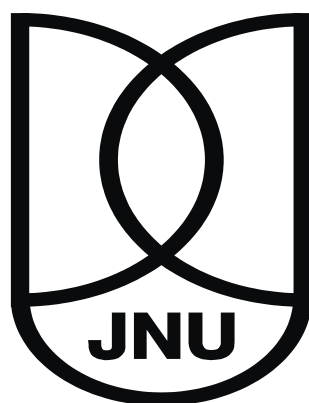# Structural and Functional analysis of Mycobacterium tuberculosis proteins EspC and EccA1 involved in ESX-1 secretion system and DprE1 and DprE2 involved in cell wall synthesis

Thesis submitted to Jawaharlal Nehru University
for award of the degree of

DOCTOR OF PHILOSOPHY

VIPIN KUMAR KASHYAP

School of Life Sciences
Jawaharlal Nehru University
New Delhi - 110067
2017

Jawaharlal Nehru University
New Delhi
India

CERTIFICATE

This is to certify that the research work embodied in this thesis entitled,

"Structural and Functional analysis of Mycobacterium tuberculosis proteins EspC and EccA1 involved in ESX-1 secretion system and DprE1 a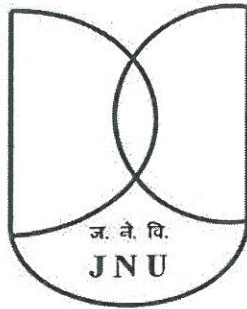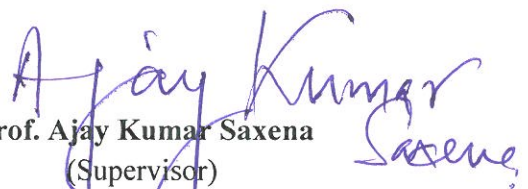nd DprE2 involved in cell wall synthesis," submitted for the award of degree of **Doctor of Philosophy** has been carried out by **Mr. Vipin Kumar Kashyap** under the guidance and supervision of **Prof. Ajay Kumar Saxena**, at Structural Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi – 110067, India.

This work is original and has not been submitted so far, in part or in full for award of any other degree or diploma of any university.

Vipin Kumar Kashayap
(Ph.D. Candidate)

Prof. Ajay Kumar Saxena
(Supervisor)

Prof. Ajay Kumar Saxena
Structural Medicine Lab
School of Life Sciences
JNU, New Delhi   110067

Prof. S. K. Goswami
(Dean)

Dedicated

to

My Parents

ACKNOWLEDGEMENT

It has been a long journey to this time where I have been able to express my gratitude towards all the persons who at some moment of the life have in some way provided me with support, enthusiasm and criticism which has been critical for my journey in general and this thesis in specific.

I thanks go to Jawaharlal Nehru University New Delhi India which provides me the ideal atmospheric condition where I had been showered upon my research carrier since August 2011. Completing my PhD degree is probably the most challenging activity of my life at this stage. I take this special moment to express my sincere appreciation and praise for those who shared the best and worst moments of my doctoral journey.

First and foremost, I would like to extend my sincerest gratitude and respect to my supervisor, Prof. Ajay Kumar Saxena, who not only gave me the opportunity to pursue my Ph.D. under his guidance, but did so most generously and wholeheartedly. I want to thank him for his unflagging encouragement and having faith in my abilities. I feel extremely grateful for his guidance and support in my life.

I expand my thanks to present Dean, Prof. S.K. Goswami and all the previous deans, prof. B C Tripathi, prof. B. N Mallick, Prof. N. B Sarin for providing excellent facilities in the School. Special thanks go to my Doctoral committee members, Prof. S. Gourinath, prof. S. S. Komath, and Dr. V. Biswal for their support and valuable suggestions. I also wish to acknowledge all the faculty of School of Life Sciences for their help, cooperation and supportive attitude towards all the students.

I am indebted to my lab Seniors Dr. Shantipal Gangwar and Dr. sitaram Meena who deserves a special mention for supervising me to carry out experiments, showering love, care and support whenever required, like an elder brother. I heartfelt and sincerely thanks to Bhanu, Sumita and Arkita for providing a healthy and scientific atmosphere as well as maintaining a refreshing atmosphere in lab. I am also thankful to Kalai and Rakesh for help in bioinformatics studies. I am immensely grateful to my SLS research scholars specially Sudhaker, Prabu, Arun, Buddhi, Sabir, Gopinath ,Gunjan, Sunil, Arpit, Kailash and Deepak for their and suggestions. I would also like to thank my all previous lab members; Sunil, Ritu, Jyoti, Nainee, ahbishek, Shiva, and Prathistha for their assistance. I would like to thank Mr. Harpal Sharma, Mr. Naveen, for their kind help and support in the lab in all regard, whenever I needed.

My special appreciation goes to Dr. Anil Verma, Dr. Shalini Gupta, Dr. Sauvik Maiti, Dr. Syed Safi and Miss. Meenal Goel for their support and encouragement.

ABBREVIATIONS

γ	:	Gamma
µ	:	Micron
µM	:	Micromolar
µl	:	Microliter
Å	:	Angstrom
ADP	:	Adenosine Diphosphate
ATP	:	Adenosine Triphosphate
AMP	:	Adenosine Monophosphate
AIRF	:	Advance instrumentation research facility
bp	:	Base pair
BCG	:	*BacilleCalmette-Guérin*
BLAST	:	Basic Logic Alignment Search Tool
bp	:	base pair
BTZ	:	Benzothiazinones
Co-IP	:	Co- immunoprecipitation
Coot	:	Crystallographic Object-Oriented Toolkit
DNA	:	Deoxyribonucleic Acid
DAPI	:	4', 6-diamidino-2-phenylindole
DPR	:	decaprenylphosphoryl-beta-D-ribose
DPA	:	decaprenylphosphoryl-beta-D-arabinofuranose
DprE	:	decaprenylphosphoryl-beta-D-ribose 2-epimerase
EDC	:	N-ethyl-N-(diethylami-nopropyl) carbodiimide
EDTA	:	Ethylenediamine Tetra Acetate
ExPASy	:	Expert Protein Analysis System

EspA	:	ESX-1 substrate protein A
EspB	:	ESX-1 substrate protein B
EspC	:	ESX-1 substrate protein C
EspD	:	ESX-1 substrate protein
EspR	:	*ESX-1-secreted protein regulator*D
HEPES	:	2-[4-(2-hydroxyethyl)-1-piperazin-1-yl]ethane sulfonate
HIV/AIDS syndrome	:	Human immunodeficiency virus/Acquired immunodeficiency
IPTG	:	Isopropyl β-D-1-thiogalactopyranoside
Kan	:	Kanamycin
Kb	:	Kilobase
KDa	:	Kilo Dalton
K_D	:	Dissociation constant at equilibrium
LAM	:	Lipoarabinomannan
LB	:	Luria Broth
LPS	:	Lipo polysaccharides
MALDI	:	Matrix-Assisted Laser Desorption/Ionization
TOF/TOF	:	Time of Flight
Mtb	:	*Mycobacterium tuberculosis*
Mb	:	Mega base pair
MOM	:	Mycolate outer membrane
ManLAM	:	Mannosylated LAM
MD	:	Molecular Dynamics
MIC	:	Minimal inhibitory concentration
MDR	:	Multi Drug Resistance
MycP	:	Mycosin-1 protease

NCBI	:	National Center for Biotechnology Information
NHS	:	N-hydroxysuccinimide
NMR	:	Nuclear Magnetic Resonance
Ni-NTA	:	Nickel Nitrilotriacetic acid
℃	:	degree centigrade
PAGE	:	Polyacrylamide Gel Electrophoresis
PIM	:	Phosphatidylinostolmannoside
PDIM	:	Phthioceroldymycocerosate
PCR	:	Polymerase Chain Reaction
PGL	:	Phenolic Glycolipid
PMSF	:	Phenyl Methyl Sulphonyl Fluoride
PubMed	:	Publication in Medical Science
RMSD	:	Root Mean Square Deviation
RPM	:	rotation per minute
SDS	:	Sodium dodecyl sulfate
SDS-PAGE	:	Sodium dodecyl sulfate-Polyacrylamide Gel Electrophoresis
TDR	:	Total Drug Resistance
Trx	:	Thioredoxin
UniProt	:	Universal Protein Resource
WHO	:	World Health Organization
XDR	:	Extensively Drug Resistance
β-ME/2-ME	:	β-Mercaptoethanol

<h1 style="text-align:center">CONTENTS</h1>

PART A:

STRUCTURE AND FUNCTION ANALYSIS OF MYCOBACTERIUM TUBERCULOSIS PROTEINS EspC AND EccA1INVOLVED IN ESX-1 SECRETION SYSTEM

Review of Literature

Bacterial pathogenicity depends on the ability to secrete virulence factors which can be exposed on bacterial cell surface, secreted into the extracellular host cytoplasm or injected directly into a host cell (Finlay and Falkow 1997). *M. tuberculosis*; the causative agent of tuberculosis is a deadliest communicable bacterium and a serious global health problem, accounting for millions of deaths every year (Organization 2009). *M. tuberculosis* is gram positive bacteria having a genome size of ~4.4 Mb with nearly 4000 genes. *M. tuberculosis* is spread through aerosol which contains bacilli and inhaled bacilli are engulfed by alveolar macrophage within the lungs. Instead of being killed, MTB survives within the phagosomal compartment of macrophage, blocks phagosome maturation and reproduce within the host cell (Feltcher, Gunawardena et al. 2015). Tuberculosis has two types, (i) Active tuberculosis where active mycobacterial bacilli can be observed in the sputum and (ii) Inactive or Latent tuberculosis where tuberculosis bacilli remain dormant and didn`t show any disease symptoms. These days tuberculosis become a great threat to health and life because of emergence MDR, TDR and XDR strains of M. tuberculosis and its occurrences with other infections like HIV, Hepatitis and metabolic disorders like diabetes etc.

M. tuberculosis creates several challenges to drug development. One challenge is to reaching drug molecules to intracellular MTB bacilli which can be present within granulomas. Another challenge is the permeability of drugs to cross unique and complex cellular boundary of *M. tuberculosis*.

All bacteria possess several protein export systems to export proteins synthesized in the cytoplasm, across the cytoplasmic membrane. These exported proteins have their functions outside of the bacilli. For this, they need specialized protein export systems. Thus these export systems are important to full virulence because of their role in exporting virulence factors that interact with the host.

MTB possess many protein export systems such as Sec pathway, Tat pathway, and ESX pathway. Among these, ESX pathway is further of five types - Esx-1 to 5. All of these pathways functions in virulence. Due to the development of drug resistant strains of MTB (MDR and XDR) and Disability of BCG vaccine to develop a sufficient protection against MTB, it is necessary to find out new drug targets and vaccine candidate molecules to provide better protection against tuberculosis. In recent, on the basis of current knowledge on MTB

Esx-1 pathway, researchers are trying to find out new drug target sites and vaccine candidate molecules, those are relevant to this export system.

Esx-1 of MTB and related tubercle bacilli is the Esx/T7SS system located in the region of difference 1 (RD-1). RD- 1 is deleted in BCG vaccine strain s of *M. bovis*. Many effectors molecules like ESAT-6   (early secreted antigenic target-6) is a 6 kDa molecule secreted by this pathway (Sørensen, Nagai et al. 1995). This protein is secreted via Esx-1, is first recognized by 10 kDa protein CFP-10. Besides these many other non RD-1 region genes are involved in secretion process like EspC, EspR, EspA, EspD etc. These molecules secreted via anEsx-1 pathway in a codependent manner and play a regulatory role in secretion process (Simeone, Bottai et al. 2009).

Various components of this export system upon secretion stimulate a strong immune response in host cells like ESAT-6, CFP-10, EspC. Some of the secreted antigens also play role in transcriptional regulation and secretion process like EspR. Immune response to these antigens are negligible or absent in BCG vaccinated persons (Millington, Fortune et al. 2011), so there is a possibility to the development of recombinant BCG vaccine supplemented with these antigens, design a drug that interrupts the mechanism by which these antigen work in pathogenicity to provide better protection against tuberculosis.

Secretion systems in M. tuberculosis

More than 20% of bacterial proteins have a function outside the cell and are reached to their proper location by several proteins export systems across the cell envelope (Schneider 1999). In M. tuberculosis also have various protein export systems to export proteins their proper locations like Sec pathway, Tat pathway, and Esx pathway. Each pathway is specialized for their substrate antigens recognition present on N or C terminal end of proteins.

Sec export system in MTB

MTB conserved Sec pathway export unfolded preproteins that are synthesized with n-terminal signal sequences. The SecA is an ATPase that recognizes and facilitates transport of preproteins through SecYFG pore complex present on mycobacterial cell envelope. SecA ATPase provides energy for this transport by repeated cycle of ATP binding and hydrolysis. Many other proteins like SecD, SecF and YajC proteins increase the efficiency of sec protein

export, and during or shortly after the translocation (Braun and Wu 1994, Economou and Wickner 1994, Paetzel, Karla et al. 2002). The signal peptide removed by a membrane bound serine protease (SP) like LspA(Sander, Rezwan et al. 2004).

Mycobacterium is unusual in having two homologs of SecA; SecA1 and SecA2. SecA1 is a house keeping gene and SecA2 is an accessory SecA involved in theSecA2 pathway. SecA gene cannot be deleted from MTB or non-pathogenic M. smegmatis unless an exogenous copy of secA1 is provided (Braunstein, Brown et al. 2001, Sassetti, Boyd et al. 2001, Guo, Monteleone et al. 2007).

Tat protein export system

MTB has a Tat pathway that export folded proteins with N-terminal Tat signal peptide sequence across the cytoplasmic membrane (Raghavan, Manzanillo et al. 2008). Tat pathway differs from Sec System because it only exports unfolded preproteins in the cytoplasm. Proteins that are exported by Tat pathway, have N- terminal sequence (Lee, Tullman-Ercek et al. 2006), but there is a presence of a pair of arginine residues (RR), that are contained within the Tat motif RRϕϕXX (where ϕ is hydrophobic residues and XX can be any residue), is distinguish feature of this pathway. The replacement of both RR residues with Lysine residues interrupt Tat- depended export (Stanley, Palmer et al. 2000).
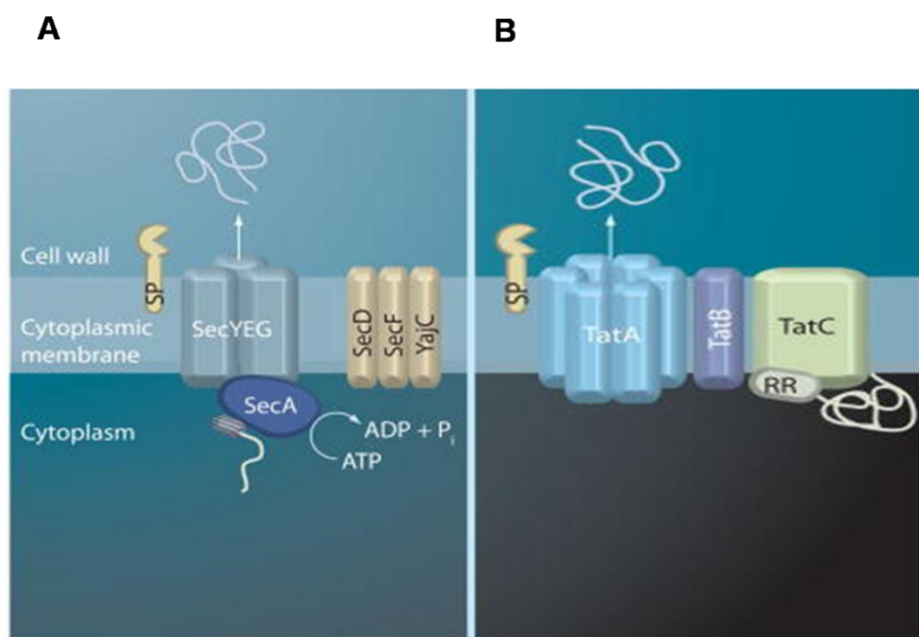
Fig.1Schematic Diagram of bacterial A, SecA secretion system B, Tat secretion pathway

MTB Tat pathway consists of TatA (integral membrane protein), TatB, and TatC (integral membrane protein) proteins. The tat signal peptide targets the folded preproteins to the Tat BC complex in the cytoplasmic membrane. Then TatA is recruited to the TatBC complex and forms a homo–oligomerictranslocase channel. Proteins are only exported by the tat system when conditions are favorable for cytoplasmic folding (DeLisa, Tullman et al. 2003).

Tat mutant of M. smegmatis has increased sensitivity to β- lactam antibiotics that indicate β-lactamase exported via this pathway(McDonough, Hacker et al. 2005, Posey, Shinnick et al. 2006).This protein has tat signal peptides. Many other proteins have been reported for the presence of tat peptide signals like phospholipaseC-A and phospholipaseC-B (PlcA and PlcB). These proteins are necessary for full virulence of MTB in mice and providing strong evidence that tat pathway contributes to the MTB pathogenesis (Raynaud, Guilhot et al. 2002). Another protein Rv2525c has tat signal peptide sequence and is suggested to have a role in infection by the demonstration of increased virulence of *M. tuberculosisRv2525c* mutant in macrophages and mice(Raynaud, Guilhot et al. 2002),(McDonough, McCann et al. 2008)).

Esx protein export system

In MTB and some gram, positive bacteria have five specialized Esx export system (esx-1 to Esx-5). The Esx systems are named for the first known secreted substrate of any Esx pathway, the 6kda early secreted antigenic target (ESAT-6) of MTB. The secreted protein lacks tat or Sec signals. More recently Esx systems have been referred to as type VII secretion system (T7SS)(Abdallah, Van Pittius et al. 2007).Each of Five Esx gene cluster contains the gene for secretion machinery and for secreted substrate protein.
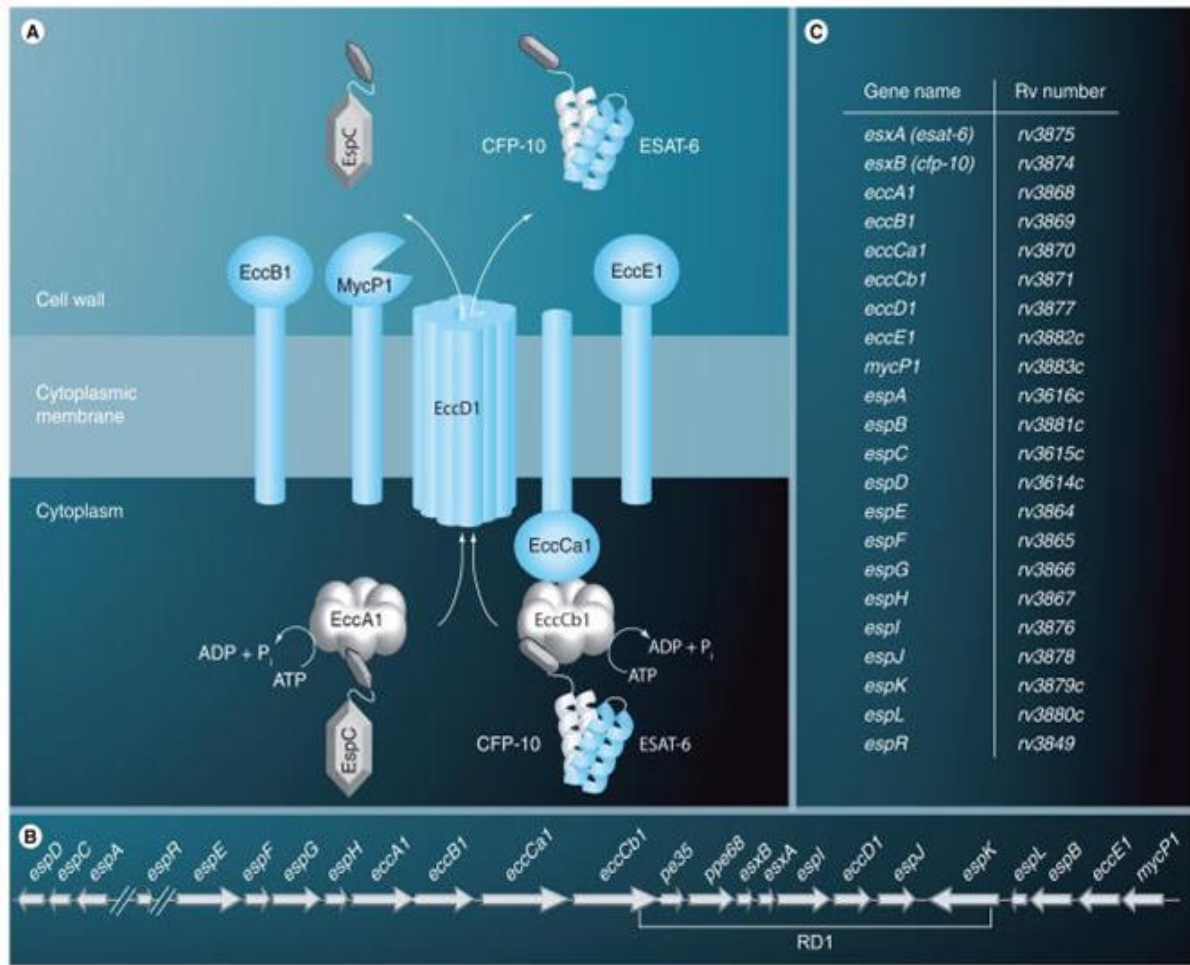
Fig. 2 Schematic Diagram of M. tuberculosis Esx-1 secretion system. A Esx-1 secretion system machinery. B. Gene locus of Esx-1secretion system arrangement in *M. tuberculosis* genome. C. list of Rv numbers of Esx-1 secretion system proteins

The best-studiedEsx system is Esx-1 system and it contains *esxA* and *esxB* genes (spanning genes are *rv3864-rv3883c*) which encode ESAT-6 and culture filtrate protein (CFP-10) respectively (Abdallah, Van Pittius et al. 2007). CFP-10 is an ESAT-6 like protein that is co secreted with ESAT-6. An additional locus involved in ESX-1 secretion is located at a distal site (*rv3616c–3614c*) (Fortune, Jaeger et al. 2005, MacGurn, Raghavan et al. 2005). The Esx-1 system is required for full virulence in MTB (MacGurn, Raghavan et al. 2005, Abdallah, Van Pittius et al. 2007).There are several reported effect of theEsx-1 system on the host; however additional research is required to determine which of these effects accounts for the Esx-1 function in virulence. During macrophage infection with MTB or *M. marinum*, Esx-1 contributes to the process of blocking phagosome maturation (MacGurn and Cox 2007). The esx-3 pathway is essential because of its role in iron acquisition. The Esx-2 and Esx-4 systems have not been directly studied in MTB; however whole genome mutagenesis studied

does not predict a requirement for Esx-2 or Esx-4 for *invitro* growth or virulence (Sassetti and Rubin 2003).Esx-5 system is also not studied in MTB and Esx-5 locus is not predicted to be important for growth or virulence of MTB. These systems export many PE/PPE proteins (Abdallah, Verboom et al. 2006, Abdallah, Verboom et al. 2009). PE/PPE proteins are highly abundant but a less understood family of proteins restricted to Mycobacteria. PE and PPE proteins contain N-terminal domain rich in Pro-Glu (PE) or Pro-Pro-Glu (PPE) repeats connected to variable c-terminal domain giving them their name.

Esx-1 secreted proteins and secretion machinery

All ESX gene cluster in MTB are characterized by a pair of esx genes, encoding homolog's of ESAT-6 and CFP-10 and gene encoding the secretion machinery (Abdallah, Van Pittius et al. 2007, Bitter, Houben et al. 2009). ESAT-6 and CFP-10 best-studied ESX secreted proteins are secreted as a heterodimer (Brodin, de Jonge et al. 2005). The complex is important for secretion because CFP-10, but not ESAT-6, has a C-terminal signal peptide that targets the protein complex to the export machinery (Champion, Stanley et al. 2006). ESAT‑6 (light blue) and CFP‑10 (dark blue) form a tight 1:1 complex (105). The subunit interface mainly contains hydrophobic residues. The carboxyl C-terminal tail of CFP‑10 is indicated and the 7 amino acids that are involved in the secretion signal are shown in pink in Figure 3. The location of the tryptophan-variable-glycine (WXG) motif is indicated for both proteins in red ink figure-3.
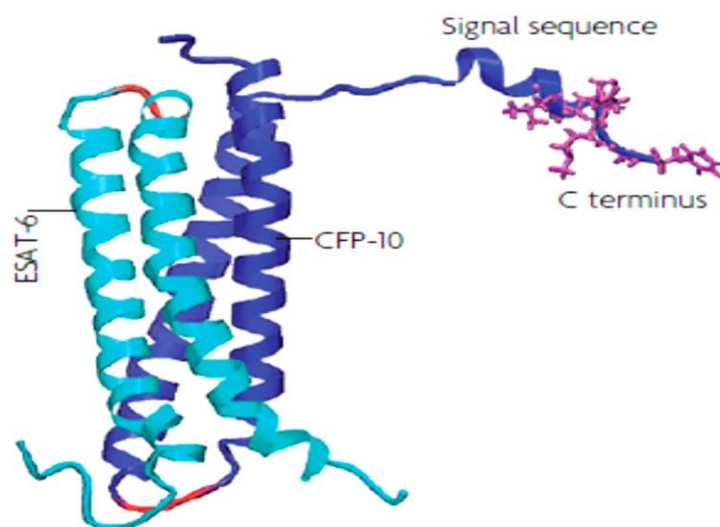
Fig.3 NMR structure of *M. Tuberculosis* proteins ESAT-6 (cyan) and CFP-10 (blue) complex.

The other proteins that are secreted by Esx-1 secretion systems are EspA, EspB, EspC, EspA, EspR(Fortune, Jaeger et al. 2005, McLaughlin, Chon et al. 2007, Raghavan, Manzanillo et al. 2008, DiGiuseppe Champion, Champion et al. 2009).These proteins are not homologous to ESAT-6 and they are larger than ESAT-6. However, like ESAT-6 and CFP-10, these Esx secreted proteins lack N-terminal Sec or Tat signal peptides. EspR is a transcriptional activator that regulates expression of the *espACD* genes, which are required for ESAT-6/CFP-10 export. When the ESX-1 pathway is active, EspR is secreted, resulting in reduced expression of *espACD*. Conversely, when the ESX-1 pathway is inactive, the cytosolic levels of EspR increase, which results in increased transcription of *espACD*(Raghavan, Manzanillo et al. 2008). There is a co-dependent secretion is observed for other ESX-1 secreted proteins (Fortune, Jaeger et al. 2005, Raghavan, Manzanillo et al. 2008). For example, EspA depends on ESAT-6/CFP-10 for its secretion and, conversely, ESAT-6/CFP-10 depends on EspA for secretion (Raghavan, Manzanillo et al. 2008). These co-dependent relationships suggest an interaction prior to secretion and it raises the possibility that the secreted proteins known so far are actually part of the ESX-1 machinery as opposed to being virulence factors that act on the host.

ESX-1 secretion system consist of many component termed EccABCDE and MycP(Pym, Brodin et al. 2002, Gao, Guo et al. 2004, Guinn, Hickey et al. 2004, MacGurn, Raghavan et al. 2005, Brodin, Majlessi et al. 2006, Raghavan, Manzanillo et al. 2008). EccA1 and EccCb1 are predicted cytoplasmic AAA ATPases, suggesting that these proteins supply energy for the secretion process (Luthra, Mahmood et al. 2008). These ATPases target proteins to the pore complex of secretion system. EccCb1 binds a seven amino acid C-terminal signal peptide of CFP-10, which is required for secretion of the ESAT-6/CFP-10 complex (Champion, Stanley et al. 2006). ESAT-6 induces metabolic flux perturbation to drive foamy macrophage differentiation (Singh, Kaur et al. 2015), inhibits TLR signaling in macrophages (Pathak, Basu et al. 2015) and exhibit membrane lysis activity (de Jonge, Pehau-Arnaudet et al. 2007) as well. Similarly, EccA1 binds a C-terminal region of EspC that is required for secretion (DiGiuseppe Champion, Champion et al. 2009).

EccD1 has 10–11 predicted transmembrane domains with short periplasmic and cytoplasmic loops, suggesting a possible role as a channel used for protein translocation across the cytoplasmic membrane. EccCa1 is a predicted integral membrane protein that interacts with the cytosolic ATPase EccCb1 (Stanley, Raghavan et al. 2003). EccB1 and

EccE1 are predicted transmembrane proteins with domains in the periplasm (Krogh, Larsson et al. 2001). EccE1 is shown to interact with EspD, a cytoplasmic protein of unknown function that is required for ESAT-6/CFP-10 secretion (MacGurn, Raghavan et al. 2005). EspB has PE/PPE like homology domain and oligomerizes to form heptameric barrel shaped structure. And EspB has bipartile secretion signal that is recognized by EccCb1 ATPase (Solomonson, Setiaputra et al. 2015). M ycP1 is a serine protease that cleaves EspB signal peptide after export (Ohol, Goetz et al. 2010).

