

https://dspace.org/

Mechanisms of mutagenesis in Mycobacterium tuberculosis: structural and functional characterisation of the DNA polymerase accessory factors encoded by Rv3394c and Rv3395c

Ndwandwe, Duduzile Edith

2013-07-29
http://hdl.handle.net/10539/12918

Downloaded from a DSpace repository. https://dspace.org/



Mechanisms of mutagenesis in *Mycobacterium tuberculosis*: structural and functional characterisation of the DNA polymerase accessory factors encoded by Rv3394c and Rv3395c

Duduzile Edith Ndwandwe

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Doctor of Philosophy

February 2013

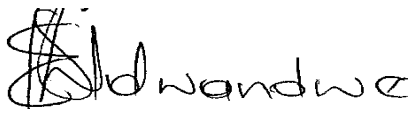
Abstract

Mycobacterium tuberculosis is presented with environmental host assaults that damage its DNA during infection. Tubercle bacilli possess mechanisms to protect against moststresses imposed by the host, including genotoxic stress. However, tolerance of DNA lesions that have escaped the normal repair processes requires the function of specialist DNA polymerases that can introduce mutations during translesion synthesis (replication by-pass), thus leading to damage-induced mutagenesis. Mycobacteria employ a novel DNA polymerase, DnaE2, for DNA damage tolerance and induced mutagenesis. DnaE2 belongs to the C-family of DNA polymerases, which are known to replicate DNA with high fidelity, and has been implicated in virulence and the emergence of rifampicin resistance of *M. tuberculosis in vivo*. In this study, DnaE2 was shown to function in the same pathway as two accessory proteins, ImuB and ImuA', for damage tolerance and induced mutagenesis in mycobacteria. In this system, DnaE2 performs the polymerase function in translesion synthesis whereas ImuB is a cryptic Y-family DNA polymerase that lacks critical active site residues. It contains a β-clamp binding motif that allows interaction with the β-clamp and presumably enables DnaE2 and ImuA' to access the replication fork. ImuB has a C-terminal region extending from the β-clamp binding motif which contains disordered regions that allow the interaction with other proteins and is important for function. ImuA' is also essential for damage tolerance and induced mutagenesis but its function remains unknown. This protein is structurally similar to *Escherichia coli* RecA protein in the N-terminus and the middle domain, but it has a distinct C-terminus that was shown to be important for the interaction with ImuB. The essential replicative, C-family polymerase, DnaE1, was shown to be upregulated in response to DNA damage and was also shown to interact with ImuB. To explore the possibility that other proteins are involved in this pathway, ImuB was C-terminally tagged for use as bait in pull-down experiments in *M. smegmatis*. However, introduction of the tag disrupted ImuB function, further reinforcing the importance of the C-terminal region of ImuB for the function of this protein, presumably via protein-protein interactions. In contrast, a variant of ImuA' which was N-terminally tagged was shown to retain functionality; however, experiments using this protein as a bait for pull-down proved to be unsuccessful. Proteomic analysis of wild type *M. smegmatis*, a *dnaE2* deletion mutant and complemented derivative was carried out on cells exposed to the same conditions as used in the pull-down assay. Base excision repair (BER) components were identified in this analysis, but did not detect ImuB and ImuA', suggesting that the levels of expression of these proteins were comparatively lower under the conditions tested resulting in failure of the pull-down

experiment. Finally, numerous attempts were made to express and purify recombinant forms of ImuB and ImuA’ in *E. coli* for use in structural studies. Both proteins were expressed in the soluble and insoluble fractions; however the levels of soluble protein were low, and as a result, purified protein preparations could not be obtained.

Declaration

I declare that this thesis is my own unaided work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.



Duduzile Edith Ndwandwe

15 May 2013

Date

Dedication

This work is dedicated to my late father **Derrick Lindani Ndwandwe** who passed on to be with the Lord on the 8th of November 2010. He was a proud father and has always been there for me, always supported and encouraged me to pursue and fulfill my dreams. I am most certain that he will be my guardian angel watching over me. May his soul rest in peace- Ndwandwe Mkhatshwa, Nxumalo, Zwide KaLanga, Sothondose, Sdinana Samaphisi, Kabanhle……

I also dedicate my work to my son, **Okuhle Amile Mshumi**. I have missed out of your first year of growing up of your life and this has made me to appreciate life even more and I hope that one day you when you read this document you will understand and appreciate the sacrifices I had to make to achieve my goals so that you can be proud of me being you mother.

Acknowledgements

I am forever grateful for the financial support I have received from the National Research Foundation (NRF), Mellon Postgraduate Mentoring Scheme and the University of the Witwatersrand. I also wish to acknowledge the following funding agencies that have enabled me to execute my duties of the project; National Health Laboratory Service (NHLS) Research Trust Project awarded to Dr. Warner as a Research Development Grant at the initiation of the study; the South African Tuberculosis and AIDS Training (SATBAT) programme (NIH/FIC1U2RTW007370/3), which has facilitated the funding of part of the project enabling the data to be published in a peer-reviewed journal; the Columbia University-South African Fogarty AIDS and TB Training & Research programme (AITRP), and KwaZulu Natal Research Institute for TB and HIV (KRITH) travel award which enabled me to travel to the United States of America (US) to learn valuable techniques that have contributed to the study.

Special thanks to my immediate supervisor **Professor Valerie Mizrahi** who gave me an opportunity to be her student. When I first walked into the laboratory, I was intimidated by Val's passion towards science which in turn drives the success of the lab. Val, you took me in like a child and mothered me and believed in me that I have the potential to succeed in this field. More than that, through your guidance and mentoring I have learnt that success does not come overnight and that experiments don't always work one's your favor but through patience there are lessons to be learnt in order to move forward. You have instilled me with specialized skills in making me a better scientist coming from a background where I did not even know the existence of science as a field but I always questioned everything. I am eternally grateful to you for having shown me that no destiny is far to be reached.

My deepest gratitude to my co-supervisor **Dr. Digby Warner,** thank you for being supportive and always helpful. Digby, you have been an absolutely extraordinary teacher and have always been courageous even when experiment did not work but you have known what to say to keep me highly motivated.

To my co-supervisor **Dr. Bavesh Kana**, a special thank you for all the financial support and motivation you gave me when times were hard. Most importantly I have learnt to be true to myself and understand that my best is all that matters in achieving my goals.

To the MMRU team then and now, thank you for all your support and faith in me. I will forever be part of the MMRU as we have made good friendships along the way learnt a lot

from each other's personal lives. To the CBTBR group, thank you for your patience and friendships that we have made. You have made me feel like a mentor to the lab when I was also in a journey of finding myself. I am most grateful to **Mrs. Fiona Peachy** for helping me with proofreading the thesis. Thank you for the never-ending support-Ngiyabonga!!!!!

I would like to thank **Dr. Sarah Fortune** (Assistant Professor of Immunology and Infectious Disease at Harvard University, School of Public Health, Boston, MA) for hosting me in your laboratory. You are the kindest person I met and always willing to lend a helping hand. To the entire Fortune lab thank you for all your support and welcoming me from Africa with warm hearts. Of particular importance is the friendship I have made with **Rupal Shah**-we connected instantly and we were able to talk about our personal lives other than the research. I will treasure our friendship and hope that we have another opportunity to meet again.

I am most grateful to **Alejandra Garces** and **Micheal Chase** (Harvard University, Fortune lab) for processing peptide samples and assisting with the statistical analysis of the raw data. Thank you to **Dr. Davis Sarracino** (ThermoScientific Center) for allowing the use of his Mass Spectrometer.

I would like to thank **Dr. Stoyan Stoychev** (CSIR Biosciences, CSIR Pretoria) for running the Peptide Mass Fingerprinting to enable the identification of proteins.

To my special husband **Sandisile Mshumi**, thank you for all your support and understanding of my studies. I am forever grateful for the sacrifices you have made for us and for being my pillar of strength. To my mother **Lettie Mhlanga**, I don't know what I could have done without you, you always knew what to say in order to lift my spirit when I felt down and kept me going. Words cannot describe the unconditional love and support you gave me as a studying mother. You took on a huge responsibility of looking after my son **Okuhle Mshumi** so that I can achieve and fulfill my goals. I would forever be thankful to you for that and hope that I have made you proud. Finally, to my baby brother **Thabani Ndwandwe**, I see myself in you encouraging me to excel.

Publication from this thesis

Warner, D. F., D. E. Ndwandwe, G. L. Abrahams, B. D. Kana, E. E. Machowski, C. Venclovas, and V. Mizrahi. Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A **107:**13093-8.

Table of Contents

Abstract ii

Declaration iv

Dedication v

Acknowledgements vi

Publication from the study viii

Table of Content ix

List of Figures xii

List of Tables xiv

Chapter 1

1. Introduction 1

1.1 The current state of TB control 2

1.1.1 Approaches to TB drug discovery 3

1.1.2 TB drugs and their mode of action against *M. tuberculosis* 4

1.2 *M. tuberculosis* encounters variable microenvironments during host infection 7

1.2.1 Evidence for genotoxic stress during host infection 8

1.2.2 A role of DNA repair in *M. tuberculosis* pathogenesis 9

1.3 DNA damage tolerance 11

1.3.1 Structural insight into Error-prone and Error-free lesion Bypass... 12

1.4 Mechanism of the SOS response 13

1.5 Importance of tolerance mechanisms 15

1.6 Emergence of drug resistant mutants through adaptation 16

1.7 SOS response in Mycobacteria 17

1.8 Working hypothesis 19

1.9 Aims and Objectives 20

Chapter 2

2. Materials and methods 21

2.1 Bacterial strains, plasmids and culture conditions 21

2.1.1 Bacterial growth conditions 25

2.2 General DNA manipulations 26

2.2.1 DNA Extraction 26

2.2.2 DNA manipulations 28

2.2.3 Agarose gel electrophoresis 29

2.2.4 DNA fragment recovery from agarose gels and quantification 29
2.2.5 Transformation of bacteria 29
2.2.6 Polymerase Chain Reaction (PCR) 31
2.2.7 DNA Sequencing 32
2.3 Construction of vectors for interaction studies 32
2.3.1 Construction of vectors for yeast two hybrid system 32
2.3.2 Construction of vectors expressing tagged proteins for pull-down assay 33
2.3.3 Construction of complementing vectors expression truncated MsImuA' 35
2.3.4 Construction of complementing vectors expressing truncated MsImuB 35
2.4 Construction of vectors for protein expression 36
2.5 Yeast two hybrid (Y2H) analysis 37
2.6 DNA damage-induced mutagenesis assay 36
2.7 DNA damage sensitive assays 37
2.8 Protein extraction, quantification and detection 38
2.8.1 Protein extraction from *M. smegmatis* 37
2.8.2 Protein analysis by SDS-PAGE and Western blotting
2.8.3 Expression and purification of recombinant ImuB and ImuA' in *Escherichia coli* 38
2.9 Proteomic analysis of *M. smegmatis* strains using mass spectroscopy 40
Chapter 3
3. Results 43
3.1 Bioinformatic analysis of DnaE2 accessory factors, ImuB and ImuA' 43
3.1.1 Prediction of ImuB and ImuA' structures 48
3.2 Interaction analysis of the proteins involved in the DnaE2-dependent induced mutagenesis pathway 50
3.2.1 Confirmation of pair-wise interactions between DnaN, ImuB, ImuA', and DnaE2 by yeast two-hybrid analysis (Y2H) 51
3.2.2 The β-clamp binding motif is required for interaction of ImuB with the β-clamp 53
3.2.3 Y2H suggests the ability of ImuB to self-interact 54
3.2.4 Assessment of the possibility that other proteins may be involved in this pathway 55
3.2.5 Mapping of interaction domains of ImuB and ImuA' 57
3.2.5.1 Identification of the interaction domain of ImuB 57
3.2.5.2 Identification of interaction domain of ImuA' 62

3.3 Analysis of *imuA'-imuB/dnaE2* cassette of *M. smegmatis* 65
3.3.1 Analysis of the DnaE2-dependent mutagenic pathway in *M. smegmatis* 67
3.3.2 ImuB and ImuA' are required for DnaE2-dependent mutagenesis in *M. smegmatis* 68
3.3.3 Analysis of truncated ImuB 69
3.3.4 Analysis of truncated ImuA' 70
3.4 Identification of other proteins involved in DnaE2-dependent damage-induced mutagenesis in M. *smegmatis* 72
3.4.1 Introduction of a tag on the C-terminus of MsImuB 72
3.4.2 Introduction of a tag on the N-terminus of MsImuB 74
3.4.3 Introduction of a tag on the N-terminus of MsImuA' 76
3.5 Pull-down analysis of N-terminally tagged ImuA' in *M. smegmatis* 78
3.6 Proteomic analysis of M. smegmatis wild type, Δ*dnaE2* and Δ*dnaE2*::*dnaE2* strains after exposure to DNA damaging conditions 80
3.7 Heterologous expression of *M. tuberculosis* ImuA' and ImuB 82
3.7.1 Small scale expression of ImuA' and ImuB 82
3.7.2 Large scale protein expression and purification of MBP-ImuB and MBP-ImuA' 84
Chapter 4
4. Discussion 88
4.1 The DnaE2-dependent mycobacterial mutasome shares functional similarity to Pol V mediated DNA repair 88
4.2 Distinguishing features of the mycobacterial mutasome 90
4.3 Insight into the DnaE2 mutagenic pathway 92
4.4 Heterologous expression of DnaE2 accessory proteins, ImuA' and ImuB 94
5. Appendices 96
6. References 100

List of Figures

Figure 1: An SOS-inducible mutagenic gene cassette 19

Figure 3.1: DnaE2-dependent mutagenesis in *M. tuberculosis* 43

Figure 3.2: Multiple sequence alignment of ImuB homologues compared to Y family polymerase from bacteria 44

Figure 3.3: Sequence alignment of ImuA proteins compared to *E. coli* RecA protein 47

Figure 3.4: ImuB shares structural similarity with Y family polymerase from *S. Solfataricus* Dpo4 48

Figure 3.5: ImuB has extensive disorder in the C-terminal region 49

Figure 3.6: ImuA' is structurally similar to *E. coli* RecA 50

Figure 3.7: Diagrammatic representation of the principle of the Yeast-two hybrid Analysis 51

Figure 3.8: Confirmation of pair-wise protein-protein interactions 52

Figure 3.9: Confirmation of the β-binding motif of ImuB 54

Figure 3.10: ImuB has the ability to self interact 55

Figure 3.11: The involvement of DnaE1 in the DnaE2-dependent mutagenic pathway 56

Figure 3.12: Schematic representation of the summary of the Y2H data 57

Figure 3.13: Sequence alignment indicating the position where artificial stop codons were introduced 58

Figure 3.14: Deletion of the C-terminal extension of ImuB eliminated interaction with other proteins but not DnaN 60

Figure 3.15: Deletion of the putative β-binding motif as well as the C-terminal extension abrogated interaction with all interacting partners 61

Figure 3.16: Removal of the little finger domain into the thumb domain abrogated interactions with all tested proteins 62

Figure 3.17: ImuA' proteins aligned indicating positions of artificial truncations 65

Figure 3.18: The C-terminal region of ImuA' is critical for the interaction with ImuB 63

Figure 3.18: Summary of the protein-protein interactions inferred fromY2H analyses 65

Figure 3.19: Analysis of *M. smegmatis* mutagenic cassette component interactions 66

Figure 3.20: DnaE2 is essential for damage tolerance and induced mutagenesis in *M. smegmatis* 67

Figure 3.21: Deletion of the C-terminal region of ImuB phenocopies the Δ*imuB* Phenotype 69

Figure 3.22: The C-terminal region of ImuA' is critical for damage tolerance and induced mutagenesis 71
Figure 3.23: Schematic representation of the TAP tag introduced on the C-terminus of MsImuB 72
Figure 3.24: Introduction of a FLAG-tag on the C-terminus of ImuB impairs protein Function 73
Figure 3.25: Schematic representation of the 3×FLAG tag introduced on the N-terminus of MsImuB 74
Figure 3.26: Introduction of a tag on the N-terminus of ImuB impaired protein function 75
Figure 3.27: Schematic representation of the 3×FLAG tag introduced on the N-terminus of MsImuA' 76
Figure 3.28: The function of ImuA' with an N-terminal 3× FLAG tag was maintained 77
Figure 3.29: Tagged ImuA' was not detected on SDS-PAGE and western blotting Analysis 78
Figure 3.30: Schematic presentation of the transcript levels of ImuA' and ImuB pre- and post UV radiation in *M. tuberculosis* 79
Figure 3.31: Identification of proteins known to be up-regulated in response to UV radiation 81
Figure 3.32: Small-scale expression of MBP-ImuB and MBP-ImuA' indicate expression in both insoluble and insoluble fractions 83
Figure 3.33: Affinity purification of MBP-ImuB 85
Figure 3.34: Affinity purification of MBP-ImuA' 86
Figure 3.35: Representation of the SDS-PAGE gel with MBP-ImuB eluted from Q-Sepharose column. 87

List of Tables

Table 2.1: Strains used in the study 22

Table 2.2: Plasmids used in the study 22

Table 2.3: Primers used to introduce artificial start and stop codons into the *ImuB* and *ImuA'* ORFs 33

Table 2.4: Incorporation of a 3xFLAG sequence on the N-terminus of MsImuA' and MsImuB 34

Table 2.5: List of primers for expression of *ImuB, ImuA'* and C-terminally truncated *ImuB* in *E. coli* 36

Chapter 1

1. Introduction

Tuberculosis (TB) was declared a global health emergency by the World Health Organization (WHO) in 1993 yet continues to claim around 2 million lives per annum. One-third of the world's population is infected with the causative agent, *Mycobacterium tuberculosis*, an obligate human pathogen. The TB burden is the highest in Asia and Africa with India and China accounting for almost 40 % of the world's TB cases. The African region has 24 % of the world's cases and the highest rates of cases and death per capita. South Africa is included amongst the countries with the highest world's cases of drug resistant TB (WHO, 2012). There were 9.4 million new TB cases in 2009, of which 3.3 million were women, and 1.1 million were in people co-infected with HIV (Barry *et al.*, 2009; Russell *et al.*, 2010; WHO, 2010; Yang *et al.*, 2007). Adding to this burden is the increasing emergence and spread of multi-drug (MDR) and extensively drug resistant (XDR) strains resistant to the major frontline and second-line anti-TB agents (Barry & Blanchard, 2010; Dye *et al.*, 2002; Sacchettini *et al.*, 2008). This pandemic is particularly alarming when one considers that TB is a curable disease.

WHO introduced the Directly Observed Therapy - Short Course (DOTS) program in 1995, and it has been implemented worldwide (WHO, 2009). This strategy is cost-effective, requiring the use of a combination therapy consisting of an intense two-month treatment phase with the first-line drugs isoniazid (INH), rifampicin (RIF), ethambutol (EMB), and pyrazinamide (PZA), followed by a four-month continuation phase of INH and RIF (WHO, 2009). It is estimated that nearly seven million lives have been saved compared with the pre-DOTS era, with high cure rates having been achieved in most countries worldwide; the result is that the global incidence of TB has been in slow decline since early 2000s (Zumla *et al.*, 2012a). However, the emergence and spread of resistant strains of *M. tuberculosis* poses a threat to global TB control. Drug-resistant TB is a man-made problem resulting primarily from patient non-compliance to the extended chemotherapy (Zumla *et al.*, 2012a). MDR-TB defined as resistance to INH and RIF is treatable with PZA, ethionamide, cycloserine, any of the injectable aminoglycosides (kanamycin, amikacin or capreomycin) and the fluoroquinolones (ciprofloxacin and ofloxacin) (WHO, 2009).This treatment is expensive, can last for up to two years and can be toxic to patients (Dye *et al.*, 2002). This exacerbates non-compliance, and can lead to the emergence of XDR strains (Naidoo, 2007). XDR-TB,

which is defined as resistance to INH and RIF plus a fluroquinolone and any of the second-line injectable agents, is virtually untreatable. The emergence of strains of *M. tuberculosis* that are resistant to a multiplicity of drugs with different mechanisms of action provides a cogent example of the capacity of the organism to adapt to, and overcome, the harsh conditions encountered during infection (Ehrt & Schnappinger, 2009; Ramakrishnan, 2012).

According to the most recent WHO report, the Millennium Development Goals (MDG) target to halt and reverse the TB epidemic by 2015 has already been achieved. New cases of TB have been falling for several years and fell at a rate of 2.2 % between 2010 and 2011; moreover, the TB mortality rate has decreased 41 % since 1990 indicating that the world is on track to achieve the global target of a 50 % reduction by 2015. Even though the fight against TB seems promising, the global burden of TB remains enormous. In 2011, there were an estimated 8.7 million new cases of TB, of which 13% were co-infected with HIV, and 1.4 million people died from TB, including almost one million deaths among HIV-negative individuals and 430 000 among people infected with HIV. Global progress also conceals major regional variations and disparities, with the African and European regions not on track to halve the levels of mortality by 2015 (WHO, 2012). This situation demands the development of new tools to control TB, including novel anti-tubercular agents which might eradicate the infection caused by drug-susceptible and drug-resistant strains of *M. tuberculosis*. However, the reality is that resistance may develop to any new anti-tubercular drug. In 2009, the first report of totally drug-resistant (TDR) strains came out of Iran (Velayati *et al.*, 2009); however, there are studies that seem to suggest that TDR was a pre-existing form of XDR-TB (Migliori *et al.*, 2012). With most of these drug-resistant strains genotyped as members of the Beijing, Haarlem and EAI super families (Velayati *et al.*, 2009), the TB epidemic has become even more alarming given that members of the Beijing super-family have also been associated with increased ability to cause and spread disease, as well as with co-infection with HIV (Caws *et al.*, 2006; Hanekom *et al.*, 2007; Johnson *et al.*, 2009).

1.1 The current state of TB control

As part of a global effort to eradicate TB, the WHO has implemented a program known as the "Global Plan to Stop TB 2006-2015", which aims at reducing the 1990 prevalence of TB by 50% by the year 2015 and the eventual eradication of the disease by 2050 (WHO, 2006). The program has the following targets: i) expanding and improving the DOTS program so that

supervised standardised treatment and effective drug management is implemented; ii) increasing TB/HIV collaborations, preventing and managing drug resistant *M. tuberculosis* strains and addressing the poor socio-economic conditions of the most vulnerable populations; iii) strengthening the primary health care sector; iv) engaging all communities (TB patients, care providers as well as the affected communities) to endorse the implementation and adherence of International Standards of Tuberculosis Care (ISTC); and v) promoting the development of novel and more effective TB diagnostic tools as well as therapies (WHO, 2006). All of these are points that the program seeks to achieve; however, a longer treatment course would lead to patient non-compliance and development of drug resistance, underscoring the need for the development of shorter drug regimens, as suggested (Salomon *et al.*, 2006). The Global Plan to Stop TB 2011-2015 further seeks to improve on the following: i) laboratory strengthening-included as a major component; ii) fundamental research and operational research; iii) strategic frameworks to set out each major component of the plan in a clear and consistent format; iv) up-to-date epidemiological projections; v) updated targets for TB care and for research and development; vi) and updated funding requirements. With no new antibiotics introduced for the treatment of TB since the discovery of rifampicin in 1967, a number of programs have recently been implemented for the discovery and development of new TB drugs in order to achieve the Global Plan to Stop TB (Russell *et al.*, 2010). Importantly, research plays a key role in efforts to eradicate the disease.

1.1.1 Approaches to TB drug discovery

As in other areas of anti-microbial drug discovery (Lamichhane, 2011; Payne *et al.*, 2007; Raviglione *et al.*, 2012; Sarker *et al.*, 2012), the discovery of novel anti-mycobacterial drugs have been carried out using two approaches: phenotypic vs. target-led. The phenotypic approach is based on the screening of small-molecule libraries for compounds with the ability to kill mycobacteria using whole-cell screening assays. Mutants that are resistant to the compound are then used to identify the molecular target of the compound by means of whole-genome sequencing. This approach has been quite successful, leading to the discovery of TMC207 (Bedaquiline) (Andries *et al.*, 2005); the novel nitroimidazole derivatives, PA-824 and OPC-67683 (Delamanid), which are in early-stage clinical development for drug-resistant and drug-susceptible TB (Barry *et al.*, 2004; Matsumoto *et al.*, 2006), and the benzothiazinone, BTZ-043 (Cole & Riccardi, 2011; Makarov *et al.*, 2009; Pasca *et al.*, 2010; Sala & Hartkoorn, 2011). Both bedaquiline and the nitroimidazoles show promising activity against non-replicating organisms. However, limitation of whole-cell screens is that it has

produced only a small number of potential hits with moderate potency (Robertson *et al.*, 2012).

The second approach, which is target-led, is based on the identification of biochemical functions (usually enzymatic) that are essential for growth and survival of the organism, and the subsequent use of high-throughput biochemical screens to identify small-molecule inhibitors of the target's function (Crunkhorn, 2012; Dover & Coxon, 2011; Zumla *et al.*, 2012b). The major challenge with this approach is that the potential hits that have potency against an essential enzyme may not efficiently penetrate the mycobacterial cell envelope. The other limitation is that the assumption that inhibition of an individual metabolic reaction will cause cell death may not be a complete description of the mode of action of an effective anti-bacterial agent (Dick & Young, 2011). Bactericidal activity may occur as a result of the downstream consequences of initial drug-target interaction which result in toxic molecules that act as true effector molecules (Dick & Young, 2011). For this reason, the identification of novel antimicrobials requires that both approaches should be used to mitigate the limitations of each approach (Abrahams *et al.*, 2012). From these approaches, a number of lead compounds with novel modes of action have been discovered in recent years (Lougheed *et al.*, 2009). In addition to these agents developed specifically for TB, four drugs developed for other conditions are being re-purposed for TB (Bald & Koul, 2012; Koul *et al.*, 2011; Ma *et al.*, 2010).

1.1.2 TB drugs and their mode of action against *M. tuberculosis*

The complete genome sequence of *M. tuberculosis* (Calver et al., 2010) has enabled the identification of essential proteins as novel targets for drug development. The success of the target-based approach relies on the quality of the target as well as the level of validation (Chen & Du, 2007; Warner & Mizrahi, 2012). Only a few validated targets have been revealed to date. RpsA is a putative target for the old first-line TB drug, PZA. PZA has constituted a key component of the short-course therapy regimen since the 1980s without its cellular target being known until very recently. The ribosomal subunit, RpsA, was found to be a target of pyrazinoic acid – the active form of the pro-drug PZA. RpsA is a vital protein involved in protein translation and the ribosome-sparing process of trans-translation (Shi *et al.*, 2011). This target may thus provide a rationale for the activity of PZA and might help to explain the treatment-shortening activity of PZA (Bald & Koul, 2012; Scorpio & Zhang, 1996; Shi *et al.*, 2011). Arabinosyltransferase, a key component enzyme in arabinogalactan

synthesis, is the target of ethambutol (EMB) (Telenti *et al.*, 1997). Combination therapy is used for the treatment of MDR-TB using the aminoglycosides, kanamycin and amikacin, together with the cyclic peptide, capreomycin. Aminoglycosides target the 16S rRNA whereas capreomycin targets rRNA methyltransferase (Gillespie, 2002; Johansen *et al.*, 2006). The *rpoB*-encoded β-subunit of RNA polymerase has been shown to be the target of rifampicin (Telenti *et al.*, 1993). Fluoroquinolones have been shown to target DNA gyrase. Isoniazid (INH) targets the InhA protein (Vilcheze *et al.*, 2006). Ethionamide is similar to isoniazid in that it is a pro-drug that requires activation to form an adduct with NAD that subsequently inhibits the NADH-dependent enoyl-(acyl-carrier protein) reductase InhA (Hazbon *et al.*, 2006). Streptomycin targets the *rpsL*-encoded S12 ribosomal protein and p-amino salicyclic acid targets the folate pathway (Chakraborty *et al.*, 2012; Gillespie, 2002; Rengarajan *et al.*, 2004). The nitroimidazoles , PA-824 and OPC-67683, are pro-drugs that require activation by a deazaflavin (cofactor F_{420}) dependent nitroreductase, Ddn (Manjunatha *et al.*, 2009; Manjunatha *et al.*, 2006; Matsumoto *et al.*, 2006). More recently, the promising drug, bedaquiline, which is active against drug-susceptible and drug-resistant *M. tuberculosis*, targets ATP synthase (Andries *et al.*, 2005; Manjunatha *et al.*, 2006; Sacksteder *et al.*, 2012) which the bacteria uses to generate energy (Bald & Koul, 2010; Hurdle *et al.*, 2011).

New TB drugs, at various stages of clinical development include rifapentine (a member of the rifamycin drug class) and the fluoroquinolones, gatifloxacin and moxifloxacin (Bald & Koul, 2012; Koul *et al.*, 2011), as well as linezolid (Koh *et al.*, 2012; Sotgiu *et al.*, 2012). Rifapentine, a bacterial RNA polymerase inhibitor, which is more effective than rifampicin *in vitro* and has a longer serum half-life, would be an ideal replacement candidate for rifampicin for use in the treatment of drug-susceptible TB. This drug is currently in Phase II clinical trials to assess if it could be given once or twice weekly as a replacement for the daily intake of the first-line antibiotic rifampicin (Ma *et al.*, 2010; Sacchettini *et al.*, 2008). Linezolid is a member of the oxazolidinones which are compounds that bind to the 50S ribosomal subunit inhibiting the subsequent formation of the 70S ribosome, thereby effectively inhibiting protein synthesis (Barry & Blanchard, 2010; Ma *et al.*, 2010).

Although linezolid, which is currently in phase II clinical trials, is associated with high toxicity levels, this drug has been successfully used off-label for the treatment of MDR- and XDR-TB (Barry & Blanchard, 2010; Lee *et al.*, 2012; Ma *et al.*, 2010; Sacchettini *et al.*, 2008). Gatifloxacin and moxifloxacin, which target the bacterial DNA gyrase, have proven to

be more effective than the current second-line drugs ofloxacin and ciprofloxacin in the treatment of TB. Currently, phase III clinical trials are underway to ascertain whether these drugs can be used to shorten first-line therapy to 4 months by replacement of ethambutol or isoniazid (Barry & Blanchard, 2010; Diacon *et al.*, 2012a; Ma *et al.*, 2010; Sacchettini *et al.*, 2008).

There are six new compounds, specifically identified for TB treatment, that are currently in clinical development. Three of these are in Phase I clinical trials. The first, PNU-100480 is an oxazolidinone that exhibits greater activity than linezolid against *M. tuberculosis* in a murine model (Ford *et al.*, 2001; Ma *et al.*, 2010; Pinon *et al.*, 2010; Sacchettini *et al.*, 2008). The second drug AZD-5847, also an oxazolidinone, is active against drug-resistant forms of *M. tuberculosis* (Ma *et al.*, 2010); and SQ-109 on the other hand, is an ethylenediamine that interacts synergistically with isoniazid and rifampicin against *M. tuberculosis* in a murine model (Barry & Blanchard, 2010; Ma *et al.*, 2010; Sacchettini *et al.*, 2008). SQ-109 targets MmpL3, which is a mycolic acid transporter required for incorporation of mycolic acid into the cell wall of *M. tuberculosis* (Grzegorzewicz *et al.*, 2012; Tahlan *et al.*, 2012). The other new drugs, namely, bedaquiline and delamanid, are currently in Phase II trials (Diacon *et al.*, 2012b; Gler *et al.*, 2012). Bedaquiline has shown great efficacy against susceptible and resistant *M. tuberculosis* strains both *in vitro* and *in vivo* (Anantharaman & Aravind, 2003; Barry & Blanchard, 2010; Diacon *et al.*, 2012a; Ma *et al.*, 2010; Sacchettini *et al.*, 2008; Tasneen *et al.*, 2011).

PA-824 and OPC-67683 are nitrofuranylamides, with OPC-67683 referred to as a delamanid that have potent activity against *M. tuberculosis.* Both drugs are pro-drugs that require enzymatic reduction to generate active forms (Maroz *et al.*, 2010; Matsumoto *et al.*, 2006; Singh *et al.*, 2008). There is substantial cross-resistance between these two agents, implying that they have a similar if not identical mechanism(s) of action (Hurdle *et al.*, 2011). Both agents have activity in the mouse model, although OPC-67683 may be more potent (Matsumoto *et al.*, 2006). When added to rifampin and pyrazinamide, either PA824 or OPC-67683 allows shortening of TB treatment in the mouse model to 3–4 months (Tasneen *et al.*, 2008). This indicates the potential to reduce the time needed to develop new anti-tuberculosis regimens (Diacon *et al.*, 2012a). It is important to note that bedaquiline, moxifloxacin and the nitroimidazoles, which target replicating and non-replicating bacilli, have also enabled the development and validation of various *in vitro* models for testing the efficacy of new TB drugs under stressful conditions encountered within the host (Cho *et al.*, 2006; Ehrt &

Schnappinger, 2009; Ramakrishnan, 2012); for example, under conditions of hypoxia in the presence and absence of reactive nitrogen intermediates (Barry & Blanchard, 2010; Hussain *et al.*, 2009; Ma *et al.*, 2010; Sacchettini *et al.*, 2008). Even though there are promising drugs currently in various stages of clinical development, there is concern that drug resistance might subvert any new chemotherapeutic regimen.

1.2 *M. tuberculosis* encounters variable microenvironments during host infection

M. tuberculosis bacilli are inhaled into the lungs inside the human host where they experience a variety of stresses from the host immune system. Inside the host, the bacilli are internalized by alveolar macrophages which should kill the bacteria (Peyron *et al.*, 2008). The interaction of the bacilli and the receptors of the immune cells results in an intracellular signaling cascade that culminates in a proinflammatory response (van Crevel *et al.*, 2002). Mycobacteria have evolve strategies that can trigger signals that dampen or modulate the innate immune response indicating that killing of the bacterium depends on the intrinsic microbicidal capacity of the host phagocytes and virulence factors of the ingested mycobacteria (Ahmad, 2011). The alveolar macrophages produce inflammatory cytokines and chemokines that serve as a signal for infection, thus allowing migration of monocytes, neutrophils and lymphocytes to the focal site of infection resulting in granuloma formation (Peyron *et al.*, 2008). During this time, the bacilli resist the bactericidal mechanisms of the macrophage by preventing phagosome-lysosome fusion, and continue to multiply inside the macrophage ultimately causing macrophage necrosis. The released bacteria are also engulfed by another macrophage and the whole process is repeated. In healthy individuals, the interplay between *M. tuberculosis* and host cellular and adaptive immune cells results in asymptomatic infection as granulomas separate *M. tuberculosis* cells from the surrounding tissue, and the mycobacteria become latent (Stewart *et al.*, 2003). The ability of *M. tuberculosis* to interact and manipulate host immune functions allows latent infection to continue. Traditionally, it was thought that granulomas were protective against TB, however recent evidence indicates that *M. tuberculosis* may actually recruit immune cells to form granulomas, wherein the bacilli are able to persist in a protected environment until conditions are favorable for reactivation and dissemination (Davis & Ramakrishnan, 2009; Ramakrishnan, 2012). Granuloma formation walls off the tubercle bacilli, providing a variable microenvironment.

Latent TB infection is associated with the formation of at least two types of tuberculous granuloma (Barry *et al.*, 2009; Lin *et al.*, 2009). The classic caseous granuloma are composed of epithelial macrophage, neutrophils, and other immune cells surrounded by fibroblasts where the centre of the caseous necrotic region is hypoxic with *M. tuberculosis* residing in the hypoxic centre (Barry *et al.*, 2009; Via *et al.*, 2008). The other kind of granuloma seen in latent TB infection is fibrotic lesion composed almost exclusively of fibroblasts that contain very few macrophages. However, it is not clear whether *M. tuberculosis* is located inside macrophage or in the fibrotic area in the lesion (Ahmad, 2011). Within the macrophages it has been believed that latency is propagated by nutrient deprivation, hypoxia, nitric oxide, carbon monoxide and acid pH (Ghosh *et al.*, 2009; Kumar *et al.*, 2008; Rustad *et al.*, 2009; Ulrichs & Kaufmann, 2006) .

1.2.1 Evidence for genotoxic stress during host infection

During infection, the bacilli are subjected to damage by exogenous and endogenously generated oxidants and radicals such as reactive oxygen and reactive nitrogen intermediates (ROI and RNI, respectively) that are toxic to the pathogen (Darwin & Nathan, 2005). ROI and RNI can damage lipids, protein and nucleic acids (Sies & Cadenas, 1985; Sies & Mehlhorn, 1986). The activated macrophages express two enzymes, phagocyte oxidase (NOX2/gp91^{phox}) and inducible nitric oxide synthase (iNOS), which produce increased quantities of ROI and RNI that constitute the major stressor of *M. tuberculosis in vivo* (Ehrt & Schnappinger, 2009). Upon phagocytosis, NOX2 subunit assembles into an enzymatically active enzyme complex that transfers electrons across the membrane from cytosolic NADPH to molecular oxygen. This produces superoxide anions (O_2^-), which dismutate into hydrogen peroxide (H_2O_2) and generate toxic hydroxyl radicals (Bedard & Krause, 2007). The toxic effects imposed by the immune system can be limited by mechanisms such as reducing the expression of the iNOS in the host, thus allowing the bacilli to minimize the damaging effects of iNOS *in vivo* (Darwin & Nathan, 2005; Ehrt *et al.*, 2001; MacMicking *et al.*, 1997). Induction of high amounts of iNOS usually occurs in an oxidative environment, and thus high levels of nitric oxide (NO) have the opportunity to react with superoxide leading to peroxynitrite formation and cell toxicity. These properties may define the roles of iNOS in host immunity, enabling its participation in anti-microbial and anti-tumor activities as part of the oxidative burst of macrophages (Cole *et al.*, 2012; Mungrue *et al.*, 2002). An *in vitro* study utilized mildly acidified nitrite, a physiological source of NO, to screen for *M. tuberculosis* mutants with an impaired ability to survive the damaging effects of this

environment (Cho *et al.*, 2006; Ehrt & Schnappinger, 2009; Ramakrishnan, 2012). Among the mutants identified in the screen, two carried mutations in the nucleotide excision repair (NER) gene, *uvrB*, which resulted in profound sensitivity to acidified nitrite which mimics the environment within the host (Ramakrishnan, 2012). Studies prior to this indicated that a *recA* mutant of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) had no detectable phenotype in a mouse model for up to 80 days post infection (Sander *et al.*, 2001) whereas a *dnaE2* mutation attenuated *M. tuberculosis* in wild type C57BL/6 mice but did not affect bacterial load until late in the infection suggesting that mycobacteria have the ability to employ tolerance mechanisms in order to withstand the ongoing DNA damaging effects imposed by the environment (Boshoff *et al.*, 2003). RecA plays a role in DNA damage response by promoting autocleave of the LexA repressor protein that regulate expression of SOS response protein whereas DnaE2 has been shown to be involved in mycobacterial SOS response by performing translesion synthesis (Boshoff *et al.*, 2003; Warner *et al.*, 2010). Thus, although *recA* and *dnaE2* mutants had increased susceptibility to DNA damaging agents *in vitro*, they had distinct phenotypes in the mouse. This indicates that there is interplay between DNA repair/ damage tolerance that plays a role in survival and pathogenesis in *M. tuberculosis*.

1.2.2 A role for DNA repair in *M. tuberculosis* pathogenesis

DNA repair plays a fundamental role in maintenance of genome stability in order to remain fit and able to establish an infection, grow and persist within a host. *M. tuberculosis* contains all DNA repair mechanisms found in other bacteria except the mismatch repair (MMR) mechanism which plays a key role in post replication mismatch repair. This finding is interesting since it suggests that the fidelity of replication is compromised or there exists an alternative mechanism(s) to correct replication errors (Mizrahi & Andersen, 1998). *M. tuberculosis* has adapted to the lack of MMR by utilizing other repair systems in which the UvrD1 protein of the nucleotide excision repair (NER) system functions in the processing of mismatches by homologous recombination (Guthlein *et al.*, 2009).

NER was first identified as a pathway that repairs DNA lesions such as thymine dimers resulting from exposure to UV radiation. DNA damage such as abasic sites, DNA cross-links, strand breaks and deaminated bases generated by ROS and RNI and are also substrates for NER. In human mycobacteria isolated from macrophages and lung samples, genes involved in NER pathway were shown to be upregulated (Graham & Clark-Curtiss, 1999; Rachman *et*

al., 2006). Recently, a *M. tuberculosis* mutant deficient in UvrA was reported to be sensitive various DNA damaging agents (Rossi *et al.*, 2011) and another key member of the NER system – the UvrB protein – was shown to be important for bacterial survival both *in vitro* and *in vivo* (Darwin & Nathan, 2005; Kurthkoti & Varshney, 2011). Mycobacteria possess most of the conserved proteins that participate in base excision repair (BER) pathway except for RecJ, reinforcing the importance of excision repair pathways which the cell can use to prevent unwanted mutations by removing the vast majority of DNA damage imposed by the host environment (Kurthkoti & Varshney, 2011; Mizrahi & Andersen, 1998; Prammananan *et al.*, 2012). A number of studies have been done in *M. smegmatis* to show that the lack of these repair processes renders the cell sensitive to genotoxic stresses modelled on that encountered by *M. tuberculosis* in the host (Cordone *et al.*, 2011; Dick *et al.*, 1998; Dutta *et al.*, 2010; Gorna *et al.*, 2010; Lim & Dick, 2001; Nyka, 1974; Wayne & Hayes, 1996).

The toxic effects of the host can also be overcome by the use of recombination and end-joining repair processes which are able to repair double-stranded breaks in the chromosome to ensure bacillary viability. Amongst other bacteria, mycobacteria have been shown to utilize the non-homologous end joining (NHEJ) pathway which was originally thought to be limited to eukaryotes (Gong *et al.*, 2005; Weller *et al.*, 2002). A recent study has reported the unexpected identification of an active single-strand annealing (SSA) pathway in *M. smegmatis* to repair chromosomal double-strand breaks (DSBs) (Gupta *et al.*, 2011). SSA is a RecA-independent mechanism which requires LigD, a component of the NHEJ system. Prior to this study, DSB repair was catalysed by two mechanisms, homologous recombination (HR) and non-homologous end-joining pathways (NHEJ). HR is generally an error-free processes since it utilizes the non-damaged sister chromatid to direct repair whereas in bacteria including mycobacteria, a remarkably compact version of NHEJ is utilized in which all of the required activities are performed by only two proteins, a Ku homodimer and the multifunctional ligase/polymerase/nuclease LigD (Della *et al.*, 2004). The *M. tuberculosis* genome also contains a number of other predicted ligases whose potential role in NHEJ processes is unclear (Aniukwu *et al.*, 2008). The organism has LigA which is an essential replicative ligase (Gong *et al.*, 2005). Moreover, recent data have shown that a Sir2-like NAD-dependent deacetylase functionally interacts with Ku during NHEJ (Li *et al.*, 2011), suggesting that multiple interacting components might mediate NHEJ activity and function.

In addition to the threat posed to genome integrity, DNA lesions which hinder the function of RNA polymerase can have detrimental consequences for cellular homeostasis and survival.

Stalling of RNA polymerase triggers the transcription-couple repair (TCR) pathway which is an important subclass of NER that ensures that actively transcribed genes are repaired more rapidly than inactive regions of the genome. In bacteria including *M. tuberculosis*, TCR is mediated by the *mfd* (*m*utation *f*requency *d*ecline)-encoded protein (Prabha *et al.*, 2011) which is a homologue of the eukaryotic transcription-repair coupling factor (TCRF). In *M. tuberculosis*, Mfd does not recognize DNA damage directly, but through its interaction with the N-terminus of the RNA polymerase β subunit (Westblade *et al.*, 2010) it initiates a cascade of events which ensures the removal of the stalled RNA polymerase from the site of damage and recruitment of the NER components to effect the necessary repair.

1.3 DNA damage tolerance

Despite the existing repair mechanisms, some DNA lesions persist in the genome and compromise survival of the organism. Bacteria have evolved mechanisms that are able to overcome the detrimental effects of the unrepaired DNA lesions, but with the consequence of generating mutations. Such mechanisms utilize specialized translesion synthesis (TLS) polymerases that replicate past lesions (Yang & Woodgate, 2007). The classical specialized DNA polymerases that belong to the Y family of polymerase are structurally distinct from normal replicative polymerases in that they are capable of accommodating and replicating across lesions in DNA that obstruct progress by the normal replicative polymerase (Andersson *et al.*, 2010; Yang & Woodgate, 2007), and thus perform an important function in maintaining bacterial viability. It is important to note that not all TLS polymerases are error-prone because some have the ability to maintain high replication fidelity by incorporating the nucleotide that normally pairs with the undamaged nucleotide when replicating across a cognate lesion (Andersson *et al.*, 2010). While these polymerases can display high fidelity of replication across cognate lesions, their replication fidelity across non-cognate lesions or on undamaged templates is often reduced (Yang & Woodgate, 2007). For this reason, TLS function is often associated with mutagenesis (Andersson *et al.*, 2010). This is a phenomenon well studied in *Escherichia coli*, where TLS polymerases form part of the SOS response which up-regulates more than 40 genes following DNA damage (Erill *et al.*, 2006; Tippin *et al.*, 2004). In *E. coli*, the polymerases that are involved in the SOS response are DNA polymerase II, IV and V. *M. tuberculosis* has homologues of the Pol IV-type Y family polymerases, but these are not to induced in response to DNA damage (Boshoff *et al.*, 2003; Brooks *et al.*, 2001; Kana *et al.*, 2010). Instead, a C-family polymerase, DnaE2 was shown to be up-regulated in response to DNA damage (Boshoff *et al.*, 2003; Kana *et al.*, 2010).

Importantly, DnaE2 belongs to the C-family of DNA polymerases, which includes replicative polymerase, DnaE1.

It is worth mentioning that damage tolerance mechanisms can result in error-free or error-prone repair. In *E. coli*, Polymerase II is also used for TLS across abasic lesions when SOS is turned on in the absence of induction of the GroELS heat shock proteins (Tessman *et al.*, 1992), while error-prone polymerase V is responsible for error-free acetylaminofluorene (AAF) adduct bypass (Napolitano *et al.*, 2000). Polymerase ζ, a B-family polymerase from mammalian cells is able to incorporate efficiently two A nucleotides opposite a TT (6-4) photoproduct *in vitro*, resulting in error-free bypass of the lesion, but its function is dependent on the presence of Rev1 protein in a two-polymerase mechanism (Goodman, 2002; Rechkoblit *et al.*, 2002). Polymerase ζ cooperates with Rev1 to accomplish TLS past abasic site, with polymerase ζ extending from the mispaired C opposite the abasic site. Rev1 exhibits a second property in addition to deoxycytidyltransferase activity by acting in combination with polymerase ζ to achieve error-free bypass. Rev1 protein is the first recognized Y-family member which acts as a deoxycytidyltransferase that incorporates dCMP opposite abasic sites in yeast and humans (Lin *et al.*, 1999; Nelson *et al.*, 1996). Another example of error-free bypass is that performed by the Y-family polymerase κ *in vitro* and in mammalian cells. This polymerase is able to bypass bulky lesions linked to the amino group of guanine that resides in the B-DNA minor groove. The capability of this polymerase has been demostrated to bypass adducts caused by an environmental carcinogenic agent benzo[*a*]pyrene (B[*a*]P) present in tobacco smoke, automobile exhaust and cooked foods (Lior-Hoffmann *et al.*, 2012; Phillips, 1983) in an error-free manner (Huang *et al.*, 2003; Rechkoblit *et al.*, 2002; Zhang *et al.*, 2000; Zhang *et al.*, 2002).

