"Role of the two ClpP protease subunits in *Mycobacterium tuberculosis*"

Yoann Personne

Queen Mary University of London

Thesis submitted for the degree of Doctor of Philosophy of the University of London

February 2012

Acknowledgements

I would like to thank the Biotechnology and Biological Sciences Research Council and Barts and The London Charity for funding.

I would like to express my thanks towards my supervisor Prof. Tanya Parish for her close supervision, guidance and expertise. I would also like to thank my second supervisor, Dr Richard Waite for his constant support.

I particularly want to thank Dr. Amanda Brown for all her advice over the years and getting all my questions answered, and the members of the TB Group (Dorothee, Julian, Lise, Melanie, Paul, and Pavitra) for their help.

I would also like to thank members of the 'TB DISCO' for making me feel welcome during my time in Seattle and more specifically thanks to Dr David Roberts, Dr Julianne Ollinger and Mai Ann Bailey for their teaching and help.

Special thanks go to Luisa, for her patience during the writing and for simply being there for me.

Abstract

Caseinolytic (Clp) proteases are the most widespread energy-dependent proteases in bacteria. They are involved in protein quality control by degrading misfolded and aggregated proteins and have a role in regulatory proteolysis. The main group of substrates of the Clp proteases is the SsrA-tagged proteins, which arise in the presence of defective translation. SsrA tagging is carried out by tmRNA, encoded by *ssrA*, together with a protein partner SmpB. While most organisms have only one ClpP, *Mycobacterium tuberculosis* has two ClpP protease subunits (ClpP1 and ClpP2) with at least one of them essential for growth. Co-expression of *clpP1* and *clpP2* was demonstrated showing that *clpP1* and *clpP2* are not expressed under different conditions. The promoter region of *clpP1P2* was identified, together with the potential ClgR binding site. A reporter system to assay ClpP1 and ClpP2 enzymatic activities was developed based on LacZ incorporating SsrA tag sequences. This showed that both ClpP1 and ClpP2 degrade SsrA-tagged LacZ, whilst only ClpP2 degrades untagged proteins. This suggests different pattern recognition for the two ClpP proteins with substrate recognition by ClpP1 dependent on the last three residues of the C-terminus of the tag sequence. Mutagenesis analysis of the accessory components demonstrated that *ssrA* is essential but SmpB deletion is viable. SmpB is not required for aerobic growth but the *smpB*Δ mutant strain was more sensitive to antibiotics targeting the ribosome as compared to wild-type cells.

Table of contents

List of abbreviations and units

ACDP	Advisory Committee on Dangerous Pathogens
BCG	Bacille Calmette-Guérin
BSA	Bovine serum albumin
cDNA	Complementary DNA
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide triphosphates
gDNA	Genomic DNA
LB	Luria-Bertani
MDR	Multi drug-resistant
MIC	Minimal Inhibitory Concentration
MTC	*Mycobacterium tuberculosis* complex
MU	Miller units
NRP1	Non-replicating phase stage 1
NRP2	Non-replicating phase stage 2
OADC	Oleic acid-albumin-dextrose-catalase
OD	Optical density
ONPG	Ortho-nitrophenyl-ß-galactoside
PCR	Polymerase chain reaction
qRT-PCR	Quantitative RT-PCR
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SDW	Sterile distilled water
Tris	Tris (hydroxymethyl) aminomethane
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside
XDR	Extremely drug-resistant

List of figures

List of tables

1 General Introduction

1.1 Mycobacteria

Mycobacteria are members of the bacterial order Actinomycetales along with the genera *Corynebacterium* and *Streptomyces*. The *Mycobacterium* genus includes over a hundred species that are found in soil, water or associated with plants, animals or humans; and the majority of these species are non pathogenic (Primm *et al.*, 2004; Tortoli, 2006). Mycobacteria are aerobic, rod-shaped, acid-fast bacteria (McMurray, 1996). They possess a GC-rich genome and a thick, waxy cell wall unique among prokaryotes containing peptidoglycan, arabinogalactan, mycolic acid, and complex lipids (Brennan and Nikaido, 1995; Clark-Curtiss *et al.*, 1985; Hett and Rubin, 2008). Based on 16S rRNA sequences, mycobacteria are divided into two taxonomic groups which also exhibit different growth rates (Stahl and Urbance, 1990); 'fast growing' mycobacteria, such as *Mycobacterium smegmatis,* form colonies within seven days; while 'slow growers', which includes most of the pathogenic mycobacteria, such as *Mycobacterium tuberculosis*, *Mycobacterium leprae* and *Mycobacterium bovis,* take weeks to months to form colonies.

The *M. tuberculosis* complex (MTC) of tuberculosis (TB)-causing mycobacteria comprises the species *M. tuberculosis*, *M. bovis*, *Mycobacterium africanum*, *Mycobacterium canetti* and *Mycobacterium microti*. Members of the MTC differ in morphology, host preference, virulence and disease patterns (Brosch *et al.*, 2002). While *M. microti* mainly causes TB in small rodents such as voles (Wells and Oxon, 1937), the host range of *M. bovis* is broad; it can cause TB in various domestic or wild animals like cattle or goats, as well as in humans (Ayele *et al.*, 2004). *M. africanum* is a major cause of human TB in West Africa while *M. canetti* is prevalent in the horn of Africa (Kallenius *et al.*, 1999; Miltgen *et al.*, 2002). The natural habitat of *M. tuberculosis* is humans, but it can also infect animals that are in contact with humans (Ocepek *et al.*, 2005).

1.2 Tuberculosis

1.2.1 History and current situation

Tuberculosis (TB) is an infectious disease that commonly affects the lungs (pulmonary TB), but it can also affect other parts of the body such as the central nervous system, bones, uro-genital tract and digestive system

(McMurray, 1996). TB is an ancient disease claiming victims throughout much of known human history (Donoghue, 2009); archaeological evidence of bone TB has been identified in Egyptian mummies and skeletal remains from the Iron Age (Taylor *et al.*, 2005; Zink *et al.*, 2003). However, it was not until 1882 that Robert Koch established that TB was a communicable disease caused by *M. tuberculosis* (with the publication of an article in the German journal Berliner Klinische Wochenschrift on 10th April 1882) (Kaufmann, 2000). With 1.2-1.5 million people dying from TB in 2010, TB is second only to HIV as the leading infectious killer of adults worldwide (WHO, 2011). Annually, approximately nine million new cases of TB are reported globally and about one third of the world's population is estimated to be latently infected (WHO, 2011), and of these individuals two to ten percent will develop active disease at some point in their life (Marks *et al.*, 2000).

In the UK, the incidence of tuberculosis has increased steadily over the past two decades. However in 2010, a decrease of 4.9% was reported in the number of cases (8,483 cases) compared to the previous year (HPA, 2010). London accounts for 39% of all TB cases in the UK and rates in some parts of the region are over 80 per 100,000, which are similar to those reported in high incidence countries in South America, Asia and North Africa (WHO, 2011).