Introduction

Mycobacterium tuberculosis requires type VII ESX-1 secretion system for host cell entry, phagosomal escape and intracellular spread and full virulence within the macrophage. ESX-1 system exports many virulence proteins like ESAT-6, CFP-10, EspR, EspA, EspC and EspB proteins (Finlay and Falkow 1997, Sassetti, Boyd et al. 2001, Organization 2009, Millington, Fortune et al. 2011). The ESX-1 secretion system is encoded by genes mostly located at ESX-1 locus of RD1 region of *Mtb*. The EccA1 (*Rv3868*) and MycP1 (*Rv3883c*) proteins are functional components of theESX-1 system. Deletion of individual genes in ESX-1 locus results in attenuation of *Mtb* as a result of blocked secretion (Sørensen, Nagai et al. 1995, Hsu, Hingley-Wilson et al. 2003, Renshaw, Lightbody et al. 2005, Feltcher, Gunawardena et al. 2015). Although, the genes cluster *espA-espC-espD (Rv3616c-Rv3615c-Rv3614c)* is not linked to theESX-1system, essential for ESX-1 dependent protein secretion. The EspC and EspA are co-secreted with EsxA and EsxB (Fisher, Plikaytis et al. 2002). The distinctive feature of theESX-1 system is the co-dependence of proteins on each other for their secretion (Solomonson, Setiaputra et al. 2015). During secretion, the interactions between substrates are critical for the function of theESX-1 system (Simeone, Bottai et al. 2009, Solomonson, Setiaputra et al. 2015).

EspC gene (Rv3615c) is encoded outside the RD1 locus and also present in BCG (Millington, Fortune et al. 2011). EspC has the Mw of ~11.7 kD in SDS PAGE and similar to ESAT-6 and CFP-10 protein as it is as immunodominant as ESAT-6 and CFP10 and contains both CD4$^+$ and CD8$^+$ epitopes. These epitopes are strongly recognized by T-cells (Millington, Fortune et al. 2011). The N-terminal residues are conserved in ESAT-6 protein family and EspC showed high sequence similarity to CFP-10 and ESAT-6 (Fisher, Plikaytis et al. 2002). The EspC is highly immunogenic similar to ESAT-6 and CFP-10 in patients suffering from active and latent TB infection. ESAT-6, CFP-10, and EspC are the most immunodominant

antigens of *M.tuberculosis* and contain multiple CD4+ epitopes. (Millington, Fortune et al. 2011).

EccA1 (*Rv3868*) is a component of the ESX-1 system and essential for *Mtb* survival in macrophages. EccA1 plays a key role in the secretion of virulence factors in *Mtb*. EccA1 protein exists as a hexamer and its C-terminal domain is involved in the AAA-ATPases activity as well in oligomerization. The exact function of the N-terminal domain of EccA1 is not known. The crystal structure of the N-terminal domain of *Mtb* EccA1 (PDB-4F3V) has been determined recently (Wagner, Evans et al. 2014). The disruption of EccA1 gene prevents the secretion of ESX-1 substrates *e.g.* ESAT-6, CFP-10, EspA, and EspB, however expression of these genes remain unaffected (Schneider 1999, Renshaw, Lightbody et al. 2005, Solomonson, Setiaputra et al. 2015). EccA1 plays a key role in translocation of secreted substrates and transferring energy to theESX-1 system. EspC protein targeted to EccA1 and CFP-10 protein targeted to EccCb1 for their secretion and it has been observed that EspC interacts with cytosolic EccA1 through its C- terminus. Yeast two hybrid analysis also confirmed that deletion of C-terminal residues of EspC abolishes EspC interaction with EccA1 (Champion, Champion et al. 2009).

The objective of our work is the structure and function analysis of *M. tuberculosis* EspC protein and its complex with *M. tuberculosis* EccA1 protein. We have expressed and purified the both EspC and EccA1 proteins using standard chromatographic techniques. Activity assays, SPR binding assays, and Circular Dichorism spectroscopy analysis were performed to investigate the structures and functions of these proteins. Extensive crystallization trials have been performed on both EspC and EccA1 proteins. We have modeled the structures of *M. tuberculosis* EspC and EccA1 proteins. The structural model of theEspC-EccA1 complex was obtained using docking and simulation dynamics techniques.

Specific objectives

1. Preparation of both M. tuberculosis proteins EspC and EccA1

Following techniques were used to achieve current objectives
- Gene cloning and protein expression of EspC and EccA1 in *E. coli.*
- Protein purification of EspC and EccA1 using standard chromatographic techniques.

2. Biochemical and Biophysical characterization of EspC and EccA1 proteins

Following techniques were used to achieve current objectives
- Secondary structure and thermal denaturation analysis using Circular dichorism spectroscopy of both EspC and EccA1 protein.
- Cell lysis, cell viability assay and Co-IP assay for EspC
- ATPase assay for EccA1 both radioactively and calorimetrically.

3. Binding analysis of EspC and EccA1 using Surface Plasmon Resonance technique.

4. Structure analysis of EspC and EccA1 alone and in the complex.
Thesis organization

C**hapter 1** describes the structural and functional analysis of M. tuberculosis EspC protein, EspC was cloned and expressed in *E. coli* expression system. The EspC was purified using standard chromatographic techniques. Cell lysis and Co-IP assays were performed to insight into its functional aspect. Molecular modeling and simulation dynamics analysis were performed on EspC to dissect the structure.

C**hapter 2** describes the structural and functional analysis of M. tuberculosis Ecca1 protein, EccA1 was cloned and expressed in *E. coli* expression system. The EccA1 was purified using standard chromatographic techniques. ATPase assay (Radioactively and calorimetrically) were performed to show the molecule is active. Molecular modeling and simulation dynamics analyses were performed on EccA1to dissect the structure.

C**hapter 3** describes the binding analysis between EspC and EccA1 proteins. Molecular docking and simulation dynamics analysis were performed on theEspC-EccA1 complex to dissect critical amino acids involved in recognition mechanism which is further validated by *in vitro* binding study using SPR technology.

CHAPTER 1: STRUCTURE AND FUNCTION ANALYSIS OF MYCOBACTERIUM TUBERCULOSIS PROTEINS EspC

Materials and Methods

Cloning, Expression, and Purification of *M. tuberculosis* EspC protein

The gene encoding Met1-Thr103 residues of EspC was amplified from *M. tuberculosis H37Rv* strain by polymerase chain reaction and cloned into a*pET21a*(+) expression vector. Following primers were used in PCR reaction to amplify EspC DNA, forward primer (5'- GATCGCTAGCATGACGGAAAACTTGACC-3') and reverse primer (5'-CATGCTC GAGGGTAAACAACCCGTCGAT-3'). The amplified PCR product was digested and ligated into the*pET21a*(+) expression vector using *NdeI* and *XhoI* restriction sites. The EspC plasmid was transformed into *E. coli* BL21 (DE3) cells for protein expression.

The cells were grown in LB media supplemented with 100 µg/ml ampicillin at 37ºC, till OD_{600} reached to 0.6-0.7. The culture was induced with 1mM IPTG at 37ºC and grown further for another 3h. The cells were harvested by centrifugation at 16,000xg and resuspended in lysis buffer containing 20mM HEPES pH 7.5, 300mM NaCl, 1mM Benzamidine-HCl, 5% Glycerol, 2mM β-mercaptoethanol, 1mM Phenylmethylsulfonyl fluoride, and 0.1 mg/ml Lysozyme. The cells were kept on the ice and disrupted by sonication. The lysate was centrifuged at 25,000xg and supernatant was collected.

The supernatant was mixed with pre-equilibrated Ni-NTA resin and incubated 1h at 360° rockers at 4°C. The Ni-NTA column was washed with buffer containing 20mM HEPES pH 7.5, 300mM NaCl, 1mM Benzamidine-HCl, 5% Glycerol, 3mM β-mercaptoethanol, 1mM Phenylmethylsulfonyl fluoride and 40mM Imidazole. Protein was eluted with buffer containing 20mM HEPES pH 7.5, 300mM NaCl, 1mM Benzamidine-HCl, 5% Glycerol, 2mM β-mercaptoethanol, 1mM Phenylmethylsulfonyl fluoride and 250mM Imidazole. The Ni-NTA eluted fractions were analyzed on SDS-PAGE. Fractions containing EspC protein were pooled, concentrated and loaded on gel filtration column (Superdex200 HiLoad 16/60, GE Healthcare). The column was pre-equilibrated with buffer containing 20mM HEPES pH 7.5, 150mM NaCl, 5mM β-mercaptoethanol. Peak fractions were pooled and concentrated to 10-12mg/ml using a 3kD cutoff centrifugal filter device (Millipore, USA).

The purity of recombinant EspC was judged on 12% SDS-PAGE and mass spec analysis (MALDI-TOF). The protein was concentrated and stored at 4ºC. The concentration of protein was measured by using absorbance at 280nm. The final recombinant EspC protein has 112 residues, Mw~11.9kD (extra Met residue at N-terminal, LEHHHHHH residues at C-terminal and 103 residues of EspC).

EspC membrane lysis experiment using human lung carcinoma A549 cells

In 12 well plate, $1x10^5$ A549 cells were cultured in RPMI medium for 24h on cover slips at 37^0C and then incubated with EspC (conc. ~20 µg/ml) (Buffer 20mM HEPES pH7.5, 50mM sodium chloride, 5% Glycerol, 0.01% SDS, 0.01% Tween20)for 2h, 4h and 6h at 37^0C. A control experiment was also performed with EspC buffer since it contains detergents (0.01% SDS, 0.01% Tween20 simultaneously) The experimental results were examined in inverted bright field microscope (Olympus) at 100x and 200x zoom at every time interval to examine the change the cell morphology and numbers with incubation time. At 2h, 4h and 6h, the A549 cells were washed three times with 1xPBS buffer (20mM potassium phosphate pH 7.4, 150mM Sodium chloride) for 5 minutes and fixed in 1xPBS buffer containing 4% formaldehyde for 30 minutes at 25^0C. Afterwards, the cells were permeabilized in 1xPBS buffer containing 0.3% Triton X-100 for 15 minutes at 25^0C and further blocked in 5%BSA blocking buffer (1x PBS buffer + 0.3% Triton X-100 + 5% Bovine Serum Albumin) for 1h.

The cells were incubated with anti-His antibody (1:100 ratio) in PBST buffer (1x PBS buffer + 0.3% Triton X-100 + 1% BSA) for 24h at 4°C. Next day, cover slips were washed three times with buffer (0.1% Triton X-100 + 1x PBS buffer) and cells were again incubated with *Alexafluor* 568 secondary antibody (1:250 ratios) for 2h. After incubation, cells were washed with PBST followed by high salt-PBS buffer (23.28 g per liter NaCl in 1x PBS) and then with PBST buffer. The cells were incubated with DAPI (conc. ~1mg/ml) for 10 minutes and further washed with high salt-PBS buffer followed by awash with PBST buffer. The cover slip was mounted on a glass slide with 10% glycerol. The edges of cover slips were covered with nail paint to avoid drying. Stained cells were imaged with Nikon confocal microscope at central instrument facility of JNU.

Cell Viability by Trypan blue dye exclusion assay

1x 10^5 A549 cells per well were seeded in six well plate (*Corning*) in regular growth culture medium for overnight. Next day, cells were incubated with incomplete growth medium as control and 20µg/ml EspC protein for 6h. After 6h, media were collected in Falcon tube and adherent cells were collected by brief trypsinization. Cells were pellet down by centrifugation at 2000 rpm for 5 min and resuspended in 1x PBS (20mM potassium phosphate pH 7.4, 150mM Sodium chloride) at 4^0C and processed for cell counting. 100µL of resuspended cells were taken into a new *Eppendorf* tube and mixed with 400µL of 0.4% Trypan Blue (final concentration 0.32%). Live and dead cells were counted with haemocytometer (cell counting chamber) using trypan blue staining under an inverted phase contrast microscope (*Olympus*).

Co-Immunoprecipitation and Mass Spectrometric analysis of peptide mixtures

PMA-differentiated THP1 cells (10×10^6 cells) were incubated with EspC protein (20 µg/ml) for 6 h at 37 ^{0}C. The cells were washed 3 times with ice chilled 1X PBS buffer (20mM potassium phosphate pH 7.4, 150mM Sodium chloride) to remove any trace of extracellular EspC protein. The Co-Immunoprecipitation experiment was performed by using Pierce C0 – Immunoprecipitation (Co – IP) kit. A mock Co-IP was also done without the EspC to identify non-specific beads interactors. The Co-IP sample was lyophilized and kept in -20^0C and then re-suspended in100mM ammonium bicarbonate buffer (100µl). The proteins were then subjected to denaturation (100µl of 6M Urea buffer in 100mM ammonium bicarbonate). And then10mM of the DTT reducing reagent was added and vortex the sample. Incubate the mixture for 60 min at room temperature followed by 40mMof the iodoacetamide alkylating reagent addition and vortex the sample. Incubate the mixture for 60 min at room temperature.Again10mMof the DTT reducing agent was added. Vortex the sample and allow the reaction to stand at room temperature for 30-60 min. Urea concentration was reduced by diluting the reaction mixture with 775µl MilliQ-H20. Trypsin digestion was done overnight at 37^0C (1:20 ratio)reaction was stopped and pH (<6) was adjusted by addition of Formic acid. Peptides were purified by the ZIP-TIP method and elutions were dried down completely in a speed–vac. Peptides were resuspended in a solution of 98% MilliQ-H20, 2% CH3CN, 0.1% FA.

The peptide mixture was cleaned using C18 columns. All the experiments were performed using EASY-nano-LC 1000 system (Thermo Fisher Scientific) coupled to QExactive mass spectrometer (Thermo Fisher Scientific) equipped with nanoelectrospray ion source.1 µg of the peptide mixture was resolved using 15 cm PicoFrit column (360 um OD, 75 um ID, 10 um tip) filled with 1.8µm C18-resin (Dr. Maeisch). The peptides were loaded with buffer A and eluted with a 0-40% gradient of Buffer B (95% acetonitrile/0.1% formic acid) at a flow rate of 300nl/min for 60 min. The QExactive was operated using the Top10 HCD data-dependent acquisition mode with a full scan resolution of 70,000 at m/z 400. MS/MS scans were acquired at a resolution of 17500 at m/z 400. Lock mass option was enabled for polydimethylcyclosiloxane (PCM) ions (m/z = 445.120025) for internal recalibration during the run.

The RAW files generated were analyzed with Proteome Discoverer software against the Homo sapiens UniProt reference proteome database (*www.uniprot.org*). For Sequence search, the precursor and fragment mass tolerances were set at 10ppm and 0.6Da, respectively. The protease used to generate peptides, i.e. enzyme specificity was set for trypsin/P (cleavage at the C terminus of "K/R: unless followed by "P") along with maximum missed cleavages value of two. Carbamidomethyl on cysteine as fixed modification and oxidation of methionine and N-terminal acetylation were considered as variable modifications for database search. Percolator node was used and both peptide spectrum match and protein false discovery rate were set to 1%.

Circular dichroism and thermal denaturation analysis

CD measurements were recorded using Chirascan^{TM} CD spectropolarimeter (*Applied Photophysics)* at AIRF JNU NEW DELHI) with a water bath to maintain the constant temperature (25^{0}C). The EspC protein was diluted to 0.1 mg/ml in 10mM potassium phosphate buffer pH 8.0 by dialysis and centricon (3KD cutoff) and loaded in 0.1 cm quartz cuvette at 25^{0}C. The blank of all experiments was 10mM potassium phosphate buffer pH 8.0. The final spectrum was an average of three sequential scans. All CD data were converted to mean residue ellipticity (deg.cm^2/dmol).

The Dichroweb server (Whitmore and Wallace 2008) was used to estimate the secondary structures from CD spectra. The SOPMA (Geourjon and Deleage 1995), GOR (Sen, Jernigan et al. 2005) and PSIPRED programs were used to estimate theoretical

secondary structure for EspC. To study the thermal stability of EspC, CD-spectra were recorded as a function of temperature. For EspC, the CD spectra were recorded from 25^0C to 85^0C in 10^0C increment.

EspC Crystallization alone and in complex with EccA1

Extensive crystallization trials were performed using different concentrations (2mg/ml, 5mg/ml, 10mg/ml and 20mg/ml) of EspC protein. Crystal Screen, Crystal Screen 2, PEG/ION and Index screens from Hampton Research, PACT premiere, JCSG, Morpheus, structure screen 1 and 2 and Stura foot print screens from the Molecular dimension and Nuc-Pro screen from Jena Biosciences were used to search the initial crystallization condition of EspC. The hanging and sitting-drop both vapor-diffusion techniques were used for crystallization experiments. In each trial, the1µl protein solution was mixed with 1µl reservoir solution and equilibrated against 100µl reservoir solution at 4°C.

Molecular Modeling of *M. tuberculosis* EspC protein

The amino acids sequence of *M. tuberculosis* EspC (gene id-885770) was retrieved from *NCBI d*atabase. PSIPRED (Jones 1999) program was used to predict the secondary structure of EspC. The EspC monomer was obtained by PHYRE2 server (Kelley, Mezulis et al. 2015) using WxG100 protein structure from *S. agalactia* (PDB-3GWK) as input and built as homodimer using the same template by COOT program(Emsley and Cowtan 2004). The EspC hexamer is obtained by ROSETTADOCK symmetry-docking server (Lyskov and Gray 2008) using EspC homodimer as input. Energy minimization was performed on EspC homodimer using GROMACS program (Hess, Kutzner et al. 2008). PROCHECK program (Laskowski, MacArthur et al. 1993) was used to check the quality of EspC model and ANOLEA program (Jones 1999) to calculate the conformational energy. The EspC modeled structure was visualized by PyMol(DeLano 2002) program.

Results and Discussion

EspC protein expression and purification

The EspC gene (Rv3615c, 312bp long) was successfully cloned into *pET21a(+)* plasmid vector with C-terminal 6x Histidine affinity tag (Fig. 4A). EspC protein was expressed in *E. coli* BL21(DE3) cells and expression was confirmed by MALDI-TOF mass spectrometry (Fig. 4B). The EspC protein expressed in soluble fraction of the cell and initially purified using Ni-NTA affinity chromatography followed by size exclusion chromatography (Superdex 200 16/60 column *GE Healthcare*) were initially it came into void volume of the column. Dynamic light scattering and small angle X-ray scattering study indicated the presence of aggregates in EspC protein (Data not shown). Several additives, for example,NaCl, $MgCl_2$, KCl, $CaCl_2$, Glucose, Sorbitol, Xylose,Sucrose, Glycerol, Arginine, and Urea were tried in buffer system to see if they could prevent aggregation of EspC protein. However, treatment with different additives has prevented EspC aggregation and improved its solubility. Addition of 0.01 % (w/v) Tween-20 and 0.01% (w/v) SDS in lysis buffer (20mM HEPES pH7.5, 50mM NaCl, 3mM βME and 5% Glycerol) prevented the EspC aggregation and protein eluted as a single peak which corresponds to hexamer on Superdex200 (16/60) column(Fig. 5B).
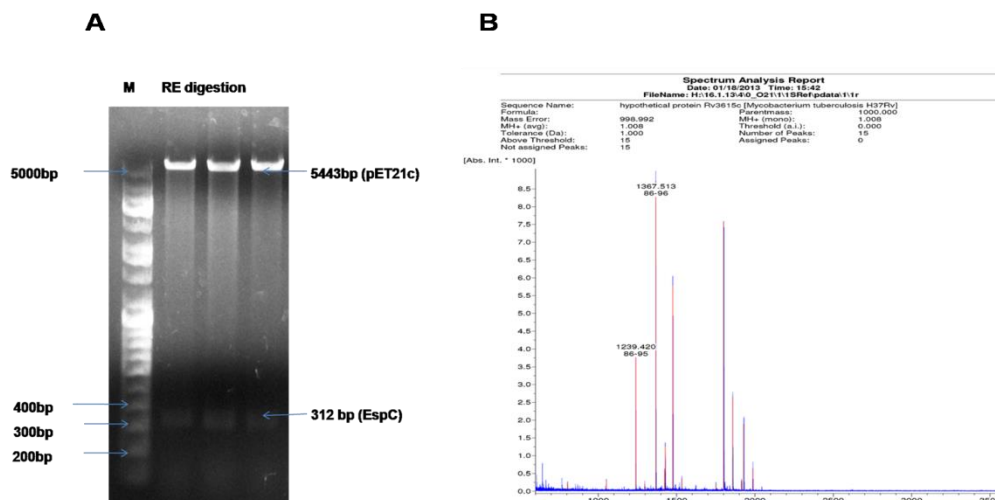
Fig.4 A. Agarose gel showing restriction digestion of EspC plasmid to confirm clone (M-DNA base pair marker). B. MALDI-TOF spectra confirming *M, tuberculosis* EspC protein expression.
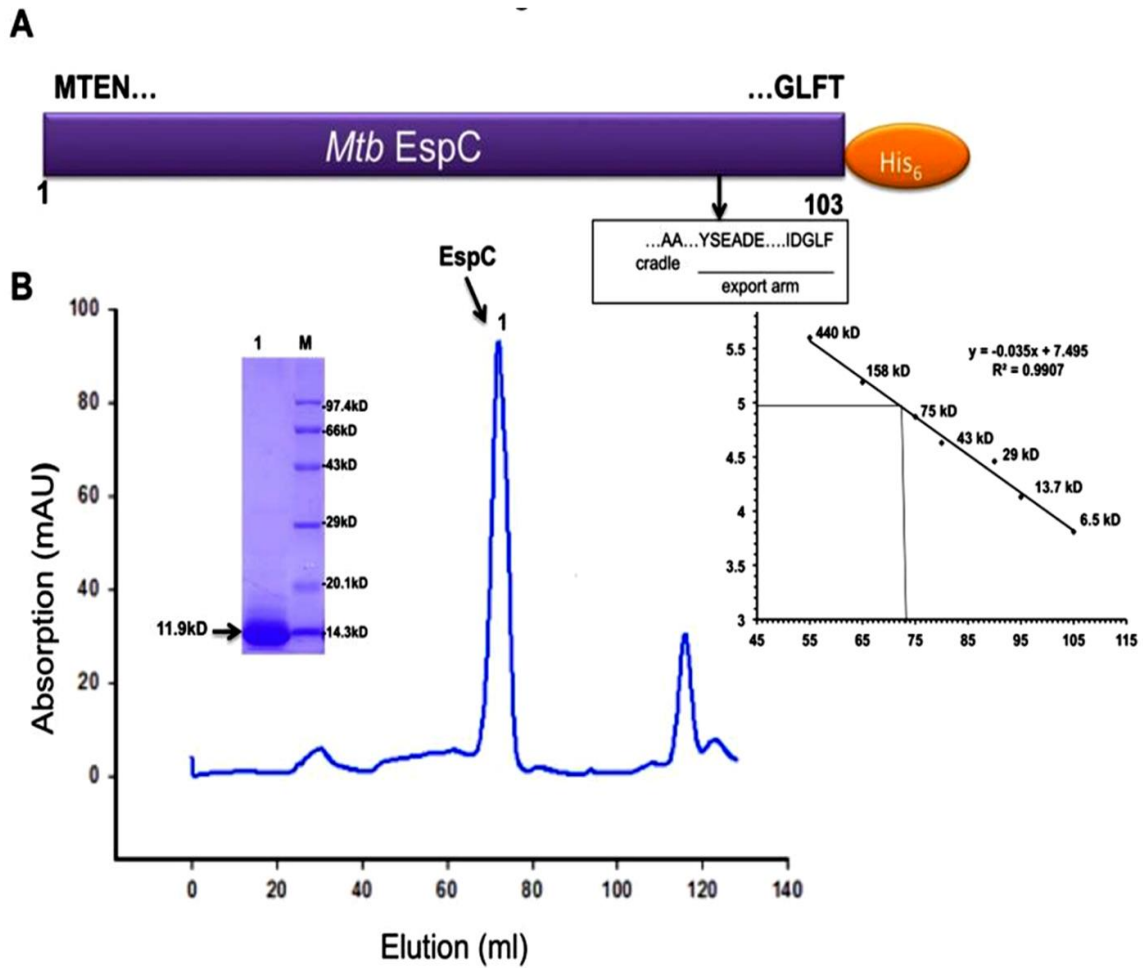
Fig.5A. Schematic view of an expression construct of EspC. The EspC export arm consists of AA cradle, YSEADE, and IDGLF motifs. B. Size exclusion chromatography and SDS-PAGE analysis of purified EspC. EspC eluted as Hexamer from Superdex200 (16/60) column and showed a single band on SDS-PAGE.

EspC exhibits membrane lysis activity

In a previous study(de Jonge, Pehau-Arnaudet et al. 2007), it is observed that ESX-1 secreted substrates form a multimer, involved in membrane pore formation and modulate the extracellular signaling pathway. To examine the EspC lysis activity, we used human lung carcinoma A549 cells and results were examined by live cell imaging and confocal microscopy. The A549 (Human lung epithelial carcinoma cell line) cells were incubated with EspC (conc.~20 µg/ml) for different time intervals (2h, 4h, and 6h). We observed that viability of A549 cells has decreased after 6h incubation with EspC (Fig. 6A).
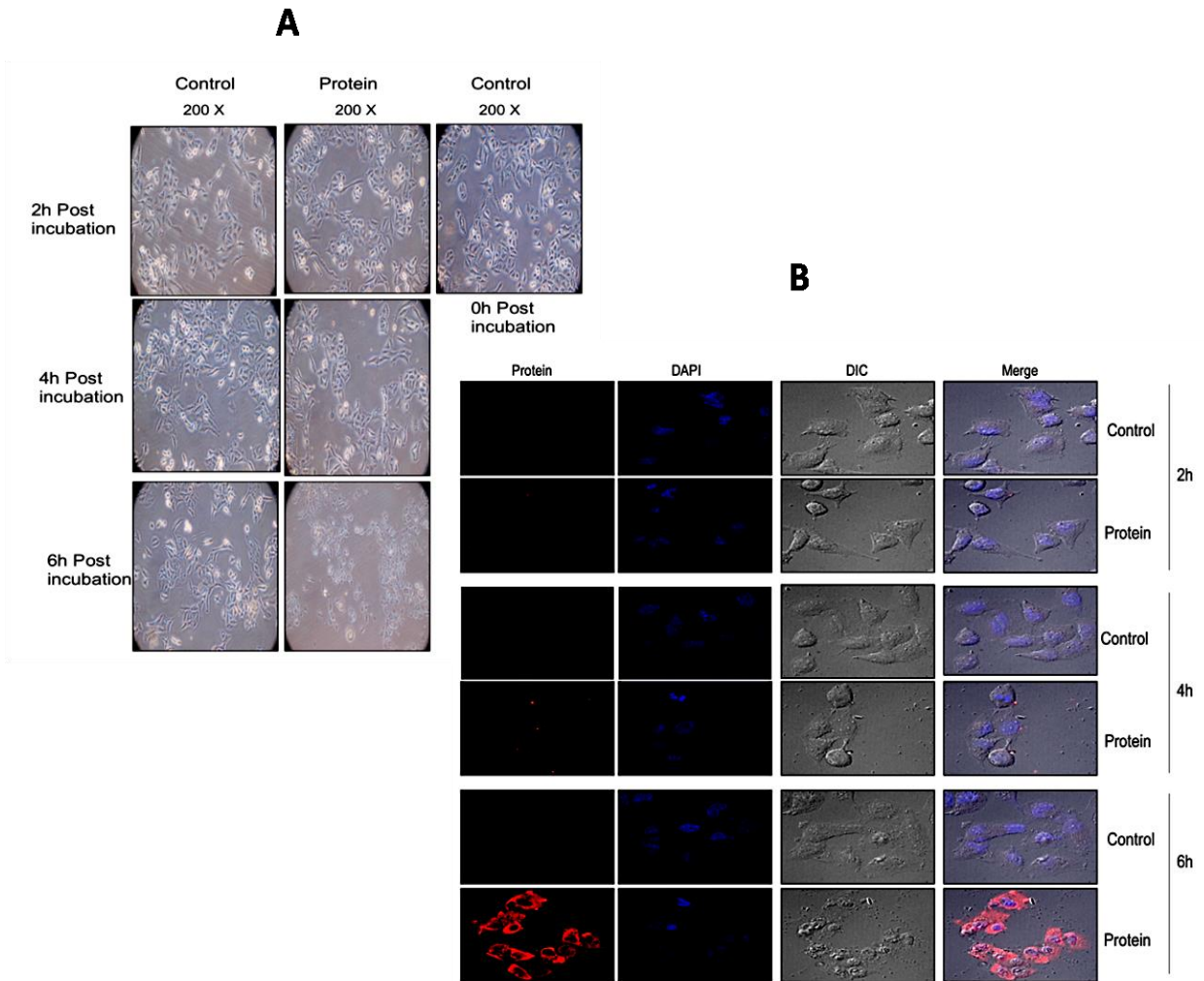
Fig.6. EspC membrane lysis and cellular localization experiments using human lung carcinoma A549 cells. A. A549 cells were incubated with EspC for 0h, 2h, 4h, and 6h. Significant changes are observed in A549 cells morphology and reduction in a number of cells after 6h of incubation with EspC. B. A549 cells were processed for immune-staining to observe the EspC localization at different time intervals. Incubation of EspC with A549 cells has caused deformed morphology, reduction in a number of cells and EspC localization in the cytoplasm after 6h (red fluorescence). Images were captured at 200x magnification using 60x Nikon confocal microscope.

