Structural Analysis of Induced Mutagenesis Protein B from *Mycobacterium tuberculosis*

UNIVERSITY *of the*
WESTERN CAPE

Jeremy Boonzaier

2849425

This thesis is submitted to the department of Biotechnology, University of the Western Cape, in fulfillment of the requirements for a Master of Science Degree in Biotechnology

Supervisor: Prof Wolf-Dieter Schubert

Co-supervisor: Prof David Pugh

Summary

Knowing the three-dimensional structure of a protein may be useful in understanding its function. In this study, induced mutagenesis protein B (ImuB) from *Mycobacterium tuberculosis* was analyzed using molecular biology and molecular modelling techniques. The *Rv3394c* gene expressing ImuB was obtained from the group of Prof. Digby Warner at the Institute of Infectious Diseases and Molecular Medicine, University of Cape Town. *Rv33974c* was amplified from an expression plasmid using polymerase chain reaction (PCR) and inserted into multiple expression vectors. The pMal-c2X-*Rv3394c* construct was most successful in producing ImuB as a fusion protein with N-terminal maltose binding protein in an *E. coli* expression systems. Attempts were undertaken to refold insoluble ImuB. Soluble MBP-ImuB was purified by affinity chromatography and size-exclusion chromatography. Purified MBP-ImuB was concentrated and used for hanging drop crystallization experiments. Crystallization of ImuB remained elusive as protein crystals did not form. A homology model of ImuB was generated based on structurally related Y-family DNA polymerases. ImuB, however, lacks the catalytic residues required for DNA replication. Sequence analysis an identified a potentially disordered C-terminal domain. Together, this would suggest that ImuB is not directly responsible for induced mutagenesis but is required as an accessory protein for induced mutagenesis to occur.

Key Words: *Mycobacterium tuberculosis*, Y-family DNA polymerase, SOS response, ImuB, induced mutagenesis, homology modeling, ImuA'-ImuB-DnaE2

Content

Acknowledgements

I would like to show appreciation to my supervisor Prof. Wolf-Dieter Schubert to whom I owe a great deal for his contribution towards my academic development and success. Secondly, I would like to acknowledge the inputs of my colleagues in the structural biology of infectious diseases group at the University of the Western Cape. I would also like to acknowledge the support of others outside of the structural biology group for sharing their knowledge and strategies used within the study. For my financial support I would like to thank the National Research Foundation.

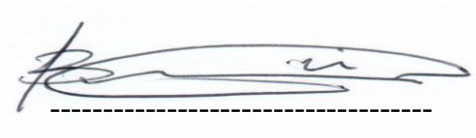
UNIVERSITY *of the*
WESTERN CAPE

General Plagiarism Declaration

Name: Jeremy Joel Boonzaier

Student number: 2849425

1. I hereby declare that I know what plagiarism entails, namely to use another's work and to present it as my own without attributing the sources in the correct way. (Refer to University Calendar part 1 for definition)
2. I know that plagiarism is a punishable offence because it constitutes theft.
3. I understand the plagiarism policy of the Faculty of Natural Science of the University of the Western Cape.
4. I know what the consequences will be if I plagiarize in any of the assignments for my course.
5. I declare therefore that all work presented by me for every aspect of my course, will be my own, and where I have made use of another's work, I will attribute the source in the correct way.



---------- 07-07-2016 ----------

Signature Date

Abbreviations

Amp	Ampicillin
APS	Ammonium persulfate
AR	Amylose resin
ATP	Adenosine tri-phosphate
DNAP	DNA polymerase
EDTA	Ethylene diamine tetra-acetic acid
FXa	Factor Xa protease
GPC	Gel permeation chromatography
GS	Glutathione sepharose
GST	Glutathione-S-transferase
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	6 x Histidine tag
IDP	Intrinsic disordered protein
II	Instability index
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilodalton
LB	Luria Bertani media
LF	Little finger
MBP	Maltose-binding protein
MMC	Mitomycin C

Mtb	*Mycobacterium tuberculosis*
MW	Molecular weight
MWCO	Molecular weight cut off
Ni-NTA	Nickel nitrilotriacetic acid
NTA	Nitrilotriacetic acid
OD	Optical density
PAD	Polymerase-associated domain
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PI	Isoelectric point
Pol	Polymerase
Pup	Prokaryotic ubiquitin-like protein
PZA	Pyrazinamide
RCF	Relative centrifugal force
RPM	Rotations per minute
RID	RNA interacting domain
RpsA	Ribosomal protein SA
RT	Room temperature
SBRG	Structural biology research group
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography

TB	Tuberculosis
TEMED	N, N, N, N-Tetramethylethylene – diamine
TLS	Translesion synthesis
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
WGS	Whole genome sequencing
WHO	World Health Organization

List of Figures

List of Tables

1. Introduction

1.1. *Mycobacterium tuberculosis* and tuberculosis

1.1.1. Tuberculosis epidemiology

Tuberculosis (TB) is a contagious human disease ranked as one of the most lethal infections alongside HIV/AIDS and malaria (Bloom and Fine, 1994). TB was declared a global health emergency by the World health Organization (WHO) in 1993 and is responsible for the death of ~2 million lives annually (Chen *et al.*, 2009). *Mycobacterium tuberculosis* (Mtb) is the causative agent of TB and is spread via aerosols from an active TB host through contact with infected bodily fluid (Wheeler and Ratledge, 1994; Warner and Mizrahi, 2006). Mtb typically infects the lungs of the human host causing pulmonary TB but is known to infect other body parts causing extra-pulmonary TB (Katz *et al.*, 2007; Chen *et al.*, 2009). During infection, Mtb bacilli permeate and enter the lung alveoli where they are exposed to the host immune response (Brennan and Draper, 1994; Dover *et al.*, 2011). The host immune response either eliminates Mtb or results in a latent/dormant state (Bloom and Fine, 1994). The latent state may be reactivated when conditions are more favourable to Mtb multiplication (Bloom and Fine, 1994). It is estimated that one third of the human populations harbours the dormant state and 8-10 million cases of newly formed active.

1.1.2. Methods in TB drug discovery

New anti-microbial drugs have been identified using either phenotypic or target-led drug discovery methods (Payne *et al.*, 2007; Dover *et al.*, 2011; Hughes *et al.*, 2011). Phenotypic drug discovery involves screening of small compound libraries for the ability to eliminate target pathogens (Dover

et al., 2011; Hughes *et al.*, 2011; Shi *et al.*, 2011). Transposon mutants are used to identify the target of compounds that have shown activity via whole genome sequencing (Dunker *et al.*, 2001; Warner and Mizrahi, 2014). Phenotypic drug discovery successfully identified the compounds bedaquiline and delamanid (Gomez and McKinney, 2004; Payne *et al.*, 2007; Dover *et al.*, 2011). Nevertheless, this method has limitations such as low probability of occurrence during screening periods (Payne *et al.*, 2007).

Target-led drug discovery involves a series of methods to identify a "target" molecule essential for survival and growth of a pathogen (Dunker *et al.*, 2001). Following the identification of the target, high throughput screening methods are used to identify candidate inhibitors (Dunker *et al.*, 2001; Warner and Mizrahi, 2014). This method relies on the ability of the candidate molecule to permeate the cell and successfully inhibit function to trigger cell death (Warner and Mizrahi, 2014).

1.1.3. Target-led TB drugs and their mode of action

Whole genome sequencing of Mtb has identified multiple novel protein targets for target-led drug discovery and development (Hughes et al., 2011) as well as Mtb strains from clinical isolates showing signs of drug resistance (Warner and Mizrahi, 2014). The first-line drug, pyrazinamide (PZA), already a key component of the short-course therapy regimen in the 1980's, was recently found to inhibit the ribosomal subunit RpsA (Njire *et al.;* Shi *et al.*, 2011), essential for ribosome-sparing and protein translation in trans-translation (Shi et al., 2011).

DNA polymerases and accessory proteins are suitable drug target candidates to inhibit microbial growth (Shi et al., 2011). Inhibiting low-fidelity, error-prone DNA polymerases may constitute one way to preventing drug resistance in Mtb. (Dunker *et al.*, 2001; Warner and Mizrahi, 2014).

Error-prone DNA polymerases such as DnaE2 are crucial for the survivability of the Mtb when exposed to DNA damaging agents. Inhibition of DnaE2 without harming the pathogens host can be achieved due to the lack of sequence similarities between it and human DNA polymerase.

1.2. DNA replication

1.2.1. Bacterial DNA replication

Chromosomal replication involves three major components: a polymerase (Pol III in bacteria), a sliding clamp processivity factor (β_2 in bacteria) and a clamp loader (DnaX in bacteria). The latter two components facilitate the continued interaction of the polymerase with the DNA strand (Baker and Bell, 1998; McHenry, 2011c) such that replicative DNA polymerases are only highly processive when the sliding clamp and clamp loader are present (Baker and Bell, 1998; McHenry, 2011a; McHenry, 2011c).

Prior to elongation, ATP-bound DnaX allows the β2 ring to encircle the ssDNA template and chaperones Pol III to the newly bound β2 ring (McHenry, 2011b; McHenry, 2011a; McHenry, 2011c). DnaX is the core of the replisome, dimerizing Pol III and interacting with replicative helicases and primase (McHenry, 2011c; McHenry, 2011b). DnaX increases the rate of replisome formation and of Okazaki fragment synthesis during chromosomal replication (Baker and Bell, 1998; McHenry, 2011a).

The replisome holoenzyme is stabilized by β_2-subunits, the ε-subunit stabilizes the α-subunit increasing its affinity for the primer (McHenry, 2011b; McHenry, 2011a; McHenry, 2011c). Once assembled, the Pol III holoenzyme replicates the entire chromosome or 5 kb without dissociating

(Boshoff *et al.*, 2003; Bruck *et al.*, 2003). The presence of all subunits except β_2 increases processivity (Baker and Bell, 1998; McHenry, 2011b; McHenry, 2011a; McHenry, 2011c).

1.2.2. DNA polymerase III

DNA polymerase III (Pol III) is the main replicative polymerase for both RNA primer extension and chromosomal replication (McHenry, 2011b; McHenry, 2011a). Low-GC Gram-positive bacteria harbor two Pol III namely homologues PolC and DnaE that subtly differ with respect to sub-domain rearrangements but retain the overall polymerase structure (Dervyn *et al.*, 2001; McHenry, 2011b; McHenry, 2011c). PolC has Mg^{2+}-dependent proofreading activity not found in DnaE polymerase families (Bruck *et al.*, 2003)

DnaE is closely related to *E. coli* Pol III in sequence and domain organization (McHenry, 2011a; McHenry, 2011c). In *Bacillus subtilis,* DnaE synthesizes the lagging strand suggesting that PolC is a leading strand polymerase (Sanders *et al.*, 2010).

Table 1: Three classes of DNA polymerase III's found in E. coli, B. subtilis *and* P. aeruginosa *with accompanying accessory factors. (McHenry, 2011b; McHenry, 2011a; McHenry, 2011c)*

Model organism	Proposed Name	Function(s)s	Accessory Factors
E. coli	PolC, α-subunit, DnaE	Extension of RNA primers High speed, high fidelity chromosomal replication	β_2, $DnaX_{cx}$
B. subtilis	PolC	High speed, high fidelity chromosomal replication	β_2, $DnaX_{cx}$
	DnaE	Extension of RNA primers	β_2, $DnaX_{cx}$
P. aeruginosa	DnaE	Extension of RNA primers High speed, high fidelity chromosomal replication	β_2, $DnaX_{cx}$
	ImuC/DnaE2	High speed, Low-fidelity replication Induced mutagenesis	β_2, ImuA, ImuB

In *E. coli*, PolC has a high elongation rate (~500 nt/s) required for leading-strand replication (Blower *et al.*, 1995; McHenry, 2011b; McHenry, 2011a; McHenry, 2011c). DnaE has a much slower elongation rate (~25 nt/s), too slow for leading-strand replication (Blower *et al.*, 1995; McHenry, 2011b; McHenry, 2011a; McHenry, 2011c). DnaE cannot functionally replace PolC (McHenry, 2011a; McHenry, 2011c) while initiation of lagging-strand replication requires both DnaE and a primase (Filee *et al.*, 2002).

PolC actively discriminates against RNA primers based on the wider diameter of the A-form RNA-DNA duplex (McHenry, 2011b; McHenry, 2011a; McHenry, 2011c). In DnaE, by contrast, two β-strands of the thumb domain interact with the minor groove of the RNA-DNA duplex allowing for efficient primer extension (McHenry, 2011b; McHenry, 2011c) suggesting that DnaE is part of the same replicative pathway (McHenry, 2011b; McHenry, 2011a).

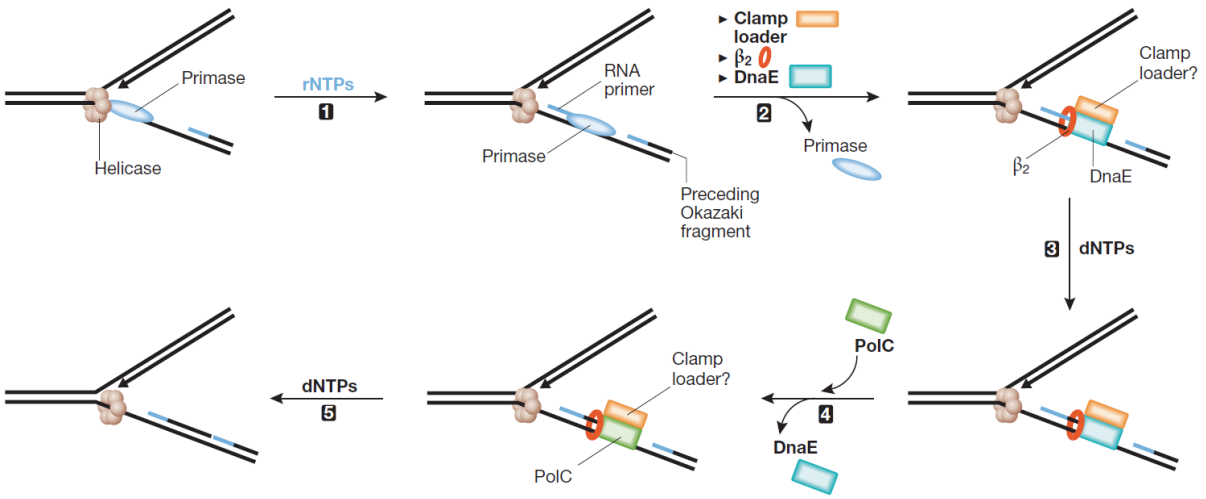


Figure 1: Model for polymerase switching in low GC Gram-positive bacteria.(McHenry, 2011b) After production of theRNA primer, DnaE-β2-clamp loader complex is formed to extend the RNA primer. Once the RNA primer is extended, DnaE is switched with PolC in order to complete the DNA replication process.

In addition to the two replicative DNA polymerase III types, mutagenic Pol III assists with DNA repair and induced mutagenesis (McHenry, 2011b; McHenry, 2011a). Numerous bacterial genomes shows DnaE homologues not involved in standard replication (McHenry, 2011b; McHenry, 2011a; McHenry, 2011c). The Mtb DnaE homologue, DnaE2/ImuC is shown to replace the main mutagenic polymerase, PolV, responsible for induced mutagenesis (Le Chatelier *et al.*, 2004; Warner *et al.*, 2010a; McHenry, 2011b).

1.3. Error-prone DNA replication and mutagenesis

1.3.1. Y-family DNA polymerases

Replication-blocking lesions pose a threat to cell survival despite the multiple DNA repair mechanisms (Collins, 1996). Y-family DNA polymerases have evolved to bypass replication blocking lesions with lower fidelity (Friedberg *et al.*, 2001).

Y-family DNA polymerases are structurally similar to high fidelity, replicative DNA polymerases merely adding a single "little finger" (LF) or "polymerase associated domain" (PAD) (Ling *et al.*, 2001). Y-family DNA polymerases replicate damaged DNA but lack exonucleolytic activity for proofreading (Ohmori *et al.*, 2001).

