The DNA Binding Protein Lsr2 from *Mycobacterium tuberculosis*

A thesis submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Emma Louise Summers

The University of Waikato

2012

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Abstract

Lsr2 is a small, basic DNA binding protein that is highly conserved in mycobacteria and related actinomycetes. *Lsr2* is essential for growth in *Mycobacterium tuberculosis* and previous studies have shown that Lsr2 is involved in down-regulating a wide range of genes involved in cell wall synthesis and metabolic functions. This regulatory function is likely to influence bacterial growth and survival.

This research investigated the biochemistry and 3D structure of Lsr2 from *M. tuberculosis*. Transmission electron microscopy (TEM) analysis of Lsr2 in complex with DNA revealed a regular fibril-like arrangement of protein coating double-stranded DNA. In addition, it was shown that Lsr2 physically protected DNA from DNase activity. The structure of the C-terminal DNA binding domain of Lsr2, determined by others part way through this research, prompted site directed mutagenesis of residues proposed to interact with DNA. Modification of arginine residues significantly reduced the binding of Lsr2 to DNA and fibril-like structures were not observed using TEM, for arginine mutants.

The first crystal structure of the N-terminal domain of Lsr2 is reported here. Two high resolution structures in monoclinic and hexagonal space groups were solved using X-ray crystallography and *ab initio* phasing. Proteolytic processing of the N-terminus of Lsr2 was revealed by the structure in $P2_1$ and this process was recreated using the protease trypsin which resulted in crystal formation in a $P3_121$ space group. Both structures show linear chains of dimeric N-terminal Lsr2, as shown by crystallographic symmetry, linked by overlapping anti-parallel β-sheets, revealing a mechanism of protein oligomerisation. Oligomerisation only occurs after the removal of the first three residues M1, A2 and K3. In solution, protein oligomerisation was recreated with trypsin, resulting in the formation of large protein complexes. A change in DNA topology after the addition of trypsin to full-length Lsr2/DNA complexes was observed using TEM. This mechanism is likely to be important to *M. tuberculosis* under "stress" conditions where proteases are known to be upregulated and where cross-linking and condensation of DNA is critical.

Acknowledgements

There are a number of people who I would like to duly thank for their support during this study. To my supervisor Associate Professor Vic Arcus, thank you for your encouragement to begin this journey, and for the opportunity to work on this project and your guidance along the way. To my co-supervisor Dr Ray Cursons, thank you for sharing your expertise and for the helpful discussions.

To my work colleagues in the C2.10 Proteins and Microbes Laboratory, past and present, it was your help and support that enabled me to make it to my destination. I am grateful for the friendships that have developed along the way, as without these, life and studies would have been much more difficult. I would like to specifically thank Ali, Ashley, Emma, Jo M, Jo H, Jonathan, Judith, Line, Mat, Marisa and Svend. I could not have done this without your positive influence.

There are a number of people in the School of Biological Sciences at The University of Auckland that without their support I would not have been able to perform the diverse range of techniques accomplished during this study. I would like to specifically thank Chris, Ivan, Kavestri, Mazdak, Verne and Associate Professor Alok Mitra.

Thank you to my family and friends for all your support – I am sure you were all wondering when I would finally finish being a student. I am also grateful for being a recipient of a University of Waikato Doctoral Scholarship for the duration of my studies.

Lastly, thank you Dean for always being there and taking the time to reload all my ammunition so that I could take out my frustrations on the shooting range on the weekends.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACN	acetonitrile
ADP	adenosine-5’ diphosphate
ATP	adenosine-5’ triphosphate
bp	base pair(s)
BLAST	basic local alignment search tool
°C	degrees Celsius
C-terminal	carboxyl terminus of peptide chain
3D	three dimensional
Da	Daltons
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
EMSA	electrophoretic mobility shift assay
ESI	electrospray ionisation
ex	excitation
FPLC	past performance liquid chromatography
g	times the force of gravity
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His tag	poly histidine tag
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio-β-D-galactosidase
kb	kilobase
kDa	kiloDaltons

L	litre
LB	Luria Bertani broth
LBT	Luria Bertani broth and tween 80
MAD	multiwavelength anomalous dispersion
MALDI-TOF	matrix assisted laser desorption ionisation time of flight
mAU	milli absorbance units
MBP	maltose binding protein
MDR	multidrug resistant
MES	2-(N-morpholino)ethanesulfonic acid
MS	mass spectrometry
MQ	milli Q water – ion exchanged purified water
Native PAGE	non denaturing polyacrylamide gel electrophoresis
nm	nanometer
NMR	nuclear magnetic resonance
N-terminal	amino terminus of peptide chain
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pI	isoelectric point
RBS	ribosome binding site
RMSD	root mean square deviation
rpm	revolutions per minute
rTEV	recombinant tobacco etch virus

SEC	size exclusion chromatography
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOC	super optimal broth with catabolite repression
SSRL	Stanford synchrotron radiation laboratory
TAE	tris acetate ethylene diamine tetraacetic acid
TB	tuberculosis or terrific broth
TBS	tris buffered saline
TBS-T	tris buffered saline and tween 20
TE	tris ethylene diamine tetraacetic acid
TEMED	tetramethylethylenediamine
TG-SDS	tris glycine sodium dodecyl sulphate
Tm	melting temperature
TOF	time of flight
Tris HCl	2-amino-2-hydroxymethyl-propane-1,3-diol-HCl
µm	micrometer
UV	ultraviolet light (280 nm)
V	volts
v/v	volume per volume
WHO	world health organisation
w/v	weight per volume
XDR	extensively drug resistant
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside

Chapter One – Mycobacteria and Disease

1.1 Tuberculosis, the global disease

One hundred and thirty years have passed since the identification of *Mycobacterium tuberculosis* by Robert Koch in 1882. Koch's studies demonstrated that the disease tuberculosis (TB) was caused by the bacterium *M. tuberculosis* and that this organism is transmissible between animals (Koch, 1882). Today, infection with *M. tuberculosis* remains a major global health problem.

Tuberculosis is the leading cause of death from a curable infectious disease (Dye, 2006). Current World Health Organisation (WHO) statistics show that globally, an estimated 8.8 million new cases and 1.1 million deaths from TB occurred in 2010 (WHO, 2011). Reassuringly, the numbers of new cases of TB and the prevalence rates have been declining and the mortality rates have fallen by more than one third since 1990. Conversely, even though the global incidence is falling, it continues to slowly increase in one of the high burden countries, South Africa, where the proportion of TB cases co-infected with human immunodeficiency virus is also highest (WHO, 2011). Of further concern is the appearance of multi-drug resistant cases of TB (MDR-TB) that are more difficult to treat. Although the trend in the global occurrence of MDR-TB cannot be reliably assessed (because of the lack of direct measurement) a survey in 2000 estimated 273,000 new cases of MDR-TB (Dye, Espinal *et al.*, 2002) and in 2008 there were an estimated 440,000 cases (WHO, 2010), which suggests that MDR-TB cases are increasing worldwide. These statistics show that TB remains a major global health problem.

In New Zealand, although the annual TB disease notification rates have dropped more than 50% over the last 30 years (1980-2010), this trend has slowed and in the last five years there has been little change (Bissielo *et al.*, 2011). Cases of MDR-TB have remained steady in New Zealand over the last ten years but

notably the first case of extensively drug resistant TB (XDR-TB) was identified in 2010 (Bissielo *et al.*, 2011). The emergence of drug resistant strains of *M. tuberculosis* (Dye, Espinal *et al.*, 2002) creates further challenges for the successful treatment of this already deadly global disease.

1.2 Tuberculosis infection

M. tuberculosis is an aerobic organism that has alternative intracellular or extracellular life-states. It is a non-motile bacilli that is characterised by its slow growth (a generation time of approximately 24 hours), dormancy and distinctive cell envelope that differs substantially from the cell wall structures of both Gram-negative and Gram-positive bacteria (Cole *et al.*, 1998; Glickman & Jacobs Jr, 2001).

Primary infection with *M. tuberculosis* will lead to the manifestation of the disease in only around 10% of cases (Lillebaek *et al.*, 2002). It is widely known that the disease outcome is linked with the host environment, that is, the immune response of the natural host – humans (Glickman & Jacobs Jr, 2001). Most infected individuals are able to generate an immune response to the bacteria which will provide partial immunity but although the response against the bacteria is highly effective in controlling the proliferation and primary infection, the organism is almost never eradicated (Glickman & Jacobs Jr, 2001). This failure to eradicate the bacteria leads to asymptomatic latent infection. *M. tuberculosis* is exceptional amongst other pathogenic bacteria in its ability to establish and maintain latency, a period where a person does not exhibit any clinical signs, but the disease may be reactivated at a later date (Glickman & Jacobs Jr, 2001; Parrish *et al.*, 1998). The majority of latent infections never develop into the clinical form of the disease (nor transmit the disease) however the resilience of *M. tuberculosis* should not be underestimated as it has the potential to reactivate after thirty three years (Lillebaek *et al.*, 2002).

Most TB infections develop in the respiratory system but can involve other organs such as the lymph nodes, bones and joints. The disease is a chronic

wasting illness characterised by fever, weight loss and coughing (with blood in later stages) – which is both a symptom and the mechanism by which the disease is spread (Glickman & Jacobs Jr, 2001; Smith, 2003). In pulmonary TB, lesions develop in the alveoli of the lungs caused by the interaction of the immune system with the bacteria and this tissue becomes damaged and necrotic, forming cavities and causing bleeding (Schluger, 2005; Smith, 2003).

For most pathogens, such as *Staphylococcus aureus*, patients treated with antibiotics for two weeks are cleared of bacteria (Connolly *et al.*, 2007). However in *M. tuberculosis* infections the host immune response packages the bacteria into structures called granulomas, physically confining them (Schluger, 2005; Wayne & Sohaskey, 2001). Here the bacteria survive by evading the host immune system and drug treatment by adapting their metabolism and shifting into a non-replicating state (Murphy & Brown, 2008). Thus treatment of TB requires long treatment times – typically six to twelve months – as most antibiotics require actively dividing bacteria for their mode of action and while in the non-replicating state *M. tuberculosis* is phenotypically resistant to otherwise bactericidal antibiotics (Arcus *et al.*, 2006; Connolly *et al.*, 2007).

1.3 *Mycobacterium tuberculosis* persistence and resistance

Many of the symptoms of TB are a result of the host immune response to *M. tuberculosis* infection rather than any directly toxic effect of the bacterium itself. Understanding the interaction between the host immune response and *M. tuberculosis* remains a major goal of research into the pathogenic mechanism of this bacterium. Based on the clinical behaviour of *M. tuberculosis* there are several tasks that it must accomplish to successfully cause infection. It must be able to grow within the host macrophage, evade or modify the host immune response and be able to persist within the host in a non-replicating state but have the ability to reactivate (Glickman & Jacobs Jr, 2001).

In contrast to other bacterial pathogens, *M. tuberculosis* infects macrophages rather than avoids them and once inside, the bacteria reside within a vacuoule

called the phagosome. If the phagosome maturation cycle occurs (phagosome-lysosome fusion), the bacteria can encounter a hostile environment that is acidic in pH, contains reactive oxygen intermediates, lysosomal enzymes, reactive nitrogen intermediates and toxic peptides (Smith, 2003; Warner & Mizrahi, 2006). However it is known that *M. tuberculosis* alters the phagosome maturation cycle by modifying the cell proteins and transport mechanisms, reducing the hostility of the environment (to a slight decrease in pH), but in this process also restricting the supply of nutrients and essential elements. Persistence of *M. tuberculosis* relies on its ability to survive and replicate in these modified phagosomes (Smith, 2003; Warner & Mizrahi, 2006).

Persistence refers to the state of the bacterium where it stops growing when it encounters unfavourable conditions but is able to resume growth when the conditions improve. It is thought that initiation of the persistent state may be due to depletion of nutrients, changes in pH, the production of growth suppressing products and the depletion of oxygen (Wayne & Sohaskey, 2001). These conditions are present in the phagosome but there is much less information known about how the bacteria survives and grows during later stages of infection in the lung (Smith, 2003). Conditions in the necrotic centre of the granulomas are acidic and low in oxygen as a result of the battle between the immune system and *M. tuberculosis,* and it is postulated that the bacteria cannot multiply and remain dormant but alive. Depending on the effectiveness of the host's immune system, an infection may be arrested at this point or progress to the next stage. An arrested infection can persist throughout a person's life and they may never develop clinical signs of disease or it may become reactivated due to a change in their immune system. Treatment of dormant infections is difficult due to the lack of information in relation to how *M. tuberculosis* exists within the environment of the lung (Smith, 2003; Wayne & Sohaskey, 2001).

1.4 Treatment of tuberculosis

The current standard chemotherapy regime for active TB, called directly observed therapy short course or DOTS, involves the supervised administration of a multidrug combination for a minimum period of six months (Warner & Mizrahi, 2006). Because the majority of antibiotics require actively dividing bacteria for their mode of action, the dormant state of *M. tuberculosis* is thought to effectively make it resistant to usually bactericidal antibiotics (Connolly *et al.*, 2007). The strict DOTS treatment protocol involves administering four different drugs (isoniazid, rifampicin, pyrazinamide and ethambutol) and is designed to overcome the phenotypic resistance that *M. tuberculosis* exhibits while in the non-replicating state (Murphy & Brown, 2008).

The need for multi-drug and long-term therapy for TB is the result of both the phenotypic and inherited genetic resistance of *M. tuberculosis* to current front-line antibiotics. Simultaneous use of multiple antibiotics makes it less likely that a mutant resistant to any one of the drugs will survive whilst the extended treatment period ensures that bacteria leaving the persistent state will be killed as they begin to replicate (Connolly *et al.*, 2007). However, complications in curing the disease may arise if the treatment is abandoned mid-course, risking selection for resistant strains and in light of the emergence of multi-drug resistant strains (resistant to at least rifampicin and isoniazid) of *M. tuberculosis,* this has occurred.

No new antibiotics against TB have been developed in the past 30 years. The antibiotics isoniazid, rifampicin and pyrazinamide are considered first-line drugs with antibiotics such as ethambutol, streptomycin and para-aminosalicylic acid considered second-line (Arcus *et al.*, 2006). There is evidence that first-line therapy is effective in treating TB and containing MDR strains of TB (Dye, Williams *et al.*, 2002; WHO, 2008). However in 2006 the WHO announced that a deadly new strain of XDR-TB (MDR plus resistant to at least three of the six classes of second-line drugs) had been detected that was virtually untreatable and subsequently had extremely high fatality rates (Singh *et al.*, 2007). An urgent need for new antibiotics towards TB is widely acknowledged and the

goals of the Global Alliance for TB Drug Development that was established in 2000 include shortening treatment times, improving persistent *M. tuberculosis* and MDR/XDR TB treatment and the development of new adjuvant drugs (Arcus *et al.*, 2006).

1.5 The genetics of *Mycobacterium tuberculosis*

The complete DNA sequence of the virulent *M. tuberculosis* strain H37Rv was completed by investigators at the Pasteur Institute and published in June 1998 (Cole *et al.*, 1998). The genome comprises 4,411,529 base pairs (bp) and contains around 4,000 genes. Some of the features of the genome include guanine + cytosine (G+C) rich sequences and a large number of genes devoted to the synthesis of fatty acids (Cole *et al.*, 1998; Schluger, 2005). A general classification of *M. tuberculosis* H37Rv genes reveals that over 500 genes are annotated as coding for enzymes involved in cell wall and cell wall processes (13%) and over 200 genes (6%) encode enzymes involved in the metabolism of fatty acids (Smith, 2003). This result is not surprising as the cell envelope of *M. tuberculosis* contains an additional layer that is rich in unusual lipids, glycolipids and polysaccharides (Cole *et al.*, 1998) and as such there are original pathways to synthesise cell wall components such as mycolic acids, mycocerosic acid, lipoarabinomannan and arabinogalactan (Chatterjee, 1997; Morita *et al.*, 2011; Norman *et al.*, 1994). The large number of possible enzymes that may be involved in fatty acid metabolism in *M. tuberculosis* could be linked to the ability of this organism to grow in an environment where fatty acids may be the main carbon source. In comparison, other organisms such as *Escherichia coli* and *Streptomyces coelicolor* have noticeably fewer genes involved in fatty acid metabolism (Smith, 2003).

The completion and publication of the whole genome sequence of the virulent *M. tuberculosis* H37Rv strain enabled Gordon and researchers (Gordon *et al.*, 1999) to perform a complete genome comparison between *M. tuberculosis* H37Rv and the vaccine strain *Mycobacterium bovis* BCG-Pasteur using bacterial

artificial chromosome libraries together with the genome sequence of *M. tuberculosis* H37Rv. Their studies revealed 10 deleted loci in *M. bovis* BCG compared with *M. tuberculosis* H37Rv. Primers for polymerase chain reaction (PCR) of these 10 loci were used to investigate the presence of these regions in the other species that make up the *M. tuberculosis* complex (*Mycobacterium africanum, M. bovis and Mycobacterium microti*) and revealed an interesting distribution amongst the group. These differences in loci deletion between species provides information on their heritage and may offer clues to phenotypic and virulence differences (Gordon *et al.*, 1999).

Prior to Gordon's studies, sequence analysis of 26 structural genes or loci from *M. tuberculosis* strains and strains from three members of the *M. tuberculosis complex* (*M. africanum, M. bovis and M. microti*) revealed an absence of neutral gene mutations indicating that *M. tuberculosis* is young on an evolutionary scale and possibly underwent a recent evolutionary bottleneck resulting in a modern global spread (Sreevatsan *et al.*, 1997). Analysis of the combination of polymorphisms at two genetic markers enabled the authors to assign the members of the *M. tuberculosis* complex into one of three distinct genotypic groups. Interestingly *M. tuberculosis* fell into each of the three groups (1-3) whereas isolates of *M. bovis*, *M. microti* and *M. africanum* all fell into group one (1). Sreevatsan and colleagues propose an evolutionary scenario: that group 1 bacteria are ancestral to, and have greater genetic diversity than, groups 2 and 3. Their results from the analysis of 842 *M. tuberculosis* complex isolates from diverse geographic locations show that the *M. tuberculosis* genome is still evolving and indicate that the use of antibiotics may be a driving factor (Sreevatsan *et al.*, 1997).

At the time of writing, the genome sequences for fifteen mycobacterial species had been completed. These include five of the *M. tuberculosis* complex species (*M. tuberculosis* H37Rv, *M. africanum, M. bovis, M. bovis BCG, M. canettii*); the commonly studied pathogenic species *Mycobacterium leprae*, *Mycobacterium avium subsp. paratuberculosis*, *Mycobacterium avium* and *Mycobacterium ulcerans;* and the common laboratory model organisms *Mycobacterium*

smegmatis and *Mycobacterium marinum* (Shiloh & DiGiuseppe Champion, 2010) with the genome sequencing of ten more species in progress (NCBI, 2012). In the future, comparisons of sequences between additional mycobacterial species will allow the construction of evolutionary trees and improve our understanding about the evolution of these bacteria that inhabit such diverse environments.

1.6 *M. tuberculosis* Proteomics and Structural Biology

Just as the study of the DNA sequence of *M. tuberculosis* H37Rv can reveal a wealth of information about the organism, so too can the study of the proteins that are encoded by this DNA. Measuring the levels of proteins during different *M. tuberculosis* H37Rv growth conditions, such as nutrient starvation (Betts *et al.*, 2002) and high temperature (Stewart *et al.*, 2002), provides information on global gene expression of the organism. Technology to identify proteins of interest is available and along with sequence annotation, the putative function of many proteins may be predicted. However predicting protein function is not always a straight forward procedure. Bioinformatic analysis of sequenced genomes reveals that around 40% of gene products have an unknown function and of the 60% of gene products that have been functionally annotated, some are likely to be incorrect because these classifications are often derived by inference rather than by experimental means (Johnston *et al.*, 2003). Furthermore, annotated predicted protein function may still not give a complete account of a proteins' role in the bacterial cell as there may be additional posttranslational modification of the protein *in vivo* (Smith, 2003).

Despite this limitation, the mass of rapidly accumulating genomic sequence data provides an opportunity for researchers such as structural biologists because, as gene annotation in combination with molecular techniques reveals genes that are essential to organisms under specific growth conditions, biologists can target specific encoded proteins for structural studies (Arcus *et al.*, 2006). Structural biologists are concerned with the three dimensional (3D) molecular structure and function of biological macromolecules such as proteins. In most cases the

3D structures of proteins are solved through X-ray crystallography. The characterization of 3D structures of proteins is technically difficult and this is reflected by the smaller number of protein structures that have been solved over time in comparison with the vast amount of genetic data that has been revealed (Hrmova & Fincher, 2009). The 3D structure of a protein enables the prediction of its function and it is possible to find out information such as the mechanism of action of an enzyme, substrate binding site, protein-protein interaction sites or the binding sites in protein-DNA interactions. This information can be used to assist the selection of amino acids for mutation studies, to further study protein function (Hrmova & Fincher, 2009).

Approximately ten years ago, only eight protein structures (only two of which were not duplicates) from *M. tuberculosis* were available via the Protein Data Bank (PDB). Currently, there are over 1000 protein structures with around 500 of these being unique proteins within *M. tuberculosis* and the numbers are increasing daily with 10 structures submitted in the month of January 2012 alone. Hence, despite the technical challenges of successfully determining the 3D structure of proteins, this field of research is proving productive in nature and is providing valuable information about the biological function of proteins from *M. tuberculosis.*

1.7 LSR2 origins from *Mycobacterium leprae* studies

The bacterium *Mycobacterium leprae* is responsible for the disease leprosy which is a chronic infectious disease that mainly affects the skin and peripheral nerves. The protein LSR2 was first identified and named in February 1991 as the result of a study investigating antigen responses in the sera isolated from lepromatous leprosy patients to an expression library of *M. leprae* (Laal *et al.*, 1991). In this study, a *M. leprae* λgt11 recombinant DNA expression library was screened using pooled sera from patients and 4 clones that coded for β-galactosidase fusion proteins were identified. These clones were given the name “LSR” but the authors do not indicate their logic for this name and so it is

speculated that it may be derived from the phrase “lepromatous sera responder”. Clone LSR2 was used for further investigation of antibody and T-cell responses and the results demonstrated that the fusion protein LSR2 was able to act as a T-cell antigen and generate the same antibody response in leprosy patients as the sonicated native *M. leprae* bacilli. DNA sequencing of the LSR2 clone revealed an open reading frame (ORF) that coded for an 89 amino acid (aa) protein with a calculated size of approximately 10 kilodaltons (kDa) and an isoelectric point (pI) of 12 (Laal *et al.*, 1991).

Later in the same year, research was published where a *M. leprae* cosmid library was screened with sera from lepromatous leprosy patients (Sela *et al.*, 1991). The use of a cosmid library enabled DNA clones to be expressed from their natural transcription-translation start signals instead of being fused with another protein. Cloned proteins that showed a strong reaction to sera from patients with lepromatous leprosy were identified as being 15 kDa in size as shown by Western blot. Sequence analysis of the clone containing the coding sequence for this protein revealed an ORF with genetic information for this 15 kDa protein that included an upstream ribosome binding site and promoter region. A comparison of the 15 kDa protein sequence with the recently isolated clone LSR2 from a λgt11 recombinant DNA expression library (Laal *et al.*, 1991) revealed that LSR2 was identical to the 89 amino acids at the C-terminal end of the 15 kDa protein (Sela *et al.*, 1991).

These two studies both confirmed the antigenic nature of this protein but their contrasting sequence results suggest that the exact protein size is probably between 10 kDa and 15 kDa. Both authors searched sequence databases available at the time and were not able to identify similar sequences in other organisms. However, through Southern blotting experiments, sequences homologous to the 15 kDa gene were found in chromosomal DNA from *M. avium*, *M. bovis* BCG and *M. tuberculosis* H37Rv (Sela *et al.*, 1991). Both authors analysed the sequence of this protein for evidence of a signal sequence - as signified by a hydrophobic domain at the N-terminal end (Oliver, 1985). However signal sequence features were not found, which indicated that it is

unlikely that this protein is secreted or is part of the cell membrane. Incredibly, it would take thirteen years for their prediction that LSR2 was not secreted to be confirmed *in vivo* when it was isolated from the soluble fraction of armadillo derived *M. leprae* bacilli (Marques *et al.*, 2004).

Meanwhile, the amino acid residues responsible for generating an immune response to LSR2 in patients with inflammatory episodes of erythema nodosum leprosum were being investigated by researchers in the same laboratory as Laal and colleagues (Singh, Jenner *et al.*, 1994; Singh, Narayanan *et al.*, 1994). In this work, LSR2 was now called "LSR" and even though the discovery of the full-length (123 aa) protein by Sela *et al* (1991) was acknowledged in the manuscripts, these researchers focussed their experimental design on the original 10 kDa truncated annotation which lacks the N-terminal amino acids 1-23 shown by Sela and colleagues. Analysis of antibody responses to overlapping peptides generated from the LSR2 amino acid sequence shows two peptides of interest. Peptides 2 (p2) and 3 (p3), with an overlapping consensus sequence of IDLTNKNAA, gave the best antibody reaction (Singh, Narayanan *et al.*, 1994). A modern amino acid multiple sequence alignment of LSR2 reveals that residues IDL and NA from the overlapping consensus sequence are highly conserved amongst Lsr2 homologues (Figure 1.1). The consensus sequence had an antibody response that was lower than with the individual peptides used alone (Singh, Jenner *et al.*, 1994) but by sequentially removing the flanking amino acids N-terminal of p2 and C-terminal of p3 and testing for antibody response, the exact residues that were important for antibody recognition in erythema nodosum leprosum patients were identified (Singh, Jenner *et al.*, 1994). These are the residues GVTYE at the N-terminal end of p2 and the residues KLRGD at the C-terminal end of p3, that extend from the consensus sequence IDLTNKNAA.

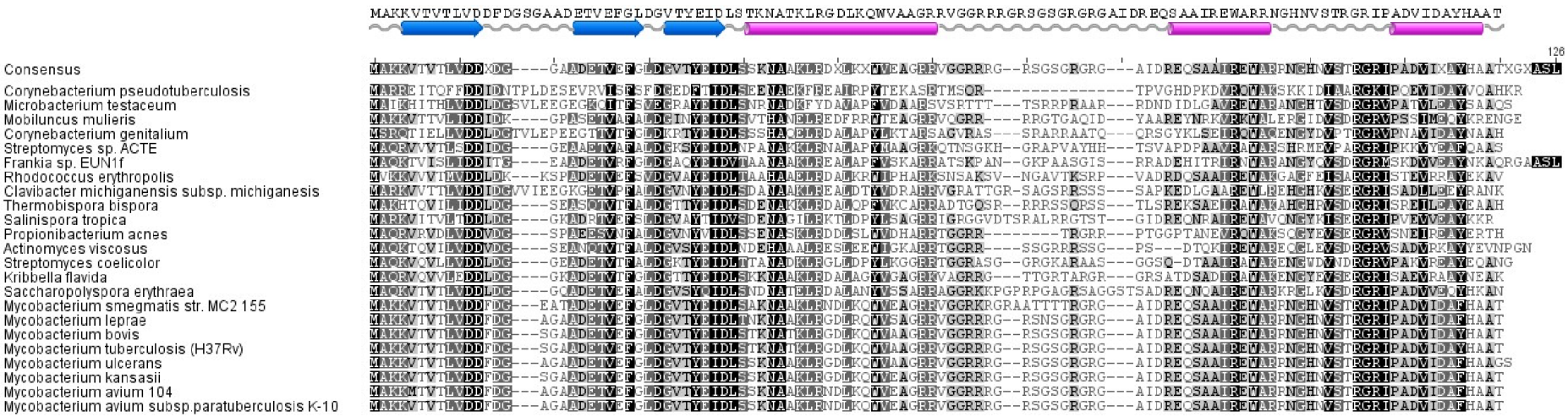


Figure 1.1. An amino acid multiple sequence alignment of Lsr2. The alignment was generated in Geneious Pro 5.4.2. (Drummond *et al.*, 2011). Dashes represent alignment gaps. The most highly conserved residues are shaded in black, followed by dark grey, then light grey with no shading for dissimilar residues. The amino acid sequence for *M. tuberculosis* H37Rv Lsr2 is shown above the alignment with predicted secondary structure elements (http://www.predictprotein.org) shown as blue arrows for β-sheets, pink cylinders for α-helices and a grey line for random coil. It was discovered that the start site of *M. avium* 104 was mis-annotated and as a consequence of this, the first four residues (MAKK) were manually added.

In contrast to these studies, a different subset of peptides gave the best T-cell response when the same overlapping peptides were used in a study with an aim to define the details of T cell recognition of LSR2 (Chaduvula *et al.*, 2012). A number of peptides were stimulatory but the best reaction across all treatment groups in this study was due to peptide 9 (residues GRSNSGRGRGAIDREQSA). An amino acid multiple sequence alignment of LSR2 reveals that the residues from peptide 9 are in a region of low conservation (Figure 1.1). It is clear that the peptides that preferentially activate T-cell responses are different from those that bind to circulatory antibodies.

In 2000, an extensive search of the newly completed *M. tuberculosis* H37Rv genome with the *M. leprae* LSR2 DNA sequence revealed a homologous *M. tuberculosis* ORF of 336 bp, translating to 112 residues and encoding a 12.1 kDa protein (Oftung *et al.*, 2000). This gene was annotated as Rv3597c within the *M. tuberculosis* H37Rv genome (Cole *et al.*, 1998) and a sequence alignment showed 93% identity between *M. leprae* LSR2 and *M. tuberculosis* Rv3597c at the amino acid sequence level (Oftung *et al.*, 2000). With the identification of a homologous counterpart to the LSR2 antigen from *M. leprae*, in *M. tuberculosis,* and evidence that a corresponding *M. tuberculosis* antigen exists at the protein level (Sela *et al.*, 1991) LSR2 had now caught the interest of researchers studying *M. tuberculosis*.

Researchers of LSR2 in *M. leprae* have used the term LSR2 (Laal *et al.*, 1991; Marques *et al.*, 2004) and LSR (Oftung *et al.*, 2000; Singh, Jenner *et al.*, 1994; Singh, Narayanan *et al.*, 1994) and more recently, Lsr2 (Chaduvula *et al.*, 2012). The term Lsr2 is characteristic with research on the equivalent *M. tuberculosis* H37Rv protein and from now on this protein will consistently be referred to as Lsr2.

1.8 Lsr2 from *Mycobacterium tuberculosis*

1.8.1 Lsr2 DNA and protein sequence analysis

Lsr2 is encoded by the 339 bp gene Rv3597c of *M. tuberculosis* H37Rv. It is 112 amino acids in length (Figure 1.2), has a molecular weight of 12.1 kDa and a theoretical pI of 10.

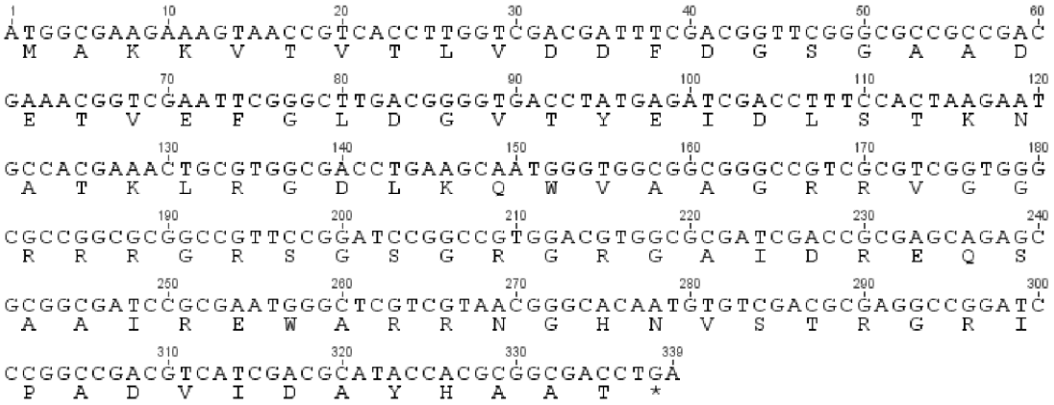


Figure 1.2. The nucleotide sequence and the amino acid translation of Lsr2 (Rv3597c). The gene for Lsr2 is 339 bp in length and encodes 112 aa. Numbering indicates the number of base pairs and the stop codon is indicated by a *. This image was prepared in Geneious Pro 5.4.2. (Drummond *et al.*, 2011).

Examination of the amino acid composition of the protein sequence reveals that it is arginine rich (13.4%) making it highly basic and positively charged. The total number of positively charged residues (Arginine-15, Histidine-2 and Lysine-5) outnumbers the number of negatively charged residues (Aspartic Acid-10 and Glutamic Acid-5). A large number of the small residues glycine and alanine are present in Lsr2 and make up 13.4% and 12.5% of the total residues, respectively (Table 1.1). There is a region of low complexity present within the unstructured region between residues 53 and 74 (Figure 1.1).

Table 1.1. The amino acid composition of Lsr2 (Rv3597c). The twenty amino acids are listed in alphabetical order and their corresponding three-letter and single-letter codes noted alongside. The dominant characteristics of each residue are listed. Lsr2 is arginine rich making it highly basic and positively charged.

Amino Acid	Codes	No. of Residues	Percentage (%)	Residue Characteristics
Alanine	Ala (A)	14	12.5%	Aliphatic, hydrophobic
Arginine	Arg (R)	15	13.4%	Basic, positive charge
Asparagine	Asn (N)	3	2.7%	Amide group, uncharged
Aspartic Acid	Asp (D)	10	8.9%	Acidic, negative charge
Cysteine	Cys (C)	0	0.0%	Sulphur-containing
Glutamine	Gln (Q)	2	1.8%	Amide group, uncharged
Glutamic Acid	Glu (E)	5	4.5%	Acidic, negative charge
Glycine	Gly (G)	15	13.4%	Aliphatic, hydrophobic
Histidine	His (H)	2	1.8%	Basic, positive charge
Isoleucine	Ile (I)	5	4.5%	Aliphatic, hydrophobic
Leucine	Leu (L)	5	4.5%	Aliphatic, hydrophobic
Lysine	Lys (K)	5	4.5%	Basic, positive charge
Methionine	Met (M)	1	0.9%	Sulphur, hydrophobic
Phenylalanine	Phe (F)	2	1.8%	Aromatic, hydrophobic
Proline	Pro (P)	1	0.9%	Aliphatic, hydrophobic
Serine	Ser (S)	6	5.4%	Polar, uncharged
Threonine	Thr (T)	8	7.1%	Polar, uncharged
Tryptophan	Trp (W)	2	1.8%	Aromatic, hydrophobic
Tyrosine	Tyr (Y)	2	1.8%	Aromatic, hydrophobic
Valine	Val (V)	9	8.0%	Aliphatic, hydrophobic

DNA sequence analysis revealed that *lsr2* orthologues are present in all sequenced mycobacterial genomes, including *M. tuberculosis*, *M. smegmatis*, *M. bovis*, *M. leprae*, *M. africanum*, *M. ulcerans*, *M. marinum*, *M. avium*, and *M. avium* subsp. *paratuberculosis*. The amino acid sequence of Lsr2 is highly conserved among mycobacteria and other actinobacteria. The high level of sequence identity among the Lsr2 homologues can be seen in a multiple amino acid sequence alignment of a selection of organisms (Figure 1.1) and suggests that these proteins perform a similar function amongst the identified group of related organisms.

Residues that are highly conserved are thought to be involved in the three dimensional or higher order structures of proteins and/or involved with protein function. The predicted secondary structure of *M. tuberculosis* H37Rv Lsr2,

alongside the multiple sequence alignment (Figure 1.1) shows that 60% of the most highly conserved residues are positioned within a secondary structure motif. Secondary structure regions make up 53% of the total residues. The remaining highly conserved residues flank a secondary structure motif. In the area of low residue conservation across species, there is lack of predicted secondary structure.

Another method of examining amino acid conservation is by generating a model based on the probability of the position of each amino acid by performing a Hidden Markov Model (HMM) analysis. A graphical representation of HMM analysis (called an HMM logo; (Schuster-Bockler *et al.*, 2004)) of Lsr2 sequences shows the conservation of amino acid positions in the alignment (Figure 1.3). Amino acid conservation is represented in stacked columns and letter height indicates residue frequency at each position. The HMM logo reveals several additional residues of significance compared with the highly conserved residues identified in the protein sequence alignment (Figure 1.1) such as alanine in position 2, aspartic acid at position 14 and leucine in positions 44 and 48. HMM logos also provide valuable information about residue characteristics at each position as colour is used to represent residue properties. For example, the other (green-coloured) aromatic residues tyrosine (Y) and phenylalanine (F) substitute tryptophan (W) more often in other family members compared with Lsr2 in *M. tuberculosis* H37Rv. Analysis and characterisation of the amino acid residues in a protein is an informative tool as it can give clues about how the protein may behave *in vitro* and *in vivo*.

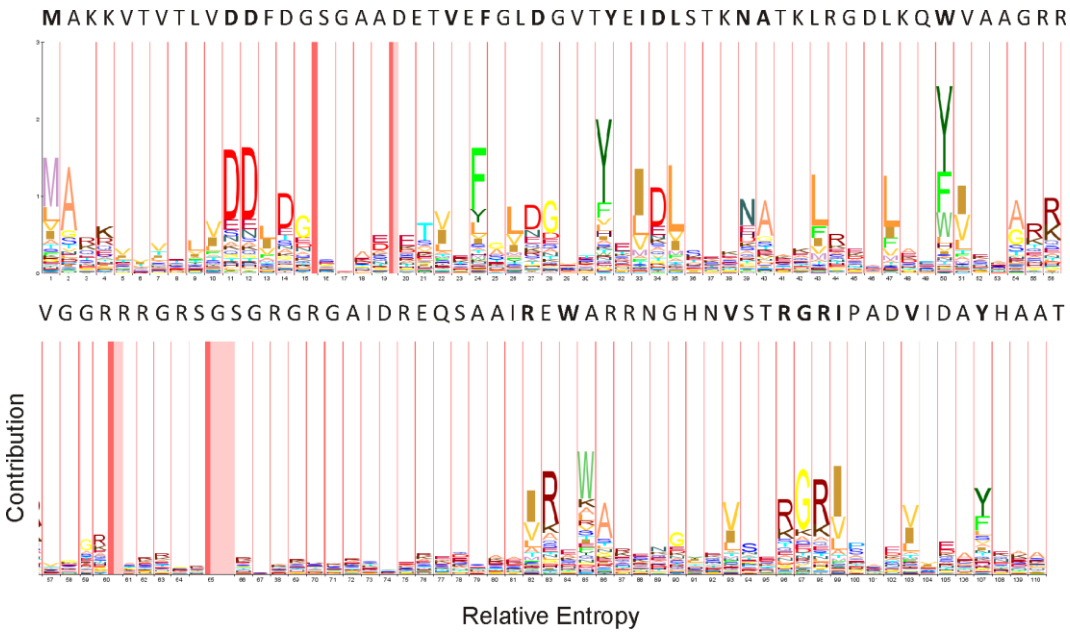


Figure 1.3. Hidden Markov Model (HMM) logo of the Lsr2 family shows the distribution of amino acids from a multiple alignment (http://pfam.sanger.ac.uk/family/PF11774). The height and width of each letter at each position is proportional to its occurrence and relative role, respectively. The pink bars represent regions of residue insertion in the alignment. Amino acid colours reflect their biological properties (red = charged; blues = polar, uncharged; yellows = aliphatic; greens = aromatic). The *M. tuberculosis* Lsr2 amino acid sequence is shown across the top of the HMM logo for comparison. Residues in bold type are highly conserved as identified in the consensus sequence alignment (Figure 1.1) generated in Geneious Pro 5.4.2.

1.8.2 Early studies on Lsr2

Lsr2 (Rv3597c) was first described as a negative regulator of transcription in studies investigating the transcriptional regulation of multi-drug tolerance in *M. tuberculosis* H37Rv (Colangeli *et al.*, 2007). Prior to this, Alland *et al.* discovered a mechanism of multi-drug tolerance in *M. tuberculosis* while investigating differential gene expression in response to the antibiotic isoniazid (INH) that inhibits cell wall synthesis (Alland *et al.*, 1998). In their study, three previously uncharacterised genes (*ini*B, *ini*A and *ini*C, organised in an operon (*ini*BAC)) were shown to be induced by INH, with the exception of *ini*A that was also induced by the antibiotic ethambutol, a cell wall synthesis inhibitor with a mechanism distinct from INH. The promoter region of this three-gene operon was further characterised, revealing genetic regulatory elements essential for gene expression and providing evidence that both activator and repressor molecules are involved in regulating this operon (Alland *et al.*, 2000). *In vivo*

studies of the individual genes of the *ini*BAC operon in *M. tuberculosis* H37Rv and *M. bovis* BCG confirmed they are induced by INH and reiterated that *ini*A is also responsive to ethambutol (Colangeli *et al.*, 2005). This *in vivo* research focused on *ini*A and showed that deleting *ini*A from *M. tuberculosis* H37Rv decreased its survival when exposed to INH and over-expressing *ini*A in *M. bovis* BCG increased its survival under INH induced stress. Analysis of two-dimensional protein crystals of *ini*A revealed a structure that could function as a transmembrane pump and led to the authors' hypothesis that this could be a mechanism that could provide multi-drug tolerance in *M. tuberculosis* H37Rv.

Colangeli and colleagues' transcription regulation research in 2007 involved generating a transposon mutagenesis library in *M. smegmatis* mc^2155 of the *M. tuberculosis* H37Rv *ini*BAC promoter fused to a lacZ reporter gene. The purpose of this library was to identify genes involved in the down-regulation of expression of the *M. tuberculosis* H37Rv *ini*BAC genes. In a colourimetric growth assay, mutants that had transposon insertions in genes involved in repressing the transcription of the *ini*BAC promoter were identified as blue colonies on media containing X-gal. The interrupted gene identified in these blue mutants in *M. smegmatis* mc^2155 was *lsr2*. In growth assays, the *lsr2* transposon mutant was shown to be more resistant to ethambutol and this was postulated to be due to lack of repression of the *ini*A gene in the *ini*BAC operon, as complementation of the gene restored antibiotic susceptibility. The researchers also performed a microarray analysis of a *lsr2* knock-out in *M. smegmatis* mc^2155 to identify genes that were up- or downregulated by deletion of *lsr2,* and showed that *ini*A and genes involved with cell wall and metabolic function were upregulated. Further experiments over expressing *lsr2* in *M. tuberculosis* H37Rv reduced the INH-induced transcription of the *ini*BAC genes, adding additional evidence to the hypothesis that Lsr2 has an important role in the regulation of multi-drug tolerance (Colangeli *et al.*, 2007).

In addition to transcription response studies, these researchers showed that overexpression of Lsr2 enabled *M. smegmatis* mc^2155 to survive for longer in culture when exposed to H_2O_2 and also within murine macrophages. The *lsr2*

knock-out in *M. smegmatis* mc^{2}155 was more susceptable when exposed to H_2O_2 and survival was significantly affected within murine macrophages This suggested that Lsr2 protects mycobacteria against reactive oxygen intermediates *in vitro* and during macrophage infection (Colangeli *et al*., 2009).

1.8.3 The role of Lsr2 in colony morphology and biofilm formation

M. smegmatis mc^{2}155 transposon insertion mutants generated by Colangeli *et al*., (Colangeli *et al.*, 2007) that successfully alleviated the repression of the *ini*BAC promoter had unusual round, wet and shiny colony morphology compared with the wild-type undulate, dry and rough colony phenotype. This altered colony phenotype was also reported previously for *M. smegmatis* mc^{2}155 *lsr2* transposon mutants obtained during the generation of colony morphology mutants in the investigation of cell wall composition (Chen *et al.*, 2006). These *lsr2* transposon (colony morphology) mutants were also defective in pellicle and biofilm formation, and apolar lipid formation. In spontaneously occurring *M. smegmatis* mc^{2}155 transposon mutants, that show smooth colony morphology, *lsr2* was again identified as the inactivated gene and varying features such as increased sliding mobility, lack of biofilm formation (Arora *et al.*, 2008), or increased biofilm formation, increased glycopeptidolipids and lack of pigment formation have been reported (Kocíncová *et al.*, 2008). These diverse morphological results nevertheless indicate that Lsr2 plays an important role in colony morphology and biofilm formation in *M. smegmatis* mc^{2}155, possibly through control of lipid production (Chen *et al.*, 2006).

It should be noted that there were a variety of techniques used in the literature (Table 1.2) to generate *M. smegmatis lsr2* knock-out mutants. Hence such varying morphological features noted in *M. smegmatis* mc^{2}155 strains could potentially be influenced by mutation technique.

Table 1.2. Description of *M. smegmatis lsr2* knock-out mutants and their construction.

Strain	Description	Reference
NJS20.1	*M. smegmatis* strain mc^{2}155 containing pG21898-12 inserted into chromosome at attP and a transposon element inserted into *lsr2*.	Alland *et al*., (2000) Colangeli *et al*., (2007)
NJS22	*M. smegmatis* strain mc^{2}155 with *lsr2* deleted (unmarked deletion).	Colangeli *et al*., (2007)
MS8444	*M. smegmatis* strain mc^{2}155 containing transposon insertion in *lsr2* gene.	Chen *et al*., (2006) Gordon *et al*., (2010)
Sm8/Sm9	Spontaneous mutants of *M. smegmatis* strain ATCC607 with mobile element inserted into *lsr2* gene.	Kocincova *et al*., (2008)
DL1215	Spontaneous mutant appearing from *M. smegmatis* mc^{2}155 Δrel_{Msm} strain. Single base pair deletion in *lsr2*.	Arora *et al*., (2008)
DL2008	*M. smegmatis* strain mc^{2}155 containing *lsr2* single base pair deletion mutant gene from DL1215.	Arora *et al*., (2008)
Δ*esx*-1 *lsr2*:Km	*M. smegmatis* strain mc^{2}155 Δ*esx*-1 containing transposon insertion in *lsr2* gene (Hygro/Kana).	Nguyen *et al*., (2010)
MKD158Δ*lsr2*	Derivative of *M. smegmatis* strain mc^{2}155 *lsr2* knock out containing kanamycin resistance cassette.	Nguyen *et al*., (2010)

1.9 Bacterial DNA binding proteins

Eukaryotic organisms contain small abundant DNA binding proteins responsible for compacting DNA into nucleosomes (Drlica & Rouviere-Yaniv, 1987). In prokaryotes, the chromosomal DNA is folded into a compact structure called a nucleoid and a number of proteins, collectively named histone-like proteins or nucleoid-associated proteins (NAP), are thought to be involved in this process (Dame, 2005; Dame & Goosen, 2002). The wide range of diverse DNA binding proteins found in bacteria suggest that there are multiple mechanisms used to solve the DNA packing conundrum (Tendeng & Bertin, 2003). The most studied and abundant members of the histone-like proteins are heat unstable protein (HU), factor for inversion stimulation (Fis), and heat stable nucleoid structuring protein (H-NS) (Dame & Goosen, 2002).

1.9.1.1 Heat unstable protein - HU

HU is similar to the eukaryotic histone protein as it is small (90 aa, 9.7 kDa), basic in nature and expressed at a high level. The amino acid sequence of HU is well conserved and homologues have been identified in all bacteria studied to date. HU is found as both a homodimer and heterodimer, binding and wrapping DNA in a non specific manner, and is capable of forming higher-order structures (Dame & Goosen, 2002; Drlica & Rouviere-Yaniv, 1987). The exact mode of DNA binding of HU is still being actively researched (Koh *et al.*, 2011) but it is known that HU prefers supercoiled over relaxed or linear lengths of DNA (Shindo *et al.*, 1992). Studies have also shown that at low concentrations HU binds to DNA and introduces bends, but as protein levels increase, polymerisation of HU occurs causing it to form rigid helical filaments that coat DNA without causing compaction (Figure 1.4)(Thanbichler *et al.*, 2005). The three dimensional structure of a *Bacillus stearothermophilus* HU protein reveals that there are several positively charged arginine and lysine residues implicated in binding in the minor groove of DNA (Tanaka *et al.*, 1984; White *et al.*, 1999).

1.9.1.2 Factor of inversion stimulation - FIS

Like the DNA binding proteins HU and H-NS, FIS is small (98 aa, 11.2 kDa), binds DNA in both a sequence-specific and non-specific manner and is abundant during exponential cell growth (Schneider *et al.*, 2001; Thanbichler *et al.*, 2005). FIS association with DNA introduces between 40° - 90° bends (Zhang *et al.*, 2004) and at low concentrations, forms aggregations at apical loops, giving a branched appearance. As the amount of FIS protein increases, the DNA becomes progressively more compacted as FIS binds to less specific sites (Figure 1.4)(Schneider *et al.*, 2001). FIS exists as a homodimer and has the common DNA binding protein "helix-turn-helix" structural motif which is found at the C-terminal end of the protein (Dame, 2005; Yuan *et al.*, 1991). The binding mechanism for FIS is proposed to cause the change in DNA topology. The three dimensional structure of FIS shows that the distance between the two helix-turn-helix motifs in the dimer is too small for binding in adjacent major

grooves in straight DNA helices but is the perfect spacing once the DNA is modelled with a 59° bend angle (Yuan *et al.*, 1991). Mutation of the positive arginine residues in the helices of the helix-turn-helix motif that interact with the phosphate backbone of the DNA results in a significant reduction in DNA binding (Yuan *et al.*, 1991).

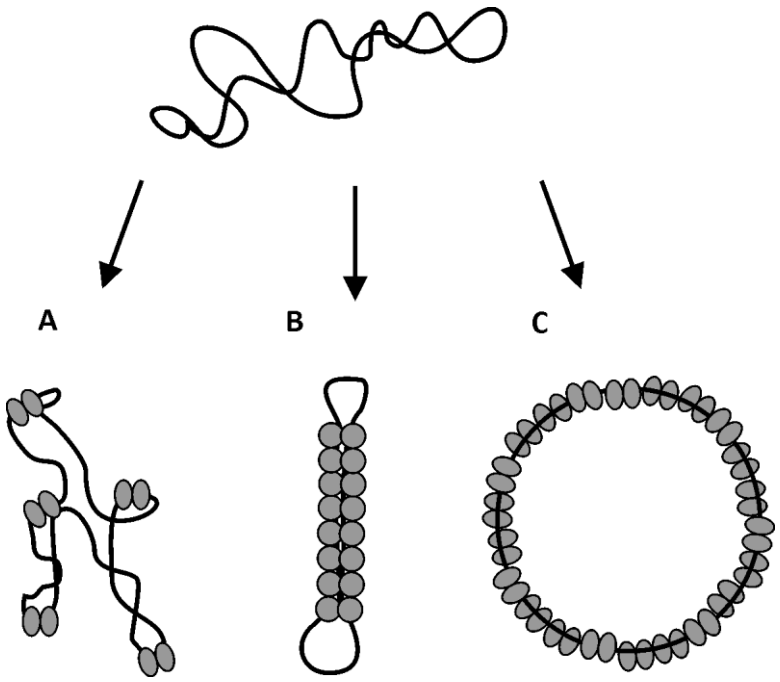


Figure 1.4. The effect of histone like proteins FIS, H-NS and HU on the conformation of a supercoiled plasmid. (A) The binding of FIS introduces bends in the DNA and promotes branching of supercoiled plasmid DNA; (B) At high concentrations, H-NS assembles into groups that interact with each other forming filament-like DNA structures; (C) HU forms rigid helical filaments that coat the DNA. This figure was recreated from (Thanbichler *et al.*, 2005).

1.9.1.3 Heat stable nucleoid structuring protein – H-NS

There is a wealth of information available on the DNA binding protein H-NS. Like HU, H-NS is small (137 aa, 15.5 kDa), binds DNA in a non specific manner and can form higher-order structures (Dame & Goosen, 2002). H-NS is organised into two domains separated by a flexible linker; the N-terminal dimerisation domain and the C-terminal DNA binding domain (Rimsky, 2004; Tendeng & Bertin, 2003). Three dimensional structures for each domain have been solved separately (Shindo *et al.*, 1995) (Renzoni *et al.*, 2001) presumably because obtaining the full

length protein structure is difficult due to the flexible linker (Dame, 2005). The most recent N-terminal dimerisation domain structure shows this domain in a super helical structure with anti-parallel dimerisation and provides an excellent model of oligomerisation and DNA compaction by H-NS (Arold *et al.*, 2010). Within the conserved sequence residues in the C-terminal domain, a motif "X-G-R" is found (situated in the loop) that is involved with DNA binding and the arginine residues have subsequently been shown to interact with the minor groove of DNA (Figure 1.6B)(Gordon *et al.*, 2011; Tendeng & Bertin, 2003). H-NS is able to assemble into groups of oligomers on DNA strands that then interact with each other, linking regions of DNA together, with subsequent protein polymerisation leading to rod like filament formation (Figure 1.4)(Schneider *et al.*, 2001; Thanbichler *et al.*, 2005).

H-NS homoloques are widespread in Gram-negative bacteria but not so in Gram-positive bacteria (Gordon *et al*., 2010). A homologue of the H-NS protein in *M. tuberculosis* H37Rv has been identified (Sharadamma *et al.*, 2010) as well as other DNA binding proteins such as Hlp (Lee *et al.*, 1998) and MsDps2 (Saraswathi *et al.*, 2009) during investigations of dormancy, and nutritional starvation.

1.9.2 DNA binding and biological function of Lsr2

Lsr2 was first described as a DNA binding protein in the 2007 study on transcriptional regulation of multi-drug tolerance in *M. tuberculosis* H37Rv (Colangeli *et al.*, 2007). In this study Lsr2 was likened to a histone-like protein as it has a small molecular weight and a high isoelectric point, properties that are also features of histone-like proteins. Even though Lsr2 shows no sequence homology with proteins in the histone-like protein class, experimental evidence showed that it had an ability to bind DNA, forming large oligomers, and could inhibit *in vitro* transcription and topoisomerase activity. This experimental evidence demonstrated that Lsr2 had an ability to interact with DNA and suggested that Lsr2 could regulate gene expression by controlling chromosomal topology (Colangeli *et al.*, 2007).

In other biochemical studies, the interaction of *M. tuberculosis* H37Rv Lsr2 with DNA provided evidence that Lsr2 binds non-specifically to DNA (but with greater preference for AT-rich DNA sequences) and forms increasingly larger complexes with DNA as DNA length increases. DNA fragments were designed to test if Lsr2 preferred curved DNA and could therefore be compared to the H-NS proteins and the results show that Lsr2 has no preference for curved or non-curved DNA, setting it apart from H-NS (Chen *et al.*, 2008). *In vivo* complementation experiments in *M. smegmatis* mc^{2}155 were used to determine the key amino acid residues of *M. smegmatis* mc^{2}155 Lsr2 by examining the ability of mutants to restore the colony morphology phenotype of an *lsr2* knockout. Lsr2 mutants with the amino acid substitutions D28A, R45A and R86A were unable to restore the colony phenotype. Residues R45 and R86 were noted to be within predicted helical regions whereas residue D28 was within a predicted random coil (Figure 1.5)(Chen *et al.*, 2008). Further experiments utilising these three mutants enabled the authors to conclude that residues D28 and R45 are involved in Lsr2 dimer formation and residue R86 and its equivalent in *M. tuberculosis* H37Rv (R84) are important for DNA binding (Figure 1.5). Although some of the features of Lsr2 noted in this study are unlike the H-NS proteins the authors also provided evidence that Lsr2 has DNA bridging ability (as shown by atomic force microscopy) which is characteristic of H-NS and they suggested that Lsr2 may have its own DNA binding domain, connecting Lsr2 to a role in nucleoid compaction and transcription repression (Chen *et al.*, 2008).

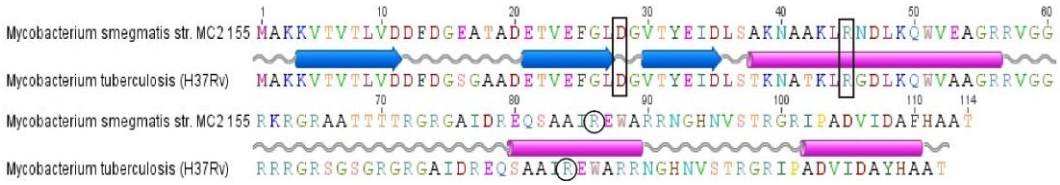


Figure 1.5. Lsr2 protein sequences from *M. smegmatis* mc^2 155 and *M. tuberculosis* H37Rv showing predicted secondary structure and residues involved in protein dimerisation and DNA binding. Mutations of the residues aspartate 28 (D28A) and arginine 45 (R45A) in *M. smegmatis* Lsr2 (boxed regions) prevented Lsr2 dimer formation; whereas mutation of the equivalent arginine 86 (R86) residue in *M. tuberculosis* Lsr2 (R84A) (circles) showed reduced DNA binding. This image was partially generated in Geneious Pro 5.4.2. (Drummond *et al.*, 2011). Predicted secondary structure elements (http://www.predictprotein.org) for *M. tuberculosis* are shown as blue arrows for β-sheets, pink cylinders for α-helices and grey line for random coil.

Direct evidence that Lsr2 is a functional homologue of H-NS was demonstrated soon after by researchers from the same lab as Chen and colleagues. In these studies, *in vivo* complementation experiments showed that the *M. tuberculosis lsr2* gene was able to complement all the phenotypes related to *hns* gene mutations in *E. coli*. Equally, *hns* from *E. coli* was able to complement the phenotype of an *lsr2* mutant in *M. smegmatis*, demonstrating the functional similarity of Lsr2 and H-NS. In addition to these phenotypic studies, it was demonstrated that Lsr2 could bind specifically to H-NS regulated genes *in vivo* suggesting similar transcriptional repression mechanisms between Lsr2 and H-NS, and similar global gene regulation functions (Gordon *et al.*, 2008).

Considering the lack of sequence similarity between Lsr2 and H-NS it is remarkable that they share apparently identical functions. Especially considering that there is an H-NS homologue in *M. tuberculosis* H37Rv. These two proteins also lack any similarity in their predicted secondary structures, however they may share similar domain organisation. All H-NS proteins are predicted to be organised into two functional domains, the N-terminal oligomerisation domain and the C-terminal DNA binding domain, separated by a flexible linker (Bertin *et al.*, 1999). More specifically, residues 1 - 83 make up the N-terminal domain and residues 91 - 137 make up the C-terminal domain of H-NS (Arold *et al.*, 2010). Recently, the three dimensional structure of the C-terminal DNA binding domain of Lsr2 was determined (Gordon *et al.*, 2010) indicating that the N-terminal dimerisation domain of Lsr2 is between residues 1-65 and the C-terminal DNA binding domain is within residues 51-112. This also confirmed the domain similarities between Lsr2 and H-NS.

The structure of the C-terminal DNA binding domain of Lsr2 from *M. tuberculosis* H37Rv was solved using nuclear magnetic resonance techniques (Figure 1.6A). The structure of this domain highlights the specific residues that may be involved in interacting with DNA and shows that there is functional similarity to H-NS but no structural similarity (Gordon *et al.*, 2010). The DNA binding domain contains two α-helices (α1, residues 79-89; α2, residues 102-112) linked by a long loop

(residues 90-101). This structure is unlike that of H-NS which has two short β-sheets and an α-helix linked by a loop (Figure 1.6B)(Shindo *et al.*, 1995).