The EspC induced the morphological changes in A549 cells and its plasma membrane was severely damaged. We also examined the EspC localization in A549 cell at 2h, 4h, and 6h using confocal microscopy. The confocal images have shown that EspC is localized in cytosol after 6h (Fig. 6B). These data indicate that EspC interacts and lysed the plasma membrane of A549 cell. After lysis, EspC has entered into the cytosol of A549 cell and caused the cell death.
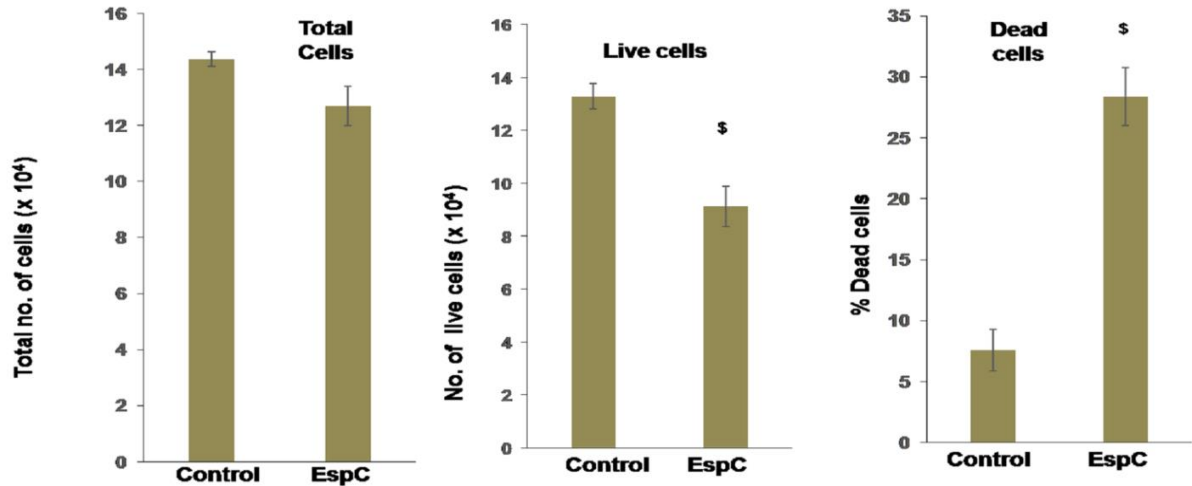
Fig.7.   A549 Cell viability assay using Trypan blue dye with EspC protein and without EspC protein (control).

After 6h post incubation of A549 cells, there was a significant change observed in viable cell count through trypan blue dye exclusion assay (Fig.7).

EspC interacts with various Macrophage cytosolic proteins

Since EspC is a mycobacterial Esx-1 secretory protein released into the cytosol of infected macrophages. To decipher the host interactome of EspC, THP-1 macrophages were incubated with rEspC and after a gentle lysis of the cells, the His_6- tagged rEspC and any interacting protein(s) were co–immunoprecipitated using Pierce CO– Immunoprecipitation (Co–IP) kit. The Co- Immunoprecipitation elutionfractions were subjected to denaturation, reduction, alkylation and trypsin digestion followed by EASY-nLC 1000 system (Thermo Fisher Scientific) coupled to a QExactive mass spectrometer (Thermo Fisher Scientific) equipped with nanoelectrospray ion source mass spectrometric analysis. The various macrophage proteins that co–immunoprecipitated with rEspC are listed in (Table 1). Many of these proteins are Cytoskeleton maintaining proteins, protease inhibitor proteins, calcium binding proteins, chemokine receptors binding proteins and with different enzymatic activities having cytosolic (for example, ADP-ribosylation factor 3, Cornulin,Involucrin,SerpinB4, Aldehyde dehydrogenase, dimeric NADP-preferring, Cystatin-B etc), cell junctions (For example; Periplakin Protein, S100-A-14 etc), extracellular exosome (For example,cornulin, Alpha-2-macroglobulin-like protein 1 etc) and membrane (For e.g.Periplakin, Neuroblast differentiation-associated protein AHNAK

20

etc) localization. To validate our findings we performed the mock (rEspC$^-$) and EspC (rEspC$^+$) Co-IP.

Table 1. EspC interactome List of proteins that co-immunoprecipitated with rEspC

No.	Accession	Protein name	Σ Coverage	Σ# Unique Peptides
1	P12035	Keratin, type II cytoskeleton	55.41	19
2	Q9UBG3	Cornulin	55.35	19
3	P07476	Involucrin	28.21	13
4	P04080	Cystatin-B	76.53	5
5	O60437	Periplakin	11.62	18
6	P48594	Serpin B4	36.67	2
7	Q9UBC9	Small proline-rich protein 3	37.87	7
8	Q09666	Neuroblast differentiation-associated protein AHNAK	11.39	12
9	A8K2U0	Alpha-2-macroglobulin-like protein 1	11.04	13
10	Q92817	Envoplakin	4.97	9
11	P35321	Cornifin-A	61.80	1
12	P27482	Calmodulin-like protein 3	31.54	3
13	P02814	Submaxillary gland androgen-regulated protein 3B	65.82	2
14	P30838	Aldehyde dehydrogenase, dimeric NADP-preferring	9.93	4

The interaction between EspC and various cytosolic proteins indeed specific as these proteins were detected only in the EspC Co-IP fraction but not in the mock Co-IP fraction. Since, we were using extracellular EspC protein in our study, to be sure that the protein was entering the macrophages and was indeed cytosolic, we performed antibody staining (anti-EspC) in A549 cells that had been treated with rEspC and analyzed the cells by confocal microscopy. We could detect intracellular EspC (Fig. 6B) in A549 cells and

further validation was provided by MS- based detection of EspC in the rEspC Co-IP fraction using THP-1 cells (not shown).

Secondary structure and thermal stability analysis of EspC

Far-UV CD spectra (260-200 nm) were obtained for EspC to estimate the contents of the secondary structure in protein. The Dichrowebserver[27] analysis yielded (~80% ☐☐helix and ~20% random coil) for EspC (Fig. 8A) These values were close to secondary structures obtained from theoretical (~79.2 ☐-helix and 21.7% random coil) analysis forEspC protein.
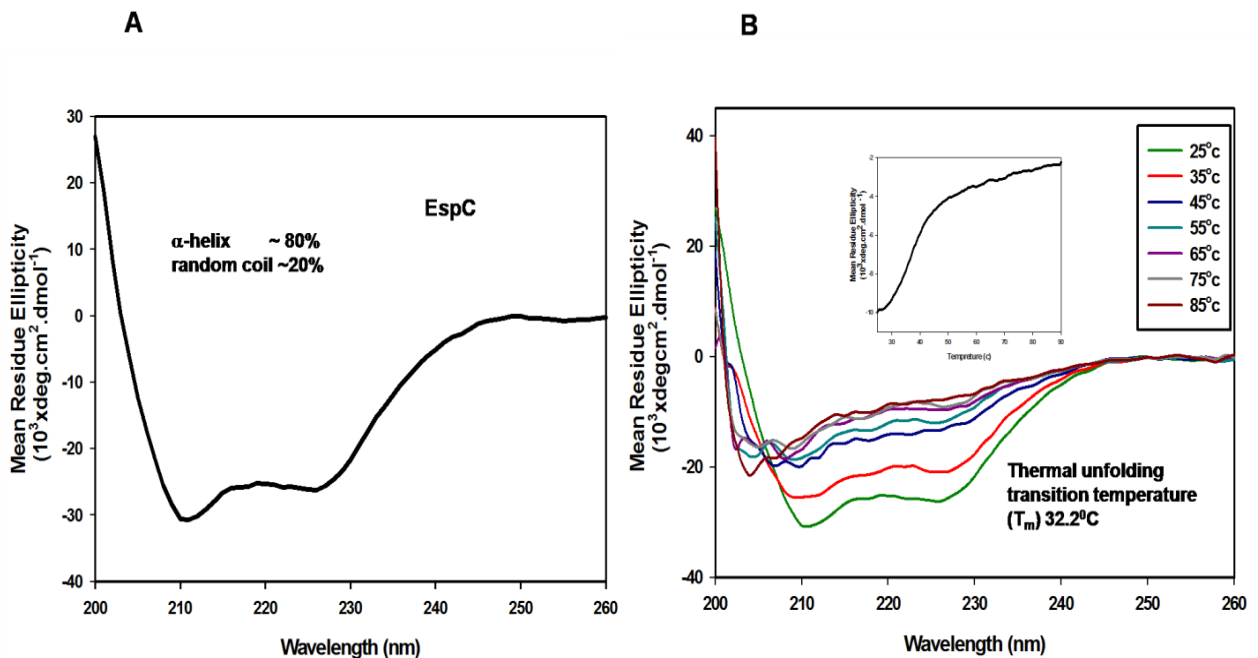
Fig.8. Circular dichorism and thermal stability profile of EspC. A. Circular dichorism and B. thermal stability profile of EspC. Spectra were recorded at room temperature and normalized to mean residue ellipticity. The secondary structure contents of matches well with secondary structural contents obtained from theoretical predictions. EspC protein is thermally unstable.

For thermal stability analysis, the CD spectra of EspC and was recorded as a function of temperature. Effect of temperature on EspC secondary structure was performed from 25^0C to 85^0C in 10^0C temperature increment (Fig. 8B) and yielded thermal unfolding transition temperature (T_m) 32.2 ± 1.0^0C which is indicating that MtbEspC protein is thermally unstable.

Crystallization of EspC and EspC-EccA1 complex

Crystallization attempts were made on purified EspC protein using numbers of screens *e.g.* PEG/Ion, Index, SaltRx, Crystal Screen2 from Hampton research, PCAT premier, JCSG, Morpheus, stura foot print, Structure screen1 and 2 screens from the Molecular dimension and Nuc-Pro screen from Jena Bioscience. Different concentrations (2mg/ml, 5mg/ml, 10mg/ml, and 20mg/ml) of MtbEspC protein at different temperatures (4^0C, 16^0C and 22^0C) were tried for crystallization. Hanging drop and sitting drop vapors diffusion methods were used in crystallization experiment. We have tried to crystallize MtbEspC protein in complex with Mtb EccA1 protein in different molar ration but failed to produce either EspC crystals or EspC-EccA1 complex crystals.

Molecular structural modeling of M. tuberculosis EspC

The *M. tuberculosis* EspC model was generated by PHYRE2 server (Kelley, Mezulis et al. 2015) using WxG100 protein structure from *S. agalactia* (PDB-3GWK) as input template and resembles with secondary structure prediction obtained by PSIPRED program (Jones 1999). *M. tuberculosis* EspC belongs to WxG100 family where all the proteins possess small helix turn helix structure. Sequence analysis of EspC using PSIPRED has indicated that EspC is well-ordered protein (Fig.9a). The EspC forms helix-turn-helix structure and composed of two helices having 36 and 56 residues (Fig.9b).

Fig. 9 A. Secondary structure prediction of *M. tuberculosis* EspC protein by PSIPRED program B. Molecular model of EspC monomer showing helix-turn-helix structure. C. Sequence alignment of the export arm of EspC with other ESX-1 virulence proteins.

EspC homodimer model (Fig.10A) has indicated that EspC molecules have a tendency to form a homodimer (interacts anti-paralleled way) through several hydrophobic interactions between the first helix and similarly to form hexamer (Fig.10B).

Fig.10. A. *M. tuberculosis*EspC homodimer model (In red-C-terminal *Yxxxd* motif). B. *M. tuberculosis*EspC hexamer model.

CHAPTER 2: STRUCTURE AND FUNCTION ANALYSIS OF MYCOBACTERIUM TUBERCULOSIS PROTEINS ECCA1

Materials and methods

Cloning, Expression, and Purification of *M. tuberculosis* EccA1 protein

The EccA1gene (Rv3868) encoding (Met1-Glu573) DNA was amplified from *M. tuberculosis H37Rv* strain by PCR and cloned into a*pET23a*(+) expression vector with C-terminal 6X Histidine affinity tag (Luthra, Mahmood et al. 2008). The EccA1plasmid was transformed into *E. coli.* BL21(DE3) cell and culture were grown at 37°C in LB media supplemented with 100 µg/ml ampicillin at 37°C, till OD_{600} reached to 0.6-0.7. The culture was induced with 0.5mM IPTG at 22°C and grown further for another 10h. The $EccA_1$ was over expressed in soluble fraction of cell. The cells were harvested by centrifugation at 16,000xg and suspended in lysis buffer containing [20mM HEPES pH 7.5, 150mM NaCl, 5% Glycerol, 4mM β-mercaptoethanol, 1mM phenylmethylsulfonyl fluoride, 4mM benzamidine-hexachloride, 0.2mM $MgCl_2$, 0.2mM ATP, 10mM Arginine and 0.2mg/ml Lysozyme]. The cell lysate was centrifuged at 25,000xg and supernatant was collected.

The supernatant was mixed with pre-equilibrated Ni-NTA resin (Sigma) and incubated 1h at 360° rocker at 4°C. The Ni-NTA column was washed with buffer containing [20mM HEPES pH 7.5, 150mM NaCl, 5% glycerol, 4mM β-mercaptoethanol, 1mM phenylmethylsulfonylfluoride, 3mM benzamidine-hexachloride, 10mM Arginine, 0.2mM $MgCl_2$, 0.2mM ATP and 40mM Imidazole]. The $EccA_1$ was eluted in buffer containing [20mM HEPES pH 7.5, 150mM NaCl, 5% Glycerol, 4mM β-mercaptoethanol, 1mM phenylmethylsulfonyl fluoride, 3mM benzamidine hydrochloride, 10mM Arginine, 0.2mM $MgCl_2$, 0.2mM ATP, and 250mM Imidazole]. The Ni-NTA eluted fractions were pooled, concentrated and loaded on Superdex^{TM}200 HiLoad (16/60) column (GE Healthcare). The column was pre-equilibrated with buffer containing [20mM HEPES pH 7.5, 150mM NaCl, 5% Glycerol, 1mM phenylmethylsulfonyl fluoride, 10mM Arginine, 0.2mM $MgCl_2$, 0.2mM ATP, 4mM β-mercaptoethanol]. The peak fractions were pooled and concentrated using a 30kD cutoff ultracentrifugal device (Millipore, USA). The purity of $EccA_1$ was analyzed by 12% SDS-PAGE and mass spectrometry. The concentration of protein was measured by absorbance at 280nm.

EccA$_1$ ATPase activity

To analyze the ATPase activity, the EccA$_1$ was dissolved in ATPase buffer containing 20mM HEPES pH 7.5, 5mM MgCl$_2$. 1µl (0.1µCi) of (γ^{32}P) ATP was added in ATPase buffer containing (1.11g of) EccA$_1$ (in a 10µl reaction) and was incubated for different time intervals at 37^{0}C. 1µl of reaction buffer was spotted on TLC plate after every 10 minutes. The TLC plate was developed in 0.5 M formic acid and 0.5 M LiCl and dried at 37^{0}C. The TLC plate was exposed to Fuji film BAS-MS 2025 imaging plate for 12h and analyzed using GE Healthcare Typhoon FLA 9500. The background obtained from reaction buffer containing no EccA$_1$ and Obtained from reaction buffer containing no(γ^{32}P) ATP was corrected in each ATP hydrolysis measurement.

A colorimetric assay was performed to analyze the EccA$_1$ ATPase activity by using ATPase assay kit (InnovaBiosciences). The ATPase assay was performed in buffer containing [20mM HEPES pH 7.5, 150mM NaCl, 5% Glycerol, 2mM β-mercaptoethanol, 1mM ATP and 1□g of EccA$_1$]. The reaction was carried out at 25°C for 5 min. The dye buffer containing [120mM malachite green, 0.06% polyvinylalcohol, 6mM ammonium heptamolybdate, 4.2% sodium citrate] was added in reaction buffer. After incubating for 15 min, 10□l of each reaction mixture was transferred to 96 well plates and the absorbance at 630nm was measured. Absorbance from reaction buffer containing no EccA$_1$ and without ATP was subtracted from each experimental data. The release of inorganic phosphate was analyzed based on absorbance from phosphate standard curve. Each assay was performed three times and average activity is calculated. The values of kinetic parameters *K$_m$ a*nd *V$_{max}$* were calculated using Prism 6.0 software (*Graph Pad Software Inc.*).

Circular dichorism and thermal denaturation analysis

CD measurements were recorded using Chirascan^{TM} CD spectropolarimeter (Applied Photophysics) with a water bath to maintain the constant temperature at 25^{0}C. The EccA1 protein was diluted to 0.1 mg/ml in 10mM potassium phosphate buffer pH 8.0 and loaded in 0.1 cm quartz cuvette. The blank of all experiments was 10mM potassium phosphate buffer pH 8.0 and subtracted from experimental data. The final spectrum was an average of three sequential scans. All CD data were converted to mean residue ellipticity (deg.cm^2/dmol).

The Dichroweb server (Whitmore and Wallace 2008) was used to estimate the secondary structures from CD spectra. The SOPMA (Geourjon and Deleage 1995), GOR (Sen, Jernigan et al. 2005) and PSIPRED (Jones 1999) programs were used to estimate theoretical secondary structure for EccA1 and To study the thermal stability of the $EccA_1$ protein, CD-spectra were recorded as a function of temperature. Analysis of the effect of temperature on the secondary structure was performed from 20^0C to 90^0C in 10^0C temperature increment for EccA1.

EccA1 Crystallization

Extensive crystallization trials were performed on Mtb EccA1 protein using different concentrations (5mg/ml, 8mg/ml and 10mg/ml). Crystal Screen, Crystal Screen 2, PEG/ION and Index screens from Hampton Research, PACT premiere, JCSG, Stura foot print, Morpheus screens from Molecular dimension were used to search the initial crystallization condition of EccA1. The hanging and sitting-drop both vapor-diffusion techniques were used for crystallization experiments. Glutaraldehyde Crosslinking, Reductive methylation of EccA1 protein was also performed to get EccA1 crystals. In each trial, the 0.4μl protein solution was mixed with 0.4μl reservoir solution and equilibrated against 100μl reservoir solution in 96 well crystallization plates at 4^0C and 16^0C.

Molecular modeling of *M. tuberculosis* EccA1 protein

The three dimensional model of *M. tuberculosis* EccA1 protein (573 residues, *gene Id*-886199) was built by comparative modeling strategy using following templates (i) X-ray structures of N- terminal domain of *M. tuberculosis* EccA1 (PDB-4F3V) (ii) crystal structure of Rubisco activase protein (PDB-3SYL). The PSIPRED server (Jones 1999)was used for predicting secondary structures present in theEccA1 protein. Alignment of query and template protein sequences were done using CLUSTALW server (Thompson, Gibson et al. 2002). The Modeler program (v. 9.11) was used and yielded five models of *Mtb*EccA1. The best model was selected based on Z- score. The $EccA_1$ hexamer was obtained using COOT program based on crystal structure of p97 hexamer (PDB-1E32). PROSA server was used to check the fitness of sequence into the model. Geometric inaccuracies in the models were evaluated by PDBsum validation server (Laskowski 2001). PROCHECK program was used to evaluate the stereochemical quality of the modeled structure (Laskowski, MacArthur et al. 1993).

Energy minimization was performed on EccA1 hexamer using GROMACS program (Hess, Kutzner et al. 2008). The EccA1 hexamer was put in a cubic box extending to 1 nm from protein surface and solvated with explicit simple point charge (SPC216) water molecules. GROMACS96 (43a2) force field was used. The system was electro-neutralized by adding 23Na$^+$ ions to the system. The final system consists of 6,250 protein atoms and 135,306 water molecules. The potential energy of the final system was -7.433x10^6 KJ/mol. PROCHECK program was used to check the quality of EccA$_1$ models and ANOLEA program to calculate the conformational energy.

Results and Discussion.

EccA1 cloning, protein expression and purification

The schematic view of EccA1 (56kDa, 573 residues) is shown in Fig. 11A. EccA1was cloned in pET23a (+) vector and over expressed in *E. coli* BL21(DE3) cell and expression were confirmed by Mass spectrometry (Fig.10).
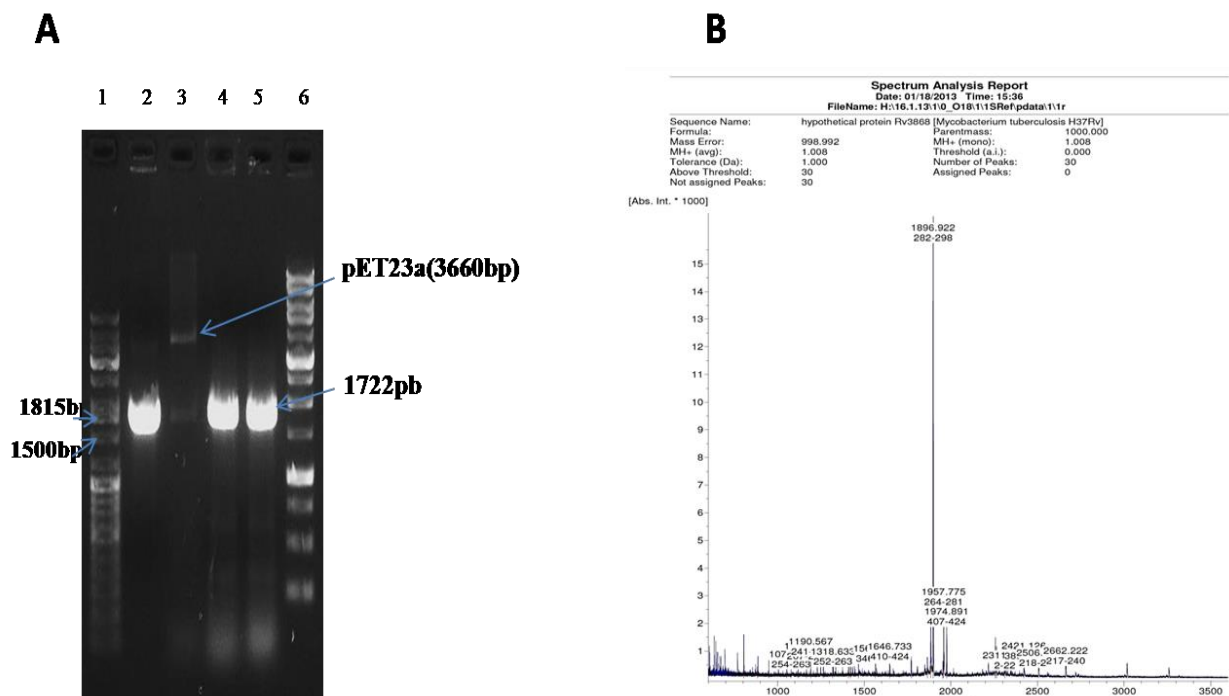
Fig. 10A. Agarose gel showing *M. tuberculosis* EccA1 clone confirmation(Lane 1 and 6-DNA base pair marker, Lane 2-PCR amplified EccA1 gene from the genome, Lane 3-double restriction digestion *Rv3868-pET23a* plasmid, Lane 4 and 5-PCR amplified product using *Rv3868-pET23a* plasmid as a template). B. MALDI-TOF spectra confirming *M. tuberculosis* EccA1 protein expression.

EccA1 was purified using Ni-NTA and size exclusion chromatography (Fig. 11B). The EccA1 eluted as a hexamer from Superdex200 (16/60) column and appeared as a single band on SDS-PAGE. Initially, EccA1 eluted in avoid volume of Superdex200 (16/60) column, however, the addition of 10mM Arginine, 0.2mM $MgCl_2$ and 0.2mM ATP in protein buffer prevented the EccA1 oligomerization and protein eluted as a hexamer. In an earlier study (25), higher ordered oligomers of EccA1 are observed and they form multiple of $EccA_1$hexamer.
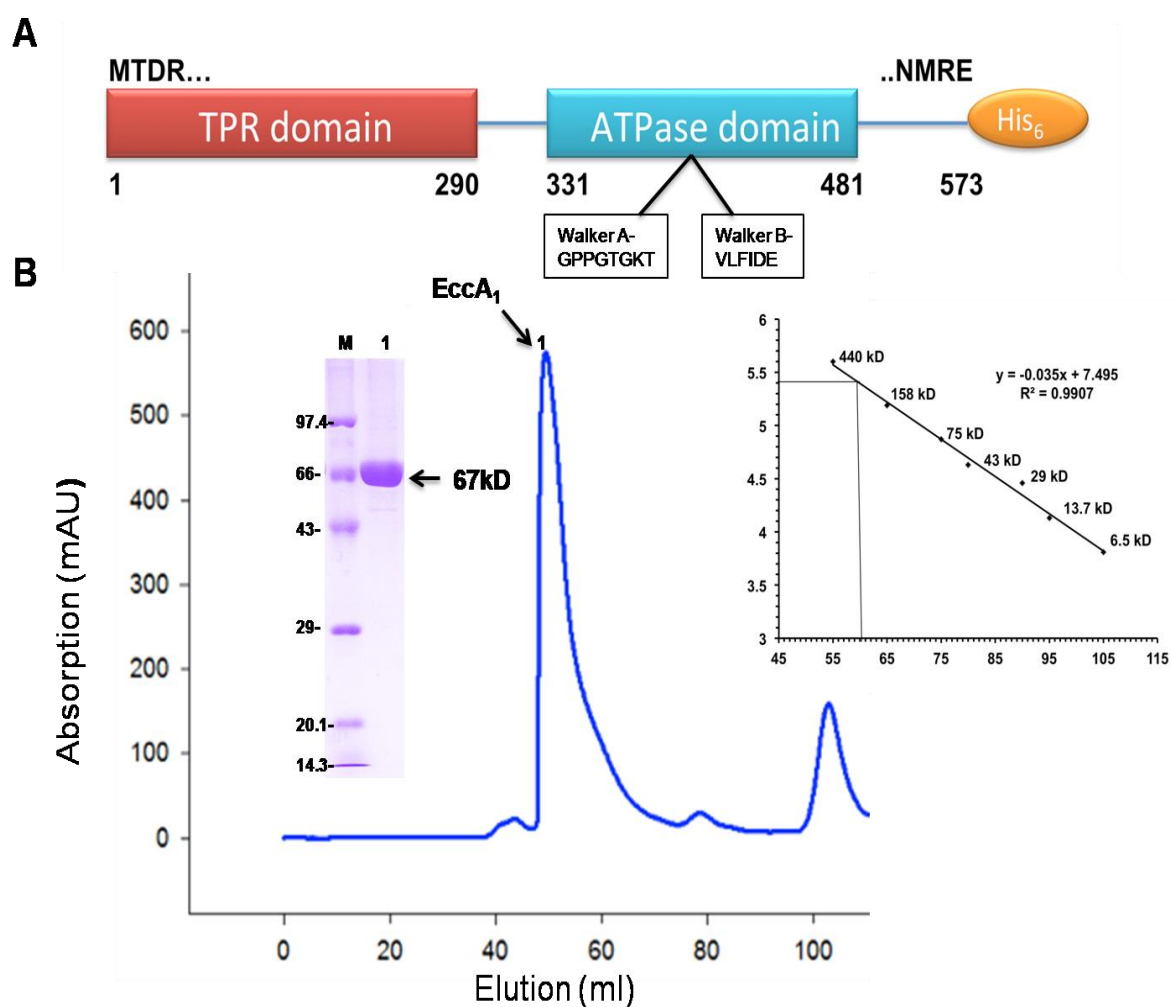
Fig. 11 A. Schematic view of various domains of $EccA_1$ and its expression construct. TPR-tetratricopeptide repeat, ATPase-ATP binding domain. B. Size exclusion chromatography and SDS-PAGE analysis of purified EccA1. The EccA1 eluted as hexamer from Superdex200 (16/60) column.

