Kinetic parameters were subsequently obtained by fitting the initial rate data to the Michealis Menton's equation. Km for the enzyme was calculated as 1.2mM and Vmax as 71.42µmoles min-1 mg-1. The molar catalytic activity [Kcat] was 26.33 sec^{-1}. The effects of temperature and pH on enzyme kinetics were also studied. Since spontaneous arrangement of chorismate to prephenate is strongly dependent on temperature, blanks of the same reaction without the enzyme were also kept at different temperatures and the readings were subtracted from those obtained in presence of the enzyme. The optimum temperature for optimum activity of *M. tuberculosis* CM ranged from 37 to 50^0C [Figure 3.9 A]. While high temperature did allow a spontaneous conversion of chorismate to prephenate, there was a distinct enhancement of the reaction rate upon addition of the enzyme. pH also has a significant role in determining the activity of CM. *M. tuberculosis* CM was found to be active between pH 6 to 9 with a pH optimum of 7 [Figure 3.9 B]. The pH optimum is similar to *E. coli* CM, where again maximum activity is at pH 7.3. CM activity of rRv1885 was greatly inhibited at acidic pH. While fungal CMs are reported to be more active at acidic pH, bacterial CMs [*E. coli* P protein, *Salmonella typhimurium* CM] are reported to be active in alkaline pH [Schnappauf *et al.*,1997; Helmstaedt *et al.*, 2001]. *Bacillus* CM has maximum activity in the range of pH 5 to 9 [Gray *et al.*, 1990]. Hence the pH range at which *M. tuberculosis* CM works efficiently is similar to that of *Bacillus subtilis* with the optimum pH being closer to that of *E. coli* CM.

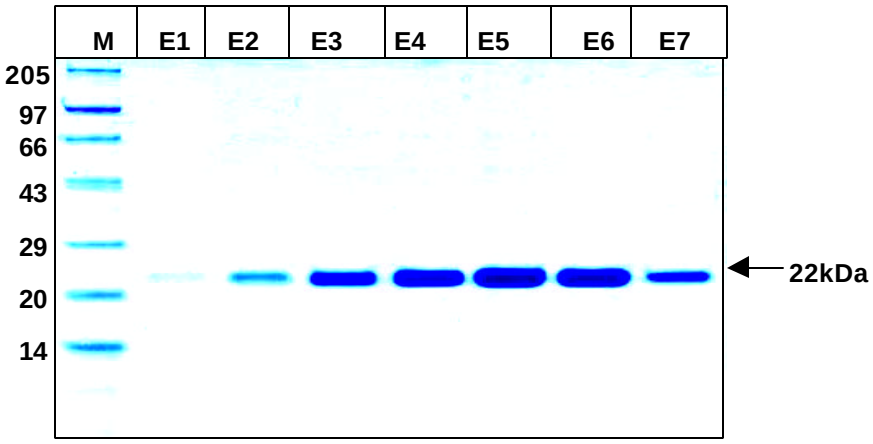


Figure 3.7: Purification of the chorismate mutase enzyme (Rv1885c) of *Mycobacterium tuberculosis* as a recombinant protein in *E. coli*. The ORF corresponding to Rv1885c was cloned in the *Nde*I and *Xho*I sites of pET23a vector with a C-terminal Histidine tag and expressed in *E. coli* BL21 cells. Affinity purification of recombinant protein was carried out using Talon resin (Clonetech). The purified protein was resolved on 10% Tris-Tricine gel. The gel was stained with Coomassie Brilliant blue dye. M represents the protein molecular size marker (Broad range, Genei), E1-E7 show the successive eluted fractions of the recombinant protein. Arrowhead indicates the position of the 22kDa protein, which is the predicted mass of *M. tuberculosis* chorismate mutase.

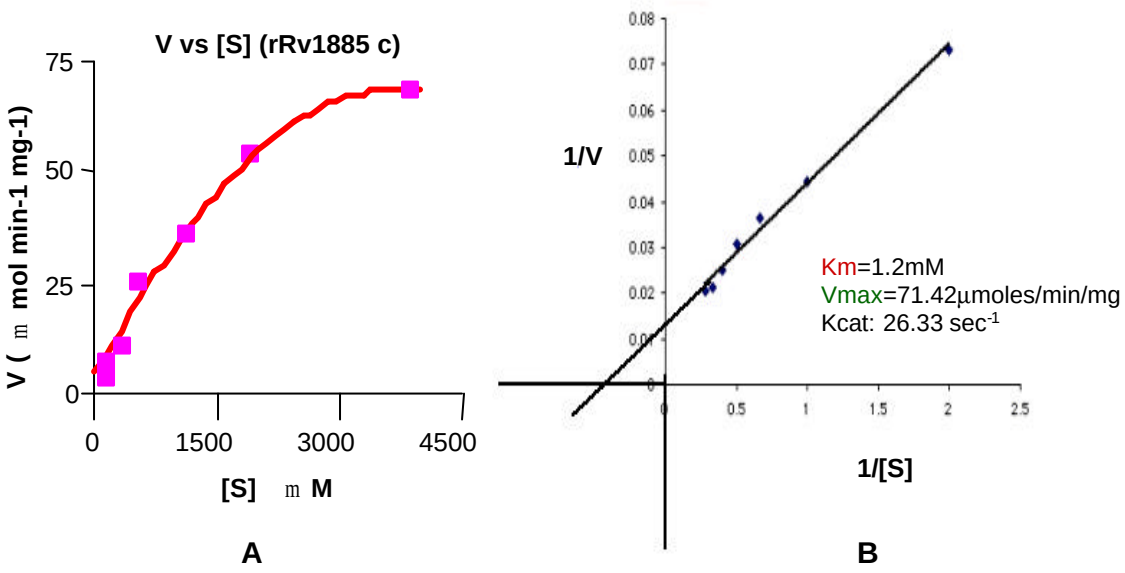


Figure 3.8 . A: Substrate saturation plot of *M. tuberculosis* chorismate mutase. Purified recombinant protein corresponding to ORF Rv1885c was assayed for chorismate mutase activity in the absence of any effector molecule. The data were fitted to functions describing Michaelis–Menton type saturation.B: Lineweaver-Burk plot of *M. tuberculosis* CM that was used to calculate Km and Vmax of the enzyme.

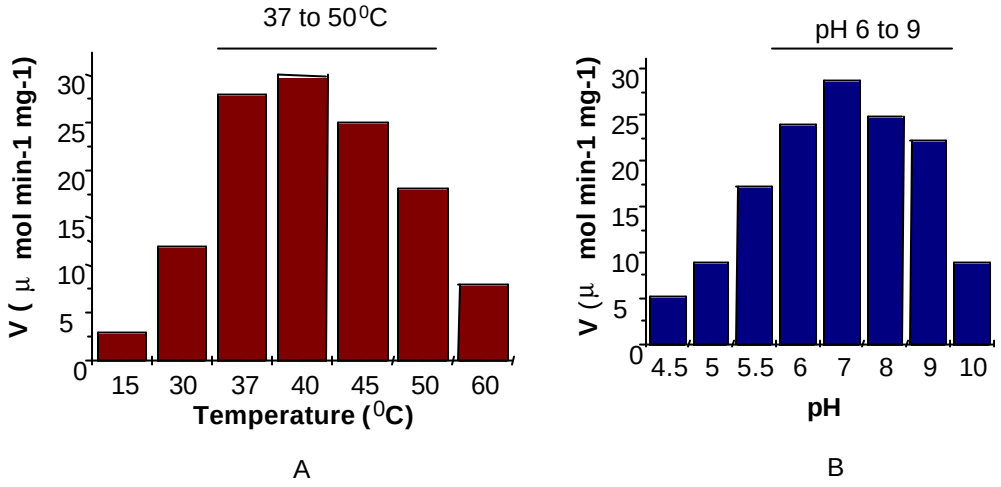


Figure 3.9 : Effect of pH and temperature on the activity of *M. tuberculosis* chorismate mutase. The temperature for optimum activity of *M. tuberculosis* CM ranged from 37 to 50^0C (A). The enzyme was found to be active between pH 6 to 9 with a pH optimum of 7 (B).

3.4.2 *M. tuberculosis* chorismate mutase [Rv1885c] does not display prephenate dehydratase [PDT] or dehydrogenase [PDH] activity

Though PDT or PDH domains have not been predicted to be present in the ORF Rv1885c, on account of several examples of divergent evolution of enzymes involved in aromatic amino acid biosynthesis [Dosselaere and Vanderleyden, 2001], it was decided to determine whether *M. tuberculosis* CM displays PDT or PDH activities. PDT and PDH assays were accordingly carried out with the recombinant protein. Results conclusively demonstrate that *M. tuberculosis* CM does not possess any intrinsic PDT or PDH activity. This is unlike the CMs of many other enteric bacteria where CM and PDT/PDH are present as fusion proteins [Ahmad and Jensen, 1988].

3.4.3 *M. tuberculosis* chorismate mutase [Rv1885c] shows allosteric regulation by tyrosine, phenylalanine and tryptophan

Having shown that the hypothetical protein coded by ORF Rv1885c encodes a functionally active chorismate mutase enzyme with no associated PDT or PDH activity, it was tested whether L-Phenylalanine, L-tyrosine, L-tryptophan, could act as potential modifiers of the CM activity of rRv1885c. This experiment revealed an unusual property of the enzyme. It was observed that while these ligands moderately enhance CM activity at low concentrations, they completely inhibit the enzyme at higher concentrations [Figure 3.10]. All known forms of CM that show allosteric regulation are usually large fusion proteins [Helmstaedt *et al.*, 2001]. The monofunctional CM of *Bacillus subtilis* is entirely resistant to feedback inhibition by aromatic amino acids [Gray *et al.*, 1990]. *E. coli* CM exists as a fusion protein and shows allosteric inhibition by only pathway specific inhibitors [Helmstaedt *et al.*, 2001]. The allosteric regulation of *M. tuberculosis* CM is also unlike yeast CM where tryptophan causes activation while tyrosine and phenylalanine cause inhibition of enzyme activity [Helmstaedt *et al.*, 2002]. Fluorimetric assays were also carried out to assess if phenylalanine binding causes a conformational change in *M. tuberculosis* CM. The blue shift in the spectra and the quenching of net fluorescence indicates that tryptophan residues of the enzyme are buried inside as a consequence of phenylalanine binding [Figure 3.11].

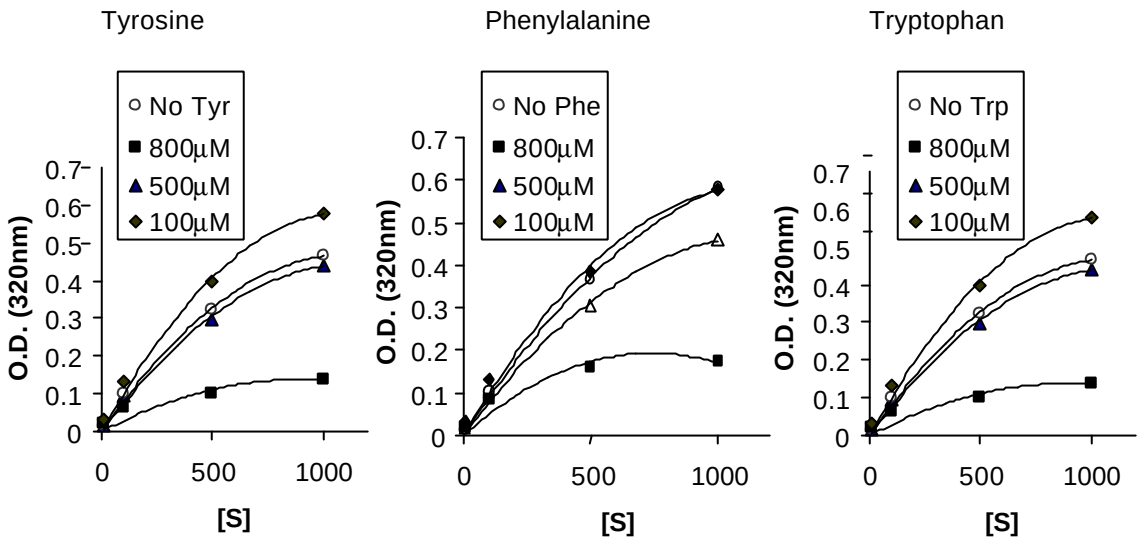


Figure 3.10: *M. tuberculosis* chorismate mutase (Rv1885c) is allosterically regulated by aromatic amino acids: The CM activty of Rv1885c was tested in the presence of various allosteric effectors (tyrosine, phenylalanine and tryptophan). Low concentration (up to 100 µM) of all the three aromatic amino acids led to a modest increase in enzyme activity. However, at higher concentrations of the effectors (800µM and above), CM activity was greatly inhibited.

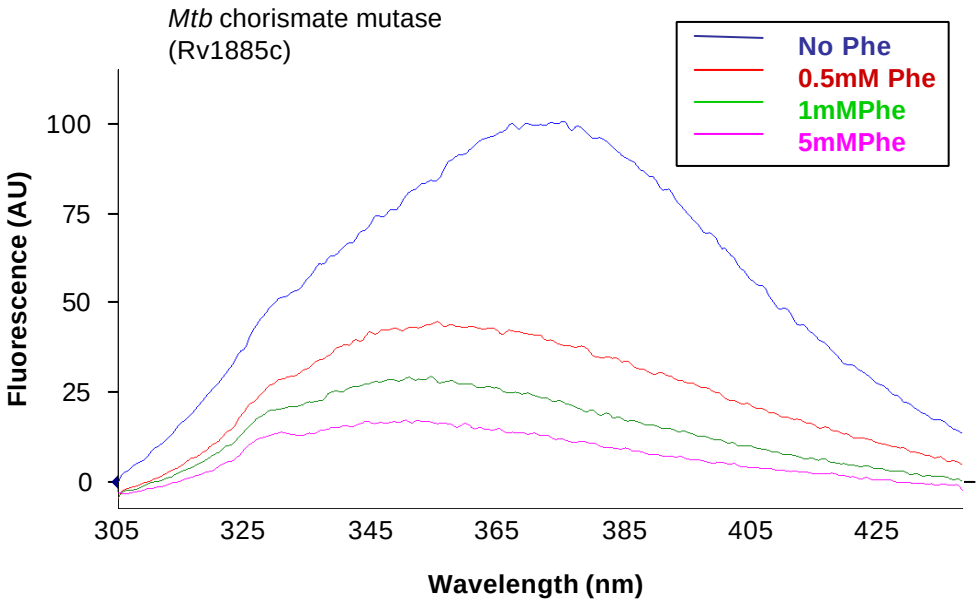


Figure 3.11:Phenylalanine binding leads to a change in the fluorescence spectrum of *M. tuberculosis* chorismate mutase (Rv1885c). Fluorescence emission spectra were recorded by exciting the protein at 280nm and recording the emission spectra in the range of 305nm-440nm. A blue shift was absorbed upon incubation of rRv1885c with phenylalanine indicating that the tryptophan residues of the protein are buried inside as a consequence of phenylalanine binding.

3.4.4 Pathway specific as well as cross pathway specific ligands protect *M. tuberculosis* CM from proteolytic cleavage

Having shown that tyrosine, phenylalanine and tryptophan are allosteric inhibitors of *M. tuberculosis* CM activity, experiments were designed to determine if binding of these ligands induce a conformational change in the enzyme that leads to inaccessibility of the active site. Limited proteolysis was employed as a structural probe to ascertain enzyme conformational change. Trypsin was taken as the protease of choice and limited proteolysis was carried out to study the amenability of the enzyme to proteolytic cleavage in the presence of various ligands. It was observed that in the presence of high concentrations [3mM and above] of effectors [Trp, Tyr and Phe], the enzyme was completely protected from proteolytic degradation [Figure 3.12 A-C]. Salicylic acid, another cross pathway secondary metabolite derived from chorismate did not confer any protection on the enzyme [Figure 3.12 D]. Higher concentrations (3mM or above) of all the three aromatic amino acids also protected the protein coded by ORF Rv0948c from trypsin cleavage [Figure 3.12, E-F].

3.4.5 *M. tuberculosis* chorismate mutase is a dimeric protein with a predominantly alpha helical structure

While catalytic and regulatory activity of *M. tuberculosis* CM point towards some novel properties of the enzyme, the study was continued to determine the biophysical parameters of the enzyme to define the actual class to which it belongs. Size exclusion chromatography was performed to determine the oligomeric state of the protein. The output was a single peak corresponding to the dimeric state of the recombinant protein [Figure 3.13]. In this context, *M. tuberculosis* CM is similar to the *E. coli* or Yeast CMs, which are also dimers of identical subunits [Lee *et al.*, 1995; Xue and Lipscomb, 1995]. To determine the secondary structure of *M. tuberculosis* CM, the CD spectrum was recorded on a spectropolarimeter [JASCO]. The data were analyzed using the online available K2D software. The results suggest a predominantly alpha helical structure for the enzyme [Figure 3.14]. This reminisces of the corresponding aroQ class of enzymes from yeast and *E. coli* [Xue Y, 1993; Strater *et al.*, 1996; Lee *et al.*, 1995] that are also

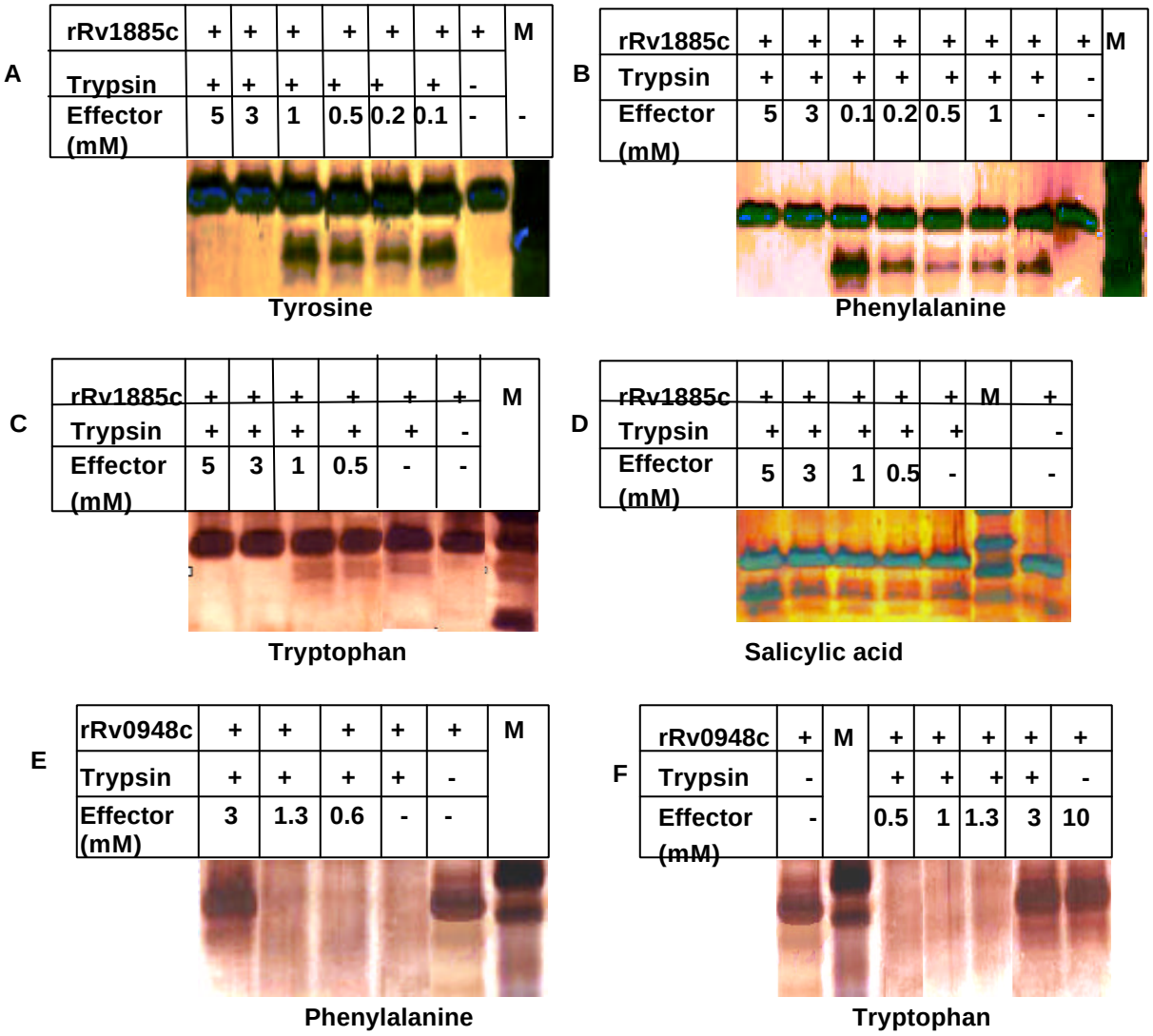


Figure 3.12 : Aromatic amino acids at high concentrations (3mM and above) provide protection to *M. tuberculosis* chorismate mutase from trypsin cleavage (Limited proteolysis studies). As evident from the above figure, tyrosine, phenylalanine and tryptophan (concentration 3mM and above) protect *M. tuberculosis* chorismate mutase enzymes (Rv1885c and Rv0948) from proteolytic cleavage (A,B,C,E,F). Salicylic acid, another cross pathway effector molecule does not protect *M. tuberculosis* CM (Rv1885c) from proteolytic cleavage (D).

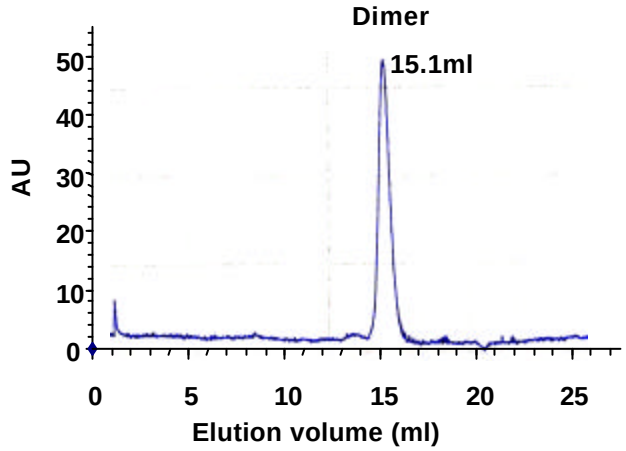


Figure 3.13: Size exclusion chromatography reveals that *M. tuberculosis* chorismate mutase (Rv1885c) is a dimeric protein (two subunits of individual MW 22kDa) rRv1885c was loaded on a Superdex 200 column (Pharmacia Biotech) and absorbance at 280nm (AU) was plotted as a function of elution volume. The elution parameter Kav vs log MW plot of protein standards was used to determine the actual MW of the recombinant protein.

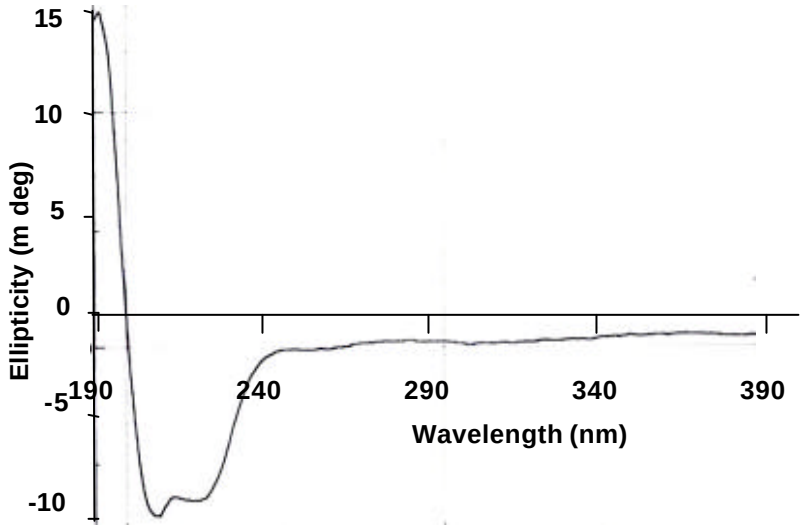


Figure 3.14: CD spectrum of *M. tuberculosis* chorismate mutase (Rv1885c) suggests a predominantly alpha helical secondary structure of the protein. The graph has been plotted as ellipticity (mdeg) as a function of wavelength (nm). Percentage helicity for secondary structure determination was calculated using the online available K2D software (http://www.embl-heidelberg.de/~andrade/k2d/) The spectrum confirms a highly helical structure of the recombinant protein with signature peaks at 208 and 222nm.

all helical proteins. The AroQ class of CMs consist of unregulated and regulated [AroQR] enzymes and are unusually divergent among closely related organisms [Hall *et al.*, 1982].