There are six Y-family DNA polymerase branches: DinB, Rad30A, Rad30B, Rev1, and two UmuC. UmuCs are exclusive to prokaryotes with distinct Gram-positive and -negative branches (Perry *et al.*, 1985; Ohmori *et al.*, 2001). DinB is widely distributed but absent from *S. cerevisiae and D. melanogaster (Ohmori et al., 2001; Godoy et al., 2007).* RAD30 and Rev1 are eukaryotic Y-family polymerases (de Groote *et al.*, 2011), with RAD30(A) or DNA polymerase η ubiquitous to eukaryotes, while RAD30B or polι is exclusive to higher eukaryotes (Ohmori *et al.*, 2001; de Groote *et al.*, 2011)

1.3.2. Y-family DNA polymerase structure

Structurally, Y-family DNA polymerases consist of a catalytic N-terminal and a DNA-binding and -positioning C-terminal domain (Jarosz *et al.*, 2007; Pata, 2010) linked by an unstructured loop (Pata, 2010). The N-terminal domain consists a finger, a thumb and a palm subdomain in analogy

to a human right hand (Trincao *et al.*, 2001; Wei, 2005). Conserved acidic residues responsible for DNA repair are located in the palm sub-domain (Trincao *et al.*, 2001). Y-family DNA polymerases share the same two-metal-ion coordination for nucleotide placement, as standard replicative polymerases (Ling *et al.*, 2001; Pata, 2010). Y-family DNA polymerases have smaller finger and thumb sub-domains resulting in an open and solvent-accessible active site (Wei, 2005; Pata, 2010) increasing the likelihood of accommodating bulky lesions (Wei, 2005; Wong *et al.*, 2008; Pata, 2010).

The C-terminal domain, unique to Y-family DNA polymerases, is referred to as the LF domain in prokaryotes and the polymerase associated domain (PAD) in eukaryotes (Lee *et al.*, 2006; Jarosz *et al.*, 2007; Pata, 2010). The LF/PAD domain follows the thumb sub-domain in sequence but structurally aligns with the finger domain (Wei, 2005; Wong *et al.*, 2008). Despite a lack of conserved residues, the LF/PAD domain it structurally highly conserved in all Y-family DNA polymerases (Ohmori *et al.*, 2001; Wei, 2005)

1.3.3. SOS response in bacteria

The SOS response is composed of transcriptional and post-transcriptional regulatory elements to detect and repair damaged DNA (Krishna *et al.*, 2007). The SOS response is best understood in *E. coli* - the standard model for bacteria (Krishna *et al.*, 2007).

The *E. coli* SOS response is activated by damaged DNA in the form of a single stranded lesion (Frank *et al.*, 1996; Belov *et al.*, 2009). RecA/ATP binds the damaged DNA to form a helical nucleoprotein filament and the active RecA* (Giese *et al.*, 2008). RecA* is a co-protease assisting the autocatalytic cleavage of the repressor LexA (Giese *et al.*, 2008; Belov *et al.*, 2009) allowing

the gene *umuDC* to be expressed producing UmuC and UmuD (Patel *et al.*, 2010). UmuD is activated to UmuD' by RecA* removing the 24 N-terminal amino acids (Frank *et al.*, 1996). Depending on the activity of RecA* different proportions of homodimers $UmuD_2$ and $UmuD'_2$ or heterodimers UmuD'D will form with distinct affinities for UmuC (Patel *et al.*, 2010; Kuban *et al.*, 2012; Belov *et al.*, 2013). The complex $UmuD'_2C$ is catalytically active while UmuDD'C and $UmuD_2C$ are not (Belov *et al.*, 2009). $UmuD'_2C$ inhibits non-mutagenic DNA replication permitting translesion synthesis (TLS) (Patel *et al.*, 2010). $UmuD'_2C$, RecA*, single-stranded binding proteins and subunits of Pol III form a polymerase V mutational complex (Pol V) for TLS (Sassanfar and Roberts, 1990; Smith and Walker, 1998; Patel *et al.*, 2010).

1.4. Induced mutagenesis proteins involved in translesion synthesis

1.4.1. ImuABC cassette

The *imuABC* cassette found in all forms of life is critical to DNA replication and mutagenesis (Erill *et al.*, 2006; Ippoliti *et al.*, 2012). The original annotations *sulA*, *dinP* and *dnaE* were renamed *imuA*, *imuB* and *imuC* to reflect their re-classification as an induced mutagenesis cassette (Abella *et al.*, 2007; Ippoliti *et al.*, 2012)

DnaE2 or ImuC is the main catalytic product of the *imuABC* cassette (Ippoliti *et al.*, 2012). The encoding gene *dnaE2* of Mtb is homologous to *E. coli dnaE* despite not being essential to DNA replication (Tippin *et al.*, 2004; Ippoliti *et al.*, 2012). The gene *dnaE2* often groups with another two genes, *imuB* and *imuA*, resulting in its renaming to *imuC* (Ippoliti *et al.*, 2012).

The *imuB* encoded protein ImuB was proposed to be structurally related to Y-family DNA polymerases (Ippoliti *et al.*, 2012). However, ImuB lacks the catalytic aspartic acid residues

normally required for TLS implying it to be catalytically inactive (Koorits *et al.*, 2007; McHenry, 2011b). In *M. tuberculosis* ImuA', a homologue of ImuA, co-exists with ImuB and DnaE2 (Warner *et al.*, 2010b; McHenry, 2011b). A homologue of *imuABC* cassette accessory proteins are found in *D. deserti* but lack similarity to known *imuB* and *imuA* gene (Ippoliti *et al.*, 2012). This homologue was annotated *imuY* due to its involvement in TLS and its sequence similarity to Y-family polymerases (Ippoliti *et al.*, 2012).

The complete *imuABC* cassette is not found in cyanobacteria or Gram-positive bacteria but can be found as single genes regulated by different *lexA* genes (Ippoliti *et al.*, 2012). Many variation of the *imuABC* cassette occur in different organisms. For an overview see Figure 2.

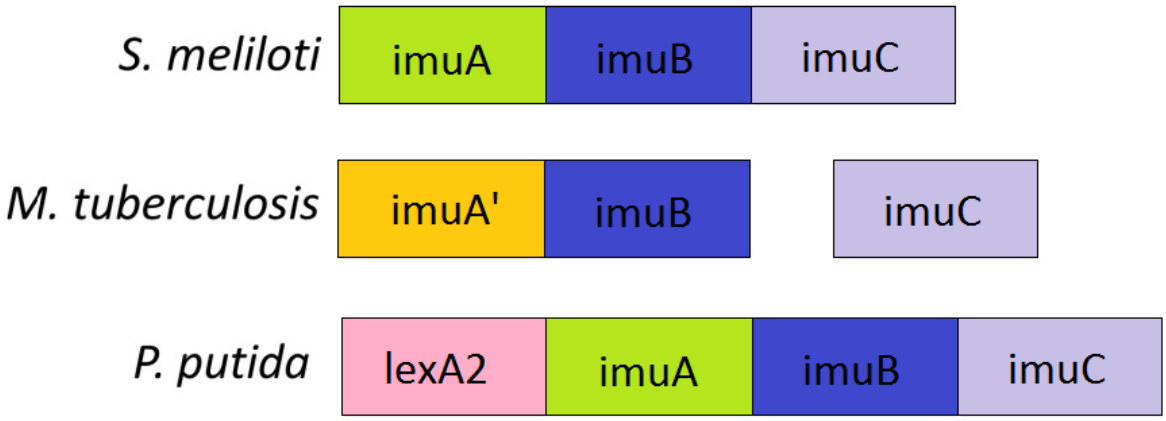


Figure 2: Configurations of the imuABC cassette in S. meliloti, M. tuberculosis *and* P. putida. S. meliloti *has a full, contiguous* imuABC *cassette regulated by the* lexA-*SOS system.*M. tuberculosis *has a split* imuABC *cassette with a variation of* imuA *annotated* imuA'. P. putida, *by contrast, has a secondary* lexA *gene annotated* lexA2.

1.4.2. *Mycobacterium* imuA'BC cassette

Mtb possesses all components for a functional SOS response mechanism (Ippoliti *et al.*, 2012). In addition mycobacterium shows evidence of hyper mutability and adaptive mutation systems

(Durbach *et al.*, 1997; Mizrahi and Andersen, 1998). The identification of putative Y-family homologues in mycobacterium indicates controlled mutagenic processes being present (Warner *et al.*, 2010a; Ippoliti *et al.*, 2012). Nevertheless, several studies failed to demonstrate that Y-family homologues are required for DNA damage-response in mycobacterium (Boshoff *et al.*, 2003; Rand *et al.*, 2003).

The Mtb gene *Rv3370c* that encodes DnaE2/ImuC is upregulated more than 10-fold when cells are exposed to mitomycin C (MMC), a potent DNA cross-linking agent (Koorits *et al.*, 2007; Warner *et al.*, 2010b). Delta *Rv3370c* mutants show reduced virulence compared to wild type strains (Warner *et al.*, 2010b) and are UV-mutagenesis deficient suggesting that DnaE2/ImuC is directly involved in induced mutagenesis similar to TLS (Boshoff *et al.*, 2003; Ippoliti *et al.*, 2012). Mtb DnaE2/ImuC has three catalytic aspartate residues similar to those of C-family polymerases. Substitution of the catalytic residues eliminates UV-induced mutagenesis and renders the cell hypersensitive to MMC treatment - mimicking the *dnaE2*-deleted mutant and suggesting that DnaE2/ImuC is responsible for UV-induced mutagenesis and survival under DNA-damaging stress conditions (Boshoff *et al.*, 2003; Koorits *et al.*, 2007; Ippoliti *et al.*, 2012).

DnaE2/ImuC requires additional proteins for induced mutagenesis (Warner *et al.*, 2010a; McHenry, 2011b) demonstrated by elevated levels of DnaE2 not leading to an increased cellular mutation rate (Ippoliti *et al.*, 2012). The genes of ImuA' and ImuB are located approximately 24.7 kB upstream from *dnaE2* on a separate locus (Warner *et al.*, 2010b).

The expression of the *imuA'BC* cassette in *Mycobacterium* is controlled by LexA, as it is in other organisms (McHenry, 2011b). Deleting any *imuA'BC* cassette gene results in comparable hypersensitivity to MMC treatment suggesting that all components are part of the same pathway (Boshoff *et al.*, 2003; Koorits *et al.*, 2007; Zeng *et al.*, 2011; Ippoliti *et al.*, 2012).

1.4.3. *Mycobacterium* ImuB

ImuB homologues form a distinct branch within the UmuC subfamily of Y-family DNA polymerases and appear to share all structural features of Y-family DNA polymerases including the defining C-terminal domain (Koorits *et al.*, 2007; Warner *et al.*, 2010b; Zeng *et al.*, 2011; Ippoliti *et al.*, 2012) as well as a β-clamp binding motif in the loop linking the N- and C-terminal domains (Warner *et al.*, 2010b).

Assumed to have a Y-family DNA polymerase structure, sequence analysis and homology modelling suggest ImuB lacks catalytic residues rendering it inactive for TLS and DNA repair (Warner *et al.*, 2010b) – in common with other ImuB homologues. ImuB would therefore appear to be an accessory protein assisting DNAE2 in TLS and DNA repair (Warner *et al.*, 2010b).

The deletion of the ImuB-encoding gene *Rv3394c* eliminates UV-mutagenesis comparable to the ImuC-deletion phenotype (Warner *et al.*, 2010a). This is similarly true for the protein ImuA' implying that both ImuB and ImuA' are essential to induced mutagenesis and that ImuA' is a homologue of ImuA (Warner *et al.*, 2010b; Ippoliti *et al.*, 2012).

ImuB readily interacts with the Mtb β-clamp supporting its accessory role in induced mutagenesis (Warner *et al.*, 2010b; Ippoliti *et al.*, 2012). As the only cassette component that interacts with the β-clamp ImuB contains a β-binding motif ^{354}QLPLWG^{359} of the DinB3 family (Warner *et al.*, 2010b). Replacing these residues eliminates the interaction to the β-clamp (Warner *et al.*, 2010b; Ippoliti *et al.*, 2012). The β-binding motif is a highly conserved in ImuB homologues implying it may be responsible for polymerase switching (Ippoliti *et al.*, 2012).

ImuB has been shown to interact with DnaE1/ImuC and ImuA' – and is, in fact, the only interaction partner known for ImuA' (Warner *et al.*, 2010b). ImuB is correspondingly a candidate for roles

such as polymerase switching and β-clamp association (Warner *et al.*, 2010b). Binding of cassette components generally involves the ImuB C-terminal domain except for the β-clamp motif (Warner *et al.*, 2010b). ImuB lacking the C-terminal domain no longer binds other cassette components interacts with the β-clamp (Warner *et al.*, 2010b).

1.5. Aims of the study

This study was a collaboration between Professor Wolf-Dieter Schubert of the Structural Biology Research Group (SBRG), Department of Biotechnology, University of the Western Cape (UWC) and Prof Digby Warner of Institute of Infectious Diseases and Molecular Medicine (IDM), University of Cape Town (UCT).

The aim of the study was to determine the three-dimensional structure of the induced mutagenesis protein B (ImuB) from *Mycobacterium tuberculosis* using X-ray crystallography. The objective for achieving the atomic structure is to provide more insight into how induced-mutagenesis in Mtb works at the molecular level. ImuB plays a role in the process of DNA repair and induced mutagenesis as it is believed to be the center of interaction between the damaged DNA molecule and all the components of the mutagenic cassette (Warner *et al.*, 2010b; McHenry, 2011b).

To achieve the aim it was necessary to 1) generate a three-dimensional homology model of ImuB, 2) generate a large-scale expression system for ImuB production in *E. coli*, 3) produce and purify ImuB, 4) crystallize and solve the structure of ImuB.

For expression, the gene of interest was cloned into commercial expression vectors namely pET, pGEX, and pMal. To obtain pure ImuB, defined production and purification methods had to be established to obtain large amounts of pure for crystallization. Purification strategies were to include amylose and immobilized metal-ion affinity as well as size exclusion chromatographies.

The hanging-drop vapour diffusion method was used during the entire crystallization experiment process.

2. Materials and Methods

2.1. Standard materials

2.1.1. General chemicals, enzymes, kits, resin and standards

If not stated otherwise, all chemicals used were purchased in "pro analysis" grade from the following companies: Fermentas, Fluka, GE-Healthcare, Invitrogen, Millipore, Promega, Roche, New England Biolabs, Finnzymes, BioRad, Novagene and Qiagen.

Table 2: Table of Enzymes

Product	**Supplier**
T4 DNA Ligase	Fermentas
FastDigest BglII restriction enzyme	Fermentas
FastDigest EcoRI restriction enzyme	Fermentas
FastDigest NcoI restriction enzyme	Fermentas
FastDigest XhoI restriction enzyme	Fermentas
FastDigest HindIII restriction enzyme	Fermentas
Factor Xa protease	New England Biolabs
3C Precession protease	UWC
TEV protease	UWC
DNAse I	Roche

Table 3: Table of kits

Product	Supplier
GeneJET gel extraction kit	Fermentas
GeneJET DNA purification kit	Fermentas
GeneJET plasmid miniprep kit	Fermentas
Phusion high fidelity PCR kit	Finnzymes

Table 4: Table of molecular weight standards

Product	Usage	Supplier
PageRuler unstained ladder	SDS-PAGE	Fermentas
Precision plus protein all blue standard	SDS-PAGE/Western Blot	BioRad
GeneRuler® 1Kb DNA ladder	Agarose gel electrophoresis	Fermentas
Unstained protein molecular weight marker	SDS-PAGE	Fermentas

Table 5: Table of resin and reagents

Products	Supplier
Amylose/agarose resin	New England Biolabs
Bugbuster protein extraction reagent	Novagen
Ni-NTA agarose	Qiagen
SnakeSkin dialysis tubing	Bio-Rad Laboratories

2.1.2. Buffers

All buffers below were prepared to their respective volumes using distilled water purified by the Millipore water purification system (Prima) unless stated otherwise. pH was adjusted using a 2^0 pH meter (Crison Instruments, Barcelona, Spain) using 3 M sodium hydroxide (NaOH) and 2 M hydrochloric acid (HCl) unless stated otherwise.

1 M IPTG: 1 M IPTG were prepared in distilled water and sterilized by filtration.

2 M Tris-HCl: 2 M Tris-HCl pH 7.2 was prepared distilled water. The pH of the buffer was adjusted using hydrochloric acid (HCl) and stored at room temperature (RT).

10% Sodium dodecyl sulphate (10% SDS): 10% SDS was prepared in distilled and stored at RT.

5 x SDS-running buffer: 5 x SDS-running buffers were prepared using 25 mM Tris pH 8.2, 250 mM glycine and 0.1% SDS in distilled water. The buffer is diluted to 1:5 with distilled water before use.

5 M NaCl solution: 5 M NaCl stock solution was prepared in distilled water to a final volume of 1 L and stored at room temperature.

50 x TAE buffer: 50 x TAE stock was prepare using 242 g/L of tris-base, 57.1 ml/L of glacial Acetic acid and 18.6 g/L of EDTA in distilled water. The buffer is diluted to 1:50 with distilled water before use.

500 mM Disodium ethylene-diamine tetra-acetate (EDTA): 500 mM EDTA was in distilled water. The pH of the solution was adjusted using NaOH to a final pH of 8.0 and stored at room temperature.

