Lsr2: an H-NS functional analog and global regulator of *Mycobacterium tuberculosis*

by

Blair Richard George Gordon

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Molecular Genetics

University of Toronto

© Copyright by Blair Gordon 2013
Mycobacterium tuberculosis (M. tb), the etiological agent of tuberculosis (TB), continues to be one of the leading global health challenges causing ~2 million deaths annually. In the majority of infected individuals, the bacteria establish a latent, asymptomatic infection capable of persisting for decades with 5-10% of infected individuals developing active disease in their lifetime. Currently it is estimated that one-third of the world’s population is latently infected, representing a large reservoir for disease reactivation and subsequent spread. Latent TB infection is a paucibacillary disease in which a small heterogeneous population of bacilli is present in the body. M. tb persisters, which are characterized by reduced or altered metabolic activity and enhanced drug tolerance, are thought to be the major contributor towards latent infection and disease relapse following chemotherapy; however, the molecular mechanisms governing persisters formation remain poorly understood.

My thesis concerns the characterization of the highly conserved DNA binding protein Lsr2 of mycobacteria. Previous biochemical study of Lsr2 revealed it exhibits DNA-bridging activity analogous to H-NS, an important nucleoid associated protein found in the proteobacteria.
Here I show using *in vivo* complementation assays that Lsr2 is functionally equivalent to H-NS, even though these proteins share no sequence similarity. I also present genetic evidence that Lsr2 is a global regulator of *M. tb* that acts primarily as a transcriptional repressor. Notably, I found that Lsr2 represses a large cohort of genes induced in *M.tb* during *in vitro* models of latency including genes implicated in persister formation. I also present evidence that *lsr2* is selectively inactivated during long-term hypoxia, a condition thought to be important for persister formation during latency. Lastly, I tested the *lsr2* deletion mutant in a mouse model of infection and found it had reduced growth relative to the WT but was still able to persist. Taken together my work implicates Lsr2 as a central regulator of persister formation and opens up exciting future research avenues on latent TB infection.
Acknowledgments

First and foremost, I would like to thank my supervisor, Dr. Jun Liu for his guidance and support over the past five years. Through the highs and the lows Jun has been there for me. Following his example, I feel I have developed as a critical thinker and problem solver.

I am also indebted to past and present Liu lab members: Vanessa Tran, Andrea Leung, Ming Li, Linru Wang, Howard Song, Alina Nakhamchik, Jackie Liu, Rachid Nessar, Anna Sintsova and Nathan Ma. I am grateful for your support, constructive criticism and sense of camaraderie we shared during our time together in the lab.

I also thank my supervisory committee members, Drs Barbara Funnell, Timothy Hughes and William Navarre. Their guidance and support has been invaluable throughout my graduate career. In particular, Will has been a great mentor and support with his expertise on H-NS. I am thankful for all of our impromptu chats as well as his career and life advice. The Navarre lab members, past and present, have been a great support and I consider many of them friends. I would like to acknowledge Sabrina Ali, Betty Zou, Steven Hersch, and Robin Imperial. As well, I also acknowledge past and present members of the Gray-Owen lab, many of whom were older than I was and served as role models, for sharing their knowledge and advice with me. This includes Michael Brooks, Nancy So, Wendy Dobson-Belaire and Gord McSheffrey.

Science is rarely an independent endeavour and so I would like to thank my collaborators along the way. This includes the Hughes Lab members who shared their technical knowledge and helped with analysis. Thank you to Drs Harm van Bakel, Atina Cote and Matthew Weirauch. I also thank Dr. Bin Xia and his graduate student Yifei Li from Peking University in Beijing for their fantastic structural work.

Lastly, I want to thank family and friends who have and continue to enrich my life outside of the lab. My parents Suzanne Fisk and Richard Gordon fostered my interest and curiosity in the natural world while growing up and have been instrumental in getting me to where I am today. Along with my siblings Graeme, Laura and Mark, they have been a great support for me through the tough times. My late Grandmother Marjorie Fisk imparted her love of learning and higher education and remains an inspiration to me. I want to thank all of the
friends I have made in Toronto both in and out of graduate school. Specifically I thank Rade Sajic, Erica Tong, Natasha Christie-Holmes, Mark Czuczman, Madhav Jagannathan (for organizing all the social events) and my mountain biking buddies Alejandro Llabres and Alex Bull. Thank you to my cat MacDuff who sat in my lap and kept me company while writing this volume.

Finally, I want to express my heartfelt gratitude to Sue Li for her unwavering support and her thoughtful insights. I could not have made it without you!
# Table of Contents

ACKNOWLEDGMENTS .......................................................................................................................... IV

TABLE OF CONTENTS ............................................................................................................................ VI

LIST OF TABLES .......................................................................................................................................... X

LIST OF ABBREVIATIONS ....................................................................................................................... XI

LIST OF FIGURES ...................................................................................................................................... XIII

CHAPTER 1: GENERAL INTRODUCTION.................................................................................................. 1

1.1 An introduction to mycobacteria ........................................................................................................... 2
  1.1.1 The pathogenic mycobacteria ........................................................................................................ 2
  1.1.2 The environmental mycobacteria .................................................................................................. 4

1.2 *Mycobacterium tuberculosis* pathogenesis and virulence factors ......................................................... 4
  1.2.1 Global burden of TB on human health ............................................................................................ 4
  1.2.2 Pathogenesis and immunology ....................................................................................................... 5
  1.2.3 Animals models of *M. tb* virulence and persistence .................................................................. 6
  1.2.4 The live attenuated vaccine BCG .................................................................................................... 9
  1.2.5 The RD1 deletion and the attenuation of BCG .............................................................................. 10
  1.2.6 Individual BCG strains are mutated for *M. tb* virulence factors ................................................ 10

1.3 Metabolism and dormancy .................................................................................................................. 12
  1.3.1 Dormancy ...................................................................................................................................... 12
  1.3.2 Evidence for dormancy and *M. tb* persisters during infection .................................................... 13
  1.3.3 Persisters and their molecular mechanisms of formation ............................................................. 13
  1.3.4 Oxygen and latency ....................................................................................................................... 15
  1.3.5 Carbon metabolism ...................................................................................................................... 17

1.4 The H-NS protein ............................................................................................................................... 21
  1.4.1 H-NS is a bacterial nucleoid-associated proteins .......................................................................... 21
  1.4.2 H-NS: a pleiotropic transcriptional regulator .................................................................................. 21
  1.4.3 Biochemical and structural properties of H-NS ......................................................................... 23
CHAPTER 2: LSR2 OF MYCOBACTERIUM REPRESENTS A NOVEL CLASS OF H-NS-LIKE PROTEINS

2.1 Abstract

2.2 Introduction

2.3 Results

2.3.1 The lsr2 gene of *M. tuberculosis* complements various phenotypes of the *E. coli* Δ*hns* mutant

2.3.2 The *hns* gene complements the phenotype of an *M. smegmatis* Δ*lsr2* mutant

2.3.3 Lsr2 binds specifically to H-NS-regulated genes

2.3.4 Repression of *hlyE* by Lsr2 is partially relieved by overexpression of *slyA*

2.4 Discussion

2.5 Experimental procedures

CHAPTER 3: LSR2 IS A NUCLEOID-ASSOCIATED PROTEIN THAT BINDS AT-RICH SEQUENCES AND REGULATES MULTIPLE GENES IN *MYCOBACTERIUM TUBERCULOSIS*

3.1 Abstract

3.2 Introduction

3.3 Results

3.3.1 Lsr2 binds AT-rich sequences globally in Mycobacterial genomes

3.3.2 Lsr2 binds genes involved in virulence and immunogenicity of *M. tb*

3.3.3 Generation of a *M. tuberculosis lsr2::hyg* mutant

3.3.4 Microarray analysis indicates Lsr2 functions predominantly as a global transcriptional repressor

3.3.5 The LSR2_UP is enriched for genes induced during persistence

3.3.6 Lsr2 regulates genes involved in stress response and persister formation

3.4 Discussion
CHAPTER 4: INVESTIGATING THE IMPACT OF LSR2 DYSFUNCTION ON IN VITRO PERSISTENCE, MACROPHAGE INFECTION, AND IN AN ANIMAL MODEL OF INFECTION ................................................................................................................................. 93

4.1 Abstract .......................................................................................................................................................... 94

4.2 Introduction .................................................................................................................................................. 95

4.3 Results .......................................................................................................................................................... 97

4.3.1 The loss of lsr2 in M. smegmatis confers a survival advantage during hypoxia-induced NRP .......... 97

4.3.2 M. smegmatis inactivates lsr2 via a transposon-mediated mechanism during hypoxia-induced NRP .......................................................... 100

4.3.3 The M. tb ∆lsr2 mutant induces rapid killing of RAW macrophages in vitro ..................................... 107

4.3.4 The M. tb ∆lsr2 mutant persists at a lower level than the WT in a murine model of infection ....... 109

4.4 Discussion .................................................................................................................................................. 112

4.5 Experimental procedures .......................................................................................................................... 116

CHAPTER 5: STRUCTURAL BASIS FOR RECOGNITION OF AT-RICH DNA BY UNRELATED XENOGENEIC SILENCING PROTEINS .................................................................................................................. 119

5.1 Abstract ...................................................................................................................................................... 120

5.2 Introduction .............................................................................................................................................. 121

5.3 Results ...................................................................................................................................................... 123

5.3.1 H-NS and Lsr2 Exhibit Similar DNA Binding Specificity ................................................................. 123

5.3.2 H-NS and Lsr2 Bind Contiguous AT Sequences ............................................................................. 126

5.3.3 DNA Minor Groove Width Determines the Binding Preference of H-NS and Lsr2 ..................... 126

5.3.4 Lsr2 has an N-Terminal Dimerization Domain and a C-Terminal DNA-Binding Domain ......... 129

5.3.5 DNA-Binding Domain of Lsr2 Exhibits a Unique Structure compared to the DNA-Binding Domain of H-NS .................................................. 132

5.3.6 Lsr2 and H-NS Bind the DNA Minor Groove using an AT-Hook–Like Loop ................................. 134

5.3.7 The Q/RGR Motif Is Essential for H-NS and Lsr2 Binding the DNA Minor Groove ................. 137

5.3.8 DNA Minor Groove Binding Reagents Compete for the Binding of H-NS to DNA .................. 137

5.3.9 Confirmation of the DNA-Binding Preference of H-NS by NMR .................................................. 138
5.3.10 Correlation with in vivo Binding Sites ................................................................. 138

5.4 Discussion ................................................................................................................. 141

5.5 Experimental Procedures .......................................................................................... 144

CHAPTER 6: SUMMARY AND FUTURE DIRECTIONS ................................................. 148

6.1 Lsr2 is functionally equivalent to H-NS ................................................................. 149

6.2 Lsr2 negatively regulates genes required for persistence ........................................ 153
   6.2.1 Does the Δlsr2 mutant of M. tb exhibit a fitness advantage over the WT strain under hypoxia?... 157
   6.2.2 Does the Δlsr2 mutant of M. tb display altered carbon metabolism and cellular redox energy
       status? ........................................................................................................................................ 160
   6.2.3 Is the Δlsr2 mutant predisposed to persister formation? ............................................... 161

6.3 Determining the role of Lsr2 during M. tb infection ............................................... 164
   6.3.1 Why are the Δlsr2-infected macrophages killed? ......................................................... 164
   6.3.2 Additional animal work ............................................................................................... 164
   6.3.3 Are lsr2 mutants present in human TB disease? ......................................................... 166

REFERENCES ............................................................................................................... 167
List of Tables

**Table 3.1.** Functional categories of Lsr2-bound genes in *M. tb* ................................. 59

**Table 3.2.** Functional categories of genes significantly up-regulated and downregulated in the *M. tb* ∆*lsr2* mutant ................................................................. 68

**Table 3.3.** Comparison of LSR2_UP with previously published microarray analysis of bacteriostasis ........................................................................... 71

**Table 3.4.** The experimentally validated TA systems that are activated in *M. tbΔlsr2* .......... 73

**Table 4.1.** Genetic analysis of spontaneous *Mycobacterium smegmatis lsr2* mutants selected from long-term hypoxia survival experiments ........................................... 104

**Table 5.1.** Summary of *in vivo* complementation experiments for chimeric proteins of Lsr2 and H-NS ........................................................................... 130
List of Abbreviations

BCG – bacilli de Calmette Guerin

CFU – colony forming units

ChIP – chromatin immunoprecipitation

CIP – ciprofloxacin

EHR – enduring hypoxic response

giv – growth in vivo mutants

HA – hemagglutinin

HGT – horizontal gene transfer

HIV – human immunodeficiency virus

H-NS – heat stable nucleoid structuring protein

INH – isoniazid

IS – insertion sequence

Lsr2 – leprosy serum reactive clone 2

M.W. – molecular weight

Mbp – megabase pair

MDR-TB – multidrug resistant tuberculosis

MOI – multiplicity of infection

MTBC – mycobacterium tuberculosis complex

NAP – nucleoid associated protein
NMR – nuclear magnetic resonance
NRP – non-replicating persistence
PAT – polyacyltrehaloses
PDIM - phthiocerol dimycocerosate
per – persistence mutants
PGL – phenolic glycolipids
RIF – rifampin
RNAP – ribonucleic acid polymerase
Rpf – resuscitation-promoting factor
RT-PCR – real-time polymerase chain reaction
SDS – sodium dodecyl sulfate
sgiv – severe growth in vivo mutants
SL-1 – sulfolipid 1
TA – toxin-antitoxin
TAG – triacylglycerol
TB – tuberculosis
TCA – tricarboxylic acid cycle
VBNC – viable but not culturable
WT – wild type
XDR-TB – extensively drug resistant tuberculosis
List of Figures

Figure 1.1. Growth characteristics of the different types of \textit{M. tb} mutants in a murine model of virulence ................................................................. 8

Figure 1.2. Overview of \textit{M. tb} carbon metabolism ................................................................. 20

Figure 1.3. H-NS oligomerization requires two dimerization interfaces ......................... 25

Figure 2.1. Western blot analysis of Lsr2 expression in \textit{E. coli} .............................................. 34

Figure 2.2. Lsr2 complements various phenotypes in an \textit{E. coli} \textit{\Delta hns} mutant ............ 36

Figure 2.3. Complementation of the colony morphology of the \textit{\Delta lsr2} mutant of \textit{M. smegmatis} by \textit{hns} ................................................................. 39

Figure 2.4. Lsr2 binds specifically to H-NS-regulated genes .............................................. 41

Figure 2.5. Overexpression of \textit{slyA} causes partial derepression of \textit{hlyE} .................... 43

Figure 2.6. Sequence alignment and secondary structure prediction of Lsr2 proteins in various actinomycetes ......................................................... 45

Figure 3.1. Lsr2 binds preferentially to high-AT-content regions of the genome, including horizontally transferred genes and virulence-associated genes in \textit{M. tb} ............................................. 58

Figure 3.2. Southern blot confirmation of \textit{M. tb} \textit{\Delta lsr2} deletion mutant ..................... 61

Figure 3.3. The \textit{M. tb} H37Rv \textit{\Delta lsr2} mutant-associated phenotypes ............................. 65

Figure 3.4. Overlap of genes upregulated in \textit{M. tb} \textit{\Delta lsr2} and genes of WT \textit{M. tb} induced by hypoxic conditions ................................................................. 72

Figure 3.5. Metabolism of cholesterol and methyl-branched or odd-chain fatty acids ....... 76

Figure 3.6. The cholesterol \(\beta\)-oxidation pathway is activated in \textit{M. tb} \textit{\Delta lsr2}. ............. 77
Figure 3.7. Methylcitrate cycle and methylmalonyl pathway are activated in ∆lsr2 of M. tb. ....................................................................................................................................................... 78

Figure 3.8. The biosynthetic pathway of PDIM and PGL................................................................. 80

Figure 3.9. Electron transport chain in M. tb under aerobic or hypoxic conditions. ........ 83

Figure 4.1. The ∆lsr2 of M. smegmatis exhibits a fitness advantage over the WT strain under long-term hypoxia........................................................................................................... 99

Figure 4.2. The loss of lsr2 confers a fitness advantage in M. smegmatis during hypoxic NRP. ........................................................................................................................................... 103

Figure 4.3. The WT is more fit than the ∆lsr2 under aerobic conditions. ......................... 106

Figure 4.4. Intracellular growth of the ∆lsr2 mutant in macrophages. ......................... 108

Figure 4.5. Growth of the M. tb ∆lsr2 in BALB/c mice.............................................................. 111

Figure 5.1. H-NS and Lsr2 exhibit similar DNA binding specificity and bind contiguous AT sequences......................................................................................................................... 125

Figure 5.2. Effects of A-tracts on H-NS and Lsr2 binding to DNA.............................. 128

Figure 5.3. Structure–function analysis of Lsr2........................................................................ 131

Figure 5.4. Solution structures of the Lsr2Ctd and H-NSCtd. ........................................ 133

Figure 5.5. Structure models of Lsr2Ctd/DNA, H-NSCtd/DNA and Bv3F Ctd/DNA complexes generated by HADDOCK 2.0.................................................................................................. 136

Figure 5.6. Correlation with in vivo binding sites.............................................................. 140

Figure 6.1. Sequence alignment of Lsr2 proteins from M. tb, M. smegmatis, and various Mycobacterium phages......................................................................................................................... 152

Figure 6.2. Lsr2’s regulatory role in a model for human M. tb infection....................... 155
Figure 6.3. Working model. ..................................................................................................... 156

Figure 6.4. The Δlsr2 of *M. tb* exhibits comparable viability to the WT strain under long-term hypoxia. ............................................................................................................................. 159

Figure 6.5. The Δlsr2 of *M. tb* exhibits a drug tolerance phenotype *in vitro*. ....................... 163
CHAPTER 1: GENERAL INTRODUCTION
1.1 An introduction to mycobacteria

The genus *Mycobacterium* consists of rod-shaped prokaryotes belonging to the class of actinobacteria which includes closely related *Streptomyces, Nocardia* and *Corynebacteria* genera. A characteristic of mycobacteria is the lipid-rich cell wall comprised of many complex lipids including the exceptionally long-chained mycolic acids. Owing to this hydrophobic cell envelope, mycobacteria are highly impermeable to antibiotics, clump when cultured in liquid media lacking detergents and cannot be identified using the Gram-stain. Instead the Ziehl–Neelsen stain (also known as the acid-fast stain) is used to test for the presence of mycobacteria, which resist decolourization with acid-alcohol and stain red, the colour of the initial carbol fushsin stain. Both the *Mycobacterium* and *Nocardia* genera are said to be acid-fast; meaning they are positive for the Ziehl–Neelsen stain. All other bacteria will stain blue with the counterstain methylene blue.

Mycobacterial genomic DNA has a high GC-content, ranging from 58-68% (Parish and Stoker 1998). Mycobacteria are slower growing than other prokaryotes with species requiring between a few days to weeks to form colonies on agar plates. Mycobacteria can be broadly classified into two groups: the slow-growers with generation times greater than 12 hours and the fast-growers with generation times ranging from 3 to 5 hours. The slow-growers include major human and animal pathogens while the fast-growers include nonpathogenic species that occur in a wide variety of environments (Parish and Stoker 1998). The defining features of this genus, along with the biosafety level-3 containment requirement for some pathogenic species, have made the study of these microorganisms difficult.

1.1.1 The pathogenic mycobacteria

The pathogenic mycobacteria include organisms that can infect most species of animals including mammals, amphibians, birds and fish. The *Mycobacterium tuberculosis* complex (MTBC) comprises several subspecies (*Mycobacterium tuberculosis, Mycobacterium bovis, Mycobacterium microti and Mycobacterium africanum*) that share 99.9% nucleotide similarity, are facultative intracellular pathogens and cause tuberculosis disease (TB) in mammals.
Mycobacterium tuberculosis (M. tb) is the most common causative agent of tuberculosis disease in humans. Its 4.4 Mbp genome is sequenced, which has facilitated deciphering the pathogenesis of this important pathogen (Cole, Brosch et al. 1998). M. africanum is most commonly found in West African countries and is observed to be less pathogenic than M. tb (Brosch, Gordon et al. 2000). In contrast, M. bovis has a broader host range causing TB in animals (such as cattle, sheep, deer, badgers etc) including humans, and was once a common source of spreading TB via unpasteurized milk (Brosch, Gordon et al. 2000). M. microti infects and causes TB in rodents but is avirulent in humans (Brosch, Gordon et al. 2000). M. leprae is a close relative of the M. tb complex. It is an extremely slow-growing obligate intracellular pathogen responsible for causing leprosy, a disfiguring disease that affects people primarily in the developing world (Cole, Eiglmeier et al. 2001; Monot, Honore et al. 2005). Comparing the smaller 3.2 Mbp M. leprae genome sequence with that of M. tb reveals it has undergone reductive evolution as evidenced by gene deletion and gene decay leading to the elimination of many metabolic activities (Cole, Eiglmeier et al. 2001). After tuberculosis and leprosy, the necrotizing skin infection Buruli ulcer is the third most common human mycobacterial infection worldwide (Silva, Portaels et al. 2009). It is an emergent disease endemic to humid tropical regions of Africa, South America and Australia and is caused by M. ulcerans, the only known toxin-producing mycobacterium (Marsollier, Brodin et al. 2007; Demangel, Stinear et al. 2009; Silva, Portaels et al. 2009).

The nontuberculous pathogenic mycobacteria are increasingly being recognized for their potential as opportunistic pathogens (e.g., M. avium and M. kansasii) in HIV-infected and immunocompromised individuals (Falkinham 1996; Primm, Lucero et al. 2004). Also M. marinum, a fish and amphibian pathogen, causes cutaneous granulomas associated with cuts or abrasions exposed to water or fish; however in AIDS patients, pulmonary and or gastrointestinal infection are often observed (Falkinham 1996; Primm, Lucero et al. 2004). Since M. marinum is the most closely related species to M. tb outside the MTBC, has a shorter doubling time (8 hours versus 24 hours for M. tb), and is a biosafety level-2 pathogen it is sometimes used as a model system for studying mycobacterial pathogenesis (Alexander, Jones et al. 2004; Tan, Lee et al. 2006; Davis and Ramakrishnan 2009; Volkman, Pozos et al. 2010).
1.1.2 The environmental mycobacteria

Examples of the fast-growing mycobacteria include \textit{M. smegmatis}, \textit{M. fortuitum}, \textit{M. phlei} and \textit{M. chelonae} (Parish and Stoker 1998). \textit{M. smegmatis} is used as a genetic host for studies of pathogenic mycobacteria because it is nonpathogenic, well-dispersed in liquid media containing detergent, has simple media requirements and grows 10 times faster than \textit{M. tb} (Jacobs Jr 2000). In fact, the early development of mycobacterial genetic tools such as phage transduction systems, transposon systems and plasmids relied upon \textit{M. smegmatis} as a surrogate system (Jacobs Jr 2000). The 6.2 Mbp genome of the most widely used \textit{M. smegmatis} strain, mc²155, has been sequenced and remains instrumental towards providing insights into mycobacterial biology (Deshayes, Perrodou et al. 2007).

1.2 \textit{Mycobacterium tuberculosis} pathogenesis and virulence factors

1.2.1 Global burden of TB on human health

Human TB remains one of the world’s major health challenges, with 8.8 million newly active cases diagnosed and 1.45 million deaths recorded in 2010 (Anonymous 2011). Furthermore, it is estimated that 2 billion people, approximately one-third of the world’s population, are latently infected representing a large reservoir for disease reactivation and spread (Corbett, Watt et al. 2003). Effective TB-control is complicated by the ability of \textit{M. tb} to become dormant and persist for extended periods in the host, but remain able to reactivate in response to decreased immune surveillance and/or other unknown factors. As well, the dormant bacilli present during latent disease are in a reduced, altered metabolic state making them innately resistant to antibiotics. As a result TB treatment requires lengthy drug-treatment regimens with multiple drugs administered for 6 to 9 months. The situation has become exacerbated by TB/HIV co-infection as well as the increasing incidences of multi-drug resistant (MDR-TB) and extremely drug-resistant (XDR-TB) (Anonymous 2011).
1.2.2 Pathogenesis and immunology

*M. tb* transmission occurs when an individual with active TB infection coughs, releasing droplets harbouring active bacilli which are inhaled by a naïve host. In a small percentage of infections, *M. tb* is effectively cleared after an adaptive immune response is mounted. However, in the majority of infected individuals, the bacteria establish a latent, asymptomatic infection that can persist for decades (Stewart, Robertson et al. 2003; North and Jung 2004) with 5-10% of latently infected individuals developing active disease in their lifetime. Latent infection is detected using the tuberculin skin test, which involves the intradermal injection of mycobacterial antigens and assaying for a positive immune reaction. During latency *M. tb* bacilli are undetectable in patient sputum, suggesting a low bacterial burden (paucibacillary state). Host immunosuppression (e.g. HIV coinfection) markedly increases the risk for reactivation (Corbett, Bandason et al. 2007) to active clinical disease with symptoms of fever, severe cough, malaise and weight loss. Left untreated, death usually occurs due to suffocation from excess fluid present in the lungs or respiratory failure due to excess tissue destruction in the lungs. In some cases *M. tb* can attack other parts of the body leading to organ failure.

Early TB infection follows a relatively reproducible course with the pathogen being phagocytosed by alveolar macrophages or dendritic cells (Leemans, Juffermans et al. 2001; Wolf, Linas et al. 2007). Once inside the phagosome, *M. tb* is able to subvert immune-mediated destruction by avoiding phagosomal acidification, inhibiting apoptosis, neutralizing toxic superoxides produced by immune cells and preventing phagosome-lysosome fusion (Russell 2007). Infected macrophages and dendritic cells then migrate to adjacent lymph nodes, where they present mycobacterial antigens to elicit a T-helper-1 (TH1)-type response (Flynn and Chan 2001; Russell 2007). The activated T-cells return to the lungs and strengthen antibacterial responses by recruiting and activating additional macrophages (Kaufmann 2001). This complex immune response results in the formation of a granuloma: a stratified structure comprised of a central area of mycobacterial-infected cells surrounded by other, non-infected phagocytes, multinucleated giant cells and foamy macrophages, with lymphocytes found at the periphery. The lesion is sealed off from surrounding tissue by a fibrotic capsule and other extracellular matrix proteins (Russell, VanderVen et al. 2010; Flynn, Chan et al. 2011). Interferon-γ (IFN-γ) and tumour-necrosis factor-α (TNF-α) are the dominant cytokines that activate macrophages and
maintain containment of *M. tb* within the granuloma (Flynn and Chan 2001; Russell 2007). It is thought that the persistent bacilli present during latent infection reside within the caseous necrotic centres of granulomas; they lack vascular markers, which is thought to benefit the host by depriving mycobacteria of oxygen and nutrients. However, this classical model in which the granuloma solely benefits the host by walling off the infection is being reexamined. Recent evidence using *M. marinum* in a zebra fish model of infection demonstrated that early granuloma formation benefits the bacteria by recruiting uninfected macrophages to provide additional environments for bacterial growth and population expansion (Davis and Ramakrishnan 2009). It has also been noted that within infected individuals, even those with active disease, that granulomas are pathologically heterogeneous in terms of their developmental stage. They range from sterility, to containment, to cavitation, with the latter occurring when the granuloma wall breaks down to released actively replicating bacteria into the airways thereby facilitating transmission (Boshoff and Barry 2005; Russell 2007). This contrasts with the classically held view that reactivation is purely the result of host immune control failure. Rather the outcomes that culminate in progression to active disease, accompanied by granuloma maturation to cavitation, seem to be determined at the level of each individual granuloma and suggest an active role for *M. tb* in the reactivation process.

### 1.2.3 Animals models of *M. tb* virulence and persistence

As previously outlined above, the natural history of *M. tb* infection in humans has distinct phases of pathogenesis: initial exposure and bacterial replication, control of infection leading to the establishment of clinical latency, and in some cases reactivation to active disease. There is no adequate animal model that recapitulates this complex disease progression, especially the latent stage (except for non-human primates but they are rarely used due to expense and ethical restrictions) (Boshoff and Barry 2005). Nonetheless, animal models have been instrumental towards understanding *M. tb* virulence and evaluating vaccine efficacy (Glickman and Jacobs 2001).

The animal models used to study *M. tb* virulence and pathogenesis include the mouse, guinea pig, rabbit, and monkey, with each having their own caveats (Chao and Rubin 2010). The murine model is the most commonly used due to its low cost, the availability of genetically
defined mouse strains and immunological reagents. Typically, mice are aerosolized with a low dose (10-100 CFU) of viable bacilli (Orme 1988; Kelly, Furney et al. 1996; North and Jung 2004; Beamer and Turner 2005). Acute infection ensues with bacteria replicating unabated for 3-4 weeks (generation time ~20 hours) (Gill, Harik et al. 2009), but then plateauing at the onset of adaptive immunity. In contrast to human latency, during this chronic stage of infection mice maintain a relatively high bacterial burden of between $10^5$ to $10^6$ CFU in the lungs. The nature of the bacilli population present in chronically infected mice is a matter of debate with some claiming it is a static non-replicating population (Kaprelyants and Kell 1993; Oliver 2005). However, a recent report, in which an unstable plasmid with a predictable loss-rate was used as a ‘molecular clock’, suggests that chronic stage bacilli exhibit reduced growth and metabolic activity (generation time ~100 hours) (Gill, Harik et al. 2009). Thus, the characteristic plateau seen during chronic infection likely represents equilibrium between the growth of bacteria and killing by the immune system. The mice are able to control the infection for 9 to 12 months (Lewis 2010; Yamaguchi, Park et al. 2011) before the bacterial burden again increases and they succumb to TB disease. This low-dose murine infection model has the advantage of approximating natural latency in that it relies solely on the host immune response for control of the infection, but the chronic stage contrasts sharply to the paucibacillary state of human latency. Nevertheless, it is reasoned that specific genes required for murine persistence are also important during human latency (Glickman and Jacobs 2001). Studies evaluating the in vivo growth of M. tb in mice have noted different growth phenotypes for mutant strains including those that: fail to grow (severe growth in vivo mutants, sgiv), replicate less than WT (growth in vivo mutants, giv) (Cox, Chen et al. 1999) or can replicate initially but cannot persist (persistence mutants, per) (McKinney, Honer zu Bentrup et al. 2000) (reviewed in (Hingley-Wilson, Sambandamurthy et al. 2003))(Figure 1.1).
Figure 1.1. Growth characteristics of the different types of *M. tb* mutants in a murine model of virulence (from Glickman and Jacobs, 2003, Cell 104, 477–485). WT *M. tb* (yellow) grows unabated until an adaptive immune response is mounted at 3 weeks. A constant bacterial burden is maintained during the chronic stage of infection. The observed mutants are as follows: *giv* (growth in vivo mutant), *sgiv* (severe growth in vivo mutant), and *per* (persistence mutant).

The Cornell model is an alternative mouse model that uses drugs to treat *M. tb*-infected mice to achieve apparent culture-negative organ sterility, analogous to the paucibacillary state of latent TB in humans (McCune, McDermott et al. 1956; McCune and Tompsett 1956). However, 3 months following cessation of antibiotics, approximately one-third of the mice spontaneously reactivate with drug-susceptible TB. This drug-induced model of latency has the advantage of achieving a very low or undetectable persister population of bacilli similar to those present in some human lesions, but has the disadvantage of being artificially induced. Despite their drawbacks, studies using either of these two murine models have yielded important information concerning TB pathogenesis (Adams, Mason et al. 1995; MacMicking, North et al. 1997; Flynn, Scanga et al. 1998), and when supplemented with *in vitro* models of latency, will continue to generate new insights into latent infection.

Aerosol infection of mice does result in granuloma formation, but they are not the well-formed structures seen in humans. In addition recent evidence indicates that TB lesions in mouse lungs are not hypoxic, nor do they exhibit the caseation and cavitation seen in human TB (Brooks, Furney et al. 1999; Aly, Wagner et al. 2006; Tsai, Chakravarty et al. 2006; Via, Lin et
The lesions observed in the guinea pig and rabbit models display caseating necrosis (McMurray, Collins et al. 1996; Dannenberg 2001) and hypoxia (Via, Lin et al. 2008). However, guinea pigs are highly sensitive to \textit{M. tb}, with no discernible latent phase, while rabbits must be infected with \textit{M. bovis} because they are resistant to \textit{M. tb} (Flynn 2006). The latter model has not been validated for studying latent disease and there are genetic differences between \textit{M. tb} and \textit{M. bovis} (Garnier, Eiglmeier et al. 2003) which may have unknown effects on disease progression (Boshoff and Barry 2005; Flynn 2006). The non-human primate model is the only model that truly mimics latent TB in humans. Low dose infection of cynomolgus macaques directly into the lungs results in latent infection (Capuano, Croix et al. 2003) as well as granulomas that are microscopically and immunologically similar to those in humans (Walsh, Tan et al. 1996; Langermans, Andersen et al. 2001; Fuller, Flynn et al. 2003). A number of factors have precluded the use of the non-human primate model such as ethical concerns, exorbitant cost and difficulty housing the animals in a BSL3 setting.

1.2.4 The live attenuated vaccine BCG

Currently the only available vaccine against TB is \textit{Mycobacterium bovis} bacilli Calmette-Guerin (BCG). It originated with the isolation of a virulent \textit{M. bovis} strain from an infected cow’s milk near the end of the nineteenth century. This strain was taken to the Pasteur institute where Albert Calmette and Camille Guerin began using the strain to study bovine tuberculosis pathogenesis. Given Louis Pasteur’s work that laboratory adaptation of pathogenic bacteria often correlates with decreased virulence, they serially passaged this strain over 13 years and monitored it for sustained decreases in virulence using animal models of infection (Brewer and Colditz 1995; Behr and Small 1999; Behr 2002). After a successful clinical trial in the early 1920’s showing that BCG protected children against TB, the vaccine was distributed worldwide with individual countries maintaining their own BCG strains via \textit{in vitro} passaging, and hence continuing the \textit{in vitro} evolution of BCG until the advent of freeze seed-stocks in 1960. As a result, many BCG daughter substrains display phenotypic and genetic heterogeneity leading to overattenuation which may have reduced their ability to protect against TB (Behr and Small 1999; Behr, Wilson et al. 1999; Behr 2002; Chen, Alexander et al. 2003; Chen, Islam et al. 2007; Leung, Tran et al. 2008). In fact, the efficacy of BCG has been called into question with protective efficacy against adult pulmonary disease ranging from 0 to 80% (Colditz, Brewer et
One hypothesis explaining the reduced efficacy that has gained traction concerns the heterogeneity of BCG strains (Behr 2002).

1.2.5 The RD1 deletion and the attenuation of BCG

In recent years, the mechanisms of BCG attenuation have been studied and it is evident that BCG strains represent natural mutants of major virulence factors of \( M. \text{tb} \). Comparative genomic analysis between \( M. \text{bovis} \) and BCG revealed that several chromosomal deletions termed regions of difference (RD) are present in the latter. Among these, RD1 is the only region absent from all BCG strains (Mahairas, Sabo et al. 1996; Behr, Wilson et al. 1999). RD1 is present in all virulent members of the \( M. \text{tb} \) complex (Gordon, Heym et al. 1999), but is partially deleted in \( M. \text{microti} \) which explains its avirulence in humans (Brodin, Eiglmeier et al. 2002). These studies strongly implicated the RD1 region as a principal MTBC virulence factor. Subsequently, this was confirmed by the observation that deletion of RD1 from \( M. \text{tb} \) caused attenuation of virulence in mice (Hsu, Hingley-Wilson et al. 2003).

The RD1 region encodes for the ESX-1 type VII secretion system that secretes two small proteins ESAT-6 and CFP-10. Deletion of \( \text{esat-6} \) and \( \text{cfp-10} \), or of genes encoding other components of the secretory apparatus in \( M. \text{tb} \) results in impaired growth in macrophages and attenuation in mouse models of infection to a similar extent as that observed for deletion of the entire RD1 (Hsu, Hingley-Wilson et al. 2003; Stanley, Raghavan et al. 2003; Guinn, Hickey et al. 2004). Conversely, introduction of the RD1 region from \( M. \text{tb} \) into BCG or \( M. \text{microti} \) only partially restores virulence, indicating that RD1 cannot fully account for the attenuation of BCG and that other genetic lesions are involved (Pym, Brodin et al. 2002; Lewis, Liao et al. 2003; Pym, Brodin et al. 2003).

1.2.6 Individual BCG strains are mutated for \( M. \text{tb} \) virulence factors

Additional mutations that contribute to the attenuation of BCG have been identified. For instance, some BCG strains harbor mutations within the biosynthetic locus for phthiocerol dimycocerosates (PDIMs) and phenolic glycolipids (PGLs) (Leung, Tran et al. 2008), two cell wall lipids known to be important for the virulence of \( M. \text{tb} \) and \( M. \text{bovis} \) (Cox, Chen et al. 1999; Reed, Domenech et al. 2004; Hotter, Wards et al. 2005), and consequently do not produce these
Polymorphisms are also present in the PhoP-PhoR two-component system in some BCG strains (Leung, Tran et al. 2008). The response regulator PhoP controls a variety of functions including: stress response, hypoxia response, respiratory metabolism, the synthesis of virulence lipids and ESX-1 secretion (Chesne-Seck, Barilone et al. 2008)(Rivero, Marquez et al. 2001; Soto, Menendez et al. 2004; Martin, Williams et al. 2006; Gonzalo-Asensio, Mostowy et al. 2008; Lee, Krause et al. 2008). This system appears to play an essential role in M. tb virulence with a phoP mutation partially accounting for the attenuation of the H37Ra strain of M. tb(Martin, Williams et al. 2006; Frigui, Bottai et al. 2008; Lee, Krause et al. 2008). Furthermore, a transposon insertion in the phoP promoter driving its overexpression was found in a clinical strain of M. bovis responsible for a severe outbreak of TB disease in humans in Spain (Rivero, Marquez et al. 2001; Soto, Menendez et al. 2004).