3.4.6 The N-terminal signal sequence of *M. tuberculosis* chorismate mutase can export *E. coli* alkaline phosphatase to the periplasmic space

While our experimental evidence in support of a dimeric alpha helical structure for *M. tuberculosis* CM classified the enzyme as an AroQ class of CM, the prediction of a signal peptide in the N-terminus of the protein was intriguing to determine if *M. tuberculosis* CM belongs to a periplasmic subclass of CMs. The CM of *Erwinia herbicola*, *Pseudomonas aeruginosa* and *Salmonella typhimurium* also possess an N-terminal signal sequence and differ in this respect from other members of the aroQ class of CMs [Xia *et al.*, 1993]. The study was therefore continued with the functional assessment of the N-terminal signal sequence of *M. tuberculosis* CM.

The approach used involved tagging *M. tuberculosis* CM signal sequence to PCR amplified *E. coli pho*A lacking its native signal sequence. This recombinant DNA molecule was cloned in pBAD18MCS2 vector [Guzman *et al.*, 1995] where expression is driven from an arabinose inducible promoter. The resultant construct was used to transform an *E. coli pho*A negative strain. In *E. coli*, Alkaline phosphatase [AP] is coded by *pho*A. AP is inactive in the cytosol due to its highly reducing environment. Wild type *E. coli pho*A has a signal sequence its N-terminus, which exports the protein to the peiplasmic space, wherein AP is functional [Brickman and Beckwith, 1975]. This property of *E. coli pho*A was used to test if the N- terminal probable signal sequence of *M. tuberculosis* CM could export *E. coli* AP [lacking the signal sequence] to the periplasmic space thereby making it active. AP activity can be detected on solid media containing X-P [a colorless compound like X-gal, that can be cleaved by AP to form a blue coloured product].

It was observed that the *E. coli pho*A negative strain, transformed with the above mentioned chimeric construct, could demonstrate *pho*A activity. Functional activity of AP

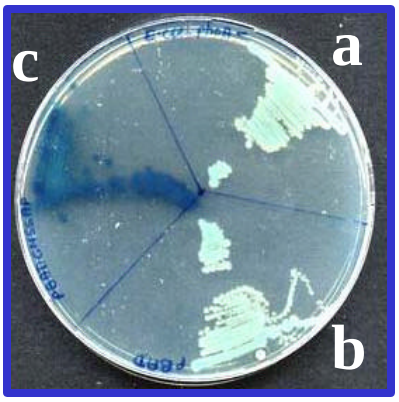
was seen in its ability to cleave the substrate XP, which could be monitored by the appearance of blue colour of the colony [Figure 3.15 A]. This indicated that *M. tuberculosis* CM ss could export AP to the periplasmic space. The *E. coli pho*A negative strain transformed with pBAD18MCS2 vector alone remained white. Localization of *Mtb* CM in the periplasmic and the cytoplasmic fractions of *E. coli* BL21 cells expressing rRv1885c were also assessed using western blot with anti-Histidine antibody. Almost the entire CM activity was confined to the periplasmic space [Fig 3.15 B].

3.4.7 Rv1885c is not the sole chorismate mutase enzyme of *M. tuberculosis*: Rv0948c also shows CM activity though with a reduced turnover

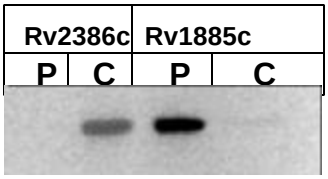
The COG [Cluster of Orthologous groups] annotation of the *M. tuberculosis* genome suggests that CM activity could be attributed to two ORFs of *M. tuberculosis*; Rv1885c and Rv0948c. Hence, *M. tuberculosis* ORF Rv0948c was also cloned and expressed in *E. coli* and the recombinant protein was assayed for CM activity. Results demonstrate that even rRv0948c shows a CM activity though its molar catalytic activity is much lower [50 fold lower than Rv1885c] [Figure 3.16]. The existence of another CM in *M. tuberculosis* genome is consistent with the hypothesis that a free monofunctional CM never exists in a single copy in any genome [Gu *et al.*, 1997]. However, it is inconsistent with the hypothesis that amongst two copies of CMs, at least one is a fusion protein. rRv0948c also showed no PDT or PDH activity indicating that it is also a monofunctional protein. rRv0948c was also found to be resistant to proteolytic digestion in the presence of phenylalanine and tryptophan [Figure 3.12 E, F] indicating that it is a regulated protein.

3.5 DISCUSSION

Chorismate mutase is known to be the sole example of an enzyme that catalyzes a pericyclic rearrangement reaction and has therefore drawn extensive attention of biologists and chemists. Earlier reports that have referred to differences in the structure and regulatory properties of the enzyme from different sources was the basis of this study of the respective properties of the equivalent protein from *M. tuberculosis*. This



A



B

Figure 3.15: The N-terminus of *M. tuberculosis* chorismate mutase (Rv1885c) encodes a functional signal sequence.

A: *E coli pho*A- strain (Sector a) was transformed with pBAD18MCS2 vector(Sector b) and pETCMssAP (Sector c) and streaked on LB-XP plates. Blue colour of colonies was observed only in Sector c, which indicates that signal sequence less *E. coli* alkaline phosphatase could be exported to the periplasmic space by the putative signal sequence in the N-terminus of *M. tuberculosis* CM.

B: Western blot demonstrating the presence of *M. tuberculosis* CM in the periplasmic space: 5 mg of protein from the cytosolic (C) and periplasmic (P) fractions of *E. coli* BL 21 cell expressing *M. tuberculosis* AS (Rv2386c) and CM (Rv1885c) were transferred to a nitrocellulose membrane and probed with monoclonal anti-Histidine antibody. As evident from the blot, only CM (Rv1885c) of *M. tuberculosis* shows a periplasmic localization.

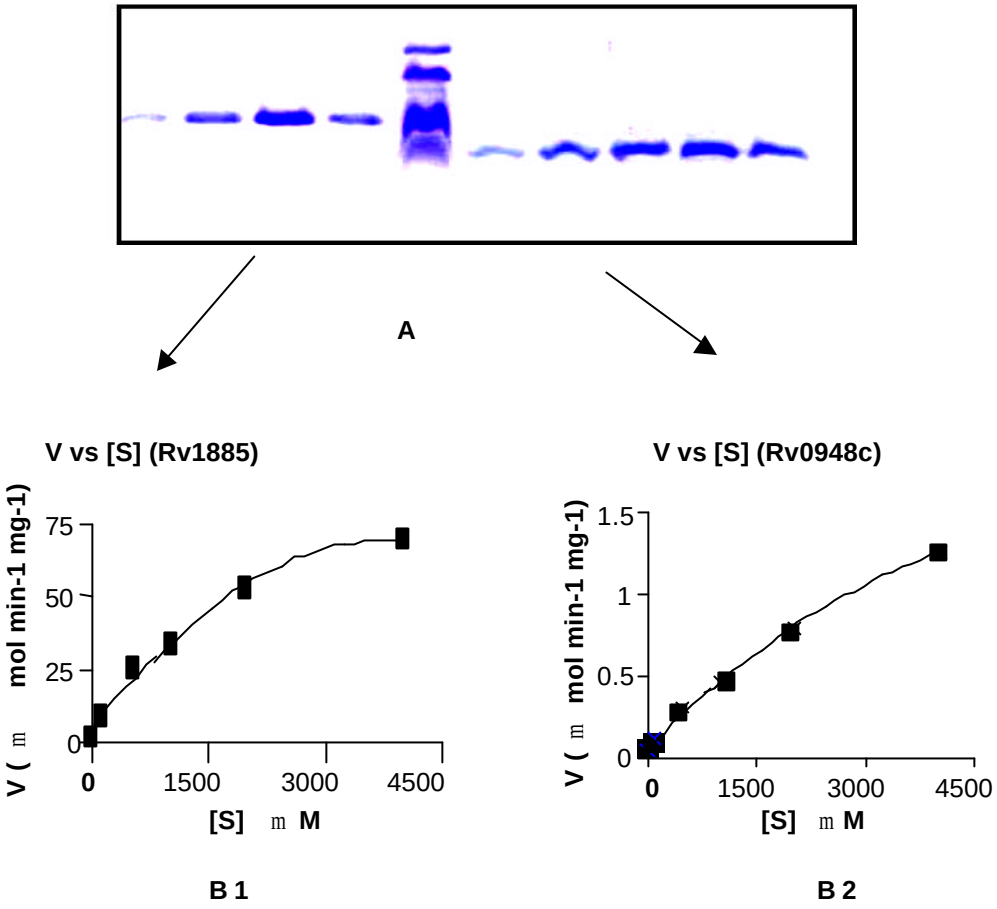


Figure 3.16: Comparative study of the chorismate mutase activity of the ORFs Rv1885c and Rv0948c. A: Purified fractions of both the proteins used for CM activity assay. B: V vs [S] plot of rRv1885c (B1) and rRv0948c (B2). As evident from the graph, Rv0948c does possess CM activity. However, the specific activity of rRv0948c was about 50 fold lower than the specific activity of Rv1885c.

study involved a detailed biochemical and biophysical characterization of recombinant *M. tuberculosis* CM, encoded by ORF Rv1885, and comparison of these properties with enzyme from other sources. The results presented in this chapter conclusively place *M. tuberculosis* CM, coded by ORF Rv1885c in the AroQ class of periplasmic chorismate mutases.

In the course of this study, several unique properties of *M. tuberculosis* CM became evident which are important in the context of differentiating the enzyme from the bacterial and fungal counterparts. The first and foremost observation with respect to the regulation of *M. tuberculosis* CM by aromatic amino acids surfaced an unusual property of the enzyme. While *M. tuberculosis* CM was found to be allosterically regulated by pathway specific [Typ and Phe] as well as cross pathway specific [Trp] ligands in the same manner, it showed a modest increase in enzyme activity at low concentrations and complete inhibition at higher concentrations of effectors. Though the reason behind this modest increase is still unclear, a logical explanation would be that, when intracellular amino acid concentrations are low, its cellular requirement is high and hence the enzyme is moderately activated. It is proposed that this unusual property of *M. tuberculosis* CM must be coordinately regulated with other enzymes of aromatic amino acid biosynthesis pathway. Here, it would be worthwhile to mention that *M. tuberculosis* PheA [Prephenate dehydratase], the enzyme that catalyzes the subsequent step in the biosynthesis of Phenylalanine, is regulated in exactly the opposite manner i.e. it is inhibited by low concentrations and activated by higher concentrations of aromatic amino acids [described in section 4.4.3, Chapter 4].

As binding of ligands causes inhibition of enzyme activity, further experiments were designed to study if ligand binding would also protect the enzyme from proteolytic cleavage. This experiment was expected to generate another line of evidence that ligand binding induces a conformational change in the enzyme. Limited proteolysis studies suggest that ligand binding imparts a much more rigid structure to the enzyme leading to enhanced stability and inaccessibility of the active site and protease cleaving sites. A cross-pathway specific secondary metabolite, salicylic acid could not protect *M. tuberculosis* CM from proteolytic cleavage. Salicylic acid is also derived from chorismate.

This suggests that only aromatic amino acids and not secondary metabolites have a binding pocket on the enzyme.

The chorismate to prephenate reaction is also known to occur spontaneously and an increase in product formation at higher temperatures was also observed. However, there was a distinct enhancement of reaction rate upon addition of the enzyme. It is possible that in spite of a spontaneous conversion of chorismate to prephenate, there is a need for enzyme catalyzed conversion as chorismate is a substrate for multiple enzymes and the fate of the compound is to be determined by the enzyme that acts upon it that in turn is dependent on the requirement of the organism.

It is also interesting to note that unlike most enteric bacteria where chorismate mutase exists in fusion with other enzymes of the aromatic amino acid biosynthesis pathway, *M. tuberculosis* CM is predicted to be a non-fusion protein [http://www.sanger.ac.uk/cgi-bin/Pfam/speciesview.pl?family=CM_2&acc=PF01817&ncbicode=83332&name=Mycobacterium%20tuberculosis]. Yet, on account of several reports of divergent evolution of aromatic amino acid biosynthesis enzymes, experiments were designed to test if the amino acid sequence of *M. tuberculosis* CM could also contribute to PDT or PDH activity. The results presented in this chapter confirm that *M. tuberculosis* CM is a monofunctional protein i.e it has no associated PDT or PDH activity. The CMs that show allosteric regulation are usually large fusion proteins. *M. tuberculosis* CM appears to be unique in the sense that inspite of being a small monofunctional protein, it shows a complex pattern of allosteric regulation. While *M. tuberculosis* does not possess a discrete ACT domain that is required for amino acid binding, it is possible that the complex regulatory activity of the enzyme rests in the uncharacterized C-terminus of the enzyme.

Among the monofunctional AroQ proteins, existence of a cleavable signal peptide has been reported in *Erwinia herbicola*. Later, AroQ CMs from *Salmonella typhimurium* and *Pseudomonas aeruginosa* were also shown to have a periplasmic location [Calhoun *et al.*, 2001]. Periplasmic CMs possess a signal peptide at the N-terminus and usually have a C-terminus of an unknown function. The C-terminus of *M. tuberculosis* CM is also

unique and its precise function is still unclear, though we suggest that it could have a regulatory role. Based upon the ability of *M. tuberculosis* CM signal sequence to export *E. coli* PhoA to the periplasmic space, it was demonstrated that *M. tuberculosis* CM also belongs to this unique periplasmic subclass of AroQ proteins. However, this study also necessitates the requirement for identification of other periplasmic enzymes of aromatic amino acid biosynthesis pathway of *M. tuberculosis*. This study would shed further light on whether there exists a periplasmic pathway of aromatic amino acid biosynthesis in *M. tuberculosis*. In the absence of existence of such a pathway, it is proposed that the periplasmic location of *M. tuberculosis* CM could either have a role to play in the modulation of the host immune response or the enzyme could have a chemotactic role, [Maddock and Shapiro, 1993]. It is significant to note that the ORF Rv1885c exists in the same operon as Antigen85A, a major secreted antigen of *M. tuberculosis*. It is also worthwhile to mention here that chorismate mutase is a virulence factor of nematode parasites that secrete CM from their oesophagal glands to alter the shikimate pathway of the plants which they parasitize [Lambert *et al.*, 1999; Bekal *et al.*, 2003]. While the shikimate pathway does not exist in humans, *M. tuberculosis* chorismate mutase might be released as a virulence factor to alter the shikimate pathway of other bacteria that would have invaded the same host cell. Such a mechanism might allow *M. tuberculosis* to be the sole invader of the host macrophages.

Finally, the evidence for the existence of another CM in *M. tuberculosis* genome is in accordance with the hypothesis that a free monofunctional CM never exists in a single copy in any genome [Gu *et al.*, 1997]. Interestingly however, this is the first report that demonstrates the existence of two monofunctional CMs in any bacterial system. In other bacteria, where CM is present in two copies, at least one of them is a fusion protein [Huang *et al.*, 1974]. Evolution of these fusion proteins was brought about with the purpose of coordination of gene expression as per the requirement of the cell. Hence, the issue of co-regulation of *M. tuberculosis* CMs and PDT/PDH still remains. While IdeR is known to govern the expression of *M. tuberculosis* prephenate dehydratase [Gold *et al.*, 2001], regulation of expression of *M. tuberculosis* CM is still not known. The apparent lack of synchrony in gene expression also suggests that *M. tuberculosis* CM

could be involved in some other pathway that is independent of the involvement of PDT/PDH.

While the data presented in this chapter support the existence of two monofunctional CMs in *M. tuberculosis* genome, yet the question remains, Why two CMs? The clue lies in the observation of a huge difference in turnover number of the two CMs, which suggests that one of these enzymes could have an additional function. Additionally, the existence of one of the CMs as a secreted protein also suggests an as yet undeciphered function of the enzyme. The second chorismate mutase of *M. tuberculosis* [Rv0948c] does not possess a signal sequence which suggests that there are two pathways for aromatic amino acid biosynthesis in *M. tuberculosis*. The role of the periplasmic pathway is still not clear. Our other novel observation that the periplasmic CM of *M. tuberculosis* shows a higher turnover number also demands elucidation of some other function of the enzyme. As chorismate utilizing enzymes are known to have divergently evolved, it is likely that these enzymes could have some other function as well. In *Pseudomonas aeruginosa*, the isochorismate pyruvate lyase [IPL] enzyme, PchB has been reported to show chorismate mutase activity [Gaille *et al*., 2003].

In summary, these results suggest that the hypothetical proteins coded by the ORFs Rv1885c and Rv0948c of *Mycobacterium tuberculosis* represent the chorismate mutase enzymes of the organism and the kinetic properties of the enzyme are under the control of strict regulatory mechanisms. Regulation is brought about by all the three aromatic amino acids *viz* Tryptophan, tyrosine and phenylalanine. This is the first detailed report that provides evidence for the existence of a monofunctional AroQ class of periplasmic CM in *M. tuberculosis* genome. Additionally, a low level sequence similarity of the protein with other known chorismate mutases make this protein an attractive molecule for the design of novel and specific compounds that would check tuberculosis.

CHAPTER 4

---

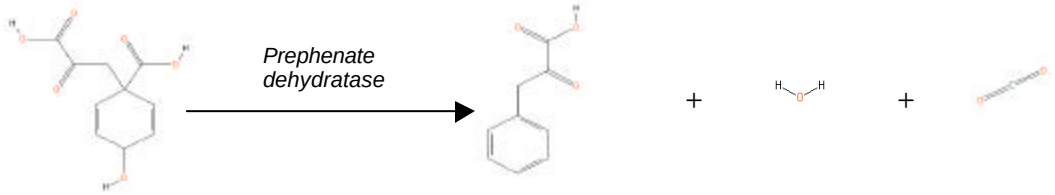
pheA, AN IdeR *REGULATED GENE OF Mycobacterium tuberculosis ENCODES A MONOFUNCTIONAL PREPHENATE DEHYDRATASE THAT REQUIRES BOTH CATALYTIC AND REGULATORY DOMAINS FOR OPTIMUM ACTIVITY*

4.1 ABSTRACT

Prephenate dehydratase [PDT] is a key regulatory enzyme in L-phenylalanine biosynthesis. In *M. tuberculosis, e*xpression of *phe*A, the gene encoding PDT has been earlier reported to be iron-dependent [Gold *et al.*, 2001; Rodriguez *et al.*, 2003]. The work described in this chapter shows that *M. tuberculosis phe*A is also regulated at the protein level by aromatic amino acids. All the three aromatic amino acids [Phenylalanine, Tyrosine and Tryptophan] are potent feedback activators of *M. tuberculosis* PDT. *In vitro* evidence that *M. tuberculosis* PDT does not show any chorismate mutase activity is also presented here which suggests that unlike many other enteric bacteria, where PDT exists as a fusion protein with chorismate mutase, *M. tuberculosis* PDT is a monofunctional and a non-fusion protein. Finally, the biochemical and biophysical properties of the catalytic and regulatory domains [ACT domain] of *M. tuberculosis* PDT were studied. It was observed that in the absence of the ACT domain the enzyme not only loses its regulatory activity but also its catalytic activity. This is the first report of a monofunctional prephenate dehydratase enzyme from a pathogenic bacteria that shows extensive feedback activation by aromatic amino acids and is absolutely dependent upon the presence of catalytic as well as the regulatory domains for optimum enzyme activity.

4.2 INTRODUCTION

*phe*A of *Mycobacterium tuberculosis* has been earlier reported to be an IdeR regulated gene [Gold *et al.*, 2001; Rodriguez *et al.*, 2003]. *phe*A codes for the enzyme prephenate dehydratase [PDT] that catalyzes the conversion of prephenate to phenylpyruvate with the elimination of water and carbon dioxide as shown below.



Prephenate Phenylpyruvate H_2O CO_2

Further, with the action of aromatic aminotransferase, phenylpyruvate is converted to phenylalanine [Figure 3.1]. Biochemical significance of the enzyme lies in its participation in the terminal steps of the biosynthesis of the aromatic amino acid, phenylalanine.

In enteric bacteria PDT usually exists as a fusion protein with chorismate mutase [Figure 4.1], thereby catalyzing the first two steps in the biosynthesis of phenylalanine. In *E. coli* as well, CM and PDT are fusion partners of the bifunctional P-protein, coded by *phe*A [Pittard, 1987]. In many other bacteria, PDT is a monofunctional protein that aligns well with the C-terminal part of P-proteins [Fischer and Jensen, 1987]. The N-terminal end of the bifunctional P-protein of *E. coli* specifies the chorismate mutase activity while the remainder of the sequence specifies the prephenate dehydratase enzymatic activity [Zhang *et al.*, 1998]. The native enzyme is a dimer of identical subunits, each containing a dehydratase active site, a mutase active site and a phenylalanine binding site. [Gething and Davidson, 1977]. PDT is reported as a highly regulated enzyme in bacterial systems [Pohnert *et al.*, 1999]. Protein engineering studies have established that the *E. coli* P-protein has three distinct domains and their functions are preserved when the individual domains are expressed separately as fragments [Zhang *et al.*, 1998]. While the CM domain rests in residues 1-109, the PDT domain includes residues 101-285. The

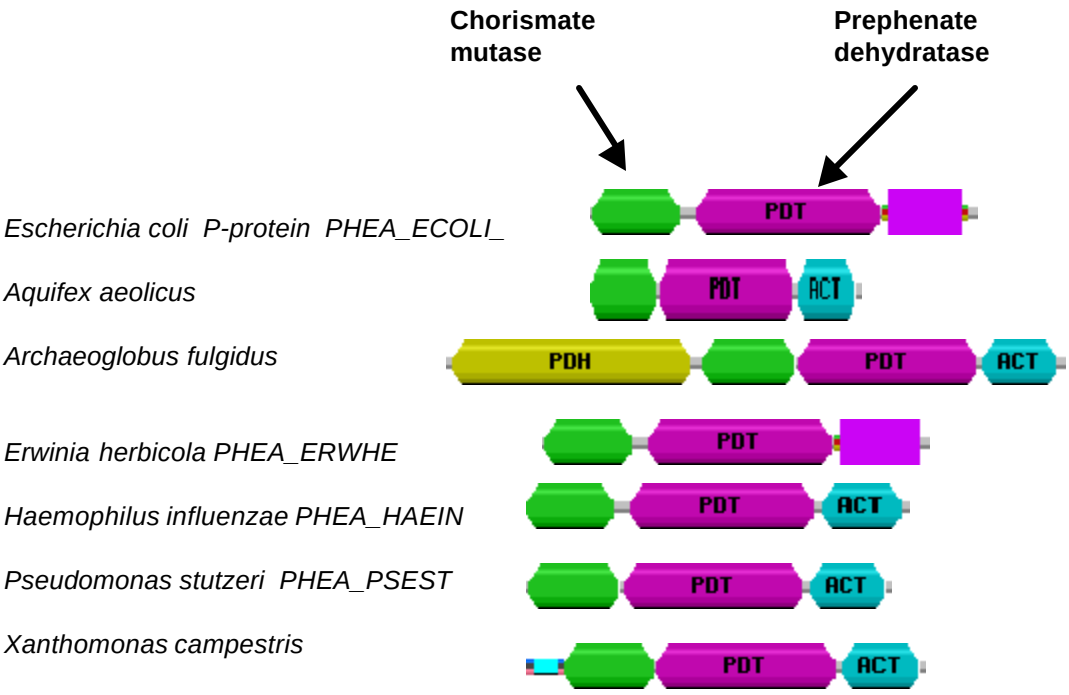


Figure 4.1: Prephenate dehydratase (PDT) exists as a fusion protein with chorismate mutase in most enteric bacteria
Source: Pfam

authors could also decipher a separate allosteric domain [residues 286–386], which was responsible for feedback inhibition by Phenylalanine. Further, the authors also showed that the Phe-binding domain of the P-protein might function as a 'modular' regulatory element when physically joined to an independent protein catalyst [Zhang *et al.*, 2003].