1.3.1 Structural insight into Error-prone and Error-free lesion Bypass

The structure of the polymerase can determine whether a particular polymerase will replicate DNA in an error-free manner by performing highly accurate and processive DNA replication together with the exonuclease activity to enhance replication fidelity. Alternatively, replication can occur in an error prone fashion, with low processivity and lacking exonuclease activity. Several crystal structures of error-prone Y-family DNA polymerases have been solved from which valuable insights into their mode of action can be drawn. A crystal structure of the N-terminal region that includes the active site of *Saccharomyces cerevisiae* pol η has been determined and apart from containing the overall shape of DNA

polymerase, it contained a novel polymerase-associated domain (PAD), which mimics an extra set of fingers with the palm domain nearly superimposable with that of the high-fidelity polymerases. The PAD domain is also referred to as the little finger domain that contains the β-clamp binding domain. The O helices believed to be important for fidelity checking are absent in pol η (Marx & Summerer, 2002). Further insight into error-prone was gained through the crystal structure of *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) which can bypass *cis-syn* CPD lesions efficiently and thus, with respect to TLS, the enzymatic properties of Dpo4 are similar to that of the eukaryotic pol η (Ling *et al.*, 2001). The Dpo4 structure was successfully obtained in ternary complexes with the DNA primer template and either a canonical or non-canonical incoming nucleoside triphosphate showing a familiar shape of a half open right hand found in several polymerases. Dpo4 contains a little finger domain in addition to the finger, palm, and thumb domains. Again, the palm domain of Dpo4 is similar to high fidelity DNA polymerases with finger and thumb domains of both pol η and Dpo4 unusually small and O helices absent (Marx & Summerer, 2002).

Apart from the structural features of the polymerases that perform DNA replication, the site of adducts in the DNA is also a determinant of whether it will be repaired in an error-free or error-prone way. NMR solution studies have shown that the B[*a*]P-dG adduct resides in the B-DNA minor groove, directing the 5’ in-coming nucleotide along the modified strand (Lior-Hoffmann *et al.*, 2012). While minor groove lesions are bypassed nearly error-free by polymerase κ, lesions residing on the major groove side tend to cause the polymerase to stall (Jia *et al.*, 2008; Rechkoblit *et al.*, 2002; Sherrer *et al.*, 2012).

1.4 Mechanism of the SOS response

During normal growth, SOS genes are negatively regulated by the LexA repressor protein which binds to a 20-bp consensus sequence (the SOS box) in the operator region of those genes under SOS regulation. Some SOS genes are expressed even in the repressed state, depending on the affinity of LexA for the corresponding SOS box (Smith & Walker, 1998). Activation of the SOS genes occurs after DNA damage by the accumulation of single-stranded DNA (ssDNA) regions generated at replication forks, where the DNA polymerase is blocked. RecA forms a filament around these ssDNA regions in an ATP-dependent manner, and becomes activated. The activated form of RecA interacts with the LexA repressor to facilitate LexA self-cleavage, thus de-repressing the genes in the SOS regulon (Smith & Walker, 1998).

Once the pool of LexA decreases, repression of the SOS genes goes down according to the level of LexA affinity for the SOS boxes. Operators that bind LexA weakly are the first to be fully expressed. In this way, LexA can sequentially activate different mechanisms of repair. In *E. coli*, genes possessing a weak SOS box (such as *lexA*, *recA*, *uvrA*, *uvrB*, and *uvrD*) are fully induced in response even to weak SOS-inducing treatments. Thus, the first SOS repair mechanism to be induced is NER, ensuring high-fidelity repair of DNA damage without commitment to a full-blown SOS response (Smith & Walker, 1998).

If, however, NER alone is not sufficient to repair the damage, the LexA concentration is further reduced, so the expression of genes with stronger LexA boxes (such as *sulA*, *umuD*, *umuC* – these are expressed late) is induced. SulA inhibits cell division by binding to FtsZ, the initiating protein in this process (Chen *et al.*, 2012; Cordell *et al.*, 2003; Trusca *et al.*, 1998). This causes filamentation, and the induction of components of polymerase V, $UmuD'_2C$, which can result in mutagenic repair. As a result of these properties, some genes may be partially induced in response to endogenous levels of DNA damage, while other genes appear to be induced only when high or persistent DNA damage is present in the cell (Patel *et al.*, 2010; Smith & Walker, 1998). Research has shown that the SOS pathway may be essential to the acquisition of bacterial mutations which lead to resistance to some antibiotic drugs (Cirz *et al.*, 2005; Cirz & Romesberg, 2007). In these studies, the increased rate of mutation was shown to be dependent on the activity of the three SOS DNA polymerases: Pol II, Pol IV and Pol V.

Genome integrity is crucial to the survival and propagation of the organism, and therefore, there is a balance between DNA repair and damage tolerance pathways. DNA repair pathways ensure error-free repair whereas damage tolerance mechanisms allow selection of a bacterial population that is able to withstand the damaging effects, sometimes generating genetic diversity. The interplay between these opposing forces in turn impacts the relative fitness of the organism. The inferred importance of DNA repair for pathogenesis suggests some pathway components as potential targets for novel antibacterials. In theory, compounds could be designed to target repair components functioning in essential repair and maintenance pathways or those providing environment-dependent damage tolerance mechanisms (Cirz *et al.*, 2005; Cirz & Romesberg, 2007; Warner, 2010). The process of replication, repair and recombination favours the maintenance of genome stability whereas adaptive mutation,

which results from low fidelity damage tolerance, can fuel the evolution of pathogenic bacteria including *M. tuberculosis* (McGuire *et al.*, 2012).

1.5 Importance of tolerance mechanisms

Genome stability is important for propagation and maintenance of bacterial population; however, the environmental factors also play a role in genome diversity and selection. Introduction of mutations in the genome allows selection of a population that has inherited those mutations to enable survival in that particular environment. In many bacteria, drug resistance determinants are carried on mobile genetic elements. In contrast, the emergence of drug resistance in *M. tuberculosis* is associated exclusively with point mutations and chromosomal re-arrangements (Musser *et al.*, 1996; Musser *et al.*, 1999; Sandgren *et al.*, 2009), a feature that is thought to result in part from the ecological isolation of *M. tuberculosis* during host infection (Gillespie, 2002; Gillespie *et al.*, 2002). There are also non-heritable tolerance mechanisms that do not require acquisition of mutations in the genome in order to withstand environmental effects. Such mechanisms include the ability to form biofilms (Donlan & Costerton, 2002), permeability barriers (Nikaido, 1994; Nikaido, 2001) and persister formation (Dorr *et al.*, 2010). Of particular importance in the case of TB is the emergence of drug resistant strains which indicate the astonishing ability of *M. tuberculosis* to survive in the presence of the antitubercular drugs and also pose a challenge for completely eradicating the disease using chemotherapy.

The appearance of antimicrobial resistance is an inevitable consequence of the selective pressure imposed by antimicrobial therapy (Cirz *et al.*, 2005) resulting in spontaneous genetic mutations which confers resistance to a drug, thus driving evolution. This Darwinian principle of selection underlies the application of combination therapy. *M. tuberculosis* is subjected to harsh environmental conditions during infection, as well as antibiotic treatment in patients, once diagnosed. If the antibiotics are not appropriately delivered into the target site and not managed appropriately, they create a sub-lethal concentration that has the potential to select for the gradual evolution of resistance through the sequential acquisition of resistance-determining mutations. A potential strategy to counter the adaptive evolution of *M. tuberculosis* during host infection – including the development of drug resistance – involves targeting mutagenic pathways underlying the evolution of drug resistance.

1.6 Emergence of drug resistant mutants through adaptation

Stationary-phase and stress-induced mutagenesis in bacterial adaptation has been associated with the emergence of drug-resistant strains of pathogenic bacteria (Phillips, 1987), and might be of particular relevance to the generation of antibiotic and other stress resistance mutations in *M. tuberculosis* within the hostile environment (Gutacker *et al.*, 2002a; Gutacker *et al.*, 2002b; Ramaswamy & Musser, 1998). Mutation in general may be detrimental to a cell as it might decrease the fitness of the organism (Friedberg *et al.*, 2002). This was demonstrated in competition experiments between *E. coli* mutants lacking any of the TLS polymerases and wild-type cells: the cells lacking the TLS polymerase suffer a severe reduction in fitness in stationary phase suggesting that TLS polymerases are essential to prevent the appearance of adaptive mutations during prolonged stationary phase where cytotoxic alkylation damage accumulates (Andersson *et al.*, 2010; Bjedov *et al.*, 2007; Yeiser *et al.*, 2002).

TLS polymerases typically have low fidelity when replicating non-cognate DNA and are therefore under tight regulation. However, they must be readily available at the stalled replication fork to prevent obstruction of replication, and hence, are often induced in response to DNA damage as part of an SOS response (Goodman, 2002; Yang & Woodgate, 2007). Notably, Y-family polymerases do not usually have exonuclease activity, with the exception of the B-family polymerase that has exonuclease activity. Furthermore, while replicative polymerases utilize a mechanism that ensures correct base pairing at the active site, Y-family DNA polymerases have flexible active sites that permit bypass of distorting lesions with a consequent loss of fidelity (Bunting *et al.*, 2003; Napolitano *et al.*, 2000; Wagner *et al.*, 1999). The C-family polymerases are primarily bacterial replicative enzymes. A well-studied polymerase belonging to this group is DNA Polymerase III. It functions as a holoenzyme comprising α (polymerase), ε and θ subunits. DNA polymerase III holoenzyme is the primary enzyme complex involved in prokaryotic DNA replication. The complex has high processivity (i.e. the number of nucleotides added per binding event) and, in *E. coli*, functions in conjunction with four other DNA polymerases (Pol I, Pol II, Pol IV, and Pol V). Being the primary holoenzyme responsible for replication, the DNA Pol III holoenzyme also has proofreading capabilities that corrects replication mistakes by means of exonuclease activity working in a 3'→5' direction. DNA Pol III is a component of the replisome, which is located at the replication fork.

The replisome is composed of a hexameric DnaB helicase that encircles the lagging strand and uses ATP to translocate along single-strand DNA and separate the parental duplex. The clamp loader pentamer contains at least two τ subunits which have C-terminal extensions that protrude from the top of the clamp loader with each of the C-terminal extensions binding to a DnaB helicase subunit. In addition, to these components, there are two Pol III cores comprising the α subunit which possess the polymerase activity, the ε subunit which has 3'→5' exonuclease activity, and the θ subunit which stimulates proofreading by the ε subunit. The two τ subunits act to dimerize the two core enzymes together with one γ unit of the clamp loader for the lagging strand Okazaki fragments, thus helping the two β subunits to form a unit and bind to DNA. The assembly containing the clamp loader and two molecules of Pol III binds a β-clamp to form a holoenzyme. Two β subunits act as sliding DNA clamps that keep the polymerase bound to the DNA. The one Pol III functions in the leading strand and the other in the lagging strand, with the lagging strand requiring synthesis of RNA primer templates by primase interacting with DnaB (Georgescu *et al.*, 2010; Langston & O'Donnell, 2006). Two distinct Pol IIIs are required for synthesis with PolC serving as the polymerase for leading strand synthesis, and DnaE synthesizing the lagging strand in the Gram+ organism, *Bacillus subtilis* (Kelman & O'Donnell, 1995; Olson *et al.*, 1995; Sanders *et al.*, 2010). The polymerases, either belonging to the Y family or C family, require interactions with the β clamp to access template DNA. This allows efficient switching between replication and repair modes for efficient lesion by-pass (Friedberg *et al.*, 2005; Langston & O'Donnell, 2006).

1.7 SOS response in Mycobacteria

Although an SOS response has been identified in mycobacteria (Davis *et al.*, 2002; Durbach *et al.*, 1997), it is unusual in that none of the predicted Y-family polymerases is included in the damage response regulon (Boshoff *et al.*, 2003; Kana *et al.*, 2010; Smollett *et al.*, 2012). Instead, Boshoff *et al.* (Boshoff *et al.*, 2003) showed that the only DNA polymerase that was induced in transcriptome analysis of *M. tuberculosis* after exposure to various DNA damaging agents was DnaE2. Subsequent to that study, DnaE1 was also shown to be up-regulated in response to DNA damage (Warner *et al.*, 2010). Although the *M. tuberculosis* genome contains two genes encoding the catalytic (α) subunit of DNA Pol III, the encoded proteins, DnaE1 and DnaE2, are not functional alternatives (Boshoff *et al.*, 2003). DnaE1 provides replicative polymerase activity and DnaE2 is part of the SOS response in *M. tuberculosis* (Tippin *et al.*, 2004). The involvement of a C-family polymerase in a bacterial

DNA damage response was novel and defined a characteristic feature of the SOS response operating in *M. tuberculosis and M. smegmatis* (Boshoff *et al.*, 2003; Kana *et al.*, 2010; Smollett *et al.*, 2012; Warner *et al.*, 2010). Subsequent to the study by Boshoff et al. (2003) several other groups identified an association between DnaE2-type C-family DNA polymerases and two other proteins, termed ImuA (for *i*nducible *mu*tagenesis) and ImuB (Abella *et al.*, 2004; Erill *et al.*, 2006; Galhardo *et al.*, 2005). In particular, these authors observed that bacterial genomes containing a DnaE2-type DNA polymerase always encode an ImuB orthologue (Figure 1**)** and, further, that the ImuA-ImuB-DnaE2 triad – often present in these genomes as an SOS-inducible gene cassette – might represent the functional equivalent of the mutagenic DNA polymerase V (encoded by *umuDC*), most well characterised in *E. coli* (Galhardo *et al.*, 2005; Koorits *et al.*, 2007). Included with *dnaE2* in the mycobacterial SOS response is an operon encoding a protein of unknown function, Rv3395c, and its partner, Rv3394c (Figure 1). The interesting feature of Rv3395c is that its orthologues are limited to Actinobacterial genomes including *M. tuberculosis.* Because Rv3394c, Rv3395c and DnaE2 are DNA damage induced, it suggests that they might function in the same pathway.

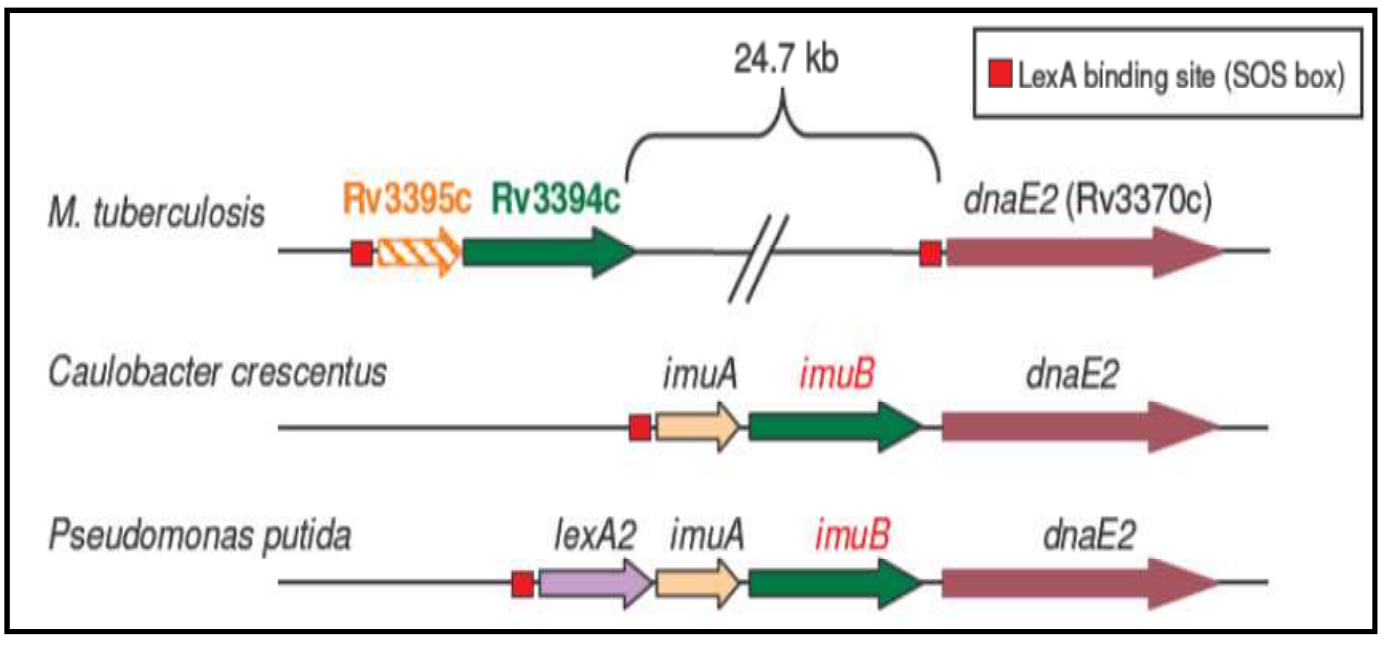


Figure 1: An SOS-inducible mutagenic gene cassette. The *M. tuberculosis* genome encodes principal components of a DNA damage-inducible cassette, DnaE2 and Rv3394c, separated by ~25kb. Orthologues of Rv3395c are found only in Actinobacterial genomes.

Rv3394c was designated the founder member of the DinB3 subfamily of Y-polymerases because of its predicted β-binding motif (Dalrymple *et al.*, 2003). The β-clamp modulates the recruitment to the replication fork of many proteins involved in various steps of DNA replication, recombination, and repair, with the current model holding that these processivity factors function as molecular “tool-belts” (Pages & Fuchs, 2002) to attach replicative and TLS polymerases to the primer-template junction interchangeably without disruption of the core assembly. It has been shown by bioinformatic analysis that DnaE2 does not have a β-binding motif and Tippin et al (2004) suggested that DnaE2 may require other proteins in order to access the replication fork. The absence of an identifiable β-clamp binding motif in DnaE2 poses an important question as to how it accesses the replication fork.

1.8 Working hypothesis

Previous work from the MMRU identified the DNA damage-inducible C-family DNA polymerase, DnaE2, and which was subsequently implicated in virulence and the emergence of rifampicin in *M. tuberculosis in vivo* (Boshoff *et al.*, 2003). However, a fundamental question at the time pertained to the ability of DnaE2 to access the DNA replication fork given the apparent absence of a predicted clamp binding motif in the mutagenic polymerase. Therefore, the failure to identify a putative β-binding motif suggested the operation of a novel mutagenic mechanism. Subsequently, the two hypothetical proteins, Rv3394c (ImuB) and Rv3395c (ImuA’) were shown to function as essential accessory factors in DnaE2-mediated mutagenesis. Specifically, disruption of either gene eliminates damage-induced mutagenesis, and rendered the mutant strains hypersensitive to DNA damaging agents, thereby

recapitulating precisely the *dnaE2* deletion phenotype. Rv3395c encodes a protein of unknown function restricted to Actinomycetes; in contrast, Rv3394c is an orthologue of ImuB-type proteins (implicated in DNA repair and mutagenesis in other organisms) and, like ImuB, possesses a predicted β-clamp binding motif. This is significant since it raises the possibility that Rv3394c might function as molecular adaptor, directly mediating the interaction of DnaE2 with template DNA. These data established a working model in which all three proteins – Rv3394c, Rv3395c and DnaE2 – must interact in order for DnaE2 to access the replication fork. However, the nature of this interaction, as well as the specific residues required for interaction were unknown, but would be critical to the design of novel compounds to disrupt complex formation.

1.9 Aims and Objectives

The aim of this study was to define key functional and structural characteristics of the replication complex responsible for DNA damage-induced mutagenesis in *M. tuberculosis*.

To achieve the aim the following objectives were pursued:

- To elucidate, by targeted (genetic) disruption of protein-binding domains, the functional interactions that are essential for DnaE2-dependent mutagenesis;
- To identify other proteins involved in DnaE2-dependent mutagenesis
- To express and purify sufficient quantities of ImuB and ImuA' protein to enable subsequent crystallization and structural analysis.
- To use proteomics as a tool to examine the role of DnaE2 in other pathways under conditions of normal growth or genotoxic stress conditions

Chapter 2

2. Materials and Methods

All general procedures and DNA manipulations were performed according to standard protocols (Sambrook, 2001; Sambrook, 1989). All culturing and molecular biological manipulations of mycobacterial were performed according to published methods (Larsen, 2000; Parish & Stoker, 2000; Snapper *et al.*, 1990). The composition of culture media and solutions are detailed in the Appendix.

2.1. Bacterial strains, plasmids and culture conditions

All bacterial strains and cloning vectors used in this study are listed in Tables 2.1 and 2.2. Glycerol stocks of bacterial strains were prepared in 33.3% glycerol (V/V) and stored at -70°C.

Table 2.1: Strains used in the study

Strain	Description	Reference/ Source
Yeast		
AH109	*HIS3, ADE2, lacZ, MEL1* containing reporter genes for screening protein interactions	Clontech
M. smegmatis		
mc^{2}155	High frequency transformation mutant of *M. smegmatis* mc^{2}6; ATCC 706	(Snapper *et al.*, 1990)
Δ*dnaE2*::*aph*	*dnaE2* knockout mutant of mc^{2}155; Km^R	(Boshoff *et al.*, 2003)
Δ*dnaE2*::*aph attB*::*dnaE2*	*dnaE2*::*aph* complemented with *dnaE2* at *attB* site; Km^{R}Hyg^R	(Boshoff *et al.*, 2003)
Δ*imuA'*	*imuA'* deletion mutant of mc^{2}155	(Warner *et al.*, 2010)
Δ*imuA attB'*::*imuA'*	Δ*imuA'* complemented with full-length *imuA'* at *attB* site; Km^R	(Warner *et al.*, 2010)
Δ*imuA attB'*::*imuA*^{N3XFLAG}	Δ*imuA'* complemented with *imuA*^{N3XFLAG} allele at *attB* site; Km^R	This study
Δ*imuB*	*imuB* deletion mutant of mc^{2}155	(Warner *et al.*, 2010)
Δ*imuB attB'*::*imuB*	Δ*imuB* complemented with full-length *imuB* at *attB* site; Km^R	(Warner *et al.*, 2010)
Δ*imuB attB'*::*imuB*^{N3XFLAG}	Δ*imuB* complemented with *imuB*^{N3XFLAG} allele at *attB* site; Km^R	This study
Δ*imuA'-imuB*	*imuA'-imuB* double knockout mutant of mc^{2}155	(Warner *et al.*, 2010)
Δ*imuA'-imuB attB'*::*imuA'-imuB*	Δ*imuA-'imuB* complemented with full-length *imuA'-imuB* at *attB* site; Km^R	(Warner *et al.*, 2010)

Δ*imuA'-imuB attB'::imuA'-imuB*^{C168}	Δ*imuA-'imuB* complemented with *imuA'-imuB*^{C168} *allele* at *attB* site; Km^R	(Warner *et al.*, 2010)
Δ*imuA'-imuB attB'::imuA'-imuB*$^{CHis\text{-}FLAG}$	Δ*imiA'-imuB* complemented with *imuA'-imuB*$^{CHis\text{-}FLAG}$ allele at *attB* site; Km^R	Dr. D. Warner
E. coli		
DH5α	*E. coli* strain used for transformation and propagation of plasmids	Laboratory stock
BL21 (DE3)	*E. coli* strain used for heterologous expression of *M. tuberculosis* proteins.	Laboratory stock

Table 2.2: Plasmids used in the study

Plasmid	Description	Source
pAINT	*E. coli–Mycobacterium* integrating shuttle vector; Km^R,Ap^R	Laboratory stock
pAMSIMUA	*M. smegmatis imuA′* complementation vector – pAINT carrying full-length *M. smegmatis imuA′*; Km^R	Dr. D. Warner
pAMSIMUAB	*M. smegmatis imuA′-imuB* complementation vector – pMC1r carrying full-length *M. smegmatis imuA'-imuB*, Km^R	Dr. D. Warner
pAINT1620stop 1	*M. smegmatis imuA′* complementation vector – pAINT carrying containing truncated *imuA′*$^{\Delta C}$; Km^R	This study
pGADT7	Y2H vector to produce AD fusions, GAL4(768-881) AD, LEU2, HA epitope tag; Ap^R	Clontech
pGBKT7	Y2H vector to produce BD fusions, GAL4(1-147) DNA-BD, TRP1, c-MYC epitope tag; Km^R	Clontech
pGADDnaE1	pGADT7 containing *M. tuberculosis dnaE1* ORF fused to GAL4 AD, Ap^R	Dr. G. Abrahams
pGADDnaE2	pGADT7 containing *M. tuberculosis dnaE2* ORF fused to GAL4 AD, Ap^R	Dr. G. Abrahams
pGADDnaN	pGADT7 containing *M. tuberculosis dnaN* ORF fused to GAL4 AD, Ap^R	Dr. G. Abrahams
pGADImuA′	pGADT7 containing *M. tuberculosis imuA′* ORF fused to GAL4 AD, Ap^R	Dr. G. Abrahams

pGADImuB	pGADT7 containing *M. tuberculosis imuB* ORF fused to GAL4 AD, Ap^R	Dr. G. Abrahams
pGADImuB stop1	pGADT7 containing *ImuB* with 168 amino acids truncated on the C-terminus fused to GAL4 AD, Ap^R	This study
pGADImuBstop 2	pGADT7 containing *ImuB* C-terminally truncated from the β-clamp binding motif fused to GAL4 AD, Ap^R	This study
pGADImuBstop 3	pGADT7 containing *ImuB* C-terminally truncated from the little finger domain fused to GAL4 AD, Ap^R	This study
pGADImuBstop 4	pGADT7 containing *ImuB* C-terminally truncated from the thumb domain fused to GAL4 AD, Ap^R	This study
pGADImuBstop 5	pGADT7 containing *ImuB* C-terminally truncated from the thumb domain fused to GAL4 AD, Ap^R	This study
pGADImuBstart 1	pGADT7 containing the C-terminal region of ImuB fused to GAL4 AD, Ap^R	This study
pGBDImuA'	pGBKT7 containing *M. tuberculosis ImuA'* ORF fused to GAL4 BD, Km^R	Dr. G. Abrahams
pGBDImuA' start1	pGBKT7 containing *ImuA'* truncated on the N-terminal domain ORF fused to GAL4 BD, Km^R	This study
pGBDImuA' start2	pGBKT7 containing *ImuA'* truncated on the N-terminal domain ORF fused to GAL4 BD, Km^R	This study
pGBDImuA'	pGBKT7 containing C-terminally truncated truncated *ImuA'* fused to GAL4 BD, Km^R	This study
pMalc2	*E. coli* expression vector containing maltose binding protein tag; Ap^R	New England Biolabs
pMal13	pMalc2 containing full-length *M. tuberculosis ImuB* N-terminally fused to Maltose binding protein tag, Ap^R	This study
pMal14	pMalc2 containing full-length *M. tuberculosis ImuA'* N-terminally fused to Maltose binding protein	This study
pMal13stop 1	pMalc2 containing C-terminally truncated *ImuB* N-terminally fused to Maltose binding protein, Ap^R	This study

2.1.1 Bacterial growth conditions

Escherichia coli strains

E. coli cells containing plasmids were grown in Luria-Bertani (LB) broth supplemented with appropriate selective agent at 37°C overnight, with vigorous shaking (Labcon Shaking Incubator) or at 30°C in a New Brunswick Scientific Innova 400 shaker. For selection on solid media, *E. coli* strains containing plasmids were grown on Luria-Bertani agar (LA) containing appropriate selective agent at 37°C overnight in an Incotherm Labotec Incubator or 48 h at 30°C in a Heraeus Instrument Incubator. *E. coli* cells carrying plasmids greater than 8 kb in size were grown at 30°C standing to limit plasmid rearrangement.

Selection agents used for strains carrying plasmids were as follows: 100 µg/ml ampicillin (Ap), 200 µg/ml hygromycin (Hyg), and 50 µg/ml kanamycin (Km). For confirmation of disruption of the *lacZ*α cassette during cloning or identification of a clone containing the *lacZ*α cassette, 40 µg/ml of 5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal) and 4µg/ml of its substrate isopropyl-β-D-thiogalactopyranoside (IPTG) were added to the growth medium.

Mycobacterial strains

Unless otherwise indicated, *M. smegmatis* strains were grown in Middlebrook 7H9 medium supplemented with 0.05% Tween 80 or on Middlebrook 7H10 solid medium both supplemented with glucose salt (0.085% NaCl, 0.2% glucose) and 0.5% glycerol.

The selection agents used for strains carrying plasmids were as follows: 50 µg/ml Hyg and 20 µg/ml Km. Resistant mutants were selected on Middlebrook 7H10 solid medium supplemented with Oleate-albumin-dextrose-catalase (OADC), 0.5% glycerol and rifampicin (Rif) 200 µg/ml.

Yeast strains

Saccharomyces cerevisiae strains, unless otherwise stated, were grown in rich YPD broth (0.1% Yeast extract, 0.2% Peptone and 0.2% Glucose) or on YPD agar (0.15% agar) both supplemented with 0.003% adenine.

Selection of strains carrying plasmids was on minimal synthetic drop-out (SD) media with leucine and tryptophan excluded. Drop-out medium (0.67 % Yeast nitrogen base without amino acids, 2 % Glucose, 2 % Bacto agar and 0.02 % drop-out mix) is used for the selection

of plasmids based on the use of auxotrophic mutant strains which cannot grow without a specific media component. Transformation with a plasmid containing the gene required for growth in the absence of the specific component enables the transformant to grow on a medium lacking the required component, allowing for selection.

2.2 General DNA manipulations

2.2.1 DNA Extraction

E. coli mini plasmid preparation

An individual bacterial colony was inoculated into 1 ml of LB containing the appropriate selective agent in a 2-ml Eppendorf tube, which was incubated at 37°C with agitation overnight. The cells were harvested by centrifugation at room temperature for 1 min. The supernatant was decanted and the pellet re-suspended in 80 µl of Solution I (50 mM glucose, 25 mM Tris hydrochloride, 10 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0) by vortexing. Then, 160µl of Solution II (0.2 M sodium hydroxide, 1% Sodium dodecyl sulphate (SDS)) was added to the cell suspension and mixed by gentle inversion of the tube. Thereafter, 120 µl of Solution III (5M potassium acetate, 11.5 % glacial acetic acid) was added and the mixture shaken vigorously and centrifuged (13200 rpm; 16 100 x*g*) at room temperature for 10 min to remove cellular debris. The supernatant was collected into a new sterile Eppendorf tube and isopropanol (220 µl) was added. The precipitation process was allowed to continue at room temperature for 5 min before the DNA was harvested by centrifugation at a speed of 13200 rpm at room temperature for 10 min. The pellet was washed with 150 µl ethanol and vacuum-dried for 20 min (SpeedVac, Savant, Farmingdale, NY, USA). The DNA was then re-suspended in 50 µl of sterile distilled water containing freshly boiled ribonuclease (RNase) (1 µl of 10 mg/ml). A small aliquot was analyzed on an agarose gel.

E. coli bulk plasmid preparation

A single colony was used to inoculate 100 ml of LB supplemented with the appropriate selective agent for maintenance of the plasmid. The culture was grown with gentle agitation at 37°C for overnight. The cells were transferred into two 50-ml Falcon tubes and harvested by centrifugation at 4500 rpm (3900 × *g*) in a Beckman Coutler Allegra X-22 (4250 rotor) centrifuge for 10 min. The cell pellets were resuspended in 1 ml of Solution I and to the suspension, 2 ml of Solution II was added followed by gently inverting the tubes for mixing.

To the mixture, 1.5 ml of Solution III was added and shaken vigorously and then 900 µl aliquots transferred into ten 1.5-ml Eppendorf tubes. Cell debris was removed by centrifugation at 13200 rpm (16 100 × *g*) in an Eppendorf Centrifuge 5415 D with standard rotor F-45-24-11(Eppendorf, Hamburg, Germany) for 10 min. The supernatant was transferred to a fresh, sterile 1.5 ml Eppendorf tube and 3 µl RNaseA (10 mg/ml stock) was added to each tube, followed by incubation at 42°C for 1 h. To each tube, 700 µl of isopropanol was added followed by centrifugation for 10 min at 13200 rpm at room temperature to precipitate the DNA. The supernatant was decanted and the DNA pellet washed with 500 µl of ice-cold 70 % ethanol. The DNA from each tube was resuspended in 100 µl sterile distilled water (sdH_2O) and the DNA from five tubes pooled together to a volume of 500 µl to which was added 50 µl of 5.3 M sodium chloride (pH 5.2) and 700 µl of phenol:chloroform (1:1). The mixture was centrifuged at room temperature for 10 min at 13200 rpm and the top aqueous layer transferred into a clean Eppendorf tube. A volume of 350 µl of chloroform:isoamyl alcohol (24:1) was added to the aqueous solution and the mixture centrifuged for 10 min at room temperature. The aqueous layer was transferred to a clean Eppendorf tube to which was added 1 ml of ice-cold 100 % ethanol. The tube was inverted for mixing and kept at -20°C for 30 min to 24 h before centrifugation at room temperature for 25 min. DNA pellet was washed with 70% ethanol and vacuum-dried with a Speed-Vac. The DNA was resuspended in 50-200 µl of sdH_2O and quantified using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific) according to the manufacturer's instructions.

Chromosomal DNA extraction from mycobacteria

Mycobacterial chromosomal DNA was isolated using a modified cetyltrimethylammonium bromide (CTAB; ICN Biomedicals, Aurora, Ohio) method (Larsen, 2000; Warner *et al.*, 2010). Briefly, mycobacterial cells were harvested and re-suspended in 500 µl of TE buffer (10 mM Tris hydrochloride pH 8.0; 1 mM EDTA). The cells were heat-killed at 65°C for 10 min for *M. smegmatis*, harvested by centrifugation (13200 rpm (16 100 × *g*) for 5 min and re-suspended in 500 µl of TE buffer. To this was added 50 µl lysozyme (10 mg/ml) and the reaction incubated at 37°C overnight. 70µl 10 % SDS and 6 µl of proteinase K (10 mg/ml) was then added and the mixture incubated at 65°C for 2 h. 100 µl of a 5 M solution of sodium chloride and 80 µl of pre-warmed CTAB/NaCl mix (10 % CTAB made in 0.7 M sodium chloride) was added to the sample and incubated at 65°C for a further 10 min. An equal

volume of chloroform:isoamyl alcohol (24:1) was added to remove residual proteins. Subsequent to centrifugation (13200 rpm for 10 min) the aqueous phase containing the DNA was precipitated by addition of 1/10 × volume of 5.3M sodium acetate (pH 5.2) and 2.5 × volumes of 100 % ethanol, and incubation at -20°C for 1 h. The DNA was then pelleted by centrifugation (13200 rpm for 20 min), washed with ice-cold 70% ethanol, dried in a vacuum centrifuge and resuspended in sdH_2O.

2.2.2 DNA manipulations

Restriction enzyme digestion

Enzymes were obtained from Fermentas, Boehringer Mannheim, New England Biolabs (NEB), Amersham or Roche Applied Science and used according to manufacturer's instructions. Plasmid DNA of up to 5 µg was digested in a 20 µl reaction volume for 1-24 h. Digested samples were incubated at the recommended temperature for maximal enzyme activity for at least 4 h. For double digestions, an appropriate buffer in which both restriction enzymes showed suitable activity was selected, otherwise the digestions were performed sequentially starting with the enzyme that require the lower pH buffer. DNA fragments were then separated on agarose gels by electrophoresis (See section 2.2.3).

Phoshorylation of DNA

Phosphorylation of blunt-ended PCR products to enable ligation with de-phosphorylated vector was performed for 30 min at 37°C using polynucleotide kinase (Roche Applied Science), as per the manufacturer's instruction. Reactions were stopped by separation on an agarose gel.

Dephosphorylation of 5' ends of plasmid DNA

Following plasmid digestion, the 5'-phosphate of linearized vector DNA was removed by treatment with Antarctic Alkaline Phosphatase (New England Biolabs) to prevent vector re-ligation. Dephosphorylation reactions were carried out for 1-12 h at 37°C, after which the enzyme was heat inactivated for 20 min at 65°C. The DNA was cleaned by using the Nucleospin kit (Macherey-Nagel) as per manufacturer's instructions. Briefly, the de-phosphorylated DNA was loaded onto a Nucleospin column, and washed. The DNA was then eluted using pre-warmed sdH_2O. The DNA was quantified either on agarose gels by

comparison to DNA molecular weight markers or using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

Ligation reactions

DNA ligations were performed using either the Fast-Link™ ligation kit (Epicentre® Biotechnologies) or T4 DNA Ligase (Roche Applied Science), as per instructions from the manufacturer. The ligation reactions were then used for transformation into *E. coli* DH5α cells (See section 2.2.5).

2.2.3 Agarose gel electrophoresis

Standard gel electrophoretic techniques were used to separate DNA fragments (Sambrook, 1989). For the separation of high molecular weight DNA fragments, 0.8 -1 % agarose gels, prepared in 1×TAE buffer (1 mM EDTA, 40 mM Tris-acetic acid, pH 8.5) were used. For low molecular weight DNA fragments of ≤1kb, 2 % agarose gels were used. All gels contained 0.5 µg/ml ethidium bromide and the DNA samples were loaded with a tracking dye (0.025 % bromophenol blue in 30% glycerol). Lambda DNA molecular weight markers (III, IV and V; Roche Applied Science) were used to assess DNA fragment sizes. The agarose gels were electrophoresed between 80 – 100 V in a Mini-Sub Cell GT mini-gel horizontal submarine unit (Bio-Rad) and visualized under UV-light using the G:Box (Syngene) gel documentation system.

2.2.4 DNA fragment recovery from agarose gels and quantification

The required DNA fragment was excised from the gel and purified using the Nucleospin kit (Macherey-Nagel) as per manufacturer's instructions. Briefly, the excised gel fragment was melted in a binding buffer by incubation at 60˚C and loaded onto a Nucleospin column, and washed. The DNA was then eluted using pre-warmed sdH_2O. The DNA was quantified either on agarose gels by comparison to DNA molecular weight markers or using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

2.2.5 Transformation of bacteria

Chemical transformation of *E. coli*

Chemically competent *E. coli* DH5α cells were used for transformation of plasmids. The cells were prepared by rubidium chloride treatment, as follows. A 1-ml aliquot of a stationary

phase overnight culture was inoculated in 100 ml of LB and grown to an OD_{600} of 0.48 - 0.55. The cells were chilled on ice for 15 min and harvested by centrifugation at 4500 rpm ($3900 \times g$) for 5 min at 4°C). The pellets were re-suspended in 20 ml of transformation buffer (Tfb) I solution (30 mM potassium acetate; 100 mM rubidium chloride; 10 mM calcium chloride; 50 mM manganese chloride; 15% v/v glycerol; pH 5.8), and chilled on ice for 15 min. The cells were harvested by centrifugation at 4500 rpm for 5 min at 4°C), re-suspended in 2 ml Tfb II solution (10 mM MOPS; 75 mM calcium chloride; 10 mM rubidium chloride; 15 % glycerol; pH 6.5) and 500 µl aliquots were flash-frozen in ethanol and stored at -80°C until further use.

For transformations, *E. coli* DH5α competent cells were thawed on ice and 100 µl of cells used per transformation. Up to 1µg of plasmid DNA was incubated with the cells on ice for 1h, heat-shocked for 90 s at 42°C and chilled on ice for 2 min. Four volumes of 2×TY (0.16 % Bactotryptone, 0.10 % Bacto yeast extract and 0.05% sodium chloride) medium was added to rescue the cells at 37°C for 1 h. These were plated on LA media containing the appropriate antibiotics, and incubated for 1 - 2 days at 37°C.

Electroporation into *Mycobacterium smegmatis*

M. smegmatis electroporations were carried out as previously described (Larsen, 2000). Briefly, 1 ml of a stationary-phase *M. smegmatis* culture was inoculated in 100 ml of LB containing 0.05% Tween 80 and grown to an OD_{600} of 0.4 - 0.7. The cells were harvested by centrifugation at 3500 rpm ($2000 \times g$) for 10 min at 4°C) and the pellet washed three times by gentle re-suspension in 10 ml of ice-cold 10% glycerol before collection by centrifugation at 3500 rpm for 10 min at 4°C. The pellet was re-suspended in 1 ml of ice-cold 10% glycerol and cells used immediately. Up to 5 µg of plasmid DNA was added to 400 µl *M. smegmatis* competent cells. The mixture was transferred to a 0.2 cm electroporation cuvette and pulsed using the following settings: 2.5 kV, 25 µF and 1000 Ω. The cells were rescued immediately with 1 ml of 2×TY for at least 5 h at 37°C before plating on Middlebrook 7H10 media supplemented with appropriate antibiotics. The plates were and incubated for 3 - 7 days at 37°C before scoring CFUs.

Lithium acetate-mediated yeast transformation

A single colony from YPD agar plate or freezer stock was inoculated into YPD broth and vigorously vortexed to disperse the clumps followed by incubated at 30˚C with shaking for

48 h. 1 ml of the culture was inoculated into 50 ml YPD broth and incubated at 30˚C for 16–18 h with shaking at 250 rpm until the culture reached stationary phase (OD_{600} > 1.5). A 30-ml aliquot of the culture was transferred to a flask containing 300 ml YPD broth and the resulting culture (OD_{600} =0.2-0.3) was incubated at 30˚C with shaking for 3 h to reach an OD_{600} of 0.4-0.6. Cells were transferred into six 50-ml tubes and centrifuged at 2278 rpm (1000 × *g)* for 5 min at room temperature. The supernatant was discarded and the pellets in each tube re-suspended in 25 ml sdH_2O. The re-suspended cells were pooled into three 50-ml tubes to a final volume of 50 ml and centrifuged at 1000 × *g* for 5 min at room temperature. The supernatant was decanted and pellets re-suspended in 1.5 ml freshly prepared, sterile 1×TE/1×LiAc solution (10 mM Tris-hydrochloride, 1 mM EDTA, pH 8.0 plus 0.1 M lithium acetate, pH 7.5).

Plasmid DNA (0.1 µg) and 0.1 mg of herring testes carrier DNA (Invitrogen) were added to a fresh 1.5 ml tube and mixed thoroughly. 100µl of competent yeast cells were added to the tube containing DNA and the sample was mixed by vortexing. To this tube, 600 µl of sterile PEG/LiAc solution (40% Polyethylene glycol 3350, 10 mM Tris-hydrochloride, 1 mM EDTA, pH 8.0 plus 0.1 M lithium acetate, pH 7.5) was added and mixed by vortexing at high speed for 10 sec followed by incubation at 30℃ for 30 min with shaking at 200 rpm. Seventy µl of dimethyl sulfoxide (DMSO) was added and mixed by gently inversion and the suspension was heat-shocked for 15 min at 42˚C then chilled on ice for 1-2 min. The cells were centrifuged for 5 s at 13200 rpm (16 100 × *g*) at room temperature and 500 µl TE buffer (10 mM Tris-hydrochloride, 1 mM EDTA, pH 8.0) was added once the supernatant was removed. The suspension was plated on synthetic drop-out (SD) agar to select for the desired transformants and the plates incubated at 30˚C for 3-5 days.

2.2.6 Polymerase Chain Reaction (PCR)

All preliminary and screening PCRs were performed using the Roche FastStart kit (Roche Applied Science), while for the amplification of fragments required for cloning, Phusion High-Fidelity DNA polymerase (Finnzymes) was used. The reactions were performed as per the manufacturer’s instructions. For reactions using the Roche FastStart Taq DNA polymerase, 20-50 µl reactions were set up containing: 1× reaction buffer, up to 50 ng plasmid or 200 ng genomic DNA (gDNA), 200 µM of each dNTP, 0.5-1.0 µM of each primer, 1.5 mM magnesium chloride, 1× GC-rich solution and 2U per 50 µl reaction of the DNA polymerase. DNA amplification was performed using the following cycling

parameters: denaturation at 94°C for 5 min; followed by 30 cycles of denaturation at 94°C for 60s, annealing for 30 s at 60˚C and extension at 72°C for 60 s; with a final extension at 72°C for 7 min. For reactions using the Phusion High-Fidelity DNA polymerase (Finnzymes), 20–50 µl reactions were set up to contain: 1× reaction buffer, up to 50 ng plasmid or 200 ng genomic DNA (gDNA), 200 µM each dNTP, 0.5 µM each primer, 3% DMSO and 0.02 U/µl of the DNA polymerase. DNA amplification was then performed using the following cycling parameters: denaturation at 98°C for 30 s; followed by 25-35 cycles of denaturation at 98°C for 10 s, annealing for 30 s, and extension at 72°C for 30 s/kb; with a final extension at 72°C for 7 min. For each amplification reaction, three control reactions (a no DNA control reaction; a reaction containing only the forward primer and another containing only the reverse primer) were used to elucidate the presence of genomic contamination and/or non-specific DNA amplification if present. All PCRs were performed using a MyCycler™ thermal cycler (Bio-Rad) with oligonucleotide primers obtained from Inqaba Biotech Ltd.

2.2.7 DNA Sequencing

DNA sequencing was outsourced to either Inqaba Biotech Ltd (South Africa) or the Central Analytical Facilities-DNA Sequencing Unit of Stellenbosch University, and was performed using the Big Dye terminator v3.1 Cycle Sequencing kit and Bioline Half Dye Mix. The EditSeq and SeqMan™ modules of the Lasergene 99 suite of software (DNASTAR, Inc., Madison, Wisconsin, USA.) were used to analyse the sequencing data.

2.3 Construction of vectors for interaction studies

2.3.1 Construction of vectors for yeast two hybrid system

Genes of interest were cloned in-fusion with activator (AD) and binding (BD) domains of the GAL4 transcription regulator (Kana *et al.*, 2010; Warner *et al.*, 2010). Artificial start and stop codons were introduced into *imuB* and *imuA'* to generate truncated proteins on the N- or C-terminus, respectively.

For *imuB*, primers in Table 2.3 were used to PCR amplify the truncated versions using pGADImuB as a template. The resulting PCR products contained 5' *Nde*I and 3' *Eco*RI restriction sites for cloning into pGADT7. For *imuA'*, primers in Table 2.3 were used to amplify the truncated version using pGBKImuA' as a PCR template and resulting PCR products cloned into pGBKT7 using the flanking 5' *Nde*I and 3' *EcoR*I restriction sites, respectively. Correct clones were confirmed by restriction analysis and sequencing.