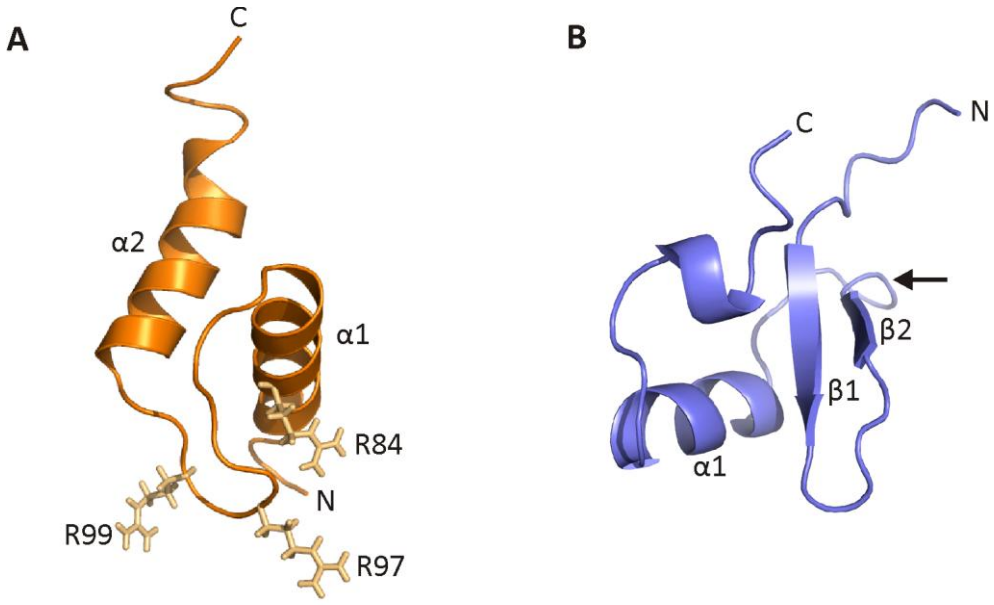


Figure 1.6. The C-terminal DNA binding domains of Lsr2 and H-NS are structurally distinct. A cartoon representation of the structure solutions of the Lsr2 (A) DNA binding domain containing two α-helices linked by a long loop (PDB; 2KNG) and the H-NS (B) DNA binding domain has two short β-sheets linked to an α-helix (PDB; 1HNR). The arginine residues R84, R97 and R99, responsible for DNA interactions in Lsr2 are depicted as sticks. The arrow indicates the loop in H-NS thought to interact with DNA. The chain termini, α-helices and β-sheets are labelled and protein depictions were rendered with PYMOL.

A structure model generated to depict Lsr2 in complex with DNA revealed that the C-terminal domain of Lsr2 can bind DNA in two different orientations by grabbing either edge of the minor groove of DNA much like a clamp, orientating and inserting residues R97, G98 and R99 into the minor groove. Outside of the minor groove, a number of arginine (e.g. R84) and serine residues interact with the phosphate backbone on either edge of the minor groove, increasing the binding affinity of Lsr2 to DNA (Gordon *et al.*, 2010). The mechanism of DNA recognition by the residues R97-G98-R99 – as a result of their positive charge and extended loop conformation - was shown to be unique in these studies. In a multiple sequence alignment of the Lsr2 family it is clear that residue R84, in addition to the R97-G98-R99 motif, are highly conserved (Figure 1.3) indicating

their significance in the mechanism of DNA binding. Even though the solution structures of the C-terminal domains of Lsr2 and H-NS are so different there is strong evidence for their functional similarity as shown with protein domain swapping experiments where the N-terminal and C-terminal domains of each protein were interchanged and complete Lsr2 and H-NS protein function was still demonstrated (Gordon *et al.*, 2010).

H-NS is a regulator of transcription, controlling the expression of a larger number of genes involved in virulence (Lucchini *et al.*, 2006). There are a large number of bacterial nucleoid associated proteins (such as histone-like proteins) that are all involved in the regulation of transcription and evidence now suggests that Lsr2 from *M. tuberculosis* H37Rv also belongs to this group. Microarray studies show that Lsr2 regulates genes involved in *M. smegmatis* mc^2155 metabolism (Colangeli *et al.*, 2007) and that Lsr2 is up-regulated at high temperatures (Stewart *et al.*, 2002) and after nutrient starvation (Betts *et al.*, 2002), highlighting its importance in global gene regulation and bacterial response to environmental changes. To date, the generation of an *M. tuberculosis* H37Rv *lsr2* knockout has been unsuccessful, suggesting that Lsr2 is an essential gene in *M. tuberculosis* H37Rv (Sassetti *et al.*, 2003) and is worthy of continued investigation into its role and function in mycobacteria.

1.9.3 PhD study objectives

Lsr2 is encoded by an essential gene in *M. tuberculosis* and is an important DNA binding protein found in *M. tuberculosis* and other mycobacteria (Sassetti *et al.*, 2003). This protein has been shown to be involved in a number of critical biological functions including chromosome organisation and gene regulation (Colangeli *et al.*, 2007). The structural organisation of this protein is able to tell us a number of things about its potential biological roles.

Proteins are commonly made up of domains that perform specific functions. Biochemical studies have illustrated that Lsr2 comprises an N-terminal dimerisation domain and a C-terminal DNA binding domain and is capable of functioning in a similar manner to other histone-like DNA binding proteins (Chen *et al.*, 2008).

The aims of this research were to determine how Lsr2 interacts with and binds to DNA, and how higher order Lsr2/DNA structures were formed. The 3D structure of Lsr2 was of great interest as it may reveal critical functional information.

Chapter Two - Materials & Methods

2.1 General materials and methods

2.1.1 Primer design, supplier and reconstitution

Primers for restriction digest and ligation cloning were designed to have 5' and 3' restriction enzyme sites with an additional number of bases (depending on the enzyme used) at each end to allow for efficient cleavage. Primers designed for site directed mutagenesis, amplification of plasmid inserts or plasmid sequencing did not contain flanking restriction enzyme sites or additional bases. All primers were designed and analysed using either VectorNTI (Invitrogen) or Geneious Pro (BioMatters Ltd, NZ) to check for primer dimer or hairpin formation. Primers were supplied by Invitrogen (USA), Sigma Aldrich (USA) or IDT (USA). Primers were supplied purified with standard desalting techniques and were reconstituted in MQ H_2O to a concentration of 100 µM with working stock concentrations prepared at 10 µM in MQ H_2O.

2.1.2 Polymerase chain reaction (PCR)

Taq DNA polymerase (Invitrogen) was used to PCR amplify inserts for ligation and genes for diagnostic purposes. PCR annealing temperature and magnesium concentration was optimised for each primer set using a temperature gradient designed around the calculated Tm and a selection of $MgCl_2$ concentrations between 0.5 mM and 1.5 mM. In the event of the amplification of non-specific DNA products, the products of expected size were excised from an agarose gel and purified.

PCR reactions were carried out in 25 µl volumes using the following concentrations and reaction conditions. 1 X PCR buffer, 0.2 mM deoxynucleotide mix (dATP, dCTP, dGTP and dTTP), 1.5 mM $MgCl_2$ (unless mentioned otherwise), 1.25 U *Taq* DNA polymerase, 0.6 µM each primer and 60 - 200 ng template DNA. The amount of template DNA from transformant colonies for PCR reactions was

unknown but was performed using 1 µl of a colony suspension which was made by re-suspending a pipette tip-full of cells in 10 µl of MQ H_2O. Colony PCR was performed as above but with either 1.0 or 1.5 mM $MgCl_2$.

Cycle conditions:		94°C	2:45 (min:sec)
	X 29	94°C	0:15
		Tm*	0:30
		72°C	0:45**
		72°C	7:00
		4°C	HOLD

*Tm determined by gradient PCR
** Elongation time dependent of the insert size (1 kb = 1 minute)

2.1.3 Purification of PCR products and restriction enzyme digests

PCR products for restriction digest and ligation reactions were purified using the QIAquick® PCR purification Kit (if purified from solution) or QIAquick® Gel Extraction Kit (if purified after gel electrophoresis) (Qiagene, Netherlands), according to the manufacturer's instructions. Restriction enzyme digested PCR products and plasmids were purified using the QIAquick® PCR purification Kit. PCR products for gel excision were separated by electrophoresis with 1 x TAE, in a 1% agarose gel, stained with SYBR Safe™ DNA dye (Invitrogen) and excised with a clean scalpel blade under blue light (Invitrogen). DNA was eluted in 30 – 50 µl of elution buffer depending on preferred final concentration.

2.1.4 Agarose gel electrophoresis

DNA was separated via agarose gel electrophoresis for visualisation. All samples were separated on a 1% agarose gel in 1 x TAE buffer unless specified otherwise. Samples were mixed with 10 x DNA loading dye prior to loading onto the gel. Gels were either stained post-electrophoresis with 0.5 µg/ml ethidium bromide in 1 x TAE or in-gel with 1 x SYBR Safe™ DNA dye (Invitrogen) and visualised by UV light on a UV transilluminator. Alternatively, SYBR Safe™ stained gels were visualised on a blue light box (Invitrogen) if the DNA bands were to be excised.

The DNA size marker 1kb Plus DNA ladder (Invitrogen) was run on each agarose gel to determine DNA fragment size.

2.1.5 DNA quantification

DNA was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA). This measures the absorbance of DNA at 260 nm.

2.1.6 Restriction enzyme digestion and ligation of DNA

Restriction enzymes were purchased from Invitrogen (USA) or Roche Applied Science (Switzerland). Restriction enzyme digestion of DNA was performed in a final volume of 100 µl using the appropriate reaction buffer as recommended by the manufacturer, at 37°C for approximately 2 hours. The reaction was performed with 20 – 30 U of each enzyme and a maximum of 2000 ng of DNA. For digestion reactions with two enzymes, an appropriate buffer that would give the maximum dual enzyme activity was chosen. Digested DNA was purified from solution using the QIAquick® PCR purification Kit (Qiagen, Netherlands) according to manufacturer's instructions and eluted in 30 - 50 µl of elution buffer depending on preferred final concentration.

DNA ligation reactions were performed in a final volume of 10 µl using T4 DNA ligase (Invitrogen), at 4°C overnight or 6 hours at room temperature. Reactions were performed using 0.5 U of T4 DNA ligase, 100 ng of plasmid and with a 2 - 5 M mass ratio of excess insert DNA to plasmid DNA, in 1 x T4 ligase buffer.

2.1.7 Plasmid DNA amplification and purification

Plasmid DNA was amplified by transformation into electrocompetent *E. coli*, subsequent growth and DNA purification. DH5α or TOP10 *E. coli* were transformed with the plasmid of interest, plated onto selective LB and low salt LB agar, respectively, and grown overnight at 37°C. Small volume (5 ml) cultures from single colonies were then grown overnight in selective LB or low salt LB media for DH5α *E. coli* and TOP10 *E. coli* respectively. Plasmid DNA was

extracted from the overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen, Netherlands) according to manufacturer's instructions and eluted in 30 - 50 µl of elution buffer depending on preferred final concentration.

2.1.8 Plasmid DNA sequencing

Purified plasmid DNA preparations were sequenced for verification of cloned genes inserted and any sequence modifications engineered. Samples were sent to The University of Waikato DNA Sequencing Facility (Hamilton, NZ) along with appropriate primers that flanked the ends of the inserted gene or the ends of the multiple cloning sites of the plasmid of interest. Sequence results were analysed using either VectorNTI (Invitrogen) or Geneious Pro (BioMatters Ltd, NZ) to check for correct in-frame insertion of the gene and engineered modifications.

2.1.9 Preparation of electrocompetent *E. coli* and *M. smegmatis*

2.1.9.1 Preparation of electrocompetent E. coli

A glycerol stock of the appropriate strain of *E. coli* was freshly streaked out on an LB agar plate and grown overnight at 37°C. A single colony was used to start a 10 ml LB overnight seeder culture grown shaking at 180 rpm at 37°C. The overnight culture was used to seed a 1 L LB culture which was grown shaking at 180 rpm at 37°C to an OD_{600} of 0.5 – 0.7. The cells were chilled on ice for at least 30 minutes before centrifugation at 4°C for 15 minutes at 4000 g in chilled sterile bottles. After discarding the supernatant, the cells were gently resuspended in 1 L of ice cold sterile 10% glycerol and centrifuged again at 4°C for 15 minutes at 4000 g. The supernatant was discarded and the cells gently resuspended in 500 ml of ice cold sterile 10% glycerol and centrifuged again at 4°C for 15 minutes at 4000 g. The supernatant was discarded and the cells resuspended in 20 ml of ice cold sterile 10% glycerol and transferred to a chilled sterile 50 ml centrifuge tube and centrifuged again at 4°C for 15 minutes at 4000 g. The supernatant was discarded and the cells resuspended in 2 to 3 ml of ice cold sterile 10% glycerol, using the smaller volume if the OD_{600} of the 1 L culture was 0.5 and using the larger volume if the OD_{600} of the 1 L culture was 0.7. A 500 ml culture may also

be used for preparation of competent cells and the volume of 10% glycerol halved at each step. Aliquots of cells of 50 µl in small tubes were flash frozen in liquid nitrogen and stored at -80°C.

2.1.9.2 Preparation of electrocompetent M. smegmatis

A glycerol stock of the appropriate strain of *M. smegmatis* was freshly streaked out on a LBT agar plate and incubated for three days at 37°C. A single colony was used to start a 2 ml LBT overnight seeder culture grown shaking at 180 rpm at 37°C. One ml of the overnight culture was used to seed a 100 ml 7H9/ADC/T culture which was grown shaking at 200 rpm at 37°C to an OD_{600} of around 0.7. The cells were chilled on ice for 1.5 hours before centrifugation at 4°C for 20 minutes at 4000 g in chilled sterile bottles. After immediately discarding the supernatant to prevent cell pellet dispersal, the cells were gently resuspended in 100 ml of ice cold sterile 10% glycerol and centrifuged again at 4°C for 20 minutes at 4000 g. The supernatant was mostly discarded but a little was left to resuspend the cells in so they could be transferred to smaller centrifuge tubes, and they were centrifuged again at 4°C for 20 minutes at 4000 g. The supernatant was discarded and the cells resuspended in approximately 0.20 ml of ice cold sterile 10% glycerol and 40 µl aliquots put into small tubes, flash frozen in liquid nitrogen and stored at -80°C.

2.1.10 DNA transformation for amplification and expression

2.1.10.1 Electroporation of Escherichia coli

All components were kept on ice prior to electroporation. One or two µl of plasmid DNA or ligation reaction was added to a freshly thawed 50 µl glycerol stock of DH5α, TOP10, BL21 or DL41 electrocompetent *E. coli* as appropriate. The mixture was placed into a 0.2 cm electroporation cuvette (Biorad Laboratories, USA), the bottom tapped on a surface to settle the mixture between the electrodes and electroporated with a BioRad Gene Pulser™ (BioRad Laboratories, USA) at 2.5 KV, 25 µF capacitance and 200 Ω resistance. One ml of SOC media was immediately added to the electroporated cells, the cells

transferred to a 1.5 ml tube and incubated while shaking at 180 rpm at 37°C for 1 hour. Two different sized aliquots (e.g. 50 µl and 100 µl) were spread onto selective LB agar plates or selective low salt LB agar plates as appropriate and grown overnight at 37°C. Resulting colonies were either used to start 5 ml cultures for plasmid DNA amplification (section 2.1.7), or 20 ml seeder cultures for protein expression, or used for colony PCR to confirm successful ligation reactions (section 2.1.2). Cultures were incubated overnight at 37°C (180 rpm) using selective LB, low salt LB or TB media as appropriate.

2.1.10.2 Electroporation of Mycobacterium smegmatis

Electroporation of *M. smegmatis* requires adjustment of transformation resistance, alternative media and increased incubation time. All components were kept on ice prior to electroporation. One µl of plasmid DNA and 260 µl of 10% glycerol were added to a freshly thawed 40 µl glycerol stock of electrocompetent mc^2155 lab strain or *lsr2* knock out mc^2155 lab strain *M. smegmatis* as appropriate. The mixture was placed into a 0.2 cm electroporation cuvette (Biorad Laboratories, USA), the bottom tapped on a surface to settle the mixture between the electrodes and electroporated with a BioRad Gene Pulser™ (BioRad Laboratories, USA) at 2.5 KV, 25 µF capacitance and 1000 Ω resistance. One ml of 7H9/ADC/T media was immediately added to the electroporated cells, the cells transferred to a 1.5 ml tube and incubated while shaking at 180 rpm at 37°C for 3 hours. Two different sized aliquots (e.g. 50 µl and 100 µl) were spread onto selective 7H10/ADC/T agar plates or selective LBT agar plates and grown for 3 days at 37°C.

2.1.11 Glycerol stock preparation for bacterial strain storage

Glycerol stocks for long term storage of expression strains, laboratory strains and transformed strains were made by combining 0.5 ml of fresh culture with 0.5 ml of sterile 50% glycerol, mixing and placing at -80°C. A fresh culture represented an overnight culture in selective LB or low salt LB media for *E. coli* strains or a 3 day culture in selective 7H9/ADC/T or LBT media for *M. smegmatis* strains.

2.1.12 Small scale protein expression of His tagged proteins and Ni Sepharose™ bead assay

A single colony of the appropriate transformed *E. coli* strain grown on a selective LB agar plate was used for a 3 ml selective LB culture grown at 37°C overnight whilst shaking at 180 rpm. This culture was used to inoculate 3.5 ml of fresh selective LB at a 1:10 dilution ratio. After growth (whilst shaking) for 1 hour at 37°C, a 0.75 ml sample of the culture was taken as an uninduced control and the remainder induced with IPTG at a final concentration of 0.5 mM and growth continued at 37°C for 3 to 4 hours. A sample of 0.75 ml for whole cell analysis and a sample of 1.5 ml for cell lysis and Ni sepharose™ bead assay were taken. All samples were centrifuged immediately after collection for 2 minutes at 13,000 rpm to pellet the cells. The supernatant was discarded and the cells resuspended in 0.1 ml and 0.2 ml 50 mM phosphate (pH 7.4), 200 mM NaCl, 20 mM imidazole buffer, for 0.75 ml and 1.5 ml samples respectively. A 10 to 20 µl aliquot of the uninduced control and whole cell samples were mixed with an equivalent amount of 4 x SDS loading dye. The resuspended 1.5 ml sample was lysed on ice using a sonicator® ultrasonic processor XL fitted with a microtip probe at level 3 for 3 x 30 seconds (Misonix Incorporated, USA). After sonication the sample was centrifuged for 2 minutes at 13,000 rpm to pellet the cell debris. The supernatant was removed for use in the Ni Sepharose™ bead assay whilst the cell pellet was resuspended in 0.2 ml of phosphate buffer, and a 10 to 20 µl aliquot was mixed 1:1 with 4 x SDS loading dye. The supernatant was mixed with 20 µl of washed Ni sepharose™ high performance beads (GE Healthcare, Sweden) and incubated at room temperature for 20 minutes with periodic mixing. Beads were pre-washed with 1 ml of phosphate buffer. The supernatant was removed and a 20 µl aliquot mixed 1:1 with 4 x SDS loading dye. The beads were then washed three times with phosphate buffer and each time a 20 µl aliquot was mixed 1:1 with 4 x SDS loading dye. The remaining Ni Sepharose™ beads were mixed with 20 µl of 4 x SDS loading dye. All samples mixed with loading dye were heated at 95°C for 5 minutes before analysis by SDS-PAGE using an appropriate percentage acrylamide gel.

2.1.13 Large scale expression of his tagged proteins and immobilised metal affinity chromatography (IMAC) fast protein liquid chromatography (FPLC)

A single colony of the appropriate transformed *E. coli* strain grown on a selective LB agar plate was used for a 20 ml selective LB culture grown at 37°C overnight whilst shaking at 180 rpm. This culture was used to inoculate 1 L of fresh pre-warmed selective LB or TB and was then grown at 37°C, shaking at 180 rpm until it reached an OD_{600} of 0.4 – 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.5 – 1.0 mM and growth was continued at 37°C for 3 – 4 hours or 30°C overnight while shaking at 180 rpm. The culture was pelleted by centrifugation at 6000 rpm for 20 minutes at 4°C, and resuspended in 30 - 40 ml 50 mM phosphate (pH 7.4), 200 mM NaCl, 20 mM imidazole lysis buffer, and a Complete Mini EDTA-free protease inhibitor cocktail tablet added (Roche Diagnostics, USA) prior to cell lysis. Cells were lysed on ice using a sonicator® ultrasonic processor XL fitted with a flat probe at level 8 for 4 x 30 seconds (Misonix Incorporated, USA). After sonication the sample was centrifuged for 20 minutes at 13,000 rpm at 4°C to pellet the cell debris. The supernatant was filtered through 1.2 µm and 0.45 µm Minisart syringe filters (Sartorius AG, Germany) prior to protein purification via IMAC.

A 5 ml HiTrap™ Chelating HP column (GE Healthcare, Sweden) was prepared by washing with 20 ml MQ H_2O, priming with 10 ml of 100 mM $NiCl_2$, and equilibrating with 20 ml of 50 mM phosphate (pH 7.4), 200 mM NaCl, 20 mM imidazole lysis buffer. The filtered supernatant was then loaded onto the prepared column using a syringe and the column was attached to either an ÄKTA Prime™ or ÄKTA Basic™ FPLC system which monitors protein elution by absorbance at 280 nm. The FPLC system was washed with MQ H_2O and lines A and B equilibrated in lysis buffer and elution buffer respectively prior to column attachment. Unless noted otherwise, the elution buffer consisted of 50 mM phosphate (pH 7.4), 200 mM NaCl, 1 M imidazole. At a flow rate of 1 ml/min, the column was washed with 15 ml of lysis buffer to remove unbound proteins and then bound proteins were eluted by running a gradient of 0 - 100% elution buffer

over 60 minutes. Fractions containing protein of interest were analysed by SDS-PAGE. After each use, the chelating column was stripped of Ni^{2+} ions using 10 ml of 100 mM EDTA (pH 8.0), washed with 20 ml of MQ H_2O and rinsed with 10 ml of 20% ethanol for storage at 4°C.

2.1.14 SDS polyacrylamide gel electrophoresis (SDS-PAGE) and native polyacrylamide gel electrophoresis of protein samples

Polyacrylamide gels were cast in a Hoefer® multiple gel casting system SE275 (Hoefer® Inc, USA). SDS-PAGE gels consisted of a pH 8.8, 15%, 16.5% or 18% resolving gel overlaid with a pH 6.8, 5% acrylamide stacking gel. Proteins of 15 kDa or larger were resolved on a 15% gel whereas proteins of approximately 11 kDa or smaller were resolved on a 16.5% or 18% gels. SDS-PAGE gels were made with 30% acrylamide/bis solution 37.5:1 (BioRad Laboratories, USA), and included 0.1% *(w/v)* SDS, 375 mM Tris (pH 8.8) in the resolving gel or 128 mM Tris (pH 6.8) in the stacking gel, and were polymerised by the addition of 0.05% *(w/v)* ammonium persulphate and 0.05% *(v/v)* TEMED. Native PAGE gels consisted of a pH 8.8, 10% resolving gel overlaid with a pH 6.8, 5% stacking gel. Native PAGE gels were made as above but omitting the SDS. Samples for native PAGE analysis were separated using a 5 x native loading dye and a TG running buffer, whereas samples for SDS-PAGE analysis were separated using a 4 x SDS loading dye and TG-SDS running buffer.

Protein samples were mixed at a 3:1 ratio with 4 x SDS or at a 4:1 ratio with 5 x native loading buffers for denaturing or native gel electrophoresis, respectively, and samples for denaturation were heated to 95°C for 5 minutes before loading onto gels. A Precision Plus Protein™ Unstained Standard (BioRad Laboratories, USA) was used as a molecular weight marker. Gels were run in appropriate running buffer at 15 mA until the dye front entered the resolving gel and then at 20 – 25 mA until the dye front reached the end of the gel.

Gels were stained using the quick coomassie blue staining method (Wong *et al.*, 2000). Gels were microwaved in a box with a loose lid, with approximately 50 ml of Fairbanks A stain, for 30 seconds, and then cooled to room temperature while

shaking gently. Gels were then rinsed with tap water and Fairbanks B stain applied and the heating and cooling repeated. This sequence was repeated for Fairbanks C and Fairbanks D stains with solution D as the destain. A rolled up tissue was placed in the warmed Fairbanks D solution to assist destaining.

2.1.15 Protein measurement, concentration and dialysis

Protein was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA). This measures the absorbance of protein at 280 nm and calculates the protein concentration using the Beer's Law equation, $A = \varepsilon\ c\ \ell$, where A is the absorbance, ε is the theoretical molar extinction coefficient ($M^{-1}\ cm^{-1}$), c is the concentration (M), and ℓ is the pathlength (cm). Theoretical molar extinction coefficients were calculated from the amino acid sequence using the online ProtParam tool (Gasteiger *et al.*, 2005).

Protein samples were concentrated using Vivaspin 500, 2 or 20 concentrators (GE Healthcare, Sweden) with a 5 kDa molecular weight cut off. Membranes were pre-wet with 1 x TBS, 10 % glycerol and rinsed with protein purification buffer prior to adding protein to help prevent protein adherence. Concentrators were centrifuged at speeds recommended by the manufacturer at 4°C routinely or at 18°C when protein solubility was noticed to be an issue.

Large scale protein dialysis was performed using 6 – 8 kDa molecular weight cut off Specta Por® dialysis tubing (Spectrum Laboratories, UDA). The tubing was cut to the appropriate length for the protein volume to be dialysed and soaked in MQ H_2O before the addition of the protein sample, and the ends were sealed with dialysis clips. Dialysis was performed in 1 L of appropriate buffer with gentle stirring for several hours at 4°C before the solution was refreshed and dialysis continued overnight. Small scale protein dialysis was performed using 20 µl or 200 µl dialysis buttons (Hampton Research, USA) and 6 – 8 kDa molecular weight cut off Specta Por® dialysis tubing. The tubing was cut to fit the buttons and soaked in MQ H_2O before use. Dialysis was performed in 20 - 50 ml volumes of appropriate buffer overnight at 4°C or for several hours at room temperature with one buffer change each.

2.1.16 Analytical and preparative size exclusion chromatography (SEC) fast protein liquid chromatography (FPLC)

Size exclusion chromatography was used to further purify proteins. Analytical SEC was performed using either a Superdex™ 75 10/300 GL (3 – 70 kDa separation range) or a Superdex™ 200 10/300 GL column (10 - 600 kDa separation range) depending on the size of the protein of interest. Preparative SEC was performed using either a HiLoad™ 16/60 Superdex™ 75 column or a HiLoad™ 16/60 Superdex™ 200 column, based on the size of the protein of interest. All columns were supplied by GE Healthcare, Sweden. Columns were used while attached to either an ÄKTA Prime™ or ÄKTA Basic™ FPLC system which monitors protein elution by absorbance at 280 nm.

The FPLC system was washed with MQ H_2O prior to column attachment. Analytical SEC columns were prepared by washing with 25 ml of 0.2 µm filtered MQ H_2O and equilibrating with 25 ml of 20 mM Tris (pH 7.4), 150 mM NaCl, at a flow rate of 1 ml/min. Preparative SEC columns were prepared by washing with 130 ml of 0.2 µm filtered MQ H_2O and equilibrating with 130 ml of 20 mM Tris (pH 7.4), 150 mM NaCl, at a flow rate of 0.3 ml/min, as an overnight method. After protein separation, SEC columns were washed with filtered MQ H_2O and rinsed with 20% ethanol with the appropriate volumes and flow rates as mentioned above.

Protein solutions were concentrated to 0.5 ml for analytical SEC and 5 ml for preparative SEC, and filtered with a 13 mm Ø, 0.2 µm Acrodisk® syringe filter (Pall Lifesciences, USA) before loading onto prepared columns. Proteins were separated with one column volume of buffer at 0.5 ml/min. Fractions containing protein of interest were analysed by SDS-PAGE (section 2.1.14).

Calibration of Superdex™ 200 10/300 GL, Superdex™ 200 16/60 and Superdex™ 75 10/300 GL SEC columns was performed using the high and low molecular weight gel filtration calibration kits (GE Healthcare, Sweden) according to manufacturer's instructions. Calibration curves were prepared by plotting the Kav (elution volume function) of the protein standards against their log molecular weight and a linear equation determined. Kav is calculated with the

equation (Ve – Vo/Vc – Vo), where Ve is the elution volume, Vo is the column void volume and Vc is the geometric column volume. Kav values of the protein of interest are substituted into the equation to determine the molecular weight of the protein. The equation for Superdex™ 75 10/300 GL was determined in this study while the other equations were previously determined.

Table 2.1. Size exclusion chromatography calibration curve molecular weight equations.

SEC Column	Equation	Vo (ml)	Vc (ml)
Superdex™ 200 10/300 GL	$MW = e^{((Kav - 1.6833)/ -0.115)}$	8.11	24.0
Superdex™ 75 10/300 GL	$MW = e^{((Kav - 1.6951)/ -0.142)}$	8.08	23.6
HiLoad™ 16/60 Superdex™ 200	$MW = e^{((Kav - 2.1739)/ -0.159)}$	46.26	124.0

2.1.17 Transmission electron microscopy (TEM) of protein samples

Specimens for negative stain electron microscopy were prepared as follows: 5 µl of IMAC purified protein sample (various concentrations) was applied to a 300 mesh copper EM grid (GCU300 ProSciTech, Australia) covered with carbon-film supported by plastic that had been rendered hydrophilic by glow discharge. Glow discharge was performed under vacuum (1 atm) at 50% apparatus voltage for 10 seconds. After 90 seconds, the sample was blotted from the edge of the drop with filter paper (Whatman No 1) and washed 3 times with filtered MQ H_2O by applying a drop onto the grid and blotting from the edge with filter paper. The sample was then stained with 5 µl of 1% uranyl acetate for 45 seconds, blotted and restained for 25 seconds, blotted, then air dried. Low-dose images were recorded at nominal magnification of 40,000 at ≈ 1 - 2 µm under focus in either a Phillips CM12 or a FEI TecNai G^2 electron microscope when higher quality images were desired. Both electron microscopes were equipped with a LaB_6 filament and operated at 120 kV. Images with scale bars of 0.2 µm, 0.5 µm and 1 µm were captured at 52,000 x, 40,000 x and 11,500 x magnification respectively.

2.1.18 Protein crystallography

2.1.18.1 Initial crystallisation screens

Concentrated protein for crystallisation screens was centrifuged at 13,000 g for 5 min at 4°C to remove particulates from the solution. Initial crystallisation screens for proteins were set up at the crystallisation facility at The University of Auckland using a high-throughput method of 480 conditions developed to minimise the protein used (Moreland *et al.*, 2005). A Multiprobe® II HTEx robot (PerkinElmer, USA) was used to dispense 75 µl of each condition into wells of a 96-well Intelli-plate (Art Robbins Instruments, USA), then a Cartesian honey bee liquid handler (Cartesian™ Dispensing system, USA) dispensed 100 nl of protein with 100 nl of precipitant solution into sitting drops in the 96-well Intelli-plates. Plates were sealed with ClearSeal film™ (Hampton Research, USA) and left at 18°C on shock-resistant shelves.

At a later time during this study, a liquid handling robot was obtained at the Proteins and Microbes research group at The University of Waikato. Subsequent crystallisation trials using Hampton crystallisation screens totalling 384 conditions (HR2-086, HR2-130, HR2-136, HR2-144; Hampton Research, USA) were used. One hundred µl of each crystallisation condition was dispensed by multi-channel pipette into 96-well Intelli-plates and a mosquito® crystal robot (TTP LabTech, Ltd, UK) was used to dispense 100 nl of protein with 100 nl of precipitant solution into sitting drops in the 96-well Intelli-plates and the plates were sealed with ClearSeal film™, and left at 18°C on shock-resistant shelves.

2.1.18.2 Fine screening and additive screening for the optimisation of crystallisation conditions

Promising crystallisation conditions revealed by initial crystallisation screens were optimised by modifying the pH, crystallisation component concentrations or the addition of specific selected components. All trials were carried out in 24-well VDX crystallisation plates (HR3-140; Hampton Research, USA) at 18°C, on shock-resistant shelves.

A hanging drop method was most commonly used for fine screening of crystallisation conditions. The top of the wells were greased using glisseal®N grease (Borer chemie, Switzerland) and 500 µl of each condition was pipetted into each well. On the surface of a 22 mm square siliconised coverslip (Hampton Research, USA) 1 -2 µl of protein was added to an equal amount of precipitant solution and the coverslip inverted and placed onto the top of the greased well. At times, an unequal amount of protein to precipitant solution was combined to alter crystallisation conditions.

Additive screens based on the most promising condition with the addition of selected components were performed using the hanging drop method above. Component selection was based on the reagent formulation details in a Hampton Additive Screen (HR2-428; Hampton Research, USA) and reagents were added at the suggested drop concentrations (Table 2.2).

Table 2.2. Reagents used in additive screening for the optimisation of crystallisation conditions.

Reagent	Drop Concentration	Reagent	Drop Concentration
$ATP.Na_2$	10 mM	**$KNaTartrate4H_2O$**	100 mM
$CaCl_2$	10 mM	**$LiCl_2$**	100 mM
$CoCl_2$	10 mM	**$(NH_4)_2SO_4$**	100 mM
$CuCl_2$	10 mM	**Acetonitrile**	4.0% (v/v)
EDTA (pH 7.6)	10 mM	**Butantol-1**	0.7% (v/v)
$Ni(II)Cl_26H_20$	10 mM	**Ethanol**	3.0% (v/v)
Nicotinamide ADP	10 mM	**Glycerol**	3.0% (v/v)
$MgCl_2$	10 mM	**D-glucose**	3.0% (w/v)
L-proline	10 mM	**Isopropyl alcohol**	3.0% (v/v)
Spermidine	10 mM	**Methanol**	3.0% (v/v)
$ZnCl_2$	10 mM	**Methyl pentanediol**	3.0% (v/v)
$CsCl_2$	100 mM	**PEG 400**	5.0% (w/v)
Glycine	100 mM	**Sucrose**	3.0% (w/v)
KCl_2	100 mM		

Sitting drop methods and streak seeding was also used with individual promising crystallisation conditions. For sitting drop preparation, the top of the wells of 24-well VDX crystallisation plates were greased using glisseal®N grease and 500 µl of each condition was pipetted into each well. A sitting drop was generated by pipetting 1 – 2 µl of protein and an equal amount of precipitant solution into a MicroBridge reservoir (Hampton Research, USA). The well was

sealed by placing a 22 mm siliconised glass coverslip onto the greased top. In preparation for streak seeding, fine screen conditions were set up using the hanging drop method and the drops allowed to equilibrate for 2 -3 hours. The technique for streak seeding involved brushing a cleaned cat's whisker against an existing crystal to pick up small crystal fragments and then lightly running it through an equilibrated drop. This was repeated for each drop.

2.1.18.3 Crystal freezing for data collection

A cryoprotectant reagent was added to a replicated sample of the crystallisation condition, for the preparation of crystal freezing if required. In this study, glycerol was routinely used unless the crystallisation condition already contained an appropriate concentration of PEG suitable to use as a cryoprotectant. For data collection at The University of Auckland home source, cryoprotectants were able to be tested to assess for ice formation prior to freezing the crystals. The crystallisation condition containing the cryoprotectant was tested by sampling it with a CryoLoop™ (Hampton Research, USA), flash freezing it in liquid nitrogen and mounting it for X-ray testing. The resulting X-ray diffraction pattern was examined for ice ring diffraction and the lowest concentration cryoprotectant that did not result in ice rings was selected for crystal freezing. For data collection at the Australian Synchrotron, crystals were frozen without testing cryoprotectants due to logistical constraints. To minimise the chance of ice formation, slightly higher cryoprotectant concentrations were used during crystal freezing unless previous conditions had been determined during testing at the home source.

2.1.18.4 Crystal testing and X-ray diffraction data collection

Crystals were transferred to the precipitant solution containing cryoprotectant using a CryoLoop™ and magnetic CrystalWand™ then immersed in liquid nitrogen inside a thick walled cryogenic vial (Hampton Research, USA) for home source data collection, or immersed in liquid nitrogen inside a SSRL automated mounting cassette (Crystal Positioning Systems, USA) for data collection at the Australian Synchrotron.

Home source data collection was performed at the School of Biological Sciences at The University of Auckland. X-rays were generated using CuKα radiation (λ = 1.5418 Å) from a rotating copper anode MicroMax™-007HF generator (Rigaku, Japan) and reflections were detected using a Mar345 detector (Mar Research GmbH, Germany). The loops were mounted on a MAR234dtb goniometer using an automated system (Mar Research GmbH, Germany) and kept cold with a stream of liquid nitrogen from a Cobra cryosystem (Oxford Cryosystems, UK). Prior to data collection, the MOSFLM's strategy function was used with 2 images 90° apart, to aid in data collection.

Synchrotron data collection was performed at the Australian Synchrotron, Melbourne, Australia using the MX1 or MX2 beamline. Reflections were measured using an ADSC Quantum 210r detector (Area Detector Systems Corporation, USA) with radiation at a wavelength of 0.9786 Å. Prior to data collection, the MOSFLM's strategy function was used with 2 images 90° apart, to aid in data collection.

2.2 Methods for Chapter Three: Lsr2 expression, purification & characterisation

2.2.1 Cloning of Lsr2 into pET30b

This expression vector was kindly supplied by Roberto Colangeli (Colangeli *et al*., 2007) at the Division of Infectious Disease and the Center for Emerging Pathogens, Department of Medicine, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, New Jersey. The *lsr2* gene (Rv3597c) was amplified by PCR from genomic DNA from *M. tuberculosis* and cloned into pET30b (Novagen) using the *NdeI*/*XhoI* restriction sites, positioning a 6 x histidine tag at the C-terminal end of the protein. The plasmid was transformed into BL21 *E. coli* for expression.

Table 2.3. Primers used for cloning *M. tuberculosis* H37Rv *lsr2* into pET30b

Primer	Sequence	Length
pET30-*lsr2*F	TTC CGA GCC GGT CCA CAT GAG A	22 bp
pET30-*lsr2*R	CTC GCT TAT GGA TTA CAT CCT GAG	24 bp

2.2.2 Cloning of Lsr2 into pMAL-c2x

The *lsr2* gene was cloned into the maltose binding protein fusion tag *E. coli* expression vector pMAL-c2x (New England Biolabs, USA) as follows. *Lsr2* was amplified from the *lsr2*-pET30b vector via PCR using forward and reverse primers containing the restriction enzyme sites *XbaI* and *PstI*, respectively.

Table 2.4. Primers used for cloning *M. tuberculosis* H37Rv *lsr2* into pMAL-c2x.

Primer	Sequence	Length
Rv3597c_Xba1_Fwd	GCT CTA GAA TGG CGA AGA AAG TAA CCG TC	29 bp
Rv3597c_Pst1_Rev	AAC TGC AGT CAG GTC GCC GCG	21 bp

Standard *Taq* DNA polymerase PCR was performed at 52°C annealing temperature and 1.0 mM $MgCl_2$ concentration after optimisation of annealing

temperature and magnesium concentration. Gel extraction of a combined 8-fold PCR reaction size was performed and the DNA product quantified. The PCR product was inserted into pMAL-c2x via appropriate enzyme digestion and ligation, transformed into DH5α for plasmid preparation for sequence verification and then transformed by electroporation into BL21 *E. coli* for expression.

2.2.3 Large scale expression and purification of His tagged Lsr2

Large scale expression was performed using standard techniques but with the inclusion of a DNase digestion step and with an elution buffer consisting of 50 mM phosphate (pH 7.4), 200 mM NaCl, 500 mM imidazole. After cell lysis, 200 µl of DNase1 (1 mg/ml; Sigma, USA) was added to the lysate and incubated at 4°C for approximately 2 hours (while gently shaking) before centrifuging for immobilised metal affinity chromatography (IMAC) purification. Fractions containing protein of interest were prepared appropriately and analysed by 1% agarose gel electrophoresis and 15% SDS-PAGE.

Size exclusion chromatography (SEC) of Lsr2 was performed using a HiLoad™ 16/60 Superdex™ 200 column according to standard techniques but in a high salt buffer to dissociate the co-purified *E. coli* genomic DNA. The buffer included 50 mM phosphate (pH7.4), 2 M NaCl, for protein characterisation studies or 20 mM Tris (pH 7.4), 2 M NaCl, when the protein was required for crystallisation studies. Phosphate purified protein was subsequently concentrated and dialysed into 20 mM Tris (pH 7.4), 150 mM NaCl buffer if required for crystallisation studies. Fractions of interest were prepared for analysis by 1% agarose gel electrophoresis and 15% SDS-PAGE.

2.2.4 Small scale expression and purification of MBP tagged Lsr2

A 3 ml overnight culture of *lsr2*-pMAL-c2X in BL21 *E. coli* in selective rich media + glucose was grown at 37°C, at 180 rpm. This culture was used to inoculate 4 ml of fresh selective rich media + glucose at a 1:10 dilution ratio. After growth (whilst shaking) for 1 hour at 37°C, a 1 ml sample of the culture was taken as an

uninduced control and the remainder induced with IPTG at a final concentration of 0.3 mM. Growth was continued at 37°C for 3 hours with 1 ml samples for whole cell analysis taken at 1, 2, and 3 hour time points. The remaining 2.5 ml was pelleted by centrifugation and resuspended in 0.5 ml of column buffer (20 mM Tris (pH 7.5), 200 mM NaCl, 1 mM EDTA) for cell lysis using sonication. Sonication was performed on ice using a sonicator® ultrasonic processor XL fitted with a microtip probe at level 3 for 3 x 30 seconds (Misonix Incorporated, USA). The sample was centrifuged for 5 minutes at 13,000 rpm to pellet the debris and 15 µl of the supernatant was added to 5 µl of 4 x SDS loading dye whilst the cell pellet was resuspended in 0.5 ml of column buffer, and a 10 µl aliquot mixed 1:1 with 4 x SDS loading dye. The remaining supernatant was mixed with 40 µl of washed amylose resin (NEB, USA) and incubated for 15 minutes in ice with periodic mixing. Resin was washed with 1 ml of binding buffer prior to use. The supernatant was removed and a 15 µl aliquot mixed with 5 µl 4 x SDS loading dye. The resin was washed once with column buffer and 15 µl of the wash solution mixed with 5 µl 4 x SDS loading dye. The resin was mixed with 20 µl of 4 x SDS loading dye. All samples mixed with loading dye were heated at 95°C for 5 minutes before analysis on a 15% SDS-PAGE gel.

2.2.5 Large scale expression and purification of MBP tagged Lsr2

A 20 ml overnight culture of *lsr2*-pMAL-c2X in BL21 *E. coli* in selective rich media + glucose was grown at 37°C, at 180 rpm. This culture was used to inoculate 1 L of fresh pre-warmed selective rich media + glucose which was then grown at 37°C, shaking at 180 rpm until it reached an OD_{600} of 0.4 – 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.3 mM and growth was continued at 37°C for 3 hours while shaking at 180 rpm. The culture was pelleted by centrifugation at 5000 rpm for 30 minutes at 4°C, and resuspended in 30 ml of column buffer (20 mM Tris (pH 7.5), 200 mM NaCl, 1 mM EDTA) and a Complete Mini EDTA-free protease inhibitor cocktail tablet added (Roche Diagnostics, USA) prior to cell lysis. Cells were lysed on ice using a sonicator® ultrasonic processor XL fitted with a flat probe at level 8 for 3 x 30

seconds (Misonix Incorporated, USA). After sonication the sample was centrifuged for 20 minutes at 13,000 rpm at 4°C to pellet the cell debris. The supernatant was filtered through 1.2 µm and 0.45 µm Minisart syringe filters (Sartorius AG, Germany) prior to protein purification.

A 5 ml MBPTrap™ HP column (GE Healthcare, Sweden) was prepared by washing with 30 ml MQ H_2O and equilibrating with 30 ml of column buffer. The filtered supernatant was then loaded onto the column using a syringe and the column was attached to either an ÄKTA Prime™ or ÄKTA Basic™ FPLC system which monitors protein elution by absorbance at 280 nm. The FPLC system was washed with MQ H_2O and lines A and B equilibrated in column buffer and elution buffer (20 mM Tris (pH 7.5), 200 mM NaCl, 1 mM EDTA, 10 mM maltose) respectively prior to column attachment. At a flow rate of 5 ml/min, the column was washed with 50 ml of column buffer to remove unbound proteins and then bound proteins were eluted with 25 ml of 100% elution buffer. Fractions containing protein of interest were prepared appropriately and analysed by 1% agarose gel electrophoresis and 15% SDS-PAGE. After use, the Dextrin Sepharose™ column was regenerated by washing with 10 ml of 0.1% SDS, and 20 ml of MQ H_2O and rinsed with 10 ml of 20% ethanol for storage at 4°C.

2.2.6 Transmission electron microscopy (TEM) of Lsr2/DNA complexes

Examination of the effect of DNase treatment on the Lsr2/DNA complexes was performed as follows. Two units of DNaseRQ1 (1U/µl; Promega, USA) was added to 20 µg of IMAC purified Lsr2 in a 20 µl volume and incubated at 37°C for the appropriate time period and then placed on ice for sampling. A 5 µl aliquot was prepared at each time point for TEM analysis as per standard procedure.

Examination of the effect of high salt buffer on the Lsr2/DNA complexes was performed as follows. IMAC purified Lsr2 sample (at 7.6 mg/ml) was mixed 1:1 with 2 M, 3 M and 4 M NaCl to create final concentrations of 1.0 M, 1.5 M and 2.0 M NaCl, and incubated for 15 minutes at room temperature. The sample was diluted 1 in 20 with MQ H_2O and a 5 µl aliquot was prepared for TEM analysis as per standard procedure with the addition of 3 extra MQ H_2O washes.

2.2.7 MALDI-TOF peptide mass analysis of Lsr2

An SDS-PAGE gel band representing Lsr2 was excised using a clean scalpel blade, cut into small fragments and transferred to a 1.5 ml tube. Gel fragments were repeatedly destained with 0.5 ml aliquots of 1:1 acetonitrile (ACN) : 25 mM NH_4HCO_3 for several hours at room temperature and left destaining overnight. The destain was removed and the gel fragments rinsed with 100 µl of 1:1 ACN : 25 mM NH_4HCO_3 and the tube was spun briefly at high speed to allow the remaining solution to be removed. Fifty µl of ACN was added for 15 minutes at room temperature to dehydrate the gel fragments and then removed. The gel fragments were dried in a vacuum desiccator for 1 hour before the addition of trypsin. A 1 µl trypsin aliquot (1 mg/ml; stored at -80°C) was diluted to 200 µl with 25 mM NH_4HCO_3 and 20 µl added to the desiccated gel fragments and the tube was incubated inverted at 37°C overnight. Ten µl of 20% ACN / 0.1% trifluoro acetic acid (TFA) solution was added to the rehydrated gel slices and left for 30 minutes at room temperature before analysis.

For analysis of peptides, a α-cyano-4-hydroxycinamic acid (HCCA) matrix (Sigma, USA) was used. Ten µg of matrix was added to 30 µl of 2:1 ACN : 0.1% TFA, mixed well, sonicated for 10 minutes in a sonicating water bath (Elma, Germany) and then centrifuged for 10 minutes at 13,000 rpm. Protein samples and peptide calibration standards (Bruker Daltonics, USA) were mixed in a 1:1 ratio (0.5 µl:0.5 µl) with HCCA matrix on an AnchorChip™ MALDI-TOF target plate (Bruker Daltonics, USA) and left to air dry.

An autoflex™ II MALDI-TOF mass spectrometer (Bruker Daltonics, USA) was used to analyse samples. Samples were analysed in RP mode, with the mass range selector adjusted as required (suppressing < 500 Da ions) and data in the range of 500 – 4000 Da collected. Settings included a pulsed ion extraction of 60 ns, a detector gain of 1400 V, an acceleration voltage of 19 kV, and a reflector voltage of 20 kV, with lazer power typically around 60%. Spectra for peptide calibration standards were first collected and the spectrometer calibrated with an automatic polynomial correction. Spectra were exported to FlexAnalysis™ software (Bruker Daltonics, USA), and the peaks identified and labelled. The processed spectra

were opened in BioTools (Bruker Daltonics, USA) and compared with masses from a predicted trypsin digest of Lsr2 with one missed cleavage.

2.2.8 Biotinylated DNA oligonucleotide design and binding assay

Short complementary DNA oligonucleotides that were high in adenine and thymine bases were designed with a 5' biotin molecule on the forward primer. Alternative biotin labelled forward primers were designed, with one containing a short 4 base pair linker for DNase cleavage that was designed based on preferred DNase I cleavage sites (Herrera & Chaires, 1994).

Table 2.5. Primers used to generate double stranded AT-rich DNA oligonucleotides.

Primer	Sequence	Length
AT-Rich_F_Biotin	Biotin - ATG CGG GTA ATA TCA CGG	18 bp
AT-Rich_R	CCG TGA TAT TAC CCG C	16 bp
AT-Rich_F_Linker_Biotin	Biotin - ATG GAT GCG GGT AAT ATC ACG G	22 bp

To generate a double stranded DNA biotin labelled oligonucleotide, 200 pmol each of forward and reverse oligonucleotide were combined with annealling buffer (10 mM Tris (pH8.0), 30 mM NaCl, 0.2 mM EDTA), in a final volume of 50 µl, heated at 95°C for 5 minutes and allowed to passively cool to room temperature. Ten µl of double stranded biotinylated oligonucleotide was incubated with 500 µl of IMAC or SEC purified Lsr2 protein fraction at 4°C overnight. The protein and biotinylated oligonucleotide mixture was added to 20 µl of washed streptavidin magnetic particles (Roche Diagnostics, Germany) and incubated at room temperature for 2 hours with intermittent mixing. Prior to use streptavidin beads were washed 3 times with 200 µl of binding buffer (50 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA), using a magnet to draw down the particles, allowing the supernatant to be removed. After incubation the solution containing unbound molecules was removed and the beads were washed twice with 200 µl binding buffer containing 1 M NaCl and resuspended in 20 µl of buffer. Fifteen µl samples of unbound molecules and wash solutions and the

resuspended beads were prepared appropriately and analysed by 1% agarose gel electrophoresis and 15% SDS-PAGE.

Attempts to remove the bound oligonucleotide and Lsr2 complex from the streptavidin beads was performed by using an elution buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA, 5 mM biotin to resuspend the beads at the last step of the method. Treatment of the streptavidin beads at the last step of the method with 1 unit of DNase RQI in 20 µl of suitable buffer (20 mM Tris (pH 8.0), 100 mM $MgCl_2$, 10 mM $CaCl_2$) at room temperature overnight or 37°C for 1 – 2 hours was also attempted to dissociate the complex.

2.2.9 Glutaraldehyde cross-linking of Lsr2

Glutaraldehyde was added to approximately 8 µg of IMAC affinity purified Lsr2 or dextrin sepharose purified Lsr2-MBP to a final 0.1% (v/v) concentration in a 15 µl volume and incubated at room temperature for 2 and 5 minutes. The reaction was stopped by the addition of 5 µl of 4 x SDS loading dye and heating for 5 minutes at 95°C. Cross-linking ability was assessed by 18% SDS-PAGE analysis.

2.2.10 Western blot analysis of mycobacterial Lsr2

Whole cell lysates of *Mycobacterium bovis* BCG and *Mycobacterium smegmatis* mc^2155 culture were subjected to SDS-PAGE along with IMAC purified Lsr2 as a positive control and 5 µl of Benchmark™ prestained protein ladder (Invitrogen, USA). Whole cell lysates were prepared by resuspending a 50 µl cell pellet of each strain in 200 µl of 1 x PBS and transferring to a 2 ml screw cap tube containing approximately 0.3 g of 0.1 mm and 2.5 mm zirconia beads (Biospec Products Ltd, USA). Beads were mixed in a Fastprep FP120 bead beater (Thermosavant, USA) for 20 seconds at setting 6, cooled for 1 minute, then the steps repeated for 30 second and 40 second time periods. Samples were centrifuged for 10 minutes at 13,000 rpm to remove froth and 2 – 8 µl used for 15% SDS-PAGE gel separation.

Antigens and other cellular proteins were electrophoretically transferred to a Protran™ nitrocellulose membrane (Whatman, UK) as follows. A transfer sandwich was set up and added to the transfer apparatus (BioRad Laboratories, USA) with an ice insert and magnetic stirring bar, and cold transfer buffer added to submerge the sandwich. Proteins were transferred for 1 hour at 100 V, after which the sandwich was disassembled and successful transfer confirmed by staining the nitrocellulose membrane with Ponceau S stain for 2 minutes and rinsing with water.

The membrane was blocked for 1 hour at room temperature or overnight at 4°C in BLOTTO solution before incubation with antibodies. Rabbit anti-Lsr2 polyclonal antibody diluted 1:5000 in BLOTTO was added to the membrane and incubated for 1 hour at room temperature and then the membrane was washed 5 x 5 minutes with 1 X TBS-T + 0.5 M NaCl. After washing, a goat anti-rabbit HRP conjugate antibody (Pierce Antibodies, USA) diluted 1:1000 in BLOTTO was added to the membrane and incubated for 1 hour at room temperature and then the membrane was washed 5 x 5 minutes with 1 X TBS-T + 0.5 M NaCl. The membrane was treated with SuperSignal chemiluminescence substrate (Pierce, USA) at a 1:1 dilution with water. The chemiluminescent signal from the membrane was visualised using a FujiFilm LAS-1000 Luminescent Image Analyzer with Intelligent Dark Box II (Fuji Photo Film Co., Ltd, Tokyo).

Lsr2 antibodies were produced at the AgResearch antibody facility, Ruakura, Hamilton. Lsr2 was expressed in *E. coli* and purified by IMAC for antibody production in four month old white rabbits. A pre-immune serum sample was taken from each rabbit and then a 0.5 ml primary immunisation consisting of 0.2 ml of 1 mg/ml Lsr2 protein and 0.3 ml Freunds Complete adjuvant was injected subcutaneously. Every two weeks for a period of 6 weeks after the primary immunisation, the rabbits were injected subcutaneously with a secondary immunisation consisting of 0.1 ml of 1 mg/ml Lsr2 protein, 0.1 ml phosphate buffered saline (PBS) and 0.3 ml Freunds Incomplete adjuvant. Rabbits were anaesthetised with sodium pentobarbitone prior to the terminal bleed.

2.2.11 Lsr2 lysine - alanine mutagenesis design, generation and cloning

Three lysine to alanine mutations (designated 2K2A, 3K3A and 5K5A), each with a different combination of the 5 lysine residues altered, were manufactured by Geneart (Geneart, Germany) to include the attB1 recombination sites used in Gateway® cloning (Invitrogen). Supplied plasmid DNA was resuspended in sterile MQ H_2O to 100 ng/µl and 50 ng was used in a Platinum *Pfx* DNA polymerase (Invitrogen) PCR reaction to amplify the inserted *lsr2* gene.

Pfx PCR reactions were carried out in 25 µl volumes using the following concentrations and reaction conditions. 1 X *Pfx* buffer, 1 x *Pfx* Enhancer, 0.3 mM deoxynucleotide mix, 1.0 mM $MgSO_2$, 0.5 U Platinum *Pfx* DNA polymerase, 0.2 µM generic forward attB primer, 0.2 µM generic reverse attB primer and 10 - 50 ng template DNA.

Cycle conditions:

	94°C	1 :00	(min:sec)
X 10	94°C	0:20	
	55°C	0:30	
	68°C	1:00	
X 20	94°C	0:20	
	65°C	0:30	
	68°C	1:00	
	68°C	2:00	

The PCR product was run on a 1% agarose gel and the amplified insert excised for purification and 10 – 15 ng was then used as the template in a further *Pfx* PCR reaction, which was then purified for a BP cloning reaction.

In the BP cloning reaction, approximately 25 femtomoles of each PCR product and 80 ng of pDONR221 vector DNA were mixed with 1 µl of 5 x BP clonase enzyme (Invitrogen) and TE buffer (pH 8.0) was added to a final volume of 5 µl. The reaction was incubated at room temperature for 6 hours and stopped by the addition of 0.5 µl Proteinase K enzyme (2 ug/ul; Invitrogen) and an incubation at 37°C for 10 minutes. Two µl of this reaction was transformed into DH5α *E. coli* and grown overnight on selective media as per standard procedure. Triplicate 5 ml selective LB overnight cultures for each mutant were grown and plasmid was purified using standard procedures. To confirm the insertion of the *lsr2* gene

into pDONR221 prior to the LR cloning reaction, a standard *Taq* PCR was performed (using 0.4 µM each of M13 forward and reverse primers and 50 ng of plasmid at a 55°C annealing temperature), and the products analysed on a 1% agarose gel as per standard procedure.

In the LR cloning reaction, approximately 75 ng each of entry clone DNA and destination vector DNA pDEST17 were mixed with 1 µl of 5 x LR clonase enzyme (Invitrogen) and TE buffer (pH 8.0) was added to a final volume of 5 µl. The reaction was incubated at room temperature for 4 hours and stopped by the addition of 0.5 µl Proteinase K enzyme (2 ug/ul; Invitrogen) and an incubation at 37°C for 10 minutes. Two µl of this reaction was transformed into DH5α *E. coli* and grown overnight on selective media as per standard procedure. Colony PCR was performed in triplicate from the resulting transformants, as per standard procedure using T7 primers, to confirm successful incorporation of the *lsr2* gene into the pDEST17 expression vector. Overnight selective LB cultures using the remaining colony suspension from positive clones were grown for plasmid preparation and sequence confirmation by DNA sequencing. Correctly cloned plasmids were transformed into BL21 *E. coli* for protein expression.

2.2.12 Small and large scale expression of Lsr2 lysine mutants

Small scale protein expression and Ni sepharose bead purification was performed using standard techniques but at a variety of expression temperatures (20°C, 25°C, 30°C and 37°C). Large scale expression of Lsr2 lysine mutants was performed using standard techniques but cultures were incubated at 20°C overnight. Large scale purification was performed using a HiTrap chelating column and following standard procedures.

2.2.13 Fluorescent staining of recombinant *E. coli* using Lsr2 antibodies

A 100 µl aliquot of each of the bacterial cultures was centrifuged for 5 minutes at 13,000 rpm to pellet the cells. The supernatant was removed and the pellet was washed with 1 ml of PBS, centrifuged again and resuspended in 50 µl of PBS. A

glass slide was cleaned by flaming in 100% ethanol and 3 x 20 µl volumes of the resuspended cells applied to the surface and left to dry. The cells were fixed to the slide by immersion in 95% ethanol for 1 minute at room temperature, draining the alcohol off the slide and dipping it in PBS and air drying before finally heating at 55°C for 10 minutes.

The fixed slides were rehydrated by rinsing with PBS and draining off the excess. The slides were incubated with rabbit anti-Lsr2 polyclonal antibody diluted 1:500 in PBS for 30 minutes at 37°C in a humidified chamber. They were then rinsed by placing in a tray of PBS at room temperature for 30 minutes and then drained. FITC conjugated goat anti-rabbit IgG fluorescent secondary antibodies (Sigma, USA) were diluted 1:100 in PBS and a solution applied to the slides for 30 minutes at 37°C in a humidified chamber. The slides were rinsed in PBS as previously and then drained and dried by careful blotting around the cell spots. Each spot was covered with a drop of mounting media (Cargille Type A) and a glass coverslip laid on top. Cells were visualised at 100 x oil immersion under blue light (ex = 485, em = 498) and photographed with a digital camera.

2.3 Methods for Chapter Four: DNA binding for three Lsr2 mutants

2.3.1 Site directed mutagenesis using QuikChange™ protocol

Site directed mutagenesis to generate four alternative mutations was performed based on the protocol from the Stratagene QuikChange™ II-E kit (Stratagene, USA). Mutagenesis primers for arginine to alanine mutations were designed with the twin base pair residue change in the centre, flanked by approximately 15 bases on either side as recommended.

Table 2.6. Primers for site directed mutagenesis to generate arginine to alanine mutations.