The chorismate mutase and prephenate dehydratase reactions in the P protein of *E. coli* occur at separate active sites. In contrast, the two active sites of the closely related enzyme chorismate mutase/prephenate dehydrogenase [*E. coli* T protein] are interacting sites [Hudson *et al.*, 1984]. Studies of the overall reaction using radioactive chorismate have shown that prephenate, which is formed from chorismate, dissociates from the mutase site and equilibrates with the bulk medium before combining at the dehydratase site. Also, the two activities are subject to differential inhibition. [Duggleby *et al.*, 1978]

In a few organisms like *Bacillus subtilis*, prephenate dehydratase is a monofunctional protein [Riepl *et al.*, 1979]. In *M. tuberculosis* as well, the ORF Rv3838c, annotated as a prephenate dehydratase appears monofunctional in the sense that it possesses only a prephenate dehydratase domain.

Regulatory domains or the ACT domains are present in many metabolic enzymes including enzymes of aromatic amino acid biosynthesis family [Chipman and Shaanan, 2001]. Amino acids that have a regulatory role in the activity of an enzyme usually bind to the ACT domain. The ACT domain is found in a variety of contexts and is proposed to be a conserved regulatory binding fold [Chipman and Shaanan, 2001]. ACT domains are linked to a wide range of metabolic enzymes that are regulated by amino acid concentration. The archetypical ACT domain is the C-terminal regulatory domain of 3-phosphoglycerate dehydrogenase [3PGDH], which folds with a ferredoxin-like topology [Schuller *et al.*, 1995]. The prephenate dehydratase of *M. tuberculosis* also possesses discrete ACT domain that is predicted to impart regulatory properties to the enzyme. In other bacteria as well, prephenate dehydratase is predicted to be a highly regulated enzyme.

PDT enzyme is known to exist in various oligomeric states in different bacterial systems. PDT of *Bacillus subtilis* has been reported to exist in three states of aggregation, an octamer, a dimer and a monomer (Riepl and Glover, 1979).

With the above background this part of the study was initiated to determine whether apart from genetic regulation, *M. tuberculosis phe*A is also regulated at the enzymatic level. The experimental approach involved expression of the *M. tuberculosis* ORF, Rv3838c in *E. coli* and determination of the biochemical parameters of the encoded protein. This chapter gives a brief description of the kinetic and regulatory properties of *M. tuberculosis* PDT and also details the use of engineered proteins to elucidate the role of the catalytic and regulatory domains of the recombinant protein. The significance of a differential feedback regulation of the terminal enzymes of the aromatic amino acid biosynthesis pathway of *M. tuberculosis* is also discussed.

4.3 EXPERIMENTAL PROCEDURES

4.3.1 Bacterial strains and plasmids:

E. coli XL1Blue and *E. coli* BL21 DE3 strains were used respectively as hosts for cloning and recombinant protein expression purposes. pET23a [Novagen] was taken as the expression vector.

4.3.2 Cloning, expression and purification of *M. tuberculosis* PDT and PDT-N and PDT-C

The ORF Rv3838c that corresponds to the prephenate dehydratase enzyme of the bacterium was amplified from *M. tuberculosis* H37Rv genomic DNA using primers carrying specific restriction enzyme sites [Table 4.1]. The amplicon were digested with *Nde*I/*Xho*I enzymes and cloned into the corresponding sites of pET23a expression

Table 4.1: Sequences of primers used for PCR amplification of *M. tuberculosis* prephenate dehydratase [coded by ORF Rv3838c] and its N and C termini [catalytic and regulatory domains]

Gene	Forward primer sequence	Reverse primer sequence
Rv3838c	*Nde*I ATA***CATATG***GTGGTGCGTATCGCTTACCTCGGT	*Xho*I AT**CTCGAG**TGCTTGCGCCCCCTGGT
Rv3838N	*Nde*I ATA***CATATG***GTGGTGCGTATCGCTTACCTCGGT	*Xho*I AT**CTCGAG**ATCGGCTCCGGTGCGCG
Rv3838C	*Nde*I ATA***CATATG***CGCACGTCTGCAGTGCTGC	*Xho*I AT**CTCGAG**TGCTTGCGCCCCCTGGT

vector. The N and the C terminus of *M. tuberculosis* PDT corresponding to the catalytic and regulatory domains were cloned separately using other sets of primers described in Table 4.1. The resultant plasmids were labeled as pET3838, pET3838N and pET3838C.

The pET: 3838/3838-N/3838-C chimeric constructs were used to transform *E coli* BL21DE3 cells and protein expression and purification was carried out exactly as described in Chapter 2, section 2.3.2.

4.3.3 Enzyme assays and kinetic studies

4.3.3.1 Prephenate dehydratase activity assay

Prephenate dehydratase activity assays were carried out as described earlier with a few modifications [Gething *et al.*, 1977]. Essentially, the dehydratase activity of rRv3838c was assayed by measuring the rate of conversion of prephenate to phenylpyruvate. The reaction mixture contained 20mM Tris-HCl, pH 8.2, 1mM EDTA, 0.01% BSA, 1mM DTT and 0.5-2mM barium prephenate in a total volume of 400µl. The sample was pre-incubated at 37^0C followed by addition of 20-100 pico moles of purified recombinant *M. tuberculosis* PDT. After a second round of incubation at 37^0C for 5 minutes, the reaction was terminated by the addition of 800µl of 1.5M NaOH. Phenylpyruvate was measured spectrophotometrically at 320nm. Appropriate blanks without the enzyme were kept as controls. Allosteric regulation of CM activity by L-Phe, L-Tyr and L-Trp was measured at 100-500µM concentrations of the effectors.

4.3.3.2 Chorismate mutase activity assay

Chorismate mutase activity assays were performed exactly as described in Section 3.3.4, Chapter 3

4.3.4 Analytical size exclusion chromatography

Gel filtration or size exclusion chromatography was used to determine the oligomeric state of *M. tuberculosis* PDT as well as the individually expressed catalytic and regulatory domains. The chromatography experiment was performed on a Superdex 200 [HP 10/30] FPLC column from Pharmacia Biotech using 10mM Tris and 100mM NaCl as the running buffer. Void volume of the column was determined using Blue Dextran 200. Elution time of all the recombinant proteins was recorded and molecular weight was calculated by estimating the elution volumes of standards of known molecular weight. The procedure is described in detail in Section 3.3.10 of Chapter 3. The recombinant proteins were loaded on the gel filtration column at a concentration of 2mg/ml in the presence of 1mM DTT.

4.3.5 Phenylalanine binding assays / Fluorimetric procedures

Binding of phenylalanine to the recombinant proteins was monitored using fluorescence spectroscopy. Fluorescence spectra of individual proteins were recorded in the presence and absence of phenylalanine on a Perkin-Elmer LS-3B spectrofluorimeter. Briefly, the recombinant proteins were excited using an excitation wavelength of 295nm and the tryptophan emission spectra were recorded from 305-440nm at 37^0C. Fluorescence slit width of excitation and emission were kept at 4nm and the scan speed was 50nm/sec. Protein concentrations were kept at 2.5μM in 10mM Tris, pH8 and the buffer signal was subtracted from the spectrum of recombinant proteins.

4.4 RESULTS

4.4.1 *phe*A (ORF Rv3838c) of *Mycobacterium tuberculosis* encodes a functional prephenate dehydratase enzyme

The recombinant protein corresponding to ORF Rv3838c, that is predicted to be the prephenate dehydratase [PDT] enzyme of the bacterium could be purified using affinity chromatography procedures [Figure 4.2]. Purified protein was quantified using Bradford assay and used for prephenate dehydratase activity assays. As evident from the substrate saturation plot, *M. tuberculosis* PDT followed Michealis Menton's kinetics [Figure 4.3A]. The double reciprocal plot of the enzyme was linear and Km for the enzyme was calculated as 1.38mM and Vmax as 270 μmoles/min/mg [Figure 4.3B]. Specific activity of the enzyme was determined as 270.27 units/mg pure protein. The effect of temperature and pH on enzyme activity was also studied. It was observed that the enzyme is maximally active between temperatures 37-42^{0}C and pH 6-7 [Figure 4.4: A,B]. Unlike the chorismate mutase of *M. tuberculosis*, *M. tuberculosis* prephenate dehydratase has a narrow range of tolerance of temperature and pH. Also, while other bacterial PDTs are active at an acidic pH [Friedrich *et al.*, 1976], *M. tuberculosis* PDT was found to be active from mildly acidic to near neutral pH.

4.4.2 Ionic interactions are required for optimum PDT activity

To determine if ionic interactions are important for the activity of *M. tuberculosis* PDT, the enzyme was assayed in the presence of various concentrations of NaCl. It was observed that the enzyme activity was completely abolished in the presence of 200mM and higher concentrations of NaCl [Figure 4.5]. This suggests that enzyme substrate interactions are brought about by ionic interactions and disruption of the same leads to inhibition of enzyme activity.

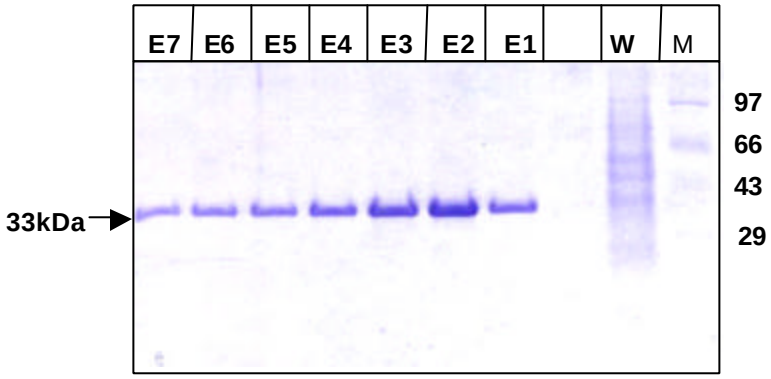


Figure 4.2: Purification of *M. tuberculosis* prephenate dehydratase (PDT) as a recombinant protein in *E. coli*. *Mtb phe*A corresponding to ORF Rv3838c was cloned in the *Nde*I and *Xho*I sites of pET23a vector with a C-terminal Histidine tag and expressed in *E. coli* BL21 cells. Affinity purification of recombinant protein was carried out using Talon resin (Clonetech). Purified protein resolved on a 10% Tris-Tricine gel and stained with Coomassie Brilliant Blue dye. M represents the protein molecular size marker (Genei, India), W represents the wash fraction during affinity purification of recombinant protein and E1-E7 show the successive eluted fractions of the 33kDa recombinant protein.

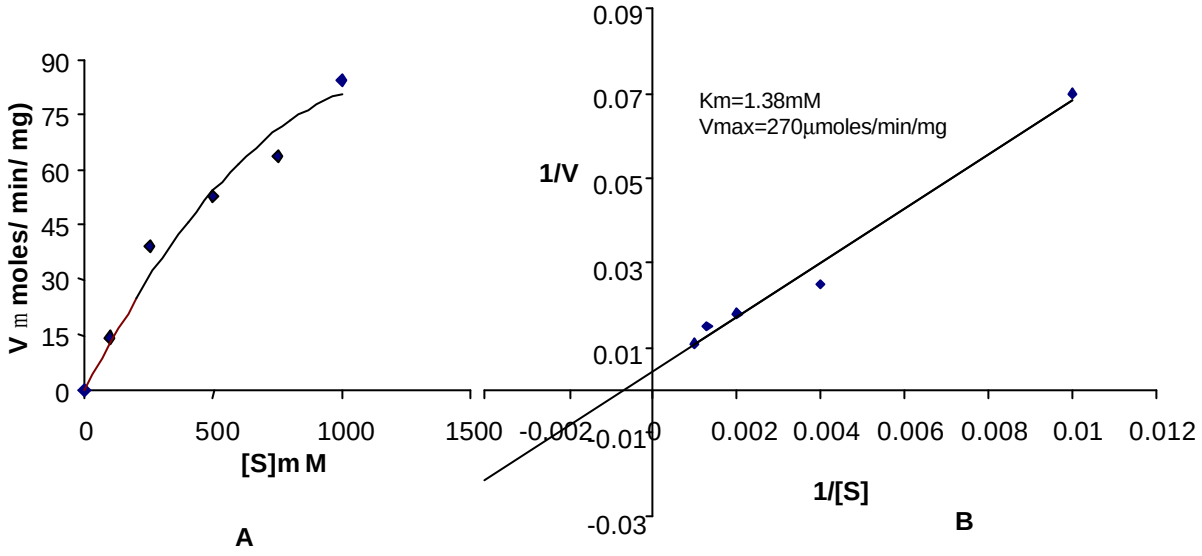


Figure 4.3: A: Substrate saturation plot of *Mtb* Prephenate dehydratase. The enzyme follows Michealis Menton's kinetics B: Double reciprocal plot of *Mtb* PDT enzyme activity that was used to calculate Km and Vmax. The specific activity of recombinant *Mtb* PDT was calculated as 270 units/mg pure protein.

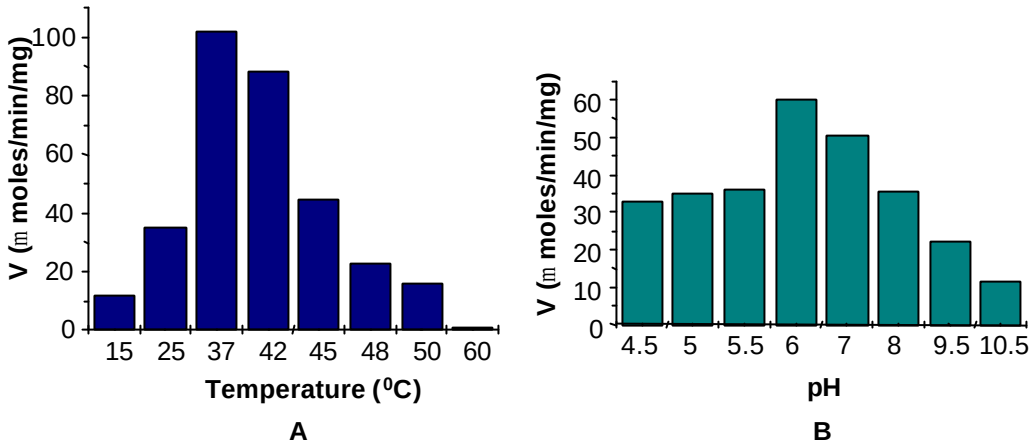


Figure 4.4: Effect of temperature and pH on the prephenate dehydratase (PDT) activity of rRv3838. Optimum pH for *M. tuberculosis* PDT activity is between pH 6-7. Optimum temperature for *M. tuberculosis* PDT activity is 37ºC. Temperatures above 50^0C completely inactivate the enzyme

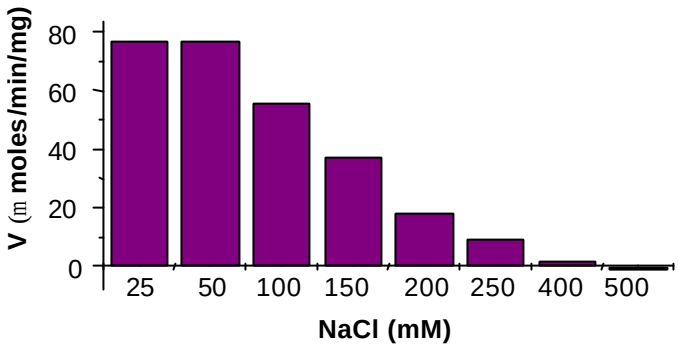


Figure 4.5: Ionic interactions are required for optimum catalytic activity of *M. tuberculosis* prephenate dehydratase (PDT) . Higher concentrations of NaCl (200mM and above completely abolish the PDT activity of *M. tuberculosis* ORF, Rv3838c.

4.4.3 Aromatic amino acids are potent feedback activators of *M. tuberculosis* prephenate dehydratase

Control of the terminal enzymes in any biosynthetic pathway is crucial for maintaining the correct balance of the end product in accordance with the requirement of the organism. This holds true for the aromatic amino acid biosynthesis as well. PDT being a terminal enzyme in phenylalanine biosynthesis, it was decided to determine the effect of phenylalanine as well as other cross pathway specific aromatic amino acids [tyrosine and tryptophan] on enzyme activity. It was observed that PDT activity is allosterically regulated in a manner exactly opposite to that of *M. tuberculosis* CM [refer section 3.4.3, chapter 3]. More specifically, *M. tuberculosis* PDT was inhibited by lower concentrations of aromatic amino acids [up to ~100μM] and highly activated at higher concentrations [Figure 4.6 A,B and C]

4.4.4 *M. tuberculosis* prephenate dehydratase does not display any chorismate mutase activity

While the chorismate mutase enzyme of *M. tuberculosis* did not show any prephenate dehydratase activity [Section 3.4.2], it was determined if the reverse also holds true i.e. whether the PDT of *M. tuberculosis* displays any CM activity. Though *M. tuberculosis* PDT does not have a predicted CM domain, on account of several examples of convergent evolution of enzyme reaction mechanisms, it was decided to determine whether PDT also has any associated CM activity. Recombinant *M. tuberculosis* PDT was therefore used in a chorismate mutase activity assay. It was observed that *M. tuberculosis* PDT is completely devoid of CM activity. This experiment confirmed that like *M. tuberculosis* CM, *M. tuberculosis* PDT is also a monofunctional protein.

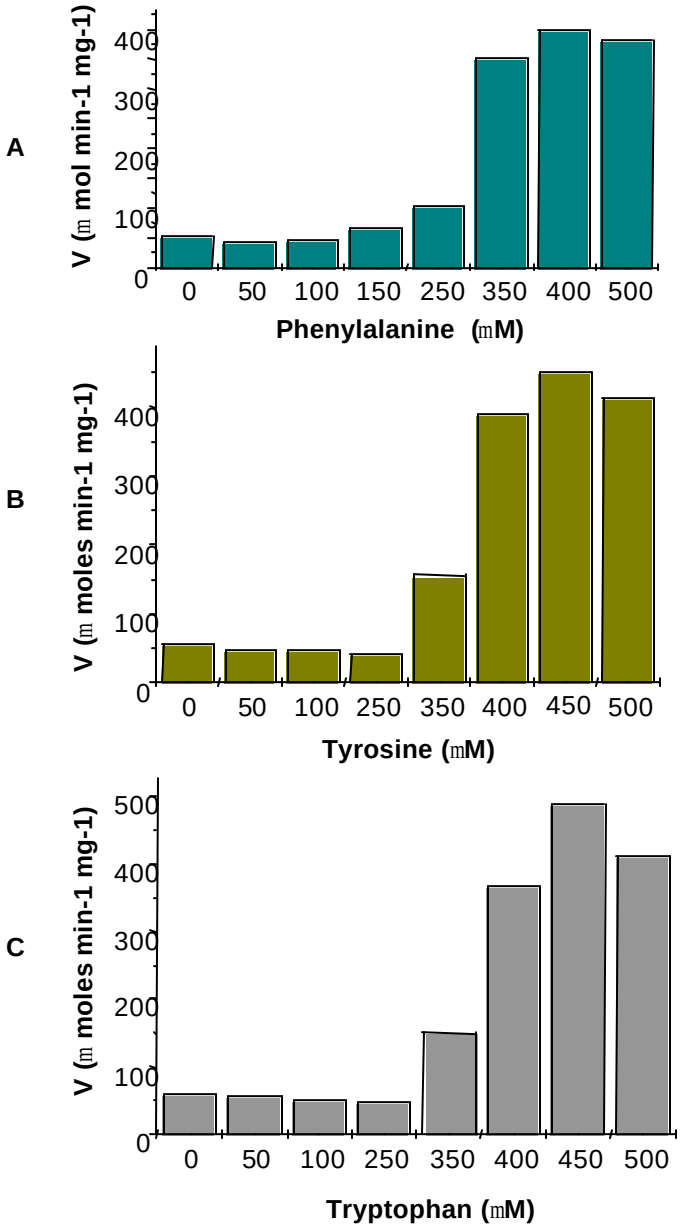


Figure 4.6: Aromatic amino acids at higher concentrations are potent feedback activators of *M. tuberculosis* PDT. Up to 100μM concentrations of aromatic amino acids negatively regulate *M. tuberculosis* PDT. However, a positive regulation was observed at higher concentrations (300μM and above).

4.4.5 The individually cloned, expressed and purified catalytic and regulatory domains of *M. tuberculosis* prephenate dehydratase are inactive in catalyzing the conversion of prephenate to phenylpyruvate

M. tuberculosis PDT has a distinct predicted catalytic and regulatory domain [Figure 4.7A]. Hence it was decided to test whether the catalytic activity of *M. tuberculosis* PDT is independent of the regulatory domain or it requires both the domains for optimum activity. For this purpose, the catalytic and regulatory domains of *M. tuberculosis* PDT were individually cloned and expressed in the *Nde*I/*Xho*I sites of pET23a vector and purified as recombinant proteins [Figure 4.7 B,C]. These proteins were labeled as PDT-N and PDT-C and were used in prephenate dehydratase activity assay. While the regulatory domain was completely inactive, it was surprising to note that even the individually cloned catalytic domain of *M. tuberculosis* PDT did not show any PDT activity [Figure 4.8]. Reconstitution of the enzyme was attempted by taking equimolar ratios of the catalytic and regulatory domains and using it in PDT activity assays. However, the activity assay demonstrated that the catalytic and regulatory domains can not be reconstituted *in vitro* to generate an active enzyme. These results are summarized in a tabular form [Table 4.2].

4.4.6 Phenylalanine binding leads to a conformational change in *M. tuberculosis* prephenate dehydratase enzyme and its regulatory domain

Fluorescence spectroscopy was used to determine if there is any conformational change in *M. tuberculosis* PDT upon ligand binding. This study was carried out with full length *M. tuberculosis* PDT as well as individual catalytic and regulatory domains. Tryptophan emission spectra for all the three recombinant proteins were monitored in the presence and absence of phenylalanine. *M. tuberculosis* PDT possesses four tryptophan residues and the individual catalytic and regulatory domains have 2 tryptophan residues each. As evident from the graph, the addition of phenylalanine leads to a change in the fluorescence emission spectrum of only the full length *M. tuberculosis* PDT and PDT-C [Figure 4.9: A, B].

