Studies on NMPylase and valine catabolism enzymes from *Mycobacterium*

A Thesis

Submitted for the Degree of

DOCTOR OF PHILOSOPHY

In the Faculty of Science

BY

Amrita Singh

Molecular Biophysics Unit

Indian Institute of Science

Bangalore- 560012, India

November 2017

Dedicated to

Dadi,

Maa-Pitaji

CONTENTS

SECTION-II

DECLARATION

I hereby declare that the work embodied in this thesis has been carried out by me under the supervision of Prof. Avadhesha Surolia and Prof. R. Varadarajan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore and that it has not been submitted for any degree or diploma to any other institution. Following prevalent scientific practice, acknowledgement has been accorded wherever due.

(Amrita Singh)

CERTIFIED

Prof. Avadhesha Surolia **Prof. R. Varadarajan**

(Research Supervisor) **(Research Supervisor)**

Acknowledgement

I take this opportunity to express my immense gratitude to all the people who have helped me get this work in this presentable shape.

First and foremost, it has been an honor for me to work with a renowned and dedicated scientist, Prof. Avadhesha Surolia. I am indebted to him for giving me the opportunity of working in his laboratory. He has been truly a mentor in all aspects of life. The "special bonding" with Prof. Surolia will always remain close to me as he has been the first one to introduce me to the field of research. Those initial days, being his first student after his return from NII, will always be cherished. His immense dedication towards science and ability to remember the information and at the same time correlate the experimental outcome has always amazed me. I am grateful to him for allowing me to explore the exciting project on AMPylation. Being in his laboratory, I had the liberty to think independently and take the project in many dimensions, which gave me the training in diverse techniques of biological science. Thank you Sir, for always being there to answer all my doubts with patience and boosting the scientific aptitude in me. I would hope to live upto at least 10% of your "dedication to science and research" one day.

I wish to express my gratitude to Prof. M. Vijayan for extending his lab facilities for me and providing guidance in my work and constant encouragement. I am grateful for the insightful discussions I have had with him regarding the structural part of my work. I'm very blessed to have a great opportunity to learn from him.

I am especially thankful to Prof. Raghavan Varadarajan for giving me MBU-studentship and for being my co-advisor, allowing me to attend his lab-meetings in my first-year, for his excellent coursework and lab facilities.

I am eternally grateful to Prof. Namita Surolia, JNCASR, Bangalore for her suggestions and kind help in the work, especially in the live-cell imaging experiments. I look up to her for her work-life balance.

I am very thankful to the past and present Chairmen of MBU, Prof. Dipankar Chatterji and Prof. R. Varadarajan and all the professors: Prof. M. Vijayan, Prof. K.

Suguna, Prof. B. Gopal, Prof. M. R. N. Murthy, Prof. N. Srinivasan, Prof. Saraswathi Vishveshwara, Prof. Manju Bansal, Prof. S. K. Sikdar, Prof. R. Narayanan, Prof. Siddhartha P. Sarma, Dr. Jayanta Chatterjee, Dr. Mahavir Singh, Dr. Aravind Penmatsa, Dr. Anand Srivastava, Dr. Rahul Roy and Dr. Somnath Dutta for their excellent course work and for maintaining a healthy research environment in MBU.

I want to thank Dr. Srikalaivani Raju, Nisha, Dr. Nidhi, Dr. Nikhil and Anju for their enthusiasm and support in some part of my work. Special thanks to Guha Arun, Ankita Shamnani, Hari, Samannaya Hazra and Badepally Nagendar Goud for working and nurturing these projects alongwith me.

I also want to thank Chanda madam for being a great support during my first year visits in NII. She is so cheerful and her vibrant nature made my stay at NII pleasant. Thanks to Valsala, Sunitha and Rajini for helping me in all paper-works. Thanks are due to the past and present MBU office staff for their helpful and cheerful attitude and for making life in MBU so much simpler- Mr. Chayapathi, Mr. Shivshankar, Mrs. Indira, Mr. Govindraju and the MBU Workshop staff- Mr. Bandrinarayana and Mr. Mani. I thank Sunitha from Proteomics facility, MBU and Anant for maintaining the MBU webmail.

I take this opportunity to express my gratitude to everyone who has taught me the various techniques: Dr. Garima for teaching me nano-ITC and MD simulations, Dr. Tandrika for compiling AMPylation project initially and teaching me cloning, Nisha and Dr. Nidhi for teaching me Autodock and Gromacs.

Garima, Tandrika, Tanushree, Rajiv, Ravi, Shweta, Nuzhat, Madhuraka and Kapil have been a great support during my visits to NII or if I have needed any chemicals. I would like to thank Tandrika and Rajiv who have been excellent seniors and I will cherish all the fun moments and treats for my entire life. I would like to thank all the present and past lab members of AS lab; Ashalatha, Thakur, Saurabh, Tejaswi, Rajiv, Nisha, Prerna, Shraddha, Sunita, Niki, Kanika, Rajanya, Rajini and Praveen. Especially to Kaushal, Padmanabh and Archita for their exceptional support during my second year of PhD and for sharing lab reagents.

I feel happy to be a part of “Chrysalis”. Thanks to Shushant, Geeta, Anju, Anil, Prachi, Raksha, Rakesh, Harsh, Manisha, Anandita, Mansi, Kritika, Santosh, Raghu, and Prasun. Thank you, Geeta for ignoring all my shortcomings and being the best of friend. Especially, thanks to Sushant for helping me during crystallography coursework and in some critical experiments.

I am extremely fortunate to have Saurabh, who was always there as my strength, an incredible support in the lab through all thick and thin patches during the phase of conducting experiments, negative results and the painful process of thesis writing. Thank you Saurabh for celebrating with me when I was at my best and handling me when I was at my worst for anything and everything in life. Thank you for convincing me for far western experiment and helping me with the leads of the project. I could not have wished for a better “colleague cum friend cum senior”.

Nisha is especially acknowledged for her companionship, care, food and good words all through. She has been the perfect partner in the project. I am indebted for the “insightful tea sessions – Gyan-Varta” with her, Rajeev and Saurabh. Thank you Nisha for being constant source of entertainment with Rajeev-Nisha and Nisha-Nidhi combinations and making lab 209 a memorable place for me. Thank you Saurabh, Nisha, Nidhi, Rajeev, Rajini, Nagendra and Tejaswi for everything.

Thanks to Spraha, Rishi, “Arvind and company” for all the fun-times that we shared. Special thanks to Shaika for food, celebrations, for being a delightful host and providing a wonderful time-out from the PhD routine.

I thank Megha, Pallavi, Swastika, and Mamta for their constant love and support.

Lastly, Nothing would have been possible without the support of my family. I am here because of the unconditional love of my parents and blind faith of my bhaiya Alok and my Babu Anurag. Especially, thank you “comrade” for standing beside me in all my good and bad decisions of life and for giving me the confidence to grow and become a better person. Thank you Dheerendra Ji, for putting up with me throughout the thesis writing.

Lastly, thanks to Almighty, the author of my life, for bringing such beautiful moments throughout.

Amrita Singh

Abbreviations

ADP	Adenosine Diphosphate
AMP	Adenosine-5'-Monophosphate
AR	Adenylyl Removase
ATB	Automated Topology Builder
AT	Adenylyl Transferase
ATP	Adenosine Triphosphate
ATase	Adenylyl Transferase
AvrB	Virulence protein B
β-ME	Beta-meracaptoethanol
BAT2	Bovine Alveolar Type 2
BCAA	Branched Chain Amino Acid
BID	Bartonella intracellular delivery
BN-PAGE	Blue Native Polyacrylamide gel electrophoresis
BSA	Bivine Serum Albumin
CD	Circular Dichroism
Cdc42	Cell division control protein 42
CDP	Cytidine Di Phosphate
CFU	Colony Forming Unit
CIAP	Calf Intestinal Alkaline Phosphate
CMP	Cytidine Mono Phosphate
CoA	Coenzyme A
Co-IP	Co-immunoprecipitation
crp	c-AMP receptor protein
CTD	C-Terminal Domain
CTP	Cytidine Tri-phospahte
DAPI	4',6 -diamidino-2-phenylindole
DLS	Dynamic Light Scattering
DMEM	Dulbecco's modified Eagle's medium
Doc	Death on curing
DSF	Dynamic Scanning Fluorimetry
DTNB	5,5'-dithio-bis-2-nitrobenzoic acid
DTT	Dithiothreitol
*Ec*Fic	*Escherichia coli Fic*
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced Green Fluorescent Protein
EEA1	Early Endosome Antigen1
EF-Tu	Elongation Factor Thermo Unstable
ESI-MS	Electro-Spray Ionization Mass Spectroscopy
FAD	Flavin adenine dinucleotide
Fic	Filamentation induced by cyclic AMP

FITC	Fluorescein isothiocyanate
FPLC	Fast Protein Liquid Chromatograpahy
FRET	Fluorescence Resonance Energy Transfer
GAPs	GTPase-Activating Proteins
GDIs	Guanine Nucleotide Dissociation Inhibitors
Gdncl	Guanidine hydrochloride
GEFs	Guanine Nucleotide Exchange Factor
GS	Glutamine Synthase
GST	Glutathione-S-Transferase
GTP	Guanosine-5'-triphosphate
GyrA	Gyrase subunit A
GyrB	Gyrase subunit B
HEK293	Human embryonic kidney cells 293
HTH	Helix Turn Helix
hVPS34	Vacuolar Protein Sorting
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
I-TASSER	Iterative Threading ASSEmbly Refinement
IbpA	Immunoglobulin-binding protein A
IgG	Immunoglobulin G
IMAC	Immobilizes Metal Affinity Chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LB	Luria Broth
LCV	*Legionella*-containing vacuoles
LC-MS	Liquid chromatography- Mass Spectrometry
MALDI-TOF	Matrix Assisted Laser Desorption/Ionisation-Time of Flight Mass
MBP	Maltose Binding Protein
MCS	Multiple Cloning Site
MD	Molecular Dynamics
MSA	Multiple Sequence Alignment
Mtb	*Mycobacterium tuberculosis*
HIBADH	3-hydroxyisobutyrate dehydrogenase
NAD^+	β -Nicotinamide Adenine Dinuleotide (oxidized form)
NADH	β -Nicotinamide Adenine Dinuleotide (reduced form)
NBS	*N*-Bromosuccinimide
NCBI	National Center for Biotechnology Information
NDP	Nucleotide Di Phosphate
NEB	New England Bio labs
NES	Nuclear Export Signal
NLS	Nuclear Localization Signals
NMP	Nucleotide Mono Phosphate
NTP	Nucleotide Triphosphate

OD	Optical Density
ORF	Open Reading Frame
PABA	Perturbation in p-Amino Benzoic Acid
PAGE	Poly-Acrylamide Gel Electrophoresis
PAK	p21 Activated Kinase 1 protein
PAMPs	Pathogen Associated Molecular Pattern
PBS	Phosphate Buffered Saline
PBST	Phosphate Buffer Saline with 0.1% Tween 20
PDB	Protein Data Bank
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reactio
p-CMB	p-chloromercuribenzoic acid
PGDH	6-Phospho Gluconate Dehydrogenase
PMA	Phorbol 12-Myristate 13-Acetate
PMSF	Phenylmethylsulfonyl Fluoride
PTM	Post-translational modification
R	Regulatory Domain
Rac	Ras-related C3 botulinum
RPM	Revolutions Per Minute
RT-PCR	Real Time Polymerase Chain Reaction
SCAD	Short Chain Acyl-CoA Dehydrogenase
SDS	Sodium Dodecyl Sulphate
SDM	Site Directed Mutagenesis
SEC	Size Exclusion Chromatography
*Sm*Fic	*Smegmatis* Fic
*So*Fic	*Shewanella oneidensis Fic*
SUMO	Small Ubiquitin like Modifier
T3SS	Type III secretion system
T4SS	Type IV secretion system
TA	Toxin-Antitoxin
TB	Tuberculosis
TCA	Tri Carboxylic Acid
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TLC	Thin Layer Chromatography
µL	microliter
µM	micromolar
UMP	Uridine Mono Phosphate
UPR	Unfolded Protein Response
UTase	Uridylyl Transferase
UTP	Uridine Tri Phosphate
VopS	*Vibrio* outer protein S

Synopsis

Post-translational modifications constitute an important arm of the regulatory mechanism in mammalian cells. Therefore, bacterial pathogens, which thrive by mimicking the host cell signaling, are proving to be difficult to be tamed. *Mycobacterium tuberculosis*, for example, the morbid pathogen manipulates the host cellular pathways such as to block the host phagosomal maturation to avoid its killing as well as creates a favourable environment for its replication. NMPylation (transfer of nucleotide monophosphate; NMP from nucleotide triphosphate; NTP) mediated by Fic domain containing proteins has been an area of immense interest in the recent years and has been established as a stratagem of bacterial effectors to manipulate host cell signaling. This thesis entitled "**Studies on NMPylase and valine catabolism enzymes from *Mycobacterium"*** presents a study of Fic domain proteins from *Mycobacterium sp.* in pathogenesis as well as in bacterial dormancy. The work covers *de novo* motif identification based on sequence analysis of Fic protein from 22 *Mycobacterium sp*. and *in vitro/ in vivo* characterization of Fic domain proteins from *Mycobacterium tuberculosis*, *Mycobacterium smegmatis* and *Mycobacterium marinum.*

The thesis begins with the literature survey on Fic mediated AMPylation. The "**Chapter 1"** commences with a brief introduction on the newly discovered PTM, "AMPylation" and its resemblance with "Adenylylation" which was discovered in 1968 by Stadtman in adenylylation mediate regulation of the reaction catalyzed by Glutamine synthase in *E. coli*. As the modification is mediated by Fic protein, belonging to Fic/Doc superfamily, the chapter further describes the characterization of Fic protein from *E. coli* where **Fic** (**F**ilamentation **i**nduced by **c**AMP) name was coined for the first time. The chapter also describes the effect of Fic domain containing bacterial effectors from *Vibrio parahaemolyticus*, *Histophilus somni*, *Legionella pneumophila and Bartonella sp.* It also discusses in detail the proposed catalytic mechanism of AMPylation and other known PTMs mediated by Fic domain proteins. Further, it summarizes the structural features for substrate recognition and

nucleotide binding. This is followed by a description of the regulation of Fic domain-containing proteins. As another SxxxEG motif containing domain regulates the activity of Fic proteins, these together represent a new type II toxin-antitoxin module. Therefore, a classification based on the antitoxin motif has been detailed. Apart from having a defined role in perturbing host cell signaling, these proteins also modify targets in the bacteria itself, the biological relevance of the same has also been discussed in this chapter. Fic domain proteins are found not only in bacteria but are distributed ubiquitously in archea and metazoans as well. Chapter 1 also covers the recent findings about the role of Fic proteins from archea and metazoans. Chapter 1 further details the characteristic features of *Mtb*, briefly describing the infection cycle in the host, the pathogen-host interplay in the lung macrophages and the strategies elicited by *Mtb* to evade the host response with a specific emphasis on the *Mtb* mediated PTMs in infected macrophages.

As set in the objectives, the principal aim was to understand the role of AMPylation in the pathogenesis of *Mycobacterium*. To this end, **Chapter 2** covers a detailed analysis of Fic domain proteins across 22 *Mycobacterium sp*. where 71% of Fic domain proteins shared non-canonical Fic motif and 14% with canonical Fic motif and 15% with Doc motif. Further, the variability in the protein sequences was utilized for *de novo* identification of linear motifs. The putative functions these Fic domain proteins disrupt in their mammalian host were deciphered from the motifs they display to localize in a particular cellular compartment. *De novo* identification was supported by the functional characterization of the localization signal in two Fic domain proteins from *M. smegmatis*. This chapter identified a new class of TA module in Fic domain proteins from *Mycobacterium*, i.e, mobile mystery protein B and identified putative small XRE family transcriptional regulator protein, which may be the antitoxin counterpart as mobile mystery protein A.

With the identification of a new set of putative regulators; TA system; multifunctional enzymes from mycobacterium, all equipped with a Fic domain, either canonical or non-canonical, we characterized Fic protein from *Mtb*, which is part of its core genome. **Chapter 3** describes the role of AMPylation activity and identifies

its cognate partner both in *Mycobacterium* and mammalian cells. The identified substrate from *Mtb* includes DNA gyrase subunit B. DNA gyrase is the only type II topoisomerase present in *Mtb,* which catalyzes negative supercoiling of DNA. AMPylation of GyrB by *Mt*Fic decreases its ATPase activity required for introduction of negative supercoiling in closed circular DNA. Further, this chapter identifies that apart from Gyrase B, there are multiple interacting partners of *Mt*Fic in the bacterial cell lysate. Further, the Chapter describes transient transfection of *Mt*Fic in HeLa and HEK293 cells result in cell rounding which eventually culminates in detachment and cell death. Subsequent investigation of sub-cellular localization of *Mt*Fic protein shows nuclear and ER lumen as the major sites of localization, which is corroborated by the identification of its interacting partners in nuclear and ER/membrane enriched fractions. A detailed characterization revealed Rac1, Cdc42, and vimentin as targets for *Mt*Fic in mammalian host cells. Altogether identification of multiple interacting partners in bacterial and mammalian host including GyraseB, Rac1, Cdc42 and vimentin, opens up new avenues for further exploration of the role of AMPylation activity in *Mycobacterium* survival, dormancy and pathogenicity.

Fic motif residues bind with phosphates of NTPs, however, the residues from flap region and core helices support interactions with the base of NTPs and place NTP in an orientation suitable for NMPylation. As Fic proteins sequences curated from the *Mycobacterium sp*. showed variabilities in these regions, nucleotide selectivity exhibited by Fic proteins were explored next. **Chapter 4**, thus covers identification of a unique motif supporting selectivity for CTP over ATP. MtFic preferred AMPylation whereas Fic from *M. smegmatis* and *M. marinum* preferred CMPylation. The auto-modifications viz. auto-AMPylation and auto-CMPylation activity were also shown by the antitoxin mutants, however the *M. marinum* Fic protein, even with antitoxin domain showed auto-CMPylation. The chapter further explores regulation of Fic activity by antitoxin motif SxxxEG. Presence of arginine/asparagine at glycine position interferes with the inhibitory effect of glutamate, antagonizing the activity of antitoxin motif, thus resulting in enhanced toxin activity even in the presence of antitoxin. Moreover, this chapter also covers

preliminary studies on residual pyrophosphatase activity exhibited by some CMPylating Fic enzymes.

Chapter 5 covers characterization a Fic protein from *M. marinum*, with winged helix turn helix domain that showed DNA binding properties and a novel PTM, CMPylation, where CMP molecule gets transferred to the cognate substrates Gyrase A and Gyrase B subunits in bacteria and PCNA and Histone1 in the mammalian host. Additionally, this chapter identifies CMPylation, as a reversible PTM, which can be reversed with the same enzyme. CMPylation and de-CMPylation mediated by *Mm*Fic is highly regulated by glutamate present in the anti-toxin domain, thus *Mm*Fic protein is well-coordinated toxin-antitoxin system also. This chapter, hence, identifies and characterizes a novel TA system from *Mycobacterium* and identifies its targets as well. Moreover, this chapter also describes functional significance of auto-AMPylation in class I Fic proteins where auto-modification opens up the N-terminus region thus enhancing the accessibility of substrate binding residues from the flap region.

While these studies were in progress, functional characterization of some of the enzymes involved in valine catabolism, were also explored. In this context, it has been shown that addition of valine in the culture media inhibits the growth of *Mycobacterium*. Since the components of valine catabolism are critical for the survival and infection cycle of *Mtb*, it constitutes a potential drug target. In **Chapter 6**, detailed structural and functional characterizations of HIBADH enzyme catalyzing the sixth step in valine catabolism are described. *Mt*HIBADH utilizes NAD^+ as a cofactor and S-HIBA as the substrate. The locations of substrates in the active site region of HIBA dehydrogenases have been defined for the first time. Using a plethora of structural and biochemical techniques, the entry and active sites of the substrate have been unambiguously characterized. The chapter also describes a plausible reaction mechanism, inferred based on the structures and modeling. Further, studies highlight Cys^{210} as the critical cysteine residue for maintenance of the conformation of the active site of the enzyme. Further, the chapter identifies a novel FRET pair in

NAD^+ bound *Mt*HIBADH enzyme, where tryptophan (Trp^{211}) acts as a donor and NAD^+ bound ({NAD}* acts as the acceptor.

Chapter 7 briefly summarizes all the findings of the research carried out and presents an overview of our present understanding of the mycobacterial Fic proteins and HIBADH enzyme.

Appendix A summarizes the preliminary results of mmsA and fadE9 enzymes involved in *Mtb* valine catabolism.

CHAPTER 1

Introduction of Fic mediated AMPylation

Table of contents

1.1 AMPylation

AMPylation also referred as "Adenylylation" is a recently re-discovered novel post-translational modification (PTM), where adenosine-5'-monophosphate (AMP) moiety from adenosine triphosphate (ATP) is covalently transferred to the threonine/tyrosine side chain of target protein substrates (Figure 1.1). This modification is performed by Fic (filamentation induced by cyclic AMP) domain-containing proteins, also referred as Fic proteins comprising HxFx(D/E)GNGRxxR motif (Woolery et al., 2010; Worby et al., 2009; Yarbrough et al., 2009).

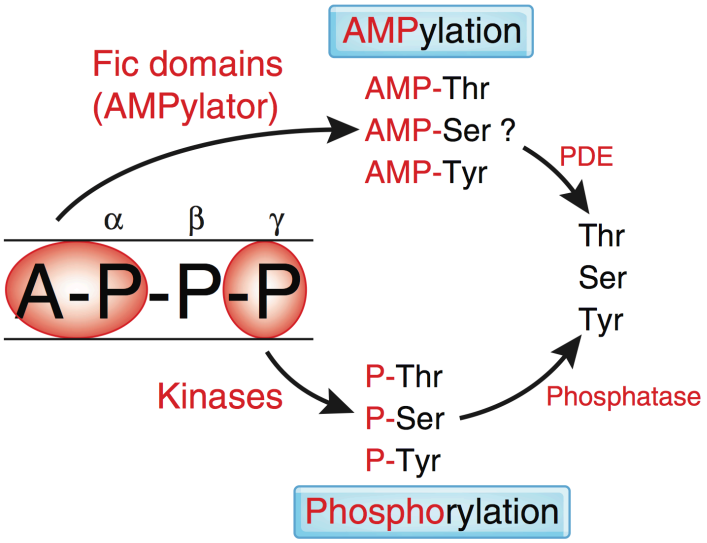


Figure 1.1. AMPylation is similar to phosphorylation PTM. *AMPylators transfer the AMP of ATP to threonine, tyrosine, and likely serine residues, and this modification can be reversed by phosphodiesterases. Fic domains containing proteins mediate AMPylation. Adapted from (Yarbrough and Orth, 2009).*

Fic proteins belong to Fido superfamily embracing Fic, Death on curing (Doc) and avirulence protein B (AvrB) proteins, all sharing a conserved α-helical fold (Kinch et al., 2009). Doc proteins share decent sequence similarity with Fic proteins, however, perform phosphorylation rather than AMPylation activity and share HxFx(D/E)GN**K**RxxR conserved doc motif. AvrB proteins, neither share sequence similarity with Fic proteins nor have a Fic linear motif, still, the conserved α helical

topology forms a flipped active site pocket similar to Fic active site (Harms et al., 2016).

Fic domains are ubiquitous and are found either as single domain proteins or as part of multi-domain proteins (Kinch et al., 2009). Fic/Doc proteins are fascinating due to remarkable variations in the Fic motif, making them versatile in performing PTMs ranging from phosphorylation, AMPylation, UMPylation to CDP-cholination, using broadly similar molecular mechanisms and provide diversity to these enzymes at both molecular and functional levels, this allows them to perform diverse biological functions (Cruz et al., 2014; Garcia-Pino et al., 2014; Yarbrough et al., 2009).

1.2 AMPylation, a re-discovered PTM "Adenylylation"

Adenylylation, re-discovered as AMPylation, was first studied by Stadtman in *E. coli* Glutamine synthetase (GS) adenylyl transferase enzyme cascade, in the early 1970s (Shapiro and Stadtman, 1968). GS catalyzes the ATP-dependent condensation of ammonia with glutamate to form glutamine, thus functions as the key enzyme in nitrogen assimilation. GS activity is regulated via a bicyclic cascade of an adenylyl transferase (ATase), a signal transduction enzyme PII and an uridylyl transferase (UTase), where ATase is the regulatory enzyme that fine-tunes the GS activity in response to the nitrogen levels in the environment. ATase is a multi-domain enzyme with N-terminal adenylyl removase (AR) domain, C-terminal adenylyl transferase (AT) domain and a regulatory domain (R) (Anderson et al., 1970).

When the level of nitrogen is high, glutamine binds to AT and activates transfer of AMP moiety to GS on a conserved tyrosine residue, placed near the glutamate binding site. Due to bulky AMP moiety, glutamate binding gets prohibited, thus adenylylation of GS effects leads to its low activity (Figure 1.2). Under nitrogen-limiting conditions, GS-AMP is deadenylylated via phosphorolysis mediated by the AR domain of ATase, re-establishing GS activity. The uridylyl transferase (UTase), a second bifunctional enzyme directly senses the nitrogen level within the cell. When the nitrogen levels are low, UTase covalently attaches an UMP moiety onto the PII enzyme. PII-UMP binds to the ATase R-domain and stimulates its deadenylylation activity.

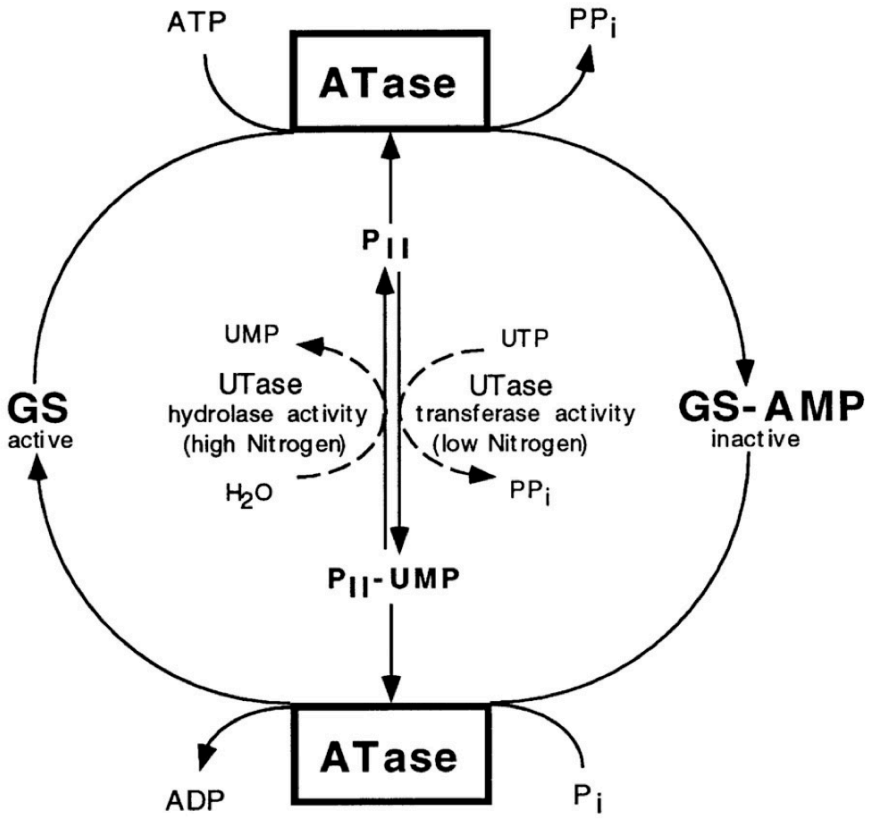


Figure 1.2. The bicyclic regulatory cascade controls glutamine synthetase activity. *The signal transduction protein PII is interchanged between its unmodified and its uridylylated form by the hydrolase and transferase activities of the UTase. The adenylylation and deadenylylation activities of adenylyl transferase (ATase) are regulated by PII and PII-UMP, respectively. Adenylylation of glutamine synthetase (GS) inhibits its activity, whereas deadenylylation restores its function. Sourced from (Jaggi et al., 1997).*

Under conditions of nitrogen excess, UTase cleaves the UMP from PII-UMP using its phosphodiesterase activity (Jaggi et al., 1997; Jaggi et al., 1996). The unmodified PII stimulates the adenylylation activity of ATase, decreasing the rate at which GS converts glutamate to glutamine (Figure 1.2).

AMPylation is recently discovered process as it involves addition of AMP in a stable manner (Woolery et al., 2010). This addition of an AMP moiety is similar to other processes such as SUMOylation, acetylation and phosphorylation.

1.3 Fic; Filamentation induced by cAMP gene in *E. coli*

Fic stands for filamentation induced by cAMP and was named after a morphological effect observed by Utsumi *et al*, in a knockout in *E. coli* K12 strain, PA3092, where increased level of cAMP at an elevated temperature (43 °C) led to inhibition of cell division, consequently the cellular morphology was altered from rod to filaments (Utsumi et al., 1983; Utsumi et al., 1982; Utsumi et al., 1981). In the filamented cells, nucleoids were regularly distributed, however, septum formation was inhibited (Figure 1.3). Interestingly, the inhibitory effect of cAMP on cell division and filamentation in a *fic* mutant was not observed in c-AMP receptor protein (*crp*) mutant, R1005 (Kawamukai et al., 1988; Kawamukai et al., 1989; Komano et al., 1991). However, when the *crp* was complemented, the cell filamentation was observed. Further investigation on *fic* mutants suggested that the filamentation effect might involve perturbation in p-aminobenzoic acid (PABA) or folate metabolism (Komano et al., 1991).

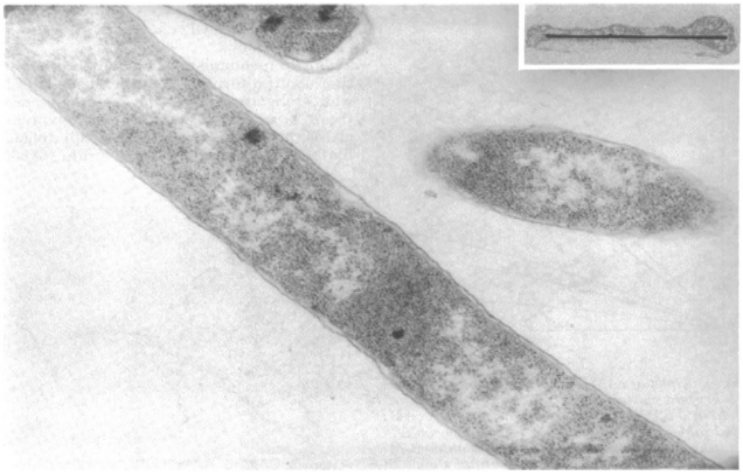


Figure 1.3. E. coli fic mutant impairs cell division *when incubated in high cAMP concentration resulting in filamentous cells. Adapted from (Utsumi et al., 1981).*

1.4 AMPylation exhibited by Fic proteins

The first report on AMPylation appeared in 2009 where *Vibrio* outer protein S (**VopS)** from *Vibrio parahaemolyticus* was characterized as the first example of a Fic domain functioning as an AMPylator (Luong et al., 2010; Yarbrough et al., 2009). *V. parahaemolyticus* is an extracellular Gram-negative bacterium that causes

gastroenteritis from eating undercooked seafood (Daniels et al., 2000). An essential virulence factor for many Gram-negative pathogens, including *V. parahaemolyticus*, is the T3SS, a needle-like structure that extends from the bacterium and penetrates a host cell to inject effectors (Park et al., 2004). VopS with other effectors, released in host cytosol through T3SS1 were implicated in causing cell rounding, induction of autophagy and cytotoxicity (Burdette et al., 2008).

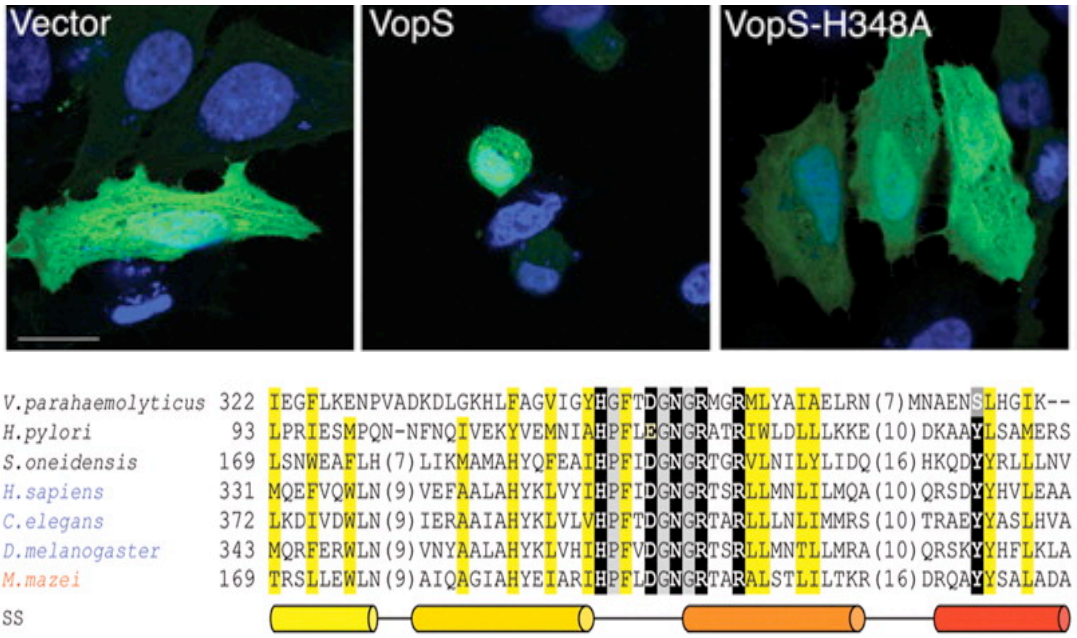


Figure 1.4. Human HeLa cells transfected with VopS shows cell rounding. *This phenotype, typical of actin cytoskeleton collapse, results from inactivation of Rho GTPase signaling. Mutant of the Fic active-site histidine of motif HxFx(D/E)GNGRxxR, displays normal cell shape, proving that the activity resides in the FIC domain. Adapted from (Yarbrough et al., 2009).*

VopS was involved in triggering cell rounding by inactivating Rho GTPases (Rho, Rac, Cdc42) (Figure 1.4). The reports by Orth K. *et al*, showed that VopS AMPylates a specific threonine (T35 in Rac1 and Cdc42; T37 in RhoA) on the switch I region of the GTP bound conformations of Rho family GTPases, preventing binding of downstream effectors like p21 activated kinase 1 protein (PAK) (Luong et al., 2010; Worby et al., 2009; Yarbrough et al., 2009). The loss of this interaction disables the host cell's control on the actin cytoskeleton, which led to cell rounding (Figure 1.4).

AMPylation of Rho GTPases further counters host immune strategy by inhibiting NFκB mediated immune response and alters production of reactive oxygen species by NADPH oxidase (Woolery et al., 2014). VopS mediated AMPylation also perturbs NLRC4 inflammasome activation, however, activates Pyrin inflammasome. Inflammasome activation stimulates caspase 1 to process pro-inflammatory cytokines into their active forms and promotes pyroptosis, a highly inflammatory form of programmed cell death of the host that is associated with an antimicrobial response (Roy and Cherfils, 2015; Zhao et al., 2014).

A parallel report uncovered another AMPylator of Rho GTPases, Immunoglobulin-binding protein A (**IbpA**) from *Histophilus somni*, an opportunistic pathogen that infects the mucosa of cattle and sheep (Worby et al., 2009; Zekarias et al., 2010). IbpA, a fibrillar surface antigen, is a major virulence factor of *H. somini* and contains two Fic domains, present at its C-terminus. As in VopS, IbpA also AMPylates Rho GTPases at switch I region and leads to cell rounding and cytoskeletal collapse in HeLa cells (Figure 1.5) as well as its most physiological relevant target- primary bovine alveolar type 2 (BAT2) cells (Worby et al., 2009).

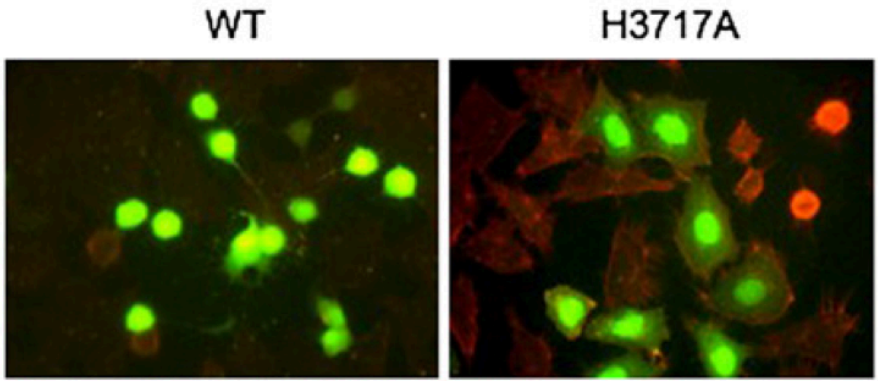


Figure 1.5. Human HeLa cells transfected with EGFP-IbpA. *Histidine 3717 is critical for Ibp-Fic2 cytotoxicity (Worby et al., 2009)*

Interestingly, the modification by IbpA was apprehended to a tyrosine of the switch I region of Rho GTPases (Y32 in Rac1 and Cdc42; Y34 in RhoA). Worby *et al*, also showed that this modification could be reversed by treating AMPylated substrates with a promiscuous snake venom phosphodiesterase. Unlike *V. parahaemolyticus,* it

lacks a specialized secretion system and IbpA functions as a conventional toxin that is secreted and then internalized by the alveolar cells.

Subsequent to the identification of VopS and IbpA as AMPylators, many genes have been annotated to encode Fic proteins, which can mediate AMPylation. It is noteworthy that the functions of fic domains have not been only restricted to the covalent transfer of AMP (Figure 1.6). As in case of; AnkX, a T4SS effector of *Legionella pneumophila*, modifies small Rab family GTPases by phosphocholine transfer (**phosphocholination**) (Mukherjee et al., 2011) and AvrAC from *Arabidopsis thaliana* that transfers UMP (**UMPylation**) to the activation loop of kinase targets BIK1 and RIPK, impeding phosphorylation required for their activation (Feng et al., 2012).

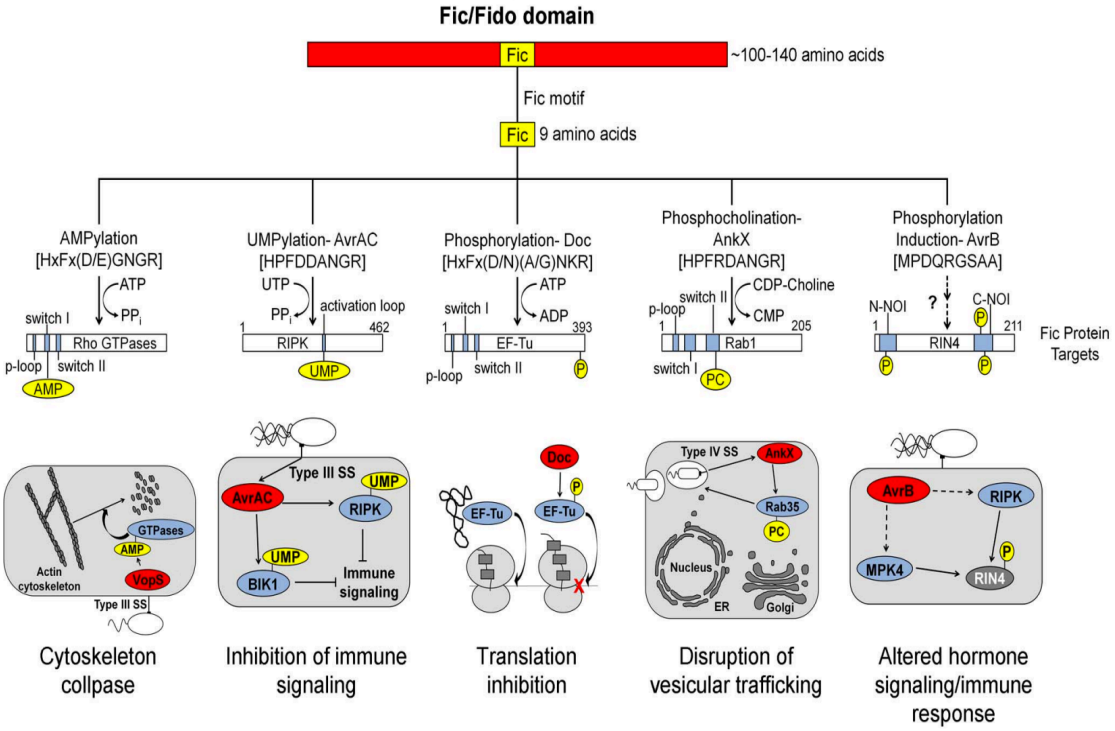


Figure 1.6. Overview of Fic domain proteins, their targets and their roles in bacterial virulence. *The Fic domain is defined by a conserved structural fold composed of six to eight a-helices (red bar) and a nine amino acid Fic motif (yellow box). Important domains of the target proteins are shown as blue bars; modifications as yellow circles (not to scale). Known physiological manifestations of each class of*

Fic domain protein are illustrated below their target. Taken from (Cruz and Woychik, 2014)

AnkX modifies Rab1 and Rab35 on serine residues located within the switch II region of the GTPase (Mukherjee et al., 2011; Tan et al., 2011). As this region is involved in binding to guanine nucleotide exchange factors (GEFs) and guanine nucleotide dissociation inhibitors (GDIs), the regulators of activation and localization of small GTPases, their phosphocholination leads to modulation of host secretory pathway by rearrangement of host vesicular trafficking such that *L. pneumophila* retains in the *Legionella*-containing vacuoles (LCV) that enable its replication inside the host cell. Briefly, GEFs stimulate the exchange of GDP for GTP to generate the active GTPase-GTP form, whereas GTPase-activating proteins (GAPs) generate the inactive GTPase-GDP form by the hydrolyzing GTP. GDIs are located in the cytoplasm and regulate the extraction of inactive GTPase-GDP from membranes and sequester it in the cytoplasm. Phosphocholination of Rab GTPase obstructs its interactions with GEFs and GDIs, thus impairing their activation by GEFs and membrane extraction by GDIs. As a consequence, phosphocholinated Rab proteins presumably remain retained on membranes in their inactive Rab-GDP form (Gavriljuk et al., 2013; Goody et al., 2012; Oesterlin et al., 2012).

AvrAC is encoded by a bacterium *Xanthomonas campestris* that infects members of the plant family Brassicaceae, including important food crops as well as model organism *Arabidopsis thaliana*. The target substrates of AvrAC, BIK1 and RIPK, mediate immune signaling in plant cells (Feng et al., 2012). As UMPylation of BIK1, and RIPK (RPM1 induced protein kinase) kinases impairs their activation, AvrAC enhances bacterial virulence by interfering with the ability of plants to induce the innate immune response against the pathogen. A recent report suggests that likewise other host targeted effectors; the activity of AvrAC can be sensed in plant cells that induce effector-triggered immunity. Detection of AvrAC depends on UMPylation of PBL2, a paralog of BIK1, which serves as a decoy substrate for the effector (Wang et al., 2015).