Primer	Sequence	Length
Lsr2Arg84-Ala84-F	AGC AGA GCG CGG CGA TCG CCG AAT GGG CTC GTC GTA	36 bp
Lsr2Arg 84-Ala84-R	TAC GAC GAG CCC ATT CGG CGA TCG CCG CGC TCT GCT	36 bp
Lsr2Arg97-Ala97-F	ACA ATG TGT CGA CGG CAG GCC GGA TCC CGG CCG ACG	36 bp
Lsr2Arg97-Ala97-R	CGT CGG CCG GGA TCC GGC CTG CCG TCG ACA CAT TGT	36 bp
Lsr2Arg99-Ala99-F	ACA ATG TGT CGA CGC GAG GCG CGA TCC CGG CCG ACG	36 bp
Lsr2Arg99-Ala99-R	CGT CGG CCG GGA TCG CGC CTC GCG TCG ACA CAT TGT	36 bp
Lsr2-R97/R99-A97/A99-F	ACA ATG TGT CGA CGG CAG GCG CGA TCC CGG CCG ACG	36 bp
Lsr2-R97/R99-A97/A99-R	CGT CGG CCG GGA TCG CGC CTG CCG TCG ACA CAT TGT	36 bp

A fresh plasmid purification of *lsr2*-pET30b for the template DNA was prepared to ensure a high degree of supercoiled DNA. Mutagenic *Pfu* Ultra DNA polymerase (Aligent Technologies, USA) PCR was performed in a total volume of 50 µl with the following final concentrations and reaction conditions. 1 X *Pfu* buffer, 0.2 mM deoxynucleotide mix, 2.5 U *Pfu* Ultra DNA polymerase, 125 ng each forward and reverse primer and 25 ng of template DNA.

Cycle conditions:

	95°C	0 :30	(min:sec)
X 17	95°C	0:30	
	55°C	1:00	
	68°C	6:00	
	4°C	HOLD	

Successful PCR amplification of the DNA was confirmed by 1% agarose gel electrophoresis of a reaction aliquot as per standard procedure prior to digestion of the DNA with 10 U of *DpnI* (New England Biolabs, USA) at 37°C for 1 hour. Following DNA digestion, 4 µl of the product was transformed into DH5α *E. coli* which were grown overnight on selective media as per standard procedure. Overnight selective cultures were grown from successful transformants for plasmid preparation and DNA sequencing as per standard procedure. Correctly mutated variants were transformed by electroporation into BL21 *E. coli* for expression.

2.3.2 Expression and purification of Lsr2 DNA binding mutants

Large scale expression was performed using a 5 ml HiTrap™ Chelating HP column according to standard techniques. Fractions containing protein of interest were prepared appropriately and analysed by 1% agarose gel electrophoresis and 15% SDS-PAGE.

Size exclusion chromatography of the Lsr2 R97A-R99A DNA binding mutant was performed using a HiLoad™ 16/60 Superdex™ 200 column using standard techniques but in a high salt buffer to dissociate the co-purified *E. coli* genomic DNA. The buffer included 50 mM phosphate (pH 7.4), 2 M NaCl, for protein characterisation studies. Phosphate purified protein was subsequently concentrated and dialysed into 20 mM Tris (pH 7.4), 150 mM NaCl buffer if required for crystallisation studies. Fractions of interest were prepared for analysis by 1% agarose gel electrophoresis and 15% SDS-PAGE.

2.3.3 Transmission electron microscopy (TEM) analysis of mutants

A 5 µl aliquot of IMAC purified Lsr2 and Lsr2 R97A, R99A and R97A-R99A DNA binding mutant protein at approximately 0.5 mg/ml was prepared for TEM analysis as per standard procedure.

2.3.4 A fluorescence assay for protein thermal stability

Protein thermal stability was assessed based on the interaction of a hydrophobic fluoroprobe SYPRO™ orange (Sigma, USA) with protein as previously published (Ericsson *et al.*, 2006). IMAC purified Lsr2, Lsr2 DNA binding mutant protein samples (at approximately 0.5 mg/ml), and buffer control samples, were prepared in triplicate in thin walled PCR tubes in a total volume of 25 µl with the following final concentrations and reaction conditions. 90 x SYPRO orange dye, 10 µl of protein, in 50 mM Na acetate buffer (pH 5.2).

Protein denaturation was performed using a 36-well rotor in a RotorGene-6000 real time PCR machine and data recorded with the supplied software (Corbett Research, Australia). Denaturation conditions were as follows. Samples were equilibrated for 90 seconds at 25°C, and the temperature ramped up to 99°C in 0.2°C increments with a 5 second delay per increment. Samples were excited at 470 nm and emission measured at 555 nm with autogain settings selected. Data was exported to an excel spreadsheet and analysed by averaging the sample data and subtracting the average buffer reading. The protein unfolding temperature was measured visually from the middle of the peak.

2.3.5 Secondary structure analysis using circular dichroism (CD)

Circular dichroism was performed at The University of Auckland, in the School of Biological Sciences, using an Applied Photophysics PiStar (π*) 180 machine fitted with a Series 800 temperature controller (AlphaOmega Instruments, USA) and B-12 water bath (Techne, UK), and data was collected using the accompanying PiStar ProDataViewer software. The instrument was purged with nitrogen at a flow rate of 8 L/min for 30 minutes prior to lamp ignition and assays were performed at a flow rate of 6 L/min. For the analysis of secondary structure, data was collected in the far UV CD spectrum (180 – 260 nm) with samples assayed in a 1 mm (type 21Q) quartz cuvette positioned with a cell spacer (Starna Scientific Ltd, UK), in a total volume of 360 µl. Baseline data was collected for no cuvette present and the empty cuvette was quickly scanned prior to collecting

buffer only data to ensure cuvette cleanliness. Samples were diluted appropriately to ensure spectra absorbance was within the maximum detection limit of 1000 MV of the instrument. The cuvette was cleaned between protein samples by filling with a 3:1 solution of concentrated sulphuric acid and nitric acid and leaving for 5 minutes, then thoroughly rinsing with dH_2O followed by 100% ethanol and drying.

SEC purified "DNA-free" protein samples were dialysed into 5 mM phosphate (pH 7.4) buffer and diluted to 6 µM – a concentration which was previously optimised to give absorbance readings within the detection limits – for analysis. Five replicate scans were collected per sample, at 20°C, with 3 seconds per time point, and a 1 nm step size and bandwidth.

Data sets were analysed by averaging the sample data and subtracting the average buffer data using ProData PiStar software and then exported in a excel compatible file format for further analysis. Raw data in millidegrees was converted into mean residue ellipticity (θ_{MRW}; $deg.cm^2/dmol$) units using the equation below and multiplied by 0.001 for secondary structure analysis using the K2D2 method (Perez-Iratxeta & Andrade-Navarro, 2008).

$$\theta_{MRW} = S/10*C*n_{pept}*L$$

S = raw signal (millidegrees)
C = protein concentration (mol/L)
n_{pept} = number of peptide bonds in protein (residues-1)
L = optical path length (cm)

For the analysis of nucleic acid conformational changes, data was collected between 230 and 360 nm following established techniques (Cary & Kneale, 2008). Samples were assayed using the same settings and buffer as for secondary structure analysis. Five replicate scans of plasmid DNA at 0.008 µM were performed to establish an elliptical pattern for DNA and then a protein titration experiment was performed with five replicate scans after the addition of approximately 1 µM, 2 µM, 3 µM, and 4 µM amounts of SEC purified Lsr2 protein.

An equivalent protein only titration experiment was performed to confirm that the protein was not contributing to the elliptical signal shown in the DNA and protein titration experiment. Change in ellipticity at 272 nm was examined against protein concentration to determine the influence of protein binding on DNA conformation.

Thermal unfolding analysis of SEC purified Lsr2 was performed at a fixed wavelength of 223 nm. Lsr2 protein was assayed at 6 µM in a 5 mm quartz cuvette (with magnetic stirring ball; Starna Scientific Ltd, UK), which required a minimum of 1.7 ml of sample. CD analysis was performed over a 35°C to 90°C temperature range, with an increase of 1°C per minute, collecting for 14.4 seconds per time point, and with a 2 nm step size and bandwidth. The raw data was analysed visually.

2.3.6 Plasmid DNA binding gel shift assay

Plasmid DNA (200 ng or 400 ng) was mixed with an excess quantity of SEC purified Lsr2 or Lsr2 R97A-R99A mutant protein (3100 ng or 3600 ng respectively) and diluted to a final volume of 20 µl using specific pH buffers. For native Lsr2 DNA binding assessment, the mixtures were required to be dialysed overnight at 4°C (using 20 µl dialysis buttons according to standard procedure), as the SEC purified protein was in a 2M NaCl buffer, however for the R97A-R99A mutant this was not required as the protein was already highly concentrated and in a low salt buffer. The following buffers were used in the assay conditions – 50 mM Na acetate (pH 5.0), 150 mM NaCl, or 50 mM Na hepes (pH 7.0), 150 mM NaCl, or 50 mM Bis Tris Propane (pH 9.0), 150 mM NaCl. The samples were analysed by 1% agarose gel electrophoresis according to standard procedures.

2.3.7 Dynamic light scattering analysis (DLS) of purified protein

Concentrated R97A-R99A Lsr2 mutant protein used in crystallography robotic screens was assessed post-use for its aggregation and polydispersity properties using dynamic light scattering. The remaining R97A-R99A Lsr2 sample was centrifuged for 15 minutes at 14,000 rpm at 4°C to remove any precipitated protein and 20 µl was then added to a cleaned and dried quartz cuvette without generating bubbles. DLS analysis was performed using a DynaPro MTSV Dynamic Light Scattering instrument (Protein Solutions Inc., UK), using 40% laser power to collect approximately 800,000 counts. Histogram analysis of the distribution of particle sizes was performed to assess protein uniformity.

2.4 Methods for Chapter Five: Structure of the N-terminal dimerisation domain of Lsr2

2.4.1 Cloning of the N-terminal domain of Lsr2 into pPROExHTb

A total of seven variations of the N-terminal domain of Lsr2 were cloned into the *E. coli* expression vector pPROExHTb (Invitrogen) as follows. The N-terminal domain was amplified from the *lsr2*-pET30b plasmid via PCR using forward and reverse primers containing the restriction enzyme sites *NcoI* and *XhoI*, respectively, positioning a 6 x histidine tag and an rTEV protease cleavage site at the N-terminal end of the protein.

Table 2.7. Primers used for cloning seven variations of the N-terminal domain of *M. tuberculosis lsr2* into pPROExHTb.

Primer	Sequence	Length
Lsr2 Nco1 Fwd	GGG GCC ATG GCG AAG AAA GTA ACC GTC ACC	30 bp
Lsr2 Nterm Xho1 Rev	GGG GCT CGA GTC AAC CGA CGC GAC GGC CC	29 bp
Lsr2 Nterm+ Xho1 Rev	GGG GCT CGA GTC AGG ATC CGG AAC GGC CGC	30 bp
Lsr2 Nterm-1 Xho1Rev	GGG GCT CGA GTC AAC GGC CCG CCG CCA CC	29 bp
Lsr2 Nterm-2 Xho1Rev	GGG GCT CGA GTC AGC GCC CAC CGA CGC GAC	30 bp
Lsr2 Nterm-3 Xho1Rev	GGG GCT CGA GTC AGC CGC GCC GGC GCC CA	29 bp
Lsr2 Nterm-4 Xho1Rev	GGG GCT CGA GTC ATC CAC GGC CGG ATC CGG	33 bp
Lsr2 Nterm-5 Xho1Rev	GGG GCT CGA GTC ACG CGC CAC GTC CAC GGC	30 bp

Standard *Taq* DNA polymerase PCR was performed at 58°C annealing temperature and with 1.0 mM $MgCl_2$. A large scale PCR was performed, the product checked on an agarose gel, and then purified from solution and quantified using standard procedures. The PCR product was inserted into pPROExHTb via appropriate enzyme digestion and ligation and transformed into DH5α *E. coli*. Colony PCR using cloning primers was performed to select positive transformants. The cell suspension from positive clones was used to inoculate selective LB media which was incubated overnight at 37°C for plasmid preparation for sequence verification. Correctly cloned variants were then transformed by electroporation into BL21 *E. coli* for expression, using standard techniques.

2.4.2 Expression and purification of His tagged N-terminal domain

Small scale expression testing of the N-terminal domain variants was performed using a Ni Sepharose™ bead assay according to standard techniques. Large scale expression and IMAC purification using a 5 ml HiTrap™ Chelating HP column was performed according to standard techniques. Fractions containing protein of interest were prepared appropriately and initially analysed both by 1% agarose gel electrophoresis and 16.5% SDS-PAGE. In subsequent purifications, protein fractions were analysed by SDS-PAGE only. Analytical SEC was performed using a Superdex™ 75 10/300 GL column as per standard techniques with a 20 mM Tris (pH 7.4), 150 mM NaCl buffer. Preparative SEC purification was performed using a HiLoad™ 16/60 Superdex™ 75 column according to standard techniques using a 20 mM Tris (pH 7.4), 150 mM NaCl buffer. Fractions of interest were prepared for analysis by 16.5% SDS-PAGE.

2.4.3 Glutaraldehyde cross-linking of Nterm and SEC analysis

Glutaraldehyde was added to approximately 15 µg of SEC purified Nterm Lsr2 to a final 0.1% (v/v) concentration in a 15 µl volume and incubated at room temperature for 2 and 5 minutes. The reaction was stopped by the addition of 5 µl of 4 x SDS loading dye and heating for 5 minutes at 95°C. Cross-linking ability was assessed by 18% SDS-PAGE analysis.

Analytical SEC was performed, with a duplicate assay reaction, using a Superdex™ 75 10/300 GL column as per standard techniques with a 50 mM phosphate (pH 7.4), 200 mM NaCl, 20 mM imidazole buffer.

2.4.4 Crystallisation screens and fine screening of N-terminal domains

SEC purified protein of the N-terminal domain constructs were concentrated using standard techniques and subjected to crystallisation screens according to standard procedures. Robotic crystallisation screens were performed at The University of Auckland for Nterm, and both robotic crystallisation screens and Hampton crystallisation screens at The University of Waikato, were performed

for Nterm2 and Nterm3. Fine screens, additive screens and crystal seeding of the N-terminal domain variants were performed according to standard procedures. Co-crystallisation Hampton screens and fine screens of Nterm2 with trypsin were performed by the addition of trypsin (immediately prior to use), to a volume of concentrated protein suitable for completing an entire screen, at mass ratios of 1:500, 1:5000 and 1: 50000 as required. For a 24-well hanging drop fine screen at a 1:500 mass ratio dilution, this typically involved the addition of 1 µl of trypsin (1 mg/ml) to 35 µl of Nterm2 (15 mg/ml).

2.4.5 Isoleucine and leucine to methionine mutagenesis in Nterm2 for selenomethionine protein generation

2.4.5.1 Site directed mutagenesis using QuikChange™ protocol

Site directed mutagenesis to generate three alternative mutations of Nterm2 was performed based on the protocol from the Stratagene QuikChange™ II-E kit (Stratagene, USA). Primers for isoleucine to methionine and leucine to methionine mutations were designed with the single base pair residue change in the centre, flanked by approximately 15 bases on either side as recommended.

Table 2.8. Primers for site directed mutagenesis to generate methionine mutations.

Primer	Sequence	Length
Lsr2 I34M34 Fwd	CGG GGT GAC CTA TGA GAT GGA CCT TTC CAC TAA G	34 bp
Lsr2 I34M34 Rev	CTT AGT GGA AAG GTC CAT CTC ATA GGT CAC CCC G	34 bp
Lsr2 L44M44 Fwd	AAG AAT GCC ACG AAA ATG CGT GGC GAC CTG AAG C	34 bp
Lsr2 L44M44 Rev	GCT TCA GGT CGC CAC GCA TTT TCG TGG CAT TCT T	34 bp
Lsr2 L48M48 Fwd	CGA AAC TGC GTG GCG ACA TGA AGC AAT GGG TGG CG	35 bp
Lsr2 L48M48 Rev	CGC CAC CCA TTG CTT CAT GTC GCC ACG CAG TTT CG	35 bp

A fresh plasmid purification of *Nterm2*-pPROExHTb for the template DNA was prepared to ensure a high degree of supercoiled DNA. Mutagenic *Pfu* Ultra DNA polymerase (Aligent Technologies, USA) PCR was performed in a total volume of 50 µl with the following final concentrations and reaction conditions. 1 X *Pfu*

buffer, 0.2 mM deoxynucleotide mix, 2.5 U *Pfu* Ultra DNA polymerase, 125 ng each forward and reverse primer and 25 ng of template DNA.

Cycle conditions:

	95°C	0 :30	(min:sec)
X 17	95°C	0:30	
	55°C	1:00	
	68°C	6:00	
	4°C	HOLD	

Successful PCR amplification of the DNA was confirmed by 1% agarose gel electrophoresis of a reaction aliquot as per standard procedure prior to digestion of the DNA with 10 U of *DpnI* (New England Biolabs, USA) at 37°C for 1 hour. Following DNA digestion, 4 µl of the product was transformed into DH5α *E. coli* which were grown overnight on selective media as per standard procedure. Overnight selective cultures were grown from successful transformants for plasmid preparation and DNA sequencing as per standard procedure. Correctly mutated variants were transformed by electroporation into BL21 *E. coli* for expression.

2.4.5.2 Site directed mutagenesis of L44M using a standard PCR protocol

Generation of the Nterm2 methionine mutant L44M was unsuccessful using the site directed mutagenesis QuikChange™ protocol and hence a standard PCR protocol using the mutagenesis primers and the original cloning primers flanking the entire Nterm2 construct were used in a two-stage PCR reaction.

Table 2.9. Primers for site directed mutagenesis to generate the L44M mutation.

Primer	Sequence	Length
Lsr2 Nco1 Fwd	GGG GCC ATG GCG AAG AAA GTA ACC GTC ACC	30 bp
Lsr2 L44M44 Rev	GCT TCA GGT CGC CAC GCA TTT TCG TGG CAT TCT T	34 bp
Lsr2 L44M44 Fwd	AAG AAT GCC ACG AAA ATG CGT GGC GAC CTG AAG C	34 bp
Lsr2 Nterm-2 Xho1 Rev	GGG GCT CGA GTC AGC GCC CAC CGA CGC GAC	30 bp

The L44M mutant was amplified in two halves from the *Nterm2*-pPROExHTb plasmid via PCR using the Lsr2 *NcoI* Fwd/Lsr2 L44M44 Rev primer pair and Lsr2 L44M44 Fwd/Lsr2 Nterm-2 Xho1 Rev primer pair. A 50 µl standard *Taq* DNA

polymerase PCR was performed at 58°C annealing temperature, with 1.0 mM $MgCl_2$, and 80 ng of template DNA. The PCR products were checked on a 1% agarose gel, and then purified from solution and quantified using standard procedures. A standard PCR was then performed using the PCR products from the two halves of the *Nterm2* gene as template and the Lsr2 Nco1 Fwd/Lsr2 Nterm-2 Xho1 Rev primer pair. A 50 µl standard *Taq* DNA polymerase PCR was performed at 58°C annealing temperature, with 1.0 mM $MgCl_2$, and 40 - 60 ng of each template DNA. The PCR product was checked on a 1% agarose gel, and then purified from solution and quantified using standard procedures.

2.4.5.3 Cloning of the L44M Nterm2 mutant into pPROExHTb

The L44M Nterm2 PCR product was inserted into pPROExHTb via appropriate enzyme digestion and ligation and transformed into DH5α *E. coli*. Colony PCR was performed using cloning primers to select positive transformants. The cell suspension from positive clones was used to inoculate selective LB media which was incubated overnight at 37°C for plasmid preparation for sequence verification. A correctly cloned mutant was then transformed into BL21 *E. coli* for expression, using standard techniques.

2.4.6 Expression and purification trials of the I34M, L44M and L48M Nterm2 methionine mutants

To confirm that each of the three methionine mutations had not compromised protein expression or behaviour, large scale expression followed by IMAC purification using a 5 ml HiTrap™ Chelating HP column, and analytical SEC purification using a Superdex™ 75 10/300 GL column was performed using the same procedures used with the native Lsr2 N-terminal domain constructs.

2.4.7 Incorporation of selenomethionine into L48M Nterm2 protein

A common procedure to express selenomethionine incorporated protein included the use of a methionine auxotroph DL41 *E. coli* strain in conjunction with autoinduction media supplemented with an appropriate ratio of methionine to selenomethionine. The *L48M Nterm2*-pPROExHTb plasmid was transformed into DL41 cells using standard procedures, and a single colony from the resulting agar plate was used to inoculate 10 ml of PA-0.5G selective media (supplemented with methionine) which was then incubated overnight at 37°C, whist shaking at 180 rpm. The 10 ml seeder culture was used to inoculate 400 ml of PASM-5052 autoinduction media (which includes selenomethionine and methionine) which was grown overnight at 37°C, whist shaking at 180 rpm.

An alternative procedure to express selenomethionine incorporated protein included the use of BL21 *E. coli* and the methionine pathway inhibition method (Doublié, 1997). This technique uses M9 minimal media supplemented with precise amounts of each amino acid. A single colony of the transformed BL21 cells was used to inoculate 10 ml of selective LB media which was grown overnight at 37°C, whist shaking at 180 rpm. The overnight culture was centrifuged at 4600 rpm for 10 minutes and resuspended in 1 ml of M9 minimal media (modified from Sambrook & Russell, 2001). This culture was used to inoculate 1 L of fresh pre-warmed selective M9 minimal media which was then grown at 37°C, shaking at 180 rpm until it reached an OD_{600} of 0.4 – 0.6. Once in mid-log phase, the following amounts of each amino acid were added. One hundred mg each of lysine, phenylalanine and threonine; 50 mg each of isoleucine and valine; and 60 mg of selenomethionine. Growth was continued for a further 15 minutes at 37°C and then protein expression was induced with 0.5 mM IPTG and incubation continued overnight.

2.4.8 IMAC purification and SEC of L48M selenomethionine Nterm2

Large scale expression cultures were purified by IMAC using a 5 ml HiTrap™ Chelating HP column according to standard techniques except that both the phosphate lysis and elution buffers were at pH 7.0 (instead of pH 7.4), and they contained 2 mM β-mercaptoethanol. Fractions containing protein of interest were prepared appropriately and analysed by 16.5% SDS-PAGE. Preparative SEC purification was performed using a HiLoad™ 16/60 Superdex™ 75 column according to standard techniques except that a 20 mM Tris (pH 7.2), 150 mM NaCl, 2 mM β-mercaptoethanol buffer was used. Fractions of interest were prepared for analysis by 16.5% SDS-PAGE.

2.4.9 L48M Nterm2 crystallisation fine screens and co-crystallisation

Concentrated SEC purified L48M Nterm2 was prepared and subjected to hanging drop fine screens and trypsin co-crystallisation screens at a 1:500 mass ratio according to standard techniques. These screens were designed based on previous successful Nterm2 crystallisation conditions.

2.4.10 X-ray diffraction data collection and processing

Details of X-ray diffraction data collection are described in section 2.1.18.4. Nterm2 protein crystals did not require an additional cryoprotectant and were flash frozen and tested for diffraction at the home source, and two data sets from two independant crystals were successfully collected. Crystals from Nterm2 co-crystallised with trypsin were bathed in the screen solution containing 20% glycerol and flash frozen and also tested for diffraction at the home source, and a single data set was collected. A further data set from this crystal was also collected at the Australian Synchrotron. Selenomethionine L48M Nterm2 crystals were bathed in the screen solution containing 20% glycerol and flash frozen and tested for diffraction at the Australian Synchrotron, and two multiwavelength data sets from two independent crystals were successfully collected.

The program MOSFLM was used to visualise and index images, and integrate the reflections (Leslie, 1992). The program auto spot finder and auto index functions were used to determine the cell parameters. Changes in RMS deviation and mosaicity were observed during MOSFLM integration. The integrated reflections were scaled and merged using SCALA within the CCP4 program (Bailey, 1994). The output using all the reflections was examined for unfavoured Rmerge, I/σ(I) and data completeness and rescaled if necessary to improve these variables.

2.4.11 Phase determination and structure solving by *ab initio* methods

Phase determination and structure solving using *ab initio* techniques was performed by K. Meindl and I. Uson at the Instituto di Biologia Molecular de Barcelona at the Barcelona Science Park, Barcelona, Spain (Rodriguez *et al.*, 2009). The $P2_1$ structure was solved with the program ARCIMBOLDO (Rodriguez *et al.*, 2012) on the supercomputer Calendula at the FCSCL, León, Spain (www.calendula.fcsc.es). A model alpha helix composed of 14 alanine residues was used as a search fragment. Location of two fragments led to partial solutions that could be distinguished from the average by a slightly increased correlation coefficient (CC) of the traces obtained by SHELXE (Sheldrick, 2010). However the solutions were stuck at mean phase errors of 73° compared to the phases of the final refined and deposited model. Iterative improvements were made by interspersing density modification and autotracing with SHELXE jobs with trimming of the partial models obtained to increase their CC until a mean phase error of 42° was reached producing an easily interpretable map.

The structure in the $P3_121$ space group was determined by molecular replacement using PHENIX (Adams *et al.*, 2010) and the earlier determined Nterm2 Lsr2 structure in the $P2_1$ space group, which had subsequently been manually built and refined, as a search model.

2.4.12 Model building and refinement

Automated model building for the $P2_1$ structure was performed using PHENIX Autobuild (Adams *et al.*, 2010) and an edited version of the *ab initio* structure as the search model. Residues A28, A29, A33 and A34 were manually removed from the "test3.pdb" coordinate file and the file was renamed "test3_edit1.pdb". Iterative rounds of manual building using COOT (Emsley & Cowtan, 2004) and refinement using PHENIX Refine (Adams *et al.*, 2010) were performed.

Automated model building for the $P3_121$ structure was performed using PHENIX Autobuild and the final refined version of the $P2_1$ structure as the search model. Iterative rounds of manual building using COOT and refinement using PHENIX Refine were performed.

Structure validation was performed using a subset of the full MolProbity analysis available within PHENIX Refine (Chen *et al.*, 2010) and KiNG analysis within PHENIX Refine (Chen *et al.*, 2009).

2.4.13 Structural analysis of Nterm2 and figure generation

All structure images and electrostatic surface potential images of structures depicted in this thesis were generated using PYMOL (DeLano, 2002).

2.4.14 MALDI-TOF mass spectrometry of whole protein

For analysis of whole protein, a 9:1 mixture of 2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid (DHBs) matrix (Sigma, USA) was used. Ten µg of matrix was added to 30 µl of 2:1 ACN : 0.1% TFA, mixed well, sonicated for 10 minutes in a sonicating water bath (Elma, Germany) and then centrifuged for 10 minutes at 13,000 rpm. Protein samples and protein calibration standards (Bruker Daltonics, USA) were mixed in a 1:1 ratio (0.5 µl:0.5 µl) with DHBs matrix on an AnchorChip™ MALDI-TOF target plate (Bruker Daltonics, USA) and left to air dry. Protein samples were usually dialysed into MQ H_2O before analysis.

An autoflex™ II MALDI-TOF mass spectrometer (Bruker Daltonics, USA) was used to analyse samples. Samples were analysed in linear mode, with the mass range

selector adjusted as required and data in the range of 2500 – 45000 Da collected. Settings included a pulsed ion extraction of 450 ns, a gain of 2500 V, and an acceleration voltage of 20 kV, with laser power typically around 60%. Spectra for protein calibration standards were first collected and the spectrometer calibrated with an automatic polynomial correction. Spectra were exported to FlexAnalysis™ software (Bruker Daltonics, USA), and the peaks identified and labelled. Peaks were assessed for the presence of full protein, protein with the tag removed, and the tag alone and for the formation of oligomers.

2.4.15 Electrospray ionisation mass spectrometry (ESI-MS)

Electrospray ionisation mass spectrometry analysis of the N-terminal domain of Lsr2 was performed using a Bruker Daltonics microTOF™ machine and the data collected and analysed with micrOTOFControl and Bruker Daltonics DataAnalysis software. Protein crystals of "matured" and trypsin co-crystallised N-terminal Lsr2 were washed briefly in acetonitrile and then dissolved in 5 µl 1:1 acetonitrile:H_2O. Dissolved crystals were diluted to 100 µl in 1:1 acetonitrile:H_2O and 25 µl of 0.2% formic acid (in 1:1 acetonitrile:H_2O) added. Calibration was performed immediately prior to analysis with infused sodium formate solution in positive ion mode and the acquired masses were matched to reference masses.

2.4.16 Full length Lsr2 trypsin and DNase digestion assay

IMAC purified Lsr2 protein (2.4 mg/ml) was mixed with trypsin (1 mg/ml, Sigma, USA) at a 1:500 mass ratio in a final volume of 400 µl and incubated at room temperature. Duplicate samples of 15 µl were taken at 1 minute, 5 minutes, 15 minutes, 30 minutes and 1 hour for 1% agarose and 16.5% SDS-PAGE gel analysis. Samples intended for SDS-PAGE analysis has 5 µl of 4 x SDS loading dye added and were heated at 95°C for 5 minutes. Samples for agarose gel analysis had the trypsin digestion stopped by adding an equal ng amount of tryspin inhibitor (0.1 mg/ml, Sigma, USA) to the aliquot and were mixed and left at room temperature for a minimum of 15 minutes before the addition of a DNA loading

dye. Samples of 30 µl were also taken at 1 minute, 5 minutes, 15 minutes, 30 minutes and 1 hour for subsequent DNase digestion. All samples for DNase treatment had trypsin inhibitor added at each sampled time point and had a minimum of 45 minutes treatment time at room temperature before the addition of DNase. To 30 µl of sample, 3 µl of 10 x DNase buffer and 2 µl of RQ1 DNase (1 U/ µl, Promega, USA) was added, mixed and left at room temperature for 90 minutes. To 15 µl of DNase treated sample, 5 µl of 4 x SDS loading dye was added and the sample heated at 95 °C for 5 minutes. DNA loading dye was added to the remaining 15 µl for agarose gel analysis. Trypsin inhibitor was added to control samples for agarose gel analysis. DNA was visualised by 1% agarose gel electrophoresis and staining with SYBR®safe DNA stain (Invitrogen).

2.4.17 Transmission electron microscopy (TEM) of trypsin digested full length Lsr2

IMAC purified Lsr2 protein (2.4 mg/ml) was mixed with trypsin (1 mg/ml, Sigma, USA) at a 1:500 mass ratio in a final volume of 200 µl and incubated at room temperature. Duplicate 5 µl samples were taken at 1 minute, 5 minutes 15 minutes, 30 minutes and 1 hour and immediately prepared for negative staining duplicate grids as per standard procedure. Electron microscopy grids were examined using a FEI TecNai G^2 electron microscope and low dose images were recorded.

2.5 Methods for Chapter Six: Pellicle and biofilm formation in *M. smegmatis* mc²155 Δ*lsr2* and Lsr2 mutant strains

2.5.1 *Mycobacterium smegmatis* strains and Lsr2 variants

The following control stains and experimental variants were generated for use in all biofilm and pellicle formation assays. The *M. smegmatis* mc²155 lab strain containing an unmarked deletion of the *lsr2* gene MSMEG_6092 was provided by R. Colangeli (NJS22; Colangeli *et al*., 2007).

Table 2.10. *Mycobacterium smegmatis* strains used and genes complemented with.

Name	Strain	Vector
mc²155	*M. smegmatis* mc²155 lab strain	none
K/O Lsr2	*M. smegmatis* mc²155 with a clean deletion of *lsr2*	none
mc²155 + pMIND	*M. smegmatis* mc²155 lab strain	Empty pMIND
K/O Lsr2 + pMIND	*M. smegmatis* mc²155 with a clean deletion of *lsr2*	Empty pMIND
R97A	*M. smegmatis* mc²155 with a clean deletion of *lsr2*	R97A-pMIND
R99A	*M. smegmatis* mc²155 with a clean deletion of *lsr2*	R99A-pMIND
R97A-R99A	*M. smegmatis* mc²155 with a clean deletion of *lsr2*	R97A-R99A-pMIND
Lsr2	*M. smegmatis* mc²155 with a clean deletion of *lsr2*	Lsr2-pMIND

2.5.2 *Mycobacterium smegmatis* biofilm culture method

Glycerol stocks of each control strain and variant strain were streaked onto LBT agar and hygromycin LBT agar plates respectively, and grown at 37°C for 3 days for use in seeder cultures. Seeder cultures of each strain were grown in 1 ml of LB media, with or without hygromycin as required, for 48 hours, shaking at 180 rpm, and were used to inoculate 4 ml of M63 biofilm media at 1:100 dilution ratio. For the control lab strain and Δ*lsr2* strain, M63 biofilm media without hygromycin or tetracycline was prepared. For all the strains containing either the empty pMIND vector or containing an inserted gene in pMIND, M63 media including hygromycin and tetracycline for induction was prepared. All strains were cultured in triplicate in 6-well CellStar polystyrene plates with lids (Greiner BioOne, Germany) at 37°C for 7 days in a humidified box, without disturbing. After 7 days the plates were visually inspected and photographed prior to quantifying pellicle and biofilm formation.

2.5.3 *Mycobacterium smegmatis* biofilm and pellicle formation assay

After growth for 7 days at 37°C in M63 biofilm media, control strains and variants were assayed for pellicle formation by examining the dry weight of the air-liquid interface cell layer and assayed for biofilm formation by crystal violet staining of the surface attached cell layer. A technique was developed to retrieve almost 100% of the pellicle cell layer. Thick blotting paper (ADVANTEC®, Japan) cut to size for the 6-well dishes with a tag on the side was dried at 50°C overnight and weighed immediately before use. A single circular blotting paper was gently laid on top of the surface pellicle and allowed to soak up media before withdrawing it and laying it pellicle side up on a tray to dry. Blotting paper was dried at 50°C for 2 hours and then weighed.

Biofilm formation was assessed after the pellicle had been removed. Each plate was inverted to discard the M63 biofilm media and 2 ml of MQ H_2O gently applied to wash the wells. The plates were inverted to discard the MQ H_2O and blotted onto paper towels to remove the last traces of liquid and then 0.5 ml of 0.1% crystal violet was applied and left at room temperature for 30 minutes with gentle swirling (60 rpm). The stain was gently removed with a pipette for disposal and the wells gently washed with 3 x 2 ml MQ H_2O by carefully adding the liquid, then gently swirling the plate by hand and inverting it and blotting onto paper towels. Crystal violet stained biofilm was dissolved by adding 2 ml of 100% ethanol to each well and incubating at room temperature for 1 hour while gently swirling (60 rpm). Staining was quantified by measuring the absorbance at OD_{570} after a 1:100 dilution of the solubilised crystal violet solution into 1 ml of 100% ethanol.

2.5.4 SYTO®9 DNA staining of surface pellicle mycobacteria

Square 22 mm glass coverslips (Deckgläser, Germany) were cleaned with a dry tissue and held bridged between plastic forceps, before carefully placing on top of the air-surface pellicle to attach it to the glass. The coverslip was placed pellicle side up on a tray and dried for 30 minutes at 50°C. The dried coverslips were stained with SYTO®9 DNA stain (Molecular Probes, USA) as follows. DNA

stain was diluted to 5 µM in TE buffer and 200 µl applied to the coverslip surface and the samples incubated at room temperature in the dark for 40 minutes. Coverslips were then rinsed by carefully submerging in a beaker of MQ H_2O for 10 seconds. A drop of mounting media was applied to a plain glass microscope slide and the stained coverslip laid facedown upon the drop. Cells were viewed at 100 x oil immersion under blue light (ex = 485, em = 498) and photographed with a digital camera.

2.5.5 Cloning into pMIND mycobacteria expression vector

Primers were designed both with and without a ribosome binding site (RBS) for cloning into pMIND to be confident that this would not influence the experimental outcome. Lsr2 and three DNA binding mutants of this protein were cloned into the mycobacteria tetracycline inducible expression vector pMIND supplied by R. O'Toole (Centre for Biodiscovery, School of Biological Sciences, Victoria University of Wellington; Blokpoel *et al*., 2005) as follows. *Lsr2* and the three DNA binding mutants, R97A, R99A and R97A-R99A were amplified from their respective pET30b plasmids via PCR using forward and reverse primers containing the restriction enzyme sites *PstI* and *SpeI*, respectively.

Table 2.11. Primers used for cloning *M. tuberculosis lsr2* and mutants into pMIND.

Primer	Sequence	Length
Rv3597c_Pst1_Fwd	GCA ACT GCA GAT GGC GAA GAA AGT AAC CGT CAC CTT G	37 bp
Rv3597c_Pst1_RBS_F	GGG GCT GCA GGG AGG AAT AAT GGC GAA GAA	30 bp
Rv3597c_Spe1_Rev	AAC TAG TTC AGG TCG CCG CGT GGT ATG CGT C	31 bp

Standard *Taq* DNA polymerase PCR was performed at 58°C annealing temperature and 1.0 mM $MgCl_2$ concentration, using 15 – 50 ng of purified plasmid as a template. Gel extraction of a combined 4-fold PCR reaction size was performed and the DNA product quantified. The PCR product was inserted into pMIND via appropriate enzyme digestion and ligation and transformed into TOP10 *E. coli* for colony PCR using cloning primers to confirm successful

transformants. The cell suspension from positive clones was used to inoculate selective low salt LB media which was incubated overnight at 37°C for plasmid preparation for sequence verification. Correctly cloned variants were then transformed into the Δ*lsr2 M. smegmatis* mc^2155 strain (NJS22; Colangeli *et al*., 2007) using standard techniques.

2.5.6 Strain typing of biofilm assay variants using colony PCR

Primers designed for confirming cloning success in the pMIND vector by DNA sequencing and primers designed for the *M. smegmatis* mc^2155 *lsr2* gene MSMEG_6092 as well as the primers designed for cloning the variants into pMIND were all used for strain verification by colony PCR.

Table 2.12. Primers used for strain typing of *M. smegmatis* biofilm variants.

Primer	Sequence	Product
MSMEG_6092 Fwd	ATG GCA AAG AAA GTG ACC GTC ACG CTT G	345 bp
MSMEG_6092 Rev	CTA AGT TGC CGC GTG GAA TGC GTC GAT C	345 bp
Rv3597c_Pst1_RBS_F	GGG GCT GCA GGG AGG AAT AAT GGC GAA GAA	360 bp
Rv3597c_Spe1_Rev	AAC TAG TTC AGG TCG CCG CGT GGT ATG CGT C	360 bp
pMIND Fwd Primer	GAT AAA GTG ACT GCT CGC TAC TCT	190/550 bp
pMIND Rev Primer	ATC ATC CGA ATC AAT ACG GTC GA	190/550 bp

After transformation of the required pMIND vector into the appropriate strain and subsequent cell growth on solid media, a single colony was removed and resuspended in 5 µl of MQ H_2O, mixed with 100 µl of PEG200/KOH solution and heated for 15 minutes at 95°C. Two µl of the lysed cell mixture was used in a standard *Taq* DNA polymerase PCR performed at a 58°C annealing temperature and with 1.0 mM $MgCl_2$. pMIND sequencing primers were used at a lower dilution of 0.3 µM compared with the standard 0.6 µM amounts. PCR products were analysed by 1% agarose gel analysis as per standard procedure. The presence of an empty pMIND vector would result in a 190 bp PCR product in comparison with a 550 bp product if the vector contained the *lsr2* gene.

Chapter Three – Lsr2 expression, purification & characterisation

3.1 Introduction

This chapter focuses on the expression and purification of the *M. tuberculosis* DNA binding protein Lsr2. DNA binding proteins control cellular processes such as DNA replication, repair and recombination. In bacteria, gene regulation is vital to be able to respond to changing environmental conditions and ultimately survive. Nuclear extracts are often used for purification of DNA binding proteins from eukaryotes but in the case of bacteria, the whole cell lysate is usually the starting material. Isolation of naturally expressed DNA binding proteins from within the cell is challenging as the proteins are normally expressed at low levels and several purification steps are required (Gadgil *et al.*, 2001). Use of an *E. coli* expression system alleviates these issues enabling a high level of protein expression and reducing the number of purification steps required.

Lsr2 has been previously expressed in *M. smegmatis* mc^2155 and *E. coli*. Purified recombinant Lsr2 protein from the cytosolic fraction of *M. smegmatis* mc^2155 and cell lysate of *E. coli* has been shown to form both dimers and tetramers, respectively, in *in vitro* glutaraldehyde cross-linking experiments (Chen *et al.*, 2006; Chen *et al.*, 2008). Further cross-linking studies with *E. coli* expressed recombinant Lsr2 in complex with a 230 base pair DNA fragment indicated progressive oligomerisation of Lsr2 (Colangeli *et al.*, 2007). These *in vitro* studies give preliminary information about the association of Lsr2. However, as this technique can allow non-specific protein interactions, it can give artificial results (Fadouloglou *et al.*, 2008).

Work presented in this chapter shows Lsr2 is a non specific DNA binding protein, is stable whilst bound to DNA (but poorly behaved without DNA), forms dimers and tetramers *in vivo* and is capable of forming large oligomeric structures when bound with DNA.

3.2 Results

3.2.1 Lsr2 expressed in *E. coli* co-purifies with genomic DNA

The *E. coli* expression plasmid pET30b containing the *M. tuberculosis* H37Rv *lsr2* gene (*lsr2*-pET30b) was kindly supplied by R. Colangeli and was cloned as follows (Colangeli *et al*., 2007). The *lsr2* gene (Rv3597c) was amplified by PCR from *M. tuberculosis* H37Rv genomic DNA and cloned into pET30b (Novagen) using the *NdeI*/*XhoI* restriction sites, positioning a 6 x histidine tag at the C-terminus of the protein. DNA sequencing of amplified and purified plasmid confirmed the insertion of *lsr2* within the *NdeI*/*XhoI* restriction sites and the positioning of the 6 x histidine tag at the C-terminal end. The lack of a cleavage site for the histidine tag gave a recombinant protein size of 13.2 kDa. During this time, research by Colangeli and colleagues demonstrated that Lsr2 was a DNA binding protein (Colangeli *et al.*, 2007).

Small scale expression testing, histidine tag binding assays and large scale expression of Lsr2 prior to the initiation this study resulted in predominantly soluble protein that was purified using immobilised metal affinity chromatography (IMAC). It was then realised that the large protein aggregates seen in the initial size exclusion chromatography (SEC) purification were Lsr2 protein in complex with *E. coli* genomic DNA (as shown by agarose gel analysis). Consequently, a DNase enzyme digestion step was incorporated into the IMAC purification method to try to remove DNA from the preparation and obtain a more uniform protein solution. Removal of the DNA from Lsr2 by enzyme digestion was unsuccessful and although the use of a urea denaturing method removed the DNA from Lsr2, attempts to subsequently refold the protein (and visualise using native gels) were not successful.

During this study, large scale expression cultures typically took around six hours to reach an appropriate optical density worthy of induction, which was three times as long in comparison with other *E. coli* cultures being used to express proteins. It appeared that the expression of Lsr2 in *E. coli* slowed the growth of the bacteria. Furthermore, during this study Lsr2 consistently co-purified with

E. coli genomic DNA during IMAC, an observation that has never been reported in other publications. The chromatogram and corresponding SDS-PAGE and agarose gels (Figure 3.1) illustrate the tight binding of Lsr2 to the column used for IMAC and the resulting peaks and fractions that include Lsr2 bound with DNA. Proteins bound non-specifically to the IMAC column are eluted early in the imidazole elution gradient, whereas Lsr2 is eluted at 100 % elution buffer which corresponds to 500 mM imidazole.

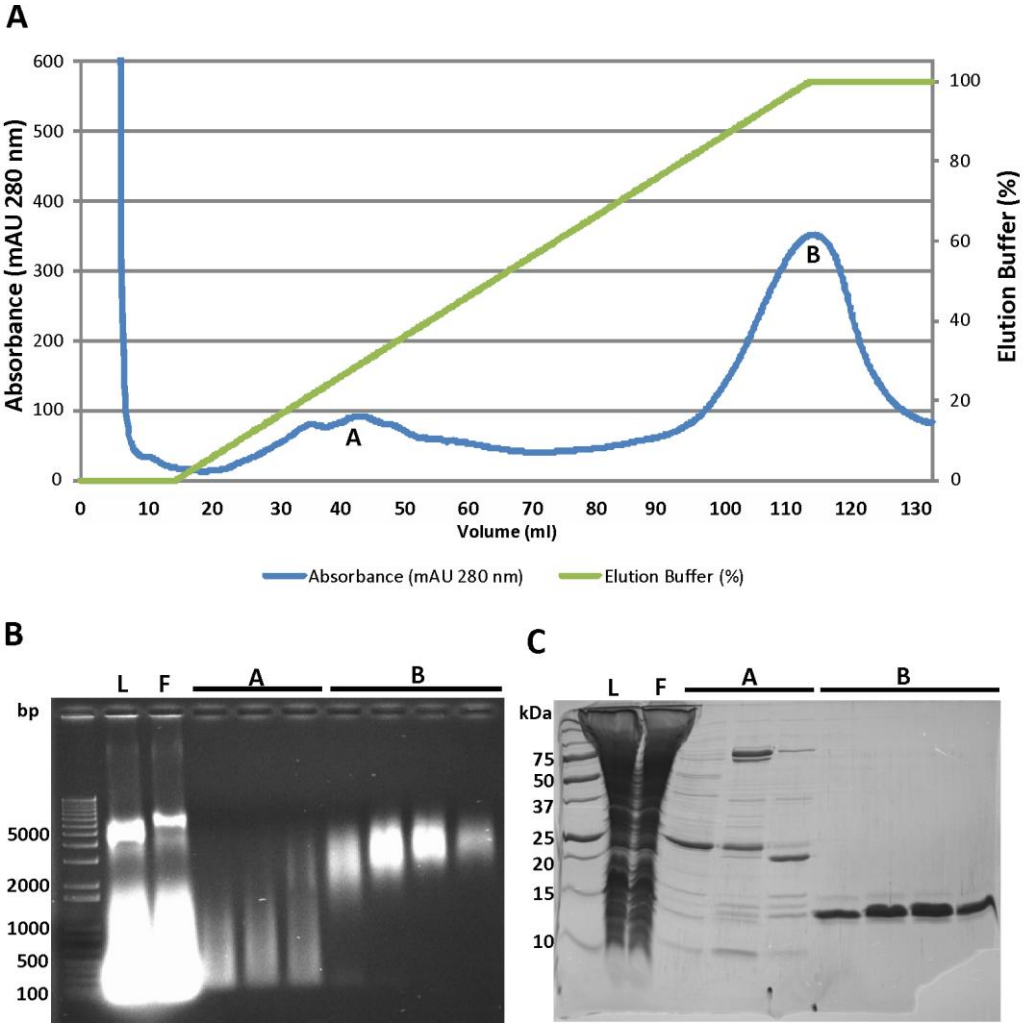


Figure 3.1. Lsr2 co-purifies with genomic DNA during IMAC purification. (A) Chromatogram of the elution buffer gradient and UV absorbance profile; (B) 1% agarose gel with sample lysate (L), column flow through (F), and fractions from peak A and B; (C) 15% SDS-PAGE gel of the corresponding lysate (L), flow through (F) and fractions from peak A and B.

The routine DNase digestion of the cell lysate prior to IMAC was later removed from the purification protocol as it was felt that it was not an efficient method for reducing DNA co-purification as evidenced by Figure 3.1. As a consequence of this, the cell lysate was viscous and difficult to filter prior to application on the IMAC column so the 0.2 µm fine filtering step was removed from the filtering process and a 0.45 µm filter was the finest used. The protein yield was uncompromised by these alterations and further purification steps were investigated to remove the co-purified DNA (section 3.2.2). IMAC purified protein was very stable at 4°C for weeks to months without showing signs of protein or DNA degradation.

Transmission electron microscopy (TEM) was used as a technique to visualise the complex of Lsr2 bound with DNA. Early images showed a complex arrangement of strands or "fibrils" that represent DNA coated by Lsr2. These images confirmed that DNA was plentiful in the purified Lsr2 protein samples but that DNase digestion of this preparation was of limited use as a purification step as digestion was not uniform and was unable to completely remove the DNA (Figure 3.2).

Throughout this study TEM was commonly used to visualise the arrangement of Lsr2 with co-purified DNA to monitor the affects of subsequent purification steps. Higher quality preparations of Lsr2 in complex with DNA were produced which allowed some preliminary measurements to be made to propose the manner in which Lsr2 was associated with the DNA. The negatively stained protein sample revealed regular fibrils of around 90 Å in diameter (Figure 3.3A), that show Lsr2 decorating the genomic DNA. It was proposed that Lsr2 is able to protect DNA from DNase treatment by binding and shielding the DNA. Single particle analysis performed by our collaborators was used to help visualise the structure of the complexes seen using TEM. Class average analysis of hundreds of short linear sections from several micrographs of these fibrils shows a helical winding pattern, suggesting that Lsr2 is bound to the DNA following its helical topology (Figure 3.3B).

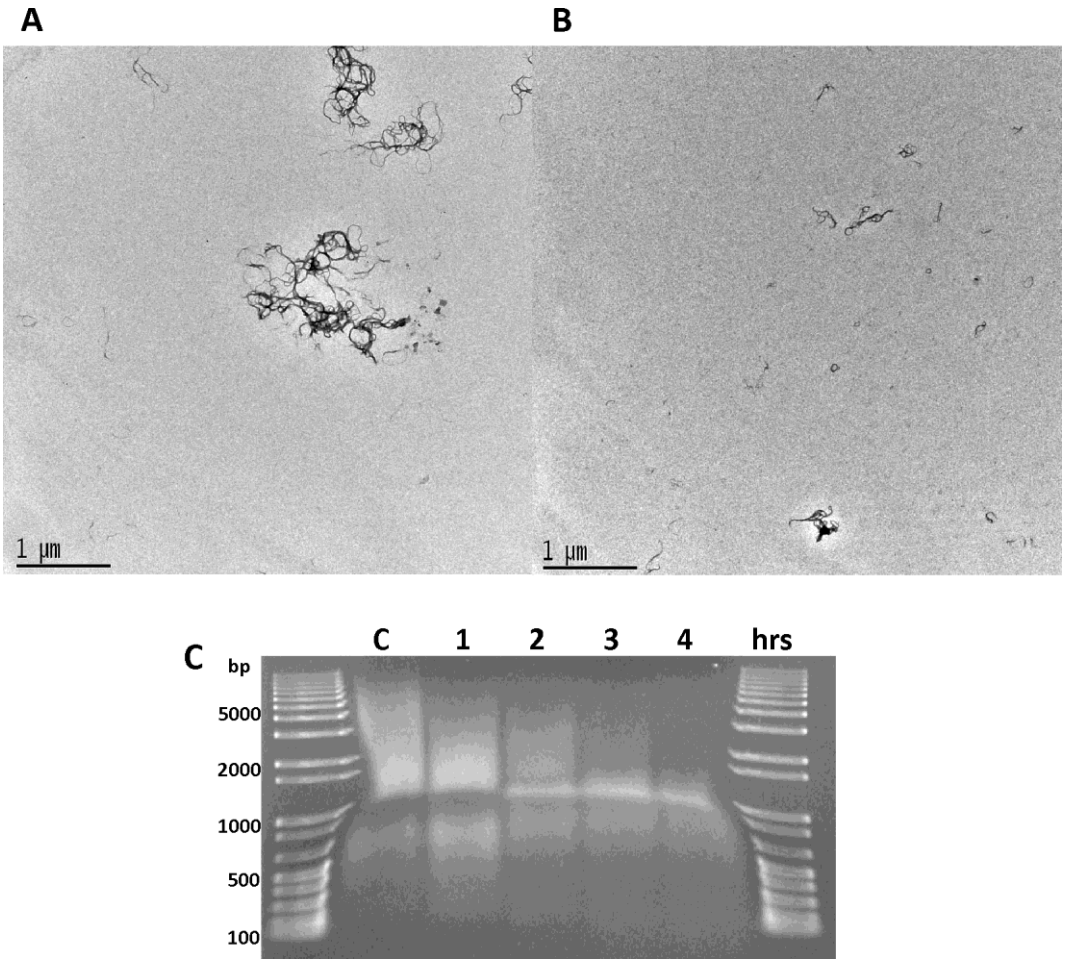


Figure 3.2. Electron micrograph images of IMAC purified Lsr2 shows co-purification with genomic DNA and limited DNase digestion. (A) Negative stained purified Lsr2 shows intertwined and coiled fibrils of DNA coated by protein; (B) after DNase digestion for 1 hour at 37°C the fibril-like appearance remains but the coils are in smaller clumps; (C) 1% agarose gel of Lsr2 digested with DNase at 37°C for up to 4 hours shows limited digestion (C = control).

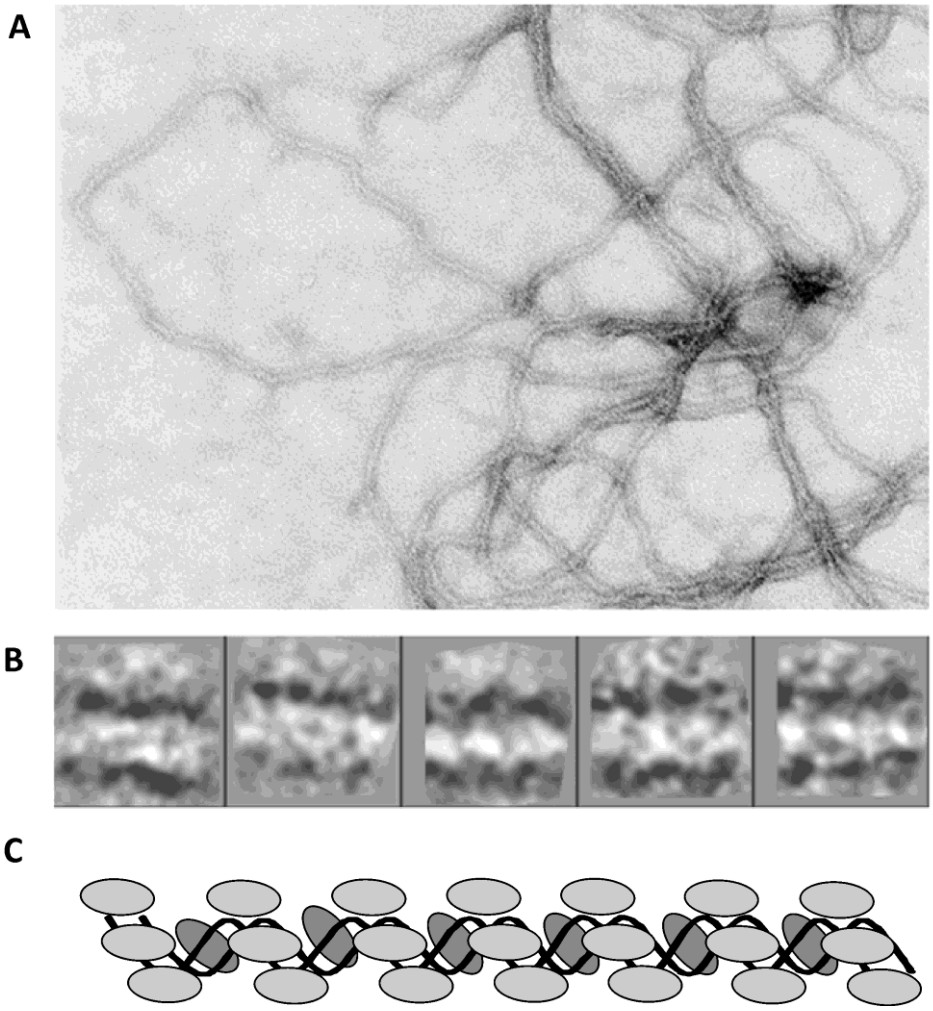


Figure 3.3. Detailed electron micrograph images reveal the structure of Lsr2/DNA complexes. (A) Negative stained purified Lsr2 displays a filamentous structure; (B) five representative class averages of Lsr2 isolated from micrographs reveals a structure with helical winding and a diameter of 90 Å; (C) A cartoon representation of Lsr2 coating a strand of DNA that illustrates how Lsr2 may coat the DNA resulting in a diameter of 90 Å (micrograph images in A and B were published (Colangeli *et al.*, 2009)).

From the detailed micrograph images a model of DNA binding is proposed. A cartoon illustration of this model (Figure 3.3C) shows how the non-specific DNA binding nature of Lsr2 would allow the formation of these large fibrils. Our proposal is that the 90 Å fibril comprises of one strand of DNA (≈ 22 Å in diameter) coated on either side by Lsr2 (≈ 34 Å in diameter; estimated to be similar in size to lysozyme).

The large amount of DNA co-purified with Lsr2 significantly influences its behaviour in *in vitro* studies. For example, theoretically, Lsr2 alone would require reversed electrode orientation to migrate through a native polyacrylamide gel due to its low isoelectric point (≈ 10). However, Lsr2 purified using IMAC and in complex with DNA migrates with the normal electrode orientation. It was also noted that the migration of the complex was retarded once it reached the higher percentage resolving gel which was thought to be due to its large mass (results not shown).

3.2.2 High salt buffers disrupt Lsr2 and DNA interaction

Initial SEC purification of recombinant Lsr2 (isolated from *E. coli* using IMAC) revealed what was thought to be large soluble aggregates of protein or incorrectly folded protein as a UV absorption peak appeared early and corresponded with the column void volume which verifies the presence of a large complex. As detailed in section 3.2.1, it was discovered that Lsr2 was co-purifying with DNA and subsequent unsatisfactory attempts using DNase treatment prior to IMAC and SEC to completely remove the DNA led to the use of high salt buffers for DNA removal.

IMAC-purified Lsr2 in complex with DNA was subjected to treatment with high salt buffers beginning with a 1 M NaCl buffer (in small gel electrophoresis assays) and increasing the concentration to buffers containing 2 M and 4 M NaCl. Small scale analytical SEC was initially used to determine the lowest NaCl concentration required to dissociate Lsr2 from the DNA. The best procedure to adjust the protein buffer solution from low to high salt and to also concentrate the protein for SEC whilst still keeping the protein soluble was also determined. At this point, the addition of an equal volume of a buffer containing 4 M NaCl to the protein sample in the standard IMAC elution buffer (resulting in a final 2 M NaCl condition) and an overnight incubation at 4°C was preferred. Centrifugation steps to concentrate the protein in high salt were also performed at 4°C.

Large scale preparative SEC was routinely used to prepare DNA-free Lsr2 protein. The complex of Lsr2 and genomic DNA in a buffer containing 2 M NaCl was efficiently separated by SEC. DNA contributed significantly to the UV absorption at 280 nm and was shown to pass quickly through the column, eluting immediately after the void volume whilst Lsr2 was retarded by the column and passed through more slowly (Figure 3.4A). Agarose gel and SDS-PAGE gel analysis of fractions collected during SEC confirmed that the protein and DNA were indeed separated (Figure 3.4B and C).

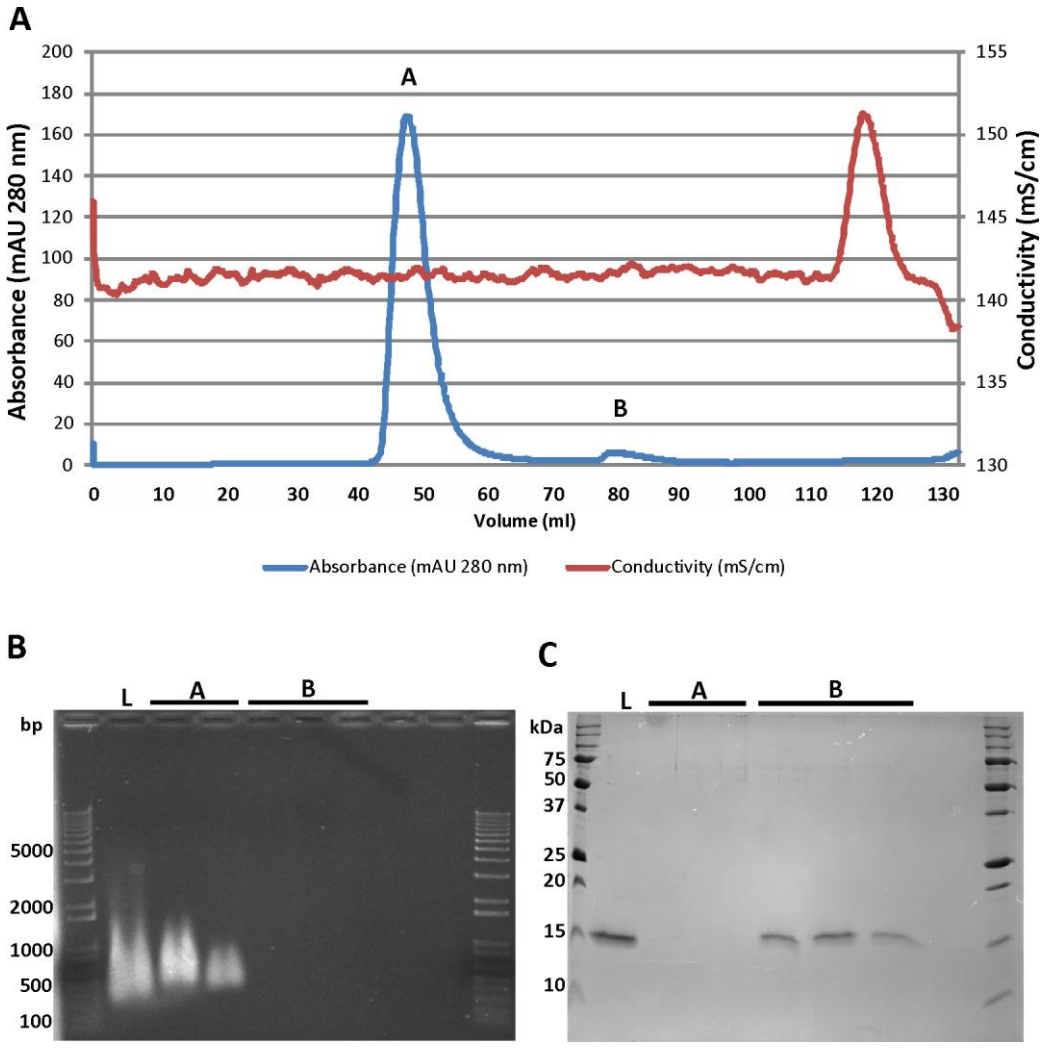


Figure 3.4. Size exclusion chromatography of recombinant Lsr2 in a 2 M NaCl buffer dissociates Lsr2 from co-purified genomic DNA. (A) Chromatogram of the UV absorbance profile of DNA(A) and Lsr2 (B); (B) 1% agarose gel with sample load (L) and fractions from peak A and B; (C) 15% SDS-PAGE gel of the corresponding sample load (L) and fractions from peak A and B.