Table 2.3: Primers used to introduce artificial start and stop codons into the *ImuB* and *ImuA'* ORFs[a]

Primer Name	Primer sequence
ImuB-NdeI-Forward	5'-ATAAAACATATGATGGCCTCCGCTCGC-3'
ImuB-EcoRI-Reverse	5'-ATAGAATTCTCATTCGTAGCTTCCCAC-3'
ImuB-Start-1-forward	5'-ATAAAACATATGATGCTCGGTGAACAGGACAGG CTT-3'
ImuB-Stop-1-Reverse	5'-ATAGAATTCTCAGCCCCATAGCGGCAACTGCAAGCCT-3'
ImuB-Stop-2-Reverse	5'-ATAGAATTCTCACGATGCGGACACCGTCTCCACCGCCT-3'
ImuB-Stop-3-Reverse	5'-ATAGAATTCTCATTCCGGTTCGCCGCGGGCGAACCG-3'
ImuB-Stop-4-Reverse	5'-ATAGAATTCTCATCGGCTCGGTGGCAAGTTGGCGGA-3'
ImuB-Stop-5-Reverse	5'- ATAGAATTCTCACGACAGAAACCGCGCATCGCCT-3'
ImuA'-NdeI-Forward	5'-ATAAAACATATGACTGCGGCCTTCGCCT-3'
ImuA'-EcoRI-Reverse	5'-ATAGAATTCTCACCGTCCACGCCCGTT-3'
ImuA'-Start1-Forward	5'-ATAAAACATATGTCCGGGGGGCCCAGC-3'
ImuA'-Start 2-Forward	5'-ATAAAACATATGTTCCCCCCGGGGGACGGTG-3'
ImuA'-Stop 1-Reverse	5'-ATAGAATTCTCAGCCGTCGGTGACCAGCA-3'

[a]Restriction sites are underlined.

2.3.2 Construction of vectors expressing tagged proteins for pull-down assay

Incorporation of a Tandem Affinity Purification (TAP) tag into MsImuB

MsImuB was C-terminally tagged with a His-FLAG TAP tag by Dr. Digby Warner. Briefly, a His-FLAG tag was introduced in the 3' end of *imuB* followed by a stop codon which was carried in pAINTPCRcomp plasmid to enable cloning of the upstream region containing the SOS box and the gene encoding MsImuA' into the complementing vector, pMCpAINT.

Incorporation of 3×FLAG sequence onto the N-terminus of MsImuB

The gene encoding MsImuB protein was PCR amplified from plasmid pAMSIMUAB using the primers listed in Table 2.2. The FLAG sequence was incorporated at the 5'-end of the

gene by nested PCR. Briefly, the first round of PCR was performed using primers 3xflag1622F1 and 1622R to incorporate half of the FLAG sequence on the 5' end and an *Acc*65I restriction site at the 3' end of *MSMEG_1622*, resulting in a PCR product of 1605 bp. The PCR product from the first round was used as a template with primers 3xflag1622F2 and 1622R to incorporate the entire sequence of the FLAG tag, including the start codon, and resulting in an amplicon of 1650 bp. An upstream region of 770 bp which includes the SOS box was amplified from plasmid pAINTPCRcomp using the primers 1620F1*Acc*65I and 1620R, with the forward primer incorporating an *Acc*65I restriction site. The PCR products were digested with *Acc*65I and ligated into pAINT on the same restriction site. Correct clones were screened by restriction digestion with *Acc*65I. Only the clones that gave an insert of 2375 bp were chosen for further restriction analysis and sequencing to confirm that the Methionine-3× FLAG sequence was correctly fused on the 5' end of *imuB* gene.

Table 2.4: Incorporation of a 3xFLAG sequence on the N-terminus of MsImuA' MsImuB[a]

Primer Name	Primer sequence
1620F1Acc65I	5'- ATAGGTACCTCGAGCAATCGAACATACAT-3'
1620R	5'-GTCTAACCGCCCACCGCACGT-3'
1622R	5'-ATAAAAGGTACCTCACTCATAGGCACC-3'
3x flag1622F1	5'-GATAAAGATTATAAGGATGACGACGATAAACCCGAGGGTTCCCGGGTGCT-3'
3x flag1622F2	5'- ATGGATTATAAGGATGACGACGATAAAGATTATAAGGATGACGACGATAAAGATTATAAGGATGA-3'
3x flag 1620F1	5'-ACGACGATAAAGATTATAAGGATGACGACGATAAACCCACAGCTGCGGCTGACCTTGATG-3'
3x flag 1620F2	5'-ATGGATTATAAGGATGACGACGATAAAGATTATAAGGATGACGACGATAAAGATTATAAG-3'
1620R	5'-ATAGGTACCGTCTAACCGCCCACCGCACGT-3'
1620FAcc65I	5'-ATAGGTACCGCTTCAGGTCAGAGGTTTCG-3'

[a]Restriction site underlined

Incorporation of 3×FLAG sequence onto the N-terminus of MsImuA’

The full-length *imuA’* gene together with the upstream region that incorporates the SOS box was PCR amplified from plasmid pAINTPCRcomp using the oligonucleotides listed in Table 2.4. The FLAG sequence was incorporated at the 5’-end of the genes by nested PCR. Briefly, the first round of PCR was performed using primers 3xflag1620F1 and 1620R to incorporate half of the FLAG sequence on the 5’ end and an *Acc*65I restriction site at the 3’ end of *imuA’* gene resulting in PCR product of 744 bp. The PCR product from the first round was used as a template with primers 3xflag1620F2 and 1620R to incorporate the entire sequence of the FLAG-tag including the start codon resulting in an amplicon of 784 bp. An upstream region of 770 bp which includes the SOS box was amplified from pAINTPCRcomp using the primers 1620F1Acc65I and 1620R, with the forward primer incorporating an *Acc*65I restriction site. The PCR products were digested with *Acc*65I and inserted into the *Acc*65I site in pAINT. Correct clones were screened by restriction digestion with *Acc*65I. Only the clones that gave an insert of 1554 bp were chosen for further restriction analysis and sequencing to confirm that the 3×FLAG sequence was fused with the gene.

2.3.3 Construction of complementing vectors expressing truncated MsImuA’

An artificial stop codon was introduced intoMsImuA’ at amino acid position 480 to truncate the C-terminal unstructured region. PCR was employed using primers 1620FAcc65I (5’-ata**ggtacc**gcttcaggtcagaggtttcg-3’) and 1620Stop1Aps718IR (5’-ata**ggtacc**ctacccgtcggtcaccag-3’) with restriction sites in bold and pAINTPCRcomp plasmid (Table 2.2) as a template. The PCR product was then digested with *Acc*65I (or the isoschizomer, *Asp*718I) and cloned in the *Acc*65I site of the pAINT vector. Clones were confirmed by restriction analysis and sequencing.

2.3.4 Construction of complementing vectors expressing truncated MsImuB

A C-terminally truncated version of ImuB was generated by Dr. D. Warner in a PCR reaction that effectively removed the 168 amino acid C-terminal region immediately after the predicted β-clamp binding motif. The truncated allele was cloned into the complementing vector, pMCpAINT (Warner *et al.*, 2010).

2.4 Construction of vectors for protein expression

Escherichia coli expression vectors

The *imuB* and *imuA'* open reading frames were PCR amplified from plasmid DNA (pGADImuB and pGADImuA') using the oligonucleotides listed in Table 2.5 The ORFs were cloned into the *Eco*RI site of the expression vector pMalc2. The C-terminal region of ImuB immediately after the predicted β-clamp binding motif ^{354}QLPLWG^{359} was removed by introduction of a stop codon. *ImuB* with the C-terminus was PCR amplified using oligonucleotides in Table 2.5 which flanked the PCR product with *Eco*RI restriction site and cloned into the *Eco*RI site of the expression vector pMalc2.

Table 2.5: List of primers for expression of ImuB, ImuA' and C-terminally truncated ImuB in *E. coli*[a]

Primer Name	Primer sequence
ImuB*Eco*RI-F	5'-ATAAAAGAATTCATGATGGCCTCCGCTCGCGTG-3'
ImuA'*Eco*RI-F	5'-ATAAAAGAATTCATGACTGCGGCCTTCGCCTCC GAC-3'
ImuB-Stop-1-Reverse	5'- ATAGAATTCTCAGCCCCATAGCGGCAACTG CAAGCCT-3'
ImuA'-*Eco*RI-Reverse	5'-ATAGAATTCTCACCGTCCACGCCCGTT-3'
ImuB-*Eco*RI-Reverse	5'-ATA GAA TTC TCA TTC GTA GCT TCC CAC-3'

[a]Restriction sites underlined

2.5 Yeast two hybrid (Y2H) analysis

Protein-protein interactions were assessed using the Clontech Matchmaker® Y2H system as per the manufacturer's instructions. Briefly, strains carrying plasmid to be assessed were grown in SD-LT broth to an OD_{600} of 0.2 then spotted on different SD media with increasing stringency to identify interacting partners. ImuB fused to GAL4 binding domain (BD) resulted in auto activation on low-stringency growth medium; therefore, all interactions involving ImuB were analyzed from experiments utilizing an ImuB GAL4 activation domain (AD) fusion construct only. All other proteins were cloned as both BD and AD GAL4 fusions.

2.6 DNA damage-induced mutagenesis assay

UV-induced mutation frequencies were determined as described previously (Boshoff *et al.*, 2003; Warner *et al.*, 2010). Briefly, *M. smegmatis* cultures were grown to mid-log phase (OD_{600} 0.4 -0.6) in 40 ml of 7H9 broth. An aliquot of 1 ml was sampled which served as the untreated control, and was serially diluted for plating. The remaining 39-ml culture was centrifuged at 4500 rpm (3900 × *g*) at room temperature for 10 min and re-suspended in 5 ml of 7H9 broth. The suspension was transferred into sterile empty Petri dishes and treated with UV at 250 mJ/cm^2 using a UV Stratalinker 1800 (Stratagene). The treated cells were transferred into a flask and fresh 7H9 broth added to 40 ml to recover at 37ºC with 1 ml of culture was sampled at 4.5 h and 24 h post-UV treatment. The sampled culture was serially diluted, and 10^{-5} to 10^{-7} dilutions were plated on 7H10 agar plates and 400 µl of the undiluted culture plated on 7H10 OADC medium supplemented with 200 µg/ml Rif. Colony forming units (CFUs) were scored after incubation at 37ºC for 3-5 days. The frequency of Rif resistance (Rif^R) was calculated as described (Warner *et al.*, 2010).

2.7 DNA damage sensitivity assays

This assay was performed in parallel with the DNA damage-induced mutagenesis assay (Section 2.6. The culture was sampled before UV treatment and ten-fold serial dilutions (up to 10^{-7}) were spotted on 7H10 medium supplemented with different concentrations of mitomycin C (MMC). The plates were incubated at 37ºC for 3-5 days and images of the plates taken using GelDoc under white light (G-Box).

2.8 Protein extraction, quantification and detection

2.8.1 Protein extraction from *M. smegmatis*

M. smegmatis strains expressing epitope-tagged proteins were grown in 100-ml cultures to mid log-phase (OD_{600} 0.4-0.6) and split into two equal 50-ml culture volumes. One aliquot was exposed to UV irradiation as described in section 2.7, and after irradiation, was allowed to recover for 1 h, whereas the other aliquot was used as an untreated control. Cells were harvested by centrifugation at 4500 rpm (3900 × *g*) for 15 min and re-suspended in 5 ml of 7H9 broth. The suspension was transferred into a sterile Petri dish and treated with UV at 250 mJ/cm^2 using UV Stratalinker 1800 (Stratagene). Both the treated cells and untreated control were incubated at 37ºC with shaking for 1 h. The method of Papavinasasundaram and

colleagues(Papavinasasundaram *et al.*, 2001) was adopted for protein extraction with minor modifications (Bhowmik *et al.*, 1987), as follows.

Cells were collected by centrifugation and washed three times in Z* buffer without β-mercaptoethanol (60 mM disodium hydrogen phosphate, 40 mM sodium phosphate, 10 mM potassium chloride, 1 mM magnesium sulfate) then re-suspended in 350 µl of Z* buffer supplemented with complete mini protease inhibitor cocktail (Roche). The cells were lysed three times for 45 s at speed 6 using the Savant Fastprep FP120 ribolyser, with 5 min intervals between pulses while cooling cells on ice. For whole-cell fractions, SDS-PAGE loading dye was added directly to the lysed cells and boiled at 95ºC for 20 min. For soluble fractions, the cells were centrifuged for 20 min at room temperature (13200 rpm; 16 100 × *g*) and supernatant (soluble fraction) collected in a fresh tube and to the pellet SDS-PAGE loading dye added for insoluble fraction. The protein concentration of each sample was quantified using the Bradford Protein Assay as per manufacture's instruction (Bio-Rad).

2.8.2 Protein analysis by SDS-PAGE and Western blotting

Equal amounts of protein from the soluble insoluble fractions (described in Section 2.8.1) were resolved on two 10% SDS-PAGE gels. One gel was stained with Coomassie brilliant blue stain (R250) to visualize protein, and the other, proteins transferred to a PVDF (Amersham) or nitrocellulose membrane (Thermo Scientific). The membrane was incubated with the anti-FLAG M2 antibody (Sigma) and FLAG-tagged proteins were detected using the ProteoQwest™ chemiluminescent Western blotting kit (Sigma).

2.8.3 Expression and purification of recombinant ImuB and ImuA' in *Escherichia coli*

Small-scale analysis of protein expression

E. coli BL21 (DE3) strains carrying vectors for expression of recombinant ImuB, ImuA' and ImuBstop1 were inoculated in 10 ml of LB broth supplemented with appropriate antibiotics for plasmid maintenance. Cells were allowed to grow to an OD_{600} of 0.5-0.7 at 37ºC with vigorous shaking (Labcon Shaking Incubator). IPTG was added to a final concentration of 0.5 mM, and the cultures incubated at 30ºC in a New Brunswick Scientific Innova 400 shaker for a further 3-5 hours. To obtain a whole-cell fraction, an aliquot of 100 µl was transferred into a tube and centrifuged to collect the cells. The cells were re-suspended in SDS-PAGE loading dye and stored at -20ºC until needed. To obtain a soluble fraction, cells were harvested from 2 ml of culture by centrifugation (2 min at 13200 rpm; 16 100 × *g*). The cells

were re-suspended in 500 µl of column buffer (20 mM Tris-hydrochloride, pH 7.5, 200 mM sodium chloride, 1 mM EDTA) and lysed by sonication (Soniprep 150 High Intensity Ultrosonic Processor, MSE, UK) with 15 s pulses separated by 15 s intervals while cooling on ice. Sonication was continued until the cells had completely lysed. The lysate was then centrifuged for 20 minutes at 4ºC to collect the cell debris. The supernatant, containing the soluble protein fraction, was transferred into a fresh tube, and the pellet was retained for analysis of the insoluble protein fraction. The concentration of the protein in the soluble fraction was determined by Bradford assay.

Equal amounts of protein from the soluble and insoluble fractions of the samples were resolved on 10% SDS-PAGE gels and proteins analysed as in Section 2.8.2 except that the fusion proteins were probed with anti-MBP primary antibody. Fusion protein was detected using the SuperSignal West Pico substrate (Thermo Scientific).

Large-scale protein expression and purification

E. coli BL21 (DE3) strains carrying vectors for expression of recombinant ImuB and ImuA' were inoculated into 10 ml of LB broth supplemented with appropriate antibiotics for plasmid maintenance. The strains were allowed to grow overnight at 37ºC with shaking on a Labcon Shaking incubator. Five ml of the overnight culture was used to inoculate 1 L of LB broth and incubated to grow to an OD_{600} of 0.5-0.7 at 37ºC with shaking. Protein expression was induced by addition of IPTG to a final concentration of 0.5 mM and incubated at 30ºC for 3-5 hours. Cells were collected by centrifugation at 4500 rpm for 30 min in a Beckman Coulter Allegra X-22 (4250 rotor) centrifuge and re-suspended in 1 ml of column buffer supplemented with complete mini protease inhibitor cocktail (Roche). Cell lysis was achieved by sonication, as described above. The lysate was clarified by centrifugation for 20 min at 4ºC and the supernatant was collected and stored at 4ºC for not more than 1 day prior to purification.

The lysate was loaded onto a 5 ml packed amylose (New England Biolabs) column pre-equilibrated with column buffer at 4ºC. Flow-through was collected and unbound protein washed with 5× column volumes of column buffer. Protein was eluted in 500 µl fractions with elution buffer (20 mM Tris-hydrochloride, pH 7.5, 200 mM sodium chloride, 1 mM EDTA, and 10 mM maltose) and protein concentration determined by monitoring absorbance at 280 nm (using a NanoDrop) to plot an elution profile. Samples from all fractions were run on 10% SDS-PAGE gel to assess the purity of the protein. Peak fractions were pooled and

dialysed against low salt column buffer (20 mM Tris-hydrochloride, pH 7.5, 50 mM sodium chloride, 1 mM EDTA) using a dialysis tubing with 15 kDa molecular weight cut-off. The dialysed protein was concentrated on Millipore concentrators with a molecular weight cut-off of 50 kDa as per manufacturer's instruction.

A small-scale experiment was initially done to determine the sodium chloride concentration at which fusion protein can elute from a Q-Sepharose column. Briefly, 50 µl of Q-Sepharose resin in a round-bottomed Eppendorf tube was pre-equilibrated with the dialysis buffer (20 mM Tris-hydrochloride, pH 7.5, 50 mM sodium chloride, 1 mM EDTA ethylenediaminetetraacetic acid). Two µl of concentrated protein extract, prepared as described above, was added and allowed to bind on ice for 2 h. The bound resin was washed with the same buffer five times, followed by centrifugation at 100 rpm for 1 min. The supernatant was collected and the bound protein eluted by sequential addition of column buffer containing sodium chloride at a concentration ranging from 200 mM to 3 M. From this analysis, a Q-Sepharose column of 2 ml packed resin pre-equilibrated with dialysis buffer was loaded with concentrated protein at 4 ˚C. Flow-through was collected and washed with 5× column volumes. Bound protein was eluted with a linear salt gradient (0.5M - 1M NaCl) and fractions of 500 µl collected. All fractions were analysed by monitoring absorbance at 280nm and by gel electrophoresis on 10% SDS-PAGE gels.

2.9 Proteomic analysis of *M. smegmatis* strains by mass spectrometry

This work was done at Harvard University, School of Public Health. Preparation of peptides were done as follows: *M.smegmatis* strains were inoculated into 5 ml 7H9 broth and allowed to grow at 37°C with shaking until they reached stationary phase (OD_{600} of 1.8-2.0). The overnight culture was used to inoculate 100 ml of 7H9 broth and incubated until they reached mid-log phase (OD_{600} 0.4-0.6). The cells were divided into two equal 50-ml aliquots where one was treated with UV as in Section 2.6 and allowed to recover for 1 hour. Both treated and untreated cells were washed three times in NSalt minimal media (100 mM Bis-Tris-hydrochloride, 5 mM potassium chloride, 7.5 mM ammonium sulphate, 0.5 mM potassium hydrogen sulphate, 1 mM potassium dihydrogen orthophosphate, 10 mM magnesium chloride, 38 mM glycerol, pH 7.0) without Tween 80 (Garces *et al.*, 2010). The pellet of washed cells was inoculated in the NSalt minimal medium and incubated overnight at 37°C. The cells were collected by centrifugation (13200 rpm; 16 100 × *g* for 10 min) and re-suspended in 350 µl of protein extraction buffer (50 mM Tris-hydrochloride, pH 7.5, 5 mM

EDTA) supplemented with complete mini protease inhibitor cocktail (Roche). The suspension was transferred into lysing matrix B tubes and lysed on a Savant Fastprep FP120 ribolyser for 3× (45 s at 6.5 m/s) with 5 min cooling on ice between pulses. After lysis, 2× tricine loading buffer (Invitrogen) was added to each sample and mixed and boiled at 95˚C for 30 min.

For mass spectrometric analyses, samples were separated on a 10-20% Tricine gel (Invitrogen), 45 min at 100 V. Gels were destained overnight by the addition of 50% ethanol and 7% acetic acid, and then allowed to hydrate for 1 h in deionized water. Gels were stained with SimplyBlue Safe Stain (Invitrogen) for 45 min, imaged, and sliced horizontally into fragments of equal size based on the molecular weight markers.

In-gel reduction, alkylation and digestion was performed after destaining and rinsing the gel sections with two washes of 50% ethanol and 7% acetic acid, followed by two alternating washes with 50 mM ammonium bicarbonate and acetonitrile. After removal of the last acetonitrile wash, 100 µl of sequencing grade trypsin (Promega,Madison, WI) was added to each gel slice at a concentration of 6.6 µg/ml in 50 mM ammonium bicarbonate/10% acetonitrile solution. The gel slices were allowed to swell for 30 min on ice, after which the tubes were incubated at 37°C for 24 h. After incubation, 100 µl of 50 mM ammonium bicarbonate/10% acetonitrile solution was added and incubation continued for another 24 h. Peptides were extracted by briefly spinning the tubes and supernatant transferred into clean Eppendorf tube followed by storage at 4°C. To the remaining gel slices 100 µl of 50% acetonitrile/0.1% formic acid was added and peptides extracted as above. The extracts were pooled and frozen at -80˚C before lyophilized to dryness. The dried peptide extracts are stable at this point and can be stored until needed.

Processing of peptides was done by Alejandra Garces Micheal Chase (Fortune Lab) and the peptide samples ran at ThermoScientific Centre with the assistance of Dr. Davis Sarracino. Dried peptide extracts were dissolved in 40 µl of 5% acetonitrile, 0.1% formic acid and then loaded into a 96-well plate (AbGene) for mass spectrometry analysis on a Thermo Fisher Scientific Orbitrap XL, Thermo Fisher Scientific LTQ-FT or a Thermo Fisher Scientific LCQ Deca XP Plus, as per manufacturer’s instruction (Thermo Fisher Scientific, Waltham, MA).For each run, 10 µl of each reconstituted sample was injected with a FamosAutosampler (Dionex, Sunnyvale, CA) and the separation was performed on a 75 µM x 20 cm column packed with C18 Magic media (Michrom Biosciences, Auburn, CA) running at 250 nl/min provided from a Surveyor MS pump (Thermo Fisher Scientific) with a flow splitter with a

gradient of 5-60 % water 0.1 % formic acid, acetonitrile 0.1 % formic acid over the course of 120 min. Between each set of samples, standards from a mixture of 5 angiotensin peptides (Michrom Biosciences) were run for 2.5 h to ascertain column performance and observe any potential carryover that might have occurred. The Orbitrap XL was run in a top eight configuration with one MS 60K resolution full scan and eight MS/MS scans. The LTQ-FT was run in a top nine configuration with one MS 200 K resolution full scan and nine MS/MS scans and the LCQ Deca XP Plus was run in a top five configuration with one MS full scan and five MS/MS scans. Dynamic exclusion was set to 1 with a limit of 180 seconds with early expiration set to 2 full scans.

Peptide identifications were done by Micheal Chase (Fortune Lab) using the database search algorithm, SEQUEST (Thermo Scientific, San Jose, CA). Spectra were searched against a composite database contained the predicted open reading frames annotated in the genome of *M. smegmatis* mc^2155. A reverse database strategy was employed to estimate false discovery rate (FDR) (Elias & Gygi, 2007). Peptides were filtered at a 1% FDR and clustered into proteins using an Occam's approach, a strategy which produced protein identifications with probabilities of 0.98 or greater in parallel analyses using ProteinProphet (Nesvizhskii *et al.*, 2003). Spectral counts were pooled across gel slices and levels of protein expression between strains compared using an extended G-test (Zhang *et al.*, 2006).

Chapter 3

3. Results

3.1 Bioinformatic analysis of DnaE2 accessory factors, ImuB and ImuA'

At the inception of this study, the working model (Figure 3.1) was that ImuB serves as a molecular adaptor, allowing DnaE2 and ImuA' to access the replication fork via its interaction with the β-clamp.

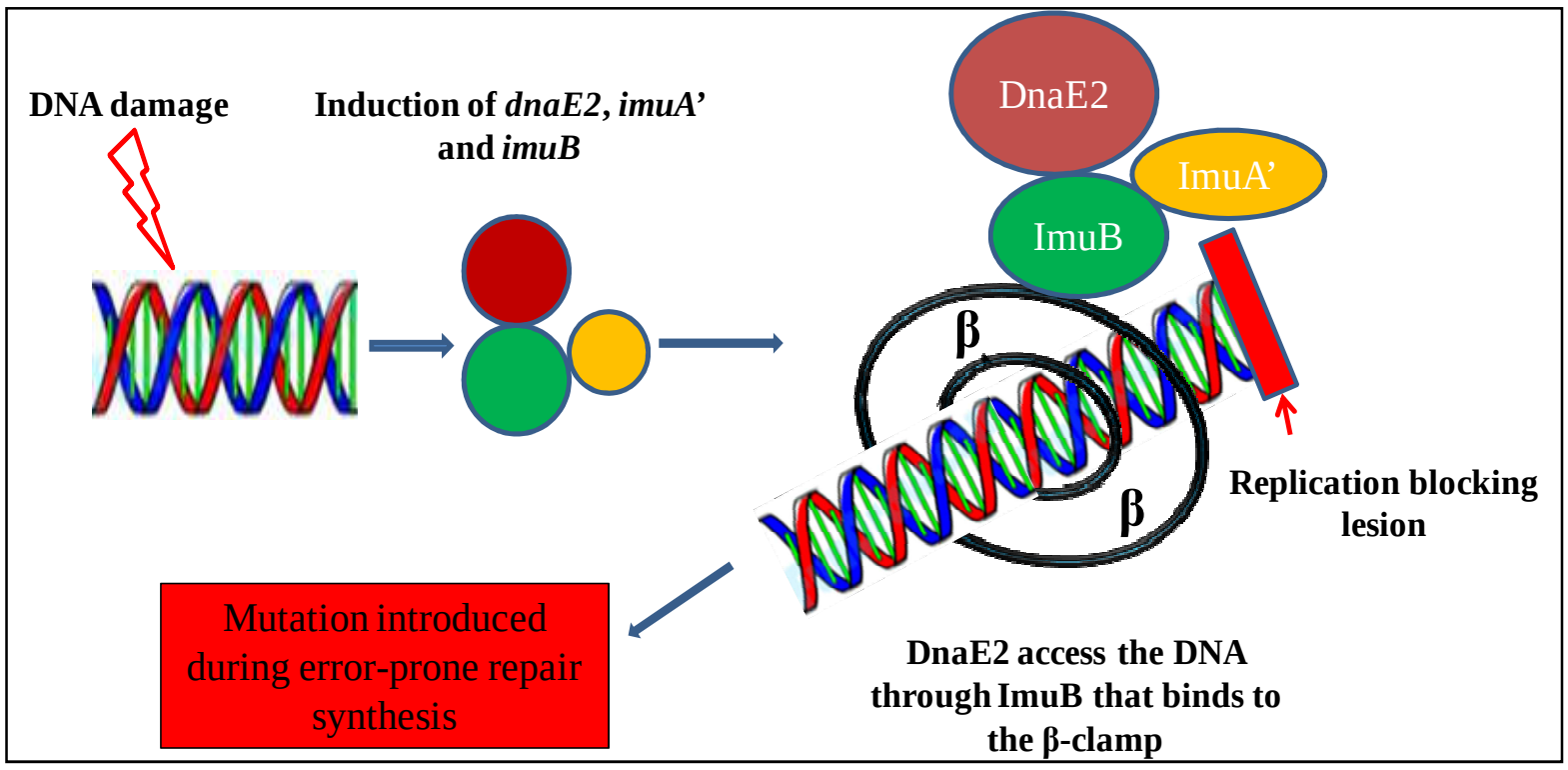


Figure 3.1: DnaE2-dependent mutagenesis in *M. tuberculosis*. According to this model, DNA damage causes normal replication to stall. In turn, this leads to the up-regulated expression of *dnaE2*, *imuB* and *imuA'*, which are recruited to the site of the lesion. DnaE2 and ImuA' access the DNA template by interacting with ImuB which binds the β-clamp, and the complex catalyzes translesion synthesis across the lesion.

As noted in the introduction, the mycobacterial C family polymerase, DnaE2, was found to be a DNA damage-inducible DNA polymerase and implicated in virulence and the emergence of antibiotic-resistant MTB mutants *in vivo* (Boshoff *et al.*, 2003). Failure to identify a β-clamp binding motif in DnaE2 suggests that it requires another protein that is able to interact with the β-clamp to enable access to the replication fork. During the course of this study, DnaE2 was demonstrated by Dr. Warner to function with two hypothetical proteins, ImuB and ImuA' (Warner *et al.*, 2010); since they are hypothetical proteins, their sequences were analysed. As shown in Figure 3.2, ImuB and its *M. smegmatis* homologue (MSMEG_1622) share significant sequence similarity with *E. coli* DinB, a Y-family polymerase. However, unlike DinB, the mycobacterial ImuB homologues are characterized

by a C-terminus that extends beyond the 3' terminal of β-clamp binding motif in DinB; moreover, ImuB lacks the catalytic residues required for polymerase function (Warner *et al.*, 2010).



Little Finger

```
Conservation:                                    6          5  5                                  5
P._putida_KT2440__PP3118_      248  ESHQALLFPLRRMLNDLAAFLAGRD--CGVQRFCLHLEHAEGPDTLLKVGLLAAERDAAMLFELARGRLE  315
C._crescentus_CB15__CC321      200  SAPEDLARVAADALALICARLEAEG--RGAKRFEVVFHRLDGRAFPVRVGLARIGRDAQRLARLVKPKLD  267
S._avermitilis__SAV6556_       193  VDPSRHRRALLSVTEELGTRLRTLE--QVCRTLTLTVRYADRSSTTRSRTLAEPTAHSAALTATAYRMYE  260
M._tuberculosis_H37Rv__Rv      254  DRVDAAAFAGRSLAAELHRALMAAG--VGCTRLAIHAVTANGEERSRVWRCAEPLTEDATADRVRWQLDG  321
M._smegmatis_mc2155__MSME      254  ERVDAAAFAGRSLAVTLHRSLEAAG--VGCTRLAIHAVTANGEQLERVWRCAEPLTEDATADRVRWQLDG  321
M._smegmatis_DinB2             258  TERREMDSAVRDLALQTLAEIVEQG--RIVTRVAVTVRTSTFYTRTKIRKLPAPSTDAGQIVDTALAVLD  325
M._tuberculosis_DinP__Din      261  TCRSEMESAVTELAQRTLNEVVASS--RTVTRVAVTVRTATFYTRTKIRKLQAPSTDPDVITAAARHVLD  328
M._smegmatis_DinB1             266  TTLAQLQDAIVPIGEHAHRRLEKDG--RGARTVTVKLKKSDMSTLTRSATLAYATTEASTLIGTARRLLL  333
M._tuberculosis_DinX__Din      253  TTMEQLHEAIDSIAEHAHQRLLRDG--RGARTITVKLKKSDMSTLTRSATMPYPTTDAGALFTVARRLLP  320
S._solfataricus_Dpo4_          256  RNLEEIKPYLFRAIEESYYKLDKRI----PKAIHVVAVTEDLDIVSRGRTFPHGISKETAYSESVKLLQK  321
E._coli__dinB_                 254  HHWSECEAIIERLYPELERRLAKVKPDLLIARQGVKLKFDDFQQTTQEHVWP--RLNKADLIATARKTWD  321
Consensus aa:                       pp.p.h..hh.phh.ph..pL........hpph.l.h...s....s....hs.s..c.s.h..hh...b.
Consensus ss:                         hhhhhhhhhhhhhhhhhhhh        eeeeeeee     eeeeeee        hhhhhhhhhhhhh
```

Little Finger

```
Conservation:                                  66 5 6
P._putida_KT2440__PP3118_      316  PLRI-------PAPVRNLRLVAEDLPPFVPQ-HQALFD-PRAQQAQPWAQLRERLRARLGDEAVKGLGAA  376
C._crescentus_CB15__CC321      268  MVDP-------GFGIEVVTVHAFAVEPMAA-AQARLDADAAASADETLAPLIDRLVNRLGENRVWRADPF  329
S._avermitilis__SAV6556_       261  ALGLQ------RARVRAIALRAEGLSPAED-AAHQLTFDPVDEKVRRIEEVADRARAKFGPRAIVPGALA  323
M._tuberculosis_H37Rv__Rv      322  WLNNRNARDRPTAAVTLLRLQAVETVSASEGLQLPLWGGLGEQDRLRARRALVRVQGLLGPEAVRVPVLS  391
M._smegmatis_mc2155__MSME      322  WLNRRTD-DRPSAPITVLRLRPVEVVS-SAALQLPLWGGIGEEDRLRARRALVRVQGLLGPEAVQVPVLS  389
M._smegmatis_DinB2             326  QFEL-------DRPVRLLGVRLELAMDDVPR-PAVTAGT----------------------------------  356
M._tuberculosis_DinP__Din      329  LFEL-------DRPVRLLGVRLELA------------------------------------------------  346
M._smegmatis_DinB1             334  DPVE-------IGPIRLVGVGFSGLSDIRQE-SLFPDL-EQPEEFPDAAPQVES--VQTGPSSPTTQWRV  392
M._tuberculosis_DinX__Din      321  DPLQ-------IGPIRLLGVGFSGLSDIRQE-SLFADS-DLTQETA-AAHYVETPGAVVPAAHDATMWRV  380
S._solfataricus_Dpo4_          322  ILEED------ERKIRRIGVRFSKFIEAIG-LDKFFDT----------------------------------  352
E._coli__dinB_                 322  ERRG-------GRGVRLVGLHVTLLDPQME-RQLVLGL----------------------------------  351
Consensus aa:                       .............lc.l.l.h..h.s........h.................................
Consensus ss:                       hh            eeeeeeee                    hhhhhhhhhhhhhh
```

```
Conservation:
P._putida_KT2440__PP3118_      377  ADHRPECAWHLAEQG---AQGNMPVA------------PGSRPGWLLSAPIALDDAA-------------  418
C._crescentus_CB15__CC321      330  ESHVPERSVVRVGPL---DPPPAARWD----------PDRPRPVRLFKRPEAIVAIAA--------ELPD  378
S._avermitilis__SAV6556_       324  A-------------------------------------------------------------------------  324
M._tuberculosis_H37Rv__Rv      392  GGHGPAERITLTVLG---LVAPEPVPQADPGQPWPGRLPDPSPAVLFDDPVDLLDAQGNPIRVTSRGMFS  458
M._smegmatis_mc2155__MSME      390  GGRGPAERITFTALG------DEPVPQADPRQPWPGQLPEPSPTVLLDDPVELLDAQGNPVRVNSRGLFS  453
M._smegmatis_DinB2                  -------------------------------------------------------------------------
M._tuberculosis_DinP__Din           -------------------------------------------------------------------------
M._smegmatis_DinB1             393  GDDVAHTELGHGWVQGAGHGVMTVRFETRASGPGPARTF--------PEDSPEISR--------------  440
M._tuberculosis_DinX__Din      381  GDDVAHPELGHGWVQGAGHGVVTVRFETRGSGPGSARTF--------PVDTGDISN--------------  428
S._solfataricus_Dpo4_               -------------------------------------------------------------------------
E._coli__dinB_                      -------------------------------------------------------------------------
Consensus aa:                       .........................................................................
Consensus ss:                            hh
```

```
Conservation:
P._putida_KT2440__PP3118_      419  ---------------YQVQGYAERIESGWWDG--------GDVRRDYYRI-ETRDGL-RGWAYRDLSQ--  461
C._crescentus_CB15__CC321      379  YPPRLFTWRG-RSHRVRRAEGPERIGQEWWRVGVEKGQTGPGKIRDYYRV-EDDTGG-RFWIFRQGLYGG  445
S._avermitilis__SAV6556_            -------------------------------------------------------------------------
M._tuberculosis_H37Rv__Rv      459  ADPARLRVRG-RDDRLRWWAGPWPDDERWWDPD----RASGRTARAQVLL-DGDPGTALLLCYRQR----  518
M._smegmatis_mc2155__MSME      454  SDPSRLSGAGKRDGVLRWWAGPWPVDERWWDPQ---NRGTGRTARAQVLLGDGEDALALLLCYRQR----  516
M._smegmatis_DinB2                  -------------------------------------------------------------------------
M._tuberculosis_DinP__Din           -------------------------------------------------------------------------
M._smegmatis_DinB1             441  ---------------------ADPVDSLDWAEY----LSSLA-------------DYQSTP---------  463
M._tuberculosis_DinX__Din      429  ---------------------ASPLDSLDWPDY----IGQLSVEGSAGAS-APTVDDVGDR---------  463
S._solfataricus_Dpo4_               -------------------------------------------------------------------------
E._coli__dinB_                      -------------------------------------------------------------------------
Consensus aa:                       .........................................................................
Consensus ss:
```

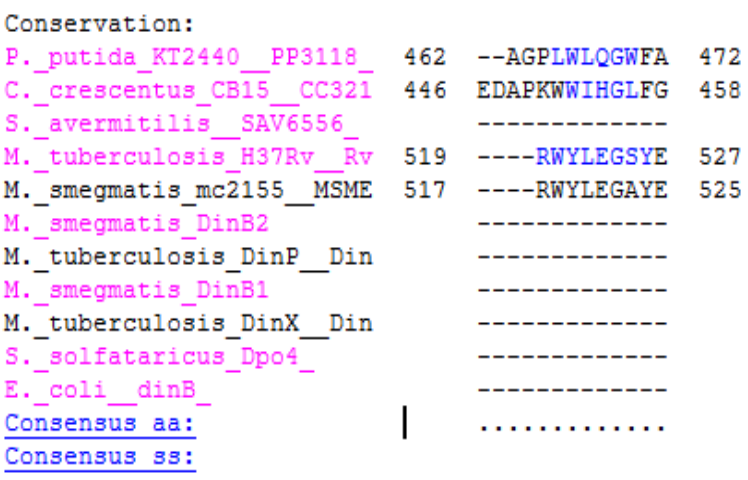


Figure 3.2: Multiple sequence alignment of ImuB homologs compared to Y family polymerase from bacteria. The proteins were aligned using the PROMALS3D multiple sequence alignment and structure server (http://prodata.swmed.edu/promals3d/promals3d.php). Sequences in bold green denote β-clamp binding motifs. Representative sequences have magenta names and are coloured according to predicted secondary structures (red: α-helix; blue: β-strand). The first line in each block indicates conservation indices (8) for those positions with a conservation index above 4. Consensus secondary structure is indicated by an ″ h″ for α-helix and an ″e″ for β-strand. Consensus amino acids are indicated by the following symbols: conserved amino acid residues, bold uppercase letter; aliphatic residues, I; aromatic residues, @; hydrophobic residues, h; alcohol residues, o; polar residues, p; tiny residues, t; small residues, s; bulky residues, b; positively charged residues, +; negatively charged residues, -; charged residues, c.

All alignments were generated with PROMALS3D bioinformatic tool (Pei *et al.*, 2008). Residues shaded grey represents active-site carboxylates highly conserved among Y family polymerase (Ling *et al.*, 2001). The β-clamp binding motifs were predicted with references to identified consensus motifs (Wijffels *et al.*, 2005): QLSLF-for DinB proteins; -QLGL- or –LF-for ImuB proteins. The underlined sequences were identified previously (Dalrymple *et al.*, 2001). The annotation of major structural domains is derived from the crystal structure of the DinB homolog in *S. solfataricus*, Dpo4 (Ling *et al.*, 2001), and the little finger domain of *E. coli* DNA polymerase IV (Bunting *et al.*, 2003). Similarly, mycobacterial ImuA and its homologues from other bacteria were aligned in comparison to *E. coli* RecA.



Figure 3.3: Sequence alignment of ImuA' proteins compared to *E. coli* RecA protein. The proteins were aligned using the PROMALS3D multiple sequence alignment tool (Pei *et al.*, 2008). Structural domains are drawn in bars as described in the *E. coli* crystal structure (Story & Steitz, 1992; Story *et al.*, 1992). The shaded bars and the underlined sequence of *E. coli* represent the position of the flexible DNA binding loops, L1 and L2, in the RecA M domain.

ImuA’ from mycobacteria exhibit sequence similarity to the N-terminal and middle domains of *E. coli* RecA, with limited similarity in the C-terminal region (Figure 3.3). *E. coli* RecA contains phosphate-binding amino acids in the P-loop motif of the M domain but ImuA’ lacks such residues. *M. tuberculosis* ImuA’ has a shorter L1 loop and an L2 that is longer. A region spanning R^{133}-R^{148} in the predicted ImuA’ M domain contains a number of positively charged residues consistent with α-helix formation and DNA binding.

3.1.1 Prediction of ImuB and ImuA’ structures

A homology model generated by a collaborator, Prof. Česlovas Venclovas (Laboratory of Bioinformatics, Institute of Biotechnology, Luthania), predicted the structure of ImuB.

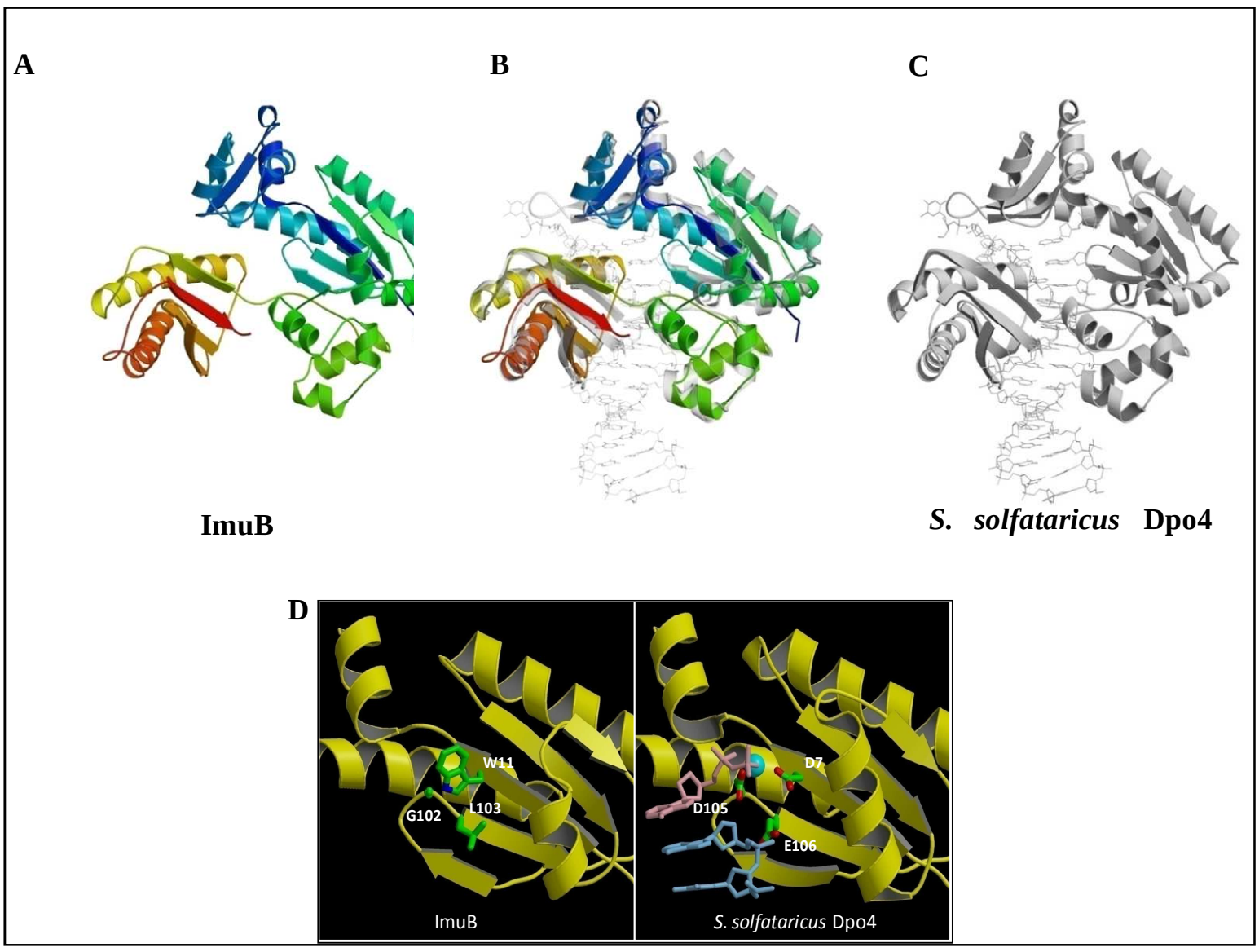


Figure 3.4: ImuB shares structural similarity with Y family polymerase from *S. Solfataricus* Dpo4. **A**. ImuB model is coloured according to protein chain progression from N (blue) to C-terminus (red), **B** ImuB model superimposed onto Dpo4 complexed with DNA, **C**. the x-ray structure of *S. Solfataricus*Dpo4 complexed with DNA and incoming nucleotide (PDB id: 1jx4), **D.** Close-up view of the active site in the palm domain.

As shown in Figure 3.4, ImuB is structurally similar to Y-family polymerase; however, it is important to note that the C-terminal region extending from the β-clamp binding motif is not depicted in the model. A close-up view of the active site confirms the absence of catalytic amino acids required for DNA polymerase activity. This suggests that ImuB is a pseudopolymerase. The absence of the catalytic residues supports the working model that ImuB may play a role in allowing DnaE2 and ImuA' to access the replication block in induced mutagenesis rather than in catalysis of translesion synthesis. The C-terminal extension of ImuB could not be modelled to any known structures suggesting that it might be a region with disorder. To test this interpretation, Prof. Česlovas Venclovas used the bioinformatic analysis tools such as, DisProt (Sickmeier *et al.*, 2007), IUPred (Dosztanyi *et al.*, 2005), DisEMBL (Linding *et al.*, 2003) and DISOPREP (Ward *et al.*, 2004) to predict regions of disorder in ImuB. These methods predicted extensive disorder in the C-terminus which includes the β-clamp binding motif and the C-terminal extension (Figure 3.5).

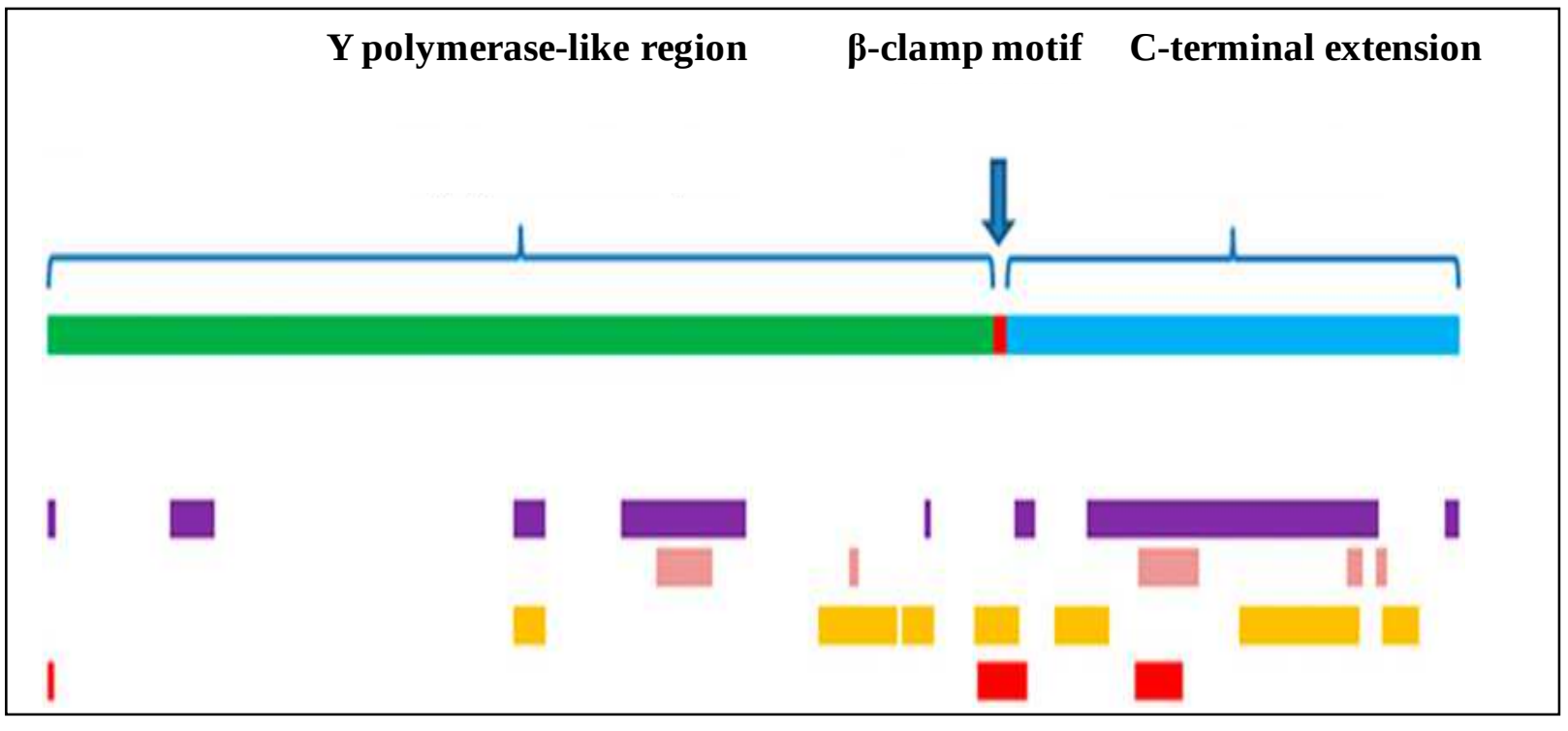


Figure 3.5: ImuB has extensive disorder in the C-terminal region. The coloured rectangles represent disordered regions in ImuB predicted by different methods: DisProt (purple), IUPred (pink), DisEMBL (yellow), and DISOPRED (red).