1.5 Phosphorylation exhibited by Doc proteins

All Doc family members are toxins of the Phd-Doc toxin-antitoxin (TA) module (Castro-Roa et al., 2013; Cruz et al., 2014; Cruz and Woychik, 2014). TA systems are small operon that encodes a stable toxin and a labile antitoxin. Under stress condition, the labile antitoxin domain is degraded, releasing the toxin free from TA protein-protein complex, which can act on its target(s), leading to growth arrest. The Doc toxin from the bacteriophage P1 TA system is the only member of the family, studied in detail. In the *E. coli* phage P1 the Doc-PhD TA module serves as a post-segregational killing system (PSK) to ensure stable inheritance of plasmid in the bacterial population (Yamaguchi and Inouye, 2011). In the absence of the Phd antitoxin, Doc induces cell growth arrest by inhibition of translation via phosphorylation of the essential elongation factor and GTPase EF-Tu. Doc phosphorylates the threonine residue of the GTPase domain of EF-Tu, which is involved in stabilization of EF-Tu in its active GTP bound form. Phosphorylation of EF-Tu at this threonine impairs its conversion into an active form, hence prevents EF-Tu binding to ribosomes and stalls translation. Thus, in the absence of antitoxin Phd in cells that have lost the Phd-Doc prophage plasmid, Doc induces growth arrest and ensures a stable inheritance of plasmid in bacterial population (Cruz et al., 2014). Translation arrest by a single phosphorylation step represents a highly effective and specific mechanism in Phd-Doc TA system for growth arrest and subsequently persister cell formation. Recently, Phd-Doc TA module has been shown to contribute persister formation in *Salmonella enterica subsp. enterica serovar, Typhimurium* persisters. Doc toxins from several other pathogens include those from *Streptococcus pneumonia, Vibrio cholera* and *Clostridium tetani* (Helaine et al., 2014).

Doc family proteins together with Fic proteins belong to Fic/Doc superfamily; however, they exhibit a remarkable difference in their protein core fold and conserved key motif. Doc proteins possess only six α helices in their core while fic fold acquires eight core helices. The conserved Doc motif, HxFx(D/N)(A/G)NKR differs slightly from the canonical fic motif, HxFx(D/E)GNGR (Castro-Roa et al., 2013).

1.6 Phosphorylation exhibited by AvrB proteins

AvrB is T3SS effector protein of a plant pathogen *Pseudomonas syringae* and enhances growth of the pathogen in the receptive plant (Desveaux et al., 2007). AvrB has an α helical core Fic fold, however, it has very different residues in place of Fic catalytic motif. AvrB interacts with the immune regulator RIN4 (RPM1 interacting proteins) and the signaling kinases RIPK and MPK4 and also induces phosphorylation of RIN4 and MPK4 (Chung et al., 2011; Mackey et al., 2002). In susceptible plants, the phosphorylation triggered by AvrB interferes with the role of RIN4 in perception of pathogen-associated molecular pattern (PAMPs). Moreover, this induced phosphorylation of RIN4 activates the plasma membrane H^+-ATPase to induce stomatal opening, which promotes bacterial invasion. The mechanism of AvrB-dependent phosphorylation of RIN4 is unclear but two structures of AvrB in complex with either a RIN4 peptide or ADP suggest a kinase activity for AvrB (Chung et al., 2011).

1.7 Legionella pneumophila DrrA ATase

Legionella pneumophila is an opportunistic human pathogen and the causative agent of Legionnaires' disease. A hallmark of *Legionella* pathogenesis is endoplasmic reticulum-derived vacuoles, called LCV, which circumvent fusion with endocytic vesicles and lysosomes, and helps in bacterial survival and proliferation (Horwitz, 1983). *L. pneumophila* manipulates intracellular vesicular trafficking through secretion of effectors via the T4SS. These effectors manipulate host factors, such as Sar1, Sec22b, and Rab1 to establish the LCVs (Kagan et al., 2004). In this context, the versatile *L. pneumophila* factor DrrA (also known as SidM) is involved in the recruitment and manipulation of the small GTPase Rab1. DrrA contains three domains, wherein the C-terminal lipid phosphatidylinositol-4-phosphate binding domain (P4M) mediates attachment to the LCVs membrane; the central guanine nucleotide exchange factor (GEF) and the N-terminal adenylyl transferase activities subvert Rab1 functions. The N-terminal domain is structurally similar to the GS adenylyl transferase (Muller et al., 2010). Likewise, it transfers AMP to a conserved tyrosine in the switch II region of Rab1-GTP (Tyr77 in Rab1b), thereby inhibiting GAP-stimulated GTP hydrolysis and

trapping Rab1 in an active state (Muller et al., 2010). Interestingly, later during infection, *L. pneumophila* also injects SidD, an effector that possesses AMP hydrolase activity to counteract DrrA function (Tan and Luo, 2011). Combination of effectors with opposite activities allows *L. pneumophila* to achieve an accurate spatiotemporal control of the Rab1 signaling pathway (Neunuebel et al., 2011; Tan et al., 2011) (Figure 1.7).

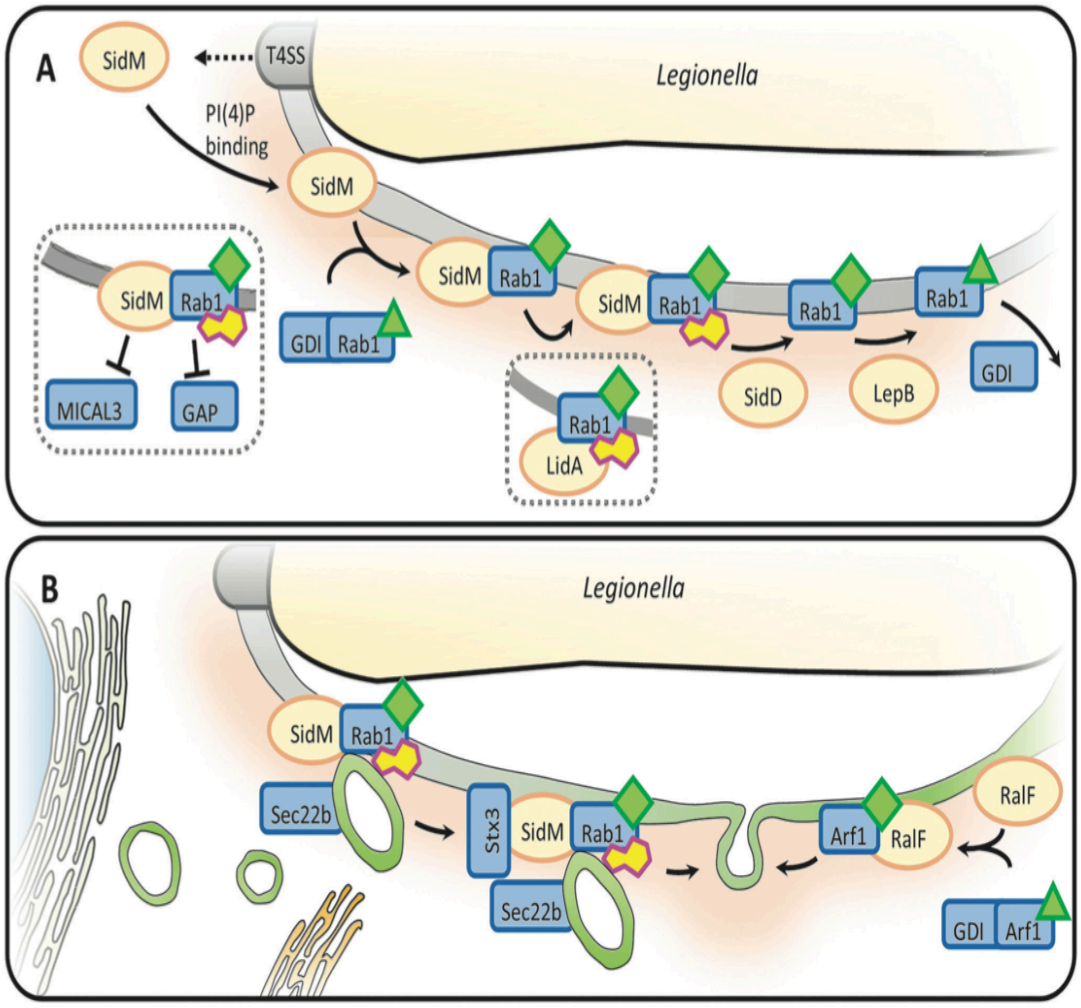


Figure 1.7. Legionella pneumophila effectors subvert Rab GTPases on the Legionella-containing vacuole (LCV). (A) Rab1 recruitment and manipulation by AMPylation. *SidM anchors onto the LCV by binding PI(4)P. It activates and recruits Rab1 onto the LCV using its guanine nucleotide exchange factor (GEF) domain and then locks Rab1 in the active state by AMPylation. SidD later deAMPylates Rab1 allowing the GTPase- activating protein (GAP) LepB to deactivate*

Rab1 and GDP dissociation inhibitors (GDIs) to extract it from the LCV. ***(B) Recruitment of Rab1 and ARF1 leads to endoplasmic reticulum (ER)-derived vesicle fusion with the LCV.*** *Rab1 activation by SidM results in tethering of Sec22b-containing ER-derived vesicles onto the LCV. In addition, SidM recruits syntaxin-3A facilitating noncanonical SNARE (soluble N-ethylmaleimide- sensitive factor activating protein receptors) pairing between Sec22b and syntaxin-3A. This induces fusion of ER-derived vesicles with the LCV. RalF acts as an ARF1 GEF to recruit and activate ARF1 onto the LCV where it promotes fusion of ER-derived vesicles. Taken from (So et al., 2015).*

1.8 *Bartonella* effectors

The genus *Bartonella* represents an increasing number of emerging bacterial pathogen that subverts cellular functions of host cells by injecting the type IV secretion system (T4SS) effector proteins, parasitizing the red blood cells of their eukaryotic hosts (Schulein et al., 2005). One of the effectors from *Bartonella sp*. is Bep protein, which are mostly composed of an N-terminal Fic domain with HPFxxGNG signature motif (Palanivelu et al., 2011) and a C-terminal *Bartonella* intracellular delivery (BID) domain, the latter being responsible for T4SS-mediated translocation into host cells. BepA is the most studied where BID domain is responsible for intracellular delivery as well as inhibition of apoptosis by the host cell (Schmid et al., 2004). Though the activity has been shown in the mammalian cell lysate, the exact function of this Fic domain remains unknown. However, vimentin, a type III intermediate filament of mesenchymal derived cells, has been recognized as a substrate of Bep2 Fic domain from *B. rochalimae* (Pieles et al., 2014).

The characterized members of Fic/Doc family are also summarized in Table 1.1.

Table 1.1. Summary of the characterized Fic/Doc effectors

Proteins	Organisms	Targets	Functions
Doc	Bacteriophage P1	Phosphorylation of EF-Tu	Bacterial cell growth arrest
AvrB	*Pseudomonas syringae*	Involved in phosphorylation of RIN4	Plant cell death. Negative regulator of plant innate immune response against pathogens
AnkX	*Legionella pbeumophila*	Ser/Thr phosphocholination of Rab1 GTPase	Disruption of membrane transport processes
VopS	*Vibrio parahaemolyticus*	Threonine AMPylation of Rho GTPase	Collapse of the host-cell actin cytoskeleton
IbpAFic2	*Histophilus soni*	Tyrosine AMPylation of Rho GTPase	Collapse of the host-cell actin cytoskeleton
FicD	*Drosophila sp.*	AMPylation of histones, HSPs and Bip	Involved in the detection of light stimulus and histamine transport.
AvrAC	*Xanthomonas campestris*	UMPylation of Arabidopsis BIK1 and RIPK, two receptor-like cytoplasmic kinase	Inhibition of plant immunity
VbhT	*Bortonella schoenbuchensis*	AMPylation of Gyrase B and ParE	Inhibits bacterial growth
NmFic	*Neisseria meningitidis*	AMPylation of Gyrase B	Inhibits bacterial growth
Bep2	*Bartonella rochalimae*	AMPylation of Vimentin	Unknown

1.9 Fic domain catalysis

Fic motif, HxFx(D/E)GNGRxxR, is present between the α3-α4 helices and the minimum core of Fic domain is defined by the set of six helices that hold the active site loop and the substrate binding residues (Roy and Cherfils, 2015). Fic mediated catalysis has been proposed to follow sequential binding mode to form a ternary complex between the Fic protein and the two substrates; ATP and the interacting target protein (Luong et al., 2010; Xiao et al., 2010). The key residue for base catalysis is invariable histidine from the Fic motif that can abstract a proton from the hydroxyl side chain of the target residue (Worby et al., 2009; Yarbrough et al., 2009). The activated hydroxyl group can then perform a nucleophilic attack on the α-phosphate of ATP (Figure 1.8). The ternary complex formation has been proposed based on the steady-state kinetic experiments of VopS (Luong et al., 2010) and IbpA and through the co-crystal of Ibp2 and Cdc42 (Xiao et al., 2010).

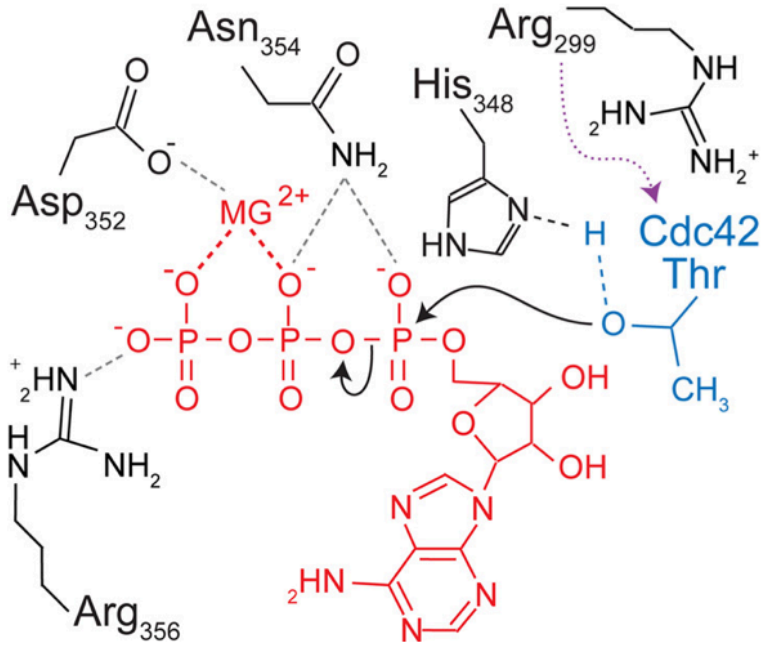


Figure 1.8. The proposed mechanism of VopS Fic domain-mediated AMPylation. *Histidine 348 acts as a general base to deprotonate the hydroxyl residue of the target protein, shown in blue. Aspartic acid residue coordinates the Mg^{2+} ion, while asparagine and arginine residues coordinate phosphates of ATP, shown in red. Adapted and reproduced from (Luong et al., 2010).*

1.10 Structural features for substrate recognition

1.10.1 Fic structural core

Fic domain is an arrangement of eight helices where α2-α5 forms the core and structure while the α1', α1, α6, and α7 are arranged in the periphery of the core α helices (Figure 1.9) (Kinch et al., 2009). However, a structural core of six α-helices defines a minimum Fic domain (Garcia-Pino et al., 2014).

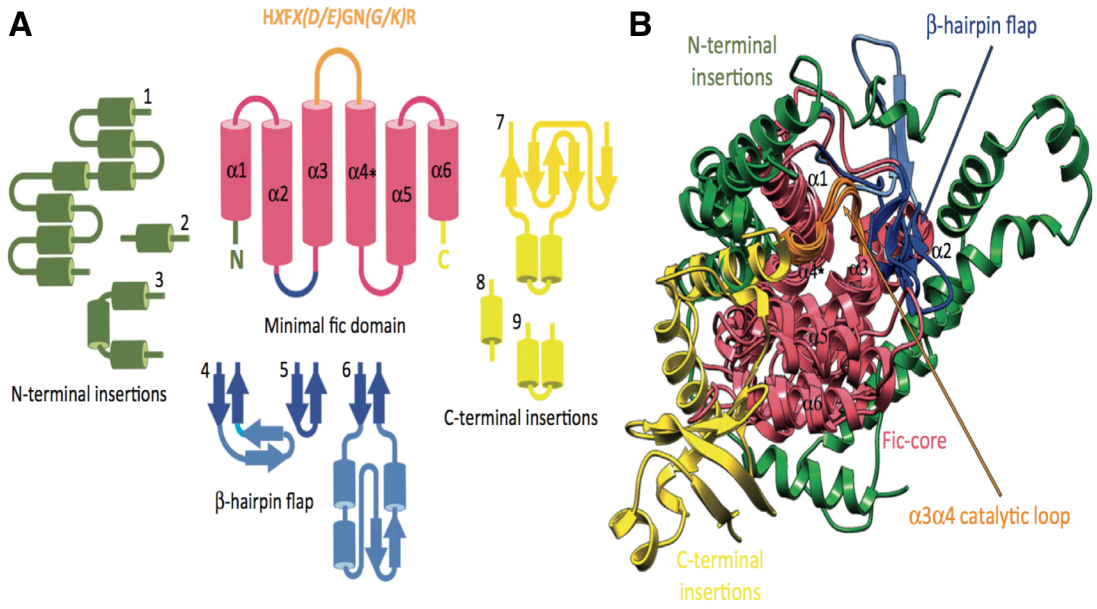


Figure 1.9. The Fic core and functional extensions. (A) A minimal Fic domain, *as observed in the toxin Doc (coloured raspberry). The six a-helices constituting the domain are labeled in black and the regulatory helix a4 is marked with a *. The three major classes of functional extensions, and their grafting sites at the N terminus, C terminus, and a1a2 loop (or β-hairpin flap) are coloured green, yellow, and blue, respectively. Additional structural elements from the β-hairpin flap are light blue. Examples of specific extensions are shown to the left, right, and below the minimal Fic domain: (1) N-terminal lobe of IbpA2 (immunoglobulin binding protein A of Histophilus somni); (2) CDP–choline binding domain from AnkX (ankyrin repeat-containing protein; Legionella pneumophila); (3) N-terminal a helix of NmFic (Neisseria meningitidis); (4) BepA (Bartonella henselae) β- hairpin flap with an additional b-hairpin insertion; (5) NmFic b-hairpin flap; (6) AnkX insert domain*

protruding from the b-hairpin flap; (7) BepA C-terminal domain; (8) NmFic C-terminal a helix; and (9) AnkX ankyrin tandem repeat. ***(B) Superposition of four representative Fic enzymes*** *represented as cartoons, Doc (death on curing; bacteriophage P1), NmFic (N. meningitidis), BepA (B. henselae), and IbpA2 (H. somni). The different structural elements of the enzymes are coloured according to the panel (A). Adapted from (Garcia-Pino et al., 2014).*

1.10.2 Nucleotide binding

Though the diverse reactions catalyzed by Fic proteins are based on the substrate orientation, the chemistry of reaction and the role of Fic motif in substrate binding appear to be conserved. Phenylalanine and asparagine of the Fic motif balance the catalytic loop conformation (Roy and Cherfils, 2015). Phenylalanine side chain interactions anchor the loop in the hydrophobic core while asparagine amide interaction stabilizes the catalytic loop in an upright conformation (Luong et al., 2010).

In NMPylating enzymes, the nucleoside moiety of NTP locks into a hydrophobic patch, formed between α3 and α5 helices and the β-hairpin of the flap. The GNG submotif forms an "anion hole" that accommodates the α-phosphate of the nucleotide, mainly through the direct hydrogen bond interaction with the polypeptide backbone (Xiao et al., 2010). The terminal conserved arginine of the Fic motif tethers the nucleotide further by interacting with ribose of the cofactor. Therefore, with the conserved Fic motif, the hydrophobic adenine binding pocket, the ribose coordination and the anion hole remain conserved in NMPylating Fic enzymes (Figure 1.10) (Garcia-Pino et al., 2014).

ATP binding canonical Fic motif has been observed in VopS, IbpA, VbhT, BepA, *So*Fic (*S. oneidensis*), *Nm*Fic (*N. meningitis*) and Doc proteins as well (Cruz et al., 2014; Engel et al., 2012; Luong et al., 2010; Xiao et al., 2010). The diphosphate moiety is central to the nucleotide selection by Fic proteins as the Fic motif cleaves the cofactor between the two phosphates of the diphosphate moiety and transfers one of the resulting phosphate to the hydroxyl group of threonine or tyrosine of the substrate (Figure 1.10). The cofactor group located towards the conserved histidine residue gets

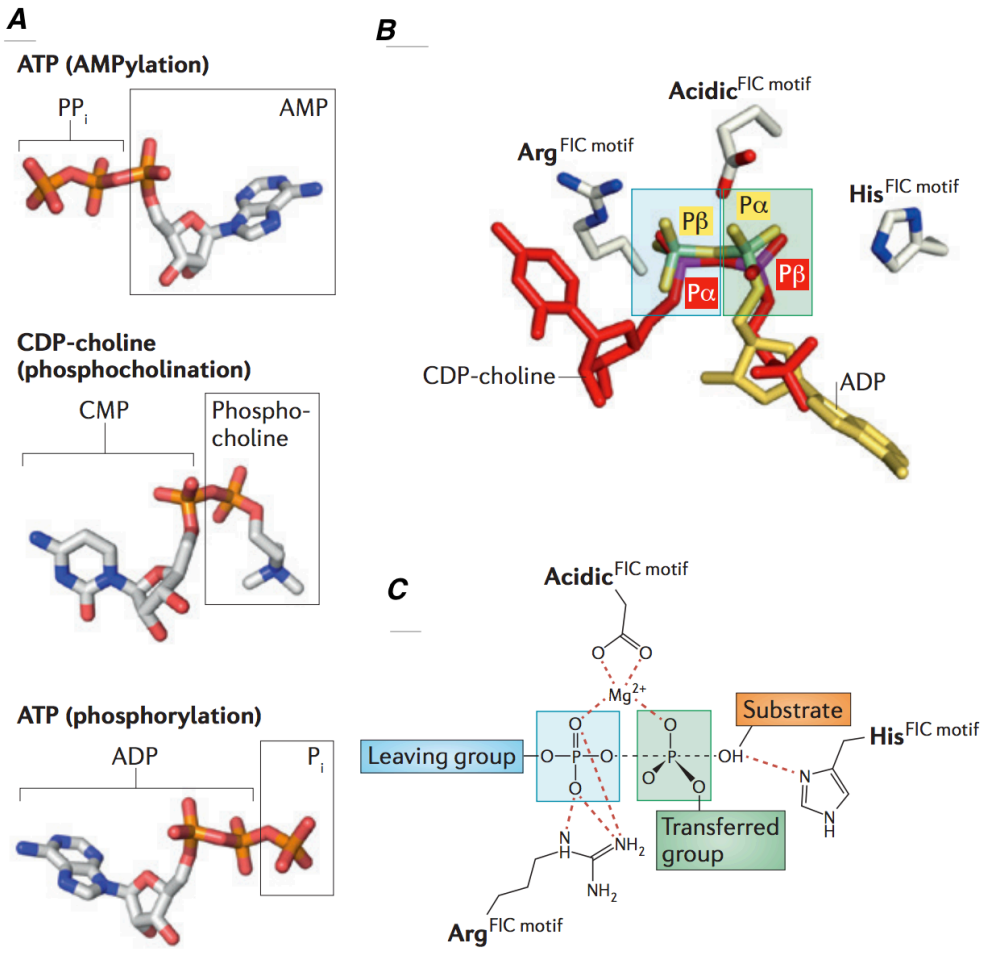


Figure 1.10. Nucleotide selection Fic proteins. (A) Fic protein cofactors and their use as donors for PTMs. *The transferred group (boxed) is AMP for AMPylating Fic proteins, phosphocholine for AnkX from Legionella pneumophila and phosphate (Pi) for the Doc component of the Doc–Phd from Escherichia coli. The leaving group is diphosphate (PPi), CMP or ADP, respectively.* ***(B) The FIC catalytic motif binds the diphosphate moiety of cofactors at a fixed position.*** *CDP-choline (red) bound to AnkX (PDB ID 4BET) and ADP (yellow) bound to human Huntingtin-interacting protein E (HYPE) (PDB ID 4U0U) are superimposed. The phosphate that belongs to the transferred group is boxed in green, and the phosphate that belongs to the leaving group is boxed in blue. The α-phosphates and β-phosphates of CDP-choline bound to AnkX and of ADP bound to HYPE are inverted with respect to the FIC motif.* ***(C) Schematic representation of the general catalytic mechanism of Fic proteins.*** *Adapted from (Roy and Cherfils, 2015).*

transferred to the substrates. Therefore, the orientation of the bound cofactor with respect to the Fic motif is the deciding factor for the group getting transferred and the group that will be released. The interesting example in this context is AMPylation in other Fic proteins (Ibp and VopS) and phosphocholination observed in AnkX. Both ATP and CDP-choline have nucleotide moiety. Ibp transfers AMP and releases PPi while AnkX transfers phosphocholine and releases CMP (Mukherjee et al., 2011). These differences arise due to the inverted positioning of α and β phosphates of ATP and CDP-choline, also providing different AMP and CMP binding pocket (Goody et al., 2012) (Figure 1.10). Thus the effector protein AnkX of *Legionella Pneumophila* provides a striking example of a Fic protein harboring a motif divergent from the consensus and exhibiting a related but diverse function. Fic motif (HxFxDANGRxxV) of AnkX shows changes in 4^{th} and 8^{th} position (glycine to alanine and arginine to valine) can accommodate phosphocholine in the cofactor-binding pocket (Goody et al., 2012; Roy and Cherfils, 2015).

AvrB, involved in phosphorylation of RIN4, binds ATP in a transposed mode, which is coupled, to the loss of the Fic signature motif of the conserved loop replaced by P_{263}DQRGSAA_{270}. However, the peptide backbone of the reaction centre and anion hole submotif involved in diphosphate binding remains conserved (Kinch et al., 2009).

In Doc proteins, the non-canonical Fic motif contains lysine in place of glycine within the anion hole-forming submotif "GNG". This change reflects in the inverted positioning of β-phosphate and γ-phosphate, similar as in AnkX α-phosphate and β-phosphate inversion. Due to the inverted conformation of ATP, the Pi moiety points towards the Fic motif histidine, thus Doc proteins can perform phosphorylation instead of AMPylation of EF-Tu elongation factor (Castro-Roa et al., 2013).

1.10.3 Substrate recognition and interaction

Different Fic proteins modify different targets, however the individual Fic protein shows specificity in substrate recognition. Based on the information of AvrB-Rin4 (Desveaux et al., 2007) and IbpA-Cdc42 (Xiao et al., 2010) co-crystals, the major contributor in the interactions of Fic protein and the target substrate seem to be the

common β-hairpin element located between helices α2 and α3. In AvrB, this segment forms anti-parallel β-strand hydrogen-bonding interactions with the RIN4 peptide and mutations of distinct β-hairpin residues impaired the ability of AvrB to promote immune response in plant. The importance of this element for target recruitment is confirmed by the structure of IbpA-Fic2 in complex with its target protein Cdc42 where a similar β-β interaction has been obtained. Notably, this β-hairpin is present in the flap region of all Fic structures solved so far, but absent in Doc, which might further confirm that Doc harbors another enzymatic function (Cruz et al., 2014). Target recognition via sequence-independent main chain-main chain hydrogen bonds implies the presence of additional structural elements to guide substrate specificity in Fic proteins. Consistently, the complex structure of IbpA-Fic2 with Cdc42 reveals a large interaction interface between the switch II region of the small GTPase and the so-called "arm domain" of IbpA. An equivalent peripheral region is found in VopS Fic protein that is likely to confer the same specificity towards the small Rho GTPases (Luong et al., 2010). Similarly, in AvrB, the Fic core β-hairpin structure is extended into a five-stranded antiparallel β sheets connected to three α-helices. This domain, referred to as upper lobe, provides a large surface for RIN4 binding. Apart from the Fic proteins of *Shewanella oneidensis and Bacteriodes thetaiotaomicron* that have large helical extensions at the N- and C-termini to possibly recognize targets, these additional satellite domains are not present in the remaining Fic proteins with known crystal structures. However, beside the four-helix bundle Fic core domain that, for catalytic reason, forms a rigid entity, the peripheral segments that regroup the helices α1, α6-α7 and α' are found in diverse dispositions in the different Fic structures. These components, together with further loop extensions are probably relevant for target specificity (Garcia-Pino et al., 2014).

1.11 Regulation of Fic proteins

Fic proteins have been described as type II toxin-antitoxin (TA) module where both the interacting partners are proteins. The antitoxin regulatory module is conserved in all Fic proteins that are bona fide nucleotidyl transferase (NMPylase). The antitoxin module is formed by an inhibitory α-helix containing a conserved (S/T)xxxE(G/N)

motif with an invariable acidic residue glutamate which obstructs the binding orientation of the nucleotide, that further prevents the in-line nucleophilic attack from the target hydroxyl residue and interferes with the Fic mediated catalysis (Figure 1.11). A-inhibitory domain (α' helix) is present near the α1 helix (Engel et al., 2012). Based on the positioning of the obstructive motif and whether it is part of the same polypeptide or not, Fic proteins have been divided into three classes; Class I where antitoxin module is part of the same operon present upstream to the Fic domain, Class II and III where the inhibitory module is present as part of the same polypeptide chain, present at N- or C-termini, respectively (Engel et al., 2012) (Figure 1.11). Conversely, Doc protein function is inhibited by Phd protein, however, Phd does not have the inhibitory motif but prevents the substrate binding when bound to Doc.

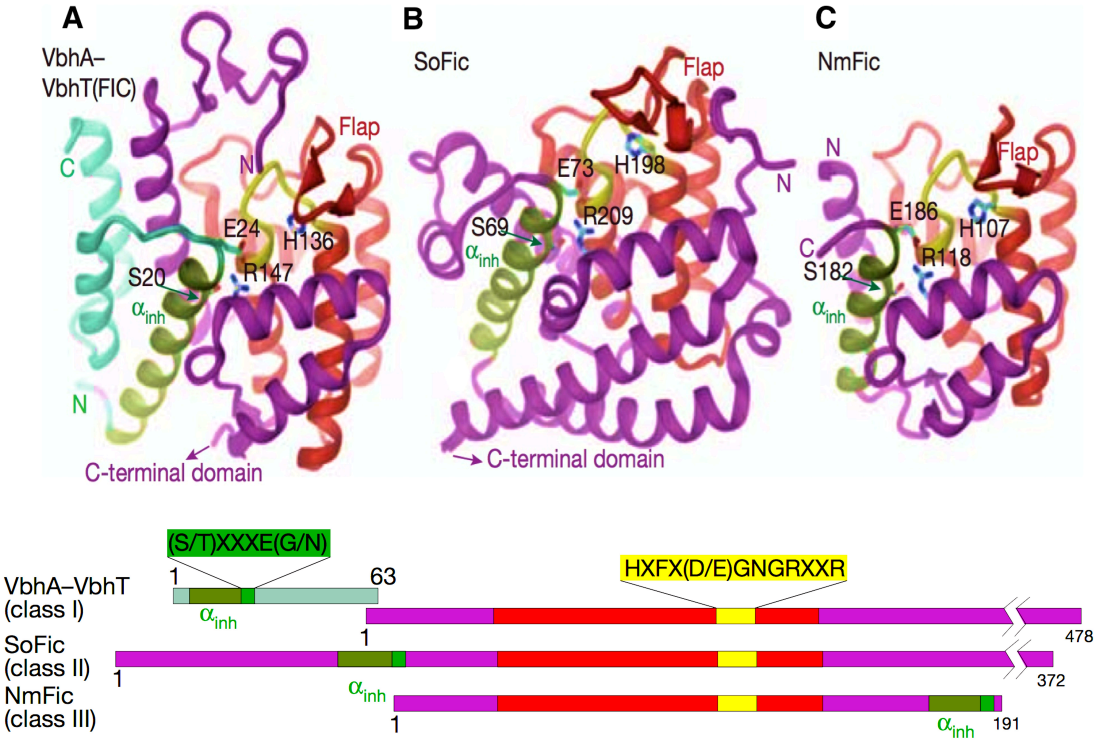


Figure 1.11. Classifications of Fic domains. (A–C) Ribbon diagrams of Fic domain structures. *Fic domain core is shown in red, active site loop is shown in yellow, inhibitory α-helix is shown in green.* ***(A)*** *Class I Fic domain VbhT is shown in complex with antitoxin VbhA.* ***(B)*** *Class II Fic domain SoFic (from S. oneidensis).* ***(C)*** *Class III Fic domain from NmFic.* ***(D)*** *Protein domain maps of VbhT/VbhA, SoFic, and NmFic. Adapted from (Engel et al., 2012).*

The activity of the Fic toxin proteins is also regulated by auto-modification (auto-NMPylation) in absence of obstructive glutamate (Goepfert et al., 2013). Fic auto-modification plays a deterministic role in activity in class III Fic proteins where auto-AMPylation at a tyrosine residue leads to de-oligomerization, subsequently, the substrate binding β-hairpin region becomes accessible (Stanger et al., 2016a). In class I Fic proteins, reports from *fic-1* from *Pseudomonas fluorescens* suggests that auto-AMPylation is probably essential for catalytic activity (Lu et al., 2016).

1.11.1 Class I Fic proteins

Class I Fic proteins are present in bacteria and are characterized by the occurrence of the inhibitory domain as a separate small protein (Figure 1.11). Likewise, in other anti-toxin proteins, the inhibitory protein perhaps would be degraded by proteolysis in stress condition to release the toxin viz. the Fic protein. The characterized members include VbhT-VbhA from *Bartonella schoenbuchensis*, PaFic TA from *P. fluorescens*, YeFic TA from *Yersinia enterocolitica* str. 8081 and EcFic-YhfG TA from *E. coli*. Growth inhibition caused by Fic toxins VbhT and YeFic correlates with the AMPylation of two target proteins GyrB and ParE (topoisomerase IV) (Figure 1.12) (Harms et al., 2015). These constitute the bacterial type II topoisomerase and are the major controller of bacterial DNA topology by introducing negative supercoiling (DNA gyrase) and by decatenating and unknotting the nucleoid (topoisomerase IV) (Figure 1.12). These Fic toxins AMPylate the specific tyrosine residue of GyrB, that forms a loop surrounding the ATP binding site. Due to the bulky AMP moiety, AMPylation impedes the ATP binding thus ATPase activity of GyrB, essential to gyrase function (Harms et al., 2015). PaFic toxin also modifies GyrB and therefore inhibits *Pseudomonas* growth (Lu et al., 2016).

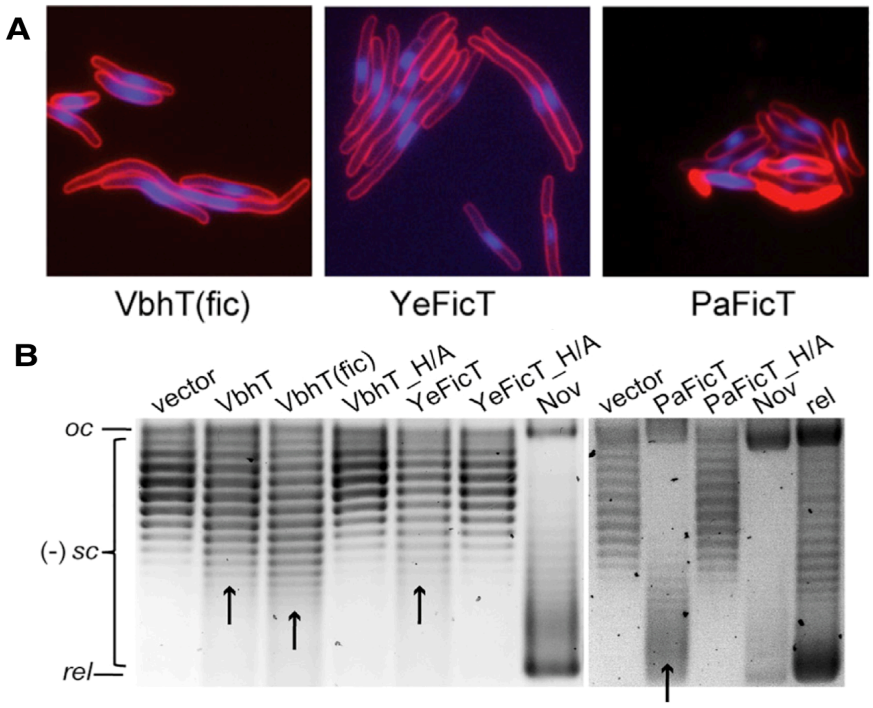


Figure 1.12. FicTs induces par phenotype and DNA gyrase inhibition. (A) *FicT expressed E. coli cells, stained with FM4064 (red) and DAPI (DNA; blue) showing a par phenotype with cell filamentation and unsegregated nucleoids, indicative of strong topo IV inhibition. The nucleoid compaction observed with VbhT(Fic), YeFicT, and PaFicT is likely a consequence of DNA gyrase inhibition.* ***(B)*** *FicTs Cause Divergent Levels of DNA Gyrase Inhibition in vivo. DNA relaxation with VbhT or YeFicT (arrows), indicative of weak inhibition of DNA gyrase, while high concentrations of novobiocin (Nov; 100 mg/ml) or the expression of PaFicT fully abrogate the negative supercoiling. Adapted from (Harms et al., 2015).*

EcFic is another interesting class I Fic protein as it has a divergent Fic motif HPFRVGSGLAQR. In the EcFic motif, a serine residue replaces conserved asparagine and the conserved negatively charged residue (D/E) that coordinates with Mg^{2+} and mediates nucleotide binding is replaced by a hydrophobic residue valine. Recent crystal structure data and biochemical data suggest that EcFic neither shows AMPylation reaction nor modifies GyrB and ParE substrates from *E. coli* (Harms et al., 2015; Stanger et al., 2016b). Though, the cell filamentation was initially attributed to a G55R mutation in EcFic protein (Komano et al., 1991), the structure of native and G55R

mutant EcFic do not show any remarkable structural differences (Stanger et al., 2016b). Further experiments are required to excavate the cofactor and the target of *E. coli* Fic to understand its role in cell division and in cell filamentation.

1.11.2 Class II Fic proteins

Approximately 95% of bona fide Fic proteins belong to class II where the antitoxin motif is part of the same protein, harboring the motif at its N-terminus (Harms et al., 2016; Khater and Mohanty, 2015). Although these proteins are by far most abundant group of Fic proteins, nearly nothing is known about their targets and biological activities. The two characterized members include *So*Fic from *Shewanella oneidensis* (Das et al., 2009) and *Cd*Fic from *Clostridium difficile* (Dedic et al., 2016). The crystal structure of *So*Fic reveals an antitoxin α helical domain, core Fic domain, and a winged helix-turn-helix (wHTH) domain. In the crystal structure, *So*Fic forms a dimer that is likely physiological. Interestingly, the arrangement of two wHTH domain in the dimer are presumed to be compatible to bind the major groove of double-stranded DNA, indicating a bi-functional nature of *So*Fic involved in both AMPylation and DNA binding. *Cd*Fic forms biological dimer both in solution as well as in crystal structure wherein the antitoxin domain is followed by Fic domain. The flap in *Cd*Fic crystal structure is unusually longer in amino acid length, which might provide diversity in different substrate recognition. Interestingly, *Cd*Fic does not follow catalytic inhibition through the inhibitory motif (S/T)xxxEG because of an arginine replacement at the place of tyrosine which coordinates with the α phosphate of the ATP, resulting in a shift of anion hole on the gamma phosphate of the nucleotide (Figure 1.13). Thus, binding of ATP in *Cd*Fic reveals a conformation of ATP unique among Fic domains, which seems to override the effect of the inhibitory helix. Conversely, *So*Fic follows activity regulation by auto-inhibition, the current dogma that all class I-III Fic domain proteins are inhibited by the inhibitory α-helix (Engel et al., 2012).

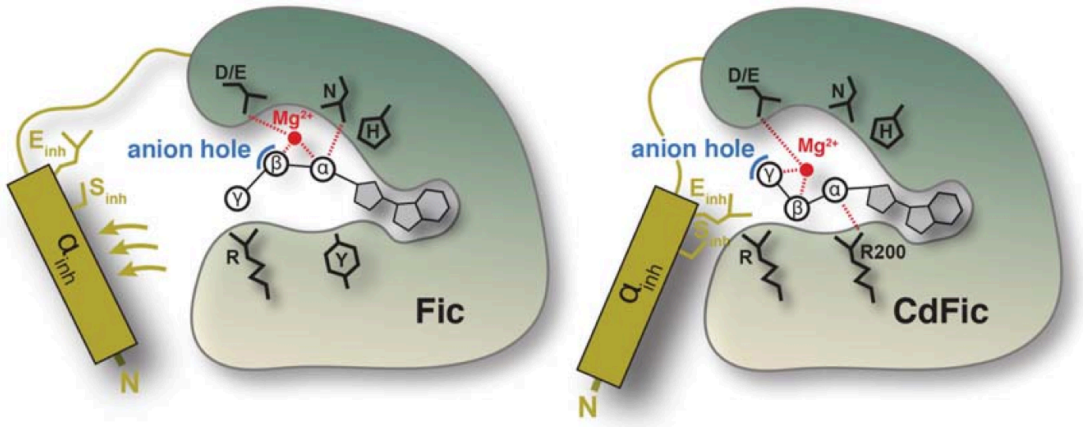


Figure 1.13. Non-obstructed ligand binding in CdFic. *A schematic model for inhibitory motif obstructed (left) and non-obstructed (right) auto-adenylylation mechanism. In former Fic proteins the inhibitory helix α-inh and/or the associated inhibitory residue, E_{inh}, obstructs a salt bridge between the terminal phosphate of ATP and the C-terminal arginine of the Fic motif. In latter, the ATP is pulled towards the opposite side of the active site cleft. Consequently, the position of E_{inh} is not obstructing proper positioning of the ATP phosphates and the auto-adenylylation activity is not inhibited. Adapted from (Dedic et al., 2016).*

1.11.3 Class III Fic proteins

Class III Fic proteins are a small group, exclusively with single domain proteins with inhibitory motif at the C-terminus. The only characterized example is *Nm*Fic protein from *Neisseria meningitides* (Engel et al., 2012; Stanger et al., 2016a). *Nm*Fic exists as a tetramer in solution as well as in crystal structure. Intriguingly, the tetramerization sequesters the monomer reservoir where they are prevented from AMPylating targets because the flap is buried in the center of the tetramer. Auto-modification in *Nm*Fic somehow helps in de-oligomerization, releasing the activated monomer to bind and AMPylate the targets (Figure 1.14). As the auto-AMPylating tyrosine is conserved in class III Fic proteins, auto-modification is considered as a conserved mechanism to activate the class III Fic protein. *Nm*Fic also AMPylates GyrB and ParE from *E. coli* and *N. meningitides*, thus inhibiting the growth.