1.2.2 Pathogenesis and immune response

Tuberculosis is caused by a complex interaction between *M. tuberculosis* and its host that can last for decades. Primary TB infection begins with inhalation of tubercle bacilli. When the bacteria reach pulmonary alveoli, they are ingested by macrophages and reside in a vacuole called the phagosome (Russell, 2001). If normal phagosomal maturation cycle occurs, i.e. phagosome-lysosome fusion, these bacteria encounter a very hostile environment that includes acid pH, reactive oxygen intermediates, lysosomal enzymes and toxic peptides. However, *M. tuberculosis* can prevent fusion of the phagosome with lysosomes, thus escaping exposure to the hostile environment (Brown *et al.*, 1969; Clemens and Horwitz, 1995). The infection of macrophages constitutes the early stage of the disease, generally occurring within the first three weeks post-infection and most patients remain asymptomatic or only present mild flu-like symptoms during this stage. *M. tuberculosis* then spreads to the lymph nodes in the lung via the lymphatic system and forms a lesion, known as the Ghon focus (Smith, 2003). Antigen-specific T lymphocytes release lymphokines

which activate macrophages and cause them to form a compact cluster or granuloma around the site of infection (Hu and Coates, 2003). The formation of the granuloma is usually sufficient to arrest the primary infection, and surrounding fibroblasts produce a dense scar tissue which encloses the lesion (Saunders and Cooper, 2000). Tuberculous granulomas are often caseous due to necrosis at the centre of the lesion but can also be fibrotic and calcified. *M. tuberculosis* is located within macrophages of granulomas and in large numbers within the central caseous region (Bouley *et al.*, 2001; Dannenberg, 1993). Granulomas are thought to limit bacterial growth by restricting access to oxygen and nutrients as well as exposing the bacteria to host factors such as nitric oxide and acidic pH (Gomes *et al.*, 1999; Schaible *et al.*, 1998; Schnappinger *et al.*, 2003; Via *et al.*, 2008). However, the bacilli are not necessarily eradicated and surviving bacteria are considered to enter a phase of non-replicating persistence (Parrish *et al.*, 1998; Wayne, 1994). Bacterial persistence during latent tuberculosis is not fully understood although latency is central to the prolonged presence of the pathogen in the host with the potential to give rise to reactivation disease (Grange, 1992). The bacterial and host factors implicated in reactivation of the disease remain to be determined but weakening of the immune system (for example by co-infection with HIV, malnutrition or ageing) is strongly associated with the onset of active tuberculosis (Dolin *et al.*, 1994; Swaminathan *et al.*, 2000).

1.2.3 The molecular response to hypoxia

Of the different conditions *M. tuberculosis* is believed to be exposed to inside the host hypoxia is probably the best studied. The main model of hypoxia was developed by Wayne, in which *M. tuberculosis* cultures are gradually depleted of oxygen and enter in a non-replicating state (Wayne, 1977; Wayne and Lin, 1982; Wayne and Hayes, 1996). Two distinct stages are observed: Non-Replicating Persistence stage 1 (NRP1), which occurs under microaerophilic conditions (at an oxygen concentration around 1%), and Non-Replicating Persistence stage 2 (NRP2), which happens at the onset of anaerobic conditions (oxygen concentration below 0.06%) (Wayne and Hayes, 1996). NRP1 is characterised by cell enlargement due to thickening of the cell wall, termination of DNA synthesis, a significant decrease in RNA synthesis and protein synthesis, and termination of cell division (Cunningham and Spreadbury, 1998; Hu *et al.*, 1998; Muttucumaru *et al.*, 2004; Wayne and Hayes, 1996). In NRP2, cell enlargement ceases and the antibiotic

susceptibility profile of the cells changes (Wayne and Hayes, 1996). Under anaerobic conditions *M. tuberculosis* rapidly reduces or completely stops protein synthesis in order to shutdown cellular metabolic activities (Hu *et al.*, 1998). However *M. tuberculosis* is not completely dormant during its state of persistence as approximately 50% of its transcriptome is active (Ehlers, 2009) and treatment with isoniazid, which is only active against replicating mycobacteria, is effective in preventing active disease in humans (Comstock, 1999; Diel *et al.*, 2005). This activity can only be explained if at least a fraction of the mycobacterial population are replicating at least from time to time.

The DosR (also called DevR) regulon is critically involved in the adaptation to hypoxia, it comprises of a set of about 50 genes that are highly induced by hypoxia, nitric oxide and during infection of macrophages, which are regulated by the DosR-DosS two-component regulatory system (Dasgupta *et al.*, 2000; Kumar *et al.*, 2007; Ohno *et al.*, 2003; Park *et al.*, 2003; Sherman *et al.*, 2001; Voskuil *et al.*, 2004). In response to hypoxia, expression of most of the genes of the DosR regulon is only transiently induced as about half of the regulon is no longer induced after 24 hours of exposure to hypoxic conditions (Rustad *et al.*, 2008). A second transcriptional response, the enduring hypoxic response (EHR), consists of more than 200 genes that are not initially induced but are upregulated after four to seven days of hypoxia. Unlike the DosR regulon, these genes overlap with genes that are upregulated under conditions of nutrient starvation suggesting that the EHR contributes to the adaptation to and maintenance of a non-replicating state (Rustad *et al.*, 2008; Rustad *et al.*, 2009a).

1.2.4 TB vaccination and treatment

M. bovis Bacillus Calmette-Guerin (BCG) is the only licensed TB vaccine and the oldest vaccine currently in use; it has been given four billion times over the last 90 years (Parida and Kaufmann, 2010). Although BCG is effective against childhood TB, its efficacy among adolescents and young adults is highly variable, especially in TB endemic areas (Antas and Castello-Branco, 2008; Fine, 1995). There are a number of hypotheses for the varying efficacy of BCG including genetic variability between different strains of BCG, exposure to environmental mycobacteria and other infections, as well as genetic and nutritional differences in human populations, and trial methods (Liu *et al.*, 2009). Since BCG may cause disease in immunocompromised individuals,

particularly those with HIV, the development of a new vaccine is essential if TB is to be controlled and ultimately eradicated. A total of 11 vaccine candidates against TB have entered clinical trials within the last several years (Kaufmann, 2010) but a better alternative to BCG has yet to be produced.

TB chemotherapy takes typically upwards of six months of daily treatment, and has two phases known as the intensive and the continuation phase. The intensive phase regimen consists of four drugs (isoniazid, rifampicin, ethambutol and pyrazinamide) to be taken daily for two months, this is then followed by the continuation phase where two drugs (isoniazid and rifampicin) are taken for a further four months. DOTS (Directly Observed Treatment Short course) is the internationally recommended strategy for TB control and was reported by the WHO in 2007 to have an 85% global success rate. However, due to inadequate therapy, poor follow-up and outdated TB control strategies, drug-resistant strains have emerged and are now widespread. Two main categories of resistance are encountered: multidrug-resistant TB (MDR-TB) is resistant to isoniazid and rifampicin and extensively drug-resistant TB (XDR-TB) is resistant to isoniazid and rifampicin plus any fluoroquinolone and at least one of the second-line anti-TB injectable drugs (amikacin, kanamycin and capreomycin). There were an estimated 650,000 cases of MDR-TB in 2010 and there are thought to be 25,000 cases of XDR-TB emerging every year; by July 2010 58 countries reported at least one case of XDR-TB (WHO, 2011). Cure rates for patients with XDR are relatively low, especially if patients are co-infected with HIV. For example, in an outbreak of XDR-TB in KwaZulu-Natal, South Africa, 98% of infected patients died within three weeks of diagnosis (Gandhi *et al.*, 2006). Virtually incurable isolates, extremely drug resistant TB (XXDR-TB) have now appeared and underline the need for new drugs (Migliori *et al.*, 2007; Velayati *et al.*, 2009).

1.3 Drug discovery and validation

The discovery and development of new drugs is a complex, lengthy and expensive process. The drug discovery pipeline starts with target identification, then followed by target validation, assay development and high-throughput screening, confirming hits to leads through to lead optimisation and development of a candidate which can then enter into the clinical trials phase (Showalter and Denny, 2008). Appropriate drug targets are genes that encode proteins or enzymes involved in processes critical for survival during growth,

infection or persistence. Potential targets for new drugs need to be identified in the context of their essentiality for survival *in vitro* and *in vivo* since a gene may be required for growth during an infection but dispensable for mycobacterial survival in defined medium and vice versa (Sassetti and Rubin, 2003). Thus, in order to identify new drug targets, essential genes need to be selected, their function identified and then validated as potential drug targets. In addition to essentiality, other properties relevant to the drug discovery process such as selectivity, suitability for structural studies and ability to monitor inhibition in whole cells are important factors to assess the value of a target for drug development (Showalter and Denny, 2008).