Due to the high salt content of the buffer used for SEC the native size of Lsr2 was not able to be accurately determined as the salt influences the mobility of the protein. The molecular weight of recombinant Lsr2 is predicted to be 13.2 kDa hence its expected dimeric form would be around 26 kDa. A calibration curve equation for the analytical SEC column predicted a size of 34 kDa for Lsr2 in 2 M NaCl buffer, whereas a calibration curve for the preparative SEC column predicted between 56 kDa to 127 kDa as the exact peak position of Lsr2 was variable between different preparations.

It was determined that Lsr2 unaccompanied by DNA was prone to binding to materials such as low protein binding PES membranes in the centrifugal concentrators used for preparing high concentrations of Lsr2 post SEC purification. Provided Lsr2 was in a buffer with a salt concentration of 1 M or above, it was able to be handled without fear of it adhering to surfaces. It therefore became routine to first concentrate Lsr2 in a 2 M NaCl buffer post SEC and then dialyse it into alternative buffers afterwards. Other modifications to the purification procedure included altering the overnight high salt incubation temperature and centrifugation temperature to 18°C to improve protein solubility and to reduce the amount of sample retained by the compulsory 0.2 μm filtration step prior to column loading.

Attempts to visualise Lsr2 devoid of DNA on native polyacrylamide gels by electrophoresis with electrodes in normal and reverse orientations and with altered buffer pH conditions as well as on native agarose gels were unsuccessful. The behaviour of Lsr2 without DNA had foiled previous attempts to discover favourable refolding conditions due to the inability to visualise the refolded protein by native gel electrophoresis and this again proved to be an issue and many experiments attempting to unite Lsr2 with a variety of designed DNA fragments were abandoned.

TEM was used to visualise how the high salt conditions used to remove the co-purified DNA would affect the structure of the fibrils seen previously with IMAC-purified Lsr2. Lsr2 samples in buffers containing increasing concentrations of salt were compared with a low salt sample using TEM (Figure 3.5). As the salt

concentration increased the fibrils became narrower in diameter and less entwined with each other (Figure 3.5B), appeared in much smaller twisted clumps (Figure 3.5C) and were very sparsely distributed, appearing as short lengths and without intertwining (Figure 3.5D). These changes in topology reflect the disruption of the electrostatic interactions between Lsr2 and the DNA.

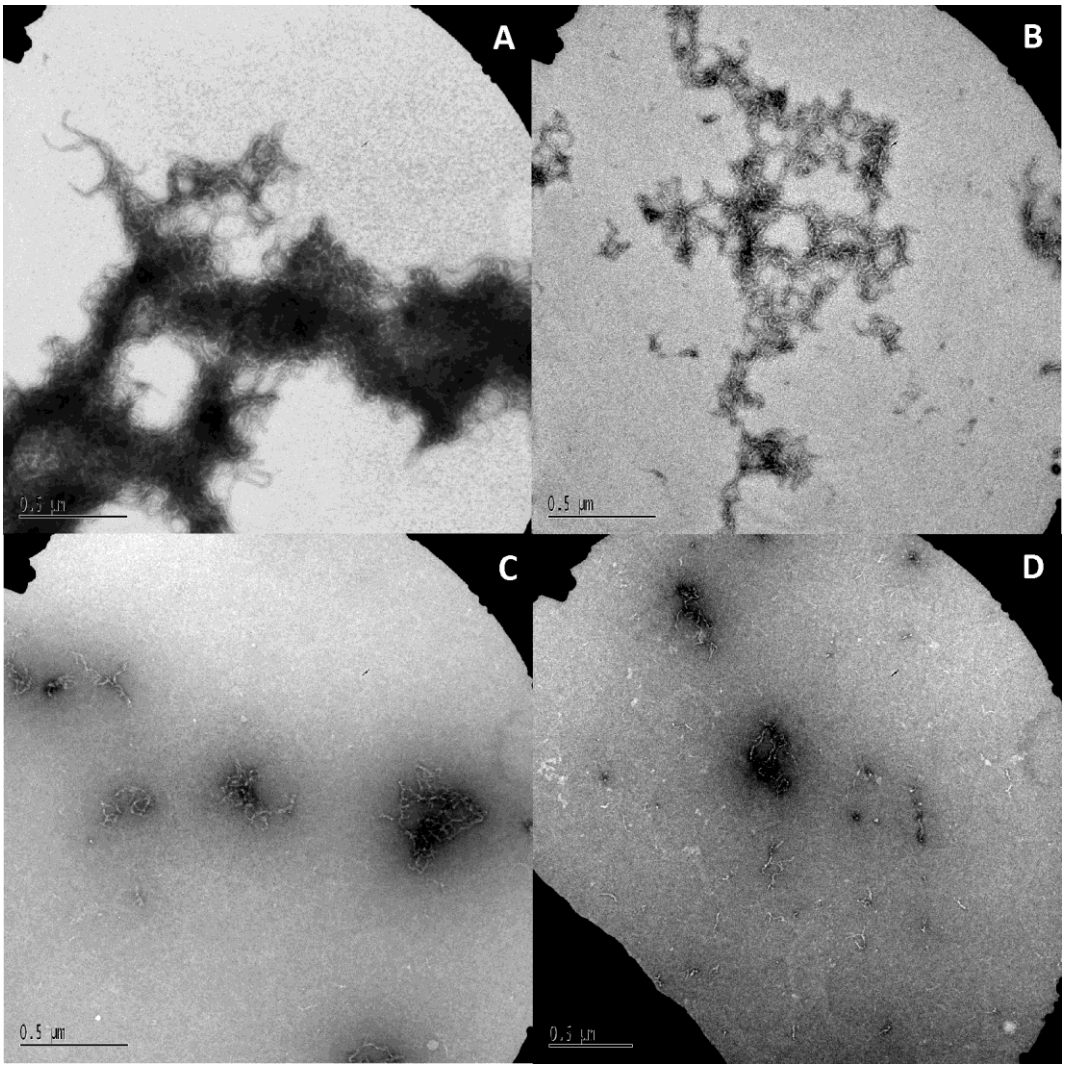


Figure 3.5. Electron micrographs of Lsr2/DNA complexes in high salt buffer reveal shorter and less entwined fibrils of protein coating DNA strands. (A) Negative stained IMAC purified Lsr2 in complex with DNA in a low salt buffer (200 mM NaCl); (B) in a 1.0 M NaCl buffer; (C) in a 1.5 M NaCl buffer; (D) in a 2.0 M NaCl buffer.

SEC purified Lsr2 was subjected to robotic crystallisation screens at concentrations of 0.1 mg/ml and 6 mg/ml. Crystals were observed in some conditions and hanging drop fine screens around the crystallisation conditions

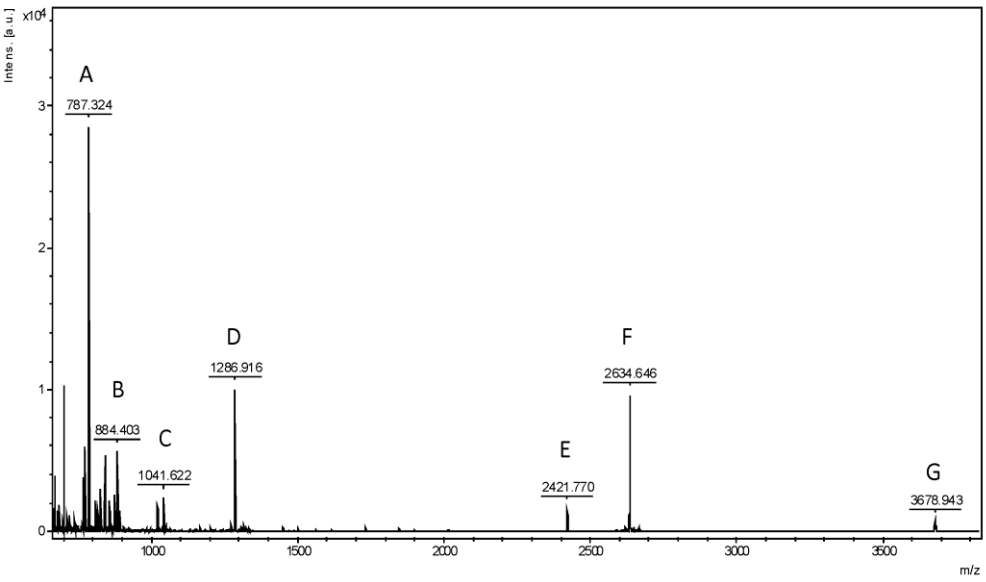
were prepared. Additive screens as well as crystal seeding was used in further fine screen attempts but all hopeful crystal candidates were revealed to be salt after analysis by X-ray diffraction.

After all attempts to completely remove contaminating genomic DNA from Lsr2, the presence of small fragments of DNA of around 100 base pairs in size was noted in a highly concentrated sample run on an agarose gel. This protein sample was the remnant of protein used in crystallography trials. This discovery highlights the tight association between Lsr2 and DNA and the difficulty in preparing DNA-free protein samples.

3.2.3 MALDI-TOF analysis confirms Lsr2 purification

Purified Lsr2 protein from both IMAC and SEC purifications was analysed by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) after in-gel trypsin digestion was used to generate peptide fragments of the protein. A number of factors prompted the use of MALDI-TOF MS to confirm the exact nature of the recombinant protein that was being studied. The protein ran at a larger mass on an SDS-PAGE gel than was predicted (15 kDa vs 13 kDa), it co-purified with large amounts of genomic *E. coli* DNA that was never reported by other researchers and attempts to replicate other researchers published DNA binding assays repeatedly failed. There was concern that the recombinant Lsr2 protein being produced and studied in our lab may have differed from that studied by other researchers.

Analysis of the peptide fragments produced by trypsin digestion of Lsr2 protein separated in an SDS-PAGE gel confirmed that recombinant Lsr2 from *M. tuberculosis* was successfully being produced and purified in the lab. Mass peaks on the spectrum with the greatest intensity matched the predicted masses for a trypsin digest of recombinant Lsr2 (Figure 3.6) with other less intense peaks also matching predictions (results not shown).



1-MAKKVTVTLVDDFDGSGAADETVEFGLDGVTYEIDLSTKNATKLRGDLKQWVAAGRRVGG-60
61-RRRGRSGSGRGRGAIDREQSAAIREWARRNGHNVSTRGRIPADVIDAYHAATLEHHHHHH-120

Peak	Amino Acids	Fragment
A	50-56	QWVAAGR
B	90-97	NGHNVSTR
C	89-97	RNGHNVSTR
D	73-84	GAIDREQSAAIR
E	100-120	IPADVIDAYHAATLEHHHHHH
F	98-120	GRIPADVIDAYHAATLEHHHHHH
G	5-39	VTVTLVDDFDGSGAADETVEFGLDGVTYEIDLSTK

Figure 3.6. Peptide mass analysis of trypsin digested Lsr2 using MALDI-TOF MS confirms purification of recombinant Lsr2. A comparison between the mass peaks on the collected spectrum (top) with a predicted trypsin digest of Lsr2 identified seven obvious peaks that match trypsin digest predictions (bottom). Mass accuracy was 300 ppm with one missed cleavage. Data was analysed by Bruker Data Analysis. The Lsr2 amino acid sequence is shown in the centre.

3.2.4 Lsr2 exchanges from co-purified genomic DNA to DNA oligos

A number of different approaches were used to examine the ability of Lsr2 to bind to DNA. Analysis of the change in mobility of DNA fragments during electrophoresis when bound with Lsr2 was attempted with a variety of double stranded DNA species including a DNA mass ladder, sequence-specific oligonucleotides, radiolabelled oligonucleotides and PCR products. Results were variable between trials and there was a lack of a convincing gel shift with these experiments which was in contrast to published gel shift experiments.

With limited evidence that Lsr2 could bind to DNA fragments but with the desire to have Lsr2 uniformly bound with short lengths of DNA for crystallisation attempts, further binding studies were performed. The rationale for these experiments was to bind Lsr2 to biotin-labelled, AT-rich DNA, and purify the complex by targeting the biotin label so that a uniform species of the complex could be used in crystallisation trials. Both SEC Lsr2 (without DNA) and IMAC purified Lsr2 successfully bound to short (16 base pair) biotin-labelled DNA oligonucleotides and were concentrated on streptavidin magnetic particles. These results showed that Lsr2 was capable of binding to very short DNA pieces and was able to exchange from co-purified genomic DNA to DNA oligonucleotides (Figure 3.7). An interesting observation from this experiment was the presence of multiple oligomeric forms of Lsr2 visible under the denaturing conditions of SDS-PAGE gel electrophoresis (Figure 3.7B).

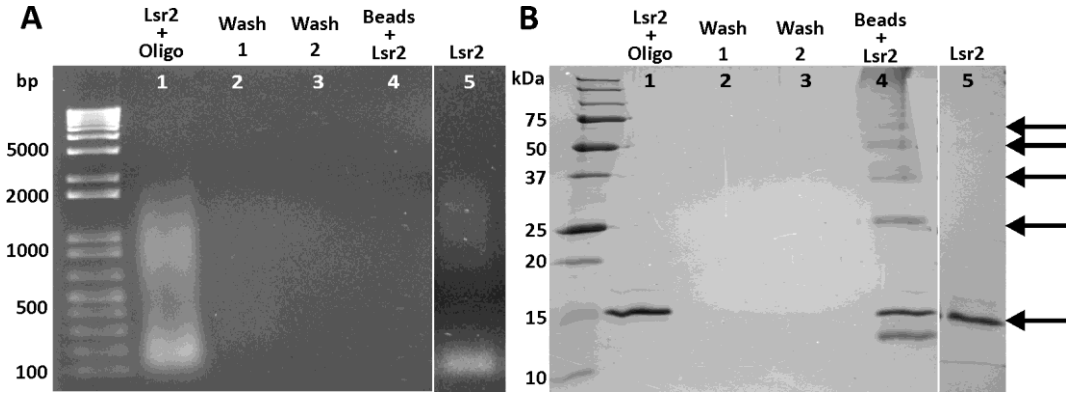


Figure 3.7. IMAC purified Lsr2 exchanges from co-purified genomic DNA to short biotinylated DNA oligonucleotides. (A) 1% agarose gel showing non-exchanged Lsr2 remains bound with genomic DNA in solution (lane 1) and protein attached to the streptavidin beads did not contain genomic DNA (lane 4). The staining represents genomic DNA not biotinylated oligonucleotide as it was too dilute to be seen; (B) 15% SDS-PAGE analysis reveals that not all Lsr2 exchanged on to the DNA oligos (lane 1) but protein that did exchange was tightly bound as it was not dislodged by washing of the streptavidin beads (lanes 2 and 3). Lane 5 of both gels represents IMAC purified Lsr2 co-purified with genomic DNA. Arrows point to the progressive oligomerisation of Lsr2 seen (monomer, dimer, trimer, tetramer and pentamer).

Removal of the streptavidin particles was necessary to use the Lsr2/oligo complex for crystallography. A number of approaches were used. The use of a buffer containing biotin to compete with the streptavidin particles and elute the

bound complex had little effect other than reducing the amount of protein oligomerisation observed on the SDS-PAGE gel. "Softlink Soft Release Avidin Resin" was used for binding the Lsr2/oligo complex as it was claimed that biotin would reversibly bind to these particles, however these claims were proven false in this case. The particles could not be separated using the procedure recommended by the manufacturer. Finally, the biotinlyated oligonucleoides were redesigned, adding additional nucleotides between the end of complementary strand and the biotin label to allow cleavage by DNase. The new biotin-labelled DNA oligonucleotides successfully bound IMAC-purified Lsr2 and DNase digestion was able to remove the complex from the streptavidin particles. However, whilst fine-tuning the DNase digestion step a flaw in this approach was revealed as DNase treatment was affecting protein integrity and consequently a uniform Lsr2/DNA complex could not be attained for use in crystallography trials.

3.2.5 Evidence of Lsr2 protein oligomerisation

Crosslinking of proteins using glutaraldehyde is often used to obtain initial information about the larger assembly of protein subunits. In this study large oligomeric structures of Lsr2 (in complex with co-purified genomic DNA) as well as dimers and trimers were detected by SDS-PAGE after a two minute incubation with glutaraldehyde (Figure 3.8A). Two dimeric Lsr2 species were observed which may correspond to dimers with and without the C-terminal histidine tag. Recombinant Lsr2 expressed with a maltose binding protein tag (section 3.2.6) (in complex with co-purified genomic DNA) is also capable of forming large oligomeric structures after crosslinking with glutaraldehyde (Figure 3.8B). This simulated large complex formation, in combination with the *in vitro* complex formation observed with the binding of Lsr2 to biotinylated DNA oligos (Figure 3.7) shows the potential for a dynamic interaction between Lsr2 and DNA.

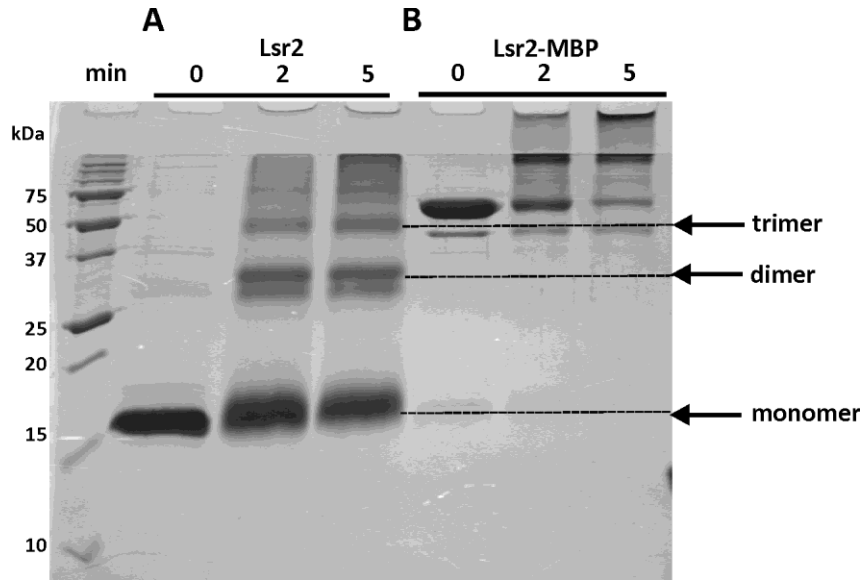


Figure 3.8. Progressive Lsr2 oligomerisation in the presence of co-purified genomic DNA. (A) IMAC purified Lsr2 forms dimers, trimers (arrows) and larger oligomeric structures when crosslinked with 0.1% glutaraldehyde and examined using 18% SDS-PAGE; (B) affinity purified Lsr2 with a MBP tag forms large oligomeric structures when crosslinked with glutaraldehyde that are unable to migrate through the stacking gel layer.

Evidence that Lsr2 exists in the bacterial cell in units larger than a monomer was revealed through Western blotting. Antibodies to *M. tuberculosis* H37Rv Lsr2 were generated at the AgResearch antibody facility, Ruakura, Hamilton. IMAC-purified Lsr2 (1 mg/ml) was used to immunise white rabbits by subcutaneous injection. Primary immunisation was performed with a solution containing 0.2 ml of Lsr2, followed by injections every two weeks for six weeks with a solution containing 0.1 ml of Lsr2. Polyclonal antibodies contained within the serum of the blood from immunised rabbits were used for Western blotting.

Whole cell lysates of *M. smegmatis* mc^2155 and *M. bovis* BCG separated by SDS-PAGE and incubated with antibodies generated to *E. coli* recombinant Lsr2 show a number of bands that correspond to the molecular weight of a dimer, trimer and oligomer (Figure 3.9). The cell lysate protein concentration of *M. bovis* BCG was higher than that of *M. smegmatis* mc^2155 and hence the intensity of the bands is greater revealing a smaller band that could represent a possible Lsr2 monomer. The band representing the monomer in *M. bovis* BCG differs in molecular weight when compared to recombinant Lsr2 expressed in *E. coli*

(Figure 3.9, left panel). Based on the intensity of all the bands in both bacterial cell lysates, it is speculated that Lsr2 is predominantly a dimer in the cell and that the protein subunits are tightly associated as they do not separate on a denaturing gel.

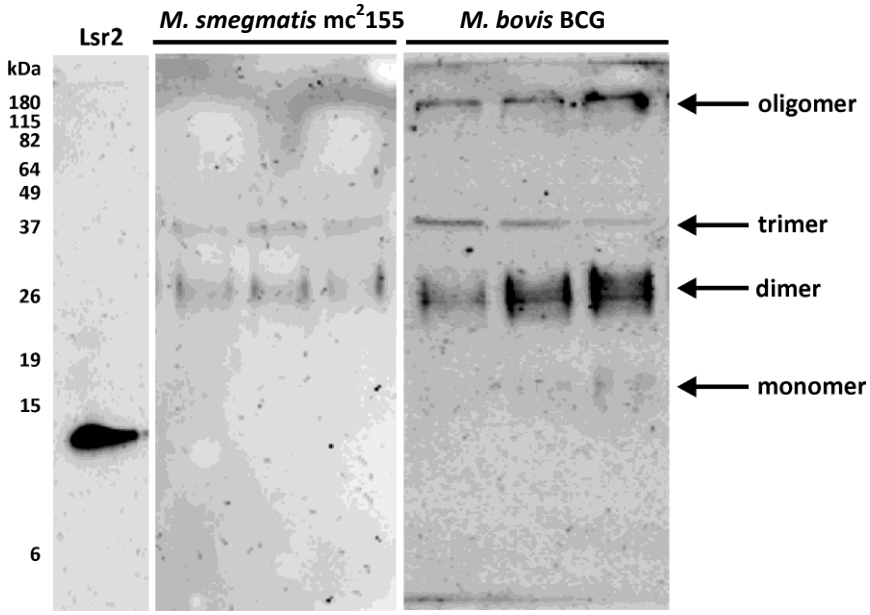


Figure 3.9. Western blot analysis of *M. smegmatis* mc²155 and *M. bovis* BCG whole cell lysates reveals multiple oligomeric forms of Lsr2. Recombinant monomer Lsr2 (left) is shown alongside triplicate samples of whole cell lysate (separated on 15% SDS-PAGE gels) from *M. smegmatis* mc²155 (centre) and *M. bovis* BCG (right). Antibodies to recombinant Lsr2 are cross-reactive in other species and detect monomer, dimer, trimer and larger oligomeric forms in both species (arrows).

The three levels of experimentation (simulated, *in vitro* and *in vivo*) presented here provide evidence that Lsr2 is capable of forming large, tightly associating structures in complex with DNA.

3.2.6 Expression of Lsr2 with alternative tag improves protein yield

Protein solubility can be improved by expression as a fusion protein with maltose binding protein (MBP). This fusion protein can be easily purified due to its affinity for amylose resin and the MBP domain can be removed using Factor Xa protease. An improvement in Lsr2 solubility was sought for further

crystallisation attempts and this expression system gave the choice to remove the fusion tag which was not an option with recombinant protein expression using pET30b. The *lsr2* gene was amplified by PCR from the *lsr2*-pET30b plasmid and cloned into the cytoplasmic expression vector pMAL-c2X using the *XbaI*/*PstI* restriction sites, positioning an MBP tag at the N-terminus of the protein. DNA sequencing of amplified and purified plasmid confirmed the insertion of *lsr2* within the appropriate restriction sites and the positioning of the MBP tag which would result in the expression of an approximately 56 kDa recombinant protein.

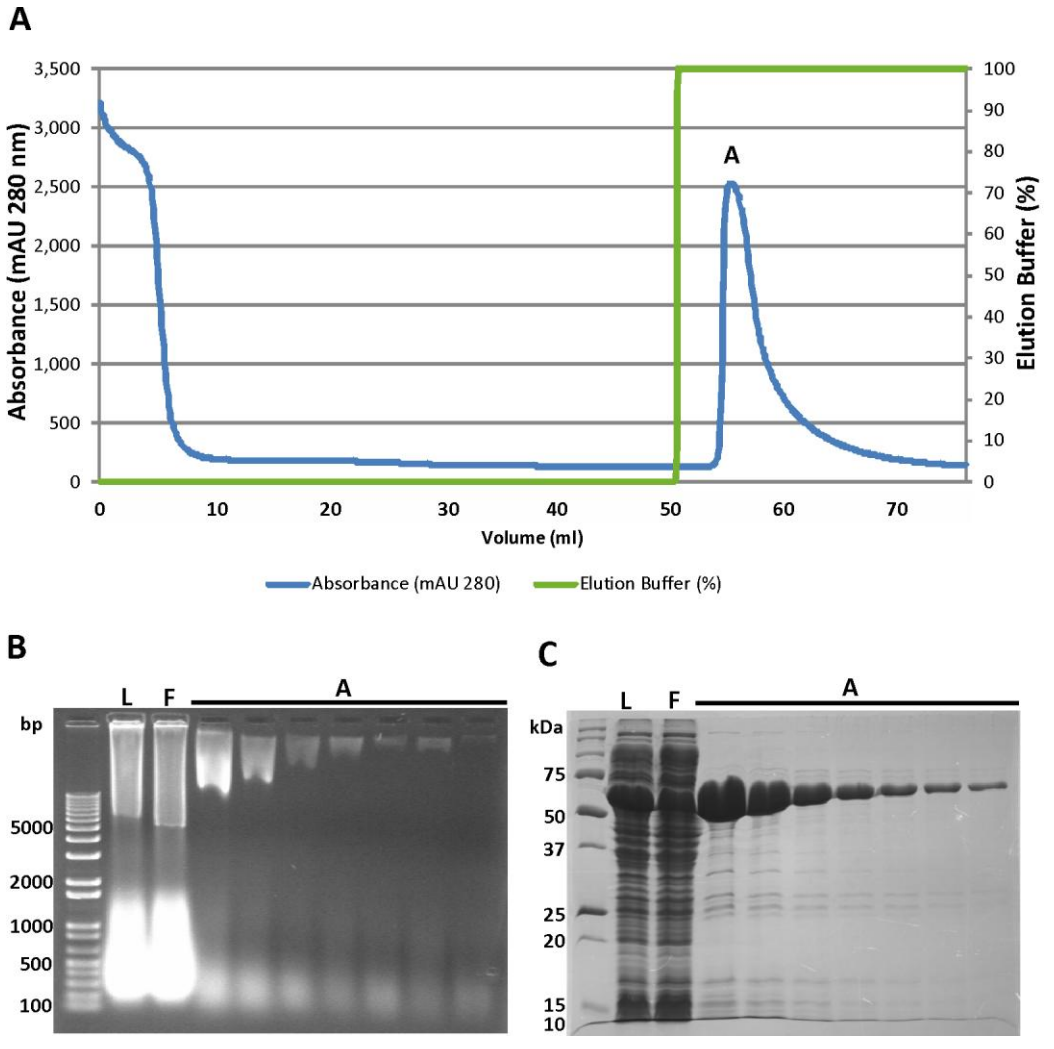


Figure 3.10. MBP tagged Lsr2 co-purifies with genomic DNA during amylase resin purification. (A) Chromatogram of the elution buffer gradient and UV absorbance profile; (B) 1% agarose gel with sample lysate (L), column flow through (F), and fractions from peak A; (C) 15% SDS-PAGE gel of the corresponding lysate, flow through and fractions from peak A.

Small scale expression testing, MBP binding assays and large scale expression of Lsr2 fused with MBP resulted in soluble protein that was purified using amylose or dextrin resin chromatography. The chromatogram and corresponding agarose and SDS-PAGE gels (Figure 3.10) illustrate the tight binding of Lsr2-MBP to the column, and easy removal with elution buffer. Gel analysis of the peak fractions reveal that Lsr2-MBP also co-purifies with genomic DNA despite MBP being fused to the N-terminal end of the protein which indicates that its DNA binding ability has not been compromised. Amylose resin purified Lsr2-MBP showed the same ability as IMAC purified recombinant Lsr2 to form large complexes after chemical crosslinking with 0.1% glutaraldehyde (Figure 3.8B).

The amylose resin purified Lsr2-MBP/DNA complex was efficiently separated by SEC in buffer containing 2 M NaCl (results not shown) as was routinely used for the purification of recombinant Lsr2 and DNA complexes (section 3.2.2).

TEM was used as a technique to visualise the complex of Lsr2-MBP bound with DNA. Micrograph images show an arrangement of strands that differ to the "fibrils" seen previously with Lsr2/DNA complexes. Although Lsr2-MBP is able to bind DNA in a non-specific manner, and coat the DNA, the arrangement of the protein on the DNA is different. Complexes of Lsr2-MBP and DNA have a knotted or "brain-like" twisted appearance and do not form a mesh-work of intertwined fibrils as seen in micrographs of Lsr2/DNA complexes (Figure 3.11). It is hypothesised that the presence of the large 42 kDa MBP at the N-terminal end of Lsr2 (13 kDa) is affecting the mechanism of DNA binding.

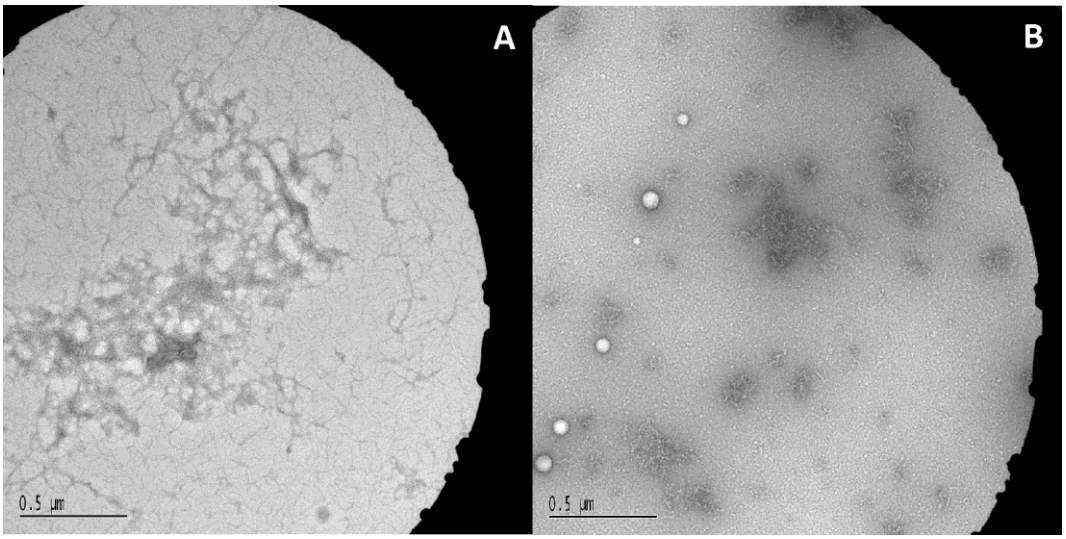


Figure 3.11. Electron micrographs of recombinant Lsr2-MBP/DNA complexes show lack of fibril-like structure and no intertwining compared with recombinant Lsr2/DNA complexes. (A) Negative stained IMAC purified Lsr2 in complex with co-purified genomic DNA; (B) amylose resin purified Lsr2-MBP in complex with co-purified genomic DNA.

Due to the behaviour of Lsr2 without co-purified DNA it was difficult to obtain a concentration suitable for successful crystallisation. By expressing Lsr2 linked to MBP, a much higher concentration was achieved and crystallisation trials were performed without the removal of the MBP. SEC purified Lsr2-MBP protein was subjected to robotic crystallisation screens at a concentration of 60 mg/ml. No crystals were observed in these screens or fine screens around the promising crystallisation conditions.

3.2.7 Alteration of positively charged residues is detrimental to *E. coli*

Five lysine residues present in Lsr2 contribute greatly to the overall basic nature of the protein. At this time, the domain structure of Lsr2 and the residues involved in DNA binding were unknown and so it was hypothesised that altering the lysine residues would reduce the overall positive charge of the protein and influence the ability of Lsr2 to bind to DNA, which in turn could improve the purification and behaviour of this protein. It was previously observed that expression of Lsr2 in *E. coli* slowed the bacteria's growth rate considerably. Perhaps altering the lysine residues would alleviate the influence Lsr2 may have on the growth rate of *E. coli*. Initially, a chemical methylation method was used

to alter the lysine residues but this method gave variable protein modification results and instead genes containing point mutations to alter each lysine to an alanine residue were manufactured (Geneart).

Three lysine to alanine mutants were designed. The first mutant comprised the two lysine residues K3 and K4 at the N-terminal end of Lsr2 (2K2A), the second included the three lysine residues K39, K43 and K49 clustered in the centre of Lsr2 (3K3A) and the third encompassed all five lysine residues (5K5A). All lysine residues are in the N-terminal dimerisation domain of Lsr2. The manufactured genes were cloned into an *E .coli* expression vector pDEST17, positioning a 6 x histidine tag (lacking a protease cleavage site) at the N-terminal end of Lsr2 which produced a 16 kDa recombinant protein.

Small scale expression testing with histidine tag binding assays were hampered by the mixed ability of the three transformed mutant strains of *E. coli* to successfully grow prior to the induction of protein expression. Mutant strain growth was attempted at three different temperatures and mutant strains were refreshed by repeating plasmid transformations but both these approaches did not improve cell growth. This change in growth of *E. coli* suggested that the pDEST17 vector had "leaky" expression of the Lsr2 mutants that this was influencing cell health. Poor cell growth was associated with cell clumping and flocculation in the liquid culture media. Smears of bacterial cells obtained from poorly-growing cultures of all mutants were stained using fluorescent antibodies to Lsr2 (Figure 3.12B-D) and an example was Gram stained (Figure 3.12F). Gram stained *E. coli* harbouring the 2K2A expression vector showed remarkable long chains of rods that have failed to dissociate after cell division.

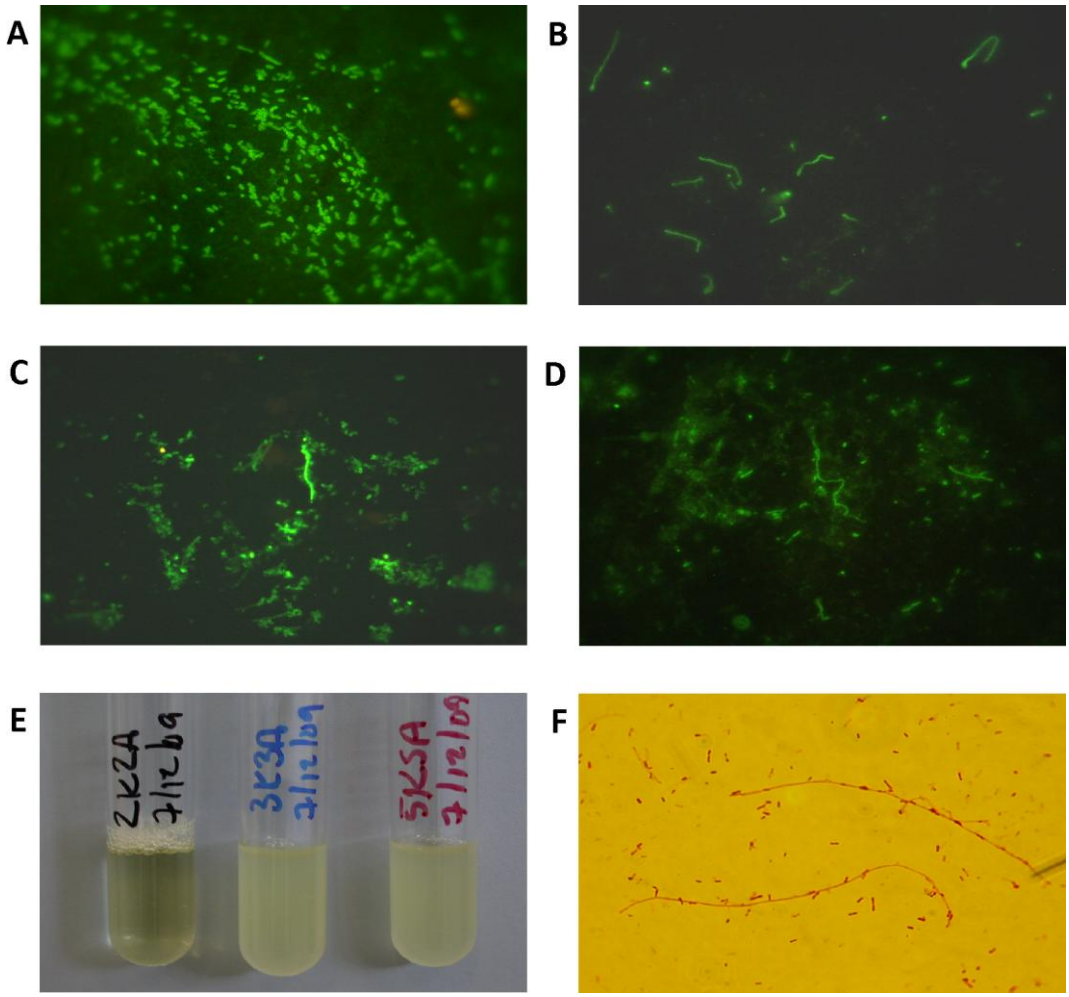


Figure 3.12. ***E. coli*** **strains containing lysine-alanine Lsr2 mutant expression vectors have impaired growth, forming long chains of cells compared with** ***E. coli*** **containing the native Lsr2 expression vector.** Fluorescent antibodies to Lsr2 illuminate the entire cell of *E. coli* harbouring the native recombinant Lsr2 expression vector (A), the 2K2A Lsr2 mutant expression vector (B), the 3K3A Lsr2 mutant expression vector (C) or the 5K5A Lsr2 mutant expression vector (D). All cell samples were from un-induced cultures. An example where the 2K2A un-induced mutant had flocculated growth which settled to the bottom (E) and the long chains of cells observed with a simple Gram stain (F). Images were captured at 100x under oil immersion on a light microscope.

A number of small scale expression tests using a variety of expression temperatures and with a number of lysis buffers were performed. Results from the small scale expression tests with the three mutants, revealed that all three recombinant proteins were detected in both the insoluble and soluble cell fractions and were present as a monomer and dimer (Figure 3.13). Large scale expression cultures of the three lysine to alanine mutants were purified using IMAC chromatography but were not successful in producing mutant Lsr2 protein (results not shown).

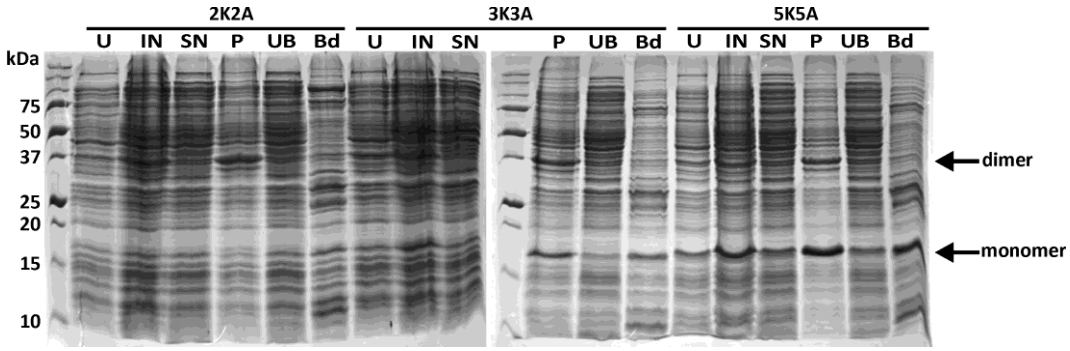


Figure 3.13. Small scale histidine tag binding assays for Lsr2 lysine mutants showed soluble and insoluble protein in both monomer and dimer forms. An example of a small scale expression test performed at 25°C. The lysine mutant 2K2A was expressed as a dimer and was present in the insoluble cell pellet only. Mutant 3K3A was expressed as both a monomer and dimer in both the soluble and insoluble fractions and the monomer was enriched on the purification beads. The mutant 5K5A was expressed as both a monomer and dimer in both the soluble and insoluble fractions with the monomer enriched on the purification beads. U = uninduced, IN = induced, SN = supernatant (soluble fraction). P = pellet (insoluble fraction), UB = unbound and Bd = bead fraction.

3.3 Discussion

The DNA binding protein Lsr2 from *M. tuberculosis* H37Rv was successfully expressed at high levels in *E. coli* however a trait, not mentioned in previous published research, of co-purifying with large quantities of genomic DNA required an additional purification step in a high salt buffer to dissociate the protein and DNA. The ability of Lsr2 to bind large amounts of genomic DNA illustrates the nonspecific DNA binding nature of this protein. Negative stained TEM images of Lsr2/DNA complexes shows large "spaghetti-like" structures with protein coating the DNA strands, entwining lengths of DNA together and crosslinking neighbouring "fibrils" confirming a nonspecific association with large quantities of DNA.

The handling of Lsr2 devoid of DNA presented further challenges as it was prone to binding to low protein binding membranes during centrifugal concentration and would aggregate at low temperatures and high concentrations. DNase treatment of the Lsr2/DNA complex was not efficient enough to be a beneficial step in DNA removal during the purification process and it was also challenging to develop a technique to exchange Lsr2 onto short designed lengths of DNA to obtain a uniform sample for protein crystallisation studies. It is hypothesised that Lsr2 is able to protect DNA from DNase digestion by extensive coating of the DNA hence explaining the limited ability of this enzyme to remove the DNA.

Protein crystallisation attempts using Lsr2 without DNA as well as using an Lsr2-MBP fusion protein to increase the protein yield were unsuccessful. It is crucial that a protein sample is uniform in its composition for crystallisation to be successful and it was discovered that after all attempts to remove co-purified genomic DNA from Lsr2, there still remained a small molecular weight pool of DNA bound to the protein which was likely to have hindered attempts to grow crystals.

Lsr2 is thought to express as a dimer however it was difficult to confirm this with the recombinant protein produced. Size exclusion chromatography may be used to measure the approximate mass of a protein through the generation of a

calibration curve and subsequent equation using protein calibration standards. However in this case due to the fact that Lsr2 was always purified using SEC in a high salt buffer, the mass was not able to be predicted. Lsr2 appeared to purify as a dimer or as other larger complexes in high salt conditions which suggests it has strongly interacting residues that enable the formation of oligomers.

The modification of lysine residues in an attempt to alter the DNA binding ability of Lsr2 had a dramatic affect on *E. coli* cell growth both prior to and during the attempted expression of the recombinant proteins. Poor cell growth was accompanied by inconsistent and insoluble protein expression. These traits are often the result of protein toxicity. Nevertheless, upon successful expression, Lsr2 mutants were observed in a dimeric form.

The formation of higher order structures of Lsr2 was seen both *in vitro* and *in vivo*. Evidence that Lsr2 naturally forms larger protein structures in complex with DNA was gathered through western blot analysis of *M. smegmatis* and *M. bovis* BCG cell lysates. Lsr2 appears predominantly as a dimer, but also as a trimer and larger complexes under denaturing gel conditions which reinforces the tight association between key residues in Lsr2.

Chapter Four – DNA binding for three Lsr2 mutants

4.1 Introduction

During the expression, purification and characterisation of Lsr2 (in Chapter Three), the tight association between Lsr2 and *E. coli* genomic DNA was noted. Consistent co-purification of Lsr2 with genomic DNA during recombinant protein purification has only been noted in our work (Colangeli *et al.*, 2009) and it creates technical challenges when working with this protein. DNA removal using DNase was only partially successful however the use of a high salt buffer enabled DNA dissociation on a large scale.

Molecular methods to reduce the DNA binding ability of Lsr2 were employed to potentially provide DNA-free purified protein. In the early phase of this study, the N-terminal dimerisation domain and C-terminal DNA binding domain arrangement of Lsr2 and the residues involved in DNA binding were unknown. So a strategy of mutating the five positively charged lysine residues of Lsr2 was performed to change the overall charge of Lsr2 in an attempt (albeit unsuccessful) to alter the DNA binding properties and to enable effortless purification (Chapter Three). A second approach to alter the DNA binding ability of Lsr2 would be to mutate arginine residues involved with DNA binding.

This chapter focuses on the generation of Lsr2 mutants that have an altered ability to bind DNA. The rationale for the production of DNA binding mutants was to achieve the purification of DNA-free protein that was homogenous and suitable for protein crystallisation studies.

Lsr2 mutant design was made possible by the previous elucidation of the 3D structure of the C-terminal DNA binding domain of Lsr2 (Gordon *et al.*, 2010). Earlier analysis of the C-terminal DNA binding domain structure of Lsr2 (using nuclear magnetic resonance (NMR) techniques) identified residues that were central to DNA binding and enabled a model of DNA binding to be proposed (Gordon *et al.*, 2010). The DNA binding domain consists of two α-helices, which

are perpendicular to each other and interact via hydrophobic residues. NMR titration experiments using a short piece of DNA revealed potential DNA binding interfaces and modelling of Lsr2 in complex with DNA showed a two-component mechanism of DNA binding, with residues interacting in the minor groove and also along the edge of the major groove. Residues R97-G98-R99 in the loop insert into the minor groove of the DNA with the two arginine side chains pointing away from each other whilst outside the major groove, side chains R77, S80, R84 and S95 interact with the sugar-phosphate backbone on either edge of the minor groove (Gordon *et al.*, 2010). The residues R84, R97 and R99 are all highly conserved among Lsr2 homologues and are proposed to be important for DNA binding and it was for these reasons that they were selected as candidates for mutagenesis and DNA binding studies (Figure 4.1).

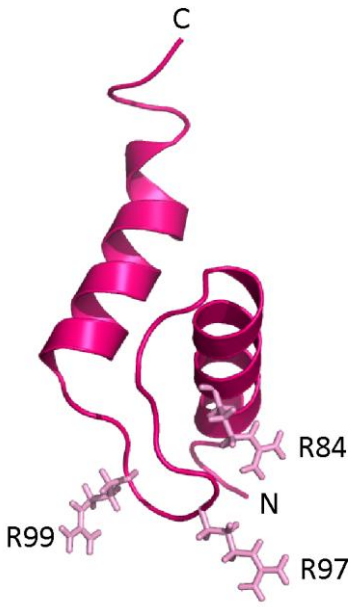


Figure 4.1. The highly conserved residues R84, R97 and R99 in the DNA binding domain of Lsr2 are proposed to interact with DNA and are candidates for mutagenesis studies. A cartoon representation of the NMR solution structure of the C-terminal DNA binding domain of Lsr2 (PDB; 2KNG) showing residues as sticks. The chain termini and residues are labelled and the protein depiction was rendered with PYMOL.

To characterise and assess the function of the mutant forms of Lsr2 a number of biophysical methods can be employed such as thermal stability assays, dynamic light scattering and circular dichroism (CD). A fluorescence-based thermal stability assay originally used for buffer optimisation and ligand-induced stabilisation of proteins can be used to assess the thermally induced melting point of proteins (Ericsson *et al.*, 2006). The theory behind this technique is that folded and unfolded proteins can be identified by their interaction with a

hydrophobic fluoroprobe. In solution the probe is quenched but will bind to the unfolding hydrophobic interior of a protein, leading to a reduction in quenching and the detection of fluorescence emission that can be studied as a function of temperature. The thermally induced unfolding process follows a two-state model of protein unfolding kinetics with a swift transition between the folded and unfolded state of a protein and the melting temperature (Tm) defined as the midpoint temperature of this transition. Some studies have shown that melting temperatures obtained with the thermally induced unfolding method correlate with temperatures obtained by other biophysical methods such as CD and differential scanning calorimetry (Ericsson *et al.*, 2006). Through protein melting temperature changes, this thermal stability assay may reveal the effect of mutations of the DNA binding residues of Lsr2 and the effect on protein stability.

CD spectroscopy is a useful technique for the analysis of the secondary structure of proteins. CD measures the difference in the absorption of the two components of polarised light – left-handed polarised light and right-handed polarised light – that arise due to structural asymmetry. The asymmetric carbon atoms present in the amino acid residues results in the protein displaying optical activity (Carpenter & Geoff Kneale, 1994). A disordered structure will show zero CD intensity whereas an ordered structure will give a spectrum which can contain both negative and positive signals. Using CD it is easy to observe changes in the secondary structure of proteins that may result from changes in pH, temperature, and ligand binding. Proteins that differ in their composition of α-helices, anti-parallel and parallel β-sheets will have CD spectra that show significant differences in shape and size (Johnson, 1988).

CD spectroscopy in the "far-UV" spectral region (190-250 nm) allows secondary structure determination whereas CD spectroscopy in the "near-UV" spectral region (250-350 nm) can provide information about protein tertiary structure. Analysis of the spectrum of a protein for secondary structure may be done by comparison to reference CD spectra for pure secondary structures as shown in Figure 4.2 (Greenfield, 2007), by the use of online modelling (Cary & Kneale,

2008), or by using reference spectra derived from collected spectra of proteins with known secondary structure (Johnson, 1988).

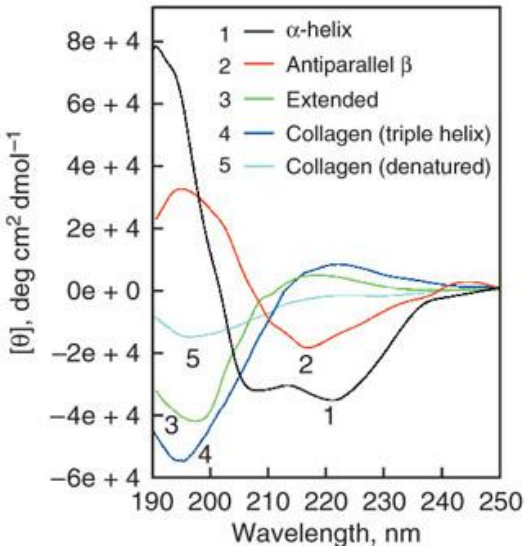


Figure 4.2 The CD spectra in the "far-UV" spectral region of pure poly-L-lysine secondary structures and placental collagen show distinct characteristic shape and size. Uniquely recognisable pieces of spectra from reference proteins can be compared against the spectrum from a tested protein to predict the proportion of each structural motif present. Poly-L-lysine α-helix and anti-parallel β-sheet conformations at pH 11.1 (1 and 2, respectively) and poly-L-lysine extended conformations at pH 5.7 (3) and placental collagen in its native triple-helix and denatured forms (4 and 5, respectively) are shown. This figure is from (Greenfield, 2007).

Work presented in this chapter demonstrates the successful production and expression of three Lsr2 mutants that show deficiencies in DNA binding. Protein stability and secondary structure of these mutants differ from native Lsr2. A double mutant with the least ability to bind DNA was a candidate for protein crystallisation attempts but poor protein homogeneity and solubility reduced the probability of yielding crystals.

4.2 Results

4.2.1 Generation of Lsr2 DNA binding mutants

Primer design was completed manually for the generation of four arginine to alanine residue mutations in Lsr2 using site directed mutagenesis. These included single residue mutations for the residues R84, R97 and R99 and a double residue mutation for both R97 and R99 residues. The recombinant Lsr2 expression plasmid *lsr2*-pET30b was used as the template for mutagenesis experiments. The double mutant R97A-R99A was successfully generated in the first round of site directed mutagenesis whilst repeat attempts at the two single mutations failed. Primers for the remaining single residue mutations were redesigned based on the double mutant primer length and composition, and further rounds of site directed mutagenesis were performed. The mutation R99A was successfully generated in the first round of site directed mutagenesis using the new primers with the R97A mutation following suit in the second round. Generation of a R84A mutant was unsuccessful. Successful residue conversion in each mutant was confirmed by DNA sequencing.

4.2.2 Expression and purification of the Lsr2 mutant R97A

Lsr2 R97A mutant was expressed in *E. coli* as a 13.2 kDa recombinant protein fused to a C-terminal 6 x histidine tag (without a tag cleavage site). Small scale expression testing utilising a histidine tag binding assay identified a protein of the correct size that was successfully captured on nickel sepharose beads (results not shown). *E. coli* growth to an optical density suitable for recombinant protein induction was accomplished in only 2 hours compared with 6 hours for native Lsr2. Large scale purification of R97A Lsr2 was performed using IMAC. The chromatogram and corresponding SDS-PAGE and agarose gels (Figure 4.3) demonstrate the two peaks that result from IMAC purification of R97A Lsr2 with the protein of interest appearing in peak B. Proteins bound non-specifically to the column were eluted during column equilibration with binding buffer and those with moderate interaction with the column were removed at around 40 %

elution buffer, whereas R97A was eluted at around 70% of elution buffer which corresponds to 700 mM imidazole (Figure 4.3A). There is a small amount of genomic DNA co-purified with R97A (Figure 4.3B). R97A Lsr2 mutant resolves as a 15 kDa recombinant protein on a 15 % SDS-PAGE gel (Figure 4.3C).

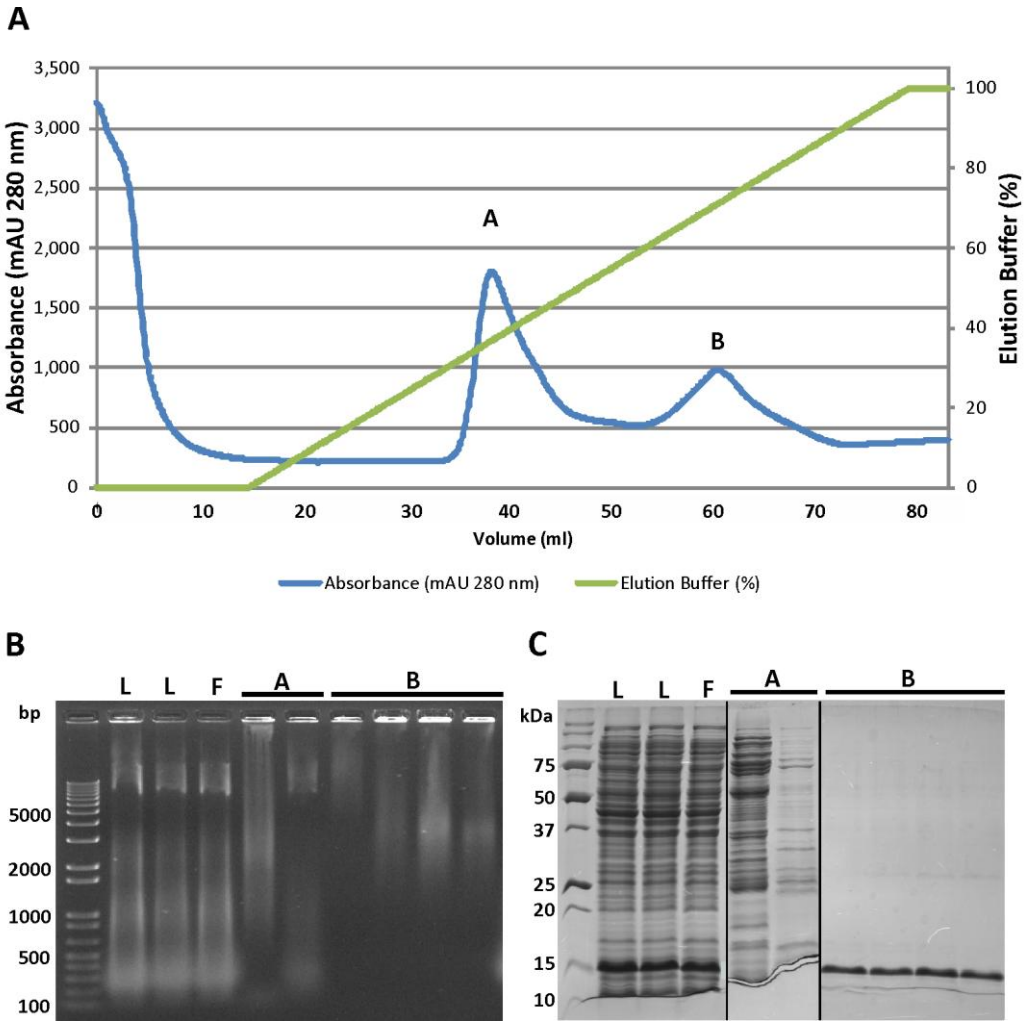


Figure 4.3. R97A Lsr2 mutant co-purifies with a reduced amount of genomic DNA during IMAC purification. (A) Chromatogram of the elution buffer gradient and UV absorbance profile; (B) 1% agarose gel with sample lysate (L), column flow through (F) and fractions from peak A and B; (C) 15% SDS-PAGE gel of the corresponding lysate, flow through and fractions from peak A and B.

4.2.3 Expression and purification of the Lsr2 mutant R99A

The Lsr2 R99A mutant was expressed in *E. coli* as a 13.2 kDa recombinant protein fused to a C-terminal 6 x histidine tag (without a tag cleavage site). Small scale expression testing was omitted in favour of large scale expression. *E. coli* growth to an optical density suitable for recombinant protein induction was accomplished in only 1.5 hours compared with 6 hours for native Lsr2. The chromatogram and corresponding SDS-PAGE and agarose gels from the large scale IMAC purification of R99A (Figure 4.4) show protein spread across two merged peaks with the protein of interest appearing in region C. Proteins bound non-specifically to the column are eluted during column equilibration with binding buffer and those with moderate interaction with the column are removed between 10 - 30 % elution buffer, whereas R99A is eluted between 30 - 50 % of elution buffer which corresponds to 300 - 500 mM imidazole (Figure 4.4A). There is a small amount of genomic DNA co-purified with R99A (Figure 4.4B). R99A Lsr2 mutant resolves as a 15 kDa recombinant protein on a 15 % SDS-PAGE gel (Figure 4.4C).

4.2.4 Expression and purification of the Lsr2 mutant R97A-R99A

The Lsr2 double mutant R97A-R99A was expressed in *E. coli* as a 13.2 kDa recombinant protein fused to a C-terminal 6 x histidine tag (without a tag cleavage site). Small scale expression testing was omitted in favour of large scale expression. *E. coli* growth to an optical density suitable for recombinant protein induction was accomplished in only 2-3 hours compared with 6 hours for native Lsr2. The chromatogram and corresponding SDS-PAGE and agarose gels from the large scale IMAC purification of R97A-R99A Lsr2 (Figure 4.5) shows two peaks with the protein of interest appearing in peak B. Proteins bound non-specifically to the column are eluted during column equilibration with binding buffer and those with moderate interaction with the column are removed at around 15 % elution buffer, whereas R97A-R99A is eluted around 50% of elution buffer which corresponds to 500 mM imidazole (Figure 4.5A). In contrast to native Lsr2 IMAC purification, there is a negligible amount of genomic DNA co-purified with R97A-

R99A that is barely visible on an agarose gel (Figure 4.5B). R97A-R99A Lsr2 mutant resolves as a 15 kDa recombinant protein on a 15 % SDS-PAGE gel (Figure 4.5C). Interestingly, the SDS-PAGE gel reveals the presence of protein bands in fractions from peak A that correspond to the molecular weight of possible dimer and hexamer forms of a protein the size of Lsr2.