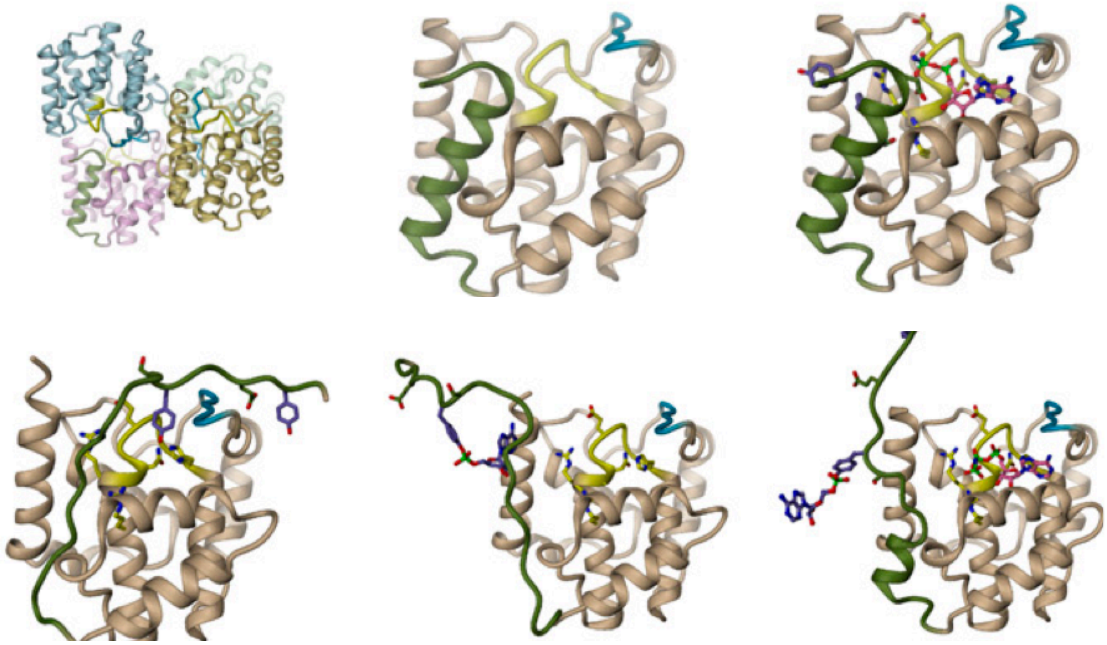


Figure 1.14. Model of cis-autoadenylylation. *NmFic is in monomer/tetramer equilibrium. Once a monomer (N) exits the tetramer, the α_{inh} can unfold and bind to the target binding site, positioning the modifiable Tyr^{183} at a specific position in the active site to allow the nucleophilic attack on the α-phosphorous of ATP resulting in auto-modification. Upon unfolding of the α_{inh}, the E_{inh} can no longer occupy its inhibitory position. Upon release of Y^{183}-AMP from the flap, the α_{inh} partially refolds. The latter conformation corresponds to the active monomeric NmFic protein. Adapted from (Stanger et al., 2016a)*

1.12 Growth inhibition by topoisomerase AMPylation

Expression of Fic toxins causes a major problem with DNA functioning, particularly chromosomal replication and segregation. In case of VbhT and YeFic proteins, which can modify both GyrB and ParE, the elongation of bacterial morphology has been attributed more to the AMPylation of ParE resulting in the collapse of replication fork and chromosomal segregation (Harms et al., 2015). AMPylation of GyrB elicited the SOS response in VbhT expressing *E. coli* cells, however, the extent of SOS induction did not correlate with the potency of growth inhibition and that SOS induction did not considerably contribute to it at relevant levels of VbhT expression. In case of *Pa*Fic toxin, the report suggests AMPylation of GyrB

target, however, it did not explore the effect on topoisomerase IV (Lu et al., 2016). The subsequent experiments also concluded that the cell filamentation induced by PaFic required the SOS pathway as in recA knockout strain i.e., the cells defective in the SOS pathway, inhibition of DNA replication by PaFic did not block the cell elongation (Figure 1.15). The model presumed in this context suggests an unidentified SOS-independent pathway that might contribute to cell morphology changes (Lu et al., 2016). Activation of the SOS pathway involves the formation of RecA filaments, the auto-cleavage of the SOS repressor LexA, and the induction of sulA, which binds FtsZ, thus blocking the completion of cell division and leading to the formation of filamentous cells. Expression levels and functional significance of sulA has been reported previously (Cox, 2007; Drlica and Zhao, 1997; Lu et al., 2016).

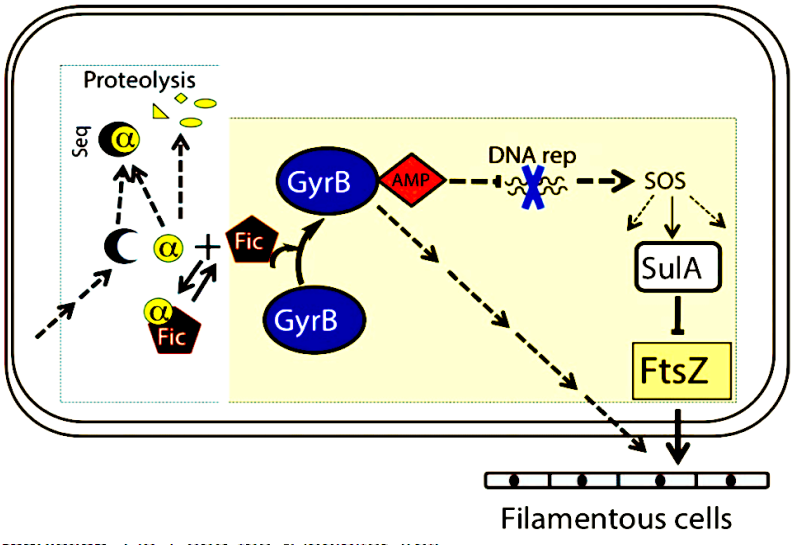


Figure 1.15. A model of Fic-1 mediated induction of bacterial cell filamentation and its regulation. *Fic-1 and the α_{inh} AntF form a dynamic complex under normal conditions. Signals from the environment activate a cascade that leads to the production of a sequestering protein (Seq) that competes for AntF or the activation of a protease that degrades the a-inhibitor. Freed or activated Fic-1 then inactivates GyrB by AMPylation, leading to the induction of SulA and the formation of filamentous cell. Alternatively, AMPylated GyrB may induce cell filamentation through an SOS response–independent pathway (dashed arrows). Adapted from (Lu et al., 2016).*

Fic toxin expression, in parallel to the cell filamentation, will also inhibit DNA replication which will most likely lead to slower metabolic rates, making bacteria resistant to some antibiotics and to certain detrimental environmental conditions. Likewise other TA modules and Phd-Doc module, which are also a factor, not the sole factor, for making cell persistent, Fic expression mediated growth arrest might also lead to bacterial persistence. Not only persistence, the regrowth of the persisters can also be replenished by addition of antitoxin domain that would deAMPylate Gyrase target and revert the inhibitory effect (Fisher et al., 2017; Kumar et al., 2017; Preissler et al., 2017).

1.13 Fic domains in archaea and eukaryotes

Fic proteins are most abundant in bacteria. Approximately 10% of the Fic domain sequences currently listed in the InterPro database (IPR003812) are encoded by archaea and eukaryotes (Engel et al., 2012). Phylogenetic analysis reveals that Fic sequences are of bacterial origin and have been acquired repeatedly and independently by archaea and eukaryotes (Khater and Mohanty, 2015).

In metazoans, a peculiar class II Fic protein is present, that also contains N-terminal trans-membrane helix followed by two tetratricopeptide repeats (TPRs), connected to the C-terminal Fic domain through a long α helical linker (Bunney et al., 2014). Human ortholog (HypE or FicD) displays a canonical Fic signature motif and canonical α-inhibitory motif. HypE is homodimer formed by Fic domains. TPR is required for membrane anchoring and providing stability and efficient catalysis to the protein. HypE ortholog localizes in ER lumen and AMPylates an ER chaperon, Bip/GRP78 protein (Preissler et al., 2015; Sanyal et al., 2015). BiP is the key component of unfolded protein response (UPR) where BiP ATPase activity is central to the processing of unfolded or misfolded proteins; therefore AMPylation of BiP has a profound effect on the regulation of UPR. The genetic and biochemical findings indicate that when the burden of unfolded ER protein is low, FicD AMPylates BiP (Ham et al., 2014). Conversely, as unfolded proteins accumulate, pre-existing AMPylated BiP is rapidly converted to the active de-AMPylated state. Reversibility of the modification is also maintained by FicD protein as its α-inhibitory motif de-

AMPylates BiP and contributes to the balance between client and active chaperons in ER (Figure 1.16) (Preissler et al., 2017). As BiP AMPylation responds directly to changes in unfolded protein load, without the need for gene expression or protein synthesis, the machinery for executing this switch may reveal something about the most proximal steps in protein folding homeostasis in the ER, which are yet to be determined.

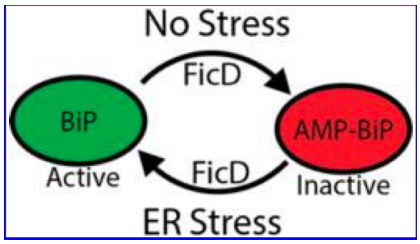


Figure 1.16. Model of AMPylation-mediated inactivation of BiP during ER homeostasis. *Adapted from (Casey and Orth, 2017).*

In *Drosophila melanogaster*, CG9532, a HypE ortholog, it was revealed that knockout flies were viable and fertile, but blind. Blindness was due to the requirement of enzymatically active CG9532 for the recycling of the neurotransmitter histamine involved in vision (Rahman et al., 2012). Various studies have contributed to the growing list of AMPylation targets of HypE, including Rho GTPases, histones, nuclear envelope proteins, regulators of gene expression and many more (Broncel et al., 2016; Truttmann et al., 2016; Truttmann et al., 2015; Yu and LaBaer, 2015), whose biological relevance needs further exploration.

1.14 *Mycobacterium tuberculosis* Fic protein, a model to understand role of AMPylation in pathogenesis

1.14.1 *Mycobacterium tuberculosis* and infection

Mycobacterium tuberculosis is a slow-growing obligate human pathogen with no natural reservoir outside humans, resides within its primary target, the macrophage cells, upon internalization and is excellent in various strategies to escape host immune surveillance. The *Mycobacterium* cell wall consists of complex layers of arabinogalactan, peptidoglycan and covalently linked long branched mycolic acids

(Figure 1.17). A high proportion of its cell wall components is lipid; tangled in a way to make the mycobacterium cell wall less porous than the outer membrane of Gram-negative bacteria. This low density of pore might cause more difficulty in absorption of nutrients and could contribute to slow growth of mycobacteria. Higher GC content of promoters, differential orientation of the gene in the relation to the direction of replication, low RNA/DNA ratio in growing mycobacteria and presence of a single ribosomal RNA operon present apart from the OriC, might also contribute to slow growth (Kumar et al., 2017).

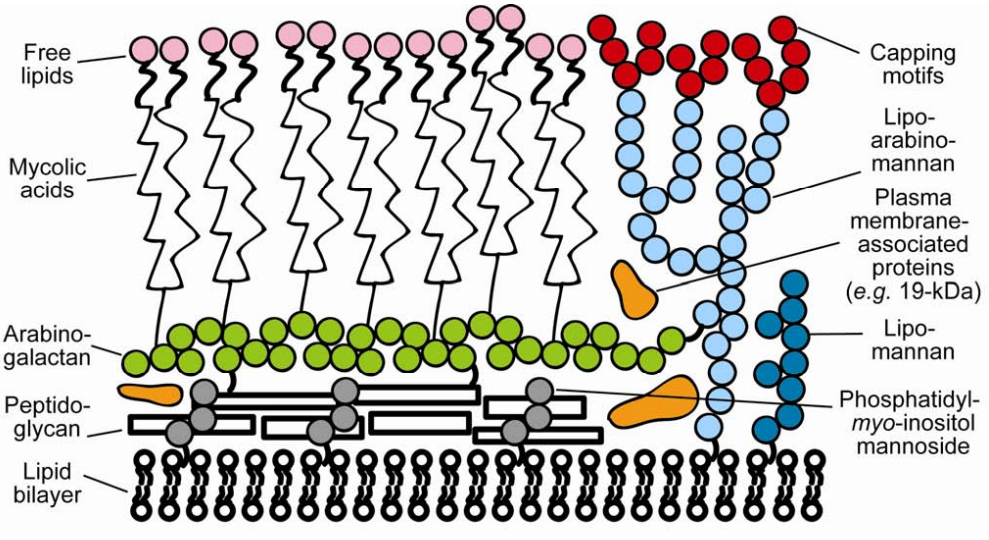


Figure 1.17. Schematic representation of the complex Mtb cell wall. *Arabinogalactan is attached to the peptidoglycan. Mycolic acids and glycolipids extend through the cell wall. Sourced from Welin A. et al, 2011.*

1.14.2 Course of infection

The infectious agent is aerosolized and exhaled by an active TB patient. These droplets remain airborne owing to their small size and are therefore inhaled by unwitting new human hosts in whom the pathogen primarily invades lungs in the initial invasion. The bacteria encounter the host cell alveolar macrophages, which are the first line of defense of the host's innate immune response. Ironically, macrophages act as both the primary site of host cellular defense as well as the primary site of bacterial

replication and dissemination. These cells recognize the pathogen and engulf them into vacuoles called phagosome (Figure 1.18).

Mycobacteria being intracellular pathogens are actively phagocytosed by these cells via several receptors expressed on the macrophage cell surface such as the complement receptors (CRs), the mannose receptor, the immunoglobulin fragment carrying the constant region of the heavy chain (Fc) receptor and the scavenger receptors, which actively recognize the pathogens (Ernst, 1998). Amongst these, the most critical receptor for is CR3 which when blocked, either by antibodies or the obstruction of its lectin site, drastically reduces mycobacterial phagocytosis (Cywes et al., 1997). This CR-3 mediated mycobacterial uptake requires the accumulation of cholesterol at the site of entry, the role of the latter being specific only for *M. tuberculosis* uptake by macrophages (Gatfield and Pieters, 2000). This cholesterol at the site of uptake is postulated to modulate the membrane cytoskeleton as well as increasing the viscosity of the membrane that is in contact with the hydrophobic mycobacterial cell wall, accelerating phagocytic uptake (Nguyen and Pieters, 2005).

In a regular infection, the internalized load of vacuoles is then delivered to lysosomes, either by budding and fusion or through the acquisition of lysosomal markers, as part of phagosome maturation. However, the mycobacterial phagosomes resist phago-lysosomal fusion, a process mediated by several molecules expressed by the pathogen. The pathogen also cleverly recruits several host molecules at this site for its devious purposes of averting the phago-lysosomal fusion, the chief molecule identified in the prevention of phagosomal maturation upon tubercular infection being the host protein TACO (tryptophan, aspartate-repeat domain containing coat protein) or Coronin I which associates exclusively with phagosomes containing live mycobacteria (Ferrari et al, 1999; Schller et al, 2001). The Coronin protein family has been shown to function in actin cytoskeleton remodeling. The mycobacteria also release several factors into the Phagosomal lumen and cytosol of the host cells that interfere/block the phagosome-lysosome transfer, a prominent protein among these being the protein kinase G (PknG).

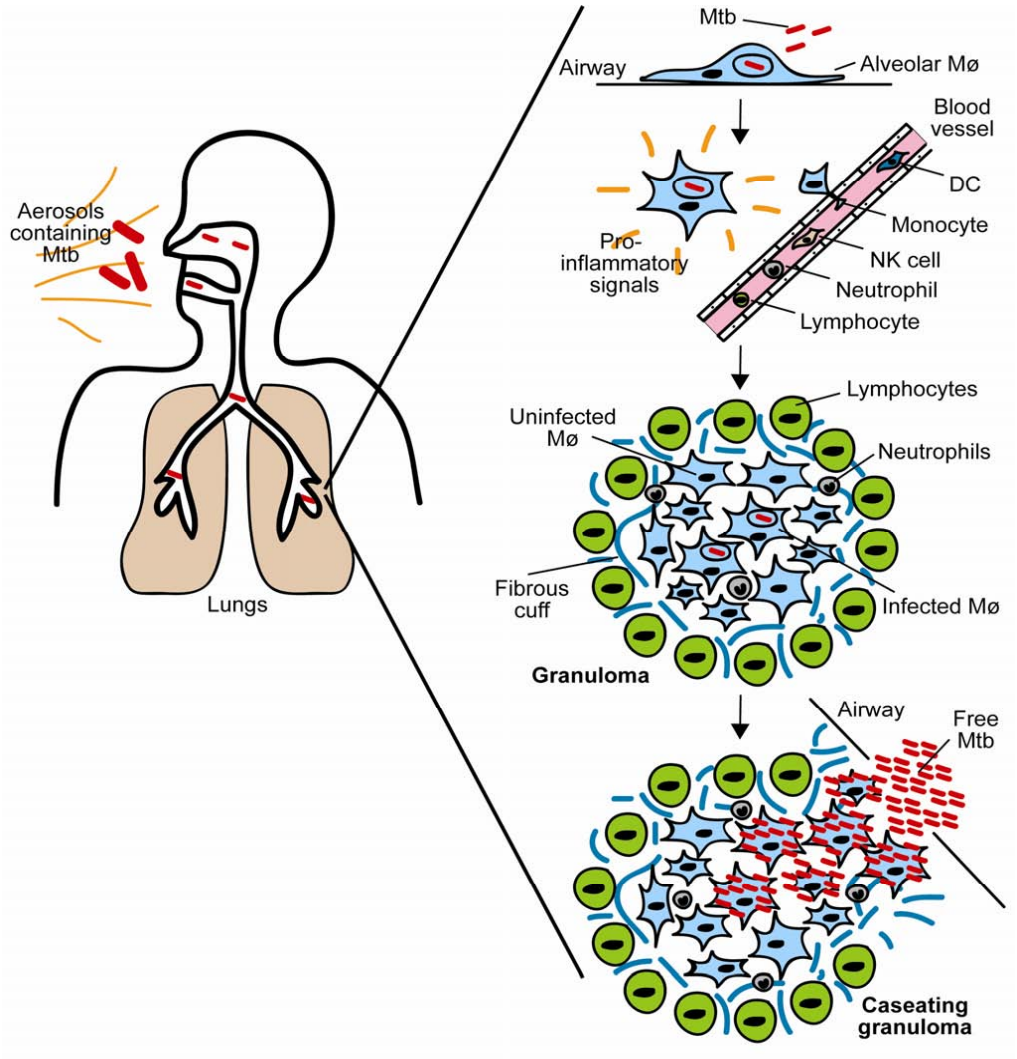


Figure 1.18. The infection route of Mtb in a human host leading to the formation of a granuloma*. Resident alveolar macrophages (Mø) phagocytose inhaled bacteria. This leads to a pro-inflammatory response and recruitment of cells of the innate and adaptive immune systems, and the formation of a granuloma. The bacilli can be contained within the structure for long periods of time, but if immune control fails, the bacilli will commence replication, and a necrotic granuloma core develops. The granuloma then ruptures and Mtb are spilled into the airways. Sourced from (Welin A. et al, 2011).*

The alveolar macrophages then invade the subtending epithelial layer of the lung, including a local inflammatory response at this site, which recruits mononuclear

cells from the neighbouring blood vessels thereby providing the burgeoning pathogen population newer cells to invade and colonize. Gradually other macrophage and host phagocytic cells accumulate at this site of infection forming a granuloma, which is pathologically defined as an organized collection of differentiated macrophages with a characteristic morphology (Adams, 1976). A granuloma begins as an amorphous cluster of a few mononuclear phagocytes surrounding individual infected macrophages. Gradually, this collection of macrophages is surrounded by T lymphocytes, some B lymphocytes, dendritic cells, neutrophils, fibroblasts and extracellular matrix components which help restrict the bacteria (Flynn et el, 2001, Peters et al, 2003). Further, the macrophages differentiate into epitheloid cells that have tightly inter-digitated cell membranes in zipper-like arrays linking adjacent cells (Cosma et al, 2003). Several macrophages may fuse to form giant cells or come to contain lipid droplets forming foamy macrophages. Over time the granulomas evolve morphologically to develop a central core of necrotic acellular debris called a "caseum", which is surrounded by epithelial cells with peripheral lymphocytes forming the outermost boundary (Figure 1.18). These granulomatous lesions may gradually become calcified or deposited with fibrin. Thus, the granuloma assumes a more organized and stratified structure over time. Mycobacteria usually reside inside the macrophages and in the central caseous region. Therefore these differentiated structures form immunologic and physical barriers that contain infection, preventing bacteria dissemination. As the fibrous sheath around the granuloma gets fortified, the number of blood vessels that penetrate the structure decreases, making the caseum hypoxic. This hypoxia induces a state of non-replicative persistence in the pathogen. In this persistent state, the organism employs unique strategies such as the restriction of biosynthetic activity to conserve energy, induction of alternate energy pathways and stabilization of essential cellular components to minimize repair or replacement (Wayne and Sohaskey, 2001). Further, in the non-replicative persistent state, the bacteria are completely intractable to drug treatment. Active patients of the disease demonstrate granulomas at various stages of development and the fate of each of these is determined locally and not systematically (Barry et al., 2009). These granulomas

may then rupture to release the infectious bacilli into the lung, from where these get released into the atmosphere upon coughing etc.

Therefore, the intricate dynamics of the granulomas in terms of their differentiated, complex organization and variability in the stages of growth within the same infected individual complicate the treatment of disease. Also, various mycobacterial populations (replicating and non-replicating) possibly occupy niche microenvironments within each granuloma, further complicating the effectiveness of drugs. Thus the unique pathology of the disease posed by the intractable mechanisms employed by the causative organism to thwart both the immune response and chemotherapeutic agents poses challenges to treatment.

1.14.3 WHY FIC PROTEINS From *M. tuberculosis*?

Numerous studies have highlighted the occurrence of numerous types of PTMs in *Mtb*. The well-known PTMs in *Mtb* are glycosylation, lipidation, or phosphorylation, known regulators of protein function or compartmentalization. Other PTMs include methylation, acetylation, and pupylation, involved in protein stability. The PTM that specifically mimics the host cell signaling are phosphorylation and dephosphorylation mediated by kinases and phosphatases, respectively. There are several ways in which *Mtb* manipulates the macrophages to avoid killing and create a favourable environment for replication, and the inhibition of phagosomal maturation is the best-characterized mechanism so far. A pair of phosphatase SapM and kinase PknG assists *Mtb* in **Phagosomal maturation arrest.**

LAM, the abundant mycobacterial cell wall glycolipid is shed by the bacterium upon entry into the cell and can thereafter be found throughout macrophage membranes. ManLAM can inhibit the recruitment of EEA1, which is necessary for subsequent fusion and fission events leading to delivery of lysosomal hydrolases and vacuolar H^+-ATPases to the phagosome. The inhibition is achieved through prevention of a Ca^{2+} surge in the cytoplasm, which is necessary for calmodulin- and Ca^{2+}/calmodulin-dependent kinase II (CaMKII)-dependent delivery of EEA1. Inhibition of the Ca^{2+} surge is preceded by an inhibition of a macrophage sphingosine

kinase (Figure 1.19). The **second mechanism** that leads to inhibition of EEA1 recruitment is a blockade of PI3P, and this is thought to occur through ManLAM-mediated inhibition of the PI3K hVPS34 as well as through SapM-mediated dephosphorylation of PI3P. The **third mechanism** of phagosomal maturation inhibition by *Mtb* in macrophages is the phosphorylation of an unknown host substrate by the serine/threonine kinase PknG, regulating phagosomal maturation, and another is activation of p38MAPK by *Mtb*, leading to reduced Rab5 activity and thus reduced EEA1 on phagosomes. A **final mechanism** of Mtb inhibition of phagosomal maturation is through retention of a molecule called coronin 1, also involved in Ca^{2+} signaling.

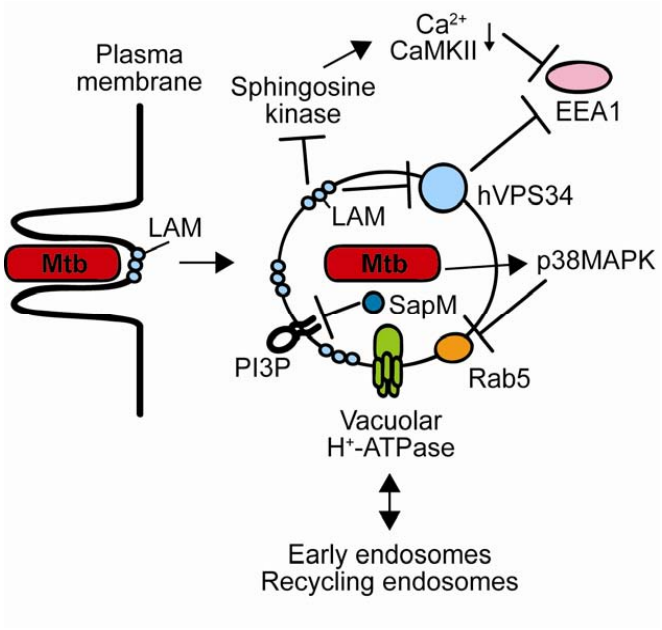


Figure 1.19. The Mtb phagosome and mechanisms of phagosomal maturation inhibition in macrophages*. LAM-mediated inhibition of the Ca2+ surge through sphingosine kinase inhibition, and of the PI3K hVPS34, as well as SapM-mediated dephosphorylation of PI3P and activation of p38MAPK by an unknown component leads to deficient recruitment of EEA1 and function of Rab5, resulting in a lack of interaction with late endosomes and lysosomes, as well as a lack of vacuolar H+-ATPase accumulation. This arrests the vacuole at the early phagosome stage, with a relatively neutral pH. Sourced from Welin A. et al, 2011.*

1.15 Objectives of the thesis

Tuberculosis, the oldest known infectious disease to man, continuous to claim thousands of lives each year, mortality from this disease alone surpassing the deaths caused by HIV and malaria put together. The main contributor to this morbid statistics is the manipulation machinery of the pathogen *M. tuberculosis* and its developed resistance towards drugs. *Mtb* manipulates the macrophages to avoid killing and create a favorable environment to thrive inside the host.

AMPylation has emerged as novel stratagem in pathogenic bacteria to perturb host cell signaling in a balanced reversible manner, thus providing another PTM to bacterial effectors, similar as phosphorylation, to mimic host like signaling. As the kinases are major regulators of signaling cascade in eukaryotic host, the accessible phosphorylation sites can be completely masked by bulky AMP upon AMPylation, thus proving this strategy as more robust and effective than the regulation through phosphorylation. Also, the substrates for bacterial AMPylators, identified from mammalian hosts till date, are the master players in cytoskeletal dynamics (Rho, Rac, Cdc42 GTPases), vesicular trafficking (Rab GTPases) and gene regulation. While Phagosomal maturation arrest in *Mtb* is well studied, the exact mechanism with late endosome fusion debility is still not clearly understood.

The primary objective of the thesis is to characterize Fic protein (Rv3641c) from *Mycobacterium tuberculosis*. Fic proteins modify their targets post-translationally, particularly AMPylation in *Mt*Fic, and block the downstream signaling. Thus the impact of AMPylation is solely attributed to the functional significance of its target(s). This thesis focuses on the target identification of Fic protein, in bacteria as well as from human host cells. The orchestration of cellular localization of Fic domain-containing proteins from *Mycobacterium sp.* and their substrate identification has been established. The second objective is to understand cofactor selectivity. Moreover, the functional characterization established Fic proteins to be part of TA module, therefore, understanding the role of antitoxin domain in inhibition as well as activation of the activity, has been endeavored. As the characterization continues, the work also describes the role of antitoxin domain in reversing the PTM, thus describing a new

mode of action of antitoxin domain, is also a new facet in this thesis. Additionally, the characterization of Fic proteins from *Mycobacterium sp*. described here, paves the way for studies on the relevance of this new TA module in pathogenesis and bacterial dormancy.

CHAPTER 2

Sequence based analysis of Fic proteins from Mycobacterium sp.

Table of contents

2.1 INTRODUCTION

As mentioned in Chapter 1, Fic proteins originated from bacteria and have evolved extensively through horizontal gene transfer. They are ubiquitous in nature covering archea and metazoans (Khater and Mohanty, 2015). Remarkably, class II Fic proteins where antitoxin domain is a part of toxin protein, present at its N termini constitute a versatile class of Fic, which dominates the phylogeny (Figure 2.1), while class I and class III Fic proteins form small, likely monophyletic, groups (Harms et al., 2016). 90% Fic domain-containing proteins which are part of class II members, display enormous sequence variability providing the protein multi-domain functionality, thus enhancing the coverage of Fic proteins from AMPylating enzymes to multi-functional enzymes (Engel et al., 2012; Harms et al., 2016).

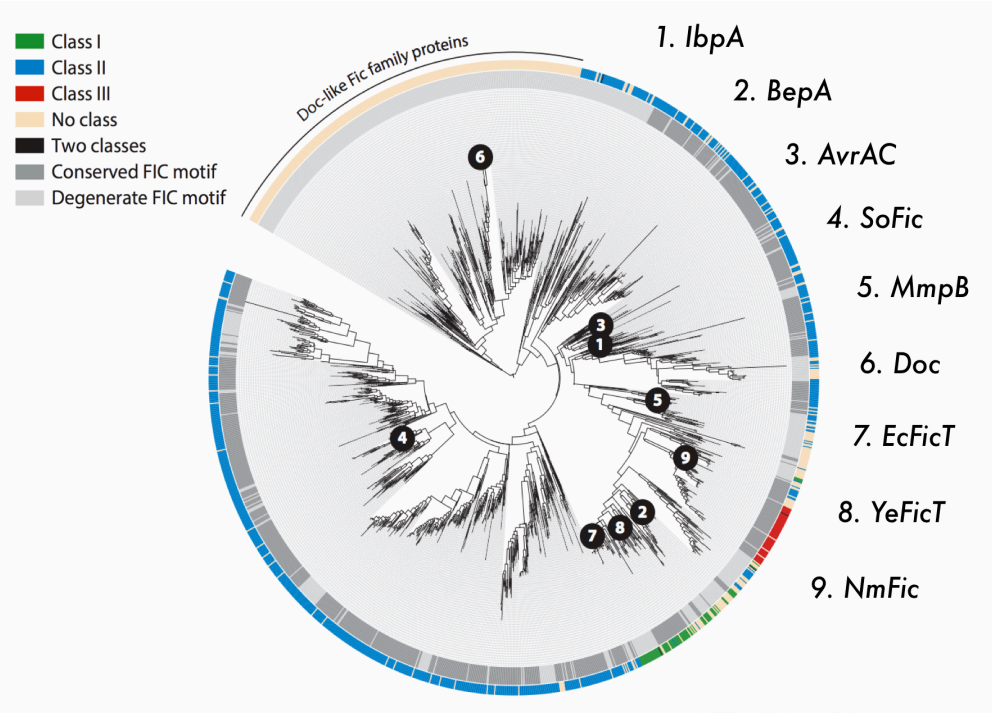


Figure 2.1. Phylogeny and functional diversity of Fic proteins. *Positions of various characterized Fic proteins in the phylogeny. Adapted from (Harms et al., 2016). IbpA of Histophilus somni, BepA of Bartonella henselae, AvrAC of Xanthomonas campestris, SoFic of Shewanella oneidensis, MmpB of Geobacter lovleyi, Doc of Salmonella typhimurium, EcFicT of Escherichia coli, YeFicT of Yersinia enterocolitica and NmFic from Neisseria meningitides.*

The major goal of these studies is to find out the AMPylators in *Mtb*. A sequence comparison of *Mycobacterium* Fic proteins revealed them to be a subset of class I and class II Fic proteins. This chapter presents studies of the analyses of Fic proteins in *Mtb* as well as across 20 other *Mycobacterium sp*, based on their amino acid sequence variability and identifies Fic, group of mobile mystery proteins B, Doc and several other non-canonical Fic motif containing mycobacterial proteins. This chapter also presents a new method to predict the putative substrates of Fic proteins by inferring their intracellular localization based on linear motif identities in them.

2.2 MATERIALS AND METHODS

2.2.1 Cloning Smeg_2181 for mammalian expression and Smeg_2140 for bacterial expression

The codon-optimized sequence of *M. smegmatis* Fic protein (Smeg_2181) for mammalian expression was obtained from Gene-Art. Cloning was performed following standard procedure (Mamiatis et al., 1985). The Smeg_2181 gene was amplified with gene-specific primers (Table 2.1) using Polymerase Chain Reaction (PCR) with the codon-optimized sequences as the template. The amplicons were digested with restriction enzymes followed by ligation with the double digested pEGFPC1 vector to obtain eGFP tagged constructs. Cloning of SMEG_2140 (*Sm*Fic) was performed with gene-specific primers (Table 2.2) in pET 28a(+) vector, following the standard cloning procedure.

For site directed mutagenesis (SDM), the pEGFPC1-2181 construct was used as template and single primer based method was followed (Edelheit et al., 2009). Linear PCR amplification with gene-specific primer (Table 2.1) having a silent mutation at one nucleotide position was followed by *Dpn*I digestion to cleave the methylated parental strands. After *Dpn*I digestion, the PCR products were transformed in *E. coli* DH5α strain. Positive colonies were screened and confirmed by sequencing.

Table 2.1. List of primers used for SMEG_2181 cloning

Primers	Sequences (5'-3')	Restriction sites
pEGFPC1_2181-FP	GCCGTT**AAGCTT**ATGGTAGCCGAACTCCGCGTCGACG	***HindIII***
pEGFPC1_2181-RP	CGTACA**CCGCGG**TCTCACCGCTCCGGATCCCTGAT	***SacII***
FP_NLS1_2181 ***RP_NLS1_2181***	GCGAAA**GAGCTC**AAGCCCATCTCCCCCGCC TTGTTGCG**CCCGGG**TTACGTACCGGGGATGAAGTACG	***SacI*** ***XbaI***
FP_NLS2_2181	ATCTCC**GAGCTC**AACGCGCGGGCCGTCG	***SacI***
RP_NLS2_2181	GTTAGG**CCCGGG**TTACGCGATGCGGCGCCG	***XbaI***
2181 H195A primer	GCCGAGTACAACGCAATCCACCCCTTCCGCGAGGGCAAC	

Table 2.2. List of primers used for SMEG_2140 cloning

Primers	Sequences (5'-3')	Restriction sites
pET28a_2140-FP	GCTATCTT**CATATG**GTGGCATCAGCCGAGGAGGGCACGGC	***NdeI***
pET28a_2140-RP	GAACAGT**GGATCC**TTAACCGGAACGGCTGCGTCGCCCG	***BamHI***

2.2.2 Mammalian cells, their maintenance and transient transfection

The mammalian cells used in this study include Human embryonic kidney cells 293 (HEK293) and human macrophages THP1 cells. Cells were cultured in their prescribed growth condition in Dulbecco's modified Eagle's medium (Invitrogen) containing 10% fetal bovine serum and 1% antibiotic-antimycotic (Sigma). Transfection of the cells with pEGFPC1-2181 constructs were carried out using Lipofectamine 2000 (Invitrogen) at $1X10^5$ cells/60 mm plates, as prescribed in the manufacturer's protocol.

2.2.3 Protein purification, labeling and treatment

The recombinant plasmid for SmFic expression was transformed into *E. coli* BL21 (DE3) cells (Novagen) and the cultures were grown in Luria-Bertani medium

supplemented with 30 μg mL^{-1} kanamycin at 37 °C until A_{600} reached ~0.7. Cultures were then induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 18 °C for another 12 hours. The cells were harvested by centrifugation (Sorvall RC-6000 rotor) for 10 min at 4 °C. The recombinant protein expression in soluble fractions was confirmed by western blotting of the purified protein using anti-His antibody. For protein purification, the pellet was resuspended in lysis buffer (20 mM Hepes, 300 mM NaCl, 10% glycerol, pH 7.4) and the cells were lysed by sonication (Vibra-Cell, Sonics and Materials, USA). Cell debris was removed by centrifugation at 14,000 rpm for 45 min. the supernatant obtained was applied to a Ni-NTA metal affinity column (His-bind resin, Novagen), equilibrated with the lysis buffer. The column was washed with 100 mL lysis buffer. Post-column washing, the protein was eluted using lysis buffer containing gradient of 50 mM-300 mM imidazole. Purity of the protein was checked on SDS-PAGE. The purified fractions were pooled and loaded onto a S-100 16/60 FPLC gel filtration column (GE healthcare) and the homogenous fractions were pooled, concentrated using centriprep-10 (Amicon) and stored in -80 °C for further use. Protein concentration was determined from A_{280}, assuming the theoretical molar extinction coefficient ε_{280} = 37930 M^{-1} cm^{-1}, which was determined by ProtParam (Gasteiger E. *et al*, 2005).

For FITC labeling of proteins, we followed an established protocol provided by FITC manufacturer (Sigma). Briefly, 10 mg/ml FITC was dissolved in sodium bicarbonate buffer (pH 9.5). 1mg *Sm*Fic recombinant protein and BSA were added to 4 mg/ml of FITC solution and allowed to bind for 6 hours at 4 °C with continuous agitation. Labeled-protein was desalted using HiTrap Desalting column-5ml (GE healthcare), followed by overnight dialysis against PBS at 4 °C.

2.2.4 Cell Immunofluorescence

HEK293 cells, grown on sterile coverslips, were transfected with pEGFPC1 native vector, pEGFPC1-2181 constructs; 2181^{H195A}, NLS1-2181 and NLS2-2181. Post-16 hours of transfection, HEK293 cells were washed twice with ice cold PBS, fixed with methanol and washed again twice with PBS. The cells were stained with

DAPI solution (5 µg/mL) (Sigma) for 5 min, washed thrice with PBS and coverslips were mounted on slides with Fluoroshield (Sigma). Imaging was performed on Leica laser-scanning confocal microscope (Germany) using 63x objective.

For internalization experiment, human macrophage THP1 cells were treated with labeled protein *Sm*Fic-FITC. In brief, after 24 hours of recovery, phorbol 12-myristate 13-acetate (PMA) differentiated THP1 cells were incubated with 20 µg of labeled protein for 30 min. All the steps were performed in the dark. Cells were fixed and examined in a confocal microscope.

2.2.5 Sequence Retrieval

Protein sequences of the genus *Mycobacterium* were retrieved from the NCBI database. 3807 protein sequences, which contained the Fic/Doc motif pattern, were shortlisted with the search pattern "HxFxxxNxR" and were further subjected to redundancy test. Percentage identity and sequence length were used as the criteria and sequences with 100% sequence identity and identical lengths were removed. Subsequently clustering of sequences having 100% identity irrespective of the lengths, were also filtered out. The final non-redundant dataset had 554 protein sequences with Fic motif pattern from *Mycobacterium* across different species. Non-canonical Fic/Doc proteins, Doc proteins and canonical Fic proteins were then separated based on pattern of their Fic motif. Our dataset contained 81 with Fic motif [HxFx(D/E)GNGR] and 84 with the Doc motif [HxFx(D/N)(A/G)NKR] and the remaining 389 sequences were classified under the non-canonical Fic protein category.

2.2.6 Alignment and Phylogenetic Analysis

Phylogenetic and molecular evolutionary analyses were conducted using MEGA v6 (Tamura et al., 2013). Molecular Phylogenetic analysis and evolutionary history was inferred using the Maximum Likelihood method, which works on the JTT matrix-based model (Jones et al., 1992). The bootstrap consensus tree was then constructed with f100 replicates to represent each of the taxa to be analyzed. Further,

replicate branches, which correspond to partitions less than 50% were collapsed and the initial tree for the heuristic search was obtained using Neighbor-Join and BioNJ algorithms to generate a matrix of pairwise distances. A final topology was selected which had superior log likelihood value. The analysis involved 81 protein sequences with canonical Fic motif. All positions with less than 95% site coverage were eliminated which included alignment gaps, missing data. There were a total of 27 positions in the final dataset.

2.3 RESULTS

2.3.1 Functional nuclear localization signals (NLS) in Smeg_2181 and Smeg_2140

M. smegmatis has been a well-studied system as the model organism to study the metabolic pathways of *Mtb*, as this organism is non-pathogenic in usual laboratory practices. *M. smegmatis* was first isolated from syphilitic chancres and its infection has been reported in a case study as fatal in a child with inherited interferon gamma receptor deficiency (Pierre-Audigier et al., 1997). Most of the other reported studies were skin or soft tissue infections (Newton Jr et al., 1993; Wallace Jr et al., 1988). There are two Fic domain-containing proteins in *M. smegmatis*. The Smeg_2181 is the closest homolog with 35% identity and 63% sequence coverage with *Mtb* putative Fic protein, Rv3641c (*Mt*Fic). However, the theoretical pI of *Mt*Fic is 4.7 and of Smeg_2181 is 10.96. The difference in pI is attributed to arginine-rich motif present at N- and C-termini of Smeg_2181. The *de novo* tools offered a high probability score (3.9-4.6) for arginine-rich motif to be involved in bipartite nuclear localization signals and nuclear export signals (Figure 2.2).

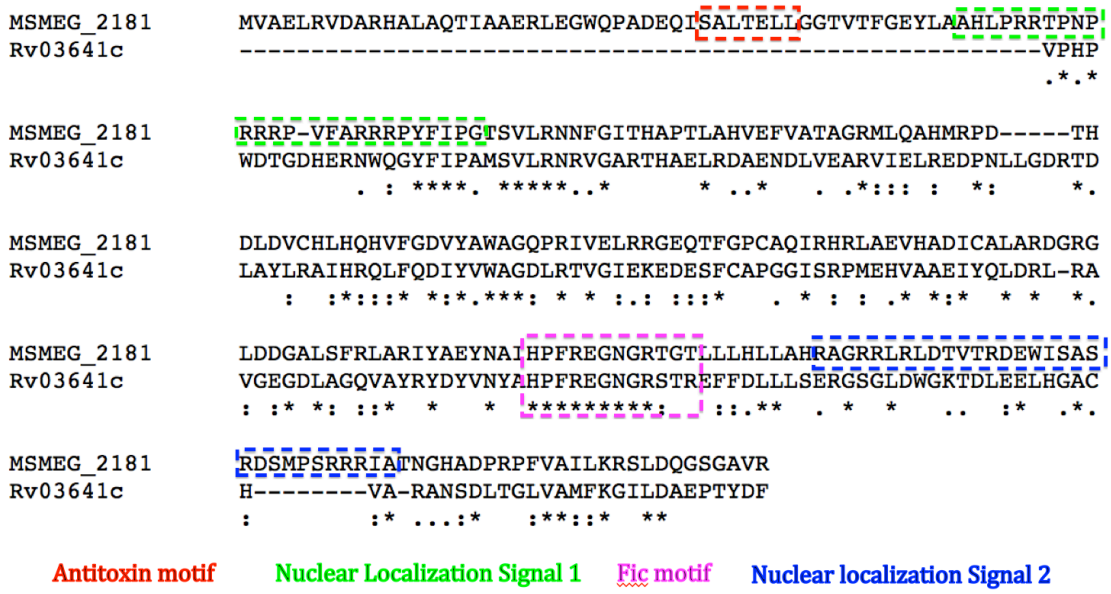


Figure 2.2. Sequence alignment of Rv3641c and Smeg_2181 showing conserved motifs. *NLS predictions were based on NLS mapper and NucPred servers.*

To characterize the nuclear localization signals, Smeg_2181^{H195A} mutant, predicted NLS1 and NLS2 were cloned in mammalian expression vector pEGFPC1, which would provide GFP at N-termini of the expressed proteins. For nuclear localization studies, HEK293 cells were transfected with the constructs. Transfection with Smeg_2181^{H195A} mutant showed GFP signal overlapping mostly with DAPI. NLS1 construct showed exact overlap and NLS2 construct showed localization both in the cytoplasm as well as in the nucleus (Figure 2.3). Nuclear localization of Smeg_2181 confirms predicted NLS to be functional.