EccA1 exhibit ATPase activity

To examine ATPase activity, EccA1 was dialyzed in protein buffer containing no ATP. We performed a sensitive radioactive assay on EccA1 using $[\gamma^{32}P]$ ATP as substrate and release of free phosphate was monitored (Fig. 11A). The free phosphate was released linearly overtime and K_m of 52.4±2.1μm and V_{max} of 1.51±0.7 μmol/min were obtained from Michaelis-Menten plot and nonlinear regression analysis using Prism 6.0 (GraphPad Software) (Fig. 11B).These data indicate that EccA1 is a weak ATPase and also reflected in its Walker-A sequence (GPPGTGKT), however, possess specific ATPase activity.
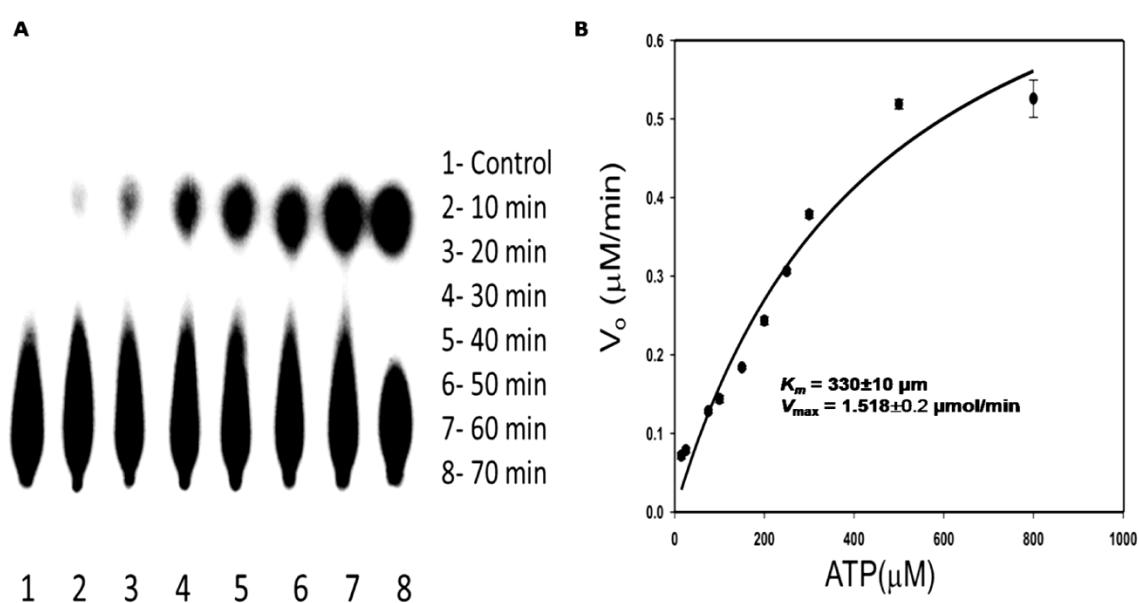
Fig.11. A. The EccA1 was incubated with $[\gamma^{32}P]$ ATP and ATPase activity was analyzed at various time intervals. The amount of hydrolyzed ATP is shown as a percentage of original $[\gamma^{32}P]$ ATP. Each point is the average of three independent experiments. The inset represents the autoradiography profiles obtained from TLC. The plot of ATP hydrolysis vs. protein concentration. B. Michaelis-Menten plot of ATP hydrolysis of EccA1.

Secondary structure and thermal stability analysis of EccA1

Far-UV CD spectra (260-200 nm) were obtained for EccA1to estimate the contents of the secondary structureof protein. The Dichroweb server analysis yielded (~60% α-helix, ~7% β-sheet and ~33% random coil) for EccA1 (Fig. 12A). These values were close to secondary structures obtained from the theoretical analysis (~60% α-helix, ~4.3% β-sheet, ~35.2% random coil) for EccA1.
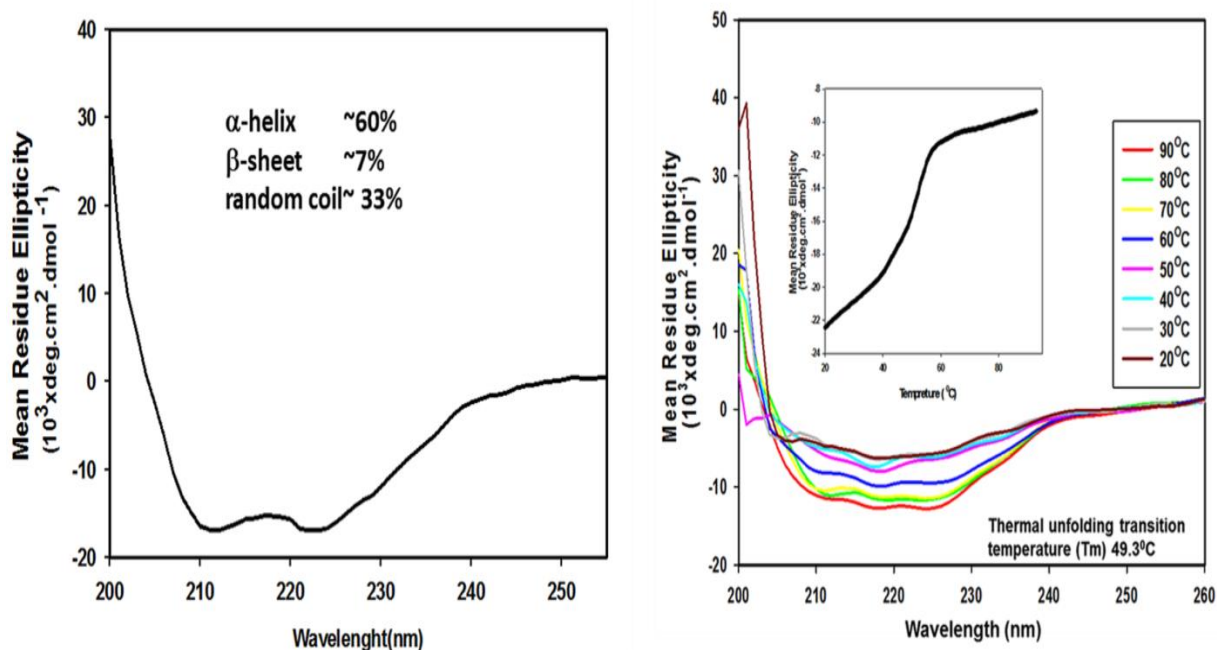
Fig. 12Circular dichorism and thermal stability profile of EccA1. A. Circular dichorism and B. thermal stability profile of EccA1. Spectra were recorded at room temperature and normalized to mean residue ellipticity. The secondary structure contents of E EccA1 matches well with secondary structural contents obtained from theoretical predictions. EccA1 proteins are quite thermostable.

For thermal stability analysis, the CD spectra of EccA1 was recorded as a function of temperature The effect of temperature on theEccA1 secondary structure was performed from 20^0C to 90^0C in 10^0C increment (Fig. 12B) and exhibited thermal unfolding transition temperature (T_m) 49.3 ± 0.1^0C which is showing that Ecca1 is thermally stable.

Crystallization of EccA1

Crystallization attempts were made on purified EccA1 protein using numbers of screens *e.g.* PEG/Ion, Index, SaltRx, Crystal Screen1 and 2 from Hampton research, PCAT premier, JCSG, Morpheus, stura foot print, Structure screen1 and 2 screens from the Molecular dimension and Nuc-Pro screen from Jena Bioscience. Different concentrations (2mg/ml, 5mg/ml, 10mg/ml and 20mg/ml) of Mtb EccA1 protein at different temperatures (4^0C, 16^0C and 22^0C) were set up for crystallization. Hanging drop and sitting drop vapors diffusion methods were used in crystallization experiment. Initially, we were able to get EccA1 crystals in several conditions of Morpheus screen but we failed to get well diffracting crystals.

Molecular modeling of EccA1

*The mtbecca*1 protein belongs to CbxX/CfxQ ATPase family and associated with various cellular activities (AAA ATPase). The available database indicated that Rubisco activase protein (PDB-3SYL) was the closest homolog of C- terminal domain of EccA1. Rubisco adopts a hexameric structure through its C-terminal (Stotz, Mueller-Cajar et al. 2011). The EccA1 also adopts a hexameric structure and its C-terminal domain is involved in oligomerization of protein (Luthra, Mahmood et al. 2008). Similar to other AAA+ ATPases, EccA1 hydrolyzes ATP *in vitro* and forms hexamer(Luthra, Mahmood et al. 2008)].

The crystal structures of N-terminal of EccA1 (PDB-4F3V) and AAA+ CbbX (PDB - 3SYL) were used as input templates to obtain the full-length EccA1 model (Fig. 13A).EccA1 modeling has indicated that large part of the protein is well structured. The N-terminal TPRs (Tetra Trico repeats) of EccA1 interacts hydrophobically back to back and together form a right-handed super helix (4F3V). Secondary structure analysis of EccA1 by PSIPRED program (Fig. 13B) indicated that EccA1 is a helical protein. Sequence analysis of EccA1 using DISOPRED program indicated that EccA1 is a well-ordered protein. The disopred program indicated that 245-255 residues, 262-277 residues and C-terminal 407-411 residues of EccA1 are disordered.

The PROCHECK program analysis on EccA1 indicated that 92.6% residues lie in the most favored region of Ramachandran plot (Data not shown). The Prosa Z-score value of EccA1 was comparable with the template structures (EccA1 model = -9.75, 4F3V = -9.77 and 3SYL = -8.25). The ProSA-web plot has indicated the good quality the model [91]. This plot shows local model quality by plotting energies as a function of amino acid position. In general, positive values correspond to problematic or erroneous parts of the input structure.

Fig.13.A. Molecular model of EccA1 monomer contains N-terminal TPR domain and C-terminal ATPase domain structure. B. Secondary structure prediction of *M. tuberculosis* EccA1 protein by PSIPRED program. C. Molecular model of EccA1 hexamer.

The EccA1 model contains 5 and 2 stranded parallel β sheets at C-terminal and one antiparallel β-sheet at the N-terminal. The model contains 5.8% β-sheet content, 54% α-helix (29 helix) and 2.4% 3-$_{10}$ helix (4 helix) (Fig.13A). The C-terminal of EccA1 has a Nucleotide-binding domain. The EccA1 monomer assembles into the hexamer through C-terminal domain (Fig. 13C)(Luthra, Mahmood et al. 2008). Previous studies have shown that full-length EccA1 assembled as hexamer for its biological activity and full-length EccA1 shows weak ATPase activity (Luthra, Mahmood et al. 2008). The EccA1 model is well structured and contains mostly helical structure as supported by another study(Luthra, Mahmood et al. 2008).

CHAPTER 3: BINDING ANALYSIS BETWEEN ESPC AND ECCA1

Material and methods

Molecular Docking of Mtb EccA1 and Mtb EspC proteins

ROSETTADOCK server with protein-protein docking function was used to build the EspC-EccA1 complex. The β-finger insertion (A81-V96) of TPR domain of EccA1and YxxxD motif of the export arm of EspC (Y87-D91) was used as input constraints in docking calculation. The docking server yielded 50 best EspC-EccA1complexes and best model was selected based on binding energy, intermolecular interactions and buried surface area in the complex. The PISA server (Krissinel and Henrick 2007) was used to analyze the interactions, binding free energy and solvent accessibility in theEspC-EccA1 complex.

Energy minimization and dynamics simulation were performed on EspC-EccA1 complex using GROMACS program. The EspC-EccA1complex was immersed in cubic box extending 0.5 nm from protein surface and solvated with explicit SPC water molecules. The chloride and sodium ions were added to neutralize the system and simulated with periodic boundary conditions. The solvated complex system consists of 467,826 protein atoms surrounded by 1000,000 water molecules. Before running the dynamics simulation, the whole system was energy minimized for 200 iterations of steepest descents and then equilibrated for 10 ps, during which protein atoms were restrained. All restraints were removed from the protein and temperature was gradually increased in 10 distinct steps of 5 ps simulation each. V- rescale (modified Berendsen thermostat) coupling was employed to maintain a constant 300K temperature with a coupling constant 300K temperature and a coupling constant of 0.2 ps. The coulomb cut off was 1.0. The time step employed was 2fs and the coordinates were saved every 4ps for MD trajectories analysis. The pressure was maintained by Parrinello-rahman pressure coupling constant and Coulomb cut off was applied during temperature coupling. The stereochemistry of simulated EspC-EccA1complex was checked by PROCHECK program, secondary structure composition by DSSP program and structure visualization by PyMOL program.

Binding analysis using native EspC and EccA$_1$ proteins through SPR

Binding affinity between EspC-EccA1was obtained by surface plasmon resonance technique using BIAcore3000 system (*Biacore Pharmacia Biosensor AB, Uppsala Sweden).* CM4 sensor chip was used for binding analysis. The experiment was performed at 25°C in HBS-P buffer containing (20mM HEPES buffer pH 7.5, 150mM NaCl, 10mM Arginine, 10 mM Glutamate, 0.02mM ATP, and 0.02 mM MgCl$_2$). The CM4 sensor chip contains carboxymethylated dextran covalently attached to gold surface and protein is covalently coupled to sensor surface via amine, thiol, aldehyde or carboxyl groups.

The EccA$_1$ was immobilized on CM4 surface activated by EDC-NHS and amine coupling chemistry. 170μl of EccA$_1$ (Conc. ~ 60μg/ml) was diluted in sodium acetate buffer pH 4.1 and injected over flow cells with 30μl/min. The 5100 RU of EccA1was immobilized on CM4 sensor chip surface and the surface was blocked by ethanolamine. Four different concentrations of EspC (0.5, 1, 2, and 4 μM) were injected over immobilized EccA1 using HBS-P buffer pH 7.5 with 30μl/min flow rate. The sensorgram was allowed to run for another 5 min. We regenerated the biosensor surface by two pulses of 50mM NaOH regeneration buffer then followed by running the buffer for another 10 min with 30μl/min flow rate to equilibrate the surface. All the data were subtracted with data obtained from reference channel. Association and dissociation kinetic constants were evaluated by BIAeveluation 3.0 software using simple 1:1 Langmuir model.

Binding analysis using native EspC and mutants EccA$_1$ protein through SPR

The binding studies between EspC and EccA1 point mutants (Y89A and D91A) were performed using BIAcore T200 system (*Biacore Pharmacia Biosensor AB, Uppsala Sweden).* The EccA$_1$ mutants were immobilized on CM4 (S series) surface activated by EDC-NHS and amine coupling chemistry. 150μl of each EccA1 Y89A (Conc. ~ 200 μg/ml) and EccA$_1$ D91A (Conc. ~ 200 μg/ml) were diluted in sodium acetate buffer pH 4.15 and injected over flow cells with 30μl/min. The 1986 RU and 2770.8 RU of EccA1Y89A and D91A were immobilized on CM4 sensor chip surface respectively and the surface was blocked by ethanolamine. Five different concentrations of EspC (1.5, 3, 6, 12 and 24 μM) were injected over immobilized EccA$_1$ mutants using HBS-P buffer pH 7.5 with 30μl/min. Association and

dissociation kinetic constants were evaluated by BIAevaluation 3.0 software using simple 1:1 Langmuir model.

Binding analysis using mutant EspC and EccA₁ protein through SPR

Three deletion mutants of EspC were generated (i) EspC-Δ5 (M1-D99 residues) (ii) EspC-Δ10 (M1-W93 residues) and (iii) EspC-Δ20 (M1-A83 residues). Two point mutants (Y87A and W94A) were also generated to study binding with native EccA1 protein. All three deletion mutants of EspC (Δ5, Δ10, and Δ20) with different concentrations (0.5, 1, 2, and 4 μM) were injected over previously immobilized native EccA₁ using HBS-P buffer pH 7.5 and kept all the experimental parameters as previous SPR experiments (see above). Similarly different concentrations (1μM, 2 μM, 5 μM, 10 μM and 20 μM of EspC Y87A and 5 μM, 10 μM, 20 μM, 40 μM and 80 μM of EspC W94A) of EspC point mutants (Y87A and W94A) were injected over immobilized native EccA1 using same using HBS-P buffer pH 7.5 and kept all the experimental parameters as previous SPR experiments (see above). Surface plasmon resonance technique was used to obtain dissociation constant (K_D) between EccA1and all mutants of EspC. Interactions between EspC and EccA1were identified. Obtained K_D values for mutants were compared with K_D obtained from native EccA1–EspC binding SPR.

Results and discussions

Molecular docking of EccA1 and EspC proteins

The EspC interacts with EccA1 through its C-terminus (DiGiuseppe Champion, Champion et al. 2009) and this information was used as input in ROSETTADOC docking program. The best EccA1-EspC docking complex was selected on the basis of binding energy, interface accessible surface area, salt bridges and a number of hydrogen bonds. The docked complex structure showed that C-terminal export arm of EspC is inserted in TPR domain of EccA1. The export arm of EspC contains YxxxD/E motif and TPR domain EccA1 contains β-hairpin. Based on the interaction between EspC-EccA₁ complexes in the best-docked complex, we observed multiple interactions between EspC and EccA1 (Table 2). On the basis of docking, we generated two mutants of β-hairpin insertion motif of TPR domain (Y89A and D91A) and probed the roles of these residues in EspC recognition. These mutations lie at the loop of β-hairpin insertion located in the concave groove of TPR. D91 of

β-hairpin of TPR forms a hydrogen bond with theY87 residue of ^{87}YxxxD/E^{91} export arm of EspC. Y89 of β-hairpin of TPR forms a hydrogen bond with W94 of the export arm and L88 of β-hairpin forms a hydrophobic interaction with I98 of the export arm of EspC.
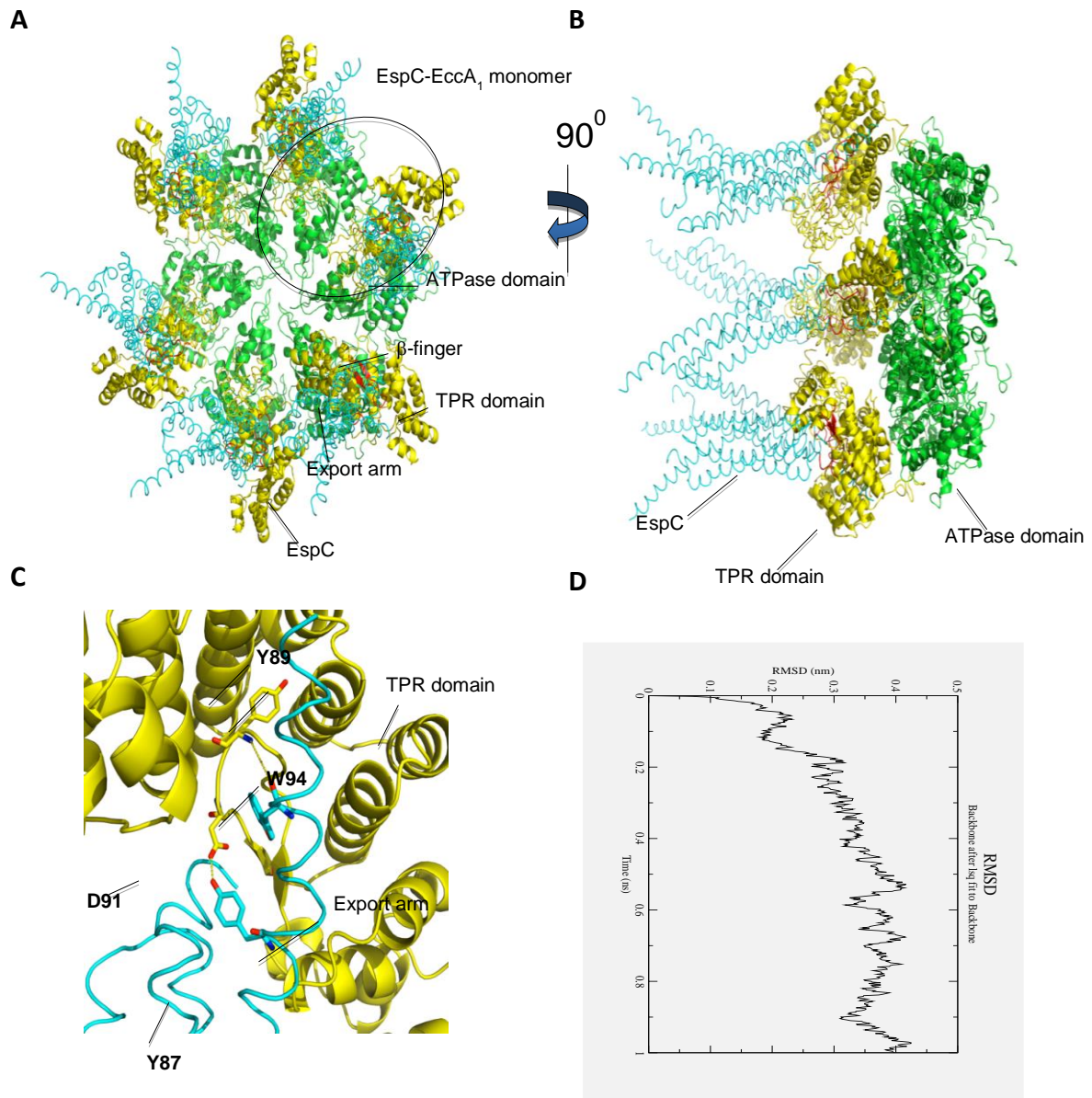
Fig.14. A. EspC-EccA1 complex obtained from molecular modeling, docking and dynamics simulation techniques. Esp Chomodimer in cyan, TPR domain in yellow and ATPase domain are shown in green color. β-hairpin insertion of TPR domain and export arm of EspC is shown in red color. B. EspC-EccA1 model after 90^0 rotations. C. Closer view of interactions between the EspC export arm and β-finger insertion of TPR domain of EccA1. D. Root mean square deviation of Cα atoms during molecular dynamics simulation, RMSD with respect to the starting structures.

Hexameric Espc-EccA1 complex consists of total 31968 atoms and The PISA server analysis on hexameric EspC-EccA1 complex (Total 12 subunits; 6 from each EspC and

EccA1), indicated the 193356.3Å^2 total surface area and 32978.2Å^2 buried surface area. The dissociation energy of the complex was -14.6kcal/mol. It indicated that EccA1-EspC complex is thermodynamically stable.

Table 2. The interacting residues at the interface of EccA1 and EspC complex

Ecca1 residues	EspC residues
Tyr89	Trp94
Leu88	Ile98
Asp91	Tyr87
Ala112	Ala93
Glu238	Tyr87
Thr234	Arg11
Arg206	Tyr87

Binding analysis using between EspC and EccA₁ proteins through SPR

We have successfully performed binding experiment between MtbEspC and EccA1 through SPR technology and we were able to get affinity (K_D~0.14) between native MtbEspC and EccA1 proteins. On the basis of our *in silico* docking study to investigate motifs and amino acids involved in the interaction we have created several C-terminal truncated mutants (Δ5, Δ10, and Δ20) of MtbEspC protein. We were able to expressed and purify all the truncated mutant proteins of MtbEspC and performed SPR experiments by injecting different concentrations of immobilized native Mtb EccA1 protein in CM4 sensor chip. We got binding different affinities between native EccA1 and truncated mutants Δ5 and Δ10 corresponding to K_D value of 6.7µM and 8.1µM respectively, which is showing several fold (~47 fold for Δ5 and ~57 fold for Δ10) decrease in affinity, while Δ20 mutant (lacking full export arm consists of AA-cradle, YSEAD, and IDGLF motifs) showed ~no binding to EccA₁. These results indicate that IDGLF motif of EspC export arm is critical for EccA1 recognition and its deletion in EspC-Δ5 and EspC-Δ10 mutants leads to ~40-60fold decrease in binding affinity. Conserved AA-cradle, YSEAD and IDGLF motifs of the EspC export arm are absolutely essential for EccA₁ recognition and deletion of these motifs in EspC-Δ20 mutant leads to ~no binding to EccA1.
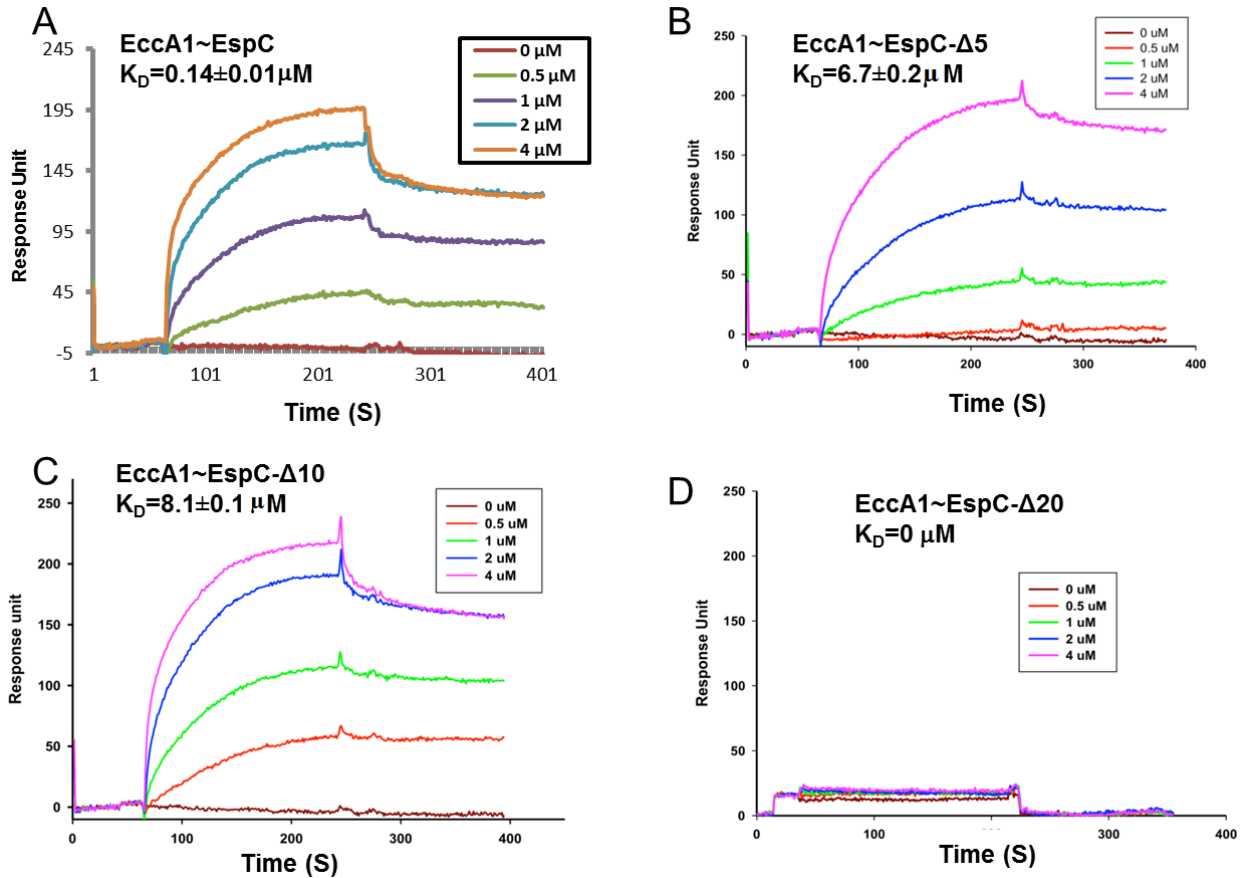
Fig. 15. Binding analysis between EccA1and EspC using surface plasmon resonance technique. A. Binding analysis of EccA1 with native EspC. B. Binding analysis of EccA$_1$ with theEspC-Δ5 mutant. C. Binding analysis of EccA1 with EspC-Δ10 mutant. D. Binding analysis of EccA1 with EspC-Δ20 mutant.