Table 4.2: Activity profile of the engineered *M. tuberculosis* PDT proteins

Clones	Protein	Residues	Purified protein activity PDT	CM
pET3838	PDT	1-321	+	_
pET3838N	PDT-N	4-192	_	NT
pET3838C	PDT-C	202-279	_	NT
	PDT-C + PDT-N		_	NT

NT: Not Tested;
+:active;
-: inactive

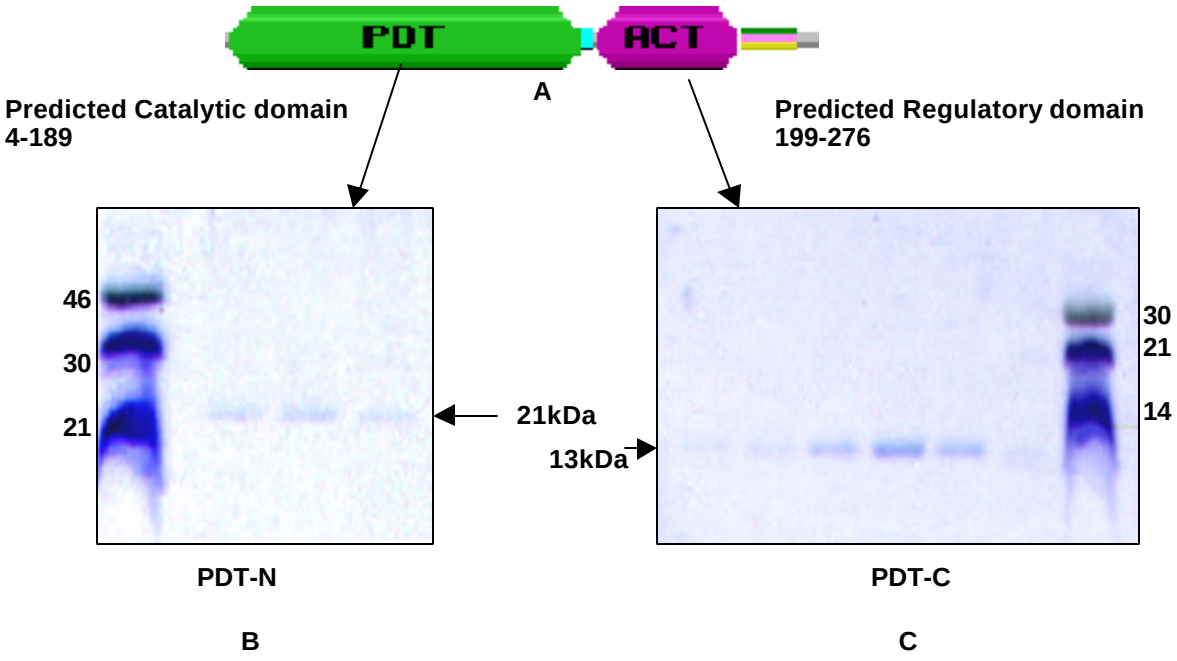


Figure 4.7: Expression and purification of the catalytic (PDT-N) and regulatory domains (PDT-C) of *M. tuberculosis* prephenate dehydratase. The individual domains were cloned in the *Nde*I/*Xho*I sites of pET23a expression vector and purified as recombinant proteins using affinity chromatography procedures. A: Schematic representation of the catalytic and regulatory domains B: Coomassie Blue stained polyacrylamide gel showing the eluted fractions of the pure recombinant proteins corresponding to catalytic and regulatory domains.

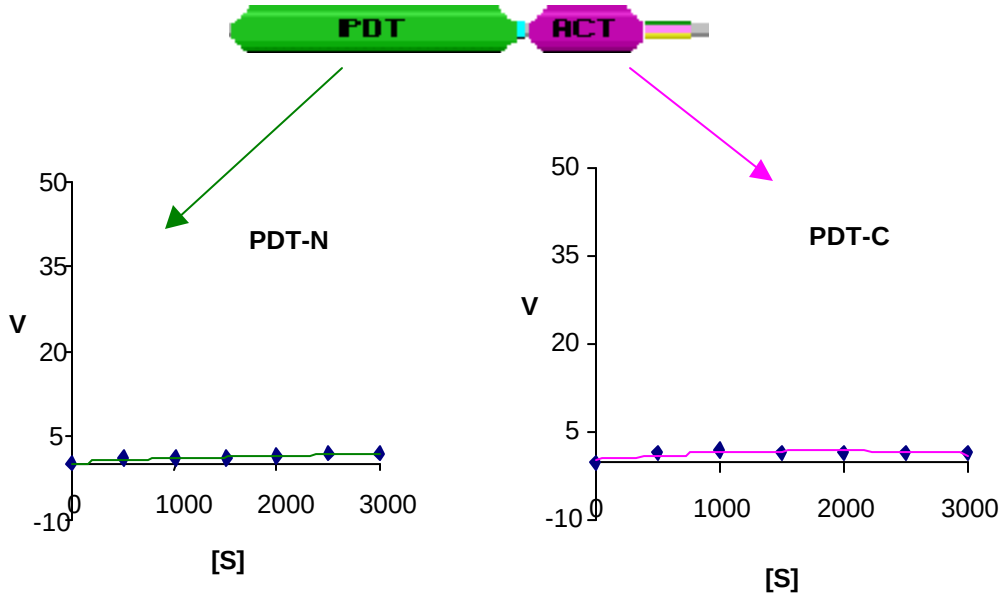


Figure 4.8: The catalytic (PDT-N) as well as the regulatory domains (PDT-C) of *M. tuberculosis* prephenate dehydratase in isolation can not catalyze the conversion of prephenate to phenylpyruvate. Lack of PDT activity of the individually purified catalytic and regulatory domains as evident from the above V vs. [S] plots suggest that interactions between both the domains are essential for enzyme activity.

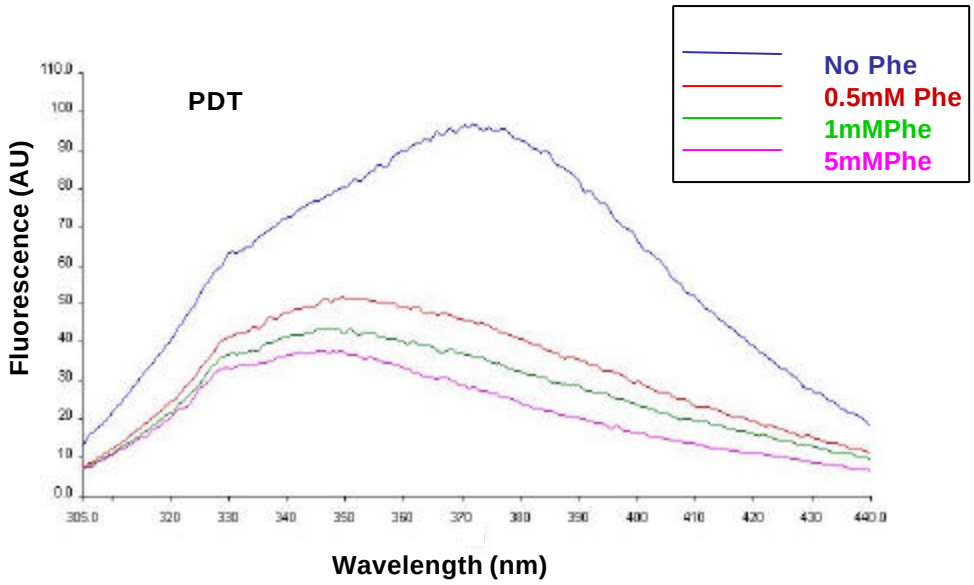
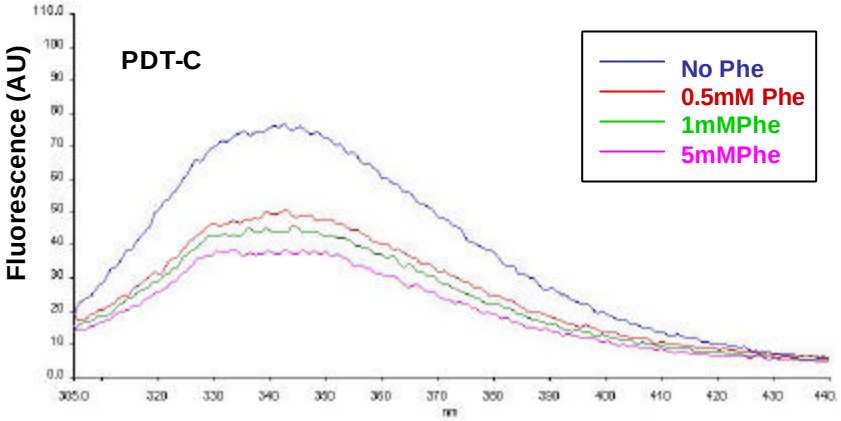
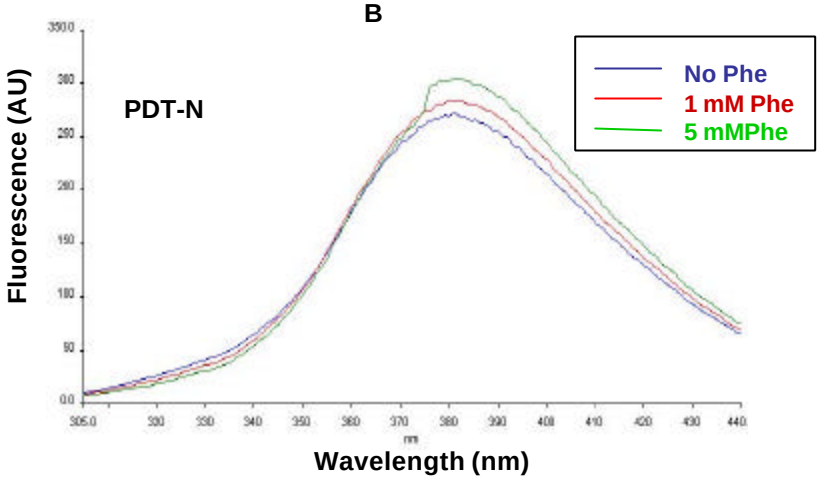


Figure 4.9: A: Fluorescence spectrum of *M. tuberculosis* prephenate dehydratase changes significantly in the presence of phenylalanine. The intrinsic fluorescence intensity decreased and the emission maxima shifted from 380nm to 350nm. This suggests that binding of phenylalanine renders a conformational change upon the enzyme. A blue shift in the spectra also indicates that tryptophan residues of the protein are buried inside upon phenylalanine binding.



B



C

Figure 4.9: B: Fluorescence emission spectrum of the PDT-C (Regulatory domain of *M. tuberculosis* prephenate dehydratase) is altered in the presence of phenylalanine. The intrinsic fluorescence intensity decreased though the emission maxima remained the same. This suggests that PDT-C undergoes a conformational change as a consequence of phenylalanine binding. C: Fluorescence emission spectrum of PDT-N (catalytic domain) remains unchanged upon the addition of phenylalanine. These experiments suggest that the phenylalanine regulatory domain resides only in the C-terminus of *Mtb* PDT.

The fluorescence emission spectrum of the catalytic domain or PDT-N remained unchanged even in the presence of phenylalanine [Figure 4.9C]. In case of full length *M. tuberculosis* PDT, a blue shift in the fluorescence spectrum along with quenching of fluorescence was seen [Figure 4.9A], while in case of PDT-C only quenching of fluorescence was seen [Figure 4.9B]. A blue shift in fluorescence spectrum and quenching indicates that the tryptophan residues of the native protein are buried inside as a consequence of the conformational change that results from the binding of phenylalanine.

4.4.7 Size exclusion chromatography reveals that *M. tuberculosis* PDT is an oligomer and PDT-N and PDT-C are monomers

Size exclusion chromatography was performed to determine the oligomeric state of *M. tuberculosis* PDT and its individual catalytic and regulatory domains. As the individual domains were catalytically inactive, this experiment was supposed to give insights into the possible reasons behind the catalytic insufficiency of the individual domain. Size exclusion was performed on a Superdex 200 column and a standard curve was drawn with standard protein MW markers. All the three proteins eluted as a single peak [Figure 4.10]. The molecular weight of all the proteins were calculated using a standard curve. The results showed that *M. tuberculosis* PDT is a hexameric protein and its individual catalytic and regulatory domains elute as monomers.

4.5 DISCUSSION

PDT exists both as a monofunctional protein and as a fusion protein with chorismate mutase. In *E. coli*, both CM and PDT activities reside in a single, bifunctional protein known as the P-protein, encoded by *phe*A [Pittard, 1987]. Amongst the monofunctional PDTs, almost all possess an ACT or amino acid binding domain that brings about feedback regulation activity of the enzyme.

In this work, *M. tuberculosis* PDT, an apparently monofunctional protein, was proved to

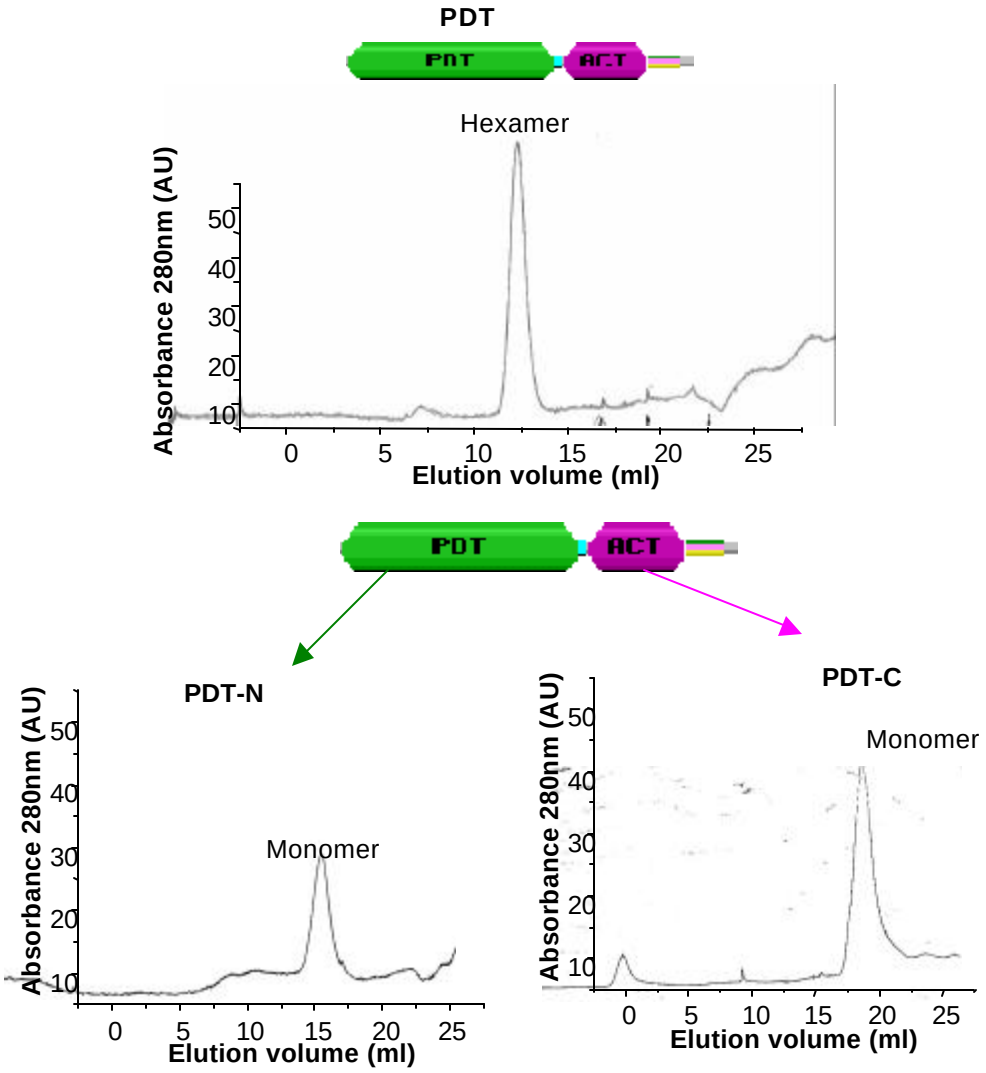


Figure 4.10: Size exclusion chromatography reveals that *M. tuberculosis* prephenate dehydratase (PDT) is a hexameric protein and its individually cloned catalytic (PDT-N) and regulatory domains (PDT-C) exist as monomers. The recombinant proteins were individually loaded on a Superdex 200 column (Pharmacia Biotech) and absorbance at 280nm (AU) was plotted as a function of elution volume. The elution parameter, Kav vs log MW plot of protein standards was used to calculate the actual MW of these proteins.

be so using a biochemical approach. The specific activity of full length recombinant PDT was very high [270 units/mg pure protein] and the enzyme was extensively feedback activated by all the three aromatic amino acids [Phe, Tyr and Trp]. Earlier reports have shown that *M. tuberculosis phe*A is an IdeR regulated gene [Gold *et al.*, 2001; Rodriguez *et al.*, 2003]. The present work demonstrates that *M. tuberculosis phe*A is also regulated at the protein level by aromatic amino acids. The high energy cost for the biosynthesis of aromatic amino acids [Atkinson, 1977] could be responsible for dual regulation [gene level as well as protein level] of *M. tuberculosis* PDT. However, dual regulation [specifically by IdeR] also suggests the involvement of the aromatic amino acid biosynthesis pathway in the biosynthesis of iron acquisition systems. Though phenylalanine as a precursor of salicylate *via* the phenyl ammonia lyase pathway is well documented in plants, this route of salicylate biosynthesis is not at all reported in bacteria [Wildermuth *et al.*, 2001]. It is therefore possible that *M. tuberculosis* could have an additional mechanism of salicylate biosynthesis that involves PDT. The dual regulation of *M. tuberculosis* PDT and specifically feedback activation by aromatic amino acids differentiates it form all other bacteria.

When the regulatory properties of *M. tuberculosis* CM and *M. tuberculosis* PDT were compared, the first observation was that both the enzymes display an exactly opposite pattern of regulation [Chapter 3, section 3.4.3]. While *M. tuberculosis* CM shows moderate activation by low concentrations of aromatic amino acids, activity was greatly inhibited at higher concentrations. The reverse holds true for *M. tuberculosis* PDT i.e. the enzyme is inhibited by low concentrations of aromatic amino acids and highly activated by higher concentrations. These results demonstrate for the first time that despite the occurrence of CM and PDT as monofunctional proteins in *M. tuberculosis*, correct balance of aromatic amino acids is brought about by opposite regulation of these two enzymes at the protein level.

Another important goal of the present study was to ascertain if the predicted catalytic domain of *M. tuberculosis* PDT, that lies in the N-terminus is independent of the feedback regulatory domain that is part of the C-terminus. Results presented in this chapter indicate that a discrete regulatory [ACT] domain along with a catalytic domain is

required for an optimally active *M. tuberculosis* PDT. These observations were found to be unlike the studies carried out on *E. coli* P protein where discrete domains of the complex P-protein retain their original activity [Zhang *et al.*, 1998]. However, in case of *E. coli* T protein wherein though CM and PDH domains can be expressed independently as functional proteins, the efficiency of the enzymes in isolation is much reduced as compared to the entire fusion protein [Chen *et al.*, 2003]. In neither case however, is the activity completely abolished, when the PDH or PDT domains are taken separately.

Various possible explanations can be considered for the complete loss of PDT activity upon removal of the regulatory domain. The dissociation of the oligomeric state of the enzyme upon removal of the C- terminus could be one possible reason. It is also possible that in the individually expressed and purified N-terminus of *M. tuberculosis* PDT, the substrate binding site is mechanistically altered leading to a loss of activity.

While the catalytic domain of *M. tuberculosis* PDT in isolation was functionally inactive, at least the individually cloned C-terminus of *M. tuberculosis* PDT retained its regulatory properties. Proof of the same was obtained using fluorimetric assays. These experiments generate sufficient line of evidence to conclude that there is a definite conformational change in the regulatory domain of *M. tuberculosis* PDT in the presence of phenylalanine. On the other hand, fluorescence emission spectra of the catalytic domain of *M. tuberculosis* PDT remain unchanged even in the presence of phenylalanine. It is possible that the C-terminus of *M. tuberculosis* PDT may serve as a modular regulatory domain that can be used to impart regulatory properties to other proteins of interest. These proteins can in turn be used as biosensors.

In summary, this study is the first report of a prephenate dehydratase enzyme that shows feedback activation by aromatic amino acids and is absolutely dependent upon the catalytic as well as the regulatory domains for optimum enzyme activity. The various levels at which the regulation of aromatic amino acid biosynthesis is brought about in *Mycobacterium tuberculosis* suggest that expression of the corresponding genes are strictly dependent upon the

requirement of the bacterium. Additionally, considering the absence of a human homologue of PDT, the enzyme might serve as a novel target for the design of chemotherapeutic compounds.

CHAPTER 5

---

*MOLECULAR DISSECTION OF THE FUNCTIONS OF trp*E *AND trp*E2 *OF Mycobacterium tuberculosis*

5.1 ABSTRACT

Anthranilate synthase [AS] and Isochorismate synthase [ICS] are other two chorismate utilizing enzymes that share extensive sequence similarity [Serino *et al.*, 1995]. In the *M. tuberculosis* genome, a tryptophan operon has been annotated with a probable anthranilate synthase or *trp*E [Rv1609]. However, in the initial genome sequence of *M. tuberculosis*, another ORF [Rv2386c], a part of the *mbt* operon [mycobactin biosynthesis operon] was reported as the second *trp*E [Cole *et al.*, 1998] and annotated as *trp*E2. Later, due to its unique location and regulation by IdeR, Rv2386c was reannotated as an isochosmate synthase [*mbt*I] [Quadri *et al.*,1998; Dussurget *et al.*, 1996]. This part of the study was carried out with the objective of actual assessment of the functions of the *trp*E and *trp*E2/*mbt*I genes of *M. tuberculosis*. The results presented in this chapter provide evidence that TrpE of *M. tuberculosis* [Rv1609] is an ammonia dependent anthranilate synthase and TrpE2/MbtI [Rv2386c] is a chorismate utilizing enzyme that catalyzes the conversion of chorismate to isochorismate. Furthermore, *trp*E as well as *trp*E2 could complement a *trp*E mutant *E. coli* strain indicating that Rv2386c possesses AS as well as ICS activities. These results have a bearing not only on the understanding of the role of these chorismate utilizing enzymes in the pathogenesis of *M. tuberculosis* but also on the behaviour and activity profile of divergently as well as convergently evolved group of enzymes.

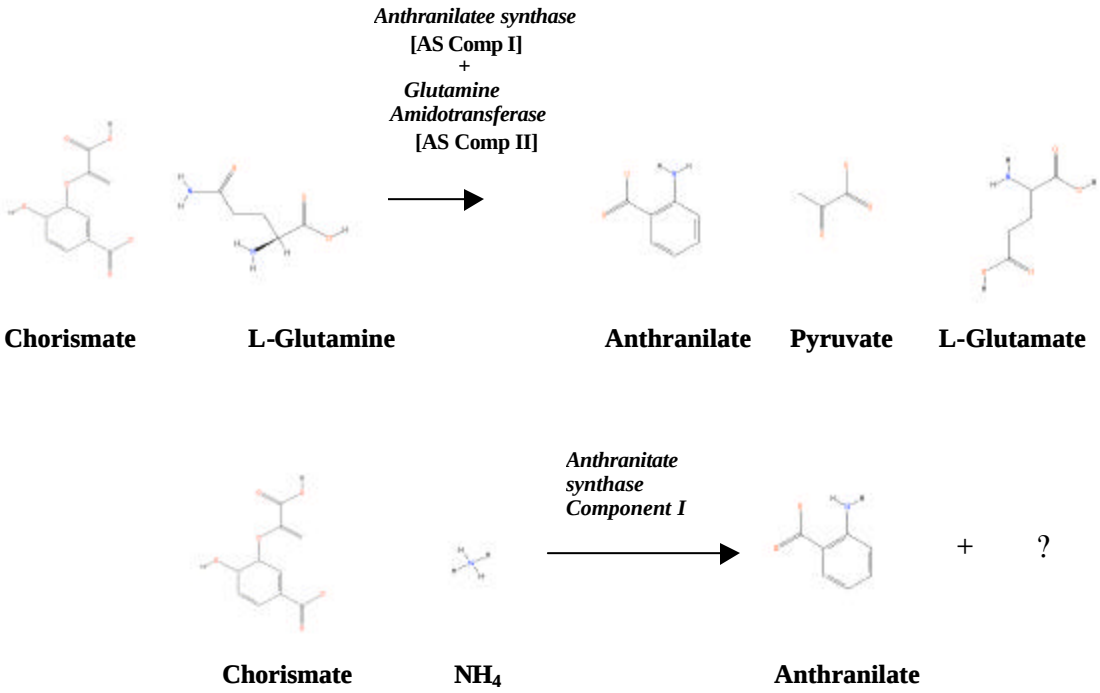
5.2 INTRODUCTION

In a continued effort to gain an insight into the physiological roles of the aromatic amino acid biosynthesis enzymes of *M. tuberculosis*, the study was extended to functional characterization of additional two chorismate utilizing enzymes, anthranilate synthase and isochorismate synthase of the bacterium. While the ORF Rv1609 is annotated as an anthranilate synthase [*trp*E], Rv2386c was earlier annotated as an anthranilate synthase [*trp*E2] and later reannotated as an isochorismate synthase [*mbt*I]. *trp*E2/*mbt*I has also been reported to be IdeR regulated [Gold *et al.*, 2001; Rodriguez *et al.*, 2003]. The proceeding sections provide a brief description of the reaction catalyzed by anthranilate synthase and isochorismate synthase.