Essential genes, by definition, cannot be deleted from the chromosome and their identification is not an easy task. The construction of a library of tagged mutants using transposon mutagenesis can be used to screen large pools of mutants for essential genes: genes that are disrupted are non-essential genes while genes that cannot be mutated by transposon insertion can be predicted to be essential. Transposon mutants can be detected by mapping the insertion sites of individual mutants by sequencing (Lamichhane *et al.*, 2003) or by using transposon site hybridisation (TraSH) (Sassetti *et al.*, 2003). This technique has been used to look at the essentiality of genes under defined growth conditions, such as growth on specific media, in macrophages, and in a mouse infection model (Rengarajan *et al.*, 2005; Sassetti *et al.*, 2003; Sassetti and Rubin, 2003). Another method called DeADMAn (Designer Arrays for Defined Mutant Analysis), where collections of archived mutants are pooled and assessed simultaneously under a particular stress condition, has been used to study essential genes during *in vivo* infections in mice (Lamichhane *et al.*, 2005), guinea pigs (Jain *et al.*, 2007) and macaques (Dutta *et al.*, 2010) .

Validation is a crucial step in the drug discovery process. For a molecular target to be validated, it requires demonstration that it is critically involved in the disease process and that its modulation will have a therapeutic effect. This can be achieved by the use of several strategies such as the creation of specific gene deletion mutants, use of site-specific recombination systems, antisense oligonucleotides and regulated promoters. It is important to determine under which conditions the gene is essential for growth, for example assay growth *in vitro* in the presence of different media, carbon sources, or exposure to environmental stresses such as hypoxic or low iron conditions, as well as

examining virulence and survival *in vivo*. *In vitro* models of nutrient starvation, oxygen deprivation (for example the Wayne model of hypoxia), as well as prolonged stationary phase can be used to replicate conditions that are believed to occur during persistence (Betts *et al.*, 2002; Hu *et al.*, 2000; Shleeva *et al.*, 2002; Sun and Zhang, 1999; Wayne and Hayes, 1996; Wayne and Sohaskey, 2001; Zhang *et al.*, 2001). For *in vivo* studies, the mouse model is widely used but TB granulomas in mice are not especially hypoxic (Aly *et al.*, 2006) and mice are not as sensitive to TB as other animals (Gupta and Katoch, 2005). Alternative animal models are more expensive and require larger facilities; TB models to date include the use of guinea pigs, rabbits and non-human primates (Dharmadhikari and Nardell, 2008). Besides essentiality, another important criteria for a good drug target is vulnerability which represents how much a protein or enzyme has to be inactivated before a therapeutic effect can be seen. This can be achieved by creating conditional knockdown mutants, where the native promoter of a target gene is replaced by one which is tightly regulated and inducible, for example (Carroll *et al.*, 2005; Ehrt *et al.*, 2005; Gandotra *et al.*, 2007). Once a drug has been designed, understanding its mechanisms of action is critical and selecting drugs that inhibit a broad spectrum of essential biochemical pathways is useful in order to minimise the appearance of resistant strains. The use of combination therapy also reduces the development of resistance and cures patients more rapidly. Thus, another challenge for drug development is the selection of optimal targets that may be inhibited without drug-drug antagonism.

Most drugs act by blocking the action of a particular protein whose function is essential in the cell. Proteins are involved in virtually every cellular function and most proteins must fold into defined three-dimensional structures to gain functional activity so, accordingly, cells contain some quality controls.

1.4 Protein quality control and proteases

Environmental conditions such as oxidative stress or extreme pH may cause proteins to misfold. Misfolded proteins expose hydrophobic polypeptide stretches, normally buried within the core of globular protein, that can coagulate and form aggregates (Alberts *et al.*, 2002a). Aggregated proteins are biologically inactive and can be toxic for the cells. Protein quality control systems contribute to the maintenance of proteome homeostasis and two strategies are used to avoid accumulation of aggregates. Molecular chaperones

stabilise protein conformations or refold misfolded proteins; however if a protein cannot be refolded by chaperones it will be degraded by specialised proteases that are often induced during adverse conditions (Dougan *et al.*, 2002; Hartl and Hayer-Hartl, 2002). Accumulation of damaged or incorrectly folded proteins also occurs during normal growth conditions with approximately 20% of newly synthesised polypeptides being degraded by proteases and a similar proportion associated with chaperones (Wickner and Maurizi, 1999). The correct balance between protein folding and degradation is critical for cell viability and multiple chaperones and proteases are present. In *E. coli*, the DnaK chaperone with its DnaJ and GrpE co-chaperones and the GroEL–GroES chaperone complex have central functions in the refolding of misfolded proteins (Hartl and Hayer-Hartl, 2002).

Proteases are enzymes that catalyse the hydrolysis of peptide bonds in polypeptides. Based on their catalytic mechanism, proteases are divided into seven groups: aspartic acid proteases, cysteine proteases, glutamic acid proteases, metalloproteases, serine proteases, threonine proteases, and the seventh group, which constitutes a number of proteases that cannot yet be assigned to any particular catalytic type (Wladyka and Pustelny, 2008). Serine proteases are one of the most abundant groups of proteolytic enzymes found in all kingdoms of life (Tripathi and Sowdhamini, 2008). The number of proteases is variable between organisms; for example *E. coli* has a large number of proteases (398), *M. leprae* has only 69 while *M. tuberculosis* encodes 142 proteases (Rawlings *et al*., 2010). In bacteria most intracellular proteolysis is performed by four energy-dependent proteases: Lon, Clp, HslV, and FtsH (Gottesman, 1996). Lon is responsible for more than half of all energy-dependent proteolysis in *E. coli* (Laskowska *et al.*, 1996; Maurizi, 1992) but is absent in *M. tuberculosis. M. tuberculosis* also lacks homologs of HslV but contains the proteasome which is absent in most bacterial species, FtsH, and two ClpPs.

1.4.1 The proteasome

Proteasomes are present in eukaryotes and archaea but are only found in some bacteria of the order actinomycetales, including *M. tuberculosis*. The eukaryotic proteasome is comprised of two complexes: the 20S core, where proteins are degraded, and the 19S regulatory cap that binds substrates to be degraded (Coux *et al.*, 1996). The 19S cap, composed of numerous proteins

including six AAA (ATPases Associated with various Activities) ATPases, is involved in the recognition, unfolding, and translocation of proteins into the proteasome core (Pickart and Cohen, 2004). The *M. tuberculosis* 20S core is composed of seven α-type and seven β-type subunits (encoded by *prcA* and *prcB* respectively) forming a barrel shaped structure which contains the active sites (Hu *et al.*, 2006; Pickart and Cohen, 2004); however only limited peptidase activity has been observed for the core particle alone. PrcA and PrcB are predicted to be essential or required for optimal growth (Darwin *et al.*, 2003; Sassetti *et al.*, 2003) and are essential for pathogenesis in *M. tuberculosis* (Darwin *et al.*, 2003; Gandotra *et al.*, 2007; Lamichhane *et al.*, 2006), although deletion of *prcBA* did not severely impact bacterial growth in *M. smegmatis* (Knipfer and Shrader, 1997). The *M. tuberculosis* proteasome system requires at least two co-factors, Mpa (Mycobaterium proteasomal ATPase) and PafA (proteasome accessory factor A). Mpa has ATPase activity (Darwin *et al.*, 2005) and directly interacts with the 20S core protease (Wang *et al.*, 2009) while PafA attaches a small protein, Pup (prokaryotic ubiquitin-like protein), to substrates targeted for degradation (Pearce *et al.*, 2006; Striebel *et al.*, 2010). The accessory proteins MpA and PafA are not essential for growth but are essential for virulence and resistance to nitric oxide (Pearce *et al.*, 2006). To date, seven proteasome substrates have been identified: FabD, PanB, MpA, PhoH2, Icl, MtrA and Ino1 (Festa *et al.*, 2010; Pearce *et al.*, 2006).

1.4.2 The Clp proteases

Caseinolytic (Clp) proteases are highly conserved serine proteases present in a wide range of bacteria as well as in plants and mammals (Gottesman *et al.*, 1990; Porankiewicz *et al.*, 1999). The Clp holoenzyme is structurally similar to the eukaryotic 26S proteasomal complex and is composed of a catalytic and a regulatory subunit. The term 'Clp protease' refers to the active Clp protease complex which consists of two protease heptamers and two ATPase hexamers (Licht and Lee, 2008).