Further on the basis of our *in silico* docking studies we recognized several critical amino acids on both proteins, those are involved in interaction, so we have created a few site directed point mutants of both EspC and EccA1 (Y87A and W94A of EspC and Y89A and D91A of EccA1) and again able to expressed, purify the site directed point mutant proteins and performed the SPR experiments (Fig.16).
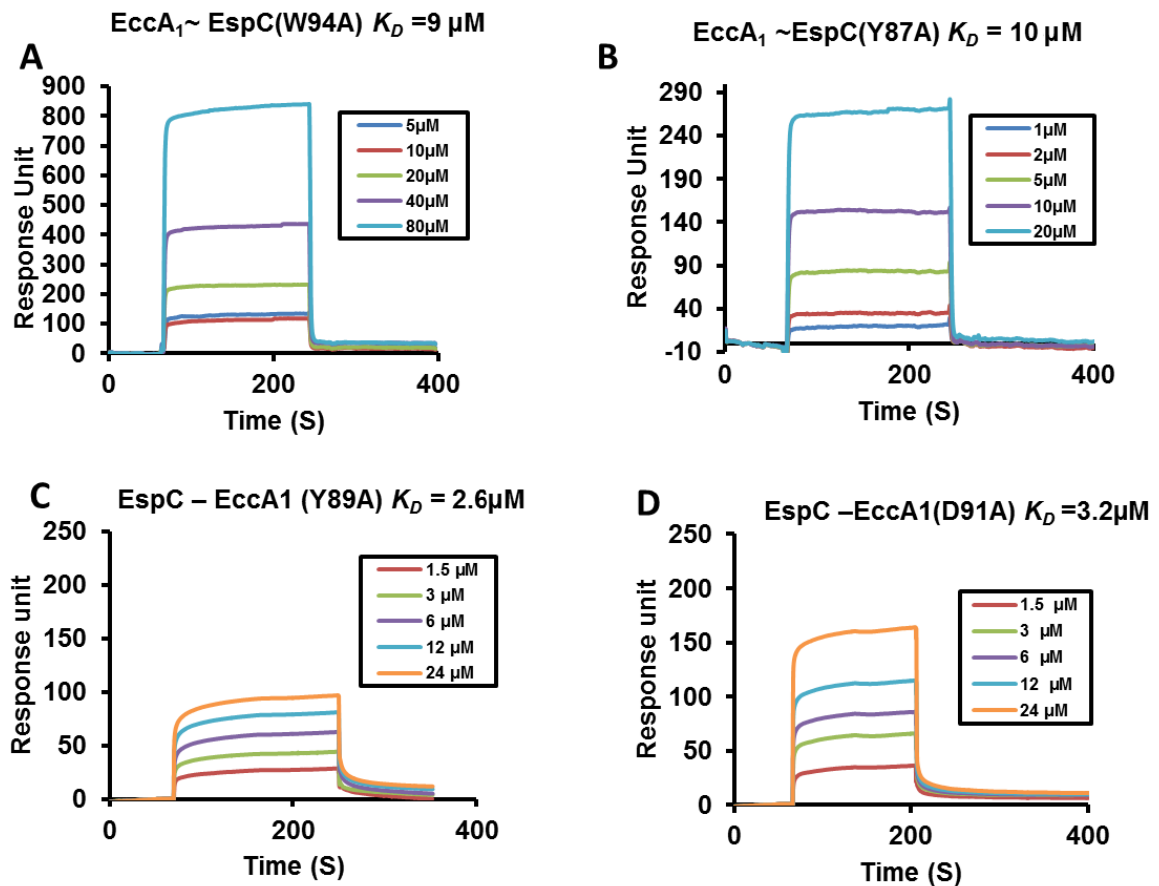
Fig.16 Binding analysis between EccA1and EspC point mutants using surface plasmon resonance technique. A. Binding analysis between native EccA1 and EspC-W94A mutant. B. Binding analysis between native EccA1 and EspC-Y87A mutant. C. Binding analysis between native EspC and Y89A-EccA1 mutant. D. Binding analysis between native EspC with D91A-EccA mutant.

We have observed the different K_D values for different mutants (K_D-9μM for EccA1-EspC(W94A), K_D-10μM for EccA1-EspC(Y87A), K_D-2.6μM for EspC-EccA1(Y89A) and K_D-3.2μM for EspC-EccA1(D91A) which is showing several fold decrease (~64 fold for EccA1-EspC(W94A), ~71 fold for EccA1-EspC(Y87A), ~18 fold for EspC-EccA1(Y89A) and ~22 fold for EspC-EccA1(D91A) in affinities in comparison to native EspC and EccA1 proteins. Thus SPR binding studies validate our *in silico* molecular docking study where D91 of β-hairpin of TPR forms a hydrogen bond with theY87 residue of ^{87}YxxxD/E^{91} export arm of EspC. Y89 of β-hairpin of TPR forms a hydrogen bond with W94 of the export arm of EspC.

Summary

In our study, we have cloned, expressed and purified both of the Mtb EspC, EccA1proteins and their mutants successfully. We have done a lot of experiments to characterize EspC and EccA1 in terms of purification, molecular weight, activity, secondary structure, binding and recognition mechanism and molecular structure modeling and docking study. To determine the molecular weight of these (EspC and EccA1) molecules we have performed size exclusion chromatography, dynamic light scattering (data not shown) intact masses (by MALDI-TOF and ESI-TOF mass spectroscopy) experiments. Among these experiments, only size exclusion chromatography technique was successfully worked out in a particular buffer condition which contains additives (see methods and materials). Size exclusion chromatography of EspC and EccA1 reported that both the proteins were eluted as a hexamer. In an earlier study (Luthra, Mahmood et al. 2008) it was also reported that EccA1 forms hexamer

Activity assays were performed to find out whether both purified molecules are active or not. On the basis another ESX-1 secreted molecule ESAT-6 (Helix-turn-helix structure) which is much similar to EspC in the term of structure, we have performed cell membrane lysis activity assay on A549 cells (Human epithelial carcinoma cell line) and were able to observe that EspC destabilize cell membrane. Confocal microscopy and Co-immunoprecipitation assay have shown that EspC interacts with a number of the host cytosolic proteins which are involved in cytoskeleton formation, metabolic pathways (see results). To determine our purified recombinant EccA1molecule is active so that we have performed ATPase assay both radioactively and calorimetrically as reported in the earlier study (Luthra, Mahmood et al. 2008) and observed that EccA1 is active and showing ATPase activity.

Secondary structure analysis by Circular Dichroism and PSIPRED have revealed that EspC is a helix-turn-helix protein as other ESX -1 proteins e.g. ESAT-6, CFP-10, EspC contains nearly ~80% α helix and ~20% random coils, whereas EccA1 is mostly a helical protein which contains ~60%α helix, ~7%β sheets and ~33% random coils in our study. Thermal denaturation analysis using CD has shown that EspC is thermally unstable with Tm value of 32.2°C whereas EccA1 is a thermally stable protein with thermal Tm value of 49.3°C.

To investigate critical amino acids involved in the interaction between EspC and EccA1, we have performed molecular docking and SPR experiments. A binding study using SPR technology has proven that EspC interacts with EccA1 *in vitro*. Moreover, one of the earlier study (Jeffery Cox et al) it has also proven *in vivo* that EspC interacts EccA1through its C-terminal recognition signal sequences by Yeast two Hybrid experiment. Further on the basis of our molecular docking study, we recognized the interacting residues of both EspC and EccA1 proteins. The binding experiments using SPR indicate that C-terminal export arm YxxxD/E motif of EspC interacts with N-terminal EccA1β-hair pin of TPR domain.

We have modeled both the EspC and EccA1 proteins using several bioinformatics tools and investigate the structural properties of these molecules. The EspC modeling by – showing that it is a helix-turn-helix two domain protein whereas EccA1 modeled by – showing it is mostly a helical protein(~60% α-helix) with ~7%β-sheet and ~33% loop region. We modeled EspC~EccA1 complex and identified that β-finger insertion of TPR domain of EccA1 accounts most of the interactions with the EspC export arm. EccA1 uses concave groove for its interaction with EspC protein, however a β-finger insertion in observed in concave groove of TPR and may form extended β-strand complementation interaction with the export arm of EspC.

In summary, we have found that EspC forms hexamer in solution and involved in A549 plasma membrane lysis and accumulated in the cytosol and caused the cell death. We propose that EspC may lyse similarly the phagosomal membrane during *mycobacterial* infection. The EccA1exists as a hexamer and is involved in translocation of ESX-1 virulence proteins out of the cell via hydrolyzing ATP. We have identified and characterized the ATPase activity of EccA1. EccA1 is essential *in vivo* for specific targeting and secretion of EspC and other co-secreted ESX-1 virulence proteins. It is the first detailed biochemical and structural characterization of recognition mechanism between EspC and EccA1proteins, essential for EspC secretion by theESX-1 system. ESX-1 secretion systems are considered potential drug target. Biochemical characterization, mutational analysis, and mechanism of recognition between EspC and EccA1 will be indispensable for the development of novel inhibitors, which will prevent the export of virulence proteins by ESX-1 secretion system.

References

Abdallah, A. M., N. C. G. Van Pittius, P. A. D. Champion, J. Cox, J. Luirink, C. M. Vandenbroucke-Grauls, B. J. Appelmelk and W. Bitter (2007). "Type VII secretion--mycobacteria show the way." Nature reviews. Microbiology **5**(11): 883.

Abdallah, A. M., T. Verboom, F. Hannes, M. Safi, M. Strong, D. Eisenberg, R. J. Musters, C. M. Vandenbroucke-Grauls, B. J. Appelmelk and J. Luirink (2006). "A specific secretion system mediates PPE41 transport in pathogenic mycobacteria." Molecular microbiology **62**(3): 667-679.

Abdallah, A. M., T. Verboom, E. M. Weerdenburg, N. C. Gey van Pittius, P. W. Mahasha, C. Jiménez, M. Parra, N. Cadieux, M. J. Brennan and B. J. Appelmelk (2009). "PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5." Molecular microbiology **73**(3): 329-340.

Bitter, W., E. N. Houben, D. Bottai, P. Brodin, E. J. Brown, J. S. Cox, K. Derbyshire, S. M. Fortune, L.-Y. Gao and J. Liu (2009). "Systematic genetic nomenclature for type VII secretion systems." PLoS pathogens **5**(10): e1000507.

Braun, V. and H. Wu (1994). "Lipoproteins, structure, function, biosynthesis and model for protein export." New comprehensive biochemistry **27**: 319-341.

Braunstein, M., A. M. Brown, S. Kurtz and W. R. Jacobs, Jr. (2001). "Two nonredundant SecA homologues function in mycobacteria." J Bacteriol **183**(24): 6979-6990.

Brodin, P., M. I. de Jonge, L. Majlessi, C. Leclerc, M. Nilges, S. T. Cole and R. Brosch (2005). "Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of Mycobacterium tuberculosis, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity." Journal of Biological Chemistry **280**(40): 33953-33959.

Brodin, P., L. Majlessi, L. Marsollier, M. I. de Jonge, D. Bottai, C. Demangel, J. Hinds, O. Neyrolles, P. D. Butcher and C. Leclerc (2006). "Dissection of ESAT-6 system 1 of Mycobacterium tuberculosis and impact on immunogenicity and virulence." Infection and immunity **74**(1): 88-98.

Champion, P. A., M. M. Champion, P. Manzanillo and J. S. Cox (2009). "ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria." Mol Microbiol **73**(5): 950-962.

Champion, P. A., S. A. Stanley, M. M. Champion, E. J. Brown and J. S. Cox (2006). "C-terminal signal sequence promotes virulence factor secretion in Mycobacterium tuberculosis." Science **313**(5793): 1632-1636.

de Jonge, M. I., G. Pehau-Arnaudet, M. M. Fretz, F. Romain, D. Bottai, P. Brodin, N. Honoré, G. Marchal, W. Jiskoot and P. England (2007). "ESAT-6 from Mycobacterium tuberculosis dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity." Journal of bacteriology **189**(16): 6028-6034.

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

DeLisa, M. P., D. Tullman and G. Georgiou (2003). "Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway." Proceedings of the National Academy of Sciences **100**(10): 6115-6120.

DiGiuseppe Champion, P. A., M. M. Champion, P. Manzanillo and J. S. Cox (2009). "ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria." Molecular microbiology **73**(5): 950-962.

Economou, A. and W. Wickner (1994). "SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion." Cell **78**(5): 835-843.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D: Biological Crystallography **60**(12): 2126-2132.

Feltcher, M. E., H. P. Gunawardena, K. E. Zulauf, S. Malik, J. E. Griffin, C. M. Sassetti, X. Chen and M. Braunstein (2015). "Label-free quantitative proteomics reveals a role for the Mycobacterium

tuberculosis SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall." Molecular & Cellular Proteomics **14**(6): 1501-1516.

Finlay, B. B. and S. Falkow (1997). "Common themes in microbial pathogenicity revisited." Microbiology and molecular biology reviews **61**(2): 136-169.

Fisher, M. A., B. B. Plikaytis and T. M. Shinnick (2002). "Microarray analysis of the Mycobacterium tuberculosis transcriptional response to the acidic conditions found in phagosomes." Journal of bacteriology **184**(14): 4025-4032.

Fortune, S., A. Jaeger, D. Sarracino, M. Chase, C. Sassetti, D. Sherman, B. Bloom and E. Rubin (2005). "Mutually dependent secretion of proteins required for mycobacterial virulence." Proceedings of the National Academy of Sciences of the United States of America **102**(30): 10676-10681.

Gao, L. Y., S. Guo, B. McLaughlin, H. Morisaki, J. N. Engel and E. J. Brown (2004). "A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion." Molecular microbiology **53**(6): 1677-1693.

Geourjon, C. and G. Deleage (1995). "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments." Bioinformatics **11**(6): 681-684.

Guinn, K. M., M. J. Hickey, S. K. Mathur, K. L. Zakel, J. E. Grotzke, D. M. Lewinsohn, S. Smith and D. R. Sherman (2004). "Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis." Molecular microbiology **51**(2): 359-370.

Guo, X. V., M. Monteleone, M. Klotzsche, A. Kamionka, W. Hillen, M. Braunstein, S. Ehrt and D. Schnappinger (2007). "Silencing essential protein secretion in Mycobacterium smegmatis by using tetracycline repressors." Journal of bacteriology **189**(13): 4614-4623.

Hess, B., C. Kutzner, D. Van Der Spoel and E. Lindahl (2008). "GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation." Journal of chemical theory and computation **4**(3): 435-447.

Hsu, T., S. M. Hingley-Wilson, B. Chen, M. Chen, A. Z. Dai, P. M. Morin, C. B. Marks, J. Padiyar, C. Goulding and M. Gingery (2003). "The primary mechanism of attenuation of bacillus Calmette–Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue." Proceedings of the National Academy of Sciences **100**(21): 12420-12425.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." Journal of molecular biology **292**(2): 195-202.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nature protocols **10**(6): 845-858.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." Journal of molecular biology **372**(3): 774-797.

Krogh, A., B. Larsson, G. Von Heijne and E. L. Sonnhammer (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." Journal of molecular biology **305**(3): 567-580.

Laskowski, R. A. (2001). "PDBsum: summaries and analyses of PDB structures." Nucleic acids research **29**(1): 221-222.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of applied crystallography **26**(2): 283-291.

Lee, P. A., D. Tullman-Ercek and G. Georgiou (2006). "The bacterial twin-arginine translocation pathway." Annu. Rev. Microbiol. **60**: 373-395.

Luthra, A., A. Mahmood, A. Arora and R. Ramachandran (2008). "Characterization of Rv3868, an essential hypothetical protein of the ESX-1 secretion system in Mycobacterium tuberculosis." Journal of Biological Chemistry **283**(52): 36532-36541.

Lyskov, S. and J. J. Gray (2008). "The RosettaDock server for local protein–protein docking." Nucleic acids research **36**(suppl_2): W233-W238.

MacGurn, J. A. and J. S. Cox (2007). "A genetic screen for Mycobacterium tuberculosis mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system." Infection and immunity **75**(6): 2668-2678.

MacGurn, J. A., S. Raghavan, S. A. Stanley and J. S. Cox (2005). "A non-RD1 gene cluster is required for Snm secretion in Mycobacterium tuberculosis." Molecular microbiology **57**(6): 1653-1663.

McDonough, J. A., K. E. Hacker, A. R. Flores, M. S. Pavelka and M. Braunstein (2005). "The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial β-lactamases." Journal of bacteriology **187**(22): 7667-7679.

McDonough, J. A., J. R. McCann, E. M. Tekippe, J. S. Silverman, N. W. Rigel and M. Braunstein (2008). "Identification of functional Tat signal sequences in Mycobacterium tuberculosis proteins." Journal of bacteriology **190**(19): 6428-6438.

McLaughlin, B., J. S. Chon, J. A. MacGurn, F. Carlsson, T. L. Cheng, J. S. Cox and E. J. Brown (2007). "A mycobacterium ESX-1–secreted virulence factor with unique requirements for export." PLOS pathogens **3**(8): e105.

Millington, K. A., S. M. Fortune, J. Low, A. Garces, S. M. Hingley-Wilson, M. Wickremasinghe, O. M. Kon and A. Lalvani (2011). "Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for Mycobacterium tuberculosis infection." Proceedings of the National Academy of Sciences **108**(14): 5730-5735.

Ohol, Y. M., D. H. Goetz, K. Chan, M. U. Shiloh, C. S. Craik and J. S. Cox (2010). "Mycobacterium tuberculosis MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence." Cell host & microbe **7**(3): 210-220.

Organization, W. H. (2009). "Global tuberculosis control: a short update to the 2009 report."

Paetzel, M., A. Karla, N. C. Strynadka and R. E. Dalbey (2002). "Signal peptidases." Chem Rev **102**(12): 4549-4580.

Pathak, S. K., S. Basu, K. K. Basu, A. Banerjee, S. Pathak, A. Bhattacharyya, T. Kaisho, M. Kundu and J. Basu (2015). "Corrigendum: Direct extracellular interaction between the early secreted antigen ESAT-6 of Mycobacterium tuberculosis and TLR2 inhibits TLR signaling in macrophages." Nature immunology **16**(3): 326.

Posey, J. E., T. M. Shinnick and F. D. Quinn (2006). "Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis." Journal of bacteriology **188**(4): 1332-1340.

Pym, A. S., P. Brodin, R. Brosch, M. Huerre and S. T. Cole (2002). "Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti." Molecular microbiology **46**(3): 709-717.

Raghavan, S., P. Manzanillo, K. Chan, C. Dovey and J. S. Cox (2008). "Secreted transcription factor controls Mycobacterium tuberculosis virulence." Nature **454**(7205): 717.

Raynaud, C., C. Guilhot, J. Rauzier, Y. Bordat, V. Pelicic, R. Manganelli, I. Smith, B. Gicquel and M. Jackson (2002). "Phospholipases C are involved in the virulence of Mycobacterium tuberculosis." Molecular microbiology **45**(1): 203-217.

Renshaw, P. S., K. L. Lightbody, V. Veverka, F. W. Muskett, G. Kelly, T. A. Frenkiel, S. V. Gordon, R. G. Hewinson, B. Burke and J. Norman (2005). "Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6." The EMBO journal **24**(14): 2491-2498.

Sander, P., M. Rezwan, B. Walker, S. Rampini, R. Kroppenstedt, S. Ehlers, C. Keller, J. Keeble, M. Hagemeier and M. Colston (2004). "Lipoprotein processing is required for virulence of Mycobacterium tuberculosis." Molecular microbiology **52**(6): 1543-1552.

Sassetti, C. M., D. H. Boyd and E. J. Rubin (2001). "Comprehensive identification of conditionally essential genes in mycobacteria." Proceedings of the National Academy of Sciences **98**(22): 12712-12717.

Sassetti, C. M. and E. J. Rubin (2003). "Genetic requirements for mycobacterial survival during infection." Proceedings of the National Academy of Sciences **100**(22): 12989-12994.

Schneider, G. (1999). "How many potentially secreted proteins are contained in a bacterial genome?" Gene **237**(1): 113-121.

Sen, T. Z., R. L. Jernigan, J. Garnier and A. Kloczkowski (2005). "GOR V server for protein secondary structure prediction." Bioinformatics **21**(11): 2787-2788.

Simeone, R., D. Bottai and R. Brosch (2009). "ESX/type VII secretion systems and their role in host–pathogen interaction." Current opinion in microbiology **12**(1): 4-10.

Singh, V., C. Kaur, V. K. Chaudhary, K. V. Rao and S. Chatterjee (2015). "M. tuberculosis Secretory Protein ESAT-6 Induces Metabolic Flux Perturbations to Drive Foamy Macrophage Differentiation." Sci Rep **5**: 12906.

Solomonson, M., D. Setiaputra, K. A. Makepeace, E. Lameignere, E. V. Petrotchenko, D. G. Conrady, J. R. Bergeron, M. Vuckovic, F. DiMaio and C. H. Borchers (2015). "Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism." Structure **23**(3): 571-583.

Sørensen, A. L., S. Nagai, G. Houen, P. Andersen and A. B. Andersen (1995). "Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis." Infection and immunity **63**(5): 1710-1717.

Stanley, N. R., T. Palmer and B. C. Berks (2000). "The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli." Journal of Biological Chemistry **275**(16): 11591-11596.

Stanley, S. A., S. Raghavan, W. W. Hwang and J. S. Cox (2003). "Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system." Proceedings of the National Academy of Sciences **100**(22): 13001-13006.

Stotz, M., O. Mueller-Cajar, S. Ciniawsky, P. Wendler, F. U. Hartl, A. Bracher and M. Hayer-Hartl (2011). "Structure of green-type Rubisco activase from tobacco." Nature structural & molecular biology **18**(12): 1366-1370.

Thompson, J. D., T. Gibson and D. G. Higgins (2002). "Multiple sequence alignment using ClustalW and ClustalX." Current protocols in bioinformatics: 2.3. 1-2.3. 22.

Wagner, J. M., T. J. Evans and K. V. Korotkov (2014). "Crystal structure of the N-terminal domain of EccA(1) ATPase from the ESX-1 secretion system of Mycobacterium tuberculosis." Proteins **82**(1): 159-163.

Whitmore, L. and B. A. Wallace (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases." Biopolymers **89**(5): 392-400.

PART B:

STRUCTURAL AND FUNCTIONAL ANALYSIS OF MYCOBACTERIUM TUBERCULOSIS PROTEINS DPRE1 AND DPRE2 INVOLVED IN CELL WALL SYNTHESIS

REVIEW OF LITERATURE

Mycobacterial cell envelope as drug target

Apart from protein export systems (Tat, SecA and Esx-1) mycobacterial cell envelope provide agreat advantage against host immune system. Mycobacterial cell envelope is a unique and exceptional structure consisting of sugars and lipids. Nearly 40% of cell dry mass consist of lipids which are accounting for mycobacterial cell impermeability to drugs and for its unique staining properties. (Arora, Foster et al. 2013)(Banerjee, Dubnau et al. 1994)(Stanley and Cox 2013). Mycobacterial cell envelope lipids play an important role in host-pathogen interactions. Isoprenoid lipids, glycerophospholipids, phosphatidylinositol mannosides, lipoarabinomannan, mycolic acids, and trehalosemycolates are essential lipids produced by all the mycobacterial species and biosynthesis pathways of cell envelope lipids always have been as apotential target for numbers of anti-tuberculosis drugs e.g. Ethionamide, Isoniazid etc.

Mycobacterial cell envelope, (Fig.1) is very complex and multi-layered structure consists of a plasma membrane, a covalently linked mycolic acid, Arabinogalactan and peptidoglycan complex and polysaccharide containing capsule. The mycolic acids are the most abundant lipid and covalently linked to Arabinogalactan (AG) play important roles in the permeability of the mycobacterial cell envelope, biofilms formation and *M. tuberculosis* pathogenicity (Draper 1998)(Ojha, Anand et al. 2005)(Sambandan, Dao et al. 2013)(Yuan, Crane et al. 1998) and functions as a major virulence determinant. During the phagocytosis of *M. tuberculosis* into dendritic macrophages several lipids PIM, TDM, and ManLAM are predominant to mediate host-pathogen interaction (Schlesinger, Azad et al. 2008)(Ishikawa, Ishikawa et al. 2009). Several *M. tuberculosis* lipids including diacylated forms of ManLAM, SL, PIM, MPM, and mycolyl lipids, represent themselves as antigens and are presented to T lymphocytes by Antigen presenting cells (APCs) through MHC-I- molecules (Ly, Kasmar et al. 2013)(Arora, Foster et al. 2013)(Guiard, Collmann et al. 2009) . A few numbers of mycobacterial lipids are known to as potential subunit vaccines (Arora, Foster et al. 2013)(Banerjee, Dubnau et al. 1994). Interestingly, it has been also reviewed that these lipids may directly insert in to host membrane and alter their fluidity and affect phagocytosis (Banerjee, Dubnau et al. 1994).
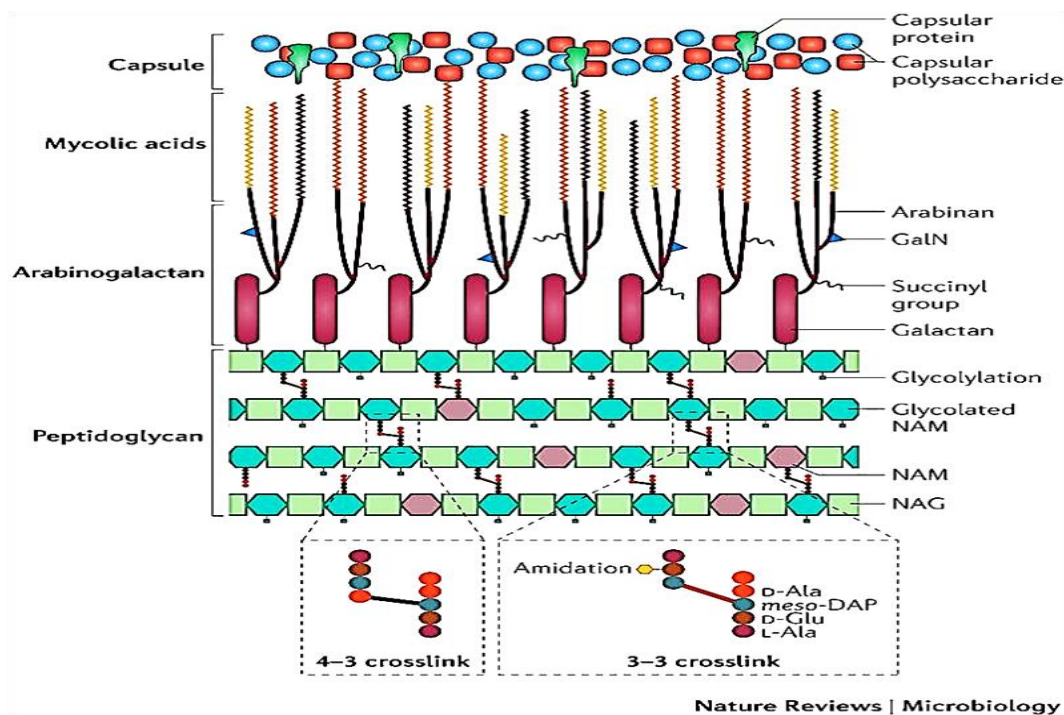
Fig. 1 Schematic representation of Mycobacterial cell envelope consisting of polysaccharide Capsule, Mycolic acid layer covalently linked to Arabinogalactan layer, Peptidoglycan layer cross-linked to thecell membrane.

Mycobacterial cell envelope component biosynthesis pathways have always a great potential as novel drug target sites. In thecurrent, a number of inhibitors targeted to these pathways, for example, themycolic acid pathway is a target of Isoniazid and Ethionamide kind of first and second line of antituberculosis drugs (Arora, Foster et al. 2013). Antigen 85 complexes, (Antigen 85, Mycoloyl transferases, Pks13 and FadD32 enzymes) have pivotal roles in assembly and transfer of mycolic acid in thecell envelope, attracting researchers as a novel drug target.

In recent DPA biosynthesis pathway has taken attention of researchers as a potential drug target. DPA is the only arabinofuranose (*araf*) donor in the biosynthesis of Arabinogalactan and ManLAM which are essential components of themycobacterial cell envelope. Biosynthesis of DPA takes place by an epimerization reaction involving DPR oxidation to DPX by enzyme DprE1 and reduction of DPX to DPA by the enzyme DprE2. Both DprE1 and DprE2 enzymes work synergistically to convert DPR to DPA. Target based and thewhole cell based screening of numbers of inhibitors against *M. tuberculosis* has produced more than 15 classes of inhibitors of *M. tuberculosis* DprE1 which is required for DPA biosynthesis. In recent, the structural basis of

DprE1 inhibition has been revealed for several inhibitors (e.g. BTZ, DNB, TCA1, nitroquinoxalinesetc.). These inhibitors inhibit DprE1 either covalently or non-covalently for example nitro group containing BTZ and DNB classes of inhibit DprE1 by making covalent adduct with Cys387 in *M. tuberculosis*(Batt, Jabeen et al. 2012) and CYS 394 in M. smegmatis(Neres, Pojer et al. 2012) while TCA1 inhibit DprE1 noncovalently in the same binding pocket with improved efficacy (Wang, Sambandan et al. 2013). Nowadays few of these inhibitors are reported to be in preclinical development trial phase.

Connection between Esx secretion system and mycobacterial cell envelope

Numbers of virulence factors are exported from Esx system across the cytoplasmic membrane then translocate through cell envelope outside of the cell either into host cytosol during infection or in culture medium during culture. Esx systems located on themycobacterial cell membrane and whole mechanism of transport of virulence factors across the MOM (mycolate outer membrane) and polysaccharide capsule is still not well understood. Although there is no direct evidence, However, there is a close relation between Esx machinery and cell envelope components to facilitate export of Esx secretion substrates across the complex MOM, apeptidoglycanlayer, and polysaccharide capsule. EccB1, a component of Esx-1 system extended throughout periplasmic space and interact with peptidoglycan layer (peptidoglycan binding protein) (Wagner, Chan et al. 2016) and similarly several MOM binding proteins may provide insight into connection between Esx-1 system and secretion mechanism and recently it was reviewed that Genes encoding the ESX systems and those required for the biosynthesis of the mycobacterial envelope are co-regulated (Bosserman and Champion 2017).