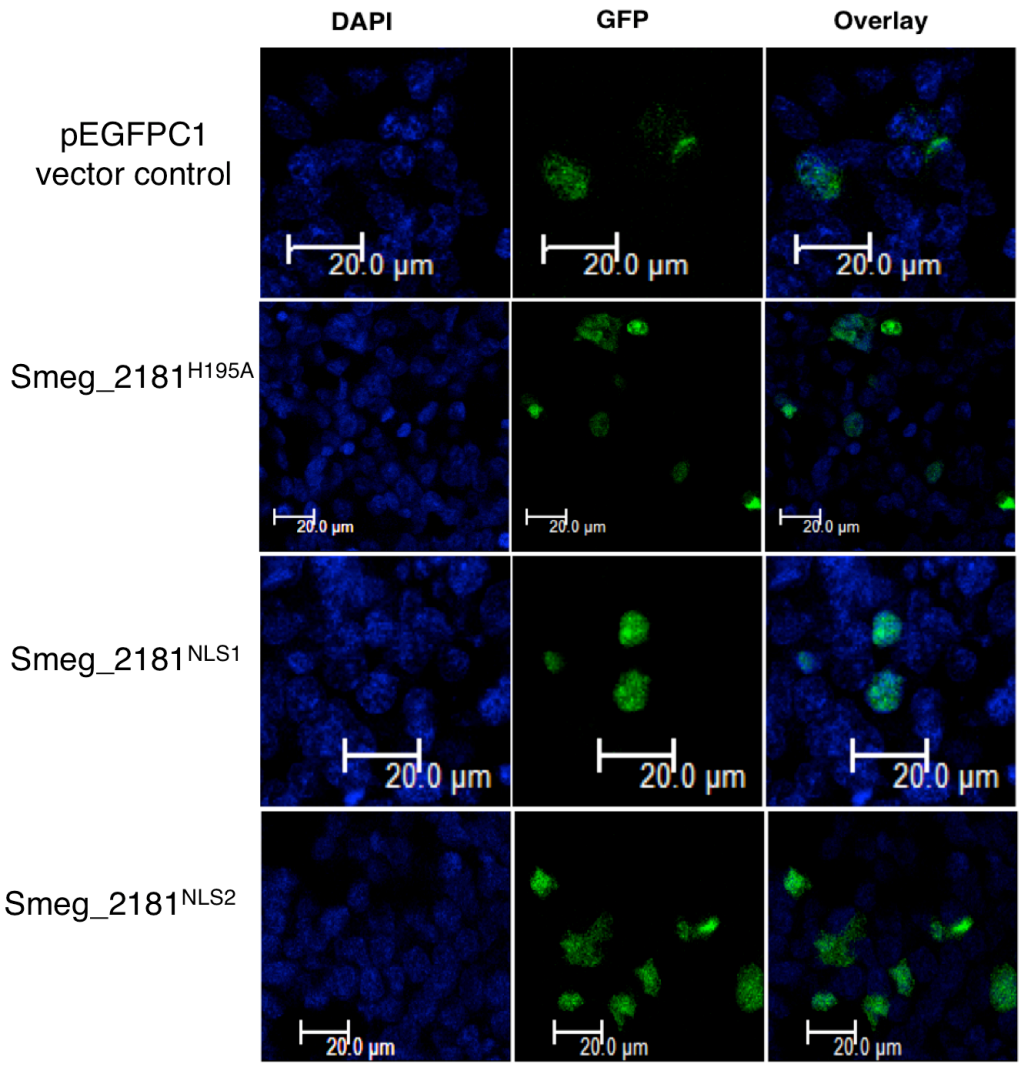


Figure 2.3. Confocal images showing co-localization of Smeg_2181 with DAPI.

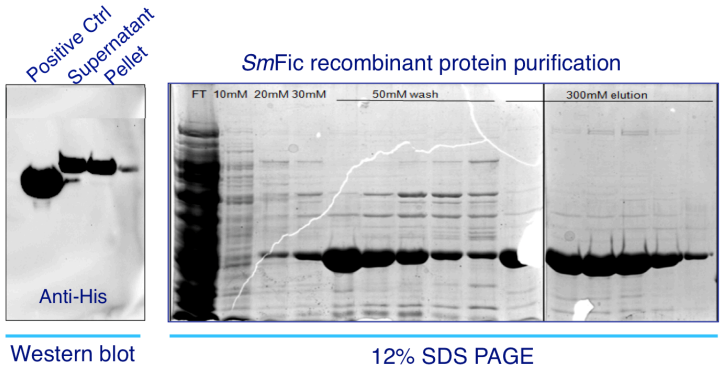


Figure 2.4. Purification of recombinant SmFic protein. *Left panel showing western blot with anti-His antibody for detection of expression in pellet and soluble*

fraction. A 42 kDa His-tagged protein was used as positive control for anti-His antibody detection. Right panel is showing 12% SDS PAGE for purified fractions.

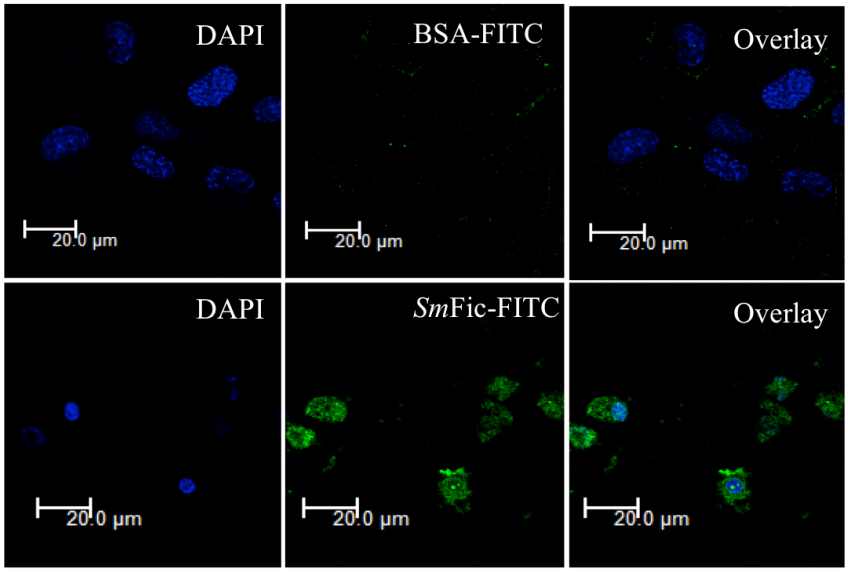


Figure 2.5. Confocal images of THP1 cells showing co-localization of FITC labeled SmFic with DAPI stained nucleus.

To validate the nuclear localization prediction of second *M. smegmatis* Fic domain-containing protein (Smeg_2140; *Sm*Fic) was purified (Figure 2.4) and fluorescently labeled with FITC. PMA differentiated THP1 cells were treated with varying concentration of labeled protein for 30 mins. Labeled BSA, used as negative control, remained in periphery of the cells in the background and labeled *Sm*Fic protein showed co-localization with DAPI (Figure 2.5). Nuclear localization validates the functional NLS present in the *Sm*Fic protein sequence. Also, internalization and localization of externally added protein signifies that the *Sm*Fic escaped from the host endocytic degradation machinery and localized into the nucleus.

2.3.2 Retrieval of proteins with Fic motif pattern from *Mycobacterium sp.*

Functional nuclear localization signals in *M. smegmatis* proteins paved the path for analysis towards the identification of Fic proteins across *Mycobacterium* including pathogenic *Mtb* strains. To this end, we curated Fic proteins from 21

Mycobacterium species with a search for Fic motif pattern "HxFxxxNxR". We obtained 3807 proteins with Fic motif pattern. After removal of redundant sequences, a dataset with 554 mycobacterial proteins was prepared (Figure 2.6). Interestingly, only 81 protein sequences revealed proteins with canonical Fic motif "HxFx(D/E)GNGR". Canonical Fic motif has been defined based on the sequence conservation shown by the Fic domain-containing enzymes, which can perform AMPylation of targets. A second subset included 84 mycobacterial proteins where the "GNGR" from the canonical motif is replaced by "GNKR", constructing a group of Phd-Doc enzyme that can perform phosphorylation instead of AMPylation. Therefore, 84 unique protein sequences with "HxFxxxNKR" motif were identified as Doc proteins from 21 *Mycobacterium sp*. Rest 389 mycobacterial proteins showed remarkable variation in amino acid sequences in the Fic motif. This subset was annotated as a group of non-canonical Fic motif "HxFxxxNxR" containing proteins from *Mycobacterium*. As the conserved "GNGR" is replaced with "xNxR", these various mutations provide space towards phosphate binding side of nucleotide and thus may provide selectivity for various nucleotide derivatives other than nucleotide triphosphates. Consequently, these identified proteins could exhibit unique in PTMs.

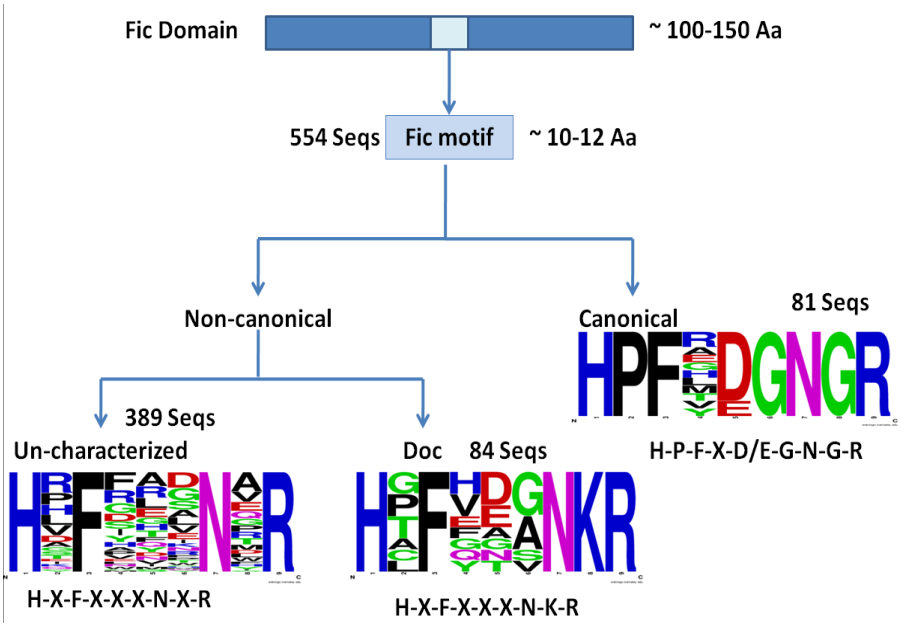


Figure 2.6. Schematic showing canonical and non-canonical Fic motif distribution across Mycobacterium. *Fic domain proteins with a minimum*

sequence length of 100-150 amino acids were retrieved with "HxFxxxNxR" motif pattern. The subgroup with 81 sequences are the AMPylating enzymes, with 84 sequences are phosphorylating enzymes and subgroup with 389 protein sequences remains unknown with respect to PTM.

2.3.3 Analysis of canonical Fic motif containing proteins from *Mycobacterium sp.*

We further analyzed the 81 protein sequences containing the canonical FIC motif. Protein similarity was analyzed by Blast tool. The percent sequence identity ranges from 11.8% to 99%. Although the phylogenetic distribution followed the core gene distribution pattern of *Mycobacterium* classification, the variability clustered the orthologous proteins of *Mycobacterium* pathogenic species such as *M. tuberculosis, M. canetti, M. kansai and M. abscessus,* together in one clade (Figure 2.7; shown in purple and blue colored clades). Variability was observed in non-pathogenic species; however, they are also clustered on similar lines being dependent on their sequence length (Figure 2.7; shown in black colored clades). When compared with the allelic distribution, the pathogenic *Mycobacterium sp*. including *Mtb strains* belong to class I Fic proteins having putative antitoxin module upstream to the toxin module. Both are present in a single operon. All the other non-pathogenic *Mycobacterium sp*. are populated with class II Fic proteins having antitoxin motif at N-terminus to the toxin motif, as a part of the same protein. The high conservation of shorter core Fic domain in pathogenic mycobacterium could be a sign of reductive evolution strategy adopted by pathogenic *Mycobacterium sp*.

2.3.4 *De novo* identification of motifs in the mycobacterial canonical Fic proteins

The phylogenetic distribution clustered the canonical Fic proteins in separate nodes, which followed the distribution of core gene in *Mycobacterium* phylogeny. However, there were outliers, which were not part of the core *Mycobacterium* gene. For example, *M. smegmatis* strains possess two canonical Fic domain proteins

wherein a closer analysis revealed that the Smeg_2181 is a part of the core gene and Smeg_2140 to be a part of the genomic island (island viewer 3). Further, we observed the variability present at their N- and C-termini. The information was obtained in four steps: I, defining the core; II, sub-grouping N- and C-termini extensions; III, prediction of mammalian localization and search for linear motifs; IV, rationalizing the functional predictions based on the overlap of predicted localization and substrate interacting sites (Figure 2.8).

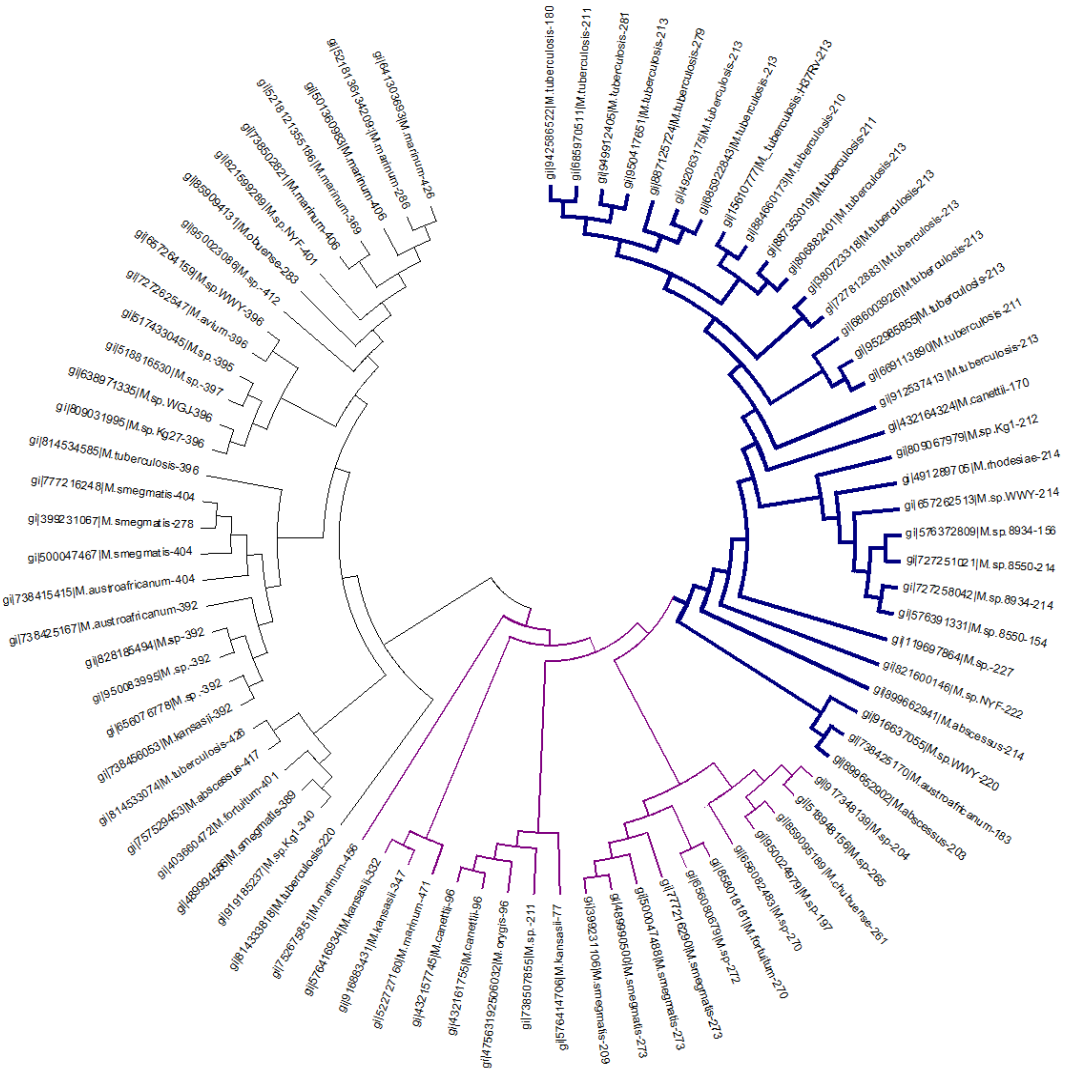


Figure 2.7. Phylogenetic distribution of canonical Fic motif in Mycobacterium sp. Phylogeny of Fic proteins from Mycobacterium sp. *demonstrating the*

distribution of class I and class II Fic proteins. Subgroups are color-coded according to sequence length (blue; 154-214 aa, black; 390-426 aa and purple; intermediate).

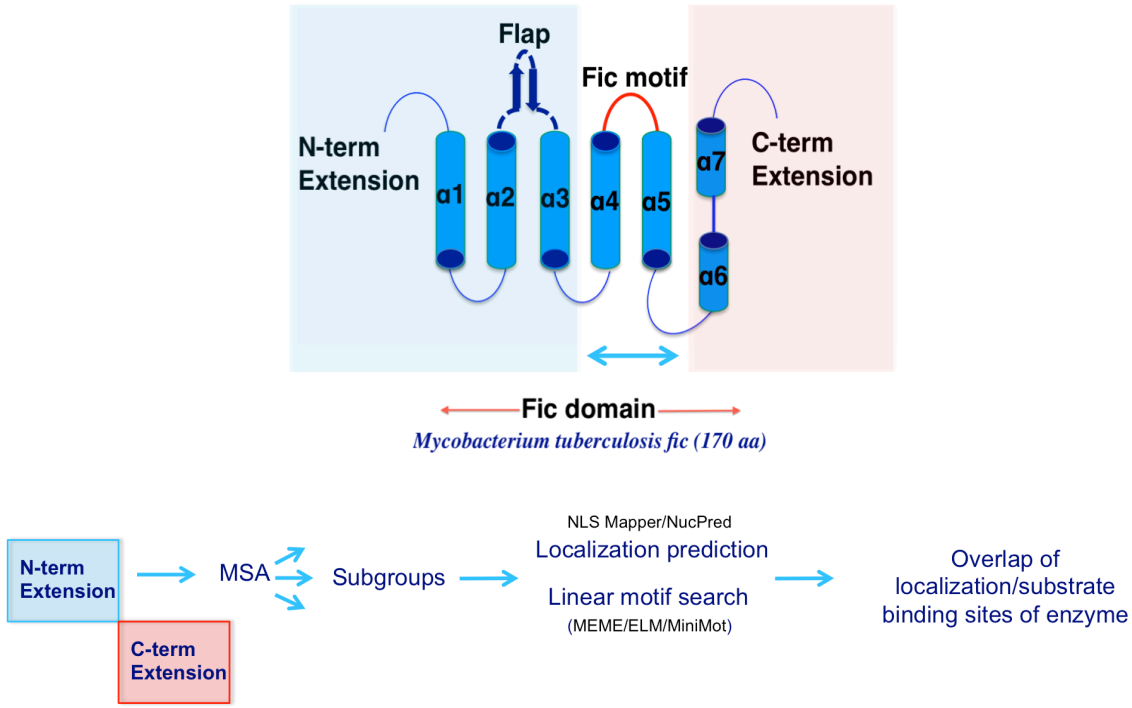


Figure 2.8. Topological representation of seven helices and schematic of the steps followed to read the variability.

Step I: Defining the core

The protein sequences were modeled using the Fold recognition software-Phyre V-2.0 (Protein Homology/analogY Recognition Engine). From the alignment of the sequences and the modeled structures, Fic domain of the protein was set on the basis of experimental evidence describing the domain as a five-helix bundle. A minimal Fic domain was considered on the basis of their sequence length and 20 residues were taken to be constant consisting of a helix (α4) and a half region (α5).

Step II: Sub-grouping N- and C-termini extensions

This region was truncated and the sequences were divided into two datasets termed as N and C for the purposes of show-casing independent roles of N- and C-termini in determining or predicting the protein's function. The N and C dataset

sequences were again submitted to MEGA6 and from the alignment and phylogenetic analysis, the sequences were further classified into groups on the basis of nodes and their distribution. There is a clear demarcation between the N and C terminal sequences which can be identified from their phylogenetic alignment wherein the sequences in the N terminal region dataset cluster differently as compared to their counterparts in the C terminal region (Figure 2.9 A & B).

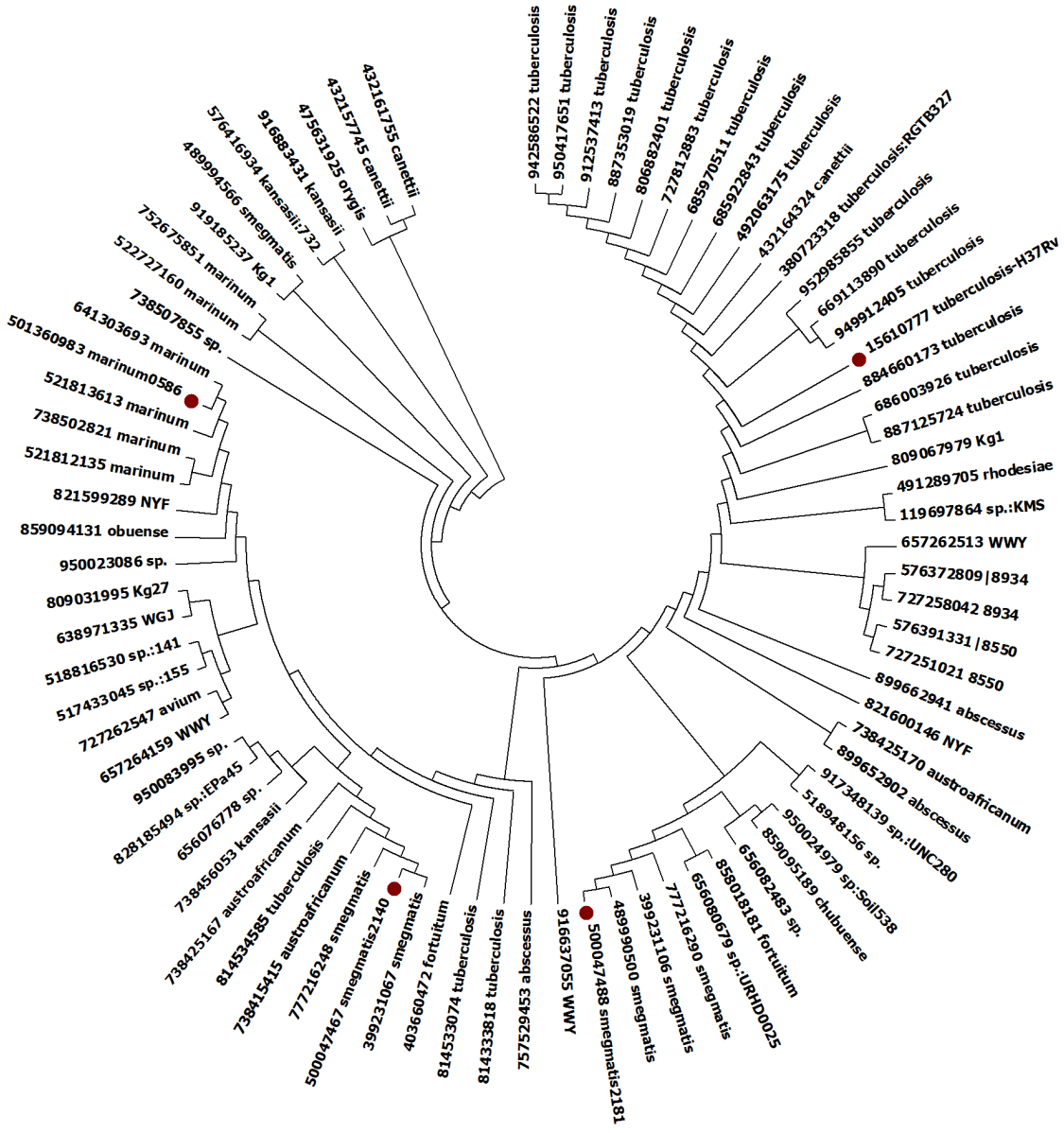


Figure 2.9A. Phylogenetic distribution based on N-termini subgroup of Fic from Mycobacterium

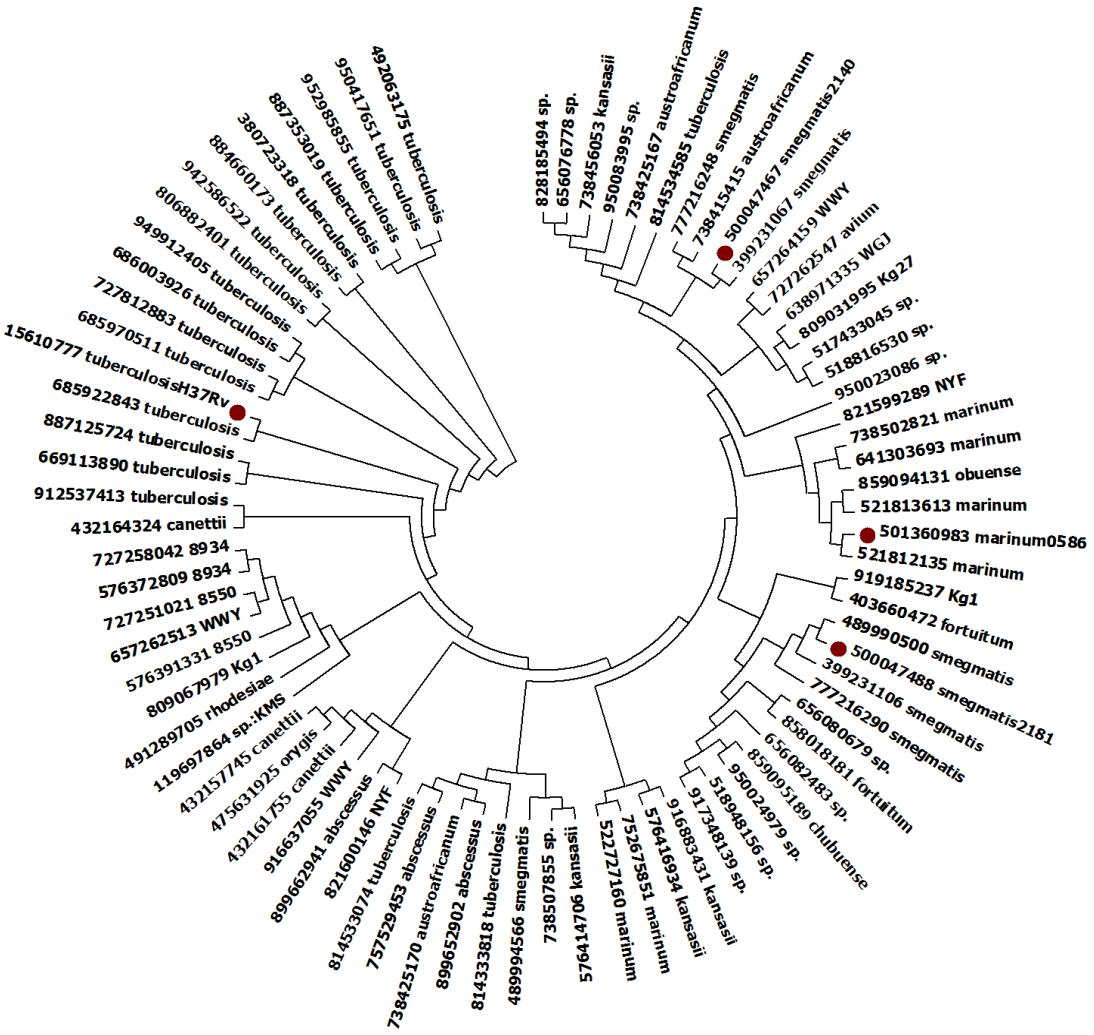


Figure 2.9B. Phylogenetic distribution based on C-termini subgroup of Fic from Mycobacterium

Step III. Prediction of mammalian localization and search for linear motifs

In view of the above observations, the sequences in the N-and C-terminal sections were further subdivided into 10 and 11 subsets respectively, on the basis of their sequence alignment and phylogenetic relationship. Consensus for each group of sequences was obtained using Jalview and this was further submitted into various Motif discovery tools like ELM - eukaryotic linear Motif, MiniMotif finder, and MEME (Bailey and Elkan, 1994). The motifs so predicted by the respective finders were again filtered on the basis of their lengths 4 - 8 amino acids long. Only the

consensus results having a minimum length of 4 amino acids were taken (the results with SH3 domain were also not included). Manual analysis of the sequence alignment was also carried out to identify a stretch of amino acids, which was conserved in that particular set of sequences and the functional association of the motifs were done using literature survey. The results were then consolidated to list the various motifs identified (Table 2.3 and 2.4). The plausible functions related to the identified linear motifs are listed in Table 2.5.

Table 2.3. N-termini dataset and the linear motifs predicted

N-termini Subsets	Predicted Motif
SN1 - H37RV	WQGYF, TxxI, ExxxLL, RxLF, [LV]xYx[RK], PxxxD, Wxx[FWY], RxGx, [DE]xxxL[LI], [DE]xxLL, Y[ILM]xx[ILV]
SN2	[LV]xYxx[RK], PxxxD, [RK]x[RK], Wxx[FWY]
SN3	RxLF, PxxxD, TxxI, DxxG
SN4	LPxP, MxxxL, RxLF
SN5 (SMEG 2181)	MxxxL, MxxE, LxxLL, PxxxD
SN6	DDxxxE, Fxx[RK]x[RK], YxxL, MxxE, PxxxD
SN7	IPXXI, WIGG, TxxD, [PSAT]x[QE]E, YxxL, Yxx[FYL], MxxxL
SN8	WIGG, Yxx[FYL], MxxxL, PxxxD, PxxP
SN9	Wxx[FW], [ST]x[LV], WxxF, ExxxLL, MxxxL, PxxxD, Wxx[FWY], DxxG, PxLxP
SN10 (M 0586)	LPXXI. WIGG, P**PP**P, LxxLxxLxL, MxxxL, LxxLL, PxxxD

Table 2.4. C-termini dataset and the linear motifs identified

C-termini Subsets	Predicted Motif
SC1 (H37RV)	KTDL
SC2	LXXL, LXXRR, TXX[IL], RXXK

SC3	RXXY, WXXE, Wxx[FWY], FxxLF, YxxL, Yxx[FYL], [RK]R[RK]L, DDxxxE, ExxxxxxE, MxxE, LxxxIxxx[IL], Wxx[FW], Txx[SA], DxxG
SC4	KXXY, LLXLL, LXXLL, PxxxD, Txx[IL], Yxx[FYL], YxxGL, RxLF, LxxLL, DxxG, Y[ILM]xx[ILV], PxxP
SC5	GLLR, YxxGL, DxxG
SC6	PXRR, Txx[SA], FxxLL, PxxxD, RxGx
SC7	PSRRR
SC8	LXXY, LLLLL, MxxxxxxxxxG, Txx[IL], [RK]xL, LxxLL, TxxxS, RxGx, [DE]xxxL[LI], [DE]xxLL
SC9	YxxL, TxxD, Yxx[FYL], [RK]xL, LxxLL, TxxxS, Txx[SA], DxxG, RxGx [PSAT]x[QE]E, [DE]xxLL
SC10 (M 0586)	GRRL, RXRXR, RRXR, WxxxE, TxxxS, TxxD, [DE]xxLL, RxGRR
SC11	LLXRR, WxxxE, Txx[SA], LxxRR

Table 2.5. Plausible functions related to the linear motifs

Predicted Motif	Plausible Function
WQGYF	Peroxisomal import - Pex14 - Pex5p
TxxI	Binds the FHA domain of Chk2 - **Checkpoint kinase 2 located on the long (q) arm of chr22**
ExxxLL	Endosomal/lysosomal compartment via interactions with clathrin adaptor protein (AP) complexes.
RxLF	Binds Cdk2 - **cyclin dependent kinase 2 - cell division**
[LV]xYx[RK]	Binds cholesterol - **Ch interacting membrane protein**
Wxx[FWY]	Binds the gamma ear domain of AP1
RxGx	Binds the phosphatase domain of PP1alpha
[DE]xxxL[LI]	Binds the Trunk domain of AP1, AP2, AP3 beta subunits
[DE]xxLL	Binds the VHS domain of GGA - **gamma adaptin ear-containing, ARF-binding (GGA) protein - golgi localized**
Y[ILM]xx[ILV]	Binds Thyroid Homone
[LV]xYxx[RK]	Binds cholesterol - **Ch interacting membrane protein**

[RK]x[RK]	Binds Sar1 - Membrane trafficking GTPase in COPII and is trafficked by Endoplasmic Reticulum Export
LPxP	Binds Cullin - **scaffold for ubiquitin ligases**
MxxxL	Binds AP2 mu2 subunit, clathrin terminal domain, PTB domain of Dab2 and is trafficked by internalization
MxxE	Binds Membrin and is trafficked to Golgi
LxxLL	Binds Oxysterol receptor LXR beta
PxxxD	Binds platelet fibrinogen receptor
DDxxxE	Binds C-Raf1 - **protooncogene - Raf kinase**
Fxx[RK]x[RK]	Binds the phosphatase domain of PP1 - **Protein phosphatase 1**
YxxL	Binds AP complex and is trafficked by internalizaiton
IPXXI	PCNA binding
TxxD	Binds the #1 FHA domain of Rad53
[PSAT]x[QE]E	Binds the TRAF domain of TRAF2 - **tnf associated protein**
Yxx[FYL]	Binds AP1,AP2,AP3, AP4 mu subunits and is trafficked by internalziation
PxxP	In SOS1 binds the SH3 domain of Grb2
Wxx[FW]	Binds the ear domain of AP2 alpha subunit
[ST]x[LV]	Binds the PDZ domain of Syntrophin - **actin binding - cytoskeletal reorganization**
WxxF	Binds Vpr - **viral protein R**
DxxG	binds the G-protien domain of Phosphate
PxLxP	Binds the MynD domain of BS69
P**PP**P	Ribosomal stalling
LxxLxxLxL	Binds Crm1 and is trafficked by Nuclear exportation - **chromosomal maintenance nuclear export protein**
KTDL	Secretory mot signal
LXXL	Nuclear Receptor-Coactivator Interaction Motif
LXXRR	Binds erk1
TXX[IL]	FHA domain of Rad53

RXXK	SH3 domian of GADS
RXXY	Factors binding - cadherins- intracellular signalling
WXXE	Type III secretion system - RhoGTPase - downstream of Cdc42
FxxLF	Binds AF2
[RK]R[RK]L	Binds Cdk2/Cyclin E complex
ExxxxxxE	Binds HIV tat
LxxxIxxx[IL]	Binds nuclear receptor
Txx[SA]	Binds the FHA domain of KAPP;
KXXY	Y of histones - chlorinated
LLXLL	Retinoblastoma protein - histone descylase associated
YxxGL	Binds AP2 mu2 subunit, clathrin terminal domain, PTB domain of Dab2 and is trafficked by internalization
GLLR	Between protein phosphatase regulatory B subunit
PXRR	DBL homology domain - guanine nucleotide exchange factors
FxxLL	Binds Androgen receptor
PSRRR	Mitosis regulation - pyrosphorylation by cdk1
LXXY	Endocytosis mot
LLLLL	AP1 DNA target protein
MxxxxxxxxxG	Binds an unknown target and is trafficked to lysosomes
[RK]xL	Binds Cdk2
TxxxS	Binds Skp1
GRRL	Part of ribosome - RNA binding protein
RXRXR	Nuclear localization
RRXR	Nuclear localization
WxxxE	Binds Rac1
RxGRR	glucose transport requires this consensus motif; target is an unknown target
LLXRR	Cytosolic tail of Golgi localization

Step IV. Rationalizing the functional extrapolations based on the overlap of predicted localization and substrate interacting sites

Next, we hypothesized that N- and C-terminal regions, by localizing these proteins in specific compartments, may play a deterministic role in their eventual function. To this end, most of the motifs identified related to mammalian cell functions were further filtered based on protein's localization prediction in specific compartments to the motifs identified for function in that particular compartment. The sorted motifs and localization prediction for *Mtb* Fic proteins are shown in Table 2.6. For *Mt*Fic (Rv3641c), the localization prediction was in ER lumen and peroxisome and the motif sorted manually, also corroborated with their involvement in ER localization, which also relates to the endocytic pathway. Also, *Sm*Fic protein was predicted to localize in the nucleus and the sorted motifs also implicate it to interact with DNA binding proteins or proteins involved in ribosomal stalling (Table 2.6).

Table 2.6. Summary of predictions for subgroups containing *Mtb* Fic proteins

Groups	Localization	Motifs related to mammalian proteins
SN-1/SC-1 Rv3641c : Class I Fic	ER lumen/peroxisome	WQGYF - peroxisomal import - Pex14 - Pex5p ExxxLL - endosomal/lysosomal compartment via interactions with clathrin adaptor protein (AP) complexes RxLF - binds Cdk2 - cyclin dependent kinase 2 - cell division Wxx[FWY] - binds the gamma ear domain of AP1 RxGx - binds the phosphatase domain of PP1alpha [DE]xxxL[LI] - binds the Trunk domain of AP1,AP2,AP3 beta subunits [DE]xxLL - binds the VHS domain of GGA - gamma adaptin ear-containing, ARF-binding (GGA) protein - golgi localized Y[ILM]xx[ILV] - binds Thyroid Hormone KTDL - secretory mot signal
SN-9/SC-9 SMEG_2140/ Mmar_0586 :Class II Fic	Nucleus	LPXXI – PCNA binding motif PPP - Ribosomal stalling LxxLxxLxL - binds Crm1 and is trafficked by Nuclear exportation - chromosomal maintenance nuclear export protein LxxLL - binds Oxysterol receptor LXR beta GRRL - part of ribosome - RNA binding protein RXRXR - Nuclear localization RRXR - Nuclear localization WxxxE - binds Rac1 TxxxS - binds Skp1 [DE]xxLL - binds the VHS domain of GGA - gamma adaptin ear-containing, ARF-binding (GGA) protein - golgi localized

2.3.5 Identification of class II Fic protein in *Mtb* strains

Mtb strain H37Rv sequenced and updated in 2013 by Cole ST *et al*., characterizes 4,411,532 bp of DNA sequence representing the whole *Mtb* chromosome (strain H37Rv) (TubercuList database), where 91.2% sequence is the protein coding sequence (Lew et al., 2011). Search for Fic protein in TubercuList shows one hit "Rv3641c" for *fic* and "Rv3642c" gene upstream to the Rv3641c represents the anti-toxin counterpart of Fic protein. Rv3641c-Rv3642c together displays the class I Fic protein. However, we identified three Fic proteins in *Mtb*; belonging to class II. All of them have the canonical Fic motif. Two Fic proteins were identified in *Mtb* strains Bir 78 (279 aa), Bir 102 (277aa) and belong to our subgroup SN-1 & SN-2 (Figure 2.10). Conversion of class I to class II Fic appear to be due to single nucleotide deletion resulting in frameshift mutation, either converting stop codon to amino acid encoding nucleotide sequence (*Mtb* Bir 78 strain Fic) or a nucleotide deletion in anti-toxin ORF resulting in frame shift making Rv3642c and Rv3641c ORFs as a single ORF (*Mtb* Bir 102 strain Fic). *Mtb* strains Bir78 and Bir102 were sequenced from the isolates of tuberculosis patients by Welcome Trust Sanger Institute in 2015.

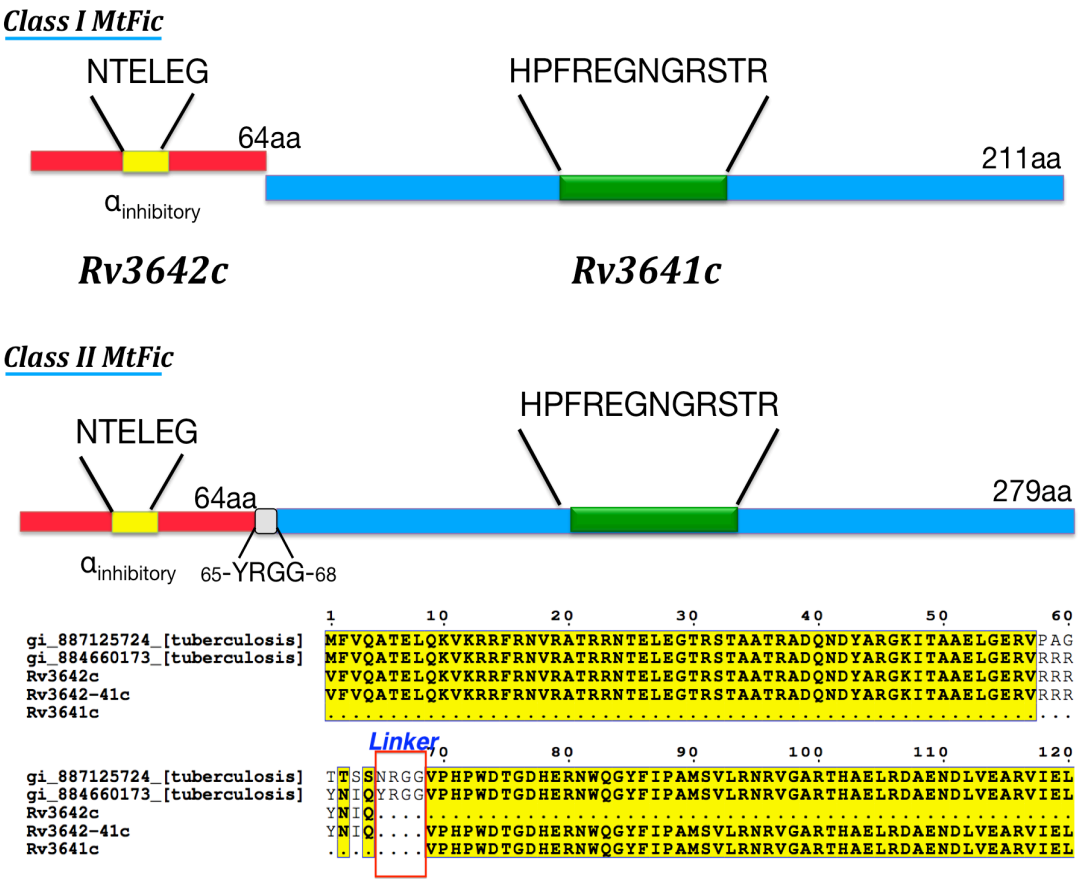


Figure 2.10. Schematic of Fic proteins from Mtb.