The unique and complex metabolism of M. tb is thought to be essential for its survival in the host, especially during the later stages of infection within granulomatous lesions. Accordingly, mutations affecting metabolism in BCG have been identified. For example, it was shown that BCG vaccine strains harbor mutations in genes required for catabolizing certain amino acids leading to misregulation of nitrogen metabolism (Chen, Alexander et al. 2003). Mutations disrupting the cholesterol import system operon (mce4) have been identified in some BCG strains (Leung, Tran et al. 2008). What has become increasingly apparent is that M. tb depends on lipids during host infection. This is reinforced by the attenuation of the M. tbΔmce4 in both activated macrophages and mice, confirming that cholesterol metabolism is important during the chronic phase of infection (Pandey and Sassetti 2008). Two closely related BCG strains contained deletions in the gene encoding for the iron-sulfur cluster containing transcriptional regulator WhiB3, which is required by M. bovis for in vivo growth in guinea pigs (Steyn, Collins et al. 2002). WhiB3 is a redox sensing protein responding to oxygen and nitric oxide, regulating carbon metabolism and virulence cell wall lipid anabolism (Singh, Guidry et al. 2007; Singh, Crossman et al. 2009). The whiB gene in Streptomyces coelicolor is essential for initiation of sporulation septation (Schwedock, McCormick et al. 1997). As such it has been postulated that M. tb persistence is analogous to bacterial sporulation and that whiB3 may play a role in mediating this switch (Soliveri, Gomez et al. 2000).
1.3 Metabolism and dormancy

1.3.1 Dormancy

Dormancy refers to a spectrum of non-replicating persistent (NRP) states, in which bacteria have reduced metabolic activity (e.g., transcription and translation), often following exposure to environmental stress conditions. For some bacteria, there are *in vitro* conditions that arrest replication but allow the bacteria to remain culturable under standard conditions whereas other species form spores via a defined developmental program, putting them into a state of metabolic inactivity and extreme resistance to adverse conditions. The viable but not culturable (VBNC) phenomenon was first described in *Vibrio* species, but is now known to occur in other nonsporulating species. VBNC refers to the inability to recover bacterial colonies by plating on standard nutrient agar, even though the bacteria remain viable, as determined by vital staining (Kaprelyants and Kell 1993). In addition these bacteria can be resuscitated via the addition of exogenous purified factors or spent media from actively dividing cultures (Oliver 2005). A VBNC state can be induced in *M. tb* (Shleeva, Bagramyan et al. 2002) and *M. smegmatis* (Shleeva, Mukamolova et al. 2004) when incubated in stationary phase for extended periods of time. In both instances the VBNC cells adopted a smaller, coccoid appearance and could be cultured with the addition of supernatant from actively growing cultures or resuscitation-promoting factor (Rpf). The metabolic conditions responsible for the VBNC state described above remain largely unknown.

There are several *in vitro* models that are thought to mimic some aspects of the *in vivo* conditions experienced by *M. tb* during chronic and latent infections, including nutrient starvation (Betts, Lukey et al. 2002) and hypoxia (Wayne and Hayes 1996; Voskuil, Visconti et al. 2004; Rustad, Harrell et al. 2008). Both dormancy models induce a rapidly reversible growth arrest and have been extensively characterized. Dormant persister cells have been observed to form stochastically in exponentially growing *M. tb* cultures and their transcriptome revealed a distinct pattern of dormancy (Keren, Minami et al. 2011). The findings and implications of these various dormant cells will be discussed in more detail below.
1.3.2 Evidence for dormancy and *M. tb* persisters during infection

TB is a complex disease caused by distinct bacterial subpopulations that reside within discrete lesions of a single host. Studies of resected lung tissue from TB patients revealed that *in vivo* *M. tb* populations are heterogeneous; ranging from actively growing, culturable bacteria recovered from open cavities, to unculturable bacteria isolated from closed cavities (granulomas) that had lost their acid-fastness (Ulrichs and Kaufmann 2006) and that could not cause disease in guinea pigs (Medlar, Bernstein et al. 1952; Wayne 1960; Boshoff and Barry 2005; Manabe, Kesavan et al. 2008; Gideon and Flynn 2011). Whether these unculturable cells are in a VBNC state is unclear since various treatments were unsuccessful in resuscitating them (Boshoff and Barry 2005). Moreover, there is mounting evidence that the dormant persister bacilli present during latency have reduced or altered metabolic activity, significantly enhancing their drug tolerance (Boshoff and Barry 2005; Zhang 2005; Keren, Minami et al. 2011). Congruent with this notion is the observation that multidrug therapy of TB patients leads to relatively quick sterilization of bacteria in patient sputum by 8 weeks (Boshoff and Barry 2005); however, a high rate of disease relapse is associated with premature cessation of treatment at this point. It is thought that these persisters most likely have altered their metabolism in response to the sequestered environment of the granuloma which restricts bacterial access to carbohydrates, oxygen and nitrogen (Parish and Stoker 1998; Boshoff and Barry 2005). Further evidence for the role of persisters in TB disease relapse following drug therapy comes from the Cornell mouse model already discussed above.

It is estimated that one-third of the world’s population is latently infected with *M. tb* and that 10% are expected to develop active TB in the future representing a large reservoir for potential reactivation and subsequent spread (Corbett, Watt et al. 2003). The prevailing hypothesis is that persisters are responsible for latent infection, the long course of drug therapy and disease reactivation (Mitchison 2000; Boshoff and Barry 2005; Garton, Waddell et al. 2008; Cardona 2009; Rustad, Sherrid et al. 2009; Chao and Rubin 2010).

1.3.3 Persisters and their molecular mechanisms of formation

Exponentially growing cultures of virtually all pathogens examined so far give rise to a small subpopulation (≈10^{-5}) of dormant persister cells (Lewis 2010). Persisters exhibit multidrug
tolerance and their presence during latent infection could explain the recalcitrance of tuberculosis to drug therapy. Furthermore, persisters are not drug resistant mutants, rather they are phenotypic variants thought to arise via stochastic changes in gene expression (Balaban, Merrin et al. 2004). These random fluctuations in gene expression result in over-expression of stasis-inducing systems causing persister formation. The best studied of these are the toxin-antitoxin (TA) systems; originally identified as plasmid maintenance factors (Lewis 2010), but since identified in most bacterial genomes. They are encoded in bicistronic operons in which the toxin is co-transcribed and co-translated with its cognate antitoxin, which together form a stable complex effectively neutralizing the toxin. However, the two proteins have differing stabilities, so that in the absence of continued expression, the unstable antitoxin is eventually degraded, leading to growth arrest by the stable toxin (Hayes 2003). Typically the toxin inhibits any one of several essential cellular processes including: DNA replication, mRNA stability, protein synthesis, cell-wall biosynthesis, and ATP synthesis (Yamaguchi, Park et al. 2011). Most often TA system toxicity is mediated through mRNA cleavage, resulting in translation inhibition (Gerdes, Christensen et al. 2005).

Pioneering work done in *E. coli* isolated persisters for transcriptome analysis from a population expressing a degradable GFP under the control of a ribosomal promoter only expressed in actively dividing cells (Shah, Zhang et al. 2006). The dim cells were sorted and shown to be resistant to the fluoroquinolone ofloxacin, confirming their status as persisters. The transcriptome of the dim cells pointed to down-regulation of biosynthesis genes and increased expression of several TA systems. Ectopic over-expression of the *E. coli* mRNA interferase toxins RelE and MazF toxins was sufficient to produce multidrug-tolerant persisters (Keren, Shah et al. 2004; Vazquez-Laslop, Lee et al. 2006). Interestingly, deletion of individual toxin genes does not yield a discernible phenotype affecting persister formation, which is likely attributable to the functional redundancy of these elements given there are at least 15 TA modules in *E. coli* (Pandey and Gerdes 2005; Alix and Blanc-Potard 2009). While persisters do form stochastically, it is recognized that there is also a deterministic component contributing to their formation as evidenced by their higher abundance as population density increases, reaching 1% in stationary-phase cultures (Keren, Kaldalu et al. 2004). Further supporting a deterministic component in persister formation is the role of the SOS response which is induced under conditions of DNA damage (Dorr, Lewis et al. 2009; Dorr, Vulic et al. 2010). Although the
SOS response causes the activation of several TA genes in *E. coli*, only a knockout for the TisB toxin had a sharply decreased level of persisters tolerant to ciprofloxacin, an antibiotic that causes DNA double stranded breaks. Moreover, cells ectopically producing TisB toxin were multidrug tolerant and had elevated levels of persisters. TisB forms an ion channel in the membrane, leading to a drop in the proton motive force and ATP levels, consistent with its role in forming dormant cells. Taken together, persisters form in a TisB-dependent manner in response to DNA damage.

TA systems are massively expanded in *M. tb* with 88 putative TA systems found in its genome (Pandey and Gerdes 2005; Ramage, Connolly et al. 2009). Surprisingly only two of these are present in *M. marinum*, which is considered to be the closest genetic relative to the MTBC. This observation strongly suggests that TA systems expanded following the divergence of the MTBC and *M. marinum* from their last common ancestor, and that these systems fulfill a prominent role in the unique biology and pathogenesis of *M. tb*. Several groups recently began to systematically delineate the roles of specific TA systems in *M. tb*. Many of the putative *M. tb* TA modules induced bacteriostasis in *M. smegmatis* or *E. coli*, confirming their roles as bona fide TA systems (Korch, Contreras et al. 2009; Ramage, Connolly et al. 2009; Singh, Barry et al. 2010). Consistent with findings from *E. coli*, microarray analysis of *M. tb* persisters revealed broad downregulation of energy and metabolism genes concomitant with higher expression levels of several TA systems (Keren, Minami et al. 2011). As with other pathogens studied, treatment of exponentially growing *M. tb* cultures with unrelated antibiotics produced a characteristic time-dependent biphasic kill curve clearly indicating the presence of drug-tolerant persisters (Keren, Minami et al. 2011). This behavior is similar to published reports for *M. tb* killing in vivo in the Cornell model, guinea pig model, and human sputum (McCune and Tompsett 1956; Jindani, Dore et al. 2003; Ahmad, Klinkenberg et al. 2009).

### 1.3.4 Oxygen and latency

Although the exact stimuli controlling latency remain unknown, changes in oxygen tension have been intimately associated with varying TB disease (Wayne 1994; Wayne and Sohaskey 2001; Gomez and McKinney 2004; Boshoff and Barry 2005; Russell, VanderVen et al. 2010). This paradigm stems from the observations that: active TB infections are preferentially associated with the most oxygenated sites within the body; within the lungs of patients failing
TB chemotherapy, high bacterial numbers are only found in lesions directly connected with open airways whereas lesions lacking direct contact with air are paucibacillary (Kaplan, Post et al. 2003); and reactivation of latent TB in humans occurs most frequently in the upper lobes of the lung, the most oxygen-rich region of the body (Adler, Finney et al. 1996). As well, inflammatory or necrotic regions are known to be in a state of depleted oxygen and human granulomas lack endothelial and blood vessel markers which taken together strongly suggest that M. tb experiences extended hypoxic periods in vivo(Wayne and Sohaskey 2001; Russell, VanderVen et al. 2010; Flynn, Chan et al. 2011). Moreover, in vitro studies revealed tubercle bacilli require oxygen for growth and that abrupt oxygen deprivation is lethal(Wayne and Diaz 1967) unless they are given time to adapt to its gradual depletion (Wayne 1976).

How M. tb adapts to hypoxia has been studied extensively in vitro using the Wayne model, which employs sealed, standing cultures incubated over an extended period of time while the bacteria deplete the limited oxygen available (Wayne 1994). M. tb then enters into NRP, demarcated by reduced metabolism and drug tolerance, reminiscent of the persister subpopulation present during latency (Wayne 1977; Wayne and Lin 1982; Wayne and Hayes 1996). In particular, early studies of hypoxic M. tb populations revealed a fourfold increase in isocitrate lyase activity, which initiates the glyoxylate shunt (discussed below) in the tricarboxylic acid cycle (TCA) and catalyzes the cleavage of isocitrate to succinate and glyoxylate (Wayne and Lin 1982). The hypoxic bacilli also exhibited a 10-fold increase in synthesis of a glycine dehydrogenase that catalyzes the reductive amination of glyoxylate while at the same time oxidizing NADH to NAD^+(Wayne and Lin 1982). Thus, diversion of some of the glyoxylate produced by isocitrate lyase into the reductive amination pathway would serve to regenerate the NAD pool, allowing for anaerobic ATP generation to perhaps allow for the continuation of viability-sustaining cellular processes during dormancy.

Gene expression studies in M.tb have identified two distinct transcriptional responses to hypoxia. There is the initial hypoxic response (~2 hours post hypoxia) consisting of 49 genes with coordinated induction under the control of two sensor kinases (DosS and DosT) and the response regulator (DosR) (Sherman, Voskuil et al. 2001; Boon and Dick 2002; Park, Guinn et al. 2003). Subsequently these genes became known as the DosR regulon which is induced by other stress conditions that impede aerobic respiration including exposure to nitric oxide
(Voskuil, Schnappinger et al. 2003), carbon monoxide (Shiloh, Manzanillo et al. 2008) and low pH (Rohde, Yates et al. 2007). The DosR regulon is also upregulated in \(M.\ tb\) isolated from macrophages (Schnappinger, Ehrt et al. 2003), as well as mice (Shi, Jung et al. 2003; Karakousis, Yoshimatsu et al. 2004). Initially it was presumed that the DosR regulon is essential for persistence; however, induction of the DosR regulon during hypoxia is transient, with most genes returning to baseline expression levels by 24 hours. Furthermore the \(\Delta dosR\) mutant of \(M.\ tb\) only showed a minor survival defect after 3 weeks incubation under hypoxia (Rustad, Harrell et al. 2008). There are conflicting reports as to the role of DosR in \(M.\ tb\) virulence and persistence in animals. They range from hypervirulence of the \(\Delta dosR\) mutant in mice (Parish, Smith et al. 2003); to no discernible phenotype (Rustad, Harrell et al. 2008; Bartek, Rutherford et al. 2009); to modest attenuation in mice, guinea pigs and rabbits (Malhotra, Sharma et al. 2004; Yam, D'Angelo et al. 2009). Interestingly, the DosR regulon is constitutively overexpressed in the \(W\)-Beijing lineage of \(M.\ tb\) associated with epidemic spread and increased drug resistance worldwide (Reed, Gagneux et al. 2007; Domenech, Kolly et al. 2010). Altogether, these results suggest that the DosR regulon may play some role in \(M.\ tb\) infection but it is unlikely to play a central role in establishing or maintaining NRP.

A second set of 230 genes induced by longer hypoxia exposure (>4 days) was identified. This gene set, termed the enduring hypoxic response (EHR), is DosR-independent (Rustad, Harrell et al. 2008). Thirty of the EHR genes are transcriptional regulators, suggesting a highly complex and possibly redundant web of regulation. The extended induction of the EHR genes suggests that they may have important roles in NRP survival. The genetic factors controlling the expression of the EHR are unknown and whether these genes play a role in \(M.\ tb\) persistence during latency remains to be determined.

### 1.3.5 Carbon metabolism

There is mounting evidence demonstrating that in the absence of carbohydrates, \(M.\ tb\) uses host derived lipids as an energy and carbon source during persistent infection (Bloch and Segal 1956; Russell, VanderVen et al. 2010). This is not surprising given that a central role for fatty acid metabolism during pathogenesis is suggested by the extensive duplication of genes involved in lipid metabolism in the \(M.\ tb\) genome (Cole, Brosch et al. 1998; Cole, Eiglmeier et al. 2001; Fleischmann, Alland et al. 2002; Van der Geize, Yam et al. 2007). In addition, genes
involved in fatty acid metabolism have been shown to be upregulated during infection of macrophages (Graham and Clark-Curtiss 1999; Dubnau and Smith 2003; Schnappinger, Ehrt et al. 2003) and mice (Timm, Post et al. 2003; Dubnau, Chan et al. 2005). Bacilli isolated from patient tissues were found closely associated with lipids (Kondo et al., 1970) and preferentially respire fatty acids rather than carbohydrates (Bloch and Segal 1956). Furthermore, it appears that host genes involved in lipid biosynthesis are induced in tissues recovered from human granulomas, which also contain high levels of triacylglycerols (TAGs), cholesterol and cholesterol esters (Kim, Wainwright et al. 2010).

The glyoxylate shunt allows \textit{M. tb} to by-pass the CO$_2$ generating steps of the TCA in order to retain carbon when growing on fatty acids as its limiting carbon source(Figure 1.2A). There are two isocitrate lyase genes ($icl1$ and $icl2$) present in the \textit{M. tb} genome, which catalyze the entry step in the glyoxylate shunt (McKinney, Honer zu Bentrup et al. 2000). The $\Delta icl1$ mutant, but not $\Delta icl2$, is impaired for persistence in chronically infected mice (McKinney, Honer zu Bentrup et al. 2000). The double mutant is incapable of growth on fatty acids or in macrophages, and is rapidly eliminated from the lungs of infected mice (Munoz-Elias and McKinney 2005).

Recent studies have highlighted that host cholesterol seems to be a major carbon source for \textit{M. tb} during persistent infection in animals. \textit{M. tb} encodes cholesterol uptake and degradation pathways (Van der Geize, Yam et al. 2007). Mutants lacking the ability to acquire or degrade cholesterol exhibit a survival defect in the persistent, but not the acute phase of animal infection (Pandey and Sassetti 2008; Chang, Miner et al. 2009; Yam, D'Angelo et al. 2009; Hu, van der Geize et al. 2010; Nesbitt, Yang et al. 2010; Griffin, Gawronski et al. 2011; Yang, Gao et al. 2011). The metabolism of cholesterol, methyl-branched fatty acids, odd-chain lipids along with catabolism of branched-chain amino acids generates the C3 product propionyl-CoA, accumulation of which is toxic to \textit{M. tb}(Savvi, Warner et al. 2008; Russell, VanderVen et al. 2010). Propionyl-CoA can be metabolized via two pathways: the methylcitrate cycle which condenses propionyl-CoA with oxaloacetate to generate pyruvate and succinate (Munoz-Elias, Upton et al. 2006; Upton and McKinney 2007), or the methylmalonyl pathway which generates methylmalonyl-CoA (Savvi, Warner et al. 2008)(Figure 1.2B). In fact, it has been shown that \textit{M. tb} uses methylmalonyl-CoA as a precursor for the biosynthesis of many methyl-branched lipids.
including the virulence-associated lipids SL-1 and PDIM; thus, lipid anabolism is thought to mitigate the toxic effects of propionate accumulation (Jain, Petzold et al. 2007). Moreover, given the strong reducing power of host fatty acids, their catabolism by \textit{M. tb} along with the inhibition of respiration by NO, CO and hypoxia during infection is predicted to result in the accumulation of reducing equivalents leading to reductive stress. A study by Singh et al. (Singh, Crossman et al. 2009) demonstrated that WhiB3 functions as an intracellular redox sensor and controls the flux of lipid precursors and reducing equivalents through the biosynthesis of various virulence polyketides (PAT, SL-1, PDIM and storage lipids (TAG)) to achieve redox balance and to modulate the host immune response. In summation, the metabolism of host lipids by \textit{M. tb} for energy and carbon, as well as the shuttling of toxic intermediates is crucial to its pathogenesis.
Figure 1.2. Overview of *M. tb* carbon metabolism. (A) The tricarboxylic acid cycle with the glyoxylate shunt shown in red. (B) Metabolism of cholesterol and methyl-branched or odd-chain fatty acids.
1.4 The H-NS protein

1.4.1 H-NS is a bacterial nucleoid-associated proteins

The nucleoid–associated proteins (NAPs) are low molecular weight proteins that bind nucleic acids in prokaryotes (Johnson, Johnson et al. 2006). They play integral roles in organizing and condensing the bacterial chromosome as well as in DNA replication, recombination and gene regulation (Johnson, Johnson et al. 2006). *E. coli* has been purported to possess up to 12 NAPs (Azam and Ishihama 1999), although this number is likely greater if taking into account proteins encoded by horizontally transferred genetic elements and other chromosomally-encoded polypeptides (Dorman 2009). One of the most intensively studied NAPs is the heat-stable nucleoid structuring protein, H-NS (Dorman 2004).

H-NS was first discovered as a small heat-stable protein capable of controlling *E. coli* RNA polymerase (RNAP)-directed transcription *in vitro* in a concentration-dependent manner (Lammi, Paci et al. 1984; Dorman 2009). H-NS was subsequently isolated during biochemical screens for bacterial ‘histone-like’ proteins using conditions that were successfully employed in isolating eukaryotic histones (Falconi, Gualtieri et al. 1988), although it should be noted that H-NS is not functionally equivalent to histones. It is a small (M.W.=15 kDa, 137 aa), neutral protein, which is highly abundant in *E. coli* at approximately 20,000 copies per cell (Lammi, Paci et al. 1984; Falconi, Gualtieri et al. 1988).

1.4.2 H-NS: a pleiotropic transcriptional regulator

Multiple studies have demonstrated that H-NS is a global regulator that controls the expression of a large number of genes (>400) in *E. coli* (Oshima, Ishikawa et al. 2006) and *Salmonella* (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006). Chromatin immunoprecipitation-on-chip (ChIP-chip) of H-NS in *S. typhimurium* indicated that H-NS binds to the most AT-rich regions of the genome (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006); including all the major virulence genes which are thought to have been acquired via horizontal gene transfer (HGT) (Groisman and Ochman 1997; Ochman, Lawrence et al. 2000). Consistent with this finding, analysis of horizontally acquired sequences from different bacteria
revealed they are AT-rich when compared to their resident genomes (Daubin, Moran et al. 2003). Complementary transcriptome analysis of a S. typhimurium Δhns mutant strain revealed that H-NS functions almost exclusively to repress its target genes (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006). In cases where it had appeared H-NS was a positive regulator for specific genes, it was later shown that the positive expression effect of H-NS was indirect (e.g., H-NS is a repressor of a repressor) (Bertin, Terao et al. 1994; Nasser and Reverchon 2002).

HGT serves a crucial role in bacterial evolution, enabling the acquisition of genetic material that may confer a selective advantage over competitors. Paradoxically, newly acquired sequences pose a challenge from a regulatory standpoint and could even be detrimental to the fitness of the recipient cell (Sorek, Zhu et al. 2007; Dorman 2009). Therefore, it was proposed that H-NS may serve to mitigate the unchecked expression of newly acquired genetic elements while at the same time facilitating their integration into existing regulatory frameworks (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006). Not only is H-NS a transcriptional repressor, but also it is now recognized as a ‘xenogeneic’ silencer of foreign DNA. The loss of H-NS is highly detrimental for cell growth, varying from growth defects in E. coli and Salmonella, to cell death in Yersinia. Improved fitness of Salmonella Δhns mutants can be obtained via compensatory deletions of certain xenogeneic genomic islands or their positive regulators (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006).

H-NS homologues have been characterized in the γ-proteobacteria along with the less well-conserved H-NS-like homologues of the α- and β-proteobacteria (Tendeng and Bertin 2003). H-NS-like molecules are diverse in their primary sequence and can often be missed based upon sequence homology, but instead can be identified by in vivo complementation (Goyard and Bertin 1997; Tendeng and Bertin 2003; Tendeng, Krin et al. 2003; Tendeng, Soutourina et al. 2003). Given the important role of H-NS in regards to silencing and integrating horizontally acquired genes, it remains to be determined how other bacteria deal with the potential challenges associated with the introduction of foreign sequences. Prior to the investigations in this thesis, H-NS homologues or functional analogues had not been identified outside of the gram-negative bacteria, including the mycobacteria.
1.4.3 Biochemical and structural properties of H-NS

H-NS consists of three distinct domains: an N-terminal dimerization domain (residues 1–64); a central secondary dimerization domain (residues 64–80), and a C-terminal DNA binding domain (residues 80–137) (Shindo, Ohnuki et al. 1999; Esposito, Petrovic et al. 2002; Arold, Leonard et al. 2010). In solution, H-NS exists as a homodimer and, even in the absence of DNA, can spontaneously oligomerize in a concentration dependent fashion (Ueguchi, Suzuki et al. 1996). DNaseI footprinting assays of several prototypical H-NS-regulated promoters showed that H-NS binding protects large stretches of AT-rich DNA, suggesting that H-NS forms higher order nucleo-protein complexes along target sequences (Lucht, Dersch et al. 1994; Tupper, Owen-Hughes et al. 1994). High affinity binding sites (Kd =15 nM) have been reported with the current paradigm being that they serve as initiation sites for nucleation of H-NS into larger oligomerized structures coating proximal DNA of lower binding affinity (Dame, Luijsterburg et al. 2005; Lang, Blot et al. 2007; Wu, Yang et al. 2008). Oligomerization activity by H-NS is requisite for silencing as demonstrated by H-NS dimerization mutants being defective for silencing even though they retain their DNA binding activity (Ueguchi, Suzuki et al. 1996). Additionally, H-NS anti-silencers have been identified that specifically abrogate its oligomerization (Williamson and Free 2005; Ali, Beckett et al. 2011). Atomic force microscopy studies of H-NS-DNA interactions revealed it bridges DNA(Dame, Luijsterburg et al. 2005), contributing to the compaction of the nucleoid and repressing transcription via the trapping or occlusion of RNA polymerase (Dame, Wyman et al. 2002; Navarre, McClelland et al. 2007). However, whether DNA-bridging truly occurs in vivo has been called into question following studies of H-NS-DNA interactions using magnetic tweezers. In these experiments the addition of H-NS to λ-DNA induced the extension of the end-to-end DNA length (Kana, Gordhan et al. 2008). This manner of DNA-binding by H-NS resulted in a more rigid complex versus naked DNA and has been referred to as the stiffening mode of DNA binding. In a subsequent study, Liu et al. (2009) showed that H-NS employs the stiffening mode of DNA binding at physiologically relevant Mg2+ concentrations but switches to the bridging mode of DNA binding at concentrations of 5 mM or higher (Biketov, Potapov et al. 2007). Further supporting the stiffening model is the finding that the transcription factor SsrB, which relieves gene silencing by
H-NS of the SPI-2 locus, could displace H-NS from DNA when bound in the stiffening mode, but not when H-NS–DNA bridges were present (Cohen-Gonsaud, Barthe et al. 2005).

Arolid et al. (2010) recently solved the crystal structure of the N-terminal S. typhimurium H-NS dimerization domain (residues 1–83) in an oligomerized state (Arolid, Leonard et al. 2010). This structure revealed the presence of the previously unrecognized central dimerization site. In this structure the H-NS monomers were arranged in an antiparallel orientation with dimerization occurring at both interfaces enabling H-NS filaments to propagate via head-to-head/tail-to-tail interactions (Figure 1.3A). The resulting higher-order structure forms a superhelical scaffold that orients the DNA binding domains of adjacent monomers on opposing faces along the superhelix (Arolid, Leonard et al. 2010). In theory, this arrangement could accommodate the binding of two DNA helices on each H-NS scaffold (Figure 1.3B), which would provide a structural rationale for how DNA bridging occurs.
Figure 1.3. H-NS oligomerization requires two dimerization interfaces. (A) H-NS monomers arranged in their multimerized state via a ‘head-to-head/tail-to-tail’ mechanism. Helices H1, H2 and H3 form the N-terminal dimerization interface. The central dimerization domain is formed by a helix–turn–helix motif formed between the C-terminal end of H3 and H4, which is shown between the yellow and blue H-NS monomers (taken from (Ali, Xia et al. 2012)). (B) A three-dimensional representation of the H-NS superhelix accommodating two DNA double helices (gray). The putative position of the C-terminal DNA binding domain is depicted in yellow while the extended helical H-NS$_{1-83}$ oligomer is shown in orange. R15, R19, and K32 form a positively charged surface capable of compensating for the repulsive effect of bridging two negatively charged DNA duplexes into close proximity and are highlighted in blue. Image modified from (Arold, Leonard et al. 2010).
1.5 Lsr2 of Mycobacteria

Lsr2 (Leprosy serum reactive clone 2) was initially identified as a potent T-cell antigen in a screen of an *M. leprae* cosmid library for positive clones that react with leprosy patient serum (Laal, Sharma et al. 1991). It is a small (M.W. = 12.5 kDa; 112 aa), basic protein that is highly conserved in mycobacteria and related actinomycetes, and bears no significant sequence homology to any known proteins. Lsr2 remained largely uncharacterized until 2006 when a *lsr2::Tn* mutant of *M. smegmatis* was isolated by our lab in a screen for altered colony morphology, which was designed to identify genes involved in mycobacterial cell wall lipid biosynthesis (Chen, German et al. 2006). Interestingly, this study determined that Lsr2 is not a cell wall component, but localizes to the cytosol and is capable of forming dimers *in vivo* (Chen, German et al. 2006). Furthermore, Lsr2 regulates the biosynthesis of cell wall lipids in *M. smegmatis* (Chen, German et al. 2006; Kocincova, Singh et al. 2008) as well as the multi-drug efflux pump, *iniBAC*, of *M. tb* (Colangeli, Helb et al. 2007). Microarray analysis of an *M. smegmatisΔlsr2* mutant identified Lsr2 as a putative negative regulator of multiple cellular processes and that most targeted genes have distinctly AT-rich 5’ untranslated sequences (Colangeli, Helb et al. 2007). Consistent with Lsr2’s role as a transcriptional regulator, electrophoretic mobility shift assays (EMSAs) demonstrated that Lsr2 binds DNA nonspecifically but with a preference for AT-rich sequences (Chen, Ren et al. 2008). Atomic force microscopic analysis revealed that Lsr2 is capable of bridging DNA and forming loop domains, providing a possible mechanism for transcriptional regulation as seen with H-NS. The abilities to oligomerize and bind DNA are both essential for Lsr2 to function, suggesting a similar domain organization as H-NS (Chen, Ren et al. 2008). As such, we postulated that Lsr2 might be an H-NS-like protein in mycobacteria (Chen, Ren et al. 2008).

1.6 Thesis Rationale and Outline

Tuberculosis is an ancient disease that still has a devastating impact on human health. Much of *M. tb*’s success as a pathogen is attributed to its ability to persist in a quiescent state for extended periods of time within host. With the completion of the *M. tb* genome sequence just
over a decade ago, along with advances in mycobacterial molecular biology, our understanding of this bacillus has greatly increased. Nevertheless the majority of virulence mechanisms involved in M. tb pathogenesis, and in particular mechanisms dictating M. tb non-replicating persistence during latent disease remain obscure.

Lsr2 is a novel DNA-binding protein ubiquitous in mycobacteria and related actinomycetes. M. smegmatislsr2 mutants were identified indicating that although strongly conserved, lsr2 is not essential for viability. Intriguingly, readily identifiable Lsr2 homologues are even present in many mycobacteriophage genomes (Chen, German et al. 2006; Hatfull, Jacobs-Sera et al. 2010) and begs the question what is the function(s) of Lsr2? Early biochemical characterization of Lsr2 revealed similar DNA-binding properties as H-NS (Chen, Ren et al. 2008). This led to the suggestion that Lsr2 could be an H-NS homologue despite the lack of sequence homology. My PhD studies address this hypothesis as well as the role of Lsr2 in mycobacterial biology and M. tb pathogenesis.

In Chapter 2, I present genetic evidence including the ability of lsr2 to functionally complement independent hns mutant phenotypes in vivo to support my hypothesis that Lsr2 is a functional analog of H-NS. This finding is significant since it is the first example of an H-NS-like protein identified outside of the gram-negative bacteria. Given H-NS’s role as an important regulator of virulence genes in bacterial pathogens such as Salmonella, Yersinia pestis and Vibrio cholerae, it was reasonable to surmise a similar role for Lsr2 in M. tb. In Chapter 3 I successfully generated and characterized an M. tb∆lsr2 mutant despite claims by others it was essential in this species. I used ChIP-chip to identify regions of the genome bound by Lsr2 as well as microarray analysis of the mutant to determine the genes controlled by Lsr2 in M. tb. Similarly to H-NS, I found that Lsr2 binds to the most AT-rich regions of the genome, including many horizontally acquired genes. Importantly, I found that Lsr2 regulates important virulence factors like ESX-1 and the persistence inducing TA systems. In fact many of the transcriptional changes I observed in the mutant coincide with genes induced in in vitro models of dormancy. This seems to indicate a prominent role for Lsr2 in controlling M. tb persistence. In Chapter 4 I provide some biological data for the function of Lsr2. I show that the M. smegmatis Δlsr2 mutant has enhanced viability in an in vitro hypoxia model. Furthermore, I describe a transposon-mediated lsr2-inactivation phenomenon in M. smegmatis, which is the first example
of such a mechanism in mycobacteria. The *M. tb* Δlsr2 mutant displayed retarded growth in immunocompetent mice but could still persist. While comparing the intracellular growth of the Δlsr2 mutant to WT, I observed that the mutant induces rapid death of infected macrophages. Lastly, in Chapter 5, I present our collaborative work which elucidates the biochemical mechanism by which H-NS and Lsr2 recognize their target DNA, and provides a unified molecular mechanism for this class of DNA binding proteins. Collectively, my work has identified Lsr2 an important pleiotropic regulator strongly implicated in mediating persister formation during TB infection.
CHAPTER 2: LSR2 OF MYCOBACTERIUM REPRESENTS A NOVEL CLASS OF H-NS-LIKE PROTEINS

A version of this chapter has been published (Gordon, B. R., R. Imperial, L. Wang, W.W. Navarre and J. Liu. 2008. *Journal of Bacteriology*. 190(21): 7052-9). I performed all experiments described in this chapter. Robin Imperial generated the *E. coli* K-12 strain N99 Δ*hns* mutant.
2.1 Abstract

Lsr2 is a small, basic protein present in *Mycobacterium* and related actinomycetes. Previous *in vitro* biochemical studies showed that Lsr2 is a DNA-bridging protein, a property shared by H-NS-like proteins in gram-negative bacteria. Here I present *in vivo* evidence based on genetic complementation experiments that Lsr2 is a functional analog of H-NS, the first such protein identified in gram-positive bacteria. I show that *lsr2* can complement the phenotypes related to *hns* mutations in *Escherichia coli*, including β-glucoside utilization, mucoidy, motility, and hemolytic activity. I also show that Lsr2 binds specifically to H-NS-regulated genes and that the repression of *hlyE* by Lsr2 can be partially eliminated by overexpression of *slyA*, suggesting that the molecular mechanisms of Lsr2 repression and depression are similar to those of H-NS. The functional equivalence of these two proteins is further supported by the ability of *hns* to complement the *lsr2* phenotype in *Mycobacterium smegmatis*. Taken together, my results demonstrate unequivocally that Lsr2 is an H-NS-like protein.
2.2 Introduction

Multiple studies have demonstrated that H-NS is a global regulator that controls the expression of a large number of genes (>400) in *E. coli* (Oshima, Ishikawa et al. 2006) and *Salmonella* (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006), most of which are derived from foreign sources via HGT. Many H-NS-repressed genes are regulated by environmental conditions such as pH, osmolarity, and temperature or are involved in bacterial virulence (Dorman 2004; Navarre, McClelland et al. 2007). The effects of H-NS on gene expression are largely inhibitory, which is partially explained by the ability of H-NS to bridge adjacent helices of DNA (Dame, Wyman et al. 2000; Dame, Noom et al. 2006).

Numerous phenotypes associated with *hns* mutations have been described, including derepression of β-glucoside metabolism (Defez and De Felice 1981), increased resistance to low Ph and high osmolarity (Hommais, Krin et al. 2001), and a loss of motility (Hinton, Santos et al. 1992; Bertin, Benhabiles et al. 1999). H-NS also regulates virulence genes in other gram-negative pathogens, including *Vibrio cholerae* (Nye, Pfau et al. 2000; Tendeng, Badaut et al. 2000), *Yersinia* (Heroven, Nagel et al. 2004; Ellison and Miller 2006), and *Shigella flexneri* (Porter and Dorman 1994; Falconi, Colonna et al. 1998; Beloin and Dorman 2003; Deighan, Beloin et al. 2003). A number of H-NS homologues have been found in γ-proteobacteria (Tendeng and Bertin 2003). They share a high level of sequence conservation and have been identified on the basis of sequence homology (Tendeng and Bertin 2003). Less well conserved H-NS homologues, termed the H-NS-like proteins, have been identified outside of this group by *in vivo* complementation, such as BpH3 of *Bordetella pertussis* (Goyard and Bertin 1997) and HvrA of *Rhodobacter capsulatus* (Bertin, Benhabiles et al. 1999), which belong to the alpha and beta subdivisions of the proteobacteria, respectively. As such, it was suggested that the H-NS-like proteins are widespread in gram-negative bacteria (Tendeng and Bertin 2003). However, H-NS-like proteins have not been identified in bacteria phylogenetically distant from proteobacteria, such as in gram-positive bacteria, either by *in silico* analysis of genomes or by *in vivo* complementation using a genomic library, as in the case of *Bacillus subtilis* (Tendeng and Bertin 2003).