5.2.1 Anthranilate synthase

Anthranilate synthase [AS] [EC 4.1.3.27] is a branch-point enzyme in the pathway for biosynthesis of aromatic amino acids that catalyzes the amination of chorismate to anthranilate. AS from various organisms has been purified and its catalytic mechanism and feedback regulation has been worked out [Caligiuri and Bauerle, 1991; Graf *et al.*, 1993].

The AS holoenzyme usually consists of two non-identical proteins and can use either glutamine or ammonia as the source of the amino group. The individual subunits are commonly referred to as Anthranilate synthase component I [α subunit] and Anthranilate Synthase Component II [β subunit]. Both the components are required for the Gln-dependent formation of anthranilate. Component II catalyzes the transfer of the amine group from Gln to component I, which in turn catalyzes the amination of chorismate [Ito and Yanofsky, 1966, 1969]. Component I alone can catalyze the conversion of chorismate to anthranilate only in the presence of very high concentrations of ammonium ions. Ammonium as well as Gln-dependent anthranilate synthase reactions require magnesium as a co-factor. The respective reactions are shown below.



In some organisms, Anthranilate synthase [TrpE] enzyme is part of a multifunctional protein, while in others only component I [TrpE] is present.

The X-ray crystal structure of the cooperative anthranilate synthase heterotetramer from *Salmonella typhimurium* has been determined at 1.9 Å resolution with the allosteric inhibitor L-tryptophan bound to the regulatory site in the TrpE subunit [Morollo *et al.*, 2001]. This structure was found to be different from the unliganded, noncooperative anthranilate synthase heterotetramer from *Sulfolobus solfataricus* [Knochel *et al.*, 1999]. These authors observed that even though the functional dimer pairs are structurally similar, they have completely different quarternary structures. Tryptophan binding to the TrpE subunit of *Salmonella typhimurium* was found to order a loop that in turn stabilizes the inactive T state of the enzyme by restricting closure of the active site cleft. This observation was consistent with differences in the cooperative behavior of the enzymes. The structural model is important for the study of the structural differences between cooperative and noncooperative anthranilate synthase homologues.

5.2.2 Isochorismate synthase [ICS]

ICS is also a chorismate utilizing enzyme that catalyzes the conversion of chorismate to isochorismate. Isochorismate is the precursor for salicylate. The enzyme is also required for enterochelin and menaquinone [Vitamin K] biosynthesis in *E. coli* [Liu *et al.*, 1990; Kaiser *et al.*, 1990; Dahm *et al.*, 1998].

Salicylate is the precursor of the siderophore pyochelin in *Pseudomonas aeruginosa*, and salicylate as well as pyochelin display siderophore activity [Ankenbauer and Quan, 1994; Serino *et al.*, 1997]. Similarly, salicylate is a precursor of the mycobacterial siderophore, mycobactin [Quadri *et al.*, 1998]. Salicylate biosynthesis in *P. aeruginosa* has been well studied after the characterization of the genes *pch*A [isochorismate synthase] and *pch*B [isochorismate pyruvate lyase], the two consecutive genes required for the conversion of chorismate to salicylate [Serino *et al.*, 1995; Gaille *et al.*, 2003]. While Rv2386c of *M. tuberculosis* appears to be a *pch*A equivalent, a gene homologous to *pch*B has not been reported in *M. tuberculosis* that can catalyze the conversion of isochorismate to salicylate. The *pch*A gene of P*. aeruginosa* encodes a protein of 52 kDa with extensive similarity to the chorismate-utilizing enzymes isochorismate synthase, anthranilate synthase [component I] and p-aminobenzoate synthase [component I]. Conversely, the 11kDa protein encoded by *pch*B does not show significant similarity with other proteins. The *pch*B stop codon overlaps the presumed *pch*A start codon [Serino *et al.*, 1995; Serino *et al.*, 1997]. Expression of the *pch*A gene in *P. aeruginosa* appears to depend on the transcription and translation of the upstream *pch*B gene. The *pch*BA genes of *P. aeruginosa* were in fact the first salicylate biosynthesis genes to be reported [Gaille *et al.*, 2002, 2003]. When *pch*BA genes were overexpressed in an *Escherichia coli ent*C mutant [lacking isochorismate synthase activity], salicylate formation was observed [Gaille *et al.*, 2003]. However, when *pch*B was expressed alone, there was no evidence of salicylate biosynthesis. In contrast, an *ent*B mutant of *E. coli* blocked in the conversion of isochorismate to 2,3-dihydroxybenzoate formed salicylate when transformed with a *pch*B expression construct. In the same study, salicylate formation was demonstrated *in vitro* when

chorismate was incubated with a crude extract of *P. aeruginosa* containing overproduced PchA and PchB proteins. Salicylate and pyruvate were formed in equimolar amounts. Furthermore, salicylate-forming activity could be detected in extracts of a *P. aeruginosa* siderophore-negative mutant grown under iron limitation, but not with iron excess. These results were able to establish a pathway leading from chorismate to isochorismate and then to salicylate catalyzed.

The structure of *Sulfolobus* anthranilate synthase has been solved and has been referred to as a prototype for the structure for other chorismate binding enzymes including isochorismate synthase.

5.2.3 The tryptophan operon of *M. tuberculosis*

While the genes involved in tryptophan biosynthesis are organized as an operon in *E. coli*, a scattered distribution is observed in the case of *M. tuberculosis* [Figure 5.1]. However, studies done so far suggest that *M. tuberculosis* possesses a homologue of all the genes that participate in tryptophan biosynthesis, suggesting the essentiality of the pathway in survival of the bacterium inside the host.

While *trp*E, B, C, A are clustered in the *M. tuberculosis* genome, *trp*G, *trp*D and *trp*F are quite distantly located. In fact, the occurrence of *trp*F was not known in *M. tuberculosis* genome until a recent report by Barona-Gomez and Hodgson [2003], which demonstrated that *his*A [Rv1603] of *M. tuberculosis*, a part of the histidine biosynthesis operon of *M. tuberculosis*, is a bifunctional protein and can complement a *trp*F mutant of *E. coli*. However, the reverse did not hold true i.e. the *E coli his*A could satisfy only histidine auxotrophy and *trp*F could satisfy only tryptophan auxotrophy in the respective *S. coelicolor* mutant strain. *M. tuberculosis his*A is therefore unique in the sense that it is a bifunctional protein with a broad substrate specificity.

In the absence of a glutamine amidotransferase [TrpG] enzyme in *M. tuberculosis trp* operon, *M. tuberculosis* TrpE is predicted to be an ammonia dependent anthranilate synthase. However, the pH requirement of this reaction is very high and far surpasses

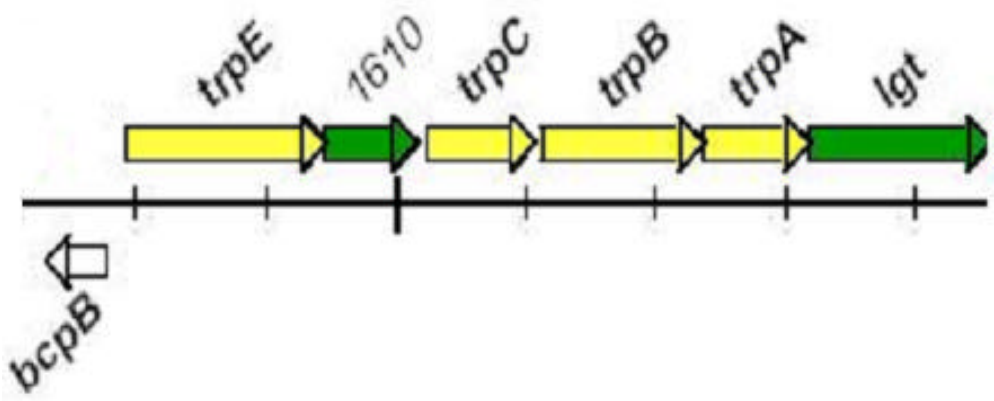


Figure 5.1: Organization of the tryptophan operon of *M. tuberculosis*. Members of this operon have been reported to be essential genes of the bacterium [*Sasseti et al, 2002*] *trp*D and *trp*G are not a part of this operon and are located elsewhere in the genome.

the physiological pH requirements. A distant location of TrpG emphasizes the need to explore the mechanism of co-regulation of these genes.

5.2.4 The *mbt* operon of *M. tuberculosis*

Iron assimilation is an important virulence attribute of *M. tuberculosis* and hence siderophore production is assumed to be associated with the virulence of the bacterium. Thus, it was obvious that the identification of genes involved in the iron sequestration pathway would lead to the identification of certain genes essential for survival of the bacterium inside the hostile environment of the host. The completion of the genome sequence of *Mycobacterium tuberculosis* has led to the identification of a gene cluster with homology to proteins of the non ribosomal peptide synthases and polyketide synthases. These genes were annotated as those possibly involved in siderophore biosynthesis [Cole *et al.*, 1998; Quadri *et al.*,1998]. This operon consists of ten genes [*mbt*A to *mbt*J], each of which have been assigned various functions. The individual roles of all these genes in the biosynthesis of siderophores has been briefly reviewed earlier [Crosa and Walsh, 2002; De Voss *et al.* 1999]. Essentially, there are seven genes [*mbt*A to *mbt*G] that appear to encode the desired modules required for the biosynthesis of the core of mycobactins. MbtA [Rv2384] activates salicylate as its acyl adenylate. The acyl adenylated salicylate is then covalently attached to a phosphopantetheine prosthetic group of MbtB. MbtB [Rv2383c] is an NRPS [non ribosomal peptide synthase] that is believed to activate serine and condense it with salicylate moiety and cyclize this product to a hydroxyphenyloxazoline. There are two other NRPSs encoded by *mbt*E [Rv2380c] and *mbt*F[Rv2379c] that have the appropriate activation, condensation and peptide carrier domains for donation of the two lysine-derived moieties of mycobactin. MbtF has a terminal domain that either functions as an epimerisation domain or a thioesterase responsible for releasing the mycobactin from the enzyme by lactamization of the terminal hydroxy lysine residue. Apart from these enzymes, the *mbt* gene cluster contains *mbt*C and *mbt*D genes, which encode proteins homologous to polyketide synthases. There are other genes in the operon [*mbt*H, *mbt*J] for which exact functions have not yet been determined.

Of all these proposed genes, only *mbt*A and *mbt*B have been experimentally shown to perform the function of acyl adenylation of salicylate and to encode the phosphopantetheinyl carrier domain, respectively [Quadri *et al.*, 1998]. Other genes require expression and characterization to solve the entire pathway of siderophore biosynthesis in *M. tuberculosis*.

Interestingly, the *mbt* operon is absent in *M. leprae*, which is a major setback to the pathogenicity of the bacterium [Cole *et al.*, 2001]. The importance of the salicylate derived siderophores in determining the pathogenicity of the bacterium was determined by generating *mbt*B knockouts of *M. tuberculosis* [De Voss *et al.* 2000]. The *mbt*B gene was replaced by a hygromycin resistant cassette through homologous recombination and the resulting bacterium was tested for production of siderophores. The assays showed that the delta *mbt*B:hyg strain of *M. tuberculosis* did not produce siderophores. The mutant bacterium was also tested for growth in macrophage like THP-1 cells under different concentrations of iron. It was observed that there was a rational decline in the growth of the mutant bacterium under iron limiting conditions. This laid the foundation for the belief that under iron limiting conditions, siderophores are essential for the survival of the bacterium.

Since salicylate derived siderophores have been shown to be required for the virulence of *M. tuberculosis*, the pathway for salicylate biosynthesis itself is vital for the bacterium.. Cole *et al.* [1998] have listed *ent*C and *ent*D genes as being possibly involved in the production of salicylate. These genes have similarity to enzymes involved in enterochelin biosynthesis pathway of *Pseudomonas aeruginosa*. However, Quadri *et al.* [1998] speculate that *mbt*I, which has similarity to anthranilate synthase, is responsible for the production of salicylate. Moreover, *ent*C and *ent*D do not contain any IdeR [Iron-dependent Regulator] binding site in their upstream sequences. Hence it is likely that MbtI could be involved indirectly in salicylate biosynthesis. In one study where *M. smegmatis* auxotroph for salicylic acid was assessed [Adilakshmi *et al.*, 2000] it was found that salicylate may even function as a signal molecule for recognition of cellular iron stress.

This Chapter gives a brief account of the studies carried on *trp*E and *trp*E2/*mbt*I of *Mycobacterium tuberculosis*. A biochemical as well as a genetic approach was employed to elucidate the function of the two. An overlap of function was observed which suggests convergent evolution of AS and ICS enzyme activities. The implications of the same are discussed.

5.3 EXPERIMENTAL PROCEDURES

5.3.1 Bacterial strains and plasmids

Table 5.1 gives a brief description of all *E. coli* and *M. tuberculosis* strains used in this study. The recombinant plasmids constructed along with their sources are also detailed.

5.3.2 Media, chemicals, buffers, and Enzymes

Luria Broth growth media supplemented with 200 μg/ml of ampicillin was used for selection and amplification of plasmids in *E. coli* hosts.

The composition of minimal media used for screening *E. coli* amino acid auxotrophs is described as follows

Minimal A medium:

K_2HPO_4	-	10.5 gm
KH_2PO_4	-	4.5 gm
$[NH4]_2SO_4$	-	1.0 gm
Sodium Citrate $2H_2O$	-	0.5 gm

Make up to 1 litre with water

After autoclaving, 1 ml of 1 M $MgSO_4$ and 0.1 ml of 1% Thiamine were added. Glucose at 0.2% was added separately. Amino acids, when required, were added to a final concentration of 40 μg/ml. Minimal A agar contained 2% bactoagar in minimal A medium, made from a separately autoclaved 4% bactoagar stock in water.

Unless otherwise specified, all chemicals and solvents, of reagent or HPLC grade, were procured from Sigma Chemicals, USA.

5.3.3 Cloning, over expression and purification of recombinant proteins

The ORFs Rv1609 and Rv2386c were PCR amplified from *M. tuberculosis* H37Rv genomic DNA using specific primers to enable directional cloning in pET23a expression vector. A brief description of the primers used is given in Table 5.2. The resultant constructs were used to transform *E. coli* BL21 DE3 competent cells and expression and purification of the recombinant protein was carried out by a method recently described [Ghosh *et al*, 2004]. The method employs the use of the dipeptide glycylglycine to enhance the solubility of a recombinant protein. In this method, the dipeptide is added to the culture media and all subsequent procedures are as described in chapter 2, section 2.3.2.

5.3.4 Anthranilate synthase activity assay [Rv1609 and Rv2386c]

The ammonia dependent AS activity of ORFs, Rv1609 and Rv2386c was studied using fluorimetric assays for anthranilate formation [Bohlmann J, 1996]. The reaction mixture contained 12.5mM Tris, 1.25% [v/v] glycerol, 0.25mM DTT, 10mM MgCl2, 1mM chorismate, 100mM NH4Cl, pH8.5. Reaction was initiated by addition of recombinant enzyme [rRv1609/ rRv2386c] to the assay mix followed by incubation at 32^0C for 30 minutes. The reaction was terminated with 10µl of 5N H_3PO_4. Anthranilate was extracted into 3ml of ethyl acetate and estimated using a fluorimeter at an excitation wavelength of 338nm and an emission wavelength of 440nm.

Table 5.1: Bacterial Strains and Plasmids

Strains/plasmids	Description	Source / Reference
M. tuberculosis		Cole *et al*., 1998
H37Rv	Laboratory Strain [Virulent]	
E. coli		
DH5α		
BL21 DE3		
E. coli BL21 *trp- tet*	A tryptophan auxotrophic derivative of *E. coli* BL21 strain	This work
E. coli BL21 *trp- rho- kan tet*	An anthranilate auxotrophic derivative of *E. coli* BL21 strain	This work
Plasmids		
pET23a	Expression vector	Novagen
pET23a1609	pET23a derivative with *M. tuberculosis trp*E [Rv1609c] cloned in *Nde*I/*Xho*I sites	This work
pET23a2386	pET23a derivative with *M. tuberculosis trp*E2/*mbt*I [Rv2386c] cloned in *Nde*I/*Xho*I sites	This work

Table 5.2: Sequences of primers used to amplify *M. tuberculosis trp*E [Rv1609] and *trp*E2 [Rv2386c]

ORF/ Gene	FORWARD PRIMER	REVERSE PRIMER
Rv1609	***Nde*I**	***Xho*I**
	AT***CATATG*** GTGCACGCCGACCTCG	AT***CTCGAG***GCAGCCACTGCGGTTC
Rv2386c	***Nde*I**	***Xho*I**
	AT***CATATG*** GTGTCCGAGCTCAGCGT	AT***CTCGAG***CTGGCGTGCAACCAG

5.3.5 Isochorismate synthase activity assay [Rv2386c]

ICS [EC 5.4.99.6] activity was determined according to the method of Poulsen *et al.* [1991] with slight modifications. The incubation mixture [250 µl] contained 0.1 M Tris-HCl, pH 7.5, 2 mM chorismate, 10 mM $MgCl_2$, and 100 pico moles of pure recombinant enzyme. The reaction mix was incubated for 60 min at 30°C followed by termination with 62.5 µl of methanol:sec-butanol [1:1, v/v]. The samples were centrifuged and loaded on a Nucleosil C-18 column. Waters 717 plus HPLC system equipped with a photodiode array detector was used to monitor isochorismate formation. Assay for direct salicylate biosynthesis was carried out using fluorimetric procedures.

5.3.6 Analytical size exclusion chromatography

The oligomeric states of the recombinant proteins were determined using size exclusion chromatographic procedures. The procedure is briefed in section 3.3.10 of Chapter 3

5.3.7 Limited proteolysis

Limited proteolysis was carried out to ascertain the existence of ligand binding pockets on *M. tuberculosis* TrpE and TrpE2. Anthranilate, salicylate and tryptophan were taken as the ligands of choice. The procedure for limited proteolysis is described in Chapter 3, section 3.3.7. The gels were viewed by silver staining.

5.3.8 Silver staining of protein gels

Silver staining was carried out as per method described in section 3.3.8 of chapter 3**.**

5.3.9 Construction of an *E. coli* BL21 *trp*E mutant strain

E. coli BL 21 was taken as the parent strain for introduction of an amber mutation in *trp*E. This strain was transduced with a P1 phage carrying *trp*E amber mutation linked to a tetracycline resistant marker [70% linkage]. The resultant strain was a tryptophan auxotroph [designated *E. coli* BL21 *trp- tet*]. A second round of transduction was carried

out in this strain with another P1 phage carrying a mutation in *rho*, linked to kanamycin resistance marker [60% linkage]. The second transduction was carried out to relieve polarity introduced by an amber mutation in *trp*E and expression of the downstream genes [Harinarayan and Gowrishankar, 2003]. The resultant strain was an anthranilate auxotroph [designated *E. coli* BL21 *trp- rho-tet kan*].

5.3.9.1 Preparation of P1 phage lysates/ P1 transduction

These procedures are detailed in section 3.3.11.1 and 3.3.11.2 of chapter 3.

5.3.9.2 Screening for auxotrophic mutants of *E. coli*

The minimal media were supplemented with appropriate amino acids and compounds [tryptophan, anthranilate] as per the requirements of the experiment.

5.4 RESULTS

5.4.1 *M. tuberculosis* TrpE [Rv1609] as well as TrpE2/MbtI [Rv2386c] show *in vitro* ammonia-dependent anthranilate synthase activity

*trp*E as well as *trp*E2 of *Mycobacterium tuberculosis* expressed in *E. coli* were purified as recombinant proteins [Figure 5.2 A and B]. The ammonia dependent anthranilate synthase activity of recombinant proteins were assessed using fluorimetric assays. It was observed that *M. tuberculosis* TrpE as well as TrpE2 display an ammonia dependent anthranilate synthase activity. However, the AS activity of TrpE2 was found to be 4 fold lower than that of TrpE [Figure 5.3]. The glutamine dependent AS activity of *M. tuberculosis* TrpE/TrpE2 was also studied. However, no evidence for the same was observed.

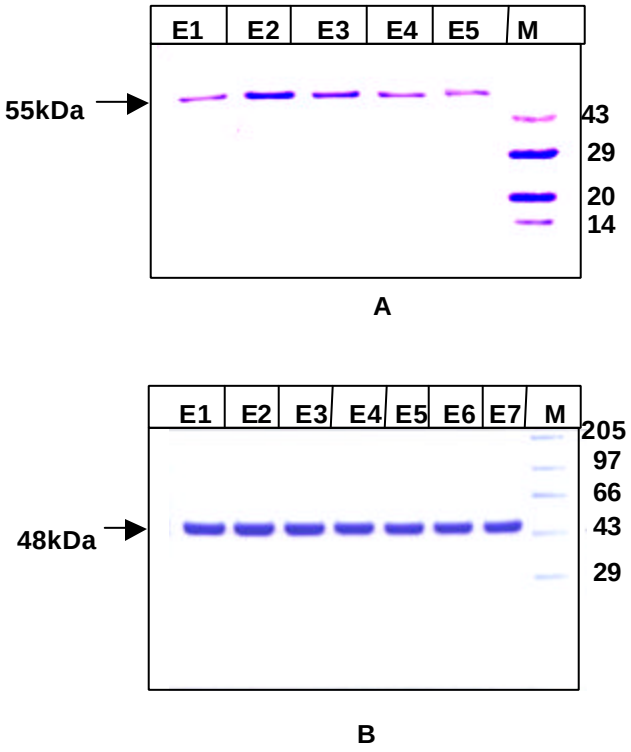


Figure 5.2: Purification of recombinant proteins corresponding to the ORFs Rv1609 (TrpE) (A) and Rv2386c (MbtI/TrpE2) (B) of *Mycobacterium tuberculosis*. Rv1609/Rv2386 were cloned in the *Nde*I and *Xho*I sites of pET23a vector with a C-terminal Histidine tag and expressed in *E. coli* BL21 cells. Affinity purification of recombinant protein was carried out using Talon resin (Clonetech, USA). The purified protein resolved on a 10% Tris-Tricine gel is shown. M represents the protein molecular size marker (Genei, India) and the E series represent the successive eluted fractions of the recombinant protein. Arrowheads indicate the position of the purified proteins.

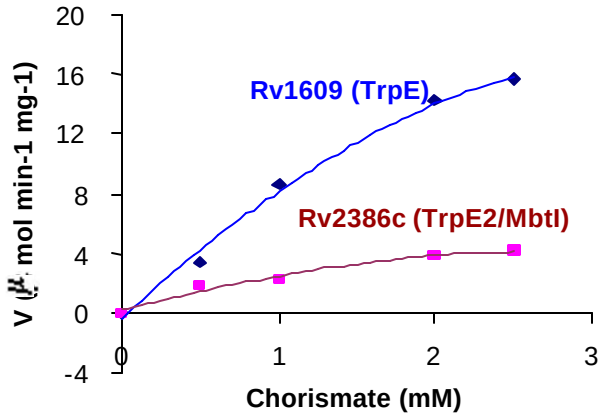


Figure 5.3: *In vitro* Anthranilate synthase (AS) activity of *M. tuberculosis* TrpE2/MbtI is 4-fold lower than that of TrpE. Ammonia dependent anthranilate synthase activity of *M. tuberculosis* TrpE (Rv1609) and TrpE2/MbtI (Rv2386c) were determined using fluorimetric assays. The above figure shows the substrate saturation plot of the two enzymes indicating that TrpE as well as TrpE2/MbtI possess AS activity. However, the AS activity of TrpE2/MbtI is ~4 fold lower than TrpE.