The third class II Fic protein from *Mtb strain 200127,* was sequenced from an isolate from antibiotic-resistant tuberculosis patient. The protein sequence shows 99% identity with Fic protein from a gram-positive bacterium *Rothia sp*. *They* are round to rod-shaped bacteria that are part of the normal community of microbes residing in the mouth and respiratory tract. Although this organism is believed to be of low virulence, it is increasingly recognized as an opportunistic pathogen mostly affecting immunocompromised hosts (Lee et al., 2008a; Morgan et al., 2010; Ramanan et al., 2014). In our subgroup division, this protein sequence belongs to SN-9 & SC-9 group where the Fic domain is preceded by an additional helix-turn-helix domain. MSA with closest homologs from *M. smegmatis* (*Sm*Fic, 43.3% identity) and *M. marinum* (*Mm*Fic, 31.5% identity) has been shown in Figure 2.11.

```
0586_MMAR       -----MRPSDAWPRHSA-ERRPWAQTQRGG--TRADRTLRSVTVSLPPYIAKVDANIDAD
Mtb             -----MPSAIIWPALTWEPQTPWVHRDALASRRATRRNSGTFSSAVPASIAALNVELSPA
2140_SMEG       MASAEEGTALTYETLTWDAPA--EAGYGLADLRAAQRQAGEYEAAIPASIAELAVSLPAA
                        :  :   :             .    : *       ::*  ** : ..:

0586_MMAR       IAVKLEDAMSEISRLDSTHGT-HLAGLSTLLLRTESVASSKIERVEASVDDYARALHGGR
Mtb             LRARLQENATAMTRFDERLHHIGGIPFSAVLLRGESASSSQIENLTVSARRLSLAVLGAS
2140_SMEG       VLADAEEASNEITRFDAELGD-EVAPFAAVLLRSESTASSNIENLTASARAIAEA--EAL
                : .  ::  . ::*:*          ::::*** **.:**:**.: .*.   : *   .

0586_MMAR       GNSSAVSMVAATTALKEMIASVNRDAPIQMTAILRAHEALMREDPTEGQHAGQVRTVQNW
Mtb             GSKVGVNAELVARNVHAMRAALNTAENISVDNILAMHRELTRGL---QQDSGVLRTQWVW
2140_SMEG       GDTSRRNAALIVSNTEAMKAAVALADRLDDRAILAMHAALMHHS--DPVSAGNWRTEQVW
                *..   .    .   . * *::     :.   **  *  * :        :*  **   *

0586_MMAR       IGGSDYSPRNALYVPPPPDTVHAYMDDLIEFANRTDIPVLIQAAIAHAQFESIHPFTDGN
Mtb             IKGD--SPVTADYVAPHHERVPAALEDLVAFMNRRDIDPLAQAAIAHAQFETIHPFTDGN
2140_SMEG       IGGGNFGPRGADYIAPRHTRVPGAIEDLLAFIRRADVPSLPQIAIAHAQFETIHPFTDGN
                * *   .*  * *: *    * . ::**: * .* *:  * * ********:********

0586_MMAR       GRIGRALINTVLRRRGATTRLVVPLASALVAHRERYFGALNTYRAGDLRPLIVTFANSSR
Mtb             GRTGRALISALLLARGVTRHVILPVSSGLLHDTDGYIQALTDYRAGDAEPIIELFIDAAQ
2140_SMEG       GRTGRALIQAMLRHKRLTRQITVPVSAGLLTDTDAYFGALSAYRDGDPAPMVERLSEASL
                ** *****.::*  :  * :: :*:::.*: . : *: **. ** **  *::  : :::

0586_MMAR       TAAAESRITAERLAEIPVEWRNMVGPIRRHSATDKLLLLLPSTPIVSSDDVASLIDAPRS
Mtb             KAITNAELLAQDIETLRDEVL-SIAQ-RKTPLLRSLTDLCCTEPAFTAHMVEEHTQGSRA
2140_SMEG       LAVANGRRLVTDLRSIRAEWDTRIAA-RRDSAVHRVADLLIQHPVFNAKLLQRELGITTG
                 * ::..  .  :  :  *    :.  *:      :  *    * ..:. :        .

0586_MMAR       SVFAAIKRLHDTGVLR-PLTNRKRDQVWGASLVLDELDDLGHRIERASAPSVRK
Mtb             SVYRLLNRMVELQILREERVKIQGQKVWTVPALNRALDDFAARAGRRG------
2140_SMEG       NARRYVDPLTAAGIIV-EFTDRARNRAWSAPEVLSALDAFAARAGRRSRSG---
                ..   :. :    ::    ..   ::.* .  :   ** :. *  * .
```

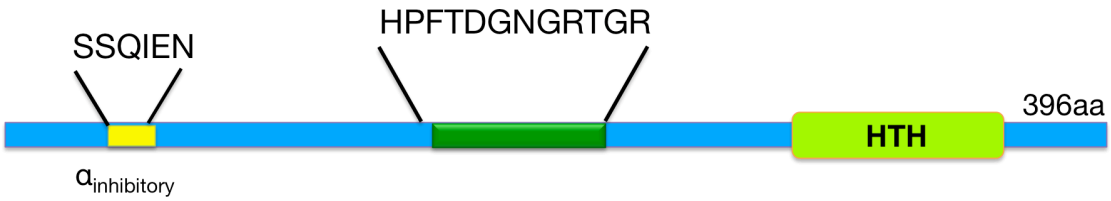


Figure 2.11. Multiple sequence alignment of MtFic with Fic proteins from M. smegmatis and M. marinum *(upper panel) and representation of motifs in class II MtFic protein from Mtb strain 200127 (lower panel).*

2.3.6 Mobile Mystery Fic proteins in *Mtb* strains

One Fic protein (gi: 620036816) from the curated sequences from *M. triplex* strain was annotated in the database as mobile mystery protein B. This sequence showed alignment with *Mtb* Fic protein only in the region covering Fic motif (66%

coverage and 26% identity), although the modeled structure showed conserved Fic fold (Figure 2.12). BlastP for this mobile mystery B protein resulted in identification of proteins with more than 60% sequence identity from *M. behemicum, M. kubicae, M. triplex, M. fortuitum subspecies acetamidolyticum, M. bacteremicum, M.ST-F2, M. tusciae, M. novocastrense, M. obuense* and *M. colombiense* (Figure 2.13). Mobile mystery protein B is the bona fide toxin of the mobile mystery protein B family (IPR013436) and their putative antitoxins are referred as mobile mystery protein A family (IPR013435), however no experimental evidence relating functions of these proteins is available (Harms et al., 2016; Mitchell et al., 2014). As the canonical Fic motif is conserved in this subgroup of mycobacterial Fic proteins, they are highly likely to perform AMPylation reaction.

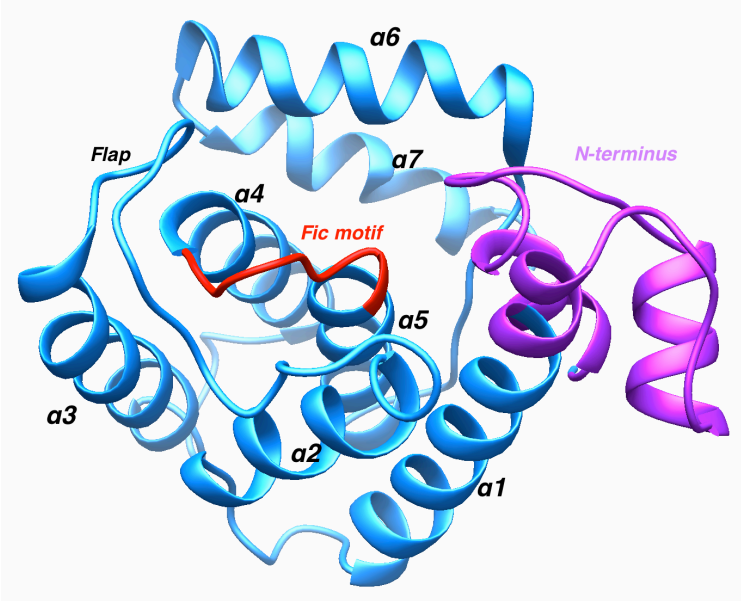


Figure 2.12. Modeled structure of mobile mystery protein B from M. triplex strain using I-TASSER server. *The model structure shows the conserved Fic topology with N-terminus extension. Fic motif region is shown in red and N-terminus extension in purple.*

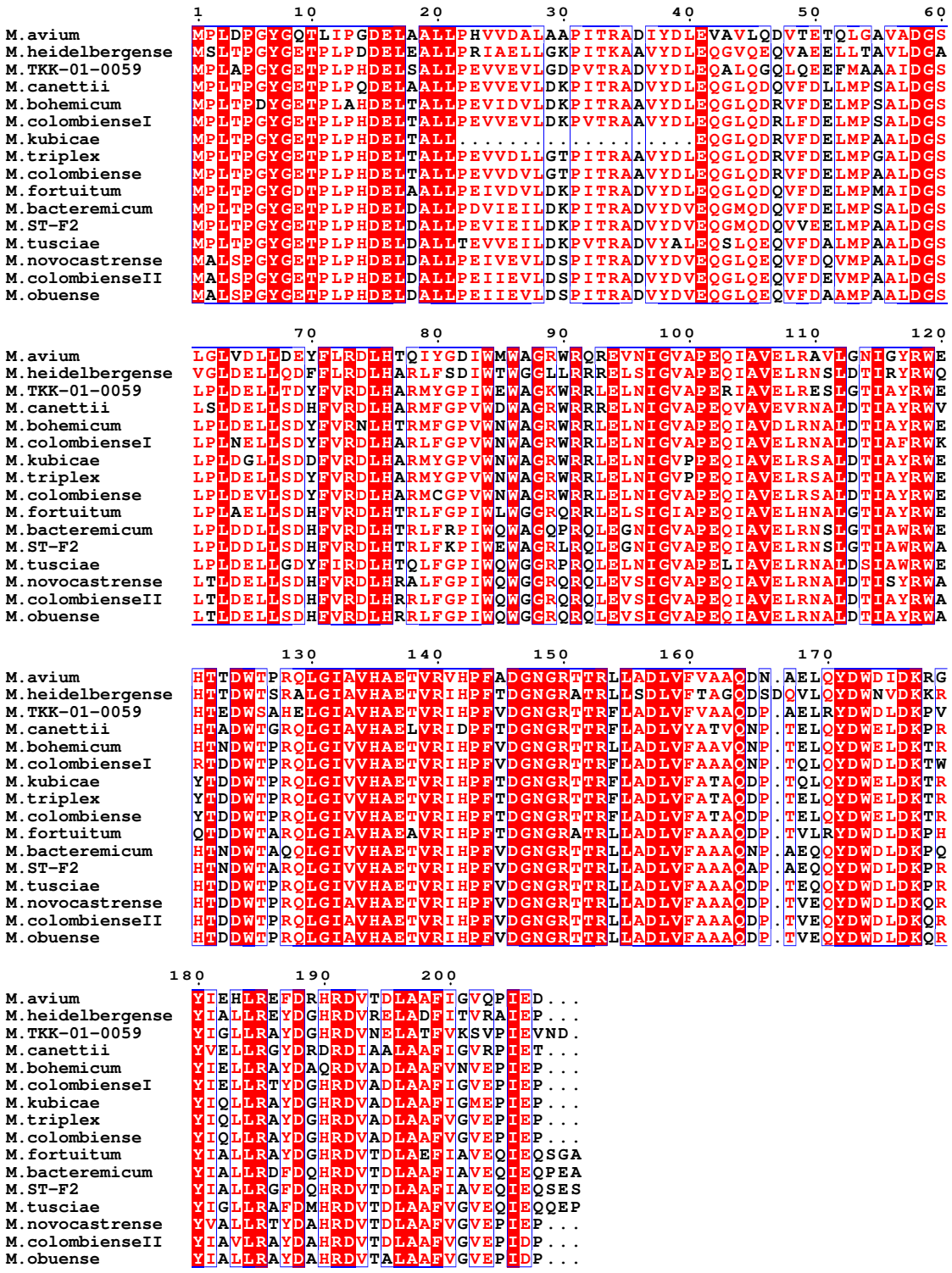


Figure 2.13. MSA of mobile mystery protein B, identified in Mycobacterium sp.

Short proteins identified in Mtb with XRE domain

>CNF55812.1 transcriptional regulator [*Mycobacterium tuberculosis*]
MTNFPKWSEVRADIVAAAGGEEAVAEARKRNQAYIDGHRLAERRKALGLFQTDVADKMGVTKSWISQI
ER GEVSTVDAIARYVQALGGEIQISAVFGDDQYILRGTDTHAA

>WP_078750489.1 XRE family transcriptional regulator, partial [*Mycobacterium tuberculosis*]
MESFGEQLRALREERKLTVNQLATYSGVSAASISRIENGKRGVPKPATIKKLAEALKIPYEGLMYKAG
YI EEVHEARATYETKCQLLEKAEAYDL

>SGD43538.1 XRE family transcriptional regulator [*Mycobacterium tuberculosis*]
MMKKYAISEAIGQVIRQYRTNAGLTTKQLAHRIGISQQQLSRYERGVNRIDVDTLLRVSLAFKLTPGR
FF EEMNMTGTGLDDILYENEEGDIQEIRMSLIADSIISPRDF

Figure 2.14. Fasta sequences of putative XRE domain proteins from Mtb

These proteins also possess the conserved Fic fold as shown by the model structure of one of the representative *M. triplex* (Figure 2.12). While the Fic motif and Fic fold remain conserved, the N-terminus extension does not show sequence similarity with the alpha inhibitory domain and canonical Fic anti-toxin motif is also absent in these sequences (Figure 2.13). Interestingly, a functional quest for this N-terminus extension identified a match with Rv1052, an uncharacterized protein, present in *Mtb* H37Rv strain where the upstream gene to it, Rv1051 encodes for a non-coding small RNA (Tsai et al., 2013). InterPro listed putative mobile mystery proteins A share putative wHTH domain and is predicted to function as XRE family transcription regulators. A search for XRE family transcription regulator / Ydil like protein identified three short (90 - 110 aa) polypeptides (Figure 2.14) belonging to this domain. A correlation of small coding RNA and Rv1052 with Fic domain-containing proteins needs experimental validation. However, as part of TA module, they can be involved in auto-regulation of mobile Fic toxin.

2.4 DISCUSSION

In *M. tuberculosis* strains, *fic* gene shows polymorphism at the gene level (Diaz et al., 2006; Sassetti et al., 2003). This chapter covers the analysis of the variability present at the level of their protein sequences. A sequence comparison of

Fic proteins *M. tuberculosis* and *M. smegmatis* revealed an arginine-rich nuclear localization signal in Fic protein from *MSmeg* which appears to be functional as the mammalian expression constructs of the *Smeg* Fic proteins showed co-localization with DAPI stain. The experiments were performed in both ways; one where the protein was ectopically expressed in mammalian cells and second where the mammalian cells were incubated with purified recombinant proteins. In both the experiments, the Fic proteins from *M. smegmatis* co-localized with DAPI stained nucleus. Further, the results suggest that these proteins can be internalized by human differentiated macrophage cells (PMA treated THP1), circumventing the proteolytic machinery of the macrophage system. Thus, validation of the nuclear localization signal from non-pathogenic *Mycobacterium* strain adds to the functional connotation of sequence-based motif present in Fic proteins. Although we identified the functional linear motifs, their physiological significance in non-pathogenic bacteria is still an open question because of their propensity to evolve with horizontal gene transfer during co-colonization with pathogenic bacteria. Next, we curated Fic domain proteins from 21 *mycobacterium sp.*

At first instance, the pathogenic *Mycobacterium* strains showed all class I Fic proteins and non-pathogenic proteins were mostly abundant in class II Fic domain proteins. However, the recently annotated *Mtb* Fic sequences presumably belong to class II. The polymorphism in Fic proteins are apparent as each species share one or more than one copy of canonical and non-canonical Fic proteins. 86 non-redundant Doc proteins sequences and 81 non-redundant canonical Fic proteins from 21 *Mycobacterium* species and strains also make these enzymes suitable as models to study a role of phosphorylation and AMPylation in pathogenic as well as in non-pathogenic mycobacterial strains.

Fic proteins from *mycobacterium sp*. showed marked variation in their N- and C-termini with localization signals for mammalian compartments. We divided the Fic proteins from 21 genome of *Mycobacterium sp*. in different subgroups based on their sequence identity at N- and C-termini, having the defined core consistent, followed by *de novo* identification of linear motifs in each subgroups. Further, the linear motifs

identified were filtered based on their presence in the loop regions as well as specifically in the flap regions of the modeled structures. While characterizing and subdividing the above mentioned groups, we identified one such case where we recognized an extra copy of fic protein in whole genome sequencing report of *Mtb* that appeared in 2015, showing 31% identity with fic protein present in genomic island of *M. marinum* (Mmar_0586) and 43% identity with that of *M. smegmatis* (Smeg_2140). As the purified Smeg_2140 (*Sm*Fic) showed functional nuclear localization signal, it appears similar to the *Mtb* Fic protein. Surprisingly, the *Mtb* Fic protein is 99% identical to Fic protein from *Rothia sp*. which is an emerging opportunistic pathogen in immunocompromised human. As these gram-positive bacteria colonize in human mouth and respiratory tract, its horizontal gene transfer to the advanced pathogenic mycobacterium cannot be ignored. These proteins also have winged helix-turn-helix domain fused at their C-termini, suggesting their role as transcriptional regulators as well.

Analyses of phylogeny of full-length proteins revealed two Fic domain proteins of ~270 aa which clustered with the proteins of ~210aa (class I). A closer inspection highlighted that in these outliers, the antitoxin domain present upstream is fused with the toxin counterpart. A comparison of nucleotide sequences confirmed a frame shift mutation due to a single nucleotide deletion in the variable region between the toxin-antitoxin sequences which altered the read to translate a single polypeptide chain, resulting in a class II Fic protein.

As an interesting observation, we also identified a new class of toxin-antitoxin system in *Mycobacterium* i.e, mobile mystery protein B and identified putative small XRE family transcriptional regulator protein which may be the antitoxin counterpart as mobile mystery protein A. The interesting feature of the mobile mystery protein B is the fused canonical Fic motif along the side, making them the AMPylators. As mobile mystery proteins are annotated to function as transcription regulators, occurrence of an AMPylation motif in them adds another level of complexity to their transcriptional regulatory functions.

Thus, we have identified a new set of putative regulators; TA system; multifunctional enzymes from mycobacterium, all equipped with a Fic domain, either canonical or non-canonical. Some of them occur as part of core gene while others emanate from genomic islands, highlighting their roles in the day-to-day life of the organisms that harbor them as well as aiding to the repertoire of adaptive responses in the bacteria. Next, we sought to characterize these Fic proteins functionally. Characterization of Rv3641c (*Mt*Fic) has been covered in Chapter 3 and subsequent Chapter 5 extends over the characterization of Fic proteins from *M. smegmatis* and *M. marinum* homologs of class II *Mt*Fic identified in antibiotic-resistant strain. Chapter 4 highlights the characterization of these proteins with respect to their nucleotide selection.

CHAPTER 3

AMPylation, a novel post-translational modification exhibited by Fic from Mtb (Rv3641c)

Table of contents

3.1 INTRODUCTION

As described in Chapters 1 and 2, structurally Fic proteins are topologically conserved (Kinch et al., 2009) and functionally belong to type II toxin-antitoxin system consisting of both the antitoxin motif (S/T)xxxE(G/N) and toxin Fic motif (HPFxxGNGRxxR). Here, the antitoxin could be part of the same operon (class I) or constitute extensions at the N-termini (class II) or extensions at the C-termini (class III) (Engel et al., 2012). The toxic effect in bacterial cells is exclusively due to the AMPylation of bacterial type II topoisomerases. As a consequence, the cell loses control on DNA topology and undergoes reversible growth arrest (Harms et al., 2015; Lu et al., 2016). The toxin effects in mammalian cells have been attributed to the AMPylation of Rho and Rab family GTPases (Muller et al., 2010; Worby et al., 2009; Yarbrough et al., 2009).

In Chapter 2, we covered the distribution of Fic domain proteins across *Mycobacterium sp*. and specifically identified Fic domain-containing proteins in *Mtb* H37Rv strain, viz. Rv3641c (Figure 3.1). To characterize its function, we looked for pathways involving small GTPases during host-pathogen interactions. In *Mtb* infection, decreased motility of the infected granuloma macrophages (Yang et al., 2012) and arrest of phagosomal maturation (Kelley and Schorey, 2003) are the two subsets of biologic processes consequent to infection where Rho GTPases and Rab GTPases are involved. Deciphering the role of AMPylation mediated by Rv3641c entails identification of a specific substrate e.g. a GTPase; thereby, the significance of such a GTPase vis-à-vis its AMPylation in phagosomal maturation could be uncovered.

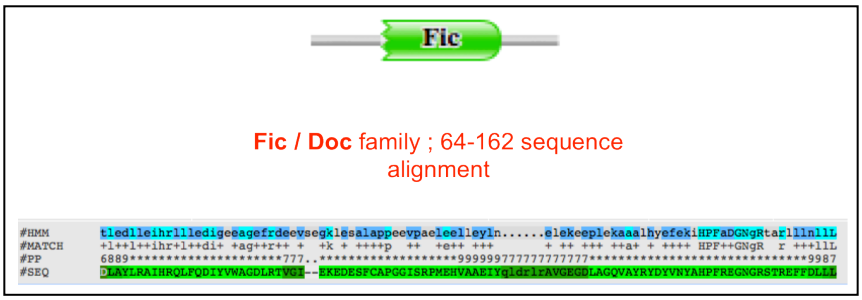


Figure 3.1. Fic domain-containing protein Rv3641c from Mtb strain H37Rv

Also, Fic polypeptide from *Mtb* has been shown to increase immunogenic response as evident from an *in vitro* detection of $CD8^+$ T cells that specifically recognize a *Mtb* polypeptide and by *in vivo* detection of a delayed type hypersensitivity reaction in individuals with latent tuberculosis (TB).

Moreover, PTMs are key players of host-pathogen interaction in *Mtb* infection. They regulate key signaling events, membrane anchoring and compartmentalization. In this chapter describes the *in-vitro* and *in-vivo* functional consequence of a recently discovered PTM - AMPylation, by a toxic Fic protein, Rv3641c, from *Mtb* (*Mt*Fic). *Mt*Fic belongs to class I Fic where the Rv03642c posses the inhibitory motif- (S/T/N)xxxE(G/N) and Rv3641c has the Fic motif- HPFREGNGRSTR (HPFxxGNGNRxxR). *Mt*Fic has been shown to hydrolyze nucleotide triphosphates (NTPs) in the absence of substrates (Mishra et al., 2012). However, the likely physiologic substrates and the function of the *Mt*Fic remain unexplored. Here, the studies reveal that *Mt*Fic can induce elongated morphology in *E. coli* cells where the morphological effect is demonstrated by its ability to AMPylate DNA Gyrase B subunit. Further, far western assay supported by co-immunoprecipitation (co-IP) and mass spectrometry affirms that the enzyme has specific interacting partners in mammalian cells and identifies them to be Cdc42, Rac1 and vimentin. Additionally, the expression of *Mt*Fic in mammalian cells leads to spotted morphology of eGFP tagged protein in ER and nucleus as well as the disruption of nuclear membrane, consistent with cytoskeletal collapse and cell rounding.

3.2 EXPERIMENTAL PROCEDURES

3.2.1 Materials

Chemicals: The enzymes used for cloning were obtained from New England Biolabs (NEB). The Ni-NTA metal affinity resin and IPTG were procured from Novagen. GST- and MBP resins were obtained from GE healthcare. Expresso™ rhamnose cloning and expression system kit was obtained from Lucigen. ATP, ADP, AMP,

CTP, kanamycin, ampicillin, imidazole, L-rhamnose, glucose, maltose and SDS-PAGE reagents were from Sigma Chemical Co., St. Louis, MO.

Strains and plasmids: *E. coli* DH5α cells (Novagen) were used for cloning and *E. coli* BL21 (λDE3), *E. coli* plys, *E. coli* pGro7cells (Novagen) were used for expression.

Antibodies and constructs: The following primary antibodies at 1:1000 dilutions were used for immunoblotting, co-IP and far western experiments: anti-His, anti-Sumo, anti-MBP, anti-GFP, anti-vimentin antibodies. Anti-AMPylated threonine (T^{AMP}) ab and anti-AMPylated tyrosine (Y^{AMP}) ab were used at 1:2000 dilutions. Secondary immunodetections were performed using HRP-conjugated anti-rabbit or anti-mouse IgG (Abcam). pGEX-2T-Cdc42-Q61L, pGEX-2T-Rac1-Q691L and pGEX-4T3-RhoA-Q63L constructs were obtained from Addgene.

3.2.2 Cloning of the mycobacterial Rv3641c gene

The Rv3641c gene was amplified with gene specific primers (Table 3.1) using PCR and *Mtb* genomic DNA as the template. PCR amplicons and vectors were digested with restriction enzyme, followed by ligation and screening to obtain the construct expressing His-tagged and MBP-tagged *Mt*Fic proteins. The gene sequence was confirmed further by sequencing. Sumo-tagged constructs were prepared in Expresso Rhamnose SUMO vector (Lucigen) using ligation independent homologous recombination based cloning method as described in the protocol. In brief, the pRham N-His SUMO vector is a linear vector. After amplification of the target gene with primers that append 18 bp sequences homologous to the ends of the vector, the PCR product was mixed with the pre-processed vector followed by transformation into *E. coli* 10G chemically competent cells where recombination within the host cells joined the insert to the vector. The nucleotide sequence of the cloned gene was confirmed by sequencing. Cdc42 and Rac1 genes were cloned in pET28a(+) vector using gene specific primers (Table 3.1) employing (pGEX-2T-Cdc42-Q61L and pGEX-2T-Rac1-Q61L) constructs as templates.

3.2.3 Site directed mutagenesis of putative substrates Rab5

The templates used for PCR amplification, pGEX-5X-Rab5, was a generous gift from Prof. Marino Zerial, Max Planck Institute of Molecular Cell Biology and Genetics, Germany. Linear PCR amplification with gene-specific primer (Table 3.1) having the mutation at one nucleotide position was followed by *Dpn*I digestion to cleave the methylated parental strands. Following *Dpn*I digestion, the PCR products were transformed in *E. coli* DH5α strain. Positive colonies were screened on the basis of restriction digestion with *Nae*I as its restriction site was introduced in primers. The similar protocol was followed for *Mt*Fic^{H144A} point mutant preparation.

Table 3.1 List of primers

Primers	Sequences (5'-3')	Restriction sites
pBAD-Sumo vector N-His-sumo-Fic	FP: CGCGAACAGATTGGAGGTGTGCCGCATCCATG RP: GTGGCGGCCGCTCTATTATCAGAAGTCGTAAGTG	
pMAL-5cx vector N-MBP-Fic	FP: AATTATCATATGGTGCCGCATCCATGGGA RP: AGGCTAGAATTCTCAGAAGTCGTAAGTGGG	***NdeI*** ***EcoRI***
pET 28 a(+) vector His-Cdc42 Q61L (1-179 aa)	FP: GGAGCCATATGATGCAGACAATTAAGTGTGTTGTTGTGGGCGA RP: GACTCGGATCCTTATGGAGGCTCCAGGGCAGCCAATATT	***NdeI*** ***BamHI***
pET 28 a(+) vector His-Rac1 Q61L (1-184 aa)	FP: TGGTTCCTCATATGCCGCAGGCCATCAAGTGTGTG RP: GCATTTTCTCTCGAGTTACTTCACGGGAGGCGGGCAGAG	***NdeI*** ***XhoI***
pEGFPC1-MtFic	FP: CTCAAGCCTCGAGCTATGGTTCCCCATCCTTGGGATACCGGCGAC RP: CCGCGGCTGCAGTCAGAAGTCGTAGGTAGGCTCGGC	***XbaI*** ***PstI***
cMtFicH144A pEGFPC1	TACGTGAACTACGCTGCCCCCTTCAGAGAAGGC	***AccI***
MtFicH144A	TACGTGAACTATGCCGCCCCGTTCCGCGAGGGC	

MtFicHT5A	CCGCATCCATGGGACGCCGGCGATCACGAACGG	***NaeI***
MtFicY17A	CGGAATTGGCAGGGCGCCTTCATCCCCGCTATG	***HaeII***

3.2.4 Protein expression and purification:

(a) VopS-WT and VopS-mutant

VopS-WT and *VopS-mutant* cloned in pGEX-TEV were kind gifts from Prof. Kim Orth, University of Texas Southwestern Medical Center, USA. Proteins were purified via GST affinity chromatography as described in (Luong et al., 2010). In brief, the induced supernatant was loaded onto the GST column, washed with 50 mM Tris-HCl pH-8, 150 mM NaCl and 0.1% Triton X-100 followed by re-equilibration with tobacco etch virus (TEV) protease cleavage buffer (50 mM Tris-HCl pH-8, 0.5 mM EDTA & 1 mM DTT). TEV protease was added to the column and incubated overnight at 4 ºC for the tag cleavage. The flow-through having TEV protease and VopS protein was collected and dialyzed against buffer containing 10 mM Tris-HCl pH-7.5, 50 mM NaCl, 1 mM DTT. Dialysis was followed by concentration and then loaded onto a 16/60 Superdex 100 column and eluted in same buffer condition. TEV protease (His-tagged) from eluted fractions was removed using Ni-NTA resin. VopS mutant protein was purified by a similar procedure. However, it was eluted in buffer containing 10 mM Tris-HCl pH-7.5, 100 mM NaCl, 1 mM DTT.

(b) N-His-MtFic

N-His-*Mt*Fic was purified by Ni-NTA affinity chromatography. *E. coli* BL21 competent cells were transformed with pET28a(+) - *Mt*Fic. 1 L LB was inoculated with overnight grown culture and induced with 0.1 mM IPTG at OD 0.6 for 6 hours at 37 °C. Cells were harvested at 5000 rpm and resuspended in 30 ml PBS pH-7.4. The cell suspension was lysed by sonication on ice followed by centrifugation at 15,000*g*. The supernatant was discarded and the pellet was washed in 30 ml PBS (pH-7.4) and subjected to centrifugation at 15,000*g* for 30 min.

For denaturation-renaturation purification, the pellet was solubilized in 25 ml of 8 M Guanidine hydrochloride (GdnCl) in PBS pH 7.4 with overnight incubation at

room temperature. The solution was centrifuged at 15,000*g* for 30 min. The supernatant was allowed to bind with Ni-NTA resin, washed with 20 mM imidazole containing 8M GdnCl in PBS pH-7.4 and finally denatured protein was eluted with 8 M GdnCl in PBS pH-7.4 containing 250 mM imidazole at room temperature. The eluted fractions were 10-fold diluted with 0.5 M arginine-hydrochloride in PBS containing 1 mM EDTA to reduce the denaturant concentration from 8 M to 0.8 M. The resulting solution was again concentrated back to the original volume in an Amicon concentrator. This solution was then dialyzed once against 0.5 M arginine-hydrochloride in PBS containing 1 mM EDTA to remove GdnCl and then dialyzed against PBS pH 7.4, 1 mM EDTA to remove arginine. During rapid dilution step as well as on dialysis, most of the protein precipitated. The recovered protein from supernatant was used for *in vitro* activity and binding analysis.

(c) N-MBP-MtFic

For MBP tagged *Mt*Fic purification, pMAL-c5X-*Mt*Fic construct was transformed in *E. coli* BL21 strain. Cells were induced with 0.1 mM IPTG at O.D. 0.8 and grown overnight at 20 °C, following cell lysis, the supernatant fraction was subjected to affinity chromatography on an amylase-sepharose column. The fractions eluted with 20 mM maltose were dialyzed against buffer (20 mM Hepes, pH-8, 300 mM NaCl, 10% glycerol) and concentrated using 10-kDa centricon.

(d) N-His-Sumo-MtFic

For N-His-Sumo tagged *Mt*Fic purification, *E. coli* G1 competent cells were transformed with the construct. A single colony from the transformed plate was inoculated as primary culture in 10 ml Luria Broth (LB) supplemented with ampicillin (100 µg/ml). 1% of an overnight grown primary culture was inoculated in 1L LB for large-scale purification. The secondary culture was induced with 0.01% rhamnose at $O.D._{0.6}$ and further incubated at 18 °C for 16 hours. The cells were harvested by centrifugation at 5000 rpm and resuspended in lysis buffer (20 mM Hepes, pH-8, 300 mM NaCl, 10% Glycerol, 10 mM imidazole, 1mM PMSF). The cell suspension was lysed by sonication at 30% amplitude for 3 cycles (on; 2s, off; 6s) following centrifugation at 14,000*g*. The supernatant collected, was then subjected to

immobilized metal affinity chromatography (IMAC) wherein the purification of *Mt*Fic was facilitated by the hexa histidine tag present at the N-terminus of Sumo tag. The protein was eluted in buffer (20 mM Hepes, pH-8, 300 mM NaCl, 10% glycerol) containing 80 mM – 200 mM of imidazole. The eluted fractions were pooled and dialyzed against buffer (20 mM Hepes-pH 8, 300 mM NaCl, 10% Glycerol). *MtGyr*B was purified using affinity chromatography (Gupta et al., 2016).

3.2.5 Protein characterization

(a) Circular dichroism (CD) studies:

CD spectrum of purified recombinant protein was carried out in 260 nm-195 nm ranges to determine secondary structure elements after refolding. CD experiments were performed on a JASCO-J715 polarimeter in a 0.2 cm pathlength cuvette, with a slit width of 1 nm, response time of 4 s and a scan speed of 50 nm/s. Each data point was an average of four accumulations.

(b) Fluorescence spectra measurement:

Fluorescence spectra of purified recombinant protein were measured for both refolded and denatured protein (5 µM) in the presence of 6 M GdnCl. The emission spectrum was recorded by exciting the protein at 295 nm and recording the emission in 300 nm - 400 nm range.

3.2.6 Activity assay

For *in-vitro* AMPylation assay, 5 µg of purified Sumo-tagged *Mt*Fic recombinant protein was incubated for 30 min at 25 ºC with 5 µg of recombinant substrates with 1µCi α-^{32}P radiolabeled ATP in reaction mix of 20 mM Hepes, pH-8, 150 mM NaCl and 1 mM TCEP. Reaction was stopped by the addition of SDS loading dye followed by resolving the samples on 12% SDS PAGE. The gel was dried, exposed to X-ray film and visualized by autoradiography using a Phosphorimager (Biorad). *In vitro* AMPylation reaction with fluorescent ATP (N6-(6-amino)hexylATP.5FAM) was performed as described above. The AMPylated protein

bands were visualized by fluorescent gel imaging by excitation at 492 nm for FAM and emission at 517 nm. AMPylation reaction was also monitored by western blots using Anti-AMPylated threonine and tyrosine antibodies (Millipore).

3.2.7 ATPase assay

MtGyr B activity assay was performed followed by AMPylation. In brief, the recombinant *MtGyr*B (50 μM) was incubated with AMPylation reaction mix with 5 mM cold ATP for 2 hours. Subsequently, it was dialyzed in eppendorf dialysis tube (30-kDa) to remove unreacted ATP and cleaved *Mt*Fic protein. The ATPase activity was performed with the wild type and AMPylated *MtGyr*B (*MtGyr*B^{AMP}). The reaction was carried out in 50 mM Hepes buffer, pH-8, 0.2 mM $MgCl_2$, 1mM DTT, and 0.5 mg/ml buffer with 5 μM GyraseB and a mixture of 0.2 mM ATP with 5 μ Ci hot ATP as tracer. The reaction mix was incubated for 30 min at 25 °C. 10 mM DTT was added to stop the reaction followed by centrifugation at 14000*g*. The supernatant collected was resolved by thin layer chromatography using PEI-cellulose F TLC sheets (Merck). Calf intestinal alkaline phosphate (CIAP) was used as control. The intensity of released orthophosphate (Pi) was quantified by Multigauge software (Fujifilm) and the relative activity of ATPase was evaluated by comparing the intensities of orthophosphate produced by unmodified *MtGyr*B. Experiments were performed in triplicates.

3.2.8 Dynamic Scanning Fluorimetry (DSF)

Ligand binding experiments were carried out using DSF method in fluorescence microplate reader (iQ5, BioRad iCycler Multicolor Real-Time PCR detection system). Reactions were carried out in 96-well PCR plate in a volume of 25 μl containing 2.5 μM protein, 5 mM $MgCl_2$ in reaction buffer of 20 mM Hepes, pH-8.0, 150 mM NaCl and 1 mM TCEP. 5X SYPRO orange dye was added to the reaction mix and incubated in the RT PCR device. Fluorescence intensities were measured at intervals of 0.5 °C over a temperature range of 25 °C to 90 °C with 0.5 °C per min dwell time. Fluorescence intensities for triplicates were plotted as a

function of temperature subsequently data points were normalized as fractions from 0-1 (0; folded state, 1; unfolded state). The melting temperatures (T_m) of proteins were determined using Boltzmann sigmoidal dose response with invariable slope (bottom=0, top=1) function in GraphPad Prism 6.

3.2.9 Immunodot blot assay

For interaction studies, immunodot blot assay was performed as described in (Cheng et al., 2014). Varying amount of recombinant *Mt*Gyr B (0.1 – 5 μg) protein was spotted as bait on nitrocellulose membrane and dried at room temperature in vacuum. The membrane was then incubated overnight at 4 °C in 5% skim milk prepared in PBST (phosphate-buffer saline with 0.1% Tween 20). Recombinant Sumo tagged *Mt*Fic protein (10 μg) in buffer; 20 mM Hepes pH-8, 100 mM NaCl, 1mM TCEP and 1mM $MgCl_2$, was used as prey. The blots were washed thrice with PBST and incubated with prey mix for overnight binding at 4 °C with gentle mixing. Subsequently, the membrane was washed thrice with PBST and incubated with primary antibody against Sumo tag (CusBio) at 1:1000 dilutions in PBST with 3% skim milk. The membrane was then washed three times and incubated for an hour with HRP conjugated anti-mouse Ig antibody at 1:10000 dilutions in PBST with 3% skim milk. Finally, the membrane was washed thrice with PBST and signals were detected using SuperSignal West Femto Chemiluminescent Substrate (Thermo scientific). Similar procedure was followed for the dot blot studies of wild type and mutant proteins interaction; where substrates viz. GyraseB and Cdc42 (0.4, 0.8 and 1.2 μg) were spotted on nitrocellulose membrane as baits and Sumo tagged *Mt*Fic and GST tagged VopS native and H348A mutant proteins were used as prey. The immunodot blots were incubated with anti Sumo-tag and anti-GST tag antibodies, respectively. For binding studies of MBP tagged and Sumo tagged *Mt*Fic proteins, 10 μg *MtGyr*B was spotted on two different nitrocellulose membranes as bait and MBP and Sumo tagged proteins were used as prey. Both dot blots were incubated and developed using anti-MBP and anti-Sumo tag antibodies respectively. For the control experiment, the immunodot blots were stripped using mild stripping solution (Sigma)

followed by blocking and immunolabeling with anti-His antibody. In the comparative studies, the separate immunoblots were developed together.

3.2.10 Growth of *E. coli* harboring *Mt*Fic constructs

E. coli BL21 cells harboring Sumo-*Mt*Fic^{WT} and *Mt*Fic^{H144A} constructs and native vector were selected on LB agar plates containing 30 μg/ml kanamycin. Single colony was inoculated from each plate and was grown overnight in LB media with 1% glucose. Cultures were inoculated in fresh LB medium in presence of 1% glucose and were grown for 2 hrs. The cells were washed gently and re-suspended in equal volume of fresh LB medium, induced with 0.05% of L-rhamnose at $OD_{600} \approx 0.4$, 0.8, 1.2 and 1.6 and subsequently were grown at 16 °C. Samples were withdrawn and plated at serial dilutions on LB agar plates with 30 μg/ml kanamycin. Plates were incubated at 16 °C and growth was measured in terms of colony forming units (CFU).

3.2.11 Bacterial-staining

To understand the morphological changes, a secondary culture of *E. coli* cells harboring *Mt*Fic constructs, were grown at 16 °C for 12 hours post-induction with 0.05% L-rhamnose. Cells from 3 ml of culture were pelleted, washed with ice cold 1X PBS pH-7.4 and incubated with DilC16 dye (5 μM; prepared in PBS) for 10 min in dark. The dye was removed by 4-5 consecutive cycles of washing and centrifugation with 1X PBS. The pelleted cells were re-suspended in 20 μl of PBS and dropped on round coverslips. The coverslips were inverted on glass slides and the mounted coverslip was sealed with clear nail polish. The images were acquired within 15 min after slide preparations using Nikon fluorescence microscope with filter for red image acquisition.

3.2.12 Mammalian cells, their maintenance and transient transfection

The mammalian cells used in this study include Human embryonic kidney cells 293 (HEK293) and Human cervical cancer cells (HeLa). Cells were cultured in their prescribed growth condition in Dulbecco's modified Eagle's medium (Invitrogen) containing 10% fetal bovine serum and 1% antibiotic-antimycotic (Sigma). Transfection of the cells with pEGFPC1 construct was carried out using Lipofectamine 2000 (Invitrogen) at $1X10^5$ cells/60 mm plates, as prescribed in the manufacturer's protocol.

3.2.13 Cell Immunofluorescence

HeLa and HEK293 cells, grown on sterile coverslips, were transfected with pEGFPC1 native vector, *Mt*Fic^{WT} and *Mt*Fic^{H144A} mutant constructs. Post-16 hours of transfection, HeLa cells were directly visualized under Nikon fluorescence microscope using 60X objective. Post 16 hours and 36 hours of transfection, HEK293 cells were washed twice with ice cold PBS, fixed with methanol and washed again twice with PBS. The cells were stained with DAPI solution (5 μg/mL) (Sigma) for 5 min, washed thrice with PBS and coverslips were mounted on slides with Fluoroshield (Sigma). Imaging was performed on Leica laser-scanning confocal microscope (Germany) using 63X objective.

3.2.14 Cell fractionation using Centrifugation

Cells were fractionated following the classical differential centrifugation method (Graham, 2002). Briefly, cells were re-suspended in an ice-cold hypotonic buffer, incubated on ice. Cells were disrupted by 10 manual strokes in homogenizer followed by ultra-centrifugation at 4 ºC (Beckman Ultracentrifuge). Centrifugation at 500*g*, 1000*g*, 10,000*g* and 100,000*g* enriched cell debris, nucleus, mitochondria and membrane/ER fractions in the pellet while the supernatant collected after 100,000*g* ultracentrifugation constituted cytosolic enriched fraction.

3.2.15 Far-western assay

The fractionated cell lysates (200 µg) were resolved on 12% SDS-PAGE and immunoblotted on nitrocellulose membrane. The procedure for far western was followed as described in (Wu et al., 2007). In brief, immunoblotting was followed by on membrane denaturation and renaturation using step gradient of 6M, 3M, 1M, 0.1M and 0M GdnCl prepared in buffer (20 mM Tris-HCl pH-7.4, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, 0.1% Tween-20, 2% skim milk and 1 mM DTT). The immunoblot was incubated overnight at 4 °C in GdnCl free buffer. The membrane was then washed three times with PBST and kept overnight at 4 °C for blocking in the buffer with 5% skim milk. For identification of binding partners of Sumo tagged *Mt*Fic, the recombinant protein was used as bait. Blocking was followed by 3 times washing and incubation of membrane with the recombinant protein (20 µg) premix with the buffer with 5 mM $MgCl_2$. The recombinant protein was allowed to bind for 16 hours at 4 °C. Following extensive washing, blots were developed with Sumo tag primary antibody (1:1000), followed by HRP conjugated secondary Mouse IgG. The bands were detected using Super-Signal West Femto Chemiluminescent Substrate (Thermo scientific). The above method was also followed for the recombinant purified putative substrates (His-Cdc42, His-Rac1, and GST-RhoA) and Sumo tagged *Mt*Fic interaction assay where 5 µg of the putative substrates were resolved on SDS-PAGE and were used as prey.

3.2.16 Protein extraction and co-immunoprecipitation assay

The interactions of *Mt*Fic and other identified substrates were confirmed by immunoprecipitation experiments. Total proteins from transiently transfected HEK293 cells, expressing eGFP tagged *Mt*Fic were extracted by lysing the cells in cell lytic M buffer (Sigma) with protease inhibitors (Sigma) and the supernatant was collected after centrifugation at 14000*g*. Similarly, control lysates were prepared with pEGFPC1 vector transfected cells. Protein concentrations were determined using Bradford colorimetric reagent (Sigma). GFP immunoprecipitation was performed using 500 µg of total cell lysate in 20 mM Hepes, pH-8, 150 mM NaCl, 1 mM TCEP

and 5 mM $MgCl_2$. Anti-GFP antibody (CST) was added in the above mix and was allowed to bind overnight at 4 ºC. Next, Protein G agarose (Pierce), pre-equilibrated with the above buffer, was added to the mix and allowed to bind further for 3 hours at 4 ºC followed by centrifugation at 3000 rpm for 3 min. The precipitated beads were washed thrice with PBS containing 0.1% TritonX-100 to remove non-specifically bound proteins. Washed beads were mixed with 5X loading dye and subjected to SDS-PAGE and were immuno-transferred onto nitrocellulose membrane for immuno-detection with respective antibodies.

3.2.17 LC-MS^2 analysis

Identification of substrates was performed by in-gel digestion (Shevchenko et al., 2006) of the corresponding band positions identified in far western experiments, followed by MS-MS analysis. Gel pieces were de-stained with 50 mM Ammonium bicarbonate, 50% acetonitrile solution, dehydrated with 100% acetonitrile and then reduced with 1 mM DTT (Sigma) for 1 hour at 56 °C. Subsequently, the samples were alkylated using 50 mM iodoacetamide (Sigma) in dark for 30 min. The reduced and alkylated samples were digested with 2 µl of trypsin (0.1 µg/µl) with overnight incubation at 37 °C. The digested peptides were extracted using formic acid (5%) and acetonitrile (60%) solution. MS-MS profile was obtained using Bruker system and was used to perform a database search for identification of the proteins, using a MASCOT search tool (Matrix Science Inc., Boston, MS, USA) in BIOTOOLS version 2.2 software (Bruker).