The central proteolytic core assembles into a tetradecamer of ClpP subunits consisting of two heptameric rings stacked on top of each other forming a cavity where protein degradation occurs (Flanagan *et al.*, 1995; Kessel *et al.*, 1995; Wang *et al.*, 1997). Inside the proteolytic chamber each monomer contains an

active site, which is formed by a Ser-His-Asp catalytic triad typical of serine proteases (Ingvarsson *et al.*, 2007; Wang *et al.*, 1997). This chamber is 51 Å in diameter and can accommodate proteins up to approximately 50 kDa (Wang *et al.*, 1997). Axial pores of 1–2 nm are present at either end of the ClpP tetradecamer where substrates enter (Baker and Sauer, 2006). These pores are only large enough for small polypeptides and unfolded proteins to enter, thus cytoplasmic proteins appear to be protected from accidental degradation (Inobe and Matouschek, 2008). ClpP by itself displays limited peptidase activity; it can degrade small peptides and is capable of hydrolysing proteins but at a very slow rate (Jennings *et al.*, 2008; Porankiewicz *et al.*, 1999; Woo *et al.*, 1989). In order to function as an effective protease ClpP associates with Clp ATPases, at either one or both ends, to make an active complex.

The Clp ATPases belong to the AAA+ superfamily of ATPase (Ogura and Wilkinson, 2001). These proteins share a common ATPase domain, the AAA domain, with a conserved sequence of 230-250 amino acid residues, containing Walker A and Walker B nucleotide binding and recognition motifs, and a C-terminal α–helical domain (Snider *et al.*, 2008; Zhang *et al.*, 2000). Clp ATPases can function both as molecular chaperones to promote protein remodelling and reactivation or as regulatory components of Clp proteases for degradation by ClpP (Wawrzynow *et al.*, 1996). The Clp ATPases are ubiquitous and highly conserved but their number and types can vary even between closely related genera (Frees *et al.*, 2007). Clp ATPases are divided into two classes based on the number of AAA domains present. Class I proteins, including ClpA, ClpB and ClpC, contain two AAA domains; whilst class II proteins, including ClpX, contain only one (Schirmer *et al.*, 1996). The first ATPase domain promotes hexamerisation of the Clp ATPase protein whereas the second ATPase domain, when present, is primarily responsible for ATP hydrolysis (Singh and Maurizi, 1994). The functional significance of having one versus two AAA+ domains for delivering substrates to ClpP is not yet known (Hanson and Whiteheart, 2005). In addition, Clp ATPases contain an N-terminal domain (N-domain) and a linker domain (also called I-domain or M-domain) but these domains are not conserved to the same degree as the AAA domain (Kojetin *et al.*, 2009).

ClpX is the most widespread Clp ATPase subunit (Kress *et al.*, 2009); ClpC is found in Gram positive bacteria and cyanobacteria, while ClpA is found in the

Gram negative proteobacteria (Kress *et al.*, 2009). ClpB is structurally similar to ClpA but does not associate with ClpP in any organisms. Thus ClpB is not involved in degradation; instead it functions exclusively as a molecular chaperone that reactivates denatured and aggregated proteins with the help of other chaperones such as DnaK or DnaJ (Motohashi *et al.*, 1999; Zolkiewski, 1999). Notably, ClpB has a heptameric ring structure (Kim *et al.*, 2000a) while the others Clp ATPases assemble into hexamers.

1.4.2.1 Protein degradation

Protein degradation by the Clp proteases is a three step-process: substrate recognition, unfolding and translocation into ClpP, degradation and peptide release (Fig 1.1). Substrate recognition by the Clp ATPases must be highly specific in order to protect cells from uncontrolled proteolysis. Clp ATPases recognise and bind to a target protein via their sensor and substrate discrimination (SSD) domain (Smith *et al.*, 1999) and a simple strategy for different ATPases to recognise specific substrates is the presence of extra domains in their N-terminal region which are missing in other ATPases (Smith *et al.*, 1999). Substrate recognition and selection by the Clp ATPase may be assisted by adaptor proteins which bind simultaneously to the substrate and to the Clp ATPase (Baker and Sauer, 2006; Kirstein *et al.*, 2006). Misfolded or aggregated proteins are recognised by exposed hydrophobic patches at their surface, whilst specific substrate proteins are recognised via degradation tags termed degrons (Varshavsky, 1991).

Following substrate recognition, the Clp ATPase unfolds and translocates the selected protein into the proteolytic cavity using energy generated by ATP hydrolysis. Clp ATPases exhibit three key structural elements for this process: a narrow gate that restricts passage of a structured protein sequence into their central channel, a long inert central channel, and ATP hydrolysis powered loops positioned along their channel that are in contact with the engaged protein (Hinnerwisch *et al.*, 2005; Martin *et al.*, 2008). Depending on the size of the protein, denaturation can require 20 to 500 molecules of ATP (Baker and Sauer, 2006). The entry of substrates into ClpP is normally blocked by the ClpP N-terminal loops that change conformation only when interacting with the Clp ATPases (Jennings *et al.*, 2008).

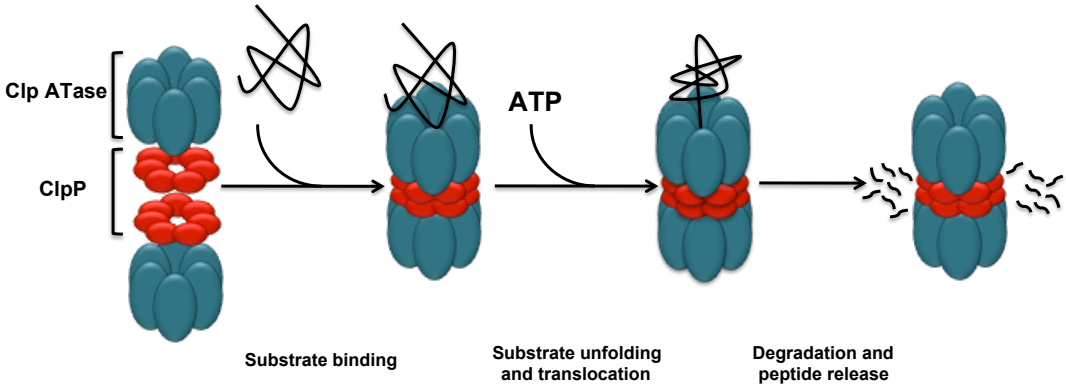

Figure 1.1 Protein degradation by the Clp proteases

In presence of ATP, ClpP and Clp ATPases form a complex. Clp ATPases recognise and bind to substrate proteins. The protein is unfolded and translocated into ClpP for degradation. Degraded peptides of approximately six to eight residues are released in the cytosol.

Proteolysis is catalysed by the active sites that line the inner surface of the internal chamber formed by ClpP (Kim *et al.*, 2000b). ClpP rapidly degrades proteins in a processive manner, peptide products have an average length of about six to eight residues (Choi and Licht, 2005; Licht and Lee, 2008; Yu and Houry, 2007), and ClpP alone controls the length of the peptide products (Jennings *et al.*, 2008). It is unclear how the peptide products are released from the chamber, the peptides may be released through the axial pores or more likely through a set of equatorial pores that open transiently at the interface of the two heptameric ClpP rings (Geiger *et al.*, 2011; Kimber *et al.*, 2010; Licht and Lee, 2008; Sprangers *et al.*, 2005).