Introduction

Mycobacterial cell envelope is very complex and multi-layered structure consist of a plasma membrane, a mycolic acid layer covalently linked to Arabinogalactan, peptidoglycan complex and outermost polysaccharide containing capsule (Fig.1). This multi-layered cellular architecture provides strong immunity to mycobacterial bacilli against host immune system and plays a key role in successful survival within the phagosomal compartments during infection. The various components of mycobacterial cell envelope contribute the ability of cell envelope to impermeable to numbers of biocides and resistance to host immune system. Because of this

complex cell envelope, it is difficult to penetrate mycobacterial cell envelope and target a drug molecule whose target site can be within the cell.

DprE1 (~50 kDa) and DprE2 (~27.8KDa) are two subunits of an enzyme complex Decaprenylphosphoribosyl-β-d-2` Epimerase, encoded by two genes*Rv3790* and *Rv3791*. DprE1 and DprE2 involved in theconversion of DPR (Decaprenylphosphoribosyl –β-D-ribose) to DPA (Decaprenylphosphoribosyl –β-D-arabinofuranose) by epimerization reaction (fig.2). The DPA is required for the synthesis of D-arabinofuranose, a building block of Arabinogalactan and LAM (Lipoarabinomannaas) in themycobacterial cell wall (Makarov, Manina et al. 2009). DprE1 is the target of BTZ (Benzothiazinone) and DNB (Dinitrobenzamide) are two important classes of drugs, is anessential enzyme in *M. smegmatis*(Neres, Pojer et al. 2012). The epimerization of DPR to DPA takes place in two sequential oxidation and areduction reaction (Crellin, Brammananth et al. 2011).
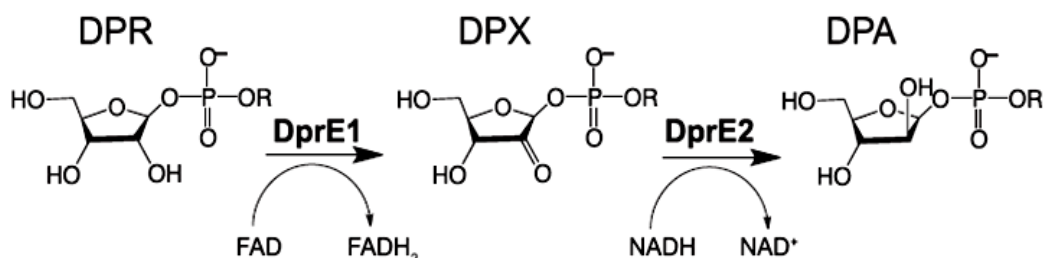
Fig.2. Chemical reaction showing two step conversion of DPR to DPA through DprE1 and DprE2 enzymes

M. tuberculosis DprE1 (Rv3790) is a 1526bp long FAD containing oxidoreductase, essential for *M. tuberculosis* viability and involved in cell wall synthesis. It converts DPR to an intermediate keto moiety DPX in an oxidation reaction. Crystal structure of Mtb DprE1 (PDB 4FDN) showed that it is a mix α-β helix sheet stable structure and divided into the substrate-binding domain (amino acid residues 197–412) and co-factor FAD-binding domain (residues 7–196 and 413–461). The FAD binding domain consists of theα+β fold, having two small β-sheets (strands β1–β4 and β5–β9, respectively), and several helices (α1–α4 and α11–α13) and co-factor FAD non-covalently deeply buried in the FAD-binding domain. The substrate-binding domain consists of an extended, antiparallel β-sheet (strands β10–β16) and contains two disordered loops

(269-297 and 316-330 residues) which possibly either involved in ligand binding or its interaction with a*Mtb*DprE2 enzyme (Batt, Jabeen et al. 2012).

M. tuberculosis DprE2 (*RV 3791*) is a 736 bp long NAD containing dehydrogenase that converts the intermediate keto moiety DPX to DPR in a reduction reaction. Thus Mtb DprE1 and Mtb DprE2 acts synergistically in the conversion of DPR to DPA. Genetic studies have shown that both DprE1 and DprE2 are essential for mycobacterial cell growth and survival and offers a good target for antituberculosis drug development (Makarov, Manina et al. 2009).

The various classes of inhibitors have potent antimycobacterial activity against MDR and XDR strains of mycobacterium. The nitro group (NO$_2$) containing BTZ and DNB classes of inhibitors (Sambandan, Dao et al. 2013) inhibit DprE1 enzyme by making covalent bond with Cys387 amino acid residue in *M. tuberculosis*(Batt, Jabeen et al. 2012) and Cys394 amino acid residue in *M. smegmatis*(Neres, Pojer et al. 2012). TCA1 is another class of small inhibitor that inhibit *M. tuberculosis* DprE1 by making noncovalent adduct with improved MIC value in comparison to BTZ and DNB classes of inhibitors. The binding site of covalent and noncovalent inhibitors overlap significantly and they lie in the same substrate binding pocket of *Mtb*DprE1. On the basis of Transcriptional analysis TCA1 down-regulates genes known to be involved in *Mtb* persistence during infection(Wang, Sambandan et al. 2013). An analogous mechanism of resistant to Benzothiazinone and Dinitrobenzamide in *M. smegmatis* has been reported but *M. tuberculosis* lacks this alternative mechanism to synthesize arabinofuranose(Ribeiro, Degiacomi et al. 2011). This action interrupts whole downward pathway and leading to blockage cell wall synthesis and ultimately *M. tuberculosis* cell death.

Recently, sulfur rich 2-mercaptobenzothiazoles ligands and their 1, 2, 3-triazoles conjugates (obtained from click chemistry) have been studied for anti-tuberculosis activity (Mir, Shafi et al. 2014). These ligands have shown *in vivo* bactericidal activity and completely inhibited the mycobacterial growth at aconcentration of 12-64 µg/ml. To examine the structure and binding mechanism of these ligands with *Mtb*DprE1, we have purified the *Mtb*DprE1 and synthesized six compounds. Binding analysis between *Mtb*DprE1 and six ligands is performed using surface Plasmon resonance technique. Molecular docking of six ligands with *Mtb*DprE1 is performed using GLIDE module of Schrödinger and analyzed the mechanism of interactions.

On the basis of extensive research performed on *M. tuberculosis* by ascreening of various drugs or inhibitors, targeting DprE1 and DprE2 can be good targets for the development of new anti-tuberculosis drugs. The structural analysis of these proteins and inhibitors and complex of these will reveal aclearer story about DprE1 and 2 catalysis and inhibition Mechanism and biology.

The objective of our work is the structure and function analysis of *M. tuberculosis* DprE1 and DprE2 proteins. We have expressed and purified the both *M. tuberculosis* DprE1 and DprE2 proteins using thestandard chromatographic technique. Extensive crystallization trials have been performed on both the proteins alone and in complex with ligands. We have modeled the structure of both *M. tuberculosis* DprE1 and DprE2 proteins and performed the interaction study with their ligands using surface Plasmon Resonance, molecular modeling, and docking and techniques.

Specific objectives

1. **Preparation of both *M. tuberculosis* proteins DprE1 and DprE2**

    Following techniques were used to achieve current objectives

    -   Gene cloning and protein expression of DprE1 and DprE2 in *E. coli.*
    -   Protein purification of DprE1 and DprE2 using standard chromatographic techniques.

2. **Binding analysis between *M. tuberculosi*s DprE1and DprE2 proteins and between *M. tuberculosi*s DprE1 protein and different inhibitors by SPR**

    Following techniques were used to achieve current objectives

    -   Binding between DprE1 and DprE2 using SPR.
    -   A binding study between DprE1 and Different inhibitors.

3. **Structure analysis of native and complexes of DprE1/DprE2 enzymes and with ligands.**

Following techniques were used to achieve current objectives

- Molecular docking study between DprE1 and its inhibitors
- DprE2 modeling and docking study with NAD

Thesis Organization

Chapter 1 describes the cloning and purification of M. tuberculosis DprE1 and DprE2 proteins, crystallization trials of DprE1 and DprE2 and secondary structure and Thermal denature analysis of DprE2 through CD. Both the proteins were cloned into pET32b and pET-SUMO vectors and expressed in *E. coli* expression system. Both the proteins were purified using standard chromatographic techniques followed by proteolytic digestion.

Chapter 2 describes the binding analysis between *M. tuberculosis*DprE1and DprE2 proteins and between *M. tuberculosi*s DprE1 protein and different inhibitors using SPR. Surface Plasmon Resonance technique was used to get K_D value.

Chapter 3 describes the *in silico* Structure analysis of native and complexes of DprE1/DprE2 enzymes and with ligands. Molecular docking and simulation dynamics analysis were performed on theDprE1-Inhibitor complex to dissect the inhibition mechanism and DprE2 molecular modeling and docking study with NAD was performed.

CHAPTER 1: CLONING, EXPRESSION, AND PURIFICATION OF M.TUBERCULOSIS PROTEINS DPRE1 AND DPRE2

Materials and Methods

Gene cloning of *M. tuberculosis* and *M smegmatis* DprE1/DprE2 in pET32b Vector

The genes encoding *M. tuberculosis* and *M. smegmatis* DprE1 and DprE2 proteins were cloned into apET32b vector containing a Thioredoxin and 6xHis tag at N-terminal. The DNA fragments encoding *M. tuberculosis* DprE1 and DprE2 genes were amplified *from M. tuberculosis H37Rv* genome using PCR. PCR amplified products were digested with *Nco1* and *Xho1* restriction enzymes and inserted into apET32b vector (*Novagen*). The successful cloning was confirmed by restriction digestion and PCR amplification. Expression of *M. tuberculosis* DprE1 was observed in BL21 (DE3) cell line as insoluble but in Arctic express cell line at 12^0C we got dprE1 in soluble fraction whereas DprE2 expressed as soluble in BL21(DE3) cell line at 16^0C.

Gene Cloning of *M. tuberculosis* and *M. Smegmatis* DprE1/DprE2 in pET-SUMO vector

The genes encoding *M. tuberculosis* and *M. smegmatis* DprE1 and DprE2 were amplified by PCR using their pET32b constructs as atemplate. PCR amplified products were ligated into apET-SUMO vector containing a SUMO and 6xHis tag at N-terminal through TA-Cloning method. The successful cloning was confirmed by PCR amplification. Expression of *M. tuberculosis* DprE1 and DprE2 proteins were observed in BL21 (DE3) cell line.

M. tuberculosis DprE1 protein expression and purification

M. tuberculosis DprE1 was cloned into apET32b vector with N–terminal Thioredoxin (17KDa) and 6x His tag having *Enterokinase* proteolytic cleavage site between protein and tags. For over expression the plasmid encoding DprE1 protein was transformed into Arctic express cells and the positive colonies were selected on the LB media containing 100μg/ml ampicillin. A single colony was inoculated in LB media containing 100μg/ml ampicillin and grown at 37^0C for 2.5-3 hrs. At OD_{600} the cultures were induced with 0.5mM IPTG for 4 more hrs. The culture was centrifuged and the pellet was re-suspended in 2X sample buffer. The samples were run on 12%

SDS-PAGE. The band corresponding to the molecular weight of DprE1 (~51kda) was detected on SDS-PAGE and Protein expression was confirmed by MALDI-TOF mass spectrometry.

For protein purification, 3 L culture was grown using LB media with several additives (Glucose, NaCl, and Glycerol). At OD 0.6 culture was added 5mM Proline then thetemperature was reduced to 12^0C and after 30 min induced with 0.12mM IPTG and incubate for next 24 hour at 12^0C. The culture was centrifuged at 6000 rpm for 5min and the cell pellet was re-suspended in 100 ml of resuspension buffer consisting of 50mM Tris pH7.5, 250mM NaCl, 5% Glycerol, 1mM PMSF, 4mM β-ME, 0.05% Triton X-100, 10mM $MgCl_2$ and 0.2 mg/ml Lysozyme. The cells were disrupted by sonication and the crude lysate was centrifuged at 13000 rpm for 15 min at 4^0C. The Obtained supernatant was allowed to bind with pre-charged Ni-NTA resin *(Sigma Aldrich)* for one and half hour at 4^0C. The column was washed with 10 ml of buffer A (consisting of 50mM Tris pH8.0, 500mM NaCl, 5mM $MgCl_2$, 4mM β-ME). Then 10ml of 5mM, 10mM, 40mM of Imidazole concentration with buffer A was given for washing. The protein was eluted with 15 ml of buffer A with 300mM Imidazole. The entire eluted fraction was run on 12% SDS-PAGE to check the pure fractions.

The eluted DprE1 fractions were pooled, concentrated using centricon (30 kDa cutoff) (Millipore, USA) till 1 ml and diluted up to 15 ml in 50mM Tris pH7.5 and 2mM EGTA solution (suspension buffer). The sample was passed through Q- Sepharose resin followed by washing with 40 ml of suspension buffer. The elutions were taken in 50mM Tris pH7.5 and different gradient of NaCl (50mM, 100mM, 150mM, 200mM, 250mM, 300mM and 500mM). Protein was started elute from 150mM NaCl to 500mM NaCl concentration.

All the elutions having protein were pooled and concentrated till 3 ml and diluted till 6 ml by adding Enterokinase buffer (50mM Tris pH 7.5, 50mM NaCl, 5mM $MgCl_2$, 2mM $CaCl_2$, 5 % Glycerol and 0.001 % NP 40). The protein was dialyzed against Enterokinase buffer for 3 hr and then add Enterokinase enzyme in 1:1000 ratio (w/w) and incubated for 12 hr by dialysis at 4^0C. After this 2mM PMSF was added to inhibit Enterokinase enzyme and diluted till 10 ml by adding buffer A. the sample was passed twice through Ni-NTA resin to allow uncleaved DprE1 and cleaved Trx-tag to bind with resin followed by 5 ml washing of buffer A. All the samples were run on the 12% SDS PAGE to check the purity, cleaved DprE1 protein came in sample

flow and wash and both were pooled and concentrated till 5 ml and loaded on gel filtration column (High load 16/60 Superdex 200 pg, *GE Healthcare*). The column was pre-equilibrated with buffer containing 50mM Tris pH 8.0, 500mM NaCl and 10% Glycerol, 4mM β-ME, 10mM MgCl$_2$, 0.5mM ATP. The peak fractions were pooled and concentrated using centrifugal filter device (*Millipore, USA*). The purity of theprotein was analyzed on SDS-PAGE.

DprE2 protein expression and purification

M. tuberculosis DprE2 protein was over expressed in BL21 (DL3) cells of E. coli. For over expression the plasmid (*DprE2-pET-Sumo*) encoding DprE2 gene was transformed into BL21(DE3) cells and the positive colonies were selected on the LB media containing 50μg/ml Kanamycin. A single colony was inoculated in LB media containing 50μg/ml Kanamycin and grown at 37C for 2.5-3 hrs. At OD$_{600}$ the cultures were induced with 1mM IPTG and further incubated for 3 more hrs. The culture was centrifuged and the pellet was re-suspended in 2X sample buffer. The samples were run on 12% SDS-PAGE. The band corresponding to the molecular weight of DprE2 (~27KDa) was detected on SDS-PAGE.

For protein purification, 3 L culture was grown and induced with 0.5mM IPTG at 16^0C and followed by overnight incubation at 16^0C. The culture was centrifuged at 8000 rpm for 10 min and the cell pellet was re-suspended in 50 ml resuspension buffer consisting of 20mM Tris pH8.0, 300mM NaCl, 5% Glycerol, 1mM PMSF, 1mM Benzamidine-HCL, 2mM β-ME and 0.3 mg/ml Lysozyme. The cells were disrupted by sonication and the crude lysate was centrifuged at 13000 rpm for 20 min at 4^0C. For Ni-NTA affinity chromatography, obtained supernatant was loaded on pre-charged Ni-NTA resin (*Sigma Aldrich*). The column was washed with buffer consisting of 20mM Tris pH8.0, 300mM NaCl, 5% Glycerol, 1mM PMSF, 1mM BHCl, 2mM β-ME and 15mM Imidazole. The protein was eluted in a buffer consisting of 20mM Tris pH8.0, 300mM NaCl, 5% Glycerol, 1mM PMSF, 1mM BHCl, 2mM β-ME and 250mM Imidazole. The entire eluted fractions were run on 12% SDS-PAGE. The eluted DprE2 fractions were pooled, concentrated using centrifugal filter device (*Millipore, USA*) and loaded on gel filtration column (High load 16/60 Superdex 200 pg, *GE Healthcare*). The column was pre-equilibrated with buffer containing 20mM Tris pH8.0, 300mM NaCl, and 5% Glycerol. The peak fractions were

pooled and concentrated to 10-12 mg/ml using centrifugal filter device (Millipore, USA). The purity of theprotein was analyzed on SDS-PAGE.

The cleavage of SUMO tag from fused DprE2 protein was performed by SUMO protease in 500µl reaction volume in SUMO protease buffer (50mM Tris-pH8.0, 0.2% Igepal (NP-40), 150mM NaCl, and 10mM DTT) at 4^0C. The aliquots were taken out at different incubation intervals like 4 hrs, 8 hrs, and 16 hrs and at 24 hrs. The tag cleaves was analyzed on SDS-PAGE.

The cleavage time 8 to 16 hour was further optimized till 12 hr. at 4^0C. After the cleavage reaction protein was passed twice through Ni-NTA column followed by 5ml wash of buffer consisting of 20mM Tris pH8.0, 300mM NaCl, 5% Glycerol, 2mM β-ME. Protein came into sample flow and wash fraction. These fractions were run on the SDS PAGE to check purity and pure fractions were pooled and loaded on Superdex 200 (16/60) gel filtration column to purify theDprE2 protein.

Crystallization of Mtb DprE1 and DprE2

Extensive crystallization attempts were made on purified DprE1 and DprE2 proteins using various screens *e.g.* PEG/Ion, Index, Salt Rx, Crystal Screen1&2 from *Hampton research* and PCAT premier, JCSG, Morpheus screens and Structure screen1&2 from the*Molecular dimension*. Different concentrations (5mg/ml, 10mg/ml and 15mg/ml) of DprE1 and DprE2 proteins and different temperatures (4^0C and 16^0C) were tried in crystallization experiment. Several attempts were also made to crystallize purified M. *tuberculosis* dprE1 protein with newly synthesized anti tuberculosis compounds e.g. Sulfur rich 2-mercaptobenzothiazole and 1, 2, 3-triazole conjugates. DprE1 protein was incubated with these compounds in 1:1 molar ratio for different time intervals (2h, 4h, 12h and 24h) at 4^0C and then 0.4µl of drops in 1:1 ratio (protein/reservoir) were aided at 4^0C using a Mosquito (*TTP Labtech*) liquid handling robot at AIRF,JNU, NEW DELHI.

Circular dichorism and thermal denaturation analysis of DprE2

CD measurements were recorded using Chirascan^{TM} CDspectropolarimeter (*Applied Photophysics*) with a water bath to maintain the constant temperature (25^0C). The DprE2 protein

was diluted to 0.1 mg/ml in 10mM potassium phosphate buffer pH 8.0 and loaded in a0.1cm quartz cuvette. The blank of all experiments was 10mM potassium phosphate buffer pH 8.0. The final spectrum was anaverage of three sequential scans. All CD data were converted to mean residue ellipticity (deg.cm^2/dmol).

The Dichroweb server (Whitmore and Wallace 2008) was used to estimate the secondary structures from CD spectra. The SOPMA (Geourjon and Deleage 1995), GOR (Sen, Jernigan et al. 2005) and PSIPRED (Jones 1999) programs were used to estimate theoretical secondary structure for DprE2 and. To study thethermal stability of DprE2, CD-spectra were recorded as a function of temperature. For DprE2, the CD spectra were recorded from 25^0C to 85^0C in 10^0C increment.

Results and discussions

Mtb DprE1 and DprE2 cloning in pET32b and pET-SUMO vector

The genes encoding *Mtb* DprE1 (Rv3790) and *Mtb* DprE2 (Rv3791) were successfully cloned into pET32b vector plasmid with N-terminal Trx solubilization tag (~17KDa) and 6x His tag followed by Enterokinase cleavage site (*Asp-Asp-Asp-Asp-Lys*). Both the gene was also cloned into apET-sumo linear vector by TA-cloning technology with N-terminal sumo solubilization tag (~11KDa) and 6X his tag (*Invitrogen* kit) followed by sumo protease cleavage site (*Gly-Gly-/-Ala-Thr-Tyr*).
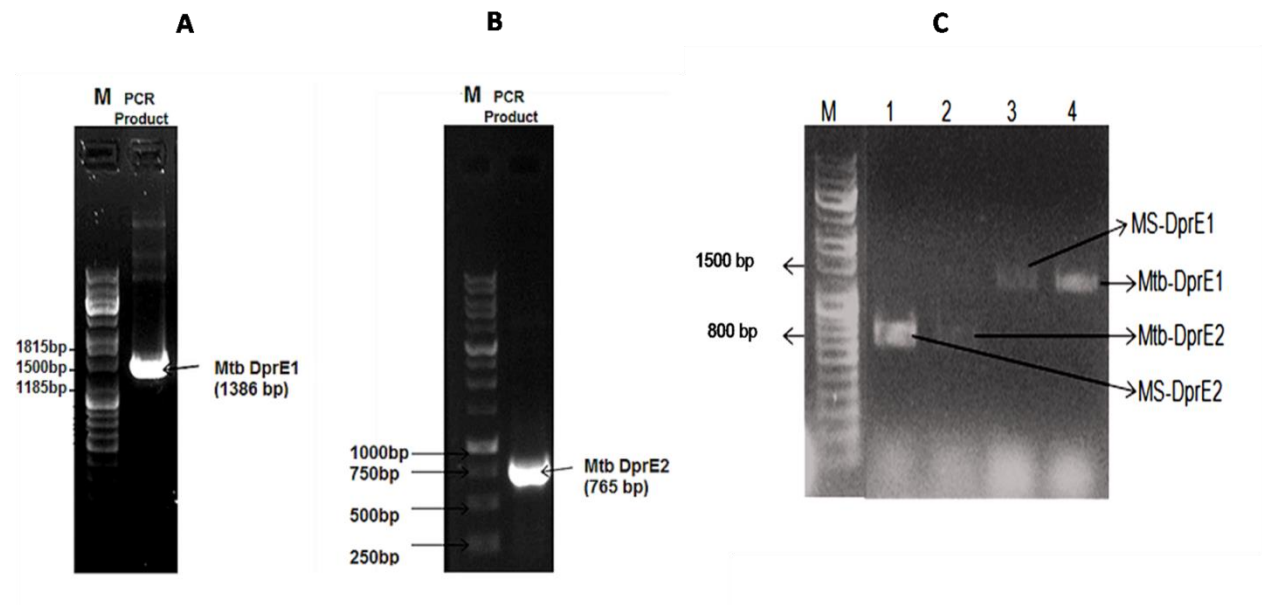
Fig.3. *M. tuberculosis* clone confirmation by PCR amplification using gene construct as template. A.*M. tuberculosis* DprE1-pET32b construct. B.*M. tuberculosis* DprE2-pET32b construct, C, *M. tuberculosis* and *M. smegmatis*DprE1 and DprE2.

Mtb DprE1 protein purification

Mtb DprE1-pET32b plasmid was transformed into different E. coli. Cell lines (BL21 DE3, Rosetta-gamy, Codon$^+$ RIL, C-41, C-43, and Arctic express cell line). In Arctic express cell line at 12^0C DprE1 was expressed as soluble protein. We have modified the media by the addition of few additives like 5mM glucose, 2mM NaCl, 1% glycerol during LB

media(*Miller,Himedia*) preparation then autoclaved. Before IPTG induction we added 5mM autoclaved Proline in the media then cool down the culture till 12^0C and induced with 0.125mM IPTG followed by 24 hr post induction incubation.

Fig.4. A. Schematic drawing of DprE1 and its expression construct.B. Gel Filtration profile cut DprE1 enzyme. The DprE1 eluted as monomer from gel filtration column as indicated by molecular weight standard. The DprE1 shows single band on SDS-PAGE after digestion. C. Enterokinase cleavage of DprE1 protein (M-Marker, Lane 1 and 2- Enterokinase digested DprE1 and lane 3-uncut DprE1 protein).

Thus we were able to get DprE1 protein as soluble. Mtb DprE1 protein purified using Ni-NTA Chromatography, anion exchange chromatography and Enterokinase proteolytic cleavage followed by gel filtration chromatography. *Mtb* DprE1 Eluted as amonomer from Superdex200 (16/60) pg column (*GE Healthcare*).

Mtb DprE2 protein purification

Mtb DprE2-pET-sumo plasmid was transformed into *E. coli*. BL21 (*DE3*) cells.In BL21 (*DE3)* cell line at16^{0}C and 0.5mM IPTG induction, we were able to get DprE2 as a soluble protein. We have successfully purified *Mtb*DprE2 protein using Ni-NTA chromatography, gel filtration chromatography, and sumo protease cleavage. The protein eluted in the void volume of Superdex200 16/60 pg column which indicated the oligomeric state of the DprE2 protein. Different additives were used to stabilize DprE2 but we never succeed.
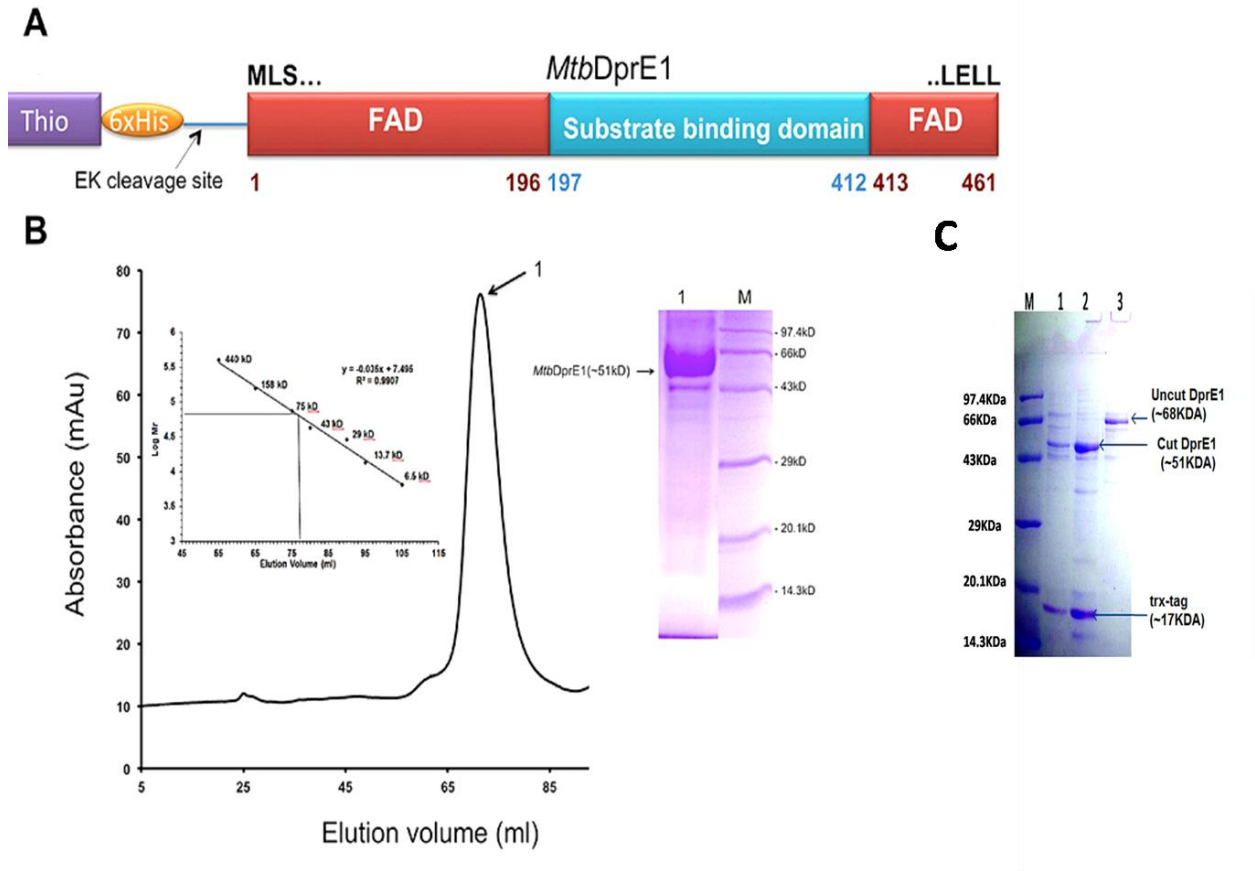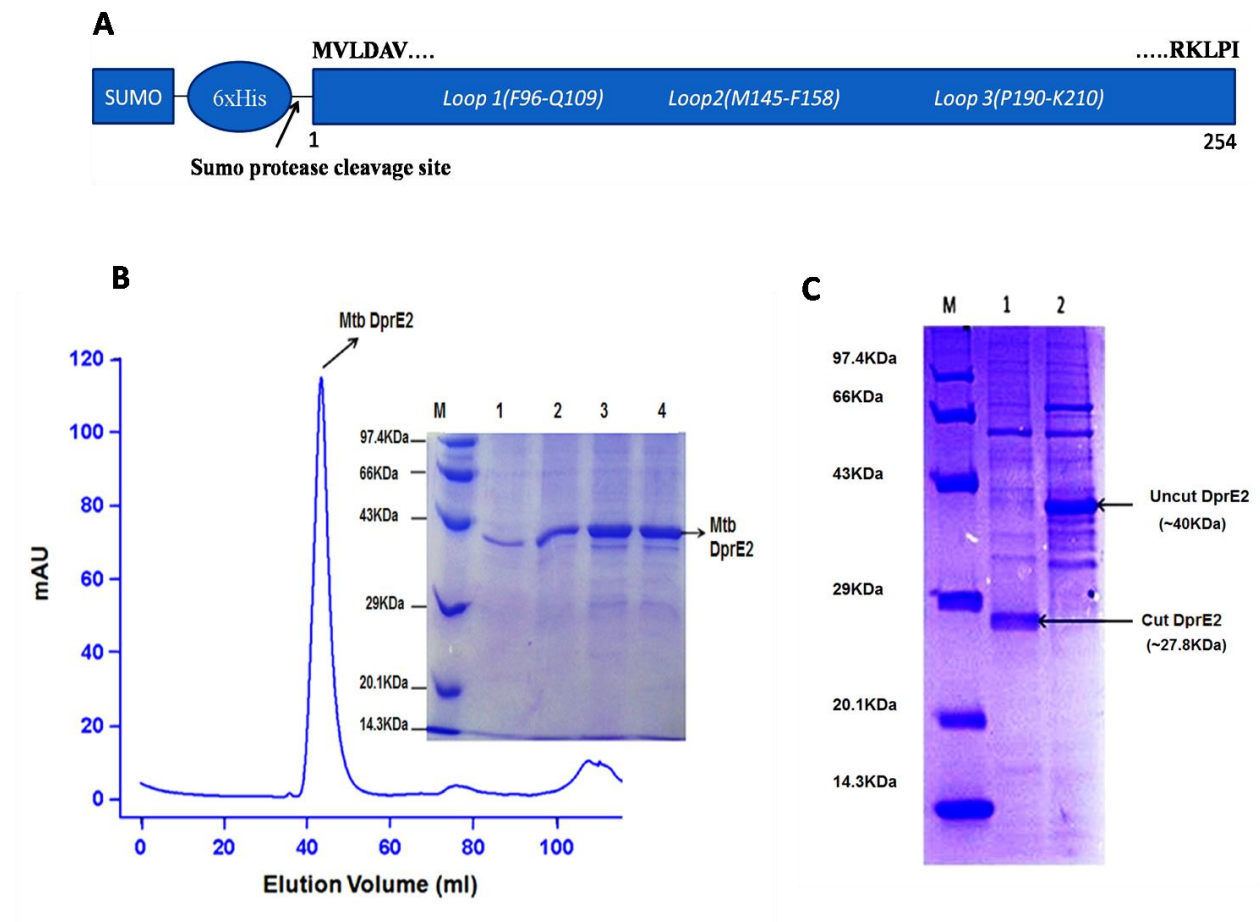

Fig.5. A. Schematic drawing of DprE2 and its expression construct.B. Gel Filtration profile cut DprE2 enzyme. The DprE2 eluted as oligomer from gel filtration column as indicated by molecular weight standard. The DprE2 shows single band on SDS-PAGE after digestion. C. Enterokinase cleavage of DprE1 protein (M-Marker, Lane 1 a- Sumo protease digested DprE2 and lane 3-uncut DprE2 protein).