8 x SDS-Sample buffer: 2.2 mL of 0.5 M Tris-HCl pH 6.8, 4 mL of 87% (v/v) glycerol, 16 mL of 10% (w/v) SDS, 200 µL of β-mercaptoethanol and 0.2 g of bromophenol blue. This is then stored in 1.5 mL microcentrifuge tubes at room temperature

Ammonium persulphate (25% APS): 25% (w/v) stocks of ammonium persulphate was prepared in distilled water and aliquoted into 1.5mL microcentrifuge tubes for storage at -20°C

Ampicillin: 100 mg/mL stocks of ampicillin was prepared using distilled water and aliquot into 1.5 mL microcentrifuge tubes for storage at -20°C

Column buffer: 20 mM Tris-HCl pH 7.4, 200 mM NaCl, and 1 mM EDTA

Coomassie staining solution: 30% (v/v) ethanol, 10% (v/v) acetic acid, 0.25% (w/v) coomassie Blue R-250

Destaining solution: 40% (v/v) ethanol, 10% (v/v) acetic acid

Kanamycin: 25mg/ml stocks of kanamycin was prepared in distilled water and aliquot into 1.5 mL microcentrifuge tubes for storage at -20°C.

Luria Bertani medium: 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl. Media was supplemented with either ampicillin or kanamycin depending on the desired expression system.

Lysis buffer: 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 0.5 mM EDTA, 1% triton X-100–, 100 µg/mL of DNase and 50 mg/mL lysozyme

Separating buffer for SDS-PAGE: 1.5 M Tris-HCl pH 8.8 was prepared in distilled water and stored at RT.

Super optimal broth (SOB): 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl and 10 mM $MgCl_2$

Super optimal broth with catabolite repressor (SOC): Super optimal broth + 20 mM filtered glucose.

Stacking Buffer for SDS-PAGE: 0.5 M Tris/HCl (pH 6.8) was prepared in distilled water and stored at RT.

2.1.3. Crystallization screens

The following commercial screens were used to screen for crystallization conditions.

Table 6: Crystallization screens

Name	Supplier
JCSG Core I Suite	Qiagen
JCSG Core II Suite	Qiagen
JCSG Core II Suite	Qiagen
JCSG Core IV Suite	Qiagen

2.1.4. Equipment

Table 7: List of equipment

Device	Model	Company
Autoclave	J.S.D. 400	Hospi
Water purification system	Prima	ELGA
Electrophoresis	Mini-PROTEAN®	Bio-Rad Laboratories
Electrophoresis power pack	Power Pac™ basic	Bio-Rad Laboratories
PCR machine gene amp® PCR	System 2700	Amersham Biosciences
Shaking incubator	Multitron II	INFORS

Shaker (roller)	Roller mixer	SRT6 Stuart®
Sonication device	Unknown	MSE
Centrifuges	Sorvall® RC6	Thermo Scientific
	5417C	Eppendorf
	Heraeus fresco17 centrifuge	Thermo Scientific
Nanodrop photometer	ND-1000 Spectrophotometer	peQLab
pH meter	Basic 20 pH meter	Crison Instruments
Nanolitre Pipette	Mosquito 4B	TTP LabTech
Incubator	Function line	Heraeus Instruments
Heating block	Thermomixer 5436	Eppendorf

2.2. Bioinformatics analysis of ImuB

2.2.1. Sequence alignment and homology modelling of ImuB

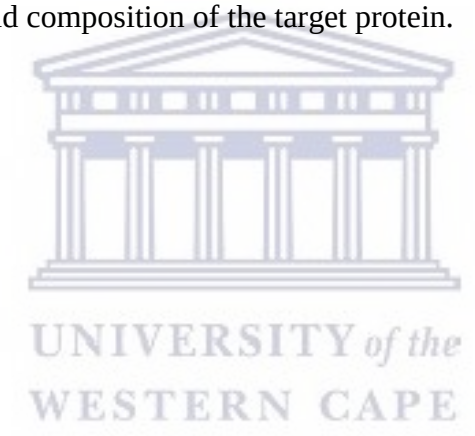
Homology modelling is the derivation of a three-dimensional model of "target' protein based on experimentally structures of homologous proteins.

A homology model of ImuB was constructed using the crystal structure of DinB from *S. solfataricus* and the crystal structure of Dpo4 in complex with DNA duplex from *S. solfataricus.* The amino acid sequence of ImuB was aligned to homologous sequences using PROMALS3D multiple sequence alignment and structure server (http://prodata.swmed.edu/promals3d/promals3d.php). PROMALS3D aligns sequences using sequence data and structural similarities based on predicted secondary structure elements. The sequence alignment was improved manually using UCSF chimera sequence alignment interface (https://www.cgl.ucsf.edu/chimera/).

The structural model of ImuB was constructed using the sequence alignment and crystal structures of the homologues in Modeller in the UCSF chimera interface. The produced model was refined to an energy minimized state using Zhang Labs High-Resolution Protein Structure Refinement servers (http://zhanglab.ccmb.med.umich.edu/ModRefiner/).

2.2.2. Disordered region prediction

Three distinctive disorder region prediction servers predicted disordered regions within the ImuB sequence. The servers used were DISOPRED (http://bioinf.cs.ucl.ac.uk/index.php?id=806), DisEMBL (http://dis.embl.de/) and DisProt (http://www.disprot.org/) which predicts disordered regions bases on the amino acid composition of the target protein.



2.3. Cloning strategies to clone Rv3394c into various protein expression vectors

Molecular cloning is a set of molecular techniques used to transfer a DNA fragment into a transferable vector, transfer it into a host organism to *inter alia* produce the desired protein. The resulting recombinant DNA molecule is usually in the form of a plasmid vector when replication occurs in a bacterial host cell.

The cloning strategies are briefly summarized in figure 3. Amplification of the *Rv3394c* gene using Phusion HF polymerase, was done with the use of the cycle conditions in table 8.

Table 8: Polymerase chain reaction (PCR) amplification conditions using Phusion HF DNA polymerase

PCR step	**PCR conditions**
Initial denaturing	98°C for 30s
Three step cycling (25 cycles)	98°C for 7s 58°C for 15s 72°C for 45s
Final elongation	72°C for 7min

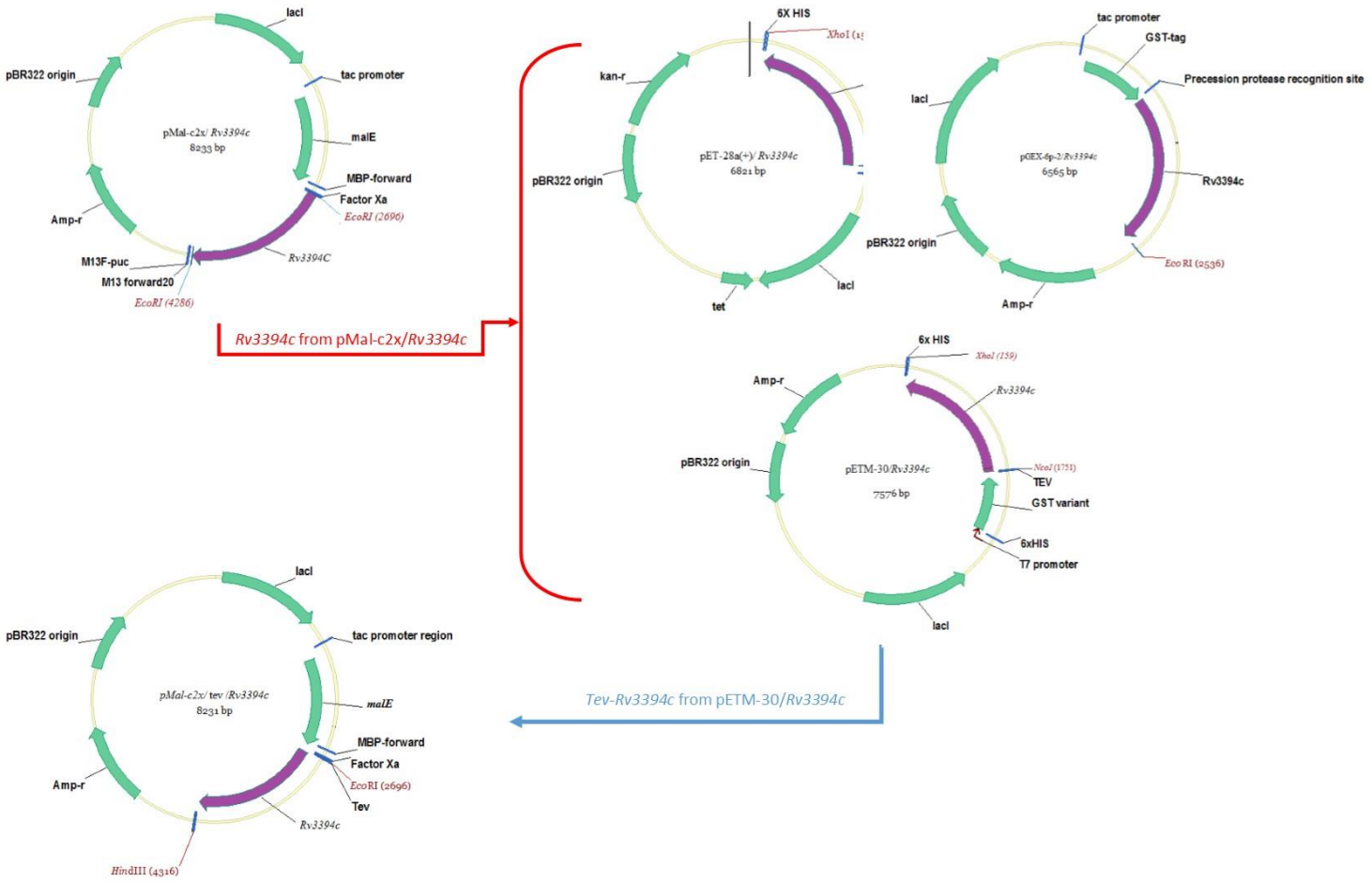


Figure 3: Cloning strategies used to produce recombinant plasmids for ImuB expression. The Rv3394c gene was amplified from the pMal-c2/Rv3394c construct (A). Rv3394c amplicons were a cloned into pET-28a(+) , pGEX-6p2 and pETM-30 for expression in E.coli expression systems. Tev/Rv3394c was amplified from the produced pETM-30/Rv3394c for insertion into the pMal-c2x expression vector to incorporate tev as the protease cut site.

2.3.1. Cloning strategy for inserting Rv3394c into pGEX-6P-2 expression vector

Full length *Rv3394c* was amplified from the pMal-c2X-*Rv3394c* construct (provided by Dr. Digby Warner) using primers

BglII
5'-GCTCATAGATCTGTCATGGCCTCCGCTCGC-3' and

5'-CCTCAGGAACCTTCATTCGTAGCTTCCCTCCAGATACC-3'.

The amplicon was cloned into the pGEX-6P-2 expression vector using restriction enzyme sites, BglII as a BamHI alternative at the 5' end and EcoRI at the 3' end. BglII produces the same sticky ends as BamHI, which could be annealed to the BamHI sticky ends of the host vector. The *Rv3394c* gene contains a BamHI recognition site at 1462 bp. BglII replaced BamHI to eliminate the digestion of the *Rv3394c* gene. The BamHI recognition site was selected to limit the linker length and produce an in-frame fusion protein. The EcoRI recognition site was already found on the host plasmid. The 3' primer was designed to anneal to the 3' end on the host plasmid, incorporating the EcoRI recognition site on the amplicon. The restriction digested amplicon was inserted into the expression vector between the BamHI site and the EcoRI site to generate the pGEX-6P-2-*Rv3394c* expression construct, which is designed to produces ImuB as an N-terminally tagged GST fusion protein.

2.3.2. Cloning strategy for inserting Rv3394c into pET expression vectors and tev-Rv3394c into pMal-c2X expression vector

Full length *Rv3394c* was amplified from the pMal-c2X-*Rv3394c* construct using primers

NcoI

5’ -GTCATACCATGGGAGTGATGGCCTCCGCTCGC- 3’ and

XhoI

5’ - CCTCAGCTCGAGTTCGTAGCTTCCCTCCAGATACC- 3’.

The amplicon for each respective construct was cloned into pETM-30 and pET-28a (+) expression vectors using the respective restriction enzymes sites. Generation of the pETM-30-*Rv3394c* expression construct, which produces GST-ImuB-His_6. pET-28a(+)-*Rv3394c* expression construct was designed to produce ImuB as an C-terminally tagged His_6.

Rv3394c and TEV protease recognition region was amplified from the pETM-30-Rv3394c construct using primers

EcoRI TEV

5’ -CAACCTCGGGGAATTCGAGAATCTTTATTTTCAGGGC- 3’ and

HindIII

5’ -CCTCAGAAGCTTTCATTCGTAGCTTCCCTCCAGATACC-3’.

The TEV-Rv3394c amplicon was cloned into the pMal-c2x vector to produce an in frame MBP-tev-imuB fusion protein. The addition of the tev protease site was done due impotency of factorXa protease to cleave MBP-ImuB and due to the ready availability of the in-house produced tev protease.

2.4. Transformation of *E. coli*

Transforming an organism like a bacterium involves introducing foreign DNA often as a plasmid into the host cell to allow its amplification and transcription. The surface of the cell is changed to allow the introduction of the recombinant DNA using external stimuli to increase the cells competence to take up foreign DNA.

2.4.1. Electro-competent cell preparation

Colonies of *E. coli* BL21, Rosetta-2 and GeneHogs were inoculated into 10 mL of SOB media in a 125 mL Erlenmeyer flask and incubated for 16-18 h at 37°C. 250 mL of pre-warmed SOB media in a 1 L flask was inoculated with 2 ml of the overnight pre-culture and allowed to grow to an optical density of 0.5- 0.7 at 600 nm. Prior to harvesting, the cells were incubated on ice for 15 min. Cells were harvested at 2800 rcf for 10 min at 4°C. The supernatant was discarded and the cells were allowed to dry on ice to remove access media. The cells were resuspended in 250 mL of chilled 10% (v/v) glycerol. The solution was centrifuged at 2800 rcf for 10 min at 4°C and the supernatant is discarded. Cells were resuspended in 250 mL of chilled 10% (v/v) glycerol and aliquot to a final volume of 50 µL. Samples were flash frozen using liquid nitrogen and stored at -80°C until future use.

2.4.2. Transformation of *E. coli* by electroporation

E. coli GeneHogs, BL21 and Rosetta-2 cells were transformed by electroporation. The electro-competent cells were incubated with 100ng of plasmid DNA for 10 min on ice. Samples were aliquoted into specialized electroporation cuvettes and pulsed with 2.5 kV using an electroporator. The electro-transformed samples were allowed to recover in 1 mL of pre-warmed SOC media for 1 h at 37°C, before being plated on LB agar with the required antibiotic.

2.5. Protein analysis

2.5.1. SDS-polyacrylamide gel electrophoresis

Protein samples were analysed using SDS-PAGE, which separates proteins by their molecular weight in an electric field. The proteins are denatured by boiling and by addition of a reducing agent. The sodium dodecyl sulphate (SDS) masks the natural charge of proteins and introduces a net negative charge proportional to the protein size. The protein samples are concentrated in the upper stacking gel which differs from the separating gel in pH, ionic strength, size and pore density.

10 µL of protein samples were mixed with 3 µL of 8 x sample buffer following incubation of the mixture for 5 min at 96°C to ensure complete denaturation. The samples were pipetted into the stacking gel pockets SDS-PAGE run at 40 mA for 30 min. After the run the gel was stained for 15 min in staining solution and excess stain removed by incubation of the gel in de-staining solution. Gels were air dried in between cellophane sheets in Perspex frames for conservation and storage.

12% separating gel:	12 mL acrylamide/bisacrylamide 30% (w/v)/0.8% (w/v)
	7.6 mL 1.5 M Tris pH 8.8
	300 µL 10% (w/v) SDS
	10 mL H_2O
	40 µL TEMED
	100 µL 25% (w/v) APS
5% stacking gel:	1.5 mL acrylamide/bisacrylamide 30% (w/v)/0.8% (w/v)
	2.5 mL 0.5 M Tris pH 6.8
	5.9 mL ddH_2O
	15 µL TEMED
	25 µL 25% (w/v) APS
8x sample buffer:	16 mL 10% SDS
	4 mL glycerol
	2.2 mL Tris pH 6.8
	800 µL β-mercaptoethanol
	1 spatula tip bromophenol blue
Staining solution:	2.5 g Coomassie Brilliant Blue R-250
	300 mL ethanol
	100 mL acetic acid
De-staining solution:	400 mL ethanol
	100 mL acetic acid

2.5.2. Concentrating protein solutions

Samples of MBP-ImuB and GST-ImuB-His_6 fusion protein were concentrated by ultracentrifugation using Vivaspin-2, -6- and -20 concentrators (Sartorius, Göttingen, Germany) with a MW cut-off of 10, 25 and 50 kDa. Samples were concentrated at 1000 rcf until the desired concentration was obtained. The protein concentration was determined using a ND-1000 Spectrophotometer by monitoring the absorbance at 280 nm throughout the concentration process. The process was stopped when the desired protein concentration of 25 mg/ml was reached.