Lsr2 is a small (~12-kDa), basic protein found in all mycobacterial genomes that have been sequenced so far. Lsr2 homologues are also present in other actinomycetes such as
Streptomyces, Nocardia, and Rhodococcus. Previous biochemical studies indicate that Lsr2 is a DNA-bridging protein (Chen, Ren et al. 2008), which suggests that Lsr2 is an H-NS-like protein. Here I show that Lsr2 is functionally homologous to H-NS using in vivo evidence based on genetic complementation experiments. This study establishes Lsr2 as the first H-NS-like protein to be identified outside the proteobacteria.
2.3 Results

2.3.1 The \textit{lsr2} gene of \textit{M. tuberculosis} complements various phenotypes of the \textit{E. coli \Delta hns} mutant

To address whether Lsr2 could functionally mimic H-NS and complement its deficiency in an \textit{E. coli \Delta hns} mutant, I began by cloning the \textit{lsr2} gene of \textit{M. tb} and \textit{hns} of \textit{S. serovar Typhimurium} into the mycobacterial-\textit{E. coli} shuttle vector pNBV1 (Howard, Gomez et al. 1995). I chose to study the \textit{lsr2} of \textit{M. tb} since it is functionally interchangeable with \textit{lsr2} of \textit{M. smegmatis} (Chen, Ren et al. 2008). The H-NS proteins of enteric bacteria are also functionally interchangeable (Bertin, Hommais et al. 2001). To perform complementation experiments in \textit{E. coli}, I first transformed the pLSR2-HA construct into a \textit{\Delta hns} mutant strain of \textit{E. coli WN582} by standard procedures and examined the expression of Lsr2 in the \textit{E. coli} recombinant strain. Western blot analysis using an anti-HA antibody indicated that the Lsr2 protein was expressed in the \textit{E. coli \Delta hns} strain harboring pLSR2-HA (Figure 2.1).

Mutations of \textit{hns} are highly pleiotropic in \textit{E. coli} and \textit{Salmonella}. Several phenotypes associated with \textit{hns} mutations have been well characterized, including mucoidy (Harrison, Pickard et al. 1994; Soutourina, Krin et al. 2002), \(\beta\)-glucoside utilization (Defez and De Felice 1981), hemolytic activity (Gomez-Gomez, Blazquez et al. 1996), and loss of motility (Hinton, Santos et al. 1992; Bertin, Terao et al. 1994). The mechanisms of \textit{hns} effects on the corresponding genes involved in these phenotypes are well documented. As such, these phenotypes have been routinely employed for identification of new \textit{hns}-like genes by in \textit{vivo} complementation experiments (Goyard and Bertin 1997) or for structure-function relationship studies of H-NS proteins (Bertin, Benhabiles et al. 1999). To determine whether Lsr2 is an H-NS-like protein, I first examined whether \textit{lsr2} could complement these various phenotypes normally associated with \textit{hns} mutations.
Figure 2.1. Western blot analysis of Lsr2 expression in *E. coli*. Lane 1, *hns* mutant WN582; lane 2, *hns* mutant WN582 with pNBV1; lanes 3 to 7, five randomly picked colonies of WN582 with pLSR2-HA; lane M, molecular weight marker.
The *hns* gene is involved in the regulation of colanic acid biosynthesis (Harrison, Pickard et al. 1994; Soutourina, Krin et al. 2002). When *E. coli* or *Salmonella* overproduces colanic acid, colonies growing on agar plates are mucoid. Consistent with previous findings, the *E. coli* \(\Delta hns\) mutant strain WN582 used in this study took on a mucoid appearance on LB agar plates (Figure 2.2A). Interestingly, colonies of the recombinant WN582 strain harboring pLSR2-HA (Figure 2.2A) or pHNS-HA (not shown) became nonmucoid and resembled the WT, parental strain MC4100. As a control, WN582 transformed with the cloning vector pNBV1 remained mucoid (Figure 2.2A).

I next examined whether *lsr2* could replace *hns* in the regulation of \(\beta\)-glucoside metabolism. In WT *E. coli* cells, the *bgl* operon encoding the gene products necessary for the uptake and fermentation of aryl-\(\beta\),D-glucosides is cryptic, primarily due to its repression by H-NS (Dole, Nagarajavel et al. 2004). As such, the ability to metabolize aryl-\(\beta\),D-glucosides such as salicin has been employed as a positive selection assay for \(\Delta hns\) mutants (Kharat and Mahadevan 2000). To assay for fermentation of aryl-\(\beta\),D-glucosides, I employed MacConkey agar plates supplemented with 0.4% salicin as an indicator medium. As expected, the WT strain MC4100, unable to ferment salicin, grew as white colonies, and the \(\Delta hns\) mutant strain WN582 grew as red colonies, since the fermentation of salicin produces acidic by-products, causing a drop in pH which gives rise to red/pink colonies (Figure 2.2B). Interestingly, the WN582 strain transformed with pLSR2-HA gave rise to white colonies, whereas the negative control WN582 transformed with the cloning vector Pnvb1 remained red (Figure 2.2B). As expected, transformation of WN582 with pHNS-HA yielded white colonies.

The third *hns*-related phenotype I assayed was hemolytic activity. H-NS represses the expression of *hlyE* (Westermark, Oscarsson et al. 2000; Wyborn, Stapleton et al. 2004), which encodes a pore-forming toxin in *E. coli*. Consequently, WT cells of *E. coli* are nonhemolytic, whereas a \(\Delta hns\) mutant strain has a hemolytic phenotype (Gomez-Gomez, Blazquez et al. 1996). Hemolytic activity was assessed on blood agar plates. As expected, the WT strain MC4100 exhibited no hemolytic activity, whereas the \(\Delta hns\) mutant strain WN582 had a translucent halo of clearing (Figure 2.2C).
Figure 2.2. Lsr2 complements various phenotypes in an E. coli Δhns mutant. (A) Complementation of the mucoidy phenotype of the hns mutant by lsr2. Bacterial strains were plated on LB agar. MC4100, WT strain; WN582, Δhns mutant; WN582/pNBV1, WN582 transformed with pNBV1; WN582/pLSR2-HA, WN582 transformed with pLSR2-HA. (B) Assay for fermentation of salicin. Bacterial strains were plated on MacConkey agar plates supplemented with 0.4% salicin. (C) Assay for hemolytic activity. Hemolytic activity was assessed on blood agar plates (see the text). (D) Assay for motility. Motility was assayed on tryptone swarm plates containing 0.3% Bacto agar. WT, WT strain N99; hns, the Δhns mutant of N99; hns/pHNS-HA, the Δhns mutant of N99 transformed with pHNS-HA; hns/pLSR2-HA, the Δhns mutant of N99 transformed with pLSR2-HA; hns/pNBV1, the Δhns mutant of N99 transformed with pNBV1.
Transformation of WN582 with PLSR2-HA or PHNS-HA restored the WT, i.e., nonhemolytic, phenotype, whereas WN582 harboring the cloning vector had obvious hemolytic activity (Figure 2.2C).

Lastly, I examined whether Lsr2 could complement the motility phenotype associated with hns mutations. Mutations in hns result in the loss of motility in Salmonella and E. coli due to the lack of flagella biogenesis (Hinton, Santos et al. 1992; Bertin, Terao et al. 1994). This is explained by the finding that H-NS is a repressor of hdfR, which negatively regulates the flagellar master-regulator operon, flhDC (Ko and Park 2000). The motility phenotype was assessed using tryptone swarm plates. Unexpectedly, the E. coli WT strain MC4100 in our collection was nonmotile, which is likely due to a laboratory-acquired mutation affecting this phenotype. To overcome this problem, I used a different E. coli K-12 WT strain, N99, for this experiment. This strain was indeed motile (Figure 2.2D), and an Δhns mutant strain was created by P1rev6 phage transduction (Thomason, Costantino et al. 2007). As expected, the Δhns mutant of N99 was nonmotile (Figure 2.2D), and transformation of this strain with pLSR2-HA or pHNS-HA restored the motility (Figure 2.2D), although not to the WT level. This partial complementation of motility phenotype by hns on a plasmid had been observed previously by different groups and is thought to be caused by the dosage effect of the hns gene (Bertin, Terao et al. 1994; Goyard and Bertin 1997; Bertin, Hommais et al. 2001). Importantly, the Δhns mutant strain transformed with the cloning vector pNBV1 was nonmotile (Figure 2.2D), confirming that the lsr2 gene by itself was responsible for the reversion of the swarming behavior of the Δhns mutant. Collectively, my results demonstrate that Lsr2 is capable of complementing the various phenotypes of E. coli related to hns mutations, which provides in vivo evidence that Lsr2 is an H-NS-like protein.
2.3.2 The *hns* gene complements the phenotype of an *M. smegmatis* ∆*lsr2* mutant

To further test the functional equivalence of Lsr2 and H-NS, I examined whether *hns* could complement the phenotype of a ∆*lsr2* mutant of *M. smegmatis*. Previously it was shown that the ∆*lsr2* mutant of *M. smegmatis* exhibited a dramatic change of colony morphology: the mutant colonies are smooth, wet, and round, in contrast to the dry, rough, and rugose morphology of the WT mc²155 strain (Chen, German et al. 2006). Although the underlying molecular mechanism remains incompletely understood, transformation of the WT *lsr2* gene from *M. smegmatis* or *M. tb* fully restored the morphological phenotype (Chen, German et al. 2006; Kocincova, Singh et al. 2008). Taking advantage of this simple phenotype, I transformed the *M. smegmatis*Δ*lsr2* mutant with the pHNS-HA plasmid by electroporation and examined the morphology of the resulting recombinant strain. Indeed, the Δ*lsr2*mutant strain harbouring the *hns* plasmid reverted to the WT colony morphology (Figure 2.3C), whereas the Δ*lsr2* mutant harboring the empty pNBV1 vector still exhibited the smooth and round colony morphology (Figure 2.3B). Transformation of the *lsr2*mutant strain with pLSR2-HA restored the colony morphology(not shown). This finding, together with the above results, clearly demonstrate that Lsr2 and H-NS are functionally interchangeable, suggesting that they are truly functional homologues.
Figure 2.3. Complementation of the colony morphology of the ∆lsr2 mutant of *M. smegmatis* by *hns*. Bacterial strains were plated on 7H11 agar and incubated at 37°C for 3 days. (A) WT *M. smegmatis* strain mc2155. (B) The ∆lsr2 mutant of mc^2^155 transformed with pNBV1. (C) The ∆lsr2 mutant of mc2155 transformed with pHNS-HA.
2.3.3 Lsr2 binds specifically to H-NS-regulated genes

Next to determine whether repression by Lsr2 was specific to H-NS-regulated genes I assayed the binding of epitope-tagged Lsr2 to *E. coli* genes by performing ChIP followed by quantitative RT-PCR. Two well-characterized H-NS-regulated genes, *proV* (Owen-Hughes, Pavitt et al. 1992) and *bglG* (Dole, Nagarajavel et al. 2004), and two H-NS-recognized genes (*yjcF* and *xapR*) identified by ChIP-on-chip experiments (Grainger, Hurd et al. 2006) were chosen as the positive H-NS binding sites. For negative H-NS binding sites, *narZ* and *phnE* were selected, both of which were found not to be recognized by H-NS in the ChIP-on-chip experiments (Grainger, Hurd et al. 2006) and have a relatively low AT-content: 45% and 37%, respectively. Cells of the WN582 strain harboring pLSR2-HA were fixed with formaldehyde and subsequently sonicated to lyse the cells and shear the chromosomal DNA to fragments of approximately 500 bp. An anti-HA antibody was then used to precipitate DNA fragments associated with the Lsr2 protein. As a negative control, the WN582 strain harboring PLSR2-FLAG, in which Lsr2 was expressed as a FLAG-tagged protein, was used and the cell lysates made in the same way were precipitated with the anti-HA antibody. Quantitative RT-PCR analyses with gene-specific primers were then performed to determine the enrichment of target sequences.

As expected, I observed that HA-tagged Lsr2 co-precipitated with all four known H-NS binding sites (*proV*, *bglG*, *yjcF*, and *xapR*) with significantly higher levels of enrichment observed than for the two negative binding sites (*narZ* and *phnE*) (Figure 2.4). These data indicate that Lsr2 binds specifically to the HNS-regulated loci *in vivo*. 
Figure 2.4. Lsr2 binds specifically to H-NS-regulated genes. The figure shows results from ChIP and quantitative RT-PCR analyses. Cultures of WN582 (Δhns mutant) harboring pLSR2-HA were subjected to ChIP experiments, and selective genes were analyzed by RT-PCR. Genes regulated by H-NS (bglG, proV, xapR, and yjcF) are highly enriched compared to genes not regulated by H-NS (narZ and phnE).
2.3.4 Repression of hlyE by Lsr2 is partially relieved by overexpression of slyA

To further characterize the functional synergy of Lsr2 and H-NS, I investigated whether mechanisms of derepression of Lsr2 are the same as for H-NS. As mentioned above, H-NS represses the expression of hlyE (Westermark, Oscarsson et al. 2000; Wyborn, Stapleton et al. 2004). Overexpression of slyA results in the derepression of hlyE (Ludwig, Tengel et al. 1995; Wyborn, Stapleton et al. 2004) due to the fact that SlyA binds to a site at the hlyE promoter that overlaps with an H-NS binding site (Wyborn, Stapleton et al. 2004). I examined whether overexpression of slyA could also eliminate the repression of hlyE by Lsr2. For this purpose, the slyA gene was cloned into the arabinose-inducible pLC2002 plasmid and transformed into the WT and the hns mutant WN582 harboring pLSR2-HA, and the hlyE transcript levels in these strains were measured by performing reverse transcriptase quantitative RT-PCR. As expected, the hns mutant strain WN582 exhibited a much higher level of hlyE expression than the WT strain (Figure 2.5, lane 2). Consistent with my phenotypic complementation studies described above, transformation of WN582 with pLSR2-HA restored the hlyE transcript to the WT level (Figure 2.5, lane 3). Interestingly, overexpression of slyA in WN582 harboring pLSR2-HA resulted in a significant increase of the hlyE transcript (Figure 2.5, compare lanes 3 and 4). The same strains grown in the absence of arabinose served as the negative control, and no enhancement of hlyE expression was observed (Figure 2.5, compare lanes 6 and 7). For the positive control, overexpression of slyA in the WT strain resulted in a 3.5-fold increase of the hlyE expression level (Figure 2.5, lane 1). Taken together, these data indicate that overexpression of slyA partially antagonizes the repression of hlyE by Lsr2, presumably by the same mechanism as it antagonizes H-NS.
Figure 2.5. Overexpression of slyA causes partial derepression of hlyE. The figure shows n-fold hlyE transcript levels determined from reverse transcriptase quantitative PCR analysis. The data were normalized to the hlyE transcript level of WT strain MC4100 to obtain n-fold expression of hlyE in various strains. The transcript level of gyrB was used as an internal control for normalization among different samples. Lane 1, WT/pSLYA; lane 2, WN582; lane 3, WN582/pLSR2-HA; lane 4, WN582/pLSR2-HA_pSLYA; lane 5, WT/pSLYA; lane 6, WN582/pLSR2-HA; lane 7, WN582/pLSR2-HA+pSLYA. +Ara, in the presence of arabinose; -Ara, in the absence of arabinose.
2.4 Discussion

In this study, I provide direct in vivo complementation evidence that Lsr2 is a functional homologue of H-NS. The lsr2 gene is able to complement all the Δhns-associated phenotypes that I have examined, including β-glucoside utilization, mucoidy, motility, and hemolytic activity. Since the various phenotypes resulted from the effects of H-NS on the expression of several unrelated genes, these findings suggest that Lsr2 functions as a global regulator in *E. coli*, just like H-NS. I also showed that Lsr2 binds specifically to H-NS-regulated genes and that the Lsr2-mediated repression of *hlyE* by can be eliminated partially by overexpression of *slyA*, suggesting that the molecular mechanisms of Lsr2 repression and depression are similar to those of H-NS. The functional equivalence of these two proteins is further supported by my finding that *hns* can complement the *lsr2* colony morphology phenotype in *M. smegmatis*. Taken together, my results clearly demonstrate that Lsr2 is an H-NS-like protein, a conclusion that is bolstered by previous in vitro biochemical studies of Lsr2 (Chen, Ren et al. 2008). Previously it was demonstrated that Lsr2 binds to DNA in a relatively sequence-independent manner, but exhibits a preference for AT-rich sequences. Analysis by atomic force microscopy revealed that Lsr2 has the ability to bridge distant DNA segments. Furthermore, it has been shown that Lsr2 exists as a homodimer in vivo (Chen, German et al. 2006) and that DNA binding and protein oligomerization are both essential for the normal function of Lsr2 in vivo (Chen, Ren et al. 2008). All of these biochemical characteristics of Lsr2 are shared by H-NS, which prompted me to examine whether *lsr2* and *hns* are functionally equivalent by performing in vivo complementation experiments. My results confirm previous findings and demonstrate that Lsr2 is indeed an H-NS-like protein. It remains to be determined whether Lsr2 plays an equivalent role with respect to other H-NS-related functions in gram-negative bacteria, e.g., maintaining the loop structure of the bacterial chromosome, contributing to nucleoid compaction, and silencing expression of foreign DNA (Navarre, Porwollik et al. 2006; Noom, Navarre et al. 2007). However, given the similar biochemical properties and in vivo functional equivalences of these two proteins, it is surmisable that such functional conservation between Lsr2 and H-NS also exists. My results therefore have strong implications for the role of Lsr2 in mycobacterial physiology and genome maintenance.
Figure 2.6: Sequence alignment and secondary structure prediction of Lsr2 proteins in various actinomycetes. The alignment was produced using the CLUSTALW program. The secondary structure was analyzed by using the Jpred program (http://www.compbio.dundee.ac.uk/~www-jpred/). The arrows represent β-strands and ovals represent α-helices. Species abbreviations: M. tb., Mycobacterium tuberculosis; M. marinum, Mycobacterium marinum; M. ulcerans, Mycobacterium ulcerans; M. avium, Mycobacterium avium; M. paratb, Mycobacterium avium subsp. paratuberculosis; M. leprae, Mycobacterium leprae; M. flavescens, Mycobacterium flavescens; M. gilvum, Mycobacterium gilvum; M. vanbaalenii, Mycobacterium vanbaalenii; M. smegmatis, Mycobacterium smegmatis; M. abscessus, Mycobacterium abscessus; N. farcinica, Nocardia farcinica; R. RH1, Rhodococcus sp. strain RHA1; S. erythraea, Saccharopolyspora erythraea; S. avermitilis, Streptomyces avermitilis; S. coelicolor, Streptomyces coelicolor; S. griseus, Streptomyces griseus subsp. griseus; J. HTCC, Janibacter sp. strain HTCC2649; A. aurescens, Arthrobacter aurescens; R. salmoninarum, Renibacterium salmoninarum; F. alni, Frankia alni; M. luteus, Micrococcus luteus; Cjw1, Mycobacterium phage Cjw1; C. sepedonicus, Clavibacter michiganensis subsp. sepedonicus; K. rhizophila, Kocuria rhizophila; T. fusca, Thermobifida fusca; Omega, Mycobacterium phage Omega. The H-NS of E. coli is included in the alignment (bottom), but only the predicted secondary structure of Lsr2 is shown.
Like H-NS, Lsr2 is likely to be a global regulator influencing multiple genes involved in stress responses and virulence. Supporting this notion are previous studies in which expression of *lsr2* is induced by environmental stress conditions, including nutrient availability, growth temperature, and antibiotic exposures (Wong, Lee et al. 1999; Stewart, Wernisch et al. 2002; Colangeli, Helb et al. 2007). Lsr2 has been experimentally shown to be involved in several unrelated cellular processes, including cell wall lipid biosynthesis (Chen, German et al. 2006; Kocincova, Singh et al. 2008) and antibiotic resistance (Colangeli, Helb et al. 2007). Microarray analysis identified a number of genes putatively regulated by Lsr2 in *M. smegmatis* (Colangeli, Helb et al. 2007).

It is remarkable that Lsr2 is functionally interchangeable with H-NS considering the lack of a significant level of sequence homology and the phylogenetic distribution of these two proteins. H-NS-like proteins have been identified only in gram-negative bacteria thus far, mostly in gamma proteobacteria, and have not been described in bacteria outside of this group. On the other hand, a mouse protein complements the phenotypes of an *E. coli hns* mutant (Timchenko, Bailone et al. 1996), suggesting that H-NS analogs are widespread and could be too divergent to be identified on the basis of sequence similarity (Tendeng and Bertin 2003). Thus far, Lsr2 proteins have only been found in actinomycetes, which are phylogenetically distant from gram-negative bacteria. Among mycobacteria, Lsr2 exhibits a high level of sequence conservation (Figure 2.6). Less-conserved homologues are present in related actinomycetes (Figure 2.6). The exclusive distribution of Lsr2 in actinomycetes mirrors the distribution of H-NS in gram-negative bacteria, which may represent an example of convergent evolution. Supporting this notion, considerable sequence divergences among H-NS-like proteins have been noted, e.g., BpH3 of *B. pertussis* (Goyard and Bertin 1997) and HvrA of *R. capsulatus* (Bertin, Benhabiles et al. 1999), which share low levels of sequence identity (20 to 30%) with the *E. coli* H-NS. Nevertheless, all H-NS-like proteins are predicted to have the same domain organization: an N-terminal domain required for oligomerization followed by a C-terminal DNA binding domain, which are separated by a flexible linker (Bertin, Benhabiles et al. 1999). The nuclear magnetic resonance (NMR) structures of the N-terminal domain of *E. coli* H-NS reveals that it contains three α-helical segments; the third and longest α-helix forms the core of a coiled-coil configuration, whereas the two remaining helices stabilize the structure (Esposito, Petrovic et al. 2002; Bloch, Yang et al. 2003). Although the N-terminal portion is not conserved in amino acid
sequences, all the H-NS-like proteins are predicted to adopt a similar coiled-coil conformation. In contrast, more conserved sequences are found in the C-terminal domain, and the NMR structure of the C-terminal domain of *E. coli* H-NS shows that it is composed of an antiparallel β-sheet, an α-helix, and a 3_10-helix structure forming a hydrophobic core (Shindo, Iwaki et al. 1995). The two-module organization of the functional domains accounts for the ability of different H-NS-like proteins to restore the WT phenotypes of the *E. coli* Δhns mutant, as well as the construction of functional chimeric proteins between different H-NS proteins (Tendeng, Krin et al. 2003). Lsr2 exhibits a low level of sequence homology (<20% identity in amino acid sequence) and has a predicted secondary structure different from that of H-NS (Figure 2.6). However, previous studies suggest that the Lsr2 protein might also be organized into two functional domains: the C-terminal DNA binding domain and the N-terminal oligomerization domain (Chen, Ren et al. 2008). It was previously found that an Arg residue at the C terminus (residue 86 of *M. smegmatis* Lsr2 or 84 of *M. tuberculosis* Lsr2) is important for DNA binding. Replacing this residue with Ala reduces the DNA binding affinity but does not affect the protein oligomerization (Chen, Ren et al. 2008). Two other mutations, R45A and D28A, likely affecting protein oligomerization were identified in the N-terminal domain (Chen, Ren et al. 2008). Furthermore, preliminary data show that a C-terminally truncated Lsr2 protein (retaining the N-terminal residues 1 to 50) exhibits a dominant-negative effect in a WT strain of *M. smegmatis*, suggesting that the N-terminal portion of Lsr2 is involved in protein oligomerization (L. Wang and J. Liu, unpublished data). It appears that although Lsr2 and H-NS do not share a similar sequence or even structural fold, they each contain two functional domains that are equivalent between these two proteins, i.e., for DNA binding and protein oligomerization. Since both functions are necessary for the normal functions of Lsr2 (Chen, Ren et al. 2008) and H-NS (Dorman 2004) *in vivo*, it is possible that they could each achieve the same purpose by preferential binding to AT-rich sequences in the genome and bridging of DNA segments via protein oligomerization. Future studies of the structure of Lsr2 to test this hypothesis are warranted.
2.5 Experimental procedures

**Bacterial strains, media, and growth conditions.** The *E. coli* K-12 strain MC4100 was used as the wild-type (WT) strain for most of our experiments. WN582 is the ∆hns mutant strain of MC4100 (*E. coli* K-12 MC4100 ∆hns::Cm), which was kindly provided by Henry Rosen (University of Washington). Another *E. coli* K-12 strain, N99, and its ∆hns mutant were used in the motility assay. The ∆hns mutant of N99 was created by P1rev6 phage transduction (Thomason, Costantino et al. 2007). P1rev6 phage was grown on a WN582 background to obtain an *hns* gene disrupted by a chloramphenicol cassette. The resulting phage lysate was incubated with N99 for 30 min in the presence of P1 salts (10 Mm CaCl2-5 Mm MgSO4). LB and sodium citrate were then added to prevent secondary infection and incubated for 1 h at 37°C with shaking. The cells were then pelleted and resuspended in LB and plated onto LB agar plates with 20 µg/ml chloramphenicol. Chloramphenicol-resistant colonies were then verified for the ∆hns::cm genotype by PCR (data not shown). *Mycobacterium smegmatis* strain mc2155 and its ∆lsr2 mutant, previously described (Chen, German et al. 2006), were used for complementation experiments. *E. coli* strains were grown in LB at 37°C with appropriate antibiotics, and *M. smegmatis* strains were grown in Middlebrook 7H9 broth or 7H11 agar supplemented with 10%OADC (oleic acid-albumin-dextrose-catalase) and appropriate antibiotics.

**Molecular cloning.** The open reading frame of *lsr2* of *Mycobacterium tuberculosis* including the 200-bp upstream sequence was PCR amplified using the forward primer 5’-CAGTCTAGAAGCCGAATGGGTATCGA-3’ and the reverse primer 5’-TTTTTAAGCTTCTAAGCGTAGTCTCGGACGTCGTATGGGTAGGGTGC-3’, which included a C-terminal hemagglutinin(HA) tag (YPYDVPDYA). The resulting PCR product was digested with *XbaI* and *HindIII* and cloned into PNBV1 pretreated with the same enzymes to yield pLSR2-HA. A second *lsr2* construct with a C-terminal FLAG tag (DYKDDDDK) was also constructed for use later in the chromatin immunoprecipitation (ChIP) experiments described below. The same forward primer as used above was used with the reverse primer 5’-GCTCTAGAAGCTTACTTTGTCATCGTCGTCCCTTTAGTCGTCCGCCGCGTGATGCGTCG-3’, which included the FLAG tag. The resulting PCR product was digested with *XbaI* and *HindIII* and cloned into pNBV1 to yield pLSR2-FLAG.
The \textit{hns}\textsuperscript{-containing} plasmid pWN426 served as the template for PCR amplification of the \textit{hns} gene of \textit{S. enteric} serovar Typhimurium with the forward primer 5’-TTGGATCCGACGACAAACCAGATACGAGAG-3’ and the reverse primer 5’-TTTTAAGCTTTTATGCGTAGTCTGTACGTCATAAGGGTA-3’. The PCR product included the 800-bp upstream sequence of the \textit{hns} open reading frame as well as a C-terminal HA tag. The \textit{hns} PCR product was digested with \textit{BamHI} and \textit{HindIII} and cloned into PNBV1, yielding pHNS-HA. The \textit{slyA} gene of \textit{S. enteric} serovar Typhimurium was PCR amplified using the forward primer 5’-AAAAGAATTCTTTATAAGGAGATTGGAATCGCCACTAGGTT-3’ and the reverse primer 5’-AAAACCTGAGACAAAGGAAAAT ACGCGTTTCTGCGC-3’. The PCR product was digested with \textit{EcoRI} and \textit{XhoI} and cloned into the arabinose-inducible expression vector pLC2002 (gift from Zhou Yu, University of Toronto, and Leslie Cuthbertson, McMaster University) pretreated with the same enzymes to generate pSLYA. All constructs were confirmed by DNA sequencing.

**B-Glucoside fermentation assay.** To assay for fermentation of aryl-\(\beta\),D-glucosides, I employed MacConkey agar (Difco) supplemented with 0.4% salicin. Bacteria were streaked on the plates and incubated at 37°C for 24 h. MacConkey agar is an indicator medium containing small peptides allowing for growth of gram-negative bacteria regardless of their ability to utilize a sugar that is supplemented into the agar. If bacteria can ferment the sugar, they will produce acidic by-products, causing a drop in pH which gives rise to red/pink colonies. If the bacteria are unable to ferment the sugar they will grow using the peptone and produce ammonia as a by-product of amino acid metabolism. In turn this leads to an increase in pH, resulting in white colonies.

**Motility assay.** The motility phenotype was assessed on tryptone swarm plates containing 1% Bacto tryptone, 0.5% NaCl, and 0.3% Bacto agar. Colonies were stabbed into the plates and incubated at 37°C for 5 h.

**Hemolysis assay.** Hemolytic activity was assessed on blood agar plates containing 3% tryptic soy broth, 1.5% agar, and 5% defibrinated sheep blood (Rockland). Three independent colonies per strain were stabbed into the plates and incubated at 37°C for 24 h. An area of clearing around a stab culture was indicative of hemolytic activity.
**ChIP assay.** Cultures of WN582 harboring PLSR2-HA or PLSR2-FLAG were grown to mid-logarithmic phase (optical density at 600 nm of 0.4 to 0.6) and treated with 1% formaldehyde for 15 min at room temperature. The cross-linking reaction was then quenched with 1.25 Mm glycine for 10 min. Cells were washed twice with ice-cold phosphate-buffered saline and sonicated to generate DNA fragments of ~500 bp. Cell lysates were precipitated with an anti-HA antibody (Sigma H9658) using agarose protein G beads (Calbiochem). The LSR2 protein tagged with the FLAG epitope served as a negative control, since it does not interact with the HA antibody.

**Quantitative RT-PCR.** Quantitative real-time PCR (RT-PCR) analyses were performed using the SYBR green mix from Sigma (S4438) according to the manufacturer’s instructions. Each primer set was done in triplicate. Primer pairs used for this analysis were as follows: forward 5’-ACACTGTTAACCAGCAAGACA-3’ and reverse 5’-GGATGAAAGCAACGCAACAGA-3’ for *bglG*; forward 5’-AGTTCCGTGCAAGAAGAGAAGA-3’ and reverse 5’-TGTTACGTTTTCGCTTACTTACT-3’ for *yjcF*; forward 5’-AATATTTGCGAGCATCCACACAGC-3’ and reverse 5’-TTTACCGAGCCGGAAATCCCAT-3’ for *proV*; forward 5’-GCATCGACGCGATTCTTCAAG-3’ and reverse 5’-GCAGCGCTTTAAAAATGTGTCGCGCC-3’ for *xapR*; forward 5’-GCAATCGACGCGATTCTTCAAG-3’ and reverse 5’-GCAATCGACGCGATTCTTCAAG-3’ for *narZ*; and forward 5’-GCGACGCAATGGCGGAGTTT-3’ and reverse 5’-AGTGGCAGGACCTTCCAGC-3’ for *phnE*.

**Reverse transcriptase quantitative PCR analysis of hlyE transcript.** The *E. coli* strains MC4100 and WN582/PLSR2-HA were transformed with the arabinose-inducible pSLYA plasmid. Cultures of MC4100, MC4100/pSLYA, WN582, WN582/pLSR2-HA, and WN582/pLSR2-HA+pSLYA were grown in LB media supplemented with 0.2% arabinose to mid-log phase (optical density at 600 nm of 0.5). The cells (0.5 ml) were mixed with 1 ml of RNAprotect bacterial reagent (Qiagen) and incubated for 15 min at room temperature. Subsequent RNA preparations were performed using the Aurum total RNA minikit (Bio-Rad 732-6820). Reverse transcription was performed using Superscript II reverse transcriptase (Invitrogen). The cDNA generated was used for quantitative RT-PCR analysis as described
above. The transcript of \textit{hlyE} was analyzed by using the primers 5’-TCCCTGGTAAGCTCACAAAGT-3’ (forward) and 5’-ACCGGCATATGCTTCCTCCTGAT-3’ (reverse). The transcript of \textit{gyrB}, a gene not regulated by H-NS (Navarre, Porwollik et al. 2006), was also analyzed and used as an internal standard for normalization among different samples. The primers for \textit{gyrB} are 5’-CACTTTCACCGAAACGACCGCAAT-3’ (forward) and 5’-TTACCAACAACATTTCCGCAGCGTG-3’ (reverse). Cultures grown in the absence of arabinose were used as the negative control.
A version of this chapter has been published (Gordon, B. R., Y. Li, L. Wang, A. Sintsova, H. van Bakel, S. Tian, W. W. Navarre, B. Xia, J. Liu. 2010. *PNAS* 107(11): 5154-5159). I performed all experiments as described.

### 3.1 Abstract

Lsr2 is a highly conserved DNA-binding protein found in mycobacteria and actinomycetes. Previous studies have established Lsr2 as a functional analog of H-NS of the Enterobacteriaceae, a global transcriptional repressor that targets AT-rich horizontally acquired DNA sequences. However, whether Lsr2 plays an equivalent role in *M. tb* biology remains unknown. In this chapter, I describe my findings that Lsr2 binds AT-rich regions dispersed throughout the *M. tb* genome, including genomic islands acquired by horizontal gene transfer. Other Lsr2-bound regions encode major virulence factors, such as the ESX secretion systems, the lipid virulence factors PDIM and PGL, and the PE/PPE families of antigenic proteins. Contrary to previous claims that *lsr2* is essential in *M. tb*, I was able to generate an *lsr2* deletion mutant. Microarray analysis revealed large-scale expression changes in the mutant with 540 genes upregulated versus 252 genes downregulated. Comparison of genome-wide binding data with expression data reveals that Lsr2 binding largely results in transcriptional repression. The
mutant’s transcriptome shows significant overlap with transcriptional responses obtained from in vitro bacteriostasis models thought to mimic the in vivo host environment. Dormant M. tb persisters are thought to be responsible for latent infection, the long course of drug therapy and disease recidivism. The genetic changes in the mutant would predispose it to being better able to adapt to dormancy and suggests that the mutant has realigned its metabolism to better suit subsisting from host-derived lipids. Taken together, I provide evidence that mycobacteria have employed a distinct molecule to achieve a function similar to H-NS, coordinating global gene regulation in this group of medically important bacteria.

3.2 Introduction

The bacterial chromosome is organized into a compact structure composed of topologically independent loops in part as a consequence of interactions with nucleoid-associated proteins (NAPs) (Mukamolova, Turapov et al. 2002). H-NS, one of the most abundant NAPs in the Enterobacteriaceae (Dorman 2004), is thought to stabilize these loops by forming long patches that bridge distant DNA helices (Noom, Navarre et al. 2007). In addition to its role in chromatin organization and compaction, H-NS plays a role in global gene regulation. H-NS exists essentially as a homodimer binding DNA nonspecifically, but with a preference for AT-rich or curved DNA (Owen-Hughes, Pavitt et al. 1992; Navarre, Porwollik et al. 2006). H-NS is responsible for binding and repressing >400 genes in Salmonella (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006) and in Escherichia coli (Grainger, Hurd et al. 2006; Oshima, Ishikawa et al. 2006), many of which are DNA sequences obtained through horizontal gene transfer and involved in adaptive stress responses and virulence (Navarre, McClelland et al. 2007). Numerous phenotypes associated with hns mutations have been described, and the effects of H-NS on gene expression are largely inhibitory (Dorman 2004; Mukamolova, Turapov et al. 2002).
This is partially explained by the ability of H-NS to bridge adjacent helices of DNA (Dame, Wyman et al. 2000; Dame, Noom et al. 2006), causing either the trapping or the occlusion of RNA polymerase in the promoter region (Dorman 2004; Dame, Luijsterburg et al. 2005). HNS homologs are widespread in the Gram-negative α-, β-, and γ-proteobacteria but have not been identified in Gram-positive bacteria or in any other groups of bacteria, leaving it unclear as to whether they employ similar mechanisms to regulate genes obtained through genetic exchange.

Lsr2 is a small, basic protein that is highly conserved in mycobacteria and related actinomycetes (Chen, German et al. 2006). Previous studies have shown that Lsr2 is involved in several cellular processes including cell wall lipid biosynthesis (Chen, German et al. 2006; Kocincova, Singh et al. 2008) and antibiotic resistance (Colangeli, Helb et al. 2007). Recently in vitro biochemical experiments demonstrated that, like H-NS, Lsr2 binds DNA in a sequence-independent manner and is capable of bridging distant DNA segments (Chen, Ren et al. 2008). Furthermore, I demonstrated through in vivo complementation assays that Lsr2 is a functional analog of H-NS; specifically, that lsr2 fully complements independent phenotypes associated with hns mutations in E. coli (Gordon, Imperial et al. 2008). These results suggest that Lsr2 may play a role in M. tb that is equivalent to that of H-NS. However, the in vivo binding sites of Lsr2 in the mycobacterial genomes and the scope of its biological function remain unknown.