3.2.18 Modeling and structural comparison

*Mt*Fic was modeled using I-TASSER server (Zhang, 2008). A comparison of structures and manual docking of ATP binding site were performed using Pymol (DeLano, 2002). The model structure is visualized using Chimera (Goddard et al., 2005).

3.3. RESULTS

3.3.1 Fic protein (Rv3641c) from *Mtb* belongs to class I Fic

Rv3641c operon encodes Fic (**f**ilamentation **i**nduced by **c**AMP) protein containing HXFX(D/E)GNGRXXR fic motif. Like all fic proteins, Rv3641c is a member of type II toxin-antitoxin (TA) module where both constituting partners of TA system are proteins. A search of the antitoxin partner with (S/T)XXXE(G/N) motif as search criteria, identified Rv3642c. The location of Rv3642c is upstream to the fic (Rv3641c) gene, where both are part of the same operon. Thus, Rv3641c belongs to class I Fic proteins (Figure 3.2) where the obstructive domain for the ATP binding is present in the upstream of the operon. Rv3642c has the inhibitory α-helix (α_{inh}) with a conserved "NTELEG" motif where a conserved glutamate competes with the ATP γ-phosphate binding site.

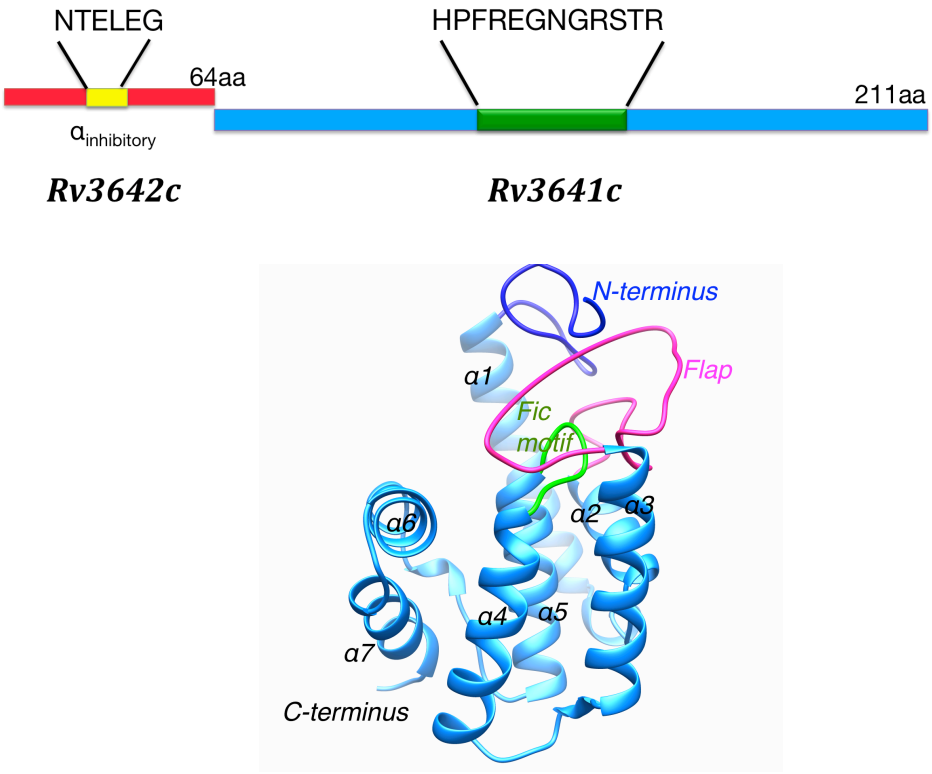
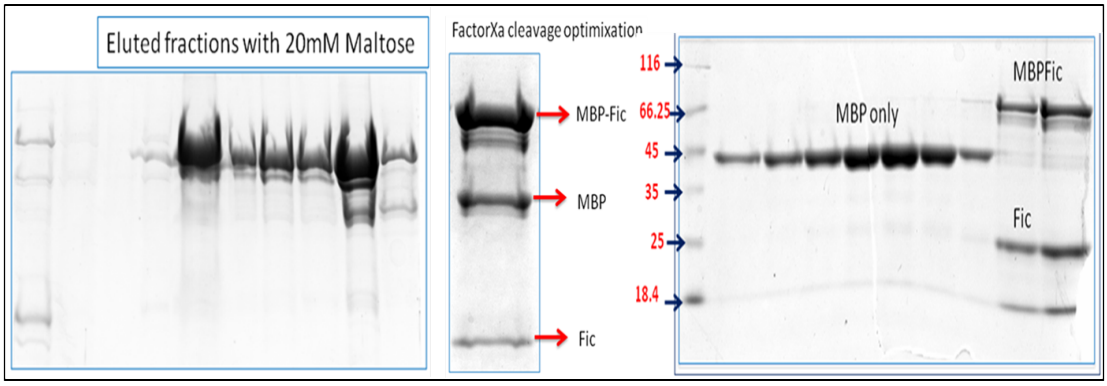


Figure 3.2. A schema representing the location Rv3642c with respect to Rv3641c *(upper panel, from Chapter 2). Modeled Rv3641c protein structure using I-TASSER (lower panel).*

Also, the modeled structure using I-TASSER revealed the conserved Fic fold topology. The predicted 3D-structure of *Rv3641c* was based on VbhT Fic from *B. schoenbuchensis* (PDB ID: 3SHG), as the template containing 27% identity with 94% coverage.

3.3.2 Purification of *Rv3641c*

Rv3641c encodes 211 amino acid protein of molecular weight 24 kDa with theoretical pI of 4.71. The gene was cloned in pET28a(+) (N terminus His-tag), pET22b(+) (six constructs were made with varying combination of deletion at N and C termini; all having hexa-His tag at C-termini), pGEX-4T1 (N-terminus GST tag), pMAL-5px (N-terminus MBP tag for expression in periplasmic space), pMAL-5cx (N-terminus MBP tag for expression in cytoplasm) and Rhamnose pBAD-Sumo (N terminus Sumo-His tag). Rv3641c was also cloned in a pRSF-Duet vector with its antitoxin counterpart (Rv3642c) where Rv3642c occupied the upstream **m**ultiple **c**loning **s**ite I (MCSI) without any tag and Rv3641c at **m**ultiple **c**loning **s**ite **II** (MCSII) with a hexa-His tag at C-terminus. Out of all the constructs, only MBP and Sumo tagged proteins were found to be in soluble fractions (Figure 3.3). However, the recombinant proteins eluted as a higher oligomer in size exclusion chromatography. The higher oligomeric nature was also observed in native and blue native PAGE analysis (Figure 3.3).



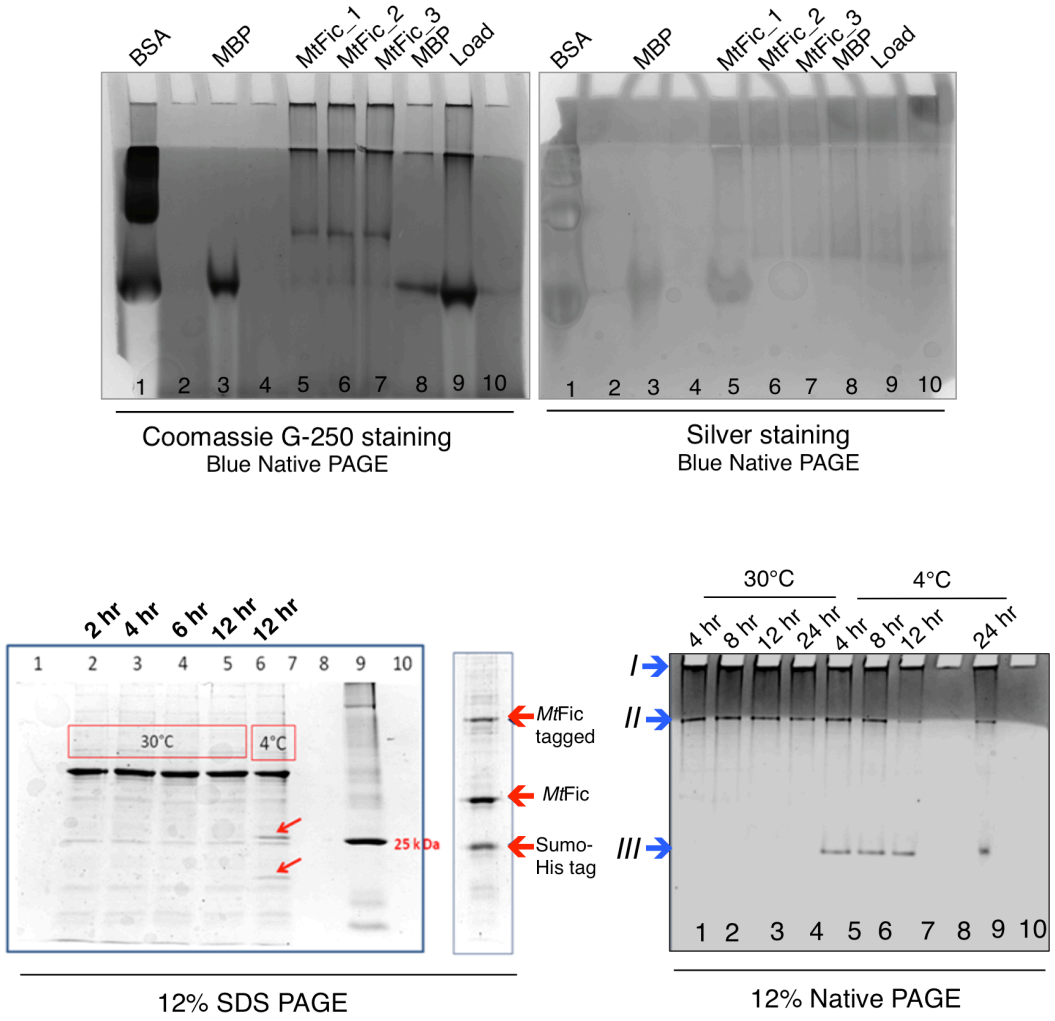


Figure 3.3. Purification and characterization of MtFic protein from soluble fractions. *Upper panel represents purification of MBP-tagged MtFic. The first 12% SDS PAGE shows the purity of eluted fractions from amylose resin using 20 mM Maltose in buffer. Second 12% SDS PAGE lane represents the optimization of MBP tag cleavage at 22 °C with protease utilizing factor Xa cleavage site present between MBP and MtFic recombinant protein, in buffer mix of 20 mM Tris-HCl pH-8 with 200 mM NaCl and 2mM $CaCl_2$. Third 12% SDS PAGE represents eluted fractions from preparative gel filtration showing that the untagged MtFic protein eluted as higher oligomer near void volume fractions when it was tagged with MBP. MBP-tag eluted in elution volume range. The middle panel shows Blue native page analysis of the tagged and untagged oligomeric status, stained with brilliant blue G250 and silver stain. MBP protein band corresponds to dimer (position above 66 kDa*

monomer of BSA). MBP-tagged MtFic showed higher oligomer with most of the protein retained in stacking and resolving interface. BSA protein with monomer and dimer bands was used as molecular weight marker in blue native page analysis. The lower panel represents purification and optimization of tag cleavage of Sumo-tagged MtFic. First 15% SDS PAGE showed purity of purified protein and tag cleavage optimization with Sumo protease at 4 °C and 30 °C. Second 15% SDS PAGE lane represents the optimized tag cleavage profile at 4 °C. Third 12% Blue Native PAGE shows Sumo tag position (III). Positions I and II are present at well-stacking interface and stacking-resolving interface correspond with the tagged and untagged MtFic protein, respectively.

*Mt*Fic was also purified from inclusion bodies to have a recombinant protein with the shortest tag (Hexa-His tag). As the protein was purified through denaturation- renaturation protocol, biophysical characterization of renatured protein was carried out to ensure protein folding. Far-UV CD spectroscopy of protein revealed that recombinant protein was rich in alpha helices. Intact tertiary structure was affirmed by monitoring tryptophan intrinsic fluorescence emission spectrum in presence and absence of denaturant (6M GdnCl). Red shift in presence of denaturant suggests that the unmodified protein contained tryptophan in buried environment implying that the recombinant protein purified through renaturation method has a tertiary structure, representative of a folded protein. However, the renatured protein showed melting temperature of 25 °C in thermal denaturation profile. Therefore, the protein could not be employed for activity and binding assay at or above 25 °C.

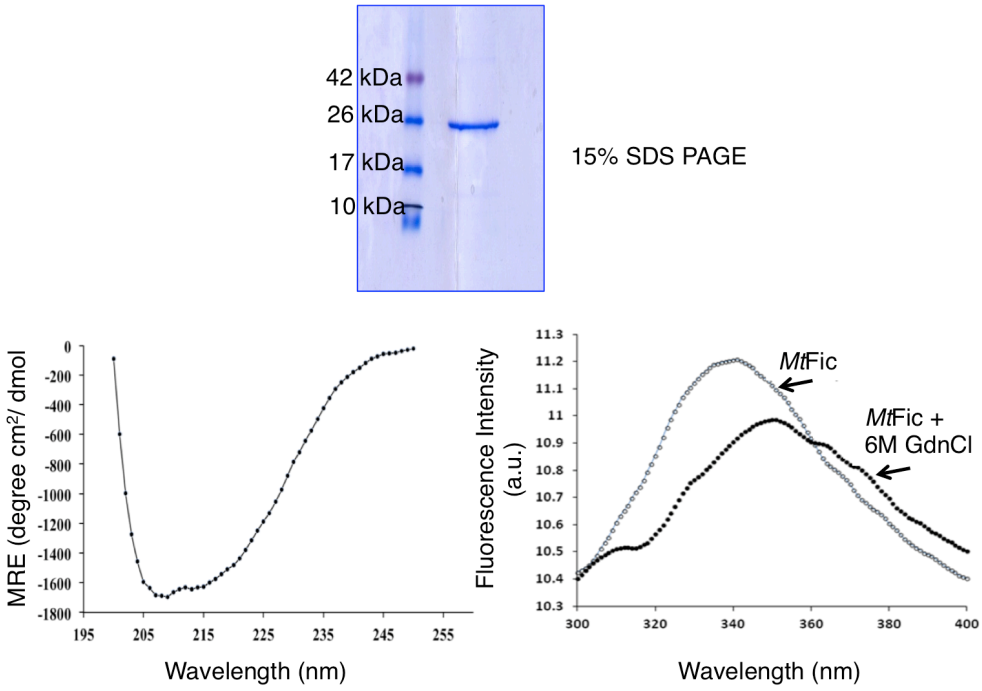


Figure 3.4. Purification and characterization of MtFic protein from insoluble fraction*. 15% SDS PAGE showing purity and molecular weight of renatured His-tagged MtFic protein to be Mr 26,000 (upper panel). Far UV CD spectrum of 10 µM renatured His-tagged MtFic protein (lower left panel). Emission fluorescence spectra of purified protein, before and after denaturation with 8M GdnCl (lower right panel).*

3.3.3 Fic protein from *Mtb* perturbs cell growth in *E. coli* and causes elongation of the cells

Fic proteins in pathogenic *Mycobacterium species* are present as single copy gene and contain Fic domain which lacks the variable termini found in other Fic domain-containing proteins from non-pathogenic or opportunistic pathogenic strains (Chapter 2). To understand the role of Fic protein from *Mtb*, Sumo tagged native and invariant His to Ala mutant *Mt*Fic proteins were expressed in *E. coli* BL21 cells in presence of inducer L-rhamnose. The expression of the functional protein in soluble fractions was achieved by inducing the *Mt*Fic constructs harboring *E. coli* culture at 16 °C. Expression of the *Mt*Fic protein resulted in the reduction in growth, shown by reduced CFU count in *E. coli* cultures, induced at different optical density. On the

other hand, the vector control and *Mt*Fic^{H144A} mutant construct expressing *E. coli* BL21 cells did not show any perturbation in their growth (Figure 3.5). The expression of the wild-type *Mt*Fic protein at the given inducer concentrations was also verified by immunoblotting with anti-Sumo antibody (Figure 3.5). To correlate the differential CFU at equal optical density of bacterial culture, *Mt*Fic expressing *E. coli* cultures were stained with red-orange fluorescent, lipophilic DilC16 dye. The expression of the *Mt*Fic native protein resulted in elongated cellular morphology (Figure 3.5). The average length of *Mt*Fic expressed *E. coli* cells were ≈10 µm whereas bacterial cells transformed with their active site mutant, H144A, were in the range of 3-5 µm (Figure 3.5). Some *Mt*Fic^{WT} transformed culture (5%) had bacterial lengths that were ≥30 µm long (Figure 3.5). These results thus indicate that *Mt*Fic is a functional toxin that inhibits growth and induces elongated morphology upon overexpression in *E. coli*.

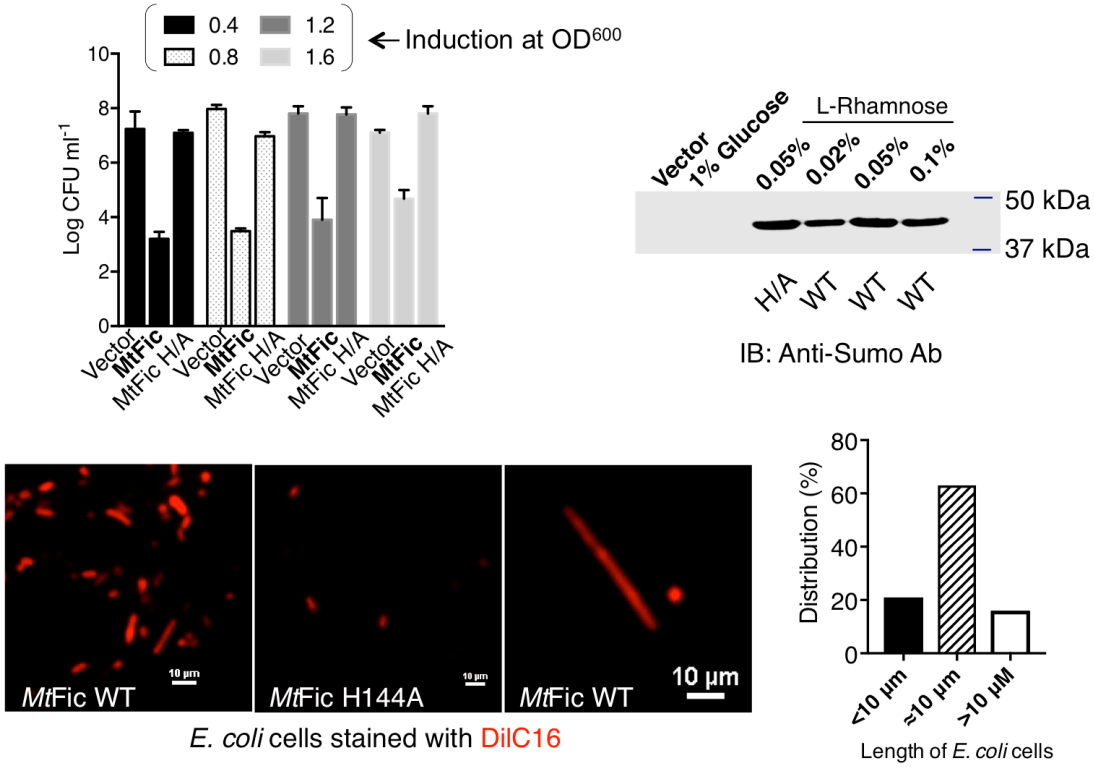


Figure 3.5. MtFic expression in surrogate E. coli BL21 bacterial cells perturbs growth and elongates the cells. *E. coli cells harboring MtFic^{WT} and MtFic^{H144A} constructs, induced for 12 hours at 16 °C at different growth phases, showed reduced*

CFUs in MtFic^{WT} expressing cells whereas catalytically inactive MtFic^{H144A} mutant does not inhibit growth (upper left panel). Western blot showing MtFic protein expression checked at varying L-rhamnose concentrations and developed with anti-Sumo tagged antibody (upper right panel). Fluorescence micrograph of DilC16 stained E. coli cells expressing MtFic^{WT} protein displayed elongated morphology (lower left panel). Bar graph representing the distribution of approximate bacterial population with reference to their length, in MtFic expressing E. coli cells (lower right panel).

3.3.4 Binding of MtGyrB with recombinant purified *Mt*Fic proteins

Members of class I Fic proteins characterized till date viz. VbhT (*Bartonella schoenbuchensis*), fic-1 (*Pseudomonas fluorescens*), *Ye*Fic (*Yersinia enterolitica*) and *Pa*Fic (*Pseudomonas aeruginosa PAO1*) have the propensity to inhibit the growth of *E. coli* strains by post-translationally modifying and abolishing the activity of DNA gyrase subunit B and ParE. *Mt*Fic is homologous to the above Fic proteins sharing sequence identity in the range of 23.76-32% (Figure 3.6).

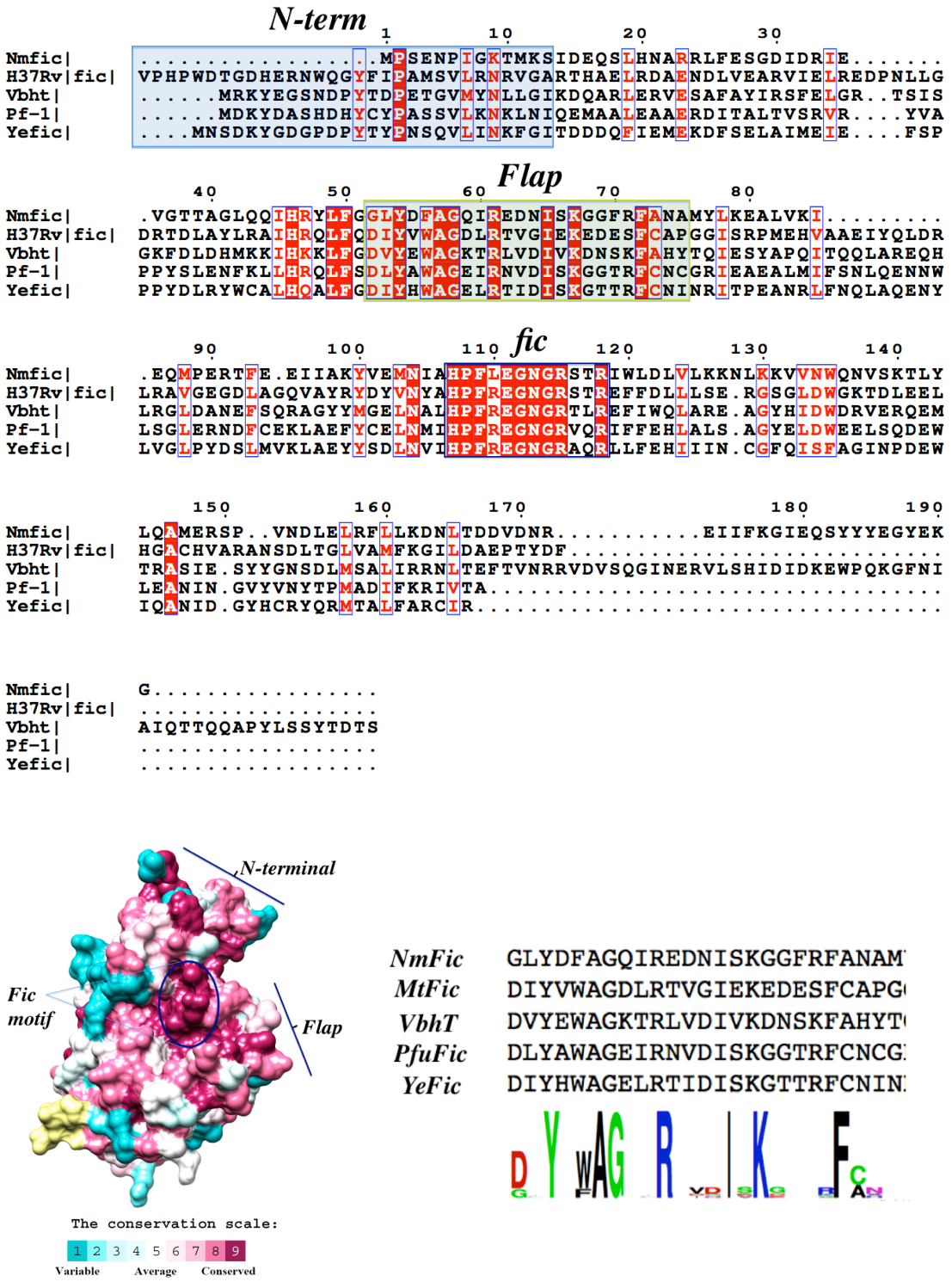


Figure 3.6. MSA of MtFic *with its homologs using Clustal Omega represented using ESPript 3.0 (Upper panel). Structural representation of amino acid conservation in class I Fic using Consurf server. The residues are color coded on the basis of their*

conservation score (lower left panel). A sequence logo generated using the WebLogo server to represent the conservation in Flap region (lower right panel).

As the substrate binding regions described by the flap and the Fic motif are conserved among them (Figure 3.6) and *Mt*Fic also inhibited growth in *E. coli*, we checked the binding of *Mt*Fic to *Mtb* type II topoisomerase, which constitutes the only DNA gyrase present in this organism. *In vitro* binding experiments by dot blot assay showed that Sumo tagged *Mt*Fic (bait) could bind to *MtGyr*B subunit (prey) in the range of 100ng - 5µg protein concentrations, substantiating DNA Gyrase B subunit as a binding partner of *Mt*Fic in *Mtb* (Figure 3.7).

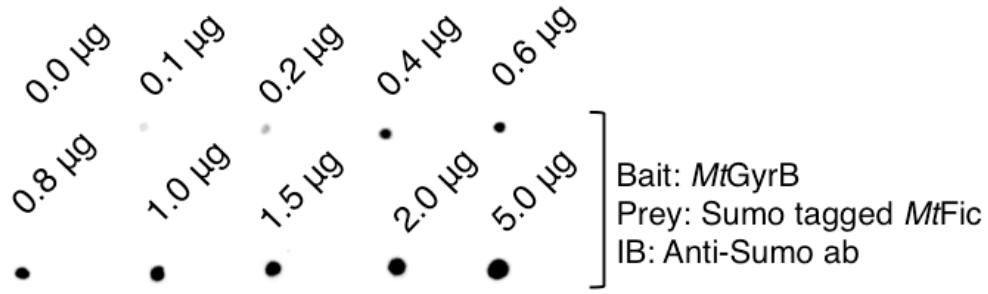


Figure 3.7. Dot blot assay representing MtFic binding with varying amounts of MtGyrB *wherein MtGyrB was spotted on nitrocellulose membrane as bait and MtFic proteins with Sumo tag was used as prey, subsequently, the blots were developed with anti-Sumo antibody.*

We also checked the binding of MBP tagged *Mt*Fic protein which has been known to hydrolyze nucleotide triphosphates (ATP, GTP, CTP, and UTP). However, it did not show binding to *MtGyr*B subunit at equimolar concentration (Figure 3.8). As purified Sumo tagged *Mt*Fic (T_m 49 °C) and MBP tagged *Mt*Fic (T_m 60.8 °C) were stable in reaction mix (Figure 3.8), the inability of MBP-tagged *Mt*Fic to bind with cognate substrate could be attributed to high molecular weight MBP tag (45 kDa), which may mask the substrate binding sites of the enzyme.

In *Mtb*, TA module viz. parDE also binds with DNA gyraseB subunit and alters the cellular morphology of bacterial cells (Gupta et al., 2016). The binding analysis with crucial *Mt*Fic^{H144A} mutant protein was studied to explore whether

binding alone is responsible for the morphological changes. VopS^{WT} and its invariant histidine mutant (VopS^{H348A}) proteins were used as positive control. In dot blot experiments, both *Mt*Fic^{WT} and *Mt*Fic^{H144A} mutant proteins showed binding with their cognate partner *Mt*GyrB (Figure 3.9), suggesting that the ability of *Mt*Fic to bind *MtGyr* subunit B alone is not sufficient to explain the changes in morphology of *E. coli* overexpressing *Mt*Fic.

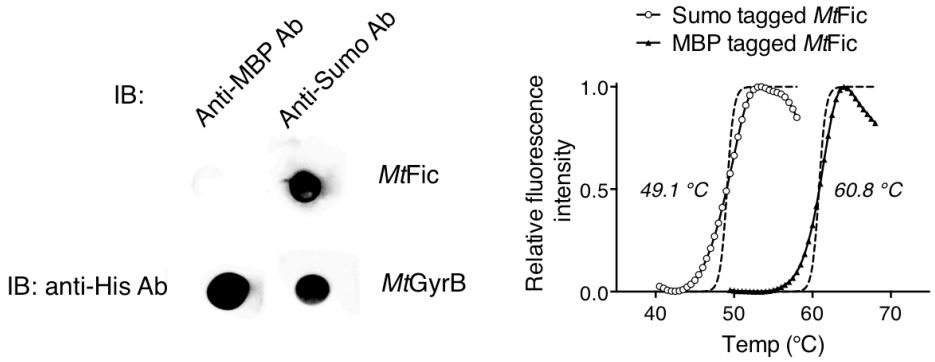


Figure 3.8. Dot blot showing binding of Sumo-tagged MtFic with MtGyrB *wherein MtGyrB was spotted on nitrocellulose membrane as bait and MtFic proteins with Sumo and MBP tags were used as prey. Blot developed subsequently with the anti-His antibody (left panel). DSF profile, representing T_m of recombinant MtFic proteins (Sumo tagged and MBP tagged) (right panel).*

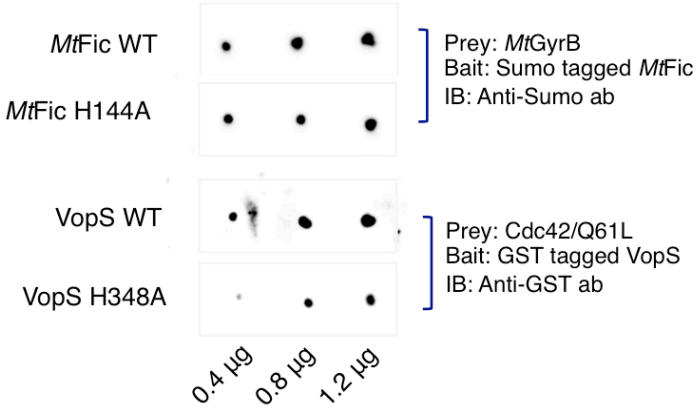


Figure 3.9. Dot blot showing that both MtFic^{WT} and MtFic^{H144A} inactive mutant interact with MtGyrB. *A similar observation was marked for VopS (WT and H348A) control proteins for their cognate substrate Cdc42.*

3.3.5 *Mt*Fic AMPylates MtGyrase B subunit and inhibits its ATPase activity

Further, AMPylation reaction was performed using *Mt*Gyr subunit B as a substrate. In activity assay, shown in figure 3.10, *Mt*Fic AMPylated *Mt*GyrB. Also, AMPylation of *Mt*GyrB reduced its ATPase activity (Figure 3.10). Since AMPylation occurs in its flap region, which is essential for its ATP binding ability, the loss of the ATPase activity of GyrB could be readily explained (Harms et al., 2015; Lu et al., 2016).

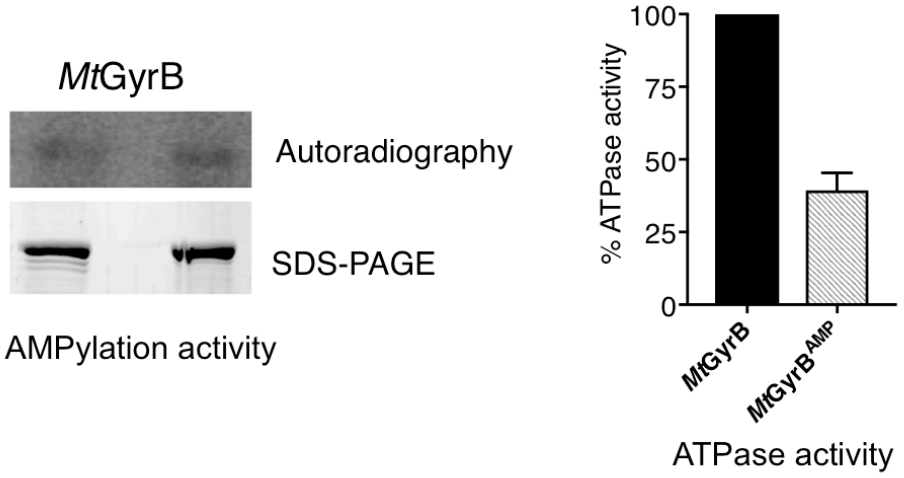


Figure 3.10. MtFic AMPylates MtGyrB and inhibits its ATPase activity. *(A) Autoradiograph of AMPylation reaction with MtFic and MtGyrB using α-32P ATP. Duplicate samples (lane 1 & 3) show AMPylation of MtGyrB by MtFic enzyme. (B) Bar graph representing relative change (in %) of ATPase activity of un-AMPylated (MtGyrB) vs AMPylated MtGyrB (MtGyrB^{AMP}).*

3.3.6 Multiple interacting partners in *M. smegmatis* cell lysate

To find out whether Gyrase is a specific target of *Mt*Fic or if there are multiple targets, far western analysis with *Mycobacterium* cell lysate was performed. The blots reveal that *Mt*Fic specifically binds to substrates of molecular weight Mr 28000, 32000, 45000, 55000 and 68000 (Figure 3.11). The far western result with *M.*

smegmatis cell lysates indicates that the enzyme has several interacting partners, however, their validation and identification in *Mtb* system needs further exploration.

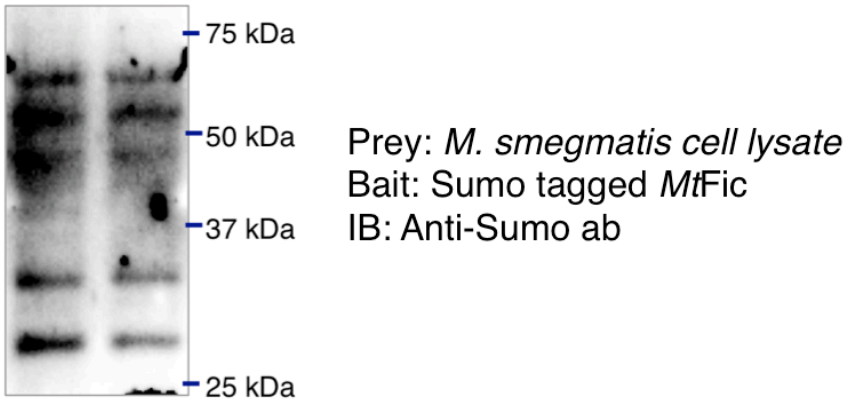


Figure 3.11. Far western blot marking molecular weight positions of MtFic interacting partners in M. smegmatis cell lysate. *M. smegmatis cell lysate was used as prey, and Sumo tagged MtFic was used as bait. Far western blot of two lanes with varied M. smegmatis lysates (100 µg and 200 µg) are shown. The blots were developed with anti-Sumo tag antibody.*

3.3.7 Expression of *Mt*Fic in HeLa cells results in cell rounding

*MtGyr*B is the physiologic substrate for *Mt*Fic wherein AMPylated GyrB showed reduced ATPase activity. To understand its role in pathogenicity, the effect of *Mt*Fic on the mammalian cell was investigated. To this end, *Mt*Fic^{WT} and *Mt*Fic^{H144A} were cloned in pEGFPC1 vector followed by transient transfection in HeLa cells subsequently examined for eGFP fluorescence and cellular morphology by fluorescence microscopy. Transient transfection of eGFP-*Mt*Fic^{WT} in HeLa cells showed remarkable changes in the cellular morphology. Transfected HeLa cells became rounded and easily detached from the substratum, observed at post 16 hours of transfection (Figure 3.12). As the *Mt*Fic^{H144A} mutant transfected cells did not show any alteration in their morphology and viability (Figure 3.12), the morphological effects are attributed to the AMPylation activity of *Mt*Fic. To check whether cell rounding observed is a consequence of the blocking of downstream signaling and

cytoskeletal collapse, we performed far western blots for binding wherein purified RhoA Q63L (GST tag), Rac1 Q61L (His tag), Cdc42 Q61L (His tag) and Rab5 (GST tag) were used as prey and purified Sumo-tagged *Mt*Fic was used as bait and immunoblotting was performed with anti-Sumo antibody. *Mt*Fic showed binding with Rac1 and Cdc42 proteins, but not with RhoA and Rab5 GTPase (Figure 3.12). Subsequently, *in vitro* AMPylation assay for the activity of the *Mt*Fic enzyme with Rac1 Q61L and Cdc42 Q61L was performed. *Mt*Fic showed weaker AMPylation activity with Rac1 and Cdc42 GTPase as compared to that of VopS used here as a positive control (Figure 3.12).

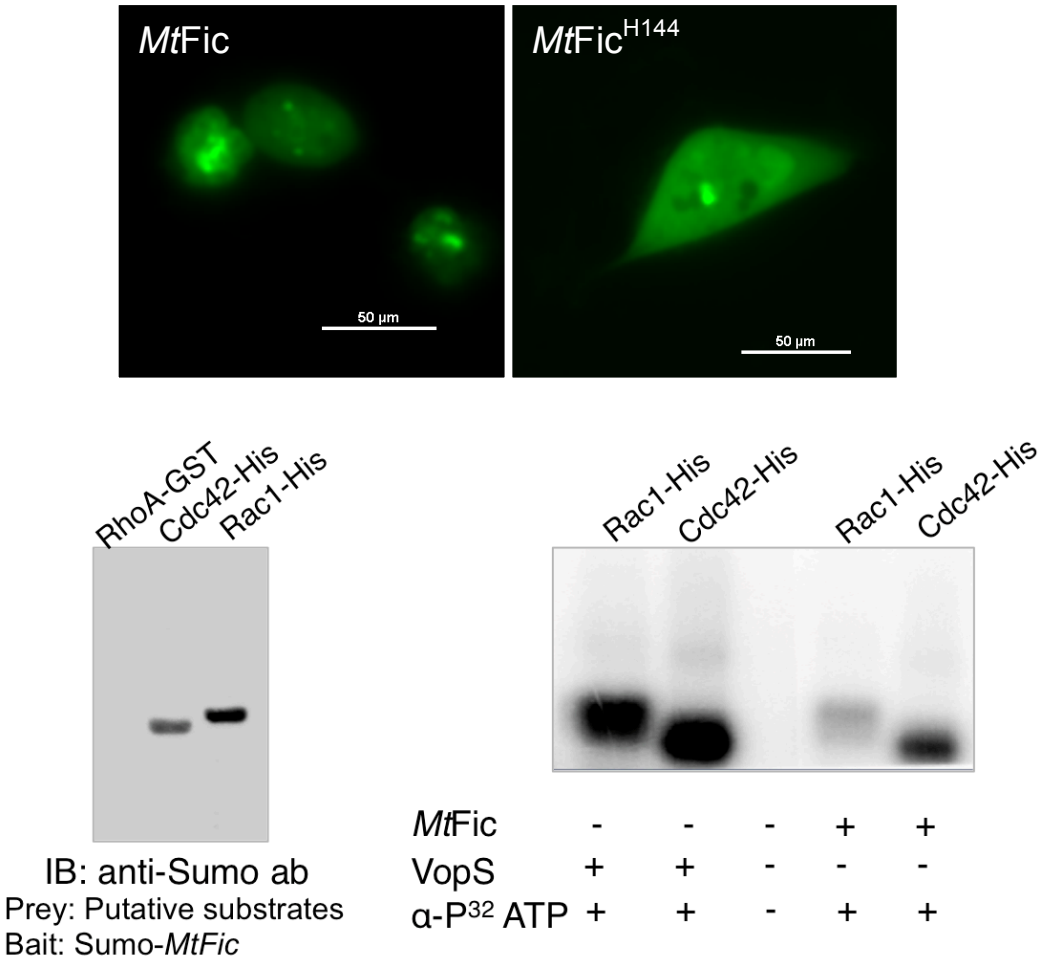


Figure 3.12. MtFic expression in HeLa cells alters morphology. *Fluorescence micrograph showing eGFP- MtFic^{WT} and MtFic^{H144A} mutant proteins showing cell rounding in HeLa cells (Upper panel). Far western blot assay for MtFic binding with*

recombinant RhoA, Cdc42, Rac1 and Rab5 proteins (lanes; 1, 2, 3) (lower left panel). Autoradiograph showing Rac1 and Cdc42 AMPylation with Sumo-tagged MtFic (lanes; 4, 5) and VopS WT activity as controls (lanes; 1, 2) (lower right panel).

In order to ensure that the cell morphological changes and death was not mammalian cell type dependent, the effect of *Mt*Fic expression in another mammalian cell line, HEK293 cells, was analyzed. The observations were conducted in two stages. Initially, at 16 hours, *Mt*Fic-eGFP shows a cytoplasmic localization with a dense spot around the nucleus (Figure 3.13), indicating its localization in a compartment. However, the spots increase with time and merge with DAPI stained nuclear compartment at later stages (36 hours) (Figure 3.13), suggesting localization of *Mt*Fic in the nuclear compartment at later stages of expression.

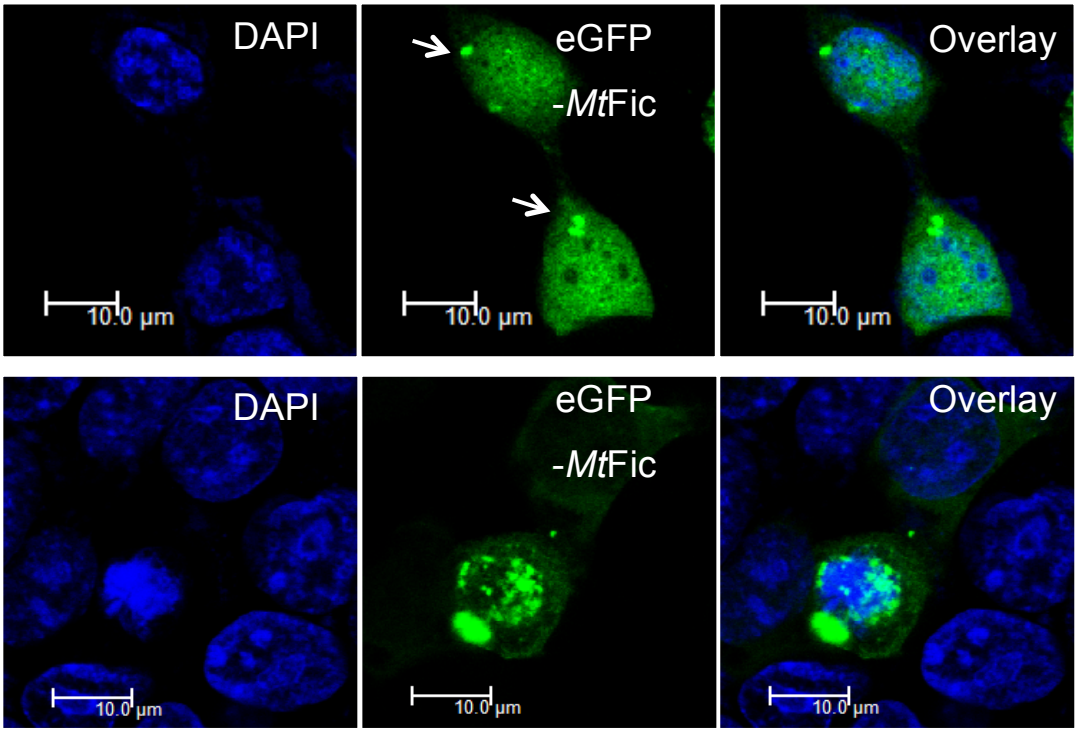


Figure 3.13. MtFic expression in HEK293 cells alters the morphology and shows a distinct expression pattern. *Confocal micrograph for expression pattern of eGFP- MtFic in transfected HEK293 cells post 16 hours of transfection; spotted expression is shown by arrow (white) (upper panel) and post 36 hours of transfection (lower panel).*

3.3.8 *Mt*Fic localizes in ER lumen in initial stages of expression

In transient transfection experiments, we observed a spotted localization of *Mt*Fic protein around the nucleus. We sought to determine whether the dense spot correlates with its localization in a compartment-surrounding nucleus. To check that, the endoplasmic reticular membrane was marked with RFP tagged Sec61β, an ER membrane protein. Co-transfection of Sec61β and *Mt*Fic in HEK293 cells revealed that Sec61β was surrounding *Mt*Fic spots, suggesting its localization in ER lumen (Figure 3.14). Moreover, in co-transfected cells, ER reticular pattern was more clumped with loss of surrounding boundary of the nucleus, though in Sec61β transfected cells, RFP-Sec61β marked the ER reticular pattern effectively surrounding the nucleus.