1.4.2.2 Interaction between ClpP and Clp ATPase

ClpP is made of heptameric rings while Clp ATPases form hexamers, thus a symmetry mismatch takes place (Bewley *et al.*, 2006; Kessel *et al.*, 1995); it has been considered that this asymmetry may facilitate rotation of the Clp ATPases for a processive degradation of substrate proteins (Beuron *et al.*, 1998). Different structural features mediate the interaction of ClpP with a Clp ATPase partner. The first involves surface loops in the Clp ATPase protein, termed the ClpP binding loops, with a highly conserved tripeptide motif [LIVMP]-G-[FL] which dock into deep hydrophobic pockets on the surface of ClpP (Bewley *et al.*, 2006; Kim *et al.*, 2001; Kim and Kim, 2003). Each of the six IGF loops in a ClpX hexamer are required for strong ClpP binding and for efficient ClpXP proteolysis (Martin *et al.*, 2007). Clp ATPases that lack the tripeptide motif do not form a complex with the ClpP protease, and transplanting the ClpA IGL loop to ClpB allows ClpB to interact with ClpP (Weibezahn *et al.*, 2004). ClpP N-terminal loops mediate a second critical interaction between ClpP and the ATPases. N-terminal truncations of *E. coli* ClpP prevents the interaction between ClpP and its ATPase partner and alignment of multiple ClpP sequences shows a very high degree of conservation in the N-terminus (Gribun *et al.*, 2005). Additionally, interaction occurs between the pore-2 loops present at the bottom of *E. coli* ClpX and the N-terminal loops of ClpP; this interaction is weak but allows fine-tuning of ClpX-ClpP transactions via changes in ATP-hydrolysis rates during substrate unfolding and degradation (Martin *et al.*, 2007, 2008).

1.5 ClpPs and Clp ATPases in *M. tuberculosis*

M. tuberculosis clpP1 (Rv2461c) and *clpP2* (Rv2460c) are co-located and presumed to form an operon. *clpP1* was recently demonstrated to be essential (Ollinger *et al*., 2011) and *clpP2* is predicted to be essential using high-density transposon mutagenesis (Sassetti *et al.*, 2003). A ClpP1-ClpP2 knockdown strain was constructed using a tetracycline inducible system and was unable to grow in absence of tetracycline confirming that the operon is essential (Carroll *et al.*, 2005; Carroll *et al.*, 2011). This knockdown strain showed reduced growth and virulence in a macrophage infection model, confirming the importance of ClpP1 and ClpP2 for growth and infection (Carroll *et al.*, 2011). *clpP1* and *clpP2* are highly expressed in both aerobic and hypoxic environments and are further up-regulated during reaeration from anaerobic conditions suggesting they are important for survival during latency and reactivation (Muttucumaru *et al*., 2004; Sherrid *et al*., 2010). ClpP1 and ClpP2 are similar in size (22 kDa and 23.5 kDa respectively) and in sequence (43% amino acid similarity). They both contain predicted active sites formed by a Ser-His-Asp catalytic triad typical of serine proteases (Rawlings *et al*., 2010). ClpP monomers can be divided into a globular domain (the head domain), and an elongated domain (the handle domain), which constitutes the equatorial wall of the tetradecamer (Wang *et al.,* 1997). The structure of *M. tuberculosis* ClpP1 revealed that the first ten residues of the handle domain are disordered which affects the formation and configuration of the tetradecamer: the equatorial pores (hypothesised to be used for peptide exit) are larger as compared to ClpP from other species (Ingvarsson *et al.*, 2007). Disorder in the handle domain has also been described for other ClpP structures (Gribun et al., 2005). The N-terminal residues of mature ClpP proteins are highly conserved, but the N-terminus sequence of *M. tuberculosis* ClpP1 significantly deviates from the consensus sequence and exhibits an extended αA helix which is unique (Ingvarsson *et al*., 2007). The active site containing the catalytic triad Ser98, His123 and Asp172 is situated at the base of the head domain at the interface with the handle domain. However, the conformation of the active site of ClpP1 is consistent with an inactive enzyme, and peptidase activity was not detected (Ingvarsson *et al*., 2007).

M. tuberculosis has three potential Clp ATPases: ClpC1 (Rv3596c), ClpC2 (Rv2667) and ClpX (Rv2457c). ClpC1 is a Class I protein and contains two

ATPase (AAA) domains while ClpX contains only one AAA domain and is classified as Class II protein. ClpX and ClpC1 are predicted to be essential for optimal growth of *M. tuberculosis* (Sassetti *et al.*, 2003). ClpC1 is required for survival in macrophages and was recently identified as a drug target (Rengarajan *et al.*, 2005; Schmitt *et al.*, 2011). ClpC1 exhibits chaperone activity; it has two repeat sequences at its N-terminus but only one of them is required for the chaperone activity (Kar *et al.*, 2008). Contrary to the other Clp ATPases, *clpC2* is predicted to be non-essential in *M. tuberculosis* although it is up-regulated during nutrient starvation (Betts *et al.*, 2002). The predicted ClpC2 protein is smaller than other Clp ATPases, does not contain any AAA domain and the purified ClpC2 did not have any ATPase activity (Benaroudj *et al.*, 2011), suggesting that this protein might not be functional. Additionally this protein is not conserved in *M. leprae*. ClpX was found to interact with FtsZ (Dziedzic *et al.*, 2010).

1.6 Gene regulation

1.6.1 Bacterial promoter and RNA polymerase

Prokaryotic promoters consist of two short sequences centred typically 10 and 35 base pairs upstream of the transcriptional start site; the *E. coli* consensus sequence consists of two hexamers: the Pribnow box (TATAAT) at -10 and TTGACA at -35 (Alberts *et al.*, 2002a). The frequency of transcription initiation events defines the strength of a promoter; *E. coli* genes with strong promoters are typically transcribed every two seconds whereas weak promoters are transcribed only about once in ten minutes (Stryer *et al.*, 2002). The bacterial RNA polymerase is a multisubunit complex composed of five subunits: 2α, β, β' and ω (Stryer *et al.*, 2002). In order to bind tightly to promoter sequences the polymerase recruits another subunit called the sigma (σ) factor. Different σ factors recognise specific sequences and respond to environmental conditions; for example sigma factor σ^{32} responds to heat shock while sigma factor σ^{S} controls expression of stationary phase genes (Watson, 2004). *M. tuberculosis* encodes 13 sigma factors (Cole *et al.*, 1998) and sigma A is the principal sigma factor of *M. tuberculosis* (Gomez *et al.*, 1998).

Promoters are commonly identified and characterised by means of reporter genes such as *lacZ* or *gfp* (Blokpoel *et al.*, 2003; Carroll *et al.*, 2005; Ehrt *et al.*, 2005). In contrast to *E. coli* promoters, mycobacterial promoters are G/C rich

and very diverse (Gomez and Smith, 2000). They are classified into three groups: group A has a -10 and -35 motif similar to the *E. coli* σ^{70} consensus, group B is characterised by the presence of a -10 hexamer similar to the consensus while group C has no resemblance to the consensus in either elements (Gomez and Smith, 2000).

1.6.2 ClpP expression

In *E. coli* the alternative sigma factor σ^{32} (RpoH) associates with RNA polymerase to activate transcription of the *clp* genes (Kroh and Simon, 1990). In several Gram positive bacteria *clpP* transcription is stress dependent; for example in *B. subtilis*, *Lactococcus lactis* and *Streptococcus mutans,* ClpP is negatively regulated by the transcriptional regulator CtsR during normal growth conditions (Derre *et al.*, 1999; Kruger *et al.*, 2001; Lemos and Burne, 2002; Varmanen *et al.*, 2000) while during stress conditions, such as heat shock, CtsR is degraded by the ClpCP complex to increase *clpP* expression (Kruger *et al.*, 2001). In *Streptococcus salivarius clpP* expression is also regulated by CtsR and a second repressor named HrcA (Chastanet and Msadek, 2003).