Tuberculosis (TB) continues to be a global health problem. M. tb infections can persist asymptomatically for decades before reactivating to cause active disease (Mukamolova, Kaprelyants et al. 1998; Parish and Stoker 1998). The two greatest challenges facing the effective treatment of TB are the large reservoir of latent disease encompassing 1/3 the world’s population, as well as the > 6 months of antimycobacterial chemotherapy required to completely sterilize infected individuals. The root cause of both challenges is thought to be an ill-defined population of non-replicating antibiotic-tolerant bacilli. Various lines of evidence suggest that persistent bacteria reside in the granuloma, a lesion of immune cells which sequesters M. tb in a harsh microenvironment where it is exposed to low pH, NO, hypoxia, and nutrient restriction (Wayne and Hayes 1996; Wayne and Sohaskey 2001; Tufariello, Mi et al. 2006). In fact, all of these conditions have been shown to produce a non-replicating persistence (NRP) in vitro for M. tb (Wayne and Hayes 1996; Betts, Lukey et al. 2002; Voskuil, Schnappinger et al. 2003; Voskuil, Visconti et al. 2004; Rustad, Harrell et al. 2008). Furthermore, there is both in vitro...
and in vivo (McCune and Tompsett 1956; Jindani, Dore et al. 2003; Ahmad, Klinkenberg et al. 2009) evidence that M. tb stochastically forms a persister subpopulation as observed for other pathogens (Lewis 2010). A better understanding of the molecular mechanisms governing the formation, maintenance and reactivation of persisters during latent infection is necessary for the development of novel drug therapies to target persistent bacteria. Uncovering the mechanisms used by M. tb to enter into, survive, and reactivate from latent disease states is critical given the global burden of TB.

In this study, I demonstrate that Lsr2 fulfills an analogous regulatory role in M. tb as H-NS does in the enteric bacteria. Lsr2 binds AT-rich regions of the genome and negatively regulates its target genes including horizontally acquired and virulence genes. In addition, many of the genes regulated by Lsr2 are induced in various models of NRP. Overall, my findings suggest that Lsr2 is a global transcriptional repressor in M. tb that regulates genes required for persister formation.

3.3 Results

3.3.1 Lsr2 binds AT-rich sequences globally in Mycobacterial genomes

To gain insight into the biological function of Lsr2, I mapped the Lsr2 binding sites in M. tb and M. smegmatis by performing chromatin immunoprecipitation (ChIP) of in vivo cross-linked Lsr2–DNA complexes, followed by microarray (ChIP-chip) analysis on a 244 000-feature oligonucleotide tiling arrays. Lsr2 coprecipitated with 21% of the M. tb genome (840 of 4009 protein-encoding ORFs) and 13% of the genome in M. smegmatis (904 of 6716 ORFs). Like H-NS (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006), Lsr2 binds genome regions with low GC-content, irrespective of the position relative to the ORF (Figure 3.1A and B). Lsr2 preferentially bound regions with a GC-content of ~47% or less, compared to the genome average GC-content of 65-67% for these species.

Foreign DNA acquired by horizontal gene transfer often displays a lower GC-content relative to the rest of the genome (Daubin, Moran et al. 2003; Jang, Becq et al. 2008). As such, I investigated whether Lsr2 has a preference for foreign genes. Of the 76 potential horizontally acquired genes in the MTBC (those genes common to M. tb and M. bovis but absent from other mycobacterial species) (Becq, Gutierrez et al. 2007), Lsr2 bound 44 genes (57.8%), including the
well-studied Rv0986-Rv0989c region (Rosas-Magallanes, Deschavanne et al. 2006) (Figure 3.1C), which is statistically significant ($p$-value = $1.1 \times 10^{-12}$). The average GC-content of these 44 genes is significantly lower than the rest of the 32 genes not bound by Lsr2 ($p$-value < 0.001) (Figure 3.1F). The 76 horizontally transferred genes are distributed in 48 genomic islands (Becq, Gutierrez et al. 2007), and Lsr2 binds at least one of the genes in 34 of these islands (70.8%); in total, 101 genes of the 267 genes in the 48 islands (37.8%). Together, these results suggest that Lsr2 preferentially bind horizontally acquired AT-rich DNA.
3.3.2 Lsr2 binds genes involved in virulence and immunogenicity of M. tb

A survey of the annotated functional categories of M. tb genes (Cole, Brosch et al. 1998) reveals that Lsr2 binds a number of genes involved in intermediary metabolism/respiration and in cell-wall/cell processes, which are proportional to their distribution in the genome (Table 1). Lsr2 binds genes involved in energy metabolism (ATP synthase atpB, -E, -F, and -H), aerobic respiration (cytochrome C oxidase subunits ctaC, -D, -E), cell-wall peptidoglycan synthesis (dacB2, murA, murl, ponA1, ponA2), and mycolic acid synthesis (fabG1,fbpC, pcaA, umaA). Genes involved in chromosomal DNA replication (dnaA, recF, dnaEl, nrdE), transcription (sigA, rho), and protein synthesis (rrf, rrs, rrl, rplI, N, O, X, infC, rpmB2, rpmH, rpsT) are also bound by Lsr2. In addition, Lsr2 binds genes involved in stress response (ahpC, cspA, dnaJ1, sodA) and many regulatory genes including two-component systems (mprB, trcR, senX3) and the whiB family of transcription factors (whiB1–4).

The ChIP-chip data also reveal that Lsr2 binds several virulence-associated genes in M. tb. The M. tb genome contains gene clusters encoding five ESX family-type VII protein-
secretion systems (Gey Van Pittius, Gamieldien et al. 2001; Brodin, Rosenkrands et al. 2004; Abdallah, Gey van Pittius et al. 2007). ESX-1 is essential for M. tb virulence (Abdallah, Gey van Pittius et al. 2007), and a role of ESX-5 in virulence has been demonstrated in M. marinum (Abdallah, Verboom et al. 2006). Lsr2 binds multiple genes in four of the five ESX regions (Figure 3.1D), except ESX-4. A second locus shown to be involved in ESX-1 secretion (Rv3614c-espA) (Fortune, Jaeger et al. 2005) is also a target of Lsr2. The equivalent ESX-1 region in M. smegmatis (MSMEG_0055-MSMEG_0083), which is also bound by Lsr2, has shown to be involved in DNA transfer (Flint, Kowalski et al. 2004; Coros, Callahan et al. 2008). Other well-established virulence factors include the cell-wall lipid phthiocerol dimycocerosates (PDIMs) (Camacho, Ensergueix et al. 1999; Cox, Chen et al. 1999) and their closely related phenolic glycolipids (PGLs) (Reed, Domenech et al. 2004). There are multiple Lsr2 binding sites throughout the PDIM/PGL biosynthetic locus (Onwueme, Vos et al. 2005) (Figure 3.1E).

Table 3.1. Functional categories of Lsr2-bound genes in M. tb

<table>
<thead>
<tr>
<th>Functional categories</th>
<th>No. of genes bound by Lsr2</th>
<th>% Lsr2 reguIon</th>
<th>No. of genes in genome</th>
<th>% genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virulence, detoxification, adaptation</td>
<td>32</td>
<td>3.8</td>
<td>228</td>
<td>5.7</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>46</td>
<td>5.5</td>
<td>247</td>
<td>6.2</td>
</tr>
<tr>
<td>Information pathways</td>
<td>40</td>
<td>4.8</td>
<td>241</td>
<td>6.0</td>
</tr>
<tr>
<td>Cell wall and cell wall processes</td>
<td>157</td>
<td>18.7</td>
<td>773</td>
<td>19.3</td>
</tr>
<tr>
<td>Insertion sequences and phages</td>
<td>13</td>
<td>1.5</td>
<td>147</td>
<td>3.7</td>
</tr>
<tr>
<td>PE/PPE</td>
<td>89</td>
<td>10.6</td>
<td>168</td>
<td>4.2</td>
</tr>
<tr>
<td>Intermediary metabolism and respiration</td>
<td>157</td>
<td>18.7</td>
<td>923</td>
<td>23.1</td>
</tr>
<tr>
<td>Regulatory Proteins</td>
<td>42</td>
<td>5.0</td>
<td>195</td>
<td>4.9</td>
</tr>
<tr>
<td>Conserved hypotheticals/unknown</td>
<td>264</td>
<td>31.4</td>
<td>1081</td>
<td>27.0</td>
</tr>
</tbody>
</table>
A surprising finding from the sequencing of the \textit{M. tb} genome is that approximately 10% of the genome’s coding capacity encodes for two large families of proteins: the acidic asparagine- or glycine-rich proteins, referred to as PE (n=99) and PPE (n=69) proteins (Cole, Brosch et al. 1998). Remarkably, I found that Lsr2 binds over half of the total PE/PPE genes (89 of 168), which is more than double the number expected by chance (\textbf{Table 3.1}, highlighted in red). Lsr2 binds overwhelmingly to PPE genes (54 of 69) and PE genes that do not belong to the PE_PGRS subclass (22/36). By contrast, Lsr2 binds fewer PE_PGRS genes (13/63), which is likely explained by the high glycine content (up to 50%) of PE_PGRS proteins that makes their corresponding genes GC-rich. PE/PPE proteins are surface-exposed antigens characterized by extensive repetitive homologous sequences (Sampson, Lukey et al. 2001; Cascioferro, Delogu et al. 2007). Considerable sequence polymorphism has been reported for members of the PE/PPE proteins among the MTBC and clinical strains of \textit{M. tb} (Fleischmann, Alland et al. 2002). As such, PE/PPE proteins are thought to represent a source of antigenic variation (Cole, Brosch et al. 1998; Choudhary, Mukhopadhyay et al. 2003). The binding of Lsr2 to the majority of PE/PPE genes suggests that this factor may negatively affect the expression of these antigenic proteins to modulate interactions with the host.

\subsection*{3.3.3 Generation of a \textit{M. tuberculosis} lsr2::hyg mutant}

Although an \textit{lsr2} mutant of \textit{M. smegmatis} had been successfully generated (Chen, German et al. 2006), two groups subsequently failed to knock out \textit{lsr2} in \textit{M. tb} leading to the claim that it is likely essential in this species (Park, Dahl et al. 2008; Colangeli, Haq et al. 2009). However, using a previously described phage-transducing system (Bardarov, Bardarov Jr et al. 2002), I successfully constructed an \textit{lsr2} deletion mutant of \textit{M. tb} as confirmed by southern blot analysis(Figure 3.2). Briefly, to make a targeted disruption of \textit{lsr2} (Rv3597c) in \textit{M. tb} I first generated an allelic exchange substrate (AES) by PCR amplifying the up and downstream DNA (500-1000 bp) flanking the \textit{lsr2} gene and incorporated these fragments into a cosmid such that they flanked a hygromycin-resistance cassette gene. Next, I cloned the recombinant cosmid into the conditionally replicating TM4 shuttle plasmid and transfected \textit{M. smegmatis} with it at the permissive temperature of 30 °C to generate a high-titre of mycobacteriophage-packaged shuttle plasmids. The final step entailed transducing \textit{M. tb} with my \textit{lsr2} knock out phage at the
Figure 3.2. Southern blot confirmation of *M. tb* *lsr2* deletion mutant. (A) Shematic representation of the digestion strategy used for southern blot analysis of suspected *lsr2* deletion mutants. Total *M. tb* DNA was digested with the restriction enzymes *BbsI* and *PvuII* and then probed with an *lsr2* PCR product produces a band of 1918 bp. (B) Representative southern blot showing the absence of the expected band in the *lsr2* deletion mutant (*Δlsr2*).
non-permissive temperature of 37°C and plating out on 7H11 agar containing hygromycin to select for successful mutants. At this temperature phage replication is restricted resulting in the accumulation of a large number of abortive transductants. Allelic exchange then occurs as a result of a double crossover between the homologous DNA arms flanking the disrupted gene.

Strikingly the \textit{lsr2::hyg} (\textit{Δlsr2}) mutant displayed altered colony morphology, producing ‘mushroom-like’ colonies in stark contrast to the flat, spread out WT colony morphology (\textbf{Figure 3.3A}). For reasons which remain unclear, I observed that the ‘mushroom-like’ colony morphology was highly unstable and was lost after subsequent passaging in liquid media or restreaking on agar plates. Nevertheless, after extended growth on 7H11 agar plates (>6 weeks) differences in colony morphology could still be observed with the mutant colonies exhibiting a ‘blebbing’ phenotype (\textbf{Figure 3.3B}). Mycobacteria are notoriously clumpy when grown in liquid culture and as such are grown with detergents to yield a more homogenous cell suspension. The \textit{Δlsr2} mutant was hyper-aggregative in liquid media, even when a higher concentration of detergent was used (\textbf{Figure 3.3C}). In order to disperse the aggregates, \textit{Δlsr2} cultures were sonicated and aspirated through needles with limited success. The ‘blebbing’ colony morphology and hyper-aggregative phenotypes are complemented by transformation of the mutant with pME-LSR2, but not pME alone, demonstrating that the absence of \textit{lsr2} is responsible for these mutant phenotypes (\textbf{Figure 3.3B and C}).

Subtle defects in cell wall components can alter the surface properties of individual cells, affecting cell-to-cell interactions, ultimately giving rise to colonies with different morphology (Alexander, Jones et al. 2004). Thus, the changes in colony morphology and aggregation of the \textit{Δlsr2} mutant may suggest that the loss of \textit{lsr2} affects some component(s) of the cell wall. I analyzed the cell wall lipid composition of the \textit{Δlsr2} mutant and WT parental strains. The apolar (\textbf{Figure 3.3D}) and polar lipid (\textbf{Figure 3.3E}) fractions of the cell wall were extracted and examined by two-dimensional thin layer chromatography (2D-TLC) as previously described (Alexander, Jones et al. 2004). The results showed that there were no observable differences in the lipid profiles between the WT and \textit{Δlsr2} strains (\textbf{Figure 3.3D and E}). It is likely that changes in either a protein or carbohydrate component of the cell wall is responsible for the colony morphology change and hyper-aggregation of the mutant. Lsr2’s supposed role in regulating PE/PPE proteins may explain the colony morphology change since they are surface
exposed proteins (Brennan, Delogu et al. 2001; Sampson, Lukey et al. 2001; Banu, Honore et al. 2002) and have been shown to cause changes in cell structure such as shape and size, as well as colony morphology (Delogu, Pusceddu et al. 2004; Kana, Mizrahi et al. 2010).

The mutant also displays a growth delay on agar media (Figure 3.3F), with colonies becoming readily visible after 3 weeks versus 2 weeks of incubation at 37°C to obtain WT colonies of comparable size. In 7H9 broth, Δlsr2 shows slightly slower growth than WT, while a more pronounced growth delay was observed in sauton broth media (Figure 3.3G). The retarded growth rate observed in Δlsr2 mutants can be rescued with lsr2 on the shuttle vector pME, but not by pME alone.
3.3.4 Microarray analysis indicates Lsr2 functions predominantly as a global transcriptional repressor

In the case of *M. smegmatis*, comparing the genome-wide binding data to previously published microarray expression data for an *M. smegmatis* ∆lsr2 mutant (Colangeli, Helb et al. 2007), revealed that Lsr2 bound to 26 of the 41 (63%) genes displaying an equal or greater than two-fold increase in expression levels in the mutant. Conversely, none of the 18 genes found to have an equal or greater than two-fold decrease in expression levels in the mutant were bound by Lsr2, indicating that, like H-NS, Lsr2 is a pleiotropic factor that has a negative effect on gene expression. Consistent with this hypothesis, Lsr2 was previously shown to bind the promoter sequence of the *mps* operon and repress its expression, preventing the production of glycopeptidolipids in *M. smegmatis* (Kocincova, Singh et al. 2008).

I identified genes regulated by Lsr2 in *M. tb* by comparing transcript levels in WT and Δlsr2 exponential-phase cultures using a 15,000 feature oligonucleotide microarray with 3 independent probes per ORF. Overall the microarray analysis revealed that 530 protein-encoding ORFs were upregulated ≥2-fold (I hereby will refer to this gene set as LSR2_UP) while 240 protein-encoding ORFs were downregulated ≥2-fold (designated LSR2_DOWN) in the Δlsr2 mutant. The *lsr2* gene itself was the most significantly downregulated gene in the mutant which serves as an excellent internal control and provides confidence in my results. Of the 530 genes in LSR2_UP, 272 (51.3%) were bound by Lsr2 or immediately adjacent to an Lsr2 ChIP-chip binding peak (176 directly bound, *p*-value = 7.73×10^-13). Conversely, 75 of the 240 (31.3%) genes of LSR2_DOWN were bound by Lsr2 or adjacent to an Lsr2 ChIP-chip binding peak (65 directly bound, *p*-value = 0.011). These results suggest that Lsr2 acts primarily as a global transcriptional repressor, which is consistent with its DNA bridging activity and its functional
equivalence to H-NS (Chen, Ren et al. 2008; Gordon, Imperial et al. 2008; Gordon, Li et al. 2010; Gordon, Li et al. 2011). Genes that were downregulated in Δlsr2 are likely controlled indirectly by Lsr2. Gene downregulation in hns mutants has also been described due to the derepression of transcriptional repressors (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006).

To determine if a particular functional category is enriched for gene expression changes in the mutant, I performed a survey of the annotated functional categories (Cole, Brosch et al. 1998) of the genes in the LSR2_UP and LSR2_DOWN gene sets (Table 3.2). I found that nearly 10% of the LSR2_UP list is comprised of putative regulatory proteins, nearly double the number expected by chance (Table 3.2, highlighted in red). Regulatory proteins of key interest are the two-component response sensor PhoP, three members of the WhiB gene family including WhiB3, as well as the alternative sigma factors SigB, SigE and SigF. Conversely, the LSR2_DOWN is largely devoid of regulatory proteins with 6-fold less than expected by chance. Based on these findings, it is tempting to speculate that Lsr2 is at the helm of a number of regulatory cascades which would be initiated when cellular concentrations of Lsr2 protein become depleted. This may partially explain why Lsr2 is such a peiotropic regulator, especially for genes displaying differential expression in the mutant but lacking direct Lsr2 binding activity as indicated by my ChIP-chip data.

Several Lsr2-bound gene clusters encoding established virulence factors are upregulated in Δlsr2, including the exs-1 and functionally linked espA-D (Abdallah, Gey van Pittius et al. 2007; Simeone, Bottai et al. 2009), mmpL4-mmpS4 (Domenech, Reed et al. 2005), Rv1501-Rv1507c and moaA1-moaD1 (Brodin, Poquet et al. 2010) which is consistent with our finding that Lsr2 is a functional analog of H-NS (Gordon, Imperial et al. 2008; Gordon, Li et al. 2010; Gordon, Li et al. 2011). H-NS serves as a central regulator of virulence gene expression in Enterobacteria. For example, H-NS silences expression of all 5 major pathogenicity islands (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006). I also found a number of the PE/PPE genes are differentially expressed in the mutant. Genomic analysis suggested that individual PE proteins are likely protein partners for PPE proteins, and the structure of a PE/PPE heterodimer has been solved (Strong, Sawaya et al. 2006). A follow up study using a computational approach predicted 289 possible PE/PPE complexes with 35 of these pairs having highly correlated mRNA expression patterns under a variety of conditions, providing additional...
evidence that these represent *bona fide* complexes (Riley, Pellegrini et al. 2008). Interestingly, within the 21 PE/PPE genes in LSR2_UP, there are 6 of the 35 predicted PE/PPE complex pairs are present (PE_PGRS11/PPE29, PE8/PPE15, PE11/PPE17, PE22/PPE36, PE27/PPE15 and PE35/PPE68). There are 9 PE/PPE genes in LSR2_DOWN with 2 of the predicted pairs present (PE31/PPE18 and PE31/PPE60). Taken together Lsr2 negatively regulates a number of genes shown to be important for survival within the host and is a virulence gene regulator like H-NS.
<table>
<thead>
<tr>
<th>Functional categories</th>
<th>No. of genes up-regulated</th>
<th>% genes up-regulated</th>
<th>No. of genes down-regulated</th>
<th>% genes down-regulated</th>
<th>No. of genes in genome</th>
<th>% genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virulence, detoxification, adaptation</td>
<td>44</td>
<td>8.2</td>
<td>12</td>
<td>4.8</td>
<td>228</td>
<td>5.7</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>29</td>
<td>5.4</td>
<td>18</td>
<td>7.1</td>
<td>247</td>
<td>6.2</td>
</tr>
<tr>
<td>Information pathways</td>
<td>10</td>
<td>1.9</td>
<td>3</td>
<td>1.2</td>
<td>241</td>
<td>6.0</td>
</tr>
<tr>
<td>Cell wall and cell wall processes</td>
<td>95</td>
<td>17.6</td>
<td>58</td>
<td>23.0</td>
<td>773</td>
<td>19.3</td>
</tr>
<tr>
<td>Insertion sequences and phages</td>
<td>24</td>
<td>4.5</td>
<td>26</td>
<td>10.3</td>
<td>147</td>
<td>3.7</td>
</tr>
<tr>
<td>PE/PPE</td>
<td>21</td>
<td>3.9</td>
<td>9</td>
<td>3.6</td>
<td>168</td>
<td>4.2</td>
</tr>
<tr>
<td>Intermediary metabolism and respiration</td>
<td>116</td>
<td>21.5</td>
<td>43</td>
<td>17.1</td>
<td>923</td>
<td>23.1</td>
</tr>
<tr>
<td>Regulatory Proteins</td>
<td>50</td>
<td>9.3</td>
<td>2</td>
<td>0.8</td>
<td>195</td>
<td>4.9</td>
</tr>
<tr>
<td>Conserved hypotheticals/unknown</td>
<td>137</td>
<td>25.4</td>
<td>48</td>
<td>19.0</td>
<td>1081</td>
<td>27.0</td>
</tr>
</tbody>
</table>
3.3.5 The LSR2_UP is enriched for genes induced during persistence

There are several *in vitro* models of persistence thought to mimic some aspects of the *in vivo* conditions encountered by *M. tb* during chronic and latent infections. I compared LSR2_UP with published transcriptome data from these models and tabulated the number of overlapping genes (Table 3.3). I analyzed the genes upregulated ≥2-fold in the following *in vitro* bacteriostasis models: (i) nutrient starvation of an exponential-phase culture that was starved in phosphate-buffered saline (PBS) (Betts, Lukey et al. 2002); (ii) the enduring hypoxic response (EHR) in which hypoxia was generated by a continuous flow of low oxygen culture media (Rustad, Harrell et al. 2008); (iii) the Wayne model which generates hypoxia using sealed culture vials with restricted headspace air (Wayne and Sramek 1994; Wayne and Hayes 1996); (iv) the 49 genes of the Dos regulon (Rustad, Harrell et al. 2008); (v) *M. tb* persisters isolated from an exponential-phase culture treated with D-cycloserine (DCS) for 14 days to lyse actively replicating cells (Keren, Minami et al. 2011); and (vi) *M. tb* treated with 10×MIC of vancomycin for 4 hours, a drug that inhibits cell wall synthesis causing envelope stress (Provvedi, Boldrin et al. 2009). To account for the varying sizes of the lists used for comparison, I estimated the number of genes expected to be shared between LSR2_UP and a given dataset based upon the size of LSR2_UP relative to the entire genome (530/4009 = 13.2%).

The LSR2_UP shows the most significant overlap with genes of WT *M. tb* induced under long-term hypoxia (models (ii) and (iii) above). Of the 230 EHR genes induced in *M. tb* after 7 days incubation under hypoxia (Rustad, Harrell et al. 2008), 117 were up-regulated in Δlsr2, which is statistically significant (*p*-value = 1.8×10^{-46}) (Figure 3.4A, Table 3.3). Another study identified hypoxia-induced genes of *M. tb* using the Wayne model for an extended time period (Voskuil, Visconti et al. 2004). In this model, days 6-8 represented early NRP (NRP1), days 10-20 represented middle NRP (NRP2), and days 30-80 represented late NRP (NRP3). Comparison between this dataset with LSR2_UP revealed that during NRP1, the overlap is equal to or slightly greater than the overlap expected by chance, and the overlap increases over the course of the NRP state reaching 57% (*p*-value =5.7×10^{-15}) by day 30 which is 4 times the overlap expected by chance (Figure 3.4B, Table 3.3). These data suggest that Lsr2 controls the expression of a significant number of genes that primarily function in the later stages of hypoxia-induced dormancy. Conversely, I found no significant overlap between LSR2_UP and genes
induced by DosR, which is an early and transient response to hypoxia (Rustad, Harrell et al. 2008). Only 8 of 49 (16%, \( p \)-value = 0.13) genes regulated by DosR were up-regulated in \( \Delta lsr2 \), which is slightly below the rate expected by random chance. This is not unexpected given that the \( dosR \) gene itself is downregulated in \( \Delta lsr2 \). Another stimulus known to induce dormancy in \( M. \ tb \) in vitro is nutrient starvation. Of the 276 genes induced following 4 days incubation in PBS (Betts, Lukey et al. 2002), 75 (27%, \( p \)-value = 6.7×10\(^{-11}\)) overlapped with the LSR2_UP which is double the number expected by chance (Table 3.3). Exposure of exponentially growing \( M. \ tb \) with vancomycin at 10×-MIC resulted in bacteriostasis (Provvedi, Boldrin et al. 2009). Microarray analysis identified 111 genes induced 4 hours post-treatment, of which 57 (51%, \( p \)-value = 9.8×10\(^{-23}\)) are upregulated in \( \Delta lsr2 \) (Table 3.3). Given that LSR2_UP shows significant overlap with genes induced during NRP in vitro, this suggests that Lsr2 regulates stress-response genes based on environmental conditions.

To date, all pathogens examined produce a small subpopulation of dormant persister cells in growing exponential cultures (Lewis 2010) including \( M. \tb \) (Keren, Minami et al. 2011). Persisters display multi-drug tolerance to antibiotics and they are thought to contribute to the recalcitrance of both acute and latent TB infections to drug therapy. Importantly, they are not resistance-mutants but are phenotypic variants of the WT (Keren, Kaldalu et al. 2004) and have been shown to form stochastically prior to the addition of antibiotics (Balaban, Merrin et al. 2004). Isolation and transcriptome analysis of \( M. \ tb \) persisters revealed wide-scale downregulation of biosynthesis and energy-producing functions (Keren, Minami et al. 2011) as observed for similar analysis of \( E. \ coli \) persisters (Shah, Zhang et al. 2006). Similarly, the \( M. \ tb \) persister transcriptome contained 98 genes whose expression was up-regulated, likely representing genes with important functions for the establishment and maintenance of persisters. Comparing the persister specific genes with LSR2_UP revealed significant overlap with 46 (47%, \( p \)-value = 1.5×10\(^{-16}\)) common to both datasets (Table 3.3), indicating that Lsr2 negatively regulates a large portion of genes important for persister formation. The significant degree of concordance between LSR2_UP and the previously reported microarray datasets from dormant \( M. \ tb \) populations strongly suggests that Lsr2 plays a critical role in stress adaptation as well as in the formation of persisters.
Table 3.3. Comparison of LSR2_UP with previously published microarray analysis of bacteriostasis

<table>
<thead>
<tr>
<th>Model</th>
<th>Condition</th>
<th>No. of genes up-regulated in study</th>
<th>LSR2_UP: 530 genes</th>
<th>Overlap (By Chance)</th>
<th>% overlap</th>
<th>p - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starvation</td>
<td>4 days in PBS</td>
<td>276</td>
<td></td>
<td>75 (36)</td>
<td>27%</td>
<td>6.7×10^{-11}</td>
</tr>
<tr>
<td>DosR</td>
<td>4 hours hypoxia</td>
<td>49</td>
<td></td>
<td>8 (7)</td>
<td>16%</td>
<td>0.13</td>
</tr>
<tr>
<td>Wayne model</td>
<td>hypoxia day 6</td>
<td>94</td>
<td></td>
<td>15 (13)</td>
<td>16%</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>hypoxia day 14</td>
<td>100</td>
<td></td>
<td>39 (13)</td>
<td>39%</td>
<td>4.1×10^{-11}</td>
</tr>
<tr>
<td></td>
<td>hypoxia day 20</td>
<td>179</td>
<td></td>
<td>94 (24)</td>
<td>53%</td>
<td>1.6×10^{-38}</td>
</tr>
<tr>
<td></td>
<td>hypoxia day 30</td>
<td>56</td>
<td></td>
<td>32 (8)</td>
<td>57%</td>
<td>5.7×10^{-15}</td>
</tr>
<tr>
<td>Enduring Hypoxic</td>
<td>hypoxia day 7</td>
<td>230</td>
<td></td>
<td>117 (31)</td>
<td>51%</td>
<td>1.8×10^{-46}</td>
</tr>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persisters</td>
<td>DCS-treated day 14</td>
<td>98</td>
<td></td>
<td>46 (13)</td>
<td>47%</td>
<td>1.5×10^{-16}</td>
</tr>
<tr>
<td>Vancomycin</td>
<td>4 hours 10×-MIC</td>
<td>111</td>
<td></td>
<td>57 (15)</td>
<td>51%</td>
<td>9.8×10^{-23}</td>
</tr>
</tbody>
</table>
Figure 3.4. Overlap of genes upregulated in *M. tb* Δ*lsr2* and genes of WT *M. tb* induced by hypoxic conditions. (A) Of the 230 EHR genes identified by (Rustad, Harrell et al. 2008), 117 genes are upregulated in Δ*lsr2* (*p*-value = 1.24×10^{-45}). (B) Overlap of genes up-regulated in Δ*lsr2* and genes induced in WT *M. tb* under long-term hypoxia (up to 80 days) identified by (Voskuil, Visconti et al. 2004). Each bar represents the total number of genes induced at different days under hypoxic conditions using the Wayne model and the fractions of genes up-regulated in Δ*lsr2* are shown in red. The overlaps are statistically significant.
3.3.6 Lsr2 regulates genes involved in stress response and persister formation

I found a number of genetic changes in the Δlsr2 mutant consistent with the notion that Lsr2 regulates genes involved in stress response and persister formation. Induction of the toxin-antitoxin (TA) modules is a potential mechanism for the formation of *M. tb* persisters and has been reported during starvation (Betts, Lukey et al. 2002) and hypoxia (Rustad, Harrell et al. 2008). TA families including HipBA, MazEF, RelBE, CcdAB, ParDE, HigBA, Phd/doc, and VapBC have been described to play a role in persister formation (Hayes 2003). The TA systems are massively expanded in *M. tb* with 88 putative TA systems found in its genome based on homology to TA systems present in other bacteria (Pandey and Gerdes 2005; Ramage, Connolly et al. 2009). As such, several groups have begun to systematically delineate the role of individual TA systems in *M. tb*. Overexpression of 32 of the putative TA systems in *M. smegmatis* or *E. coli* induced bacteriostasis and/or drug tolerance, confirming their roles as bona fide TA systems (Korch, Contreras et al. 2009; Ramage, Connolly et al. 2009; Singh, Barry et al. 2010); nevertheless, their role in *M. tb* persistent infection has not been investigated. My transcriptional analysis indicates that 31 of the 88 putative TA systems are activated in *M. tbΔlsr2*, including 15 TA systems belonging to the 32 confirmed systems (Table 3.4). The most common mechanism of TA system toxicity is mediated through mRNA cleavage, resulting in translation inhibition (Gerdes, Christensen et al. 2005). Consistent with this observation, ribosomal proteins have been reported to be downregulated in dormancy models including starvation (Betts, Lukey et al. 2002), the Wayne model (Voskuil, Visconti et al. 2004), and persisters (Keren, Minami et al. 2011). In fact, a persister subpopulation was successfully isolated from an *E. coli* population expressing a degradable GFP under the control of a ribosomal promoter. As a result, dim cells could be sorted and were confirmed as drug tolerant persisters (Shah, Zhang et al. 2006). I noticed that are 16 of the 60 (27%) genes encoding for ribosomal proteins present in the *M. tb* H37Rv genome are downregulated ≥2-fold in Δlsr2 compared to WT, while none are upregulated. Taken together with the activation of TA systems these findings suggest that like persisters, the mutant displays limited translation activity, indicating a metabolic downshift.

Table 3.4. The experimentally validated TA systems that are activated in *M. tbΔlsr2*
<table>
<thead>
<tr>
<th>Rv #</th>
<th>Gene</th>
<th>Annotation</th>
<th>Fold Change</th>
<th>q-value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0277A</td>
<td>vapB25</td>
<td>VapBC family antitoxin</td>
<td>2.13</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv0277c</td>
<td>vapC25</td>
<td>VapBC family toxin</td>
<td>2.46</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv0456B</td>
<td>mazE1</td>
<td>MazEF family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv0456A</td>
<td>mazF1</td>
<td>MazEF family toxin</td>
<td>2.05</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv0596c</td>
<td>vapB4</td>
<td>VapBC family antitoxin</td>
<td>2.32</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv0595c</td>
<td>vapC4</td>
<td>VapBC family toxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv0748</td>
<td>vapB31</td>
<td>VapBC family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv0749</td>
<td>vapC31</td>
<td>VapBC family toxin</td>
<td>2.43</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv1103c</td>
<td>mazE3</td>
<td>MazEF family antitoxin</td>
<td>2.47</td>
<td>0.59</td>
</tr>
<tr>
<td>Rv1102c</td>
<td>mazF3</td>
<td>MazEF family toxin</td>
<td>2.05</td>
<td>0.88</td>
</tr>
<tr>
<td>Rv1956</td>
<td>higA1</td>
<td>HigAB family antitoxin</td>
<td>4.52</td>
<td>0.47</td>
</tr>
<tr>
<td>Rv1955</td>
<td>higB1</td>
<td>HigAB family toxin</td>
<td>3.31</td>
<td>0.38</td>
</tr>
<tr>
<td>Rv1960c</td>
<td>parD1</td>
<td>ParDE family antitoxin</td>
<td>3.69</td>
<td>0.47</td>
</tr>
<tr>
<td>Rv1959c</td>
<td>parE1</td>
<td>ParDE family toxin</td>
<td>3.45</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv1991A</td>
<td>mazE6</td>
<td>MazEF family antitoxin</td>
<td>No change</td>
<td>0.59</td>
</tr>
<tr>
<td>Rv1991c</td>
<td>mazF6</td>
<td>MazEF family toxin</td>
<td>3.62</td>
<td>0.00</td>
</tr>
<tr>
<td>Rv2021c</td>
<td>Rv2021c</td>
<td>RelBE family antitoxin</td>
<td>3.67</td>
<td>0.47</td>
</tr>
<tr>
<td>Rv2022c</td>
<td>Rv2022c</td>
<td>RelBE family toxin</td>
<td>2.49</td>
<td>0.74</td>
</tr>
<tr>
<td>Rv2142A</td>
<td>parD2</td>
<td>ParDE family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv2142c</td>
<td>parE2</td>
<td>ParDE family toxin</td>
<td>2.47</td>
<td>0.31</td>
</tr>
<tr>
<td>Rv2801A</td>
<td>mazE9</td>
<td>MazEF family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv2801c</td>
<td>mazF9</td>
<td>MazEF family toxin</td>
<td>2.18</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv2830c</td>
<td>vapB22</td>
<td>VapBC family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv2829c</td>
<td>vapC22</td>
<td>VapBC family toxin</td>
<td>3.62</td>
<td>0.10</td>
</tr>
<tr>
<td>Rv2865</td>
<td>relB2</td>
<td>RelBE family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv2866</td>
<td>relE2</td>
<td>RelBE family toxin</td>
<td>3.13</td>
<td>0.00</td>
</tr>
<tr>
<td>Rv3385c</td>
<td>vapB46</td>
<td>VapBC family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv3384c</td>
<td>vapC46</td>
<td>VapBC family toxin</td>
<td>2.24</td>
<td>0.30</td>
</tr>
<tr>
<td>Rv3407</td>
<td>vapB47</td>
<td>VapBC family antitoxin</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Rv3408</td>
<td>vapC46</td>
<td>VapBC family toxin</td>
<td>2.55</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Dormant M. tb persisters pose a major health concern since this population of bacilli is innately resistant to current drugs that only target actively replicating cells. Furthermore, persisters can undergo resuscitation which involves the reversal of NRP into an actively replicating pathogenic state (Chao and Rubin 2010). Analysis of the M. tb genome revealed that it encodes five genes with substantial homology to the resuscitation-promoting factor (Rpf) of Micrococcus luteus (Cole, Brosch et al. 1998). The Rpf of M. luteus is a small secreted protein...
capable of restoring active growth to viable but not culturable (VBNC) *M. luteus* cultures after prolonged incubation in stationary phase (Mukamolova, Kaprelyants et al. 1998). The five *M. tb* Rpf-like proteins (RpfA-E) have also been demonstrated to be potent (pM-fM concentrations) growth stimulators of *in vitro* VBNC persistent mycobacteria (Mukamolova, Turapov et al. 2002). Addition of recombinant mycobacterial Rpf facilitates the growth of freeze dried BCG (Wu, Yang et al. 2008) and significantly increases the recovery of *M. tb* from the sputum of active tuberculosis patients, indicating that the majority of sputum-borne bacilli are in a dormant state and can reactivate in an Rpf-dependent manner (Mukamolova, Turapov et al. 2010). *M. tb rpf* deletion mutants have been found to be defective for *in vivo* reactivation in a chronic murine infection model (Tufariello, Mi et al. 2006; Russell-Goldman, Xu et al. 2008). All five *rpf* genes are downregulated in the mutant (*rpfA* and *rpfB* are downregulated >2-fold) suggesting its expression profile is conducive to a persistent cellular state.