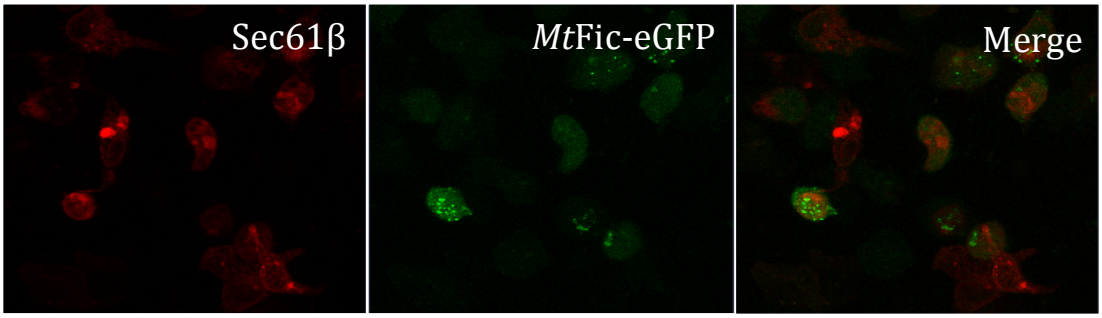


Figure 3.14. Dense spot localizes in ER lumen and nucleus. *RFP-Sec61β is shown in red and MtFic-eGFP is shown in green.*

3.3.9 *Mt*Fic disrupts nuclear Lamin B1, a marker for nuclear envelope

To visualize the effect of *Mt*Fic on the nucleus, the nuclear envelope was marked with immunofluorescence method using an antibody against lamin B1 and fluorescence micrograph was visualized post 16 hours of transient transfection. Interestingly, the un-transfected cells were stained with lamin B1 antibody marking the lamin B1 encircling the DAPI stain (nucleus) (Figure 3.15). However, the nuclear periphery could not be marked in transfected HEK293 cells (both transfected and un-transfected cells are shown in the same field).

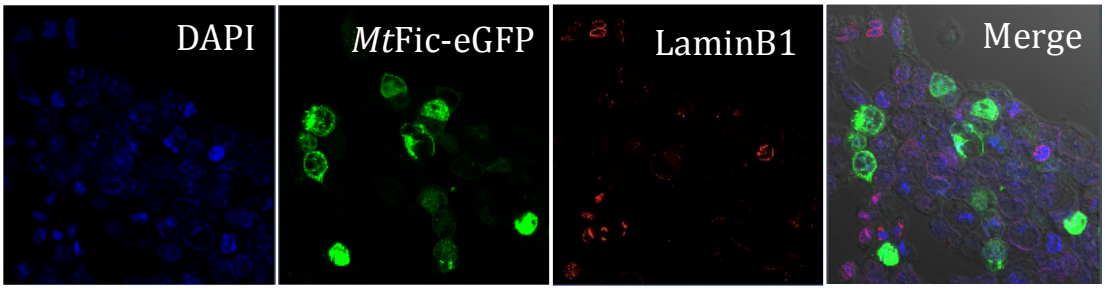


Figure 3.15. Staining of Lamin B1 in transfected HEK293 cells. *Lamin B1 is shown in red, nucleus with DAPI is colored blue and MtFic-eGFP is shown in green.*

3.3.10 *Mt*Fic expressed in HEK293 cells was active

Spotted morphology in ER could be either due to functional localization signal or aggregated protein that needs to be degraded via cargo. The spotted morphology also appeared in HeLa cells transfected with *Mt*Fic^{H144A} mutant where the detachment, rounding of cells and the deleterious effects of *Mt*Fic activity were not observed. This, in turn, suggests that the *Mt*Fic expressed in the mammalian cells must be due to its catalytic activity rather than due to its non-specific aggregation. To affirm this, AMPylation activity assay was performed using cell lysate prepared from transfected cells. Equal concentration of cell lysate from lipofectamine untransfected HEK293 cells was used as a control. Subsequently, the activity assay was performed with cell lysates in presence of a fluorescent ATP analogue, N6-(6-amino) hexyl-ATP-5-FAM (FAM-ATP), followed by separation of bands on 12% SDS-PAGE. Fluorescent imaging ($\lambda_{excitation}$ at 492 nm and $\lambda_{emission}$ at 517 nm) of the gel led to the identification of two intense band positions corresponding to Mr 72000 and 35000 in *Mt*Fic-eGFP expressed cell lysate (Figure 3.16). Lysate from HEK293 cells transfected with control vector, pEGFPC1, served as a negative control. Altogether, the activity assay confirms the activity of *Mt*Fic and identifies two substrate positions, which did not appear in control lane.

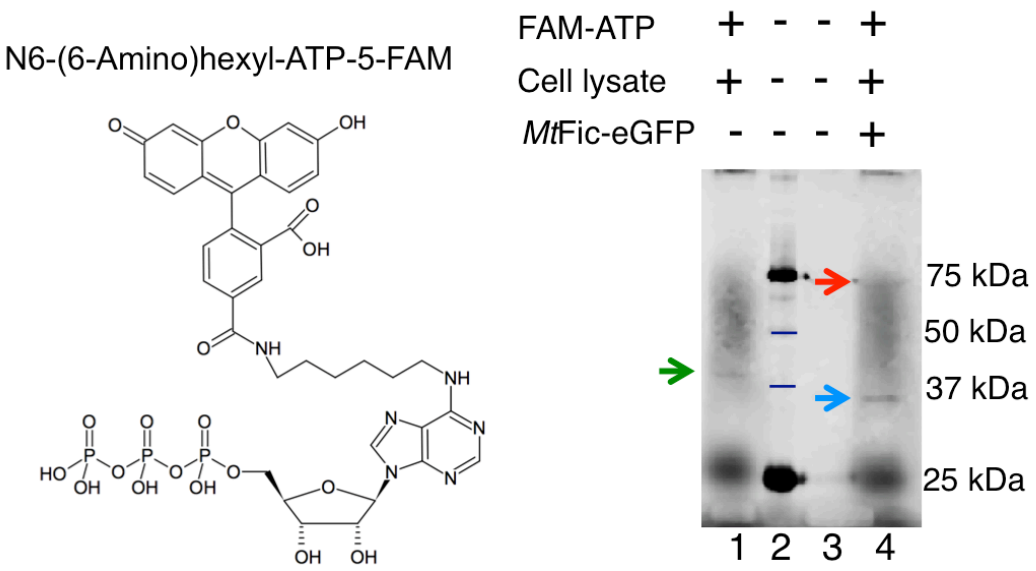


Figure 3.16. MtFic AMPylates at specific positions in HEK283 cell lysate*. Fluorescent gel image of SDS-PAGE post-AMPylation assay with FAM-ATP (Ex 492 nm and Em 517 nm) as nucleotide, in the presence of cell lysate from transiently transfected HEK293 cells having both substrates and eGFP tagged MtFic enzyme. Mr 40000 position appeared in control cell lysate (lane1; green arrow) while Mr 72000 (red arrow) and 34000 (blue arrow) developed in MtFic-eGFP expressed cell lysate (lane 4). A SDS-PAGE marker is shown in lane 2.*

3.3.11 Identification of interacting partners for *Mt*Fic in mammalian cell lysates

Identification of interacting partners in mammalian cells is crucial to understand the role of *Mt*Fic in *Mt*b pathogenicity. The purified *Mt*Fic was difficult to handle because of low yield, auto-tag cleavage propensity and instability of tag cleaved protein. Therefore, for *in-vitro* detection of the interacting partners, we probed the interactions by far western blotting method where all the steps were conducted at 4 °C. HEK293 cell lysate was fractionated by ultracentrifugation method, and each of the enriched fractions (nuclear, mitochondrial, ER/membrane and cytoplasmic enriched fractions) was used as prey and Sumo tagged *Mt*Fic was

used as bait followed by immuno-blotting with anti-His tag antibody. As shown in Figure 3.17, substrate positions were identified in ER and membrane-enriched fractions, with the molecular weight ranging from Mr 16,000 to 55,000. Two molecular weight positions Mr 18,000 and 34,000 were noted to emanate from the enriched fraction of nucleus. *Mt*Fic interacting proteins with Mr 50000, 34000, 25000 and 22000 in the cytoplasmic fraction were detectable at higher concentration (>200 µg). Far western blot thus indicates that *Mt*Fic has interacting partners of Mr 18000, 25000, 34000, and 55000 in ER/membrane, nuclear and cytoplasmic fractions of HEK293 cell lysate.

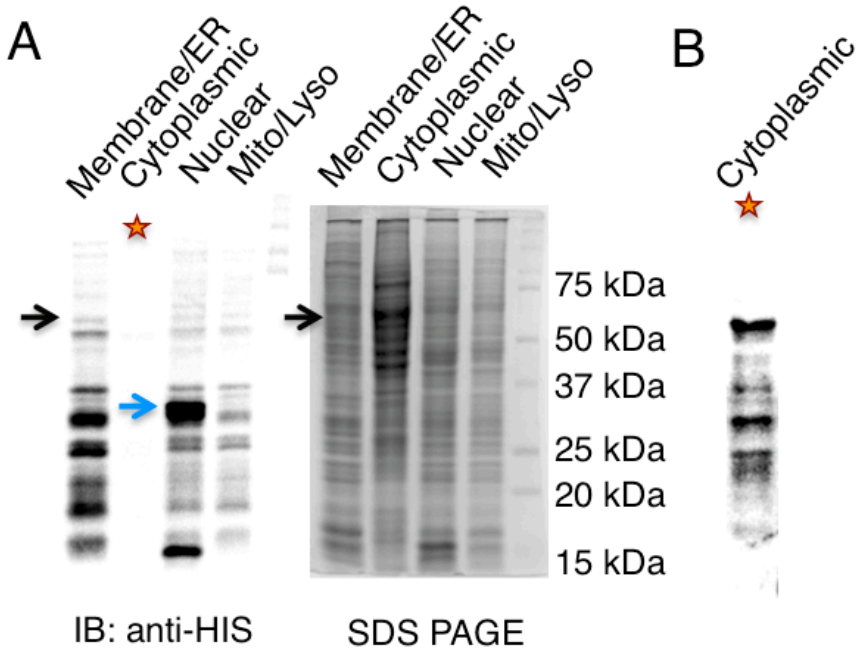


Figure 3.17. MtFic interacts with proteins in HEK293 cell lysate. *Far western blot showing molecular weight positions (Mr) of interacting partners in fractionated HEK293 cell lysates (bait), cytoplasmic fraction (bait), incubated with Sumo-tagged MtFic (prey), subsequently developed with anti-His antibody. Cytoplasmic fraction was developed separately (shown by star). Duplicate samples of fractionated cell lysate resolved on the same SDS-PAGE and arrow (black) shows the band positions for mass spectrometry analysis.*

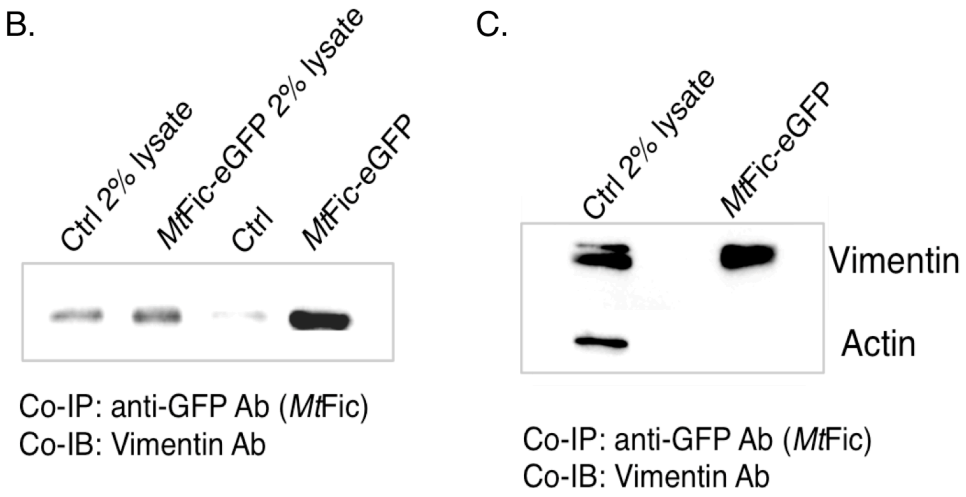
3.3.12 Vimentin identified as potential target of *Mt*Fic

In the fluorescent gel image of AMPylation reaction, the molecular weight range of Mr 40000-70000 appeared as a smear and hence we processed the molecular weight position coinciding with the prominent bands around Mr 50,000, observed in the far western analysis, for its identification by MS^2 analysis. Mascot search identified vimentin (3 unique peptides) with confidence (Figure 3.18A). We confirmed the interaction also by co-immunoprecipitation (Co-IP) with anti-GFP antibody followed by immunoblotting with vimentin antibody (Figure 3.18B). As vimentin is an intermediate filament and is enriched in cellular lysates, specificity of the GFP antibody in the co-IP experiment was confirmed by checking a lack of interaction with actin, another cytoskeletal element, in the immuno-precipitated sample (Figure 3.18C).

A.

Protein	No of unique peptides
Vimentin	3

LLQDSVDFSLADAINTEFK
TNEKVELQELNDRFANYIDKVR
NLQEAEEWYKSKFADLSEAANRNNDALR



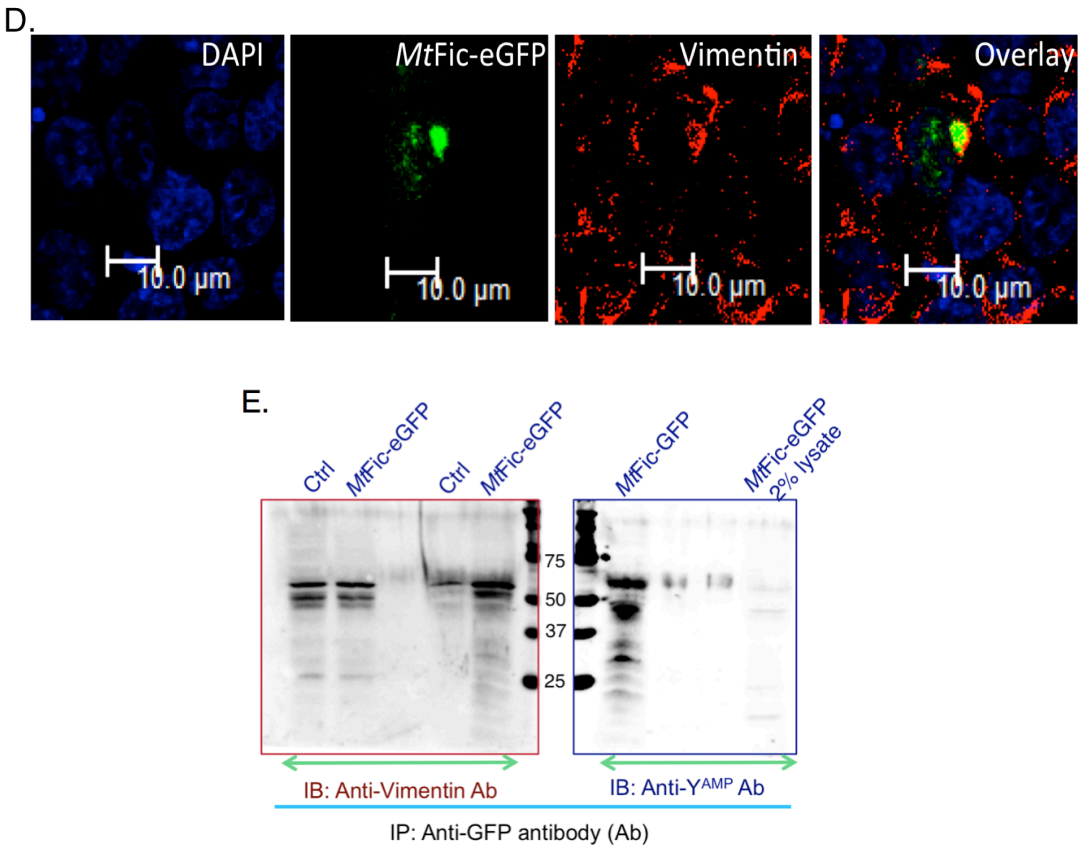


Figure 3.18. MtFic interacts Vimentin. ***(A)*** *Mr 50k-55k band of far western in LC-MS^2 analysis corresponds to vimentin with three unique peptide matches.* ***(B)*** *Western blots of co-IP samples with anti-GFP antibody followed by probing with anti-vimentin antibody lanes; (1) 2% cell lysate of control sample (without MtFic- eGFP protein) (2) 2% cell lysate of MtFic-eGFP expressed cell lysate (3) co-IP sample from control lysate with anti-GFP ab (4) co-IP sample from MtFic-eGFP expressed cell lysate with anti-GFP ab.* ***(C)*** *Co-IP control experiment where co-IP samples were probed with vimentin and subsequently with actin antibody. Lanes; (1) 2% cell lysate of MtFic-eGFP expressed cells (2) co-IP sample from MtFic-eGFP expressed cell lysate with anti-GFP ab.* ***(D)*** *Confocal images probed for vimentin, DAPI and eGFP tagged MtFic.* ***(E)*** *Western blot of co-IP samples with anti-vimentin and anti-Y^{AMP} antibodies.*

To demonstrate the interaction in a cellular system, co-localization studies were performed. Confocal images showed that the dense *Mt*Fic-eGFP spot is surrounded by immunolabeled vimentin puncta, thus representing partial co-

localization in the periphery of the dense spot (Figure 3.18D). Vimentin has been identified as a substrate of Bep2 Fic protein wherein the enzyme AMPylates a tyrosine residue from vimentin substrate. Bep2 also belongs to class I Fic protein and is the closest homolog to *Mt*Fic. Henceforth, the AMPylation activity of *Mt*Fic was probed by western blot method using anti-tyrosine AMPylation (Y^{AMP}) antibody. The band position corresponding to vimentin molecular weight was identified by anti-Y^{AMP} antibody in the co-IP sample (Figure 3.18E). Taken together, the data indicate that vimentin is a potential substrate of *Mt*Fic. However, identification of vimentin needs validation by *in vitro* AMPylation activity assay.

3.4 DISCUSSION

After 38 years of the discovery of adenylylation, reports appeared on AMPylation activity of VopS from *Vibrio parahemolyticus* (Yarbrough ML. *et al*., 2009) and Ibps from *Histophilus somni* (Worby CA., *et al*., 2009) uncovering the significance of this PTM as a strategy evolved in pathogenic bacterial effectors where AMP transfer on switch I region of Cdc42, Rac1 and Rho small GTPase blocks the binding of downstream effectors, resulting in perturbed host signaling events. This study highlights the characterization of Rv3641c (*Mt*Fic) from the pathogen *Mtb,* with respect to its AMPylation activity in the bacterial and mammalian system. *Mt*Fic belongs to class I Fic where the Rv3642c has a pI of 12 with the inhibitory motif NTELEG and Rv3641c (*Mt*Fic) with pI of 5.6 contains the Fic motif HPFREGNGRSTR. *In-vitro*, the purified *Mt*Fic is highly unstable therefore its induction in *E. coli* cells leads to its expression in insoluble fractions. To achieve the soluble expression of *Mt*Fic, the gene was cloned under pBAD promoter in p-Rhamnose expression system and induced at a lower temperature. The expressed protein had Sumo-tag at its N-terminus, as a solubilization tag, followed by His-tag for its purification. Reduced CFU at an equal optical density in *E. coli* cells expressing $MtFic^{WT}$ alone shows that the Sumo tagged recombinant protein is active and is a functional toxin in bacterial cells. The transformation by the wild-type *Mt*Fic led to the elongation of *E. coli*. Further, based on the sequence similarity in substrate

binding regions and because of the reminiscence of the morphologic and toxic effect of *Mt*Fic with those of homologous Fic proteins viz., VbhT, *Ye*Fic, *Pa*Fic, and *Nm*Fic, we reasoned that the cognate partner should also be a topoisomerase as in the other Fic toxin effectors. Owing to the fact that the only topoisomerase II present in *Mtb* system is Gyrase, we performed *in vitro* binding and AMPylation activity assay with GyraseB. Dot blot results clearly indicate that though both MBP-tag and Sumo-tag could solubilize the *Mt*Fic protein, MBP-tagged *Mt*Fic did not show binding to GyraseB. Whereas, the Sumo tagged *Mt*Fic showed binding with GyraseB. Additionally, it was seen that the observed binding of *Mt*Fic to GyraseB alone was not sufficient for the morphological effect, as the *Mt*Fic^{H144A} mutant that also binds to GyraseB, neither perturbed bacterial cell growth nor its elongation. Overall, the study demonstrates the AMPylation activity of *Mt*Fic for the first time. AMP transfer on recombinant GyraseB caused by *Mt*Fic resulted in its decreased ATPase activity of the former. The data are consistent with the AMPylation of a tyrosine residue of GyraseB since this modifiable tyrosine is present near the loop participating in ATP binding leading to a loss of its activity. Thus the consequence of the Fic protein function is defined by the substrate it modifies. Therefore, using the AMPylation activity for GyraseB from *Mt*Fic, in parallel with the perturbed growth and elongation in the surrogate *E. coli* host cells, we postulate *Mt*Fic to be a physiologically functional toxin in *Mtb*.

Next, we sought to determine the toxic effect of AMPylation activity on mammalian cells wherein *Mt*Fic AMPylated Rac1 and Cdc42 GTPases and altered the morphology of HeLa and HEK293 cells. The rounded morphology leading to cell death indicates that *Mt*Fic, if released in mammalian cells during infection, plays a role in pathogenesis through its AMPylation activity. However, the AMPylation of Rac1 and Cdc42 was found to be less competent when compared with that of VopS protein, which might be ascribed to the absence of arm domain present in VopS and Ibp proteins wherein the hydrophobic interactions of switch II of Rho GTPases and arm domain of the Fic proteins facilitate the interaction. Binding of Rac1 and Cdc42 with *Mt*Fic was also shown by far western method.

The expression of eGFP tagged *Mt*Fic in HEK293 cells showed distinct spatio-temporal expression pattern. In the initial time points, the expression was homogenous both in cytoplasmic and nuclear fraction, while in later stages, dense eGFP-*Mt*Fic spots accumulated in ER followed by its accumulation inside the nucleus, disruption of lamin B1 and eventual detachment of cells leading to cell death. A distinct expression pattern and cell death is apparently due to AMPylation activity, as its mutant counterpart, *Mt*Fic^{H144A} devoid of enzymatic activity did not show the accumulation pattern and cell death. This led us to hypothesize that the spatial pattern might represent itinerary of *Mt*Fic as well as presence of variable cognate substrates in different cellular compartment in mammalian cells. Identification of its substrates proved challenging because of the instability of *Mt*Fic upon auto-tag cleavage and due to the fact that AMPylation activity involves transient interaction between the enzyme and substrate proteins. To probe this transient interaction, we optimized the far western blot protocol and identified the interacting partners in the fractionated lysate of HEK293 cells. Identification of more prominent bands in ER and nuclear fractions validates the hypothesis that *Mt*Fic has several cognate substrates in fractionated cell lysate. Also, AMPylation activity with fluorescent ATP and eGFP-*Mt*Fic expressed cell lysate identified two substrate positions around Mr 34000 and Mr 72000. A correlation was established between the far western identified substrate positions and the fluorescent ATP aided activity assay, which led to reaffirmation of Mr 34000 band in both the assays. However, Mr >60000 could not be identified in far western method which may be due to limitation of the proper renaturation of larger proteins on the blots.

Remarkably, the identified position in the range of smear in activity assay and prominent bands in membrane/ER-enriched and cytoplasmic fractions was found to be vimentin. Vimentin has also been shown to be a substrate for Bep2 Fic protein from *B. rochalimae* (Pieles et al., 2014). The *in vitro* identification of vimentin as the binding partner for *Mt*Fic in addition to Rac1 and Cdc42 also signifies that *Mt*Fic has the propensity to modulate the operation of cellular cargo. Overall findings of these studies are summarized in Figure 3.19.

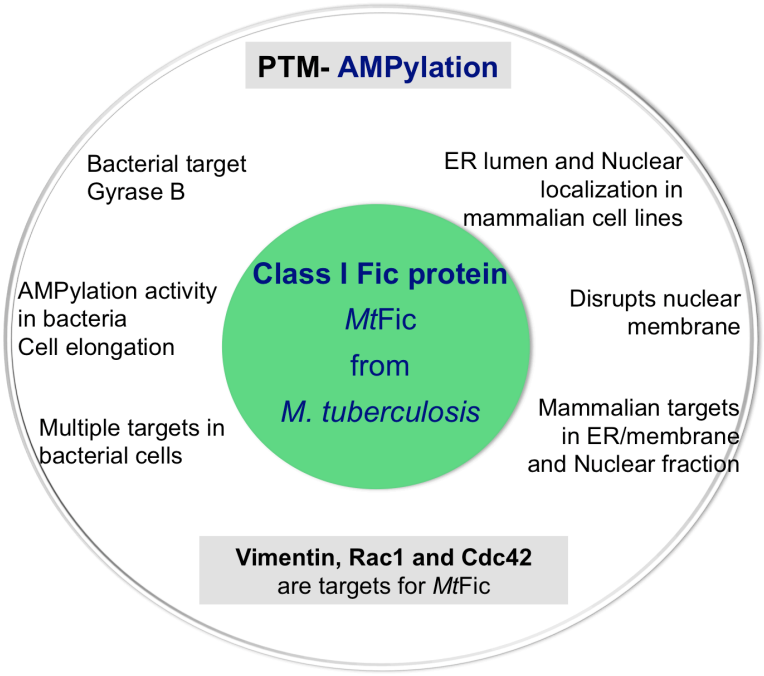


Figure 3.19. Illustration of list of findings of the chapter.

3.5 CONCLUSION AND IMPLICATION OF THE STUDY

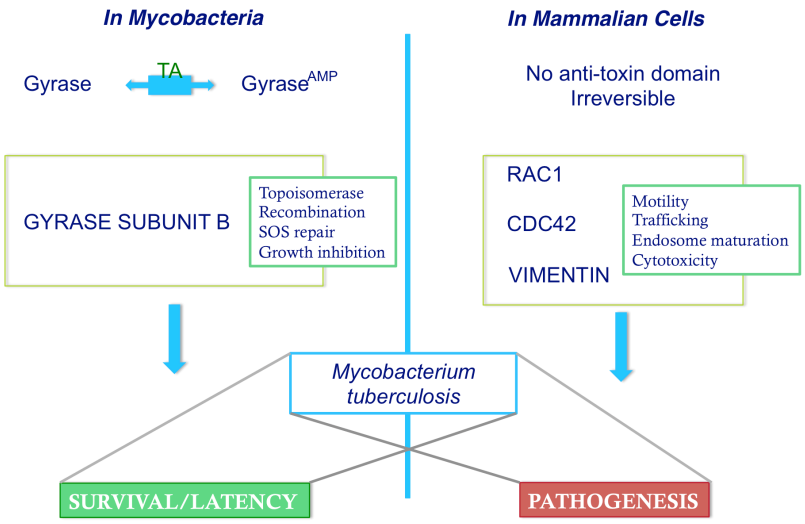


Figure 3.20. A schematic diagram showing probable outcomes due to MtFic AMPylation.

The study is the first report on the AMPylation activity of Fic protein from *Mtb* and identifies its cognate partner both in *Mycobacterium* and mammalian cells (Figure 3.20). DNA gyrase is the only type II topoisomerase present in *Mtb,* which catalyzes negative supercoiling of DNA and is essential for efficient DNA replication, transcription and recombination. AMPylation of GyrB by *Mt*Fic decreases its ATPase activity required for introduction of negative supercoiling in closed circular DNA. As hypothesized in the introduction section, AMPylation of Rac1 and Cdc42 should be correlated with the decreased motility of *Mtb* infected macrophage cells. Vimentin is essential type III intermediate filament and is also involved in internalization of various exoenzymes from pathogenic bacteria as from *Clostridium botulinum C3 (Adolf et al., 2016)*. In addition, vimentin interacts with Rab7a, which controls vesicular membrane traffic to late endosomes and lysosomes, thus important for maturation of phagosomes and autophagic vacuoles (Cogli et al., 2013). Vimentin phosphorylation elicits redistribution of vimentin in the soluble fraction and also controls spatial reorientation of vimentin filament inside the cells (Garg et al., 2006). Moreover, vimentin surface expression increases, as well as the maturation of phagosome is inadequate during *Mtb* infection. Presumably, vimentin modification may alter the internalization of mycobacterial effectors as well as spatial localization pattern of vesicles. Thus these studies provide new avenues for further exploration of the role of AMPylation activity in *Mycobacterium* survival, dormancy, and pathogenicity.

CHAPTER 4

Nucleotide selectivity in Mycobacterium Fic proteins AMPylation/CMPylation

Table of Contents

4.1 INTRODUCTION

As described in Chapter 1, Fic proteins show remarkable specificities for nucleotides, highly characteristic PTMs attributable to the orientation of bound nucleotide and the nature of their corresponding substrates. In brief, Fic proteins, so far, have been reported to display selectivity towards ATP, CDP-choline and UTP, to AMPylate, phosphocholinate, and UMPylate respectively, their specific substrates. Proteins belonging to Doc family, on the other hand, mostly bind to ATP and phosphorylate their substrates. The substrates identified are Rho GTPases (both GDP and GTP bound conformations for diverse Fic proteins), Rab family GTPase, histones, heat shock proteins, vimentin and bacterial class topoisomerase II proteins (gyrase B and topoisomerase IV). In Chapter 3, we have shown AMPylation activity exhibited by *Mtb* H37Rv Fic protein, Rv3641c. As model structure illustrated the conserved residues for ATP binding, AMPylation activity was anticipated for Rv3641c. Residues shown for ATP binding in the closest homolog is represented in the sequence alignment of *Mt*Fic and VbhT proteins (Figure 4.1).

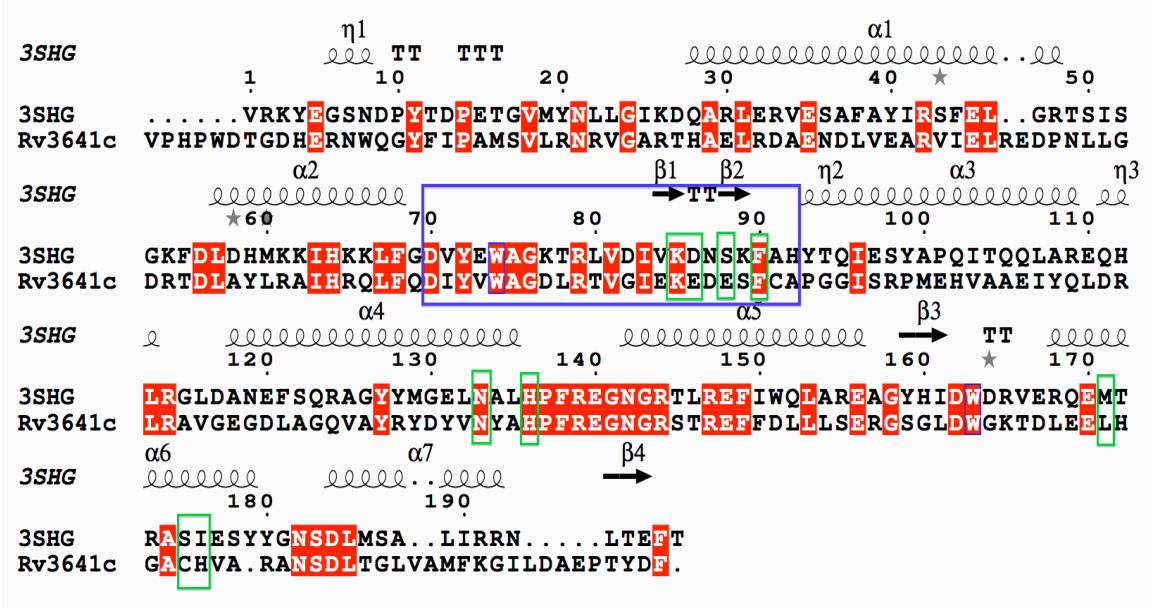


Figure 4.1. Sequence alignment of MtFic and VbhT Fic proteins. *MtFic and VbhT share 29% sequence identity. Residues marked in Green box are from ATP binding pocket, surrounding 4 Å to the adenine residue. Flap region is shown in a*

blue box. Aspartic acid and serine residues from flap show direct interaction with adenine moiety.

Nucleoside binding pocket

Adenosine moiety of bound ATP is surrounded by residues from α4-α5 helices and flap; mainly β-hairpin residues (Figure 4.2). While the nature of residues is mostly hydrophobic, one/two residues from either of the structural units show hydrogen bond interactions with adenosine moiety. In case of IbpFic2 and HYPE (human Fic) the hydrogen-bonding residue is an asparagine from α5 helix (Figure 4.2).

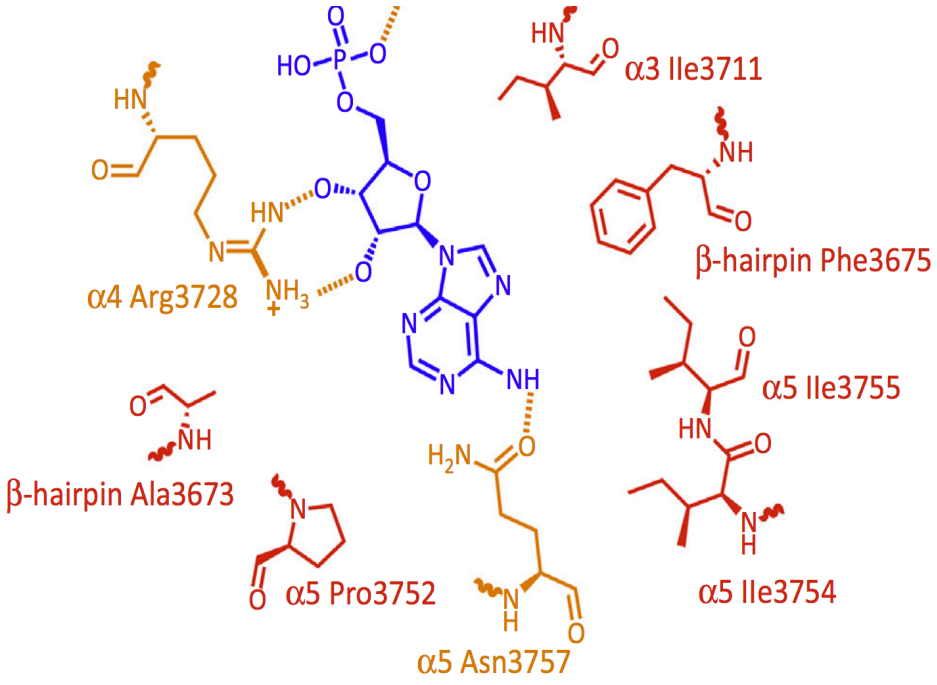


***Figure 4.2. Adenosine surrounding residues from IbpFic crystal structure**. Adapted from (Garcia-Pino et al., 2014)*

In *Nm*Fic, VopS and VbhT, charged residues from flap and α4 helix such as in VbhT toxin, aspartic acid and serine – residues from flap and α4-helix residue asparagine participate in hydrogen bond interactions with amine groups of adenosine, while the residues from α6 helix provides the hydrophobic niche for adenosine moiety (Figure 4.3). Mainly GNG submotif and last arginine (xxR) of the motif are

major participants in placing the phosphates and sugar of nucleotide in an orientation suitable for AMPylation. As the phosphates and sugar accommodating residues are from Fic motif, the orientation of bound nucleotide triphosphate is fairly conserved. Altogether, the flap is highly variable and so the nucleoside surrounding residues from the β-hairpin region of the flap. Though the nucleoside-binding pocket is not conserved, the nature of amino acids contributed by the flap and α4-6 helices complement the hydrophobic and hydrogen bond interactions, respectively, required for adenosine orientation.

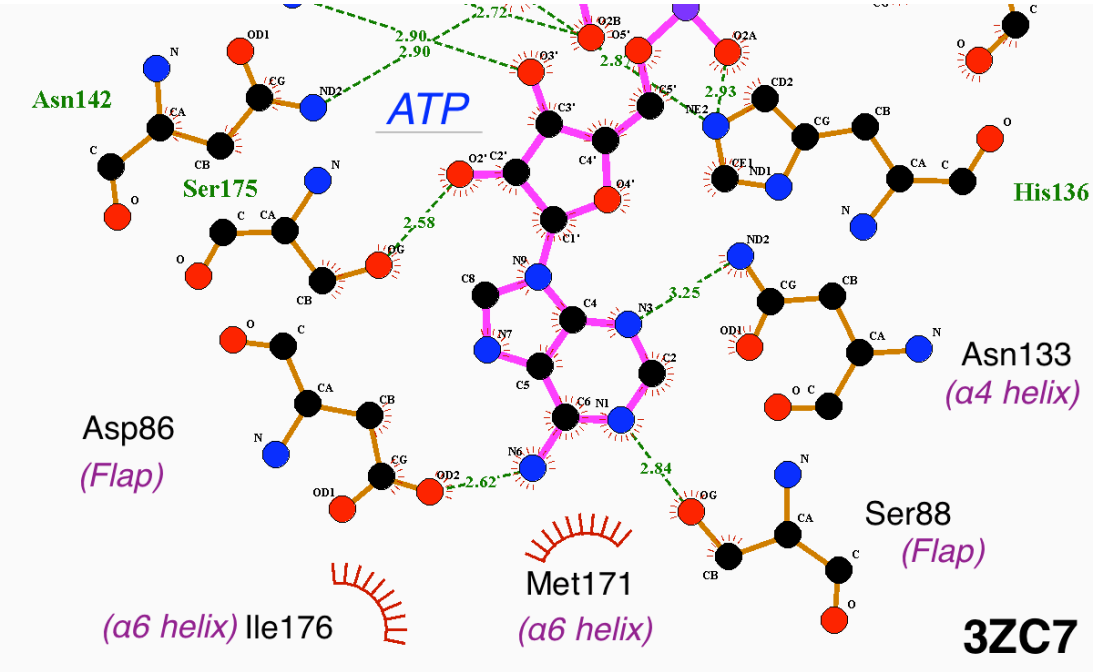


Figure 4.3. Adenosine surrounding residues from the VbhT crystal structure. *The 2D representation of Adenosine interacting residues of Vbht (PDB ID 3ZCN) using LigPlus. The output of Ligplus with 3ZCN, crystal structure of VbhT, as input.*

In case of AnkX, the variability originates in the Fic motif itself. The AnkX Fic motif [HxFxDANGRxxV] shows variation in its anion hole "ANG" and the last arginine is replaced by valine. Valine in place of arginine in the motif results in accommodation of choline in the active site. The "ANG" anion hole accommodates interaction with the alpha phosphate (Figure 4.4). The cytosine position thus inverts and gets stabilized by interactions from CMP domain residues. A tyrosine[41] stabilizes

the amine and stacks the cytosine ring. The keto group of cytosine is stabilized by an arginine[44].

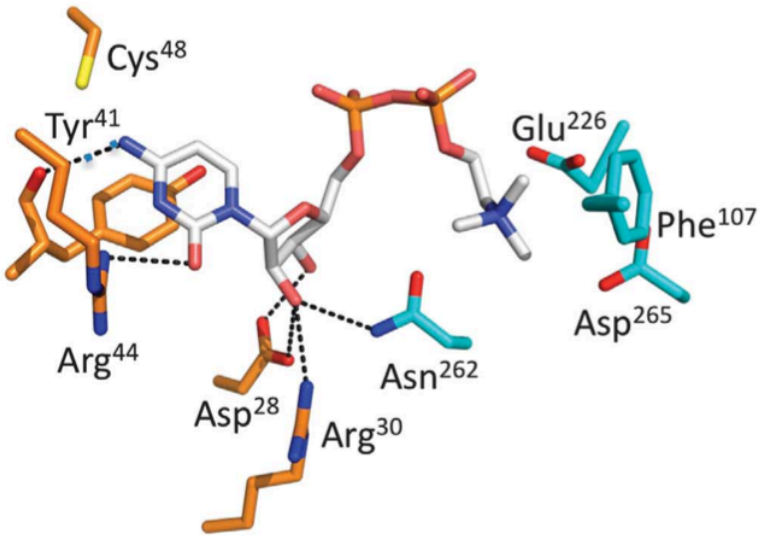


Figure 4.4. Cytosine surrounding residues from AnkX crystal structure. *Adapted from (Goody et al., 2012). Interactions of the CMP domain (in orange) and FIC domain (in cyan) residues with the cytidine and choline moieties of CDP- choline are shown. Hydrogen bonds are shown in dotted lines.*

The Fic protein dataset prepared in Chapter 2, also showed variability in the flap region thereby distinguishing *Mycobacterium* Fic proteins into different subgroups. This chapter uncovers the significance of this variability in terms of nucleotide selection. *Mycobacterium* Fic proteins could be clearly sub-grouped in three sections, first that AMPylate, second that CMPylate and third that can NMPylate as they are promiscuous about nucleotide donors.

4.2 MATERIALS ARE METHODS

4.2.1 Materials

Restriction enzymes and cloning reagents were procured from NEB (New England Biolabs). NTPs (ATP, CTP, GTP and UTP), nucleotide diphosphates (ADP,

CDP, GDP, and UDP), other nucleotide derivatives, SYPRO orange dye, SDS PAGE reagents were purchased from Sigma Chemicals Co., St Louis, MO. α-^{32}P ATP and α-^{32}P CTP were obtained from BRIT (Mumbai, India). RT PCR tubes were purchased from Biorad. Trypsin, FPLC grade was acquired from Promega. Purification reagents, Ni-NTA resin, and IPTG were obtained from Novagen. Genomic DNA of *M. marinum* was procured from BEI resources.

Strains and plasmid: *E. coli* DH5α cells were used for amplification of the constructs during cloning and *E. coli* BL21 (DE3) and Rosetta strains were used for purification of the proteins. Native constructs were prepared in a pET22b vector.

4.2.2 Cloning of class II Fic protein from *M. marinum* (Mmar_0586)

M. marinum strains have three Fic proteins, wherein only Mmar_0586 belongs to class II Fic and homologous to *Sm*Fic (Smeg_2140). Mmar_0586 gene was cloned in pET22b(+) vector following standard cloning procedure. The inserts were amplified using gene specific primers (Table 4.1).