Again using homology modelling, the structure of ImuA' was predicted by Prof. Venclovas using *E. coli* RecA as template as these proteins share significant sequence similarity. Figure 3.6 shows that, as predicted from the sequence similarity, ImuA' is structurally similar to a RecA protein in the N-terminus and middle domains, but possesses an unrelated C-terminus.

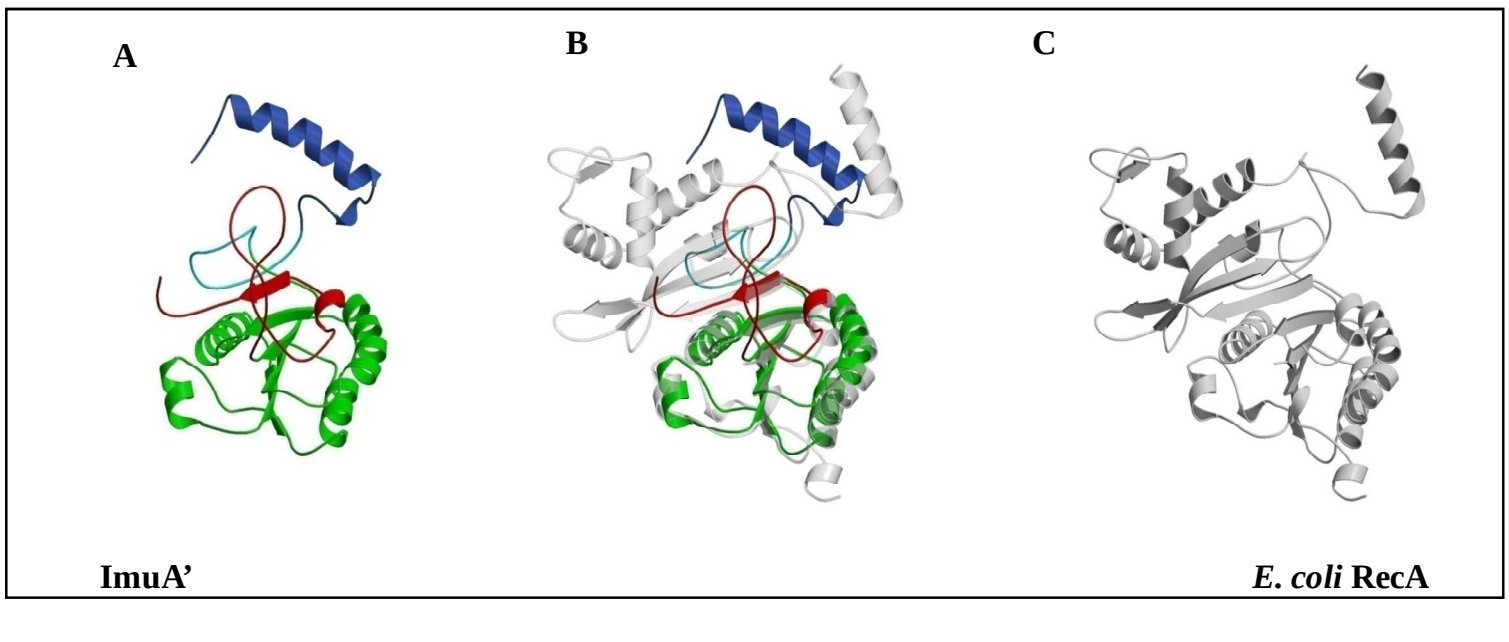


Figure 3.6: ImuA’ is structurally similar to *E. coli* RecA. **A**. ImuA’ model is coloured according to protein chain progression from N (blue) to C terminus (red), **B** ImuA’ model superimposed onto *E. coli* RecA **C**. the x-ray structure of *E. coli* RecA (PDB id: 1u94).

The data presented in this section indicate that ImuB is structurally similar to a Y family polymerase but has a C-terminal extension that is disordered. Moreover, the absence of active site acid residues characteristic of nucleotidyl transfer suggests that the protein is catalytically non-functional. In turn, this suggests that ImuB may function as a scaffold protein, consistent with the presence of a disordered C-terminal region which might allow interactions with other proteins. Although it has a distinct C-terminal domain, the fact that ImuA’ is structurally similar to RecA protein in the N-terminus and the middle domain suggests that ImuA’ may perform functions similar to that of RecA such as recruiting polymerase and/or other proteins to the site of damage (Patel *et al.*, 2010).

3.2. Interaction analysis of the proteins involved in the DnaE2-dependent induced mutagenesis pathway

Yeast two-hybrid (Y2H) analysis (Dolan *et al.*, 1989; Vidal & Legrain, 1999; Walhout & Vidal, 2001) was employed to explore the possibility that protein-protein interactions might determine the function of the different cassette components. In this system, genes encoding proteins of interest are cloned in-frame with the GAL4 activation (AD) and binding domain (BD) plasmids, which are then co-transformed into the yeast strain AH109. When the proteins interact, they bring the two domains (AD and BD) into close proximity, thus reconstituting a functional GAL4 transcriptional activator which drives expression of the downstream gene. Restoration of GAL4 function is observed by reversal of the auxotrophic phenotype of the host strain during growth on minimal media lacking a specific amino acid or

acids, as represented in Figure 3.7. For the experiments described below, two reporters used in the screen were *ADE2* and *HIS3*. *HIS3* and *ADE2* are endogenous yeast genes which are required for the synthesis of histidine (His) and adenine (Ade), respectively. *HIS3* expression can be scored in a growth-based assay by plating strains on media that lack histidine. Very small levels of expression are required for growth, and amounts of expression can be exquisitely modulated by exposing the yeast strain to increasing concentrations of 3-amino-triazole (3-AT), a competitive inhibitor of the product of the *HIS3* gene product, HIS3 protein (His3p) (Joung *et al.*, 2000; Kennedy, 2002). Yeast strains that lack *ADE*2 form red colonies after several days of growth on plates lacking adenine. Graded expression of *ADE2* changes the colony colour from red to pink to white, giving a colourimetic readout that is simple to monitor. The drawback of using the *ADE2* and *HIS3* reporters is that expression levels cannot be easily quantified. The *ADE2* reporter alone provides strong nutritional selection and the use of *HIS3* reporter reduces the incidence of false positives, thus allowing the control of the stringency of selection (James *et al.*, 1996).

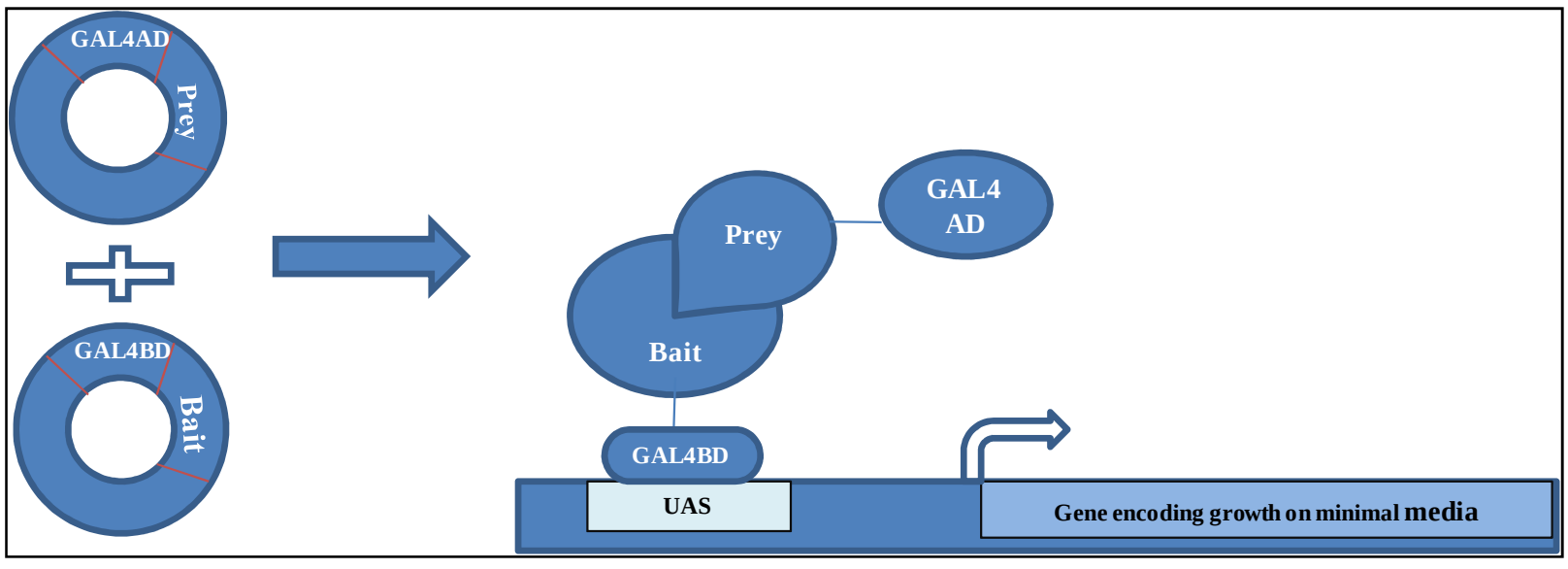


Figure 3.7: Diagrammatic representation of the principle of the Yeast-two hybrid analysis. Y2H is based on the functional reconstitution of an intact transcriptional factor that activates reporter expression. The bait protein is fused to the DNA binding domain (BD) of a GAL4 transcription factor and the prey protein is fused to the GAL4 activation domain (AD) of the transcription factor. When the bait and the prey protein physically interact when co-expressed in the host, a functional transcription factor is reconstituted by bringing the BD and AD domains in close proximity to activate expression of reporter gene by binding to the upstream activation sequence (UAS).

3.2.1 Confirmation of pair-wise interactions between DnaN, ImuB, ImuA', and DnaE2 by yeast two-hybrid analysis (Y2H)

In preliminary studies, the fusion of ImuB to the BD was found to result in auto-activation in the Y2H system, thus yielding a false positive signal in the plating assay. In contrast, no auto-activation was observed when ImuB was cloned as an AD fusion protein. Therefore, in

all of the interaction analyses described below, ImuB was used solely as an AD fusion. Since auto-activation was not observed with any of the other constructs, all other proteins were used in Y2H assays in both AD and BD configurations. In previous work in our laboratory, Dr. Garth Abrahams had determined that the optimal concentration of 3-AT required to restrain background growth on synthetic drop-out (SD) medium lacking His (Durfee *et al.*, 1993; Kirkman-Correia *et al.*, 1993) was 1.25 mM.

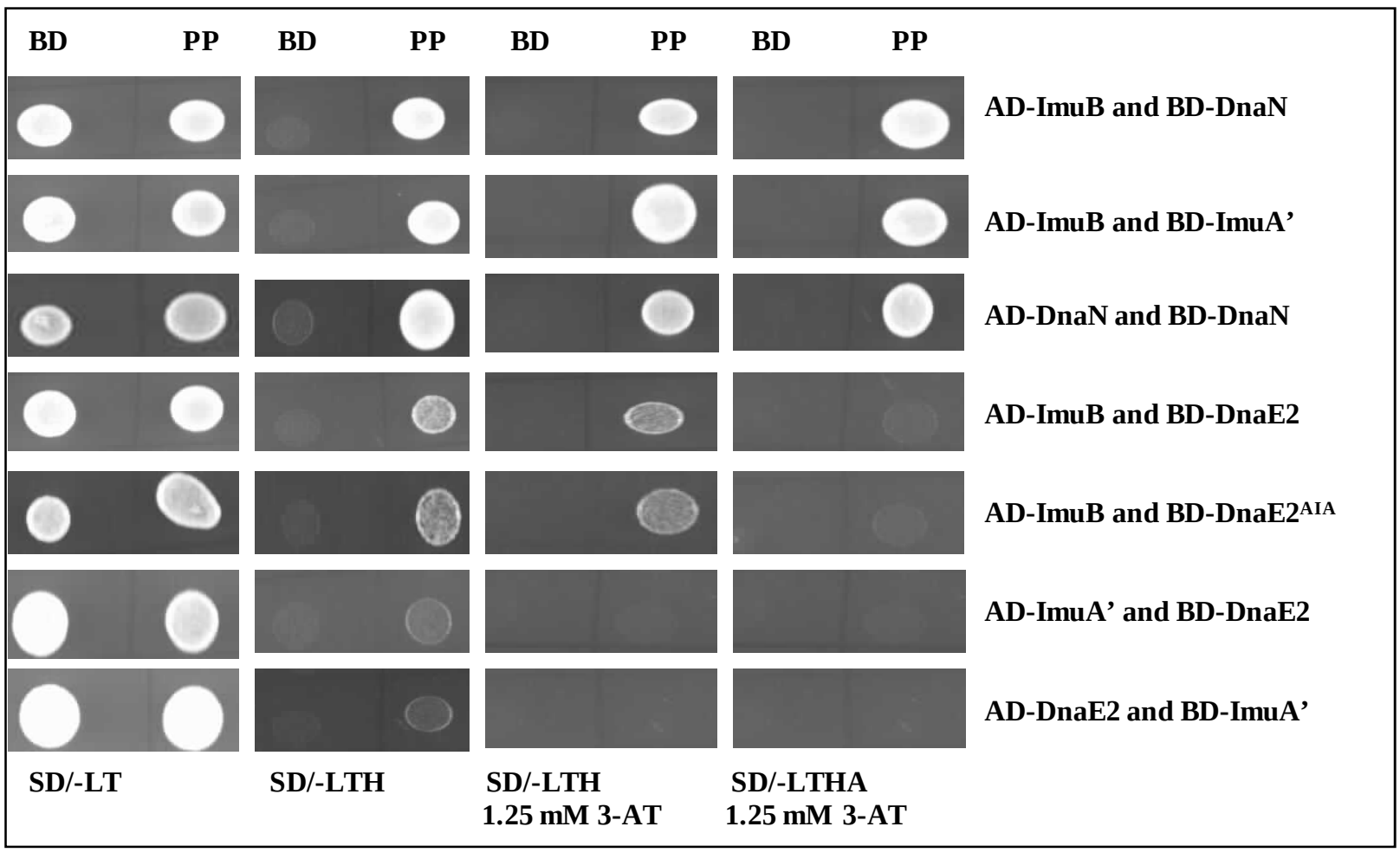


Figure 3.8: Confirmation of pair-wise protein-protein interactions. PP indicates the specific protein-protein interaction and BD indicates the negative control where empty BD plasmid was co-transformed with the AD-fusion plasmid. Proteins expressed in-frame with GAL4 are presented as AD and BD fusions, respectively. Protein interactions are observed on synthetic drop-out (SD) medium lacking leucine and tryptophan (SD/-LT); leucine, tryptophan and histidine (SD/-LTH); and leucine, tryptophan, histidine and adenine (SD-LTHA), increasing the stringency of selection. 3-AT was added to medium-stringency and high-stringency media in order to increase the confidence in the inferred interaction.

The data in Figure 3.8 indicated that ImuB interacts with DnaE2, ImuA' and DnaN. In addition, DnaN exhibits self-interaction, consistent with the function of the clamp protein as a homodimer (Neuwald, 2003; Stewart *et al.*, 2001; Stukenberg *et al.*, 1991). From these results, it appears that ImuA' does not interact with DnaE2, and neither ImuA' nor DnaE2 interact with DnaN. This result is consistent with the working model, since it suggests that the interaction of ImuB with ImuA' and DnaE2 might serve to bring these proteins to the replication fork through their interaction with the *dnaN*-encoded β-clamp. In parallel work,

the catalytic aspartic acid residues of DnaE2 were mutated to alanines (Dr. Warner) to confirm that DnaE2 provides polymerase function. The mutant lacking catalytic residues eliminated UV-induced mutagenesis and was also hypersensitive to the genotoxic agent, mitomycin C (MMC), strongly implicating a direct role of DnaE2 in TLS (Warner *et al.*, 2010). The loss of DnaE2 function as a result of mutating an acidic active site residue was a very important result since it provided the first direct evidence that the C-family polymerase was directly involved in TLS. This finding was also consistent with the fact that ImuB lacks critical active site residues required for catalysis (DNA replication) which would preclude it from catalyzing TLS. A Y2H construct was thus generated in which the catalytic residues of DnaE2 were mutated ($DnaE2^{AIA}$) order to investigate the possibility that this mutation affected the ability of DnaE2 to interact with ImuB. If this were the case, it would invalidate the inference that DnaE2 itself catalyzes TLS. The data in Figure3.8 show that mutating the catalytic residues in DnaE2 does not disrupt interaction with ImuB, further supporting the conclusion that the catalytic residues of DnaE2 are required for TLS function. Again, these results were consistent with the idea that ImuB provides the means for DnaE2 and ImuA' to access DNA template through its interaction with the β-clamp. However, the nature of these interactions remains unknown.

3.2.2 The β-clamp binding motif is required for interaction of ImuB with the β-clamp

ImuB has a putative β-clamp binding motif, ^{354}QLPLWG^{359}, which is located in the C-terminal region, upstream of the predicted disordered region. The presence of this motif identified ImuB as founder member of the DinB3 family of Y polymerases (Dalrymple *et al.*, 2003). To confirm the role of this motif in the interaction of ImuB with the β clamp, a mutated AD-ImuB fusion construct was generated in which the Gln354 residue in ImuB was replaced by an alanine. This construct was used to assess the effect of disrupting the canonical β-binding motif in ImuB on the inferred interaction with DnaN, DnaE2 and ImuA'.

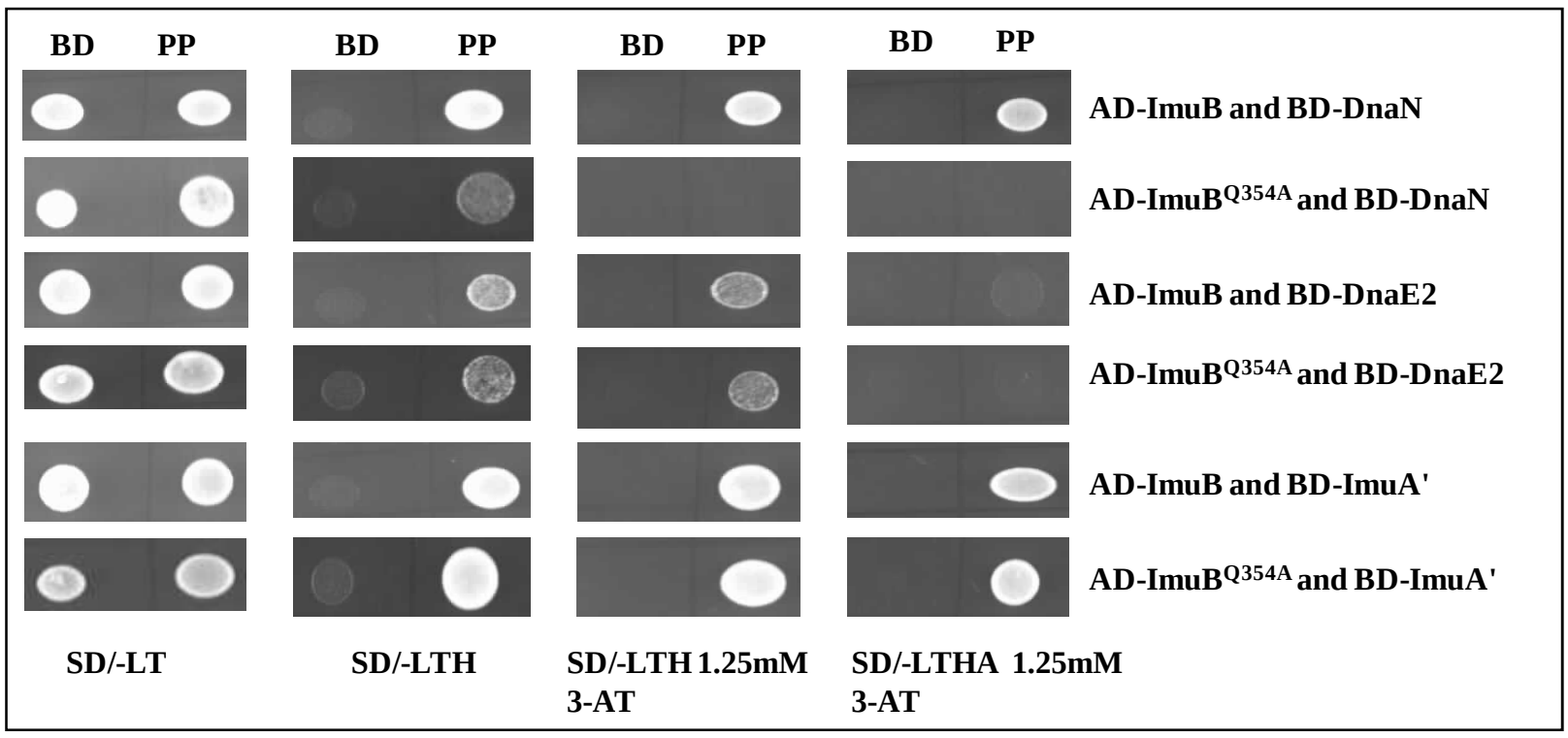


Figure 3.9: Confirmation of the β-binding motif of ImuB. PP indicates the protein-protein interaction and BD indicates the negative control where empty BD plasmid was co-transformed with AD-fusion plasmid. Proteins expressed in-frame with GAL4 are presented as AD and BD fusions.

As shown in Figure 3.9, this mutation specifically abolished the interaction with DnaN, thereby confirming the role in β-binding of the first residue in the predicted motif. Notably, however, this mutation did not affect the interaction of ImuB with ImuA' and DnaE2.

3.2.3 Y2H suggests the ability of ImuB to self-interact

DNA Pol V is composed of two UmuD′ subunits and a single UmuC subunit (Bruck *et al.*, 1996; Woodgate *et al.*, 1989). The absence of a Pol V homologue in organisms possessing ImuA-ImuB-DnaE2 suggests that this cassette may represent a non-orthologous replacement for PolV (Jiang *et al.*, 2009). In order to explore this possibility further, the ability of DnaE2, ImuB and ImuA' to self interact was assessed by Y2H.

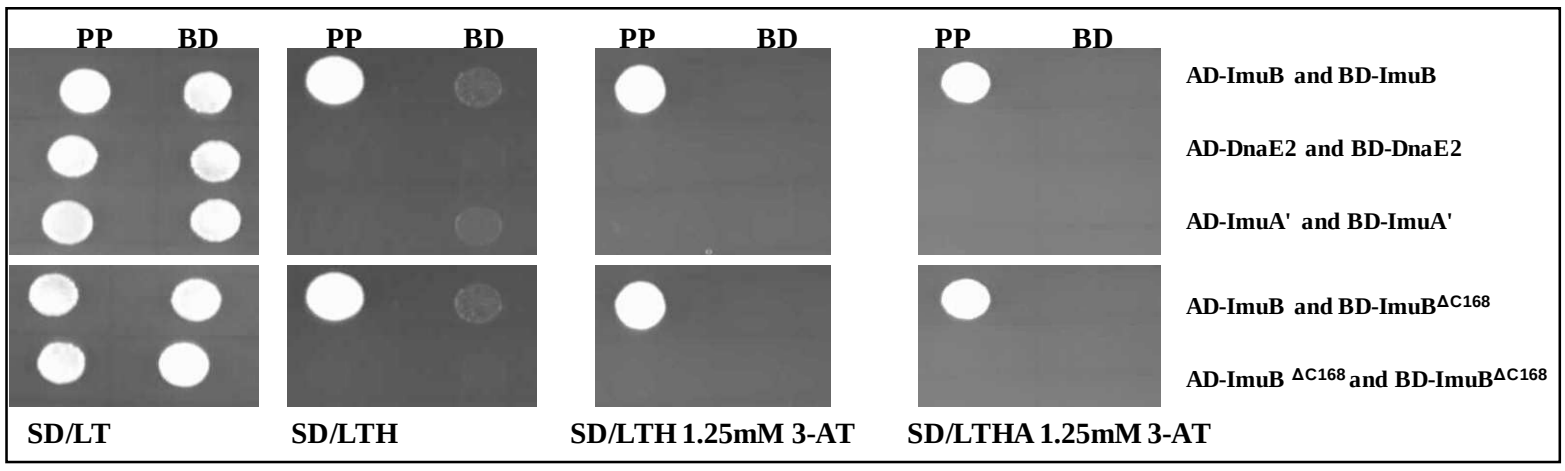


Figure 3.10: ImuB has the ability to self interact. ImuB, DnaE2 and ImuA' were used in both AD- and BD-fusions. As a control for BD-ImuB auto-induction the C-terminal region of ImuB in both AD- and BD-fusions was eliminated.

As shown in Figure 3.10, only ImuB appears capable of self-interaction based on the Y2H data. As described above (section 3.2.1), the BD-ImuB construct is associated with auto-activation, the self-interaction observed in Figure 3.10 may be as a result of the auto-activation of the BD-ImuB. In order to rule out this possibility a deletion of the C-terminal extension in the BD-fusion was tested against full length AD-ImuB and retained the ability to self-interact. However, when the C-terminal extension of ImuB was deleted in both AD- and BD-fusions eliminated the ability to self-interact. This result also indicates that ImuB self-interaction occurs via the C-terminal extension. Since ImuB appears to be the only cassette component that interacts with the β-clamp, this raises the possibility that ImuB might function similarly to the two UmuD´ components which are required to interact with UmuC to give a functional polymerase V in *E. coli* (Sutton *et al.*, 2000; Sutton *et al.*, 2001; Sutton & Walker, 2001).

3.2.4 Assessment of the possibility that other proteins may be involved in this pathway

The C-family polymerase, DnaE1, is essential for replicating the mycobacterial genome with high fidelity (Boshoff *et al.*, 2003). DnaE1 was shown to be up-regulated in response to DNA damage although it contains no identifiable SOS box (Warner *et al.*, 2010), suggesting that it might also be involved in this pathway. To investigate this possibility at the level of protein-protein interactions, the ability of DnaE1 to bind members of the mutagenic complex was assessed by Y2H.

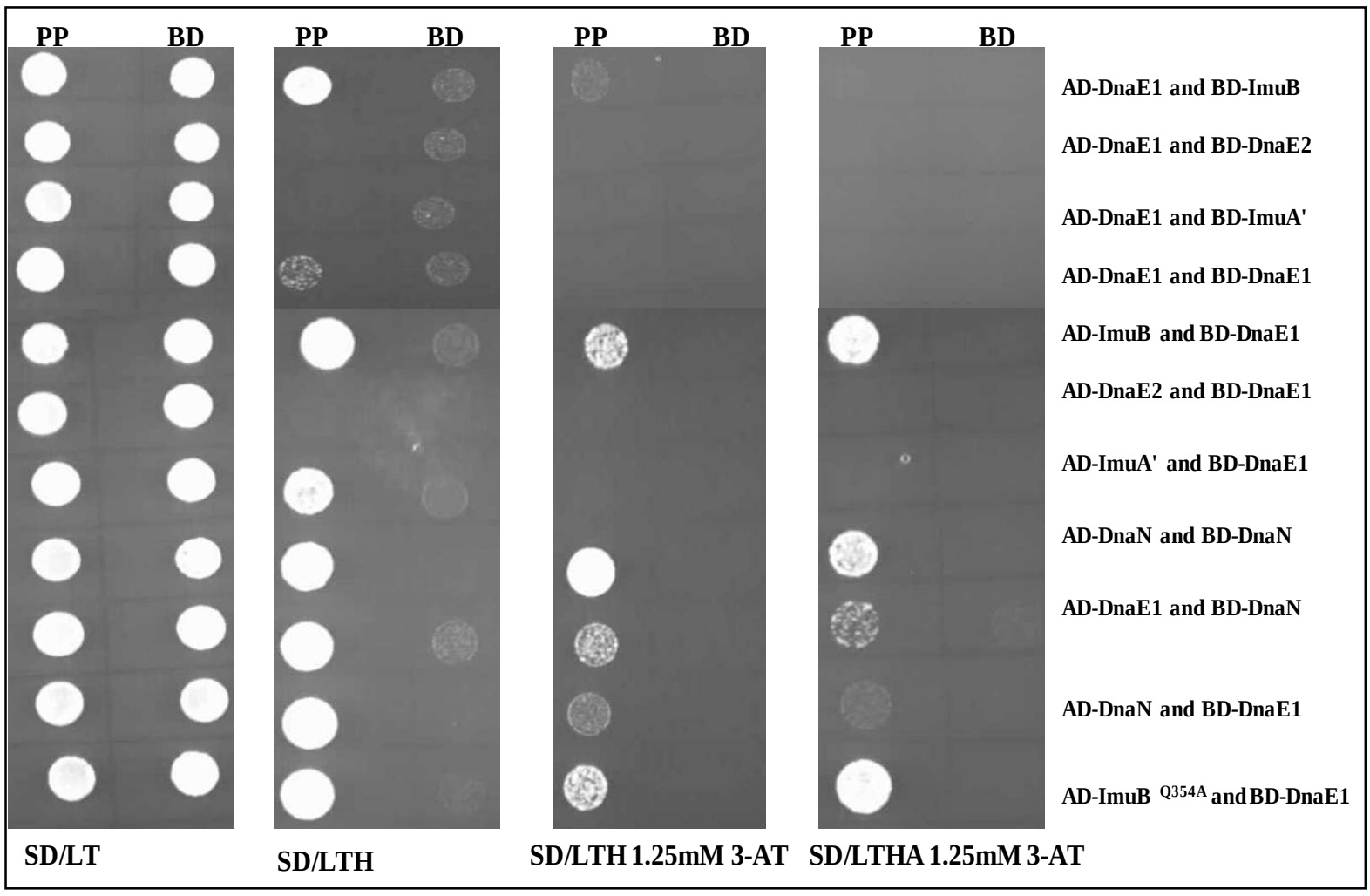


Figure 3.11: The involvement of DnaE1 in the DnaE2-dependent mutagenic pathway. DnaE1 was tested in both the AD and BD fusions against the other proteins.

Consistent with its known β-binding motif, DnaE1 was able to interact with DnaN in both AD- and BD-fusion configurations (Figure 3.11). Interestingly, these results suggested that while DnaE1 does not interact with ImuA' or DnaE2, it is able to interact with ImuB. The results also show that disrupting the β-binding motif of ImuB by mutating the first residue (Gln354Ala) does not interfere with the interaction of DnaE1, thus indicating that the interaction of ImuB with DnaE1 does not involve the β-binding motif. This is surprising given that DnaE1 can interact with the β-clamp directly. As represented in Figure 3.12, ImuB is able to self-interact and there is a possibility of involvement of DnaE1 in this pathway, however the nature of the interactions requires further investigation.

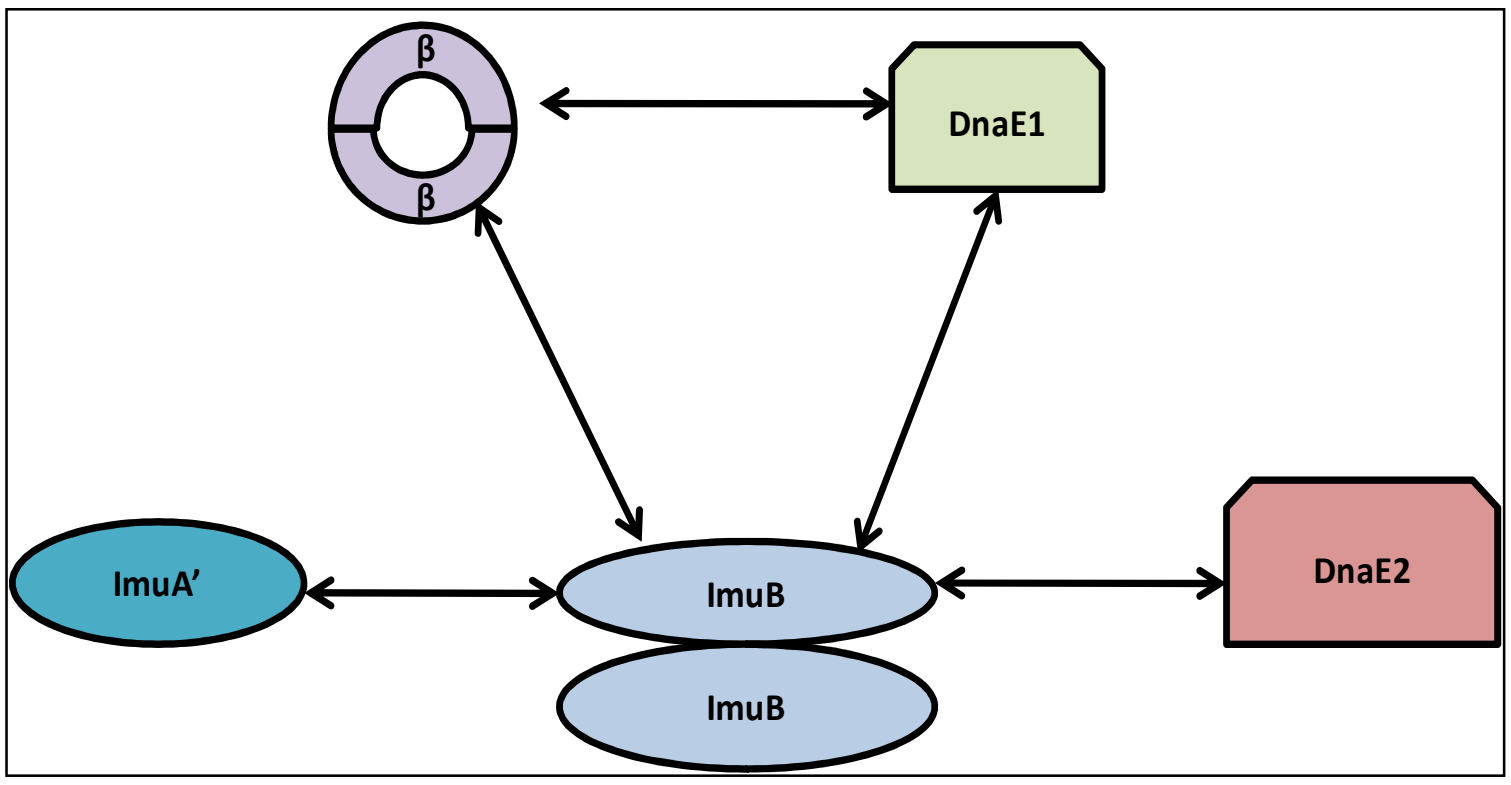


Figure 3.12: Schematic representation of the summary of the Y2H data. ImuB enables DnaE2 and ImuA' to access the DNA template by interacting with the β-clamp and is able to self interact. DnaE1 also interact with ImuB when it can by itself interact with the β-clamp.

3.2.5 Mapping of the interaction domains of ImuB and ImuA'

3.2.5.1 Identification of the interaction domain of ImuB

A defined approach was adopted to map the interaction domain(s) of ImuB by progressively truncating ImuB from the C-terminus towards the N-terminus, thus generating a panel of constructs deficient in specific domains. The truncations were created by introducing artificial start and stop codons depicted in Figure 3.13. These constructs were then tested in the Y2H system for interaction with DnaE1, DnaE2, DnaN and ImuA'.

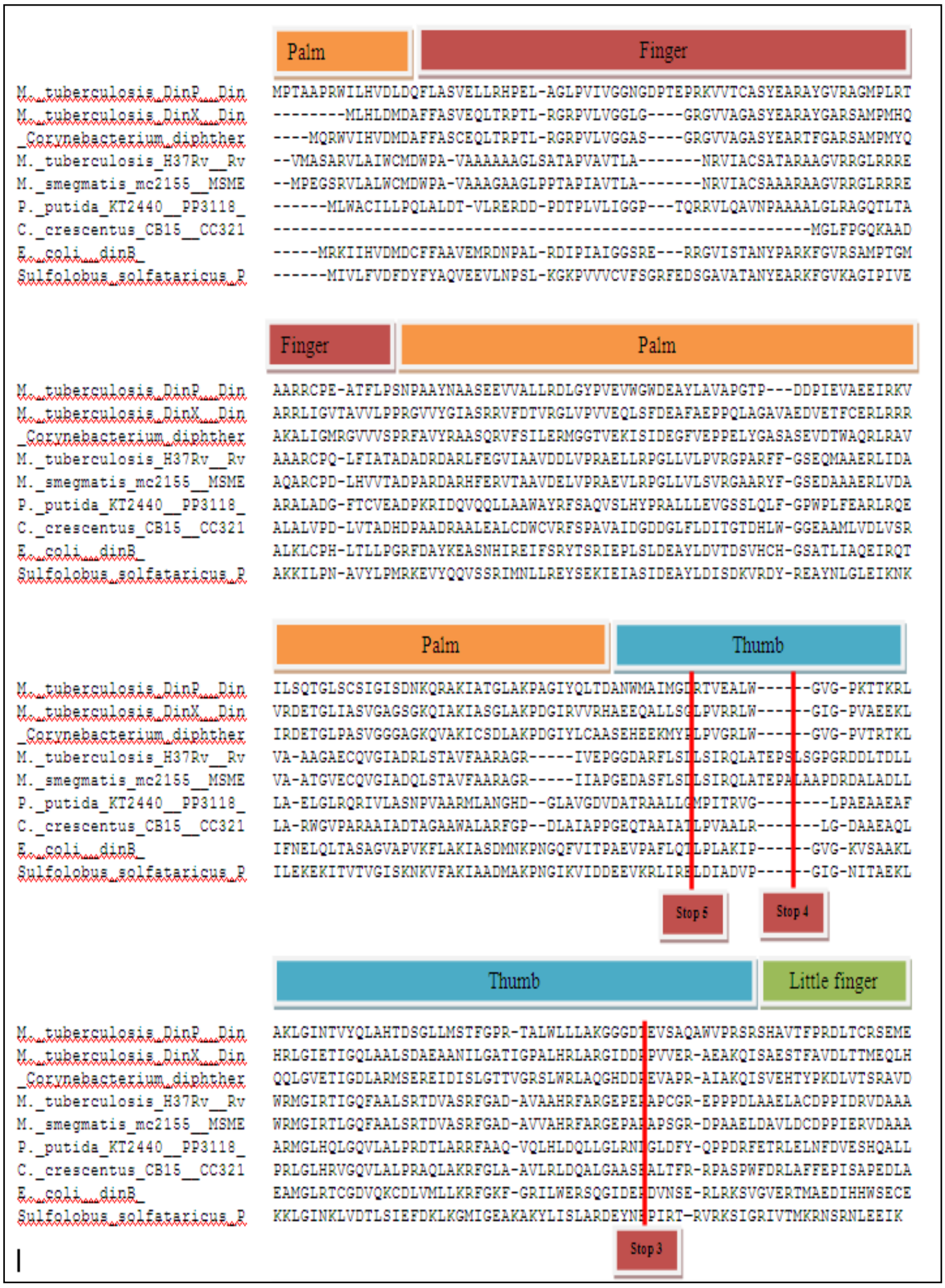

Palm
Finger
M._tuberculosis_DinP__Din MPTAAPRWILHVDLDQFLASVELLRHPEL-AGLPVIVGGNGDPTEPRKVVTCASYEARAYGVRAGMPLRT
M._tuberculosis_DinX__Din --------MLHLDMDAFFASVEQLTRPTL-RGRPVLVGGLG----GRGVVAGASYEARAYGARSAMPMHQ
_Corynebacterium_diphther ----MQRWVIHVDMDAFFASCEQLTRPTL-RGRPVLVGGAS----GRGVVAGASYEARTFGARSAMPMYQ
M._tuberculosis_H37Rv__Rv --VMASARVLAIWCMDWPA-VAAAAAAGLSATAPVAVTLA-------NRVIACSATARAAGVRRGLRRRE
M._smegmatis_mc2155__MSME --MPEGSRVLALWCMDWPA-VAAAGAAGLPPTAPIAVTLA-------NRVIACSAAARAAGVRRGLRRRE
P._putida_KT2440__PP3118_ ------MLWACILLPQLALDT-VLRERDD-PDTPLVLIGGP---TQRRVLQAVNPAAAALGLRAGQTLTA
C._crescentus_CB15__CC321 --MGLFPGQKAAD
E._coli__dinB_ -----MRKIIHVDMDCFFAAVEMRDNPAL-RDIPIAIGGSRE---RRGVISTANYPARKFGVRSAMPTGM
Sulfolobus_solfataricus_P ------MIVLFVDFDYFYAQVEEVLNPSL-KGKPVVVCVFSGRFEDSGAVATANYEARKFGVKAGIPIVE
Finger
Palm
M._tuberculosis_DinP__Din AARRCPE-ATFLPSNPAAYNAASEEVVALLRDLGYPVEVWGWDEAYLAVAPGTP---DDPIEVAEEIRKV
M._tuberculosis_DinX__Din ARRLIGVTAVVLPPRGVVYGIASRRVFDTVRGLVPVVEQLSFDEAFAEPPQLAGAVAEDVETFCERLRRR
_Corynebacterium_diphther AKALIGMRGVVVSPRFAVYRAASQRVFSILERMGGTVEKISIDEGFVEPPELYGASASEVDTWAQRLRAV
M._tuberculosis_H37Rv__Rv AAARCPQ-LFIATADADRDARLFEGVIAAVDDLVPRAELLRPGLLVLPVRGPARFF-GSEQMAAERLIDA
M._smegmatis_mc2155__MSME AQARCPD-LHVVTADPARDARHFERVTAAVDELVPRAEVLRPGLLVLSVRGAARYF-GSEDAAAERLVDA
P._putida_KT2440__PP3118_ ARALADG-FTCVEADPKRIDQVQQLLAAWAYRFSAQVSLHYPRALLLEVGSSLQLF-GPWPLFEARLRQE
C._crescentus_CB15__CC321 ALALVPD-LVTADHDPAADRAALEALCDWCVRFSPAVAIDGDDGLFLDITGTDHLW-GGEAAMLVDLVSR
E._coli__dinB_ ALKLCPH-LTLLPGRFDAYKEASNHIREIFSRYTSRIEPLSLDEAYLDVTDSVHCH-GSATLIAQEIRQT
Sulfolobus_solfataricus_P AKKILPN-AVYLPMRKEVYQQVSSRIMNLLREYSEKIEIASIDEAYLDISDKVRDY-REAYNLGLEIKNK
Palm
Thumb
M._tuberculosis_DinP__Din ILSQTGLSCSIGISDNKQRAKIATGLAKPAGIYQLTDANWMAIMGDRTVEALW------GVG-PKTTKRL
M._tuberculosis_DinX__Din VRDETGLIASVGAGSGKQIAKIASGLAKPDGIRVVRHAEEQALLSGLPVRRLW------GIG-PVAEEKL
_Corynebacterium_diphther IRDETGLPASVGGGAGKQVAKICSDLAKPDGIYLCAASEHEEKMYRLPVGRLW------GVG-PVTRTKL
M._tuberculosis_H37Rv__Rv VA-AAGAECQVGIADRLSTAVFAARAGR-----IVEPGGDARFLSRLLSIRQLATEPSLSGPGRDDLTDLL
M._smegmatis_mc2155__MSME VA-ATGVECQVGIADQLSTAVFAARAGR-----IIAPGEDASFLSRLLSIRQLATEPALAAPDRDALADLL
P._putida_KT2440__PP3118_ LA-ELGLRQRIVLASNPVAARMLANGHD--GLAVGDVDATRAALLGMPITRVG--------LPAEAAEAF
C._crescentus_CB15__CC321 LA-RWGVPARAAIADTAGAAWALARFGP--DLAIAPPGEQTAAIATLPVAALR--------LG-DAAEAQL
E._coli__dinB_ IFNELQLTASAGVAPVKFLAKIASDMNKPNGQFVITPAEVPAFLQTLPLAKIP------GVG-KVSAAKL
Sulfolobus_solfataricus_P ILEKEKITVTVGISKNKVFAKIAADMAKPNGIKVIDDEEVKRLIRELDIADVP------GIG-NITAEKL
Stop 5
Stop 4
Thumb
Little finger
M._tuberculosis_DinP__Din AKLGINTVYQLAHTDSGLLMSTFGPR-TALWLLLAKGGGDTEVSAQAWVPRSRSHAVTFPRDLTCRSEME
M._tuberculosis_DinX__Din HRLGIETIGQLAALSDAEAANILGATIGPALHRLARGIDDRPVVER-AEAKQISAESTFAVDLTTMEQLH
_Corynebacterium_diphther QQLGVETIGDLARMSEREIDISLGTTVGRSLWRLAQGHDDREVAPR-AIAKQISVEHTYPKDLVTSRAVD
M._tuberculosis_H37Rv__Rv WRMGIRTIGQFAALSRTDVASRFGAD-AVAAHRFARGEPERAPCGR-EPPPDLAAELACDPPIDRVDAAA
M._smegmatis_mc2155__MSME WRMGIRTLGQFAALSRTDVASRFGAD-AVVAHRFARGEPARAPSGR-DPAAELDAVLDCDPPIERVDAAA
P._putida_KT2440__PP3118_ ARMGLHQLGQVLALPRDTLARRFAAQ-VQLHLDQLLGLRNIGLDFY-QPPDRFETRLELNFDVESHQALL
C._crescentus_CB15__CC321 PRLGLHRVGQVLALPRAQLAKRFGLA-AVLRLDQALGAASEALTFR-RPASPWFDRLAFFEPISAPEDLA
E._coli__dinB_ EAMGLRTCGDVQKCDLVMLLKRFGKF-GRILWERSQGIDERDVNSE-RLRKSVGVERTMAEDIHHWSECE
Sulfolobus_solfataricus_P KKLGINKLVDTLSIEFDKLKGMIGEAKAKYLISLARDEYNEPIRT-RVRKSIGRIVTMKRNSRNLEEIK
Stop 3



Figure 3.13: Multiple sequence alignment of ImuB homologues indicating the position in *M. tuberculosis* ImuB where artificial stop codons were introduced. The red line indicates the position where stop codons were introduced to sequentially truncate the ImuB protein. The letters in red and underlined indicate the β-clamp binding motif in mycobacterial ImuB and *S. solfataricus* Dpo4.