Evidence of metabolic reprogramming has been reported in *M. tb* under hypoxia and during persistent infection. A growing body of evidence suggests that *M. tb* uses host lipids rather than carbohydrates as a source of carbon and energy during persistent infection (Russell, VanderVen et al. 2010). To address whether the *M. tb Δlsr2* has altered metabolism favoring persistence within the host, I employed the Pathway Tools Software (Wang, Slayden et al. 2000) which contains manually curated and verified metabolic pathways of *M. tb* from the TB database ([http://www.tbdb.org/](http://www.tbdb.org/)). This program allows input of microarray expression data to identify metabolic pathways that are affected in specific experiments. Using this program, I found that three pathways related to cholesterol catabolism are activated in Δlsr2: the β-oxidation of cholesterol, the methylcitrate cycle, and the methylmalonyl pathway (Figure 3.5-Figure 3.7). β-Oxidation of cholesterol, methyl-branched fatty acids and odd-chain lipids yields propionyl-CoA (Russell, VanderVen et al. 2010), which is toxic to *M. tb* unless metabolized (Savvi, Warner et al. 2008). In turn, propionyl-CoA is metabolized via two pathways, the methylcitrate cycle (Munoz-Elias, Upton et al. 2006; Upton and McKinney 2007) and the methylmalonyl pathway (Savvi, Warner et al. 2008). Nearly all genes involved in these three pathways are overexpressed in Δlsr2 (Figure 3.6 and Figure 3.7).
Figure 3.5. Metabolism of cholesterol and methyl-branched or odd-chain fatty acids. This figure is from a recent review article (Russell, VanderVen et al. 2010). The cholesterol degradation pathway, the methylcitrate cycle, and the methylmalonyl pathway are activated in Δlsr2 of M. tb(indicated by red arrow). By contrast, the biosynthetic pathway of PDIM is downregulated in Δlsr2 (indicated by red cross). The details of these pathways are shown in Figure 3.6 through to Figure 3.8.
Figure 3.6. The cholesterol β-oxidation pathway is activated in *M. tuberculosis* Δlsr2. This figure was generated using the Pathway Tools software suite. The fold changes in expression of genes in the pathway are indicated. All genes except *hsaC* (Rv3568c) are upregulated ≥2 fold in *M. tuberculosis* Δlsr2. The expression of *hsaC* in Δlsr2 is 1.66 fold of that in the WT strain.
Figure 3.7. Methylcitrate cycle and methylmalonyl pathway are activated in Δlsr2 of M. tb. This figure is modified from a recent review article (Rhee, de Carvalho et al. 2011). Genes involved in the methylcitrate cycle (prpC, prpD, icle) are upregulated by ≥2 fold in Δlsr2, and the expression of acn is 1.1 fold of that in WT. For genes involved in the methylmalonyl pathway, accA2 and accD2, which could also be involved in the first step of this pathway (http://tbcyc.tbdb.org/), and mutA are upregulated by ≥2 fold in Δlsr2. The glyoxylate shunt pathway may also be activated since icle, which also functions in the first step of this pathway, is up-regulated.
Propionyl-CoA is also a precursor for the biosynthesis of methyl-branched lipids in *M. tb* (Figure 3.5) (Jain, Petzold et al. 2007). *M. tb* produces many methyl-branched lipids in the cell wall, including virulence-associated sulfolipid (SL-1), phthiocerol dimycocerosates (PDIM), and phenolic glycolipids (PGL) (Jackson, Stadthagen et al. 2007). The metabolism of propionyl-CoA by the methylmalonyl pathway generates methylmalonyl-CoA precursors for the synthesis of these lipids (Figure 3.5) (Jain, Petzold et al. 2007). The biosynthetic locus of PDIM and PGL contains more than 30 genes (Figure 3.8) (Onwueme, Vos et al. 2005). Intriguingly, many genes known to be involved in PDIM and/or PGL biosynthesis are downregulated in Δlsr2 (Figure 3.8), suggesting that the Δlsr2 shuts down PDIM/PGL biosynthesis, which may redirect the carbon flux for more essential functions (e.g., TCA cycle) under carbon limiting conditions. The reduction of lipid biosynthesis is likely preserved in *in vivo* infections where it has been reported that *M. tb* loses its acid-fastness and becomes difficult to detect (Ulrichs and Kaufmann 2006).
Figure 3.8. The biosynthetic pathway of PDIM and PGL. (A) Genetic organization of the biosynthetic locus of PDIM and PGL (taken from (Onwueme, Vos et al. 2005)). (B) Major biosynthetic pathways for PDIM and PGL (taken from (Mukamolova, Yanopolskaya et al. 1998)). Multiple genes in this locus (fadD26, ppsD, ppsE, drrA, drrC, papA5, fadD28, lppX, fadD22, fadD29) are downregulated by ≥2 fold in M. tb Δlsr2.
A crucial feature in the adaptation of *M. tb* to alternative energy sources (e.g., cholesterol) and changing environmental conditions (e.g., oxygen tension) is the cellular balance of oxidative and reductive reactions in the metabolic scheme. In *M. tb*, there are two respiratory routes. The major aerobic respiratory route is comprised of the type-I NADH dehydrogenase (encoded by *nuoA*-N) along with the *bc1*-aa3 cytochrome pathway (encoded by *qcrCAB* and *ctaBCDE*, respectively), while the major route under hypoxia is comprised of the type-II NADH dehydrogenase (*ndh* and *ndhA*) and the *bd* cytochrome pathway (*cydABCD*) (Figure 3.9) (Matsoso, Kana et al. 2005; Shi, Sohaskey et al. 2005; Cook, Berney et al. 2009). Transition from the acute to the persistent phase of infection in *M. tb*-infected mice is demarcated by a switch in bacterial respiration from the aerobic to the hypoxic route (Shi, Sohaskey et al. 2005). In the *M. tb∆lsr2*, I observed that the type-I NADH dehydrogenase genes (*nuoD* and *nuoE*) are downregulated while the type-II NADH dehydrogenase (*ndh*) is up-regulated. As well, the genes of the cytochrome *bd* oxidase (*cydA* and *cydD*) were also up-regulated in ∆lsr2. Other reducing equivalent regeneration pathways are also upregulated in the mutant. The isocitrate lyase gene *icl1* is significantly upregulated in the mutant. This enzyme cleaves isocitrate to yield succinate and glyoxylate. Isocitrate lyase activity increases in hypoxic *M. tb* cultures (Wayne and Lin 1982) and is required for *in vivo* persistence (McKinney, Honer zu Bentrup et al. 2000). In turn glyoxylate can reenter the TCA via malate synthase, which catalyzes the condensation of glyoxylate with the acetate from acetyl-CoA (Figure 3.7). Alternatively, in *M. tb* glyoxylate can be utilized via another pathway involving the reductive amination of glyoxylate by glycine dehydrogenase with the concomitant oxidation of NADH to NAD+. Interestingly, hypoxic *M. tb* cultures were reported to have a 10-fold increase in glycine dehydrogenase activity while malate synthetase activity did not increase, suggesting that the main function of the glyoxylate-to-glycine shunt is to replenish NAD+ to support minimal metabolism during NRP hypoxia when oxygen is not available as a terminal electron acceptor (Wayne and Lin 1982). I found that the gene encoding for glycine dehydrogenase, *gcvB*, is significantly upregulated in ∆lsr2. Regeneration of NAD+ during hypoxia may also occur via other dehydrogenases such as alanine dehydrogenase, which couples the reductive amination of pyruvate to alanine with the oxidation of NADH (Hutter and Dick 1998; Wayne and Sohaskey 2001; Koul, Arnoult et al. 2011). The gene encoding for alanine dehydrogenase, *ald*, is induced during starvation (Betts, Lukey et al. 2002), and is also upregulated in the mutant.
Taken together, the activation of the TA systems and suppression of growth initiation factors in the \textit{M. tb}∆\textit{lslr2} indicate it is likely predisposed to persistence. Moreover it appears to have realigned its redox and energy status to favor growth using lipid substrates in oxygen limiting conditions, both of which are thought to predominate in the host environments during latent infection.
Figure 3.9. Electron transport chain in *M. tb* under aerobic or hypoxic conditions (modified from (Cook, Berney et al. 2009)). Type-I NADH dehydrogenase and the *bc1–aa3* cytochrome pathway constitute the major respiratory route under aerobic conditions. Type-II NADH dehydrogenase and the *bd* cytochrome pathway are the major route under hypoxic conditions.
3.4 Discussion

In this study, I provide evidence that mycobacteria have employed a distinct molecule (Lsr2) to achieve an H-NS equivalent function. Lsr2 is similar in function to H-NS in several aspects, including its ability to target AT-rich sequences and silence gene expression by trapping or occluding RNA polymerase at the promoter region via DNA looping or bridging activity (Dorman 2004; Chen, Ren et al. 2008). H-NS also interacts with DNA via the DNA stiffening mode of DNA binding, but it is unknown whether Lsr2 can also bind DNA in the stiffening mode. As a consequence of their binding preference, both proteins have a predilection for xenogeneic genetic material, including multiple loci important for virulence and stress response. My ChIP-on-chip analysis using high-resolution oligonucleotide tiling arrays revealed a strong correlation between the percentage of GC-content and Lsr2 binding. Lsr2 preferentially targets genome regions with a GC-content of ~47% or less, which is similar to that of H-NS-repressed genes; the average GC-content of an H-NS-repressed ORF in Salmonella is 46.8% (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006). This provides an explanation as to why Lsr2 can fully complement independent phenotypes associated with hns mutations in E. coli and vice-versa that H-NS is capable of complementing an lsr2 mutation in M. smegmatis (Gordon, Imperial et al. 2008). Similar to H-NS (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006), Lsr2 binding is not restricted to promoter regions, and many of the Lsr2-binding sites are within coding regions, suggesting that, like H-NS (Dame, Wyman et al. 2002), Lsr2 can silence gene expression by polymerizing along DNA and bridging adjacent helices, a biochemical property that has previously been demonstrated for Lsr2 (Chen, Ren et al. 2008). My transcriptome analysis of the M. tb Δlsr2 revealed large-scale changes in gene expression with 540 genes upregulated and 252 genes downregulated, which is similar in scale with changes observed in a Salmonella Δhns strain (535 genes upregulated and 330 genes downregulated) (Lucchini, Rowley et al. 2006). Analyzing the microarray data with ChIP-chip data revealed that a large proportion of the genes bound by Lsr2 are de-repressed in the mutant.

The ability of M. tb to enter, maintain and reactivate from dormancy is central to its pathogenicity and underscores the difficulty eradicating this pathogen. Unfortunately, the stimuli and molecular mechanisms governing the dormancy program remain largely unknown. Because Lsr2 has such an overarching regulatory role in M. tb, I compared my microarray data to previously published microarray datasets for M. tb from in vitro models of bacteriostasis to
determine if Lsr2 regulates genes involved in dormancy. The host granuloma is thought to constrain bacterial replication by limiting oxygen (Wayne and Sohaskey 2001; Rustad, Sherrid et al. 2009), which is required for growth of M. tb (Wayne and Diaz 1967; Wayne 1976). I found that of the 230 genes activated in the EHR (Rustad, Harrell et al. 2008), 51% are up-regulated in the mutant. Conversely, the Dos regulon, which comprises the initial hypoxic response, does not show significant overlap with LSR2_UP, suggesting Lsr2 negatively regulates genes required for long-term hypoxia, and not the short-term response. Nutrient limitation is another condition thought to arrest bacterial replication in the granuloma. I found 75 of the 276 starvation-induced genes (Betts, Lukey et al. 2002) are in common with LSR2_UP which is twice the overlap expected by chance. Considerable overlap has been noted previously between the EHR and starvation transcription profiles (Rustad, Harrell et al. 2008). These findings indicate that Lsr2 negatively regulates a core set of stress response genes that promote persistence as a survival strategy. Further supporting this notion is that 57 out of the 111 activated genes identified from bacteriostatic M. tb treated with the antibiotic vancomycin were also upregulated in the mutant. Moreover, a recent study isolated drug-tolerant M. tb persisters from an actively growing culture (Keren, Minami et al. 2011). The persisters’ transcriptome had 46 of its 98 activated genes in common with LSR2_UP. Based on these comparative analyses, this suggests the presence of a general transcriptional signature of M. tb growth arrest that is negatively regulated by Lsr2.

I identified specific genetic changes in the mutant that are consistent with my hypothesis that Lsr2 suppresses a stress response/dormancy program. Overall I found that 31 of the 88 putative TA systems in M. tb are activated in the mutant. In addition, 6 of the 10 TA systems found to be induced in M. tb persisters (Keren, Minami et al. 2011) are upregulated in the mutant (Rv2021c/Rv2022c, Rv1989c/Rv1990c, Rv2865/Rv2866, Rv1955/Rv1956, Rv2034/Rv2035 and Rv2021c/Rv2022c). Both the toxin and antitoxin of Rv1955/Rv1956, Rv2034/Rv2035, and Rv2021c/Rv2022c are also upregulated during adaptation to starvation (Betts, Lukey et al. 2002) and the EHR (hypoxia) (Rustad, Harrell et al. 2008). As well, Rv1989c/Rv1990 and Rv2865/Rv2866 were also upregulated in the EHR. Therefore, these 6 TA systems appear to play a prominent role in the dormancy program and are negatively regulated by Lsr2. The majority of the TA system toxins are mRNA interferases, resulting in translation inhibition (Gerdes, Christensen et al. 2005). Removing the inhibition imposed by Lsr2 on the TA systems would allow M. tb to quickly erase its transcriptional profile when faced with adverse
conditions, allowing for a rapid metabolic shift and dormancy. The role of these TA systems in M. tb persistence in vivo has not been explored and remains to be determined. I also found that 16 ribosomal proteins are downregulated ≥2-fold in Δlsr2, signaling a lower rate of translation in the mutant and is consistent with the mutant being primed for NRP. The resuscitation-promoting factors (Rpfs) occur in a subset of actinomycetes and play an important role in reactivation of dormant M. tb in vitro(Mukamolova, Kaprelyants et al. 1998; Mukamolova, Turapov et al. 2002; Shleeva, Bagramyan et al. 2002), as well as in vivo(Downing, Mischenko et al. 2005; Kana, Gordhan et al. 2008; Mukamolova, Turapov et al. 2010). I found all five of the M. tb rpf genes are downregulated in theΔlsr2 with rpfA and rpfB expression levels ≤2-fold that of WT. In one study, deletion of rpfB alone led to a reactivation defect in chronic mouse infection (Tufariello, Mi et al. 2006), whereas a subsequent study(Russell-Goldman, Xu et al. 2008) found that a double knockout of rpfA and rpfB was dramatically attenuated for reactivation. Moreover, the finding that up to 99% of the bacteria found in human sputum are dormant and can only be cultured with the addition of exogenously added Rpf (Mukamolova, Turapov et al. 2010) indicates that in vivo the absence of these factors is critical for maintaining dormancy. Although the precise mechanism by which Rpfs stimulate growth remains unknown, a crystal structure of the domain conserved among the Rpf proteins revealed a c-type lysozyme-like fold, an enzymatic activity predicted to cleave the glycosidic bond between N-acetyl glucosamine and N-acetylmuramic acid in cell wall peptidoglycan (Cohen-Gonsaud, Barthe et al. 2005). Site-directed mutagenesis studies revealed that Rpf enzyme activity is indispensible for the resuscitation and growth-stimulatory effects of these proteins indicating that the remodeling of peptidoglycan is integral for reactivation (Mukamolova, Murzin et al. 2006). Since peptidoglycan is essential to maintain the integrity of the cell wall, disruptions in its structure alter sensitivity to physical agents. Indeed, strains deleted for several rpf genes in M. tb show an increased susceptibility to SDS(a known cell wall–damaging agent) (Kana, Gordhan et al. 2008), a phenotype I would expect to observe for the M. tb Δlsr2 mutant. Cell wall peptidoglycan hydrolysis has been shown to be critical for spore germination, a process that may conceptually resemble resuscitation (Chao and Rubin 2010). For instance, Bacillus anthracis spores deficient for several peptidoglycan hydrolases display significant germination defects(Giebel, Carr et al. 2009; Heffron, Lambert et al. 2010). Deletions of various peptidoglycan hydrolases, including an rpf gene, in the actinomycete Streptomyces coelicolor also resulted in defective or delayed
germination (Haiser, Yousef et al. 2009). Alternatively, growing bacteria release peptidoglycan fragments of the cell wall into the extracellular milieu, which has been reported to be sufficient for signaling \textit{B. subtilis} spore resuscitation (Shah, Laaberki et al. 2008). This mechanism is dependent on the eukaryotic-like serine/threonine kinase PrkC, which possesses three extracellular PASTA domains believed to bind peptidoglycan fragments (Jones and Dyson 2006). \textit{M. tb} possesses a PrkC homolog PknB, shown to be involved in cell shape (Kang, Abbott et al. 2005), although I did not observe differential expression of this gene in the mutant. Taken together peptidoglycan degradation and signaling appears to be important for resuscitation from spores in \textit{Bacillus} species. This resuscitation paradigm may hold true for \textit{M. tb} reactivating from dormancy. Indeed, stationary phase \textit{M. tb} peptidoglycan is cross-linked differently from that isolated from actively growing cells (Lavollay, Arthur et al. 2008), presumably to generate a more stable morphology for survival under stress conditions, accompanied by the adoption of a coccoid shape due to peptidoglycan remodeling (Shleeva, Bagramyan et al. 2002).

The LSR2\_UP was highly enriched for transcriptional regulators. Transcriptional regulators such as sigma (\(\sigma\)) factors play a role in the adaptation of bacteria to environmental stresses. The \(\sigma\) factors are components of RNA polymerases that bind to the enzyme’s core subunits and direct promoter binding specificity in a context-dependent manner. The presence of 13 \(\sigma\) factors in \textit{M. tb} reflects its ability to adapt to various stress conditions likely encountered during host infection. I found that in the absence of Lsr2, the expression of the alternative sigma factors \textit{sigB}, \textit{sigE} and \textit{sigF} are significantly upregulated, while \textit{sigC} is significantly downregulated. Transcription of \textit{sigB} has been shown to be upregulated during stress conditions including stationary phase, nutrient starvation and hypoxia (Hu and Coates 1999; Manganelli, Dubnau et al. 1999; Betts, Lukey et al. 2002), and an \textit{M. tb sigB} mutant was defective for \textit{in vitro} hypoxia survival (Fontan, Voskuil et al. 2009). \textit{SigE} is an extracytoplasmic function sigma factor (a family of sigma factors which responds to extracellular stimuli) (Missiakas and Raina 1998) that positively regulates \textit{sigB} expression (Manganelli, Voskuil et al. 2001). \textit{SigE} expression is activated during stress conditions such as heat shock, SDS treatment, starvation and hypoxia (Manganelli, Dubnau et al. 1999; Betts, Lukey et al. 2002; Rustad, Harrell et al. 2008). The \textit{sigF} gene occurs only in pathogenic mycobacteria and bears homology to a sporulation-specific sigma factor in \textit{Streptomyces coelicolor} and \textit{Bacillus subtilis} (DeMaio, Zhang et al. 1997). It is upregulated during stationary phase and starvation (Manganelli, Dubnau et al. 1999;
Betts, Lukey et al. 2002), while there are conflicting reports as to whether or not it is induced during hypoxia (Michele, Ko et al. 1999; Chen, Ruiz et al. 2000). It is interesting that sigC is downregulated in the Δlsr2 mutant given SigF positively regulates its expression (Geiman, Kaushal et al. 2004; Lee, Karakousis et al. 2008); however, it has been reported to be downregulated during stationary phase and in response to heat shock and SDS-induced surface stress (Manganelli, Dubnau et al. 1999) leaving it unclear what the role of SigC is. Nevertheless, given that 3 of the 13 sigma factors are significantly upregulated in my Δlsr2 mutant, this would suggest that Lsr2 plays an important role repressing the expression of genes involved in adaptation to stress conditions.

In this study, I successfully generated an M. tb Δlsr2 mutant using a phage-transduction system (Bardarov, Bardarov Jr et al. 2002) indicating that contrary to previous claims (Park, Dahl et al. 2008; Colangeli, Haq et al. 2009), lsr2 is not essential in M. tb. Initially, the mutant produced ‘mushroom-like’ colonies in contrast to the flat, spread out WT colonies. This phenotype was lost after re-streaking or passaging in liquid media, suggesting that a compensatory mutation likely occurred to counter the negative fitness cost associated with the over-production of a cell surface factor. Even though the ‘mushroom’ colony morphology was unstable, a ‘blebbing’ colony morphology was observed for the mutant after prolonged growth as well as a hyper-aggregative phenotype when grown in liquid media. Subtle defects in cell wall components can alter the surface properties of individual cells, affecting cell-to-cell interactions, ultimately giving rise to colonies with different morphology (Alexander, Jones et al. 2004). I compared the apolar and polar lipid as well as the mycolic acid profiles of the Δlsr2 mutant and WT parental strains but was unable to identify a difference to account for the mutant phenotypes. This leads me to speculate that perhaps a protein or carbohydrate component of the cell wall is responsible for the colony morphology change and hyper-aggregation of the mutant. For example, I observed a downregulation of the rpf peptidoglycan hydrolases in the mutant. Both the rpfA and rpfB deletion mutants display altered colony morphology and the rpfA mutant is also hyper-aggregative in liquid media (Downing, Betts et al. 2004; Russell-Goldman, Xu et al. 2008). In addition, Lsr2’s supposed role in regulating PE/PPE proteins may explain the colony morphology change since they are surface exposed proteins that have been shown to cause changes in cell structure and colony morphology (Brennan, Delogu et al. 2001; Sampson, Lukey et al. 2001; Banu, Honore et al. 2002). The PE and PPE proteins form heterodimers and the
hyper-aggregation could be due to intercellular interactions between cognate PE and PPE proteins upregulated in the mutant.

The Δlsr2 mutant exhibits a growth delay phenotype on agar plates and in liquid media. A fitness decrease associated with hns mutations is partly due to derepression of foreign DNA; thus, given my finding that Lsr2 negatively regulates xenogeneic islands in M. tb the unmitigated expression of these genes is a likely explanation for the growth defect. Interestingly, the M. smegmatisΔlsr2 strain displays an equivalent growth rate as the WT strain (Chen, German et al. 2006). M. tb is purported to have up to 88 TA systems while M. smegmatis only harbours 3 such systems. Activation of TA systems has been shown to induce bacteriostasis. Thus, the upregulation of several TA systems in the mutant could very well explain its retarded growth rate. Furthermore, a quintuplet rpf deletion mutant also exhibited a growth delay phenotype (Kana, Gordhan et al. 2008). It is also possible that the mutant’s retarded growth rate may be due to the fact that it has realigned its metabolism to utilize lipids, whereas the in vitro growth media used in these studies contained glycerol as a carbon source.

Overall, the findings presented in this chapter extend our understanding of Lsr2. I have shown that like H-NS, Lsr2 is global transcriptional repressor which targets horizontally acquired genetic islands and virulence factor genes. In addition, the widespread genetic changes within the M. tbΔlsr2 suggest it is pre-adapted for persistence within the host. Therefore, inactivation of lsr2 results in the induction of genes necessary for persistence and alteration of the overall metabolic scheme for growth using lipids under oxygen-limiting conditions. Future studies are warranted to evaluate the mutant’s redox state and carbon source preference. As well, investigating the ability of the M. tbΔlsr2to persist in in vitro and in vivo models of persistence will strengthen the link between Lsr2 and latency.
3.5 Experimental procedures

Bacterial strains, media, and growth conditions. To generate an \( M. \text{tb} \Delta lsr2 \) mutant strain the \( lsr2 \) gene in \( M. \text{tb} \) H37Rv was replaced with a hygromycin-resistance cassette using TM4 phage-mediated specialized transduction (Bardarov, Bardarov Jr et al. 2002). The primer pairs used to PCR amplify the left and right fragments to generate the allelic exchange substrate were (5’-CGGCTTCCATATAATTGGGCAGCTGGATACCTGCTGGCGCAC-3’ and 5’-CGGCTTCCATTCTTTGGCATTGGCTACCGGCCGCCCAGCGCA-3’) and (5’-CGGCTTCCATAGATTGGTGGCTTACCCTGGCTTTCTCTCTCTTGTG-3’ and 5’-CGGCTTCCATCTTCTTTGGGGTGAAGAGATCACACCGCAGACGAC-3’). The resulting PCR products were digested with the restriction enzyme \( PfIMI \) and ligated with the 1600 bp and 1760 bp fragments of the p0004 plasmid pretreated with the same enzyme to generate pKO-lsr2. pKO-lsr2 and phLR plasmid DNA were digested with \( PacI \), ligated and packaged using the MaxPlax Lambda packaging extract (Epicentre), followed by transduction into \( E. \text{coli} \) NM759. The resulting phLR/pKO-lsr2 plasmid DNA from the transductants was electroporated into \( M. \text{smegmatis} \) mc\(^2\)155 and plated for mycobacteriophage plaques at the permissive temperature of 30 °C. A high-titre phage lysate, prepared from one temperature-sensitive phage plaque, was used to infect H37Rv at the non-permissive temperature of 37°C as described previously (Bardarov, Bardarov Jr et al. 2002). Hygromycin-resistant colonies were selected at 6-weeks post-transduction and confirmed by Southern hybridization. Cultures were grown at 37°C with shaking in Middlebrook 7H9 medium supplemented with 10% OADC, 0.05% Tween80. Hygromycin was added (75 μg/ml) for cultures of \( M. \text{tb} \Delta lsr2 \).

ChIP-on-Chip Experiments. Cultures (50 mL) of \( M. \text{tb} \) H37 Rv/pLSR2-HA or \( M. \text{smegmatis} \) MS8444/pLSR2-HA grown to \( \text{OD}_{600} \) 0.6–0.8 were treated with 1% formaldehyde and incubated for 15 min at room temperature. The cross-linking reaction was quenched with 1.25mM glycine for 10min. Cells were washed twice with ice-cold PBS and sonicated to generate DNA fragments of ~500–700 bp. Cell lysates were precipitated with an anti-HA antibody (Sigma H9658) using agarose protein G beads (Calbiochem). Input and ChIP DNA were amplified and labeled with monofunctional reactive Cy3 or Cy5 dyes on the basis of the T7-based protocol (Liu, Schreiber et al. 2003). Subsequently, 6.5 μg each of labeled ChIP cRNA and input cRNA were hybridized to a 244,000 \( M. \text{tb} \) H37Rv or \( M. \text{smegmatis} \) mc\(^2\)155 whole-genome tiling array (Agilent
Technologies) and scanned using the Genepix Professional 4200A scanner. Feature intensity ratios were acquired using Imagene v7.5 (Biodiscovery) and lowess-normalized using the marray R software package from Bioconductor. Lsr2-binding peaks were determined using a custom perl script with a peak defined as five or more neighboring features having intensity ratios ≥1.5 and extended with adjacent features with ratios ≥1.15. The result presented is representative of at least two independent experiments (independent bacterial cultures and experiments were performed at a different time).

**Isolation of chromosomal DNA, restriction endonuclease cleavage and Southern blot analysis.** *M. tb* colonies of interest were grown to OD$_{600}$ ~1.0. Cultures were then centrifuged, resuspended in 1 mL ice-cold 1×PBS and transferred to a 2 mL screw cap tube containing 0.5 g of 0.1-mm zirconia/silica beads (BioSpec). Cells were lysed by bead beating (Mini-Beadbeater; BioSpec Products) at the high speed setting for 2 × 1 min pulses. Cell debris was removed by centrifugation at 13 000 rpm for 30 minutes. The supernatant was then transferred to a 1.5-mL Eppendorf tube and boiled for 5 min. The crude DNA preps(10µg) were then digested overnight in a thermocycler at 37°C with the restriction enzymes *BbsI* and *PvuII*. The digested DNA was separated on a 1.0% agarose gel and blotted onto Hybond N+ nylon membranes (Amersham Biosciences). The *lsc* of *M. tb* was PCR amplified using the forward primer 5’-AAAAGAATTCATGGCGAAAG TAACCGTCACC-3’ along with the reverse primer 5’-TTTTCTCGAGTCAGGTCGCCGTTGTATGC-3’. The AlkPhos Direct Labeling and Detection System with CDP-Star was used to label 100 ng of the *lsc* PCR product and used to detect the target fragment.

**TLC analysis of cell wall lipids.** All thin-layer chromatography (TLC) analyses were performed on Silica Gel 60 plates (Whatman). The apolar and polar lipids were prepared from *M. tb* cells (50 mg dry biomass) according to previously published procedures (Alexander, Jones et al. 2004). These lipids were analyzed by two-dimensional TLC (2D-TLC) using the following solvent systems. Apolar lipids were developed with petroleum ether-ethyl acetate (98:2; three times) in the first dimension and petroleum ether-acetone (98:2) in the second dimension. Polar lipids were separated with chloroform-methanol-water (60:30:6) in the first dimension and
chloroform-acetic acid-methanol-water (40:25:3:6) in the second dimension. Lipids were detected by charring with α-naphthol or 5% phosphomolybdic acid.

**RNA extraction.** Cultures (50 ml) of *M. tb* H37Rv WT and *M. tb* Δlsr2 were harvested at an OD$_{600}$ ~0.4. Cells were pelleted and transferred to 2-ml screw cap tubes containing 1 ml RNA protect Bacterial Reagent (Qiagen) and incubated for 5 min at room temperature. Cells were again pelleted and resuspended in 400 µl lysis buffer (20 mM NaCH$_3$COOH, 0.5% SDS, 1mM EDTA, pH 4) and 1 ml phenol/chloroform (pH 4.5, Sigma). Cells were disrupted by bead beating with glass beads by three 30-sec pulses. They were then incubated at 65°C for 4 min and then at 4°C for 5 min before being centrifuged at 13,000 rpm for 5 min. The supernatant was then extracted with 300 µl of chloroform/isoamyl alcohol (24:1) and precipitated with isopropanol. Precipitated nucleic acids were collected by centrifugation and the pellets were washed with 70% ethanol and air dried. Crude RNA samples were treated with DNaseI (Fermentas) for 2 hours at 37°C and purified further using an RNeasy kit (Qiagen) according to the manufacturer’s instructions. The quality of purified total RNA was assessed by gel electrophoresis.

**Microarray analysis.** For cDNA production 25µg total RNA was reverse transcribed at 42°C overnight using 2 µl Superscript II reverse transcriptase (Invitrogen), 25µg 9-mer random primers and 2 µl dNTP mix (0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.25 mM dTTP, 0.25 mM 5-(3-aminoalyl)-dUTP) in a total volume of 100 µl (25 mM Tris pH 8.4, 37.5 mM KCl, 3 mM MgCl$_2$, and 0.1 M DTT). RNA hydrolysis was performed by adding 15µL 1M NaOH and then neutralized with 15µL 1M HCl after incubating for 20 min at 65°C. The cDNA was purified using a QIAquick column (Qiagen). The microarray protocol has been described previously (Zhang, Morris et al. 2004). Briefly, samples were labeled for 1 hr at RT and then quenched with 4 M hydroxylamine. The labeled cDNA was purified and 1 µg per sample was hybridized to a 15 000 feature *M. tb* H37Rv ORF array with three distinct probes per ORF (Agilent Technologies) and scanned using the Genepix Professional 4200A scanner. Feature intensity ratios were acquired using Imagene v7.5 (Biodiscovery) and lowess-normalized using the marray R software package from Bioconductor. Significance Analysis of Microarrays (SAM) (Tusher, Tibshirani et al. 2001) was performed to identify genes that are significantly upregulated or downregulated.
CHAPTER 4: INVESTIGATING THE IMPACT OF LSR2 DYSFUNCTION ON IN VITRO PERSISTENCE, MACROPHAGE INFECTION, AND IN AN ANIMAL MODEL OF INFECTION
4.1 Abstract

A non-replicating persistent *Mycobacterium tuberculosis* (*M. tb*) subpopulation is thought to be present during human tuberculosis (TB) infection. This subpopulation predominates during latent infection and is thought to be responsible for the lengthy course of drug therapy required to treat TB. *In vitro* models have shown that *M. tb* can exist in a non-growing, drug-resistant state that is thought to mimic *in vivo* persistence. Previously, I identified Lsr2 as a pleiotropic transcriptional regulator in *M. tb* that regulates genes found to be induced during *in vitro* persistence; hence, Lsr2 might play an important role in persister formation and TB disease progression. In this study I show that the Δ*lsr2* mutant of *Mycobacterium smegmatis* exhibits a survival advantage during hypoxia-induced persistence. During long-term hypoxia I observed what appeared to be *lsr2* mutant colonies among the WT colonies recovered. I showed that the *lsr2* gene in most of these altered colonies had been inactivated due to disruption of its promoter by the IS1096 insertion sequence. Aerobic passaging in liquid media of some of the *lsr2* mutants gave rise to revertant WT colonies with sequencing analysis confirming the full excision of IS1096. I also evaluated whether the loss of *lsr2* affected *M. tb* pathogenesis. Infection of macrophages indicated that the mutant induces more rapid macrophage death. The *lsr2* mutant strain displays a compromised replication capacity and is maintained at a lower bacterial burden compared to WT during chronic infection in mice. Taken together, these results indicate that the downregulation of *lsr2* seems to confer increased mycobacterial viability during *in vitro* hypoxia-induced persistence, and may allow *M. tb* to persist at a lower bacterial burden for longer periods of time *in vivo*.
4.2 Introduction

The vast majority of individuals infected with *M. tb* do not progress to active disease. However, the bacillus can remain in the host for decades in a dormant non-replicating persistent state, with the potential for future revival and initiation of clinical disease (Wayne 1994). During latent infection the bacilli are thought to reside within a mass of macrophages and other immune cells which is collectively referred to as the granuloma. Granulomas are thought to limit *M. tb* growth by restricting access to oxygen and nutrients (Boshoff and Barry 2005). Resection of blocked lesions from human lung tissue (e.g., no access to bronchial airways) revealed the presence of viable tubercle bacilli (Salkin and Wayne 1956), which lead to the development of the *in vitro* Wayne model for studying hypoxia-induced mycobacterial persistence (Wayne 1976; Wayne 1977; Wayne 1994; Wayne and Hayes 1996). Mycobacteria are obligate aerobes and the abrupt shift of an aerobic *M. tb* culture to anaerobic conditions causes rapid death of the culture (Wayne and Diaz 1967). In contrast, the gradual depletion of oxygen allows *M. tb* to adapt and survive anaerobiosis for an extended period of time by entering into non-replicating persistence (NRP), but upon aeration can resume synchronous growth (Wayne 1977; Wayne and Hayes 1996). There is further evidence suggesting that persisting *M. tb* in lung lesions are nutritionally starved. These recovered bacilli differ with respect to their cellular morphology and staining properties when compared to actively growing bacilli *in vitro* (Salkin and Wayne 1956; Nyka 1974). Mycobacterial cultures, including *M. tb*, starved *in vitro* for two years developed these same properties. Upon addition to nutrient broth they regained their acid fastness and resumed growth (Nyka 1974).

Lsr2 is highly conserved amongst Mycobacteria (>85% identity in amino acid sequences) suggesting its core function(s) is likely to be conserved (Chen, German et al. 2006). My previous work established that Lsr2 is functionally analogous to H-NS (Gordon, Imperial et al. 2008; Gordon, Li et al. 2010; Gordon, Li et al. 2011), an important nucleoid-associated protein that serves as a central regulator of virulence gene expression in the enterobacteriaceae (Dorman 2004; Dorman 2007). Furthermore, I found that Lsr2 regulates the expression of many genes involved in both establishing and exiting from NRP. To gain insight into the role of Lsr2 during bacteriostasis I compared the viability of the WT and the *lsr2::Tn (Δlsr2)* mutant of *M. smegmatis* (Chen, German et al. 2006) under two well characterized *in vitro* persistence models for mycobacteria: nutrient starvation in PBS and hypoxia using sealed tubes with no air
headspace. My rationale for choosing to study *M. smegmatis* is its rapid growth rate, it is non-pathogenic and it is far less clumpy than *M. tb*. Also it has been demonstrated that *M. smegmatis* exhibits highly similar physiological characteristics as *M. tb* under hypoxia, which suggests that the dormancy response to oxygen depletion is not a unique characteristic of the pathogenic mycobacteria (Dick, Lee et al. 1998). Similarly nutrient starved cultures of *M. smegmatis* enter dormancy and display the same altered morphology and staining properties observed in patient lesions (Nyka 1974; Smeulders, Keer et al. 1999).