Crystallization of Mtb DprE1 and DprE2

Several crystallization attempts were made on both purified DprE1 and DprE2 proteins using various screens to get initial crystallizations buffer conditions. Different concentrations of proteins, temperature variation, sitting and hanging drop vapor diffusion methods, all were failed to produce crystals of DprE1 and DprE2 proteins.

Secondary structure and thermal stability analysis of DprE2

Far-UV CD spectra (260-200 nm) were obtained for DprE2 to estimate the contents of the secondary structure of the protein. The Dichroweb server[27] analysis yielded (~46% α-helix, 28% β sheets and ~26% random coil) for DprE2 (Fig. 6A) These values were close to secondary structures obtained from theoretical (~49% α-helix, ~30% β sheets and 21.7% random coil) analysis for DprE2.
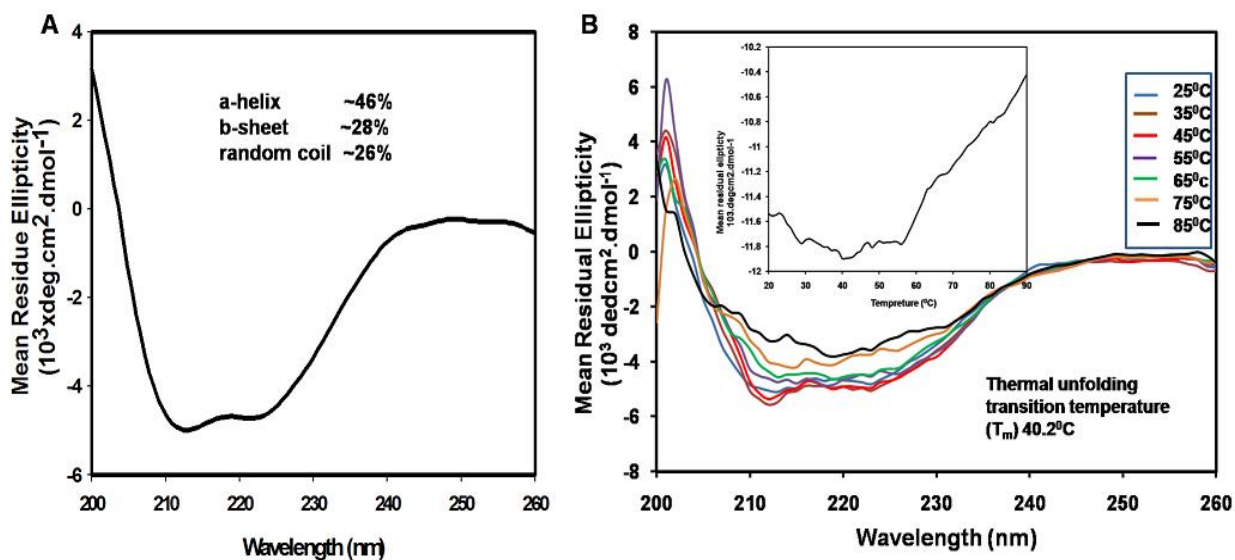
Fig. 6 Circular dichorism and thermal stability profile of *M. tuberculosis* DprE2. A-B. Circular dichorism and thermal stability profile of EspC. Spectra were recorded at room temperature and normalized to mean residue ellipticity. The secondary structure contents of DprE2 matches well with secondary structural contents obtained from theoretical predictions. DprE2 proteins is quite thermostable

For thermal stability analysis, the CD spectra of DprE2 and was recorded as a function of temperature. Effect of temperature on theDprE2 secondary structure was performed from 25^0C to 85^0C in 10^0C temperature increment (Fig. 6B) and yielded thermal unfolding transition temperature (T_m) 40.2 ± 1.0^0C.

CHAPTER 2 BINDING ANALYSIS BETWEEN OF M. TUBERCULOSIS PROTEINS DPRE1-DPRE2 AND BETWEEN DPRE1- DIFFERENT INHIBITORS USING SPR

Materials and Methods

Mtb DprE1-DprE2 binding study using SPR

Auto lab SPR was used at the Advanced Instrumentation Research Facility, Jawaharlal Nehru University, New Delhi to assess the binding of MtbDprE1 and MtbDprE2 in real time. The Binding of Mtb DprE2 (15.625μM to 100 μM) with previously immobilized DprE1 was done in running buffer consisting of 20 mM Sodium-Phosphate buffer pH 7.4, 150 mM NaCl, 5mM $MgCl_2$, 2mM $CaCl_2$ and 2% Glycerol- at 25 °C. The association kinetics was monitored for 300 seconds followed by dissociation for the next 300 seconds. Different concentrations (3.125 μM, 6.25 μM, 12.5 μM, 25 μM, 50 μM and 100 μM) of Mtb DprE2 were injected over immobilized Mtb DprE1 with a flow rate of 20μl per minute across the sensor chip. The binding surface was then regenerated using 50 mM NaOH. Before analysis, all the SPR data were subtracted with data obtained from theblank. The results were plotted as a differential response to binding of different inhibitors to Mtb DPRE1. All the Binding experiments were performed in triplicates at 25^0C.

Synthesis and inhibition analysis of six sulfur rich 2-mercaptobenzo-thiazole and its 1, 2 ,3-triazole conjugates ligands against *Mtb*DprE1

All six ligands in Table 1 are synthesized according to material & method section of the published paper (Mir, Shafi et al. 2014).^{1}H NMR spectra on each ligand are determined using Bruker (400 MHz) spectrometer and chemical shifts are expressed as ppm against TMS as aninternal reference. Mass spectra on ligands are recorded at 70eV (Shimadzu, Japan) and column chromatography is performed on silica gel 60.

Autolab SPR was used at the Advanced Instrumentation Research Facility, Jawaharlal Nehru University, New Delhi to assess the binding of different inhibitors with Mtb DPRE1 in real time. For this, the surface (a self-assembled monolayer of 11-mercaptoundecanoic acid [11-MUA] on agold surface, Autolab) was first activated with N-hydroxysuccinimide (NHS, 0.05 M)/N-ethyl-N-(diethylamino-propyl) carbodiimide (EDC, 0.2 M). Then, M.Tb DPRE1 was immobilized to

the activated sensor surface at a concentration of 30µM in filtered (0.22µm) and degassed 20 mM sodium acetate buffer (pH 4.2). After ligand immobilization, the surface was blocked with 100 mM ethanolamine at pH 8.5 followed by regeneration using 50 mM NaOH. Binding of different inhibitors (15.625µM to 500 µM) with immobilized DPRE1 was done in SPR running buffer consisting of 20 mM Sodium-Phosphate buffer pH 7.4, 150 mM NaCl, 5mM $MgCl_2$, 2mM $CaCl_2$ and 2% Glycerol with aflow rate of 30µl per minute at 25^0C. The association kinetics was monitored for 300 seconds followed by dissociation for the next 300 seconds. The binding surface was then regenerated using two pulses of 50 mM NaOH (20 sec for each). Before analysis, all the SPR data were subtracted with data obtained from theblank. The results were plotted as a differential response to binding of different inhibitors to Mtb DPRE1. Binding experiments were performed in triplicates.All the inhibitors were dissolved in 20% methanol and 20mM stock of all the inhibitors was maintained at 4^0C. All dilutions were prepared in SPR running buffer.

Results and discussions

Mtb DprE1-DprE2 interacts in SPR binding study

The obtained SPR binding data were analyzed with Autolab SPR Kinetic Evaluation software. Different amounts of Mtb DprE2 sensorgram showing(Fig.6) different intensity of SPR signals. The SPR sensorgrams are showing thespecific binding affinity between Mtb DprE1 and Mtb DprE2 with K_D value of 29.2µM. The experimental evidence suggested that Both Mtb DprE1 and Mtb Dpre2 form complex to work synergistically for epimerization of DPR to DPA via DPX keto intermediate formation.
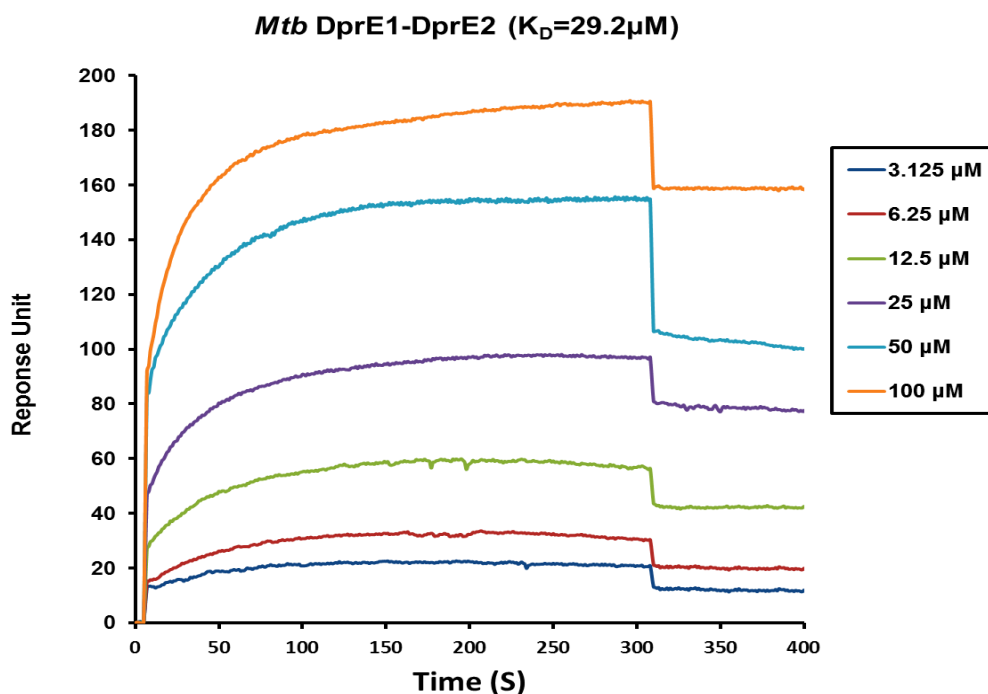
Fig.7 Binding analysis between the *M. tuberculosis* DprE1 and DprE2 using surface plasmon resonance where the different concentration of DprE2 protein was passed over the surface immobilized DprE1.

Synthesis and inhibition analysis of six sulfur rich 2-mercaptobenzo-thiazole and its 1, 2 ,3-triazole conjugates ligands against *Mtb*DprE1

Six different benzofused heterocyclic ligands containing 1, 2 ,3-triazole moiety (Table 1, Fig. 2) are synthesized(according to Materials and Methods section), as thiol rich heterocycles are effectiveanti-mycobacterial compounds. All ligands are characterized using ^{1}HNMR, IR and MALDI-MS/ESI-MS. These six ligands have been tested earlier for their anti-mycobacterial activities against *M. tuberculosisH37Rv* strain using microdilution assay(Arora, Foster et al. 2013). As seen in Table 1,except BTZ-4F (MIC<128 µg/ml), all fives conjugates showed minimal MIC values. Lowest MIC value (8 µg/ml) is observed for nitro derivatives of BTZ heterocycles *e.g.,* BTZ-4N, BTZ-3N, and BTZ-3Cl ligands.
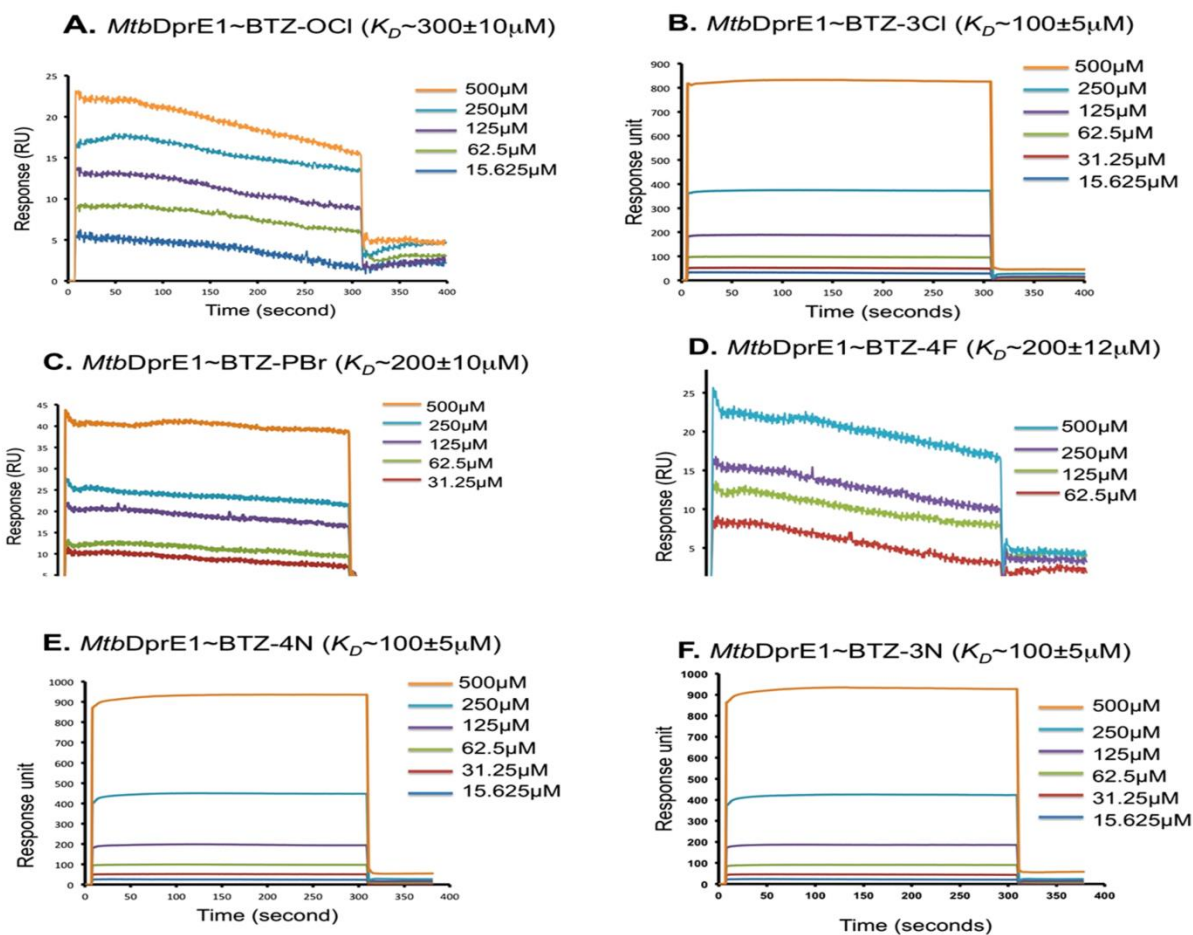
Fig. 8 Binding analysis of all six ligands with purified *Mtb*DprE1 is performed using surface plasmon resonance technique(Fig. 8A-F). The $K_D$value of each compound against *Mtb*DprE1 is shown in Table 1.As seen from Table 1,BTZ-Cl ligand showsthe lowest K_D value of 100μM against *Mtb*DprE1. Changing chloride from meta to ortho position in case of BTZ-OCl increased the K_D value to 3-folds (300μM) towards *Mtb*DprE1. In microdilution assay with *MtbH37Rv* strain, the BTZ-OCl ligand also shows the MIC value of 16μg/ml.Whenchloride group is replaced fromortho to meta position in case of BTZ-Cl, decreased the MIC value to 8 μg/ml consistent with binding data observed in surface plasmon resonance experiment.

Lowest K_D(~100μM) is obtained for BTZ-4N ligand against *Mtb*DprE1. Replacing nitro group from meta to para position did not change the K_D value for BTZ-3N (K_D~100μM). In earlier microdilution experiment, replacing nitro group from meta to para position in case of

69

BTZ-3N and BTZ-4N ligands did not bring any change in MIC value (~8μg/ml), consistent with data observed in our binding experiment.

The BTZ-PBr and BTZ-4F ligands show the K_D values of 200μM. No change in K_D is observed when bromide group is replaced with fluoride group. Substitution of bromide group at para position in BTZ-PBr increased the MIC value to 16 μg/ml, consistent with binding data. In thecase of BTZ-4F ligand, the MIC value is increased significantly to <128 μg/ml, inconsistent with binding data.These results indicate that nitro derivatives of 1, 2, 3-triazole conjugates have shown potential anti-mycobacterial activity rather than their halo-derivatives.

Table 1. Minimal inhibitory concentration (MIC) and $K_D$values of six DprE1 ligands

Compound	Structure	Molecular weight (Dalton)	MIC (μg/mL)	Equilibrium Dissociation constant (K_D)
1. BTZ-OCl		358.5	16	300μM
2. BTZ-3Cl		358.5	8	100μM
3. BTZ-PBr		403	16	200μM
4. BTZ-4F		342	<128	200μM
5. BTZ-4N		369	8	100μM
6. BTZ-3N		369	8	100μM

CHAPTER 3:IN SILICO STRUCTURE ANALYSIS OF NATIVE AND COMPLEXES OF DPRE1/DPRE2 ENZYMES AND WITH LIGANDS

Materials and methods

MtbDprE1-FAD structure preparation

The crystal structure of the *Mtb*DprE1-FAD-QN129 complex (PDB-4P8T) is used as aninitial model as it contains no disordered loop in *Mtb*DprE1 structure and theQN129 ligand is noncovalently bound to *Mtb*DprE1. Before thevirtual screening, raw *Mtb*DprE1 structure (PDB-4P8T) is converted to complete receptor model by using protein preparation wizard in *Maestro*program (*Maestro v.9.2*). Default settings are used, except missing side chains are filled and all crystallographic waters and QN129 ligands are removed. Exhaustive sampling is done to obtain better H-bond assignment at refinement step of protein structure. The docking receptor grid is created by Glide's receptor grid generation module (Glide v.5.7). The 5 Å grid box is generated and center was set to theCys325 residue of *Mtb*DprE1 binding pocket. The *Mtb*DprE1-FAD structure is subjected to constrained energy minimization step to remove any steric clashes.

MtbDprE1-FAD~ligand docking

Three-dimensional structures of all six ligands (**Table 1**)are obtained usingProDrg server (Rashmi and Swati 2015)(Pulaganti, Banaganapalli et al. 2014).The possible ionization state of each ligand is generated atpH7.0±2.0 using Epik module and 32 possible stereoisomers are generated for each compound. Docking of all six ligands isperformed using GLIDE module of SchrÖdinger-9.4 program (Bessonov, Vassall et al. 2017)(Ali, Kumar et al. 2015) using prepared *Mtb*DprE1-FADstructure as areceptor. QN129 ligand is used as areference for docking analysis with all six ligands. Structures of six ligands in complex with *Mtb*DprE1-FADare generated using induced fit docking (IFD) protocol of Schrödinger program (33). Default settings are used, except the XP (Extra Precision) scoring function is used in docking calculation.

The grid box and center are set at Cys387 residue of substrate binding pocket of *Mtb*DprE1 and no constraints are applied. During ligand docking using GLIDE program, the vander Waals radii of protein and ligand are scaled by afactor of 0.5. Only residues within 5 Å of

the ligands are refined forprime active site optimization. GLIDE uses two scoring functions *e.g.,*Emodel and Glide score. Emodel mainly comprises of protein ligand coulomb-van der Waals energy and Glide score is capable of selecting best binding pose of ligand from many poses in the active site. Glide score is an empirical scoring function and comprises of lipophilic-lipophilic terms, rotatable bond penalties, hydrogen bonds terms and contributions from protein-ligand coulomb-vander Waals energies. The empirical scoring function also includes the terms, which accounts for hydrophobic enclosures. For scoring each structure, 10,000 cycles of scoring is performed to identify the best binding pose of each ligand. After docking, 1000 cycles of post-docking minimization is performed and best pose for each ligand together with thescore is saved for further analysis.

The binding pose with lowest IFD score is considered the correct docked structure. Each ligand complex with an X-score (binding energy) and docking score is used to identify the correct docked structure (Table 2,Figures 9A-F).The PISA server (Krissinel and Henrick 2007) is used to analyze the interactions, binding free energy and solvent accessibility in the*Mtb*DprE1-FAD-ligand complex. The stereochemistry of all six *Mtb*DprE1-FAD-ligand complexes is checked by PROCHECK (Laskowski, MacArthur et al. 1993)program and structure visualization by PyMOL program (DeLano 2002).

Mtb DprE2 molecularModeling and docking with NAD

Mtb DprE2 protein structural homology modeling was performed using various programs like I-TASSER (Zhang 2008), SWISS-MODEL (Benkert, Biasini et al. 2010), PHYRE2 (Kelley, Mezulis et al. 2015) and MODELLER (Webb and Sali 2014). The model generated by SWISS-MODEL was selected for further study.The SWISS-MODELs are built based on the target-template alignment using ProMod3. Coordinates which are conserved between the target and the template are copied from the template to the model (Guex and Peitsch 1997).The QMEAN score (Benkert, Tosatto et al. 2008) was used to evaluate the quality of the generated model.

The modeled protein structure is energy minimized. The calculations were performed with GROMACS 4.5.3 (Hess, Kutzner et al. 2008) package, using the GROMOS 96 force field. The box dimensions ensured that any protein atom was at least 1.5 nm away from the wall of the box with periodic boundary conditions, solvated by simple point charge (SPC) water molecules.

NaCl counter ions were added to satisfy the electroneutrality condition. Energy minimization was carried out using the steepest descent method. The energy minimized structure is taken as an initial structure for equilibration. The energy minimized structure is validated using Ramachandran plot through PROCHECK (Fig. 10D).

The binding site of NADH in Mtb DprE2 was predicted by AutoDockVina software (Morris, Huey et al. 2009) and Gold docking software (Jones, Willett et al. 1997). We have selected binding site cavity on Mtb DprE1 the basis of Autodock software tool. The binding site cavity for *Mtb* DprE2 was defined by Asp101, Glu103, Ser147, Arg152, Tyr 160,Gly191, Tyr236 based on the active site residues of the template (PDB-1fmca) (Tanaka, Nonaka et al. 1996). The coordinates of NADH were taken from zinc online database (2).Both receptor and ligand molecule was prepared in Autodock Tools software in which polar hydrogen atoms were added to the receptor molecule and nonpolar hydrogen atoms were merged. The protein receptor (*Mtb DprE2*) and the ligands were converted from PDB to PDBQT format and vice-versa with the help of Open Babel tool. In the ligand molecule, all bonds were set to allow rotation and configuration file of the AutodockVina software, a grid box of $40 \times 40 \times 40$ points was used around the active site having x, y and z coordinates of 38.660, 34.370, 27.060 respectively to specifically cover the active residues in the binding pocket of the receptor or protein molecule that allowed ligands to behave as flexible.

Results and Discussions

The binding of six ligands with *Mtb*DprE1

The crystal structures of *Mtb*DprE1 and *Msmeg*DprE1 alone as well as in complex with several inhibitors have been determined (Batt, Jabeen et al. 2012)(Neres, Pojer et al. 2012). The *Mtb*DprE1 (461 residues) structure contains two major domains,(i) FAD binding domain involving residues from 7-196 and 413-461 regions and (ii) substrate binding domain involving residues from 197-412 region of *Mtb*DprE1. Two disordered loops *e.g.,* 269-297 and 316-330 residues are observed in substrate binding site of *Mtb*DprE1, which may be involved in ligand binding as well as interaction with a*Mtb*DprE2 enzyme (Piton, Foo et al. 2017).
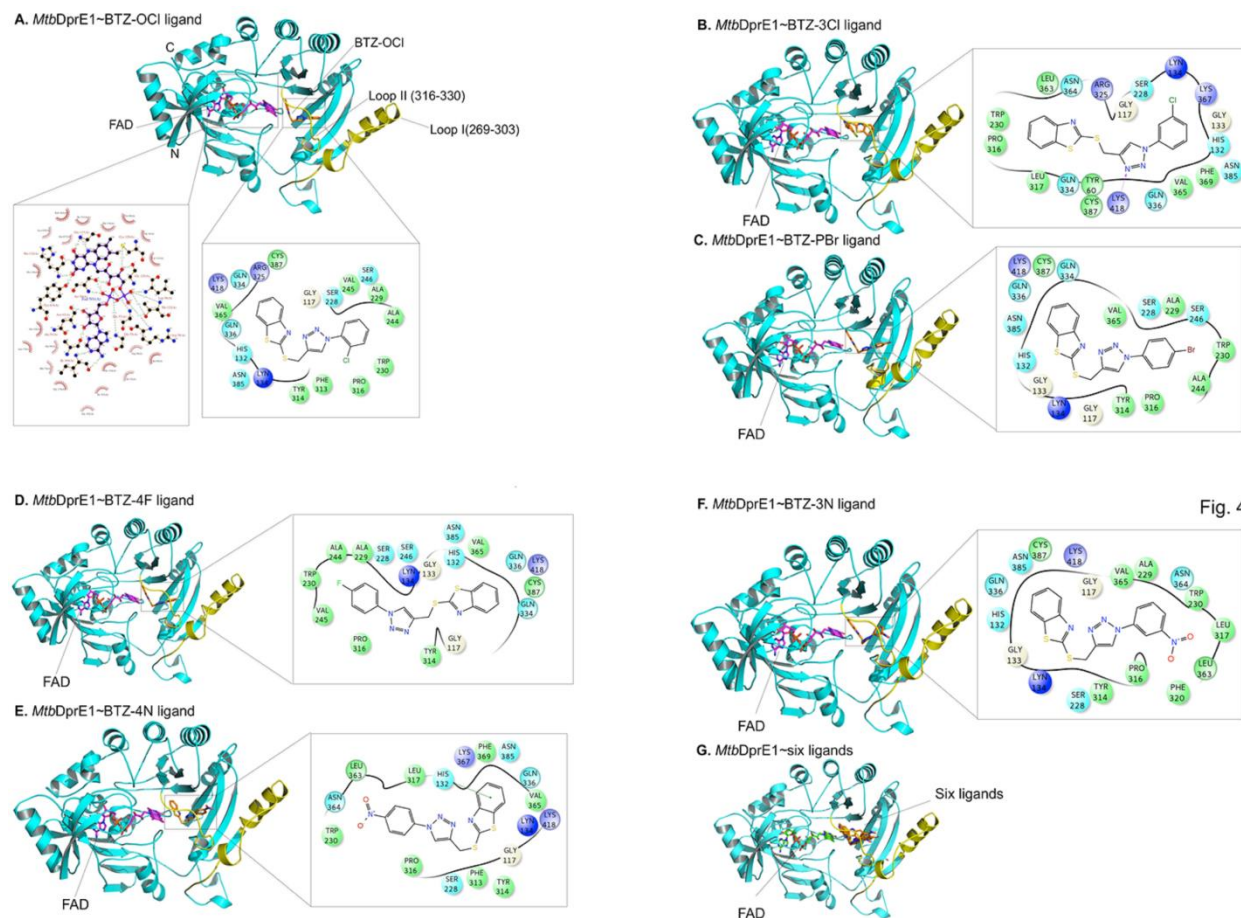
Fig. 9 Structural and interaction map of DprE1 with inhibitors. A-F. Carbon a trace atom is represented in gray and inhibitors by orange sticks. To facilitate description, the *M. tuberculosis* numbering is used. The hydrophobic groove is represented by sticks and surface is blue. FAD, G117, and V365 (yellow sticks) form the middle part of the binding cavity and the side of the pocket. In hot pink are amino acids implicated in the stabilization of inhibitors in the pocket with van der Waals contacts. Amino acids involved in H-bonding with inhibitors are represented in lime green and the H-bonds by black dots. Cysteine 387 is represented as a red stick when it forms a covalent bond with the inhibitors and in salmon pink when it is not involved in inhibitor binding.