In the first construct, an artificial stop codon (Stop 1) was introduced immediately after the β-clamp binding motif, creating a truncated protein which more closely resembled the DinB-

type Y-family polymerases of *E. coli* and *S. solfataricus* (Figure 3.13). Elimination of the C-terminal ImuB region abrogated all interactions with other proteins, except for the β-clamp (Figure 3.14). This was a key result, since it suggested that the C-terminal region is necessary for all protein-protein interactions except the ImuB-β-clamp interaction. Moreover, it reinforced the inferred role of the QLPLWG motif in β-clamp binding.

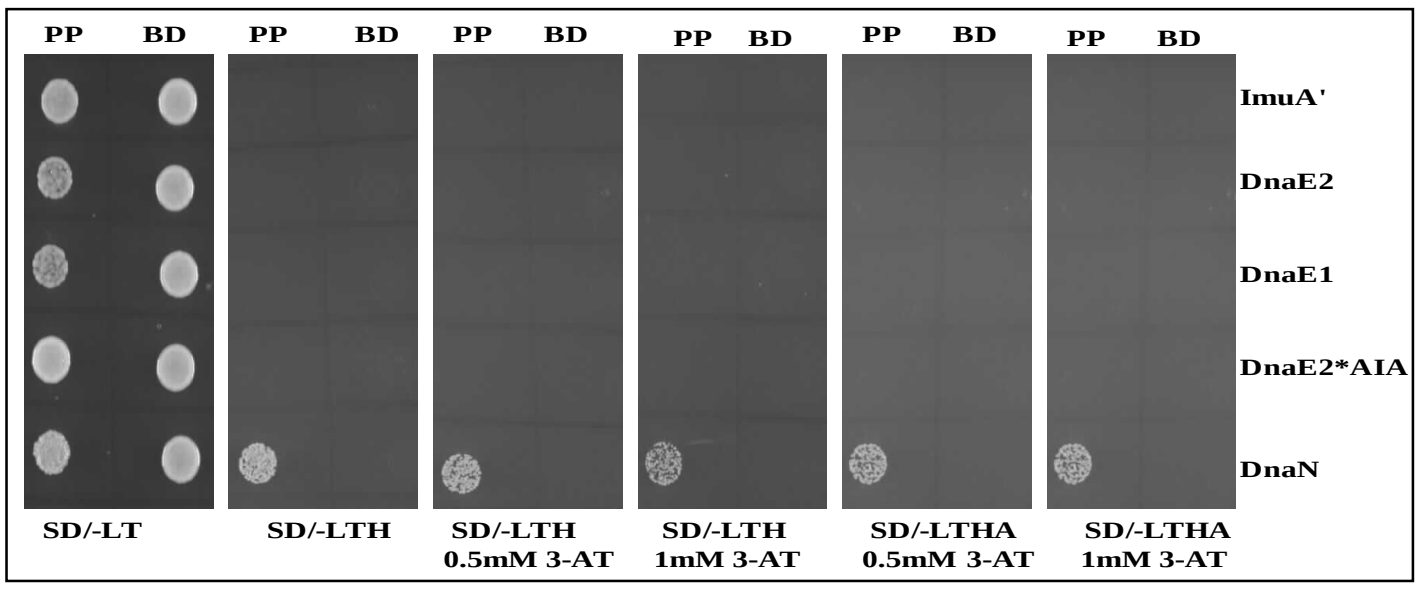


Figure 3.14: Deletion of the C-terminal extension of ImuB eliminated interaction with other proteins but not DnaN. The interaction of ImuB stop 1 against other proteins was tested with ImuB stop 1 in the AD domain and the putative interacting partners in the BD domain.

The second construct had all other domains except that both the β-clamp binding motif and the C-terminal extension were deleted. As shown in Figure 3.15, this abolished the interaction with the *dnaN*-encoded β-clamp as well as the interactions with the other proteins. This result confirmed the identity of the β-binding motif and again suggesting that that the interaction of ImuB with other proteins is mediated by the extended C-terminal region.

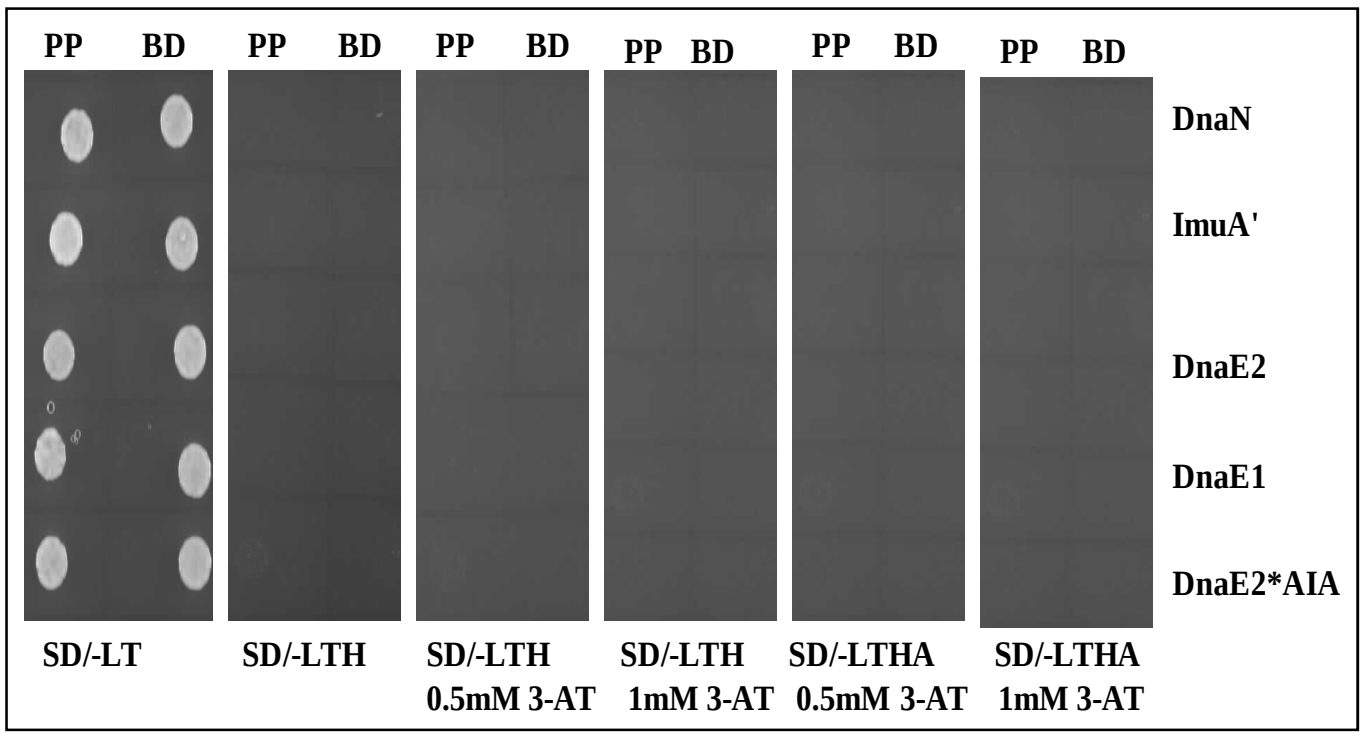


Figure 3.15: Deletion of the putative β-binding motif as well as the C-terminal extension abrogated interaction with all interacting partners. The interaction of ImuB stop 2 with other proteins was tested with ImuB stop 2 as an AD domain fusion and the other proteins as BD domain fusions.

ImuB was then sequentially truncated further towards the N-terminus deleting the little finger domain and into the thumb domains generating three constructs. As above, the interaction with these three constructs was assessed against the other proteins. These extensively truncated forms of ImuB indicated no interaction with DnaE1, DnaE2, ImuA’ or the $DnaE2^{AIA}$ mutant (Figure 3.16).

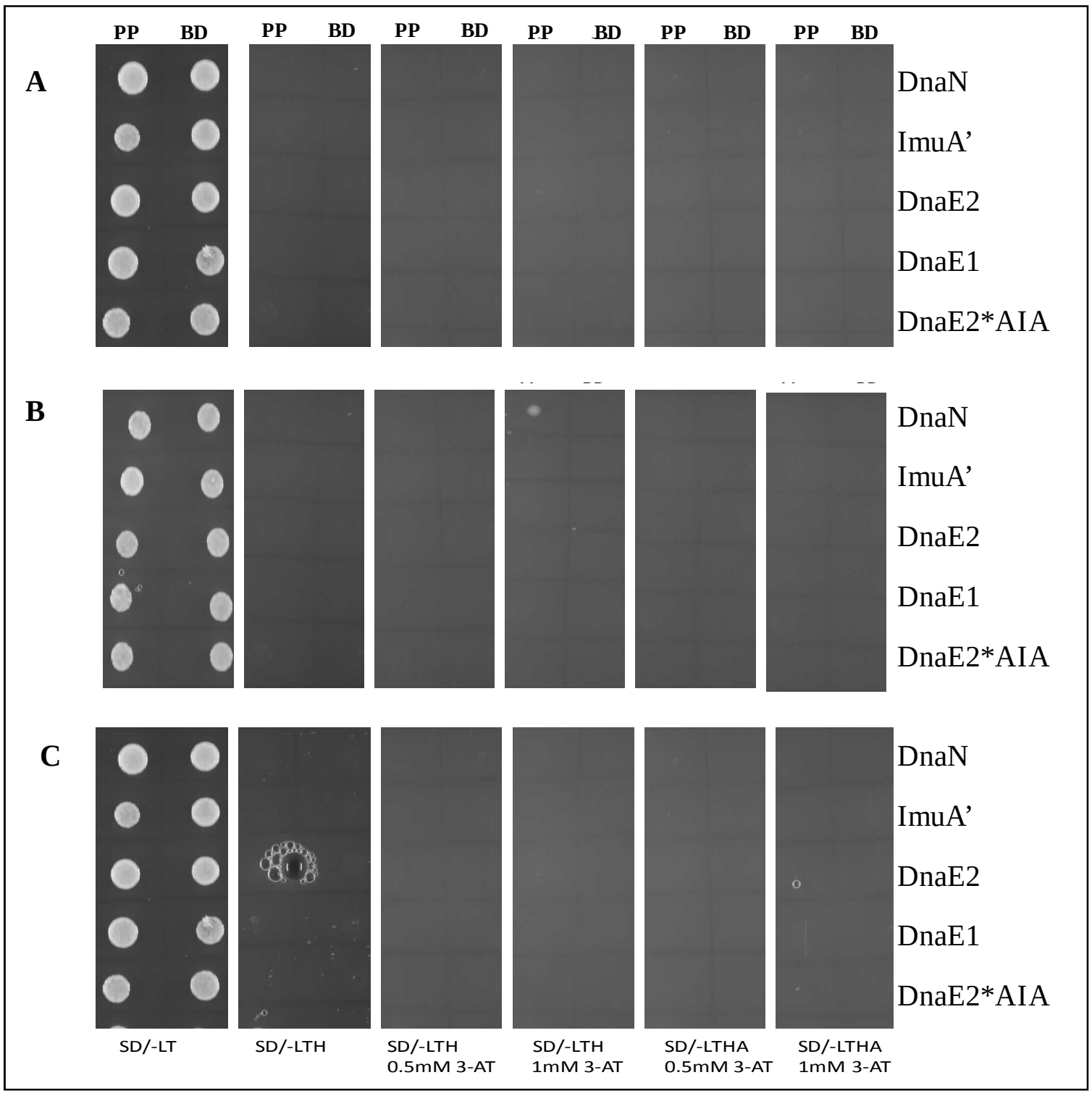


Figure 3.16: Removal of the little finger domain into the thumb domain abrogated interactions with all tested proteins. **A.** Deletion of the little finger domain at position 232 (Stop 3), **B** deletion into the thumb domain at position 176 (Stop 4) and **C** deletion into the thumb domain at position 165 (Stop 5) amino acids.

3.2.5.2 Identification of interacting domains of ImuA'

A similar approach was used to map the site/domain of ImuA' with ImuB. Constructs were generated in which specific domains were deleted by introducing artificial start and stop codons (Figure 3.17).

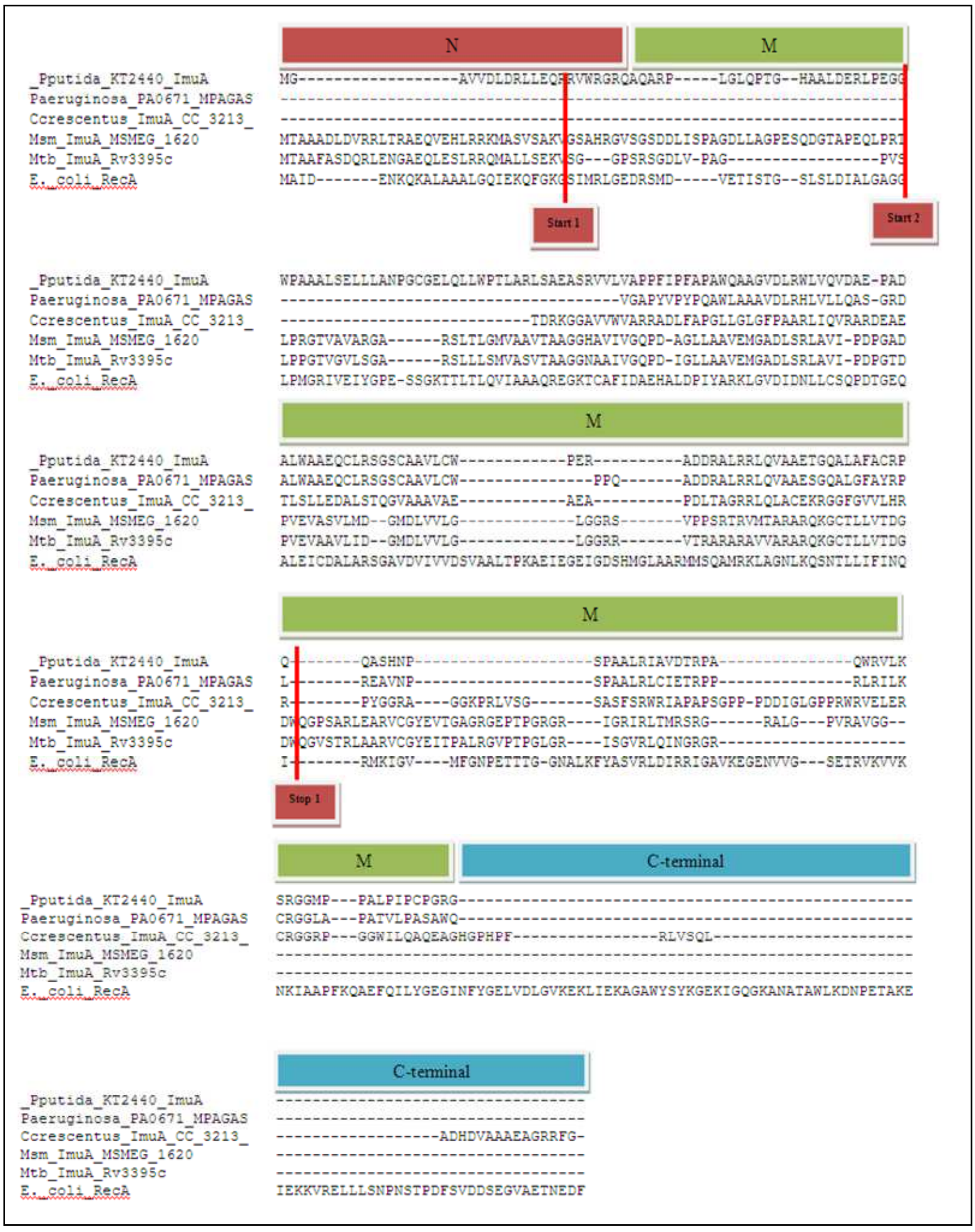


Figure 3.17: ImuA' proteins aligned indicating positions of artificial truncations. ImuA' was sequentially deleted from the N-terminus at positions 32 (Start 1) and 49 (Start 2). The C-terminal region deleted at position 160 (Stop 1) thus creating a C-terminal deficient construct.

The first artificial start codon (start 1) was introduced to remove the N-terminal domain, but leaving the core and the C-terminal domain intact. The second start codon (start 2) was introduced to remove part of the N-terminus into the core domain. The C-terminal domain was truncated by generating a stop codon (stop 1), leaving intact N-terminal and core domains (Figure 3.17).

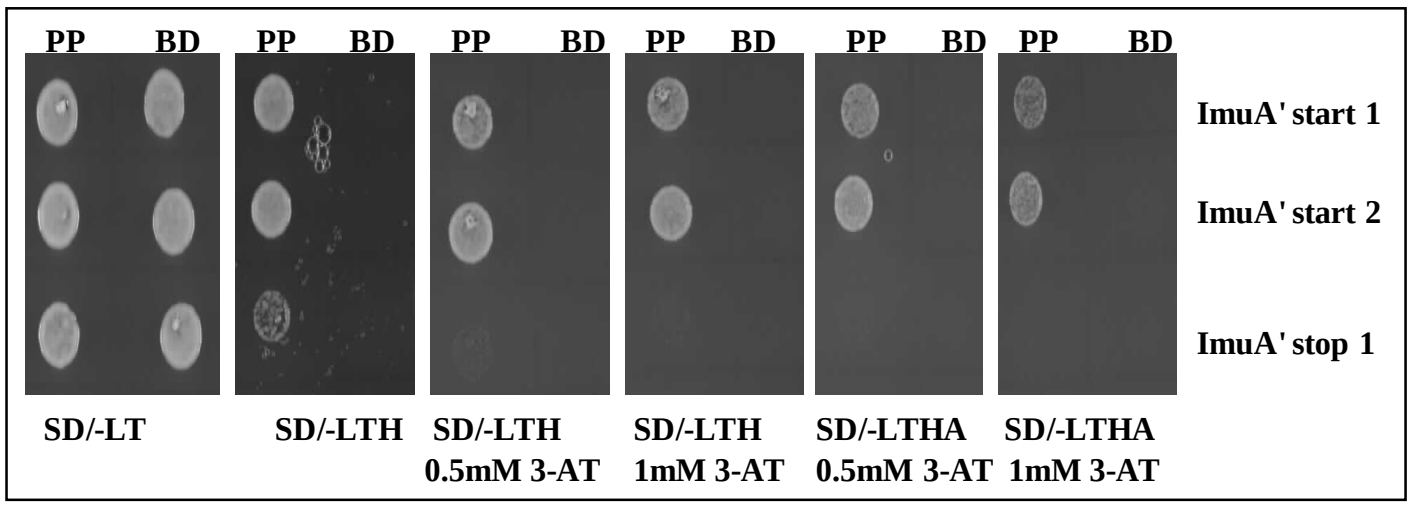


Figure 3.18: The C-terminal region of ImuA' is critical for the interaction with ImuB. Interactions of ImuB with ImuA' truncations were assessed using AD-ImuB and fusions of the ImuA' constructs with the BD domain.

From the results shown in Figure 3.18, it appears that the interaction with ImuB occurs in the C-terminal region of ImuA'. Mapping the interaction domain of ImuB on the C-terminus of ImuA' provides some understanding of the role of ImuA' in induced mutagenesis in that the N- and middle domains are similar to RecA and the C-terminus responsible for interaction with ImuB, suggesting a divergent form of a RecA protein that is required to form a mycobacterial mutasome through interaction with another component of the system.

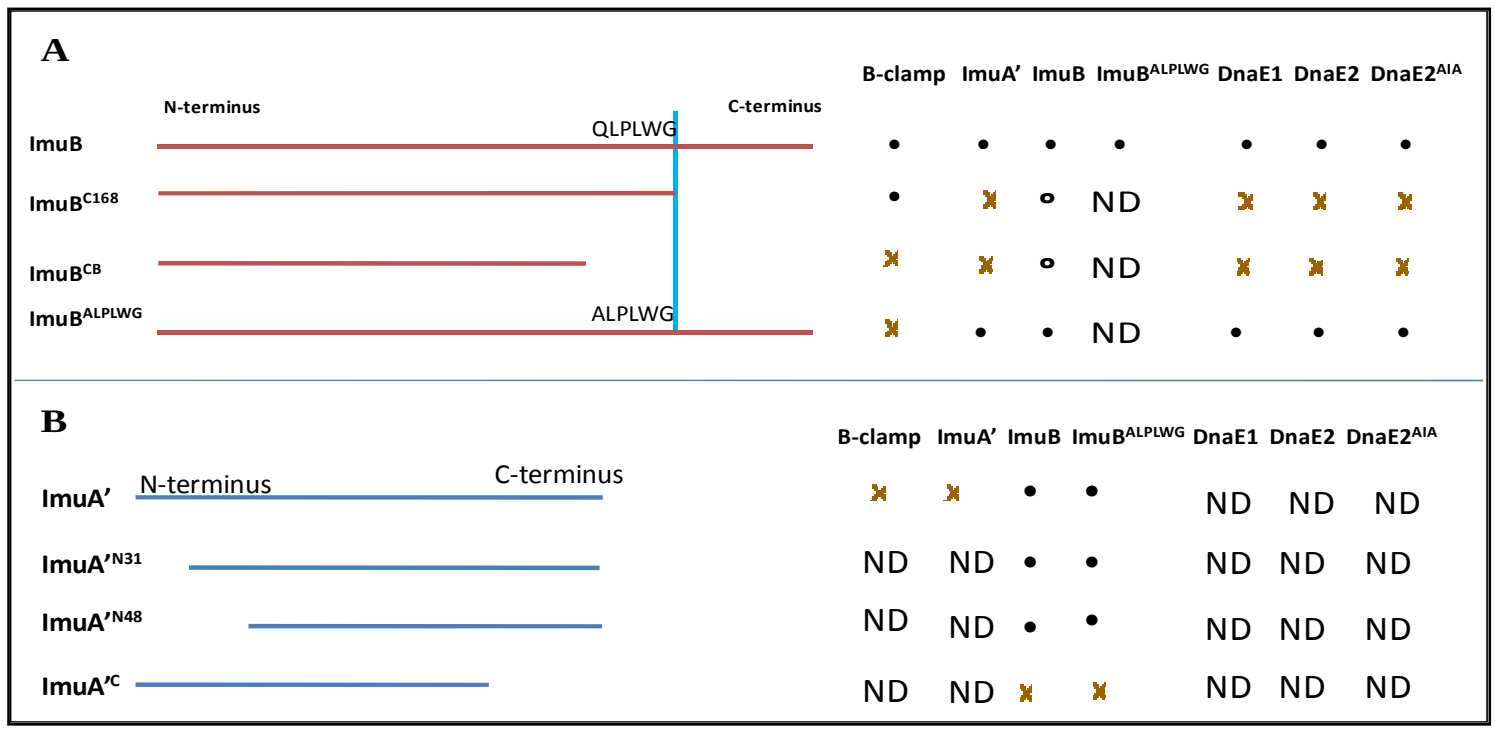


Figure 3.18: Summary of the protein-protein interactions inferred from Y2H analyses. A. Interaction of ImuB with other proteins. Full-length ImuB is able to interact with the other cassette components, DnaE2 and ImuA', as well as DnaE1 and the β-clamp. Removal of the 168 amino acid C-terminal region eliminates the interaction with other proteins except the β-clamp. Deletion of the β-clamp binding motif together with the C-terminus abrogates all interactions while a single mutation of the first amino acid of the β-clamp binding motif eliminates the interaction with the β-clamp but not the other proteins. **B.** Full length ImuA' interacts with ImuB only, and mutating the β-clamp binding motif does interfere with this interaction. Deletion of the C-terminal region of ImuA' disrupted the interaction with ImuB but not the deletion of 31 and 41 amino acids on the N-terminus of ImuA'. ND indicates not determined, • indicates interaction, × indicates no interaction observed and º indicates interactions where full-length ImuB interacted with truncated alleles.

The data presented in Section 3.2.5 are summarised in Figure 3.18, indicating that the C-terminal regions of ImuB and ImuA' are important for interaction with other proteins and also confirming the identity of the β-clamp binding motif. However, mapping the regions of interactions of the accessory factors using Y2H does not indicate whether these regions are critical for protein function in the mycobacterial cell.

3.3 Analysis of the *imuA'-imuB/dnaE2* cassette of *M. smegmatis*

In order to assess the role of the inferred protein-protein interactions in DnaE2 function, a combination of DNA damage-induced mutagenesis and DNA damage survival assays was applied. In order to survive DNA damage, cells employ TLS which can, in some cases, be mutagenic. This provides the rationale for assessing both survival and mutagenesis in cells exposed to genotoxic agents. The assays were performed in a model organism, *M. smegmatis*, which is a rapidly growing non-pathogenic mycobacterium. The genomic context of the *imuA'-imuB/dnaE2* cassette is loosely conserved between *M. tuberculosis* and *M. smegmatis*.

In *M. smegmatis* mc^{2}155, *imuA*' and *imuB* are situated 10.8 kb upstream of *dnaE2* whereas in *M. tuberculosis imuA*' and *imuB* are located ~24.7 kb upstream of *dnaE2*. However, the encoded proteins are highly conserved (Figure 3.2 and 3.3), and a LexA-binding site is located in the promoter regions of upstream of *imuA*' both in *M. tuberculosis* (Davis *et al.*, 2002) and *M. smegmatis*.

In order to establish whether the interactions inferred from Y2H analyses of *M. tuberculosis* proteins apply similarly to *M. smegmatis*, the Y2H analysis was repeated using the *M. smegmatis* ImuA', ImuB, β-clamp and DnaE2 using both AD and BD domain constructs. In order to distinguish the proteins from *M. smegmatis* from those of *M. tuberculosis*, they are named MsImuA', MsImuB, Msβ-clamp and MsDnaE2. In addition, the first catalytic residue of the β-clamp binding motif of MsImuB Q354A was mutated to confirm the β-clamp binding motif.

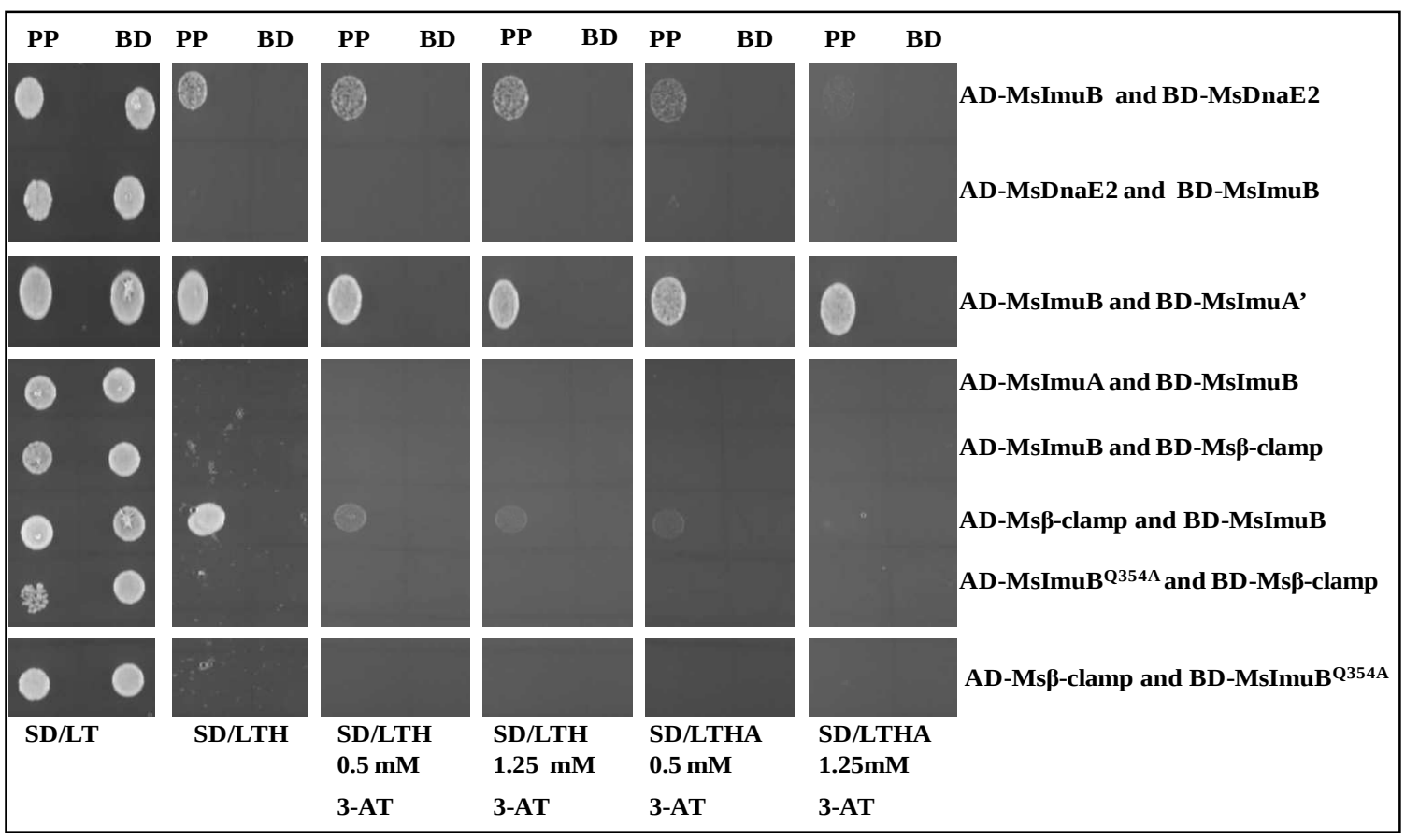
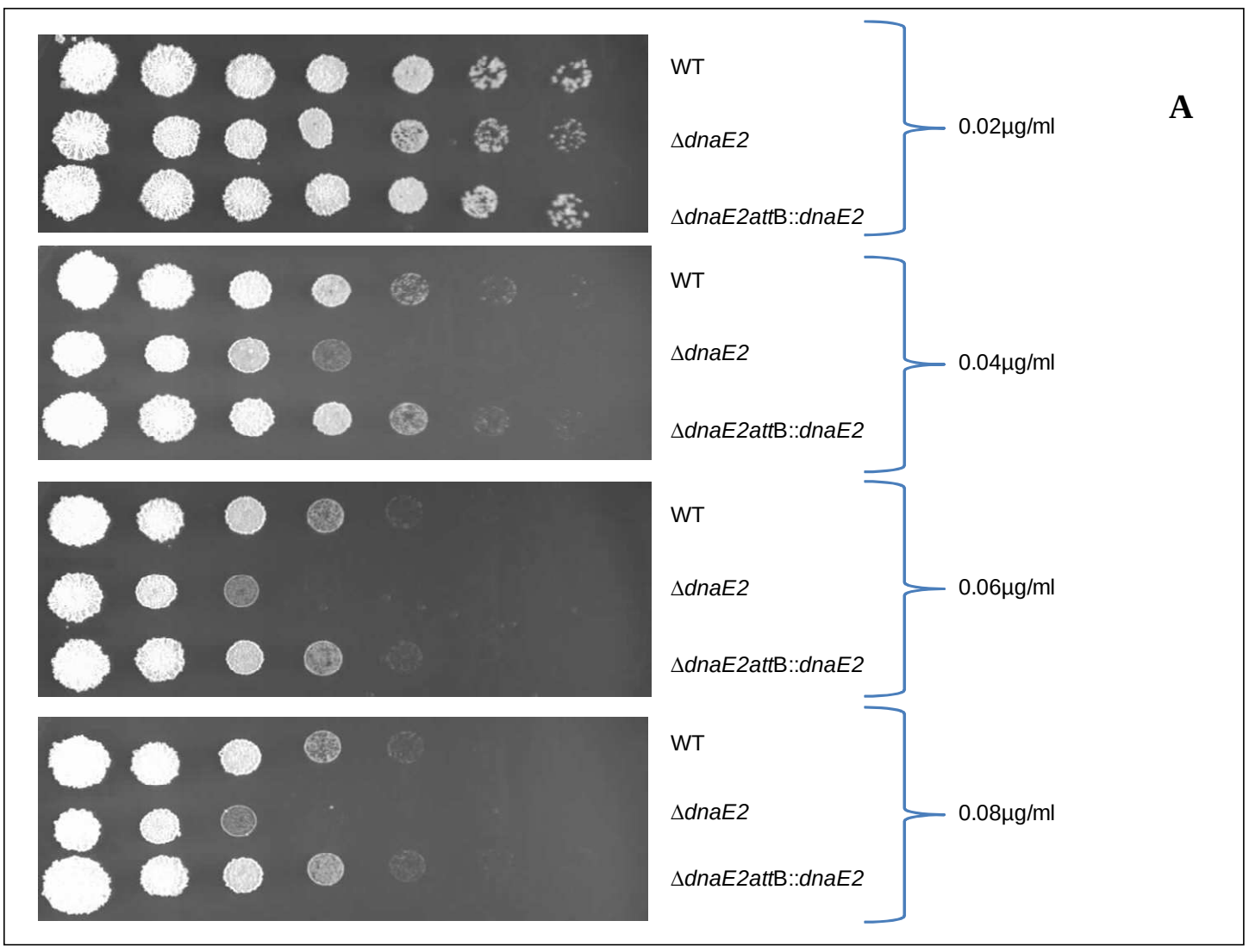


Figure 3.19: Analysis of *M. smegmatis* mutagenic cassette component interactions. MsImuB as an AD domain fusion showed interaction with MsImuA', MsDnaE2 and Msβ-clamp but did not show interaction with the other protein when used as a BD domain. This is contrary to the BD-ImuB of *M. tuberculosis* where auto-activation was observed. In another twist, the MsImuB as a BD domain fusion retained the interaction with Msβ-clamp but at highest stringency, the interaction was eliminated. The results also show that mutating the first catalytic residue of the β-clamp binding motif abrogates the interaction with the β-clamp, recapitulating the observations with the *M. tuberculosis* proteins.

3.3.1 Analysis of the DnaE2-dependent mutagenic pathway in *M. smegmatis*

In accordance with established methods (Boshoff *et al.*, 2003), DNA damage-induced mutagenesis was determined by measuring the frequency of emergence of rifampicin resistant (Rif^R) mutants following DNA damage. Initial analyses were performed utilizing the Δ*dnaE2* deletion mutant and its complemented derivative, Δ*dnaE2 attB*::*dnaE2* constructed by Dr. H. Boshoff and described previously (Boshoff *et al.*, 2003) as controls to validate the assays. The data presented in Figure 3.20 indicate that the Δ*dnaE2*deletion mutant is sensitive to DNA damage and that this phenotype is reversed in the complemented strain, confirming that DnaE2 is required for DNA damage tolerance in *M. smegmatis*. These data also indicate that the Δ*dnaE2* mutant is impaired in the ability to form Rif^R mutants, and the phenotype is restored to wild type by complementation.



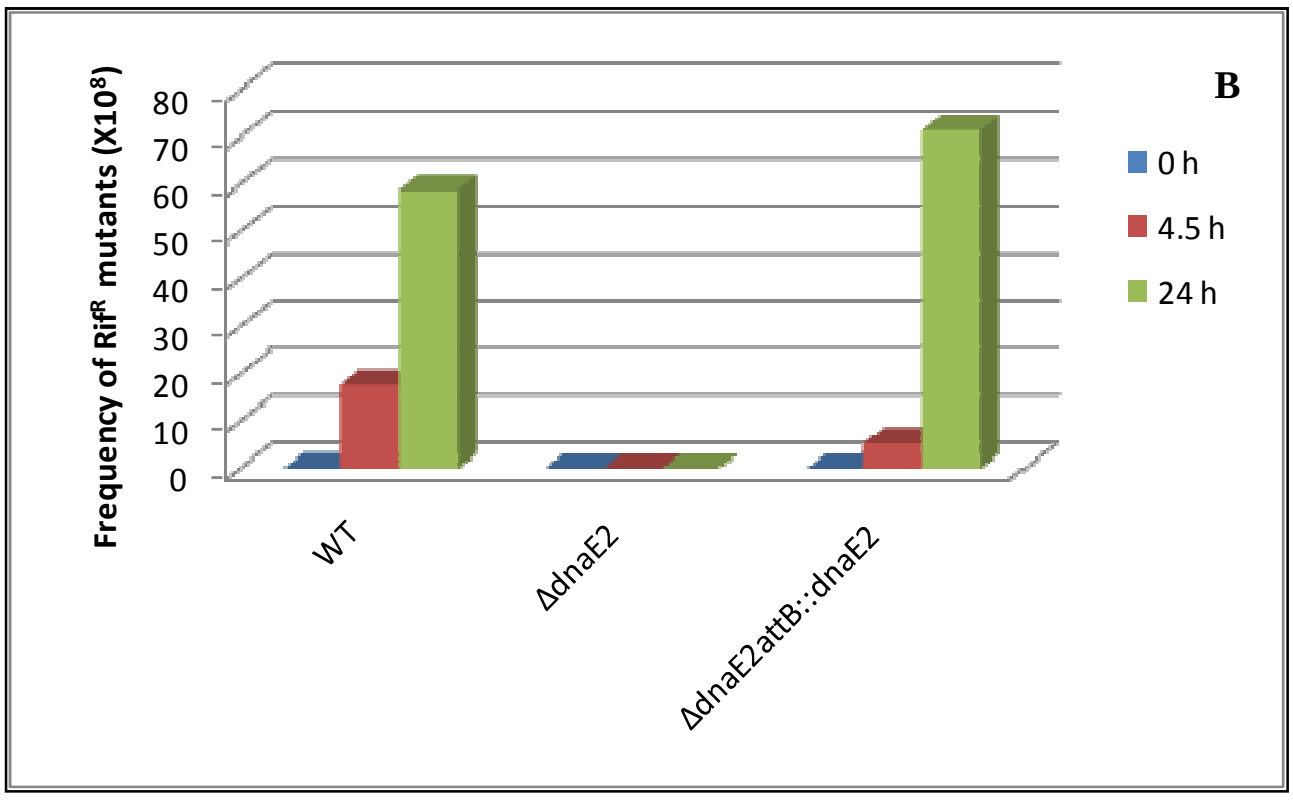


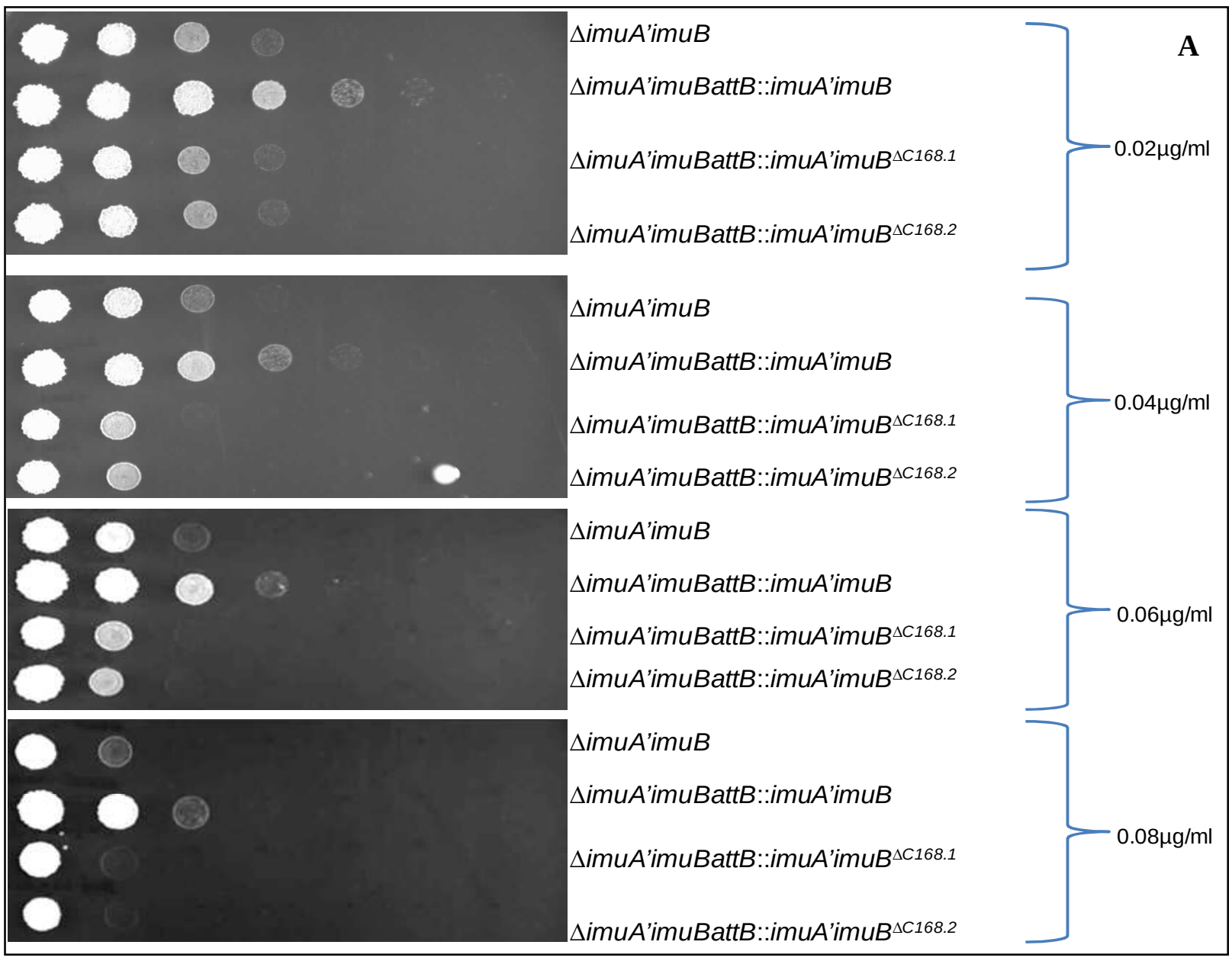
Figure 3.20: DnaE2 is essential for damage tolerance and induced mutagenesis in *M. smegmatis*. **A.** DNA damage tolerance assay of the wild type mc^2155 (WT), *dnaE2* deletion mutant and *its* complemented derivative, (Δ*dnaE2ttB*::*dnaE2*, following exposure to mitomycin C (MMC). Cells were grown to mid-log phase and ten-fold dilution's spotted on solid medium containing different concentrations (0.02, 0.04, 0.06 and 0.06 µg/ml) of MMC. **B**. UV-induced mutation frequency to rifampicin resistance (Rif^R). Cells were grown to mid-log phase and irradiated with UV at 25 J/cm^2 before being allowed to recover in fresh medium for 24 h. At 4.5 h and 24 h post irradiation, cells were sampled and plated on solid medium supplemented with Rif (200 µg/ml) and the frequency of Rif^R mutants determined as a function of total viable cells. Data are from a representative experiment performed in triplicate.

3.3.2 ImuB and ImuA' are required for DnaE2-dependent mutagenesis in *M. smegmatis*

The data in section 3.2 enabled the identification of interacting domains predicted to be necessary for DnaE2-dependent damage tolerance. Single knockout mutants of *M. smegmatis* lacking the homologs of ImuB and ImuA', Δ*imuA'* and *imuB*, respectively, as well as the double knockout mutant Δ*imuA'*-Δ*imuB* were constructed by Dr. Warner. In parallel with the study described here, these mutants were used to show that deletion of either ImuB or ImuA' abrogated induced mutagenesis and resulted in hypersensitivity of *M. smegmatis* to the DNA damaging agent, MMC, a phenotype that was reversed by complementation with a full-length copy of the relevant gene (*imuA*' or *imuB*, depending on the deletion allele). This provided evidence that ImuB and ImuA' are individually essential for DnaE2-dependent mutagenesis, and informed the experiments described below.

3.3.3 Analysis of truncated ImuB

Y2H analyses utilizing artificially truncated proteins (Figure 3.18) indicated that the C-terminal extension was critical for interactions with the other cassette components (DnaE2, ImuA') as well as DnaE1. To determine if the C-terminus of ImuB is important for function, a stop codon was introduced in ImuB encoding gene to remove the 168 amino acids of the unstructured region immediately after the β-clamp binding motif. This allele was introduced by complementation at the *att*B site in the Δ*imuA'imuB* mutant strain. DNA damage induced mutagenesis and survival was assessed in this strain. Deleting the C-terminal unstructured extension of ImuB phenocopies the Δ*imuB* phenotype in the presence of ImuA' and DnaE2 (Figure 3.21). This indicates the equal requirement of ImuA', ImuB and DnaE2 for induced mutagenesis as well as the importance of the interaction region of ImuB to allow interaction with other protein for function. This result is significant as it further shows that ImuB is the adaptor that mediates access of the other protein to template DNA through the C-terminal extension. In combination, these results imply that the ability of ImuB to interact with other proteins might be critical for its biological function.