In actinobacteria, including *C. glutamicum* and *Strep. lividans*, the *clpP1P2* operon is controlled by the transcriptional activator ClgR (*clp* gene Regulator) (Bellier *et al.*, 2006; Bellier and Mazodier, 2004; Engels *et al.*, 2004). ClgR both induces and is degraded by the Clp proteases in these organisms, this degradation is mediated by two C-terminal alanine residues of the ClgR protein sequence (Bellier *et al.*, 2006; Engels *et al.*, 2004). ClgR is not involved in the heat-triggered activation of *clpP1P2* expression in *C. glutamicum,* indicating that the corresponding promoter can be recognised by different sigma factors; one active at ambient temperatures and requiring ClgR, and the other activated by severe heat stress and acting independently of ClgR (Engels *et al.*, 2004). *clgR* homologs are present in all actinomycete genomes, except *M. leprae*, at a conserved genomic location downstream of the *ftsK* and *pgsA3* genes, although the identity and number of genes inserted between *ftsK*, *pgsA3* and *clgR* varies to some degree (Engels *et al.*, 2004). The *M. tuberculosis clgR* gene has been shown to replace its counterpart in *C. glutamicum* for activating *clpP1P2* expression, suggesting a conserved role for ClgR actinomycetes (Engels *et al*., 2004). Upon binding to ClgR, it has been suggested that DNA undergoes considerable bending which is an important feature for transcription activation (Russo *et al.*, 2009). An alignment of all ClgR operator regions

allowed the deduction of a consensus motif in *C. glutamicum* (WNNWMGCYNNNRGCGWWS, where M=A or C; R = A or G; S= C, T or G; W = A or T; Y = C or T and N stands for any nucleotide) (Engels *et al.*, 2005).

M. tuberculosis ClgR homologue (Rv2745c) binds upstream of the *clpP1P2* promoter and directly activates transcription of the ClpP proteases (Mehra and Kaushal, 2009*;* Sherrid *et al.,* 2010). Disruption of the *M. smegmatis* ClgR homologue prevents up-regulation of *clpP1P2* upon exposure to vancomycin, suggesting ClgR also controls *clpP1P2* expression in *M. smegmatis* (Barik *et al.*, 2009). A ClgR deletion mutant confirmed that ClgR was activating the expression of *clpP1* and *clpP2* in *M. tuberculosis* and this strain exhibited a reduced capacity to replicate in macrophages compared to the wild-type organism (Estorninho *et al.*, 2010). Microarray analyses showed that ClgR was induced after heat shock (Stewart *et al*., 2002), starvation (Betts *et al.*, 2002) and under anaerobic conditions (Muttucumaru *et al*., 2004) in *M. tuberculosis.* Moreover ClgR is involved in the reaeration response as it is induced during the transition of *M. tuberculosis* from bacteriostasis to growth (Sherrid *et al.*, 2010). At least ten genes have been proposed to be directly regulated by ClgR in *M. tuberculosis,* with the majority involved in protein stabilisation, disassembly or degradation (Estorninho *et al.*, 2010).

1.7 ClpP activators

A novel class of compounds, the acyldepsipeptides (ADEPs), which target ClpP, are active against *M. tuberculosis* and also have potent activity against various Gram positive bacteria, including multidrug resistant isolates such as methicilin-resistant *Staphylococcus aureus* (MRSA) (Brotz-Oesterhelt *et al.*, 2005; Ollinger *et al.*, 2011; Socha *et al.*, 2010). Unusually, ADEPs are activators rather than inhibitors of the protease function of ClpP (Brotz-Oesterhelt *et al*., 2005). ADEPs prevent the interaction of ClpP with its ATPase partner by competing for the same binding site and trigger a conformational change in ClpP that widens the entrance pores (Lee *et al.*, 2010; Li *et al.*, 2010). ADEPs induce oligomerisation of ClpP monomers and activate the resulting tetradecamer to bind and degrade unfolded, nascent polypeptides (Kirstein *et al.*, 2009). This unregulated proteolysis by the protease ultimately leads to cell death, which could be caused by the shortage of essential cellular proteins or by the accumulation of considerable amounts of diverse protein fragments. Additionally, ADEP treatment induces FtsZ degradation and

prevents the formation of FtsZ rings, thus inhibiting cell division in Gram positive bacteria (Sass *et al.*, 2011). Gram negative bacteria are resistant to ADEP due to efflux pumps that remove the drug from the cell; this has been demonstrated as the deletion of those efflux pumps or the use of permeabilising agents made them susceptible to ADEP treatment (Brotz-Oesterhelt *et al*., 2005). ADEPs activity against *M. tuberculosis* is enhanced by the addition of efflux pump inhibitors, demonstrating that export occurs as seen with Gram negative bacteria (Ollinger *et al.*, 2011). Resistance strains can arise by deletion of the *clpP* gene in species where *clpP* is not essential such as *B. subtilis* (Brotz-Oesterhelt *et al.*, 2005). Resistance was also observed in *S. lividans* which contains five ClpP proteins, ClpP3 is insensitive to ADEP and can substitute for ClpP1; other resistance mechanisms are also involved but remain to be characterised (Gominet *et al.*, 2011). Since both of the ClpP proteases may be essential in *M. tuberculosis* (Ollinger *et al.,* 2011) targeting ClpP activity may be an attractive choice for new antimycobacterial drugs.

1.8 Role of the Clp proteases

The number of Clp proteases varies between organisms; while most bacteria such as *E. coli*, *B. subtilis,* and *Y. enterocolitica* (Gottesman *et al.*, 1993; Pederson *et al.*, 1997; Volker *et al.*, 1994) have a single *clpP* gene, other organisms such *Bacillus thuringiensis* and *Corynobacterium glutamicum* have two ClpP (Engels *et al.*, 2004; Fedhila *et al.*, 2002). Four *clpP* genes are present in the cyanobacterium *Synechococcus* (Schelin *et al.*, 2002) and most *Streptomyces* species have five ClpP (Viala *et al.*, 2000).

Clp proteases participate in protein quality control, by degrading misfolded and aggregated proteins potentially toxic for the cell, which contributes to cellular protein homeostasis (Frees and Ingmer, 1999; Kruger *et al.*, 2000; Thomsen *et al.*, 2002). Degradation of these non-functional proteins is useful to protect the cell but also serve as a way to efficiently recycle the cell limited resources (Keiler *et al.*, 1996). During periods of stress levels of proteins such as heat shock proteins are adjusted to allow the cell to adapt to hostile environmental conditions. Protein levels can be adjusted by modulating the rate of synthesis or the rate of degradation; regulation by proteolysis is more rapid than transcriptional induction since continuous proteolysis can be inhibited quickly (Jenal and Hengge-Aronis, 2003). Clp proteases play a role in various

regulatory processes via controlled proteolysis of key regulatory proteins (Frees *et al.*, 2007; Inobe and Matouschek, 2008).

1.8.1 Role of the Clp proteases in *E. coli*, *B. subtilis* and virulence

More than 50 Clp substrate proteins have been identified in *E. coli* (Flynn *et al.,* 2003). Several substrates are transcriptional regulators including FnR (fumarate-nitrate reductase) and an iron-sulfur cluster regulator (IscR), while others are involved in cell division (GTPase, and FtsZ). The ClpXP complex degrades the stationary phase sigma factor σ^s and six proteins (Fnr, IscR, IscU, AcnB, MoaA, and LipA) which are sensors of oxidative stress. Six ribosomal proteins were identified as ClpXP substrates suggesting that the ClpXP complex may degrade ribosomes when nutrients become limiting, releasing amino acids for new protein synthesis. Nearly all identified proteins reside in the cytoplasm except one outer membrane protein (OmpA) and one inner membrane protein (RseA).

Bacterial toxin–antitoxin (TA) systems are diverse and widespread in prokaryotes. They typically consist of a complex including a small stable toxin and an unstable antitoxin that neutralises the toxin. Reduction of antitoxin levels releases its associated toxin, which can then interfere with a specific cellular target such as mRNA or DNA helicase. TA complexes are used by low-copy or antibiotic resistance plasmids for example to ensure their maintenance (Hayes, 2003). Clp proteases have been implicated in the degradation of antitoxins of several TA complexes. In *E. coli* the ClpAP complex degrades the antitoxin MazE (Aizenman *et al.*, 1996) while the ClpXP complex degrades the antitoxin Phd (Lehnherr and Yarmolinsky, 1995). Additionally, in *Staph. aureus* the ClpCP complex degrades antitoxins of *mazEF*, *axe1-txe1* and *axe2-txe2* TA systems (Donegan *et al.*, 2009).