In silico binding of BTZ-OCl and BTZ-Cl ligands

The BTZ-Cl ligand is involved in hydrogen bonding with backbone O' of Lys418 and shows the K_D value of 100µM.BTZ ring is held in substrate binding pocket and several polar and hydrophobic interactions are involved in stabilization of BTZ-Cl ligand in substrate binding pocket of *Mtb*DprE1 (Fig. 9A-B). With BTZ-OCl ligand, only polar and hydrophobic

interactions are observed at *Mtb*DprE1 substrate binding pocket and show the $K_D$value of 300µM. These results show that additional hydrogen bonding between BTZ-Cl with Lys418 of *Mtb*DprE1make ~3 fold higher affinity for BTZ-Cl than BTZ-OCl.

As seen from Table 1, docking score of -5.631 andX-score of -7.82 are observed for BTZ-OCl ligand while docking score of -5.987 and X-score of -8.29 are observed for BTZ-Cl ligand. BTZ-Cl ligand is involved in hydrogen bonding and hydrophobic interactions with the*Mtb*DprE1 binding site, while BTZ-OCl ligand is involved only in hydrophobic interactions.Higher docking and X-score in thecase of BTZ-Cl ligand may lead to higher binding affinity towards *Mtb*DprE1.

In silico binding of BTZ-3PBr and BTZ-4F

Both BTZ-PBr and BTZ-4F ligands bind to *Mtb*DprE1 with K_D value of 200µM (Fig. 9C-D).Both ligands are held by hydrophobic interactions with substrate binding pocket of *Mtb*DprE1. These ligands binding affinities are less than BTZ-Cl (K_D~100µM), however more than BTZ-OCl ligand (K_D~300µM).As seen from Table 1, the docking score of -5.107 and X-score of -8.02 are observed for BTZ-3PBr, while docking score of -5.444 and X-score of -7.75 are observed for theBTZ-4F ligand.

In silico binding of BTZ-4N and BTZ-3Nligands

Both BTZ-4N and BTZ-3N ligands have equal affinities for *Mtb*DprE1 with K_D values of 100µM. In BTZ-4N ligand complex, the BTZ ring forms π-π stacking interaction with His132, polar and hydrophobic interactions with several residues of substrate binding pocket of *Mtb*DprE1(Fig. 9E-F).However, theBTZ-3N ligand is held only by polar and hydrophobic interactions with residues of substrate binding pocket. Superposition of all six ligands (Fig. 9F) with standard ligand (QN127) shows the overall similar orientation in *Mtb*DrpE1 substrate binding pocket. As seen from Table 2, the docking score of -5.290 and X-score of -8.27 are observed for theBTZ-4N ligand, while docking score of -4.926 and X-score of -8.03 are observed for theBTZ-3N ligand, which isquite comparable to each other.

Table 2.Glide docking score obtained from GLIDE program

Ligand	Docking score
BTZ-OCl	-5.631
BTZ-3Cl	-5.987
BTZ-PBr	-5.107
BTZ-4F	-5.444
BTZ-4N	-5.290
BTZ-3N	-4.926

Mtb DprE2 Molecular Modelling and docking study with NAD

Mtb DprE2 homology structure (Fig.10B) was obtained by SWISS-MODEL program using Abeta-bound human ABAD crystal structure (PDB-1SO8) as input template. Mtb DprE2 belongs to short dehydrogenase/reductase (SDR) protein family and contain Rossmann fold domain (Seven parallel β-sheets connected with α-helices). Secondary structure analysis of DprE2 by PSIPRED server (Fig. 10A) indicated that DprE2 is a mixed helix-sheet-loop protein and gives 48.97% α-helices, 20.99% extended β-sheets and 30.04 % loops making it a flexible protein. Mtb DprE2 adoptedaα-β-α sandwich structure where all seven parallel β sheets were layered between two layers of α-helices. Modeled structure has revealed that in the loop regions between residues Phe96-Gln109, Met145-Phe158 and Pro190–Lys210 was highly disordered. Alignment of energy minimized and without energy minimized structures (fig.10C) give 0.389 RMSD and major fluctuations observed mainly in loop regions (Fig.10D) (Met145-Phe158 and Pro190–Lys210). Like other SDR family proteins, these disordered loops give functional diversity to the proteins and may be binding sites of Substrates molecule and get stabilized after substrate or ligand binding.
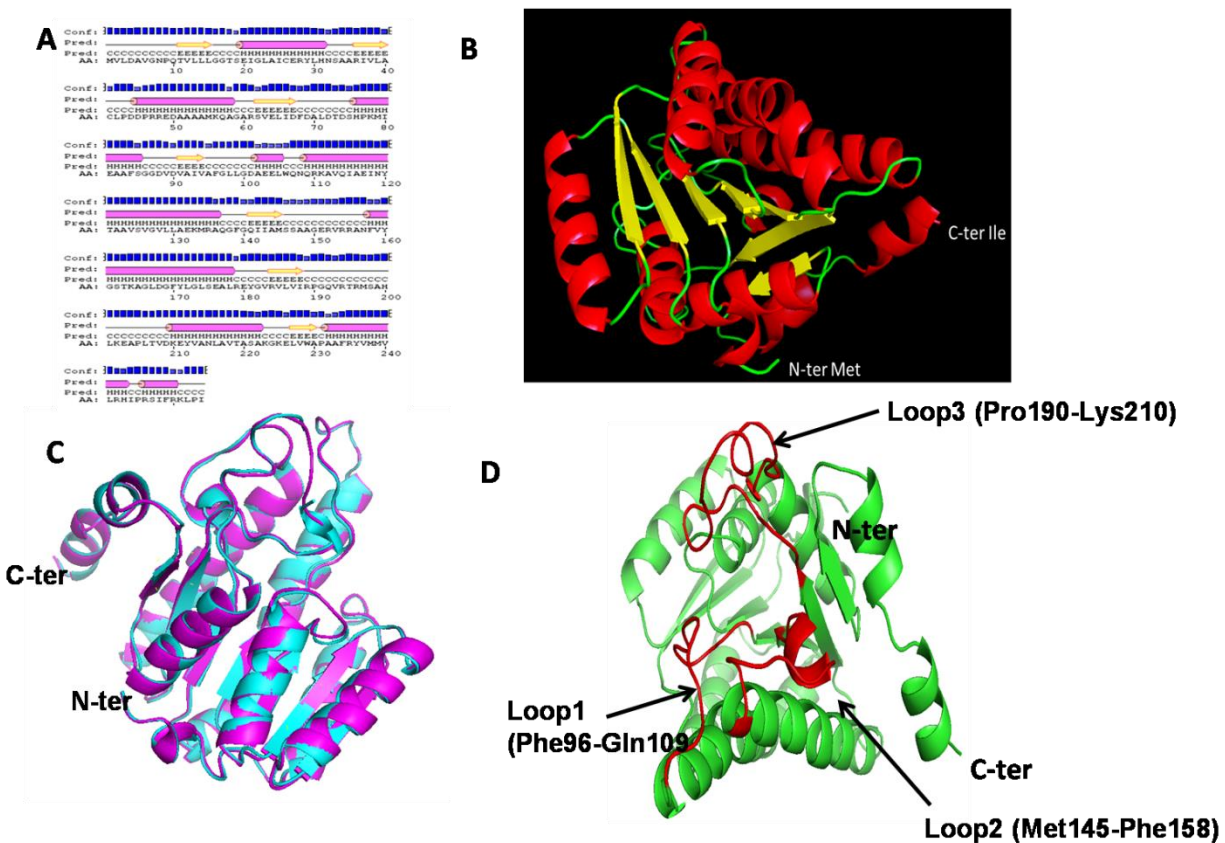
Fig. 10A. Secondary structure prediction of *M. tuberculosis* DprE2 protein by PSIPRED program.B. 3D Homology Model of DprE2 by SWISS Model.C. A superimposition of DprE2 models with (Magenta) and without (Cyan) energy minimization.D. DprE2 model showing three highly disorder loop regions (Red).

In our molecular docking study of *M. tuberculosis* DprE2 modeled structure with NADH, it has been observed that NADH interacts with DprE2 in the loop regions.(Fig.11A) . Numbers of amino acids are involved in the binding through hydrophobic and wander walls interactions.

Fig.11A. 3D model of theDprE2-NAD complex after molecular docking using Auto Dock.B. Close view of residues interacting with NAD in Docked 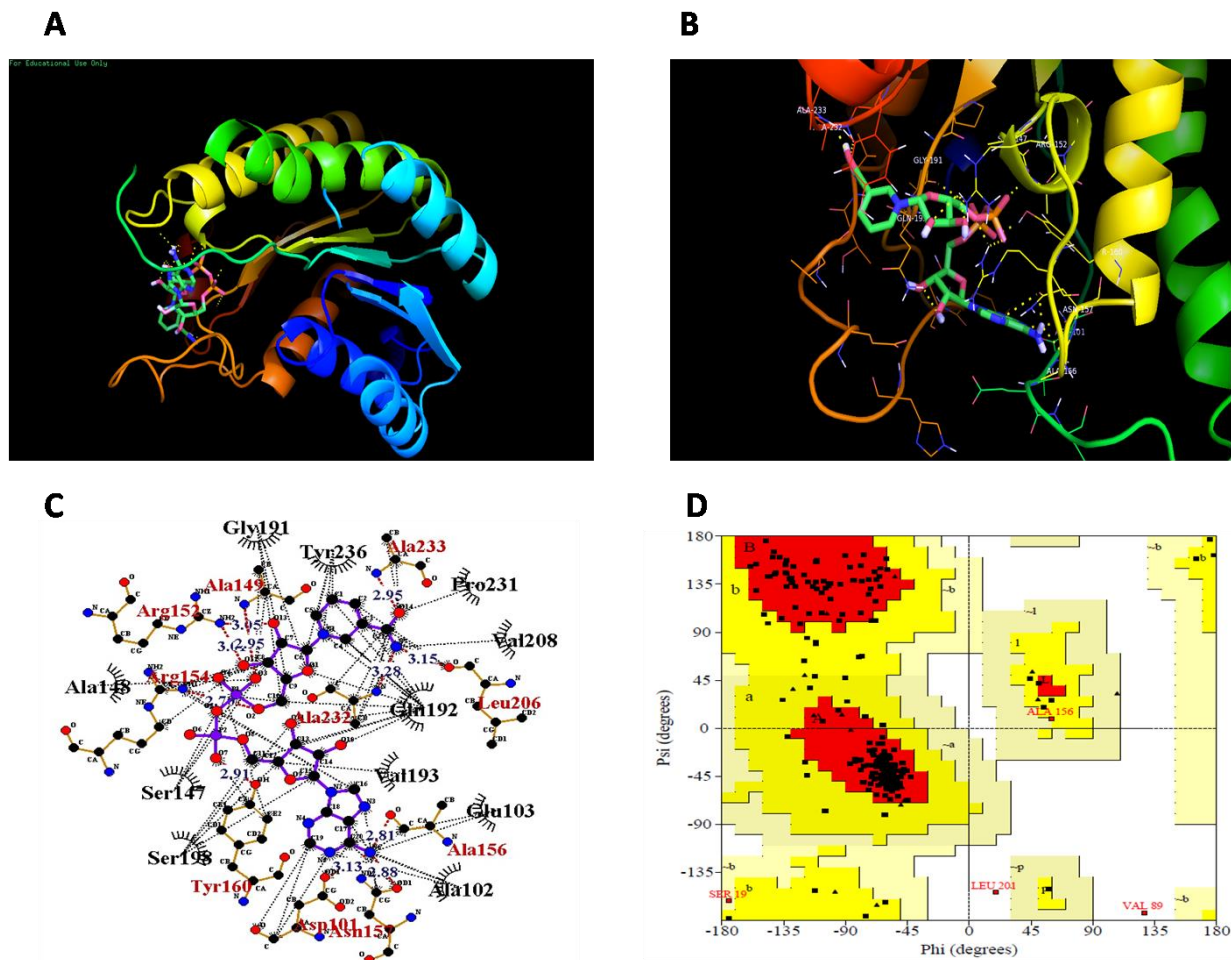complex.C.Ligplot diagram showing multiple hydrophobic and van der Waals interactions between DprE2 and NAD.D. Ramachandran plots for the model of DprE2 protein after energy minimization showing maximum numbers of amino acids are in thefavored region.

Summary

In our study, we have successfully cloned both *M. tuberculosis* DprE1 and DprE2 genes in pET32b and pET-SUMO vectors with N-terminal Trx-tag (~17KDa) and Sumo-tag (`11KDa) respectively. We have expressed *M. tuberculosis* DprE1 as asoluble protein in *E. coli* BL21 (DE3) *Arctic express* cell line at 12^0C using additive media (see methods and materials) whereas *M. tuberculosis* DprE2 become soluble at 16^0C in *E. coli* BL21 (DE3) cell line.

M. tuberculosis DprE1 purified by standard chromatographic techniques (Ni- NTA affinity chromatography, Q- sepharose anion exchange chromatography and size exclusion chromatography) and Enterokinase cleavage to remove N-terminal Trx fusion tag. A size exclusion chromatography column (superdex 200 16/60pg) has shown that *M. tuberculosis* protein eluted as amonomer. M. tuberculosis DprE2 was purified by Ni-NTA affinity chromatography and Size exclusion chromatography followed by sumo protease cleavage to remove N-terminal sumo tag(~11KDa). *M. tuberculosis* DprE2 was always eluted into void volume of the Superdex 200 16/60 pg column which corresponds to theoligomeric state of the protein. Although we have applied various strategies to stabilize DprE2 in a definite state but never succeeded.

SPR binding studies between *M. tuberculosis* DprE1 and DprE2 showing specific interaction with the kD value of 29.2µM. Our *M. tuberculosis* DprE1 –DprE2 binding study using SPR provides strong evidence as reported in previous studies that *M. tuberculosis* DprE1 and DprE2 are the two subunits of an enzyme complex decaprenyl phosphoribosyl –D – epimerase which is involved in the epimerization of DPR to DPA in a sequential oxidation and reduction reaction.

Following strategies have been employed in *M. Tuberculosis* drug discovery efforts (i) *invivo* identification of drug targets and its validation in cell based study (ii) validation of drug targets using biochemical studies and (iii) structure analysis of drug targets in complex with inhibitors/cofactors. Once these inhibitors are characterized, they will be tested *invivo* system. In thecurrent study, we have characterized six sulfur rich 2-mercaptobenzothiazole and 1,2,3-triazole conjugated ligands, which show anti-tuberculosis activity against the*MtbH37Rv* strain. These ligands bind substrate binding pocket of the*Mtb*DprE1 enzyme, a potential drug target. All

six ligands have shownanti-tuberculosis activities against *MtbH37Rv* strain in an earlier microdilution assay. These ligands have been verified in dose response and prioritized based on the MIC value (Table 1).For *invitro*analysis, we have purified the *Mtb*DprE1 enzyme and its binding analysis is performed with all six ligands using surface plasmon resonance technique. The obtained K_D values of six ligands against *Mtb*DprE1 are consistent with MIC values obtained for all six ligands against *MtbH37Rv*strainin microdilution assay (Table. 1).

In absence of crystal structure, molecular docking approach is widely used to understand the protein-ligand interaction and drug discovery. This knowledge helps in protein modification as well as in changing the chemistry of ligand binding. In *Mtb*DprE1, the FAD binding site is quite conserved and interacts with several residues as shown in Fig. 9A.All six ligandsoccupy quite similarly the substrate binding pocket of *Mtb*DprE1 (Fig. 9A-F). Observed K_D values of all six ligands are attributed to thefavorable position of ligand scaffolds in the substrate binding pocket of *Mtb*DprE1, which results in important protein-ligand interactions. A closer look at *Mtb*DprE1-FAD-six ligands interactions reveals that network of hydrophobic, polar and hydrogen bonds are involved in stabilizationof six ligands in substrate binding pocket and resultsin specific binding affinity. The BTZ-OCl ligand show low binding affinity (300μM) irrespective of more hydrophobic and polar contacts and similar orientation like other ligands in *Mtb*DprE1 substrate binding pocket.

The docking score and predicted binding affinities of all six ligands against *Mtb*DprE1 are approximately close to each other. These data indicate these ligands may have asimilar binding potential towards *Mtb*DrpE1. It is observed that lower Xscore may lead to better predicted binding affinity for theligand. The addition of R groups (NO2,Cl,Br,FandN)at adifferent position in BTZ phenyl ring has resulted in aslight change in its orientation, which results the difference in binding towards *Mtb*DrpE1. It has also resulted in aslight shift in G-score, X-score and interaction profile of respective ligand with *Mtb*DprE1.Hydrophobic interactions are the key interactions involved in binding of BTZ derived ligands to *Mtb*DrpE1.BTZ-3Cl ligand has the best G‑score(-5.987kcal/mol) and X‑score (-8.29kcal/mol)against*Mtb*DrpE1 and seemsto be the best ligand for *Mtb*DprE1. Knowledge about the structures and inhibition mechanism of these new sulfur rich 2-mercaptobenzothiazole

and 1,2,3-triazole conjugated ligands will provide new avenues for anti-tuberculosis drug development.

Structural analysis of Mtb DprE2 by Molecular modeling indicated that it is a Rossmann fold containing dehydrogenase/reductase family protein. It contains three highly disordered loops. The Docking with NADH illustrated that these loops are involved in the interaction with NAD.

References

Ali, R., S. Kumar, O. Afzal and S. Bawa (2015). "In-vitro antimicrobial screening and molecular docking studies of synthesized 2-chloro-N-(4-phenylthiazol-2-yl) acetamide derivatives." <u>Drug Development and Therapeutics</u>**6**(2): 79.

Arora, P., E. L. Foster and S. A. Porcelli (2013). CD1d and natural killer T cells in immunity to Mycobacterium tuberculosis. <u>The New Paradigm of Immunity to Tuberculosis</u>, Springer**:** 199-223.

Banerjee, A., E. Dubnau, A. Quemard, V. Balasubramanian, K. S. Um, T. Wilson, D. Collins, G. De Lisle and W. R. Jacobs Jr (1994). "inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis." <u>Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information</u>**263**(5144): 227-229.

Batt, S. M., T. Jabeen, V. Bhowruth, L. Quill, P. A. Lund, L. Eggeling, L. J. Alderwick, K. Fütterer and G. S. Besra (2012). "Structural basis of inhibition of Mycobacterium tuberculosis DprE1 by benzothiazinone inhibitors." <u>Proceedings of the National Academy of Sciences</u>**109**(28): 11354-11359.

Benkert, P., M. Biasini and T. Schwede (2010). "Toward the estimation of the absolute quality of individual protein structure models." <u>Bioinformatics</u>**27**(3): 343-350.

Benkert, P., S. C. Tosatto and D. Schomburg (2008). "QMEAN: A comprehensive scoring function for model quality assessment." <u>Proteins: Structure, Function, and Bioinformatics</u>**71**(1): 261-277.

Bessonov, K., K. A. Vassall and G. Harauz (2017). "Docking and molecular dynamics simulations of the Fyn-SH3 domain with free and phospholipid bilayer-associated 18.5-kDa myelin basic protein (MBP)—Insights into a noncanonical and fuzzy interaction." <u>Proteins: Structure, Function, and Bioinformatics</u>.

Bosserman, R. E. and P. A. Champion (2017). "ESX systems and the Mycobacterial Cell Envelope: What's the connection?" <u>J Bacteriol</u>.

Crellin, P. K., R. Brammananth and R. L. Coppel (2011). "Decaprenylphosphoryl-β-D-ribose 2'-epimerase, the target of benzothiazinones and dinitrobenzamides, is an essential enzyme in Mycobacterium smegmatis." <u>PloS one</u>**6**(2): e16869.

DeLano, W. L. (2002). "Pymol: An open-source molecular graphics tool." <u>CCP4 Newsletter On Protein Crystallography</u>**40**: 82-92.

Draper, P. (1998). "The outer parts of the mycobacterial envelope as permeability barriers." <u>Front Biosci</u>**3**: D1253-D1261.

Geourjon, C. and G. Deleage (1995). "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments." <u>Bioinformatics</u>**11**(6): 681-684.

Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling." <u>electrophoresis</u>**18**(15): 2714-2723.

Guiard, J., A. Collmann, L. F. Garcia-Alles, L. Mourey, T. Brando, L. Mori, M. Gilleron, J. Prandi, G. De Libero and G. Puzo (2009). "Fatty acyl structures of Mycobacterium tuberculosis sulfoglycolipid govern T cell response." The Journal of Immunology**182**(11): 7030-7037.

Hess, B., C. Kutzner, D. Van Der Spoel and E. Lindahl (2008). "GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation." Journal of chemical theory and computation**4**(3): 435-447.

Ishikawa, E., T. Ishikawa, Y. S. Morita, K. Toyonaga, H. Yamada, O. Takeuchi, T. Kinoshita, S. Akira, Y. Yoshikai and S. Yamasaki (2009). "Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle." Journal of Experimental Medicine**206**(13): 2879-2888.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." Journal of molecular biology**292**(2): 195-202.

Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor (1997). "Development and validation of a genetic algorithm for flexible docking." Journal of molecular biology**267**(3): 727-748.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nature protocols**10**(6): 845-858.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." Journal of molecular biology**372**(3): 774-797.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of applied crystallography**26**(2): 283-291.

Ly, D., A. G. Kasmar, T.-Y. Cheng, A. de Jong, S. Huang, S. Roy, A. Bhatt, R. P. van Summeren, J. D. Altman and W. R. Jacobs (2013). "CD1c tetramers detect ex vivo T cell responses to processed phosphomycoketide antigens." Journal of Experimental Medicine: jem. 20120624.

Makarov, V., G. Manina, K. Mikusova, U. Möllmann, O. Ryabova, B. Saint-Joanis, N. Dhar, M. R. Pasca, S. Buroni and A. P. Lucarelli (2009). "Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis." Science**324**(5928): 801-804.

Mir, F., S. Shafi, M. Zaman, N. P. Kalia, V. S. Rajput, C. Mulakayala, N. Mulakayala, I. A. Khan and M. Alam (2014). "Sulfur rich 2-mercaptobenzothiazole and 1, 2, 3-triazole conjugates as novel antitubercular agents." European journal of medicinal chemistry**76**: 274-283.

Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson (2009). "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility." Journal of computational chemistry**30**(16): 2785-2791.

Neres, J., F. Pojer, E. Molteni, L. R. Chiarelli, N. Dhar, S. Boy-Röttger, S. Buroni, E. Fullam, G. Degiacomi and A. P. Lucarelli (2012). "Structural basis for benzothiazinone-mediated killing of Mycobacterium tuberculosis." Science translational medicine**4**(150): 150ra121-150ra121.

Ojha, A., M. Anand, A. Bhatt, L. Kremer, W. R. Jacobs and G. F. Hatfull (2005). "GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria." Cell**123**(5): 861-873.

Piton, J., C. S.-Y. Foo and S. T. Cole (2017). "Structural studies of Mycobacterium tuberculosis DprE1 interacting with its inhibitors." Drug discovery today**22**(3): 526-533.

Pulaganti, M., B. Banaganapalli, C. Mulakayala, S. K. Chitta and C. Anuradha (2014). "Molecular modeling and docking studies of O-succinylbenzoate synthase of M. tuberculosis—a potential target for antituberculosis drug design." Applied biochemistry and biotechnology**172**(3): 1407-1432.

Rashmi, M. and D. Swati (2015). "In silico drug re-purposing against African sleeping sickness using GlcNAc-PI de-N-acetylase as an experimental target." Computational biology and chemistry**59**: 87-94.

Ribeiro, A. L. d. J. L., G. Degiacomi, F. Ewann, S. Buroni, M. L. Incandela, L. R. Chiarelli, G. Mori, J. Kim, M. Contreras-Dominguez and Y.-S. Park (2011). "Analogous mechanisms of resistance to benzothiazinones and dinitrobenzamides in Mycobacterium smegmatis." PloS one**6**(11): e26675.

Sambandan, D., D. N. Dao, B. C. Weinrick, C. Vilchèze, S. S. Gurcha, A. Ojha, L. Kremer, G. S. Besra, G. F. Hatfull and W. R. Jacobs (2013). "Keto-mycolic acid-dependent pellicle formation confers tolerance to drug-sensitive Mycobacterium tuberculosis." MBio**4**(3): e00222-00213.

Schlesinger, L. S., A. K. Azad, J. B. Torrelles, E. Roberts, I. Vergne and V. Deretic (2008). "Determinants of phagocytosis, phagosome biogenesis and autophagy for Mycobacterium tuberculosis." Handbook of tuberculosis: immunology and cell biology: 1-22.

Sen, T. Z., R. L. Jernigan, J. Garnier and A. Kloczkowski (2005). "GOR V server for protein secondary structure prediction." Bioinformatics**21**(11): 2787-2788.

Stanley, S. A. and J. S. Cox (2013). Host–pathogen interactions during Mycobacterium tuberculosis infections. Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism, Springer**:** 211-241.

Tanaka, N., T. Nonaka, T. Tanabe, T. Yoshimoto, D. Tsuru and Y. Mitsui (1996). "Crystal structures of the binary and ternary complexes of 7α-hydroxysteroid dehydrogenase from Escherichia coli." Biochemistry**35**(24): 7715-7730.

Wagner, J. M., S. Chan, T. J. Evans, S. Kahng, J. Kim, M. A. Arbing, D. Eisenberg and K. V. Korotkov (2016). "Structures of EccB 1 and EccD 1 from the core complex of the mycobacterial ESX-1 type VII secretion system." BMC structural biology**16**(1): 5.

Wang, F., D. Sambandan, R. Halder, J. Wang, S. M. Batt, B. Weinrick, I. Ahmad, P. Yang, Y. Zhang and J. Kim (2013). "Identification of a small molecule with activity against drug-resistant and persistent tuberculosis." Proceedings of the National Academy of Sciences**110**(27): E2510-E2517.

Webb, B. and A. Sali (2014). "Protein structure modeling with MODELLER." <u>Protein Structure Prediction</u>: 1-15.

Whitmore, L. and B. A. Wallace (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases." <u>Biopolymers</u>**89**(5): 392-400.

Yuan, Y., D. D. Crane, R. M. Simpson, Y. Zhu, M. J. Hickey, D. R. Sherman and C. E. Barry (1998). "The 16-kDa α-crystallin (Acr) protein of Mycobacterium tuberculosis is required for growth in macrophages." <u>Proceedings of the National Academy of Sciences</u>**95**(16): 9578-9583.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." <u>BMC bioinformatics</u>**9**(1): 40.