2.6. Protein production and cell disruption

2.6.1. Small-scale protein production

ImuB was produced on a small scale from each of the cloned constructs in *E.coli*, to determine optimal expression conditions. A maximum volume of 25 mL was used to limit wastage. The cells containing the construct of interest were inoculated with 2 mL of pre-culture and allowed to grow to a desired OD_{600nm} of 0.4 – 0.6. The incubation temperature were varied between 15°C - 25°C to alter the rate of protein production. The culture was induced with concentration of IPTG (20 µM to 500 µM). The bacterial culture was incubated overnight and samples were taken before induction, 2 and 4 h after induction, and overnight. Cells were harvested at 4000 rcf for 30 min and resuspended in Bugbuster protein extraction reagent. For every OD_{600} = 0.1 in 100 mL, 100 µL of Bugbuster protein extraction reagent would be added. The resuspended sample was incubated at room temperature for 30 min to rupture the cells. The cell lysate of each sample was centrifuged to separate the soluble fraction from the insoluble material. Both soluble and insoluble fractions were analyzed using SDS-PAGE.

2.6.2. Soluble MBP-ImuB production

The pMal-c2X-*Rv3394c* vector with an ampicillin (Amp) resistance gene, was used to produce N-terminal MBP-ImuB fusion protein. The pMal-c2X-Rv3394c construct was transformed into *E. coli* BL21 cells. 1 L of plasmid-bearing bacteria was incubated at 37°C and 180 rpm in LB-amp broth until the culture reached an OD_{600nm} of 0.4 - 0.6. The incubation temperature was decreased to 25°C to reduce the rate of protein production and the culture induced with IPTG to a final concentration of 0.5 mM. The bacterial culture was incubated overnight and harvested by centrifugation at 4000 rcf for 30 min and resuspended in 30 mL of lysis buffer. The resuspended sample was incubated at room temperature for 30 min and ruptured via sonication. Cells were exposed to sonication for 20 s and rested for 20 s intervals for up to four cycles. The ruptured cells were centrifuged at 16000 rcf for 1 h to separate the insoluble material from the soluble materials, which contained the MBP-ImuB fusion protein.

2.6.3. Insoluble ImuB-His6 production

The pET-28a(+)-*Rv3394c* construct bears a kanamycin (Kan) resistance gene and produces a C-terminal His_6-ImuB fusion protein after transformation of *E. coli* BL21 cells. 1 L of plasmid-bearing bacteria was incubated at 37°C and 180 rpm in LB-amp broth until the culture reached an OD_{600nm} of 0.4 - 0.6. The incubation temperature is then decreased to 30°C and the culture is induced with IPTG to a final concentration of 0.1 mM. The bacterial culture was allowed to grow overnight and harvested at 4000 rcf for 30 min. Cells were resuspended in 30 mL of lysis buffer and incubated at RT for 30 min prior to sonication. The ruptured cells were centrifuged at 16000 rcf for 1 h to separate the insoluble material from the soluble material. The insoluble material contained the insoluble ImuB-His_6 fusion protein.

2.7. Chromatographic methods

2.7.1. Immobilized metal affinity chromatography (IMAC) of solubilized ImuB-His6

The pET-28a(+)-*Rv3394c* produces a C-terminally His_6-tagged ImuB fusion protein to assist with purification via IMAC. Histidine readily binds to immobilized metal ions through its imidazole ring.

Prior to purification, a 10 µL sample of the resin is taken for SDS-PAGE analysis to ensure that the resin does not have bound proteins. The cell lysate containing solublilzed ImuB-His_6 fusion protein, is allowed to couple with Ni-NTA resin for 1 h at room temperature. 3 mL of Ni^{+2}-NTA resin was equilibrated with 5 column volumes (CV) of column buffer prior to coupling of the soluble cell lysate. Coupling of the cell lysate to the Ni^{+2}-NTA resin was done at 4°C. After coupling, the unbound cell lysate is collected as the flow-through and resin was washed with 3x 100 mL wash buffer. Wash buffer fraction were collected and tested on SDS-PAGE. 20 µL of resin was taken after the wash step for later analysis via SDS PAGE. The fusion protein was eluted using column buffers with imidazole concentrations of 10 mM, 20 mM, 50 mM, 80 mMand 100 mM. ImuB-His_6 was eluted at an approximate imidazole concentration of 40 mM. Eluted protein samples were collected in fractions of 15 mL each. 20 µL of each sample was set aside for analysis by SDS-PAGE.

2.7.2. Amylose affinity chromatography of soluble MBP-ImuB fusion protein

The cell lysate containing soluble MBP-ImuB fusion protein was incubated with the amylose resin for 1 h at 4°C temperature. The MBP-tag binds amylose with high affinity. 3 mL of amylose resin was equilibrated with 5 CV of column buffer. 20 µL of the resin was collected for later analysis via SDS PAGE. The cell lysate was incubated with the amylose resin on a roller mixer overnight at 4°C. After coupling of the MBP-ImuB to the amylose resin, the unbound cell lysate is collected as the flow-through. The amylose resin was washed with 3x 100 mL each with wash buffer. The wash fractions were collected and analysed on SDS-PAGE. 20 µL of resin are taken after the wash step for later analysis via SDS PAGE. The fusion protein was eluted using a gradient concentration of maltose between 5 mM and 50 mM. Protein was eluted at an approximate maltose concentration of 10 mM. Eluted protein samples were collected in fractions of 15 mL each. 20 µL of each sample was collected for analysis on SDS-PAGE. A final resin sample was collected after the elution of the protein. This was done to ensure that the resin does not contain any of the fusion protein

2.7.3. Size exclusion chromatography

Size exclusion chromatography (SEC) separates protein based on their hydrodynamic volume. The column matrix consists of porous agarose beads with irregular pores with a diameter up to 34 µM. During SEC, larger molecules are excluded from the pores, which reduces their accessible volume. This results in earlier elution times. Smaller particles, experience larger assessable volumes due to their accessibility of the pores on the column matrix. Smaller particles elute later due to the increased retention time in the porous matrix. MBP-ImuB was purified using a gel filtration column (Superdex 200 16/60, GE Healthcare, Pittsburgh, USA) and 1x PBS as the running buffer.

The protein solutions were collected and pooled, the pooled samples were concentrated to a maximum of 15mg/mL prior to crystallization experiments.

2.8. Protein refolding

2.8.1. Washing of insoluble protein

The insoluble ImuB-His_6 protein pellet was purified using three wash steps. During the first step, the insoluble pellet is resuspended in buffer containing 1-2% Triton X-100 to solubilizes and hence remove membrane fragments and membrane. Following ultracentrifugation at 16000 rcf, the supernatant containing solubilized membrane/membrane proteins were discarded.

The insoluble pellet was resuspended in a buffer containing 10 mM tris pH 7.4 and 1 M NaCl to solubilize negatively charged oligonucleotides such as DNA and RNA. In the final step, the pellet was washed with standard buffer containing 10 mM tris pH 8.0, 50 mM NaCl, and 8 M Urea to denature and solubilize all remaining protein. At each step, a sample was taken to analyze on SDS-PAGE.

2.8.2. Refolding using dialysis

Dialysis is a method of changing buffer components by allowing selective diffusion through a semi-permeable membrane (figure. 4). Here, dialysis was used to potentially allow the refolding of a target protein – here ImuB – by gradually removing urea as the denaturation agent. The protein within the dialysis tube is dissolved in high concentrations of urea. The tube is successively placed in buffers with stepwise lower urea concentrations, allowing for the gradual removal of the denaturant. Under ideal conditions, this procedure may allow for the refolding of denatured and hence unfolded protein to an energetically minimal “native” state.

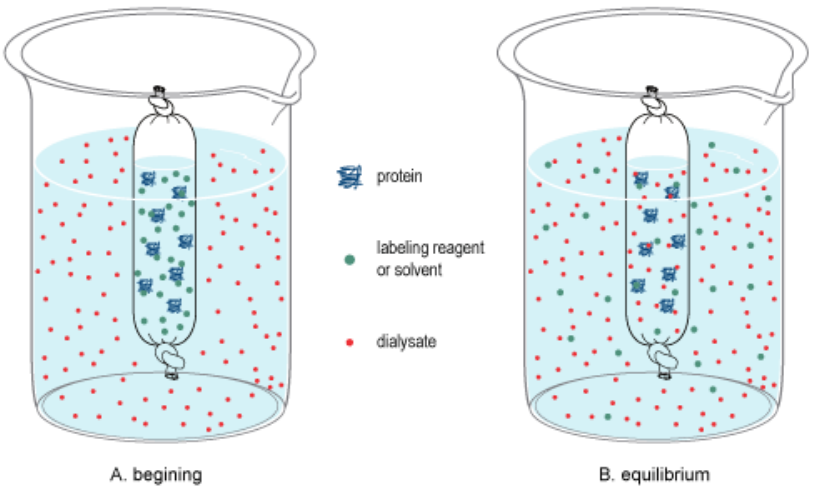


Figure 4: Principle of dialysis ("Dialysis Figure" by Potcherboy at English Wikipedia. Licensed under CC BY 3.0 via Commons - https://commons.wikimedia.org/wiki/File:Dialysis_Figure.png#/media/File:Dialysis_Figure.png). During the dialysis process, molecules diffuse through semi-permeable membrane to produce an equilibrium between molecules in the buffer and the solution within the dialysis tubing.

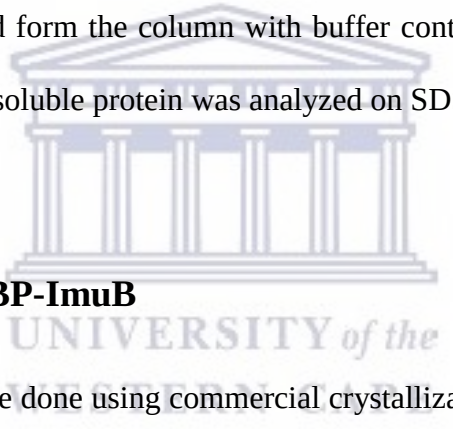
Solubilized target protein was transferred to a Bio-Rad SnakeSkin dialysis tube with a 3500 MWCO. The target protein-containing dialysis tube was placed in 2 L of the buffer containing 50 mM Tris pH 8.0 and 50 mM NaCl contain 6 M urea and allowed to incubate for 4 h at 4°C to allow the gradual removal of urea. The buffer was replaced with buffer containing 3 M urea and allowed to incubate for another 4 h at 4°C. The final buffer change was done using a buffer containing no urea. The sample in the dialysis tubing was analyzed using SDS-PAGE at each interval.

2.8.3. IMAC on-column refolding

The solubilized protein was purified using an IMAC column containing 2 mL of Ni-NTA resin. The sample was washed with buffer containing 8 M urea and 5 mM imidazole. A resin sample was collected after the wash step for SDS-PAGE analysis.

The immobilized protein was washed with gradient wash steps, which decreased the urea concentration. The slow removal of the protein denaturant should in principle allow parameters such as hydrophobicity and bond formation to refold the protein to its native, energy minimalized state.

The refolded protein was eluted form the column with buffer containing 50 mM imidazole and 0.1% Triton X-100. The eluted soluble protein was analyzed on SDS-PAGE.



2.9. Crystallization of MBP-ImuB

Crystallization experiments were done using commercial crystallization screens (Qiagen, Hilden, Germany) in 96-well sitting drop plates. 400 ηL drops composed of equal amounts of protein- and crystallization solution were pipetted using the Mosquito 4B nanolitre pipetting robot (TTP LabTech, UK). Plates were incubated at constant temperatures (4°C, 16°C and 20°C). Protein concentration varied between 5 mg/mL to 25 mg/mL

3. Results

ImuB has been proposed to serve as an adaptor linking the β-clamp, DnaE2 and ImuA' in a transcription complex essential for translesion synthesis in Mtb (Warner *et al.*, 2010a; McHenry, 2011b). This role of ImuB has, however, not been demonstrated experimentally..

3.1. Theoretical parameters of ImuB

ImuB has a theoretical MW of 56.6 kDa and a theoretical pI of 8.16. According to ProtParam (http://web.expasy.org/cgi-bin), a protein analysis tool, the estimated half-life of ImuB in *E.coli* is more than 10 h and its instability index is 44, classifying the protein borderline unstable *in vitro* – as a value below forty generally indicates a stable protein (Guruprasad *et al.*, 1990). Removing the C-terminal extended region, which extends from the β-clamp motif of ImuB, would reduce the instability index to 37 producing a theoretically stable polypeptide. The research group of Prof. Digby Warner (IDM, University of Cape Town) previously generated an expression construct for full-length ImuB by cloning the gene *Rv3394c* into the pMal-c2X plasmid. This successfully produces soluble ImuB N-terminally fused to MBP in *E. coli*. pMal-c2x-*Rv3394c* was not initially used due to the inaccessibility of amylose resin for affinity chromatography. This plasmid was used as a template for amplification *Rv3394c* by PCR.

3.2. Bioinformatics analysis of ImuB

3.2.1. Sequence alignment of ImuB

Structural homologues of ImuB were identified by a protein basic local alignment search tool on the UniProt website (http://www.uniprot.org/blast/). A crucial step in homology modelling is sequence alignment of the target protein to the sequence of the homologues protein template (Lee *et al.*, 2006). ClustalW2 was used to align the *Rv3394c* sequence to Y-family polymerase sequences of *M.smegmatis, P.putida, E.coli,* and *S. solfataricus* (See Fig. 5). Further sequence alignments were generated using PROMALS3D bioinformatics tool, which aligns protein sequences based on structural and predicted sequence identities (Figure 5). Red rings marks acidic active side residues conserved in Y-family DNA polymerases. Domain regions were inferred from the crystal structure of Dpo4 from *S. solfataricus* (Pdb Id: 1JX4) and the LF domain region was derived using the crystal structure and sequence of *E. coli* DinB (polymerase IV) (http://www.uniprot.org/uniprot/Q47155).

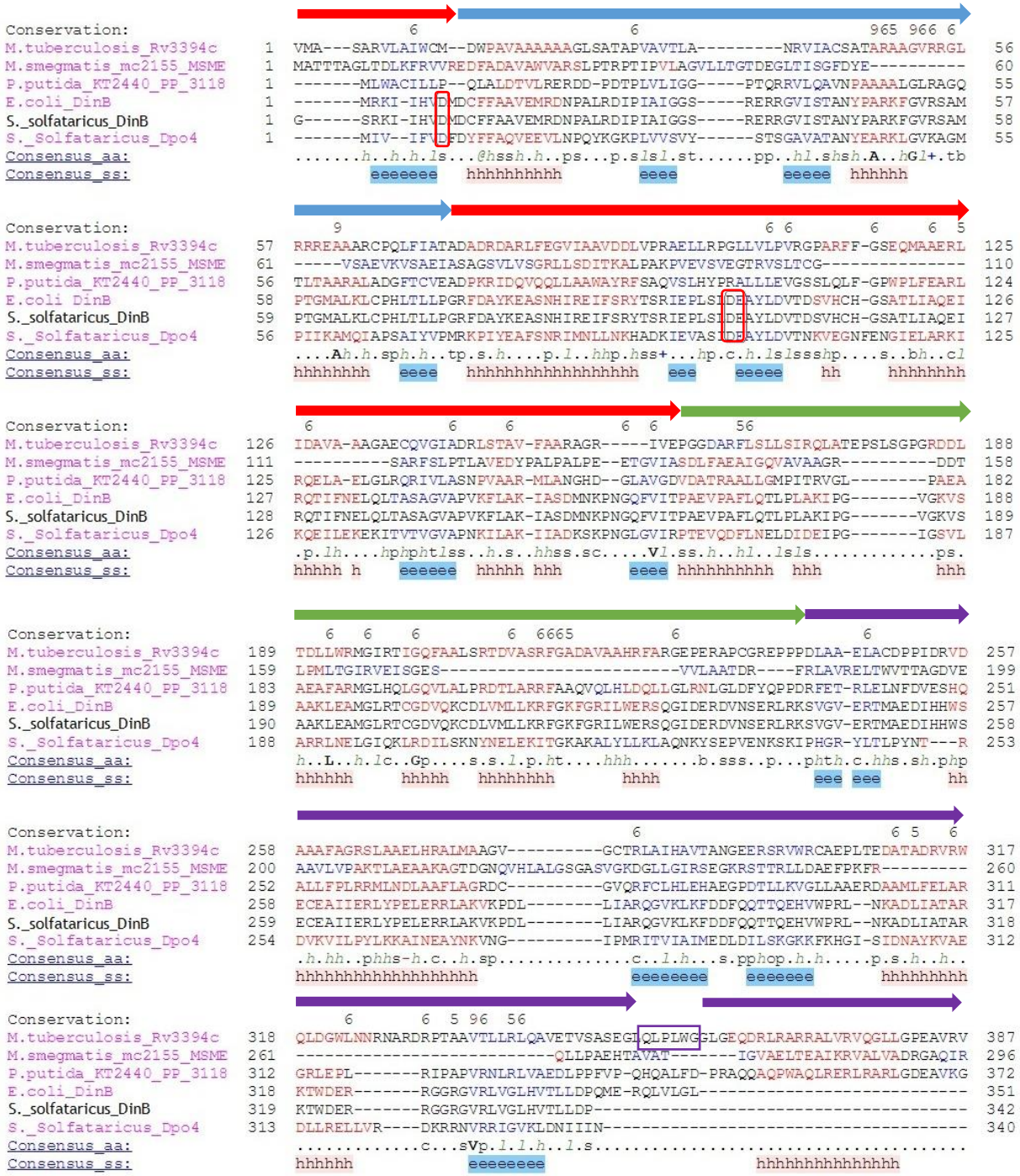


Figure 5: Multiple sequence alignment of ImuB homologues and Y-family DNA polymerases from S. solfataricus. The sequences were aligned using the PROMALS3D multiple sequence alignment server (http://prodata.swmed.edu/promals3d/promals3d.php). Consensus secondary structures (ss) are indicated by "h" for helices and "e" for β-strands. Consensus amino acids (aa) are indicated with the following symbols: aromatic residues "@"; hydrophobic residues "h"; polar residues "p"; compact residues "t"; bulky residues "b"; small residues "s"; positively charged residues "+"; negatively charged residues "-". The domain structure was annotated using the known crystal structure of Dpo4 from S. solfataricus (Pdb Id: 1JX4) ; Palm Domain "red"; Finger domain "Blue"; Thumb domain "green"; and LF domain "purple". The β-binding motif is show using a purple rectangle following the LF domain.