General protein turnover in *B. subtilis* depends almost exclusively on the Clp protease (Kock *et al.*, 2004). In response to environmental conditions *B. subtilis* undertakes complex developmental processes that either lead to the manifestation of competence (i.e. the ability to be genetically transformed) or to the formation of endospores (Msadek *et al.*, 1998). To decide which of these processes occurs and to ensure their correct relative timing, cells have to integrate a number of signals such as DNA replication status and cell density.

ClpP proteolysis is crucial in the information processing of both developmental pathways since strains lacking *clpP* fail to develop competence and display defects in sporulation (Msadek *et al.*, 1998). The ClpCP complex controls sporulation by degrading the anti-sigma factor SpoIIAB, which prevents expression of sigma factor F (Pan *et al.*, 2001). Additionally *B. subtilis* ClpP mutants are unable to grow at high temperature and are non-motile (Msadek *et al.*, 1998). Spx, a global transcriptional regulator of oxidative stress in several Gram positive bacteria, is a substrate of the ClpXP protease in *B. subtilis* (Nakano *et al.*, 2002). Additionally, a microarray study showed that ClpP has an impact on multiple central regulons of *B. subtilis* involved in virulence, oxidative stress response, metal homeostasis, and SOS DNA repair (Michel *et al.*, 2006). The number of regulators whose activity is controlled by ClpP mediated proteolysis might therefore be underestimated at present. In contrast to *B. subtilis*, *B. thuringiensis* contains two ClpP homologs which control different cellular regulatory pathways: ClpP1 is essential for normal cell division at low temperature, whereas ClpP2 is required for motility and sporulation (Fedhila *et al.*, 2002).

Clp proteases have been linked to virulence and pathogenesis in many organisms. For example, ClpP is essential for intracellular survival and replication of *Listeria monocytogenes* and modulates the expression of listeriolysin O, a major virulence factor required to escape from the phagosomes of macrophages (Gaillot *et al.*, 2000). ClpP is important for *Salmonella typhimurium* to grow under various stress conditions, such as low pH, elevated temperature, and high salt concentrations (Thomsen *et al.*, 2002). In *Staph. aureus* ClpP has an impact on several regulons involved in virulence, heat shock response, oxidative stress response, DNA repair, autolysis, and anaerobic growth (Frees *et al.*, 2003; Michel *et al.*, 2006; Savijoki *et al.*, 2003). Loss of *clpP* in *Strept. pneumoniae* results in attenuation in virulence (Robertson *et al.*, 2002) and ClpP regulates expression of the virulence factors pneumolysin and pneumococcal surface antigen A (Kwon *et al.*, 2003).

Additionally, ClpP is also involved in biofilm formation in various species such as *Pseudomonas fluorescens* and *Staphylococcus epidermidis* (O'Toole and Kolter, 1998; Wang *et al.*, 2007). Either enhancement or diminution of biofilm formation can be observed in *clp* mutants. In *Staph. aureus*, biofilm formation is increased in the absence of ClpP but is reduced in the absence of ClpX or ClpC

(Frees *et al.*, 2004). Conversely *Strept. mutans* strains lacking ClpP, but not strains lacking ClpC, have a reduced capacity to form biofilms (Lemos and Burne, 2002). ClpP therefore may act by controlling the stability or activity of transcriptional regulators of biofilm maturation.

1.8.2 Role of the Clp proteases in actinobacteria

Corynebacterium glutamicum contains two *clpP* genes arranged in an apparent operon and are predicted to be essential (Engels *et al.*, 2004). ClpP1 and ClpP2 are induced upon heat and osmotic stress suggesting a role in stress conditions (Engels *et al.*, 2004). ClpCP and ClpXP complexes degrade GlnK, a protein involved in nitrogen control (Strosser *et al.*, 2004).

In addition to the *clpP1P2* operon most *Streptomyces* species have three other *clpP* genes: *clpP3* and *clpP4* are arranged in an operon, and *clpP5* is monocistronic. ClpP1 degrades the transcriptional activator PopR; PopR is the activator of the *clpP3P4* operon, so when *clpP1* is present this operon is not expressed (Viala *et al.*, 2000). ClpP3 however does not fully perform the role of ClpP1 as *clpP1* deletion mutants fail to form aerial mycelium on various media (de Crecy-Lagard *et al.*, 1999). The role of ClpP5 is, as yet, unknown (Gominet *et al.*, 2011). *Streptomyces* have a complex growth cycle with three major stages: substrate mycelium, aerial mycelium, and sporulation. Disruption of the *clpP1* gene in *S. lividans* and *S. coelicolor* blocks differentiation at the substrate mycelium step, while over-expression of *clpP1* and *clpP2* accelerates aerial mycelium formation in *S. lividans*, *S. albus* and *S. coelicolor,* indicating that *clpP* genes are involved in the formation of aerial mycelium (de Crecy-Lagard *et al.*, 1999). In *S. lividans* strains lacking ClpP1, cell cycle progression is blocked at the early stages of growth (Viala and Mazodier, 2003). Overproduction of ClpX activates production of the antibiotic actinorhodin in *S. lividans* and accelerates its production in *S. coelicolor* (de Crecy-Lagard *et al.,* 1999). Altogether Clp proteases regulate morphological and metabolic differentiation in *Streptomyces* species.

To date, only one substrate protein has been identified in *M. tuberculosis*: RseA, the SigE anti-sigma factor (Barik *et al.*, 2009). SigE regulates expression of genes important for virulence and for responding to environmental stresses and its expression is linked to RseA. In the presence of vancomycin, RseA becomes phosphorylated and the ClpC1P2 complex degrades phosphorylated

RseA so SigE becomes activated. This cleavage is specific to the phosphorylated state of the protein, as degradation was blocked when the protein was dephosphorylated, and exclusive to the ClpC1P2 complex as ClpP1 could not substitute ClpP2 and ClpX could not substitute ClpC1 (Barik *et al*., 2009).

1.9 Recognition signals

The Clp proteases degrade a large range of proteins, however substrate proteins possess recognition signals also called degradation tags or degrons (Varshavsky, 1991). Two main degradation tags are found in the cell: the SsrA tag which is added to the C-terminus of proteins during the process of trans-translation, and the N-degron.

1.9.1 Trans-translation

In bacteria, ribosome stalling is a serious issue for cell survival since cells need to maintain a pool of active ribosomes for translation, and the release of incomplete proteins may be toxic for the cell. Transfer-messenger RNA (tmRNA), encoded by the *ssrA* gene, liberates stalled ribosomes and tag incomplete nascent proteins for degradation (Fig 1.2). Trans-translation is an effective system to release stalled ribosomes at the end of a truncated mRNA that lacks a stop codon, at stretches of rare codons, or at inefficient termination codons (Keiler *et al.*, 1996; Li *et al.*, 2006; Roche and Sauer, 1999). tmRNA is a large molecule (260 to 430 nucleotides depending on bacterial species) with both tRNA and mRNA activities. It contains a tRNA like domain, a series of pseudoknots and a short internal open reading frame (mRNA like domain) of 8-35 codons ending with a stop codon (Keiler *et al*., 1996; Williams and Bartel, 1996). A two-piece tmRNA is present in *Caulobacter crescentus* and some cyanobacteria as a result of gene permutation and rearrangement and displays a reduction in pseudoknot number and a linear tag reading frame (Gaudin *et al.*, 2002; Keiler *et al.*, 2000; Williams, 2002). tmRNA is one of the most abundant RNAs in the cell and its half-life exceeds the doubling time of the bacterium in *E. coli* (Hallier *et al.*, 2004). *E. coli* has 500-1000 copies of tmRNA per cell, corresponding to one tmRNA for every ten to 20 ribosomes (Moore and Sauer, 2005); tmRNA levels in *M. smegmatis* are equivalent to *E. coli*, which suggests a high rate of trans-translation in mycobacteria (Andini and Nash, 2011; Lee *et al.*, 1978). In *E. coli* it is estimated that the synthesis of