Mycobacterial virulence is assessed by the ability to cause progressive infection and death in animal models (Dunn and North 1995). Macrophages are the primary target cells invaded by *M. tb* and provide an essential niche to establish infection in the host (Russell, VanderVen et al. 2010). Various studies have established that necrosis of host macrophages is a strategy used by virulent *M. tb* to avoid innate host defenses, while attenuated strains cause macrophage apoptosis (Balcewicz-Sablinska, Keane et al. 1998; Keane, Remold et al. 2000; Divangahi, Chen et al. 2009). In this study, I sought to assess the biological role of Lsr2 in the context of NRP and during *M. tb* infection. In addition to the experiments mentioned above, I also examined if the loss of *lsr2* in *M. tb* affects intracellular survival or virulence in a mouse model of infection.
4.3 Results

4.3.1 The loss of *lsr2* in *M. smegmatis* confers a survival advantage during hypoxia-induced NRP

In the PBS starvation model, cultures of WT, *lsr2::Tn* (∆*lsr2*) and the complemented strains were grown to midlog phase when they were resuspended in 1×PBS before being aliquoted to 2 mL screwtop vials with ample headspace air to facilitate atmospheric oxygen diffusion. For all three strains I observed a steady decrease in CFU that reached a low point approximately 1 log below the starting CFU levels followed by a steady increase in CFU back to the initial levels (Figure 4.1A). Fluctuations in CFU levels have been reported previously during long-term starvation experiments with *M. smegmatis* (Smeulders, Keer et al. 1999) and no loss of viability was observed in *M. tb* during long-term PBS starvation (Betts, Lukey et al. 2002). Taken together, the *M. smegmatis ∆lsr2* mutant responds similarly to nutrient limitation as the WT strain.

The hypoxia dormancy model was performed in the same manner as the nutrient starvation model except that midlog phase cultures in 7H9 broth liquid media were aliquoted directly to tubes with no air headspace. In contrast to my findings in the nutrient starvation model, for the hypoxia dormancy model I observed enhanced viability of the ∆*lsr2* mutant over the WT (Figure 4.1B). For the first half of the experiment, the CFU levels of all strains steadily decreased. However, while the WT CFU continued to decrease, the ∆*lsr2* levels began increasing and by day 156 had increased by over 1 log the original level and 3 logs greater than the WT. This finding suggests that the *M. smegmatis ∆lsr2* mutant has enhanced viability during long-term hypoxia over the WT.

In both the nutrient depletion and the hypoxia model, methylene blue, a blue dye that decolourizes when it is reduced, was used as a visual indication of oxygen depletion. The 7H9 hypoxia cultures showed complete dye decolourization and, hence oxygen depletion after 10 days. In contrast, the methylene blue dye in the nutrient starved cultures did not decolourize but remained the same colour as the control solution containing no bacteria. This would suggest that the starved cells significantly decreased their respiration rate while remaining viable in the presence of oxygen. These findings are consistent with observations of nutrient starved *M. tb*
cultures, which rapidly decrease their respiration rate over the first 96 hours of starvation (Loebel, Shorr et al. 1933; Betts, Lukey et al. 2002).
Figure 4.1. The \( \Delta lsr2 \) of \( M. \ smegmatis \) exhibits a fitness advantage over the WT strain under long-term hypoxia. The WT, \( \Delta lsr2 \) and complemented strains were each incubated under (A) nutrient starvation and (B) hypoxia using the Wayne model. The viable bacteria counts at different time points were determined by plating and numerating CFU for 3 independent tubes per strain. Each point represents the mean CFU ± SD. For those points without error bars, the SD was too small to be visualized.
4.3.2 *M. smegmatis* inactivates *lsr2* via a transposon-mediated mechanism during hypoxia-induced NRP

Strikingly, when plating the WT culture for CFU determination from the hypoxic cultures, I observed colonies that resembled the Δ*lsr2* mutant, exhibiting the same smooth colony morphology (Figure 4.2A) (Chen, German et al. 2006). These colonies began to appear with a high frequency, 52% at day 77, and reached 63% at day 156. To rule out contamination with the *lsr2*:Tn mutant which is kanamycin-resistant, I streaked several of the supposed mutant colonies on 7H11 agar with kanamycin and found them all to be kanamycin-sensitive. I repeated the experiment with different batches of starting cultures which generated similar results. To test whether these colony morphology variants are *lsr2* mutants, I randomly picked 29 morphological variants from different time points from two independent experiments and performed Southern blot and PCR sequencing analyses. For several of the colonies I was unable to obtain a PCR product for the *lsr2* gene’s promoter even though the product was obtained when WT mc²155 template DNA was used. I hypothesized that the inability to PCR amplify the *lsr2* promoter region was likely due to the presence of a transposable element, specifically IS1096 which is highly prevalent in the *M. smegmatis* genome (e.g., The *M. smegmatis* mc²155 genome contains 24 copies of IS1096 while its parental strain ATCC 607 has only 11 copies) (Dame, Wyman et al. 2002). I demonstrated that all 29 colonies were *lsr2* mutants: 25 have IS1096 inserted in the promoter region of the *lsr2* gene in either direction (Figure 4.2B and Figure 4.2C) and five have different point mutations within the ORF of *lsr2* (Table 4.1). I was able to PCR amplify a hybrid PCR product using one of two IS1096 forward primers (one for each possible orientation of the transposon) and a reverse primer near the beginning of the *lsr2* ORF demonstrating that the transposon was in fact IS1096. I sequenced the hybrid PCR fragments which confirmed the identity of IS1096 and the insertion location (Table 4.1). In addition, I observed an even higher conversion frequency to ∆*lsr2*-like colonies for the complemented strain (Δ*lsr2* + pLSR2).

Taken together, the *lsr2*-inactivation phenomenon suggests that the loss of functional *lsr2* may confer a fitness advantage under hypoxia in *M. smegmatis*. To address this hypothesis I performed competition experiments in which the WT strain carrying a hygromycin resistance marker was mixed at equal numbers with the Δ*lsr2*, which harbours a kanamycin resistance marker and incubated under hypoxia. CFUs of the WT and Δ*lsr2* were determined by counting the hygromycin or kanamycin-resistant colonies, respectively. As expected, I found
that Δlsr2 outgrew WT after 80 days under hypoxia (Figure 4.2D). Interestingly, I did not observe the spontaneous conversion of WT to lsr2 mutants in the competition assay, presumably because Δlsr2 in the mixture outcompeted the WT for nutrients, reducing the replication of WT and preventing the generation of mutations.

To determine if the lsr2 mutants isolated during hypoxia could revert back to WT, I sequentially passaged some of the mutants in 7H9 broth, each time plating serial dilutions of the culture to visually inspect the colony morphology for reversion. For colony 28 (Table 4.1) which has an lsr2 point mutation I observed WT-like colonies after the second passage and their frequency increased with each additional passage. After 5 passages, the majority of the colonies were WT-like (Figure 4.3A), suggesting that the lsr2 point mutation reverted back to WT, which was confirmed for some colonies by DNA sequencing. Moreover, I observed reversion of mutants with the IS1096 insertion in the lsr2 promoter, but not to the same degree as the point mutant. These revertants were confirmed by southern blot and sequencing analysis. I also performed aerobic competition experiments with the Δlsr2 and WT strains carrying different antibiotic resistance makers. In three independent experiments, the WT outgrew the lsr2::Tn after several passages under aerobic growth (Figure 4.3B). Taken together, my data indicate that in M. smegmatis, Lsr2 is advantageous for oxygenic growth and that the loss of lsr2 function confers a fitness advantage under hypoxia.
Figure 4.2. The loss of lsr2 confers a fitness advantage in *M. smegmatis* during hypoxic NRP. (A) Examples of spontaneous lsr2 mutants that appeared in WT cultures under long-term hypoxia. These colonies exhibit a smooth morphology in contrast to the rough, rugose WT colonies. (B & C) Randomly picked Δlsr2-like colonies were subjected to southern blot and PCR sequencing analysis and results showed that most colonies have IS1096 inserted at 70bp 5’ of lsr2 ORF (C). The IS1096 was inserted in either orientation, resulting in different restriction fragment sizes (B). (D) The hybrid PCR strategy used to identify IS1096 including if it inserted in the reverse (Scenario 2) or the same (Scenario 3) orientation relative to lsr2. (E) The Δlsr2 outcompeted the WT in two independent hypoxia growth competition experiments. The numbers of Δlsr2 and WT CFU at each time point were determined by plating out three independent tubes on corresponding antibiotic-containing plates for both strains. Each point on the graph is the ratio of the mean Δlsr2CFU to the mean of WT CFU ± SD.
Table 4.1. Genetic analysis of spontaneous *Mycobacterium smegmatis lsr2* mutants selected from long-term hypoxia survival experiments

<table>
<thead>
<tr>
<th>Colony</th>
<th>Set*</th>
<th>Time Point (days)</th>
<th>Genetic Lesion**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>59</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>59</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>92</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>145</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>145</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>145</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>145</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>190</td>
<td>C insertion at position 56 causing premature stop codon</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>190</td>
<td>G238C transversion causing E80Q amino acid substitution</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>190</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>190</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>190</td>
<td>G218T transversion causing G73V amino acid substitution</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>77</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>77</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>77</td>
<td>IS1096 insertion in forward orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>77</td>
<td>C111T transition causing L37F amino acid substitution</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>77</td>
<td>IS1096 insertion in reverse orientation 70bp 5’ of <em>lsr2</em> ORF</td>
</tr>
</tbody>
</table>

*Set refers to which of two independent experiments a particular mutant was isolated.

**IS1096 insertion direction is relative to *lsr2*. For those clones that did not have an IS1096 insertion, the numbering corresponds to the affected nucleotide in the 345 bp *lsr2* ORF.
Figure 4.3. (Previous page) The WT is more fit than the Δlsr2 under aerobic conditions. (A) One of the spontaneous lsr2 mutants which contains a point mutation causing a L37F substitution underwent reversion to WT when it was passaged in 7H9 broth under aerobic conditions several times and the culture was plated on 7H11 agar plates after each passage. Wild type colonies began to appear after two passages and its number increased with each passage. After 5 passages, the majority of the colonies displayed wild type-like morphology. (B) Three independent colonies of both the Δlsr2 and WT strains carrying different antibiotic resistance makers were mixed at equal number and grown aerobically in 7H9 broth until stationary phase. A fraction of the culture was passaged in fresh media and the growth under aerobic conditions was repeated. The numbers of Δlsr2 and WT CFU after each passage were determined by plating on corresponding antibiotic-containing plates. Each point on the graph is the ratio of the mean WT CFU to the mean of Δlsr2 CFU ± SD.
4.3.3 The *M. tb* ∆lsr2 mutant induces rapid killing of RAW macrophages in vitro

Macrophages are the primary target cells for intracellular invasion by *M. tb* and provide an essential niche to establish infection in the host (Russell, VanderVen et al. 2010). I performed macrophage infection assays to probe whether loss of *lsr2* affects *M. tb* intracellular survival and replication. In these experiments, RAW 264.7 macrophage monolayers were infected with WT, ∆lsr2 or ∆lsr2 + pLSR2 at a multiplicity of infection (MOI) of ~1. My preliminary macrophage infection experiments showed that during a 6 day infection, the WT and complemented strains grew steadily whereas the ∆lsr2 mutant showed only slight increases in intracellular CFU by day 6 (Figure 4.4A). Previously in Chapter 3 I showed that the ∆lsr2 mutant has slightly slower growth than WT in 7H9 broth, but this difference cannot account for the apparent lack of growth of ∆lsr2 inside macrophages. These initial macrophage experiments were hampered by the extreme clumpiness of the ∆lsr2 which resulted in a substantially lower MOI than the other strains. In a subsequent experiment in which the MOI of the ∆lsr2 was increased to similar levels as the WT and complemented strains, I found that the CFU of ∆lsr2 recovered from infected macrophages markedly decreased over time, while the WT and complemented strains showed significant intracellular growth during the same period (Figure 4.4B). An even greater decline in the ∆lsr2 CFU was observed when a higher MOI was used for infection (Figure 4.4B).

By the 96 hour time point the cell culture medium of the macrophages infected with the WT or complemented strains had turned yellow, while the culture medium in the wells containing the ∆lsr2-infected cells remained fuchsia (Figure 4.4C). Cell culture medium contains phenol red which serves as a pH indicator. A colour change to yellow often indicates that cells have grown confluent and produced an abundance of acidic waste products. This stark difference in colour of the culture medium suggested that the WT and complemented strain-infected macrophages significantly outnumbered the mutant infected macrophages. This was confirmed by microscopic observation which revealed that macrophages infected by ∆lsr2 were mostly killed, as evidenced by the detachment of RAW cells from wells. Thus, the loss of dead macrophages during the washing step explains the poor recovery of ∆lsr2 in this assay.
Figure 4.4. Intracellular growth of the ∆lsr2 mutant in macrophages. (A) RAW 264.7 macrophage monolayers in 24-well plate (5×10⁴ cells/well) were infected for 3 hours with WT (2.2×10³ CFU/well), ∆lsr2 (4.6×10² CFU/well), complemented strain (2.1×10³ CFU/well). Intracellular bacterial numbers were determined over 6 days. (B) RAW macrophages were infected with a higher MOI. RAW cells (5×10⁴ cells/well) were infected with WT (1.0×10⁵ CFU/well), ∆lsr2 (3.2×10⁵ CFU/well), ∆lsr2 (3.3×10⁶ CFU/well) (high MOI), and the complemented strain (1.8×10⁵ CFU/well). Intracellular bacterial numbers were determined at 48 hour time points over 6 days. (C) The WT and complement-infected macrophage monolayers were fully confluent by 96 hours resulting in the acidification of the culture medium as indicated by the yellow colour. The ∆lsr2-infected macrophages were mostly dead and the medium maintained a pink-red colour indicating a near neutral pH.
4.3.4 The *M. tb* ∆*lsr2* mutant persists at a lower level than the WT in a murine model of infection

To determine the role of Lsr2 in virulence, I aerosol-infected BALB/c mice with 30 CFU of WT, ∆*lsr2* or ∆*lsr2* + pLSR2 and monitored the growth of bacteria for 8 weeks. Similar to previous reports, the WT strain grew exponentially in the lungs for the first 4 weeks, and reached a plateau after the immune system was activated (Figure 4.5A). During the persistent stage of infection, *M. tb* reduces its growth and metabolic activity, doubling only once every 100 hr (Gill, Harik et al. 2009). An equilibrium between the growth of bacteria and killing by the immune system is reached resulting in a constant bacterial burden (Gill, Harik et al. 2009). Interestingly, although the ∆*lsr2* replicated at a slower rate and plateaued at a lower bacterial burden during the chronic phase of infection (about 1 log lower than that of WT), the mutant appears to be fully capable of persisting in animals (Figure 4.5A). This type of mutant in murine infection has been termed ‘giv’ mutant (‘growth in vivo’) (Glickman and Jacobs 2001). Dissemination to the spleen and liver also occurred for the mutant, albeit at a lower level or at a later time point (Figure 4.5B and C). The slow replication of the ∆*lsr2* at the acute phase of infection may be due to its intrinsic slow growth rate observed in vitro. Alternatively, the rapid death of macrophages induced by ∆*lsr2* may promote activation of the immune response, resulting in a more effective control of the bacterial burden.

Interestingly, at all timepoints starting from 2 weeks onward I observed the presence of colony morphology variants for all 3 strains recovered from animal tissues which resembled the unstable ‘mushroom’ colony phenotype of the ∆*lsr2* mutant (Figure 4.5D). These colonies were most often recovered from the liver but were sometimes seen in lung and spleen homogenates. I performed sequencing and southern blot analysis for several of the colony morphology variants to determine if they were *lsr2* mutants. I did not identify any mutations within the *lsr2* ORF or insertion events that would perturb *lsr2* expression indicating the colony morphology phenotype appears to be *lsr2*-independent. Re-streaking of these colonies resulted in a mixed population of colonies with either WT-like colony morphology or the ‘mushroom’ phenotype. Colony morphology changes in *M. tb* recovered from murine tissues during standard aerosol-infection experiments have not been previously reported. *M. tb* recovered from mice in a modified Cornell model were reported to exhibit atypical small, smooth colony morphology on 7H10 agar.
(Scanga, Mohan et al. 1999). The cause(s) of the altered colony morphology phenotypes in the previous study as well as my work remains unknown.
**Figure 4.5. Growth of the *M. tb* Δlsr2 in BALB/c mice.** Mice were aerosol infected with 30 CFU of WT, Δlsr2 or Δlsr2 + pLSR2 and at each time point, 6 mice per group were sacrificed and the bacterial burden in lung (A), spleen (B) and liver (C) was determined. (D) An example of the ‘mushroom’ colony morphology variants recovered from mouse tissues.
4.4 Discussion

My previous work led me to hypothesize that the H-NS functional analogue Lsr2 of mycobacteria may suppress genes necessary to orchestrate a genetic response to bacteriostatic conditions. In this study I assessed the biological role of Lsr2 by first comparing the viability of the \textit{M. smegmatis mc}^{2155} and its \textit{Δ}lsr2 mutant in two widely used models of mycobacterial dormancy: nutrient starvation and hypoxia. My results indicate that loss of \textit{lsr2} does not affect nutrient starvation survival. However, under conditions in which oxygen is the limiting factor, the loss of \textit{lsr2} is associated with increased fitness, as evidenced by the \textgreater{}100-fold increase in survival of the \textit{Δ}lsr2 versus the WT over 156 days of hypoxia; and by the mutant outcompeting its parental strain in a hypoxia competition experiment. Further supporting the selective advantage of the \textit{Δ}lsr2 mutant under hypoxia is the \textit{lsr2}-inactivation phenomenon I observed in the WT cells. I found a high frequency of the WT colonies becoming \textit{lsr2} mutants during the course of the hypoxia experiment, but never during the nutrient-starvation assay, suggesting that \textit{lsr2}-inactivation occurs specifically in response to hypoxia in \textit{M. smegmatis}. In a panel of 29 spontaneous \textit{lsr2} mutants from the WT background isolated from separate time points as well as from independent experiments for genetic characterization, I found that 25 contained an IS1096 insertion 70 bp upstream of \textit{lsr2} and the remaining 4 had various point mutations within \textit{lsr2}. Spontaneously arising \textit{lsr2} mutants have previously been isolated with an \textit{lsr2} point mutant isolated from an \textit{M. smegmatis} mutant for the stringent response regulator relA background during \textit{in vitro} dormancy studies (Arora, Whiteford et al. 2008) as well as two \textit{lsr2} mutants containing different IS1096 insertions within the \textit{lsr2} ORF (Kocincova, Singh et al. 2008). In the latter study, the frequency of spontaneous \textit{lsr2} mutants was estimated to be in the order of \textasciitilde{}10^{-5}. In contrast, I consistently found that \textgreater{}50\% of the WT and complemented strains’ colonies converted to \textit{lsr2} mutants under hypoxia, suggesting that this is not a random occurrence.

Many bacteria possess the ability to switch between two phenotypes in a heritable and reversible manner in a process termed phase variation [reviewed extensively in (van der Woude and Baumler 2004)]. The mechanisms underlying phase variation include DNA rearrangements/inversions, point mutations, frame shift mutations, altered methylation patterns and insertion sequences. In this chapter I was able to demonstrate the reversibility for some of the \textit{lsr2} IS1096 insertion and point mutants. In the case of the IS1096 revertants, sequencing
data indicated complete excision of the IS1096 along with the 8 bp target sequence duplication. Interestingly, reversible inactivation of virulence genes by an insertion sequence has been reported to occur in other bacteria including *Neisseria* and *Staphylococcus*. The capsule is an important virulence factor in *Neisseria meningitidis* which promotes survival and dissemination within the host; however, encapsulation hinders the attachment and invasion of host cells during the initiation of infection (Andreu and Gibert 2008). Hammerschmidt *et al.* demonstrated the reversible inactivation of a capsule biosynthesis gene by the insertion sequence IS1301, which is thought to provide the bacteria with the ability to both pass the mucosal barrier and to resist the host's immune defense (Hammerschmidt, Hilse *et al.* 1996). *Staphylococcus epidermis* is another human pathogen shown to undergo IS-mediated phase variation. Biofilm-producing *S. epidermidis* strains are known to be more virulent than biofilm-negative strains (Downing, Mischenko *et al.* 2005). Ziebuhr *et al.* showed that IS256 reversibly inserts into a 350 bp hotspot in the biofilm synthesis operon (Ziebuhr, Krimmer *et al.* 1999). It is unclear why biofilm-producing strains switch off biofilm production but it is known to occur *in vivo* (Downing, Betts *et al.* 2004) and was hypothesized to allow single bacterial cells to detach from the biofilm and disseminate to novel habitats (Ziebuhr, Krimmer *et al.* 1999). The overarching theme in these examples is an IS-mediated phase variable mechanism which leads to adaptation including changes in cell wall structure, as denoted by a colony morphology change, thought to be essential for mediating different phases of the respective bacteria’s life cycle. My findings seem to indicate transposon-mediated phase variation of the *lsr2* gene when alternating between aerobic versus anaerobic conditions in *M smegmatis*. The observation that *M. smegmatis* undergoes reversible inactivation of *lsr2* under hypoxia was intriguing. *M. smegmatis* is found throughout the environment including water where it can form biofilms. A loss of *lsr2* in this strain has been reported to cause the inability to form biofilms as well as become hypermotility (Chen, German *et al.* 2006; Arora, Whiteford *et al.* 2008); thus, perhaps reversible inactivation of *lsr2* allows *M. smegmatis* to assume two different lifestyles depending on the environment. It could be that cells lacking a functional *lsr2* gene allows single cells to break away in search of more favourable environments, similar to what was suggested for *S. epidermidis* (Ziebuhr, Krimmer *et al.* 1999).
TB infection is initiated after tubercle bacilli are deposited in the distal alveoli, where they are phagocytosed by alveolar macrophages. In this study I determined the virulence of the \textit{M. tb}\textDelta{}lsr2 mutant by assessing its ability to replicate within macrophages and an immunocompetent mouse model of infection. When I infected macrophages with the \textit{M. tb}\textDelta{}lsr2 mutant I observed a steady reduction of intracellular bacteria which could be explained by the extensive mortality of these macrophages. While the WT and complemented strain-infected macrophages were nearly 100\% confluent by 96 hours post infection, the \textDelta{}lsr2-infected macrophages had all detached which leads me to believe they underwent apoptosis. Infection of macrophages by nonpathogenic mycobacteria including avirulent \textit{M. tb} strains, leads to apoptosis which results in efficient killing of intracellular bacilli and acts as a bridge from the innate immune response to the adaptive immune response by facilitating antigen presentation for CD8 T-cell priming (Molloy, Laochummoonvorapong et al. 1994; Schaible, Winau et al. 2003). In contrast, virulent mycobacteria exhibit faster intracellular growth rates than avirulent strains and actively suppress macrophage apoptosis in favour of necrosis, a process characterized by plasma membrane lysis and release of more bacilli into the surrounding tissue to propagate the infection (Keane, Balcewicz-Sablinska et al. 1997; Keane, Remold et al. 2000). Both virulent and avirulent mycobacteria cause macrophage plasma membrane microdisruptions, which when left unrepaired leads to necrosis. In turn the macrophage responds by resealing these lesions via translocation of lysosomal and Golgi apparatus–derived vesicles to the plasma membrane in an attempt to tip the scales in favour of apoptosis; however, virulent \textit{M. tb} is able to inhibit this vesicle-mediated plasma membrane repair (Divangahi, Chen et al. 2009). Furthermore, \textit{M. tb} is able to block phagosome maturation by interrupting acidification and lysosome fusion, which creates a suitable niche for bacterial replication (Armstrong and Hart 1971). There is evidence that phagosome-lysosome fusion and membrane repair require phosphatidylinositol 3-phosphate (PI3P) production, and that \textit{M. tb} secretes SapM, a PI3P phosphatase capable of blocking phagosome-late endosome fusion in an \textit{in vitro} system (Vergne, Chua et al. 2005; Divangahi, Chen et al. 2009). Interestingly, I found that \textit{sapM} is significantly downregulated in the \textDelta{}lsr2 mutant in my microarray data. Furthermore, phagosome maturation also requires the conversion of Rab5 to active, GTP-bound Rab7 and \textit{M. tb} secretes two effectors, PtpA and NdkA, thought to be responsible for the absence of Rab7 in the mycobacterial phagosome. PtpA is a tyrosine phosphatase that dephosphorylates the host Vps33B, a component of the homotypic fusion and
protein sorting complex, which is critical for Rab7 activation (Bach, Papavinasasundaram et al. 2008). NdkA is a guanosine triphosphatase GTPase-activating protein that may directly catalyze the conversion of GTP to GDP on Rab7 to render it inactive (Sun, Wang et al. 2010). Both ptPA and ndkA are also significantly downregulated in the mutant. It has been shown that the SecA2-dependent secretion of the mycobacterial superoxide dismutase A (SodA) is required to avert host cell apoptosis (Hinchey, Lee et al. 2007) and that overproduction of superoxide anions is proapoptotic (Kahl, Kampkotter et al. 2004). The soda gene is also significantly downregulated in the mutant. Although, it is unknown by what mechanism the macrophages are dieieng, the downregulation of these four important host cell effector proteins in Δlsr2 could lead to the inability to inhibit host macrophage apoptosis. Future study is warranted to investigate this hypothesis.

Lastly, I evaluated the ability of the mutant to survive in a mouse model of infection. I found that the Δlsr2 replicated at a slower rate resulting in a lower bacterial burden during the chronic stage of infection, along with delayed dissemination to secondary sites of infection. These observations may be attributed to the slower growth rate of the mutant observed during in vitro growth as well as the intracellular growth defect of the mutant. An increase in apoptosis of host macrophage would facilitate better clearance of the mutant and a more robust adaptive immune response. It remains to be seen how the mutant would survive in a longer-term infection. I expect that the lsr2mutant would persist in the lungs of infected animals longer with less severe tissue damage due to a reduction in the recruitment of proinflammatory cytokines to infected organs, ultimately prolonging host survival. In this study, I described the recovery of altered colonies which resembled the Δlsr2mutant I recovered immediately after phage transduction to generate the strain. These colony morphology variants were most readily observed in the liver, an organ rife with cholesterol; thus, given my findings that the mutant’s metabolism is realigned to utilize cholesterol the inactivation of lsr2 in this tissue would not be surprising. I was however unable to identify any genetic lesions affecting lsr2 in a panel of variants tested. A possible scenario could be that during the in vitro culturing of these colonies that they somehow restored the lsr2 gene. Or perhaps lsr2 expression is abrogated due to an indirect mechanism such as a mutation in an activator of lsr2. Determining the expression level of lsr2 and Lsr2 protein levels will be able to distinguish between these two possibilities.
4.5 Experimental procedures

**Bacterial strains, media, and growth conditions.** All strains of *Mycobacterium smegmatis* strain mc²155 and *Mycobacterium tuberculosis* H37Rv were routinely grown in Middlebrook 7H9 broth or Middlebrook 7H11 agar (Difco) supplemented with 10% OADC (oleic acid, bovine serum albumin [fraction V], dextrose, and catalase; Difco). When appropriate kanamycin was added at a concentration of 25 µg/ml and hygromycin at a concentration of 50 µg/ml.

**Persistence and competition assays.** Cultures were grown to OD<sub>600</sub>~0.5 at which point they were aliquoted to 2 mL screw-top tubes. For the hypoxia NRP assay, 2 mL per tube was added so that there was no headspace air as a source of oxygen. In the hypoxia competition assay, separate aerobically grown cultures of *M. smegmatis*+ pME and ∆lsr2 were mixed in a 1:1 ratio to give an OD<sub>600</sub> of 0.6 before being aliquoted to the sealed tubes. For the nutrient starvation NRP assay log phase cultures were pelleted and washed twice with 1×PBS before being resuspended in 1×PBS and then 0.2 µL was aliquoted to 2 mL tubes. In both assays, tubes were incubated without shaking at 37°C. At various time points over a 156 day period the CFU/ml was determined by plating serial dilutions onto 7H11 agar from triplicate tubes for each strain. Oxygen depletion was monitored using the oxygen indicator dye methylene blue as previously described (Wayne and Hayes 1996). Control tubes containing either PBS or 7H9 broth medium with methylene blue but no bacteria were also set up. For the aerated competition experiments separately grown cultures of *M. smegmatis*+ pME and ∆lsr2 were mixed in a 1:1 ratio and diluted to an OD<sub>600</sub> of 0.005 in a 5 mL volume. Cultures were grown to stationary phase (OD<sub>600</sub>~1.5) when serial dilutions and plating to assess the CFU for each strain was performed. This culture was then used to repeat the passaging cycle.

**Isolation of chromosomal DNA, PCR amplification, and restriction endonuclease cleavage.** *M.smegmatis* and *M. tb* colonies of interest were grown to OD<sub>600</sub> ~1.0. Cultures were then centrifuged, resuspended in 1 mL ice-cold 1×PBS and transferred to a 2 mL screw cap tube containing 0.5 g of 0.1-mm zirconia/silica beads (BioSpec). Cells were lysed by bead beating (Mini-Beadbeater; BioSpec Products) at the high speed setting for 2 × 1 min pulses. Cell debris was removed by centrifugation at 13 000 rpm for 30 minutes. The supernatant was then transferred to a 1.5-mL Eppendorf tube and boiled for 5 min. The crude DNA preps were then
used for PCR or endonuclease digestion. The PCR primers used to amplify the *M. smegmatis lsr2* ORF were the forwards primer 5’-ATGGCAAAAGAAAGTGACCGTCACG-3’ and the reverse primer 5’-CTAAGTTGCCGCTGGGAATGCGTC-3’. The primers used for the amplification of the hybrid PCR products between IS1096 and the *lsr2* promoter were either the forward primer 5’-TAACCGCCGATCGTACCAGCAGCAGAAA-3’ (IS1096 inserted in the same orientation as the *lsr2* ORF) or 5’-CTCTGTGCTTCCGTCTTCTCCTCCG-3’ (IS1096 inserted in the reverse orientation as the *lsr2* ORF), and the reverse primer 5’-CGTGACGTCACCTTTCTTGTCCAT-3’ targeting the first 24 bases of *lsr2*. The *lsr2* of *M. tb* was PCR amplified using the forward primer 5’-AAAAGAATTCATGGCGAAAGTAACCGTCACC-3’ along with the reverse primer 5’-TTTTCTCGAGTCAGGCCTGTGATGCTG-3’. *M. smegmatis* DNA (10µg) for southern blot analysis was digested overnight in a thermocycler at 37°C with the restriction enzymes *PvuI* and *Acc651*. *M. tb* DNA for southern blot analysis was digested in the same manner except with the restriction enzymes *BbsI* and *PvuII*.

**Southern blot analysis.** The digested DNA was separated on a 1.0% agarose gel and blotted onto Hybond N+ nylon membranes (Amersham Biosciences). The AlkPhos Direct Labeling and Detection System with CDP-Star was used to label 100ng of the respective *lsr2* PCR product with the above primers according to the manufacturer’s instructions.

**Intracellular survival.** Infection of murine RAW 264.7 macrophage and enumeration of intracellular *M. tb* CFU were performed as previously described (Ramakrishnan and Falkow 1994). Briefly, infecting inocula was sonicated 2×1 minute pulses and passed through a 24 gauge needle to disperse bacterial aggregates. Approximately 12 h prior to infection, macrophages were seeded into 24-well tissue culture plates at a density of 5×10^4cells per well in 1 ml RPMI medium supplemented with 10% FBS. RAW 264.7 cells were infected by *M. tb* at an MOI of ~1 (e.g., one bacterium to one macrophage). The actual inoculum used was determined by plating on 7H11 agar plates and enumerating CFU/mL. The actual CFU/mL values are given in the text. The infection was allowed to proceed for 3 h at 37°C in 5% CO₂, and the extracellular bacteria were then removed by washing the macrophage monolayer 3× with RPMI. Subsequently, cells were incubated with fresh RPMI containing 10% FBS and 20 µg/ml gentamicin at 37°C in 5% CO₂ with the media replaced on alternate days. On days 0 (3 hours), 2, 4 and 6, the infected
macrophage monolayers (three wells per strain) were washed twice with fresh RPMI media and then lysed with 0.1 ml of 1% Triton X-100 (Sigma) to release intracellular mycobacteria, which were then enumerated by plating serial dilutions on 7H11 agar plates.

**Mouse infection.** Frozen stocks of *M. tb* H37Rv WT, Δlsr2 or Δlsr2 + pLSR2 in PBS-0.05% Tween80, were thawed overnight at 37°C and sonicated before use. BALB/c mice were infected with 30-50 CFU using the Glass-Col Inhalation Exposure System (Glax-Col, LLC). The concentration of bacteria suspended in PBS-0.05% required to deliver 30-50 CFU per mouse was standardized for each frozen stock prior to performing the actual experiment. Four mice were sacrificed 24 h post-infection to confirm the correct dose was delivered. At time points 0 (24 hours), 2 weeks, 4 weeks, 6 weeks and 8 weeks, six mice per strain were sacrificed and the lung, liver and spleen were harvested. Tissues were homogenized in 2 mL PBS-0.05% Tween80 using the OMNI TH homogenizer. Organ homogenates were serially diluted, and plated in triplicate on 7H11 at 37°C, with colonies counted 4 weeks later.
CHAPTER 5: STRUCTURAL BASIS FOR RECOGNITION OF AT-RICH DNA BY UNRELATED XENOGENEIC SILENCING PROTEINS

A version of this chapter has been published (Gordon, B. R., Y. Li, A. Cote, M. T. Weirauch, P. Ding, T. R. Hughes, W. W. Navarre, B. Xia, J. Liu. 2011. PNAS.108(26): 10690-10695) with data from (Gordon, B. R., Y. Li, A. Cote, A. Sintsova, H. van Bakel, S. Tian, W. W. Navarre, B. Xia, J. Liu. 2010. PNAS.107(11): 5154-5159). This chapter includes data performed by others. I made the constructs for the PBM experiments, participated in the data analysis and did some of the RT-PCR. Atina Cote performed the Protein Binding Microarray experiments. The domain swapping experiments and assessment of chimeric proteins were performed by Linru Wang and Anna Sintsova. Structure determinations, NMR titration experiments, docking models and the netropsin competition assays were performed by our collaborator Bin Xia’s lab (Yifei Li, Songhai Tian and Pengfei Ding) from Peking University in Beijing, China.
5.1 Abstract

H-NS and Lsr2 are nucleoid-associated proteins from Gram-negative bacteria and Mycobacteria, respectively, that play an important role in the silencing of horizontally acquired foreign DNA that is more AT-rich than the resident genome. Despite the fact that Lsr2 and H-NS proteins are dissimilar in sequence and structure, they serve apparently similar functions and can functionally complement one another. The mechanism by which these xenogeneic silencers selectively target AT-rich DNA has been enigmatic. We performed high-resolution protein binding microarray analysis to simultaneously assess the binding preference of H-NS and Lsr2 for all possible 8-base sequences. Concurrently, we performed a detailed structure-function relationship analysis of their C-terminal DNA binding domains by NMR. Unexpectedly, we found that H-NS and Lsr2 use a common DNA binding mechanism whereby a short loop containing a “Q/RGR” motif selectively interacts with the DNA minor groove, where the highest affinity is for AT-rich sequences that lack A-tracts. Mutations of the Q/RGR motif abolished DNA binding activity. Netropsin, a DNA minor groove-binding molecule effectively outcompeted H-NS and Lsr2 for binding to AT-rich sequences. These results provide a unified molecular mechanism to explain findings related to xenogeneic silencing proteins, including their lack of apparent sequence specificity but preference for AT-rich sequences. Our findings also suggest that structural information contained within the DNA minor groove is deciphered by xenogeneic silencing proteins to distinguish genetic material that is self from nonself.
5.2 Introduction
The H-NS protein is one of the most intensively studied members of the bacterial nucleoid-associated proteins (Dorman 2004). Initially discovered under conditions for isolating eukaryotic histones (Varshavsky, Nedospasov et al. 1977), H-NS is thought to play a role in nucleoid structure (Falconi, Gualtieri et al. 1988). The recent identification of in vivo H-NS binding sites by genome-wide studies revealed that H-NS also plays a major role in selectively silencing expression from sequences with significantly higher AT-content than the core genome (Grainger, Hurd et al. 2006; Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006; Oshima, Ishikawa et al. 2006). AT-rich genomic islands are usually obtained via horizontal gene transfer and are associated with adaptive stress responses and virulence (Ochman, Lawrence et al. 2000). The silencing of such sequences by H-NS, termed xenogeneic silencing, is thought to allow bacteria to safely acquire new genetic material without compromising their genomic and regulatory integrity (Navarre, McClelland et al. 2007). As a result, H-NS serves as a central regulator of virulence gene expression in the enteric bacteria, including Yersinia, Escherichia coli, and Salmonella (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006; Cathelyn, Elliston et al. 2007; Banos, Pons et al. 2008; Muller, Schneider et al. 2010).