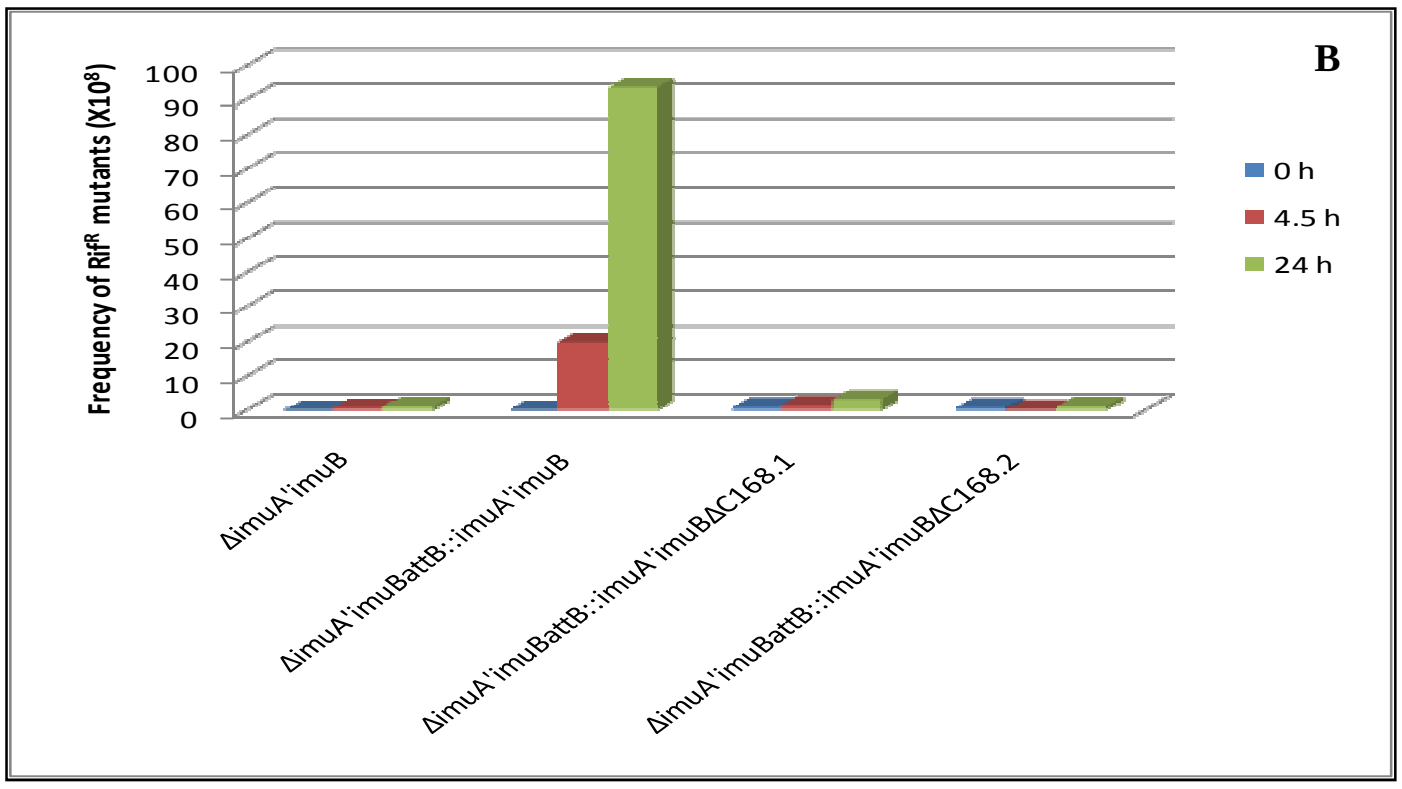
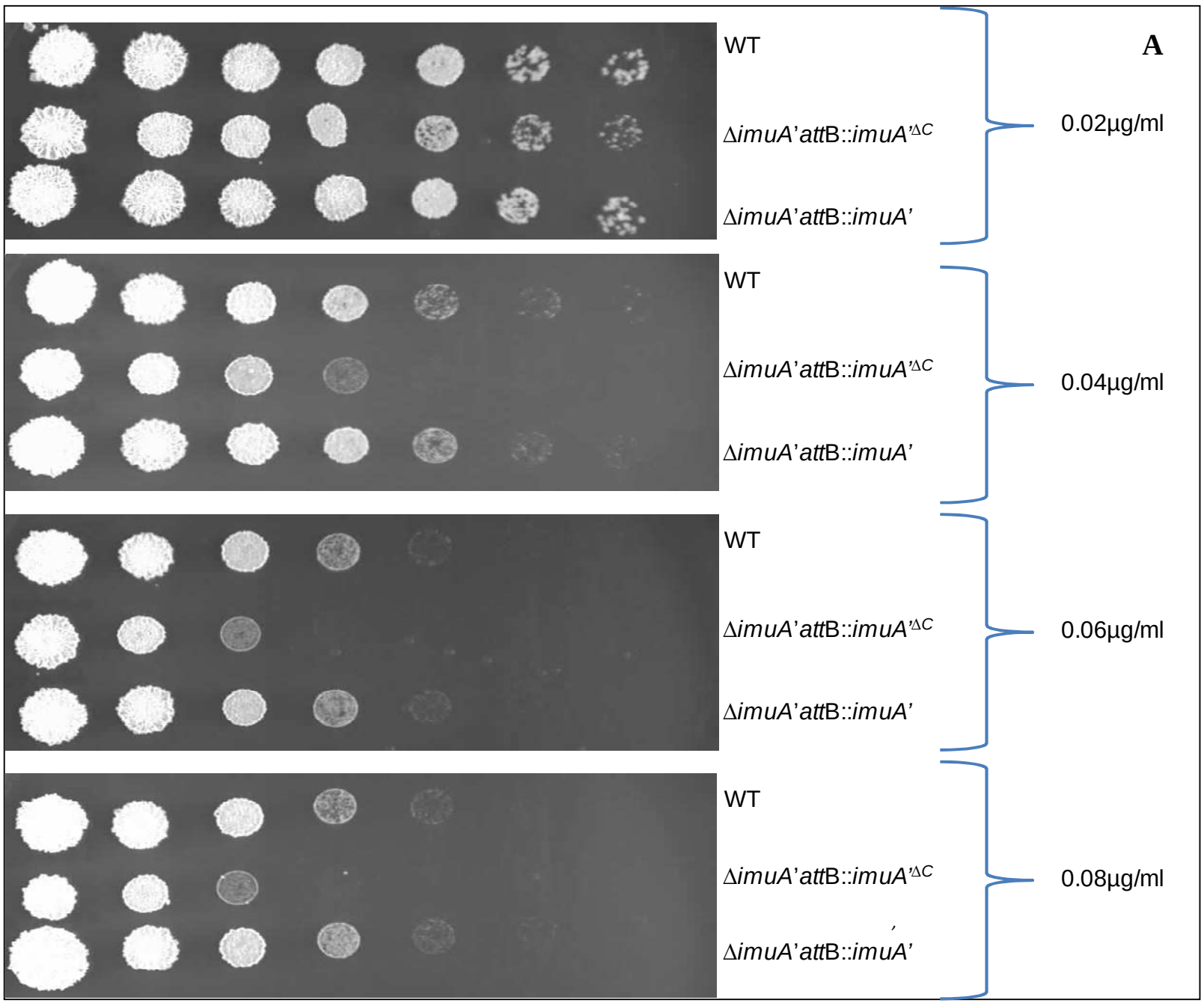


Figure 3.21: Deletion of the C-terminal region of ImuB phenocopies the Δ*imuB* phenotype. (A) DNA damage tolerance assay and (**B**) UV-induced mutagenesis of the double deletion mutant (Δ*imuA'-imuB*), the complemented Δ*imuBatt*B::*imuA'-imuB* derivative and two independent clones in which the complementing allele carried a 3-terminally truncated form of *imuB* that gives rise to with 168 amino acids C-terminal deletion (Δ*imuA'imuBatt*B::*imuA'imuB*$^{\Delta C168.1}$ and Δ*imuA'imuBatt*B::*imuA'imuB*$^{\Delta C168.2}$, respectively). Data are from a representative experiment performed in triplicate.

3.3.4 Analysis of truncated ImuA'

The Y2H analysis described in Section 3.2.5 suggested that the C-terminal region of ImuA' is important for the interaction with ImuB. To determine of the interaction observed by Y2H was necessary for function in *M. smegmatis*, the consequence of C-terminal truncation of ImuA' ImuB on damage tolerance and induced mutagenesis was investigated. To do this, a complementation vector was constructed in which an artificial stop 1 was introduced to truncate ImuA' after the middle domain eliminating the C-terminus (Figure 3.17). The ability of this construct to complement an ImuA' deletion mutant was then determined in both induced mutagenesis and DNA damage tolerance assays.



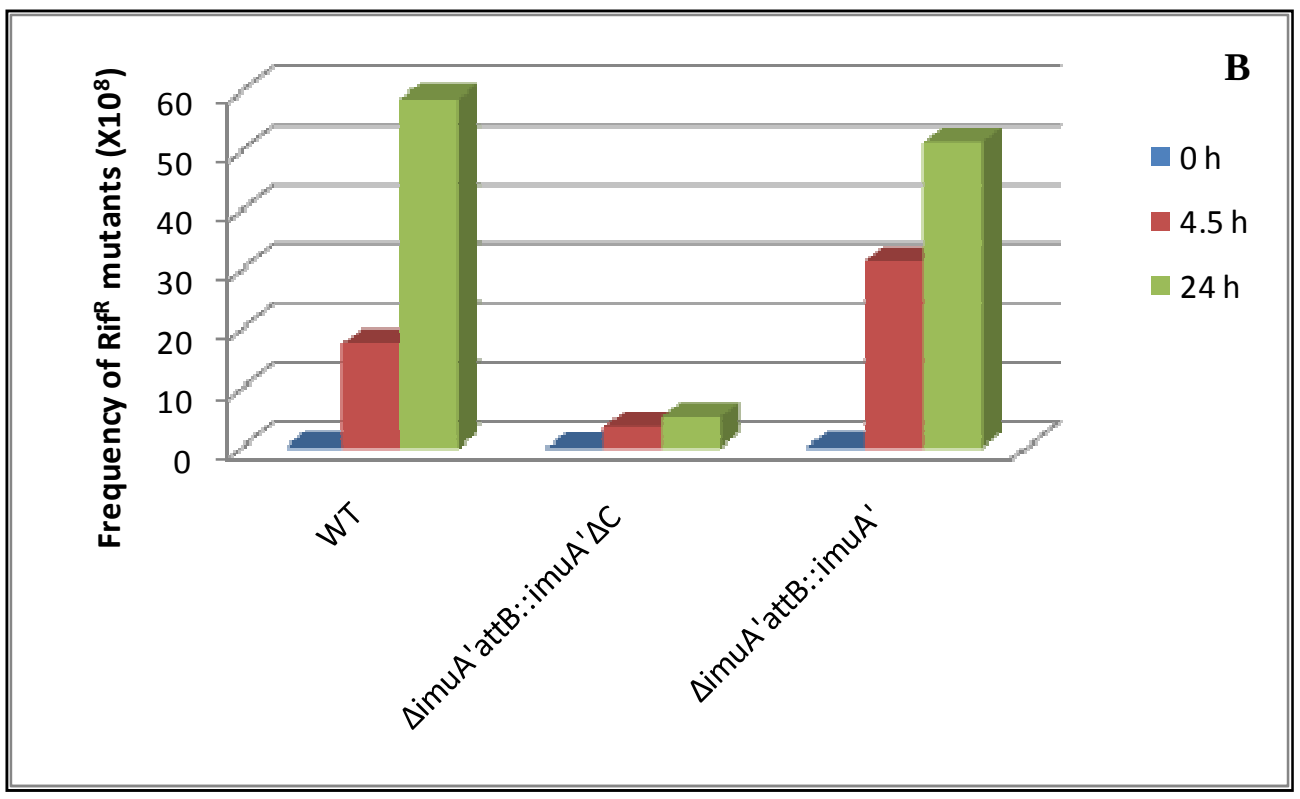


Figure 3.22: The C-terminal region of ImuA' is essential for damage tolerance and induced mutagenesis. (A) DNA damage tolerance assay and (**B**) UV-induced mutagenesis of the wild type mc^{2}155 (WT), a complemented derivative of the Δ*imuA'* mutant carrying full-length imuA' (Δ*imuA'att*B::*imuA'*), or a derivative carrying 3'-truncated imuA' that encodes a truncated form of ImuA' with 44 amino acids deleted from the C-terminus (Δ*imuA'att*B::*imuA'*$^{\Delta C}$). Data are from a representative experiment performed in triplicate.

The data indicate that deletion of the C-terminal region of ImuA' renders the cells sensitive to DNA damage; moreover, the truncated ImuA' is unable to support induced mutagenesis (Figure 3.22) and so phenocopies the *imuA'* gene deletion mutant (Warner *et al.*, 2010). These results suggest that the C-terminal regions of both ImuB and ImuA' are critical for function. Since the protein-protein interactions identified in this study also map to the C-termini of ImuB and ImuB, the implication is that the protein-protein interactions are essential for function of the split mutagenesis cassette.

3.4 Identification of other proteins involved in DnaE2-dependent damage tolerance and induced mutagenesis in *M. smegmatis*

The unexpected protein-protein interaction observation between DnaE1 and ImuB suggested the possibility that other proteins might be involved in DnaE2-dependent damage tolerance. To explore this possibility further, a pull-down approach was designed to analyse the proteins involved in the DNA damage response in *M. smegmatis*. These methods requires tagging the protein of interest inside the cell with an epitope, and then subjecting the host cell to conditions known to induce the protein of interest and interacting partners (Phizicky & Fields, 1995).

3.4.1 Introduction of a tag on the C-terminus of MsImuB

The working model underlying this study proposed that ImuB serves as a molecular adapter allowing other proteins to access the damaged template DNA. Therefore, for the pull-down experiments, ImuB was tagged on the C-terminus with a His-FLAG tandem affinity purification (TAP) tag (Figure 3.23).

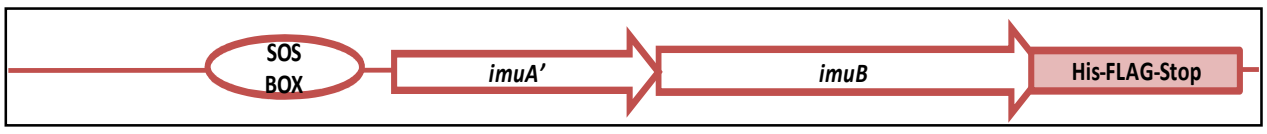


Figure 3.23: Schematic representation of the TAP tag introduced on the C-terminus of MsImuB. A TAP tag was introduced into the C-terminus of *M. smegmatis* ImuB (MsImuB) in order to use it as *in vivo* bait to identify other interacting partners. The tag was introduced into a complementing vector in the double mutant Δ*imuA'imuB* background. The SOS box denotes the region in the operator where the LexA protein binds to control expression of down stream genes.

Previous results had established that the ImuB C-terminal region is critical for protein-protein interactions (Figure 3.14) and, in turn, for DNA damage tolerance and induced mutagenesis (Figure 3.21). In order to evaluate the impact of the C-terminal TAP tag on ImuB function,

the mutant carrying the tagged ImuB protein was applied in DNA damage tolerance and induced mutagenesis assays.

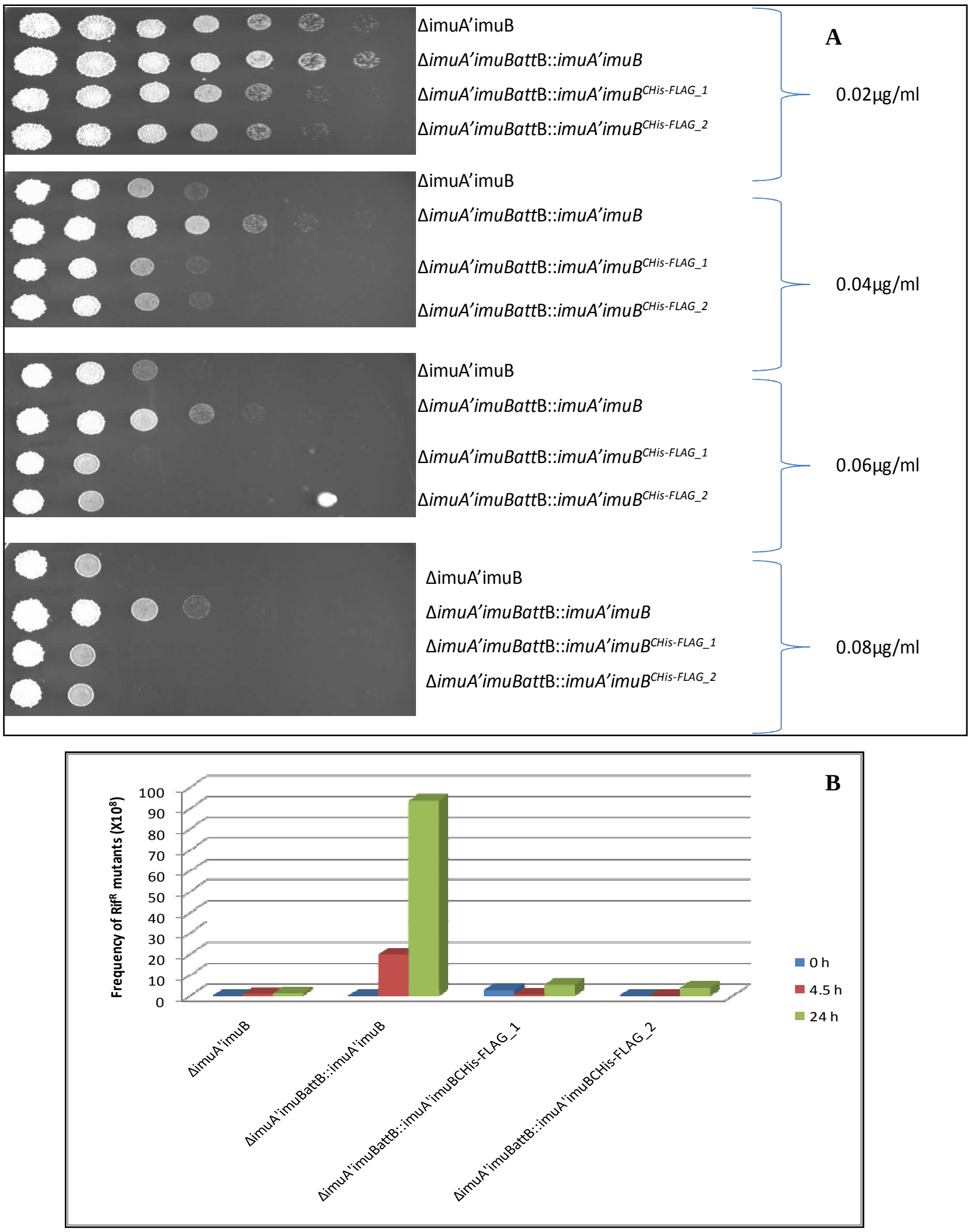


Figure 3.24: Introduction of a FLAG-tag on the C-terminus of ImuB impairs protein function. **(A)** DNA damage tolerance assay and (**B**) UV-induced mutagenesis of the double deletion mutant, Δ*imuA'-imuB*, the complemented strain carrying the full length *imuA'-imuB* region

(Δ*imuBatt*B::*imuA'imuB*), and two independent clones carrying the His-FLAG tag on the C-terminus of ImuB in the complementing vector (Δ*imuA'imuBatt*B::*imuA'imuB*$^{\text{CHis-FLAG_1}}$ and Δ*imuA'imuBatt*B::*imuA'imuB*$^{\text{CHis-FLAG_2}}$, respectively). Data are from a representative experiment performed in triplicate.

The results shown in Figure 3.24 suggested that the introduction of the TAP tag appeared to disrupt protein function as complementation was not observed when ImuB was C-terminally tagged. This result was consistent with the importance of the C-terminal region of ImuB and suggested that the tag might interfere with the ability of ImuB to interact with other proteins via its C-terminal region. Based on this result, an alternative strategy, involving the N-terminal labelling of ImuB with a 3×FLAG tag, was then adopted.

3.4.2 Introduction of a tag on the N-terminus of MsImuB

3× FLAG tag comprising the 3× DYKDDDDK (Einhauer & Jungbauer, 2001) sequence was introduced immediately after the start codon of MsImuB (Figure 3.25).

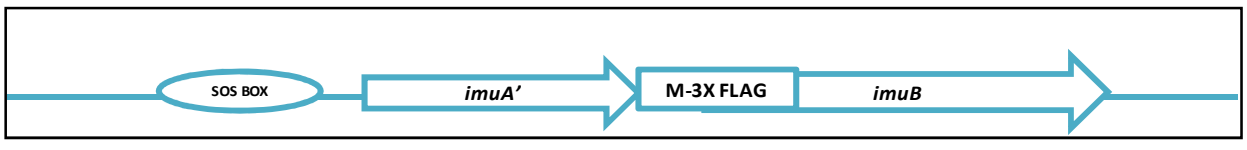
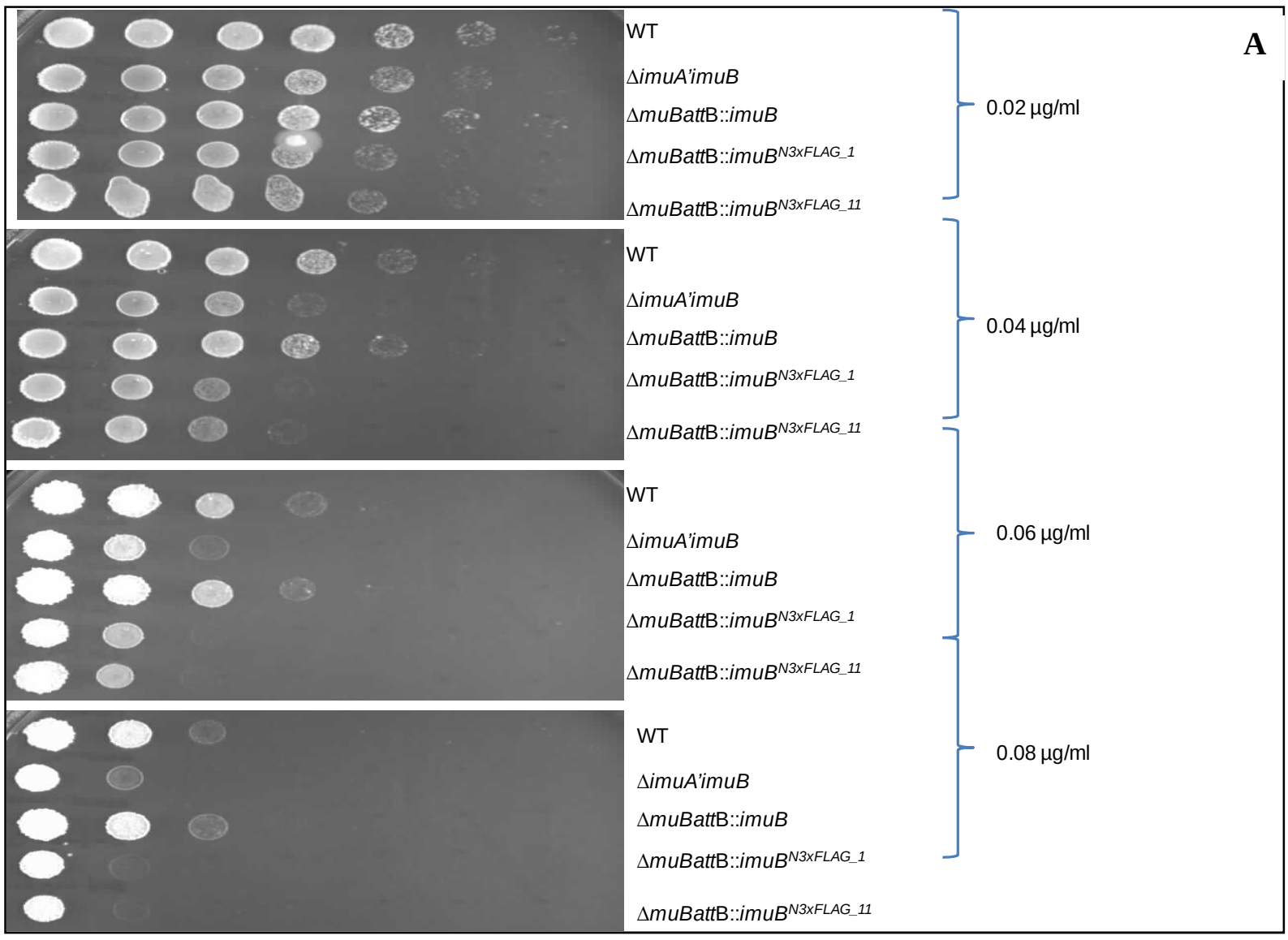


Figure 3.25: Schematic representation of the 3×FLAG tag introduced on the N-terminus of MsImuB. A 3×FLAG tag was introduced on the N-terminus of MsImuB in order to use it as *in vivo* bait to identify other interacting partners. The tagged allele was introduced by complementation at *att*B site on the chromosome carried on complementing vector pAINT into the single deletion mutant Δ*imuB*. The SOS box denotes the region in the operator where the LexA protein binds to control expression of down stream genes.

The effect of the 3×FLAG tag on MsImuB function was again evaluated in DNA damage tolerance and induced mutagenesis assays. Unfortunately, as with the C-terminal tag, introduction of the tag on the N-terminus also appeared to disrupt MsImuB function (Figure 3.26) as shown by the failure of the N-terminally tagged form of ImuB to complement the phenotype of the Δ*imuA'-imuB* mutant when carried on an integration vector with functional (full-length) *imuA'*.



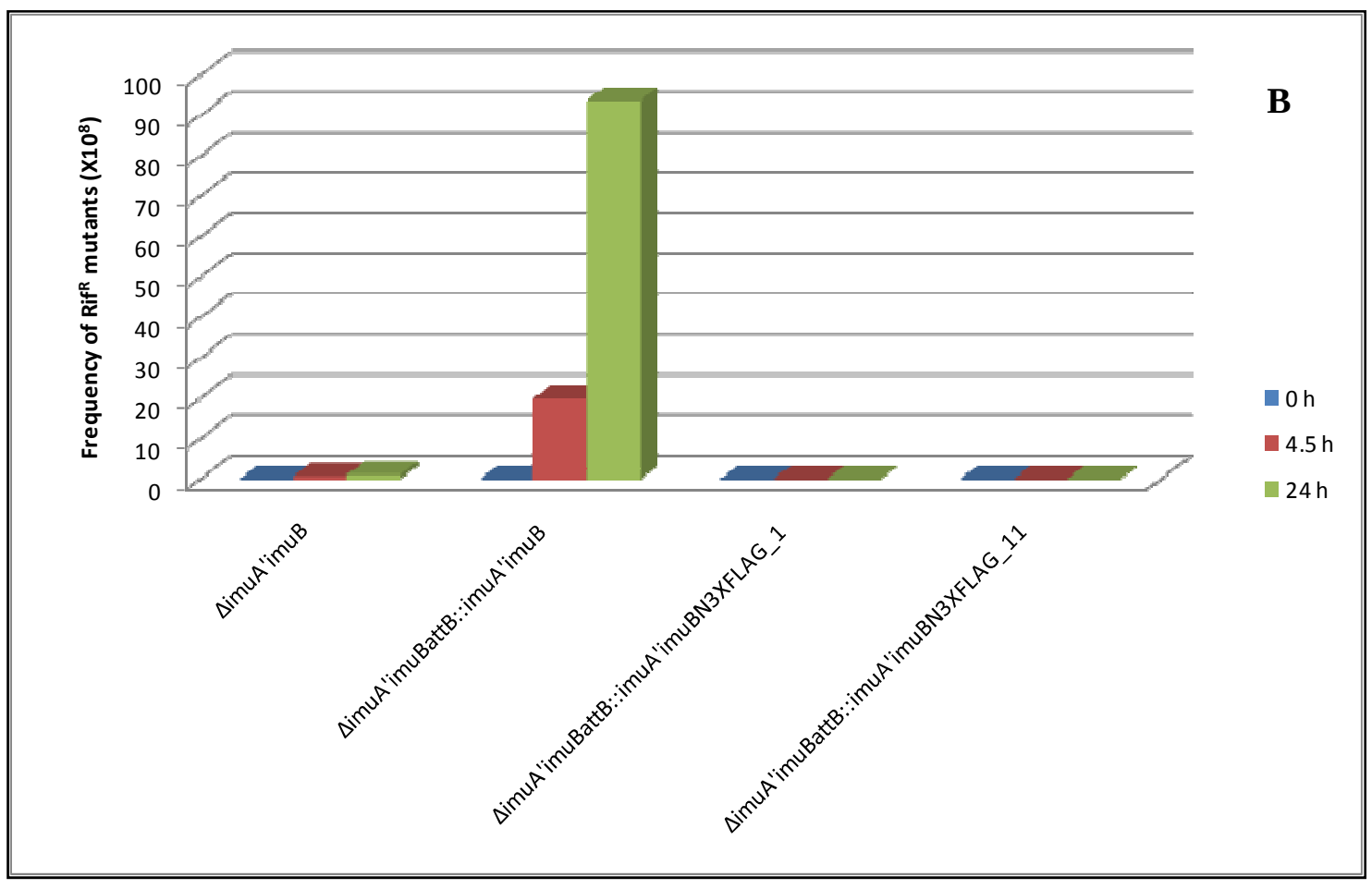


Figure 3.26: Introduction of a tag on the N-terminus of ImuB impaired protein function. (A) DNA damage tolerance assay and (**B**) UV-induced mutagenesis of the wild type mc^2155 (WT), the Δ*imuA'-imuB* deletion mutant, its complemented derivative, Δ*imuA'-imuB att*B::*imuA'-imuB* and two independent clones carrying the 3×FLAG tag on the N-terminus of MsImuB (Δ*imuA'imuB att*B::*imuA'imuB*$^{N3\times FLAG_1}$ and Δ*imuBatt*B::*imuA'imuB*$^{N3\times FLAG_11}$, respectively) inserted between the first and second codons of ImuB. Data are from a representative experiment performed in triplicate.

Therefore, adding a tag either on the N- or C-terminus of *M. smegmatis* ImuB renders the protein non-functional. Therefore, there was no option but to select another protein as bait in the pull-down assay.

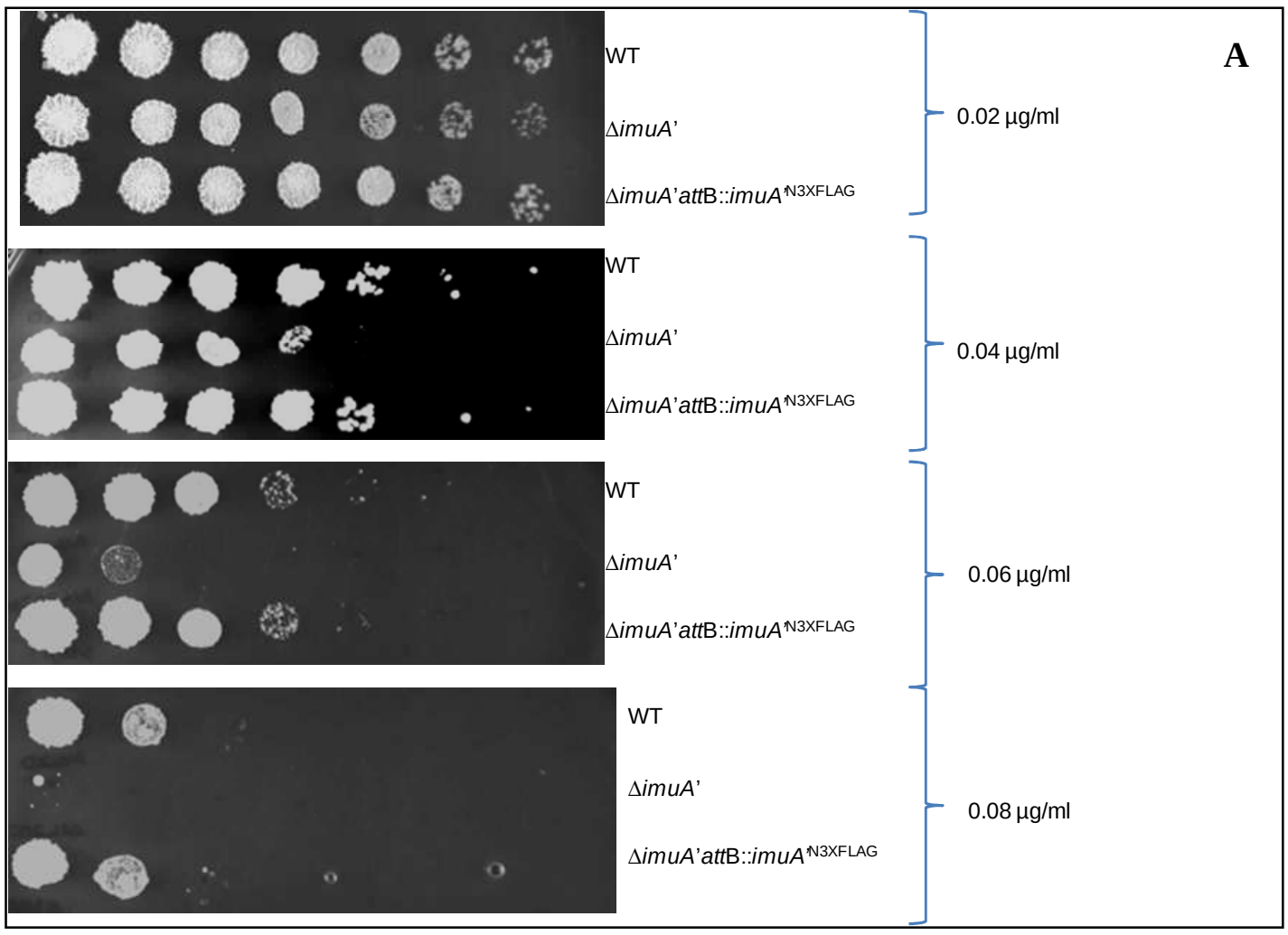
3.4.3 Introduction of a tag on the N-terminus of MsImuA'

MsImuA', the other member of the split *imuA'-imuB/dnaE2* cassette in *M. smegmatis*, was tagged with a 3× FLAG tag preceded by a start codon on the N-terminus. The tag was introduced on the N-terminus because the Y2H data had implicated the C-terminus in the interaction of ImuA' with ImuB, and the tagging experiments described above had shown that a C-terminally tagged form of MsImuB was not functional. In order to limit the impact of the 3×FLAG tag on potential interactions occurring on the N-terminus, it was separated from the N-terminus by a spacer of 7 amino acids between the tag and the second codon. The construct carrying the 3×FLAG tag-encoding sequence on the 5' end of *imuA'* was introduced into the Δ*imuA'* mutant strain by integration at the *att*B site.



Figure 3.27: Schematic representation of the 3×FLAG tag introduced on the N-terminus of MsImuA'. The tagged allele was introduced by complementation at *att*B site on the chromosome carried on complementing vector pAINT into the single deletion mutant Δ*imuA'*. The SOS box denotes the region in the operator where the LexA protein binds to control expression of downstream genes.

The impact of the N-terminal tag on MsImuA' function was again evaluated in DNA damage tolerance and induced mutagenesis assays. In this case, introduction of the tag on the N-terminus of MsImuA' did not disrupt the function of the protein (Figure 3.28) as shown by the restoration of MMC sensitivity and UV-induced mutagenesis to wild type levels by N-terminal tagged MsImuA'. Incorporation of the tag as an in-frame fusion with ImuA' was confirmed by DNA sequencing.



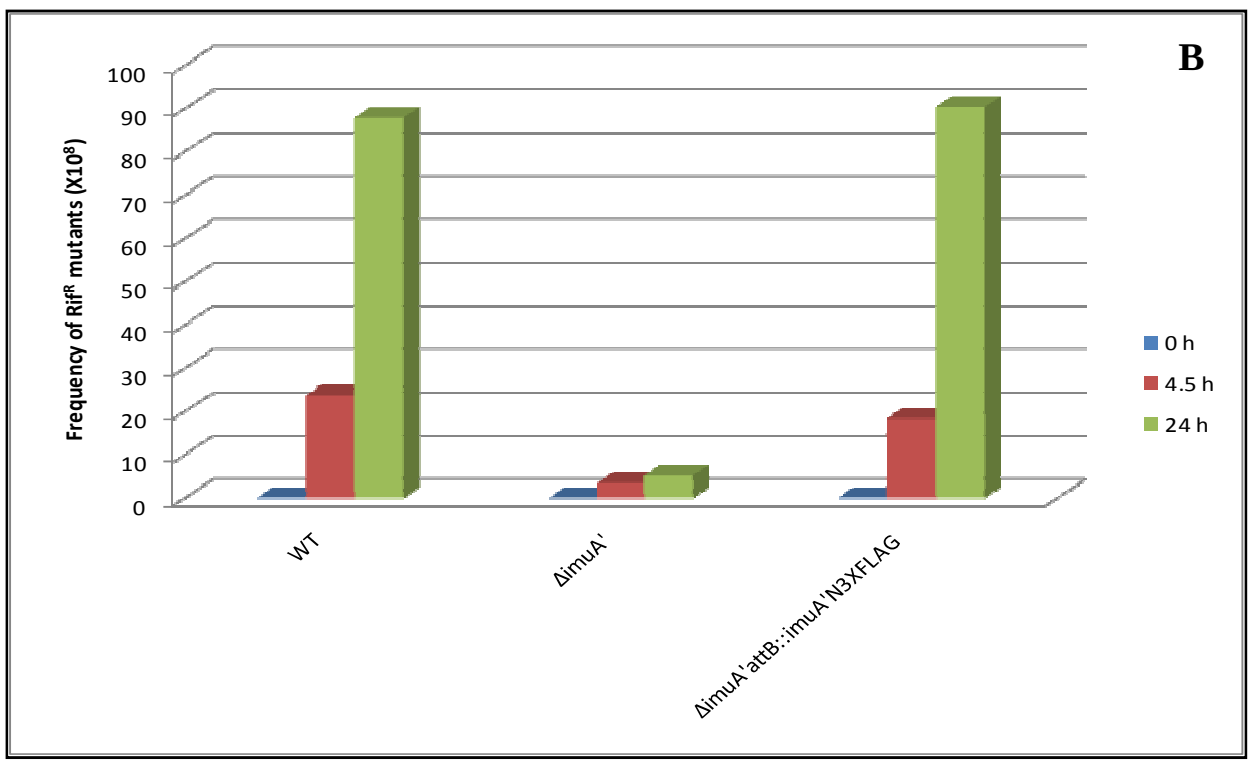


Figure 3.28: The function of ImuA' with an N-terminal 3× FLAG tag was maintained. (A) DNA damage tolerance assay and (**B**) UV-induced mutagenesis of the wild type mc^2155 (WT), single deletion mutant (Δ*imuA'*) and the strain carrying the tagged allele on the N-terminus Δ*imuA'att*B::*imuA'*$^{N3\times FLAG}$ preceded by a start codon. Data are from a representative experiment performed in triplicate.

3.5 Pull-down analysis of N-terminally ImuA' in *M. smegmatis*

Since the introduction of the tag did not appear to affect MsImuA' function, the strain expressing the N-terminally tagged protein was selected for pull-down assays with the aim of identifying other interacting partners of this protein. Prior to conducting the pull-down experiment, a Western blot analysis of the tagged ImuA' was performed to determine whether the tagged protein could be detected in a cell extract of *M. smegmatis*. The tagged strain was grown to mid-log phase and divided into two equal volumes: one of these was exposed to UV radiation to induce expression of genes, while the other was not exposed to UV and served as an untreated (non-induced) control. Both UV-irradiated and control samples were allowed to recover in normal media for 1 h, after which proteins were extracted, and separated by SDS-PAGE. Western blot analysis using M2 anti-FLAG antibody as probe was performed (Figure 3.29). The only protein detected in the western blot was the 49-kDa FLAG-BAP control (Figure 3.29, lane 7). A protein of the expected size of 34-kDa of the tagged ImuA' was not detected in any of the protein extracts from cultures of the various strains, with or without UV irradiation.

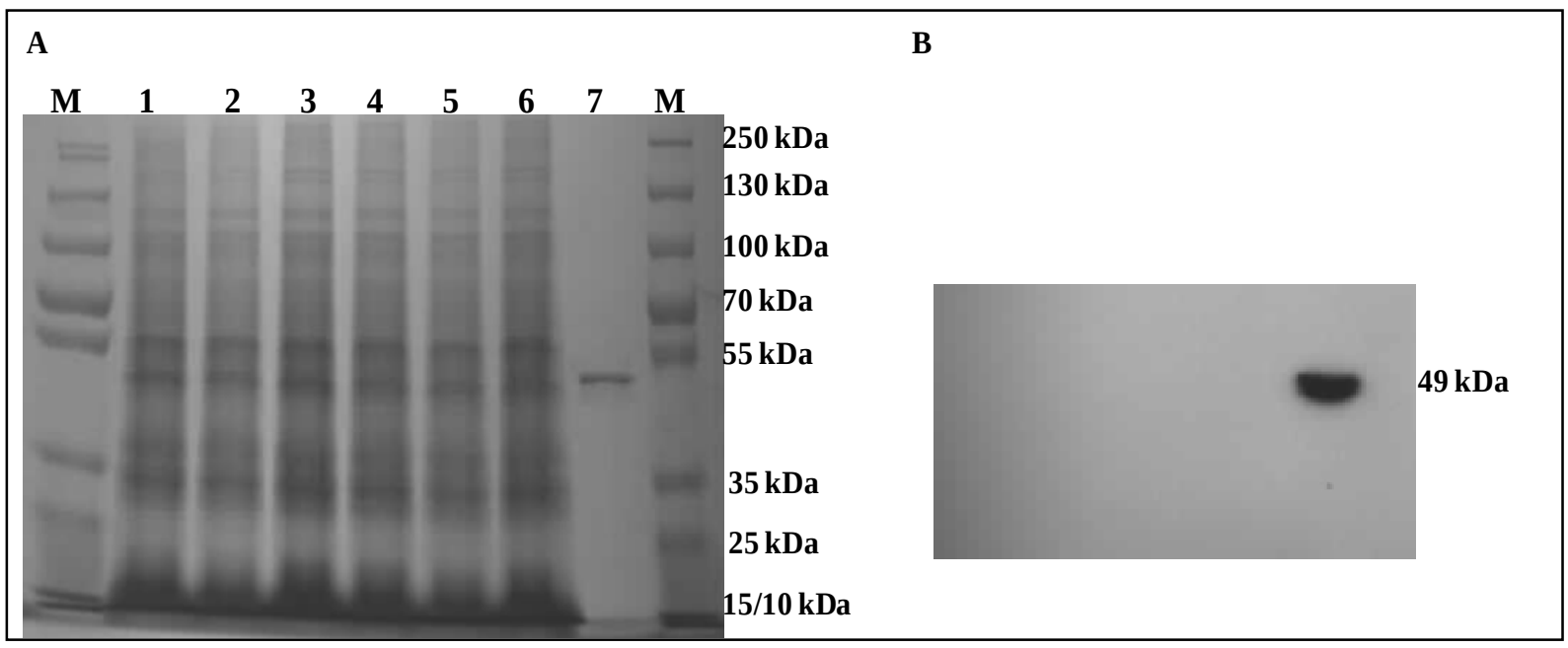


Figure 3.29: Tagged ImuA' was not detected by western blot analysis of whole-cell extracts of *M. smegmatis*. A. SDS_PAGE of whole-cell extracts of the various strains. **B.** Western blot analysis of the proteins transferred to nitrocellulose membrane and probed with M2 anti-FLAG antibody. Lane 1, WT, untreated control; lane 2, WT, UV-irradiated; lane 3, Δ*imuA'*, untreated control; lane 4, Δ*imuA'*, UV-irradiated; lane 5, Δ*imuA'att*B::*imuA'*$^{N3\times FLAG}$ untreated control; lane 6, Δ*imuA'att*B::*imuA'*$^{N3\times FLAG}$, UV-irradiated; lane 7, 49-kDa FLAG-BAP control. M denotes the molecular weight marker.

Failure to detect the protein by western blot could be attributed to the level of expression since *imuA'* was expressed from its native promoter. Previous work in the lab (Warner *et al.*, 2010) had established that transcript levels of *M. tuberculosis imuA'* are very low, even

following UV irradiation, in contrast to *imuB* which has a much higher basal level of expression (2% of that of *sigA*) and is very significantly induced following DNA damage (Figure 3.30).

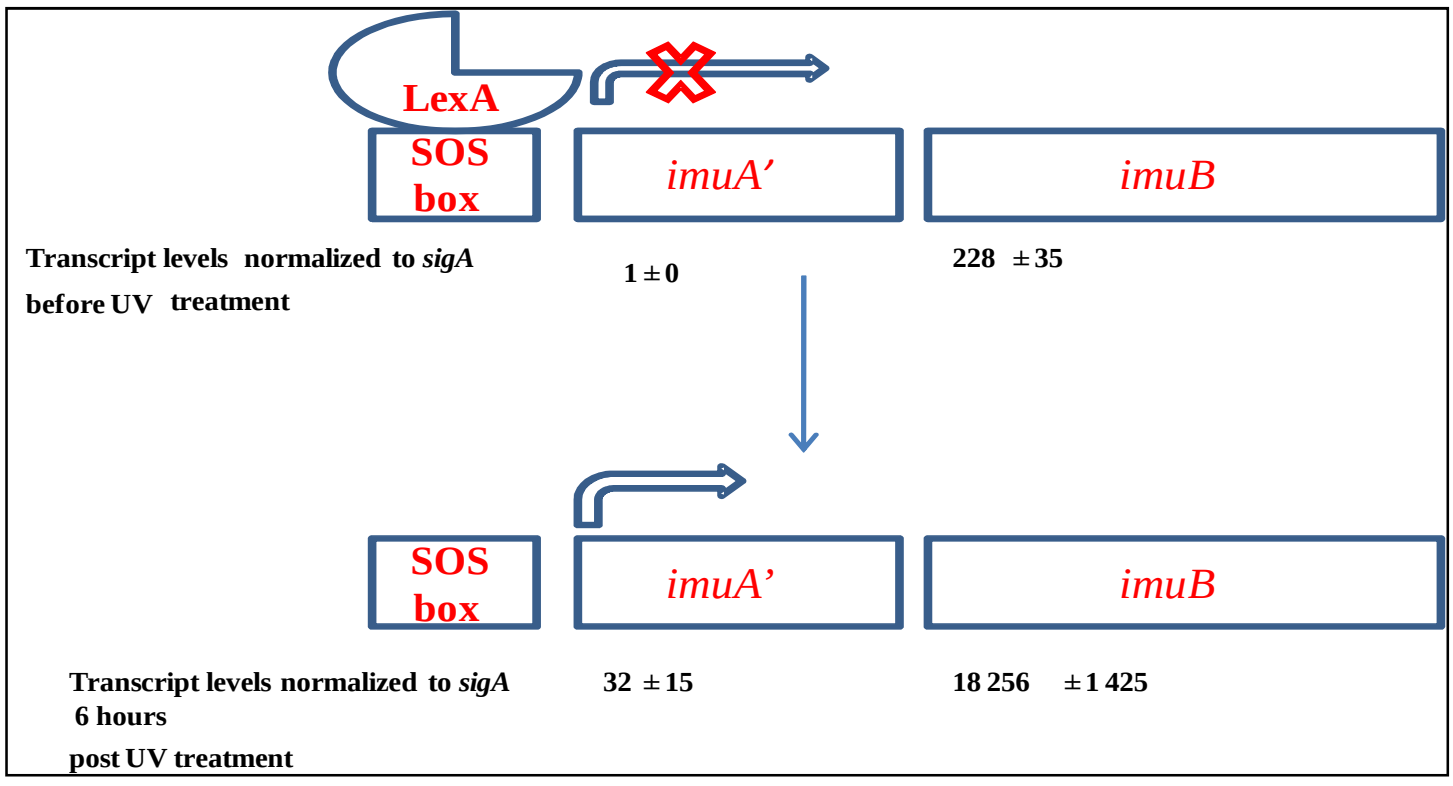


Figure 3.30: Schematic presentation of the transcript levels of ImuA' and ImuB pre- and post UV radiation in *M. tuberculosis*. The values below the *imuA'* and *imuB* genes represent normalized transcript levels of the corresponding genes in *M. tuberculosis*, and are based on the qRT-PCR analysis performed by Dr. B. Kana, as reported (Warner *et al.*, 2010). A value of "100" in this context corresponds to a transcript level that is 1% of that of *sigA*. Therefore, the level of *imuA'* transcript in *M. tuberculosis*, 6 hours after UV irradiation is only 0.3% of that of *sigA*.