5.4.2 *M. tuberculosis* TrpE and TrpE2 are regulated proteins

Limited proteolysis was carried out to determine whether *M. tuberculosis* TrpE/TrpE2 are regulated by their end products. Tryptophan and anthranilate were used as the ligands of choice for TrpE and tryptophan and salicylate were used as ligands for TrpE2/MbtI [considering the possible involvement of TrpE2 in salicylate biosynthesis]. Tryptophan was found to protect TrpE [Figure 5.4B] as well as TrpE2 from trypsin cleavage [Figure 5.4 C]. Anthranilic acid also protected *M. tuberculosis* TrpE [Figure 5.4A] from proteolytic cleavage but salicylate could not do the same for *M. tuberculosis* TrpE2/ MbtI [Figure 5.4 D]. Phenylalanine binding to TrpE/TrpE2 were studied using fluorimetry. As there was no change in the fluorescence spectra of the proteins, it was concluded that these enzymes do not have phenylalanine binding pockets and are completely insensitive to the influence of cross pathway effector molecules [Figure 5.5: A, B].

5.4.3 *M. tuberculosis* ORFs Rv1609 [*trp*E] as well as Rv2386c [*mbt*I/*trp*E2] can complement an *E. coli trp*E mutant strain

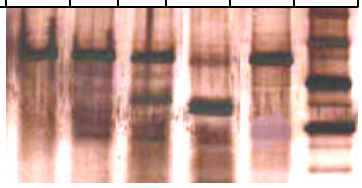
Introduction of an amber mutation in *trp*E introduces polarity in the *trp* operon. Hence, as described in section 5.3.9, mutation in *rho* was introduced to relieve polarity. This strain was used as host for complementation tests with pET23a chimeric constructs, where *M. tuberculosis trp*E and *trp*E2 expression were driven from T7 promoter. It was observed that upon IPTG induction, both the constructs could enable *E. coli trp*E mutant strain to grow on minimal media [Figure 5.6]. This genetic complementation evidence demonstrated that Rv2386c has AS activity as well.

5.4.4 MbtI/TrpE2 [Rv2386c] shows *in vitro* isochorismate synthase activity

Taking into account the unique location of *M. tuberculosis* t*rp*E2/*mbt*I as part of the siderophore biosynthesis operon, and considering earlier reports by Quadri *et al.* [1998], it was decided to determine whether *M. tuberculosis* TrpE2 also shows isochorismate synthase activity. If indeed so, this would define the first step towards the biosynthesis of

Rv1609 (TrpE)

rRv1609	+	+	+	+	+	M
Trypsin	+	+	+	+	-	
Effector (mM)	5	3	2	-	-	



Anthranilic acid

A

rRv1609	M	+	+	+	+	+
Trypsin		+	+	+	+	+
Effector (mM)		-	-	1	3	5

Tryptophan

B

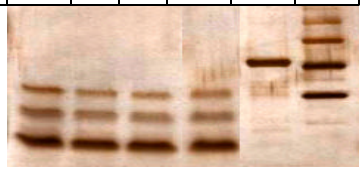
Rv2386c (TrpE2/MbtI)

rRv2386	M	+	+	+	+	+	+	+
Trypsin		-	+	+	+	+	+	+
Effector (mM)		-	-	0.5	1	1.5	3	5

Tryptophan

C

rRv2386	+	+	+	+	+	M
Trypsin	+	+	+	+	-	
Effector (mM)	5	3	1	-	-	



Salicylic acid

D

Figure 5.4: *M. tuberculosis* TrpE and TrpE2/MbtI are regulated at the protein level by effectors. The pathway specific effectors, Anthranilic acid as well as Tryptophan protect *M. tuberculosis* Anthranilate Synthase (TrpE/Rv1609) from proteolysis (A,B). *Mtb* TrpE2/MbtI is protected from proteolytic cleavage by tryptophan (C). Salicylic acid does not protect the protein from tryptic cleavage (D). M represents protein molecular size marker.

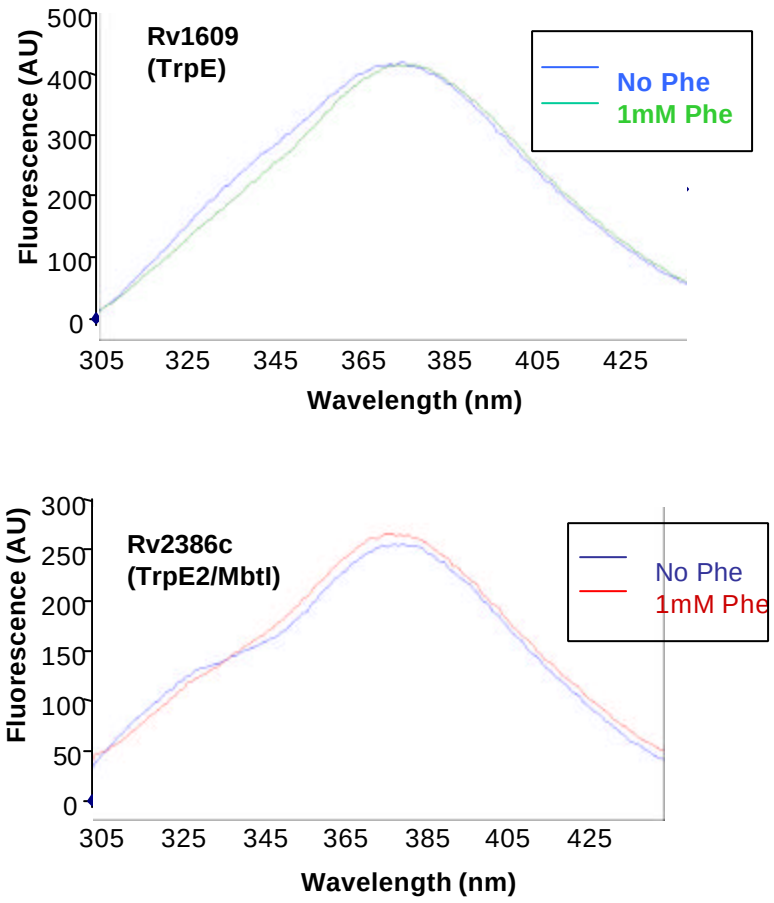
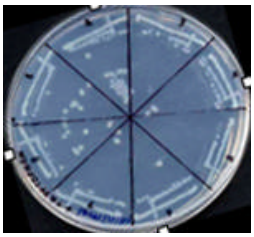


Figure 5.5: The cross pathway regulator phenylalanine does not change the fluorescence spectrum of *Mtb* TrpE (A) and TrpE2/MbtI (B). Phenylalanine does not have any modulatory effect on the activity of the enzyme.

Rv1609 **Rv2386**
***trp*E** ***trp*E2/*mbt*I**



Minimal medium

Figure 5.6: *M. tuberculosis trp*E (Rv1609) and *trp*E2/*mbt*I (Rv2386c) can complement an *E. coli* BL21 *trp*E mutant strain:

E. coli BL21 *trp*E mutant strain was transformed with pET23a constructs carrying *M. tuberculosis* genes Rv1609c (*trp*E) and Rv2386c (*trp*E2/*mbt*I) and its ability to grow on minimal mediium containing IPTG was tested. As evident from the figure above, Rv1609 (*trp*E) as well as Rv2386c (*trp*E2) could allow the *trp*E mutant *E. coli* BL21 strain (*E. coli trp- rho- kan tet)* to grow on minimal media.

salicylate, the precursor for salicylate. For this purpose, an HPLC assay was carried out to determine whether chorismate could be converted to isochorismate in the presence of rRv2386c. Results based on the retention time of isochorismate suggest that *M. tuberculosis* TrpE2 can convert chorismate to isochorismate [Figure 5.7]. The oligomeric state of *M. tuberculosis* TrpE2 was also determined using size exclusion chromatography. The native protein was found to be a dimer [Figure 5.8]. This observation was similar to other ASs which are also reported to be dimeric proteins. However, this is unlike the equivalent ICS enzyme PchA of *Pseudomonas aeruginosa* that exists as a monomer.

5.4.5 Coupled assays indicate that Rv2386c and Rv1885c are involved in the conversion of chorismate to salicylate

Earlier reports have documented the ability of isochorismate pyruvate lyase [IPL] enzyme of *P. aeruginosa* to catalyze a chorismate mutase [CM] reaction as well. It was accordingly decided to check if the CM of *M. tuberculosis* could execute an IPL reaction. Hence, both the enzymes were used in a coupled assay to monitor salicylate formation. While salicylate could not be directly formed by TrpE2/MbtI, detectable levels of salicylate was formed in a coupled assay involving recombinant TrpE2/MbtI and CM proteins [Figure 5.9]

5.4.6 TrpE2/MbtI [Rv2386c] does not show any evidence for direct salicylate biosynthesis

Recombinant TrpE2 was used in a fluorimetry based biochemical assay to assess if the enzyme could directly convert chorismate to salicylate. This experiment was done taking into account the absence of an isochorismate pyruvate lyase enzyme in *M. tuberculosis* genome, which is required for conversion of isochorismate to salicylate. It was observed that there is no evidence for direct salicylate biosynthesis by TrpE2 alone [Figure 5.9].

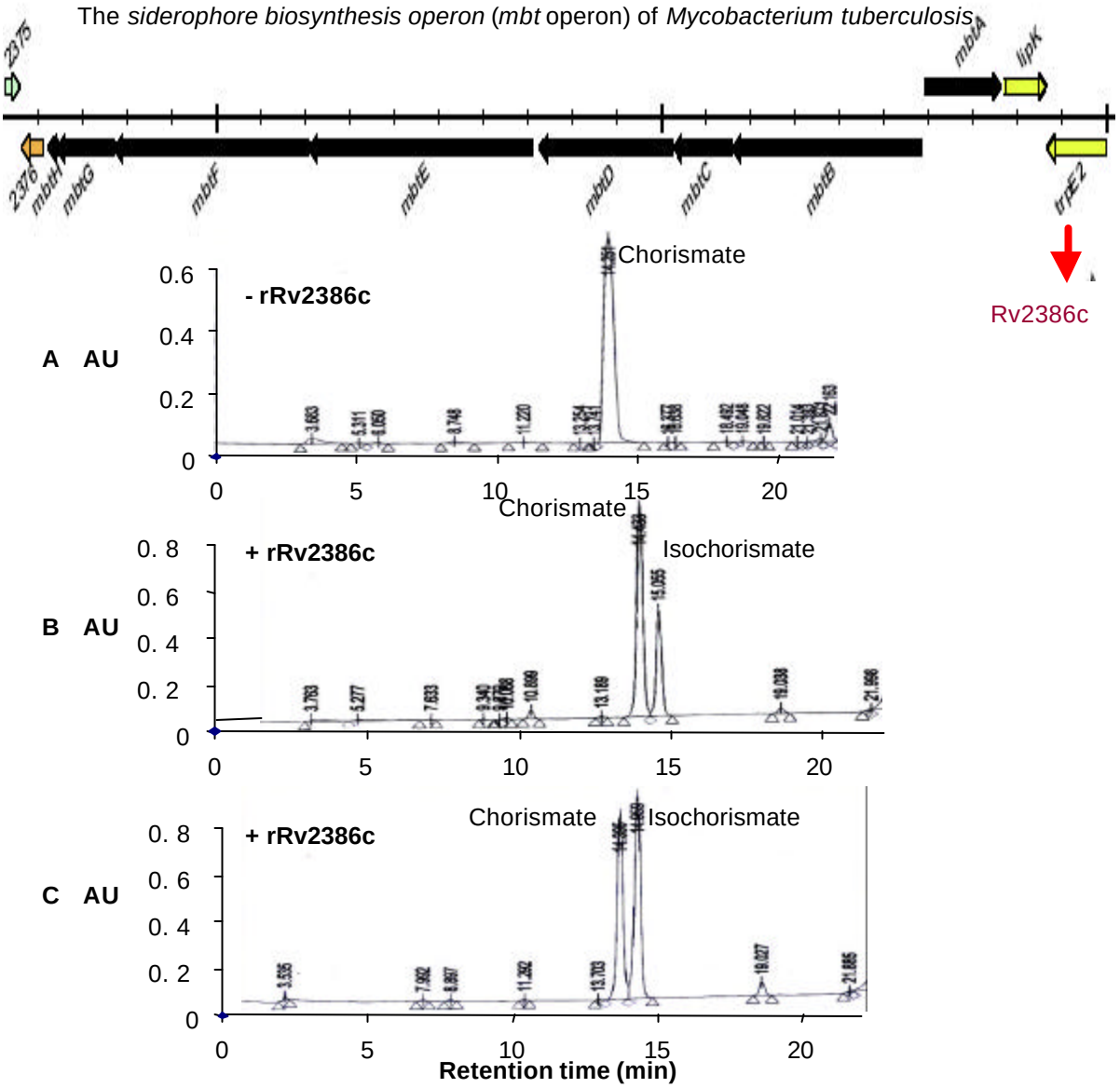


Figure 5.7: Recombinant Rv2386c (MbtI/TrpE2) can catalyze the conversion of chorismate to isochorismate. The reaction mixture containing 0.1M Tris-HCl, pH7.5, 2mM chorismate, 10mM $MgCl_2$ and rRv2386c was incubated at 30^0C for 60 minutes, terminated with the addition of 62.5µl pf MeOH: sec-BuOH (1:1 v/v), centrifuged and analysed by HPLC. As evident from the above figure, in the absence of rRv2386c, a single peak of chorismate was observed (A) and in the presence of rRv2386c, two peaks corresponding to chorismate as well as isochorismate could be seen (B and C, with increasing concentrations of chorismate). These results demonstrate the isochorismate synthase activity of *M. tuberculosis* MbtI/TrpE2

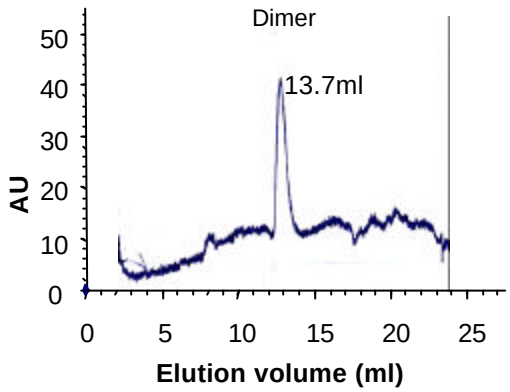


Figure 5.8: *M. tuberculosis* TrpE2/MbtI (Rv2386c) is a dimeric protein (two subunits of individual MW of 50kDa). rRv2386c was loaded on a Superdex 200 column (Pharmacia Biotech). Absorbance at 280nm (AU) was plotted as a function of elution volume. The elution parameter Kav vs log MW plot of standard proteins was used to determine the actual MW of TrpE2/MbtI .

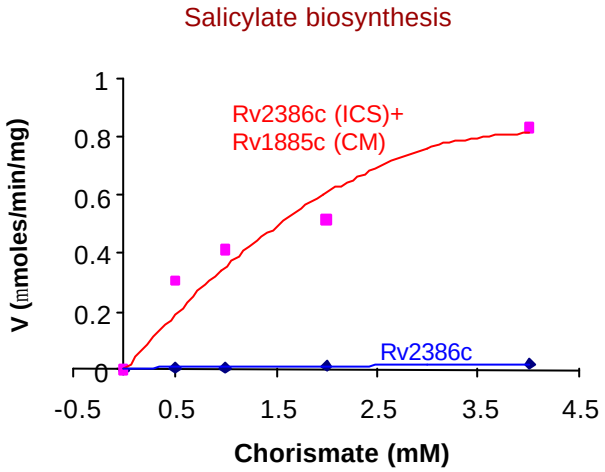


Figure 5.9: *M. tuberculosis* Chorismate mutase (CM) can substitute for an isochorismate pyruvate lyase (IPL) enzyme. Fluorimetric estimation of salicylate formation was carried out with *M. tuberculosis* MbtI/TrpE2 alone and in MbtI/TrpE2 and CM coupled reactions. While salicylate could not be formed by MbtI/TrpE2 (rRv2386c) alone, detectable amounts of salicylate were formed in a coupled assay involving both MbtI/TrpE2 and CM.

5.5 DISCUSSION

It has earlier been suggested that Anthranilate synthase [AS] and Isochorismate synthase [ICS] share a common reaction mechanism [He, 2004]. Extensive sequence similarity between AS and ICS enzymes of *M. tuberculosis* also suggest a commonality in function. *M. tuberculosis* ORFs 1609 and Rv2386c share about 40% aa sequence similarity amongst themselves as well as with the *E. coli* TrpE. Genetic and biochemical studies carried out on *M. tuberculosis* ORFs, Rv1609 and Rv2386 point to a convergent evolution of AS and ICS enzyme activities. A high sequence similarity of the two groups of enzymes is perhaps responsible for an overlap of function.

A glance at the biochemical profiles of TrpE and TrpE2 reveals that the AS activity of TrpE2 is 4 fold lower than TrpE. However, genetic evidence suggests that in spite of a weaker AS activity, even *trp*E2 could complement an *E. coli trp*E mutant strain. This observation could perhaps be attributed to the use of a strong promoter [T7], which in this case was sufficient for complementation purposes.

Chorismate can be converted to salicylate by the sequential action of the enzymes isochorismate synthase [ICS] and isochorismate pyruvate lyase [IPL] [Young *et al.*, 1969]. *M. tuberculosis* genome does not have an annotated IPL enzyme, and hence it has been suggested that Rv2386c [an ICS] could directly convert chorismate to salicylate [Gaille *et al.*, 2003]. However, results presented in this chapter indicate lack of evidence for direct salicylate biosynthesis by rRv2386c. It has been reported earlier that the isochorismate pyruvate lyase and chorismate mutase enzymes of *P. aeruginosa* are structurally and even functionally similar [Gaille *et al.*, 2002, 2003] as they possess a common chorismate binding domain. As *P. aruginosa* ICS also shows CM activity [Gaille C, 2002], the same analogy was used in *M. tuberculosis* to observe that chorismate mutase of *M. tuberculosis* could be a partner in salicylate biosynthesis. However, in *in vitro* coupled assays more than 2-3 μM salicylate could never be produced.

Considering the weak activity of *M. tuberculosis* CM as an IPL, it is still unclear whether *M. tuberculosis* CM participates in the conversion of chorismate to salicylate. The weak

IPL activity could be a mere relic of the broad substrate specificity of *M. tuberculosis* CM. However, in the absence of any *in vitro* evidence of the direct conversion of chorismate to salicylate by Rv2386c, it is still open for discussion whether *M. tuberculosis* depends upon the spontaneous conversion of isochorismate to salicylate or it depends on the weak IPL activity of CM [Young *et al.*,1969; Marshall and Ratledge, 1971, 1972]. As Rv2386 [mbtI/*trpE2*] even complements an *E. coli trp*E mutant strain, it also seems likely that *mbt*I could generate anthranilate, an amino analogue of salicylate which could get converted to salicylate by deamination.

In summary, the results presented in this chapter suggest that TrpE and TrpE2/MbtI *of M. tuberculosis* are interchangeable for their anthranilate synthase activity. Additionally TrpE2 can also execute the function of an isochorismate synthase. The evolutionary insight that could be drawn from the present findings is that despite the high sequence similarity at the amino acid level, these two chorismate utilizing enzymes can perform different functions. These results are consistent with the hypothesis that most ancestral enzymes have a broad substrate affinity which gradually get refined during the course of evolution to form enzymes with higher specific activities [Barona-Gomez and Hodgson, 2003, Xie *et al.*, 2003]. Additionally, the absence of a human homologue makes this protein an attractive drug target for the design of novel compounds against TB.

CHAPTER 6

Mycobacterium tuberculosis PPE/MPTR ORF, Rv2608 SHOWS A DIFFERENTIAL B CELL RESPONSE AND A LOW T CELL RESPONSE IN VARIOUS CATEGORIES OF TB PATIENTS

6.1 ABSTRACT

About 10% of the coding capacity of the *Mycobacterium tuberculosis* genome is contributed by two large, unrelated and functionally uncharacterized families of acidic, glycine rich proteins, the PE and the PPE families. The existence of polymorphic repetitive sequences, PGRS [Polymorphic GC Rich Sequences] and MPTRs [Major Polymorphic Tandem Repeats] in the C-terminal region of most of these genes suggest that these genes could be important in imparting antigenic variability to the bacterium. Earlier report from the present lab described a PPE family member Rv2430c as an immunodominant antigen [Choudhary *et al.*, 2003; 2004]. The present work describes the evaluation of a representative PPE_MPTR gene, Rv2608 for its ability to elicit humoral as well as well as T cell responses. Rv2608 was also found to be polymorphic in different clinical isolates as determined by PCR-RFLP analysis. 51 clinically confirmed TB patients belonging to fresh infection [n=22], relapsed infection [n=21], extrapulmonary infection [n=8], and 10 healthy controls were included in the study. The recombinant Rv2608 protein showed reactivity with sera of different category TB patients. Synthetic peptides corresponding to predicted regions of high antigenicity from the MPTR region of Rv2608 protein were used to assess cellular versus humoral immune response elicited by the protein. Interestingly, while the full length Rv2608 failed to show a differential humoral response as a function of patient infection category, the synthetic peptides elicited a predominantly humoral response in category II patients [relapsed TB]. The T cell response to the peptides as determined by T cell proliferation assays was low in all patient categories. The predominance of antibodies to this protein in relapsed infection cases and a relatively low cellular immune response indicate that this PPE member could be playing an important role in evading host immune responses by directing it towards a more humoral type that is not important to contain the infection. Additionally, for the first time the role of Gly-X-Gly-Asn-X-Gly repeat motifs present in a PPE family protein in eliciting a humoral immune response has been demonstrated.

6.2 INTRODUCTION

In the quest for identification of other attributes that would enable *M. tuberculosis* to establish a successful infection, the study was continued with the identification of novel antigens that would help the bacterium to evade host immune responses. In this context, it is significant to note that the only source of variation in the otherwise conserved genome of *M. tuberculosis* resides in the PE and the PPE gene families [Cole *et al.*, 1998]. The existence of PE/PPE gene families was evident even before the *Mycobacterium tuberculosis* genome was sequenced with occasional reports of occurrence of glycine and alanine rich multiple repetitive sequences in the genome [Poulet and Cole, 1995] or the identification of a few fibronectin binding proteins [Abou-Zeid *et al.*, 1991]. Sequencing categorized the PE/PPE gene families as two large unrelated families of highly acidic glycine rich proteins that constitute about 10% of the coding capacity of the genome [Cole *et al.*, 1998]. Comparative genome sequencing in various mycobacterial species revealed that by and large PE and PPE gene families are unique to *Mycobacterium tuberculosis* with few homologues in *M. leprae, M. bovis, M marinum etc* [Cole, 2002]. Amongst the *M. leprae* homologues, a major serine rich antigen is expressed in leprosy patients [Vega-Lopez *et al.*, 1993].

It is generally believed that the PE and PPE genes could be a source of antigenic variability. A recombinant PE_PGRS [Rv1759c] protein was shown to possess fibronectin binding properties and was also recognized by patient sera [Espitia *et al.*, 1999]. The same group also reported immense intra-strain variability in the PGRS domain with the N-terminal region staying constant. Transposon insertion in the PE_PGRS gene [Rv1818c] was shown to reduce macrophage infection ability of *M. tuberculosis* [Brennan *et al.*, 2001]. Surface localization of a PPE protein [Rv1917c] and many other PE_PGRS proteins has been reported [Sampson *et al.*, 2001; Banu *et al.*, 2002]. Few PE_PGRS genes have also been shown to be expressed during preclinical infection [Singh *et al.*, 2001]. Dissection of the PE_PGRS genes into PE and the PGRS domains to study their specific immunological response during mice infection revealed that the PE region can elicit an effective cellular immune response and the humoral response is largely directed against the Gly-Ala rich PGRS domain [Delogu *et al.*, 2001].