Rv3394c and its *M. smegmatis* homologue (MSMEG-1622) have similarities in sequence to the Y-family DNA polymerases DinB of *E. coli.* Significant differences include a C-terminal extension in ImuB beyond the β-clamp binding motif predicted to have few secondary structural features (Warner *et al.*, 2010b). In addition, ImuB lacks the acidic catalytic residues required for polymerase activity (Indicated by red circulated residues on fig. 5).

3.2.2. Disordered region prediction of ImuB

Disordered region prediction used three servers, which predict disordered regions on ImuB sequence. Prediction performed using DISOPRED shows high probability on the C-terminal extended region.

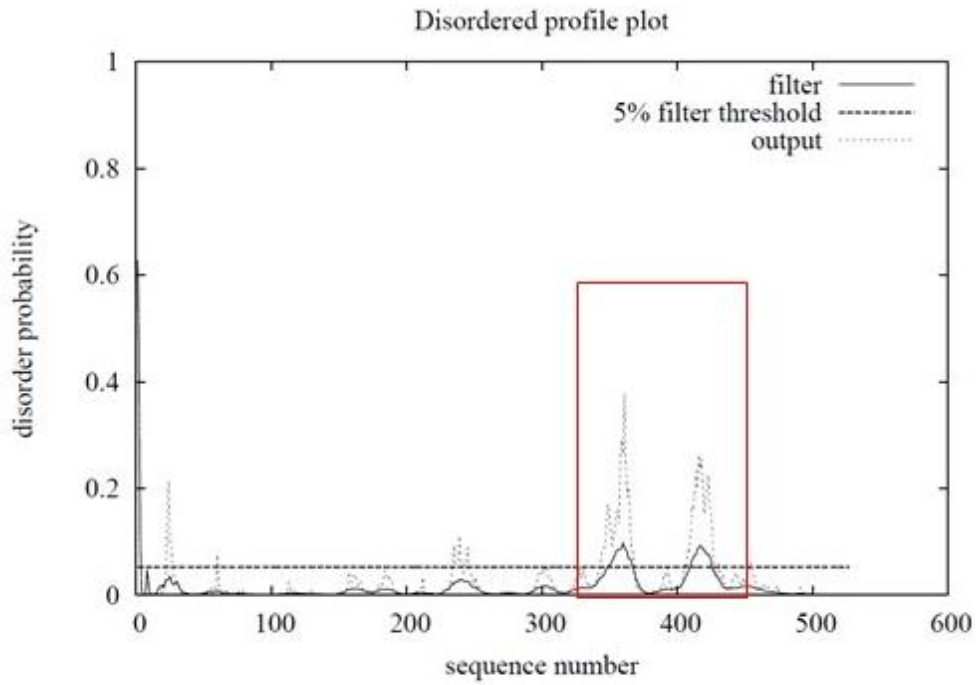


Figure 6: Disordered profile plot of ImuB (DISOPRED). The red box shows high probability of disordered regions on ImuB. The disordered regions indicated by the red box is located C-terminus of ImuB.

A summary showing the disordered regions on ImuB, was compiled using the data obtained from DisProt, DisEMBL and DisoPRED servers (Figure 9). Prediction of disordered regions shows that the C-terminus has a high probability of have unstructured features.

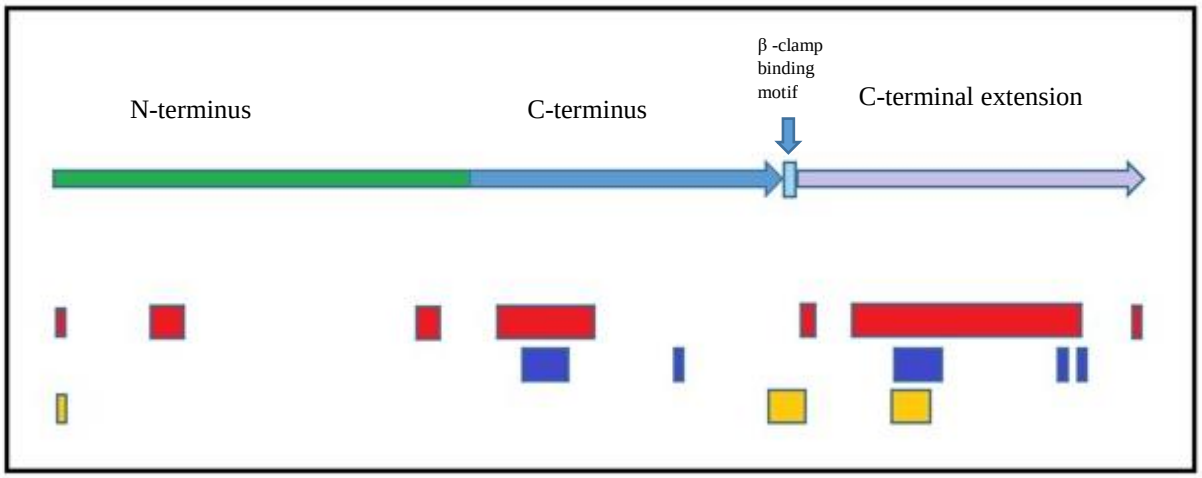


Figure 7: Summary of disordered region prediction of ImuB using DisProt, DisEMBL, and DISOPRED prediction servers. Predicted disordered regions are indicated with coloured rectangles based on the method of disordered region prediction used. DisProt (Red), DisEMBL (blue) and DISOPRED (yellow)

3.3. Generation of constructs for ImuB production

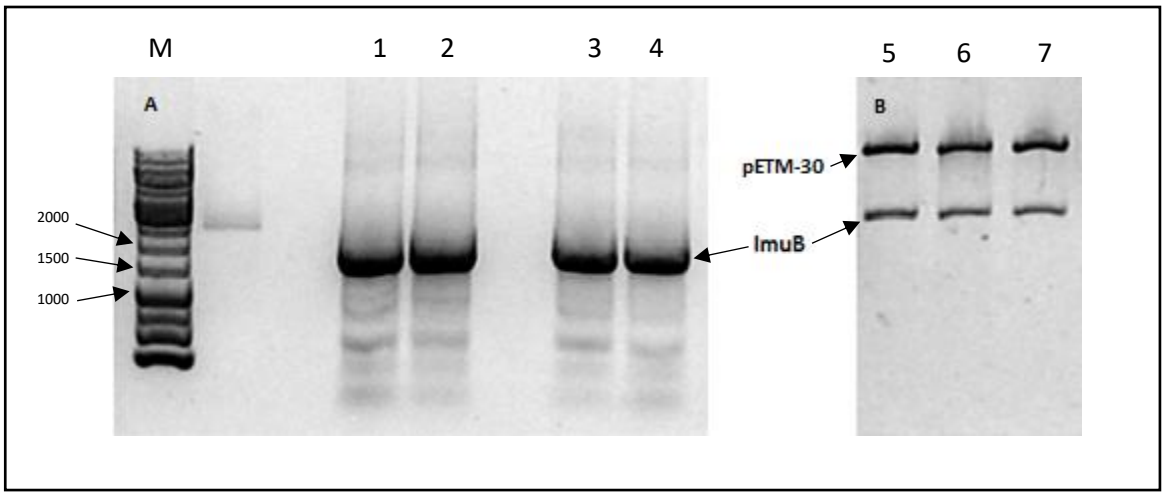


Figure 8: Agarose electrophoresis gel depicting PCR amplification of ImuB and verification of pETM-30-Rv3394c clone via restriction digestion. A: Amplification of full length Rv3394c for insertion into expression vectors (pETM-30, pET-28a (+) and pGEX-6P-2). Lanes 1&2: Duplicate PCR amplification of ImuB for the insertion into pETM-30 and pET-28a (+). Lanes 3&4: Duplicate PCR amplification of ImuB for the insertion into pGEX-6P-2. Lanes 5,6 & 7: Verification of pETM-30/Rv3394 clones with the use of Nco1 and Xho1 double digestion.

Amplification of *Rv3394c* by PCR resulted in the generation of a 1586 bp amplicon (figure 10.A.) pMal-c2X-*Rv3394c* was used as template DNA for the amplification of *Rv3394c*. Constructs were generated to enable expression of *Rv3394c* in *E. coli* expression systems. *Rv3394c* was cloned into pETM-30, pET-28a (+), pGEX-6P-2 and pMal-c2X. The clones were verified using PCR, restriction digestion with one exception (pGEX-6P-2 construct) and sequence analysis. *Rv3394c* was successfully cloned into each expression vector to produce ImuB in *E.coli* expression strains (Table 9). The sequence was analysed by Inqaba Biotech and Stellenbosch University. The pGEX-6P-2 was not verified using restriction analyses due to the elimination of BamHI restriction site. The elimination of the BamHI site occurred when ImuB was ligated using a "sticky end" produced with BglIII.

3.4. Protein production

3.4.1. Small scale expression of ImuB

Rv3394c expression constructs were transformed into BL21 (DE3), and BL21(DE3) derivatives Arctic express and Rosetta expression strains for the production of ImuB as a recombinant protein in *E. coli*. *E.coli* Rosetta was optimized to use codons for expression of eukaryotic proteins whereas *E.coli* arctic express was optimized to produce protein at lower expression temperatures. Each transformed strain was subjected to numerous expression conditions, which varied in IPTG concentration, expression times, temperature variation and media composition. The ImuB fusion protein was visible as a distinct induced band corresponding to the theoretical size of the fusion protein on SDS-PAGE.

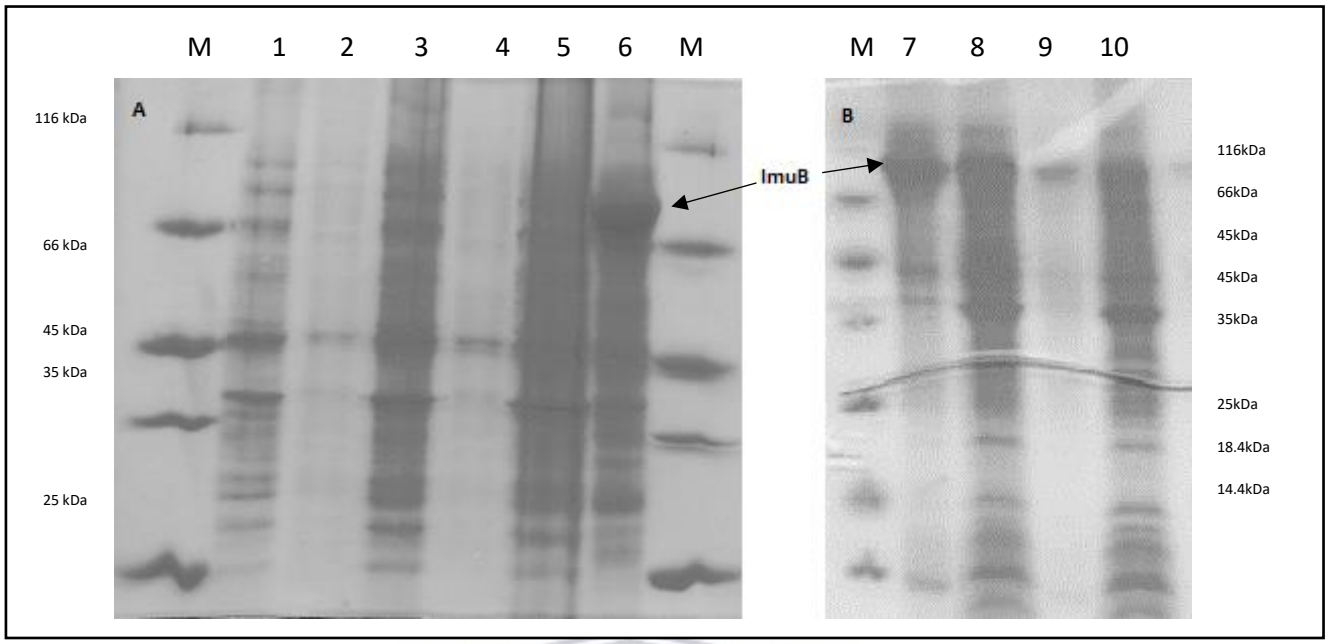


Figure 9: Small-scale production trail of ImuB. A: Small scale production of ImuB from pETM-30-Rv3394c in construct in E. coli BL21 expression strain. B: Small-scale production of ImuB from pMal-c2X-Rv3394c in E. coli BL21 expression strain. Induction of ImuB was done with 0.5 mM of IPTG at an optical density of 0.5. Cells were then grown overnight at a temperature of 25°C. M: Marker, 1: 2 h soluble fraction, 2: 2 h insoluble fraction, 3: 4 h soluble fraction, 4: 4 h insoluble fraction, 5: Overnight soluble fraction, 6: Overnight insoluble fraction. 7: Overnight insoluble fraction, 8: Overnight soluble fraction, 9: 4 h insoluble fraction, 10: 4 h soluble fraction.

Expression of soluble MBP-ImuB fusion protein was achieved using 0.5 mM IPTG (Figure 9). Expressed GST-ImuB-His_6 (~83 kDa) was only visible in the insoluble fraction during expression trials of the pETM-30-*Rv3394c* expression system. Insoluble protein extract containing GST-ImuB-His6 was collected via centrifugation at 16000 rcf for purification and refolding experiments. MBP-ImuB (~99.1 kDa) is produced as both soluble and insoluble protein. Protein production using the pMal-c2X-*Rv3394c* construct was up-scaled to produce larger amount of the soluble fusion protein.

The pMal-c2X-*Rv3394c* expression system was the only successful construct to produce soluble N-terminal tagged MBP-ImuB. All protein production systemsproduced large quantities of

insoluble ImuB with exception to pMal-c2x-*Rv3394c (see table 9)*. Insoluble fusion proteinwas later used for refolding experiments. Once soluble protein production was achieved from small scale expression trials, protein production experiments were up-scaled to larger quantities.

Table 9: Summary of protein expression tests from various protein expression systems.