The implication, based on the *M. tuberculosis* expression data, is that MsImuA' is present at very low levels in *M. smegmatis*, even following DNA damage induction, and could therefore not be detected by Western blot analysis. Adding to the limitation of detection is that the 3×FLAG sequence incorporated was based on the original FLAG tag sequence (DYKDDDDK), whereas the M2 anti-FLAG antibody has been optimized for the modified FLAG sequence (DYKDHD) (Hernan *et al.*, 2000). In spite of these limitations, a preliminary pull-down experiment was nonetheless attempted to ascertain whether tagged ImuA' might still allow the identification of its interacting proteins. In this experiment, wild type *M. smegmatis* and the Δ*imuA'attB*::*imuA'*$^{N3\times FLAG}$ mutant were grown to mid-log phase, split into equal volumes, and then either treated with UV to induce the SOS response, or left untreated as a control. Both induced and non-induced cultures were allowed to recover for 1 h. The cells were lysed and protein was extracted and quantified. Equivalent amounts of total protein (20 µg/µl) from each sample were loaded onto M2 anti-FLAG beads and allowed to bind. After several washes, bound protein was eluted by boiling the beads in the presence of

SDS-PAGE loading dye SDS-PAGE analysis suggested that all sample preparations were heavily contaminated with antibody, and no significant differences were observed between samples. Nonetheless, bands were excised from the lanes containing eluate from UV-irradiated samples of the wild type and tagged strains for Peptide Mass Fingerprinting. The whole lane of each strain was divided into five slices per strain making sure that the dominant bands were in different gel slices. The PMF analysis (performed at the Molecular and Biomedical Technologies Platform of the CSIR in Pretoria) did not yield any conclusive results owing to contamination with antibody, which obscured the data (data not shown). Unfortunately, owing to time limitations, this work could not be pursued further.

3.6 Proteomic analysis of *M. smegmatis* wild type, Δ*dnaE2* and Δ*dnaE2att*B::*dnaE2* strains after exposure to DNA damaging conditions.

At the inception of this study, DnaE2 was known to be responsible for induced mutagenesis while subsequent data indicate that DnaE2 functions with two proteins, ImuA' and ImuB, to form a mutagenic cassette (Boshoff *et al.*, 2003; Warner *et al.*, 2010). One of the cassette components, ImuB, was shown to be present under normal growth conditions and massively up-regulated at the transcript level when exposed to DNA damage in *M. tuberculosis* (Warner *et al.*, 2010), whereas with the other components- DnaE2 and ImuA'- are only moderately up-regulated. The inability to detect tagged ImuA' by Western blot analysis of whole-cell extracts (Section 3.4) prompted the idea to investigate the presence of the cassette components by mass spectroscopy (MS) in a label-free approach. Wild type, the Δ*dnaE2* deletion mutant and the Δ*dnaE2att*B::*dnaE2* complemented strain were used for this analysis to determine the ability to detect these proteins under the conditions tested. Again, the cells were grown to mid-log phase and split into two equal volumes, one of which was treated with UV and the other used as a control. Following recovery of 1 h, whole-cell fractions of both samples were prepared for in-gel tryptic digestion to run on LC-MS. Peptide counts of identified proteins were determined. Only proteins that have been shown to be transcriptionally up-regulated in response to DNA damage by genome-wide expression analysis (Boshoff *et al.*, 2003), are shown in Figure 3.34.

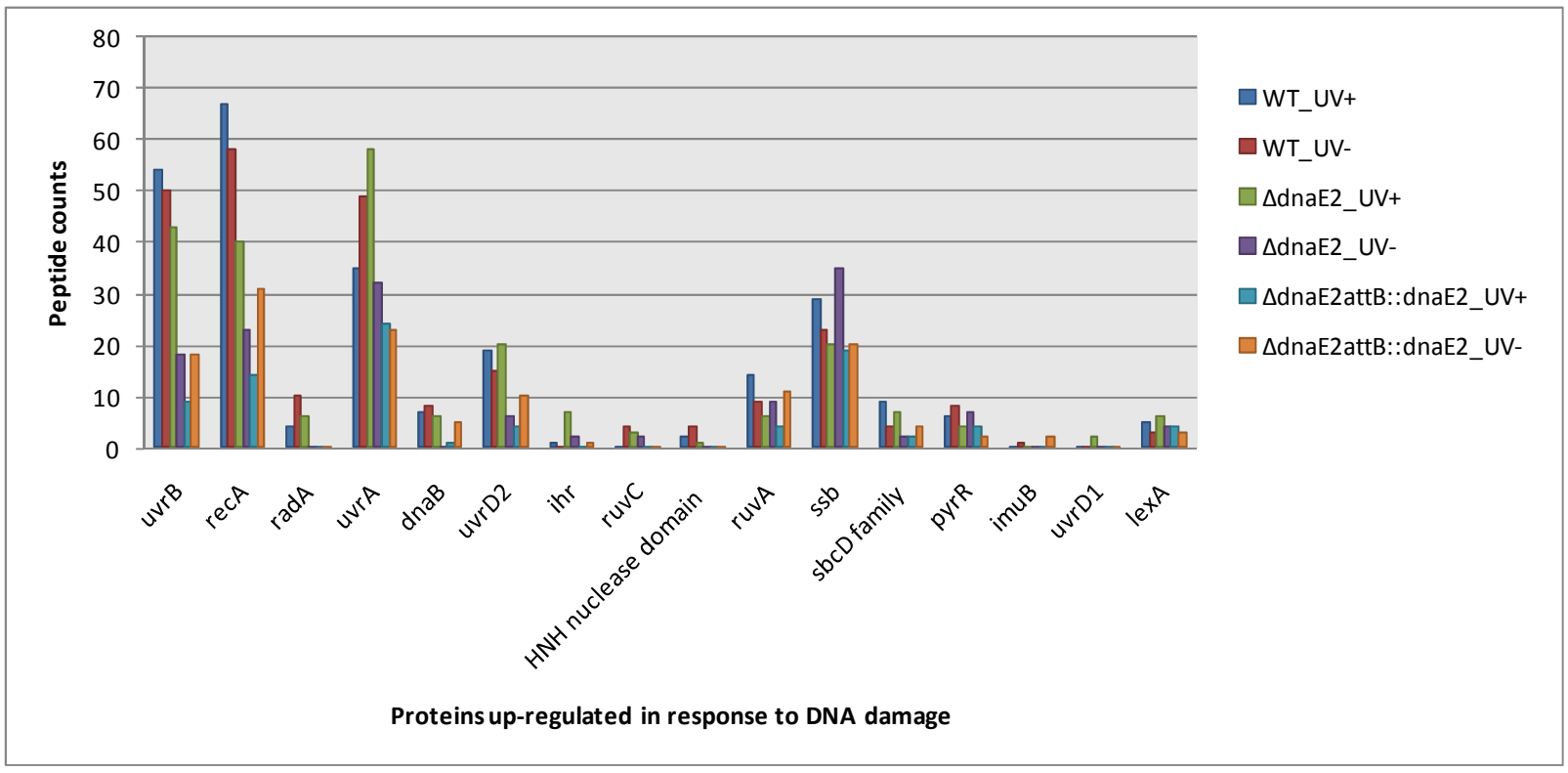


Figure 3.31: Identification of proteins known to be up-regulated in response to UV radiation. Proteins involved in nucleotide excision repair proteins (NER),UvrA, UvrB, UvrD1 and UvrD2; Homologous recombination (HR), RecA, RuvA, RuvC; RadA is known to be involved in DNA repair; replicative DNA helicase DnaB; single-stranded DNA binding protein ssb protein; possible DNA repair exonuclease belonging to sbcD family; pyrimidine biosynthesis regulatory protein PyrR; ImuB; a protein containing HNH nuclease domain and LexA were identified. UV + denotes cells treated and UV- denotes untreated cells. Only proteins that were previously shown to be up-regualated in response DNA damage are shown (Boshoff *et al.*, 2003).

The data in Figure 3.31 show that, of the proteins identified, the majority are involved in NER and HR DNA repair pathway processes. The Ssb protein and helicases are required in replication to unwind the double-stranded DNA and to bind unwound (single-stranded) DNA during repair and replication. RecA and LexA were identified in this analysis. These proteins are the key regulators of the mycobacterial SOS response. ImuB, another component of the SOS regulon, was also identified. DnaE2 peptides were not detected in the wild type and complemented strains, and ImuA' peptides were not detected in any of the strains. In contrast, ImuB was detected in all strains, with and without UV irradiation. These observations are consistent with expression data in *M. tuberculosis*, which showed that *imuB* is expressed at a comparatively high basal level (Warner *et al.*, 2010). Interestingly, however, the DNA repair proteins that were identified in this experiment were not significantly up-regulated in response to the imposed damage under the conditions tested.

3.7 Heterologous expression of *M. tuberculosis* ImuA' and ImuB in *E. coli*

ImuB and ImuA' are annotated as hypothetical proteins, Rv3394c and Rv3395c, respectively. However, the combination of Y2H interaction data, as well as genetic and microbiological assays, had established a role for these proteins in the DNA damage response. This suggested that structural analysis of these proteins might further aid in elucidating their role in DnaE2-dependent DNA damage tolerance.

3.7.1 Small scale expression of ImuA' and ImuB

Constructs were generated to enable the expression of the *M. tuberculosis* proteins as *E. coli* recombinants. Genes encoding ImuB and ImuA' were cloned in the pMALc2 expression vector for N-terminal tagging with a maltose binding protein (MBP). The fusion proteins, MBP-ImuB and MBP-ImuA' were expressed in *Escherichia coli* BL21 (DE3) by induction with IPTG at 30ºC. Fusion proteins were identified as distinct bands corresponding to the theoretically determined size of the fusion proteins on SDS-PAGE gels. Further to that, the identified bands were confirmed by probing the fusion proteins with anti-MBP primary antibody, and observing bands by chemiluminescence released through the peroxidise reaction with the conjugated secondary antibody. According to theses analyses, both proteins were present in the soluble and insoluble fractions of whole-cell extracts.

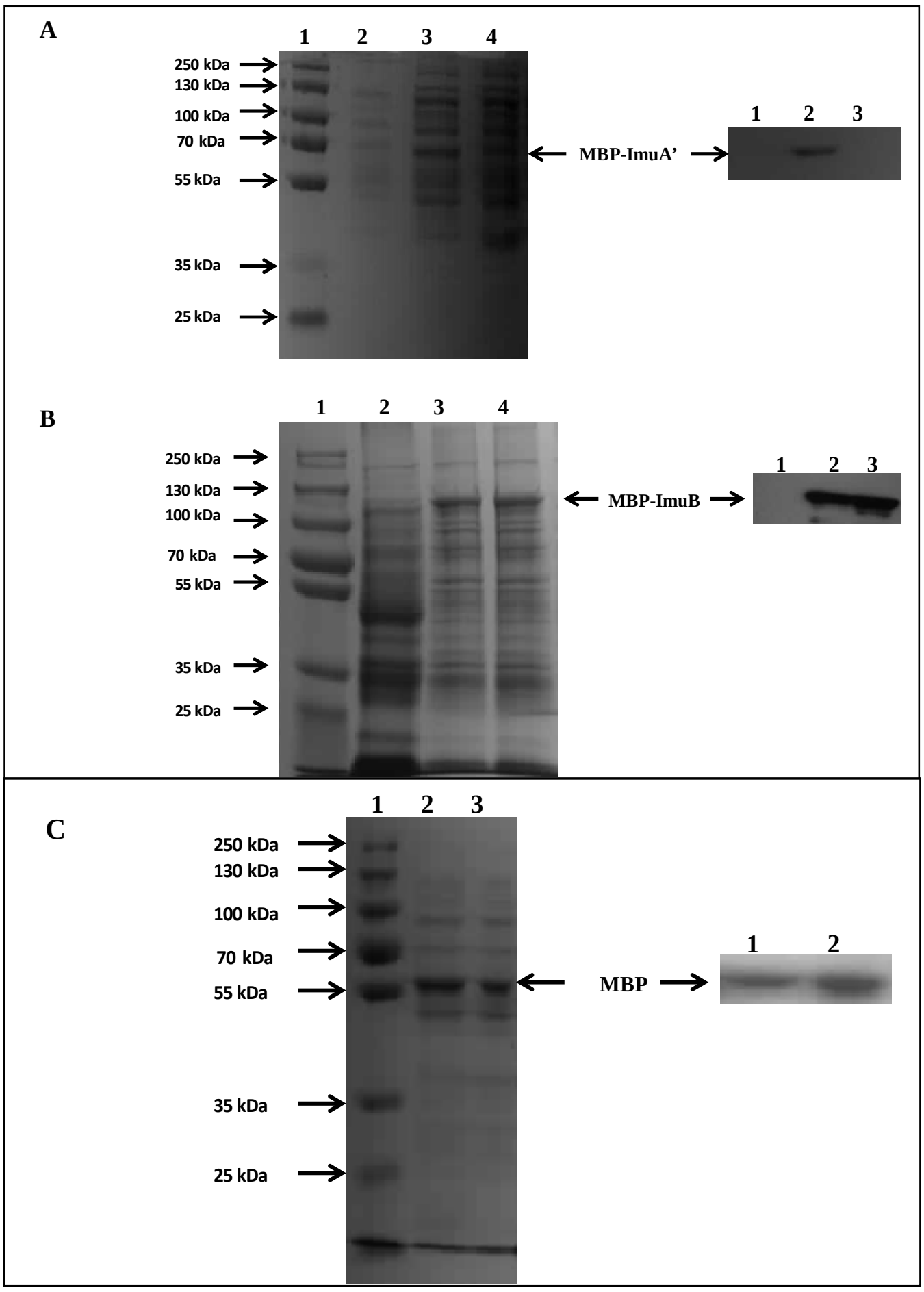


Figure 3.32: Small-scale analysis of MBP-ImuB and MBP-ImuA' indicates expression in both insoluble and insoluble fractions. A. MBP-ImuA' expression in the insoluble (lane 3) and soluble fraction (lane 4). **B.** MBP-ImuB expression in the insoluble (lane 3) and soluble (lane 4) fraction (lane 3). **C.** Protein extract from the empty vector control showing expression of the MBP tag in the insoluble (lane 3) and soluble fraction (lane 4). In all gels, lane 2 is the uninduced control. On the right of each gel are corresponding Western blotting analyses of the fusion proteins.

The predicted molecular weights of the MBP, MBP-ImuA' and MBP-ImuB were 52 kDa, 83 kDa and 115 kDa, respectively. Bands on SDS-PAGE gel corresponding to the predicted

sizes of the fusion proteins were excised for identification by PMF. In this method proteins separated on SDS-PAGE gels are trypsin digested in the gel and the peptide fragments generated are then identified by determination of molecular masses which are used to search for protein identity in databases (Thiede *et al*., 2005). According to this analysis, MBP-ImuB was confirmed with 95% confidence, but MBP-ImuA' was not confirmed.

3.7.2 Large scale protein expression and purification of MBP-ImuB and MBP-ImuA'

Based on the results from the small-scale experiment, large-scale expression was then carried out for both recombinants even though the PMF analysis had only confirmed the MBP-ImuB fusion protein. The fusion proteins were expressed as in described in Section 3.7.1. Induced cells were lysed and the corresponding soluble fractions were loaded onto an amylose column for affinity purification chromatography. Thirty two 0.5-mL fractions were collected by elution with 10 mM maltose in column buffer. Two µl aliquots of each fraction used to determine the protein concentration by measuring absorbance at 280 nm. An elution profile was plotted from the determined concentration.

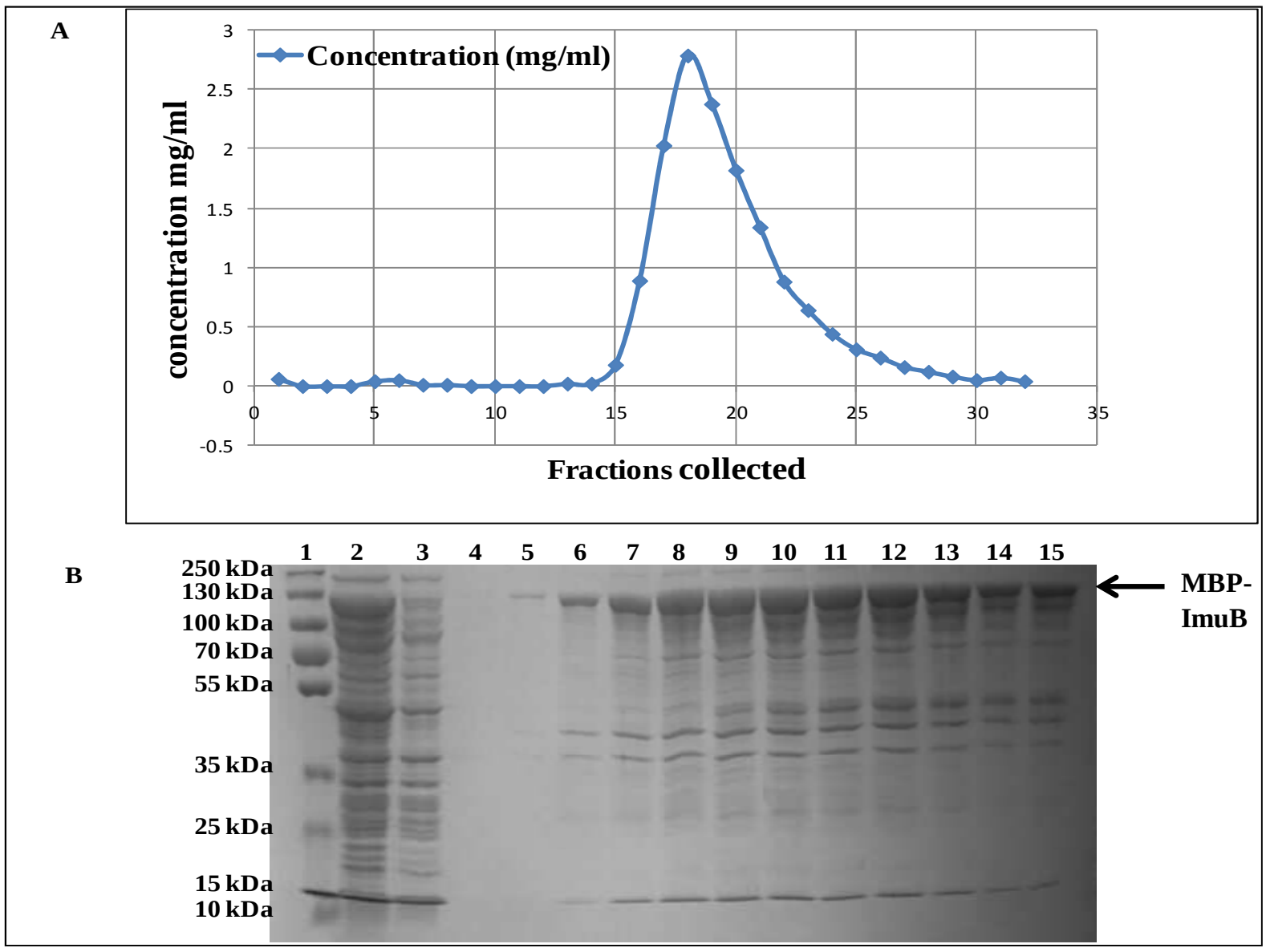


Figure 3.33: Affinity purification of MBP-ImuB. A. Elution profile of MBP-ImuB from the amylose column and **B**. SDS-PAGE gel of the soluble fractions containing the highest concentration of protein. Lane 1, molecular weight marker; lane 2, soluble fraction from lysate of induced cells; lane 3, flow-through; and lanes 4-15, eluted fractions 14-24

Figure 3.33 indicates that the fusion protein can bind to the amylose resin; however, many contaminating proteins co-eluted with the fusion protein. Even those fractions that were enriched for the MBP-ImuB (e.g., in lane 5 of Figure 3.33) were relatively impure. Putative MBP-ImuA' recombinant protein purified from the soluble fraction of induced cells by affinity chromatography was similarly associated with an excess of contaminating protein (Figure 3.34). In this case, a dominant band of lower molecular weight (~50 kDa) was present in both soluble and flow-through fractions as well as the eluted fractions. Based on its size, and apparent binding to the amylose column, this lower band was thought to be the MBP protein. This hypothesis was tested by excising this protein from the gel and subjecting it to PMF analysis. This analysis showed that the ~50 kDa was identified as MBP indicating that there might be partial degradation of the fusion protein.

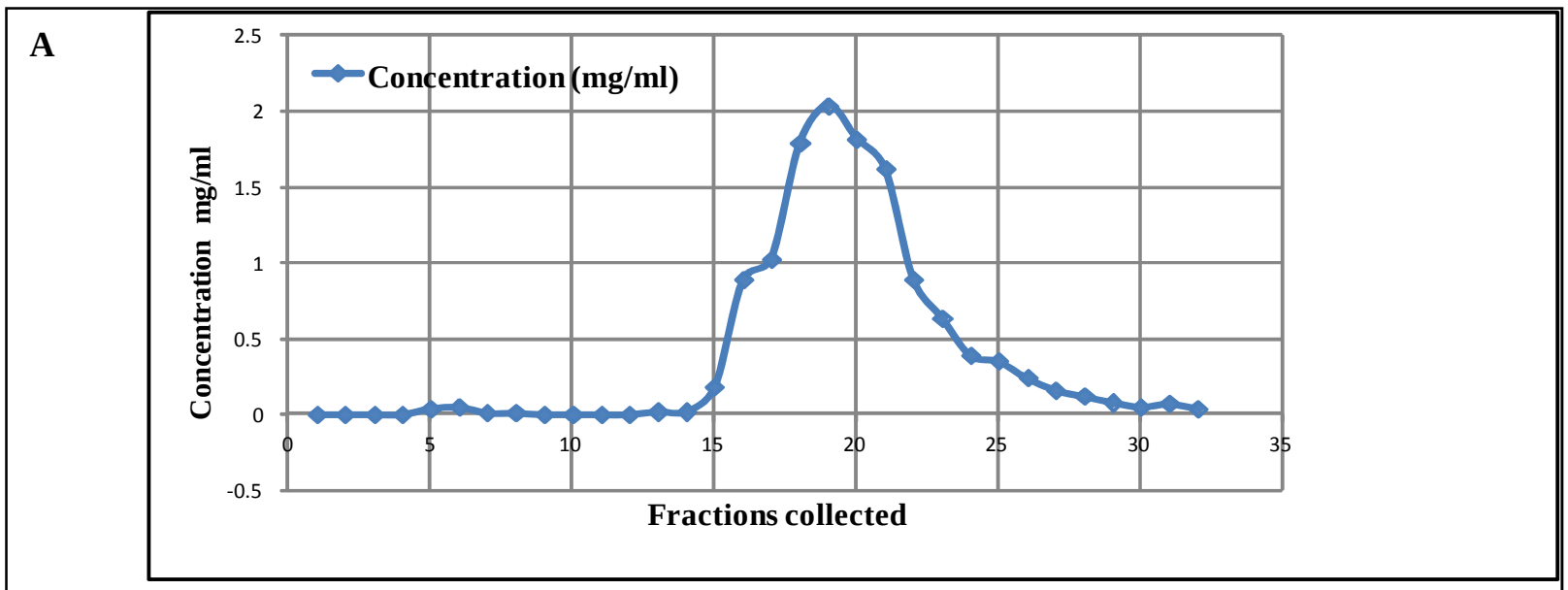


Figure 3.34: Affinity purification of MBP-ImuA’. A. Elution profile of MBP-ImuA’ from the amylose column and **B**. SDS-PAGE gel of the soluble fractions containing the highest concentration of protein. Lane 1 is molecular weight marker, lane 2 is the soluble fraction, lane 3 is the flow through fraction and lane 4-15 is eluted fractions 14-24.

Since both MBP-ImuA’ and MBP-ImuB appear to elute with contaminating proteins, further purification was attempted using Q-Sepharose ion exchange chromatography was attempted. The peak fractions were and dialysed against a column buffer containing 50 mM NaCl, after which the dialysed protein was concentrated using Amicon ultra centrifugal filters with a molecular weight cut-off (MWCO) of 50kDa In the case of MBP-ImuB, a small-scale experiment was done to establish the approximate concentration of NaCl required to elute the protein from Q-Sepharose. Based on this analysis, a 0.5 – 1 M NaCl gradient was used to elute the Q-Sepharose column.

The dialysed fraction was loaded onto Q-Sepharose column pre-equilibrated with the buffer used for dialysis. A gradient of NaCl from 500 mM to 1M concentration was used for elution after washing the bound protein. Thirty five fractions of 500 µl of eluate were collected and 2 µl used to determine the protein concentration by absorbance at 280 nm.

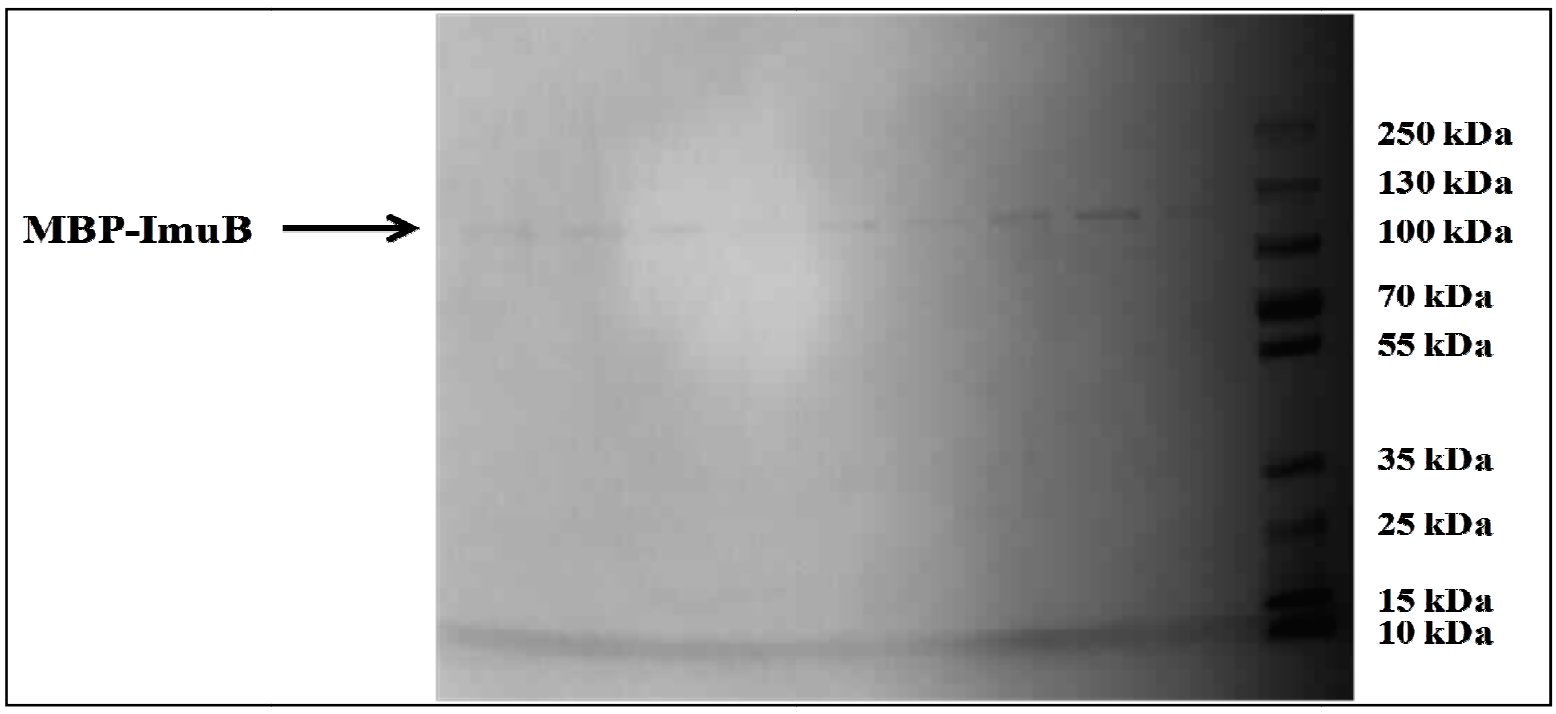


Figure 3.35: Representation of the SDS-PAGE gel with MBP-ImuB eluted from Q-Sepharose column. Thirty five 0.5-ml fractions collected and 20 µl of fractions 1-9 from the marker were run).

The identification of IMBP-ImuB in many fractions from the Q-Sepharose column, suggesting that the NaCl gradient used to elute the column was probably too shallow (Figure 3.35). Nonetheless, this experiment suggested that ion exchange chromatography using a steeper elution gradient potentially could be used to purify MBP-ImuB. Other systems that were attempted were pGEX3X which allowed N-terminal fusion with glutathione S-transferase (GST), pET15b and the protein was in the insoluble fraction which cannot be taken forward for crystallography. One way of overcoming the challenges involved in purifying proteins with disordered regions is to remove the disordered region. The C-terminal region of ImuB extending beyond the β-clamp binding motif was removed and this truncated version of ImuB expressed in the pMalc2 being the system that gave some soluble protein proved to also be unsuccessful as not visible band was observed on SDS-PAGE (Data not shown).

Chapter 4

4. Discussion

4.1 The DnaE2-dependent mycobacterial mutasome shares functional similarities to Pol V

The data presented here have afforded some insight into the DnaE2-dependent mutagenic pathway in mycobacteria. In particular, these analyses have established that DnaE2 function is dependent on the accessory factors, ImuB and ImuA', both of which are essential for DNA damage survival and induced mutagenesis (Warner *et al.*, 2010). This is consistent with the observation that DnaE2 generally occurs together with homologs of ImuB in organisms lacking Pol V (Abella *et al.*, 2004; Erill *et al.*, 2006; Galhardo *et al.*, 2005). In mycobacteria, ImuB and ImuA' are in an operon and DnaE2 is located 24.7 kb upstream of this operon of which their expression is regulated by LexA indicating that they all form part of the SOS response. These findings therefore suggest a split mutagenic cassette-*dnaE2/imuA'imuB*-operating in mycobacteria that may have diverged from other organisms (Warner *et al.*, 2010). In many respects, these data presented here and previously provide compelling suggestion of functional analogy with *E. coli* DNA polymerase V (Boshoff *et al.*, 2003). All three genes of the split *imuA*'-*imuB*/*dnaE2* cassette are induced as part of the mycobacterial SOS response (Boshoff *et al.*, 2003), indicating that the activity of the "mycobacterial mutasome" might be limited to circumstances of extreme damage.

ImuB shares considerable structural similarity with Y-family DNA polymerases, but possesses a C-terminal extension containing disordered regions that are characteristic of unstructured protein domains (Wright & Dyson, 1999). In addition, unlike "true" Y-family polymerases, ImuB lacks the catalytic residues required for polymerase function; instead, consistent with the involvement of unstructured regions in protein-protein interactions (Dyson & Wright, 2005; Wright & Dyson, 1999), Y2H analyses imply that ImuB plays a key role – via its extended C terminal domain – in enabling DnaE2 to access the replication fork for TLS. Interestingly, sequence alignment shows that ImuB possess the conserved residues, C66 and P67, identified recently for complex formation with UmuD and RecA (Cafarelli *et al.*, 2013). These residues are located in the N-terminal region that resembles structural similarity to Y family polymerase (Ling *et al.*, 2001). Consistent with the identification of ImuB from *M. tuberculosis* as a member of the DinB3-type protein, these residues are conserved amongst DinB-like proteins (Cafarelli *et al.*, 2013).

The observation that ImuB has the ability to self-interact is reminiscent of the Pol V subunit, UmuD, which interacts with UmuC to constitute a functional Pol V. UmuD plays a role in protecting the cell from the deleterious effects of the error-prone DNA damage response pathway, a function that is genetically distinct from its role in SOS mutagenesis (Napolitano *et al.*, 2000; Pages & Fuchs, 2002). $UmuD_2$, together with UmuC, may act as a primitive DNA damage checkpoint, as they specifically inhibit DNA replication without affecting transcription or translation when present at high levels in cells (Murli *et al.*, 2000). UmuD and UmuC slow the resumption of DNA replication after UV irradiation and therefore, the presence of both proteins acts in a non-catalytic manner to delay SOS mutagenesis and give accurate pathways such as NER time to proceed (Opperman *et al.*, 1996; Opperman *et al.*, 1999). A comparatively high basal level of *imuB* transcript was previously observed during growth of *M. tuberculosis* under standard *in vitro* conditions (Warner *et al.*, 2010). The massive induction of *imuB* by DNA damage together with the observation that it has the ability to self-interact suggests that ImuB might have a similar function to $UmuD_2$, which is processed to $UmuD'_2$ to constitute an active Pol V to fully engage in SOS DNA repair. The finding that ImuB protein was not up-regulated in UV-irradiated *M. smegmatis* cells that were recovered for an hour post-irradiation before protein extraction either suggests that the DNA-damaging conditions used were not sufficient to induce ImuB or that the damage was repaired during the recovery period and the cells were in the process of recovering at the time of sampling. The latter explanation is consistent with the very tight regulation of TLS polymerase activity observed in *E. coli*, where the Pol V components are rapidly degraded by ClpXP when no other mechanisms of survival are available, thus limiting the potential to introduce mutations as that would compromise the integrity of the genome (Patel *et al.*, 2010).

To further extend this analogy, UmuD has been shown to share structural characteristics with intrinsically disordered proteins (IDPs), which have significantly less secondary or tertiary structure *in vitro* than other proteins (Dyson & Wright, 2005; Ollivierre *et al.*, 2010). The IDPs often have important biological roles in regulation as they assume precise structures upon interaction with their binding partners (Dyson & Wright, 2005). Structural analysis of $UmuD'_2$ has revealed that it possesses extended shorter arms that are disordered from the unbound C-terminal globular domain (Ferentz *et al.*, 2001; Peat *et al.*, 1996). The data presented here further suggest that ImuB could be a functional analog of UmuD because of its apparently disordered C-terminal extension which is involved in key protein-protein

interactions with other components of the mutasome. Therefore, ImuB might play a role in regulating the commitment step to fully engage SOS DNA repair through interaction with other proteins involved in this pathway via its C-terminal region.

If DNA damage persists, the $UmuD_2$ interacts with the RecA nucleoprotein that stimulates processing to remove the N-terminal domain of UmuD components to UmuD', signaling commitment to SOS mediated repair by Pol V. UmuD has a DNA-binding domain as well as a LexA-like domain that undergoes autolytic cleavage in the presence of RecA and ssDNA. During this process, LexA also undergoes autodigestion de-repressing the expression of SOS regulated proteins committing to full-on SOS response. The data presented in this study demonstrated that ImuB is able to self-interact and also showed that the introduction of a tag at the N-terminus of ImuB disrupted its function in DNA damage tolerance and induced mutagenesis. Therefore, a key question that remains to be addressed is whether ImuB undergoes processing in mycobacteria in order to constitute a functional mutasome.

ImuA' shares similarity with *E. coli* RecA specifically in the N-terminal domains that contain the two DNA binding sites and the ATP binding region, but this protein lacks the characteristic C-terminal RecA domain, and instead, has a distinct C-terminus. ImuA' also lacks the canonical, RecA-like ATP binding motif. ImuA' also differs significantly from RecA in the DNA binding loops, L1 and L2 (Figure 3.3.) suggesting that ImuA' is unable to bind DNA in a RecA manner. However, present in the predicted M domain of ImuA' is a region spanning from residues Arg^{133} to Arg^{148} in *M. tuberculosis* ImuA' that contains a large number of positively charged residues and is predicted to form α-helix and might be a candidate motif for DNA binding (Warner *et al.*, 2010). The results presented here also confirm that the C-terminal region of ImuA' is crucial for interaction with ImuB for induced mutagenesis, and thus serves some kind of regulatory role analogous to the C-terminus of RecA, which serves as autoregulatory role, and when deleted, enhances RecA function (Cox, 2007).

4.2 Distinguishing features of the mycobacterial mutasome

ImuB was originally identified as founder member of the DinB3 subfamily of Y family polymerases, on the basis of a defining-clamp binding motif in the C-terminal region (Dalrymple *et al.*, 2003). This, together with the absence of a β-binding motif in DnaE2, suggested that ImuB has the potential to serve as an adaptor protein allowing DnaE2 to access the replication fork. The experimental data presented here are consistent with the

identity of the β-binding motif as ^{354}QLPLWG^{359} as the Q354A mutation eliminated the interaction of ImuB with the β-clamp in the Y2H system. Interestingly, however, work carried out by Dr. Warner showed that the same mutation in *M. smegmatis* ImuB had no impact on the function of the mutasome {Warner, 2010 #14; in his parallel study, mutation of five of the six residues was necessary in order to cripple the function of the mutasome. This discrepancy is most likely due to the fact that the Y2H system only provides limited information on the interaction between two proteins, which in this case, form part of a larger protein-protein interaction network in the mycobacterial cell.

ImuB comprises an obvious, Y family-like N-terminal region, and a C-terminal region that has characteristics of unstructured proteins which are often involved in protein-protein interactions. The disordered C-terminal extension of ImuB suggested by homology modeling is consistent with the proposed role of ImuB as a "hub" protein. This type of structural arrangement (defined N-terminus and undefined C-terminus) is not unique, and has been observed in other mycobacterial proteins. For example, prokaryotic ubiquitin-like protein (Pup) has been shown to interact with mycobacterium proteosomal ATPase (Mpa) via the C-terminal half of Pup which has a disordered region (Liao *et al.*, 2009).

The data further suggest that *M. tuberculosis's* response to DNA damage is distinct in that DnaE1 is also able to interact with ImuB, although this replicative polymerase has a β-binding motif and has been shown to interact with the β-clamp (Kana *et al.*, 2010). Warner *et al.* (Warner *et al.*, 2010) and O'Sullivan *et al.* (O'Sullivan *et al.*, 2008) have shown an induction of DnaE1 after exposure to UV radiation and treatment to ciprofloxacin, respectively. The interaction between DnaE1 and ImuB is surprising. One possible explanation for this observation could simply be that the disordered region of ImuB is highly promiscuous in terms of protein-protein interaction, and that the inferred interaction with ImuB is a false positive result. However, if there is a genuine interaction between DnaE1 and ImuB, this may facilitate polymerase switching to occur after TLS has taken place. There is also emerging evidence of the involvement of replicative polymerases in mutagenic processes especially the involvement of eukaryotic high fidelity replicase Pol δ. Pol δ's involvement in mutagenic processes was largely shown by sharing of subunits, Pol 31 and Pol32, which are essential for TLS performed by Pol ζ (Johnson *et al.*, 2012). This could be explained by a recent study done in Drosophila that showed show that replicative polymerases and TLS polymerases compete in double strand break repair (Kane *et al.*, 2012). In this study, the

authors showed that TLS polymerases can function during the initial synthesis stage of homologous recombination repair and that they compete with Pol δ during the repair process, and further, that Y-family polymerase, Rev1, acts to coordinate the initial recruitment of TLS polymerase η, thus prevent the replicative polymerase from acting during early repair synthesis (Kane *et al.*, 2012).

4.3 Insight into the DnaE2 mutagenic pathway

Having shown that DnaE1 also interacts with ImuB, it was plausible that other proteins may be involved in this pathway. Y2H was a limited technique as it allows the assessment of binary interactions which have been defined by the researcher, and so are biased. In order to confirm the interactions established by Y2H and to identify other interacting partners, a pull-down assay was attempted. The preferred bait protein for use in the pull-down assay was ImuB. The gene encoding MsImuB was therefore tagged on the C-terminus with a tandem affinity tag. However, it was discovered that the introduction of the tag disrupted the function of the protein, as confirmed by the damage tolerance and induced mutagenesis phenotype of a mutant strain of *M. smegmatis* carrying C-terminally tagged ImuB. This observation accentuated the importance of the C-terminus of ImuB and suggested that the tag interfered with one or more interactions between ImuB and other protein/s that are essential for function of the mutasome. Switching the tag to the N-terminus also resulted in a loss of ImuB function. The reason for this observation is unclear, but could include interference with processes such as post-translational modification, analogous to that observed for UmuD. Studies have shown that there is competition between the RecA/ssDNA-mediated and Lon-mediated degradation of UmuD products (Gonzalez *et al.*, 1998; Patel *et al.*, 2010). These two degradation processes play a vital role because the RecA/ssDNA degradation provides the processed UmuD for TLS whereas the Lon/ClpXP degradation degrades the protein when not required by the cell, thus keeping the level of mutagenically active UmuD' protein at a minimum. Both degradation mechanisms occur on the N-terminus of UmuD removing 24 amino acids. Introduction of a phleyomycin resistance protein (PRP) at the N-terminus of UmuD indicated that the Lon degradation site is within the N-terminal region of the protein. In this study, the wild-type PRP itself was stable when assayed in a wild-type background, but when fused to UmuD, the PRP-UmuD fusion protein was unstable as it was rapidly degraded (Gonzalez *et al.*, 1998). Taken together with the idea that ImuB has a function similar to UmuD, it is tempting to speculate that introduction of a tag on the N-terminus of ImuB may have disrupted posttranslational processing– if it occurs. Therefore, a high priority

for future work will be to test the hypothesis that ImuB undergoes post-translational modification.

As an alternate strategy for the pull-down experiment, ImuA' was tagged instead. In this case, N-terminal tagging of ImuA' did not affect its function. However, the tagged protein could not be detected by Western blotting using anti-FLAG antibody in extracts of *M. smegmatis* cells that had been exposed to UV irradiation. This problem is probably due to the low level of expression of ImuA' even after damage induction (Warner *et al*., 2010), which would reduce the sensitivity of the pull-down assay. Unfortunately, the one pull-down experiment that was attempted using this protein was not successful, with heavy contamination with antibody being one technical problem that could potentially addressed using a more gentle elution method (*e.g*., with free FLAG petide), and low sensitivity being another.

A further challenge with regards to the pull-down assay is the time at which cells can be harvested for protein extraction as studies in *E. coli* have shown that the SOS response is dynamic at a molecular level. Two studies have looked at promoter activity of LexA-repressed promoters as a measure of the dynamics of the SOS response (Friedman *et al.*, 2005; Shimoni *et al.*, 2009). The first monitored promoter activity of *recA*, *lexA* and *umuDC* using a low-copy reporter plasmid in which the promoters were fused to *gfp* gene. On the basis that the rate of accumulation of GFP in a cell is proportional to the rate of transcript production from the promoter, the authors found that the response to DNA damage in individual cells is highly structured: the damaged cell accurately times and synchronizes the repair process, it allows temporal activation of various promoters to efficiently repair the damage by modulation of the level LexA repressor. This therefore limits the response level, thereby avoiding damage response that is too high at early stages (Friedman *et al.*, 2005).

The second study also looked at promoter activity of *recA* and concluded that predicting the dynamics of regulatory systems suggests a stochastic rather than deterministic process (Shimoni *et al.*, 2009). Harvesting the cells at 1 h post DNA damage was taken from a study that identified the pattern of RecA expression in mycobacteria as a measure of induction of the SOS response {Papavinasasundaram, 2001 #124}. In this study the authors have shown that in *M. smegmatis* there are detectacle levels of RecA protein at 1 h post treatment with MMC. Both these studies show that the SOS response is dynamic and complex such that the

time at which cells should be harvested for protein extraction is a challenge for pull-down experiment.

Trying to identify other proteins that associate with (components of) the mutasome using ImuA'-mediated pull-down is challenging in that DnaE2, ImuB and ImuA' are all required for function of the mutasome, and based on Y2H analysis, may have to interact at some point for DNA damage survival as well as induced mutagenesis to occur. However, it still remains to be determined when these proteins interact during this dynamic process, which may also involve post-translational modification. In other words, do these (and other) proteins interact with one another simultaneously, or is there a temporal order of interactions? Therefore, the time at which the proteins are extracted for proteomic analysis could have a major influence on the outcome of a pull-down assay using one of the mutasome components as bait.

4.4 Heterologous expression of DnaE2 accessory proteins, ImuA' and ImuB

The expression of recombinant forms of ImuA' and ImuB in *E. coli* suitable for purification turned out to be very challenging. Of the expression system tested, the most promising results were obtained using the pMAL-c2 system, which allowed ImuB and ImuA' to be expressed in *E. coli* as N-terminal MBP fusions. However, in both cases, the majority of the fusion protein was found in the insoluble fraction. Amylose affinity chromatography was used in an attempt to purify MBP-ImuA' and MBP-ImuB proteins from the soluble fraction of *E. coli* cell extracts. However, the peak fractions that eluted from the affinity column were relatively impure, and contained many contaminating proteins. Preliminary data suggest that ion exchange chromatography could be used to further purify the MBP-ImuB protein; however, this remains to be confirmed. Proteins with disordered regions, such as ImuB, are known to be prone to aggregate or express in inclusion bodies therefore making it an enormous task to find appropriate in vitro conditions to obtain sufficient quantities in a soluble form, suitable for downstream processing (Goldstone *et al.*, 2008; Noens *et al.*, 2011). Various strategies have been employed to improve the production of problematical recombinant proteins. For example, proteins with disordered regions are best purified and maintained in soluble form in complex with their partners. An attempt was therefore made to co-express ImuB and ImuA' using the pETDuet system (Novagen), but this was not successful (data not shown). There have also been studies that show improve expression of mycobacterial proteins in *E. coli*, such as DprE1, when co-expressed with chaperones from *E. coli* (GroES) and *M. tuberculosis*

(CPN60.2) (Batt *et al.*, 2012). It will be worth testing whether this approach will work for ImuB and ImuA'.

5. Appendices

Appendix A: List of abbreviations

Ap	Ampicillin
ATc	Anhydrotetracycline
ATCC	American Type Culture Collection
bp	Base pairs
CFU	Colony forming unit
d	Days
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
DOTS	Directly Observed Therapy – Short Course
h	Hours
HIV	Human Immunodeficiency Virus
Hyg	Hygromycin
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kb	Kilo base pair
Km	Kanamycin
LA	Luria-Bertani agar
LB	Luria-Bertani broth
MDR-TB	Multidrug-Resistant Tuberculosis
min	Minutes
ml	Mililitre
OADC	Oleic acid-albumin-dextrose-catalase
OD600	Optical density at 600 nanometre wavelength
ORF	Open reading frame
PCR	Polymerase Chain Reaction
R	Resistant
RNA	Ribonucleic acid
Rpm	Revolutions per minute
s	Seconds
sdH20	Sterile distilled water
TB	Tuberculosis
TDR-TB	Totally Drug Resistant Tuberculosis
U	Units
XDR-TB	Extensively Drug Resistant Tuberculosis
v/v	Volume per volume
w/v	Weight per volume
WHO	World Health Organization
Y2H	Yeast two hybrid
BD	Binding domain
AD	Activation domain
SD	synthetic drop-out
T	tryptophan
L	leucine
H	histidine
A	Adenine
HIS3	Gene encoding imidazoleglycerol-phosphate dehydratase
ADE2	Gene encoding phosphoribosylaminoimidazole carboxylase

lacZ	Gene encoding β-galactosidase
attB	$tRNA^{Gly}$ attachment site
SD	Synthetic Drop out medium

Appendix B: Culture media and solutions

Culture media

All media was made up in one litre de-ionized water and except otherwise stated sterilized by autoclaving (121°C for 20 mins).