Clearly identifiable H-NS homologs are found only in a subset of proteobacteria (Tendeng and Bertin 2003). However, proteins lacking sequence similarity to H-NS may function as xenogeneic silencers in other bacteria. The Pseudomonas MvaT/MvaU can complement phenotypes of E. coli hns mutants and selectively bind AT-rich sequences in a manner similar to H-NS (Tendeng, Soutourina et al. 2003; Castang, McManus et al. 2008). The Rok protein of Bacillus subtilis was also found to be a functional analog of H-NS (Smits and Grossman 2010). Lsr2 of Mycobacterium tuberculosis (M. tb) can complement hns phenotypes in E. coli (Gordon, Imperial et al. 2008) and specifically binds and silences AT-rich regions of the mycobacterial genome (Chen, Ren et al. 2008; Kocincova, Singh et al. 2008; Gordon, Li et al. 2010). The functional equivalence of H-NS and Lsr2 appears to be a result of convergent evolution, as Lsr2 exhibits <20% sequence identity with H-NS and has a different predicted secondary structure (Gordon, Imperial et al. 2008).

An important feature that is critical for H-NS function is its ability to multimerize to form higher-order nucleoprotein complexes (Badaut, Williams et al. 2002; Stella, Spurio et al. 2005), a property that is also important for the function of the MvaT/MvaU proteins (Castang and Dove
The 80 N-terminal amino acids of H-NS contain two distinct dimerization domains that form extended higher-order structures via head-to-head/tail-to-tail interactions (Dorman 2004; Arold, Leonard et al. 2010), and are joined through a flexible linker to the C-terminal DNA-binding domain (residues 91–137) (Sette, Spurio et al. 2009; Arold, Leonard et al. 2010).

The mechanism by which H-NS and analogous proteins selectively bind AT-rich DNA remains unclear. Although early biochemical studies suggested that a major binding determinant is intrinsically curved DNA caused by repeated A-tracts (Owen-Hughes, Pavitt et al. 1992; Tupper, Owen-Hughes et al. 1994), recent genome-wide binding studies revealed that H-NS binding correlates much more strongly with the degree of AT-content (Lucchini, Rowley et al. 2006). Footprinting and ChIP studies indicate that H-NS binds DNA with little sequence specificity. However, a high-affinity binding sequence present in some promoters was recently identified (Wu, Yang et al. 2008), which may serve as a nucleation site for subsequent polymerization on lower affinity AT-rich sites. The critical feature for the specific H-NS high-affinity sequence examined is the presence of a T-A step that may cause a structural anomaly (Lang, Blot et al. 2007), but this mode of binding does not explain how H-NS binds the majority of sites mapped in vivo and by footprinting of a number of promoters in vitro.

In this study we sought to explore the mechanism by which the H-NS and Lsr2 proteins selectively target AT-rich sequences.
5.3 Results

5.3.1 H-NS and Lsr2 Exhibit Similar DNA Binding Specificity

To characterize the DNA binding specificity of H-NS and Lsr2, we used protein binding microarray (PBM), whereby each protein was individually applied to microarray slides containing double-stranded oligonucleotide target sequences to simultaneously assess their affinity for various sequences in an unbiased manner. HNS and Lsr2 were produced as N-terminal GST-tagged proteins by T7-driven in vitro transcription and translation and applied to arrays containing 41,944 features that were designed such that all nonpalindromic 8-mers are represented 32 times (16 times for palindromic 8-mers) on each array, so that the assay provides a robust estimate of relative preference to each 8-mer (Berger, Philippakis et al. 2006). The combined data (average) from two independent array experiments using arrays of different design were used to analyze the relative binding preferences.

For each 8-mer, we quantified the relative binding preference by two measures, the Z-score and E-score, which capture essentially the same information (relative binding preference of a protein for a given 8-base sequence) (Berger, Philippakis et al. 2006; Badis, Berger et al. 2009). The Z-score is calculated from the average signal intensity across the 32 spots containing each 8-mer on each microarray and scales almost linearly with binding affinity (Berger, Philippakis et al. 2006; Badis, Berger et al. 2009). The E-score is a nonparametric statistical measure that essentially reflects the relative ranking of the signal intensities of the 32 probes that contain each 8-mer, relative to the remaining ~41,000 probes (Berger, Philippakis et al. 2006). The E-score ranges from +0.5 (most favored) to −0.5 (most disfavored), and on the basis of random permutations of the array data, there should be no random 8-mer sequence that achieves an E-score above 0.45 (Berger, Badis et al. 2008).

PBM experiments with both H-NS and Lsr2 were successful, in that we obtained multiple 8-mers with E-scores above 0.45 for each protein with the highest score around 0.49, indicating clear binding preferences. Moreover, the most highly preferred sequences had an obvious relationship to each other (see below). There is a striking correlation between the PBM data for H-NS and Lsr2, not only for the most preferred sequences, but also for moderately and less-preferred sequences. This finding is evident from the scatter plot using the Z-score (\(R^2 = 0.859\))
(Figure 5.1A) or E-score ($R^2 = 0.809$) (data not included) of all8-mers. Therefore, H-NS and Lsr2 bind DNA with essentially the same sequence preference.
Figure 5.1. H-NS and Lsr2 exhibit similar DNA binding specificity and bind contiguous AT sequences. (A) Scatter plots comparing DNA binding of Lsr2 and H-NS; the Z-score of individual 8-mers were plotted. The Z-score was calculated from the average signal intensity across the 16 or 32 spots containing each 8-mer. Examples of 8-mers containing the indicated 6-mer sequences are highlighted. (B) The Z-score of individual 8-mers were plotted against the AT-content of each 8-mers, showing binding preference of H-NS and Lsr2 for AT-rich sequences. A total of 32,896 DNA sequences were used for calculation for each protein and the results are mean ± SD. (C) The Z-score (mean ± SD) of individual 8-mers that are 87.5% AT was plotted against the position of G or C. (D) The Z-score (mean ± SEM) of individual 8-mers containing 75% AT was plotted against the number of contiguous AT sequences.
5.3.2 H-NS and Lsr2 Bind Contiguous AT Sequences

Previous ChIP-chip analysis suggested that H-NS and Lsr2 exhibit preferential binding for AT-rich sequences in the genome (Lucchini, Rowley et al. 2006; Navarre, Porwollik et al. 2006; Gordon, Li et al. 2010). This finding is confirmed by our PBM data. A positive correlation between the Z-score (Figure 5.1B) or E-score (data not included) and the AT-content of 8-mers was observed. The data for H-NS and Lsr2 are nearly identical and essentially overlap.

For each group of 8-mers with identical AT-content, the position of G or C within the sequence affects binding. For 8-mersthat contain a single G or C residue (87.5% AT), sequences containing G or C at the central position show markedly lower preference by H-NS and Lsr2 than those where the G or C is located at the periphery positions (Figure 5.1C). Extending the analysis to 8-mers that are 75% AT (two Gs or Cs) revealed that there is a positive correlation between the binding preference and the maximum length of contiguous AT sequence (no G or C) (Figure 5.1D), indicating that H-NS and Lsr2 bind contiguous AT sequences.

5.3.3 DNA Minor Groove Width Determines the Binding Preference of H-NS and Lsr2

To identify DNA sequence features other than AT-content that influence H-NS and Lsr2 binding, we next focused on the 136 8-mer sequences that are 100% AT. H-NS and Lsr2exhibit a clear preference for these sequences but they are not equally favored; the Z-score and E-score for the 100% AT 8-mers range from 2.4 to 10.6 and 0.28 to 0.49, respectively, for both H-NS and Lsr2. Strikingly, many of the less-preferred100%-AT 8-mers contain A-tracts, which are defined as stretches of four or more As or Ts (Aₙ, Tₙ, n ≥ 4, or AₙTₙ, m + n ≥ 4, on the same strand) that do not contain T-A steps (reviewed in (Haran and Mohanty 2009)). The average Z-score for the 8-mers that do not contain A-tracts are significantly higher than that of 8-merscontaining A-tracts (Figure 5.2A), suggesting that A-tracts have a negative impact on H-NS and Lsr2 binding for these DNA oligomers. This finding is consistent with previous observation that a T-A step is present in some H-NS high affinity sequences(Lang, Blot et al. 2007) because T-A steps and A-tracts are exclusive to each other. Notably, the effect of A-tracts on H-NS and Lsr2 binding is dependent on the overall AT-content of the sequence. For 8-mers containing 50% to 75% AT,
the presence of A-tracts positively influences H-NS and Lsr2 binding, but for 8-mers that contain 87.5% AT, A-tracts do not have a significant impact on binding (Figure 5.2A).

The simplest explanation for the PBM data is that H-NS and Lsr2 exhibit optimal binding to DNA with appropriate minor groove width. A-tracts have narrower minor grooves than other DNA sequences including AT-rich sequences containing T-A steps (Haran and Mohanty 2009), and AT-rich sequences have narrower minor grooves than GC-rich sequences (Rohs, West et al. 2009; Stella, Cascio et al. 2010). For relatively GC-rich sequences (e.g., 50–75% AT), the occurrence of A-tracts may result in increased binding for H-NS and Lsr2 because it serves to narrow minor groove width. The positive impact of A-tracts on relatively GC-rich sequences (e.g., 50% AT) is not simply a result of the contiguousness of the AT sequence in A-tracts, because such sequences containing A-tracts are still more favorable binding sites than similar sequences containing contiguous AT-rich sequences that lack A-tracts and contain T-A steps (Figure 5.2B). DNA sequences with higher AT-content have narrower minor grooves (Rohs, West et al. 2009; Stella, Cascio et al. 2010) and the narrowing effect of A-tracts becomes insignificant upon reaching 87.5% AT. For sequences that are 100% AT, the minor groove may reach optimal geometry for binding, and the presence of A-tracts, which further narrows the minor groove, becomes unfavorable.
Figure 5.2. Effects of A-tracts on H-NS and Lsr2 binding to DNA. (A) The Z-score (mean ± SEM) of individual 8-mers were plotted for each AT-content, and the 8-mers that contain an A-tract DNA were compared with 8-mers that do not contain A-tracts but have the same AT-content. (B) The Z-score (mean ±SEM) of individual 8-mers containing 50% AT was plotted, and 8-mers containing A-tract DNA (AAAA, AAAT, AATT, ATTT, TTTT, filled bar) were compared with 8-mers that contain contiguous AT sequences (ATAA, ATAT, ATTA, AATA, TAAA, TAAT, TATA, TATT, TTAA, TTAT, TTTA, empty bar) but not A-tracts (they contain the T-A step). ***P < 0.0001; ns, not significant.
5.3.4 Lsr2 has an N-Terminal Dimerization Domain and a C-Terminal DNA-Binding Domain

The 80 N-terminal amino acids of H-NS contain two distinct dimerization domains that form extended higher-order structures via head-to-head/tail-to-tail interactions (Dorman 2004; Arold, Leonard et al. 2010), and are joined through a flexible linker to the C-terminal DNA-binding domain (residues 91–137) (Sette, Spurio et al. 2009; Arold, Leonard et al. 2010). Less is known about Lsr2, but it has been shown to form dimers in vivo (Chen, German et al. 2006). To map the functional domains of Lsr2, six chimeric protein constructs containing different combinations of the N- and C-terminal sequences of Lsr2 and H-NS were generated, and their ability to complement phenotypes associated with hns mutations in E. coli and lsr2 mutations in M. smegmatis was tested (Gordon, Imperial et al. 2008). Three constructs (C1, C4, and C6) could complement several phenotypes in hns mutants but were ineffective at complementing an lsr2 mutation in M. smegmatis (Table 5.1). One construct (C5) containing the N-terminal region (residues 1–65) of Lsr2 and the C-terminal region of H-NS (residues 77–137) complemented all phenotypes associated with hns and lsr2 mutations (Table 5.1). To determine whether repression by the chimeric proteins was specific to H-NS regulated genes, the binding specificity of HA-tagged chimeric proteins in E. coli was analyzed by ChIP followed by quantitative real-time PCR. C1, C5, and C6 co-precipitated with known H-NS binding sites (proV, bglG, yjcF, and xapR) with significantly higher levels of enrichment than the sites known not to interact with H-NS (narZ and phnE) (Grainger, Hurd et al. 2006) (Figure 5.3A). These results indicate that Lsr2 contains an N-terminal dimerization domain between residues 1–65 and a C-terminal DNA-binding domain within residues 51–112. Moreover, the purified N-domain protein (residues 1–65) formed dimers but failed to bind DNA, whereas the purified C-domain (51–112) bound DNA but was deficient in dimer formation (Figure 5.3B–D). Thus, Lsr2 and H-NS share a similar overall domain organization, which provides a molecular explanation for their functional similarity.
Table 5.1. Summary of *in vivo* complementation experiments for chimeric proteins of Lsr2 and H-NS

<table>
<thead>
<tr>
<th>Chimeric proteins*</th>
<th>Composition</th>
<th>Phenotypes associated with <em>E. coli hns</em> mutants</th>
<th>Colony morphology of <em>M. smegmatis lsr2</em> mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MucoYd</td>
<td>Motility</td>
</tr>
<tr>
<td>C1</td>
<td>H-NS (aa 1-89)-Lsr2 (aa 51-112)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>C2</td>
<td>Lsr2 (aa 1-50)-H-NS (aa 77-137)</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>C3</td>
<td>Lsr2 (aa 1-50)-H-NS (aa 65-137)</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>C4</td>
<td>H-NS (aa 1-89)-Lsr2 (aa 74-112)</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>C5</td>
<td>Lsr2 (aa 1-65)-H-NS (aa 77-137)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>C6</td>
<td>Lsr2 (aa 1-65)-H-NS (aa 65-137)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+ , phenotype complemented; − , not complemented.

*The expression of the chimeric proteins in *E. coli* and *M. smegmatis* were confirmed by Western blot analysis.
Figure 5.3. Structure–function analysis of Lsr2. (A) Binding of Lsr2-H-NS chimeric proteins to *E. coli* genes. The binding of chimeric proteins (C1–C6; see text for description) to H-NS-regulated genes (*proV, bgIG, yjcF, and xapR*) and genes not regulated by H-NS (*narZ* and *phnE*) was assayed by ChIP and RT–PCR. (B and C) Cross-linking of purified Lsr2 proteins. Glutaraldehyde (1%) was added to purified protein (6 μg/sample). Aliquots were removed at the indicated time points (min) and analyzed by Western blotting with an anti-His antibody. (B) N-terminal domain Lsr2 protein (residues 1–65). (C) C-terminal domain Lsr2 protein (residues 51–112). (D) Electrophoretic mobility shift assays. DNA fragment P104 *lmo* (50 ng) was incubated with the indicated amounts (μg) of N- and C-terminal Lsr2 proteins and analyzed on 4% polyacrylamide gel.
5.3.5 DNA-Binding Domain of Lsr2 Exhibits a Unique Structure compared to the DNA-Binding Domain of H-NS

The structure of the C-domain (Lsr2\textsubscript{Ctd}, residues 66–112) was solved by our collaborator, Bin Xia and colleagues, using nuclear magnetic resonance (NMR) methods. I have summarized the results of their work below. For a more in depth presentation of the data see our publications (Gordon, Li et al. 2010) and (Gordon, Li et al. 2011). The domain consists of two $\alpha$-helices ($\alpha_1$, residues 78–89; $\alpha_2$, residues 102–112) linked by a long loop (residues 90–101) (Figure 5.4A and B). The first nine residues (66–74) are completely flexible. The two helices are perpendicular to and packed against each other through hydrophobic interactions among residues Ile83, Arg84, Ala87, Val94, Ile100, and Val104 and aromatic stacking between Trp86 and Tyr108. A protein structure database search using DALI did not identify any structure with a z-score over 3.0. The structure of Lsr2\textsubscript{Ctd} is unique and distinct from that of H-NS which was also solved by our collaborator. The C-terminal DNA binding domain (H-NS\textsubscript{Ctd}, residues 91–137) of the Salmonella H-NS consists of a two-stranded antiparallel $\beta$-sheet ($\beta_1$, residues 97–100; $\beta_2$, 105–109), linked by a well-defined loop (loop2, residues 110–117) to an $\alpha$-helix (residues 117–126), and followed by one $3_{10}$ helix (residues 130–133) (Figure 5.4C and D).
Figure 5.4. Solution structures of the Lsr2<sub>Ctd</sub> and H-NS<sub>Ctd</sub>. (A) Superimposition of backbone traces for the ensemble of 20 structures of Lsr2<sub>Ctd</sub>. (B) Ribbon representation of the Lsr2<sub>Ctd</sub> mean structure; the flexible N-terminal region (residues 66–74) is not included. Side-chains of residues that form the hydrophobic core are shown in red, except those of W86 and Y108, which are shown in orange. (C) Superimposition of backbone traces for the ensemble of 20 structures of H-NS<sub>Ctd</sub>. (D) Ribbon representation of H-NS<sub>Ctd</sub> mean structure. Only well-structured regions are included. Side chains of residues that form the hydrophobic core are shown in red. (E) Sequence alignment of the C-terminal domains of H-NS, Bv3F and Lsr2. (F) Superimposing structures of H-NS<sub>Ctd</sub> (red), Bv3F<sub>Ctd</sub> (blue), and Lsr2<sub>Ctd</sub> (yellow). The loop consisting of the conserved “T/SXQ/RGRXPA” motif adopts nearly identical conformation in these three structures.
5.3.6 Lsr2 and H-NS Bind the DNA Minor Groove using an AT-Hook–Like Loop

The DNA-binding site of Lsr2Ctd was mapped by NMR titration experiments using a 27-mer DNA containing 9 consecutive A-T base pairs. Comparison of 2D 1H-15N HSQC spectra of Lsr2Ctd, free and in complex with DNA, reveals that significant chemical shift changes occur at residues located mainly on the α1 helix and the nearby linker loop, constituting the DNA-binding sites of Lsr2Ctd. A structure model for the Lsr2C/DNA complex was calculated using HADDOCK 2.0 (Smits and Grossman 2010), and the result revealed that Lsr2 can bind DNA by grabbing either edge of the minor groove like a clamp (Figure 5.5A and B). There are two major components of Lsr2 involved in DNA binding. Residues Arg97-Gly98-Arg99 of the linker loop adopt an extended conformation and are inserted into and oriented parallel to the minor groove, and the two Arg side-chains point away from each other and occupy a region covering five A-T base pairs. This resembles the central Arg-Gly-Arg core conformation of the DNA-binding AT-hook motif (i.e., Pro-Arg-Gly-Arg-Pro, flanked by positively charged residues) of the mammalian nonhistone chromatin protein HMGA (Badaut, Williams et al. 2002) (Figure 5.5A). HMGA proteins bind AT-rich DNA and are involved in regulation of chromatin structure and gene expression (Badaut, Williams et al. 2002). The conformation of the Arg-Gly-Arg core in the AT-hook motif and the nature of its interactions with bases of AT pairs are the main determinants for HMGA binding to the minor groove of AT-rich DNA (Badaut, Williams et al. 2002). Therefore, the presence of an AT-hook Arg-Gly-Arg core-like conformation in the Lsr2 structure provides a molecular explanation for the preferential binding of Lsr2 to AT-rich DNA and is consistent with the previous finding that poly(dI-dC), with hydrogen bond patterns in the minor grooves identical to that of AT-rich DNA, is preferentially bound by Lsr2 (Chen, Ren et al. 2008). Outside of the minor groove, side-chains of Arg77, Ser80, Arg84, and Ser95 interact with the sugar–phosphate backbone on either edge of the minor groove. These additional interactions increase the binding affinity of Lsr2 for DNA, which is in the micromolar range on the basis of NMR titration experiments, compared to that of short AT-hook peptides (e.g., Pro-Arg-Gly-Arg-Pro), which is in the millimolar range (Castang and Dove 2010). As a result, Arg77 and Arg80, along with Arg97-Gly98-Arg99 of the linker loop, form a positively charged cleft and act like a clamp to grab one strand of DNA (Figure 5.5B), representing a unique mechanism of
DNA recognition. These residues, particularly Arg84 and Arg97-Gly98-Arg99, are highly conserved among Lsr2 homologs (Gordon, Imperial et al. 2008).

Although H-NS does not exhibit sequence homology to Lsr2, the C-terminal domain of H-NS family proteins contains an XGR motif imbedded in the conserved TWTXGRXP sequence that has been implicated in DNA binding (Figure 5.4E) (Tendeng and Bertin 2003; Dorman 2004). In H-NS, the X residue of the XGR motif is Gln, whereas in some other H-NS homologs, such as the Bv3F protein of *Burkholderia vietnamiensis*, the X residue is Arg. Furthermore, the loops containing the Q/RGR motif in H-NS Ctd and Bv3F Ctd adopt almost identical conformation, which is also quite similar to the RGR motif in Lsr2 Ctd (Figure 5.4F), even though the overall tertiary structure of Lsr2 Ctd is quite different from the other two. This result suggested that the H-NS proteins likely bind DNA by a similar mechanism and that the loop structure is critical for the observed similarities in target specificity.

NMR titration experiments for both H-NS Ctd and Bv3F Ctd with DNA confirmed that the loop containing the XGR motif is intimately involved in DNA binding. In the docking model for H-NS Ctd/DNA, the loop region consisting of Gln112-Gly113-Arg114 is inserted into the minor groove of the DNA duplex. The side chains of Gln112 and Arg114 are oriented parallel to the minor groove pointing away from each other, and occupy a region covering 5 base pairs (Figure 5.5C). Similarly, in the case of Bv3F Ctd, the Arg89-Gly90-Arg91 loop motif inserts into the minor groove of DNA, and the side chains of the two Arg residues are in a conformation that resembles that of the AT-hook motif (Figure 5.5D) (Badaut, Williams et al. 2002). Taken together, these results indicate that Lsr2 and H-NS both bind the DNA minor groove through an AT-hook–like motif which likely explains their similar DNA binding preferences.
Figure 5.5. Structure models of Lsr2Ctd/DNA, H-NSCtd/DNA and Bv3FCtd/DNA complexes generated by HADDOCK 2.0. (A) Ribbon representation of Lsr2C/DNA complex in which Lsr2C clamps the A21–A25 edge of DNA minor groove. Residues of Lsr2Ctd involved in DNA binding are in pink, and residues of DNA affected by protein binding are in yellow. Side-chains of Arg77, Ser80, Arg84, Ser95, Arg97, and Arg99 are in red. An AT-hook motif of HMG-I (Protein Data Bank code: 2ezd) is superimposed onto the Lsr2Ctd linker loop with the backbone of residues 7–15 and two Arg side-chains shown in blue. (B) A different view of the same complex. The electrostatic potential surface of Lsr2Ctd is shown, and covalent bonds between heavy atoms are illustrated for DNA. Blue: positively charged residues; red: negatively charged residues; gray: uncharged residues. (C and D) Ribbon representation of (C) H-NSCtd/DNA and (D) Bv3FCtd/DNA docking models. Protein backbone and side chains that are involved in DNA-binding are labeled in red. Residues without NH signals but may be involved in DNA binding are labeled in orange. The binding interface of DNA is shown in blue. Residue E102 of H-NS that showed chemical-shift changes but is not involved in DNA binding is labeled in purple.
5.3.7 The Q/RGR Motif Is Essential for H-NS and Lsr2 Binding the DNA Minor Groove

The docking models predict that the conserved Q/RGR motif is the primary site interacting with DNA. Specifically, side chains of the first and last residues (Q/R and R) directly contact the DNA minor groove. To confirm this, we performed site-directed mutagenesis to replace the Gln/Arg and Arg residues with Ala and examined the interaction of mutant proteins with a DNA duplex by NMR. For all three proteins, the mutations did not change their gross structure. For H-NS_{Ctd}, the Q112A/R114A double mutations essentially abolished DNA binding (data not included). The single Q112A or R114A mutations each reduced DNA binding, but the R114A mutation caused a much more pronounced effect than Q112A, suggesting a more important role for R114 in DNA binding. Similarly, for both the R97A/R99A mutant of Lsr2_{Ctd} and the R89A/R91A mutant of Bv3_{Ctd}, DNA binding activity was completely abrogated (data not included). The two Arg residues in Lsr2_{Ctd} appear to play equivalent roles; single R97A or R99A mutations each reduced DNA binding activity to a similar extent. These results demonstrate that the Q/RGR motif in all three proteins studied is critical for DNA binding.

5.3.8 DNA Minor Groove Binding Reagents Compete for the Binding of H-NS to DNA

To confirm that H-NS binds the minor groove of AT-rich DNA, we performed competition experiments using netropsin, a naturally occurring polypyrrolecarboxamide that binds to the minor groove of AT-rich DNA (Tabernero, Verdaguer et al. 1993). Addition of netropsin to the preformed H-NS_{Ctd}/DNA complex released H-NS_{Ctd} from the DNA; the NH signals of H-NS_{Ctd} residues in complex with DNA shifted back toward that of free H-NS_{Ctd}, and this effect is concentration-dependent (data not included). At a netropsin/DNA ratio of 2.5:1, the 2D 1H-15N HSQC spectrum is essentially identical to that of free H-NS_{Ctd}, indicating that netropsin completely displaces H-NS_{Ctd} from DNA. Similar results were found for Lsr2_{Ctd} (data not included). These data provide strong support for our conclusion that H-NS, Bv3F and Lsr2 bind to the minor groove of DNA.
5.3.9 Confirmation of the DNA-Binding Preference of H-NS by NMR

The PBM analysis indicates that H-NS and Lsr2 exhibit optimal binding activity for contiguous AT sequences that do not contain A-tracts. To substantiate this finding, we analyzed the binding of H-NS\textsubscript{Ctd}, Lsr2\textsubscript{Ctd}, and Bv3\textsubscript{Ctd} to several DNA substrates by NMR. Four DNA substrates were tested, including the CGCATATATGCG duplex, which was used throughout the experiments described above, a CGCATGCATGC duplex in which a “GC” was introduced at the center of the sequence, an A-tract containing sequence CGCAAAAAAAAAAGCG/CGCTTTTTGCG, and a CGCGCGCGCGCG duplex that is 100% GC. These DNA substrates caused similar NH chemical-shift change patterns in each protein and the extent of changes reflects the binding affinity. Consistent with our finding from the PBM experiments, NH signals of H-NS\textsubscript{Ctd} residues were most affected by the binding of the CGCATATATGCG DNA, followed by CGCAAAAAAAAAAGCG/CGCTTTTTGCG, CGCATGCATGC, and CGCGCGCGCGCG (data not shown). For Lsr2\textsubscript{Ctd} the A-tract containing DNA appears to be equivalent to the AT-rich sequence in binding affinity (data not shown). For Lsr2\textsubscript{Ctd} and Bv3\textsubscript{Ctd}, similar results were obtained except that the A-tract containing DNA appears to be equivalent to the AT-rich sequence in binding affinity (data not shown).

5.3.10 Correlation with \textit{in vivo} Binding Sites

Finally, we asked whether preferred sequences identified by PBMs \textit{in vitro} reflect genome binding sites of H-NS and Lsr2 \textit{in vivo}. The high-affinity binding sequence (TCGATATATT) of H-NS identified in the \textit{E. coli} proU operon (Wu, Yang et al. 2008) contains an 8-mer (GATATATT). Its complement strand sequence (AATATATC) is the 19th ranked 8-mer of the H-NS binding identified by PBM, with an E-score of 0.467. I scored the occurrences of each of the preferred 8-mers in all bound fragments identified by ChIP-seq or ChIP-chip analysis (Gordon, Li et al. 2010; Kahramanoglu, Seshasayee et al. 2011), and compared with that in randomly selected fragments. High-scoring 8-mers with E-score $\geq 0.45$ were used for H-NS analysis. Because of the high GC-content of the \textit{M. tb} genome, I used 8-mers with E-score $\geq 0.40$ for Lsr2 analysis. In both cases, I observed a significant enrichment for the preferred sequences in the neighborhood of bound fragments, with a peak close to the center (Figure 5.6A).
and B). Therefore, a large proportion of in vivo binding events apparently involve sequence preferences that can be derived from in vitro experiments. *Salmonella* has a genome-wide AT-content of 47.8%, whereas the mean AT-content of *M. tb* is 34.4%. I reanalyzed the ChIP-chip data by plotting the fraction of bound sequence (i.e., sequence with binding ratio $\geq 2$) at each AT-content. The fraction of the bound sequence begins to increase when the AT-content reaches $\sim 38\%$ for Lsr2 and $\sim 50\%$ for H-NS (Figure 5.6C), which corresponds to the mean AT-content of the corresponding genomes, respectively. Normalization of the data against the mean AT-content of the respective genome shows that H-NS and Lsr2 exhibit nearly identical binding when the AT-content of the bound sequence is compared relative to the AT-content of the whole genome (Figure 5.6D). It remains a question how each silencer targets sequences that are comparatively AT-rich with respect to their corresponding genomes. For example, H-NS avoids sequences with an AT-content of 50%, whereas Lsr2 targets such sequences as foreign (Figure 5.6C). Perhaps differences between the RGR and QGR motifs in binding affinity for less AT-rich sequences may provide an explanation. Thus, the RGR motif, used primarily in silencers (Lsr2) from bacteria with low AT-content ($\sim 34\%$) genomes (*Mycobacterium*), may enable tighter binding to sequences of lower AT-content, whereas the QGR motif, found in silencers (H-NS) from bacteria with higher AT-content ($\sim 50\%$) genomes (*Salmonella and Escherichia*), may lower the affinity of these proteins for DNA that is only mildly AT-rich. Alternatively, however, the relative abundance of AT-rich sequences in these genomes and the competitive advantage of preferred sequences for recognition by silencers may suffice as an explanation.
Figure 5.6. Correlation with in vivo binding sites. (A and B) Enrichment of sequences preferred in vitro within genomic sequences bound in vivo by the same protein. Comparison of bound to randomly selected sequences for Lsr2 (A)(Gordon, Li et al. 2010) or H-NS (B)(Kahramanoglou, Seshasayee et al. 2011) showing the relative enrichment of high-scoring 8-mers (cutoff: $E \geq 0.45$ and $E \geq 0.40$ for H-NS and Lsr2, respectively). $P$ value was calculated for the interval (−1,000 bp to +1,000 bp) by the Wilcoxon-Mann-Whitney rank sum test, comparing the number of occurrences per sequence in the bound set versus the background set. (C and D) Reanalysis of the ChIP-chip data. The AT-content of individual 60-mers used in the tiling array for ChIP-chip analysis of Lsr2 (Gordon, Li et al. 2010) and H-NS (Navarre, Porwollik et al. 2006) were calculated, respectively. The fraction of sequences that exhibit ≥ twofold binding ratio was calculated for each AT-content and plotted against the AT-content. (D) The AT-content from (C) was normalized against the mean AT-content of the respective genomes: 47.8% AT for the Salmonella genome and 34.4% AT for M. tb genome.
5.4 Discussion

We find that structural information contained within the DNA minor groove is deciphered by nonspecific DNA binding proteins to distinguish genetic material that is self from nonself. In this study, we have performed a comprehensive analysis of the DNA binding specificity and binding mechanism of both the Lsr2 and H-NS proteins. We present multiple lines of evidence that these proteins, unrelated in sequence and structure, use a common mechanism to interact directly with the minor groove of DNA, and propose that the specific geometry of the DNA minor groove largely dictates the degree of binding. Our conclusions provide a unified molecular mechanism to explain various findings related to xenogeneic silencing proteins, including their lack of apparent sequence specificity and their ability to bind even nonoptimal (i.e., GC-rich) sequences with affinities only an order of magnitude less than some preferred sites (Wu, Yang et al. 2008). It remains to be determined whether other H-NS analogs, such as the MvaT/MvaU of Pseudomonas (Tendeng, Soutourina et al. 2003; Castang, McManus et al. 2008) and Rok of Bacillus (Smits and Grossman 2010), exploit a similar binding mechanism.

The PBM analysis provided a comprehensive and unbiased approach for analyzing sequence specificity at high resolution. Insights gained from the PBMs allowed us to generate predictive models regarding the specific structural features within AT-rich DNA that are preferentially targeted by H-NS and Lsr2. The PBM data suggest that the shape of DNA, specifically that of the DNA minor groove, is a primary determinant of H-NS and Lsr2 binding specificity. The presence of GC base pairs in the center of AT-rich 8-mer sequences is unfavorable for H-NS and Lsr2 binding, which is likely because of the less optimal electrostatic potential for binding to arginine residues (Rohs, West et al. 2009). The presence of a 2-NH₂ group on G that protrudes into the minor groove may also disturb the binding of H-NS and Lsr2 (Badaut, Williams et al. 2002). The effect of A-tracts on the binding preference of H-NS and Lsr2 depends on the AT-content of the flanking sequences. A-tracts have a unique structure, which is distinct from that of B-DNA and is cooperatively formed whenever there are four or more adjacent As or Ts (Haran and Mohanty 2009). In contrast to typical B-DNA, where bases are perpendicular to the helical axis and thus have a wider minor groove, the bases within the A-tracts are negatively inclined relative to the global helical axis, and are highly propeller twisted. This propeller-twisted conformation results in a highly narrow minor groove. Thus, H-NS and
Lsr2 exhibit low binding affinity for DNA sequences with a minor groove width that is too wide (i.e., GC-rich sequences) or too narrow (i.e., A-tract sequences within the context of an AT-rich region), and exhibit the highest affinity for DNA sequences with an ideal minor groove width (i.e., mixed AT-rich sequences or a short A-tract within a GC-rich sequence). The binding of H-NS and Lsr2 to the DNA minor groove is further supported by the demonstration that netropsin, a DNA minor groove-binding molecule (Tabernero, Verdaguer et al. 1993), effectively outcompeted H-NS, Lsr2 and Bv3F for binding to an AT-rich sequence. An earlier study using distamycin, another minor groove-binding compound, showed similar results but was interpreted to be the result of disruption in DNA curvature (Yamada, Muramatsu et al. 1990). The DNA substrate whose binding to H-NS, Lsr2, and Bv3F was successfully competed with netropsin in our study, however, does not contain significant intrinsic curvature (Yoon, Prive et al. 1988). It is likely that DNA curvature is not used for recognition by this class of proteins, even though recognition is for an AT-rich minor groove. This also appears to be the case for the C-terminal domain of the α-subunit of RNA polymerase, which also recognizes AT-rich DNA in the minor groove (Aiyar, Gourse et al. 1998; Ross, Ernst et al. 2001; Benoff, Yang et al. 2002).

H-NS binds DNA in a cooperative manner, generating extended nucleoprotein filaments (Donato, Lelivelt et al. 1997; Falconi, Colonna et al. 1998; Arold, Leonard et al. 2010), but the relatively short PBM probes presumably detect binding of individual monomers, which, for H-NS, is relatively weak compared with the affinities of most sequence-specific DNA-binding proteins for most sites (Wu, Yang et al. 2008). However, for both H-NS and Lsr2, the preferred monomer-binding 8-mer sequences identified by PBM experiments in vitro are enriched at the center of the genomic fragments bound by the same protein in vivo. Therefore, the preferred monomer binding sequences are likely to be a component of the targeting mechanism, and may serve as initiation sites for nucleation of H-NS to form higher-order nucleoprotein structures.

Our structural analysis of H-NS, Lsr2, and Bv3F reveals that all three proteins bind the minor groove using a loop containing a Q/RGR motif. The high-quality structures for each protein’s DNA binding domain enabled the construction of docking models that implicate the Q/RGR motif as the primary DNA interacting site. Accordingly, engineering mutations where the Q/RGR motif is replaced with “AGA” abolished the DNA binding activity of all three proteins. This mode of binding is reminiscent of the eukaryotic chromatin protein HMGA. HMGA proteins are required for the assembly of higher-order transcription enhancer complexes.
critical for the transcriptional activation of a number of important genes involved in diverse cellular processes (Reeves and Beckerbauer 2001). Each HMGA protein possesses a set of three AT hooks, which are unstructured while free in solution. However, each assumes a planar, crescent-shaped conformation when bound to DNA, which is dictated by the shape of a narrow minor groove of the AT-rich DNA substrate (Huth, Bewley et al. 1997). As such, HMGA proteins recognize substrate structure rather than nucleotide sequence, and the presence of multiple AT-hook peptides is necessary to confer high-affinity binding (Huth, Bewley et al. 1997). Our finding that H-NS and Lsr2 contain an AT-hook-like motif with a similar conformation in their respective DNA-binding domain (Figure 5.4) provides a molecular mechanism to explain these proteins’ preferential binding to AT-rich DNA sequences.

Compared with Gln, the positively charged Arg is more likely to bind negatively charged DNA. This finding may explain why Gln is less critical for DNA binding than Arg in the QGR motif of H-NS, and why Lsr2 and Bv3F have apparently higher affinity for less AT-rich sequences than H-NS. The second Arg residue of the Q/RGR motif is conserved in all three proteins, and site-directed mutagenesis analysis shows that it is critical for DNA binding. Our data agree with recent findings that several classes of transcription factors bind to the minor groove of AT-rich sequences by inserting Arg residues into the minor groove, presumably because of their ability to form favorable electrostatic interactions with the relatively narrow groove formed by these sequences (Rohs, West et al. 2009; Rohs, Jin et al. 2010). These studies reveal that common mechanistic themes underlie DNA recognition in the absence of “typical” highly sequence-specific interactions that occur via contacts with bases in the major groove.
5.5 Experimental Procedures

Molecular Cloning and Protein Expression for PBM experiments. The ORF of *lsr2* from *Mycobacterium tuberculosis* H37Rv was PCR-amplified with the forward primer 5′-AAAAGAGCTCATGGCGAAGAAAGTAACCGTCACCT-3′ and reverse primer 5′-TTTTGGATCCTCAGGTCGCCGCTTATGCGTCG-3′). The *hns* gene of *Salmonella enterica* serovar Typhimurium LT2 was PCR-amplified using the forward primer 5′-AAAAGAGCTCATGAGCGAAGCACTTAAAATTCTGA-3′ and reverse primer 5′-TTTTGGATCCTTATTCCTTGATCAGGAAATCTTCC-3′. The respective PCR products were digested with SacI and BamHI and cloned into the T7-GST vector pTH5325 (pDEST15-Magic_modified) pretreated with the same enzymes. Proteins for PBM experiments were expressed via *in vitro* translation (IVT) reactions (PURExpress; New England BioLabs) without purification. The GST-tag (at the N terminal of the fusion protein) is necessary for the detection of binding signals in the protein binding microarray (PBM) experiments (see below).