The involvement of PE/PPE genes in the virulence of the pathogen has also been reported [Ramakrishnan *et al.*, 2000]. Work done from the present laboratory has also shown that a PPE gene family member, Rv2430c is an immunodominant antigen of *M. tuberculosis* [Choudhary *et al.*, 2003, 2004].

The results presented in this chapter describe an *in-silico* approach to identify probable antigens from the PPE_MPTR [Major Polymorphic Tandem Repeat] subfamily. The humoral and cellular immune response to the same was studied using well characterized patient samples. Synthetic peptides corresponding to regions of high antigenic index of the protein were used to map the antigenic domains and assess the antigenic potential of the Gly-X-Gly-Asn-X-Gly repeat motif in eliciting a differential immune response. These results suggest that the PPE_MPTR ORF Rv2608 could be involved in directing the host towards development of a more humoral type of immune response.

6.3 EXPERIMENTAL PROCEDURES

6.3.1 PCR-RFLP analysis of the PPE ORF, Rv2608

PCR-RFLP was carried out to examine if Rv2608 exhibited polymorphism in different clinical isolates of *M. tubercuolsis*. Briefly, Rv2608 was PCR amplified from about 30 different clinical isolates and the amplified product was digested with *Sau*3AI enzyme. The digested product was separated on a 10% polyacrylamide gel and visualized under UV after ethidium bromide staining.

6.3.2 Cloning, overexpression and purification of Rv2608, a PPE MPTR subfamily member of *M. tuberculosis*

The PPE ORF, Rv2608 with predicted *in-silico* regions of high antigenic index was selected for cloning, expression and immunological characterization of the recombinant protein. The ORF was amplified from H37Rv genomic DNA using primers having specific restriction enzyme sites to enable directional cloning [Forward primer: AATGGATCCATGAATTTCGCCGTTTTG with a *Bam*HI overhang and Reverse primer:

AATAAGCTTGAAAAGTCGGGGTAGCGC with a *Hind*III overhang]. The amplified gene was first cloned in pGEMT easy vector followed by subcloning in pRSETa expression vector. The positive clones were used to transform *E. coli* BL21 cells. Protein expression was studied by initially growing the BL21 cells, transformed with the plasmid construct, overnight with appropriate antibiotics. 1% inoculum was taken from this primary culture from which a 5ml secondary culture was set up. When the absorbance value reached about 0.4, cells were induced to produce the recombinant protein by the addition of 1mM IPTG. 100µl of the culture was taken after every one hour to study the expression kinetics of the protein. A separate aliquot of uninduced culture was kept as a control. The cells were suspended in 1X SDS PAGE sample buffer and analyzed on a 12% SDS polyacrylamide gel to check for the expression of the 59.6kDa recombinant protein. The recombinant protein was then purified to homogeneity using the QIAExpressionist kit (Qiagen, USA). Cells harvested from 50 ml of induced culture were resuspended in lysis buffer containing 100 mM NaH_2PO_4, 10 mM Tris.Cl and 8M urea [pH 8.0]. The lysate was loaded onto a Ni-NTA column pre-equilibriated with the lysis buffer. The column was washed with wash buffer containing 100 mM NaH_2PO_4, 10 mM Tris.Cl and 8 M urea [pH 6.3]. Protein was eluted with elution buffer containing 100 mM NaH_2PO_4, 10 mM Tris.Cl and 8 M urea [pH 4.5], and resolved by electrophoresis in a 12% SDS polyacrylamide gel. A single 60kDa protein band was observed upon staining with Coomassie Brilliant Blue dye. The recombinant protein was tested for reactivity with patient sera after dialysis. Recombinant Hsp10 [Heat shock protein], a major secreted antigen of *M. tuberculosis* was used as control antigen.

6.3.3 Design of synthetic peptides

The PPE ORF, Rv2608 was scanned to identify regions of high antigenic index using the Protean software of Lasergene Navigator™ [DNA STAR]. Ten synthetic peptides of varying lengths corresponding to *in-silico* predicted regions of high antigenic index were commercially obtained as lyophilized powders. Peptide stocks of concentration 0.1 mg/ml were prepared in carbonate bicarbonate buffer and stored in aliquots at -70^0C.

6.3.4 Human patient sera

Fifty one TB patients confirmed by tuberculin skin test, radiographic examination and observation of Acid Fast Bacilli [AFB] in sputum for pulmonary TB and at the site of presumed TB in case of extrapulmonary infection were selected for this study. These patients were reporting to the Out Patient Department of the Mahavir Hospital and Research Centre Hyderabad, India. All the patients with confirmed diagnosis of TB were culture positive as well. The patients were categorized as follows: Category I: Individuals [n=22] diagnosed for TB for the first time; Category II: Individuals [n=21] with a relapsed TB and Category III: Extrapulmonary TB patients [n=8]. Sera were collected from all the subjects during early stage of infection when chemotherapy had just started. Healthy control [n=10] sera were taken from the laboratory staff of CDFD. These were individuals who did not have a prolonged direct contact with a TB patient. As this study was carried on a PPE gene family member of *M. tuberculosis*, members of which are unique to mycobacteria [3], cross reactivity to this protein would not be expected and therefore control subjects with other bacterial infection were not considered necessary for inclusion in our study.

6.3.5 Enzyme Linked Immunosorbent Assay [ELISA]

Serum samples from infected as well as control subjects were used to screen the recombinant protein and synthetic peptides using indirect ELISA technique**.** 96 well polystyrene microtiter plates [Costar] were incubated overnight with 2μg/ml of the synthetic peptides and 5 tuberculin units/ml of *M. tuberculosis* PPD. Carbonate bicarbonate buffer was used as the antigen coating buffer and each peptide was coated in triplicate. The plates were washed with PBS Tween [0.05% Tween 20 in PBS] and blocked by incubating with 250 μl of 2% BSA in PBS tween for one hour at 37^0C. Diluted Serum [1:200 in blocking buffer] was added to the antigen coated plate and the plates were incubated at 37^0C for another 60 minutes. After incubation, the plates were washed extensively with PBS tween followed by the addition of 100ul of secondary antibody [anti-Human IgG linked to the horse radish peroxidase enzyme]. The plates were

incubated again at 37^0C for one hour. After washing 6 to 8 times, the peroxidase activity was assessed by adding 100 µl of OPD [ortho phenylene diamine], the specific substrate of the enzyme dissolved in citrate phosphate buffer. OPD was added in dark and the plates were incubated at 37^0C for 30 minutes to allow colour formation. The reaction was stopped by the addition of 1N H_2SO_4. Absorbance was taken at 405nm.

6.3.6 Lymphocyte proliferation assay

Lymphocyte proliferation assays were carried out essentially as per method described earlier with a few modifications [Van de Loosdrecht *et al.*, 1994]. Heparinised blood was drawn and diluted with equal volume of RPMI1640 medium without serum. Diluted blood was layered on Ficoll gradient in 1:3 proportion. After a low speed [800g] centrifugation for 30 minutes, the peripheral blood mononuclear cells [PBMCs] were isolated and washed twice for 10 minutes at 800g to remove debris and platelets. Cell concentration was adjusted to 10^6/ml. Viability of the cells was checked using Trypan Blue. To each well of the microtiter plates, 0.1ml of cell suspension and 0.1 ml of antigen [2µg/ml] was added. ConA [Concanavalin A] was used as a positive control antigen. Control and experimental cultures were run in triplicate. The plate was incubated at 37^0C with 5% CO_2 for a period of 72 hours. At the end of the 3rd day, 15µl of the tetrazolium salt MTT [2mg/ml] was added and incubated for another 4 hours. The culture was terminated and the MTT crystals were dissolved in 100 µl of acidified isopropanol. After one hour, the optical density was recorded using ELISA plate reader using a dual wavelength of 570 nm and 620 nm reference filter. Data were expressed as Stimulation Index [S.I.] i.e. ratio of the mean O.D. of experimental cultures [with test antigen] to the mean O.D. of control cultures [without antigen]. S.I. greater than or equal to 2 was considered as positive stimulation index.

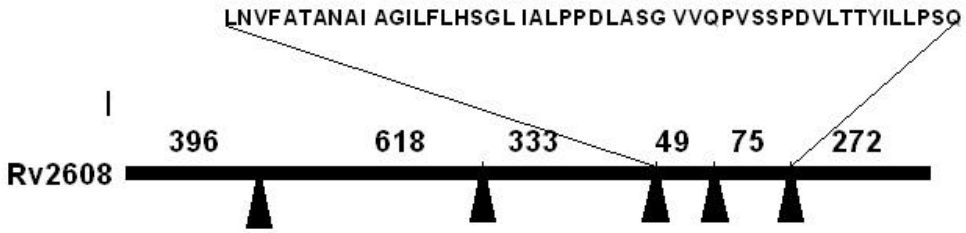
6.3.7 Statistical methods

Analysis of variance [ANOVA] as a test of statistical significance was performed using an online software [http://www.physics.csbsju.edu/stats/ANOVA.html] to calculate the p values and determine if there was any difference between different patient categories with respect to each antigen tested. The 95% confidence intervals for means were also determined for each set of data. Differences between groups were considered statistically significant if the 95% confidence interval limits did not overlap. To ascertain the results obtained by ANOVA, Kruskal Wallis non-parametric test [http://department.obg.cuhk.edu.hk/ResearchSupport/KruskallWallis.ASP] was also carried out. t tests were performed for paired comparison of means. $p<0.05$ was considered statistically significant.

6.4 RESULTS

6.4.1 Genetic variation in the PPE ORF, Rv2608

PE/PPE genes are predicted to be a source of antigenic variability of *M. tuberculosis* and polymorphism in a few of them based upon variation in the number of repeat sequences has already been reported [Espitia *et al.*, 1999]. A member of the PPE gene family, belonging to the major polymorphic tandem repeat [MPTR] subclass was analyzed by PCR-RFLP to score for the presence of genetic variation in different clinical isolates. The 1.7kb amplicon [Figure 6.1A] was digested with *Sau*3AI and the digest was electrophoresed on a 10% polyacrylamide gel. 16% of the clinical isolates showed a deviation from the normal band pattern. Figure 6.1 B gives the complete summary of the polymorphism obtained in 30 different clinical isolates. The disappearance of restriction fragments was restricted to the C terminus of the ORF, which is the predicted variable region of the PPE ORFs. It was therefore important to further evaluate Rv2608 in terms of its ability to elicit B and T cell response so as to study its role as a possible antigen for immune surveillance.



A

Location of *Sau*3AI sites	Expected bands (bp)	Additional bands (bp)	25 isolates	2 isolates	3 isolates
396	396		+	+	+
1014	618		+	+	+
1347	333		+	+	+
1396	49		+	+	-
1471	75		+	-	-
1753	272		+	+	+
		~120	-	+	+
		~100	-	+	+

B

Figure 6.1:A: *Sau*3AI Restriction map of PPE ORF, Rv2608. Arrowheads point to the *Sau*3A sites in the 1743bp Rv2608 ORF. Numbers above the line indicate the size of the restriction fragments (in base pairs) generated after *Sau*3AI digestion. B: Summary of *Sau*3AI PCR-RFLP pattern of 30 different clinical isolates of *Mycobacterium tuberculosis.*

6.4.2 Expression of Rv2608 in *E. coli* and purification of the recombinant protein

To evaluate the antigenic ability of Rv2608, the corresponding gene was expressed in *E. coli* BL21 cells and purified as a 6X His-tag fusion protein. Purified recombinant Rv2608 was fractionated by electrophoresis on a 12% polyacrylamide gel. A single band corresponding to 59.6kDa protein was observed upon staining the gel with Coomassie Brilliant Blue dye [Figure 6.2]. The expression of the gene was confirmed by probing the membrane containing the total cellular protein of *E. coli* BL21 cells harboring the Rv2608 construct with anti-Histidine antibody. There was no leaky expression of the protein in uninduced cells. The recombinant protein was largely present in the insoluble fraction and was therefore purified in the presence of 8M urea [Figure 6.2, LaneE]. The yield of the protein was 6mg/litre of culture. The recombinant protein was dialyzed overnight and used for immunoreactivity analysis.

6.4.3 Design of synthetic peptides based on antigenicity prediction of Rv2608

In-silico analysis of Rv2608 revealed the presence of two regions of high antigenicity: Two amino acid stretches [37 amino acids and 25 amino acids] corresponding to important antigenic epitopes within Rv2608 were selected for peptide synthesis [Figure 6.3 a]. Additional eight overlapping regions [Figure 6.3 b] which were essentially the subsets of the two main peptides were also selected for peptide synthesis. These peptides were used to map the antigenic domains of the protein. Table 6.1 shows the amino acid sequences of all the 10 synthetic peptides used in the present study. The peptides were part of the C terminal region of Rv2608 and apart from the high antigenic index also possessed the repeat motif Gly-X-Gly-Asn-X-Gly, characteristic of the PPE_MPTR gene family.

Table 6.1: Amino acid sequence of the synthetic peptides

PEPTIDE ID	AMINO ACID SEQUENCE*
P1	D**N**I**GN**A**N**I**G**F**GN**R**G**DA**N**I**G**I**GN**I**G**DR**N**L**G**I**GN**T**GN**WK [37]
P2	RP**G**LDELSFTLT**GN**P**N**RPD**GG**ILTK [25]
P1a	D**N**I**GN**A**N**I**G**F**GN**K [13]
P1b	**N**I**G**F**GN**R**G**DA**N**IK [13]
P1c	R**G**DA**N**I**G**I**GN**I**G**K [13]
P1d	**G**I**GN**I**G**DR**N**L**G**IK[13]
P1e	DR**N**L**G**I**GN**T**GN**WK [13]
P2a	RP**G**LDELSFTLTK [13]
P2b	LSFTLT**GN**P**N**RPK [13]
P2c	**GN**P**N**RPD**GG**ILTK [13]

*Residues in bold represent the Glycine-Asparagine repeat motifs

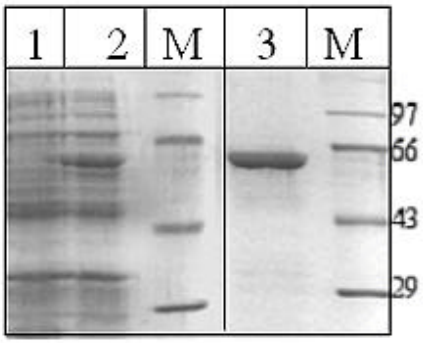


Figure 6.2: Expression and purification of *M*. *tuberculosis* protein corresponding to the PPE_MPTR ORF Rv2608. The left panel shows the uninduced and induced cell lysates and protein molecular size marker (Lanes 1, 2, M). The right panel shows the purified recombinant protein (Lane 3) and the protein molecular size marker (M).

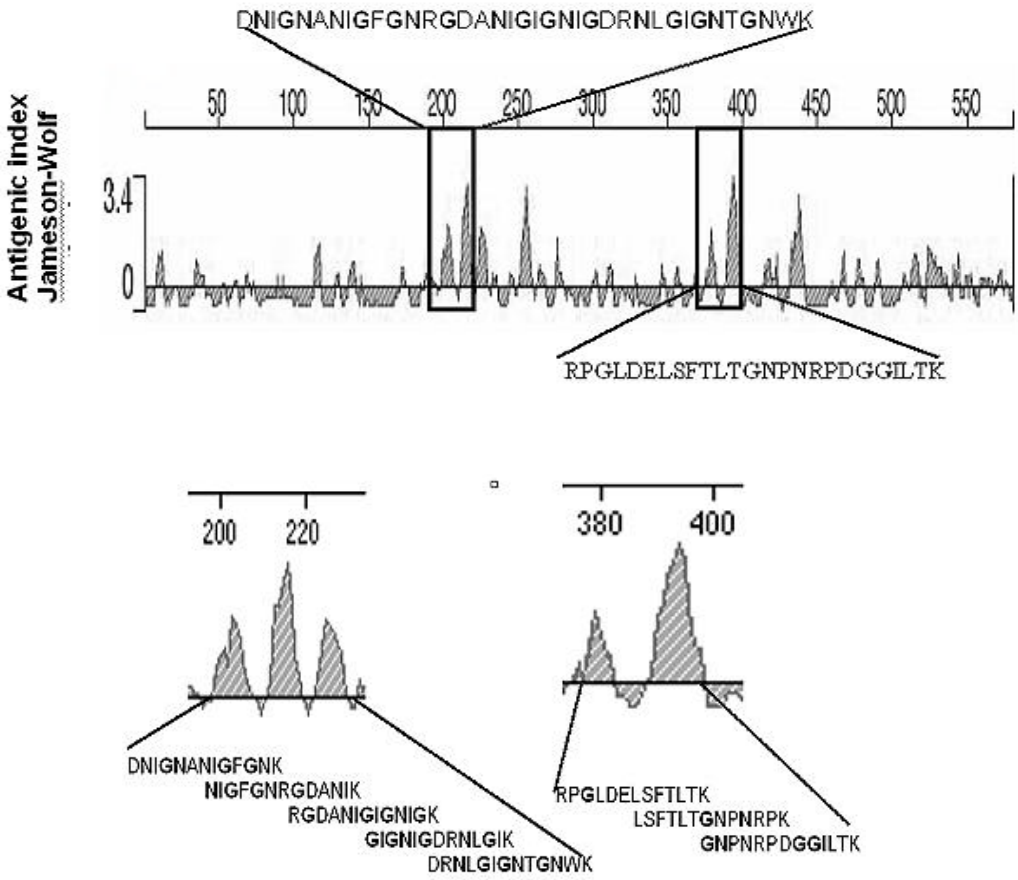


Figure 6.3:*In-silico* analysis of Rv2608 reveals regions of high antigenic index (potential antigenic determinants). Overall antigenic index of the protein was calculated using the Jameson Wolf method of the Protean software of Lasergene Navigator™. The boxed areas indicate the regions selected for designing synthetic peptides to map the region that was actually eliciting a variable immune response. As can be seen, one of the peptides (37mer) is largely composed of Gly-Asn repeats which is lesser in number in the other peptide (23mer). B: Stretches of overlapping peptides used for ELISA and T cell proliferation assay. These peptides were used to further map the region that was antigenic.

6.4.4 rRv2608 shows positive reactivity to sera from different categories of TB patients

The humoral response to the recombinant PPE protein was characterized by measuring serum IgG antibodies to the protein using ELISA. Antibody response was analyzed as a function of mean absorbance at 492nm. Recombinant Hsp10, a major secreted antigen of *M. tuberculosis* was used for comparison of the response to the rPPE protein. It was observed that for all the patient categories, serum reactivity to rRv2608 was equal to or higher than the response to Hsp10 [$p>0.05$, indicating no difference between the response to HSP10 and Rv2608] [Figure 6.4]. Healthy controls also showed some reactivity to the recombinant protein, however the response was significantly less when compared to that of patients [$p=0.0002$ using student's t test as a test of statistical significance for paired comparison of means between the patients and healthy controls].

6.4.5 Synthetic peptides corresponding to regions of high antigenicity within Rv2608 elicit strong humoral immune response in patients with relapsed TB infection:

Having shown that the recombinant protein coded by Rv2608 elicited an antibody response which was equal to or higher than that elicited by Hsp10 antigen in all the categories of TB patients selected for the study, an attempt was made to dissect differential responses if any as a function of patient category. For this, synthetic peptides spanning the two major antigenic regions within Rv2608 [P1 and P2] were used in ELISA [Table 6.1]. The results suggest that these peptides strongly react with patient sera [Figure 6.5] and hence the protein must be generating a strong humoral response in the host. Since a positive response was obtained with the peptides 1 and 2, patient sera were also tested for reactivity against the short overlapping peptide sequences 1a, 1b, 1c, 1d, 1e which were all components of peptide 1 and 2a, 2b and 2c which were a part of peptide 2. The results obtained indicate that even these overlapping peptide stretches react equally well with patient sera. Exact mapping of the antigenic region was not possible as most of the peptides showed a similar response. This was a reflection of

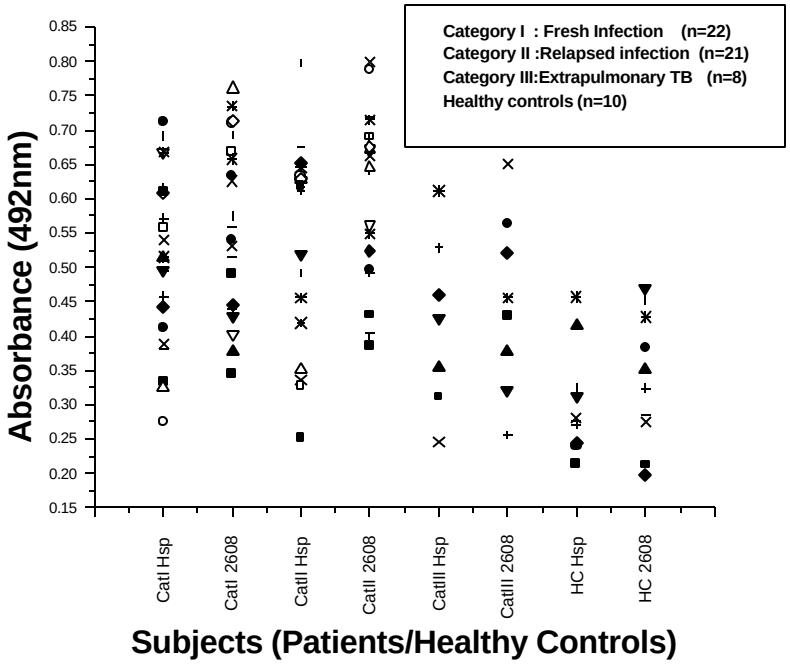


Figure 6.4:Antibody response of different categories of TB patients to rRv2608 is equivalent to the response to rHsp10, a well documented antigen of *M. tuberculosis*. Serum reactivity was measured by ELISA and the graph was plotted as patient response (O.D. at 492nm) to rHsp10 and rRv2608. The difference between patient response to Hsp10 and rRv2608 was not significant for all patient categories ($p>0.05$ using paired t tests). However, the response of healthy controls was lower and differed significantly from the patients ($p=0.0002$ using paired t test). (HC= Healthy Controls, Cat= Category)

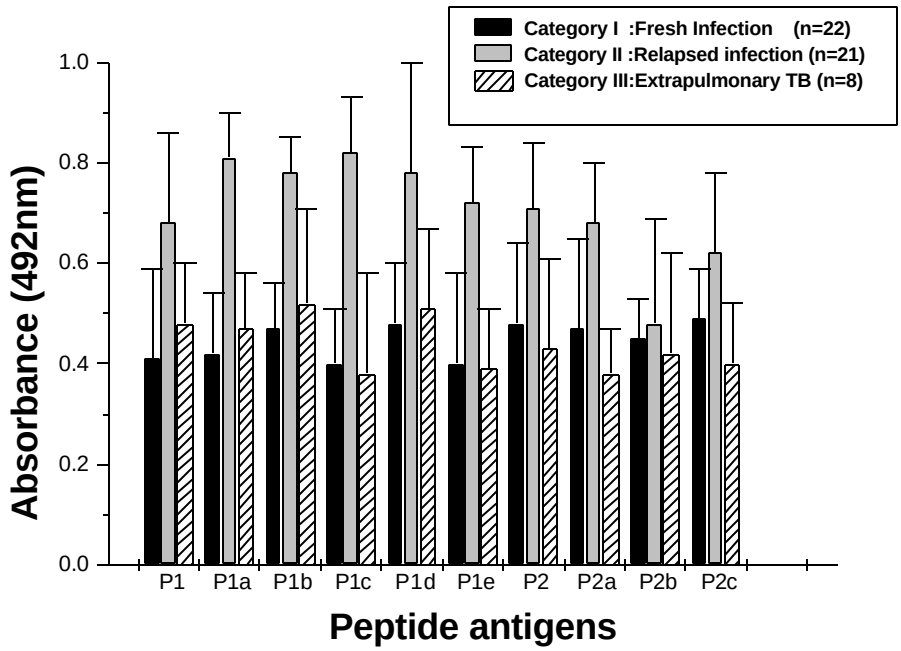


Figure 6.5:Antibody response of different categories of TB patients to different synthetic peptides (regions of high antigenicity within Rv2608) as determined by ELISA. Response to all the peptides was plotted as absorbance at 492nm (mean ± SD). The response of Category II patients was significantly higher than Category I or III ($p < 0.001$ for both, using paired t tests) with respect to each peptide antigen.

the fact that the Gly-X-Gly-Asn-X-Gly repeat motifs were present in all the peptides. Very interestingly, there was a significantly varied response to the peptides in different category of TB patients which was not so when the complete recombinant Rv2608 protein was used. The peptides could clearly distinguish between the patient categories [$p<0.001$ using ANOVA for each peptide antigen] [Table 6.2]. While humoral response observed in case of fresh infection cases [Category I] was similar to that of extrapulmonary TB patients [Category III], category II or the relapsed cases showed an unusually high antibody response to all the peptides. The response of Category II patients was significantly higher than Category I or III [$p< 0.001$ for both, using t test as a test of statistical significance].