Constructs	**Cell type**	**Fusion Protein**	**Protein production outcome**
pMal-c2X-*Rv3394c*	BL21	MBP- *FactorXa* -ImuB	Fusion protein produces at a seemingly equal rate in both soluble and insoluble fraction
	Rosetta		Fusion protein produces in soluble and insoluble fraction but in lesser amounts when compared to BL21 cell strain
	Arctic Express		Protein production was equivalent to that of the BL21 strain.
pETM-30-Rv3394c	BL21	GST-TEV-ImuB-His_6	Protein produced largely in the insoluble fraction
	Rosetta		Protein produced largely in the insoluble fraction
	Arctic Express		Protein produced in the insoluble fraction
pET-28a(+)-Rv3394c	BL21	ImuB-His_6	Protein produces on a very low scale in the insoluble fraction
	Rosetta		No indication of induction in both the soluble and insoluble fraction
	Arctic Express		No indication of induction in both the soluble and insoluble fraction
pGEX-6P-2-Rv3394c	BL21	GST-3C Precision Protease -ImuB	Protein produced in the insoluble fraction
	Rosetta		Protein produced in the insoluble fraction
	Arctic Express		Protein produced in the insoluble fraction
pMal-c2X-tev-Rv3394c	BL21	MBP-FXa-TEV-ImuB	Protein produced in the insoluble fraction
	Rosetta		Protein produced in the insoluble fraction
	Arctic Express		Protein produced in the insoluble fraction

MBP-ImuB from the pMal-c2x-Rv3394c was produced in *E. coli* BL21, Rosetta and Arctic Express cells with the use of LB and various IPTG concentrations. *E. coli* BL21 and Arctic Express produced the highest yields of soluble protein production during the expression trials phase. *E. coli* BL21 was chosen as the main protein expression system to produce MBP-ImuB due to its ability to produce fusion protein rapidly when compared to *E.coli* Arctic Express. *E.coli* Arctic Express requires low incubation temperature to produce protein on higher yields. Protein expression at low incubation temperatures requires more time to receive the same yields of protein as the *E.coli* BL21. Increasing of the linker between the MBP-tag and ImuB with the insertion of a TEV protease recognition site resulted in the production of insoluble protein. Introduction of the TEV protease recognition site was done due to the inability of Fxa to cleave MBP from the ImuB fusion protein. This favors the notion that ImuB solubility may be highly dependent on the availability of the MBP-tag.

3.5. Protein purification

3.5.1. Affinity purification of MBP-ImuB

MBP-ImuB was purified using amylose resin as the affinity matrix. Total soluble protein extract from lysed cells and amylose resin was incubated for 2 h of a rolling incubator at 4°C to allow for optimal binding of MBP-ImuB to the amylose resin. The fusion protein was washed using a high salt buffer to remove most of the impurities prior to elution. MBP-ImuB was eluted with a buffer containing maltose in the range of 50-500 mM. Eluted fraction were concentrated prior to additional purification steps.

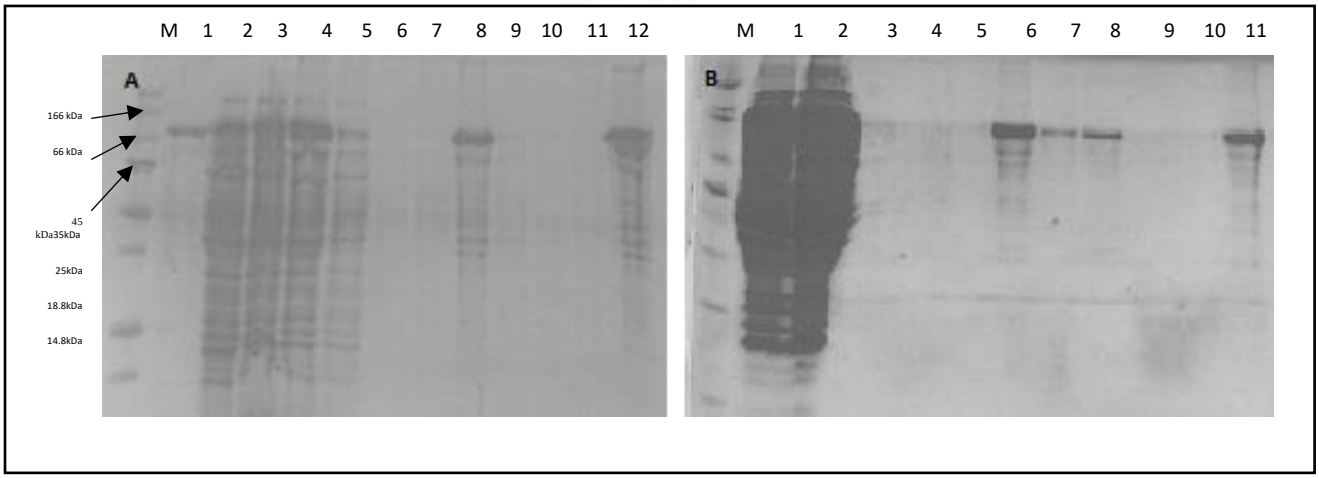


Figure 10: A: Purification of MBP-ImuB using amylose resin affinity chromotography. Lane M: Molecular Weight Marker, Lane 1: Purified MBP-ImuB. Lane 2: Total soluble fraction, Lane 3: Unbound protein Flow-through, Lane 4: Resin with bound protein before wash, Lanes 5-7: Wash fractions with wash buffer (20 mM Tris-HCl, 200 mM, NaCl, 0.5 mM EDTA pH 8.1), Lane 8: Resin sample before cleavage with Fxa, Lane 9-11: Elution with elution buffer (20 mM Tris-HCl, 200 mM, NaCl, 0.5 mM EDTA pH 8.1), lane 12: Resin sample after Elution. B: Purification of MBP-ImuB using amylose resin and Maltose. Lane M: Marker, Lane 1: Total soluble Fraction. Lane 2: Unbound protein flow through, Lanes 3-5: Wash with wash buffer (20 mM Tris-HCl, 200 mM, NaCl, 0.5 mM EDTA pH 8.1) Lane 6: Resin sample before elution, Lanes 7-10: Elution of bound protein with elution buffer (20 mM Tris-HCl, 200 mM NaCl, 50 mM Maltose, and 0.5 mM EDTA) Lane 11: Resin sample after protein elution.

Affinity purification of MBP-ImuB resulted in approximately 50% of eluted protein (As seen in Fig. 10.B). Various challenges occurred during the initial purification process. During the affinity

purification experiments, cleavage of the MBP-tag from ImuB with the use of Factor Xa protease resulted in immediate precipitation of ImuB. This fortified the assumption that the MBP-tag is required to solubilize ImuB. Maltose was used to elute the fusion protein from the amylose resin. ~ 50% of the fusion protein was eluted using buffer containing maltose in the ranges of 20 – 500 mM. High concentration of MBP-ImuB fusion was found to be "stuck" on the resin. Elution attempts with high concentrations of salt and maltose was done but resulted in protein precipitation.

3.5.2. MBP-ImuB purification by size-exclusion chromatography

Size-exclusion chromatography (SEC) was used to separate MBP-ImuB from unknown impurities. Protein fractions for the SEC were analyzed using SDS-PAGE to confirm purity of the sample fraction.

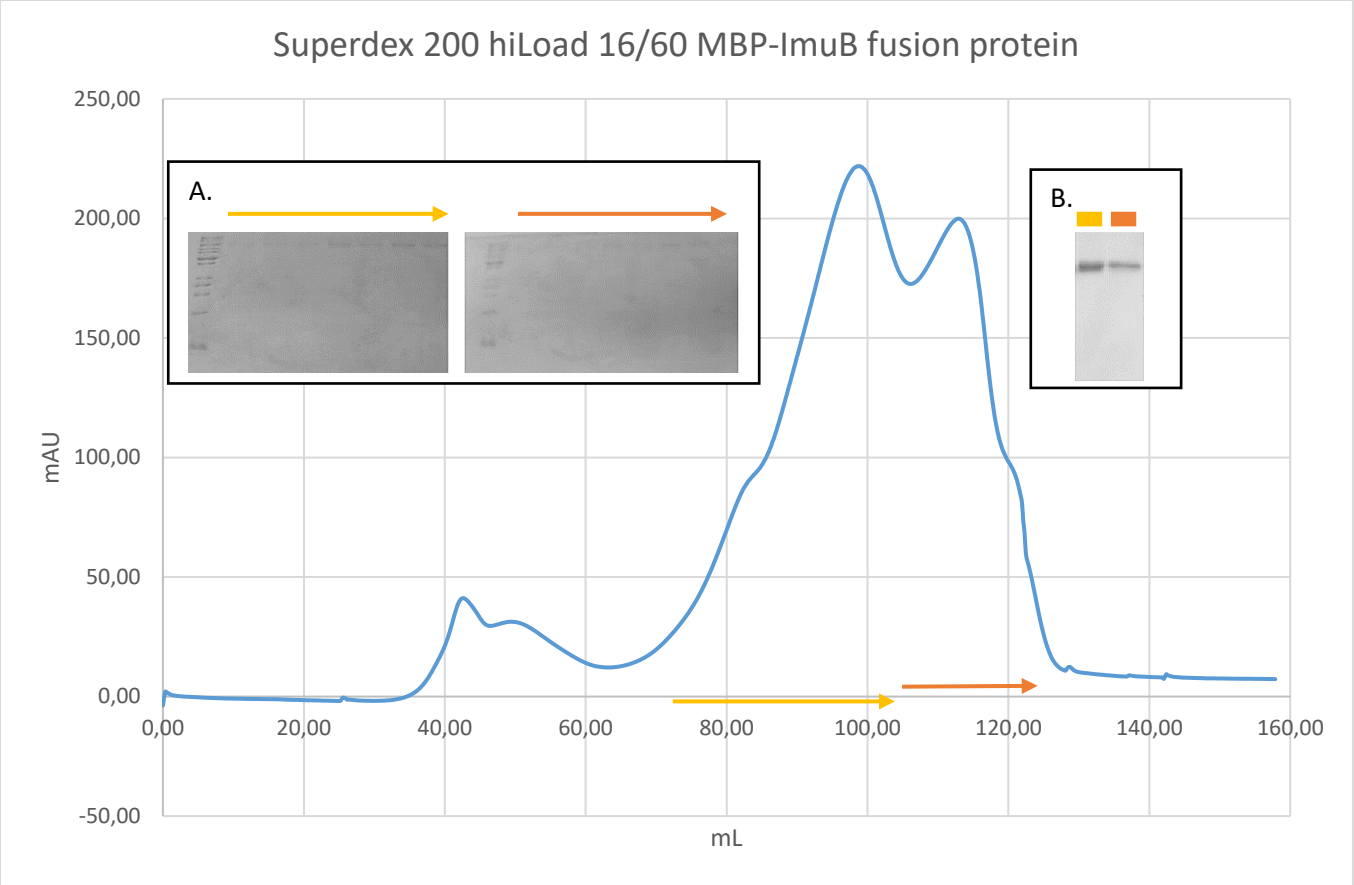


Figure 11: Chromatogram of MBP-ImuB SEC. A: SDS-PAGE image of the eluted peaks. Coloured arrow corresponds to peak eluted on Chromatogram. B: Pooled and concentrated elution fractions from SEC purification.

Two peaks (indicated by yellow and orange bar lines in fig. 11) contained a protein the size of MBP-ImuB (figure 11). The eluted protein was ~100 kDa, which is the theoretical size of the fusion protein. Each of the peaks were colour-coordinated with the indicated bands on the SDS-PAGE. SEC chromatograms shows a prevalent peak eluting between 80 – 100 mL with the addition of a second overlapping peak between 100 – 120 mL of elution buffer. This pattern has occurred often when purifying ImuB using SEC purification. This occurrence gives rise to the assumption that MBP-ImuB may exist in two states. In this study, these states will be referred to as the dimeric state (orange) and oligomeric state (yellow). These states were pooled and concentrated separately for crystallization experiments.

3.6. Refolding of ImuB-His6

ImuB-His6 was purified from insoluble lysate produced using the pET-28a (+)/Rv3394c expression vector. The insoluble fraction was washed and solubilized using 8 M Urea. Coupling of the insoluble fraction with Ni-NTA resin removed unwanted protein impurities prior to undergoing protein refolding. Elution of the insoluble fusion protein for the dialysis refolding experiment was done using an elution buffer containing imidazole (500 mM) and NaCL (200 mM). A random variety of Tween 80 and Triton-X100 concentrations were used to assist with the refolding of the fusion protein. The adjustment of temperature (4°C, 16°C and 20°C) and buffer composition was explored to retrieve possible refolding conditions. Refolding of ImuB-His6 was inconclusive using both on-column refolding and dialysis. On-column refolding often resulted in precipitated protein, which could not be eluted from the Ni-NTA column. Refolding via dialysis often resulted in rapid precipitation within the Bio-Rad SnakeSkin dialysis tubing. No trace of fusion protein was found within the soluble fractions for each of the conducted experiments.

3.7. Crystallization

MBP-ImuB fusion protein was the only soluble candidate that was used for crystallization experiments. Purified MBP-ImuB fractions obtained from size exclusion chromatography were pooled, concentrated to 25 mg/mL and stored at 4 ˚C waiting crystallization setup. Distinctive crystallization conditions using the JCSG Core Suites were tested with different protein concentrations (5-25 mg/mL). Variety of temperatures were exploited for each of the crystallization experiment (4°C, 16°C and 20°C). Most of the crystallization conditioned prepared resulted in no plausible crystal “hits”. Precipitation of the fusion protein and phase separation were common occurrences during crystallization experiments of MBP-ImuB. Protein precipitation

generally occurred within a week of crystallization preparation but could take up to two weeks at 4°C. Precipitant was collected using centrifugation and tested using SDS-PAGE after wash steps with salt free buffer. The precipitant was confirmed to be MBP-ImuB. No useable protein crystals "hits" were observed during the crystallization protocol.

3.8. Homology modelling of ImuB using Y-family DNA polymerase crystal structures

Detailed structural data on biological macromolecules still require experimentally verifiable data as obtained by x-ray crystallography, nuclear magnetic resonance spectroscopy and electron microscopy. However, in the absence of such data, homology modelling can provide useful information about a target protein based on features it shares with related proteins and may indicate the basic function of the target protein.

A homology model of ImuB was constructed using the protein modelling software "Modeller". The ImuB model was based on the crystal structures of Dpo4 (PDB id: 1JX4) and DinB (PDB id: 4IR1) from *S. solfataricus* and *E. coli*, respectively. The Zhanglab ModRefiner algorithm was used to refine the homology model, which encompasses residues 1-360 only as the homologues lack an equivalent C-terminal region.

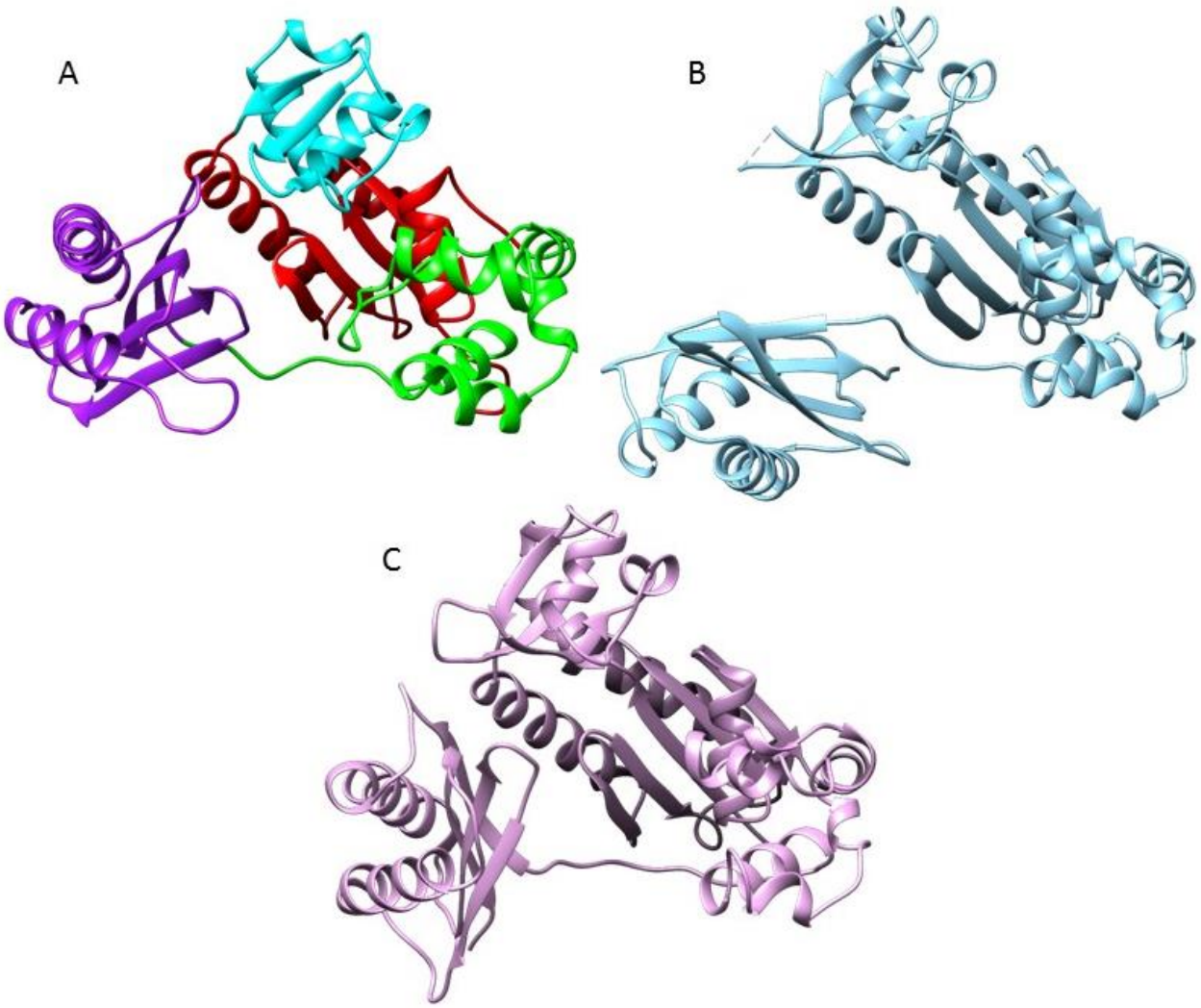


Figure 12: Comparison of the ImuB homology model with the crystal structure of Dpo4 (PDB id: 1JX4) from S. solfataricus. A: Truncated ImuB homology model coloured to protein chain progression from Palm Domain "red"; Finger domain "Blue"; Thumb domain "green"; and LF domain "purple". B: Crystal structure of DinB from S. solfataricus (PDB Id: 1K1S) C: Dpo4 from S. solfataricus in a tertiary complex with DNA duplex (PDB Id: 1JX4)

The ImuB homology model is structurally similar to the Y-family DNA polymerase DinB and Dpo4 from *S. solfataricus* with minor rearrangement of secondary structural elements. The model retains an overall right-hand structure reminiscent of Y-family DNA polymerases. Comparison of domain structure of the palm, finger, thumb and LF shows that each domain retains the secondary structural features in each domain (See figure 12). The palm domain on the ImuB model is similar

to the crystal structures of Dpo4 and DinB. The palm domain retains the β-sheet, which harbours the catalytic residues in Y-family DNA polymerase.