Table 4.1. List of primers for 0586_22b construct

Primers	Sequences (5'-3')	Restriction sites
pET22b_0586-FP	AACCTTCTCATATGCGACCCAGTGATGCGTGGCCTAGAC	***NdeI***
pET22b_0586-RP	AGGATGGACTCGAGTTTCCGAACGCTGGGTGCGCTT	***XhoI***

PCR was followed by restriction digestion (restriction enzymes are listed above). Double digestion for vectors and inserts were followed by ligation using T4 DNA ligase. The positive constructs were screened by double digestion. The sequences was confirmed by DNA sequencing.

4.2.3 Overexpression of *Sm*Fic and *Mm*Fic proteins

The recombinant plasmids were transformed in *E. coli* BL21 (DE3) and cultures were grown in LB medium supplemented with 30 µg/mL kanamycin (*Sm*Fic) and 100 µg/mL ampicillin (*Mm*Fic) till OD_{600} reaches ≈0.6; cultures were then incubated on ice for cold shock for 30 min, followed by induction at 0.4 mM IPTG at 18 °C for 12 hours at 160 rpm. The cells were harvested by centrifugation (Sorvall $RC6^{+}$) for 10 min at 4 °C.

4.2.4 Purification of Fic and substrate proteins

The cell pellet was resuspended in lysis buffer (20 mM Hepes, pH-7.4, 200 mM NaCl, 5% glycerol, 1 mM PMSF) and cells were lysed by sonication (Vibra-cell, Sonics and Materials, USA). Cell debris was removed by centrifugation at 14000 rpm, 4 °C and supernatant obtained was set for binding with pre-equilibrated Ni-NTA resin (Novagen) for 30 mins. The column was washed with 100 mL of lysis buffer. Post column washing, the protein was eluted with a gradient of 100-200 mM imidazole containing lysis buffer. The purity of the fractions were checked on 12% SDS PAGE. Pure fractions were concentrated using centricon-30 (Amicon) and injected on a S100 10/300 GL gel filtration column. The monomer fractions were pooled, concentrated, and checked for its intact N or C-termini by western blotting using anti-His antibody. Protein concentration was determined by taking A_{280}; keeping molar extinction coefficient ε=37590 M^{-1} cm^{-1} (*Sm*Fic) ε=37590 M^{-1} cm^{-1} (*Mm*Fic), wherein molar extinction coefficient (ε) were determined by submitting protein sequence at Expasy ProtParam (http://web.expasy.org/protparam/).

4.2.5 Multiple sequence alignment and generation of conserved profile

Sequence alignment was performed using program CLUSTAL omega (Thompson et al., 1994) *via* the clustal Ω web-service at the European Bioinformatics institute (http://www.ebi.ac.uk/Tools/msa/clustalo/).

4.2.6 Site directed mutagenesis of *Sm*Fic and *Mm*Fic proteins

A two-step single primer based method was used to generate the mutants by amplifying the *Sm*Fic-pET28a and *Mm*Fic-pET22b constructs. The primers used for SDM are listed in Table 4.2.

Table 4.2. List of primers

Primers	Sequences (5'-3')	Restriction sites
SmFic_H227A	CAGTTCGAGACAATAGCTCCCTTCACCGACGGC	***AluI***
SmFic_E100G	CATATGGCATCAGCTGGGGAGGGCACGGCGCTC	***PvuII***
SmFic_W173A	CGCACTGAGCAGGTTGCGATCGGCGGCGGGAAT	***PvuI***
MmFic_H225A	CAGTTCGAATCGATCGCCCCCTTCACAGACGGC	***PvuI***
MmFic_E94G	GCGTCTTCGAAAATTGGCCGGGTTGAAGCGAGC	***BstNI***
MmFic_E94G/R95G	GCGTCTTCCAAAATTGGGGGCGTTGAAGCGAGCG TG	***BslI***
MmFic_W171A	GTAAGGACTGTTCAGAATGCGATCGGTGGAAGCG ACTACTCT	***NaeI***
MmFic_W171L	GTAAGGACTGTTCAGAATTTGATCGGTGGATCCGA CTACTCT	***BamHI***
MmFic_I172A	GTAAGGACTGTTCAGAATTGGGCCGGCGGAAGCG ACTACTCT	***NaeI***
MmFic_W171A/I172A	AAGGACTGTTCAGAATGCGGCCGGCGGAAGCGAC TACTC	***NaeI***

4.2.7 Activity assay for self- AMPylation/CMPylation

The autocatalytic activity of *Sm*Fic proteins was evaluated by taking a varying concentration of proteins (2.5 µg, 1.5 µg and 150 ng) with 10µCi α-P^{32} labeled ATP in reaction buffer containing 20 mM Hepes pH-8, 150 mM NaCl, 1 mM TCEP, 1 mM PMSF and 20 mM $MgCl_2$. The reaction was incubated at 25 °C for 1 hour followed by stopping it by adding 1X SDS loading dye. The reactions mixture was loaded onto 12% SDS-PAGE and visualized by autoradiography using phosphorimager (Biorad). The similar procedure was followed for measuring auto-CMPylation activity assay using α-P^{32} labeled CTP.

4.2.8 Binding and stability analysis by nano-dynamic scanning fluorimetry (DSF)

Sumo-tagged *Mt*Fic, incubated with equimolar concentrations of varying nucleotide derivatives were loaded into nanoDSF grade capillaries, which were then placed into the Prometheus NT.48. Each sample was measured in duplicates. Unfolding of ligand unbound and bound *Mt*Fic samples were detected during heating in a linear thermal ramp (1 °C/min, 25 - 90 °C) at low detector sensitivity and with an excitation power of 10%, and unfolding transition points were determined from changes in the emission wavelengths of tryptophan fluorescence at 350 and 330 nm. Unfolding transition-points were automatically identified by the Prometheus NT. control software. Unfolding onset temperatures were determined after baseline correction and normalization of the data. The unfolding onset was defined as the point at which 1% of the protein was unfolded.

4.2.9 Binding and stability analysis by DSF using extrinsic fluorophore

DSF experiments were done on a fluorescence microplate reader (iQ5, BioRad iCycler Multicolor Real-Time PCR detection system). Protein unfolding was monitored as a function of temperature using a fluorophore, SYPRO Orange. Protein samples (4 μM) in 50 mM Hepes, pH-8, 300 mM NaCl were mixed with 5X SYPRO orange dye and were incubated in a 96-well PCR microplates (BioRad) in the RT-PCR device. Dwell time was kept 0.5 °C per min, a temperature ranging from 20 °C to 90 °C and the fluorescence intensity was measured at each interval of 0.5 °C. Fluorescence intensities for triplicates were plotted as a function of temperature followed by normalization of the data points as fractions from 0-1 (0; folded state, 1; unfolded state). The melting temperatures (T_m) of proteins were determined using Boltzmann sigmoidal dose response with invariable slope (bottom=0, top=1) function in GraphPad Prism 6. Binding parameter K_d was calculated using the single site binding parameter.

4.2.10 Limited proteolysis assay

Mycobacterial Fic proteins were subjected to proteolytic cleavage with trypsin protease. 4 ng/µl trypsin was used to cleave 10 µM protein (1:1000 molar ratio) at incubation temperature at 25 °C. Prior to incubation, the mix was subdivided into 8 aliquots; each one of them was incubated for different time points. The samples from different time points were resolved on 12% SDS-PAGE. Fragmentation pattern was visualized by Coomassie-staining followed by destaining with the methanol-acetone solution. To study the cleavage protection in present of ligand (CTP), 10 µM protein was incubated with 5 mM CTP followed by incubation and resolving as the procedure above. The fragmentation pattern was identified by in-gel digestion followed by LC-MS^2 analysis of the peptides for mascot search.

4.2.11 In gel digestion and LC-MS^2 analysis for (AMP/CMP)ylated peptides

For MS^2 analysis, gel pieces were de-stained (50 mM Ammonium bicarbonate, 50% acetonitrile solution), dehydrated with 100% acetonitrile and then reduced with 1 mM DTT (Sigma) for 1 hour at 56 °C. Subsequently, the samples were alkylated using 50 mM iodoacetamide (Sigma) in the dark for 30 min. The reduced and alkylated samples were digested with 2 µl of trypsin (0.1 µg/µl) with overnight incubation at 37 °C. The digested peptides were extracted using formic acid (5%) and acetonitrile (60%) solution. MS^2 profile was obtained using Bruker system and was used to perform a database search for identification of the proteins, using a MASCOT search tool (Matrix Science Inc., Boston, MS, USA) in BIOTOOLS version 2.2 software (Bruker).

4.2.12 Malachite green based activity assay for Pi detection

The malachite green assays were performed in 96-well plates at 30°C. The final reaction consisted of 5 µM *Sm*Fic enzyme, 10 mM pyrophosphates in CMPylation reaction buffer. The assay and reagent preparations were performed

according to the standard protocol from Perkin-Elmer. All general chemicals were purchased from Sigma- Aldrich (St. Louis, MO).

4.3 RESULTS

4.3.1 Cloning and purification of *Sm*Fic and *Mm*Fic

To characterize the proteins *in vitro*, we initiated the work with cloning and purification of the *Sm*Fic and *Mm*Fic. SmegmaList lists *2140* gene (1209bp) to encode a 402 aa protein of molecular weight 43kDa. MarinoList lists 0586 gene (1215 bp) that encodes a 405 aa protein of 44kDa molecular weight. The cloning and purification of *Sm*Fic are covered in Chapter 2. The *Mm*Fic was PCR amplified using *M. marinum* E9 strain genomic DNA and cloned in pET 22b(+) vector. The clone was confirmed by double digestion with *Nde*I and *Xho*I (Figure 4.5).

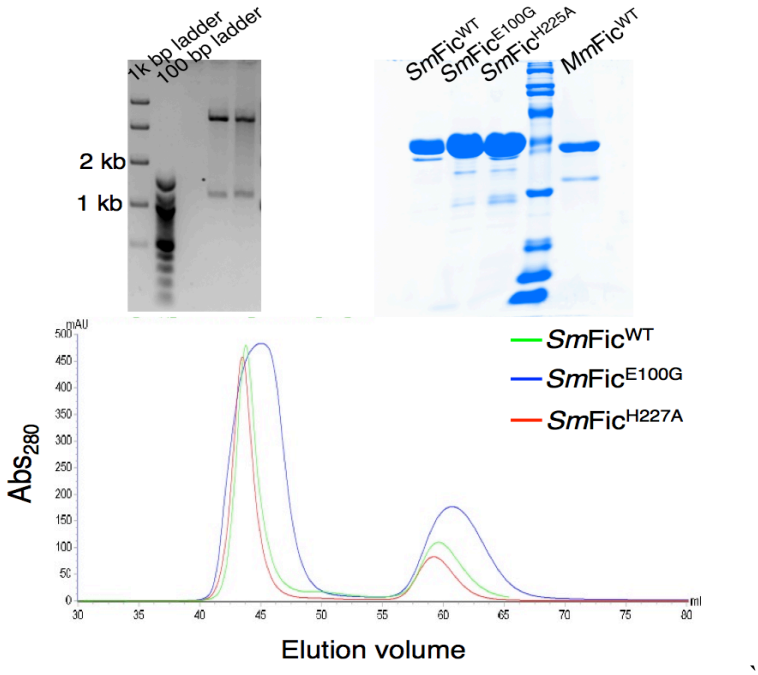


Figure 4.5. Cloning, purification and characterization of SmFic and MmFic. *Upper left panel shows a screening of MmFic construct based on double digestion and fall out release. The fall out corresponds to 1200 bp, the size of MmFic 0586*

gene. Upper right shows purified protein resolved on 12% SDS-PAGE for MmFic and its mutants. Lower left panel shows purified protein resolved on 12% SDS-PAGE for MmFic and its mutants. Lower right panel shows gel filtration profile of SmFic and its mutants. The major fraction elutes in the void (nucleic acid bound fraction) and elution volume correspond to monomer position.

Both the proteins were purified with immobilized metal affinity chromatography method using Ni-NTA resin and eluted at varying concentrations of imidazole. pI of *Sm*Fic is 5.86, and the protein was eluted at 200 mM imidazole concentration. The purification result has been covered in Chapter 2. The $Abs_{260/280}$ ratio was always higher in desalted *Sm*Fic protein, therefore elution was followed by gel filtration with Sephacryl S100 using FPLC (Figure 4.5). The nucleic acid bound fraction gets eluted in void ($OD_{260/280}$ ratio >2) and a nucleic acid unbound fraction eluted as monomer. The monomer fractions were pooled and concentrated till 10 mg/ml and stored at -80 °C.

Though *Mm*Fic is homologous to *Sm*Fic, *Mm*Fic was challenging to purify as the antitoxin domain (1-110 amino acid) has pI of 11 and toxin domain (111-405 amino acid) has pI of 5.6 resulting in overall pI of this type II fic protein 9.5. Any change from pH-8 as well as higher imidazole concentrations (>300 mM) precipitated the protein. Therefore, after affinity purification, the eluted protein was immediately subjected to rapid buffer exchange using desalting column chromatography. The highest solubility could be achieved for *Mm*Fic to 3 mg/ml. Size exclusion chromatography showed the proteins to be monomeric. Sumo tagged *Mt*Fic was purified as described in Chapter 3.

4.3.2 AMPylation is preferred over CMPylation in *Mt*Fic

The nucleotide selectivity for Sumo tagged *Mt*Fic was determined based on differential protein stability obtained upon different ligand binding using nano-DSF method. In this method, the protein stability is addressed by thermal unfolding experiments employing intrinsic tryptophan/tyrosine fluorescence. Therefore, change in emission wavelength at 350 nm with respect to temperature has been utilized to

calculate melting temperature of the protein upon different ligand binding. As *Mt*Fic belongs to class I Fic, the absence of antitoxin domain can lead to enhanced activity and in turn auto-modification, hence the nonhydrolyzable ATP (ApCpp) and CTP (CpCpp) nucleotides (Jena Biosciences) were used for the binding assay. At equimolar concentration (5 mM) of nonhydrolyzable ATP and CTP, ATP showed more T_m shift than CTP (Figure 4.6). Interestingly, the melting profile appeared smooth and symmetrical when equimolar concentration (5 mM) of ADP and AMP were bound to *Mt*Fic than PPi bound and unbound protein. Together with the homogenous melting profile, ADP and AMP ligands binding provided more thermal stability to the protein (Figure 4.6). The data is consistent with the conserved ATP surrounding residues in *Mt*Fic.

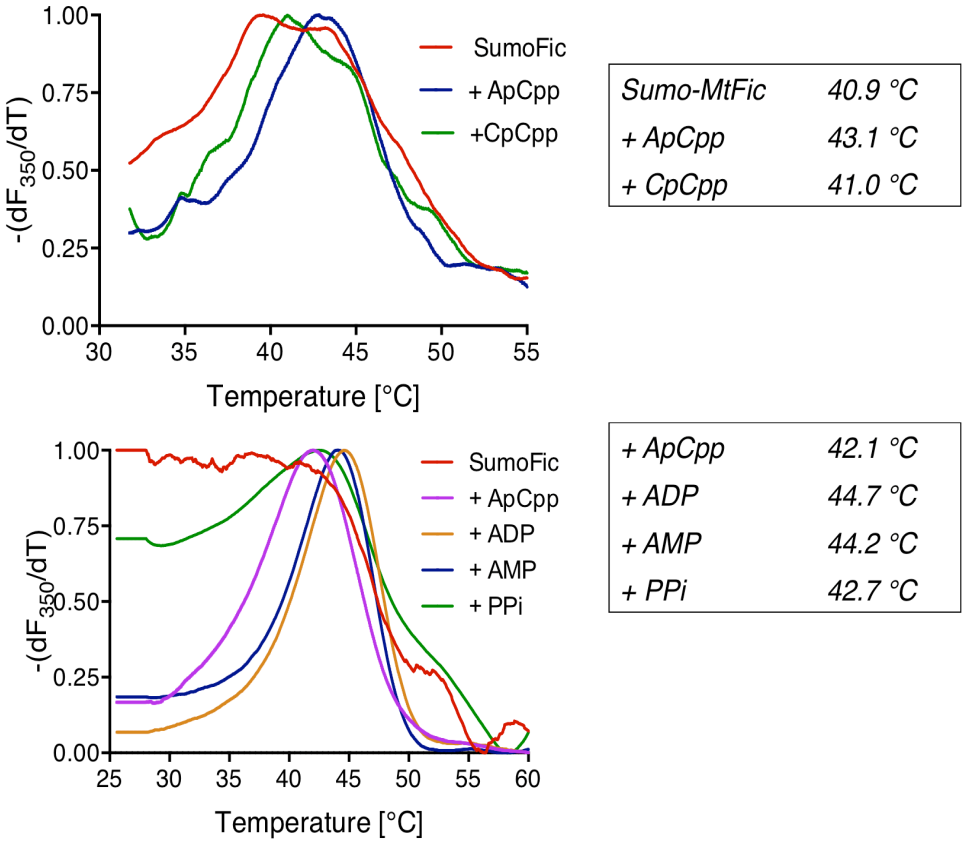


Figure 4.6. nano-DSF profile of MtFic with nucleotides *in terms of temperature and a first derivative of fluorescence intensity at the emission wavelength of 350 nm. Upper panel shows a nano-DSF profile for an equimolar concentration of non-*

hydrolysable ATP and CTP. The lower panel shows nano-DSF profile with equimolar concentrations of ApCpp, ADP, AMP, and PPi.

4.3.3 CMPylation is preferred over AMPylation in *Sm*Fic

As *Sm*Fic shares the canonical Fic motif, it was expected to display AMPylation activity. However, subgroup containing *Sm*Fic protein showed variation in protein length as well as the sequence identity when compared to the subgroup having *Mt*Fic protein. To distinguish the preferred nucleotide for *Sm*Fic protein, DSF experiments were performed using extrinsic fluorophore. This method again exploits the difference in protein folding and stability in the absence and presence of a bound ligand where the unfolding is monitored by an amphiphilic fluorescent dye, typically SYPRO Orange that exhibits a strong fluorescence enhancement when bound to hydrophobic areas which become exposed upon gradual thermal unfolding of the protein. The experiment performed with *Sm*Fic native protein, unbound and bound with equimolar concentration (1 mM) of NTPs (ATP, CTP, GTP, and UTP) showed not much stability change in the protein, however among them, CTP appeared to be a marginally more stabilizing ligand (ΔT_m=2 °C) (Figure 4.7). As *Sm*Fic is class II Fic protein with fused antitoxin domain, inhibitory glutamate obstructs the binding orientation of γ-phosphate of NTP. Therefore, only a slight change in protein stability upon NTP binding is expected. To comprehend it further, we identified the antitoxin motif (S/T)xxxE(G/N) towards the N-terminus; SSNIEN (Figure 4.7) and generated point mutant at inhibitory glutamate position (*Sm*Fic^{E100G}). Recombinant *Sm*Fic^{E100G} protein was unstable as compared to the native *Sm*Fic protein, which might arise due to slight conformational change in the mutant protein itself due to loss of intra-molecular interactions of glutamate, supposedly with arginine of the submotif xxR. However, an equimolar concentration of NTPs showed remarkable changes in the stability of the mutant protein. Consistent with *Sm*Fic^{WT}, the *Sm*Fic^{E100G} mutant gained maximum stability upon CTP binding (ΔT_m=8 °C). ATP and GTP binding showed intermediate stabilization ($\Delta T_m \approx$ 6 °C). UTP binding appeared to be the least stabilizing (ΔT_m=4 °C) in all the four NTPs assayed (Figure 4.7).

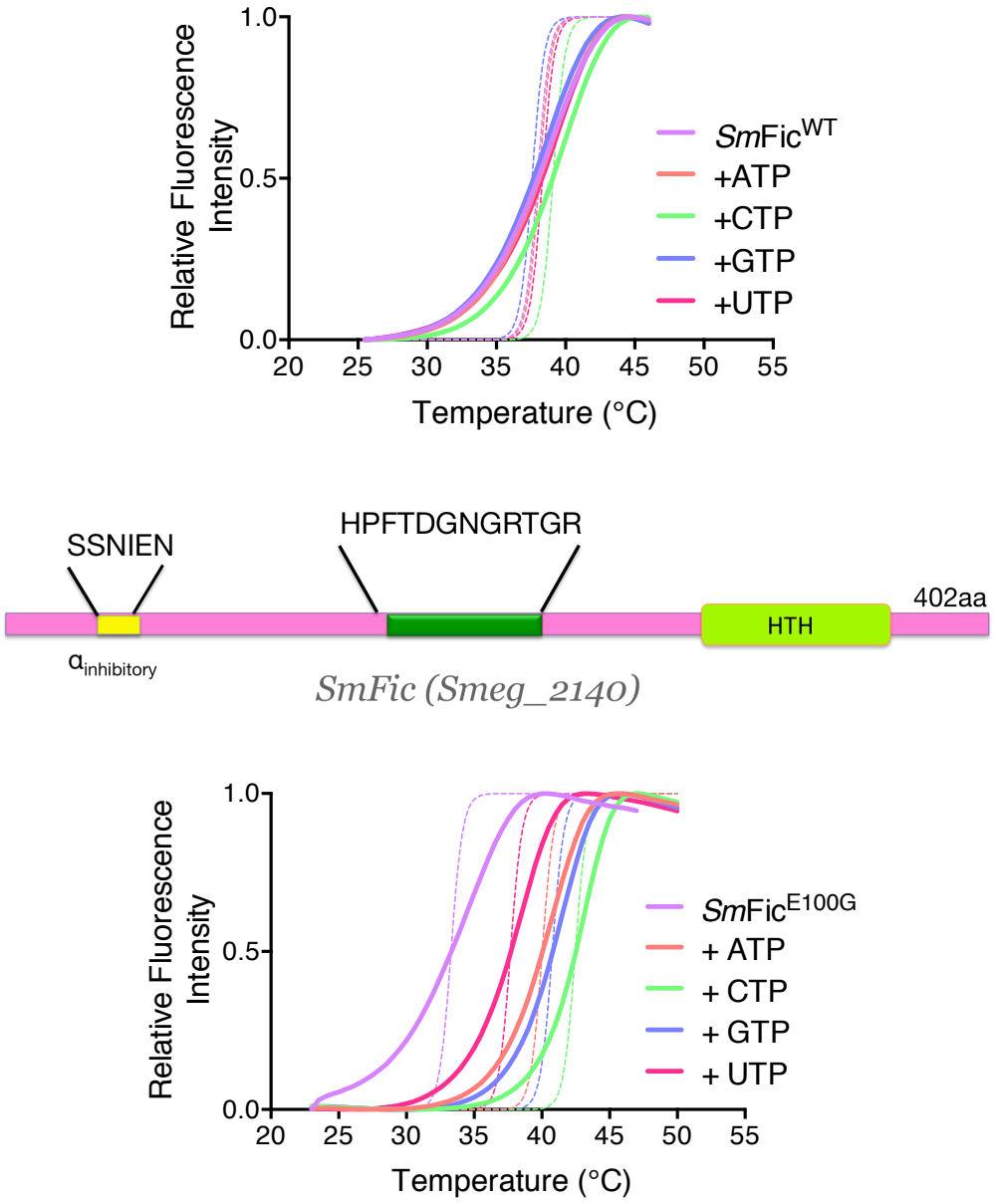


Figure 4.7. DSF profile of class II fic protein SmFic (Smeg_2140) with NTPs. *Upper panel represents two state thermal denaturation profile of SmFic^{WT} in the presence of equimolar (1mM) NTPs. Middle panel represents a schema of class II SmFic protein with its antitoxin motif, toxin motif and HTH domain. The lower panel represents two state thermal denaturation profile of SmFic^{E100G} mutant protein in the presence of equimolar (1mM) NTPs. The data points are normalized between 0 to 1. The solid lines are the data points and the dotted lines are curve fit for two state unfolding with slope =1.*

*Sm*Fic is a multi-domain protein and along with toxin Fic and antitoxin domain, it also has a winged helix-turn-helix domain (wHTH domain) at the C-terminus (Figure 4.7). As HTH domain is involved in DNA binding, further assays were performed with equimolar concentrations of CTP and its deoxy-counterpart dCTP. As CTP binding stabilized the protein more than the dCTP (Figure 4.8), we deduce that CMPylation is the preferred activity of *Sm*Fic protein.

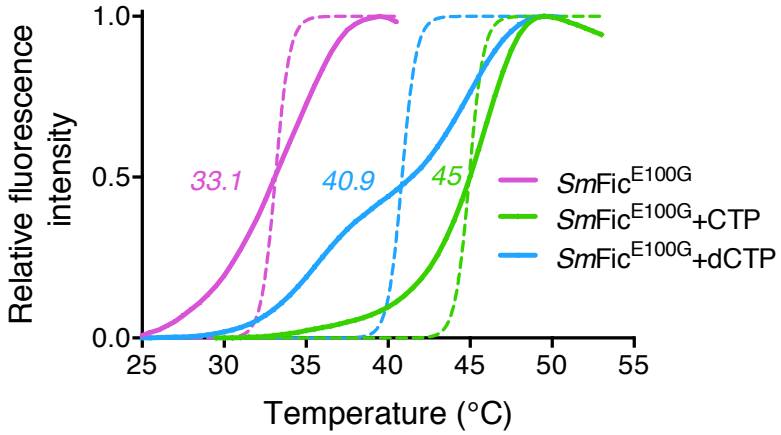


Figure 4.8. DSF profile of SmFic (Smeg_2140) with CTP and dCTP*. The graph represents two state thermal denaturation profile of SmFic E100G mutant in presence of equimolar (1mM) UTP, CTP, and dCTP. The data points are normalized between 0 to 1. The solid lines are the data points and the dotted lines are curve fit for two state unfolding with slope =1.*

Moreover, CDP binding to *Sm*Fic^{E100G} protein imparted additional stabilization when compared to equimolar CTP binding. However, for *Sm*Fic native protein, CMP was the most stabilizing ligand (Figure 4.9). Binding of Pi, PPi, CTP and CDP with *Sm*Fic glutamate mutant protein showed a gradual order of stability change (Figure 4.9). Moderate stabilization with PPi (Figure 4.9) validates the fact that interactions accomplished between the Fic motif and the phosphates of NTPs are requisite for the contact during cofactor binding. Preferably, the nucleotide selection follows discrimination between purines or pyrimidines. However, the *Sm*Fic^{E100G} mutant protein marginally stabilizes further upon uracil binding whereas the stability shift upon UDP and PPi bindings were similar, suggesting that uracil binding has a

marginal contribution in stabilizing the protein (Figure 4.10). While CMPylation is the most preferred and UMPylation is the least, implying that a negatively charged residue from the nucleoside surrounding residue(s) that stabilizes the amine of cytosine might be repelling the keto group of uracil.

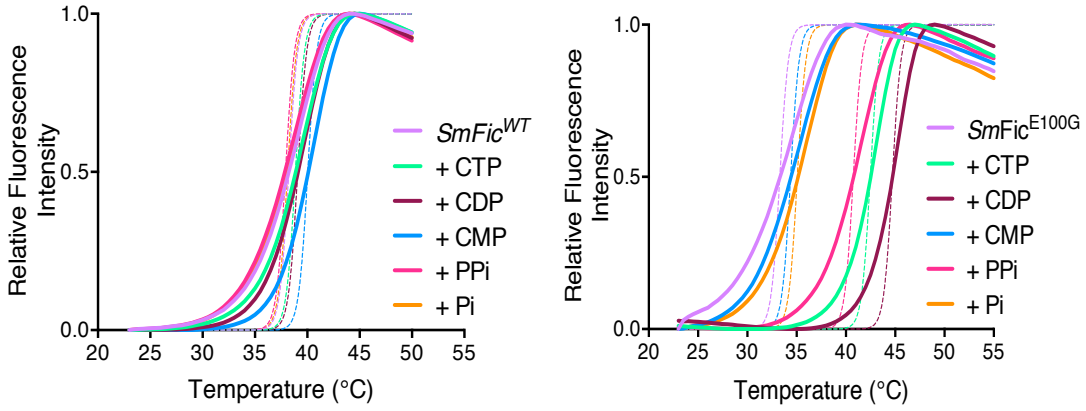


Figure 4.9. DSF profile of SmFic (Smeg_2140) with CTP and the derivatives. *The profiles represent two state thermal denaturation of SmFic^{WT} and SmFic^{E100G} mutant in presence of equimolar (1 mM) CTP, CDP, CMP, PPi and Pi. The data points are normalized between 0 to 1. The solid lines are the data points and the dotted lines are curve fit for two state unfolding with slope =1.*

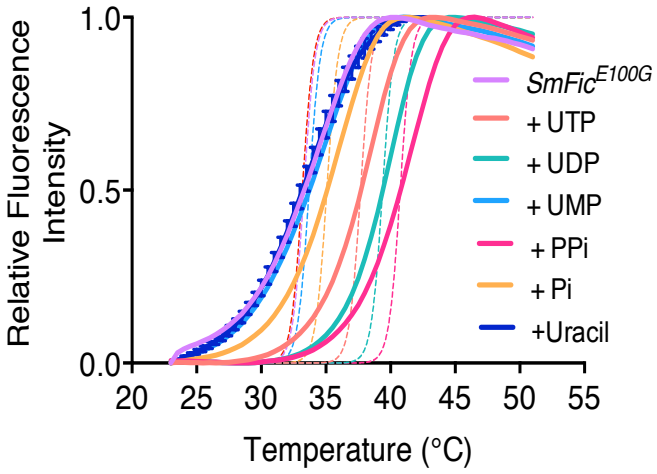


Figure 4.10. DSF profile of SmFic (Smeg_2140) with UTP and the derivatives. *The profiles represent two state thermal denaturation of SmFic^{E100G} mutant in the presence of equimolar (1mM) UTP, UDP, UMP, PPi, Pi, and uracil base. The data*

points are normalized between 0 to 1. The solid lines are the data points, and the dotted lines are curve fit for two state unfolding with slope =1.

4.3.4 Determination of auto-modification (AMPylation/ CMPylation) activity of *Sm*Fic

If obstruction due to glutamate from SxxxEG motif is obviated, the γ-phosphate of NTP interacts with arginine of submotif xxR (Figure 4.11). This interaction allows the proper orientation of bound NTP a prerequisite for the nucleophilic attack and activity, thus allowing the toxin to be relatively more active and promiscuous for auto-modification at serine, threonine or tyrosine residue of the toxin enzyme. Therefore auto-modification activity also provides a strategy to check the NMPylation activity in absence of substrate. Auto-modification activity of the recombinant *Sm*Fic was assayed with α-P^{32} ATP as nucleotide in a reaction mix with 5 mM $MgCl_2$ as a divalent ion in 50 mM Hepes, pH-8. The activity assay was performed with *Sm*Fic^{WT}, *Sm*Fic^{E100G} point mutant and double mutant *Sm*Fic$^{E100G/H227A}$ proteins, as enzymes, where E^{100} is the inhibitory glutamate from antitoxin motif and H^{227} is the conserved histidine, the key residue from toxin motif. While the mutant protein *Sm*Fic^{E100G} showed auto-AMPylation activity at ≥1.5 μg amount, *Sm*Fic^{WT} and double mutant (*Sm*Fic$^{E100G/H227A}$) proteins could not perform auto-modification (Figure 4.12). The auto-AMPylation data show that the purified monomeric *Sm*Fic protein can bind to nucleotide substrate and is active. Inactivity of invariable histidine mutant also suggests the His^{227} is the key residue for catalysis. Inactivity towards auto-modification of *Sm*Fic^{WT} and *Sm*Fic$^{E100G/H227A}$ double mutant is consistent with the characteristic of the antitoxin motif as well as toxin motif. The preferred CMPylation activity was further shown with activity with α-P^{32} CTP in similar condition with an equimolar concentration of nucleotide as well as the radioactivity. In similar conditions and incubation time, *Sm*Fic$^{\mathbf{E100G}}$ mutant showed more activity with α-P^{32} CTP as compared to α-P^{32} ATP (Figure 4.12).

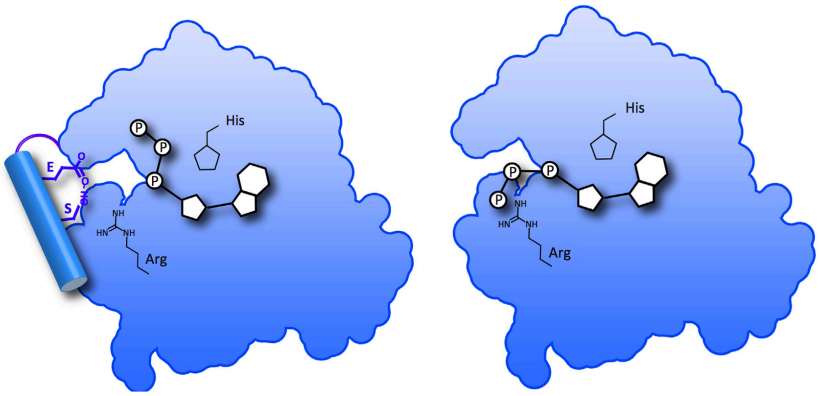


Figure 4.11. Schematic of the mechanism of toxin inhibition by antitoxin in class II Fic proteins. *Adapted from (Garcia-Pino et al., 2014). Antitoxin domain serine and glutamate interacts with arginine from fic motif, making this arginine unavailable to stabilize the γ-phosphate of ATP. Moreover, the binding of γ-phosphate is perturbed due to repulsion from glutamate negative charge.*

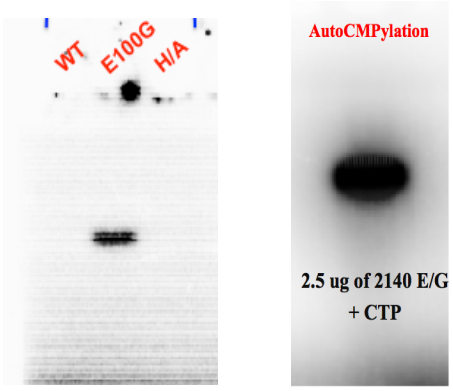


Figure 4.12. Autoradiograph showing auto-NMPylation activity of SmFic. *Reaction mixes contain equal concentrations of SmFic^{WT}, SmFic^{E100G} and SmFic^{H227A} mutants, as enzymes. Radiolabel α-^{32}P ATP and α-^{32}P CTP were used as nucleotides.*

4.3.5 Analysis to understand basis of the nucleotide selectivity

The closest homolog of *Sm*Fic crystallized and characterized so far is *So*Fic (29% identity) which has been shown to perform AMPylation and has also been crystallized with ATP, ADP and PPi. To deduce CTP preference of *Sm*Fic, the ATP surrounding residues in the structural overlap of modeled structure of *Sm*Fic and *So*Fic crystal structure were compared. In *So*Fic, the adenine ring of bound ATP is

surrounded by tyrosine from conserved Fic motif and leucine from β-hairpin of the flap. However, in *Sm*Fic, the flap region showed an insertion of 12 amino acid residues [GNWRTEQVWIGG]. Specifically, VWIGG is the inserted sequence when structural alignments were compared (Figure 4.13).

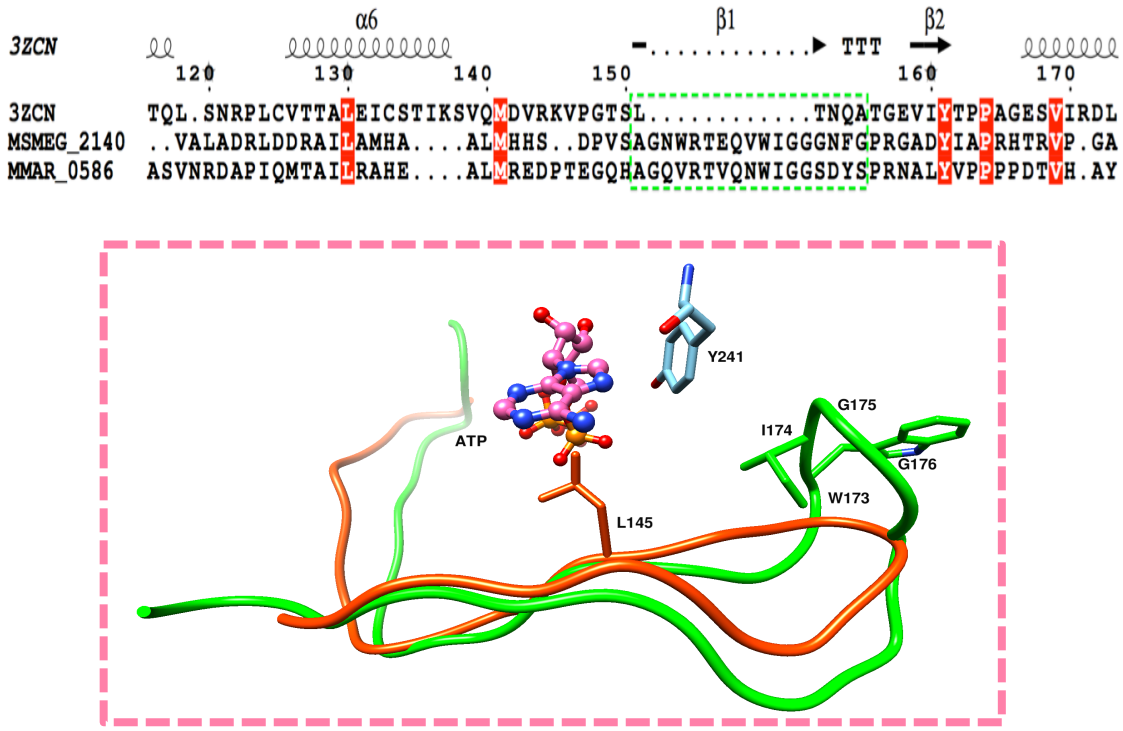
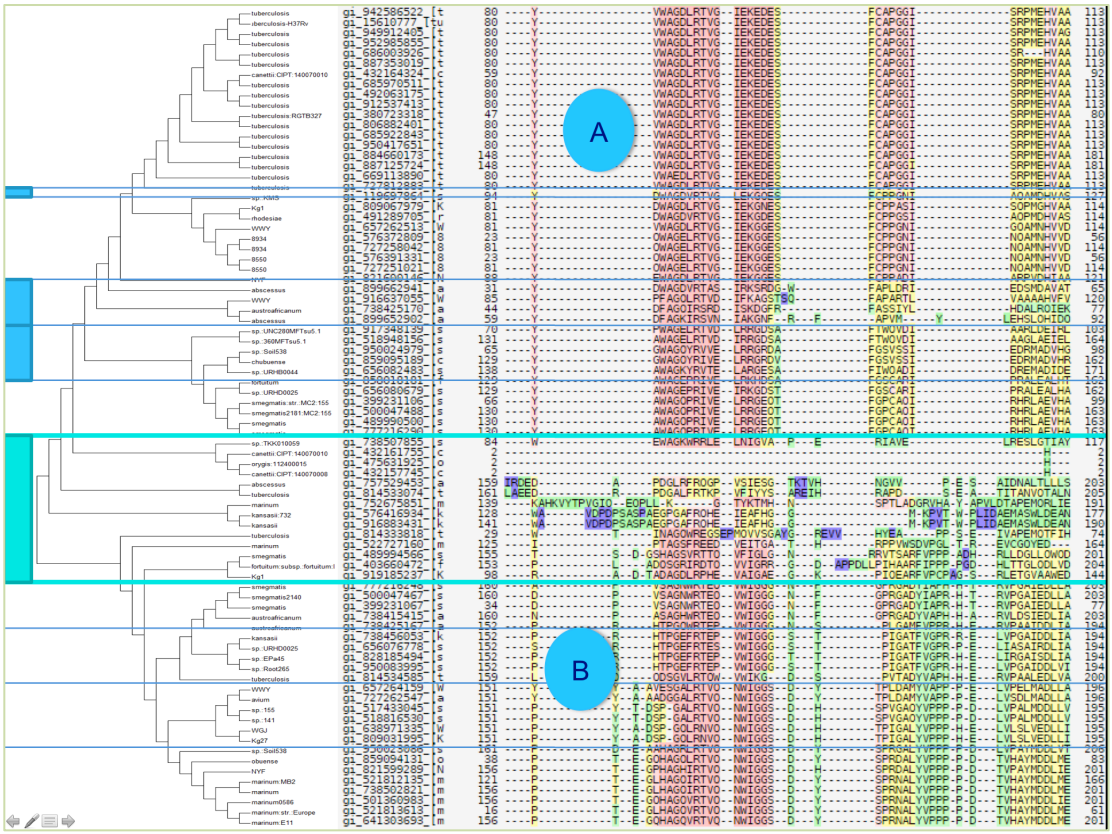


Figure 4.13. Comparison of flap residues participating in nucleotide binding from SoFic and SmFic proteins. *Upper panel represents the MSA profile of ATP bound SoFic (PDB ID 3ZCN) with SmFic and MmFic proteins. The flap regions exhibit amino acid insertion in flap region, common for MmFic when compared to SoFic protein sequence. Lower panel represents structural alignment showing the flap from SmFic (model structure) and SoFic. The insertion in the flap region of SmFic extends in the section corresponding to SoFic β-hairpin of flap.*

In CDP-choline binding, a cytosine gets accommodated in the CMP binding site (Chapter 1 & Figure 4.4. It is present in opposite orientation to adenosine binding site and also, the flap inserted amino acid sequences were unstructured in the *Sm*Fic model structure. Therefore, the prediction of an orientation of cytosine as well as the supporting interactions from the flap could not be deduced with docking study.

4.3.6 AMPylation activity is assigned to other members of the subgroup A

Next, we examined the residues from the flap region in the multiple sequence alignment results, obtained in Chapter 2. This MSA represents a comparison of Fic proteins from 21 different *Mycobacterium species*. The dataset of 81 canonical Fic motif-containing proteins had two distinctions (Figure 4.14). The First group shared the residues similar to *Mt*Fic, with a consensus of K(D/E)xSxF, wherein the aspartic acid, serine and phenylalanine residues from the flap region surround the adenosine of ATP. As *Mt*Fic showed the preference for ATP over CTP in ligand binding assay (Figure 4.6), the preference of this group was assigned for ATP ligand. The second group shared consensus of Gx[W/F]RT[E/Q]xxWI[G/K]G (Figure 4.14).



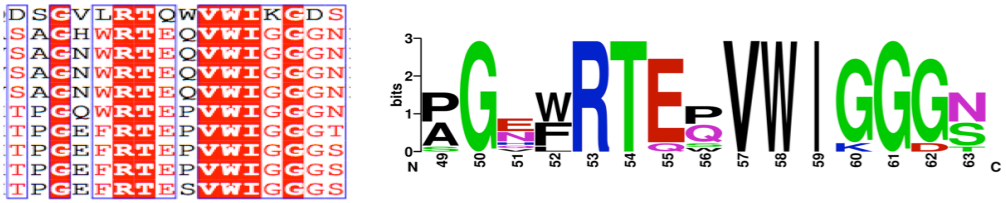


Figure 4.14. Comparison of flap region from Fic proteins from 21 Mycobacterium sp using clustal omega. The *upper panel represents the MSA profile of flap region of Mycobacterial Fic proteins. Upper half of the alignment (section A) represents homologs of MtFic with conserved residues surrounding adenosine moiety. The lower half of alignment with SmFic homologs shows a drastically different set of residues. The lower panel represents the consensus followed and identified from the flap region.*

4.3.7 CMPylation activity is preferred over AMPylation for other members of the subgroup B

Further, we asked the question if CMPylation activity is limited to *Sm*Fic only or other proteins belonging to the same group would also show CMPylation preference over AMPylation. To this end, we cloned and purified another member of the subgroup, a homolog of *Sm*Fic, Fic protein from *M. marinum (MmFic; Mmar_0586)*. The preference for nucleotide ligand was determined using DSF method as described for *Sm*Fic. At equimolar concentrations (1 mM and 10 mM) of ATP and CTP, recombinant *Mm*Fic showed additional protein stability upon CTP binding as compared to that on ATP binding (Figure 4.15).