6.4.6 The T cell response of TB patients to Rv2608 peptide antigens was low and the differences between various categories of patients were not evident

T cell proliferation assays were carried out to evaluate the response to different synthetic peptides. The overall T cell response of patients to these peptides was very low [S.I.<2] and the response could not distinguish between patient categories [$p>0.05$, using ANOVA and Kruskal Wallis test] at least for peptide 1 and its derivatives [Figure 6.6]. Peptide 2 and its derivatives exhibited a higher response in fresh infection cases as against relapsed and extrapulmonary cases [$p<0.05$ for both, using t test for paired comparison of means]. As can be noted from the amino acid sequence of the peptides [Table 6.1], peptide 2 has lesser number of glycine asparagine repeats and shows a higher T cell proliferative response in fresh infection cases.

6.5 DISCUSSION:

The ORF Rv2608 selected for the present study is a member of the PPE_MPTR class which is characterized by the presence of a conserved N-terminal region and a C-terminal domain with major polymorphic tandem repeats [MPTR] of Gly-X-Gly-Asn-X-Gly

Table 6.2: Summary of the results of statistical analyzes to estimate differences in humoral immune response to different peptide antigens

Peptide Antigen		Mean [O.D. at 492nm]	95% confidence interval of Mean	Degree of freedom	F value	P value	Difference between categories [Significant [S] / Not Significant [NS]
P1	I	0.412	0.338 to 0.486	2	12.69	<0.0001	S
	II	0.675	0.598 to 0.750				
	III	0.483	0.359 to 0.606				
P1a	I	0.426	0.360 to 0.492	2	29.69	<0.0001	S
	II	0.770	0.704 to 0.836				
	III	0.473	0.363 to 0.582				
P1b	I	0.469	0.416 to 0.521	2	35.47	<0.0001	S
	II	0.775	0.720 to 0.828				
	III	0.520	0.432 to 0.607				
P1c	I	0.416	0.364 to 0.467	2	70.48	<0.0001	S
	II	0.810	0.756 to 0.862				
	III	0.380	0.293 to 0.464				
P1d	I	0.482	0.427 to 0.535	2	33.47	<0.0001	S
	II	0.787	0.731 to 0.842				
	III	0.527	0.437 to 0.617				
P1e	I	0.407	0.348 to 0.464	2	31.26	<0.0001	S
	II	0.711	0.651 to.0.770				
	III	0.405	0.308 to 0.501				
P2	I	0.480	0.413 to 0.546	2	14.51	<0.0001	S
	II	0.706	0.637 to 0.774				
	III	0.441	0.331 to 0.551				
P2a	I	0.473	0.412 to 0.533	2	18.63	<0.0001	S
	II	0.689	0.627 to 0.750				
	III	0.392	0.291 to 0.491				
P2b	I	0.450	0.379 to 0.521	2	0.7688	0.4	NS
	II	0.498	0.425 to 0.570				
	III	0.422	0.304 to 0.540				
P2c	I	0.491	0.433 to 0.548	2	10.05	0.0002	S
	II	0.632	0.572 to 0.690				
	III	0.410	0.314 to 0.505				

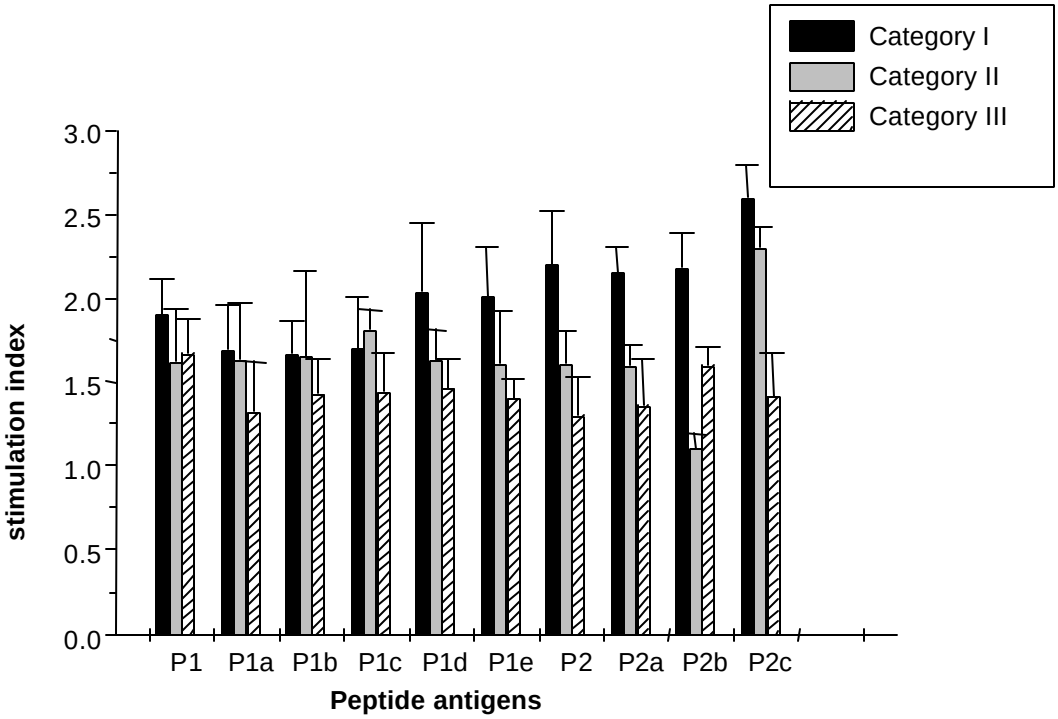


Figure6:The T cell proliferative response of different categories of TB patients as a result of stimulation by various synthetic peptides (regions of high antigenicity within Rv2608). Stimulation index(S.I.) = O.D. in test well (with antigen) /O.D. in control well (without antigen)> 2 was taken as a positive response, where test antigens are the peptides. The overall T cell response of TB patients to the peptide antigens was low. Only peptide 2 and its subsets show a positive S.I. Maximum S.I. is seen in Category I patients which progressively decreases in category II and extrapulmonary cases.

residues. Apart from this, the ORF also possesses regions of high antigenic index, which is a measure of overall hydrophilicity and surface probability. To test if polymorphism of the C-terminal region of this ORF exists in different clinical isolates of *M. tuberculosis*, PCR amplified Rv2608 was subjected to PCR-RFLP analysis. The observed variation in the band pattern lends weight to the hypothesis that PE/PPE genes, notably Rv2608 are perhaps a source of antigenic variability in the otherwise conserved genome of *M. tuberculosis*.

The rRv2608 protein was used in ELISAs to determine its reactivity to patient sera. The primordial observation that the recombinant protein reacted with patient sera indicates that this protein is definitely expressed during infection. Serum response of patients as well as healthy controls to rRv2608 was equivalent to or greater than the response to Hsp10, a well documented antigen of *M. tuberculosis* [Young and Garbe, 1991]. While category wise differentiation of serum reactivity towards the full length recombinant protein was not very apparent, it was significant to note that the extrapulmonary TB patients showed less reactivity with rRv2608 protein as compared to Category I or II [p=0.048]. It will be worthwhile to explore whether Rv2608 represents a protein[s] required by the bacterium to establish a pulmonary infection.

Since the serum response to the recombinant PPE protein was equal to or greater than Hsp10, it was decided to possibly map the antigenic domains of the probable PPE antigen using a synthetic peptide approach [Miguez *et al.*, 1998; Liljeqvist *et al.*, 1998; Benitez *et al.*, 1998]. Peptides corresponding to regions of high antigenic index were accordingly designed. The analyzes of the comparative humoral immune responses indicate that the serum response of patients to all the ten peptides is similar. This could be explained by the fact that all the peptides have a common repeat motif thereby eliciting similar response. While this negated the efforts to map the immunodominant epitope required for eliciting a strong humoral immune, a difference in the response of patients categorized according to different states of infection was surprisingly evident. Category II patients [relapsed infection cases] demonstrate the highest B cell response to the peptides followed by extrapulmonary TB cases.

The synthetic peptides were also used for T cell proliferation assays with the peripheral blood mononuclear cells of all category patients. It has been earlier shown that in about 90% of patients with active TB, there is a significant antibody response and/or T cell proliferative response to peptide specific single antigens of *M. tuberculosis* [Wilkinson *et al.*, 1998]. The 38kDa antigenic protein of *M tuberculosis* is a potent stimulus for both T cell and B cell responses in humans [Young *et al.*, 1986; Anderson *et al.*, 1989]. The T cell proliferative response to the synthetic peptides was of the order of fresh infection cases> relapsed TB > extrapulmonary TB cases at least for peptide 2 and its derivatives. However, the observed Stimulation Index [SI] with all the peptides was very low in all categories of TB patients [S.I.<2]. A high humoral response and a low cellular immune response to the peptides in category II patients points to an important possible function of the PE/PPE gene families. It is likely that these antigens play a role in evading host immune response and prevent the establishment of an effective cellular response, which is required to contain the disease. The positive T cell response in some cases could be explained by the fact that IgG antibody responses again require the involvement of helper T cells.

Antibody levels usually decrease in cured TB cases but dramatically increase in patients showing poor compliance [Bothamley *et al.*, 1993]. High antibody response to the peptides and a low T cell response hence explain the relapse of infection in category II patients. *In vivo*, it is possible that the responsive T cells are not able to expand as the glycine, asparagine repeat motifs somehow prevent antigen processing. The situation can be equated with the Epstein Barr Virus Nuclear antigen, where again the Gly-Ala repeat regions play an important role in preventing antigen processing [Levitskaya *et al.*, 1995]. Peptides 2 and 2c, which have lesser number of Gly-Asn repeats show a comparatively higher T cell response.

In conclusion, these results establish a relationship between immune responses to the PPE antigen and the status of the disease [fresh or relapsed TB]. The present study is the first report wherein, in a clinical setting, it has been demonstrated that the repeat sequences present within Rv2608 elicit a high humoral immune response and a low T cell response. Since PPE_MPTR is a gene family of *M. tuberculosis* of which Rv2608 is

a member sharing the MPTR motif, it is likely that other members of the same family may also serve the same function in the bacterium. These data contribute towards a better understanding of humoral as well as cellular immune responses elicited by PPE antigens. The practical utility of using these peptides for differentiating fresh infection from relapsed or reactivation cases is another interesting proposition and deserves large scale validation.

SUMMARY

SUMMARY

The present study represents the first systematic effort to characterize the aromatic amino acid biosynthesis pathway enzymes of *Mycobacterium tuberculosis* and establish its relationship with iron metabolism of the bacterium. While earlier reports have demonstrated the essentiality of the pathway [Parish and Stoker, 2002; Sassetti *et al.*, 2003], this study has been able to surface the intricacies of the catalytic and regulatory mechanisms employed by few member enzymes of the pathway.

The work was initiated with the identification of IdeR regulated genes of *M. tuberculosis*, which resulted in the identification of previously known as well as unknown targets of IdeR. The as yet unknown targets of IdeR identified in this study include *fec*B, a member of the ferric dicitrate uptake system; Rv1404, a MarR equivalent transcription regulator and the urease operon. Regulation of the urease operon by IdeR was demonstrated by *in vitro* binding assays. The known targets of IdeR that drew attention include the aromatic amino acid biosynthesis enzymes. Subsequently, the study was continued with the functional characterization of a few aromatic amino acid biosynthesis enzymes that might help the bacterium to survive inside the nutrient deficient environment of the host and possibly play a role in the pathogenicity of the disease.

Chorismate mutase [CM], prephenate dehydratase [PDT], anthranilate synthase [AS] and isochorismate synthase [ICS] enzymes of *M. tuberculosis* were selected for a detailed characterization. The corresponding genes were individually cloned and expressed in *E. coli* and the recombinant proteins were purified using metal affinity chromatography. Chorismate mutase activity was assigned to a hypothetical protein corresponding to ORF Rv1885c of *M. tuberculosis*. *In vitro* assays showed that the enzyme followed Michealis Menton's kinetics and the activity was temperature and pH dependent. Other biochemical assays demonstrated that *M. tuberculosis* CM does not have any associated PDT [prephenate dehydratase] or PDH [prephenate dehydrogenase] activity, indicating that it is a monofunctional protein. It was also demonstrated that *M. tuberculosis* CM, in spite of being a small protein is a highly regulated enzyme that shows feedback inhibition by pathway specific [tyrosine and phenylalanine] as well as cross pathway specific [tryptophan] aromatic amino acids.

Additionally, it was also shown that *M. tuberculosis* CM is a dimeric, alpha helical protein and possesses a functional signal sequence at its N-terminus. These studies could generate sufficient line of evidence to place the enzyme coded by *M. tuberculosis* ORF Rv1885c in the aroQ class of periplasmic chorismate mutases.

M. tuberculosis genome was scanned for the presence of another copy of chorismate mutase, as redundancy of the enzyme has been reported in many bacterial systems. According to the COG [Cluster of Orthologous Groups] database, Rv0948c could be the second chorismate mutase enzyme of *M. tuberculosis*. Hence, the gene corresponding to Rv0948c was expressed in *E. coli* and the recombinant protein was purified. It was observed that rRv0948c also possessed chorismate mutase activity though the specific activity was much lower than that of Rv1885c. This appears to be the first report of the existence of two chorismate mutase enzymes in a bacterium, both of which are apparently non-fusion proteins, having differential enzymatic activity.

Prephenate, the end product of the reaction catalyzed by CM is the substrate for the enzyme prephenate dehydratase [PDT]. Studies carried out on *M.tuberculosis* PDT indicate regulation of the enzyme in a manner exactly opposite to that of *M. tuberculosis* CM. While aromatic amino acids could bring about inhibition of activity of *M. tuberculosis* CM, they caused extensive activation of *M. tuberculosis* PDT. The other unique property of *M. tuberculosis* PDT that this study could decipher was the absolute requirement of the catalytic as well as the regulatory domains of the protein for optimum enzyme activity. Fluorescence spectroscopy could demonstrate that phenylalanine binding induces a conformational change only upon the full-length protein and the individually cloned and purified regulatory domain. The fluorescence spectrum of the catalytic domain was insensitive to the addition of phenylalanine. It was also proved that *M. tuberculosis* PDT does not possess any associated CM activity indicating that like CM, PDT is also a monofunctional protein. These studies on PDT and CM have therefore provided a definite mechanism for the control of terminal branches of aromatic amino acid biosynthesis in *Mycobacterium tuberculosis*.

Anthranilate synthase and isochorismate synthase are other two aromatic amino acid biosynthesis enzymes that were characterized in the present study. While the ORF

Rv1609 of *M. tuberculosis* is annotated as an anthranilate synthase [*trp*E] [Cole *et al*., 1998], Rv2386c was earlier annotated as ananthranilate synthase [*trp*E2] [Cole *et al*, 1998] and later re-annotated as an isochorismate synthase [*mbt*I] [Quadri *et al*., 1998]. These two ORFs were expressed in *E. coli* BL21 cells and the corresponding recombinant proteins were purified. Functions of these genes were studied using biochemical assays as well as using a genetic approach. A tryptophan auxotrophic *E. coli* BL21 strain was constructed that specifically lacked AS activity. This strain could not grow on minimal media but could grow on minimal media supplemented with anthranilate. The anthranilate auxotrophy of this strain could be satisfied with pET23a chimeric constructs carrying *M. tuberculosis* ORFs Rv1609 as well as Rv2386c. This experiment suggested that while Rv1609 [*trp*E] does code for an anthranilate synthase, Rv2386c also possesses AS activity. The ICS [Isochorismate synthase] activity of rRv2386c [*trp*E2/*mbt*I] was also tested using HPLC equipped with a diode array detector. It was observed that rRv2386c could convert chorismate to isochorismate, the precursor for salicylate. A direct proof for the involvement of *M. tuberculosis* TrpE2/MbtI [Rv2386c] and CM [Rv1885c] in salicylate biosynthesis was also provided.

In vitro ammonia-dependent AS activity assays for the protein coded by Rv1609c and Rv2386c were carried out using fluorescence spectroscopy. AS activity of Rv2386c was found to be four fold lower than Rv1609. It was also observed that Rv1609 [TrpE] as well as Rv2386c [TrpE2] could be protected from tryptic cleavage in the presence of tryptophan. Genetic and biochemical studies carried out on *M. tuberculosis* ORFs, Rv1609 and Rv2386c point to a convergent evolution of AS and ICS enzyme activities. The high sequence similarity of the two groups of enzymes is perhaps responsible for an overlap of function.

These studies suggests that the flux of chorismate into the biosynthesis of amino acids and secondary metabolites is regulated at both genetic and enzymatic levels in *M. tuberculosis*. While Rv2386c [*trp*E2/*mbt*I] and Rv3838c [*phe*A] have been reported to be under the regulatory control of IdeR [Gold B *et al.*, 2001], a 3 fold induction of *M. tuberculosis* ORF, Rv1885c has been reported in a *sig*E mutant *M. tuberculosis* strain [Manganelli *et al.*, 2001]. According to the present study, regulation of all these enzymes is also brought about by intracellular concentration of various ligands.

While assimilation of nutrients from the hostile environment of the host is an important virulence attribute of pathogenic microbes, evasion of the host immune response is another factor that determines the outcome of an infection. Off late, a lot of evidence has come up in support of the PE/PPE family proteins of *M. tuberculosis* as dominant antigens [Choudhary *et al.*, 2003]. The hypothesis whether the PPE family proteins of *M. tuberculosis* play a role in eliciting a variable host immune response was tested. A representative PPE_MPTR family ORF, Rv2608 was selected that possessed regions of high antigenic index. The corresponding recombinant protein showed positive reactivity to patient sera. Further, a synthetic peptide approach was used to map the antigenic domains of the protein. These peptides were used in ELISAs as well as lymphocyte proliferation assays. This study could identify the PPE_MPTR motif Gly-X-Gly-Asn-X-Gly as being responsible for eliciting a predominantly B cell response in patients suffering from relapsed tuberculosis [Chakhaiyar *et al*., 2004]. This finding is important in the context of identification of novel antigens of *M. tuberculosis* that can be used as subunit vaccines.

In summary, the work presented in this thesis describes for the first time, the detailed characterization of few enzymes [Chorismate mutase, prephenate dehydratase, anthranilate synthase and Isochorismate synthase] of *M. tuberculosis* that participate in the biosynthesis of aromatic amino acids and/or iron acquisition systems. In light of essentiality of these genes and the absence of a human homologue, these enzymes could serve as novel and specific drug targets to check the growth of the tubercle bacillus. Additionally, the study also improves the current understanding of the antigenic proteins of *M. tuberculosis* that are used by the bacterium to establish a successful infection inside the host.

VITAE

PRACHEE

Laboratory of Molecular and Cellular Biology,
Centre for DNA Fingerprinting and Diagnostics,
HYDERABAD, INDIA

EDUCATIONAL QUALIFICATIONS

Ph.D (Thesis submitted to University of Hyderabad), 2004
Centre for DNA Fingerprinting and Diagnostics, Hyderabad, INDIA
Title of thesis: Studies on biomedically important *Mycobacterium tuberculosis* genes involved in nutrient assimilation and immune response

Master of Science (M.Sc), 1999
Department of Botany,
University of Delhi, Delhi, INDIA
Specialization: Genetics and Plant Breeding
First Division with 75% Marks

Bachelor of Science (B.Sc), 1997
University of Delhi, Delhi, INDIA
Subjects; Botany (Hons), Zoology and Chemistry
First Division with 77% Marks

PUBLICATIONS

1. **Prachee Prakash**, Niteen Pathak and Seyed E Hasnain. 2005. *phe*A of *Mycobacterium tuberculosis* encodes a monofunctional prephenate dehydratase that requires both catalytic and regulatory domains for optimum activity. *Journal of Biological Chemistry* 280:19641-8
2. **Prachee Prakash**, Bandi Aruna, Abhijit A. Sardesai, and Seyed E. Hasnain. 2005. Purified recombinant hypothetical protein coded by ORF Rv1885c of *Mycobacterium tuberculosis* exhibits a monofunctional AroQ class of periplasmic chorismate mutase activity. *Journal of Biological Chemistry* . 280:20666-71
3. **Prachee Prakash**, Sailu Yellaboina, Akash Ranjan and Seyed E. Hasnain. 2005. Computational prediction and experimental verification of novel IdeR binding sites in the upstream sequences of *Mycobacterium tuberculosis* ORFs. *Bioinformatics* 21:2161-6
4. **Prachee Chakhaiyar**, Nagalakshmi Y, Aruna B, Murthy KJ, Katoch VM, Hasnain SE. 2004. Regions of high antigenicity within the hypothetical PPE Major Polymorphic Tandem Repeat Open-Reading Frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. *Journal of Infectious Diseases. 190:1237-44.*
5. **Prachee Chakhaiyar** and Hasnain SE. 2004. Defining the mandate of tuberculosis research in a post genomic era*. Medical Principles and Practice. 13:177-184.*
6. Rohini Qamra, **Prachee Prakash**, Bandi Aruna, Seyed. E. Hasnain and Shekhar. C. Mande 2005. Crystallizatio and preliminary X-ray crystallographic studies of *Mycobacterium tuberculosis* chorismate mutase. *Acta Crystallograph Section F*. 61-473-475
7. Rohini Qamra, **Prachee Prakash**, Bandi Aruna, Seyed. E. Hasnain and Shekhar. C. Mande 2006. The 2.15 A crystal structure of *Mycobacterium tuberculosis* chorismate mutase reveals an unexpected gene duplication and suggests a role in host-pathogen interactions. *Biochemistry*.45(23):6997-7005.

8. Yellaboina S., Ranjan S., **Prachee Chakhaiyar** , Hasnain SE, and Ranjan A. 2004. Prediction of DtxR regulon: Identification of binding sites and operons controlled by Diphtheria toxin repressor in *Corynebacterium diphtheriae*. *BMC Microbiology. 4:38*

9. Ghosh S, Rasheedi S, Rahim SS, Banerjee S, Choudhary RK, **Prachee Chakhaiyar**, Ehtesham NZ, Mukhopadhyay S and Hasnain SE. 2004. A novel method for enhancing solubility of the expressed recombinant proteins in *E. coli*. *BioTechniques. 37: 418-423*

10. Choudhary RK, Mukhopadhyay S, **Prachee Chakhaiyar**, Sharma N, Murthy KJR, Katoch VM, and Hasnain SE. 2003. PPE antigen Rv2430c of *Mycobacterium tuberculosis* induces a strong B-cell response. *Infection Immunity. 71: 6338-6343.*

11. Siddiqi N, Shamim M, Hussain S, Choudhary RK, Ahmed N, **Prachee**, Banerjee S, Savithri GR, Alam M, Pathak N, Amin A, Hanief M, Katoch VM, Sharma SK, Hasnain SE. 2002. Molecular characterization of multidrug- resistant isolates of *Mycobacterium tuberculosis* from patients in North India. *Antimicrobial Agents And Chemotherapy. 46:443-50.*