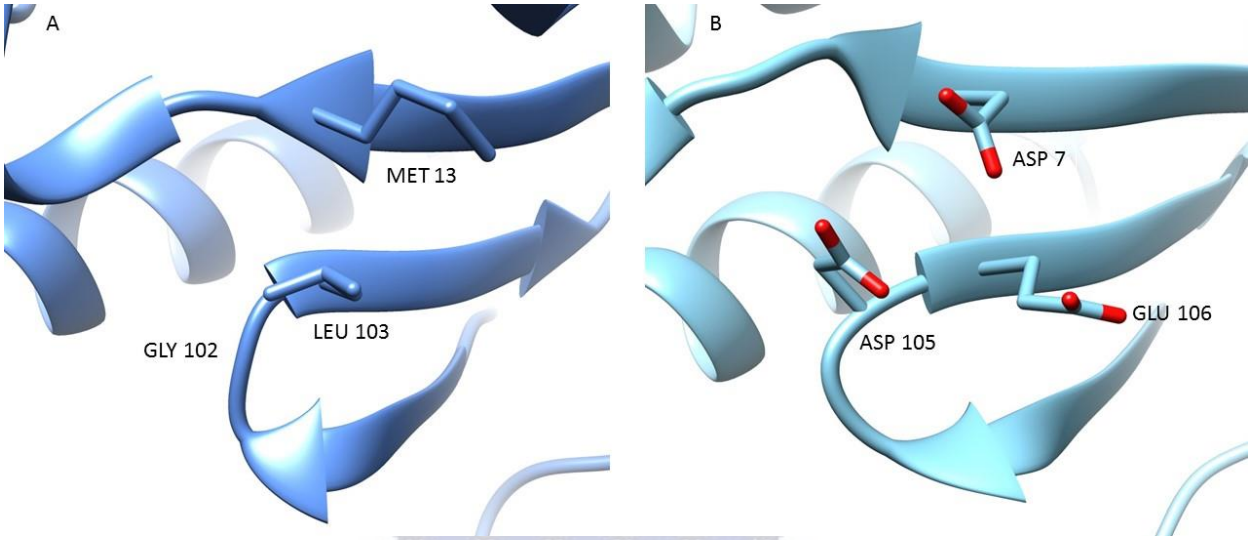


Figure 13: Close up view of active sites residues on palm domain. A: Close up view of active site residues on the ImuB model. B: Close up view of active site on the crystal structure of Dpo4 from S. solfataricus.

Active sites comparison of ImuB model and Dpo4 crystal structure shows that the ImuB model does not share similar carboxylated active site residues (ASP7, ASP105 and GLU106, Dpo4 numbering) in its palm domain required for two metal ion coordination, which binds and positions incoming nucleotide bases(Fig. 13) (Lee *et al.*, 2006; Lehmann *et al.*, 2007). Instead, the residues are replaced by MET13, GLY102 and LEU103, all of which lacks a catalytically active carboxylate group. The lack of catalytic residues suggests that ImuB is not required as polymerase but required for the positioning of DnaE2 and the DNA lesion.

4. Discussion:

4.1. Rv3394c clone generation

Rv3394c was amplified using PCR and cloned in expression vectors pMal-c2x, pET-28a(+), pETM-30 and pGEX-6p2. Cloning of Rv3394c into pGex-6p2 required the use of BglIII restriction enzyme as a BamHI alternative. BglIII produces the same restriction overhangs as BamHI, which allows for ligation using the BamHI cloning site on pGEX-6p2. The use of BglIII eliminated the cleavage of the BamHI site at 1462bp of the Rv3394c gene.

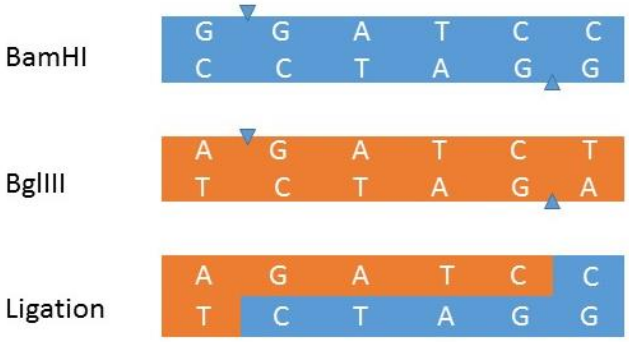


Figure 14: Comparison of BglIII and BamHI cleavage recognition sites and ligation result of BglIII and BamHI overhangs.

Ligation of BgIII and BamHI overhangs produces a non-palindromic site, which could not be cleaved using any known restriction enzyme (Figure 14). The cloning of pGEX-6p2-*Rv3394c* could only be verified using sequencing and PCR. Cloning of *Rv3394c* into pETM-30 and pET-28a(+) could be achieved using the same *Rv3394c* amplicon as pET-30 and pET-28(+) has similar multiple cloning sites. All cloned constructs were successfully verified using restriction digestion with exception to the pGEX-6p2-*Rv3394c* construct.

4.2. Production and purification of MBP-ImuB

Production of recombinant ImuB in *E. coli* turned out to be a challenging process as many of the constructs produced, resulted in the fusion protein aggregates and insoluble fusion protein. During protein expression experiments, most rewarding result was produced from the pMal-c2X-*Rv3394c* expression construct, which produced soluble ImuB in *E. coli* as an N-terminal MBP fusion protein. This result was previously achieved by Dr. Digby Warner's research group but was not explored immediately due to the lack of purification material and factorXa to cleave ImuB from MBP affinity tag. As generally seen during expression experiments, the MBP-ImuB fusion protein was predominantly produced as insoluble protein with a small amount of soluble protein, which was purified for further analysis. Generally, Mtb proteins are difficult to produce solubly in *E. coli* due to the high GC content of their genes, bias codon usage between Mtb and *E. coli*, protein size, hydrophobicity, acidity and the lack of essential chaperones (Lakey *et al.*, 2000).

pMal-c2X-Rv3394c was used to produce MBP-ImuB on a large scale for processing and crystallization. Amylose resin affinity chromatography was the first step to purify MBP-ImuB from the soluble *E. coli* cell lysate. MBP-ImuB was eluted from the resin using maltose, although ~50 % of the fusion protein remained bound to the resin. Cleavage of the MBP-tag from ImuB using FactorXa restriction enzyme always resulted in precipitation implying that the solubility of ImuB was directly linked to the presence of the MBP tag. MBP is reported to increase the solubility of a fusion protein when compared to other fusion tags (Smyth *et al.*, 2003). Following initial purification step, MBP-ImuB was purified using SEC resulting in overlapping peaks containing protein corresponding to the theoretical size of MBP-ImuB when analysed using SDS-PAGE. These peaks could be observed as MBP-ImuB exhibiting two protein states within solution. ImuB is known to have the ability to self-adhere (Warner *et al.*, 2010b). Fractions under each peak were

pooled, and concentrated separately prior to crystallography experiments. Substantial amounts of protein, especially the assumed "dimeric" state, precipitated upon concentration. Constantly occurring protein precipitation could be due to the instability created by the assumed disordered C-terminal region. Proteins with disordered regions are known to aggregate *in vitro,* making it a daunting task to purify the subject for downstream processes (Dunker *et al.*, 2001; Dobson, 2004; Cromwell *et al.*, 2006).

4.3. Crystallization of MBP-ImuB:

Sitting-drop vapor diffusion crystallization of MBP-ImuB was unsuccessful as no protein crystal formation was observed during the screening process using commercial screening kits (JCSG crystallization suites). MBP-ImuB rapidly precipitated in most screen conditions. A common occurrence throughout the study was the precipitation of MBP-ImuB during concentration processes. Precipitation of ImuB could primarily be due to hydrophobic aggregation, which subtly disrupts the folded structure exposing more of the hydrophobic interior to the solution, which is common when working with proteins containing disordered regions such as ImuB. Crystal growth could be hindered by conformational heterogeneity between ImuB and the fusion protein. ImuB is predicted to have a highly disordered C-terminal extension, which could hinder crystal growth by decreasing homogeneity of the fusion protein molecule in solution. Flexibility on both C-terminal and N-terminal regions could be the major factor for the lack of crystal formation and nucleation.

4.4. Homology modelling of ImuB:

A homology model of ImuB was produced due to the inability to produce an atomic structure experimentally.

Homology modelling and sequence alignment shows that ImuB consists of three sections; 1) an N-terminal regions, which harbours three polymerase-like domains (palm, finger and thumb domain); 2) a C-terminal region, which consists solely of the Y-family polymerase-like LF domain; 3) and the disordered C-terminal extension, which extends from the β-clamp binding motif. Disordered region prediction and lack of structured homologues predicts that the C-terminal extension contains high probability of disordered regions. The C-terminal extension is excluded from the model due to the lack of homologues with experimentally determined structural information.

4.5. Structural similarity between ImuB and Y-family DNA polymerase

ImuB is structurally similar to Y-family DNA polymerases such as Dpo4 from *E.coli* by retaining domain structures such as the palm, LF, finger and thumb domains (Ohmori *et al.*, 2001; Ippoliti *et al.*, 2012). These domains are positioned in a right-hand-like structure, which favours interaction with the DNA duplex. This right-hand-like structure is a well-known feature of polymerases but the C-terminal LF domain is unique to Y-family DNA polymerases.

ImuB is predicted to uniquely possess a disordered C-terminal extension as a feature of ImuB and ImuB homologues (Aravind *et al.*, 2013). ImuB furthermore lacks the catalytic residues on the β-sheet of the palm domain (Asp 7, Asp 105, and Glu 106), disqualifying it as a functional Y-family DNA polymerase (Figure 13) (Rechkoblit *et al.*, 2006). Sequence alignment with ImuB and DinB from *E. coli* and *S. solfataricus*, indicated that ImuB possesses the conserved residues (C66 and P67) mainly responsible for the formation of the RecA/UmuD complex during the SOS response (figure 5) (Cafarelli *et al.*, 2014). The presence of C66 and P67, classifies ImuB as a member of

the DinB3-type Y-family polymerase although ImuB itself does not perform any polymerase activity (Cafarelli et al., 2014). Previously ImuB was classified as a member of DinB3-type Y-family polymerase due to the presence of the β-clamp binding domains (^{354}QLPLWG^{359}) (Bunting *et al.*, 2003; Cafarelli *et al.*, 2014).

4.6. ImuB C-terminal extension Vs intrinsic disordered proteins

Highly disordered regions such as ImuB's C-terminal extension, are featured domains of intrinsic disordered proteins (IDP) (Dunker *et al.*, 2001; Dunker *et al.*, 2008; Turoverov *et al.*, 2010). It is common for a protein to contain both structured and unstructured regions therefore this type present in ImuB is not unique (Structured N-terminal and unstructured C-terminal regions) (Dunker *et al.*, 2001; Dunker *et al.*, 2008). There is a wide number of distinctive proteins exhibiting the same structural characteristic such as the prokaryotic ubiquitin-like protein (Pup), which has been indicated to be involved in the interaction with the mycobacterial proteosomal ATPase (MPa) via its disordered C-terminal region (Chen *et al.*, 2009). The Y-family DNA polymerase, UmuD, is also known to have features, which have significantly less secondary and tertiary structural aspects reminiscent of IDP's. IDP's are often found to have specified biological roles within regulatory systems, as they are more likely to assume exact structures upon interaction with their regulatory binding partners (Godoy *et al.*, 2007; Ollivierre *et al.*, 2013). ImuB and $UmuD_2$ share an extended shorter arms structure that is disordered from the unbound globular C-terminal domain (Peat *et al.*, 1996; Wang, 2001; Hawver and Beuning, 2013). Previous binding analysis of ImuB with its cassette counterparts pin-points the mechanism of binding to occur using the C-terminal extension reminiscent of IDP's. Structural comparison of the modelled structure of ImuB and the

crystal structure UmuD suggests that they may be distant homologues because of their very apparent C-terminal region, along their designated Y-family DNA polymerase N-terminal structure (Ling *et al.*, 2001; Dunker *et al.*, 2008; Ippoliti *et al.*, 2012; Hawver and Beuning, 2013). This would propose that ImuB could function as an analog of UmuD with the use of the C-terminal extension which would have key involvement in protein-protein interaction with other components in the mutasome such as ImuA' and DnaE2 (Warner *et al.*, 2010a).

4.7. Proposed role of ImuB:

ImuB interacts with ImuA', DnaE2 and DNA implying that ImuB functions as an organizing scaffold (Ohmori *et al.*, 2001; Jarosz *et al.*, 2007; Lehmann *et al.*, 2007; Choi *et al.*, 2010). ImuB is important in enabling DnaE2 to access the replication fork using its β-clamp binding domain upstream of the C-terminal extended region (Warner, 2010; Warner *et al.*, 2010b).

ImuB is able to form self-interact oligomeric states reminiscent of the UmuD (Sassanfar and Roberts, 1990; Smith and Walker, 1998; Warner *et al.*, 2010a). Similar to UmuD, ImuB is induced on a large scale when in the presence of damaged DNA in Mtb (Warner *et al.*, 2010b). ImuB induction in the presence of DNA damage and the ability of ImuB to self-interact would suggest that it may have similar functions to that of UmuD in the *E. coli* SOS response (Belov *et al.*, 2009; Warner *et al.*, 2010a; Warner, 2010).

Put together, assumptions based on structural features using the homology model and disorder prediction suggests that 1) ImuB may plays a key role in the regulation of the initial step of the SOS DNA repair mechanism, 2) ImuB may act as an adaptor protein that is able to bind both DnaE2 and ImuA' using its C-terminal region, and binds DNA using its N-terminal non-active

polymerase domain, 3) ImuB may be required during the process of polymerase switching through the interaction with the β-binding Clamp domain.

4.8. Conclusion and future strategies

The aim of this study was to produce and purify recombinant ImuB for crystallization and structure determination. The *Rv3394c* gene was successfully amplified by PCR and cloned into protein production vectors (see table 9). The vectors were transformed into *E. coli* host strains and analyzed for recombinant protein production. Most expression systems produced substantial amount of insoluble ImuB with exception to the *pMal-c2x-Rv3394c* expression construct. The fusion protein, produced from the *pMal-c2x-Rv3394c* expression vector, was successfully purified by amylose affinity chromatography and size exclusion chromatography. Upon cleavage of ImuB from the fusion tag, the protein unfortunately precipitated almost immediately. The purified fusion protein was concentrated and setup for crystallization, but no useful crystal hits were observed. The protein in most wells during crystallization experiments either precipitated or produced phase separation within the solution. A comparative model of ImuB was built using homologues with available crystal structures. The model structures shows that the N-terminus of ImuB may have a characteristic Y-family polymerase-like structure, which lacks the active site residues, conserved in other Y-family polymerases. The C-terminus was shown to be uncommon among Y-family polymerase yet is conserved amongst ImuB homologues (Koorits *et al.*, 2007; Ippoliti *et al.*, 2012). Strategies for protein production and refolding experiments do not result in the same degree of success for different protein species and consequently the production of soluble, correctly folded protein remains a trial by error process (Soto, 2001). It needs to be noted that the proteolytic

stability and solubility of a specific protein cannot be confidently predicted in advance (Smyth *et al.*, 2003). Despite being able to recover functional protein *in vitro* after preparative refolding from inclusion, the protein optimization for a specific protein species required time-consuming efforts with results that being irregular, may not always result in positive process for a convenient downstream application such as macromolecule x-ray crystallography and structure determination (Soto, 2001; Smyth *et al.*, 2003). Although a straight soluble protein is desirable, limited knowledge about the molecular basis of protein aggregation does not permit successful protein folding when it does not occur spontaneously. In the course of the experiment, the production and purification of ImuB has become quite a daunting task as the full length protein was speculated to not be produced as a fully soluble protein. High rate of insolubility could be due to the fact that ImuB has an unstructured C-terminal region, reminiscent of intrinsic disordered protein regions. Truncation of ImuB could be a possible solution to the insolubility problem. It is also shown that the N-terminal region is predicted to adopt a well-defined Y-family-like polymerase three dimensional structure, which could allow for soluble protein production, purification and structure analysis. Introduction of a binding partner for ImuB during protein production could stabilize the C-terminal domains and produce a soluble macromolecule, which could be further processed and analyzed. For future purposes, the co-production of ImuB with its binding partners, DnaE2 and ImuA' should be pursued. This may result in stable complex which can be processed further for downstream application and crystallization.