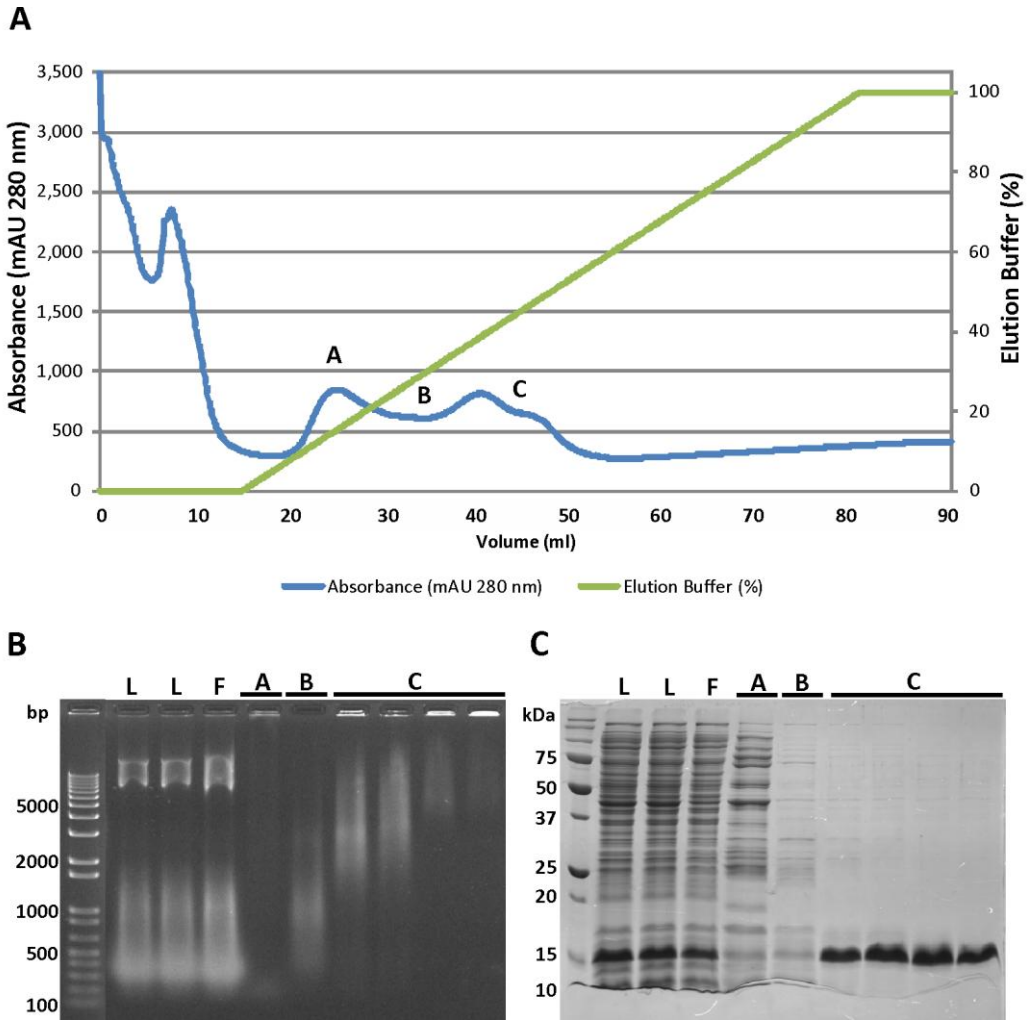


Figure 4.4. R99A Lsr2 mutant co-purifies with a reduced amount of genomic DNA during IMAC purification. (A) Chromatogram of the elution buffer gradient and UV absorbance profile; (B) 1% agarose gel with sample lysate (L), column flow through (F) and fractions from regions A, B and C; (C) 15% SDS-PAGE gel of the corresponding lysate, flow through and fractions from regions A, B and C.

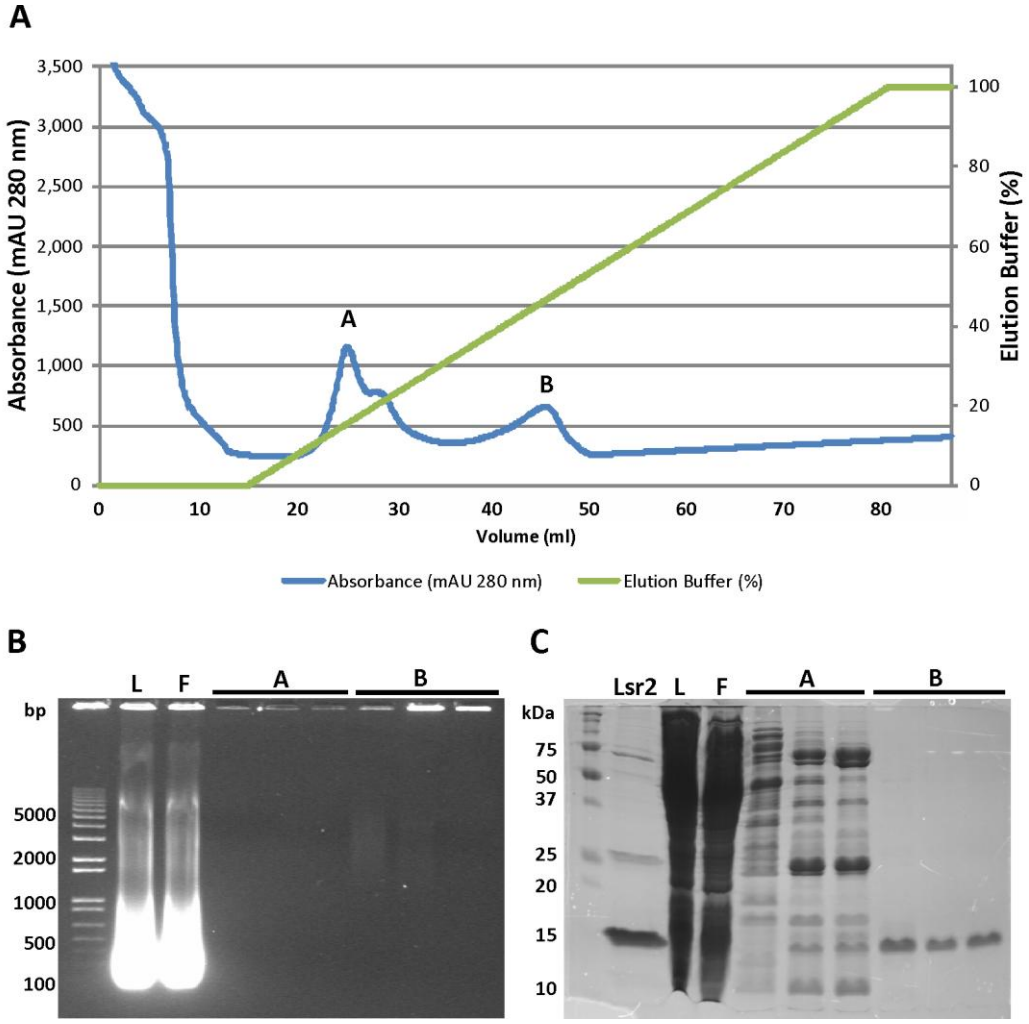


Figure 4.5. The Lsr2 double mutant R97A-R99A barely co-purifies with genomic DNA during IMAC purification. (A) Chromatogram of the elution buffer gradient and UV absorbance profile during IMAC purification; (B) 1% agarose gel with sample lysate (L), column flow through (F) and fractions from peak A and B; (C) 15% SDS-PAGE gel of the corresponding lysate, flow through and fractions from peak A and B and a control native Lsr2 sample for comparison.

4.2.5 Lsr2 DNA binding mutants have reduced DNA binding ability

It was apparent from the initial IMAC protein purification procedure that each residue mutation in Lsr2 had affected the ability of the protein to bind to DNA. A comparison of the amount of co-purified DNA between each mutant and native Lsr2 by agarose gel analysis visibly shows this (Figure 4.6). A bright stained DNA band represents the amount of co-purified DNA from a native Lsr2 purification in comparison to a slight smear of DNA for the single R97A and R99A mutants and a barely visible dusting of DNA for the double R97A-R99A Lsr2 mutant. Practically it was also noted that bacterial cell lysis using sonication was quicker with Lsr2

mutants and loading of the cell lysate onto the column was easier and resulted in less column compaction than typically occurred with native Lsr2 lysate.

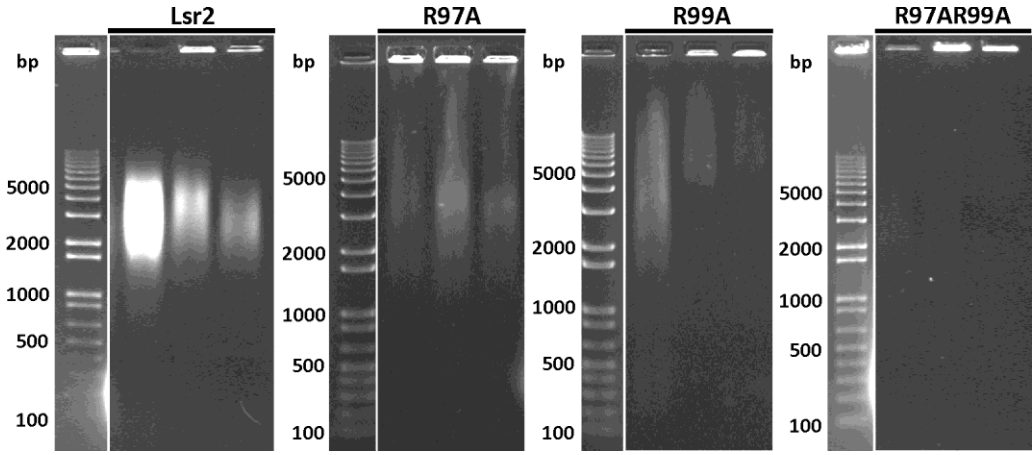


Figure 4.6. The amount of DNA co-purified with each Lsr2 mutant is much less than with native Lsr2. A comparison of 1% agarose gel analysis of IMAC purified native Lsr2 (left), R97A mutant (second from left), R99A mutant (second from right) and R97A-R99A double mutant of Lsr2 (right).

Transmission electron microscopy (TEM) was used previously as a technique to examine the structures formed and the arrangement of native Lsr2 with co-purified DNA (in Chapter Three). Native Lsr2 bound with DNA shows a complex arrangement of strands or "fibrils" that represent DNA completely coated by Lsr2. Since a reduction of co-purified DNA was observed during protein purification of the Lsr2 mutants, it was of interest to use this technique to visualise the effects of residue mutation. Analysis of negatively stained IMAC purified mutant Lsr2 protein revealed striking differences between them (Figure 4.7). The single mutant R97A showed an absence of an ability to form any of the fibril-like structures usually seen with native Lsr2 (Figure 4.7B) whereas the other single mutant R99A was capable of limited structure formation in the form of short independent DNA-coated strands (Figure 4.7C). The double R97A-R99A Lsr2 mutant was also incapable of forming any fibril-like structures (Figure 4.7D). The difference in behaviour between the two single Lsr2 mutants may indicate that the role of each of the arginine residues in DNA binding and arrangement is significantly different.

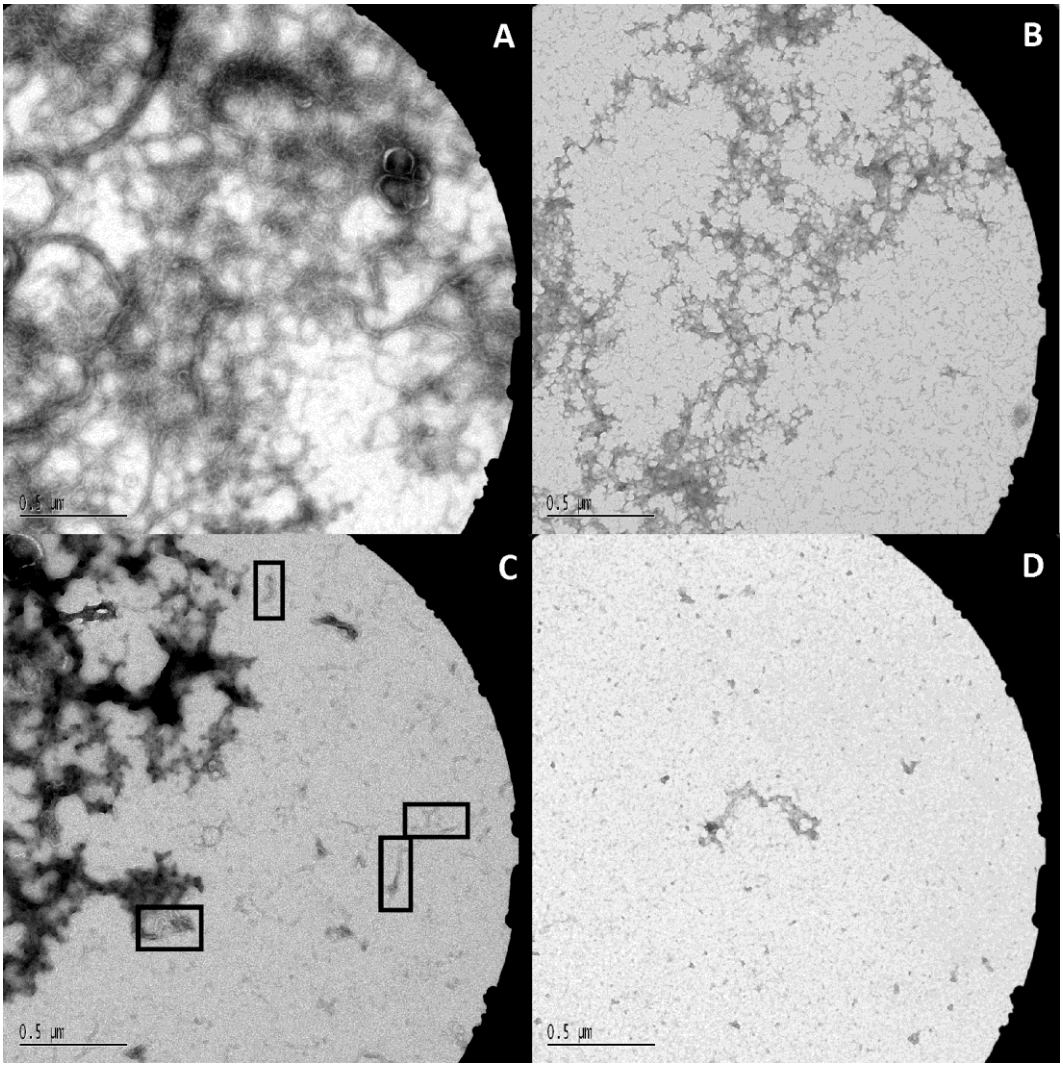


Figure 4.7. Electron micrograph images of IMAC purified Lsr2 mutants show a failure to form protein/DNA structures as normally observed with native Lsr2. (A) Negative stained purified native Lsr2 shows intertwined and coiled fibrils of mixed sizes representing DNA coated by the protein; (B) R97A mutant fails to form any noticeable fibril-like structures - staining artefacts seen only; (C) R99A mutant shows a limited number of single thin fibrils (examples in boxes areas) scattered throughout staining artefacts (upper left); (D) the double R97A-R99A Lsr2 mutant fails to form any DNA/protein structures – background staining seen only.

4.2.6 Mutant protein stability and structure differs from native Lsr2

Protein stability and secondary structure of the Lsr2 DNA binding mutants was characterised using two biophysical methods – fluorescence based thermal stability assay and circular dichroism.

IMAC-purified recombinant native Lsr2, R97A, R99A and R97A-R99A mutant Lsr2 proteins were subjected to thermal denaturation in the presence of a fluorescent dye (SYPRO orange) that has preference for hydrophobic residues that become exposed during protein unfolding. In this assay fluorescence is detected over a temperature range with the fluorescence peak representing the point of complete protein unfolding and the following decrease in intensity indicating protein/probe complex precipitation.

Results from the thermal stability assay show a reduction in protein stability with residue mutation (Figure 4.8). The double mutant R97A-R99A has the most pronounced reduction in stability (52.0°C) compared to native Lsr2 (58.6°C) than either of the individual Lsr2 mutants R97A or R99A (54.8°C and 55.4°C, respectively).

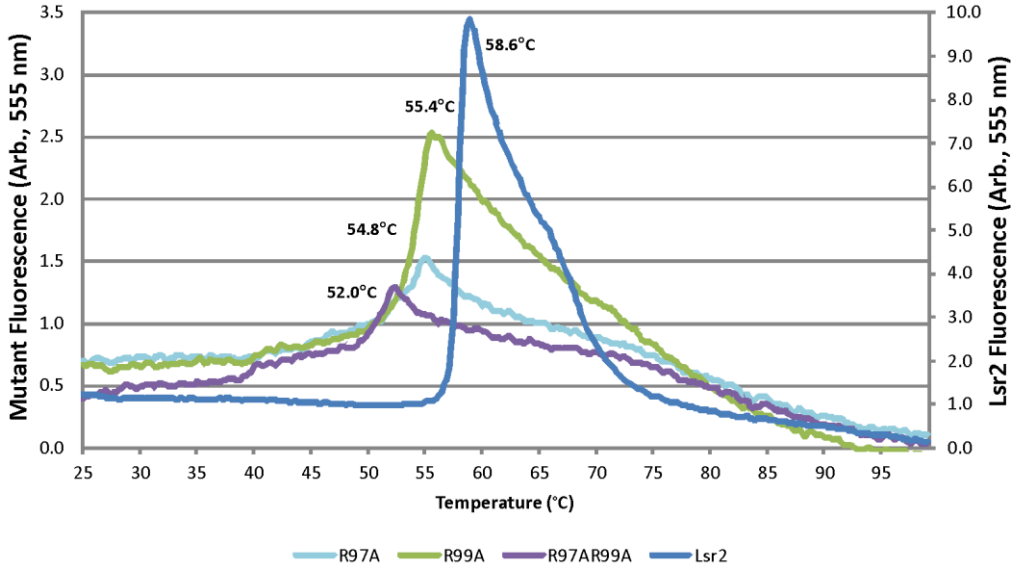


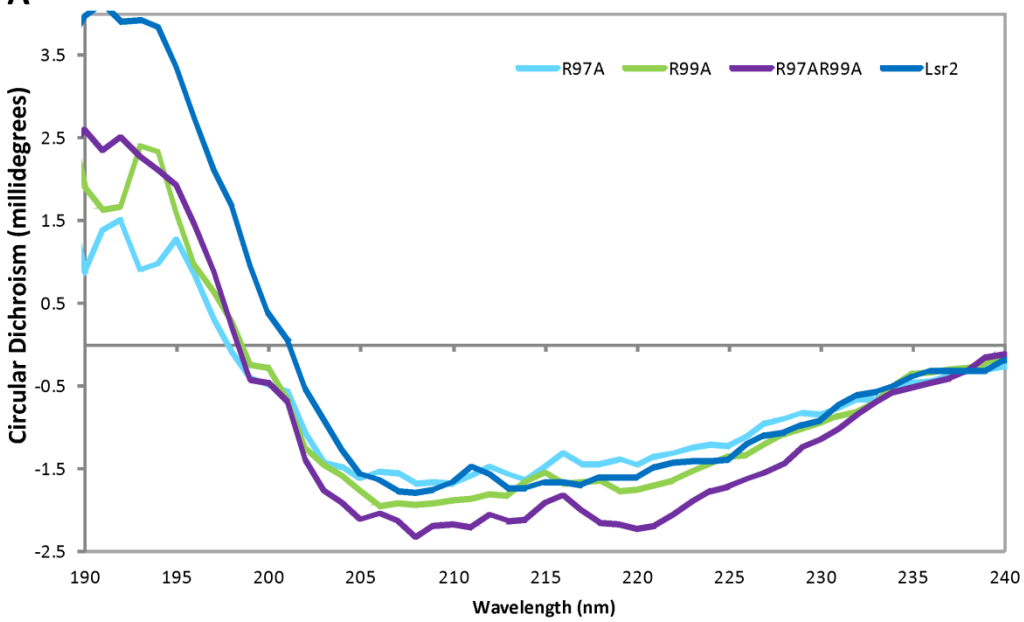
Figure 4.8. Thermal shift assay results indicate a reduction in protein stability in Lsr2 mutants. A shift in the fluorescence peak towards lower assay temperatures indicates a reduction in protein stability. In order of protein stability, native Lsr2 unfolds at 58.6°C, whilst the mutant R99A unfolds at 55.4°C, the mutant R97A unfolds at 54.8°C and the double Lsr2 mutant R97A-R99A unfolds at 52.0°C.

However, these results should be interpreted with caution, as the prepared protein samples contained co-purified genomic DNA. The thermal stability of each protein would be influenced by the amount of DNA that it was bound with. It is possible that thermal stability may be positively correlated with the amount of bound DNA.

Secondary structure characterisation of native Lsr2 and the three DNA binding mutants R97A, R99A and R97A-R99A was undertaken using CD to determine if these mutations would affect protein conformation. For this analysis, protein was further purified by SEC in buffer conditions containing 2 M NaCl to dissociate co-purified DNA from the protein as previously described in Chapter Three and dialysed into a buffer suitable for CD (5 mM phosphate (pH 7.4)). Dialysis into phosphate buffer caused protein precipitation for all purified proteins however some soluble protein remained and was used for analysis. Replicate CD spectra were collected, averaged and baseline and buffer spectra removed.

The raw data from the collected CD spectrum for all protein samples shows subtle differences in the optical activity between native Lsr2, R97A, R99A and R97A-R99A Lsr2 mutants over the appointed wavelength range (Figure 4.9A). All proteins were analysed for secondary structure at a consistent concentration (6 uM) and over a wavelength range of 190 – 240 nm using the K2D2 method (Perez-Iratxeta & Andrade-Navarro, 2008). Secondary structure analysis using the K2D2 method translated the optical activity observed between the appropriate wavelengths into the proportion of α-helix and β-sheet structures (Figure 4.9B). This analysis showed some differences in the percentage of α-helix and β-sheet structures between native Lsr2 and the Lsr2 DNA binding mutants. However, the secondary structure estimation error is high in the analysis of native Lsr2 which indicates the prediction was not very accurate. Furthermore, as the optical activity detected for all proteins was small it is not wise to give too much weight to outcome of the secondary structure analysis. These results may suggest that mutations of the proposed DNA binding residues of Lsr2 had an effect on protein stability and structure.

A



B

Protein	% Alpha Helix	% Beta Sheet	Max Error	% Disorder*
Lsr2	29	16	0.32#	55
R97A	13	35	0.23	52
R99A	14	34	0.23	52
R97AR99A	23	25	0.23	52
Predicted Lsr2	34	19	na	47

Figure 4.9. The CD spectra for native Lsr2, R97A, R99A and R97A-R99A Lsr2 mutants show subtle differences in optical activity which translate into secondary structure variation. (A) Raw data from the collected CD spectrum; (B) a table summarising the estimated proportion of α-helix and β-sheet structures in Lsr2 and the mutants R97A, R99A, R97A-R99A. CD spectrum analysis was performed using K2D2 (Perez-Iratxeta & Andrade-Navarro, 2008). Non-experimental native Lsr2 α-helix and β-sheet secondary structure prediction proportions are also listed for comparison (http://www.predictprotein.org). * is calculated as a percentage of the remaining residues that are neither a α-helix nor a β-sheet; # total error observed is high and structure prediction is not accurate.

CD was also used in an attempt to characterise the DNA binding ability of native Lsr2 and the DNA binding mutants R97A, R99A and R97A-R99A. CD spectra between 250 and 300 nm is dominated by the optical activity of nucleic acids and changes in conformation are attributed to the polynucleotide chain. A suitable quantity of plasmid DNA that gave strong optical activity was established for this assay. It was confirmed that Lsr2 did not contribute to the optical activity over

the appropriate wavelength and hence a titration experiment with native Lsr2 was performed. A reduction in the optical activity of DNA was observed between the DNA only control and the highest amount of protein added however analysis of the change in ellipticity with protein concentration did not demonstrate a typical binding curve (results not shown). Further studies of DNA binding using CD spectra analysis were therefore abandoned for the Lsr2 DNA binding mutants.

Due to the lack of observable binding kinetics between Lsr2 and DNA, a thermal unfolding assay using CD was attempted with the theory that protein unfolding may be influenced by bound DNA. Initial experiments with DNA-free native Lsr2 demonstrated that using this technique, the protein did not unfold following a typical two-state model with a sharp transition between folded and unfolded protein. Rather it showed a gradual unfolding over the temperature range (results not shown).

4.2.7 Lsr2 mutant R97A-R99A is a crystallography candidate

Results from IMAC purification of all three Lsr2 DNA binding mutants revealed that the double R97A-R99A mutant co-purified with the least amount of *E. coli* genomic DNA. Hence, the R97A-R99A Lsr2 mutant was the best candidate for continued exploration and placing into crystallography trials, due to the increased chance of being able to completely remove the co-purified DNA and the possibility that the subsequently purified protein would not adhere to membrane surfaces like its native Lsr2 counterpart.

Large scale preparative SEC in a buffer containing 2 M NaCl was used to prepare DNA-free R97A-R99A Lsr2 protein as was routinely used with native Lsr2 (Figure 4.10). The two peaks on the resulting chromatogram correspond to the DNA (A) and protein (B) which are eluted at the same volumes as observed for native Lsr2. During SEC, DNA is detected by UV absorption at 280 nm and is eluted immediately after the column void volume whist the protein is retarded by the column and passes through at a slower rate.

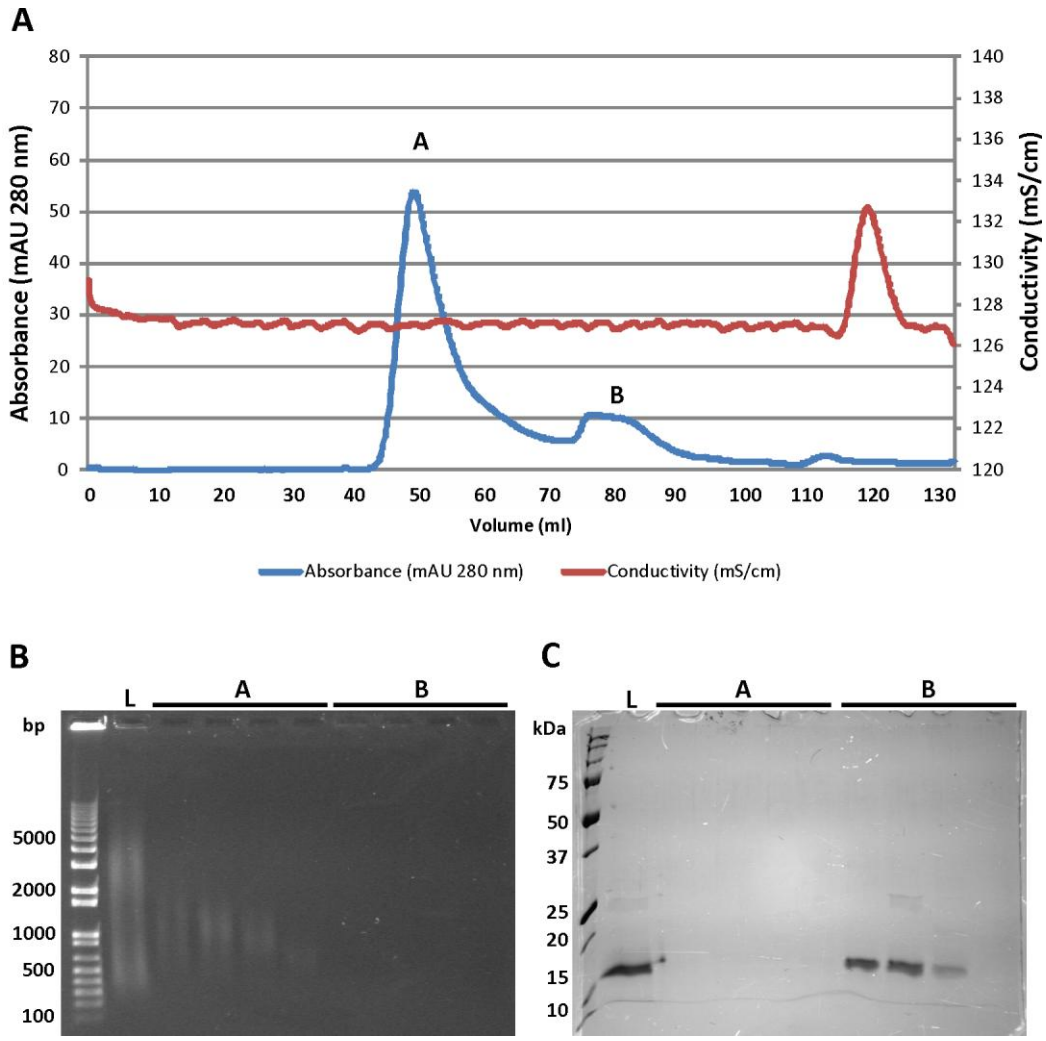


Figure 4.10. Size exclusion chromatography of R97A-R99A mutant Lsr2 in a 2 M NaCl buffer dissociates the protein from the co-purified genomic DNA. (A) Chromatogram of the UV absorbance profile of DNA (A) and R97A-R99A Lsr2 (B); (B) 1% agarose gel with sample load (L) and fractions from peak A and B; (C) 15% SDS-PAGE gel of the corresponding sample load (L) and fractions from peak A and B.

Agarose gel and SDS-PAGE gel analysis of fractions collected during SEC confirm that the mutant protein R97A-R99A is separated from the DNA (Figure 4.10B and C). Due to the high salt content of the buffer used for SEC the size of the Lsr2 mutant R97A-R99A was not able to be accurately determined as the salt influences the mobility of the protein. However it was eluted at the same volume as native Lsr2 did in high salt buffer.

The ability of R97A-R99A mutant Lsr2 to bind to plasmid DNA was examined briefly. Concentrated DNA-free R97A-R99A Lsr2 in a low salt buffer was combined with plasmid DNA in buffers of differing pH and analysed in a gel shift

assay (Figure 4.11B). In comparison with native Lsr2, the R97A-R99A mutant was unable to demonstrate an ability to bind to plasmid DNA and retard its movement through an agarose gel.

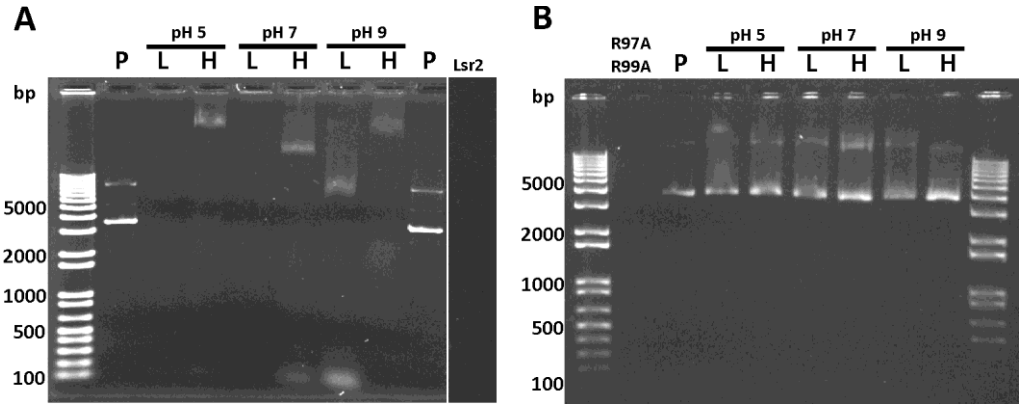


Figure 4.11. The DNA binding mutant R97A-R99A is unable to bind to plasmid DNA. A gel shift assay using 1 % agarose gels demonstrates that native Lsr2 is able to retard the movement of plasmid DNA through the gel (A) as opposed to mutant Lsr2 R97A-R99A which cannot (B). Plasmid only controls (P) of 100 ng (in A) or 50 ng (in B) and protein only (Lsr2 and R97A-R99A) samples (3 ug) are shown. Reactions were setup under different pH conditions with 200 ng (L) and 400 ng (H) plasmid concentrations.

These gel shift results in conjunction with the observation of a negligible amount DNA co-purified with the mutant and the successful removal of this DNA using SEC in high salt, provided support that the double mutant was a promising candidate for protein crystallisation with DNA-free homogenous protein. Purified R97A-R99A Lsr2 was subjected to robotic crystallisation screens at concentrations of 2 mg/ml on two separate occasions. It was difficult to obtain higher protein concentrations. Crystals were observed in some conditions and fine screens and additive screens were set up to screen around the crystallisation conditions. All promising crystals were revealed to be salt after analysis by X-ray diffraction. Dynamic light scattering analysis of concentrated protein used for crystallography attempts showed that the protein sample was not homogeneous.

4.3 Discussion

Lsr2 is a basic protein rich in positively charged arginine residues (13.4 %). It is hypothesised that these positively charged residues may interact with the negatively charged phosphates of the DNA backbone, which would explain the non-specific binding of Lsr2 to DNA. The experimental substitution of a number of positively charged residues has been performed with *M. smegmatis* mc^2155 Lsr2 (Chen *et al.*, 2008). In these studies the *M. smegmatis* mc^2155 Lsr2 arginine residues R86, R99 and R101 were individually mutated to alanine residues and the ability of the mutant recombinant protein to complement the phenotype of an *M. smegmatis* mc^2155 Δ*lsr2* strain was assessed. These residues are the equivalent of the *M. tuberculosis* H37Rv Lsr2 residues R84, R97 and R99 assessed in this study. The *M. smegmatis* mc^2155 study showed that changing residues R99 and R101 did not affect the *in vivo* function of Lsr2 however altering residue R86 did. For *in vitro* work the equivalent R84A *M. tuberculosis* H37Rv Lsr2 mutant was expressed and was shown to be critical for DNA binding (Chen *et al.*, 2008).

In this study cloning of a *M. tuberculosis* H37Rv R84A Lsr2 mutant was unsuccessful due to a reoccurring base pair deletion in the gene and so the DNA binding function of this substitution was not assessed. However, three Lsr2 mutants comprising R97A, R99A and R97A-R99A were cloned, expressed and assessed for their DNA binding ability. This study showed that the DNA binding ability of all three mutants was compromised with the double R97A-R99A mutant showing the least ability to interact with both genomic *E. coli* DNA and plasmid DNA. These results are in contrast to the studies reported by researchers investigating *in vivo* complementation of an *M. smegmatis* mc^2155 *lsr2* knockout (Chen *et al.*, 2008).

This research suggests that the residues R97 and R99 of Lsr2 are significant in the binding of DNA and if their mutation is not significant enough to disrupt the normal function of Lsr2 *in vivo* then this suggests that there are other crucial DNA binding residues that work in synergy with R97 and R99 and could

compensate in some way. A hierarchy of residue DNA binding ability may be suggested by the DNA/Lsr2 structures seen using TEM.

The double R97A-R99A Lsr2 mutant lacked the ability to form fibril-like structures that correspond to protein coating the DNA (in IMAC purified protein samples) as was usually seen in the native Lsr2 micrographs. However it was interesting that a difference in structure formation was observed between the two single mutants R97A and R99A. The mutant R97A was not capable of forming coiled-strands or fibrils as was typically observed, however in the micrographs of R99A this protein showed a limited number of short strands that were fibril-like in structure. The TEM images suggest, by their lack of DNA/Lsr2 structures, that the residue R97 has a stronger interaction with DNA than R99 and that altering both residues has an additive effect. As the *in vivo* function of a R86A Lsr2 mutant in a *M. smegmatis* mc^2155 Δ*lsr2* was compromised it is possible that this residue has an even stronger interaction with DNA. However if this was the case, we would see fibril-like structures in TEM images despite mutations of both R97 and R99 residues.

Analysis of protein thermal stability of the three DNA binding mutants showed that their denaturation temperature may have been influenced directly by the amount of bound DNA in the sample preparation. Secondary structure analysis of DNA-free mutant proteins showed subtle differences between the proportion of α-helix and β-sheet structures. However the buffer conditions required for CD analysis were not ideal for optimal protein solubility and this may have influenced the results.

Protein crystallisation was attempted using the double R97A-R99A mutant in the belief that its reduced DNA binding ability would improve sample homogeneity and behaviour, thereby increasing the chances of successful crystallisation. However protein crystals were not successfully generated after thorough fine screening attempts.

Chapter Five – Structure of the N-terminal dimerisation domain of Lsr2

5.1 Introduction

One of the main objectives of this PhD research was to determine the 3D structure of the DNA binding protein Lsr2 from *M. tuberculosis* H37Rv as, at the time, this structure was unknown and this information would reveal critical functional information. However, in 2010 an NMR structure of the C-terminal DNA binding domain of Lsr2 was published (Gordon *et al.*, 2010). The structure of the C-terminal DNA binding domain revealed a mechanism for DNA binding and identified residues crucial for this interaction. A model of DNA binding by Lsr2 postulated that Lsr2 could bind DNA in two orientations by grabbing either edge of the minor groove like a clamp, orientating and inserting the conserved residues R97, G98 and R99 into the minor groove. The publication of the structure of the C-terminal domain of Lsr2 forced a change in the focus and direction of this PhD research. Further studies were therefore directed towards determining the 3D structure of the N-terminal dimerisation domain of Lsr2.

Lsr2 is predicted to be organised into two functional domains, the N-terminal dimerisation domain and the C-terminal DNA binding domain, separated by a flexible linker. Based on the elucidation of the 3D structure, the C-terminal DNA binding domain is thought to comprise residues 51-112 and the N-terminal dimerisation domain is between residues 1-65 based on *in vivo* complementation experiments (Gordon *et al.*, 2010). A multiple amino acid alignment of Lsr2 across a number of bacterial species revealed that a number of highly conserved residues were present within the two aforementioned domains whereas there was little conservation in the proposed linking region (Figure 1.1). Residue conservation is commonly indicative of the residue having a role in protein function or structure and as the conserved residues of Lsr2 were found within the two domains this supported the description of the domain organisation.

Secondary structure predictions showed that there is likely to be three sequential β-sheets followed by one α-helix within the proposed N-terminal dimerisation domain (residues 1-65) of Lsr2 (Figure 1.5). The α-helix terminates at residue 56 which coincides with a reduction in residue conservation and a number of residues with low conservation and no predicted secondary structure follow this domain until residue 80. Residue 80 is the beginning of the predicted α-helix secondary structure of the C-terminal domain. The NMR structure of the residues 66-112 showed that residues 66-74 were flexible in nature and that the more rigid structure began at residue 78 with the first α-helix (Gordon *et al.*, 2010). Therefore, it was proposed that a disordered and flexible region lies between the two domains (residues 56 and 80).

A number of attempts to crystallise the full-length construct of Lsr2 by using alternative protein tag purification techniques, residue modification (Chapter Three) and site directed mutagenesis (Chapter Four) were made during the course of this investigation, all of which were unsuccessful in producing protein crystals. It was proposed that the presence of a highly disordered and flexible region between the two domains of Lsr2 was the reason why crystallisation of the full construct was so difficult. Flexibility in a protein is known to make crystallisation difficult (Benvenuti & Mangani, 2007). Consequently, it is often beneficial to work with individual domains of proteins to remove poorly conserved and flexible regions and therefore increase the likelihood of successful crystallisation (Dale *et al.*, 2003). To increase the probability of crystallising Lsr2, the cloning, expression and purification of a number of different constructs of the N-terminal domain was undertaken.

The domain organisation of Lsr2 is thought to be similar to that of the *E. coli* nucleoid-like DNA binding protein H-NS (Gordon *et al.*, 2008). These two proteins have been functionally compared using domain swapping experiments, which determined that the N-terminal domain of Lsr2 is capable of enabling protein dimerisation (Gordon *et al.*, 2010). *In vitro* complementary studies have shown that Lsr2 can complement phenotypes related to H-NS gene mutations in *E. coli* and that the reverse complementation of H-NS in the Lsr2 phenotype in

M. smegmatis mc^2155 is possible, demonstrating functional similarity between these two DNA binding proteins (Gordon *et al.*, 2008). The H-NS N-terminal oligomerisation domain has recently been solved using crystallography, revealing that the protein is capable of higher order complexes via sequential interactions of this domain, forming oligomers (Arold *et al.*, 2010).

Crystallographic data of the N-terminal dimerisation domain of Lsr2 was obtained during this study and the work presented in this chapter shows the first 3D structure of this domain of Lsr2 from *M. tuberculosis* H37Rv. The structural detail shows the mechanism of complex oligomerisation of this protein. Evidence of the function of Lsr2 in DNA packaging was provided by further *in vitro* studies. The structure shows that the mechanism of oligomerisation of Lsr2 is distinct from that of H-NS and that although there is functional similarity between these proteins, their mode of DNA packaging is dissimilar.

Work presented in this chapter was recently published in PLoS ONE (Summers *et al.*, 2012).

5.2 Results

5.2.1 Multiple N-terminal dimerisation domain constructs improve protein yield and solubility

A number of constructs of the N-terminal domain of Lsr2 were designed (Table 5.1) and cloned into the *E. coli* expression vector pPROExHTb using the *NcoI*/Xho1 restriction sites, positioning a 6 x histidine tag and rTEV protease cleavage site at the N-terminal end of the expressed protein. Lsr2 N-terminal domain variants were designed to range from 56 to 74 amino acids in length, with the shortest variant terminating at the end of the predicted α-helix region and the longest alternative terminating 5 residues prior to the α-helix of the C-terminal DNA binding domain. The molecular weight of the expressed proteins ranged between 9.1 kDa and 10.9 kDa. Out of interest, a single construct of the C-terminal DNA binding domain was also designed but it failed to express after a number of attempts at various expression temperatures and repeat transformations into an expression strain (results not shown).

Table 5.1. Seven N-terminal Lsr2 domain constructs were designed, cloned, expressed and purified. The C-terminal boundaries of Lsr2 N-terminal constructs were terminated at random points between the two domains of Lsr2. Exp. MW = expressed molecular weight in Daltons.

Construct	Amino Acid Sequence	Residues	Exp. MW
Nterm	55 - G R R V G	59	9413
Nterm+	55 - G R R V G G R R R G R S G S	68	10383
Nterm1	55 - G R	56	9100
Nterm2	55 - G R R V G G R	61	9626
Nterm3	55 - G R R V G G R R R G	64	9996
Nterm4	55 - G R R V G G R R R G R S G S G R G	71	10653
Nterm5	55 - G R R V G G R R R G R S G S G R G R G A	74	10938

Small scale expression testing and histidine tag binding assays were successful for all seven N-terminal Lsr2 constructs and the results revealed that Nterm1 was poorly expressed compared to the other constructs (results not shown). Large scale expression and IMAC purification was performed for all but the Nterm1 construct due to its reduced expression level (Figure 5.1).

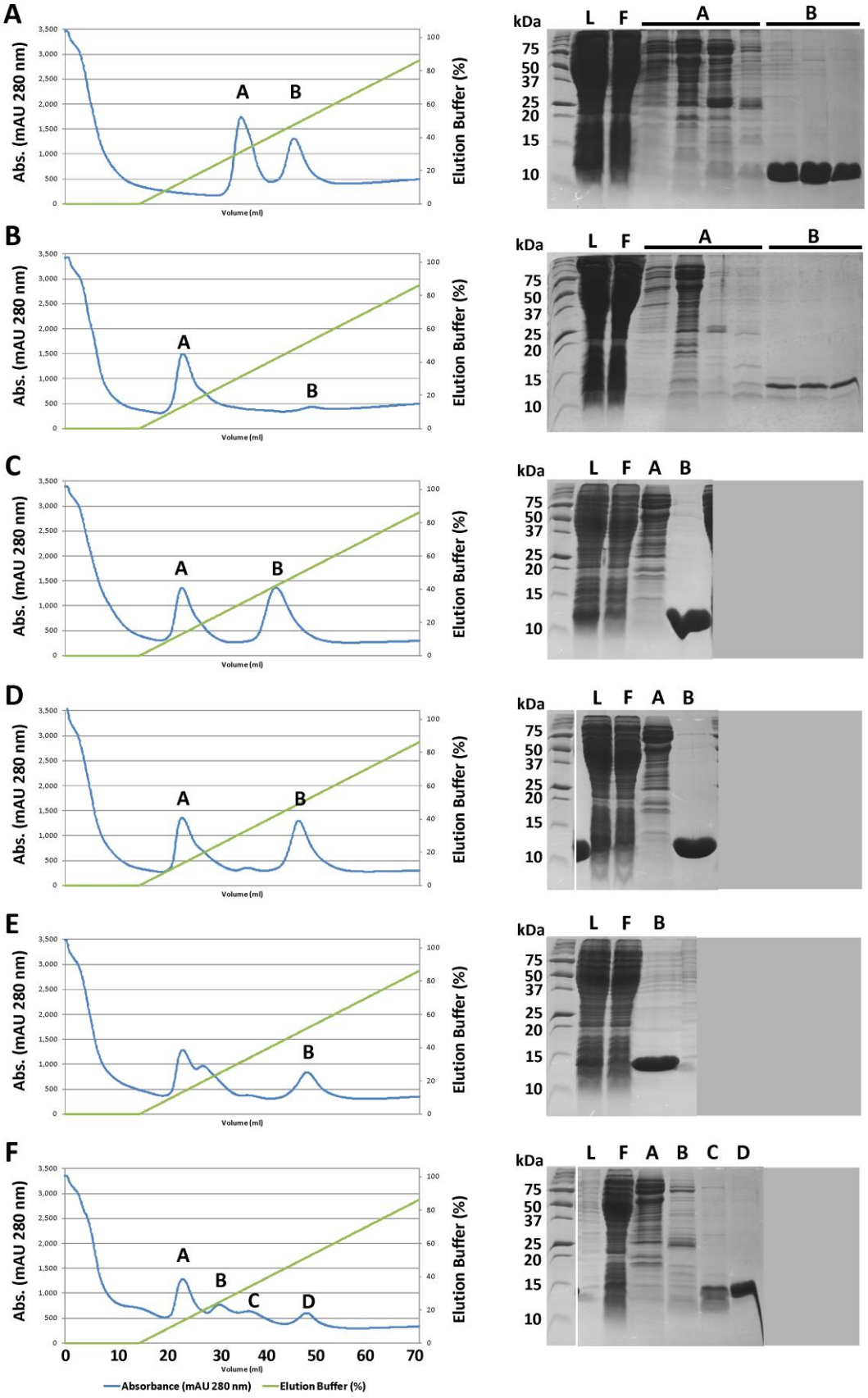


Figure 5.1. Variation in IMAC purification of Nterm (A), Nterm+ (B), Nterm2 (C), Nterm3 (D), Nterm4 (E) and Nterm5 (F) N-terminal domains of Lsr2. Chromatograms (left panel) and 18% SDS PAGE gels (right panel) show cell lysate (L), flow (F) and fractions from peaks (A-D).

The protein of interest was eluted toward the end of the elution gradient at approximately 40-50% elution buffer which corresponded to a buffer containing approximately 400-500 mM imidazole. Variation in the protein expression level for each construct was apparent after purification. Based on the absorbance level of the protein peak during IMAC purification, Nterm+ had the lowest protein yield, followed by Nterm5, and Nterm4 at a higher yield, and Nterm, Nterm2 and Nterm3 having similarly highest protein yields.

SEC of the six purified N-terminal domain constructs was performed in a standard low salt buffer as it was determined by agarose gel electrophoresis of protein fractions, that there was no co-purified *E. coli* genomic DNA as had been typically observed with the full length Lsr2 protein (results not shown). The four constructs Nterm, Nterm+, Nterm2 and Nterm3 were purified successfully by size exclusion chromatography (Figure 5.2), whereas the constructs Nterm4 and Nterm5 aggregated prior to chromatography, during protein concentration, leaving no soluble protein in solution. Based on the absorbance level and SDS-PAGE gel results, a very low yield of protein for Nterm+ was obtained after size exclusion (Figure 5.2B) in comparison to the higher yields of Nterm3, Nterm and Nterm2 in increasing yield respectively (Figure 5.2 D, A and C).

Overall, the three Lsr2 N-terminal domain constructs Nterm, Nterm2 and Nterm3 were expressed at high levels in *E. coli*, behaved favourably during preparation for size exclusion chromatography and were able to be purified and concentrated to amounts appropriate for protein crystallography trials.

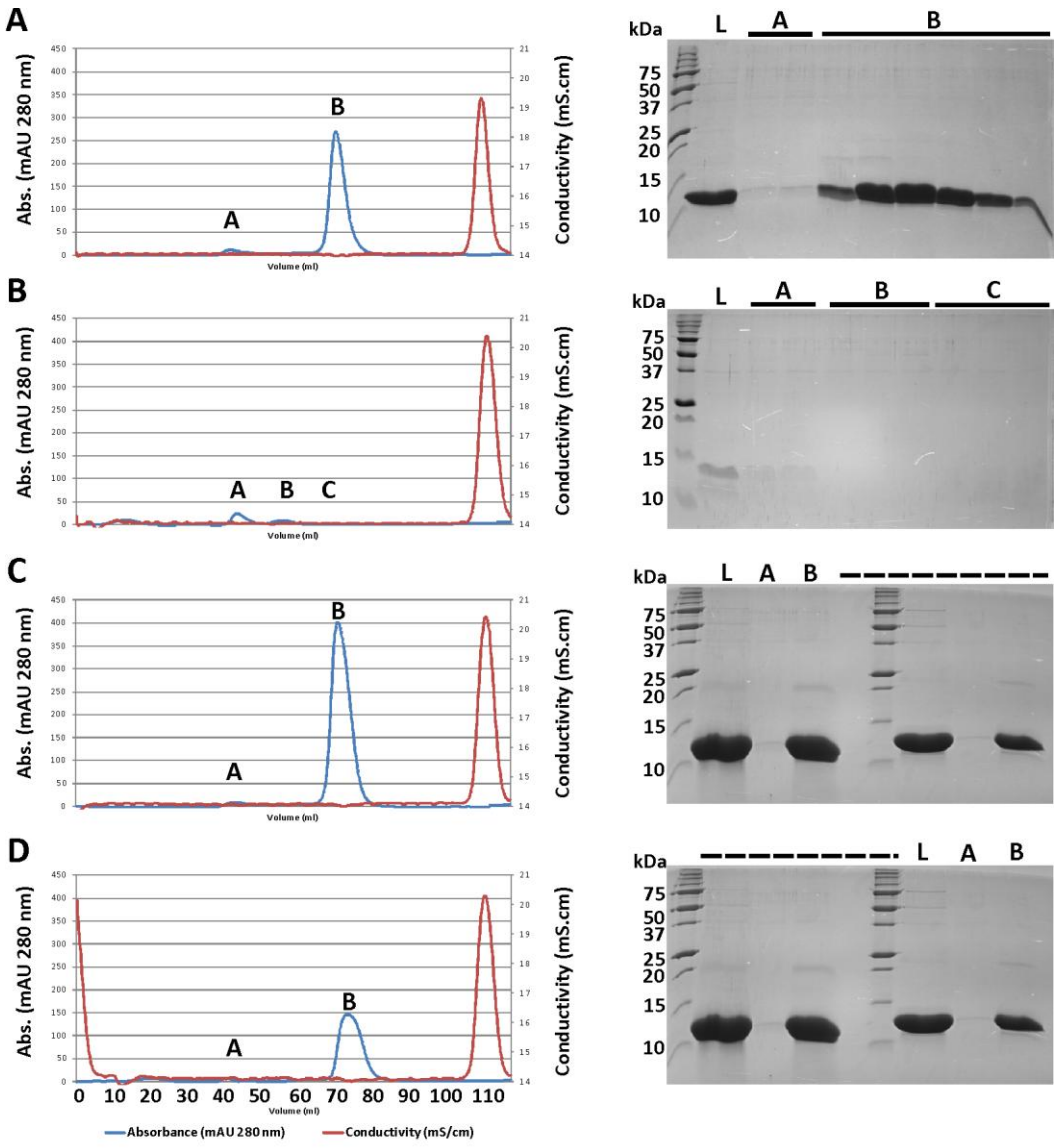


Figure 5.2. Size exclusion chromatography of N-terminal domains Nterm (A), Nterm+ (B), Nterm2 (C) and Nterm3 (D) shows a variation in protein solubility and yield. Chromatograms (left panel) and 18% SDS PAGE gels (right panel) which show loaded sample (L) and fractions from appropriate peaks (A-C). IMAC purified Nterm4 and Nterm5 precipitated while concentrating and were unable to be purified by SEC while Nterm+ precipitated soon after concentration.

To obtain initial information regarding the ability of the N-terminal domain to assemble into larger subunits, a glutaraldehyde cross-linking experiment was performed which showed that Nterm had an ability to form multimers (Figure 5.3A). Further analysis of cross-linked Nterm using a calibrated size exclusion column illustrated a number of protein species that showed an approximate doubling in size, suggesting that a dimer was the most common smallest denomination used in the assembly of larger structures (Figure 5.3B).

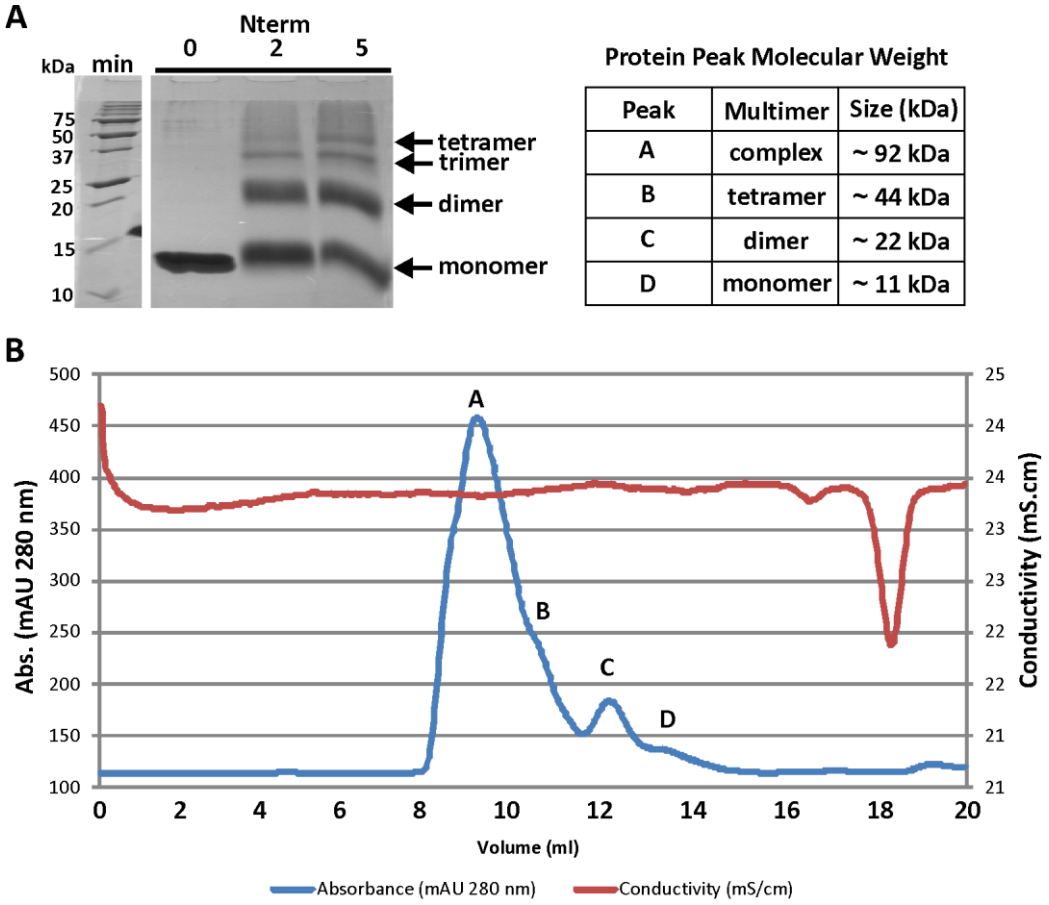


Figure 5.3. The construct Nterm of the N-terminal domain of Lsr2 forms oligomeric structures when incubated with 0.1% glutaraldehyde. (A) IMAC purified Nterm forms dimers, trimers, tetramers and larger structures after 2 minutes with 0.1% glutaraldehyde (18% SDS-PAGE); (B) SEC analysis of Nterm crosslinked for 5 minutes illustrates larger complexes from over time (peak A) as smaller species such as dimers and tetramers diminish (peaks C and B, respectively). The table lists calculated approximate molecular weight of species observed during SEC analysis (Superdex75 10/300 column).

5.2.2 Multiple N-terminal protein crystallography trials required for successful protein crystallisation

Size exclusion chromatography purified Nterm protein was subjected to robotic crystallisation screens at a concentration of 7 mg/ml however no crystals were observed.

Size exclusion chromatography purified Nterm2 and Nterm3 were subjected to robotic crystallisation screens at concentrations of 21 mg/ml and 2 mg/ml respectively and tiny crystals were observed in three conditions for Nterm2 only. Hanging drop fine screens around the crystallisation conditions and additive screens were prepared but resulted in limited crystal development. At this time it was noted that limited crystal development only occurred when an older Nterm2 protein preparation was used and not with recently purified protein. SDS-PAGE analysis confirmed that there was still full length protein remaining in the older Nterm2 protein preparation (results not shown).

The arrival of a liquid handling robot in the lab prompted new Hampton crystallisation screens to be completed with 3-week old Nterm2 at 15 mg/ml and this resulted in crystal development in a variety of conditions. Extensive fine screening around a wide range of PEG conditions, including crystal seeding, use of microbridges and non-siliconised coverslips and additive screening was performed over a number of weeks using Nterm2 from the same purification and resulted in crystal growth in a greater variety of conditions.

Further Hampton crystallisation screens were performed with Nterm2 at 63 mg/ml after the removal of the histidine tag using rTEV protease and further purification but were unsuccessful in producing crystals. Hampton crystallisation screens were also performed with Nterm2 at 15 mg/ml co-crystallised with trypsin at a ratio of 1:500 as detailed in section 5.2.3, resulting in crystals in a wide variety of conditions.

5.2.3 Protein ''maturation'' or protease digestion facilitates successful diffraction quality crystal formation

The development of diffraction quality crystals of the Nterm2 construct was a slow iterative process that took a number of months from the initial robotic crystallisation screens laid down in March 2011. Initial crystals did not diffract however these results ruled out the formation of salt crystals and renewed efforts to produce better crystals (Figure 5.4A). Subsequent testing of crystals at the Australian Synchrotron revealed that higher quality diffracting protein crystals were developing (Figure 5.4B) but that the data was not suitable due to multiple crystal plates forming together. After further fine screening, crystals that were more singular and also diffracted reasonably well were tested using The University of Auckland home source X-ray beam (Figure 5.4C).