PBM Experiments. The design of “all 10-mer” universal PBMs using a de Bruijn sequence of order 10 has been described previously (Berger, Philippakis et al. 2006) and is described in detail in a separate publication (Philippakis, Qureshi et al. 2008). A de Bruijn sequence of order *k* is a circular string of length 4^k^ that contains every *k*-mer exactly once when overlaps are considered, which can be generated by linear feedback shift registers (Berger, Philippakis et al. 2006). The two de Bruijn sequences for our PBMs differ by cyclic permutations of A, C, G, and T. The exact sequences of the two PBM “all 10-mer” designs (designated ME and HK) used in our study have been described previously (Lam, van Bakel et al. 2011).

PBM experiments were conducted and analyzed as previously described (Berger, Philippakis et al. 2006; Berger, Badis et al. 2008; Badis, Berger et al. 2009; Lam, van Bakel et al. 2011). Briefly, single-stranded oligonucleotide microarrays were double-stranded by primer extension. Before protein incubation, double-stranded microarrays were first blocked with PBS/2% skim milk and then washed. Protein from the IVT reaction was diluted to 20 ng/μL in a final 150-μL reaction mixture containing: 2% milk, 51.3 ng/μL salmon testes DNA, 0.2 μg/μL BSA, 50 μM zinc acetate, and 0.1% Tween-20. The assembled microarrays were incubated with protein binding mixtures for 1 h at 20 °C. Microarrays were then washed and incubated with Alexa488-conjugated rabbit polyclonal antibody to GST (50 μg/mL in PBS, 2% milk) for 1 h at
20 °C. Finally, microarrays were washed and dried. Protein-bound microarrays were scanned to detect Alexa488-conjugated antibody using an Agilent microarray scanner. Image spot intensities were quantified using ImaGene software (BioDiscovery).

**PBM Z-Score and E-Score.** Every nonpalindromic 8-mer occurs 32 times in each of our PBM arrays. Each protein was tested on two different universal microarrays (designated ME and HK). Because of this redundancy, we are able to provide a robust estimate of the relative preference of a DNA binding protein for every contiguous 8-mer. To estimate the relative preference for each 8-mer, two different scores, the Z-score and E-score, were calculated as described previously (Berger, Philippakis et al. 2006). Briefly, the distribution of log (median intensity) over all 8-mers is used to compute a Z-score for each 8-mer according to the following formula:

\[
\text{Z-score} = \frac{\log \text{median intensity of specific 8-mer} - \log \text{median intensity of all 8-mers}}{\text{robust estimate of standard deviation}}
\]

Here, median intensity refers to the median normalized signal intensity for the set of probes containing a match to each 8-mer (32 probes for nonpalindromic 8-mer or 16 probes for palindromic 8-mer). The robust estimate of the SD is the median absolute deviation, multiplied by 1.4826 for normally distributed data.

It was shown previously that higher PBM median signal intensity corresponds to stronger protein-DNA binding affinity (Berger, Philippakis et al. 2006). Therefore, both the median signal intensity and Z-score are advantageous because they retain information regarding relative differences in signal intensity, and thus probe occupancy and relative affinity as well. However, experimental variability and differences in absolute signal intensities and nonspecific binding can make these measures difficult to compare for different transcription factors.

The E-score is a rank-based, nonparametric statistical measure that is invariant to protein concentration and readily allows different experiments to be compared on the same scale. This enrichment score is a variation on area under the ROC curve and has been described previously in detail (Berger, Philippakis et al. 2006). Briefly, for each 8-mer we consider the collection of all probes containing a match as the “foreground” feature set and the remaining probes as the “background” feature set. We compare the ranks of the top half of the foreground with the ranks of the top half of the background by computing a modified form of the Wilcoxon-Mann-Whitney
statistic scaled to be invariant of foreground and background sample sizes. The E-score ranges from +0.5 (most favored) to −0.5 (most disfavored).

**Comparison Between PBM Data and ChIP-chip or ChIP-seq Data.** I analyzed the ChIP-chip dataset for Lsr2 or ChIP-seq dataset for H-NS available in the literature (Gordon, Li et al. 2010; Kahramanoglu, Seshasayee et al. 2011). Bound sequences were scanned to determine enrichment of highly-bound PBM 8-mers. To standardize the length of the bound sequences across the datasets, I took the center of the ChIP peak and added 1 kb on each side. Enrichment ratio was determined with respect to 2-kb length random genomic regions taken from the genome (same version as the one used for the chip experiments) of the corresponding organism. For each dataset, the number of random genomic regions sampled was 10 times the number of bound sequences. Random and bound sequences were scanned with the PBM E-score 8-mer profile using a 250-bp length moving window with a 50-bp tiling distance. Enrichment ratio was calculated as the ratio of the total number of 8-mers above the cutoff value (E ≥ 0.45 for H-NS and E ≥ 0.40 for Lsr2) found in the bound sequences to the number of 8-mers above the cutoff value found in the random genomic sequences. P values were calculated for the interval (−1,000 bp to +1,000 bp) by the Wilcoxon-Mann-Whitney rank sum test, comparing the number of occurrences per sequence in the bound set versus the background set. A modified perl script written by Lourdes Pena Castillo was used for this analysis.

**Molecular Cloning and Generation of Chimeric Proteins.** The construct pLSR2-HA used for expression of HA-tagged Lsr2 in *E. coli* and mycobacteria was described previously in Chapter 2. The *lsr2* gene was expressed under its endogenous promoter, and the 200-bp upstream region of the start codon together with the ORF of *lsr2* was cloned into pNBV1 to generate pLSR2-HA. Six pNBV1-based chimeric constructs (pC1–C6) composed of the Lsr2 of *M. tb* H37Rv and the H-NS of *Salmonella enterica* serovar Typhimurium were generated by a two-step PCR strategy. Briefly, the N-terminal parts of the *lsr2* and *hns* ORFs together with their own upstream promoter regions (200 bp for *lsr2* and 838 bp for *hns*) were amplified by PCR using pLSR2-HA and pHNS-HA as templates with appropriate primers. Similarly, the C-terminal regions containing an HA tag were PCR amplified using the same templates. The N- and C-terminal fragments were then pieced together by a secondary PCR using specific primers, giving rise to
fusion DNA fragments, which were subsequently cloned into the XbaI and HindIII sites of the vector pNBV1. All constructs were confirmed by DNA sequencing.

**Expression and Purification of Recombinant Protein for Domain Mapping Experiments.** The expression constructs pN1–65 and pC51–112 encoding the N-terminal (residues 1–65) and C-terminal (residues 51–112) regions of Lsr2 were generated by PCR amplification using appropriate primers and cloned into the XhoI and HindIII sites of pET21d (Novagen). The construct pLSR2 (Chen, Ren et al. 2008) was used to express the full-length Lsr2 protein of *M. tb*. These plasmids were transformed into *E. coli* BL21 (DE3), and protein expression was induced with 1mM IPTG at OD600 = 0.6 for 3 h at 37 °C. The recombinant proteins were purified using Ni-NTA columns by standard methods (Qiagen). The purity of the recombinant proteins was estimated to be >95%. The procedures for in vitro cross-linking experiments and gel-mobility shift assays were described previously (Chen, Ren et al. 2008).

**Chromatin Immunoprecipitation and Quantitative Real-Time PCR.** Cultures of *E. coli* WN582 harboring one of the chimeric protein expression constructs, pC1–C6, grown to OD600 = 0.6 were treated with formaldehyde, and protein–DNA complexes were immunoprecipitated using an anti-HA antibody as previously described in Chapter 2. Quantitative real-time PCR analyses were performed for four H-NS-bound genes (*bglG, proV, xapR, yjcF*) and two negative control genes (*narZ and phnE*).

**Structure determination, NMR Titration Experiments, and Docking Model.** These experiments were performed by Bin Xia and colleagues at Peking University in Beijing China. The experimental procedures for this work can be found in the supplementary materials for (Gordon, Li et al. 2010; Gordon, Li et al. 2011).
CHAPTER 6: SUMMARY AND FUTURE DIRECTIONS
6.1 Lsr2 is functionally equivalent to H-NS

I have established through in vivo complementation assays that Lsr2 of mycobacteria and related actinomycetes is functionally analogous to H-NS. The functional similarity of Lsr2 and H-NS is further explained by the finding that Lsr2 and H-NS share the same overall domain composition: the N-terminal domain is involved in the formation of higher order oligomers while DNA binding activity is confined to the C-terminal domain (Chen, Ren et al. 2008). In fact these domains can be swapped between Lsr2 and H-NS to generate functional chimeric molecules despite the fact that these two proteins share <20% identity in amino acid sequences and exhibit different tertiary structures. Work by our collaborator Bin Xia’s laboratory at Peking University has revealed that H-NS and Lsr2 use a common mechanism to selectively target AT-rich DNA. Both proteins possess a prokaryotic AT-hook motif (‘RGR’ in Lsr2 and ‘QGR’ in H-NS) located within a loop region of both proteins that inserts into the minor groove of AT-rich DNA. Nevertheless, despite our progress in determining the molecular mechanisms governing the functional equivalence of these proteins, further questions regarding Lsr2 biology remain to be answered and the following are some suggestions I have for addressing them.

H-NS is one of the most abundant nucleoid-associated proteins in E. coli with approximately 20 000 copies per genome equivalent. H-NS protein levels are thought to remain constant, however, hns expression may increase under certain stress conditions (Dorman 2004). Determining the cellular concentration of Lsr2 protein in M. tb as it progresses through its entire growth phase (early log to stationary phase) as well as during various stress conditions will provide further insight into the biological role of Lsr2. Purified Lsr2 protein can be used to generate anti-Lsr2 antibodies, or alternatively they could be purified from TB patient serum since anti-Lsr2 antibodies are known to be produced (Laal, Sharma et al. 1991). I have found that Lsr2 negatively regulates genes involved in stress response, persistence and virulence. I predict that under conditions of stress or during host infection, Lsr2 protein levels will decrease allowing for derepression of these loci, which would represent a key difference between Lsr2 and H-NS.

H-NS has been shown to form heterologous protein-protein interactions. For example, the Hha-family of H-NS accessory molecules physically interact with H-NS (Nieto, Madrid et al. 2002) to significantly enhance repression at many horizontally acquired genomic islands.
(Navarre, Porwollik et al. 2006; Vivero, Banos et al. 2008). H-NS also forms heterodimers with its paralogue StpA and physically interacts with the FlIG flagellar motor protein (Donato and Kawula 1998) as well as the RNA-binding protein Hfq (Kajitani and Ishihama 1991). Coimmunoprecipitation along with mass spectrometry on Lsr2 should be performed to identify any Lsr2-interacting proteins, which would expand our knowledge of Lsr2’s biological role. It also remains to be determined exactly how Lsr2 blocks gene expression as well as how it is countersilenced. Based upon Lsr2’s functional equivalence to H-NS, it is assumed that Lsr2 binding regulates gene expression by simply occluding or trapping RNA polymerase (RNAP) at transcriptional start sites. Antagonizing H-NS repression at many loci is mediated by different transcription factors with specific DNA binding activity that disrupts H-NS-DNA interactions (Navarre, McClelland et al. 2007). The identities of specific Lsr2 antagonist proteins remain enigmatic. I noticed in my microarray data that many genes that are activated by phoP are repressed by Lsr2 suggesting that phoP may be an Lsr2 countersilencer. As well, the ability of RNAP to transcribe some H-NS silenced genes depends upon the σ factor with which RNAP is associated. Candidate Lsr2 countersilencers and alternative σ factors can be tested for their ability to disrupt Lsr2-DNA interactions using EMSAs. In vitro transcription assays could be used to determine their ability to drive transcription from Lsr2 repressed promoters.

I performed a BLAST search of Lsr2 against the mycobacteriophage database (http://phagesdb.org/) which revealed the presence of Lsr2 homologs present within a number of recently sequenced mycobacteriophage genomes (Figure 6.1). Given what is known about the role of Lsr2 in mycobacteria, it is perplexing as to why these phages also harbor an lsr2 gene. Did the phages acquire lsr2 from a progenitor mycobacterial cell or conversely is lsr2 of phage origin? The amino acid identity of the phage Lsr2 proteins with the M. tb Lsr2 protein ranges from 34-42%. The prokaryotic AT-hook ‘RGR’ motif is conserved in some of these homologs or the slight variation ‘KGR’ also occurs and would be expected to be functionally similar. Nevertheless, there are some Lsr2 homologs with the motif ‘RGY’ (Figure 6.1) whose DNA binding ability may be greatly altered or reduced. Furthermore, a number of the homologs have an aspartate residue immediately in front of the first residue of the AT-hook motif whereas in the mycobacterial Lsr2 the same residue is a threonine. It is possible that the presence of the negatively charged aspartate could lead to electrostatic repulsion with the negatively charged
DNA phosphate backbone which would abolish binding activity. To address these questions, I propose cloning these \textit{lsr2} phage genes into the pNBV1 plasmid and assaying for their ability to complement the \textit{M.smeegmatis} \textit{Δlsr2} colony morphology like I did in Chapter 2. If they are unable to complement the phenotype, the phage \textit{lsr2} genes should be expressed in a WT background to assess whether they are dominant negative mutants, indicated by a conversion to an\textit{Δlsr2}-like colony morphology. H-NS and Lsr2 mutants with a functional oligomerization capacity but with an impaired DNA binding ability exhibit a dominant negative phenotype when introduced into a WT background (Williamson and Free 2005; Banos, Pons et al. 2008; Chen, Ren et al. 2008), presumably because they interfere with the ability to dimerize and form higher-order complexes. Thus since Lsr2 is known to target foreign DNA, it is surmisable that the phage Lsr2 proteins function as antagonists of Lsr2 mediated repression to allow for phage gene expression. In fact, H-NS-like proteins which function as countersilencers against the endogenous H-NS have been identified within horizontal genomic islands in the enteric bacteria (Mellies, Elliott et al. 1999; Williamson and Free 2005; Muller, Schneider et al. 2010). The phage \textit{lsr2} genes could also be deleted from their respective genomes and the mutant phages could be assayed for plaque formation. If indeed they are functioning as host Lsr2 antagonists, I would expect the phage \textit{lsr2} mutants to be defective for plaque formation.
Figure 6.1. Sequence alignment of Lsr2 proteins from *M. tb*, *M. smegmatis*, and various *Mycobacterium* phages. The alignment was produced using the CLUSTALW. Species abbreviations are as follows: *M. tb*, *Mycobacterium tuberculosis*; *M. smegmatis*, *Mycobacterium smegmatis*; JoeDirt, *Mycobacterium* phage JoeDirt; LeBron, *Mycobacterium* phage LeBron; Faith, *Mycobacterium* phage Faith; Optimus, *Mycobacterium* phage Optimus; LittleE, *Mycobacterium* phage LittleE; Pumpkin, *Mycobacterium* phage Pumpkin; Rey, *Mycobacterium* phage Rey; Bongo, *Mycobacterium* phage Bongo. The ‘RGR’ motif occurs at positions 112-114 in this alignment.
6.2 Lsr2 negatively regulates genes required for persistence

I have demonstrated that Lsr2 negatively regulates a significant portion of the genes that are induced during various models of persistence. The Δlsr2 mutant of *M. tb* bears the transcriptional hallmarks of non-replicating persistent bacilli thought to be present during latent TB disease. Furthermore, evidence from *M. smegmatis* indicates that abrogation of *lsr2* is beneficial for long term viability during hypoxia (Chapter 4). I hypothesize that Lsr2 represses a stress response program that initiates persistence as a survival strategy during various stress conditions, and that this genetic program is essential for maintaining bacterial persistence during latency (Figure 6.2). In my working model, under normal conditions Lsr2 is abundant in the cell; however, when persistence-signaling conditions (i.e. hypoxia) are experienced, Lsr2 levels are depleted in order to allow for the induction of these Lsr2-repressed systems (Figure 6.3). Moreover, I found that nearly 10% of the 540 genes repressed by Lsr2 are transcriptional regulators (Chapter 4). I suspect that a number of the Lsr2-repressed transcription factors may act as Lsr2 countersilencers to further amplify the genetic response to environmental cues promoting NRP.

I predict that the Lsr2 regulon is regulated in an analogous fashion as the SOS response is regulated by LexA. In the case of the SOS response when DNA damage occurs it stimulates the proteolysis of LexA, lowering its concentration, allowing for LexA-repressed DNA repair mechanisms to be expressed (Shinagawa 1996; Butala, Zgur-Bertok et al. 2009). The degree of the response can be adjusted so that high-fidelity repair mechanisms are employed first before error-prone repair mechanisms. LexA has different affinities for the various operator/promoter sequences so that the more ‘dangerous’ error-prone repair system genes are turned on last because LexA has a higher affinity for these genes. Like the SOS response, I predict that expression of the Lsr2 regulon would be modular, depending upon the cellular concentration of Lsr2 protein. In Chapter 5 I found that Lsr2 binding affinity correlates with AT content, and that there are additional factors which also influence Lsr2’s binding affinity (e.g., there are a wide range of binding affinities for all 8-mers with the same % AT). Comparing my PBM binding data with my ChIP-chip data, I found that the preferred sequences identified by PBM
vitro reflect genome binding sites of Lsr2 in vivo. Therefore as cellular concentrations of Lsr2 decrease, there would be a hierarchy of gene expression from the Lsr2 regulon which would depend upon the relative binding affinity of Lsr2 for each gene. In *M. tb* Lsr2 protein levels could be reduced by several means including genetic inactivation via phase variation, induction of an *lsr2* transcriptional repressor, or proteolysis of the Lsr2 protein. I propose several lines of experimentation to explore the role of the Lsr2 regulon during persistence.
Figure 6.2. (previous page) Lsr2’s regulatory role in a model for human *M. tb* infection proposed by (Chao and Rubin 2010). I have added my proposed roles of Lsr2 in this model (*lsr2* regulon genes highlighted in green represents Lsr2-regulated genes). *M. tb* enters a host and establishes a primary infection, which is dominated by actively dividing cells. Some dormant cells can be generated through the induction of toxin-antitoxin (TA) loci or other factors that alter metabolism (this could be mediated by inactivation of *lsr2*, leading to the activation of TA systems and alternative metabolic and respiratory routes) Cell-mediated immunity clears out actively dividing cells preferentially, leaving a predominantly dormant population (persisters) to account for the paucibacillary latent state. Standard chemotherapy will also lead to persisters which are largely responsible for disease relapse (drug tolerance mediated by inactivation of *lsr2* would contribute to the formation of these persisters). A similar state can be achieved through *in vitro* studies using a variety of host stress conditions. Treatment with isoniazid, which targets actively dividing bacteria, reduces reactivation risk by targeting the resuscitated scouts. Rifampin or other dormancy-active antibiotics may help clear the remaining dormant subpopulation and lead to tissue sterilization.
I hypothesize that during unfavourable conditions within the host (e.g., hypoxia), repression by Lsr2 is lost either by $lsr2$ gene inactivation and/or Lsr2 proteolysis. This in turn leads to activation of genes involved in fatty acid metabolism, toxin-antitoxin systems, as well as genes important for persistence in the host. Activation of these systems induces the formation of $M. tb$ persisters, which is characterized by reduced or altered metabolic activity (e.g., preference for host cholesterol as carbon source), enhanced drug tolerance, and the long-term persistence in the host. These persisters survive in the host at a reduced level and cause less pathology, thereby prolonging the survival of host. These persisters lead to latent infection (Figure 6.2).
6.2.1 Does the $\Delta$lsr2 mutant of \textit{M. tb} exhibit a fitness advantage over the WT strain under hypoxia?

The fitness advantage of the \textit{M. smegmatis} $\Delta$lsr2 during long-term hypoxia as well as the lsr2-inactivation phenomenon are compelling findings with broad implications for mycobacterial biology. Nevertheless, \textit{M. smegmatis} is non-pathogenic making it an imperfect model for studying \textit{M. tb} pathogenesis; thus, similar studies should be performed with \textit{M. tb}. I have performed a preliminary experiment comparing the survival of the \textit{M. tb} WT, $\Delta$lsr2 and the complemented strains during hypoxia-induced NRP. I initially observed growth for all 3 strains which peaked around day 20, and was then followed by a steady decrease in CFUs (Figure 6.4). Interestingly the $\Delta$lsr2 mutant appears to outcompete the WT, with $\sim$2 logs higher CFUs observed at 50 days. However, the time course is too short compared to that of \textit{M. smegmatis} since the doubling time of \textit{M. tb} is much longer than that of \textit{M. smegmatis} (18-24 hours versus 4 hours). A more appropriate time course for this experiment would be $>$600 days. During this experiment a thorough effort should be undertaken to determine whether WT \textit{M. tb} undergoes spontaneous lsr2-inactivation under hypoxia. In the \textit{M. tb} genome IS6110 is the most common mobile element and is frequently associated with spontaneous insertions (McEvoy, Falmer et al. 2007). Candidate lsr2 mutants could be confirmed by performing southern blot and PCR sequencing analysis using probes specific to IS6110 and lsr2. In the event that lsr2 mutations cannot be identified, the following experiments should be performed. It is possible that epigenetic mutations have resulted in a loss of lsr2 expression (e.g., an lsr2 activator has been mutated or an lsr2 repressor is upregulated). Alternatively, Lsr2 may be proteolysed. To distinguish between the two possibilities, RT-PCR and western blot analysis should be performed to determine lsr2 transcript and protein levels, respectively. Transposon mutagenesis studies of an \textit{M. tb} strain containing the lsr2 promoter fused to a lacZ reporter could be performed to identify repressors and/or activators of lsr2 transcription. The mutants would be cultured on media containing X-gal, which would allow for colour-based screening for mutations affecting lsr2 expression. Mutations in lsr2 repressors would be dark blue whereas mutations in lsr2 activators would be white. In addition it would be informative to perform a similar long-term hypoxia competition experiment as the one with \textit{M. smegmatis} to determine if the $\Delta$lsr2 mutant of \textit{M. tb} also exhibits a fitness advantage over the WT strain under long-term hypoxia. This could be performed using the \textit{M. tb} WT strain I generated harbouring a plasmid with a
kanamycin resistance marker and the Δlsr2 mutant, which contains a hygromycin resistance cassette.

I found significant overlap between genes upregulated in Δlsr2 and the genes activated in *M. tuberculosis* during several models of persistence. However, the highest degree of overlap was observed with the hypoxia-induced NRP, suggesting to me that the mutant is predisposed to surviving long-term hypoxia. Abrupt anaerobiosis can be induced by the copper-activated iron wool method (Parker 1955) and was shown to be lethal to actively growing *M. tuberculosis* cultures (Wayne and Lin 1982). I predict that unlike WT, the mutant would exhibit tolerance to abrupt anaerobiosis, which would be observed as enhanced survival compared to the WT.
Figure 6.4. The \( \Delta lsr2 \) of \( M. \ tb \) exhibits comparable viability to the WT strain under long-term hypoxia. The WT, \( \Delta lsr2 \) and complemented strains were each incubated under hypoxia using the Wayne model for 100 days. The viable bacteria counts at different time points were determined by plating and numerating CFU for 3 independent tubes per strain. Each point represents the mean ± SD.
6.2.2 Does the $\Delta lsr2$ mutant of $M. tb$ display altered carbon metabolism and cellular redox energy status?

I found that three pathways related to cholesterol catabolism are activated in the $\Delta lsr2$ mutant: the β-oxidation of cholesterol, the methylcitrate cycle, and the methylmalonyl pathway. My microarray data indicates that nearly all genes involved in these three pathways are overexpressed in the $\Delta lsr2$ mutant suggesting that it is better suited to subsisting from host-derived lipids versus carbohydrates. This may explain why the mutant exhibits a growth defect in vitro since glycerol is supplied as the carbon source. The $\Delta lsr2$ mutant should be tested for its ability to grow in modified 7H9 broth containing cholesterol, propionate, or valeric acid (a fatty acid with an odd number of carbons) as the sole carbon source to determine if it grows better on these lipids than WT. The finding that $\Delta lsr2$ prefers cholesterol for growth would be significant since host cholesterol is required for $M. tb$ persistent infection (Pandey and Sassetti 2008; Chang, Miner et al. 2009; Yam, D'Angelo et al. 2009; Hu, van der Geize et al. 2010; Nesbitt, Yang et al. 2010; Griffin, Gawronski et al. 2011; Yang, Gao et al. 2011).

A crucial feature in the adaptation of $M. tb$ to alternative energy sources (e.g., cholesterol) and changing environmental conditions (e.g., reduced oxygen tension) is the balance of oxidative and reductive reactions in the metabolic scheme. My microarray data seems to suggest that the $\Delta lsr2$ may have realigned its redox and energy status to favor the host environment during persistent infection. To confirm this, the NAD$^+$/NADH ratio and ATP level of the $\Delta lsr2$ and WT should be ascertained during hypoxia. Switching from aerobic conditions to hypoxia is associated with an initial decrease of the NAD$^+$/NADH ratio in $M tb$, which is recovered at a later time and maintained (Leistikow, Morton et al. 2010; Watanabe, Zimmermann et al. 2011). I anticipate that the $\Delta lsr2$ would have a shorter adjustment period than WT and achieve a stable NAD$^+$/NADH level more rapidly than the WT during the transition. The intracellular ATP level of $M. tb$ also drops in response to reduced oxygen tension and stabilizes at 25% of aerobic levels (Leistikow, Morton et al. 2010; Watanabe, Zimmermann et al. 2011), which is adequate to sustain the persisters in NRP. The $\Delta lsr2$ mutant may adjust the ATP level more rapidly than WT under these conditions. There are published protocols for these measurements in mycobacteria (Leistikow, Morton et al. 2010). Briefly, nucleotides are extracted using either acid or base and measured by using an alcohol dehydrogenase-based
NAD⁺-NADH cycling assay. ATP is extracted using a chloroform heat-based method and measured by a luciferase-based ATP assay (Promega)(Leistikow, Morton et al. 2010).

Isocitrate lyase activity has been shown to be important for M. tb persistence in vivo, allowing for carbon retention when using host lipids as a carbon source (McKinney, Honer zu Bentrup et al. 2000). I found the icl1 gene was significantly upregulated in the mutant. I propose testing the isocitrate lyase enzymatic activity of the Δlsr2 and WT cell lysates using previously described methods (Roche, Williams et al. 1970; Wayne and Lin 1982). In turn the enzyme involved in the reentry of glyoxylate into the TCA is malate synthase, which catalyzes the condensation of glyoxylate with the acetate from acetyl-CoA. However, in M. tb glyoxylate can be utilized via another pathway involving the reductive amination of glyoxylate by glycine dehydrogenase with the concomitant oxidation of NADH to NAD⁺. Interestingly hypoxic M. tb cultures were reported to have a 10-fold increase in glycine dehydrogenase activity (Wayne and Lin 1982) and I found that the gene encoding for glycine dehydrogenase, gcvB, is significantly upregulated in Δlsr2(2.6-fold, q-value = 0.3%). I predict that the increased glyoxylate synthesis in the mutant serves to provide a substrate for the regeneration of NAD⁺, especially when oxygen is not available as a terminal electron acceptor. The mutant lysates can be tested for glycine dehydrogenase activity using previously described methods (Goldman and Wagner 1962). Another enzyme that contributes to maintaining the NAD⁺ pool is alanine dehydrogenase (encoded by ald) which recycles NADH through the conversion of pyruvate to alanine when oxygen, as a terminal electron acceptor becomes limiting (Hutter and Dick 1998). Alanine dehydrogenase activity is reported to increase in anaerobic M. smegmatis cultures(Hutter and Dick 1998; Usha, Jayaraman et al. 2002) and I found the ald gene is significantly upregulated (5.4-fold, q-value = 0.0%) in the M. tb Δlsr2 mutant. The alanine dehydrogenase activity of the M. tb and Δlsr2 should be determined using previously described methods (Ohashima and Soda 1979).

6.2.3 Is the Δlsr2 mutant predisposed to persister formation?

My transcriptional analysis indicates that the Δlsr2 mutant transcriptome bears key signatures of persisters such as activated TA system expression and downregulated ribosomal gene expression. Virtually all pathogens examined thus far, including M. tb, produce dormant
persister cells in growing exponential cultures (Keren, Kaldalu et al. 2004; Keren, Minami et al. 2011). The presence of drug-tolerant persisters in a growing bacterial culture is indicated by a biphasic killing pattern following antimicrobial exposure. To determine the role of Lsr2 in persister formation I would compare the persister formation ability of the WT and ∆lsr2. In addition, I would construct a strain with constitutive lsr2 expression using the pMV261 mycobacterial expression vector which uses an hsp60 promoter (Stover, de la Cruz et al. 1991). I would test the respective persister formation levels of these three strains over a 15 day period across an array of antibiotics with different mechanisms of action. These would include the first-line drugs isoniazid, ethambutol, pyrazamide, rifampicin and D-cycloserine which was used by Keren et al. for M.tb persister transcriptome determination (Keren, Minami et al. 2011). I predict that the ∆lsr2 would have several logs higher levels of persisters compared to WT, whereas the strain overexpressing lsr2 would be defective for persister formation. In fact my preliminary data showed that after exposure to 1 µg/ml of isoniazid, rifampin, or ciprofloxacin for 4 days, the numbers of isoniazid- and ciprofloxacin-tolerant ∆lsr2 CFU were higher than that of the WT or complemented strains (Figure 6.5). This is consistent with my working model that TA systems activated in ∆lsr2 contribute to drug-tolerant persister formation.
Figure 6.5. The Δlsr2 of *M. tb* exhibits a drug tolerance phenotype *in vitro*. The percentage of surviving bacteria after 4-day incubation with 1 μg/ml of each drug was determined. Δlsr2 survived at higher levels than the wild type (WT) and the complemented (lsr2C) strains after treatment with ciprofloxacin (CIP), and isoniazid (INH), but not rifampin (RIF).
6.3 Determining the role of Lsr2 during *M. tb* infection

6.3.1 Why are the ∆lsr2-infected macrophages killed?

In Chapter 4 I found that macrophages infected with ∆lsr2 died soon after being infected. Virulent *M. tb* induces mostly macrophage necrosis after sufficient bacterial growth has occurred, whereas attenuated *M. tb* strains cause apoptosis which is thought to be a host defense mechanism limiting bacterial replication (Balcewicz-Sablinska, Keane et al. 1998; Keane, Remold et al. 2000; Divangahi, Chen et al. 2009). It is unknown at this time why and how the macrophages are killed. The macrophage infection experiments should be repeated to determine the nature of macrophage killing by ∆lsr2 using the Cell Detection ELISA kit (Roche), which allows measurement of apoptosis and necrosis by quantifying cytoplasmic (apoptosis), and extracellular (necrosis) histone-associated DNA fragments. I predict that the ∆lsr2 mutant induces mostly macrophage apoptosis, allowing it to persist at a lower bacterial number for extended periods without prematurely killing the host. My microarray data suggests that the expression of host effector proteins secreted by *M. tb* important for blocking apoptosis is downregulated in the mutant. Western blot analysis of secreted protein fractions for SapM, PtpA, SodA and NdkA should be performed. If these analyses corroborate my expression data, attempts should be made to complement the mutant for WT-like intracellular growth with each gene of interest constitutively expressed from pMV261.

6.3.2 Additional animal work

I aerosol-infected BALB/c mice with 30 CFUs of each respective strain and monitored the growth of bacteria for 8 weeks. I found that although ∆lsr2 replicated at a slower rate and plateaued at a lower bacterial burden during chronic infection (about 1 log lower than WT), the mutant appears to be fully capable of persisting in mice. Nevertheless two months is not long enough to draw definitive conclusions about the mutant’s ability to persist. Future experiments should be performed but with death as the endpoint. Additional analysis should also be included and are outlined below. The immune response elicited by the mutant can be assessed using supernatants from homogenized lungs to determine the presence of proinflammatory cytokines (e.g., INF-γ, TNF, IL-6, IL-1β, IL-4, IL-12p70, TGF-1β) using commercially available ELISA kits. Histological studies of lung sections should also be performed for both pathological
analysis and Ziehl-Nielsen staining to evaluate the presence of acid fast bacilli. I anticipate that the \( \Delta lsr2 \) can persist in the lungs of infected animals long-term without causing severe tissue damage by reducing the recruitment of proinflammatory cytokines to infected organs, thereby prolonging host survival. It could even be the case that the bacterial burden of the mutant will decrease over the course of chronic infection to pucibacilliary levels, and the Ziehl-Nielsen staining can determine if persistent bacteria are still present. Furthermore, I propose using the same ‘molecular clock’ plasmid used by Gill et al. (2009) to assess the mutant’s replication rate within mice. In the aforementioned study, they determined that during the persistent stage of murine infection, \( M. \text{tb} \) reduces its growth rate, doubling once every 100 hr (Gill, Harik et al. 2009). The equilibrium between the growth of bacteria and killing by the immune system results in a constant bacterial burden during the persistent phase of infection (Gill, Harik et al. 2009). This system could be used to distinguish whether 1) the \( \Delta lsr2 \) bacterial burden is the result of a balance between replication and killing, or 2) the mutant is able to evade killing and is a static population.

The Cornell mouse model of latent TB could also be performed to more directly assess the \( \Delta lsr2 \) mutant’s ability to persist (Cheigh, Senaratne et al. 2010). In this model, mice are aerosolized with \( \sim 100 \) CFU of \( \Delta lsr2 \), WT or complemented strain. At 4 weeks post-infection, the mice would be treated with isoniazid and pyrazynamide delivered ad libitum in drinking water. The duration of drug treatment is 8 weeks and the mice would continue to be assessed for organ CFU levels and histology for 24 weeks post cessation of drug treatment(Cheigh, Senaratne et al. 2010). I would expect to see more persisters present in the \( \Delta lsr2 \)-infected mice following antibiotic treatment as well as a higher rate of disease relapse amongst this cohort of mice. \( M. \text{tb} \) recovered from mice in the Cornell model sometimes exhibit atypical small, smooth colonies on 7H10 agar (McCune, Feldmann et al. 1966; McCune, Feldmann et al. 1966; Scanga, Mohan et al. 1999). These bacilli were negative for acid fast staining and did not grow in liquid 7H9 media upon subculturing, yet PCR amplification of IS6110 confirmed these bacilli to be \( M. \text{tb} \)(Scanga, Mohan et al. 1999). I suspect some of these colonies may contain lsr2 mutations and may prefer fatty acids as a carbon source for growth. Morphological variants should be subcultured in 7H9 broth supplemented with cholesterol, propionate, or valeric acid as a carbon source. Once obtained, southern blot and DNA sequencing analysis should be performed to determine if they
are \textit{lsr2} mutants. In the event that they are not \textit{lsr2} mutants, whole genome sequencing of these colonies should be done to identify novel mutations associated with persisters. Moreover, I recovered altered colonies resembling the \textit{\Delta lsr2} mutant from mouse tissues during my infection experiment but could not identify \textit{lsr2} mutations. These clones should be analyzed for \textit{lsr2} expression and their genomes should also be sequenced.

\textbf{6.3.3 Are \textit{lsr2} mutants present in human TB disease?}

It was recently shown that the majority of the bacilli present in patient sputum are in a VBNC state, but can be cultured if pretreated with exogenously added Rpf (Mukamolova, Turapov et al. 2010). It is possible that these may be naturally arising \textit{lsr2} mutants since I found all the \textit{rpf} genes are downregulated in the mutant. Human TB patient samples could be obtained and treated with Rpf. The resulting colonies could be analyzed for their \textit{lsr2} mutations and/or their \textit{lsr2} expression levels. In addition, organ homogenates from mouse experiments should be treated with Rpf proteins prior to plating to determine if a similar phantom subpopulation of VBNC bacilli is present.
REFERENCES


Cathelyn, J. S., D. W. Ellison, et al. (2007). "The RovA regulons of Yersinia enterocolitica and Yersinia pestis are distinct: evidence that many RovA-regulated genes were acquired more recently than the core genome." Mol Microbiol66(1): 189-205.


Hotter, G. S., B. J. Wards, et al. (2005). "Transposon mutagenesis of Mb0100 at the ppe1-nrp locus in Mycobacterium bovis disrupts phthiocerol dimycocerosate (PDIM) and glycosylphenol-PDIM biosynthesis, producing an avirulent strain with vaccine properties at least equal to those of M. bovis BCG." J Bacteriol 187(7): 2267-2277.