References

ABELLA, M., CAMPOY, S., ERILL, I., ROJO, F. & BARBÉ, J. (2007). Cohabitation of Two Different lexA Regulons in Pseudomonas putida. Journal of Bacteriology, 189, 8855-8862.

ARAVIND, L., ANAND, S. & IYER, L. (2013). Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. Biology Direct, 8, 20.

BAKER, T. A. & BELL, S. P. (1998). Polymerases and the Replisome: Machines within Machines. Cell, 92, 295-305.

BELOV, O. V., CHULUUNBAATAR, O., KAPRALOV, M. I. & SWEILAM, N. H. (2013). The role of the bacterial mismatch repair system in SOS-induced mutagenesis: A theoretical background. Journal of Theoretical Biology, 332, 30-41.

BELOV, O. V., KRASAVIN, E. A. & PARKHOMENKO, A. Y. (2009). Model of SOS-induced mutagenesis in bacteria Escherichia coli under ultraviolet irradiation. Journal of Theoretical Biology, 261, 388-395.

BLOOM, B. R. & FINE, P. E. M. 1994. Tuberculosis: Pathogenesis, Protection, and Control.

BLOWER, S. M., MCLEAN, A. R., PORCO, T. C., SMALL, P. M., HOPEWELL, P. C., SANCHEZ, M. A. & MOSS, A. R. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. Nat Med, 1, 815-821.

BOSHOFF, H. I. M., REED, M. B., BARRY III, C. E. & MIZRAHI, V. (2003). DnaE2 Polymerase Contributes to In Vivo Survival and the Emergence of Drug Resistance in Mycobacterium tuberculosis. Cell, 113, 183-193.

BRENNAN, P. J. & DRAPER, P. 1994. Tuberculosis: Pathogenesis, Protection, and Control.

BRUCK, I., GOODMAN, M. F. & O'DONNELL, M. (2003). The essential C family DnaE polymerase is error-prone and efficient at lesion bypass. The Journal of biological chemistry, 278, 44361-8.

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. & GODOY, V. G. (2014). The DinB•RecA complex of Escherichia coli mediates an efficient and high-fidelity response to ubiquitous alkylation lesions. Environmental and Molecular Mutagenesis, 55, 92-102.

CHEN, X., SOLOMON, W. C., KANG, Y., CERDA-MAIRA, F., DARWIN, K. H. & WALTERS, K. J. (2009). Prokaryotic Ubiquitin-Like Protein Pup Is Intrinsically Disordered. Journal of Molecular Biology, 392, 208-217.

CHOI, J.-Y., LIM, S., KIM, E.-J., JO, A. & GUENGERICH, F. P. (2010). Translesion Synthesis across Abasic Lesions by Human B-Family and Y-Family DNA Polymerases α, δ, η, ι, κ, and REV1. Journal of molecular biology, 404, 34-44.

COLLINS, A. R. 1996. Chapter 11 DNA Damage and its repair. In: BITTAR, E. E. & NEVILLE, B. (eds.) Principles of Medical Biology. Elsevier.

CROMWELL, M. M., HILARIO, E. & JACOBSON, F. (2006). Protein aggregation and bioprocessing. The AAPS Journal, 8, E572-E579.

DE GROOTE, F. H., JANSEN, J. G., MASUDA, Y., SHAH, D. M., KAMIYA, K., DE WIND, N. & SIEGAL, G. (2011). The Rev1 translesion synthesis polymerase has multiple distinct DNA binding modes. DNA Repair, 10, 915-925.

DERVYN, E., SUSKI, C., DANIEL, R., BRUAND, C., CHAPUIS, J., ERRINGTON, J., JANNIERE, L. & EHRLICH, S. D. (2001). Two essential DNA polymerases at the bacterial replication fork. Science, 294, 1716-9.

DOBSON, C. M. (2004). Principles of protein folding, misfolding and aggregation. Seminars in Cell & Developmental Biology, 15, 3-16.

DOVER, L. G., COXON, G., DOVER, L. G. & COXON, G. (2011). Currrent status and research strategies in tuberculosis drug development. Journal of Medicinal Chemistry, 54, 6157-6165.

DUNKER, A. K., LAWSON, J. D., BROWN, C. J., WILLIAMS, R. M., ROMERO, P., OH, J. S., OLDFIELD, C. J., CAMPEN, A. M., RATLIFF, C. M., HIPPS, K. W., AUSIO, J., NISSEN, M.

S., REEVES, R., KANG, C., KISSINGER, C. R., BAILEY, R. W., GRISWOLD, M. D., CHIU, W., GARNER, E. C. & OBRADOVIC, Z. (2001). Intrinsically disordered protein. Journal of Molecular Graphics and Modelling, 19, 26-59.

DUNKER, A. K., SILMAN, I., UVERSKY, V. N. & SUSSMAN, J. L. (2008). Function and structure of inherently disordered proteins. Current Opinion in Structural Biology, 18, 756-764.

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. Molecular Microbiology, 26, 643-653.

ERILL, I., CAMPOY, S., MAZON, G. & BARBÉ, J. (2006). Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. Nucleic Acids Research, 34, 66-77.

FILEE, J., FORTERRE, P., SEN-LIN, T. & LAURENT, J. (2002). Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. Journal of molecular evolution, 54, 763-73.

FRANK, E. G., ENNIS, D. G., GONZALEZ, M., LEVINE, A. S. & WOODGATE, R. (1996). Regulation of SOS mutagenesis by proteolysis. Proceedings of the National Academy of Sciences, 93, 10291-10296.

FRIEDBERG, E. C., FISCHHABER, P. L. & KISKER, C. (2001). Error-Prone DNA Polymerases: Novel Structures and the Benefits of Infidelity. Cell, 107, 9-12.

GIESE, K. C., MICHALOWSKI, C. B. & LITTLE, J. W. (2008). RecA-Dependent Cleavage of LexA Dimers. Journal of Molecular Biology, 377, 148-161.

GODOY, V. G., JAROSZ, D. F., SIMON, S. M., ABYZOV, A., ILYIN, V. & WALKER, G. C. (2007). UmuD and RecA Directly Modulate the Mutagenic Potential of the Y Family DNA Polymerase DinB. Molecular Cell, 28, 1058-1070.

GOMEZ, J. E. & MCKINNEY, J. D. (2004). M. tuberculosis persistence, latency, and drug tolerance. Tuberculosis, 84, 29-44.

GURUPRASAD, K., REDDY, B. V. B. & PANDIT, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering, 4, 155-161.

HAWVER, L. A. & BEUNING, P. J. 2013. UmuC D Lesion Bypass DNA Polymerase V. In: LANE, W. J. L. D. (ed.) Encyclopedia of Biological Chemistry. Waltham: Academic Press.

HUGHES, J. P., REES, S., KALINDJIAN, S. B. & PHILPOTT, K. L. (2011). Principles of early drug discovery. Br J Pharmacol, 162, 1239-49.

IPPOLITI, P., DELATEUR, N., JONES, K. & BEUNING, P. (2012). Multiple Strategies for Translesion Synthesis in Bacteria. Cells, 1, 799-831.

JAROSZ, D. F., BEUNING, P. J., COHEN, S. E. & WALKER, G. C. (2007). Y-family DNA polymerases in Escherichia coli. Trends in microbiology, 15, 70-77.

KATZ, D., ALBALAK, R., WING, J. S. & COMBS, V. (2007). Setting the agenda: A new model for collaborative tuberculosis epidemiologic research. Tuberculosis, 87, 1-6.

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, 6, 863-868.

KRISHNA, S., MASLOV, S. & SNEPPEN, K. (2007). UV-Induced Mutagenesis in Escherichia coli SOS Response: A Quantitative Model. PLoS Comput Biol, 3, e41.

KUBAN, W., VAISMAN, A., MCDONALD, J. P., KARATA, K., YANG, W., GOODMAN, M. F. & WOODGATE, R. (2012). Escherichia coli UmuC active site mutants: Effects on translesion DNA synthesis, mutagenesis and cell survival. DNA Repair, 11, 726-732.

LAKEY, D. L., VOLADRI, R. K. R., EDWARDS, K. M., HAGER, C., SAMTEN, B., WALLIS, R. S., BARNES, P. F. & KERNODLE, D. S. (2000). Enhanced Production of Recombinant Mycobacterium tuberculosis Antigens in Escherichia coli by Replacement of Low-Usage Codons. Infection and Immunity, 68, 233-238.

LE CHATELIER, E., BECHEREL, O. J., D'ALENCON, E., CANCEILL, D., EHRLICH, S. D., FUCHS, R. P. & JANNIERE, L. (2004). Involvement of DnaE, the second replicative DNA

polymerase from Bacillus subtilis, in DNA mutagenesis. The Journal of biological chemistry, 279, 1757-67.

LEE, C. H., CHANDANI, S. & LOECHLER, E. L. (2006). Homology modeling of four Y-family, lesion-bypass DNA polymerases: The case that E. coli Pol IV and human Pol κ are orthologs, and E. coli Pol V and human Pol η are orthologs. Journal of Molecular Graphics and Modelling, 25, 87-102.

LEHMANN, A. R., NIIMI, A., OGI, T., BROWN, S., SABBIONEDA, S., WING, J. F., KANNOUCHE, P. L. & GREEN, C. M. (2007). Translesion synthesis: Y-family polymerases and the polymerase switch. DNA Repair, 6, 891-899.

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.

MCHENRY, C. S. (2011a). Bacterial replicases and related polymerases. Current Opinion in Chemical Biology, 15, 587-594.

MCHENRY, C. S. (2011b). Breaking the rules: bacteria that use several DNA polymerase III's. EMBO reports, 12, 408-14.

MCHENRY, C. S. (2011c). DNA Replicases from a Bacterial Perspective. Annual Review of Biochemistry, 80, 403-436.

MIZRAHI, V. & ANDERSEN, S. J. (1998). DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? Molecular Microbiology, 29, 1331-1339.

NJIRE, M., TAN, Y., MUGWERU, J., WANG, C., GUO, J., YEW, W., TAN, S. & ZHANG, T. Pyrazinamide resistance in Mycobacterium tuberculosis: Review and update. Advances in Medical Sciences.

OHMORI, H., FRIEDBERG, E. C., FUCHS, R. P. P., GOODMAN, M. F., HANAOKA, F., HINKLE, D., KUNKEL, T. A., LAWRENCE, C. W., LIVNEH, Z., NOHMI, T., PRAKASH, L., PRAKASH, S., TODO, T., WALKER, G. C., WANG, Z. & WOODGATE, R. (2001). The Y-Family of DNA Polymerases. Molecular Cell, 8, 7-8.

OLLIVIERRE, J. N., SIKORA, J. L. & BEUNING, P. J. (2013). Dimer exchange and cleavage specificity of the DNA damage response protein UmuD. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1834, 611-620.

PATA, J. D. (2010). Structural diversity of the Y-family DNA polymerases. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, 1804, 1124-1135.

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, 171-84.

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[prime] protein and its regulation in response to DNA damage. Nature, 380, 727-730.

PERRY, K. L., ELLEDGE, S. J., MITCHELL, B. B., MARSH, L. & WALKER, G. C. (1985). umuDC and mucAB operons whose products are required for UV light- and chemical-induced mutagenesis: UmuD, MucA, and LexA proteins share homology. Proceedings of the National Academy of Sciences, 82, 4331-4335.

RAND, L., HINDS, J., SPRINGER, B., SANDER, P., BUXTON, R. S. & DAVIS, E. O. (2003). The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. Molecular Microbiology, 50, 1031-1042.

RECHKOBLIT, O., MALININA, L., CHENG, Y., KURYAVYI, V., BROYDE, S., GEACINTOV, N. E. & PATEL, D. J. (2006). Stepwise Translocation of Dpo4 Polymerase during Error-Free Bypass of an oxoG Lesion. PLoS Biol, 4, e11.

SANDERS, G. M., DALLMANN, H. G. & MCHENRY, C. S. (2010). Reconstitution of the B. subtilis Replisome with 13 Proteins Including Two Distinct Replicases. Molecular Cell, 37, 273-281.

SASSANFAR, M. & ROBERTS, J. (1990). Nature of the SOS-inducing signal in Escherichia coli. The involvement of DNA replication. J Mol Biol, 212, 79 - 96.

SHI, W., ZHANG, X., JIANG, X., RUAN, H., BARRY, C. E., WANG, H., ZHANG, W. & ZHANG, Y. (2011). Pyrazinamide inhibits trans-translation in Mycobacterium tuberculosis: a potential mechanism for shortening the duration of tuberculosis chemotherapy. Science (New York, N.Y.), 333, 1630-1632.

SMITH, B. T. & WALKER, G. C. (1998). Mutagenesis and More: umuDC and the Escherichia coli SOS Response. Genetics, 148, 1599-1610.

SMYTH, D. R., MROZKIEWICZ, M. K., MCGRATH, W. J., LISTWAN, P. & KOBE, B. (2003). Crystal structures of fusion proteins with large-affinity tags. Protein Science : A Publication of the Protein Society, 12, 1313-1322.

SOTO, C. (2001). Protein misfolding and disease; protein refolding and therapy. FEBS Letters, 498, 204-207.

TIPPIN, B., PHAM, P. & GOODMAN, M. F. (2004). Error-prone replication for better or worse. Trends in microbiology, 12, 288-95.

TRINCAO, J., JOHNSON, R. E., ESCALANTE, C. R., PRAKASH, S., PRAKASH, L. & AGGARWAL, A. K. (2001). Structure of the Catalytic Core of S. cerevisiae DNA Polymerase η: Implications for Translesion DNA Synthesis. Molecular Cell, 8, 417-426.

TUROVEROV, K. K., KUZNETSOVA, I. M. & UVERSKY, V. N. (2010). The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. Progress in Biophysics and Molecular Biology, 102, 73-84.

WANG, Z. (2001). Translesion synthesis by the UmuC family of DNA polymerases. Mutation Research/DNA Repair, 486, 59-70.

WARNER, D., NDWANDWE, D., ABRAHAMS, G., KANA, B., MACHOWSKI, E., VENCLOVAS, C. & MIZRAHI, V. (2010a). Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. Proc Natl Acad Sci USA, 107, 13093 - 13098.

WARNER, D. F. (2010). The role of DNA repair in M. tuberculosis pathogenesis. Drug Discovery Today: Disease Mechanisms, 7, 5-11.

WARNER, D. F. & MIZRAHI, V. (2006). Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. Clinical microbiology reviews, 19, 558-70.

WARNER, D. F. & MIZRAHI, V. (2014). Translating genomics research into control of tuberculosis: lessons learned and future prospects. Genome Biology, 15, 514.

WARNER, D. F., NDWANDWE, D. E., ABRAHAMS, G. L., KANA, B. D., MACHOWSKI, E. E., VENCLOVAS, C. & MIZRAHI, V. (2010b). Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America, 107, 13093-8.

WEI, Y. (2005). Portraits of a Y-family DNA polymerase. FEBS Letters, 579, 868-872.

WHEELER, P. R. & RATLEDGE, C. 1994. Tuberculosis: Pathogenesis, Protection, and Control.

WONG, J. H., FIALA, K. A., SUO, Z. & LING, H. (2008). Snapshots of a Y-Family DNA Polymerase in Replication: Substrate-induced Conformational Transitions and Implications for Fidelity of Dpo4. Journal of molecular biology, 379, 317-330.

ZENG, Y.-H., SHEN, F.-T., TAN, C.-C., HUANG, C.-C. & YOUNG, C.-C. (2011). The flexibility of UV-inducible mutation in Deinococcus ficus as evidenced by the existence of the imuB–dnaE2 gene cassette and generation of superior feather degrading bacteria. Microbiological Research, 167, 40-47.