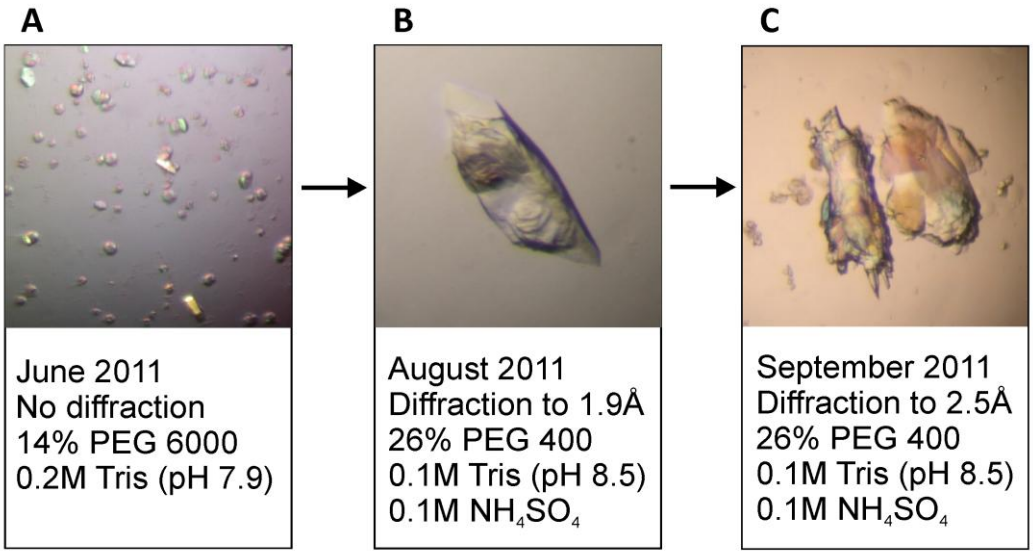


Figure 5.4. Iterative crystal improvements generate crystals of Nterm2 capable of diffracting strongly during X-ray analysis. (A) Initial crystals were confirmed to be non-diffracting protein; (B) high resolution but smeary data was achieved with a crystal from an additive screen; (C) crystal refinement gave lower resolution but less smeary data. Protein crystal X-ray analysis was completed at the Australian Synchrotron for crystals A and B, and at The University of Auckland home source for crystal C.

Initial robotic screens performed at The University of Auckland with Nterm2 did not yield any crystals, however with the arrival of a liquid handling robot in the lab and the use of Hampton crystal screens, a number of promising crystallisation conditions were discovered. Unexpectedly, it was difficult to produce good crystals in subsequent fine screen conditions. It was proposed that the protein preparation was not fresh enough as it was initially used in the Hampton screens at an age of 3 weeks old and perhaps now contained a mixture of protein species. Size exclusion chromatography was used to re-purify the protein sample in an attempt to "refresh" the preparation. Surprisingly, the results revealed that instead of expected protein degradation, that a number of larger multimers had naturally formed over time during protein aging, and protein peaks with calculated sizes of dimers, tetramers and larger complexes were seen (Figure 5.5A). Analysis of protein from each peak on a 10% native gel illustrated that protein in these different complexes had a variety of electrophoretic properties however there were only two alternative protein sizes as shown by 16.5% SDS-PAGE analysis (Figure 5.5B and C).

This protein aging or "maturation" was recreated by purifying fresh Nterm2 and repeating identical fine screen conditions every two weeks in an attempt to try to mimic the conditions originally present for successful crystal generation. At the same time, trypsin co-crystallisation techniques were used in fine screens at a dilution mass ratio of 1:500. Samples of trypsin treated fresh Nterm2 protein were analysed on 10% native and 16.5% SDS-PAGE gels and compared to the results observed for "matured" Nterm2 (Figure 5.5B and C). In a similar outcome to naturally "matured" Nterm2, trypsin treated Nterm2 produced a number of different native protein species but still resulted in only two main protein sizes.

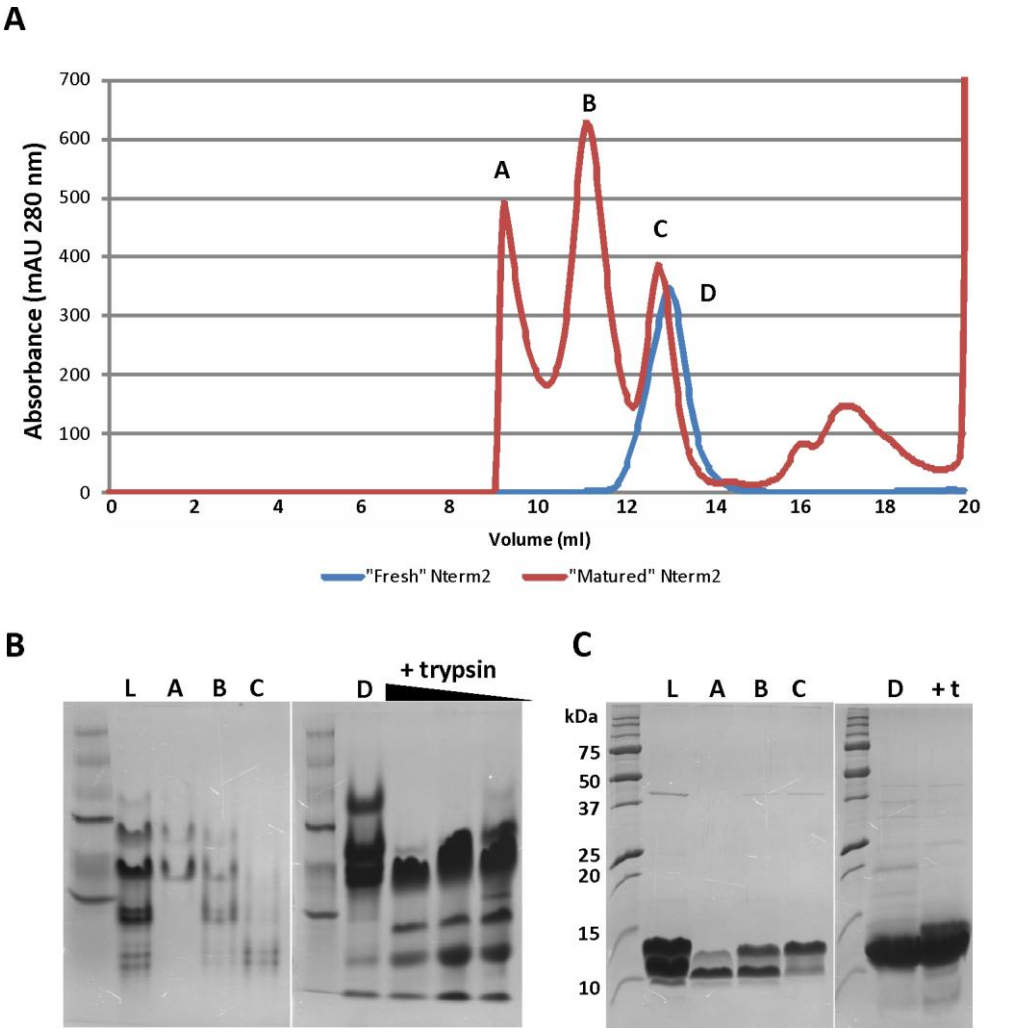


Figure 5.5. Re-purification of "matured" Nterm2 by size exclusion chromatography revealed that a number of multimers had naturally formed. (A) Analytical size exclusion chromatography shows multimeric (A) tetrameric (B) and dimeric (C), forms of "matured" protein in comparison to "fresh" dimeric Nterm2 (D); (B) protein loaded (L) and fractions from each peak (A-D) display different electrophoresis properties on a 10% native gel; (C) analysis of the corresponding fractions on 16.5% SDS-PAGE shows two protein sizes. Addition of trypsin in decreasing amounts (1:550, 1:5000, 1:50000 mass rations) to "fresh" Nterm2 (for protein crystallisation) artificially produces a number of protein species as seen on a 10% native gel (B), and results in only two protein sizes on 16.5% SDS-PAGE (C).

Nterm2 trypsin co-crystallisation fine screens at a dilution ratio of 1:500 resulted in rapid forming but small crystals. Dilution ratios were optimised in an attempt to improve crystal quality. It was hypothesised that reducing the amount of trypsin added to Nterm2 would slow the rate of crystal formation and produce higher quality crystals however dilution mass ratios of 1:5,000 and 1:50,000 resulted in poorer quality crystals instead (Figure 5.6). The original mass ratio of 1:500 was subsequently selected for all further co-crystallisation attempts and further Hampton crystallisation screens with Nterm2 and trypsin produced crystals in a wide variety of conditions, leading to promising fine screen crystal development.

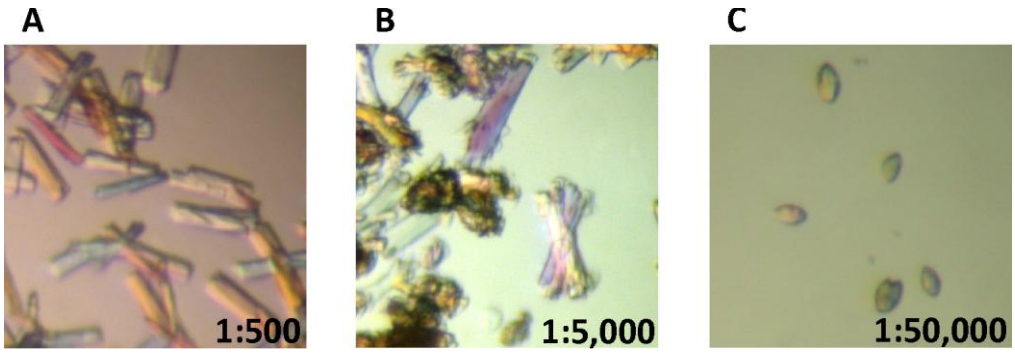


Figure 5.6. Crystal growth optimisation using a range of trypsin:Nterm2 dilution mass ratios. Crystals formed rapidly using trypsin co-crystallisation techniques. A dilution ratio of 1:500 trypsin to Nterm2 produced better quality crystals (A) compared with dilution ratios of 1:5,000 (B) and 1:50,000 (C).

The successful collection of X-ray diffraction data from high quality Nterm2 crystals was a culmination of the discovery of protein “maturation” being an essential factor in crystal development and the investigation of the ability of trypsin to artificially create this protein aging, in combination with the use of the Australian Synchrotron facility. Crystals of Nterm2 from two different crystallisation conditions using “matured” and trypsin-added protein preparations provided suitable data for solving the structure of Nterm2.

Two singular diamond shaped Nterm2 crystals that grew within a mixture of crystal shapes successfully gave native data sets using the home source X-ray beam at The University of Auckland. These crystals grew in 26% PEG 400, 0.1 M Tris (pH 8.5), 0.12 M $(NH_4)_2SO_4$, in a hanging drop after four days, using "mature" Nterm2 protein that was approximately 10 weeks old (Figure 5.7A). Data sets were collected to 1.73Å and 1.87Å in the $P2_1$ space group.

A data set for a single hexagonal shaped Nterm2 crystal co-crystallised with trypsin was also collected at The University of Auckland however the slight smearing of the diffraction image spots made indexing of the data difficult. This crystal grew in 0.1 M MES (pH 6.5), 0.01 M $CoCl_2$, 1.4 M $(NH_4)_2SO_4$, 7.5% 1-4-dioxane, in a hanging drop fine screen after six days, using protein that was 5 weeks old. After transportation to the Australian Synchrotron, a data set to 1.96Å resolution in the $P3_121$ space group was collected for this crystal (Figure 5.7B).

The successful collection of a diffraction pattern data set from a protein crystal is not all that is required to be able to solve the structure of a protein. The phases associated with each diffraction spot also need to be determined. A number of techniques are available to determine the phases. Molecular replacement is a common technique that utilises the phase relationships from a related molecule and an attempt is made to fit the structure under investigation to a previously solved structure. In the case of the N-terminal domain of Lsr2, there were no other known related structures solved at this point in time and hence molecular replacement could not be used. The heavy atom selenium may be incorporated into a protein as selenomethionine enabling anomalous scattering techniques to be used for phase determination.

The data collection statistics for the native and hexagonal data sets are shown in Table 5.2.

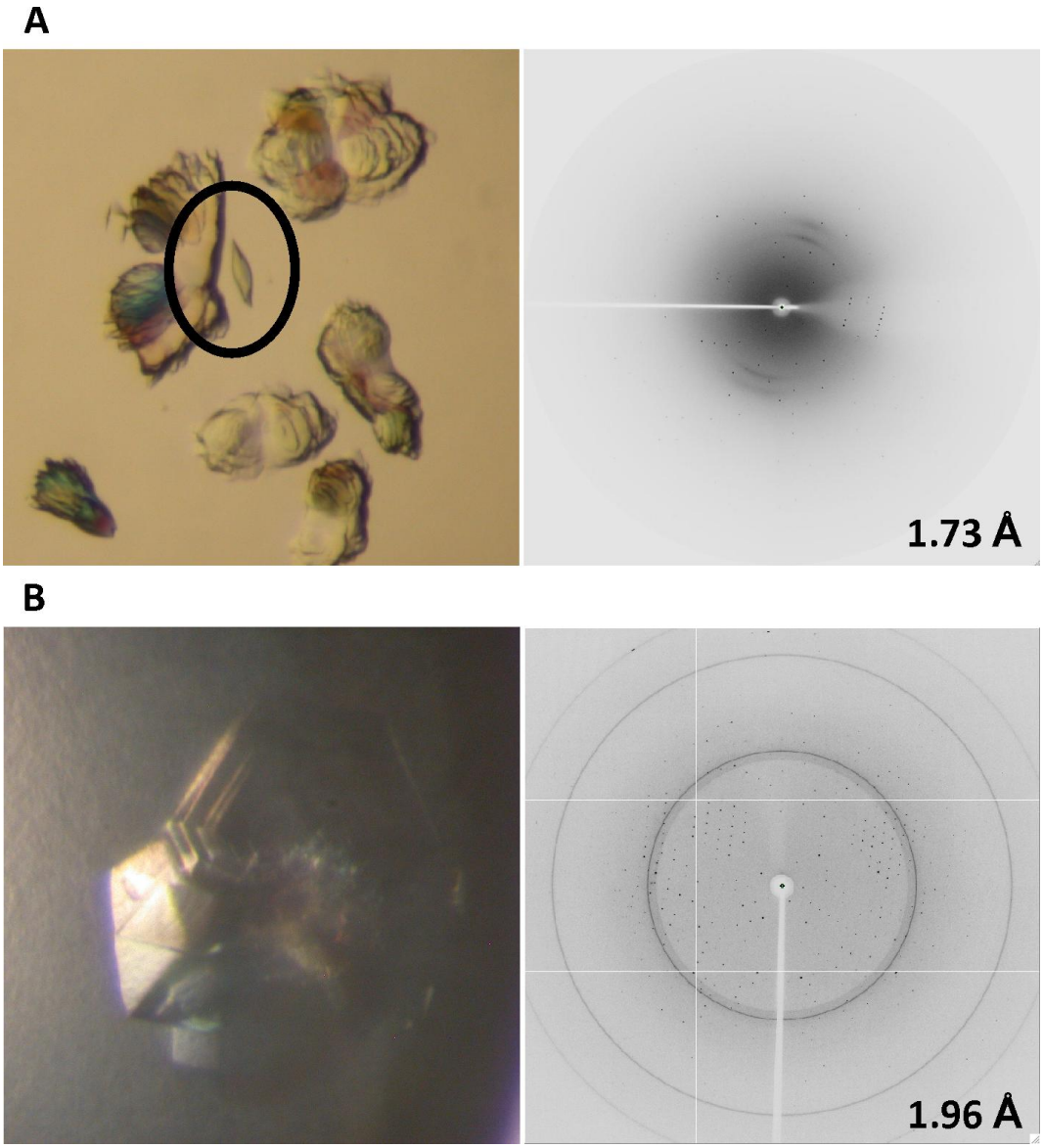


Figure 5.7. Crystals generated from "matured" Nterm2 and Nterm2 co-crystallised with trypsin diffracted to a high resolution and provided high quality data. (A) Two singular diamond shaped crystals grown in 26% PEG 400, 0.1 M Tris (pH 8.5), 0.12 M $(NH_4)_2SO_4$, provided native data sets to 1.73Å and 1.87Å in a $P2_1$ space group; (B) a hexagonal shaped crystal co-crystallised with trypsin (1:500), grown in 0.1 M MES (pH 6.5), 0.01 M $CoCl_2$, 1.4 M $(NH_4)_2SO_4$, 7.5% 1-4-dioxane, provided a data set to 1.96Å in a $P3_121$ space group. The corresponding diffraction pattern images are displayed on the right of the crystal photos. Native data sets and the hexagonal data set were collected at the home source at The University of Auckland however a second higher quality hexagonal data set was subsequently collected at the Australian Synchrotron.

Table 5.2. Crystallographic data collection statistics.

	Nterm2	Nterm2 + Trypsin
Data Collection		
Wavelength (Å)	1.5417	0.9786
Space Group	P1 2_1 1	P3_1 2 1
Unit Cell (Å)	32.51, 27.03, 56.83	57.39, 57.39, 105.31
Unit Cell (°)	90.0, 94.3, 90.0	90.0, 90.0, 120.0
Resolution (Å)[a]	32.42-1.73 (1.82-1.73)	52.65-2.04 (2.15-2.04)
R-merge	0.092 (0.308)	0.096 (0.562)
Completeness (%)	99.4 (96.3)	100.0 (100.0)
Redundancy	9.77 (9.18)	20.43 (20.45)
No. of Observations	102957 (13617)	272612 (38858)
No. of Unique Reflections	10538 (1483)	13341 (1900)
Mean I/σ(I)	17.1 (6.9)	22.6 (6.2)

[a] numbers in parentheses correspond to the highest resolution shell

5.2.4 Generation of methionine mutants for selenomethionine labelling provides diffraction quality crystals

The heavy atom selenium has been successfully used for phase determination via multiwavelength anomalous dispersion (MAD) after its incorporation into protein as the modified residue selenomethionine. Mutant variants of Nterm2 were designed to incorporate the residue methionine into the protein for selenomethionine labelling, as there were only two existing methionine residues at the very N-terminal end of the protein and these were predicted to be unsuitable for labelling due to their location.

Three alternative residues within Nterm2 (excluding the expressed tag and linker) were chosen for mutation to a methionine residue. These residues were selected based on a number of variables that help predict residues that would least likely disrupt protein tertiary structure if altered. Residues that have hydrophobic properties are likely to be buried within the protein structure and this also coincides with positioning within highly ordered regions of the protein such as in α-helices and β-sheets. The residues I34, L44 and L48 were selected for mutation to methionine as they were highly conserved, hydrophobic, and positioned in structurally ordered regions as illustrated by the HMM logo and

secondary structure predictions (Figure 5.8). It was predicted that each of these residues had less than 5% exposure to solvent which supported their selection as buried residues (www.predictprotein.org). The presence of other hydrophobic residues valine and phenylalaine at the same positions as I34, L44 and L48 in the HMM logo confirmed that these were positions in the protein that showed conservation of hydrophobic, buried residues.

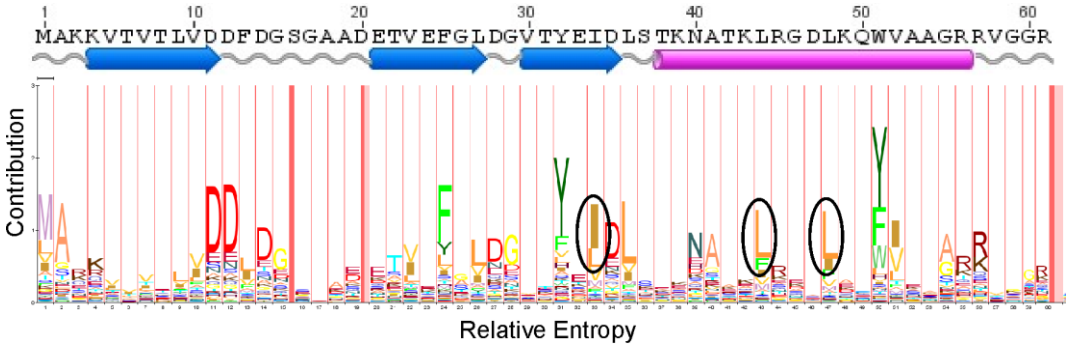


Figure 5.8. Hidden Markov Model (HMM) logo of the Nterm2 protein construct region shows highly conserved and structurally important residues which assisted in residue selection for site directed mutagenesis. Residues I34, L44 and L48 (circled) are highly conserved (as demonstrated by the large letter height and width), hydrophobic (yellow) and are found in structurally ordered regions as predicted by the secondary structure. Letter height and width at each position is proportional to its occurrence and relative role, respectively. Secondary structure prediction is shown underneath the *M. tuberculosis* Lsr2 amino acid sequence with β-sheets depicted by blue arrows and a α-helix by the pink cylinder. HMM logo - http://pfam.sanger.ac.uk/family/PF11774.

Site directed mutagenesis (using a QuikChange™ protocol) was successfully used to generate two of the methionine mutants, I34M and L48M, however this technique was unsuccessful for the mutant L44M and hence standard PCR and cloning techniques were employed to generate this methionine Nerm2 mutant (results not shown).

The three methionine mutants were expressed in *E. coli* and purified using IMAC with an elution gradient. The protein of interest eluted in the second peak at approximately 40% elution buffer which corresponded to a buffer containing approximately 400 mM imidazole (Figure 5.9). The protein yield (determined by peak absorbance) of L44M was similar to that of the native Nterm2 (Figure 5.1) whereas I34M produced less, and L48M expression produced much larger

amounts, than native Nterm2. These yields transferred through to SEC purification, and resulted in the mutant I34M with the lowest yield, L48M with the highest yield and L44M with a yield in between (Figure 5.10). Analytical SEC confirmed that mutant variants of Nterm2 were still capable of forming and purifying as dimers (Figure 5.10).

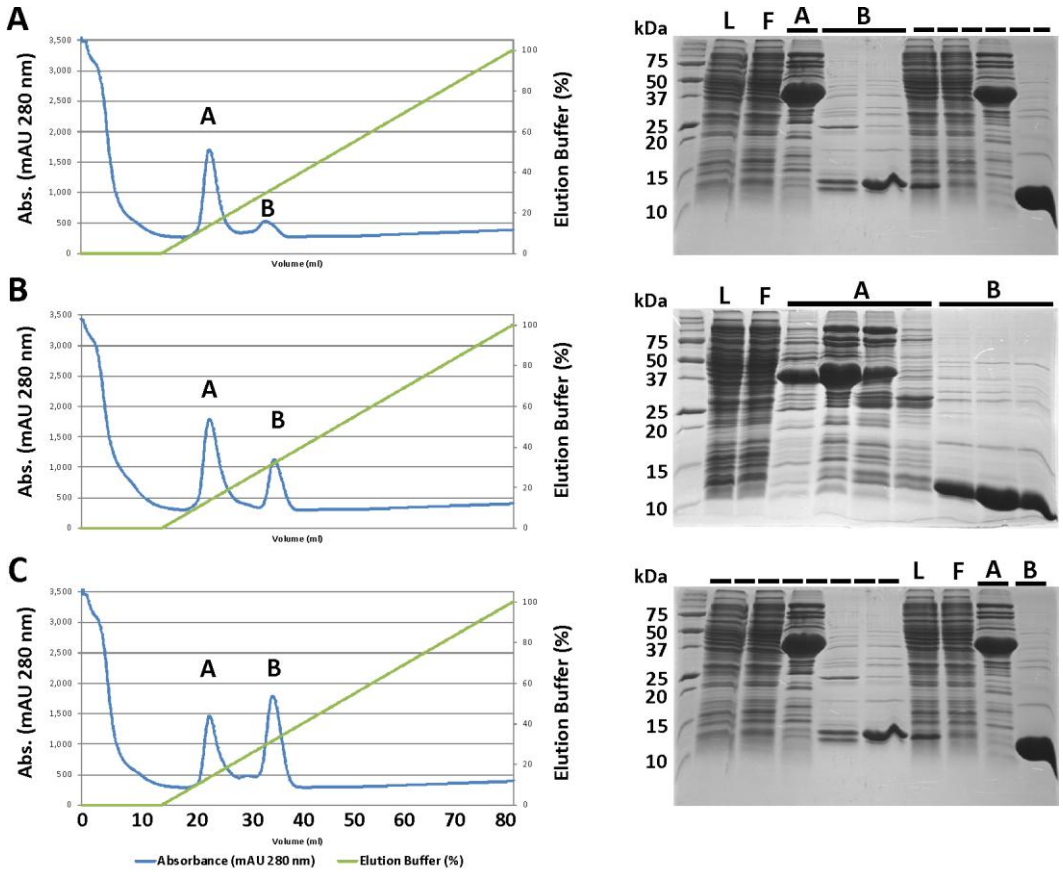


Figure 5.9. Mutation of three alternative residues to methionine resulted in varying protein yields after expression in *E. coli* and purification by IMAC. Chromatograms of the elution buffer gradient and UV absorbance profile from the purification of Nterm2 mutants I34M (A), L44M (B), L48M (C), and corresponding 16.5% SDS-PAGE gels showing sample lysate (L), column flow through (F) and fractions from peak A and B from each purification.

For the production of selenomethionine labelled protein, gene expression in a methionine auxotroph, that is, a bacterial strain that is unable to synthesize the methionine required for its growth and metabolism, is required. The *E. coli* expression strain DL41 was created by researchers for this purpose. The L48M Nterm2 methionine mutant was selected over the other mutants as it gave the

highest yield. After transformation into the DL41 strain, expression of selenomethionine protein was attempted using an auto-induction media method however this resulted in a failure to express recombinant protein. After repeating the growth assay and omitting the toxic selenomethionine ingredient in addition to sampling over a time course, it was determined that this expression method was not suitable for this protein (results not shown).

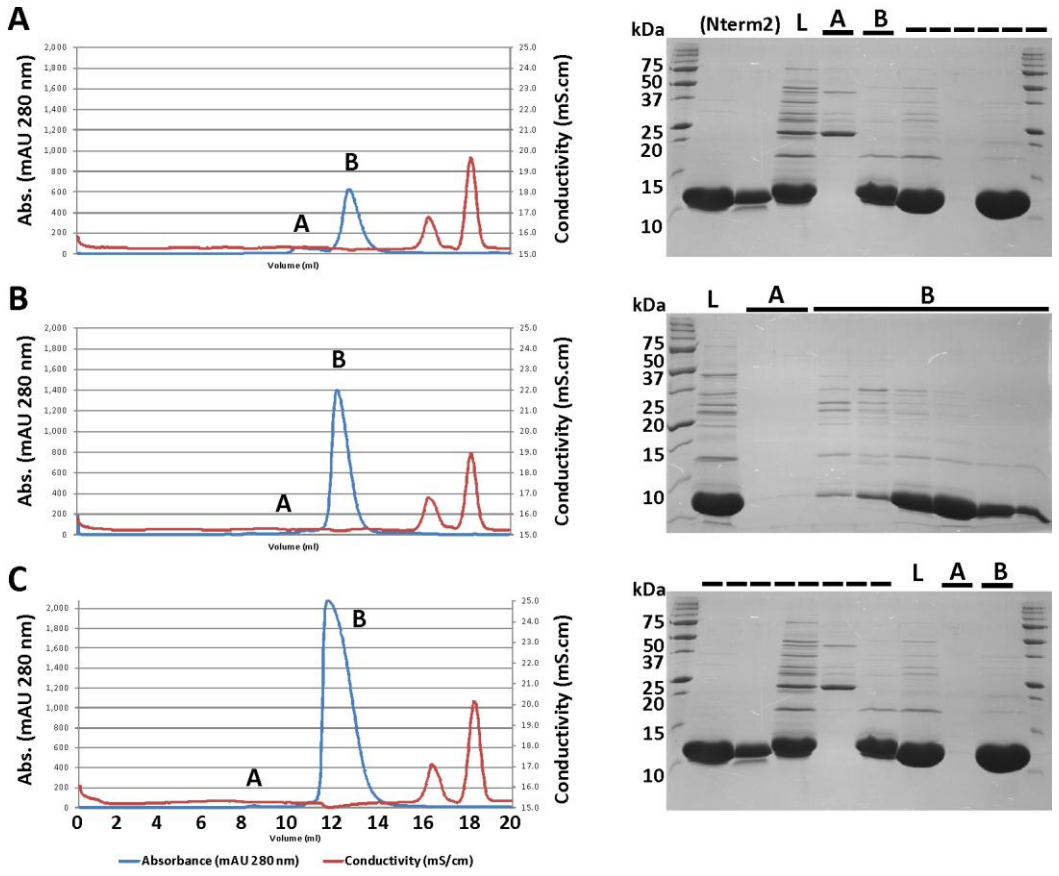


Figure 5.10. Analytical size exclusion chromatography of the three alternative methionine mutants illustrates the variation in protein yield and purification of the expected dimeric form. Chromatograms of the UV absorbance and buffer conductivity profile from the purification of Nterm2 mutants I34M (A), L44M (B), L48M (C), and corresponding 16.5% SDS-PAGE gels showing sample load (L), and fractions from peak A and B from each purification.

An alternative technique based on the inhibition of methionine biosynthesis that did not require the use of a methionine auxotroph strain was then used to express L48M Nterm2 recombinant protein. The BL21 *E. coli* expression strain and a methionine pathway inhibition method was utilised for expression of selenomethionine labelled protein. An adequate, but lower yield of L48M

Nterm2 protein compared with the native Nterm2 was achieved using this method after protein purification using IMAC (Figure 5.11A). The L48M Nterm2 mutant eluted in the second peak at approximately 40% elution buffer which corresponded to a buffer containing approximately 400 mM imidazole. The mutant protein was further purified using preparative SEC column in a low salt buffer, providing protein of a single species that appeared to show tight dimerisation under SDS-PAGE denaturing conditions (Figure 5.11B).

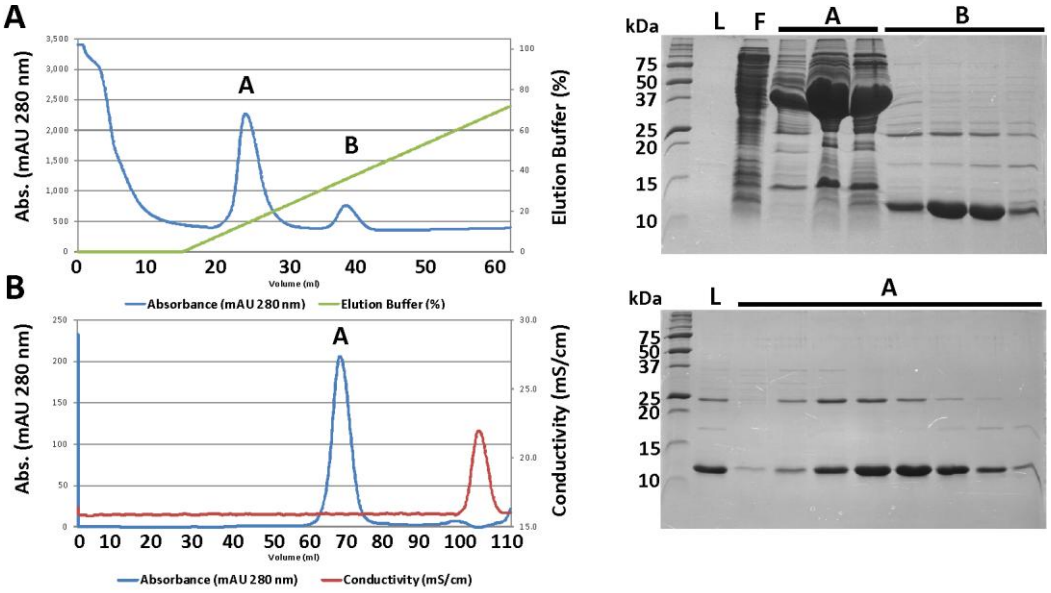


Figure 5.11. Production of selenomethionine labelled L48M Nterm2 mutant protein by using a methionine pathway inhibition method provides an adequate yield of dimeric protein. (A) Chromatogram of the UV absorbance and elution buffer profile from IMAC purification and the corresponding 16.5% SDS-PAGE gel showing sample lysate (L), column flow through (F) and fractions from peak A and B; (B) chromatogram of the UV absorbance and conductivity profile from preparative size exclusion chromatography and the corresponding 16.5% SDS-PAGE gel showing sample load (L) and fractions from peak A. Selenomethionine incorporation was confirmed by MALDI-TOF-MS analysis (results not shown).

Hanging drop crystallisation fine screens of L48M selenomethionine labelled Nterm2 were prepared based on the original native data set crystal conditions but with a range of lower pH buffers recommended for selenium labelled protein. These conditions were unsuccessful in producing crystals. Later, after trypsin co-crystallisation conditions had been optimised, hanging drop fine screens replicating the solutions that gave native crystals in the space group $P3_121$ were

prepared with selenomethionine labelled L48M Nterm2. Crystal growth was abundant in these conditions and a number of crystals grown in 0.1 M MES (pH 6.5), 0.01 M $CoCl_2$, 1.5-1.6 M $(NH_4)_2SO_4$, and 0-5% 1-4-dioxane were transported to the Australian Synchrotron (Figure 5.12). Multiwavelength data sets with resolution up to 1.63Å were collected on two occasions for the purposes of MAD structure determination however the data was twinned and was unable to be processed. On close examination, crystal surfaces were not completely smooth and were slightly pitted.

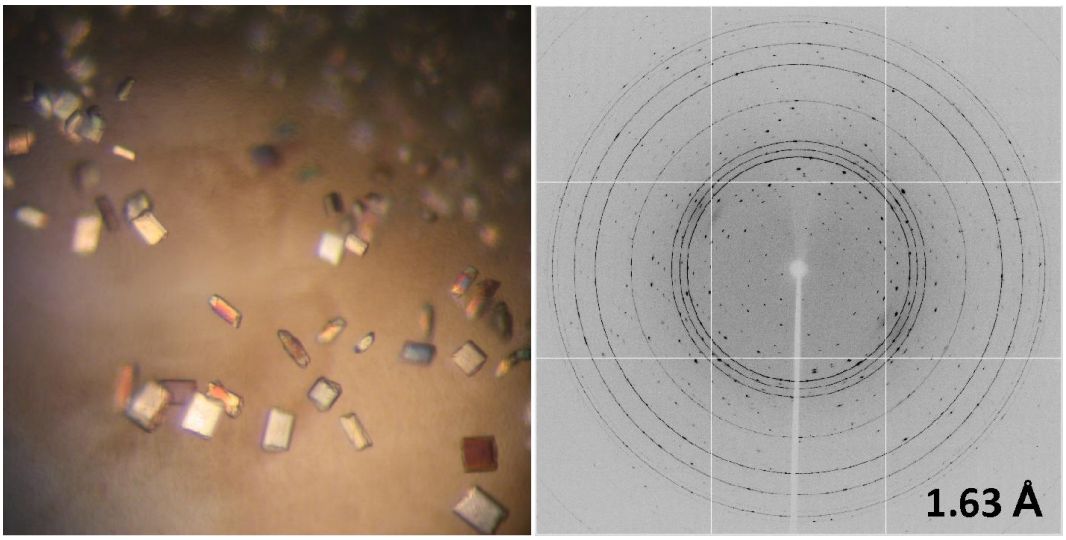


Figure 5.12. Crystals from selenomethionine labelled L48M Nterm2 mutant diffracted to a high resolution but provided data that was "twinned". Prism-like shaped crystals co-crystallised with trypsin (1:500) grown in 0.1 M MES (pH 6.5), 0.01 M $CoCl_2$, 1.5-1.6 M $(NH_4)_2SO_4$, and 0-5% 1-4-dioxane provided data sets to 1.63Å in an undetermined space group. A corresponding diffraction pattern image is displayed on the right of the crystal photo. Data sets suitable for MAD analysis were collected at the Australian Synchrotron.

5.2.5 *Ab initio* techniques solve the phase problem

Both X-ray data sets collected at the home source from "matured" Nterm2 in the $P2_1$ space group were sent to Isabel Uson at Instituto de Biologia Molecular de Barcelona to see if they were suitable for *ab initio* phasing methods. This technique is usually limited to small proteins diffracting at atomic resolution however advancement in methodology has enabled phase information to be determined for larger structures with diffraction data to 2Å resolution (Rodriguez *et al.*, 2009).

Ab initio methods successfully solved the phase problem and a structure of two short α-helix strands with β-sheet fragments composed of alanine residues was revealed (Figure 5.13A). Briefly, PHASER (McCoy *et al.*, 2007) was used with short helical poly-alanine fragments to search for molecular replacement solutions. Then for a large number of solutions, SHELXE (Sheldrick, 2010) was used to try to build longer fragments. Solutions with correlation coefficients above the noise were then identified. Initial rounds of structure building were attempted with the *ab initio* structure provided using molecular replacement techniques but the model did not refine well. Residues A28, A29, A33 and A34 were deleted from the *ab initio* structure, "freeing" the β-sheet fragment that was thought to be possibly incorrectly joined between the two α-helical strands (Figure 5.13; boxes). This modified poly-alanine structure was successfully used in repeated rounds of structure building using molecular replacement techniques and structure refinement (Figure 5.13B). A second crystal structure of the N-terminal domain of Lsr2 was determined at 2.04Å from the crystal in the hexagonal space group $P3_121$, by molecular replacement using the previously determined Nterm2 structure as a search model and its structure subsequently refined.

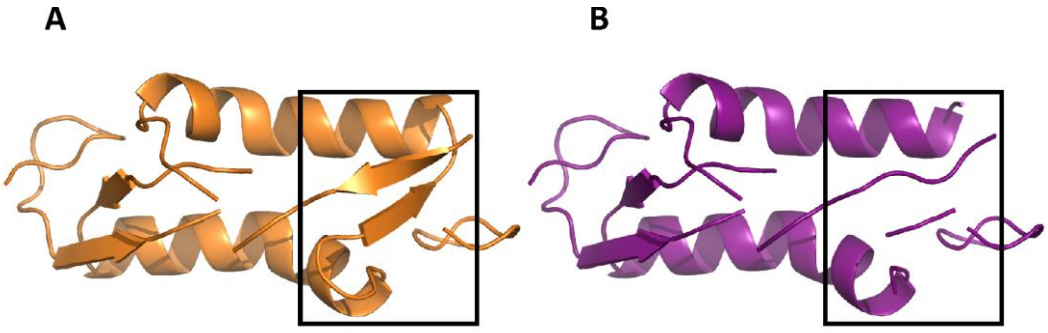


Figure 5.13. Cartoon diagrams of the *ab initio* poly-alanine structure of Nterm2 solved using two native data sets in the $P2_1$ space group. (A) The original *ab initio* solved structure that was used in initial structure building and refinement cycles; (B) the *ab initio* structure after removal of residues that were calculated to be linking the two α-helical strands inappropriately. The "trimmed" structure was used in the successful structure building and refinement of Nterm2. The boxed area illustrates the views before (A) and after (B) the removal of residues A28, A29, A33 and A34 which joined the bottom-most β-sheet to each of the α-helices. Protein depictions were generated in PYMOL.

5.2.6 Lsr2 N-terminal domain structure reveals a mechanism of oligomerisation

Each of the two crystal structures of the N-terminal dimerisation domain of Lsr2 contained two copies of the protein in the asymmetric unit forming an intimate dimer that comprises residues K4 to V58 in chain A and K4 to G59 in chain B. No interpretable electron density was observed for the N-terminal fusion tag, the N-terminal residues M1, A2, K3 or for the C-terminal residues G60 and R61. The final models have a crystallographic R-factor of 0.16 and 0.25 and R-free factors of 0.18 and 0.28, for the $P2_1$ and $P3_121$ space groups respectively. The model has excellent geometry, with all residues in the favoured region of the Ramachandran plots as defined by MOLPROBITY and average root mean square deviations (RMSD) from ideal values for bond angles of 0.914° and 0.997° and bond lengths of 0.006Å and 0.008Å for $P2_1$ and $P3_121$ space groups respectively. A summary of the refinement statistics is given in Table 5.3.

The Lsr2 dimerisation domain forms a compact homodimeric α/β structure (Figure 5.14). The monomer displays an extended N-terminus which is one strand of an anti-parallel β-sheet (the other stand provided by a symmetry related molecule), an anti-parallel β-sheet of two strands followed by an α-helix with a significant kink. The residues L48 and K49 are responsible for this kink. The secondary structure elements in this crystal structure closely agree with the predicted secondary structure of the dimerisation domain. The β-sheet is at an angle of approximately 45° to the α-helix and lies with its face flanking the N-terminal end of the helix (Figure 5.14A). The dimer is formed in a "tail-to-tail" fashion, with the anti-parallel β-sheet of chain A alongside the anti-parallel β-sheet of chain B, both on the same plane and approximately 45° to the two α-helices (Figure 5.14B).

Table 5.3. Crystallographic refinement statistics.

	Nterm2	Nterm2 + Trypsin
Refinement		
R-factor[a]	0.16 (0.19)	0.25 (0.28)
R-free	0.18 (0.25)	0.28 (0.33)
Protein Atoms	834	838
Solvent Atoms	206	135
Ramachandran Favoured	99.1	100.0
Ramachandran Outliers	0.0	0.0
Average B-value (Å^2)	13.6	36.7
RMSD		
Bond Angles (°)	0.914	0.997
Bond Lengths (Å)	0.006	0.008

[a] numbers in parentheses correspond to the highest resolution shell

The kinked C-terminus of the helix allows it to pack against the β-strands of the neighbouring molecule in the dimer. Residues that have been previously implicated in dimerisation – D28 and R45 (Chen *et al.*, 2008) - anchor the anti-parallel β-sheet of one chain to the α-helix of the other chain via a salt bridge (Figure 5.14C). The dimer interface buries 1548.2 Å^2, corresponding to 31% of the total surface area of each monomer. The C-terminal residues of each monomer point downwards and away from the β-sheets of the compact homodimer. Presumably the DNA binding domain of Lsr2 could be in a number of orientations but would be linked to the dimerisation domain via the flexible (approximately 16 amino acid residue) inter-domain region. Ten residues are well-conserved across the N-terminal domain of the Lsr2 protein family (D11, D12, F25, Y32, I34, D35, L36, L44, L48 and W51, (Figure 5.8 and Figure 5.15). These cluster in the core of the dimer and provide a large number of tertiary hydrophobic and hydrogen bonding interactions that connect the β-sheet, α-helix and a large N-terminal loop (Figure 5.15). This residue conservation in the loop region suggests that this conformation is of importance in the structure. Residues D12 and D35 point towards the solvent and are not involved in tertiary interactions in the dimer. Rather, they are critical in linking the dimers together to form oligomers (Figure 5.16D).

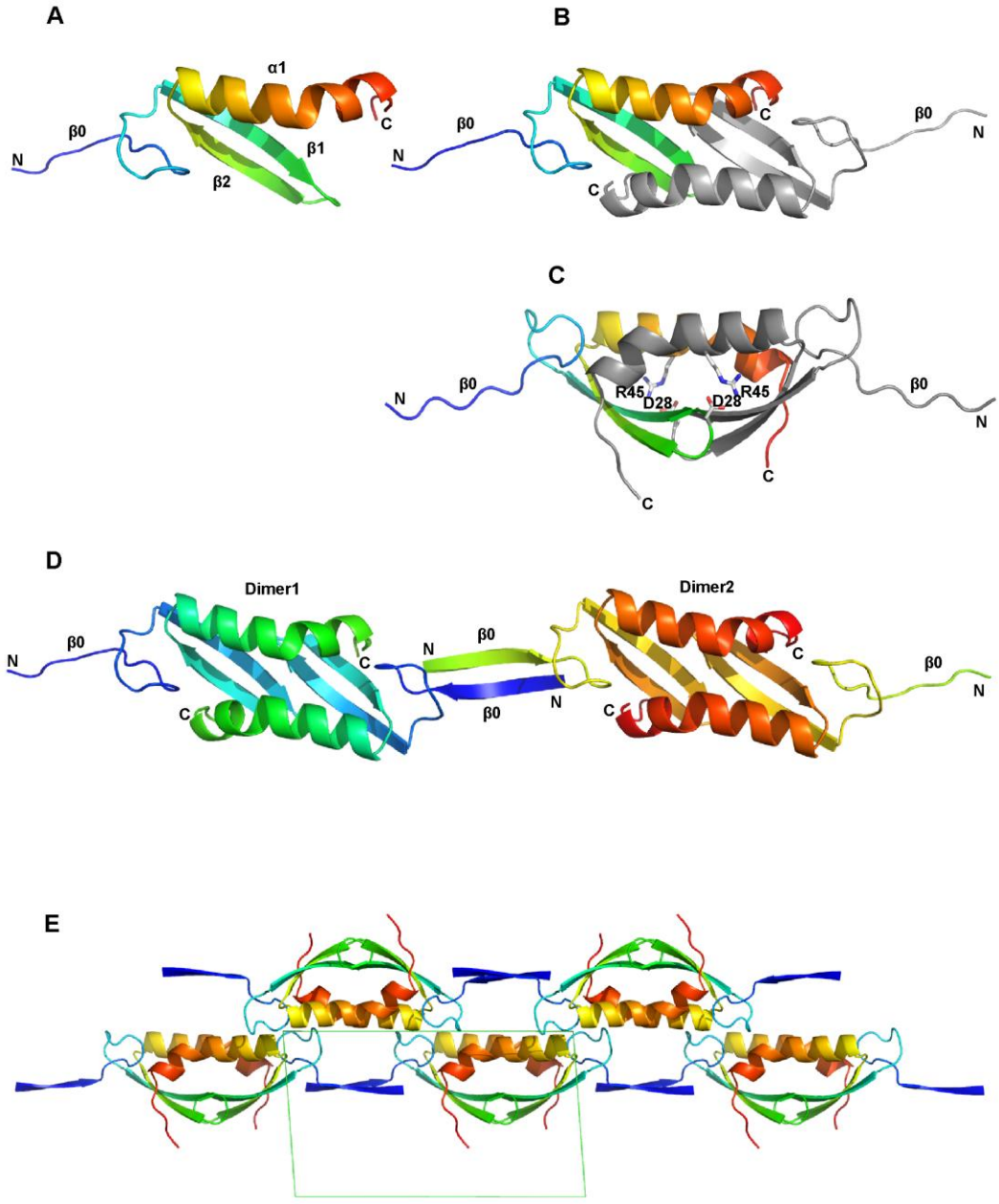


Figure 5.14. Cartoon diagrams of Lsr2 N-terminal domain in the $P2_1$ crystal structure. (A) A depiction of the monomer: the α-helix, β-sheets and chain termini are labelled and the chain is coloured blue-red (N-terminus to C-terminus); (B) the dimer as seen in the crystal structure; (C) a 90° rotation of figure B showing residues critical for dimerisation as grey, blue and red sticks; (D) Lsr2 N-terminal domain oligomerisation as generated by crystallographic symmetry. The N-terminus of one dimer becomes one strand of an anti-parallel β-sheet. The second strand is presented by a neighbouring dimer. The sheet linking two dimers is shown as β0; (E) crystallographic symmetry (in space group $P2_1$) showing the unit cell (in green) projected perpendicular to the b-axis. The two-fold screw axis generates alternating Lsr2 chains that lay back to back in the horizontal plane this view. For all figures protein depictions were rendered with PYMOL.

The primary protein interface is between the two interlocking protein monomers however there is a second interface of 380.9 $Å^2$, corresponding to 7.7% of the total surface area of each monomer. This second interface is between the extended N-terminus of one strand of a dimer and the extended N-terminus of another strand of a dimer, forming an anti-parallel β-sheet connecting the two dimers (Figure 5.14D).

The crystal structure reveals a chain of dimers formed by crystallographic symmetry (Figure 5.14D and E). In both crystallographic space groups ($P2_1$ and $P3_121$) there is two-fold symmetry (for simplicity illustrated for $P2_1$ in Figure 5.14E) that generates alternating Lsr2 chains that lie back-to-back. Each N-terminal end contributes one strand of an anti-parallel β-sheet (Figure 5.14D). This additional third β-sheet was predicted in secondary structure analysis. In both structures from "matured" and trypsin co-crystallised Nterm2, the N-terminal histidine tag and rTEV cleavage site and residues M1, A2 and K3 are missing. The fact that the structures are identical and one was the result of trypsin treatment led to the suspicion that the "matured" sample successfully crystallised due to traces of a contaminating protease.

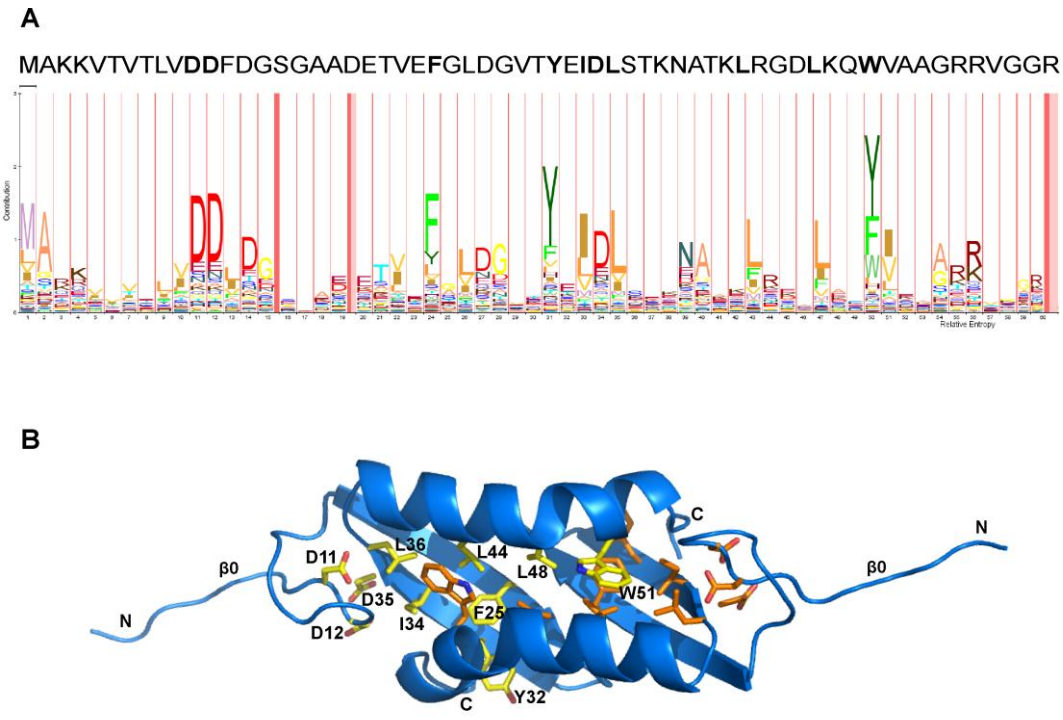


Figure 5.15. Conserved amino acids of the N-terminal domain of Lsr2 important for structural integrity. The structure is for the $P2_1$ space group. (A) Relative conservation of amino acids at different positions from a multiple sequence alignment are depicted using HMM Logo (http://pfam.sanger.ac.uk/family/PF11774.2). The height of the letters represents the relative entropy and the width represents the relative contribution of the position to the overall protein family. The pink bars represent regions of insertion in the alignment. Amino acid residue colours reflect their biological properties (red = charged; blues = polar, uncharged; yellows = aliphatic; greens = aromatic). The amino acid sequence for Lsr2 from *M. tuberculosis* is shown across the top of the HMM logo for comparison and residues that are well-conserved are in bold. (B) Cartoon diagram of Lsr2 N-terminal dimerisation domain showing conserved residues. The structure is for the $P2_1$ crystal form. Conserved residues for each chain are shown in yellow and orange respectively. Yellow residues from chain A are labelled. The chain termini are also labelled. Depiction was generated in PYMOL.

Electrostatic surface potential analysis of the crystal structure illustrates the electrostatic complementarities between dimers (Figure 5.16). The N-terminal ends of each dimer and the lysine residues along the α-helixes display a positive charge (Figure 5.16A) whereas the β-sheet side of the dimer has an overall negative charge that extends into the loop region (Figure 5.16B). In a chain of dimers formed by crystallographic symmetry, it was noticed that the positively charged N-terminal ends of each overlapping β-strand interact with the negatively charged loop regions of the adjacent dimer (Figure 5.16C).

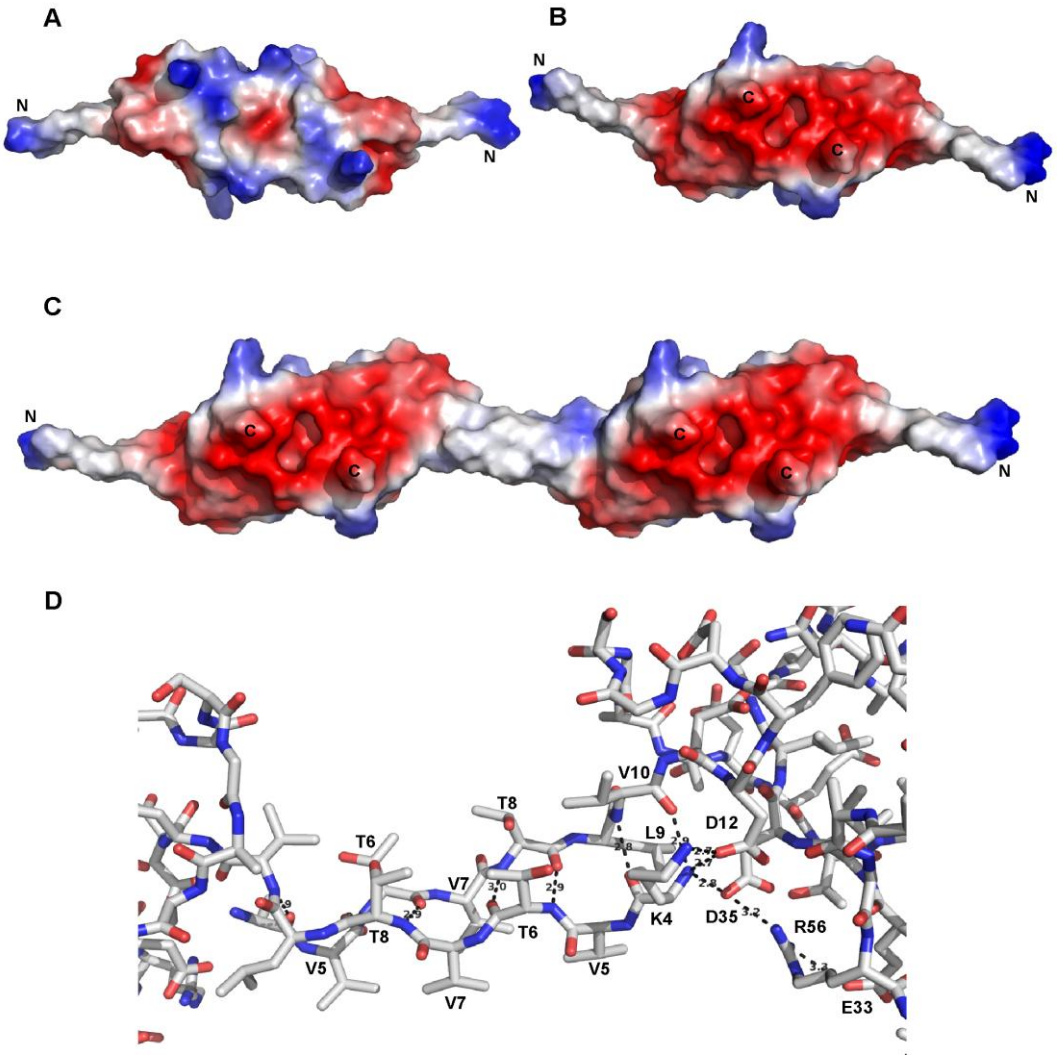


Figure 5.16. Electrostatic surface potential diagrams of Lsr2 N-terminal domain and details of residues that interact for oligomerisation. A dimer viewed from (A) the α-helices aspect and (B) the β-sheet aspect; (C) an oligomer of two dimers viewed from the β-sheet aspect; (D) an enlargement of the residues (shown as sticks) in the anti-parallel β-sheets that interact between two adjoining dimers. Interacting residues are labelled and hydrogen bonding is represented by black dashed lines. Colours indicate the surface electrostatic potential: blue = positive (or above 0.1 Volt), white = neutral and red = negative (or below 0.1 Volt). The chain termini are labelled where visible. Surface potential diagrams and protein depictions were generated in PYMOL.

Observation of the interaction between residue K4 (the N-terminus after truncation) of one dimer and those residues involved in the twisted loop conformation of the second dimer shows a number of non-covalent bonds (Figure 5.16D). The oligomer is not only stabilised by interstrand hydrogen bonds between the overlapping β-strands, but also by a network of polar interactions between K4 and its oligomeric neighbour. Positively charged amino groups of K4 form salt bridges with carbonyl groups and acidic groups from residues V10, D12 and D35 from the neighbouring dimer nestled alongside (Figure 5.16D). This structural alignment could not occur if the N-terminal residues M1, A2 and K3 were present as the polar interactions between the K4 backbone amide and V10, D12 and D35 could not form if a peptide bond were present (Figure 5.16D). Within an individual dimer, R56 from one monomer in the dimer forms a salt bridge with both E33 and D35 from the other monomer (Figure 5.16D).

The addition of trypsin in crystallisation conditions facilitated crystallisation in a different space group, $P3_121$ (Figure 5.17). The dimerisation and the linear oligomerisation of the N-terminal domain of Lsr2 is identical in this hexagonal space group when compared to the structure in the $P2_1$ crystals. The difference between the two space groups is the arrangement of the linear chains of the molecules. The linear oligomer strands lie side by side in the monoclinic crystals whereas the hexagonal crystals are made up of stacked layers of linear strands that are rotated by 120 degrees (Figure 5.14 and Figure 5.17A). This provides further evidence that the tendency to form linear chains of dimers is a result of truncation of the N-terminus and is not a crystallographic artefact.

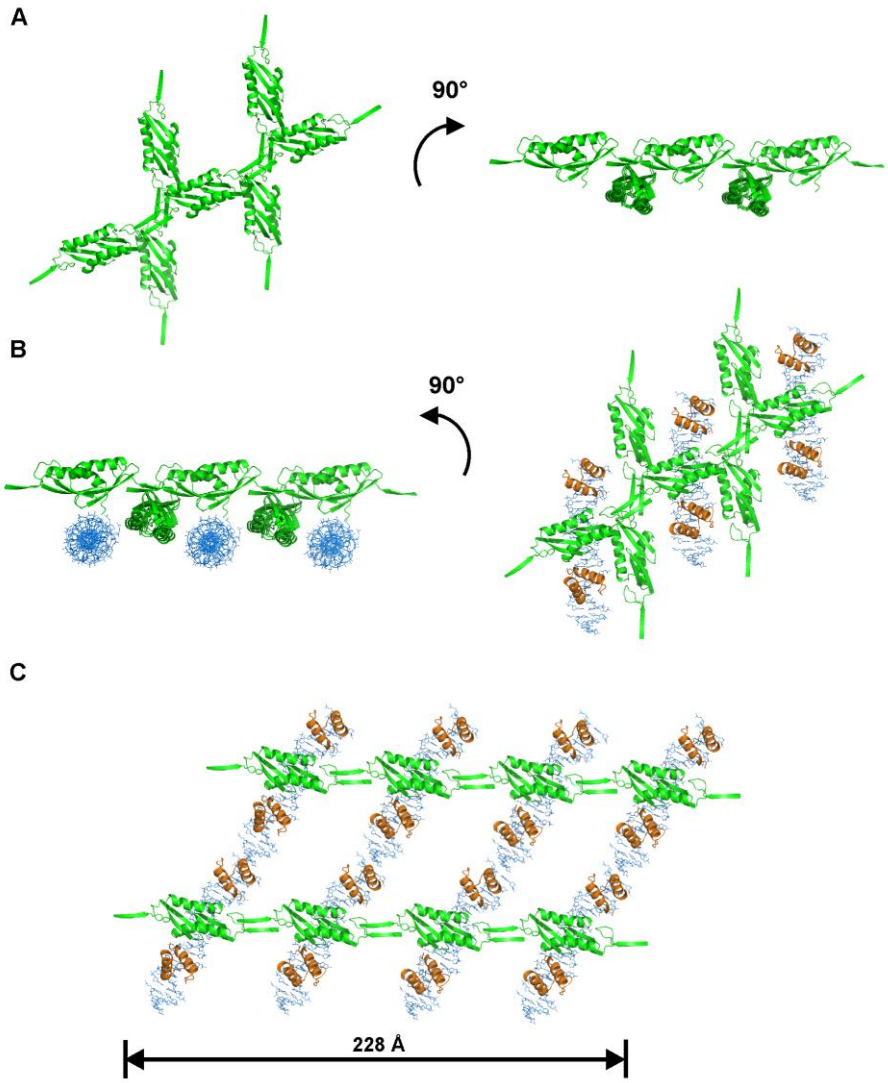


Figure 5.17. A model for the chromosomal DNA organisation by Lsr2 as suggested by hexagonal crystal packing. (A) The N-terminal Lsr2 dimerisation domain in the $P3_121$ space group in two orientations; (B) the proposed orientation of DNA (in blue) between the N-terminal chains and a 90° rotation illustrating this from above. The C-terminal DNA binding domains (in orange) have been positioned according to (Gordon *et al.*, 2010); (C) our model of full-length Lsr2 binding to DNA and forming chains, cross-linking multiple strands of DNA. The dimension at the bottom of the figure is consistent with the compact fibrils seen in TEM analysis of Lsr2 with DNA (Figure 5.21). Protein depictions were generated in PYMOL.