Luria-Bertani Broth (LB)

5g yeast, 10 g tryptone, 10 g sodium chloride

Luria-Bertani Agar (LA)

5 g yeast, 10 g tryptone, 10 g sodium chloride, 15g agar

2TY
5g sodium chloride, 10g yeast extract, 16g tryptone

Middlebrook 7H9

2 ml glycerol, 4.7g Difco™ Middlebrook 7H9 broth

Middlebrook 7H10

5 ml glycerol, 19 g Difco™ Middlebrook 7H10 agar

YPD broth

1.0 g Yeast extract, 2.0 g Peptone and 2.0 g Glucose

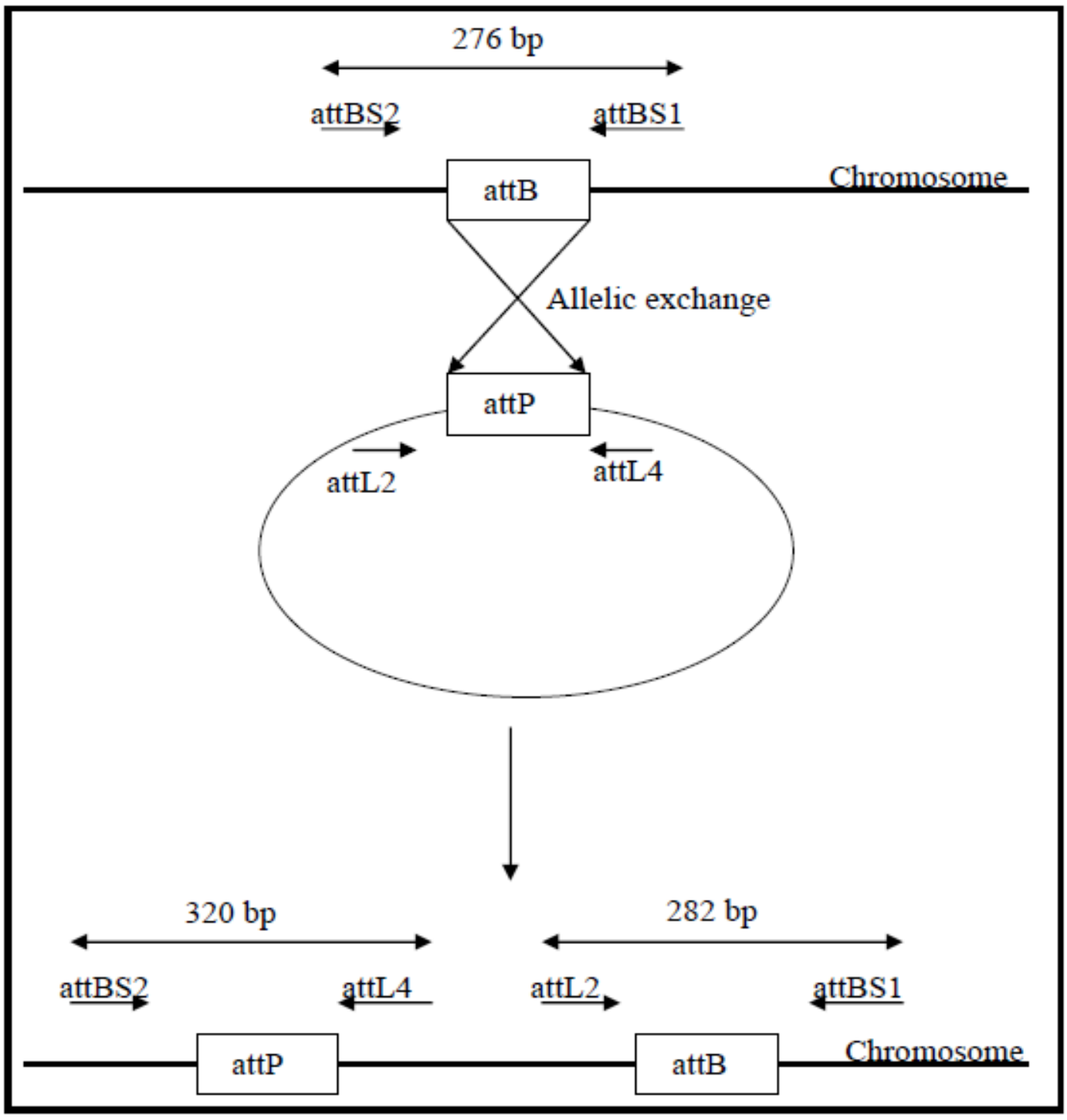
YPDA

YPD plus 0.15% agar both supplemented with 0.003% adenine.

Synthetic Drop out medium (SD)

6.7g Yeast nitrogen base without amino acids, 0.2 g Glucose, 20 g Bacto agar and 0.2 g drop-out mix

Appendix C: *attB* PCR strategy



Schematic representation of the PCR strategy used to confirm successful integration at the tRNAGly attachment site following allelic exchange

7. References

Abella, M., Erill, I., Jara, M., Mazon, G., Campoy, S. & Barbe, J. (2004). Widespread distribution of a lexA-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum. *Mol Microbiol* **54**, 212-222.

Abrahams, G. L., Kumar, A., Savvi, S. & other authors (2012). Pathway-selective sensitization of Mycobacterium tuberculosis for target-based whole-cell screening. *Chem Biol* **19**, 844-854.

Ahmad, S. (2011). Pathogenesis, immunology, and diagnosis of latent Mycobacterium tuberculosis infection. *Clin Dev Immunol* **2011**, 814943.

Anantharaman, V. & Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol* **4**, R81.

Andersson, D. I., Koskiniemi, S. & Hughes, D. (2010). Biological roles of translesion synthesis DNA polymerases in eubacteria. *Mol Microbiol* **77**, 540-548.

Andries, K., Verhasselt, P., Guillemont, J. & other authors (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science* **307**, 223-227.

Aniukwu, J., Glickman, M. S. & Shuman, S. (2008). The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev* **22**, 512-527.

Bald, D. & Koul, A. (2010). Respiratory ATP synthesis: the new generation of mycobacterial drug targets? *FEMS Microbiol Lett* **308**, 1-7.

Bald, D. & Koul, A. (2012). Advances and strategies in discovery of new antibacterials for combating metabolically resting bacteria. *Drug Discov Today* **pii:s1359-6446**, 00331-00335.

Barry, C. E., 3rd, Boshoff, H. I. & Dowd, C. S. (2004). Prospects for clinical introduction of nitroimidazole antibiotics for the treatment of tuberculosis. *Curr Pharm Des* **10**, 3239-3262.

Barry, C. E., 3rd, Boshoff, H. I., Dartois, V., Dick, T., Ehrt, S., Flynn, J., Schnappinger, D., Wilkinson, R. J. & Young, D. (2009). The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* **7**, 845-855.

Barry, C. E., 3rd & Blanchard, J. S. (2010). The chemical biology of new drugs in the development for tuberculosis. *Curr Opin Chem Biol* **14**, 456-466.

Batt, S. M., Jabeen, T., Bhowruth, V., Quill, L., Lund, P. A., Eggeling, L., Alderwick, L. J., Futterer, K. & Besra, G. S. (2012). Structural basis of inhibition of Mycobacterium tuberculosis DprE1 by benzothiazinone inhibitors. *Proc Natl Acad Sci U S A* **109**, 11354-11359.

Bedard, K. & Krause, K. H. (2007). The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol Rev* **87**, 245-313.

Bhowmik, T., Johnson, M. C. & Ray, B. (1987). Factors influencing synthesis and activity of beta-galactosidase in *Lactobacillus acidophilus*
J Ind Microbiol

2, 1-7.

Bjedov, I., Dasgupta, C. N., Slade, D., Le Blastier, S., Selva, M. & Matic, I. (2007). Involvement of Escherichia coli DNA polymerase IV in tolerance of cytotoxic alkylating DNA lesions in vivo. *Genetics* **176**, 1431-1440.

Boshoff, H. I., Reed, M. B., Barry, C. E., 3rd & Mizrahi, V. (2003). DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell* **113**, 183-193.

Brooks, P. C., Movahedzadeh, F. & Davis, E. O. (2001). Identification of some DNA damage-inducible genes of Mycobacterium tuberculosis: apparent lack of correlation with LexA binding. *J Bacteriol* **183**, 4459-4467.

Bruck, I., Woodgate, R., McEntee, K. & Goodman, M. F. (1996). Purification of a soluble UmuD'C complex from Escherichia coli. Cooperative binding of UmuD'C to single-stranded DNA. *J Biol Chem* **271**, 10767-10774.

Bunting, K. A., Roe, S. M. & Pearl, L. H. (2003). Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. *EMBO J* **22**, 5883-5892.

Cafarelli, T. M., Rands, T. J., Benson, R. W., Rudnicki, P. A., Lin, I. & Godoy, V. G. (2013). A single residue unique to DinB-like proteins limits formation of the Pol IV multi-protein complex in Escherichia coli. *J Bacteriol* **195**, 01349-01312.

Calver, A. D., Falmer, A. A., Murray, M. & other authors (2010). Emergence of increased resistance and extensively drug-resistant tuberculosis despite treatment adherence, South Africa. *Emerg Infect Dis* **16**, 264-271.

Caws, M., Thwaites, G., Stepniewska, K. & other authors (2006). Beijing genotype of Mycobacterium tuberculosis is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. *J Clin Microbiol* **44**, 3934-3939.

Chakraborty, S., Gruber, T., Barry, C. E., 3rd, Boshoff, H. I. & Rhee, K. Y. (2012). Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in Mycobacterium tuberculosis. *Science* **339**, 88-91.

Chen, X. P. & Du, G. H. (2007). Target validation: A door to drug discovery. *Drug Discov Ther* **1**, 23-29.

Chen, Y., Milam, S. L. & Erickson, H. P. (2012). SulA inhibits assembly of FtsZ by a simple sequestration mechanism. *Biochemistry* **51**, 3100-3109.

Cho, S. H., Goodlett, D. & Franzblau, S. (2006). ICAT-based comparative proteomic analysis of non-replicating persistent Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **86**, 445-460.

Cirz, R. T., Chin, J. K., Andes, D. R., de Crecy-Lagard, V., Craig, W. A. & Romesberg, F. E. (2005). Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol* **3**, e176.

Cirz, R. T. & Romesberg, F. E. (2007). Controlling mutation: intervening in evolution as a therapeutic strategy. *Crit Rev Biochem Mol Biol* **42**, 341-354.

Cole, C., Thomas, S., Filak, H., Henson, P. M. & Lenz, L. L. (2012). Nitric oxide increases susceptibility of Toll-like receptor-activated macrophages to spreading Listeria monocytogenes. *Immunity* **36**, 807-820.

Cole, S. T. & Riccardi, G. (2011). New tuberculosis drugs on the horizon. *Curr Opin Microbiol* **14**, 570-576.

Cordell, S. C., Robinson, E. J. & Lowe, J. (2003). Crystal structure of the SOS cell division inhibitor SulA and in complex with FtsZ. *Proc Natl Acad Sci U S A* **100**, 7889-7894.

Cordone, A., Audrain, B., Calabrese, I., Euphrasie, D. & Reyrat, J. M. (2011). Characterization of a Mycobacterium smegmatis uvrA mutant impaired in dormancy induced by hypoxia and low carbon concentration. *BMC Microbiol* **11**, 231.

Cox, M. M. (2007). Regulation of bacterial RecA protein function. *Crit Rev Biochem Mol Biol* **42**, 41-63.

Crunkhorn, S. (2012). Trial watch: Novel antimicrobial fights TB resistance. *Nat Rev Drug Discov* **11**, 590.

Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E. & Jennings, P. A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc Natl Acad Sci U S A* **98**, 11627-11632.

Dalrymple, B. P., Wijffels, G., Kongsuwan, K & Jennings, P. (2003). Towards understanding of protein-protein interaction network hierarchies. Analysis od DnaN Beta binding peptide motifs in members of protein families interacting with the eubacterial processivity clamp, beta subunit of DNA polymerase III. . *Proceedings of the First Asia-Pasific bioinformatics conference on Bioinformatics* **19**, 153-162.

Darwin, K. H. & Nathan, C. F. (2005). Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun* **73**, 4581-4587.

Davis, E. O., Dullaghan, E. M. & Rand, L. (2002). Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in Mycobacterium tuberculosis. *J Bacteriol* **184**, 3287-3295.

Davis, J. M. & Ramakrishnan, L. (2009). The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell* **136**, 37-49.

Della, M., Palmbos, P. L., Tseng, H. M. & other authors (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306**, 683-685.

Diacon, A. H., Dawson, R., von Groote-Bidlingmaier, F. & other authors (2012a). 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. *Lancet* **380**, 986-993.

Diacon, A. H., Donald, P. R., Pym, A. & other authors (2012b). Randomized pilot trial of eight weeks of bedaquiline (TMC207) treatment for multidrug-resistant tuberculosis: long-term outcome, tolerability, and effect on emergence of drug resistance. *Antimicrob Agents Chemother* **56**, 3271-3276.

Dick, T., Lee, B. H. & Murugasu-Oei, B. (1998). Oxygen depletion induced dormancy in Mycobacterium smegmatis. *FEMS Microbiol Lett* **163**, 159-164.

Dick, T. & Young, D. (2011). How antibacterials really work: impact on drug discovery. *Future Microbiol* **6**, 603-604.

Dolan, J. W., Kirkman, C. & Fields, S. (1989). The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc Natl Acad Sci U S A* **86**, 5703-5707.

Donlan, R. M. & Costerton, J. W. (2002). Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev* **15**, 167-193.

Dorr, T., Vulic, M. & Lewis, K. (2010). Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS Biol* **8**, e1000317.

Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434.

Dover, L. G. & Coxon, G. D. (2011). Current status and research strategies in tuberculosis drug development. *J Med Chem* **54**, 6157-6165.

Durbach, S. I., Andersen, S. J. & Mizrahi, V. (1997). SOS induction in mycobacteria: analysis of the DNA-binding activity of a LexA-like repressor and its role in DNA damage induction of the recA gene from Mycobacterium smegmatis. *Mol Microbiol* **26**, 643-653.

Durfee, T., Becherer, K., Chen, P. L., Yeh, S. H., Yang, Y., Kilburn, A. E., Lee, W. H. & Elledge, S. J. (1993). The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev* **7**, 555-569.

Dutta, N. K., Mehra, S., Didier, P. J. & other authors (2010). Genetic requirements for the survival of tubercle bacilli in primates. *J Infect Dis* **201**, 1743-1752.

Dye, C., Williams, B. G., Espinal, M. A. & Raviglione, M. C. (2002). Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science* **295**, 2042-2046.

Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208.

Ehrt, S., Schnappinger, D., Bekiranov, S., Drenkow, J., Shi, S., Gingeras, T. R., Gaasterland, T., Schoolnik, G. & Nathan, C. (2001). Reprogramming of the macrophage transcriptome in response to interferon-gamma and Mycobacterium tuberculosis: signaling roles of nitric oxide synthase-2 and phagocyte oxidase. *J Exp Med* **194**, 1123-1140.

Ehrt, S. & Schnappinger, D. (2009). Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol* **11**, 1170-1178.

Einhauer, A. & Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J Biochem Biophys Methods* **49**, 455-465.

Elias, J. E. & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214.

Erill, I., Campoy, S., Mazon, G. & Barbe, J. (2006). Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic Acids Res* **34**, 66-77.

Ferentz, A. E., Walker, G. C. & Wagner, G. (2001). Converting a DNA damage checkpoint effector (UmuD2C) into a lesion bypass polymerase (UmuD'2C). *EMBO J* **20**, 4287-4298.

Ford, C. W., Zurenko, G. E. & Barbachyn, M. R. (2001). The discovery of linezolid, the first oxazolidinone antibacterial agent. *Curr Drug Targets Infect Disord* **1**, 181-199.

Friedberg, E. C., Wagner, R. & Radman, M. (2002). Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* **296**, 1627-1630.

Friedberg, E. C., Lehmann, A. R. & Fuchs, R. P. (2005). Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol Cell* **18**, 499-505.

Friedman, N., Vardi, S., Ronen, M., Alon, U. & Stavans, J. (2005). Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria. *PLoS Biol* **3**, e238.

Galhardo, R. S., Rocha, R. P., Marques, M. V. & Menck, C. F. (2005). An SOS-regulated operon involved in damage-inducible mutagenesis in Caulobacter crescentus. *Nucleic Acids Res* **33**, 2603-2614.

Garces, A., Atmakuri, K., Chase, M. R. & other authors (2010). EspA acts as a critical mediator of ESX1-dependent virulence in Mycobacterium tuberculosis by affecting bacterial cell wall integrity. *PLoS Pathog* **6**, e1000957.

Georgescu, R. E., Yao, N. Y. & O'Donnell, M. (2010). Single-molecule analysis of the Escherichia coli replisome and use of clamps to bypass replication barriers. *FEBS Lett* **584**, 2596-2605.

Ghosh, J., Larsson, P., Singh, B., Pettersson, B. M., Islam, N. M., Sarkar, S. N., Dasgupta, S. & Kirsebom, L. A. (2009). Sporulation in mycobacteria. *Proc Natl Acad Sci U S A* **106**, 10781-10786.

Gillespie, S. H. (2002). Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. *Antimicrob Agents Chemother* **46**, 267-274.

Gillespie, S. H., Billington, O. J., Breathnach, A. & McHugh, T. D. (2002). Multiple drug-resistant Mycobacterium tuberculosis: evidence for changing fitness following passage through human hosts. *Microb Drug Resist* **8**, 273-279.

Gler, M. T., Skripconoka, V., Sanchez-Garavito, E. & other authors (2012). Delamanid for multidrug-resistant pulmonary tuberculosis. *N Engl J Med* **366**, 2151-2160.

Goldstone, R. M., Moreland, N. J., Bashiri, G., Baker, E. N. & Shaun Lott, J. (2008). A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in Mycobacterium smegmatis. *Protein Expr Purif* **57**, 81-87.

Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S. & Glickman, M. S. (2005). Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol* **12**, 304-312.

Gonzalez, M., Frank, E. G., Levine, A. S. & Woodgate, R. (1998). Lon-mediated proteolysis of the Escherichia coli UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis. *Genes Dev* **12**, 3889-3899.

Goodman, M. F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **71**, 17-50.

Gorna, A. E., Bowater, R. P. & Dziadek, J. (2010). DNA repair systems and the pathogenesis of Mycobacterium tuberculosis: varying activities at different stages of infection. *Clin Sci (Lond)* **119**, 187-202.

Graham, J. E. & Clark-Curtiss, J. E. (1999). Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc Natl Acad Sci U S A* **96**, 11554-11559.

Grzegorzewicz, A. E., Kordulakova, J., Jones, V. & other authors (2012). A Common Mechanism of Inhibition of the Mycobacterium tuberculosis Mycolic Acid Biosynthetic Pathway by Isoxyl and Thiacetazone. *J Biol Chem* **287**, 38434-38441.

Gupta, R., Barkan, D., Redelman-Sidi, G., Shuman, S. & Glickman, M. S. (2011). Mycobacteria exploit three genetically distinct DNA double-strand break repair pathways. *Mol Microbiol* **79**, 316-330.

Gutacker, M., Valsangiacomo, C., Bernasconi, M. V. & Piffaretti, J. C. (2002a). RecA and glnA sequences separate the bacteroides fragilis population into two genetic divisions associated with the antibiotic resistance genotypes cepA and cfiA. *J Med Microbiol* **51**, 123-130.

Gutacker, M. M., Smoot, J. C., Migliaccio, C. A. & other authors (2002b). Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **162**, 1533-1543.

Guthlein, C., Wanner, R. M., Sander, P., Davis, E. O., Bosshard, M., Jiricny, J., Bottger, E. C. & Springer, B. (2009). Characterization of the mycobacterial NER system reveals novel functions of the uvrD1 helicase. *J Bacteriol* **191**, 555-562.

Hanekom, M., van der Spuy, G. D., Streicher, E. & other authors (2007). A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family is associated with an increased ability to spread and cause disease. *J Clin Microbiol* **45**, 1483-1490.

Hazbon, M. H., Brimacombe, M., Bobadilla del Valle, M. & other authors (2006). Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **50**, 2640-2649.

Hernan, R., Heuermann, K. & Brizzard, B. (2000). Multiple epitope tagging of expressed proteins for enhanced detection. *Biotechniques* **28**, 789-793.

Huang, X., Kolbanovskiy, A., Wu, X., Zhang, Y., Wang, Z., Zhuang, P., Amin, S. & Geacintov, N. E. (2003). Effects of base sequence context on translesion synthesis past a bulky (+)-trans-anti-B[a]P-N2-dG lesion catalyzed by the Y-family polymerase pol kappa. *Biochemistry* **42**, 2456-2466.

Hurdle, J. G., O'Neill, A. J., Chopra, I. & Lee, R. E. (2011). Targeting bacterial membrane function: an underexploited mechanism for treating persistent infections. *Nat Rev Microbiol* **9**, 62-75.

Hussain, S., Malik, M., Shi, L., Gennaro, M. L. & Drlica, K. (2009). In vitro model of mycobacterial growth arrest using nitric oxide with limited air. *Antimicrob Agents Chemother* **53**, 157-161.

James, P., Halladay, J. & Craig, E. A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425-1436.

Jia, L., Geacintov, N. E. & Broyde, S. (2008). The N-clasp of human DNA polymerase kappa promotes blockage or error-free bypass of adenine- or guanine-benzo[a]pyrenyl lesions. *Nucleic Acids Res* **36**, 6571-6584.

Jiang, Q., Karata, K., Woodgate, R., Cox, M. M. & Goodman, M. F. (2009). The active form of DNA polymerase V is UmuD'(2)C-RecA-ATP. *Nature* **460**, 359-363.

Johansen, S. K., Maus, C. E., Plikaytis, B. B. & Douthwaite, S. (2006). Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Mol Cell* **23**, 173-182.

Johnson, R., Warren, R. M., van der Spuy, G. D. & other authors (2009). Drug-resistant tuberculosis epidemic in the Western Cape driven by a virulent Beijing genotype strain. *Int J Tuberc Lung Dis* **14**, 119-121.

Johnson, R. E., Prakash, L. & Prakash, S. (2012). Pol31 and Pol32 subunits of yeast DNA polymerase delta are also essential subunits of DNA polymerase zeta. *Proc Natl Acad Sci U S A* **109**, 12455-12460.

Joung, J. K., Ramm, E. I. & Pabo, C. O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A* **97**, 7382-7387.

Kana, B. D., Abrahams, G. L., Sung, N. & other authors (2010). Role of the DinB homologs Rv1537 and Rv3056 in Mycobacterium tuberculosis. *J Bacteriol* **192**, 2220-2227.

Kane, D. P., Shusterman, M., Rong, Y. & McVey, M. (2012). Competition between replicative and translesion polymerases during homologous recombination repair in Drosophila. *PLoS Genet* **8**, e1002659.

Kelman, Z. & O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem* **64**, 171-200.

Kennedy, B. K. (2002). Mammalian transcription factors in yeast: strangers in a familiar land. *Nat Rev Mol Cell Biol* **3**, 41-49.

Kirkman-Correia, C., Stroke, I. L. & Fields, S. (1993). Functional domains of the yeast STE12 protein, a pheromone-responsive transcriptional activator. *Mol Cell Biol* **13**, 3765-3772.

Koh, W. J., Kang, Y. R., Jeon, K., Kwon, O. J., Lyu, J., Kim, W. S. & Shim, T. S. (2012). Daily 300 mg dose of linezolid for multidrug-resistant and extensively drug-resistant tuberculosis: updated analysis of 51 patients. *J Antimicrob Chemother* **67**, 1503-1507.

Koorits, L., Tegova, R., Tark, M., Tarassova, K., Tover, A. & Kivisaar, M. (2007). Study of involvement of ImuB and DnaE2 in stationary-phase mutagenesis in Pseudomonas putida. *DNA Repair (Amst)* **6**, 863-868.

Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature* **469**, 483-490.

Kumar, A., Deshane, J. S., Crossman, D. K., Bolisetty, S., Yan, B. S., Kramnik, I., Agarwal, A. & Steyn, A. J. (2008). Heme oxygenase-1-derived carbon monoxide induces the Mycobacterium tuberculosis dormancy regulon. *J Biol Chem* **283**, 18032-18039.

Kurthkoti, K. & Varshney, U. (2011). Distinct mechanisms of DNA repair in mycobacteria and their implications in attenuation of the pathogen growth. *Mech Ageing Dev* **133**, 138-146.

Lamichhane, G. (2011). Novel targets in M. tuberculosis: search for new drugs. *Trends Mol Med* **17**, 25-33.

Langston, L. D. & O'Donnell, M. (2006). DNA replication: keep moving and don't mind the gap. *Mol Cell* **23**, 155-160.

Larsen, M. H. (2000). Appendix 1, Some common methods in Mycobacterial genetics. In *Molecular genetics of mycobacteria*, pp. p. 313-320. Edited by G. F. H. a. J. W. R. & Jacobs. Washington,
D.C.: ASM Press.

Lee, M., Lee, J., Carroll, M. W. & other authors (2012). Linezolid for treatment of chronic extensively drug-resistant tuberculosis. *N Engl J Med* **367**, 1508-1518.

Li, Z., Wen, J., Lin, Y. & other authors (2011). A Sir2-like protein participates in mycobacterial NHEJ. *PLoS One* **6**, e20045.

Liao, S., Shang, Q., Zhang, X., Zhang, J., Xu, C. & Tu, X. (2009). Pup, a prokaryotic ubiquitin-like protein, is an intrinsically disordered protein. *Biochem J* **422**, 207-215.

Lim, A. & Dick, T. (2001). Plate-based dormancy culture system for Mycobacterium smegmatis and isolation of metronidazole-resistant mutants. *FEMS Microbiol Lett* **200**, 215-219.

Lin, P. L., Rodgers, M., Smith, L. & other authors (2009). Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. *Infect Immun* **77**, 4631-4642.

Lin, W., Xin, H., Zhang, Y., Wu, X., Yuan, F. & Wang, Z. (1999). The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Res* **27**, 4468-4475.

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453-1459.

Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91-102.

Lior-Hoffmann, L., Wang, L., Wang, S., Geacintov, N. E., Broyde, S. & Zhang, Y. (2012). Preferred WMSA catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase kappa elucidates error-free bypass of a bulky DNA lesion. *Nucleic Acids Res*.

Lougheed, K. E., Taylor, D. L., Osborne, S. A., Bryans, J. S. & Buxton, R. S. (2009). New anti-tuberculosis agents amongst known drugs. *Tuberculosis (Edinb)* **89**, 364-370.

Ma, Z., Lienhardt, C., McIlleron, H., Nunn, A. J. & Wang, X. (2010). Global tuberculosis drug development pipeline: the need and the reality. *Lancet* **375**, 2100-2109.

MacMicking, J. D., North, R. J., LaCourse, R., Mudgett, J. S., Shah, S. K. & Nathan, C. F. (1997). Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc Natl Acad Sci U S A* **94**, 5243-5248.

Makarov, V., Manina, G., Mikusova, K. & other authors (2009). Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis. *Science* **324**, 801-804.

Manjunatha, U., Boshoff, H. I. & Barry, C. E. (2009). The mechanism of action of PA-824: Novel insights from transcriptional profiling. *Commun Integr Biol* **2**, 215-218.

Manjunatha, U. H., Boshoff, H., Dowd, C. S. & other authors (2006). Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, 431-436.

Maroz, A., Shinde, S. S., Franzblau, S. G., Ma, Z., Denny, W. A., Palmer, B. D. & Anderson, R. F. (2010). Release of nitrite from the antitubercular nitroimidazole drug PA-824 and analogues upon one-electron reduction in protic, non-aqueous solvent. *Org Biomol Chem* **8**, 413-418.

Marx, A. & Summerer, D. (2002). Molecular insights into error-prone DNA replication and error-free lesion bypass. *Chembiochem* **3**, 405-407.

Matsumoto, M., Hashizume, H., Tomishige, T., Kawasaki, M., Tsubouchi, H., Sasaki, H., Shimokawa, Y. & Komatsu, M. (2006). OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. *PLoS Med* **3**, e466.

McGuire, A. M., Weiner, B., Park, S. T. & other authors (2012). Comparative analysis of Mycobacterium and related Actinomycetes yields insight into the evolution of Mycobacterium tuberculosis pathogenesis. *BMC Genomics* **13**, 120.

Migliori, G. B., Centis, R., D'Ambrosio, L., Spanevello, A., Borroni, E., Cirillo, D. M. & Sotgiu, G. (2012). Totally drug-resistant and extremely drug-resistant tuberculosis: the same disease? *Clin Infect Dis* **54**, 1379-1380.

Mizrahi, V. & Andersen, S. J. (1998). DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Mol Microbiol* **29**, 1331-1339.

Mungrue, I. N., Husain, M. & Stewart, D. J. (2002). The role of NOS in heart failure: lessons from murine genetic models. *Heart Fail Rev* **7**, 407-422.

Murli, S., Opperman, T., Smith, B. T. & Walker, G. C. (2000). A role for the umuDC gene products of Escherichia coli in increasing resistance to DNA damage in stationary phase by inhibiting the transition to exponential growth. *J Bacteriol* **182**, 1127-1135.

Musser, J. M., Kapur, V., Williams, D. L., Kreiswirth, B. N., van Soolingen, D. & van Embden, J. D. (1996). Characterization of the catalase-peroxidase gene (katG) and inhA locus in isoniazid-resistant and -susceptible strains of Mycobacterium tuberculosis by automated DNA sequencing: restricted array of mutations associated with drug resistance. *J Infect Dis* **173**, 196-202.

Musser, J. M., Amin, A. & Ramaswamy, S. (1999). Mutations in genes associated with drug resistance in Mycobacterium tuberculosis isolates from Italy. *J Infect Dis* **180**, 1751-1753.

Naidoo, R. (2007). Active pulmonary tuberculosis: experience with resection in 106 cases. *Asian Cardiovasc Thorac Ann* **15**, 134-138.

Napolitano, R., Janel-Bintz, R., Wagner, J. & Fuchs, R. P. (2000). All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *EMBO J* **19**, 6259-6265.

Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **382**, 729-731.

Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658.

Neuwald, A. F. (2003). Evolutionary clues to DNA polymerase III beta clamp structural mechanisms. *Nucleic Acids Res* **31**, 4503-4516.

Nikaido, H. (1994). Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264**, 382-388.

Nikaido, H. (2001). Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria. *Semin Cell Dev Biol* **12**, 215-223.

Noens, E. E., Williams, C., Anandhakrishnan, M., Poulsen, C., Ehebauer, M. T. & Wilmanns, M. (2011). Improved mycobacterial protein production using a Mycobacterium smegmatis groEL1DeltaC expression strain. *BMC Biotechnol* **11**, 27.

Nyka, W. (1974). Studies on the effect of starvation on mycobacteria. *Infect Immun* **9**, 843-850.

O'Sullivan, D. M., Hinds, J., Butcher, P. D., Gillespie, S. H. & McHugh, T. D. (2008). Mycobacterium tuberculosis DNA repair in response to subinhibitory concentrations of ciprofloxacin. *J Antimicrob Chemother* **62**, 1199-1202.

Ollivierre, J. N., Fang, J. & Beuning, P. J. (2010). The Roles of UmuD in Regulating Mutagenesis. *J Nucleic Acids* **2010**.

Olson, M. W., Dallmann, H. G. & McHenry, C. S. (1995). DnaX complex of Escherichia coli DNA polymerase III holoenzyme. The chi psi complex functions by increasing the affinity of tau and gamma for delta.delta' to a physiologically relevant range. *J Biol Chem* **270**, 29570-29577.

Opperman, T., Murli, S. & Walker, G. C. (1996). The genetic requirements for UmuDC-mediated cold sensitivity are distinct from those for SOS mutagenesis. *J Bacteriol* **178**, 4400-4411.

Opperman, T., Murli, S., Smith, B. T. & Walker, G. C. (1999). A model for a umuDC-dependent prokaryotic DNA damage checkpoint. *Proc Natl Acad Sci U S A* **96**, 9218-9223.

Pages, V. & Fuchs, R. P. (2002). How DNA lesions are turned into mutations within cells? *Oncogene* **21**, 8957-8966.

Papavinasasundaram, K. G., Anderson, C., Brooks, P. C., Thomas, N. A., Movahedzadeh, F., Jenner, P. J., Colston, M. J. & Davis, E. O. (2001). Slow induction of RecA by DNA damage in Mycobacterium tuberculosis. *Microbiology* **147**, 3271-3279.

Parish, T. & Stoker, N. G. (2000). Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. *Microbiology* **146 (Pt 8)**, 1969-1975.

Pasca, M. R., Degiacomi, G., Ribeiro, A. L. & other authors (2010). Clinical isolates of Mycobacterium tuberculosis in four European hospitals are uniformly susceptible to benzothiazinones. *Antimicrob Agents Chemother* **54**, 1616-1618.

Patel, M., Jiang, Q., Woodgate, R., Cox, M. M. & Goodman, M. F. (2010). A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V. *Crit Rev Biochem Mol Biol* **45**, 171-184.

Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40.

Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R. & Hendrickson, W. A. (1996). Structure of the UmuD' protein and its regulation in response to DNA damage. *Nature* **380**, 727-730.

Pei, J., Kim, B. H. & Grishin, N. V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36**, 2295-2300.

Peyron, P., Vaubourgeix, J., Poquet, Y. & other authors (2008). Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence. *PLoS Pathog* **4**, e1000204.

Phillips, D. H. (1983). Fifty years of benzo(a)pyrene. *Nature* **303**, 468-472.

Phillips, I. (1987). Bacterial mutagenicity and the 4-quinolones. *J Antimicrob Chemother* **20**, 771-773.

Phizicky, E. M. & Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiol Rev* **59**, 94-123.

Pinon, M., Scolfaro, C., Bignamini, E., Cordola, G., Esposito, I., Milano, R., Mignone, F., Bertaina, C. & Tovo, P. A. (2010). Two pediatric cases of multidrug-resistant tuberculosis treated with linezolid and moxifloxacin. *Pediatrics* **126**, e1253-1256.

Prabha, S., Rao, D. N. & Nagaraja, V. (2011). Distinct properties of hexameric but functionally conserved Mycobacterium tuberculosis transcription-repair coupling factor. *PLoS One* **6**, e19131.

Prammananan, T., Phunpruch, S., Jaitrong, S. & Palittapongarnpim, P. (2012). Mycobacterium tuberculosis uvrC essentiality in response to UV-induced cell damage. *Southeast Asian J Trop Med Public Health* **43**, 370-375.

Rachman, H., Strong, M., Schaible, U., Schuchhardt, J., Hagens, K., Mollenkopf, H., Eisenberg, D. & Kaufmann, S. H. (2006). Mycobacterium tuberculosis gene expression profiling within the context of protein networks. *Microbes Infect* **8**, 747-757.

Ramakrishnan, L. (2012). Revisiting the role of the granuloma in tuberculosis. *Nat Rev Immunol* **12**, 352-366.

Ramaswamy, S. & Musser, J. M. (1998). Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. *Tuber Lung Dis* **79**, 3-29.

Raviglione, M., Marais, B., Floyd, K. & other authors (2012). Scaling up interventions to achieve global tuberculosis control: progress and new developments. *Lancet* **379**, 1902-1913.

Rechkoblit, O., Zhang, Y., Guo, D., Wang, Z., Amin, S., Krzeminsky, J., Louneva, N. & Geacintov, N. E. (2002). trans-Lesion synthesis past bulky benzo[a]pyrene diol epoxide N2-dG and N6-dA lesions catalyzed by DNA bypass polymerases. *J Biol Chem* **277**, 30488-30494.

Rengarajan, J., Sassetti, C. M., Naroditskaya, V., Sloutsky, A., Bloom, B. R. & Rubin, E. J. (2004). The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria. *Mol Microbiol* **53**, 275-282.

Robertson, B. D., Altmann, D., Barry, C. & other authors (2012). Detection and treatment of subclinical tuberculosis. *Tuberculosis (Edinb)* **92**, 447-452.

Rossi, F., Khanduja, J. S., Bortoluzzi, A. & other authors (2011). The biological and structural characterization of Mycobacterium tuberculosis UvrA provides novel insights into its mechanism of action. *Nucleic Acids Res* **39**, 7316-7328.

Russell, D. G., Barry, C. E., 3rd & Flynn, J. L. (2010). Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**, 852-856.

Rustad, T. R., Sherrid, A. M., Minch, K. J. & Sherman, D. R. (2009). Hypoxia: a window into Mycobacterium tuberculosis latency. *Cell Microbiol* **11**, 1151-1159.

Sacchettini, J. C., Rubin, E. J. & Freundlich, J. S. (2008). Drugs versus bugs: in pursuit of the persistent predator Mycobacterium tuberculosis. *Nat Rev Microbiol* **6**, 41-52.

Sacksteder, K. A., Protopopova, M., Barry, C. E., 3rd, Andries, K. & Nacy, C. A. (2012). Discovery and development of SQ109: a new antitubercular drug with a novel mechanism of action. *Future Microbiol* **7**, 823-837.

Sala, C. & Hartkoorn, R. C. (2011). Tuberculosis drugs: new candidates and how to find more. *Future Microbiol* **6**, 617-633.

Salomon, J. A., Lloyd-Smith, J. O., Getz, W. M., Resch, S., Sanchez, M. S., Porco, T. C. & Borgdorff, M. W. (2006). Prospects for advancing tuberculosis control efforts through novel therapies. *PLoS Med* **3**, e273.

Sambrook, J., and D. W. Russell (2001). *Molecular cloning: a laboratory manual*, 3rd. ed edn. New York: Cold Spring Harbor.

Sambrook, J., E. F. Fritsch, and T. Maniatis (1989). *Molecular cloning. A laboratory manual*, Second Edition edn. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

Sander, P., Papavinasasundaram, K. G., Dick, T., Stavropoulos, E., Ellrott, K., Springer, B., Colston, M. J. & Bottger, E. C. (2001). Mycobacterium bovis BCG recA deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model. *Infect Immun* **69**, 3562-3568.

Sanders, G. M., Dallmann, H. G. & McHenry, C. S. (2010). Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases. *Mol Cell* **37**, 273-281.

Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B. K., Church, G. M. & Murray, M. B. (2009). Tuberculosis drug resistance mutation database. *PLoS Med* **6**, e2.

Sarker, M., Talcott, C., Madrid, P., Chopra, S., Bunin, B. A., Lamichhane, G., Freundlich, J. S. & Ekins, S. (2012). Combining cheminformatics methods and pathway analysis to identify molecules with whole-cell activity against Mycobacterium tuberculosis. *Pharm Res* **29**, 2115-2127.

Scorpio, A. & Zhang, Y. (1996). Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med* **2**, 662-667.

Sherrer, S. M., Maxwell, B. A., Pack, L. R., Fiala, K. A., Fowler, J. D., Zhang, J. & Suo, Z. (2012). Identification of an unfolding intermediate for a DNA lesion bypass polymerase. *Chem Res Toxicol* **25**, 1531-1540.

Shi, W., Zhang, X., Jiang, X., Yuan, H., Lee, J. S., Barry, C. E., 3rd, Wang, H., Zhang, W. & Zhang, Y. (2011). Pyrazinamide inhibits trans-translation in Mycobacterium tuberculosis. *Science* **333**, 1630-1632.

Shimoni, Y., Altuvia, S., Margalit, H. & Biham, O. (2009). Stochastic analysis of the SOS response in Escherichia coli. *PLoS One* **4**, e5363.

Sickmeier, M., Hamilton, J. A., LeGall, T. & other authors (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* **35**, D786-793.

Sies, H. & Cadenas, E. (1985). Oxidative stress: damage to intact cells and organs. *Philos Trans R Soc Lond B Biol Sci* **311**, 617-631.

Sies, H. & Mehlhorn, R. (1986). Mutagenicity of nitroxide-free radicals. *Arch Biochem Biophys* **251**, 393-396.

Singh, R., Manjunatha, U., Boshoff, H. I. & other authors (2008). PA-824 kills nonreplicating Mycobacterium tuberculosis by intracellular NO release. *Science* **322**, 1392-1395.

Smith, B. T. & Walker, G. C. (1998). Mutagenesis and more: umuDC and the Escherichia coli SOS response. *Genetics* **148**, 1599-1610.

Smollett, K. L., Smith, K. M., Kahramanoglou, C., Arnvig, K. B., Buxton, R. S. & Davis, E. O. (2012). Global analysis of the regulon of the transcriptional repressor LexA, a key component of SOS response in Mycobacterium tuberculosis. *J Biol Chem* **287**, 22004-22014.

Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol* **4**, 1911-1919.

Sotgiu, G., Centis, R., D'Ambrosio, L. & other authors (2012). Efficacy, safety and tolerability of linezolid containing regimens in treating MDR-TB and XDR-TB: systematic review and meta-analysis. *Eur Respir J* **40**, 1430-1442.

Stewart, G. R., Robertson, B. D. & Young, D. B. (2003). Tuberculosis: a problem with persistence. *Nat Rev Microbiol* **1**, 97-105.

Stewart, J., Hingorani, M. M., Kelman, Z. & O'Donnell, M. (2001). Mechanism of beta clamp opening by the delta subunit of Escherichia coli DNA polymerase III holoenzyme. *J Biol Chem* **276**, 19182-19189.

Story, R. M. & Steitz, T. A. (1992). Structure of the recA protein-ADP complex. *Nature* **355**, 374-376.

Story, R. M., Weber, I. T. & Steitz, T. A. (1992). The structure of the E. coli recA protein monomer and polymer. *Nature* **355**, 318-325.

Stukenberg, P. T., Studwell-Vaughan, P. S. & O'Donnell, M. (1991). Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J Biol Chem* **266**, 11328-11334.

Sutton, M. D., Smith, B. T., Godoy, V. G. & Walker, G. C. (2000). The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu Rev Genet* **34**, 479-497.

Sutton, M. D., Farrow, M. F., Burton, B. M. & Walker, G. C. (2001). Genetic interactions between the Escherichia coli umuDC gene products and the beta processivity clamp of the replicative DNA polymerase. *J Bacteriol* **183**, 2897-2909.

Sutton, M. D. & Walker, G. C. (2001). Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc Natl Acad Sci U S A* **98**, 8342-8349.

Tahlan, K., Wilson, R., Kastrinsky, D. B. & other authors (2012). SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **56**, 1797-1809.

Tasneen, R., Tyagi, S., Williams, K., Grosset, J. & Nuermberger, E. (2008). Enhanced bactericidal activity of rifampin and/or pyrazinamide when combined with PA-824 in a murine model of tuberculosis. *Antimicrob Agents Chemother* **52**, 3664-3668.

Tasneen, R., Li, S. Y., Peloquin, C. A., Taylor, D., Williams, K. N., Andries, K., Mdluli, K. E. & Nuermberger, E. L. (2011). Sterilizing activity of novel TMC207- and PA-824-containing regimens in a murine model of tuberculosis. *Antimicrob Agents Chemother* **55**, 5485-5492.

Telenti, A., Imboden, P., Marchesi, F., Schmidheini, T. & Bodmer, T. (1993). Direct, automated detection of rifampin-resistant Mycobacterium tuberculosis by polymerase chain reaction and single-strand conformation polymorphism analysis. *Antimicrob Agents Chemother* **37**, 2054-2058.

Telenti, A., Philipp, W. J., Sreevatsan, S., Bernasconi, C., Stockbauer, K. E., Wieles, B., Musser, J. M. & Jacobs, W. R., Jr. (1997). The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol. *Nat Med* **3**, 567-570.

Tessman, I., Liu, S. K. & Kennedy, M. A. (1992). Mechanism of SOS mutagenesis of UV-irradiated DNA: mostly error-free processing of deaminated cytosine. *Proc Natl Acad Sci U S A* **89**, 1159-1163.

Thiede, B., Hohenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F. & Jungblut, P. R. (2005). Peptide mass fingerprinting. *Methods* **35**, 237-247.

Tippin, B., Pham, P. & Goodman, M. F. (2004). Error-prone replication for better or worse. *Trends Microbiol* **12**, 288-295.

Trusca, D., Scott, S., Thompson, C. & Bramhill, D. (1998). Bacterial SOS checkpoint protein SulA inhibits polymerization of purified FtsZ cell division protein. *J Bacteriol* **180**, 3946-3953.

Ulrichs, T. & Kaufmann, S. H. (2006). New insights into the function of granulomas in human tuberculosis. *J Pathol* **208**, 261-269.

van Crevel, R., Ottenhoff, T. H. & van der Meer, J. W. (2002). Innate immunity to Mycobacterium tuberculosis. *Clin Microbiol Rev* **15**, 294-309.

Velayati, A. A., Masjedi, M. R., Farnia, P., Tabarsi, P., Ghanavi, J., Ziazarifi, A. H. & Hoffner, S. E. (2009). Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. *Chest* **136**, 420-425.

Via, L. E., Lin, P. L., Ray, S. M. & other authors (2008). Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* **76**, 2333-2340.

Vidal, M. & Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* **27**, 919-929.

Vilcheze, C., Wang, F., Arai, M. & other authors (2006). Transfer of a point mutation in Mycobacterium tuberculosis inhA resolves the target of isoniazid. *Nat Med* **12**, 1027-1029.

Wagner, J., Gruz, P., Kim, S. R., Yamada, M., Matsui, K., Fuchs, R. P. & Nohmi, T. (1999). The dinB gene encodes a novel E. coli DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol Cell* **4**, 281-286.

Walhout, A. J. & Vidal, M. (2001). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**, 297-306.

Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138-2139.

Warner, D. F. (2010). The role of DNA repair in M. tuberculosis pathogenesis. *Drug Discov Today Dis Mech* **7**, e5-e11.

Warner, D. F., Ndwandwe, D. E., Abrahams, G. L., Kana, B. D., Machowski, E. E., Venclovas, C. & Mizrahi, V. (2010). Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **107**, 13093-13098.

Warner, D. F. & Mizrahi, V. (2012). Approaches to target identification and validation for tuberculosis drug discovery: a UCT perspective. *S Afr Med J* **102**, 457-460.

Wayne, L. G. & Hayes, L. G. (1996). An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. *Infect Immun* **64**, 2062-2069.

Weller, G. R., Kysela, B., Roy, R. & other authors (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* **297**, 1686-1689.

Westblade, L. F., Campbell, E. A., Pukhrambam, C., Padovan, J. C., Nickels, B. E., Lamour, V. & Darst, S. A. (2010). Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res* **38**, 8357-8369.

WHO (2006). The Stop TB strategy.

WHO (2009). Global TB Control Report.

WHO (2012). Global tuberculosis report 2012.

Wijffels, G., Dalrymple, B., Kongsuwan, K. & Dixon, N. E. (2005). Conservation of eubacterial replicases. *IUBMB Life* **57**, 413-419.

Woodgate, R., Rajagopalan, M., Lu, C. & Echols, H. (1989). UmuC mutagenesis protein of Escherichia coli: purification and interaction with UmuD and UmuD'. *Proc Natl Acad Sci U S A* **86**, 7301-7305.

Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321-331.

Yang, W. & Woodgate, R. (2007). What a difference a decade makes: insights into translesion DNA synthesis. *Proc Natl Acad Sci U S A* **104**, 15591-15598.

Yeiser, B., Pepper, E. D., Goodman, M. F. & Finkel, S. E. (2002). SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci U S A* **99**, 8737-8741.

Zhang, B., VerBerkmoes, N. C., Langston, M. A., Uberbacher, E., Hettich, R. L. & Samatova, N. F. (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res* **5**, 2909-2918.

Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J. S., Geacintov, N. E. & Wang, Z. (2000). Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro. *Nucleic Acids Res* **28**, 4138-4146.

Zhang, Y., Wu, X., Guo, D., Rechkoblit, O., Geacintov, N. E. & Wang, Z. (2002). Two-step error-prone bypass of the (+)- and (-)-trans-anti-BPDE-N2-dG adducts by human DNA polymerases eta and kappa. *Mutat Res* **510**, 23-35.

Zumla, A., Abubakar, I., Raviglione, M. & other authors (2012a). Drug-resistant tuberculosis--current dilemmas, unanswered questions, challenges, and priority needs. *J Infect Dis* **205 Suppl 2**, S228-240.

Zumla, A., Hafner, R., Lienhardt, C., Hoelscher, M. & Nunn, A. (2012b). Advancing the development of tuberculosis therapy. *Nat Rev Drug Discov* **11**, 171-172.