The crystallographic data in the more complex $P3_121$ space group reveals a possible mechanism of protein alignment and binding to DNA. Figure 5.17A and B shows the arrangement of the Lsr2 N-terminal domain in this space group, the potential positioning of DNA strands and C-terminal DNA binding domain of Lsr2. Within the constraints of crystal packing, the stands of DNA can be uniformly

positioned underneath chains of N-terminal domains and the C-terminal DNA binding domain can sit adjacent to the N-terminal domain (Figure 5.17B). Interestingly, the protein is at a 120° angle to the DNA strand which would off-set each strand of DNA to each other. Crystal packing gives us an idea of how Lsr2 may be positioned on, and bind to, the DNA and oligomerisation of the nature seen in the crystal structure shows how many DNA strands, already bound by Lsr2, may be drawn together in a very uniform manner (Figure 5.17C).

The model presented here provides a structural basis for chromosomal DNA condensation. Evidence from the $P3_121$ space group crystal packing of the N-terminal domain and TEM images of trypsin-activated Lsr2 oligomerisation on genomic DNA (Figure 5.21) shows a dynamic mechanism for DNA compaction. The modelled crystallographic dimensions of four strands of DNA in complex with the full length Lsr2 protein are consistent with the compact fibrils seen in TEM analysis of Lsr2 with DNA which is discussed further in section 5.2.7.1.

Coordinate files and structure factors for structures in both monoclinic and hexagonal crystal forms have been deposited in the Protein Data Bank with PDB codes of 4E1P and 4E1R respectively.

5.2.7 Proteolytic digestion of Nterm2 is necessary for formation of oligomeric structures

Purification of Nterm2 by both analytical and preparative SEC after IMAC showed that the N-terminal domain was a dimer (Figure 5.18A). Further SEC purification of concentrated Nterm2 protein that had previously been used in crystallography trials but had been stored at 18°C for some months revealed that much larger oligomeric forms had developed naturally over time and it was proposed that this was probably due to small amounts of contaminating proteases slowly degrading the protein (Figure 5.18B). The crystal structures of the N-terminal domain in both the native monoclinic data set and the hexagonal data set strongly suggested that the removal of the first three residues was crucial for crystal organisation and protein oligomerisation. Trypsin was added to fresh

dimeric Nterm2 protein at a 1:500 mass ratio (shown in Figure 5.18A) for 5 minutes at room temperature and purified by SEC. The results showed that the majority of the dimeric Nterm2 had now developed into larger multimers that are seen in peaks A and B (Figure 5.18C) with little of the dimer remaining (peak C). This result was contradictory to what would normally be expected to occur to protein digested using trypsin but is in agreement with previous observations of "matured" Nterm2 protein preparations.

Molecular weight analysis of equivalent Nterm2 protein preparations using MALDI-TOF-MS enabled clarification of the size of peptides present in "fresh", trypsin-treated and "matured" samples in comparison to the crystallised protein retrieved by dissolving a single crystal (Figure 5.19). The mass to charge ratio (m/z) represents the molecular weight of the protein. Although the experimental masses given using this technique were not exactly as predicted (6199.8 Da), the results nonetheless confirmed that a molecular weight species of approximately 6137 Da representing Nterm2 without the N-terminal tag, spacer and the first three residues (M1, A2, and K3) was present in "fresh", crystallised and trypsin treated preparations (Figure 5.19A-C). Peptides of approximately 3254 Da representing the entire N-terminal tag through to the first three residues were present in the "fresh" and trypsin treated samples but not in the crystalline form. A mass of 9431 Da, representing the full length Nterm2 protein was observed in the "fresh" preparation but was absent from the trypsin treated sample presumably as proteolysis had progressed further in this sample (Figure 5.19A and C). Analysis of "matured" Nterm2 revealed that it had formed oligomeric structures up to seven dimers long (Figure 5.19D). Electrospray ionisation investigation of dissolved crystals accurately predicted a molecular weight of 6198.9 Da and 6200.1 Da for monoclinic and hexagonal crystals respectively (Figure 5.19E). These figures compare closely to the calculated molecular weight of Nterm2 minus the N-terminal tag and the first three residues and are comparable with the protein species observed in MALDI-TOF-MS investigations.

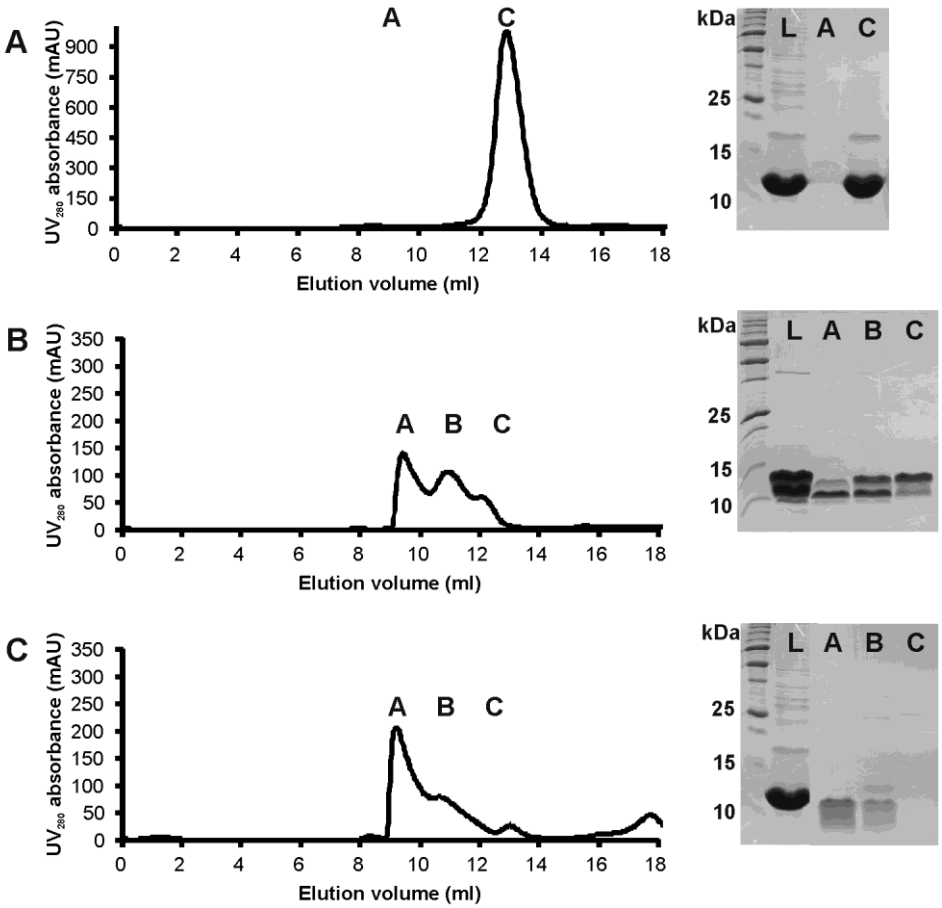


Figure 5.18. The addition of trypsin to the N-terminal domain of Lsr2 (Nterm2) facilitates oligomerisation. (A) Size exclusion chromatography of freshly purified Nterm2 showing a single dimer peak; (B) a chromatogram of the same protein construct used for crystallography (now 14 weeks old) demonstrating larger oligomeric forms in solution; (C) the addition of trypsin (1:500) for 5 minutes at room temperature to a fresh sample of Nterm2 accelerates the formation of large oligomeric species in solution. Corresponding 16.5% SDS-PAGE gels are on the right and show the sample load (L), and fractions from peaks A, B and C.

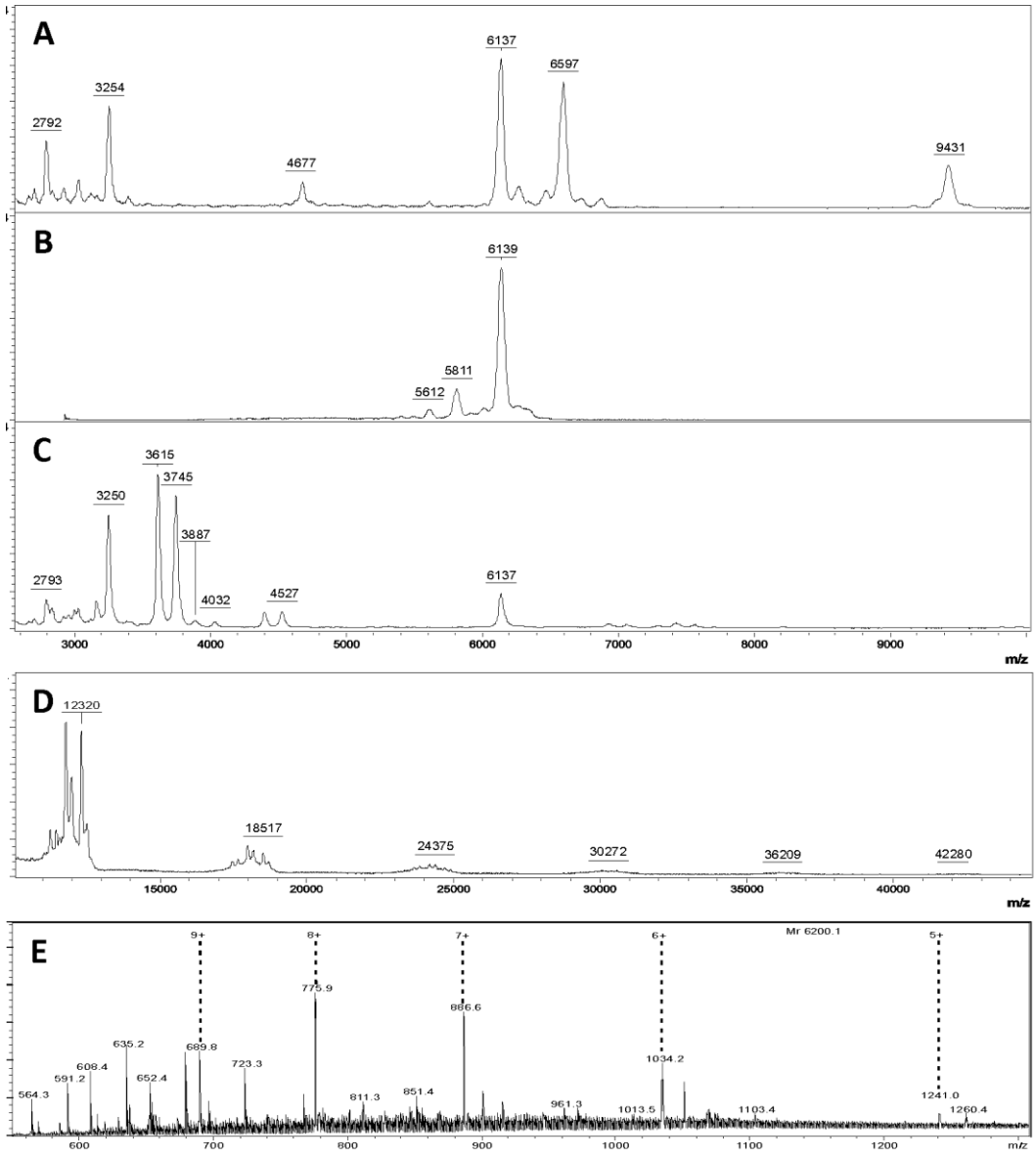


Figure 5.19. MALDI-TOF-MS analysis of fresh and trypsin treated Nterm2 protein illustrates that species of the same size as those from dissolved crystals are seen across all samples. (A) In a relatively fresh (4 week old) Nterm2 protein preparation several peaks corresponding to the full-length protein (9431 m/z), the protein with the histidine tag plus the first three residues (MAK) removed (6137 m/z) and the tag itself (3254 m/z) are seen; (B) a single dissolved crystal from identical crystallisation conditions as the $P2_1$ space group crystal structure shows that almost the entire sample is a species with the histidine tag plus the first three residues (MAK) removed (6139 m/z); (C) the addition of trypsin (1:500) accelerates the removal of the histidine tag and first three residues (3250 m/z), leaving the smaller 6137 m/z species; (D) "matured" Nterm2 (3 months old) illustrates the oligomerisation potential of the N-terminal dimerisation domain once the tag and first three residues are absent. Mass to charge ratio peaks show dimers (12320 m/z), trimers (18517 m/z), tetramers (24375 m/z), pentamers (30272 m/z), hexamers (36209 m/z) and heptamers (42280 m/z); (E) electrospray ionisation investigation of dissolved hexagonal crystals gives an accurate molecular weight of 6200.1 Da the of the protein species observed in MALDI-TOF-MS investigations.

5.2.7.1 Lsr2 cleavage by trypsin allows DNA restructuring, compaction and protection

The crystal structure of the dimerisation domain of Lsr2 revealed an interesting mechanism of oligomerisation but it would be more relevant to see if the full length protein was also capable of forming long chains. The effect of N-terminal proteolytic processing and oligomerisation on full length Lsr2 bound to DNA was investigated. Full length Lsr2 that is expressed in *E. coli* and purified using IMAC chromatography, results in a preparation of Lsr2 bound to *E. coli* genomic DNA (Colangeli *et al.*, 2009). When this sample was treated with DNAse for 1 hour, the majority of the DNA was digested and Lsr2 provides little or no protection (Figure 5.20, lanes 1 and 2). However, proteolytic treatment of Lsr2 using trypsin in small amounts (1:500 mass ratio) for increasing periods of time, followed by DNase treatment, resulted in DNA condensation, providing increasing levels of protection of DNA against DNAse degradation (Figure 5.20, lanes 3-12). Intriguingly, processed Lsr2 appeared to protect stretches of DNA of a relatively uniform small size and this may hint at localised zippering of DNA by Lsr2 as opposed to long-range crosslinking of the DNA. SDS PAGE analysis of equivalent samples showed that prior to proteolytic treatment, Lsr2 was relatively versatile and ran as a monomer, a dimer and higher oligomers on the gel. However, after processing, Lsr2 ran exclusively as a dimer and higher order oligomers despite the denaturing challenge of SDS and heating prior to running the gel (Figure 5.20). Full length dimeric Lsr2 (from Figure 5.20, lane 1) was analysed by MALDI-TOF-MS peptide mass analysis and confirmed the presence of a dimeric form of Lsr2 (results not shown).

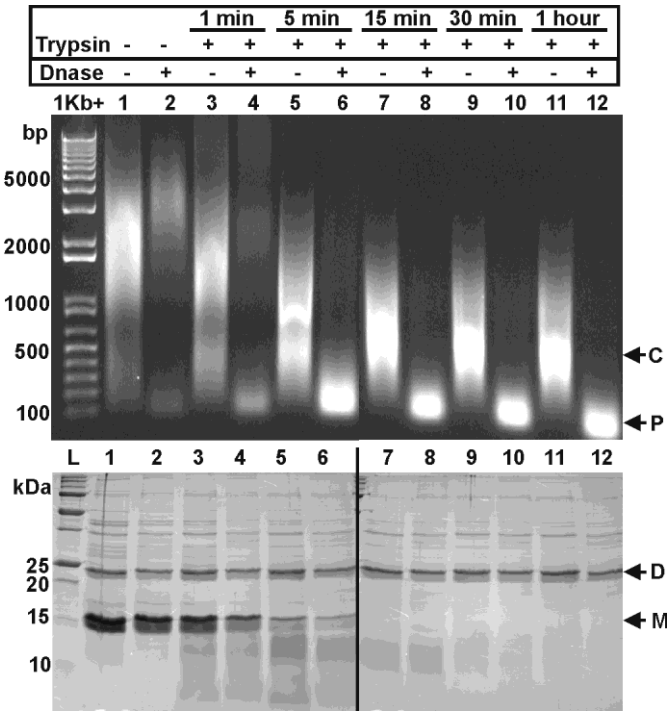


Figure 5.20. Trypsin digestion of full length Lsr2 co-purified with *E. coli* genomic DNA shows DNA compaction and DNase protection. Trypsin (1:500) initiates the compaction of co-purified DNA by Lsr2 over 1 hour as visualised on a 1% agarose gel (SYBR® safe DNA stain). Subsequent digestion of the compacted DNA using DNase shows protection of the DNA by Lsr2. Equivalent samples were run on an SDS-PAGE gel to show the presence of Lsr2 (both monomer and dimer forms) in the sample after trypsin digestion. Components are labelled: C = condensation of DNA upon treatment with trypsin; P = protection of DNA; D = Lsr2 dimer; M = Lsr2 monomer.

The effect of trypsin-induced oligomerisation of Lsr2 on DNA topology mentioned above was visualized using negative-stain TEM (Figure 5.21). Previous experimental evidence using TEM showed that IMAC purified Lsr2 decorated the length of DNA strands (Colangeli *et al.*, 2009). Trypsin was combined with IMAC purified Lsr2 and sampled over the same time course as above. It was hypothesised that trypsin would cleave residues suitable to promote oligomerisation and would allow a change in protein organisation and hence DNA packing. The effect of trypsin on the C-terminal histidine tag was unknown. Prior to proteolytic treatment, Lsr2/DNA strands were visible with diameters ranging between 40Å and 120Å (Figure 5.21A). This was consistent with single lengths of double stranded DNA coated with Lsr2 dimers and in some cases, pairs of strands lying side-by-side. After treatment, larger, well ordered fibres (with diameters of ~200Å) were present along with condensed, knotted bundles of Lsr2/DNA strands (Figure 5.21B, arrows A and B, respectively). A change in

Lsr2/DNA strand structure was evident after only 1 minute, and larger ordered fibres began to appear after 5 minutes of trypsin treatment (results not shown). The diameter of the fibrils suggests higher-order structuring of DNA. The condensed knotted bundles became so dense after 1 hour that defocusing and image capture was challenging due to the uneven topology (results not shown).

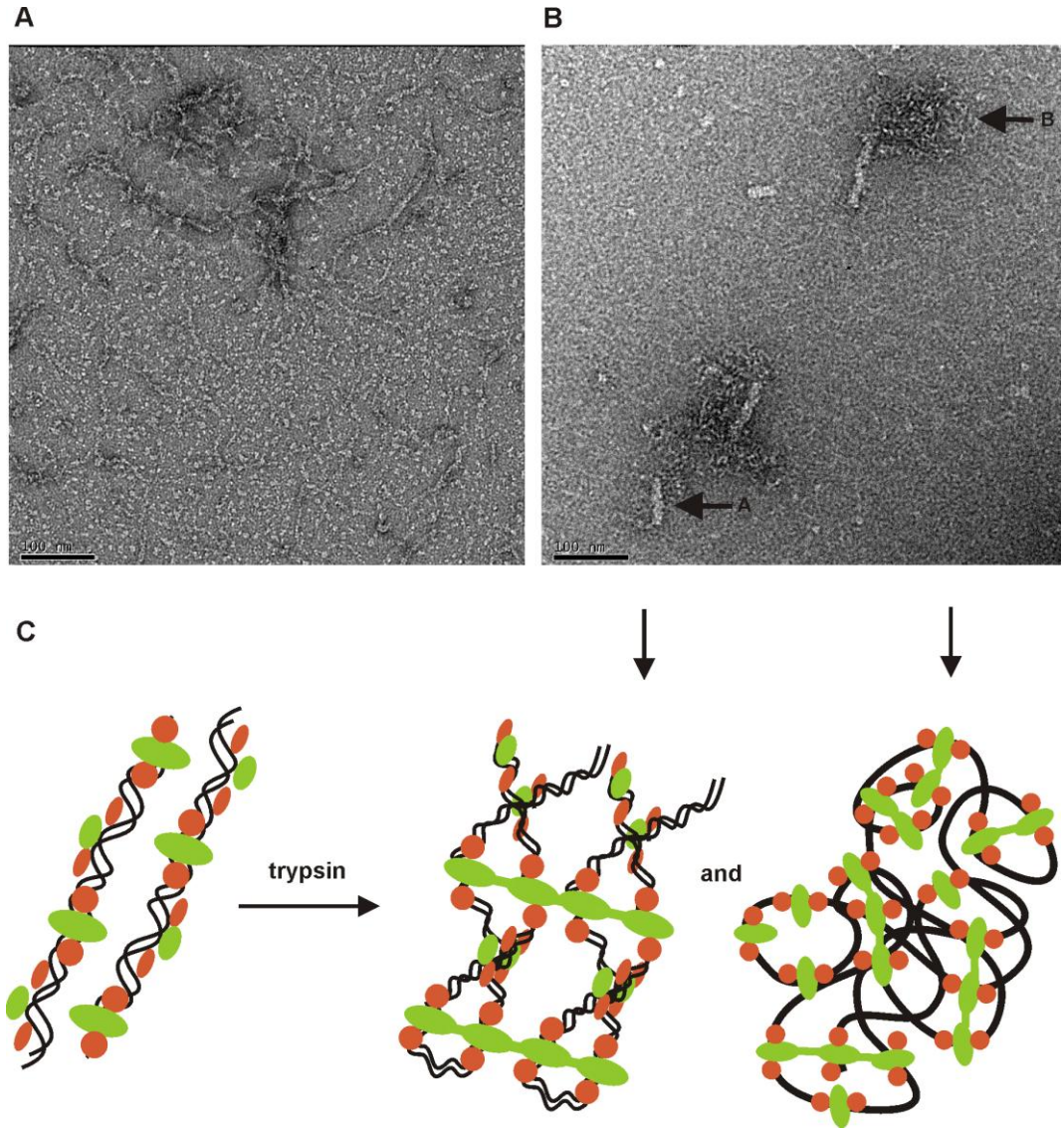


Figure 5.21. Negatively stained Lsr2/DNA complexes visualised by transmission electron microscopy show obvious morphological changes upon trypsin treatment. (A) Lsr2 co-purified with *E. coli* genomic DNA; (B) Lsr2 after trypsin digestion (1:500 ratio) for 30 minutes; Arrows A and B point to compacted structures commonly seen during the trypsin digestion time series (scale bars = 100 nm); (C) a cartoon representation of protein oligomerisation and DNA compaction after trypsin digestion. Lsr2 N-terminal dimerisation domain is shown in green and C-terminal DNA binding domain in orange. The downward pointing arrows correlate the features in (B) with their cartoon representations in (C).

5.3 Discussion

Lsr2 from *M. tuberculosis* H37Rv is a DNA binding protein that likely influences gene regulation and the organisation of bacterial chromatin by binding to DNA and indeed it has been shown to be able to crosslink double stranded DNA *in vitro* (Chen *et al.*, 2008). Lsr2's relationship to the *E. coli* nucleoid-associated protein H-NS discovered through domain swapping and *in vitro* complementation experiments also suggests a role in chromosome organisation (Gordon *et al.*, 2010). Although Lsr2 and H-NS share similarities in domain organisation and function, they share less than 20% amino acid sequence identity and they have different predicted secondary structure elements (Gordon *et al.*, 2008). The 3D structure of the DNA binding domain of Lsr2 has recently been solved by NMR methods and revealed that the structure is unique and distinct from that of H-NS. A model of DNA binding of Lsr2 calculated that Lsr2 can bind DNA in two orientations by grabbing either edge of the minor groove like a clamp (Gordon *et al.*, 2010). Whilst the mode of DNA binding by Lsr2 (via its C-terminal domain) is well characterized, the dimerisation and oligomerisation mechanisms have not been established. This aspect of Lsr2 function is thought to be important with respect to the mechanism of DNA crosslinking, chromosome organization and gene regulation.

In this study, the mechanism of dimerisation and oligomerisation of Lsr2 was established by focusing on determining the 3D structure of the N-terminal domain. Earlier attempts to crystallise the full-length Lsr2 were unsuccessful. Lsr2 co-purifies with *E. coli* genomic DNA and whist the DNA can be dissociated by buffer exchange into high salt buffer and removed via SEC, the protein is not completely free of DNA and is prone to binding to other material afterwards. Also, the disordered and flexible region between the two domains was predicted to make protein crystallography difficult due to protein flexibility. Seven N-terminal domain constructs differing in length and including a range of residues reaching into the unstructured inter-domain region were expressed and purified. From these, three variants reached crystallisation trials and Nterm2 successfully grew diffraction quality crystals that yielded a 1.73Å resolution

dataset. Crystals grew only after many weeks and at that time, required an unknown "maturation" process for crystallization to proceed. Nevertheless, these crystals provided high quality X-ray diffraction data that were suitable for structure determination via newly developed *ab initio* phasing techniques (Rodriguez *et al.*, 2009). The Nterm2 structure is a dimer that has no homologues in the structural databases and is unlike the oligomerisation domain of H-NS from *E. coli* despite its previously identified functional similarities. The structure provides a rationalisation for the sequence conservation seen for this domain – the majority of the conserved residues form tertiary interactions in the structure.

The structure showed linear chains of Lsr2 N-terminal domain oligomers that were the result of crystal packing. The chain "links" were constructed by the tight association of a β-strand from each of the N-termini of neighbouring dimers. These links can only form if the N-terminus of the domain is truncated by three residues, and this provided a rationale for the "maturation" required for crystallization. It also presented a hypothetical mechanism for protease-driven oligomerisation of Lsr2. Artificial oligomerisation was promoted using trypsin and a second crystallographic structure was determined in a new hexagonal space group. This also showed linear chains of Lsr2 N-terminal domain oligomers very similar to those seen in the monoclinic crystals. However the linear oligomer strands were packed in the crystal in a hexagonal format as compared to the side-by-side format of the monoclinic crystals.

Lsr2 oligomers were not only seen in the crystals of the N-terminal domain but also in solution for both purified Nterm2 and full-length Lsr2 protein. The IMAC purified full-length Lsr2 protein, which co-purifies with DNA, oligomerised on treatment with trypsin. This oligomerisation increasingly protects DNA from DNAse degradation. Further, DNA compaction and crosslinking was induced by trypsin and could be visualized using electron microscopy. Based on the TEM images, the hexagonal crystal lattice structure for Nterm2 was used to speculatively build a model for DNA organisation by full-length Lsr2.

The hexagonal crystals of the Lsr2 N-terminal domain had a higher solvent content when compared to the monoclinic crystals. Pores running through the hexagonal crystal lattice were of the appropriate dimensions to accommodate long chains of double stranded DNA and adjacent spaces were appropriate to accommodate the Lsr2 C-terminal DNA binding domains and a model was developed accordingly. The proposed model (Figure 5.17), is consistent with the dimensions of the fibres observed under TEM and the mode of DNA binding predicted previously (Gordon *et al.*, 2010).

In summary, this study revealed the first crystal structure of the N-terminal dimerisation domain of Lsr2 from *M. tuberculosis* H37Rv. The arrangement of Lsr2 N-terminal domain oligomers differed from other known oligo-forming DNA binding proteins such as *E. coli* H-NS (Arold *et al.*, 2010). Oligomerisation was dependent upon proteolytic processing of the N-terminus of Lsr2 and both protein oligomerisation and DNA compaction could be induced by treating Lsr2 with low concentrations of the protease typsin. There are many trypsin-like proteases encoded in the genome of *M. tuberculosis* H37Rv which may indicate that proteolytic processing of proteins to cause oligomerisation could occur. This mechanism may help to explain how Lsr2 is involved in regulating the expression of a large number of genes involved in cellular processes.

Chapter Six – Pellicle and biofilm formation in *M. smegmatis* mc²155 Δ*lsr2* and Lsr2 mutant strains

6.1 Introduction

Bacteria that form biofilms are abundant and present in most environments. Biofilm formation by the bacterial pathogen *Pseudomonas aeruginosa* has been the most extensively studied individual species but other gram-negative bacteria such as *Pseudomonas fluorescens* and *E. coli* have also been studied in detail (O'Toole *et al.*, 2000; O'Toole & Kolter, 1998). It is unclear if the human pathogen *M. tuberculosis* is capable of forming a biofilm (Ha *et al.*, 2005) however it is well established that *M. smegmatis* and other environmental bacteria such as *M. avium* are capable of producing a biofilm (Carter *et al.*, 2003; Ojha & Hatfull, 2007).

Bacteria growing in a biofilm are noted to be more resistant to antibacterial compounds than their planktonic counterparts and this is a trait established in *M. smegmatis* mc²155 which is able grow within its biofilm unaffected by the antibiotic isoniazid at concentrations that are usually inhibitory (Teng & Dick, 2003). In other studies investigating isoniazid resistance in planktonic *M. tuberculosis* H37Rv, drug induced specific gene upregulation was responsible for conferring antibiotic resistance of the organism (Alland *et al.*, 2000; Colangeli *et al.*, 2005). Continued investigation into the regulation of these antibiotic resistance genes identified the DNA binding protein Lsr2 which was capable of repressing these specific genes (Colangeli *et al.*, 2007).

Lsr2 has also been shown to play an important role in colony morphology and biofilm formation in *M. smegmatis* mc²155 (Chen *et al.*, 2006). When grown on either mycobacteria-specific (7H10/11 agar) or rich (LB agar) media, Δ*lsr2* mutant colonies were smooth, wet and round which was in contrast to the dry, rough and rugose morphology of the wild type parental strain (Arora *et al.*, 2008; Chen *et al.*, 2006; Colangeli *et al.*, 2007; Kocíncová *et al.*, 2008). These Lsr2 mutants had differing abilities to form biofilms with some researchers observing a

reduction or lack of biofilm formation (Arora *et al.*, 2008; Chen *et al.*, 2006) whilst others saw an increase in biofilm formation (Kocíncová *et al.*, 2008). Cell aggregation studies in liquid cultures (without the addition of the detergent Tween 80) also showed contrasting behaviour of Δ*lsr2* mutants. In one study, the growth of Δ*lsr2* mutants in liquid media resulted in increased cell aggregation (Arora *et al.*, 2008) but in another there was less cell aggregation compared with the original *M. smegmatis* mc^2155 strain (Kocíncová *et al.*, 2008). These contradictory morphological results suggested that further research was warranted in this area.

A biofilm may be defined in simplistic terms as a community of microorganisms that are attached to a surface (O'Toole *et al.*, 2000) or more comprehensively as a bacterial population embedded within a matrix which adheres to each other and to surfaces and interfaces (Costerton *et al.*, 1995). The formation of these bacterial communities depends upon the creation of a matrix which is self-produced (Costerton *et al.*, 1999) by the synthesis of extracellular substances (Branda *et al.*, 2005). Matrices are diverse in their construction and their composition varies depending on the local environmental conditions (Branda *et al.*, 2005). A complex mixture of extracellular polysaccharides, protein and DNA produced by biofilm bacteria form the matrix which holds the biofilm together (Costerton *et al.*, 1995; Whitchurch *et al.*, 2002). Usually the addition of a detergent such as Tween 80 is required for growth of mycobacteria as a uniform cell suspension however Tween is omitted for biofilm formation to allow necessary cell-cell interaction (Arora *et al.*, 2008).

Studies on the effect of Lsr2 in *M. smegmatis* biofilm formation typically focus on surface-adhered bacterial populations however air-liquid floating biofilms or pellicles represent another type of biofilm (Branda *et al.*, 2005). There appears to be a lack of convention in using the term biofilm or pellicle to represent surface-adhered or air-liquid biofilms respectively. Three publications to my knowledge have investigated the role of Lsr2 in pellicle formation however only one clearly used the term pellicle which leads to confusion when comparing and contrasting experimental results from a variety of researchers (Arora *et al.*, 2008;

Chen *et al.*, 2006; Nguyen *et al.*, 2010). Lsr2 mutants in the three aforementioned studies were defective in pellicle formation. Results presented in this chapter from the examination of biofilm and pellicle formation in *M. smegmatis* mc^2155 attempt to establish a convention of clearly distinguishing between the development of surface adhered biofilm and air-liquid pellicle formation.

Examination of biofilm formation ability in *M. smegmatis* mc^2155 is a common research technique used for assessing the impact of gene deletions and transposon mutagenesis in this bacterial species (Klepp *et al.*, 2012; Recht *et al.*, 2000; Röse *et al.*, 2004). It is also used in studies to characterise mycobacteria species and strains (Esteban *et al.*, 2008). Examination of the methodology used in the study of *M. smegmatis* mc^2155 biofilm formation reveals some similarities and notable differences. The use of a M63 minimal media supplemented with amino acids, a carbon source (glucose, glycerol or dextrose), $MgSO_4$ and $CaCl_2$ was routine in all but one examined research method, which used Middlebrook 7H9 media instead (Esteban *et al.*, 2008). The study that used 7H9 media also differed in its use of basic fuchsine to stain cells for visual biofilm quantification whereas all other research methods used a crystal violet biofilm staining technique and quantified by stain extraction and absorbance measurement (Recht *et al.*, 2000).

A point of difference in culture technique was in the use of PVC or polystyrene plastic dishes and the dish well size. Three publications examined biofilm formation in PVC 96-well plates (Recht & Kolter, 2001; Recht *et al.*, 2000; Teng & Dick, 2003), whereas five used polystyrene plastic-ware (Esteban *et al.*, 2008; Gopalaswamy *et al.*, 2008; Ojha *et al.*, 2005; Ojha & Hatfull, 2007; Röse *et al.*, 2004) and one did not mention what was used (Klepp *et al.*, 2012). Furthermore, of those studies that used polystyrene plastic the well size varied between 60 mm dishes, 12-well, 24-well and tissue culture treated 96-well plates. The incubation temperature used in these studies also varied between room temperature, 30°C and 37°C. Another point of difference in these biofilm studies was that the researchers that used 60 mm polystyrene dishes were only

interested in pellicle formation (and did not perform biofilm staining) and as previously mentioned, they confusingly used the term biofilm instead of pellicle (Ojha *et al.*, 2005; Ojha & Hatfull, 2007). Overall, after reviewing research on biofilm formation in *M. smegmatis* mc^{2}155, it was concluded that it was generally difficult to decipher if investigators were presenting pellicle or biofilm formation results and if their biofilm staining results were solely due to the surface-adhered biofilm or included the pellicle and biofilm together.

This chapter focuses on the exploration of the *M. smegmatis* mc^{2}155 Δ*lsr2* mutant phenotype, establishing conditions suitable for pellicle and biofilm growth and the development of assays for quantifying pellicle and biofilm formation. The rationale for this work was that once a phenotype was clearly established in the Δ*lsr2* mutant, the role of Lsr2 in pellicle and biofilm formation could be investigated. If the mutant phenotype could be rescued by the introduction of *M. tuberculosis* H37Rv Lsr2 on a vector, then the role of specific DNA binding residues in the C-terminal domain could be investigated by the introduction of Lsr2 DNA binding mutants into *M. smegmatis* mc^{2}155, and their effect on pellicle and biofilm formation examined. Three Lsr2 DNA binding mutants were created and characterised earlier which showed deficiencies in DNA binding (Chapter Four). These Lsr2 mutants in addition to native Lsr2 were conditionally expressed using a tetracycline inducible mycobacteria vector pMIND (Blokpoel *et al.*, 2005) and their ability to rescue the Lsr2 mutant phenotype examined.

Work presented in this chapter demonstrates that a limited phenotype for the *M. smegmatis* mc^{2}155 Δ*lsr2* mutant was established and a technique to accurately examine both pellicle and biofilm growth was developed. Analysis of pellicle and biofilm formation between wild type and Lsr2 mutant strains demonstrated only qualitative and not quantitative differences. The expression of native Lsr2 and Lsr2 DNA binding mutants had no effect on pellicle and biofilm formation.

6.2 Results

6.2.1 *M. smegmatis* Δ*lsr2* shows similar morphology to wild type

M. smegmatis mc^{2}155 containing an unmarked chromosomal deletion of the *lsr2* gene (MSMEG_6092) was supplied by R. Colangeli (Colangeli *et al.*, 2007). The Δ*lsr2* mutant colony morphology was similar to the *M. smegmatis* mc^{2}155 lab strain when grown with both Middlebrook and LBT media. The Δ*lsr2* mutant grown on Middlebrook 7H10 agar exhibited cream coloured, irregular, raised, dull, rough and dry colony morphology and in Middlebrook 7H9 media formed white aggregates (results not shown). Grown on LBT agar, the Δ*lsr2* was more yellow in colour and colonies were raised, irregular, rough but slightly shiny (Figure 6.2A) and in LBT media formed yellow aggregates. In this study the Δ*lsr2* mutant did not exhibit a smooth mucoid colony type nor have a lack of cell aggregation as reported by others. In shaking LB media cultures (grown overnight for seeding biofilm assays) both the lab strain and Δ*lsr2* begin growth in aggregate forms but grew to identical optical densities (Figure 6.2B). In a trial growth curve assay performed in supplemented M63 biofilm media, the Δ*lsr2* showed cell aggregation compared with the lab strain (Figure 6.2C) but had no observed reduction in growth. Overall, the Δ*lsr2* strain appeared and behaved in a similar manner to the lab strain. The Δ*lsr2* strain was checked for the deletion of the *lsr2* gene by PCR analysis which confirmed that the gene was indeed absent (Figure 6.1).

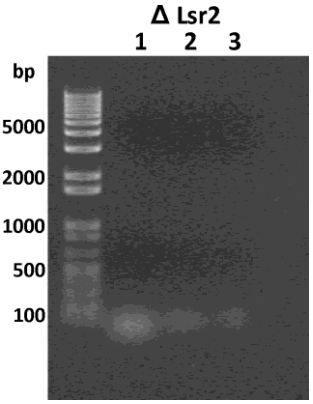


Figure 6.1. PCR analysis of *M. smegmatis* mc^{2}155 Δ*lsr2* strain. Primers were designed for *M. smegmatis Lsr2* (MSMEG_6092; Lane 1), *M. tuberculosis Lsr2* (Rv3597c; lane 2) and pMIND multiple cloning site (lane 3). Control PCR products are shown in Figure 6.7 where this image was excised from.

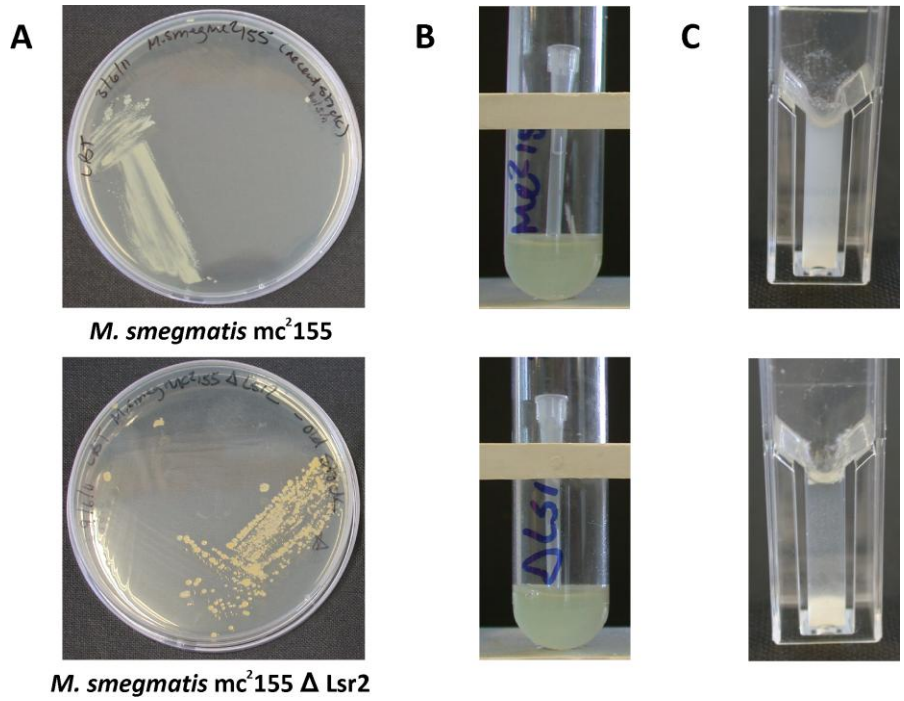


Figure 6.2. *M. smegmatis* Δ*lsr2* mutant did not show mucoid colony appearance and a lack of cell aggregation as reported by others. (A) Colonies on LBT agar were more yellow in colour compared to the lab strain but similar in their raised, irregular and rough appearance; (B) the Δ*lsr2* strain showed identical growth to the lab strain in an overnight LB media culture; (C) in M63 biofilm media the Δ*lsr2* showed greater cell aggregation compared with the lab strain however growth was unaffected.

6.2.2 Culture method and media affects pellicle and biofilm formation

For the optimisation of biofilm formation a number of media were extensively trialled; LB medium, auto-induction media (Studier, 2005), Hartman's de Bont, M63 minimal media and finally M63 supplemented with 0.2% glycerol, 0.5% bacto-tryptone, 1 mM $MgSO_4$ and 0.7% mM $CaCl_2$ based on published media (Chen *et al.*, 2006). Biofilm assays were also optimised by trials in tissue culture flasks, 60 mm polystyrene dishes and a number of multi-well plates. Tissue culture flasks and 60 mm polystyrene dishes failed to consistently form an obvious pellicle or biofilm (results not shown). Tissue culture treated 6-well dishes showed development of both pellicle and biofilm structures (Figure 6.3A) and some difference between the lab and mutant strain were observed. Tissue culture treated 24-well dishes were trialled to improve throughput and efficiency. Pellicle and biofilm structures were observed in 24-well dishes however the assay error within replicates was much higher due to the practicality of using a

smaller well size (Figure 6.3C). An intermediate sized non tissue-culture treated 12-well dish showed striking differences in pellicle morphology compared with the tissue culture treated 6- and 24-well dishes (Figure 6.3B). It was decided that the pellicle and biofilm assays would proceed using 6-well tissue culture dishes due to their ease of use and their accurate and consistent assay results.

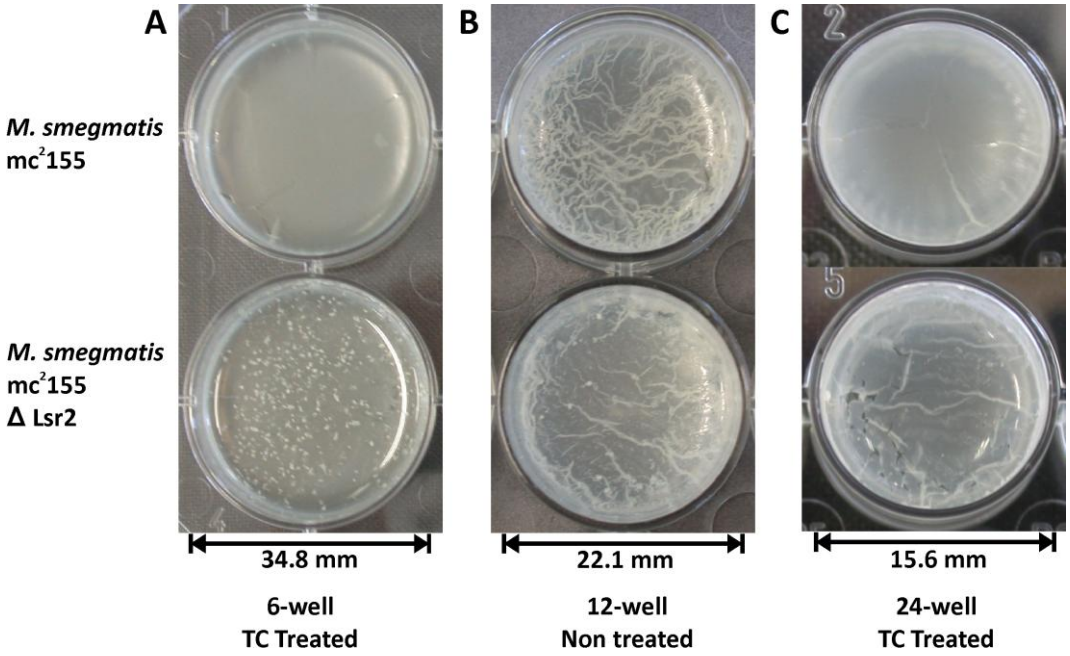


Figure 6.3. Pellicle and biofilm formation assays in a variety of tissue culture treated and non treated multi-well plates was assessed and revealed subtle differences. (A) The use of 6-well tissue culture (TC) treated plates and supplemented M63 media was settled upon due to their ease of use and consistent assay results; (B) cultures in 12-well non tissue culture treated dishes showed differences in pellicle morphology compared with treated dishes; (C) smaller 24-well dishes developed a suitable pellicle and biofilm but were not practical for accurate biofilm analysis.

6.2.3 A modified technique for pellicle and biofilm estimation

Results from the biofilm growth optimisation experiments demonstrated that *M. smegmatis* mc^{2}155 was capable of developing both a pellicle and a biofilm and a technique to quantify both was developed. Pellicle wet weight was initially measured by using a wide scalpel blade to scrape the surface of the media however it was obvious that some material remained behind on the surface. Instead, thick blotting paper cut to size for the 6-well dishes was used to collect the entire pellicle in one manoeuvre by gently laying it on top of the pellicle

(Figure 6.4A) for a few seconds and then removing the paper using the side tag (Figure 6.4B). Blotting paper was dried at 50°C and pre-weighed prior to use and subsequently dried and weighed to record pellicle dry weight. Very little pellicle material remained on the surface after removal and the underlying biofilm remained undisturbed (Figure 6.4C) allowing a commonly used crystal violet staining technique to be performed (Figure 6.4D). Figure 6.4 follows the same three wells of *M. smegmatis* mc^{2}155 during pellicle removal and at the end of CV staining and ready for stain dissolving and quantification by OD_{570}. During biofilm formation optimisation in a range of multi-well dishes it was noted that there was a tendency to accidentally mechanically disturb the biofilm while removing spent media and washing the wells. This was a particular problem when using 24-well dishes and this issue contributed to the final selection of 6-well dishes for experimental assays.

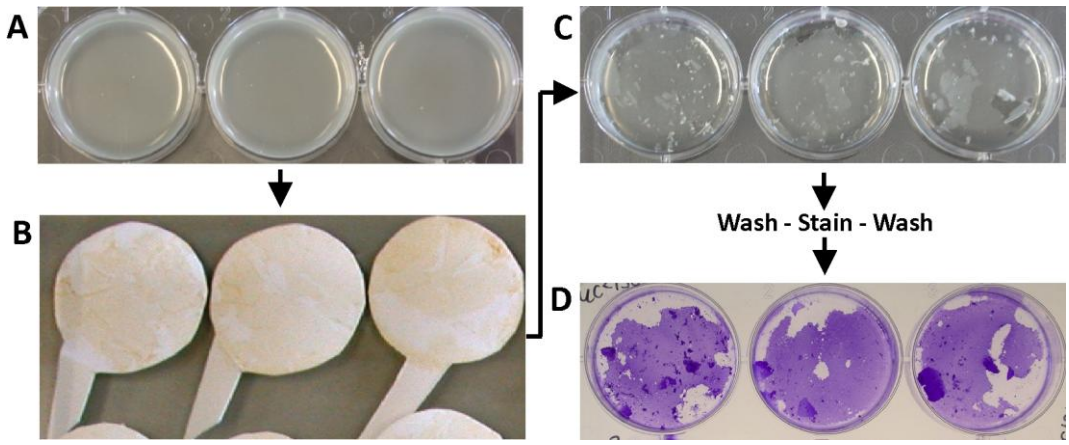


Figure 6.4. The use of custom made paper discs with tabs enabled distinct pellicle and biofilm quantification. (A) Typical pellicle formation of *M. smegmatis* mc^{2}155 lab strain in triplicate wells of a 6-well dish prior to pellicle removal; (B) the underside of dried paper discs that have pellicle from the above wells attached; (C) almost no pellicle remains in the wells after pellicle removal and the biofilm is undisturbed; (D) the resulting crystal violet stained biofilm.

6.2.4 *M. smegmatis* Δ*lsr2* exhibits different pellicle morphology

Pellicle and biofilm growth of *M. smegmatis* mc^{2}155 and *M. smegmatis* mc^{2}155 Δ*lsr2* in biofilm media appeared very similar however differences in pellicle morphology were noted. In contrast to the very smooth and glossy pellicle

exhibited by the lab strain, the Δ*lsr2* strain had a thinner pellicle with clumps of cellular material dispersed on top (Figure 6.5A). Examination of the lab strain pellicle under a light microscope showed a uniform distribution of cells in comparison to the Δ*lsr2* which had clumped regions (out of focus) overlaid on more uniformly distributed cells (Figure 6.5B). DNA staining of the pellicle illustrates a more uniform matrix in the lab strain in contrast to irregular thin and thick matrix regions in the Δ*lsr2* (Figure 6.5C).

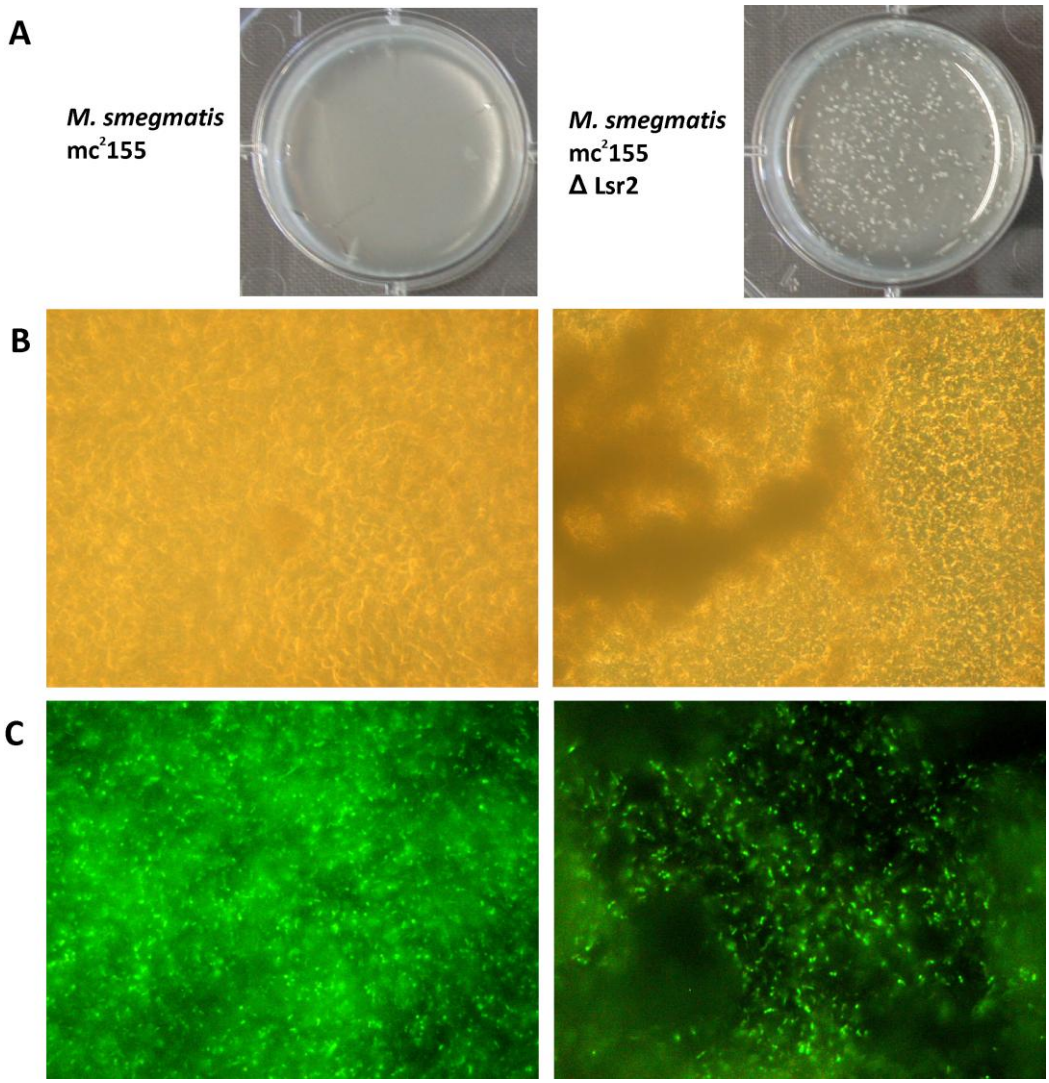


Figure 6.5. The pellicle of *M. smegmatis* Δ*lsr2* grown in M63 biofilm media is irregular and clumped. (A) Photographs illustrate the smooth and glossy lab strain pellicle, in contrast to the Δ*lsr2*; (B) light microscopy images show cell clumping in the Δ*lsr2*; (C) SYTO®9 DNA staining of the pellicle highlight the uneven matrix distribution in the Δ*lsr2* strain.

6.2.5 Complementation with native Lsr2 or Lsr2 DNA binding mutants fails to rescue the pellicle morphology phenotype

An investigation into the ability of *M. tuberculosis* H37Rv Lsr2 and Lsr2 mutants deficient in DNA binding to complement the morphological difference in pellicle formation between *M. smegmatis* mc^2155 lab strain and the Δ*lsr2* was undertaken. *M. tuberculosis* H37Rv *lsr2* (Rv3597c) and DNA binding mutants were cloned into the tetracycline inducible mycobacteria expression vector pMIND supplied by R. O'Toole (Blokpoel *et al.*, 2005). The transcripts of the introduced genes, cloned versions of Lsr2 and the three DNA binding mutants were generated both with and without a ribosome binding site (RBS) ahead of the gene. Lsr2 and mutants were induced in the *M. smegmatis* mc^2155 Δ*lsr2* strain and all strains (including controls) were grown in triplicate in M63 biofilm media in 6-well dishes and assayed for pellicle and biofilm growth. The unusual clumped pellicle morphology observed in the Δ*lsr2* was unable to be restored to that of the smooth and glossy lab strain by the expression of Lsr2 or any of the Lsr2 mutants (results not shown) cloned either with or without an RBS. The reason for this was unkown. There was no difference in pellicle dry weight (Figure 6.6A) or biofilm formation (Figure 6.6B) between control and experimental strains of *M. smegmatis* mc^2155. Figure 6.6 represents the combined data from three replicate non-RBS pMIND cloned experiments and five replicate RBS-pMIND cloned experiments. Three experiments were also completed for RBS-pMIND cloned strains in 12-well and 24-well dishes which also had the same outcome (results not shown).

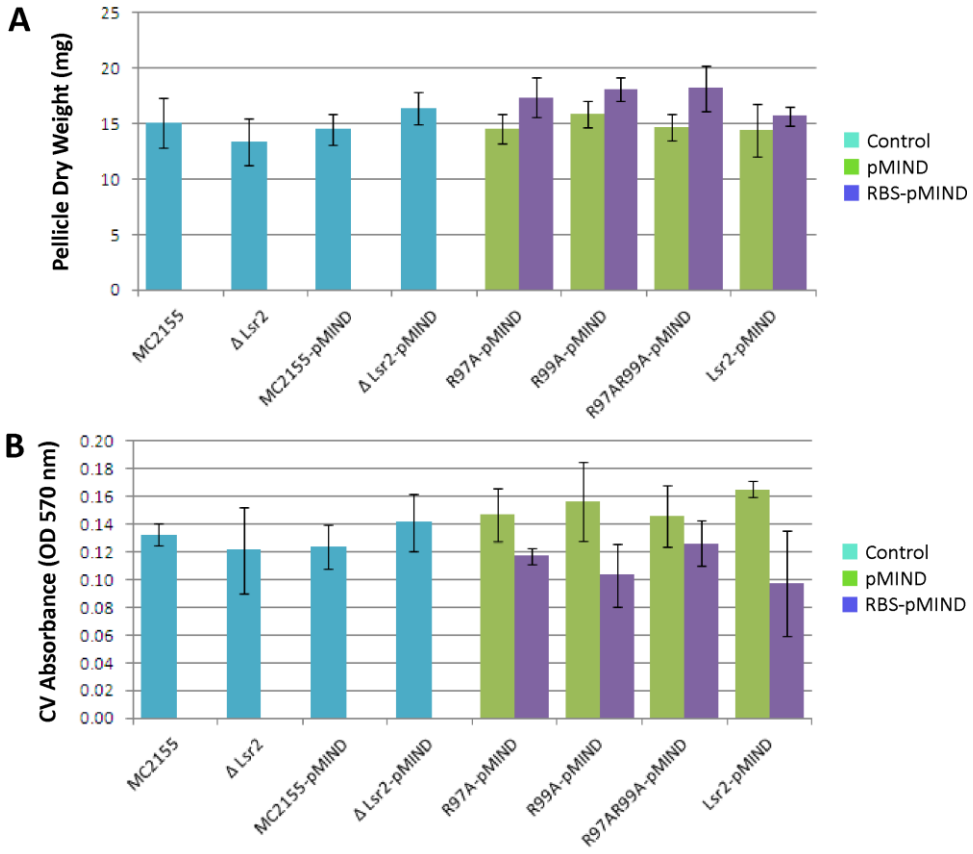


Figure 6.6. *M. smegmatis* Δ*lsr2* is not deficient in pellicle or biofilm formation and is not influenced by complementation with *M. tuberculosis* Lsr2 or DNA binding variants. (A) Dry pellicle weight of control strains, Lsr2 and DNA binding mutant Lsr2 genes introduced in a vector excluding a RBS (pMIND) and including a RBS (RBS-pMIND) (standard deviation error bars shown); (B) biofilm growth quantified by crystal violet staining and UV absorption of control strains, Lsr2 and DNA binding mutant Lsr2 genes introduced in a vector excluding a RBS (pMIND) and including a RBS (RBS-pMIND) (standard deviation error bars shown).

Molecular techniques were used to confirm that the genes for Lsr2 and Lsr2 mutants were present in the pMIND vector, that the pMIND vector was present in the appropriate *M. smegmatis* mc^{2}155 strain and that the correct lab strain or Δ*lsr2* strain was used for each experimental condition. PCR analysis of crudely lysed *M. smegmatis* mc^{2}155 cells confirmed the correct strain, vector and insert for each experimental condition (methods section 2.5.6). Primers designed for *M. smegmatis* mc^{2}155 *lsr2* (MSMEG_6092; lane 1) amplified this gene (345 bp) in the lab strain "mc^{2}155" and lab strain containing pMIND "mc^{2}155 + pMIND" only. Primers used in cloning *M. tuberculosis* H37Rv *lsr2* (Rv3597c; lane 2) successfully amplified *lsr2* (360 bp) and the mutant forms (lower panel) and apart from some

non-specific priming, did not amplify a gene from the four control strains (upper panel). Primers designed to encompass either side of the multiple cloning site of pMIND (lane 3) successfully amplified this region devoid of a gene (190 bp) in the control strains "mc^{2}155 + pMIND" and "Δ Lsr2 + pMIND" (upper panel) and amplified this region inclusive of *lsr2* (550 bp) in the experimental strains (lower panel). PCR analysis was performed for both the RBS inclusive and exclusive strains - the figure below illustrate results of the analysis of the RBS-pMIND strains. These techniques were used a number of times to repeatedly confirm the correct strains.

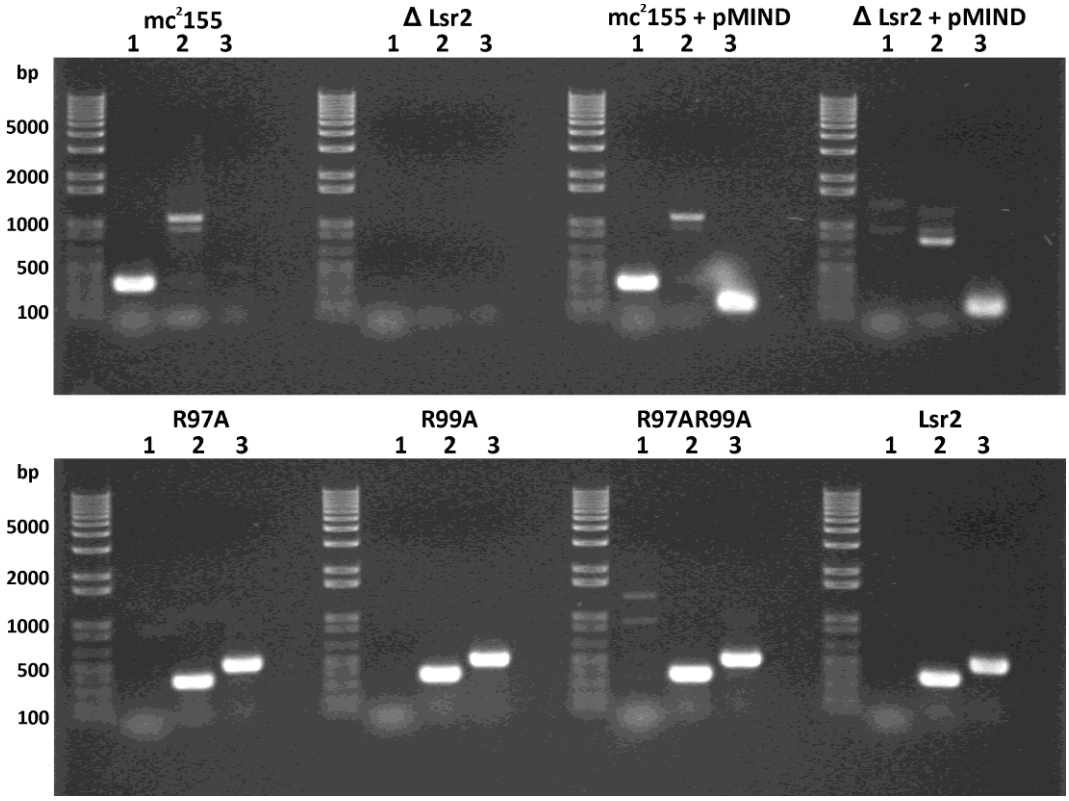


Figure 6.7. PCR analysis of control and experimental strains confirms correct *M. smegmatis* mc^{2}155 strain, vector and gene insert. Primers were designed for *M. smegmatis* mc^{2}155 *Lsr2* (MSMEG_6092; Lane 1), *M. tuberculosis* H37Rv *Lsr2* (Rv3597c; lane 2) and pMIND multiple cloning site (lane 3). The upper panel illustrates the two *M. smegmatis* mc^{2}155 strains including empty vector controls and the lower panel illustrates the four experimental variants.

6.3 Discussion

It is well established that *M. smegmatis* mc^{2}155 is capable of producing a biofilm (Ojha & Hatfull, 2007). A number of researchers have shown that the DNA binding protein Lsr2 is involved in biofilm formation and have observed that *M. smegmatis* mc^{2}155 Δ*lsr2* mutants grow as smooth, wet and round colonies on mycobacteria-specific media which is in contrast to the dry, rough and rugose colony morphology displayed by the wild type strain (Arora *et al.*, 2008; Chen *et al.*, 2006). These Δ*lsr2* mutants were generated by the insertion of a transposon element into the *lsr*2 gene (Chen *et al*., 2006) and a single base pair deletion arising from a spontaneous mutation that resulted in a truncated Lsr2 product (Arora *et al*., 2008)(Table 1.2).

An *M. smegmatis* mc^{2}155 Δ*lsr2* mutant containing a clean deletion of the gene was obtained for this study and its ability to form a biofilm examined. In contrast to other published research, examination of this mutant strain showed colony morphology that was similar to the lab strain *M. smegmatis* mc^{2}155. A modified technique for pellicle quantification was developed during this study to be able to clearly define the differences between growth of the air-surface pellicle and the surface-attached biofilm as published studies were ambiguous in this regard. After extensive experimental development to establish conditions for pellicle and biofilm growth, it was concluded that there was no difference in pellicle and biofilm growth between the lab strain and Δ*lsr2*. These results are in contrast to other studies (Arora *et al.*, 2008; Chen *et al.*, 2006). However a difference in pellicle morphology between the lab strain and Δ*lsr2* was noted. The usually smooth and glossy pellicle of the lab strain became thin and clumped in the mutant. Alterations in the surface of the pellicle between an *M. smegmatis* wild type strain and mutant such as observed in this work have been reported in a recent study which similarly discovered no differences in biofilm formation between the examined strains (Klepp *et al.*, 2012).

To examine if Lsr2 could rescue the morphological pellicle phenotype observed in the Δ*lsr2* strain, *lsr2* was introduced in a tetracycline inducible vector. Also, a number of DNA binding mutants of Lsr2 had been generated in previous studies

and it was of interest to see what influence these proteins may have on pellicle morphology. The mycobacteria expression vector pMIND was selected as it had been successfully used for inducible expression in our lab (Robson *et al.*, 2009) and the genes of interest were cloned into the vector in two ways - with and without an RBS - to account for any issues with translation in *M. smegmatis*. The expression of Lsr2 and DNA binding mutants in the mutant strain did not rescue the pellicle morphology phenotype and did not influence pellicle or biofilm formation.

Chapter Seven – Discussion

7.1 General background

Lsr2 from *M. tuberculosis* H37Rv is a small (12 kDa) and basic protein (pI of 10) that is highly conserved in mycobacteria and related actinomycetes (Gordon *et al.*, 2008). Lsr2 was first identified as an immunodominant T-cell antigen of *M. leprae* (Laal *et al.*, 1991; Oftung *et al.*, 2000; Singh, Narayanan *et al.*, 1994) and it continues to be studied in an effort to identify critical residues that are responsible for generating an immune response. DNA sequence analysis reveals that *lsr2* orthologues are present in all sequenced mycobacterial genomes, including *M. smegmatis, M. bovis and M. bovis* BCG, and the high amino acid sequence similarity among the Lsr2 homologues suggests that they perform a similar function.

Lsr2's highly basic, arginine rich and positively charged nature implies an ability for it to interact with negatively charged phosphate groups such as those on the DNA backbone. Lsr2 has been shown to be able to down-regulate the transcription of genes involved in cell wall synthesis and metabolic functions in *M. tuberculosis* H37Rv and it has been suggested that it could do this by controlling chromosomal topology (Colangeli *et al.*, 2007). Knock-out mutants of Lsr2 in *M. smegmatis* mc^2155 show altered colony morphology and reduced biofilm formation (Arora *et al.*, 2008; Chen *et al.*, 2006), however to date, an Lsr2 knock-out in *M. tuberculosis* H37Rv has been unsuccessful suggesting that Lsr2 is an essential gene in *M. tuberculosis*.

This project was undertaken to investigate the biochemical and structural characteristics of the DNA binding protein Lsr2 from *M. tuberculosis* H37Rv.

7.2 Lsr2 is a non-specific DNA binding protein

Lsr2 was expressed as a recombinant protein in *E. coli* to generate large amounts of protein that could be purified using affinity chromatography. It was known that Lsr2 had been expressed by other researchers in this manner for experimental work however, there was no mention in published work that Lsr2 co-purified with large amounts of *E. coli* genomic DNA, as was discovered during this study. The presence of this DNA with Lsr2 was beneficial at times for characterising the behaviour of Lsr2 however in general the DNA was a hindrance in most of the techniques used throughout this project.

It was determined that the DNA could be removed most efficiently by an incubation step with a buffer containing 2 M NaCl, followed by a separation of the DNA from protein using SEC. After DNA removal, the highly positively charged nature of Lsr2 influenced its behaviour and it would stick to centrifugal concentrator membranes and aggregated when it was highly concentrated. As a consequence of this, the majority of DNA-free Lsr2 handling was performed in a high salt buffer. In contrast to these technical difficulties, purified Lsr2 (co-purified with DNA) was stable even after several months of storage at 4°C.

TEM analysis of negatively stained Lsr2 in complex with co-purified with DNA revealed a regular fibril-like arrangement of protein coating the DNA strands of around 90 Å in diameter. The results suggested that Lsr2 bound to the DNA following its helical topology. Changes in Lsr2 and DNA topology by preparation of Lsr2 in high salt buffers were examined using TEM. As salt concentration increased, weakening the interaction between Lsr2 and the DNA, fibrils became more dispersed and less entwined suggesting that Lsr2 was also capable of drawing multiple strands of DNA together.

After obtaining concentrated, DNA-free, native Lsr2 protein from two different recombinant protein expression systems, crystallisation trials were performed but protein crystals were not forthcoming.

7.3 Properties of the C-terminal DNA binding domain of Lsr2

An NMR structure of the C-terminal DNA binding domain of Lsr2 from *M. tuberculosis* H37Rv was published part way through this study and this prompted a change in experimental focus. Authors of the NMR structure publication also identified and defined the boundaries of the N-terminal and C-terminal domains of Lsr2 that they had determined through domain swapping experiments with another DNA binding protein, H-NS (Gordon *et al.*, 2010). Prior to their research, the domain structure of Lsr2 had not been established. To bind to DNA, Lsr2 interacts with the minor groove of the DNA via the residues R97, G98, and R99.

Three Lsr2 DNA binding mutants were generated in this study to investigate the influence of these specific residues on DNA binding; denoted R97A, R99A and R97A-R99A. The DNA binding ability of these three mutants was reduced considerably, especially by the mutation of both arginine residues. TEM analysis of affinity purified Lsr2 mutants visually showed that the formation of fibrils of DNA coated by Lsr2 was minimal after mutating single arginine residues and was abolished by the double mutation.

Protein thermal stability was affected by residue mutation. However, the order of reduction in thermal stability suggested that the amount of DNA bound with each protein in the preparations used may have had the greater influence in thermal stability than residue mutation itself.

Protein secondary structure was slightly affected by residue mutation. Circular dichroism analysis of SEC purified Lsr2 and mutant proteins showed that the proportion of α-helix and β-sheet structures was altered. However, as the difference in optical activity between proteins was small the outcome of the secondary structure analysis is not conclusive.

Crystallisation trials were performed with the R97A-R99A mutant but protein crystals were not forthcoming.

7.4 Lsr2 DNA binding mutants and *Mycobacterium smegmatis* mc²155 biofilm and pellicle formation

It is unclear if the human pathogen *M. tuberculosis* is capable of forming a biofilm (Ha *et al.*, 2005) however it is well established that *M. smegmatis* mc^{2}155 is capable (Ojha & Hatfull, 2007). Lsr2 has been shown to play an important role in biofilm formation in *M. smegmatis* mc^{2}155 (Chen *et al.*, 2006) with Lsr2 mutant strains having a range of biofilm forming abilities (Arora *et al.*, 2008; Kocíncová *et al.*, 2008). Access to a *M. smegmatis* mc^{2}155 Δ*lsr2* lab strain was provided by R. Colangeli and this strain was used to explore the biofilm phenotype and to investigate the role of Lsr2 and three Lsr2 DNA binding mutants in biofilm formation.

It was established that the *M. smegmatis* mc^{2}155 lab strain was capable of producing both an air-surface pellicle and a surface-adhered biofilm after growth in M63 biofilm media. Specific techniques were developed for the consistent individual measurement of both the pellicle and biofilm which had not been reported previously. Custom-made filter paper sampling devices were used to remove the pellicle prior to biofilm staining (using crystal violet), and the dry pellicle weight and stain optical density was measured to quantify pellicle and biofilm formation, respectively.

A limited pellicle phenotype was established for the Δ*lsr2* strain whereby the pellicle formed in grainy patches instead of a smooth surface. The introduction of *M. tuberculosis* H37Rv *lsr2* and the three *lsr2* DNA binding mutants on a tetracycline-inducible mycobacterial vector (pMIND) failed to rescue the phenotype observed in the Δ*lsr2*. The reason for this was unknown.

7.5 Crystallisation and characterisation of the N-terminal dimerisation domain of Lsr2

After the publication of the C-terminal DNA binding domain NMR structure of Lsr2 part way through this study, the research focus switched to the expression and purification of the N-terminal dimerisation domain of Lsr2. The technique used to prepare protein for NMR analysis of the C-terminal domain of Lsr2 included expression of the full-length protein followed by limited trypsin digestion to remove the N-terminal domain. This approach is not suited to protein crystallographic techniques and instead separate protein domain expression was used. The publication of the crystallographic structure of the dimerisation domain of an *E. coli* DNA binding protein, H-NS, encouraged focus in this undertaking as the higher order structure of the N-NS dimerisation domain was intriguing (Arold *et al.*, 2010).

Seven alternative constructs of the N-terminal dimerisation domain of Lsr2 were expressed and purified with three of these proving sufficiently well-behaved to proceed into crystallisation trials. SEC estimation of the molecular weight of all the N-terminal domain constructs indicated that this domain was purifying as a dimer, however i*n vitro* glutaraldehyde cross-linking experiments also showed that the N-terminal domain was capable of forming much larger structures. Crystallisation trials were performed with Nterm, Nterm2 and Nterm3 N-terminal domain constructs of Lsr2, with Nterm2 successfully producing diffraction quality crystals.

Crystals of Nterm2 formed only after a “maturation” period where some unknown modification of the protein was taking place in solution prior to use in crystallisation techniques, or after the addition of trypsin (at a 1:500 mass ratio) to the protein immediately before crystallisation screens were performed. Analysis of “matured” Nterm2 using SEC showed that a number of larger multimeric forms were developing over time and it was determined that this effect could be recreated by the addition of trypsin to purified Nterm2.

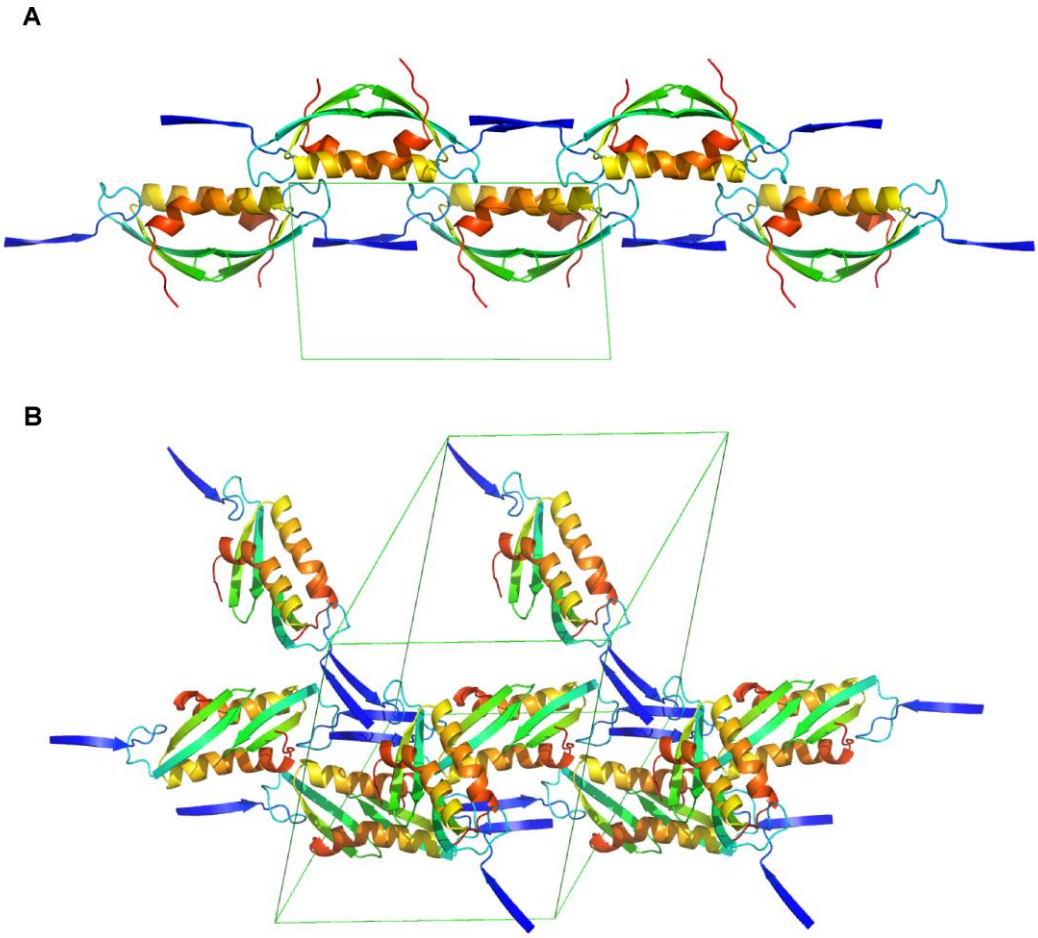


Figure 7.1. The 3D structure of the N-terminal domain of Lsr2 was determined in two different crystallographic space groups. (A) A cartoon diagram of the crystallographic symmetry of the N-terminal domain in the $P2_1$ crystal structure showing the unit cell (in green) perpendicular to the b-axis. The two-fold screw axis generates two Lsr2 chains that lay back to back in the horizontal plane shown here; (B) a cartoon diagram of the crystallographic symmetry of the N-terminal domain in the $P3_121$ crystal structure, showing the unit cell (in green). A two-fold symmetry (as a result of this space group) is visible in the horizontal plane and the first tier of Lsr2 chains in the three-fold vertical screw c-axis is shown.

An X-ray diffraction data set was collected from a crystal from "matured" Nterm2 and a crystal from trypsin co-crystallised Nterm2, both at a high resolution. Due to the fact that there were no structures similar to the N-terminal domain of Lsr2 in the protein data bank, molecular replacement techniques for structure solving were not available. A partnership between ourselves and scientists using *ab initio* techniques for solving the crystallographic phase problem was formed and the phase problem was completed, and an initial poly-alanine structure was solved and provided. The initial structure was modified to remove residues

suspected of being incorrectly positioned and it was then used in completing Nterm2 structure building and refinement. This study produced the first crystal structure of the N-terminal dimerisation domain of Lsr2 from *M. tuberculosis* H37Rv.

The crystal from "matured" Nterm2 provided a structure in the $P2_1$ space group, whereas the crystal formed by co-crystallistion with trypsin, provided a structure in the $P3_121$ space group (Figure 7.1). The Lsr2 dimerisation domain forms a compact homodimeric α/β structure that locks together in a head-to-head fashion. Both structures reveal chains of the N-terminal domain of Lsr2 (shown by crystallographic symmetry), with the N-terminal end of each monomer forming one strand of an anti-parallel β-sheet that allows oligomerisation to occur. In both structures the N-terminal expression tag and the first three residues M1, A2 and K3 are missing and it is proposed that these residues are required to be absent for oligomerisation to occur. The fact that the two structures have identical residues missing (due to either "maturation" or trypsin treatment) has led to the hypothesis that oligomerisation of the N-terminal domain could be an *in vivo* mechanism initiated by the activation of a site-specific protease.

The crystallographic data in the higher symmetry $P3_121$ space group reveals a possible mechanism of protein alignment and binding to DNA (Figure 7.2). Within the constraints of crystal packing, strands of DNA can be uniformly positioned at a 120° angle to and underneath chains of the N-terminal domain. The C-terminal DNA binding domain (PDB 2KNG) can be positioned adjacent to the N-terminal domain dimer. This model provides a possible mechanism for the binding and compaction of chromosomal DNA.

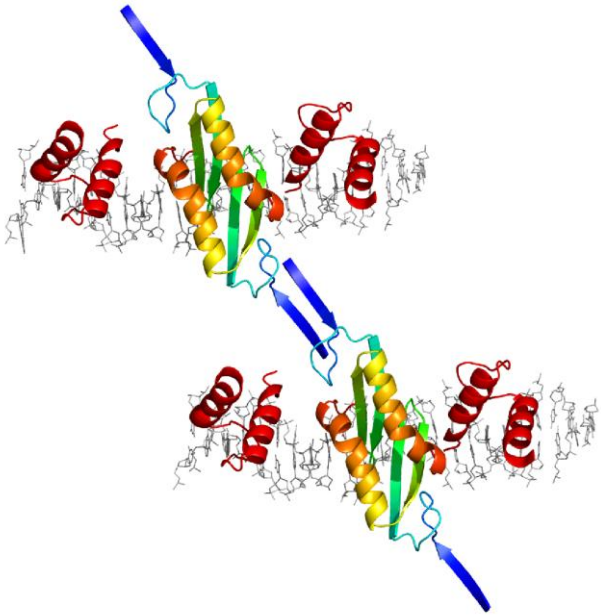


Figure 7.2. The $P3_121$ crystal packing allows modelling of a possible arrangement of a chain of the full-length Lsr2 protein across multiple strands of DNA. DNA strands (in grey) could be bridged by dimeric Lsr2 after its oligomerisation via anti-parallel β-sheet formation (in blue) in the N-terminal domains (in rainbow colour). Crystallographic spacing allows the placement of a single DNA strand below the N-terminal domain at a 120° angle. The C-terminal DNA binding domain (PDB 2KNG) is able to be positioned flanking either side of the N-terminal domain and binds to the DNA via its conserved RGR residues in the loop between the two helices.

This model was tested by the addition of trypsin to full-length affinity purified Lsr2 protein co-purified with genomic DNA. Proteolytic treatment of Lsr2 with small amounts of trypsin (1:500 mass ratio) resulted in DNA condensation and the formation of thick helical structures which were visualised using TEM, and which also provided protection of the DNA from DNase digestion as observed by agarose gel electrophoresis. Typically, protease treatment of protein would result in protein degradation rather than the formation of larger structures and therefore this result appears to be a specific feature of Lsr2.

7.6 Future research

Research presented in this thesis provides a hypothesis for the mechanism of DNA binding and chromosome compaction of the non-specific DNA binding protein Lsr2 from *M. tuberculosis* H37Rv. Further characterisation of the N-terminal domain could be undertaken by the mutation of specific residues as identified by the 3D structure, to identify those that are essential for the formation of higher order structures. It would also be of interest to further examine the compact structures observed with TEM after the addition of trypsin to Lsr2, to build a 3D model of the arrangement of Lsr2 on the strands of DNA. For those researchers with the ability to perform *in vivo* experiments, the affect of creating Lsr2 mutants that are unable to form oligomers due to residue modification or due to a "non-cleavable" N-terminus could be investigated. Alternatively, controlled oligomerisation of Lsr2 *in vivo* could be examined to investigate the influence of Lsr2 on chromosome packaging and cell growth. Continued work on the mechanism of DNA binding and compaction by Lsr2 may present a research avenue to interrupt gene regulation and DNA protection by this protein, and elucidate the details surrounding its biological function.

Appendices

Appendix A: Buffers, solutions, growth media, bacterial strains, plasmids and antibiotics

A1: Buffers and solutions

Electrophoresis buffers were made with distilled water whereas chromatography buffers were made with MQ water and were then filtered with a 0.2 µm filter prior to use. Loading dye was made using MQ water.

BLOTTO	10% (*w/v*) non fat milk powder in 1 x TBS-T
10 x DNA loading dye	0.4% (*w/v*) bromophenol blue, 0.4% (*w/v*) xylene cyanol, 50% (*v/v*) glycerol
Fairbanks A stain	0.05% (*w/v*) coomassie blue R-250, 25% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks B stain	0.005% (*w/v*) coomassie blue R-250 10% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks C stain	0.002% (*w/v*) coomassie blue, 10% (*v/v*) acetic acid
Fairbanks D stain	10% (*v/v*) acetic acid
Mounting media	45 ml glycerol, 5 ml H_2O, 0.0008 g Na_2CO_3, 0.036g $NaHCO_3$
5 x native loading dye	300m M Tris-HCl pH 6.8, 50% *(v/v)* glycerol and 0.05% (*w/v*) bromophenol blue
PEG200/KOH	30 g PEG 200, 19.5 ml dH_2O, 465 µl 2M KOH, pH 13.49
PBS (10 x)	0.2 M phosphate (0.038 M NaH_2PO_4, 0.162 M Na_2HPO_4), 1.5 M NaCl, pH 7.4
PBS (1 x)	100 ml 10 x PBS + 900 ml dH_2O

50 mM phosphate (pH 7.4)	A 500 ml volume of 50 mM phosphate buffer at pH 7.4 was made by combining 38.7 ml of 0.5 M Na_2HPO_4 and 11.3 ml of 0.5 M NaH_2PO_4. NaCl and imidazole were then added to appropriate concentrations
Ponceau S stain	0.2% (*w/v*) ponceau red, 1% (*v/v*) acetic acid
Resolving buffer	1.5 M Tris-HCl, pH 8.8
4 x SDS loading dye	200 mM Tris-HCl pH 6.8, 8% (*w/v*) SDS, 40% (*v/v*) glycerol, 0.4% (*w/v*) bromophenol blue, 400 mM β-mercaptoethanol
Stacking buffer	1.0 M Tris-HCl, pH 6.8
50 x TAE	242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA (pH 8.0)
10 x TAE	400 mM Tris-acetate, 10 mM EDTA
1 x TAE	100 ml 10 x TAE + 900 ml H_2O
1 x TBS	20 mM Tris-HCl pH 7.6, 150 mM NaCl
1 x TBS-T	20 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% *(v/v)* tween 20
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0
TG-SDS running buffer	25 mM Tris-HCl pH 8.3, 250 mM glycine, 0.1% (*w/v*) SDS
TG running buffer	25 mM Tris-HCl pH 8.5, 250 mM glycine
Transfer buffer	25 mM Tris-HCl, 192 mM glycine, 20% methanol, 0.01% (*w/v*) SDS

A2: Growth media

Dehydrated Difco™, BDL™ and Bacto™ media and ADC enrichment media was supplied by Becton, Dickinson and Company, USA. Growth media was made up in distilled water and sterilised by autoclaving at 121°C for 15 minutes.

Hartman's de Bont	Tween 80 omitted for biofilm assay. Add in order: 1 x trace metal stock*, 15 mM $(NH_4)_2SO_4$, 0.2% (*v/v*) glycerol, 0.05% (*v/v*) tween 80, 50 mM MOPS, pH 7.0, autoclave then add 1 x phosphate stock **
LB	1% (*w/v*) bactotryptone or bactopeptone, 0.5% (*w/v*) yeast extract, 1% (*w/v*) NaCl, pH 8.0, autoclave
LB agar	1% (*w/v*) bactotryptone or bactopeptone, 0.5% (*w/v*) yeast extract, 1% (*w/v*) NaCl, 15 g/L agar, pH 8.0, autoclave
LBT	1% (*w/v*) bactotryptone or bactopeptone, 0.5% (*w/v*) yeast extract, 1% (*w/v*) NaCl, pH 8.0, autoclave then add 0.05% (*v/v*) sterile tween 80 at 45°C
LBT agar	1% (*w/v*) bactotryptone or bactopeptone, 0.5% (*w/v*) yeast extract, 1% (*w/v*) NaCl, 15 g/L agar, pH 8.0, autoclave then add 0.05% (*v/v*) sterile tween 80 at 50°C
Low salt LB	1% (*w/v*) bactotryptone or bactopeptone, 0.5% (*w/v*) yeast extract, 0.5% (*w/v*) NaCl, pH 8.0, autoclave
Low salt LB agar	1% (*w/v*) bactotryptone, or bactopeptone 0.5% (*w/v*) yeast extract, 0.5% (*w/v*) NaCl, 15 g/L agar, pH 8.0, autoclave
M63 minimal	100 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 1.7 µM $FeSO_4 \cdot 7H_2O$, pH to 7.0 with KOH, autoclave, then add 1 mM $MgSO_4$, 0.2% glycerol (v/v)
M63 biofilm media	100 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 1.7 µM $FeSO_4 \cdot 7H_2O$, pH to 7.0 with KOH, autoclave, then add sterile 1 mM $MgSO_4$, 0.2% (v/v) glycerol, 0.5% (w/v) bactotryptone and 0.7 mM $CaCl_2$
M9 salts	42 mM Na_2HPO_4, 22 mM KH_2PO_4, 19 mM NH_4Cl, 8.5 mM NaCl, pH to 7.4 with 10 M NaOH, autoclave
M9 minimal	42 mM Na_2HPO_4, 22 mM KH_2PO_4, 19 mM NH_4Cl, 8.5 mM NaCl, pH to 7.4 with 10 M NaOH, autoclave, then add sterile 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 0.4% (w/v) glucose

PA-0.5G	50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 0.5% (*w/v*) glucose, 0.1 x metals mix***, 100 µg/ml each of 17 amino acids (no Cys, Tyr or Met). For methionine auxotroph DL41 add 100 µg/ml methionine. Individual components autoclaved or sterile filtered before adding to sterile dH_2O
PASM-5052	50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 0.5% (*w/v*) glycerol, 0.05% (*w/v*) glucose, 0.2% (*w/v*) α-lactose, 1 x metals mix***, 100 nM vitamin B12, 200 µg/ml each of 17 amino acids (no Cys, Tyr or Met), 10 µg/ml methionine, 125 µg/ml selenomethionine
Rich media + glucose	1% (*w/v*) bactotryptone, 0.5% (*w/v*) yeast extract, 0.5% (*w/v*) NaCl, 0.2% (*w/v*) glucose
SOC	2% (*w/v*) bactotryptone or bactopeptone, 0.5% (*w/v*) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, autoclave. Once cooled add sterile glucose to 20 mM
TB	1.2% (*w/v*) bactotryptone, 2.4 % (*w/v*) yeast extract, 0.4% (*w/v*) glycerol in 9 part volume, autoclave. Once cooled, combine with a 1 part volume of sterile 0.17 M KH_2PO_4, 0.72 M K_2HPO_4
7H9/ADC/T	0.47 g 7H9 powder, 0.2% *(w/v)* glycerol in 90 ml H_2O, autoclave then add 10 ml ADC enrichment and 0.05% (*v/v*) tween 80 at 45°C
7H10/ADC/T agar	1.9 g 7H10 powder, 0.5% (*w/v*) glycerol in 90 ml H_2O, autoclave then add 10 ml ADC enrichment and 0.05% (*v/v*) tween 80 at 50°C

*100 x trace metals stock for HdB media: EDTA [0.1 g], $MgCl_2.6H_2O$ [1 g], $CaCl_2.2H_2O$ [10 mg], $NaMoO_4.2H_2O$ [2 mg], $CoCl_2.6H_2O$ [4 mg], $MnCl_2.2H_2O$ [10 mg], $ZnSO_4.7H_2O$ [20 mg], $FeSO_4.7H_2O$ [50 mg], $CuSO_4.5H_2O$ to 100 ml with dH_2O.

**100 x phosphate stock for HdB media: K_2HPO_4 [15.5g], $NaH_2PO_4.2H_2O$ [11.1g] to 100 ml with dH_2O.

***1000 x metals mix made up from the sterile stock solutions of each component to give the following concentrations: 50 µM $FeCl_3$ in 0.12 M HCl (filter sterile), 20 µM $CaCl_2$, 10 µM $MnCl_2$, 10 µM $ZnSO_4$, 2 µM $CoCl_2$, 2 µM $CuCl_2$, 2 µM $NiCl_2$, 2 µM Na_2MoO_4, 2 µM Na_2SeO_3, 2 µM H_3BO_3.

A3: Bacterial strains, plasmids and antibiotics

Table A.1. Bacterial strains used in the study.

Strain	Description
Escherichia coli	
TOP10	High efficiency transformation and plasmid copy number. Inactivated endonuclease I for high plasmid DNA yield. *LacZ* for blue/white colony screening.
DH5α	High efficiency transformation and high plasmid copy number. Inactivated intracellular endonucleases for high plasmid DNA yield.
BL21 (DE3)	DE3 lysogen contains T7 polymerase upon IPTG induction. This strain is deficient of lon and omp-t proteases and is therefore suitable for expression of non-toxic genes.
DL41	Methionine auxotroph suitable for expression of selenomethionine labelled protein.
Mycobacterium smegmatis	
mc^{2}155	Electrocompetent wild-type *M. smegmatis* mc^{2}155 lab strain.
mc^{2}155 Δ*lsr2*	The wild-type *M. smegmatis* mc^{2}155 lab strain with a clean unmarked deletion of MSMEG_6092 *lsr2* (NJS22; Colangeli *et al*., 2007).

Table A.2. Plasmids used in the study.

Plasmid	Description
pET30b	*E. coli* expression vector. T7 promoter. Encodes either an N-terminal 6 x histidine tag-thrombin-S·tag or C-terminal 6 x histidine expression tag with no cleavage site. Kanamycin resistance (Colangeli *et al*., 2007).
pMAL-c2x	*E. coli* expression vector. Tac promoter. Encodes *malE* for expression of an N-terminal MBP fusion tag with a Factor Xa protease cleavage site. Ampicillin resistance (NEB, USA).
pPROExHTb	*E. coli* expression vector. Trc promoter. Encodes an N-terminal 6 x histidine tag with an rTEV protease cleavage site. Ampicillin resistance (Invitrogn, USA).
pMIND	*M. smegmatis* tetracycline inducible expression vector. TetRO promoter. Kanamycin and hygromycin resistance (Bolkpoel *et al*., 2005).
pDEST17	*E. coli* expression vector. T7 promoter. Encodes an N-terminal 6 x histidine tag with no protease cleavage site. Ampicillin and chloromphenicol resistance (Invitrogen, USA).

Table A.3. Antibiotic stocks and concentrations used in the study.

Antibiotic	Stock	Final Concentration
Ampicillin	100 mg/ml	100 μg/ml
Hygromycin	50 mg/ml	50 μg/ml
Kanamycin	50 mg/ml	50 μg/ml
Tetracyline	0.2 mg/ml	0.02 μg/ml

Table A. 4. Plasmid constructs used in the study.

Plasmid	Description	Reference
***lsr2*-pET30b**	pET30 with *M. tuberculosis* H37Rv *lsr2* inserted into *NdeI/XhoI* restriction sites.	Colangeli *et al*., (2007)
***lsr2*-pMAL-c2x**	pMAL-c2x with *M. tuberculosis* H37Rv *lsr2* inserted into *XbaI/PstI* restriction sites.	This study
***lsr2*-pMIND**	pMIND with *M. tuberculosis* H37Rv *lsr2* inserted into *PstI/SpeI* restriction sites.	This study
R97A-pET30b	Site directed mutagenesis mutant of *lsr2*-pET30b with R97A mutation.	This study
R99A-pET30b	Site directed mutagenesis mutant of *lsr2*-pET30b with R99A mutation.	This study
R97AR99A-pET30b	Site directed mutagenesis mutant of *lsr2*-pET30b with R97AR99A mutation.	This study
R97A-pMIND	pMIND with R97A *lsr2* mutant inserted into *PstI/SpeI* restriction sites.	This study
R99A-pMIND	pMIND with R99A *lsr2* mutant inserted into *PstI/SpeI* restriction sites.	This study
R97AR99A-pMIND	pMIND with R97AR99A *lsr2* mutant inserted into *PstI/SpeI* restriction sites.	This study
R97A-RBSpMIND	pMIND with RBS and R97A *lsr2* mutant inserted into *PstI/SpeI* restriction sites.	This study
R99A-RBSpMIND	pMIND with RBS and R99A *lsr2* mutant inserted into *PstI/SpeI* restriction sites.	This study
R97AR99A-RBSpMIND	pMIND with RBS and R97AR99A *lsr2* mutant inserted into *PstI/SpeI* restriction sites.	This study
***Nterm*-pPROExHTb**	pPROExHTb with Lsr2 N-terminal residues 1-59 inserted into *NcoI/XhoI* restriction sites.	This study
***Nterm+*-pPROExHTb**	pPROExHTb with lsr2 N-terminal residues 1-68 inserted into *NcoI/XhoI* restriction sites.	This study
***Nterm1*-pPROExHTb**	pPROExHTb with Lsr2 N-terminal residues 1-56 inserted into *NcoI/XhoI* restriction sites.	This study
***Nterm2*-pPROExHTb**	pPROExHTb with Lsr2 N-terminal residues 1-61 inserted into *NcoI/XhoI* restriction sites.	This study
***Nterm3*-pPROExHTb**	pPROExHTb with Lsr2 N-terminal residues 1-64 inserted into *NcoI/XhoI* restriction sites.	This study
***Nterm4*-pPROExHTb**	pPROExHTb with Lsr2 N-terminal residues 1-71 inserted into *NcoI/XhoI* restriction sites.	This study
***Nterm5*-pPROExHTb**	pPROExHTb with Lsr2 N-terminal residues 1-74 inserted into *NcoI/XhoI* restriction sites.	This study
I34M-pPROExHTb	Site directed mutageneis mutant of *Nterm2*-pPROExHTb with I34M mutation.	This study
L44M-pPROExHTb	Site directed mutageneis mutant of *Nterm2*-pPROExHTb with L44M mutation.	This study
L48M-pPROExHTb	Site directed mutageneis mutant of *Nterm2*-pPROExHTb with L48M mutation.	This study

Appendix B: Publications

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D - Biological Crystallography, 66*(2), 213-221.

Alland, D., Kramnik, I., Weisbrod, T. R., Otsubo, L., Cerny, R., Miller, L. P., Jacobs, W. R., Jr., & Bloom, B. R. (1998). Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): The effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Science, 95*(22), 13227-13232.

Alland, D., Steyn, A. J., Weisbrod, T., Aldrich, K., & Jacobs, W. R., Jr. (2000). Characterization of the *Mycobacterium tuberculosis* iniBAC promoter, a promoter that responds to cell wall biosynthesis inhibition. *Journal of Bacteriology, 182*(7), 1802-1811.

Arcus, V., Lott, J., Johnston, J., & Baker, E. (2006). The potential impact of structural genomics on tuberculosis drug discovery. *Drug Discovery Today, 11*(1/2), 28-34.

Arold, S. T., Leonard, P. G., Parkinson, G. N., & Ladbury, J. E. (2010). H-NS forms a superhelical protein scaffold for DNA condensation. *Proceedings of the National Academy of Sciences, 107*(36), 15728-15732.

Arora, K., Whiteford, D., Lau-Bonilla, D., Davitt, C., & Dahl, J. (2008). Inactivation of *lsr2* results in a hyper-motile phenotype in *Mycobacterium smegmatis*. *Journal of Bacteriology, 190*(12), 4291-4300.

Bailey, S. (1994). The CCP4 suite - programs for protein crystallography. *Acta Crystallographica Section D, 50*, 760-763.

Benvenuti, M., & Mangani, S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols, 2*(7), 1633-1651.

Bertin, P., Benhabiles, N., Krin, E., Laurent-Winter, C., Tendeng, C., Turlin, E., Thomas, A., Danchin, A., & Brasseur, R. (1999). The structural and functional organization of H-NS-like proteins is evolutionarily conserved in Gram-negative bacteria. *Molecular Microbiology, 31*(1), 319-329.

Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A., & Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Molecular Microbiology, 43*(3), 717-731.

Bissielo, A., Lim, E., & Heffernan, H. (2011). *Tuberculosis in New Zealand: Annual Report 2010*. Wellington, New Zealand: Institute of Environmental Science and Research Ltd (ESR)

Blokpoel, M. C. J., Murphy, H. N., O'Toole, R., Wiles, S., Runn, E. S. C., Stewart, G. R., Young, D. B., & Robertson, B. D. (2005). Tetracycline-inducible gene regulation in mycobacteria. *Nucleic Acids Research, 33*(2), e22.

Branda, S. S., Vik, Å., Friedman, L., & Kolter, R. (2005). Biofilms: the matrix revisited. *TRENDS in Microbiology, 13*(1), 20-26.

Carpenter, M. L., & Geoff Kneale, G. (1994). Circular dichroism for the analysis of protein-DNA interactions. In G. G. Kneale (Ed.), *DNA-Protein Interactions* (Vol. 30, pp. 339-345): Humana Press.

Carter, G., Wu, M., Drummond, D. C., & Bermudez, L. E. (2003). Characterization of biofilm formation by clinical isolates of *Mycobacterium avium*. *Journal of Medical Microbiology, 52*(9), 747-752.

Cary, P. D., & Kneale, G. G. (2008). Circular dichroism for the analysis of protein–DNA interactions. In T. M. a. B. Leblanc (Ed.), *DNA-Protein Interactions* (Vol. 543, pp. 613-624): Humana Press.

Chaduvula, M., Murtaza, A., Misra, N., Narayan, N. P. S., Ramesh, V., Prasad, H. K., Rani, R., Chinnadurai, R. K., & Nath, I. (2012). Peptides of Lsr2 of *Mycobacterium leprae* show hierarchical responses in lymphoproliferative assays with selective recognition by anergic lepromatous leprosy patients. *Infection and Immunity, 80*(2), 742-752.

Chatterjee, D. (1997). The mycobacterial cell wall: structure, biosynthesis and sites of drug action. *Current Opinion in Chemical Biology, 1*(4), 579-588.

Chen, J., German, G., Alexander, D., Ren, H., Tan, T., & Liu, J. (2006). Roles of Lsr2 in colony morphology and biofilm formation of *Mycobacterium smegmatis*. *Journal of Bacteriology, 188*(2), 633-641.

Chen, J. M., Ren, H., Shaw, J. E., Wang, Y. J., Li, M., Leung, A. S., Tran, V., Berbenetz, N. M., Kocincova, D., Yip, C. M., Reyrat, J.-M., & Liu, J. (2008). Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Research, 36*(7), 2123-2135.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D, 66*, 12-21.

Chen, V. B., Davis, I. W., & Richardson, D. C. (2009). KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. *Protein Science, 18*, 2403-2409.

Colangeli, R., Haq, A., Arcus, V. L., Summers, E., Magliozzo, R. S., McBride, A., Mitra, A. K., Radjainia, M., Khajo, A., Jacobs, W. R., Salgame, P., & Alland, D. (2009). The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. *Proceedings of the National Academy of Sciences, 106*(11), 4414-4418.

Colangeli, R., Helb, D., Sridharan, S., Sun, J., Varma-Basik, M., Hernado Hazbon, M., Harbacheuski, R., Megjugorac, N. J., Jacobs, W. R., Holzenburg, A., Sacchettini, J. C., & Alland, D. (2005). The *Mycobacterium tuberculosis* iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Molecular Microbiology, 55*(6), 1829-1840.

Colangeli, R., Helb, D., Vilchze, C., Hazbn, M. H., Lee, C.-G., Safi, H., Sayers, B., Sardone, I., Jones, M. B., Fleischmann, R. D., Peterson, S. N., Jacobs, W. R., & Alland, D. (2007). Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in *M. tuberculosis*. *PLoS Pathogens, 3*(6), e87.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., III, C. E. B., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., & Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature, 396*(12 November), 537-544.

Connolly, L. E., Edelstein, P. H., & Ramakrishnan, L. (2007). Why is long-term therapy required to cure tuberculosis? *PLoS Medicine, 4*(3), e120.

Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R., & Lappin-Scott, H. M. (1995). Microbial biofilms. *Annual Review of Microbiology, 49*(1), 711-745.

Costerton, J. W., Stewart, P. S., & Greenberg, E. P. (1999). Bacterial biofilms: A common cause of persistent infections. *Science, 284*(5418), 1318-1322.

Dale, G. E., Oefner, C., & D'Arcy, A. (2003). The protein as a variable in protein crystallization. *Journal of Structural Biology, 142*(1), 88-97.

Dame, R. T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology, 56*(4), 858-870.

Dame, R. T., & Goosen, N. (2002). HU: promoting or counteracting DNA compaction? *FEBS Letters, 529*(2-3), 151-156.

DeLano, W. L. (2002). The PyMOL molecular graphics system *DeLano Scientific*.

Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. In Charles W. Carter, Jr. (Ed.), *Methods in Enzymology* (Vol. 276, pp. 523-530): Academic Press.

Drlica, K., & Rouviere-Yaniv, J. (1987). Histonelike proteins of bacteria. *Microbiological Reviews, 51*(3), 301-319.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., & Wilson, A. (2011). Geneious v5.4, Available from http://www.geneious.com.

Dye, C. (2006). Global epidemiology of tuberculosis. *The Lancet, 367*(9514), 938-940.

Dye, C., Espinal, Marcos A., Watt, Catherine J., Mbiaga, C., & Williams, Brian G. (2002). Worldwide incidence of multidrug resistant tuberculosis. *The Journal of Infectious Diseases, 185*(8), 1197-1202.

Dye, C., Williams, B. G., Espinal, M. A., & Raviglione, M. C. (2002). Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science, 295*(5562), 2042-2046.

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

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry, 357*(2), 289-298.

Esteban, J., Martín-de-Hijas, N., Kinnari, T., Ayala, G., Fernández-Roblas, R., & Gadea, I. (2008). Biofilm development by potentially pathogenic non-pigmented rapidly growing mycobacteria. *BMC Microbiology, 8*(1), 1-8.

Fadouloglou, V. E., Kokkinidis, M., & Glykos, N. M. (2008). Determination of protein oligomerization state: Two approaches based on glutaraldehyde crosslinking. *Analytical Biochemistry, 373*(2), 404-406.

Gadgil, H., Jurado, L. A., & Jarrett, H. W. (2001). DNA affinity chromatography of transcription factors. *Analytical Biochemistry, 290*(2), 147-178.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571-607). Totowa: Humana Press.

Glickman, M. S., & Jacobs Jr, W. R. (2001). Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline. *Cell, 104*(4), 477-485.

Gopalaswamy, R., Narayanan, S., Jacobs, W. R., & Av-Gay, Y. (2008). *Mycobacterium smegmatis* biofilm formation and sliding motility are affected by the serine/threonine protein kinase PknF. *FEMS Microbiology Letters, 278*(1), 121-127.

Gordon, B. R. G., Imperial, R., Wang, L., Navarre, W. W., & Liu, J. (2008). Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins. *Journal of Bacteriology, 190*(21), 7052-7059.

Gordon, B. R. G., Li, Y., Cote, A., Weirauch, M. T., Ding, P., Hughes, T. R., Navarre, W. W., Xia, B., & Liu, J. (2011). Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proceedings of the National Academy of Sciences, 108*(26), 10690-10695.

Gordon, B. R. G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., Navarre, W. W., Xia, B., & Liu, J. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences, 107*(11), 5154-5159.

Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., & Cole, S. T. (1999). Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Molecular Microbiology, 32*(3), 643-655.

Greenfield, N. J. (2007). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols, 1*(6), 2876-2890.

Ha, K. Y., Chung, Y. G., & Ryoo, S. J. (2005). Adherence and biofilm formation of *Staphylococcus epidermidis* and *Mycobacterium tuberculosis* on various spinal implants. *Spine, 30*(1), 38-43.

Herrera, J. E., & Chaires, J. B. (1994). Characterization of preferred deoxyribonuclease I cleavage sites. *Journal of Molecular Biology, 236*(2), 405-411.

Hrmova, M., & Fincher, G. B. (2009). Functional Genomics and Structural Biology in the Definition of Gene Function. In D. J. Somers, P. Langridge & J. P. Gustafson (Eds.), *Methods in Molecular Biology, Plant Genomics, Methods and Protocols* (Vol. 513, pp. 199-227). New York: Humana Press.

Johnson, W. C. (1988). Secondary structure of proteins through circular dichroism spectroscopy. *Annual Review of Biophysics and Biophysical Chemistry, 17*(1), 145-166.

Johnston, J., Arcus, V., Morton, C., Parker, M., & Baker, E. (2003). Crystal structure of a putative methyltransferase from *Mycobacterium tuberculosis*: Misannotation of a genome clarified by protein structural analysis. *Journal of Bacteriology, 185*(14), 4057-4065.

Klepp, L. I., Forrellad, M. A., Osella, A. V., Blanco, F., Stella, E. J., Bianco, M. V., Santangelo, M. d. l. P., Sassetti, C., Jackson, M., Cataldi, A. A., Bigi, F., & Morbidoni, O. (2012). Impact of the deletion of the six *mce* operons in *Mycobacterium smegmatis*. *Microbes and Infection,* (in press).

Koch, R. (1882). Die aetiologie der tuberculose. *Berliner Klinische Wochenschrift, 19*, 221-230.

Kocíncová, D., Singh, A. K., Beretti, J.-L., Ren, H., Euphrasie, D., Liu, J., Daffé, M., Etienne, G., & Reyrat, J.-M. (2008). Spontaneous transposition of IS1096 or ISMsm3 leads to glycopeptidolipid overproduction and affects surface properties in *Mycobacterium smegmatis*. *Tuberculosis, 88*(5), 390-398.

Koh, J., Shkel, I., Saecker, R. M., & Record Jr, M. T. (2011). Nonspecific DNA binding and bending by HUαβ: Interfaces of the three binding modes characterized by salt-dependent thermodynamics. *Journal of Molecular Biology, 410*(2), 241-267.

Laal, S., Sharma, Y., Prasad, H., Murtaza, A., Singh, S., Tangri, S., Misra, R., & Nath, I. (1991). Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy sprectrum. *Proceedings of the National Academy of Science, 88*, 1054-1058.

Lee, B. H., Murugasu-Oei, B., & Dick, T. (1998). Upregulation of a histone-like protein in dormant *Mycobacterium smegmatis*. *Molecular and General Genetics MGG, 260*(5), 475-479.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, 26*.

Lillebaek, T., Dirksen, A., Baess, I., Strunge, B., Thomsen, V. Ø., & Andersen, Å. B. (2002). Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. *Journal of Infectious Diseases, 185*(3), 401-404.

Lucchini, S., Rowley, G., Goldberg, M. D., Hurd, D., Harrison, M., & Hinton, J. C. D. (2006). H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathogens, 2*(8), e81.

Marques, M. A. M., Espinosa, B. J., Xavier da Silveira, E. K., Pessolani, Maria C. V., Chapeaurouge, A., Perales, J., Dobos, K. M., Belisle, J. T., Spencer, J. S., & Brennan, P. J. (2004). Continued proteomic analysis of *Mycobacterium leprae* subcellular fractions. *Proteomics, 4*(10), 2942-2953.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography, 40*(4), 658-674.

Moreland, N. J., Ashton, R., Baker, H., Ivanovic, I., Patterson, S., Arcus, V., Baker, E., & Lott, J. (2005). A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallographica Section D, 61*, 1378-1385.

Morita, Y. S., Fukuda, T., Sena, C. B. C., Yamaryo-Botte, Y., McConville, M. J., & Kinoshita, T. (2011). Inositol lipid metabolism in mycobacteria: Biosynthesis and regulatory mechanisms. *Biochimica et Biophysica Acta (BBA) - General Subjects, 1810*(6), 630-641.

Murphy, D. J., & Brown, J. R. (2008). Novel drug target strategies against *Mycobacterium tuberculosis*. *Current Opinion in Microbiology, 11*(5), 422-427.

NCBI. (2012). Complete microbial genomes (Publication no. www.ncbi.nlm.nih.gov/genomes/lproks.cgi).

Nguyen, K. T., Piastro, K., Gray, T. A., & Derbyshire, K. M. (2010). Mycobacterial biofilms facilitate horizontal DNA transfer between strains of *Mycobacterium smegmatis*. *Journal of Bacteriology, 192*(19), 5134-5142.

Norman, E., De Smet, K. A., Stoker, N. G., Ratledge, C., Wheeler, P. R., & Dale, J. W. (1994). Lipid synthesis in mycobacteria: characterization of the biotin carboxyl carrier protein genes from *Mycobacterium leprae* and *M. tuberculosis*. *Journal of Bacteriology, 176*(9), 2525-2531.

O'Toole, G., Kaplan, H. B., & Kolter, R. (2000). Biofilm formation as microbial development. *Annual Review of Microbiology, 54*(1), 49-79.

O'Toole, G. A., & Kolter, R. (1998). Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Molecular Microbiology, 28*(3), 449-461.

Oftung, F., Mustafa, A. S., & Wiker, H. G. (2000). Extensive sequence homology between the *Mycobacterium leprae* LSR (12 kDa) antigen and its *Mycobacterium tuberculosis* counterpart. *FEMS Immunology and Medical Microbiology, 27*(1), 87-89.

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

Ojha, A., & Hatfull, G. F. (2007). The role of iron in *Mycobacterium smegmatis* biofilm formation: the exochelin siderophore is essential in limiting iron conditions for biofilm formation but not for planktonic growth. *Molecular Microbiology, 66*(2), 468-483.

Oliver, D. (1985). Protein secretion in *Escherichia coli*. *Annual Review of Microbiology, 39*(1), 615-648.

Parrish, N. M., Dick, J. D., & Bishai, W. R. (1998). Mechanisms of latency in *Mycobacterium tuberculosis*. *Trends in Microbiology, 6*(3), 107-112.

Perez-Iratxeta, C., & Andrade-Navarro, M. A. (2008). K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Structural Biology, 8*(25).

Recht, J., & Kolter, R. (2001). Glycopeptidolipid acetylation affects sliding motility and biofilm formation in *Mycobacterium smegmatis*. *Journal of Bacteriology, 183*(19), 5718-5724.

Recht, J., Martinez, A., Torello, S., & Kolter, R. (2000). Genetic analysis of sliding motility in *Mycobacterium smegmatis*. *Journal of Bacteriology, 182*(15), 4348-4351.

Renzoni, D., Esposito, D., Pfuhl, M., Hinton, J. C. D., Higgins, C. F., Driscoll, P. C., & Ladbury, J. E. (2001). Structural characterization of the N-terminal oligomerization domain of the bacterial chromatin-structuring protein, H-NS. *Journal of Molecular Biology, 306*(5), 1127-1137.

Rimsky, S. (2004). Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Current Opinion in Microbiology, 7*, 109-114.

Robson, J., McKenzie, J. L., Cursons, R., Cook, G. M., & Arcus, V. L. (2009). The vapBC Operon from *Mycobacterium smegmatis* is an autoregulated toxin-antitoxin module that controls growth via inhibition of translation. *Journal of Molecular Biology, 390*(3), 353-367.

Rodriguez, D., Sammito, M., Meindl, K., de Ilarduya, I. M., Potratz, M., Sheldrick, G. M., & Uson, I. (2012). Practical structure solution with ARCIMBOLDO. *Acta Crystallographica Section D, 68*(4), 336-343.

Rodriguez, D. D., Grosse, C., Himmel, S., Gonzalez, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M., & Uson, I. (2009). Crystallographic *ab initio* protein structure solution below atomic resolution. *Nature Methods, 6*(9), 651-653.

Röse, L., Kaufmann, S. H. E., & Daugelat, S. (2004). Involvement of *Mycobacterium smegmatis* undecaprenyl phosphokinase in biofilm and smegma formation. *Microbes and Infection, 6*(11), 965-971.

Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual* (Third ed.). New York: Cold Spring Harbor Laboratory Press.

Saraswathi, R., Pait Chowdhury, R., Williams, S. M., Ghatak, P., & Chatterji, D. (2009). The mycobacterial MsDps2 protein is a nucleoid-forming DNA binding protein regulated by sigma factors σA and σB. *PLoS ONE, 4*(11), e8017.

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology, 48*(1), 77-84.

Schluger, N. W. (2005). The pathogenesis of tuberculosis: the first one hundred (and twenty-three) years. *American Journal of Respiratory Cell and Molecular Biology, 32*(4), 251-256.

Schneider, R., Lurz, R., Lüder, G., Tolksdorf, C., Travers, A., & Muskhelishvili, G. (2001). An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Research, 29*(24), 5107-5114.

Schuster-Bockler, B., Schultz, J., & Rahmann, S. (2004). HMM Logos for visualization of protein families. *BMC Bioinformatics, 5*(1), 7.

Sela, S., Thole, J. E., Ottenhoff, T. H., & Clark-Curtiss, J. E. (1991). Identification of *Mycobacterium leprae* antigens from a cosmid library: characterization of a 15-kilodalton antigen that is recognized by both the humoral and cellular immune systems in leprosy patients. *Infection and Immunity, 59*(11), 4117-4124.

Sharadamma, N., Harshavardhana, Y., Singh, P., & Muniyappa, K. (2010). *Mycobacterium tuberculosis* nucleoid-associated DNA-binding protein H-NS binds with high-affinity to the Holliday junction and inhibits strand exchange promoted by RecA protein. *Nucleic Acids Research, 38*(11), 3555-3569.

Sheldrick, G. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallographica Section D, 66*(4), 479-485.

Shiloh, M. U., & DiGiuseppe Champion, P. A. (2010). To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? *Current Opinion in Microbiology, 13*(1), 86-92.

Shindo, H., Furubayashi, A., Shimizu, M., Miyake, M., & Imamoto, F. (1992). Preferential binding of *E. coli* histone-like protein HUα to negatively supercoiled DNA. *Nucleic Acids Research, 20*(7), 1553-1558.

Shindo, H., Iwaki, T., Ieda, R., Kurumizaka, H., Ueguchi, C., Mizuno, T., Morikawa, S., Nakamura, H., & Kuboniwa, H. (1995). Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli*. *FEBS Letters, 360*(2), 125-131.

Singh, J., Upshur, R., & Padayatchi, N. (2007). XDR-TB in South Africa: No time for denial or complacency. *PLoS Medicine, 4*(1), e50.

Singh, S., Jenner, P., Shanker Narayan, N., Ramu, G., Colston, M., Prasad, H., & Nath, I. (1994). Critical residues of the *Mycobacterium leprae* LSR recombinant protein discriminate clinical activity in erythema nodosum leprosum reactions. *Infection and Immunity, 62*(12), 5702-5705.

Singh, S., Narayanan, N. P., Jenner, P. J., Ramu, G., Colston, M. J., Prasad, H. K., & Nath, I. (1994). Sera of leprosy patients with type 2 reactions recognize selective sequences in *Mycobacterium leprae* recombinant LSR protein. *Infection and Immunity, 62*(1), 86-90.

Smith, I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical Microbiology Reviews, 16*(3), 463-496.

Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S., & Musser, J. M. (1997). Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences, 94*(18), 9869-9874.

Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Young, D. B., & Butcher, P. D. (2002). Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology, 148*(10), 3129-3138.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification, 41*(1), 207-234.

Summers, E., Meindl, K., Uson, I., Mitra, A., Radjainia, M., Colangeli, R., Alland, D., & Arcus, V. (2012). The structure of the oligomerization domain of Lsr2 from *Mycobacterium tuberculosis* reveals a mechanism for chromosome organization and protection. *PLoS ONE*,7(6), e38542.

Tanaka, I., Appelt, K., Dijk, J., White, S. W., & Wilson, K. S. (1984). 3-Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature, 310*(5976), 376-381.

Tendeng, C., & Bertin, P. N. (2003). H-NS in Gram-negative bacteria: a family of multifaceted proteins. *TRENDS in Microbiology, 11*(11), 511-518.

Teng, R., & Dick, T. (2003). Isoniazid resistance of exponentially growing *Mycobacterium smegmatis* biofilm culture. *FEMS Microbiology Letters, 227*(2), 171-174.

Thanbichler, M., Wang, S. C., & Shapiro, L. (2005). The bacterial nucleoid: A highly organized and dynamic structure. *Journal of Cellular Biochemistry, 96*(3), 506-521.

Warner, D. F., & Mizrahi, V. (2006). Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. *Clinical Microbiology Reviews, 19*(3), 558-570.

Wayne, L. G., & Sohaskey, C. D. (2001). Nonreplicating persistence of *Mycobacterium tuberculosis*. *Annual Review of Microbiology, 55*(1), 139-163.

Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C., & Mattick, J. S. (2002). Extracellular DNA required for bacterial biofilm formation. *Science, 295*(5559), 1487.

White, S. W., Wilson, K. S., Appelt, K., & Tanaka, I. (1999). The high-resolution structure of DNA-binding protein HU from *Bacillus stearothermophilus*. *Acta Crystallographica Section D, 55*(4), 801-809.

WHO. (2008). *WHO report 2008: Global tuberculosis control: surveillance, planning, financing*. Geneva: World Health Organisation

WHO. (2010). *WHO report 2010: Global tuberculosis control*. Geneva

WHO. (2011). *WHO report 2011: Global tuberculosis control*. Geneva

Wong, C., Sridhara, S., Bardwell, J. C. A., & Jakob, U. (2000). Heating greatly speeds coomassie blue staining and destaining. *Biotechniques, 28*(3), 426-432.

Yuan, H. S., Finkel, S. E., Feng, J. A., Kaczor-Grzeskowiak, M., Johnson, R. C., & Dickerson, R. E. (1991). The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. *Proceedings of the National Academy of Sciences, 88*(21), 9558-9562.

Zhang, J., Zeuner, Y., Kleefeld, A., Unden, G., & Janshoff, A. (2004). Multiple site-specific binding of Fis protein to *Escherichia coli* nuoA-N promoter DNA and its impact on DNA topology visualised by means of scanning force microscopy. *ChemBioChem, 5*(9), 1286-1289.