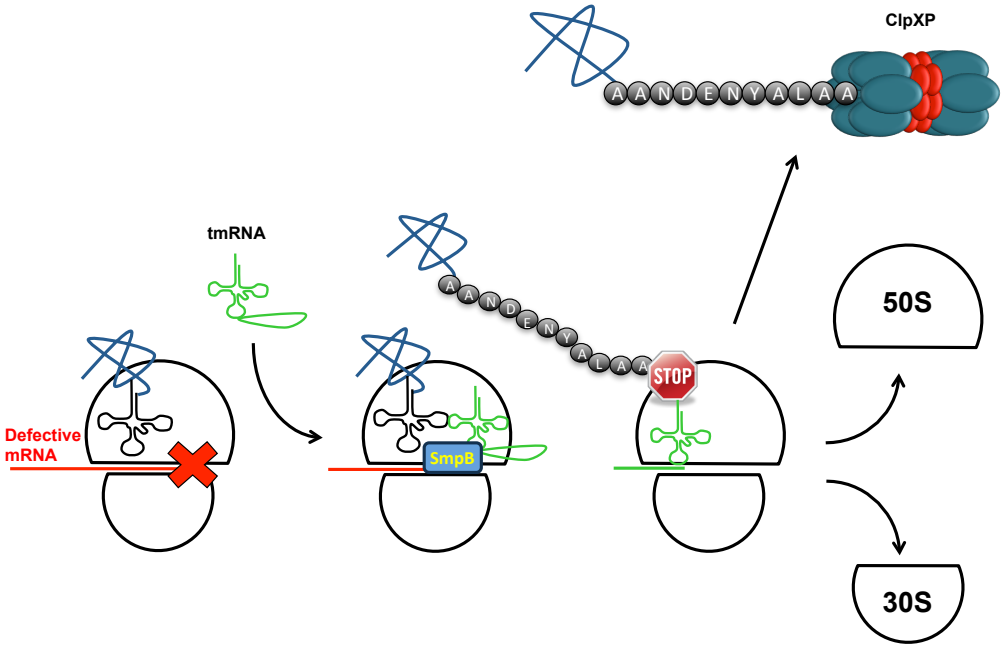

Figure 1.2 The trans-translation model of tmRNA activity

Transfer-messenger RNA (tmRNA) enters in the vacant A site of the stalled ribosome with small protein B (SmpB). The defective mRNA is released and the mRNA region of tmRNA becomes the template for translation until a stop codon is reached. The ribosome is subsequently released and the protein now harbouring an 11 amino acids tag is degraded by the ClpXP complex or other proteases.

about 0.4% of all proteins terminates with tagging during normal exponential growth (Moore and Sauer, 2005)

1.9.1.1 The trans-translation model

The majority of tmRNA is believed to be in a complex with the small basic protein B (SmpB) wich binds tmRNA with high affinity (Keiler, 2008). tmRNA can be charged with alanine by alanyl-tRNA synthetase, the $tmRNA^{ala}$-SmpB complex is recognised by elongation factor Tu (EF-Tu) and GTP to form a complex that recognises stalled ribosomes. GTP is hydrolysed, EF-Tu is released and the tRNA domain of tmRNA enters the unoccupied A site of a stalled ribosome. tmRNA first functions as a tRNA to transfer its pre-charged alanine to the nascent polypeptide in a transpeptidation reaction. The defective mRNA is then replaced by the mRNA region of tmRNA so translation resumes. Ribosomal protein S1 binds tmRNA with high affinity to the mRNA domain and is possibly required for switching the reading frame from the original mRNA to the mRNA domain of tmRNA (Wower *et al.*, 2000). Termination of translation occurs at the stop codon of the new mRNA and the ribosome is released. The nascent protein, now harbouring a C-terminal tag, usually 10-11 amino acids long, is degraded by proteases.

SmpB is an essential component for trans-translation. SmpB is important for tmRNA stability (Hallier *et al.*, 2004; Hanawa-Suetsugu *et al.*, 2002; Moore *et al.*, 2003); in *C. crescentus* SmpB protects tmRNA from degradation *in vitro* and the *in vivo* levels of SmpB during the cell cycle correlate with tmRNA stability (Hong *et al.*, 2005). Additionally SmpB enhances tmRNA aminoacylation by making contacts with alanyl-tRNA synthetase (Hanawa-Suetsugu *et al*., 2002; Shimizu and Ueda, 2002) and is required for stable association of tmRNA with the ribosome, it overcomes the absence of codon-anticodon interactions (Felden and Gillet, 2011; Nonin-Lecomte *et al.*, 2009).

The SsrA tag does not affect the structure or thermodynamic stability of attached proteins, suggesting that it causes degradation simply by providing a recognition site for protease binding (Karzai *et al.*, 2000). Five different protease complexes (Tsp, FtsH, ClpXP, ClpAP and Lon) degrade SsrA-tagged proteins in *E. coli* (Choy *et al.*, 2007; Gottesman *et al.*, 1998; Herman *et al.*, 1998; Spiers *et al.*, 2002). However the ClpXP complex is thought to be the major protease responsible for their degradation in the cell due to the binding of

the adaptor protein SspB which binds to ClpX and a part of the tag (Lies and Maurizi, 2008; Wiegert and Schumann, 2001). SspB binds to the first seven amino acids of the tag sequence while ClpX binds to the terminal three residues and enhances ClpXP degradation while inhibiting degradation by ClpAP (Farrell *et al.*, 2005; Flynn *et al.*, 2001; Levchenko *et al.*, 2000). No similar adaptors have been identified in Gram positive bacteria, and in *Streptococcus pneumoniae* SsrA-tagged proteins were found to be degraded by ClpXP without the participation of an adaptor protein (Ahlawat and Morrison, 2009).

1.9.1.2 Functions of the tmRNA pathway

SmpB and tmRNA have been identified in all bacterial genomes including those with reduced genome sizes (Mao *et al.*, 2009; Williams, 2002b) suggesting a conserved role amongst bacteria. However, tmRNA is not required for growth under normal conditions of many bacteria including *E. coli* (Oh and Apirion, 1991), *B. subtilis* (Muto *et al.*, 2000), *C. crescentus* (Keiler and Shapiro, 2003), *Salmonella enterica* (Julio *et al.,* 2000) and *Yersinia pseudotuberculosis* (Okan *et al.*, 2006); but it does play an important role in cell growth and resistance to adverse conditions. For example *E. coli ssrA* deletion cells recover more slowly from carbon starvation, have reduced motility, are more sensitive to sublethal concentrations of antibiotics that inhibit protein synthesis and display a constitutive heat shock response (Abo *et al.*, 2002; Komine *et al.*, 1994). *B. subtilis ssrA* deletion strains have growth defects in high temperature, low temperature and sporulation (Muto *et al*., 2000). Disruption of *ssrA* in *Y. pseudotuberculosis* and *Sal. enterica* reduces their virulence and in *Streptomyces ssrA* and *smpB* mutant cells show growth and sporulation defects and are more sensitive to hygromycin (Yang and Glover, 2009). In contrast, *ssrA* is essential in *Neisseria gonorrhoeae* (Huang *et al.*, 2000) *Helicobacter pylori* (Thibonnier *et al.*, 2008) and is predicted to be essential in *Mycoplasma pneumoniae*, *Mycoplasma genitalium* and *Haemophilus influenzae* (Akerley *et al.*, 2002; Hutchison *et al.*, 1999).

The tmRNA-SmpB system mainly serves two functions: it releases stalled ribosomes and provides a general quality control system that promotes the degradation of incomplete proteins. A tmRNA variant which mediates the addition of a peptide tag (tmRNA-DD) that does not result in degradation can be used to distinguish the two functions of tmRNA when expressed in *ssrA*

deletion mutants; from this it has been identified that it is the ribosomal recycling and not the proteolysis tagging function that is required for viability in *N. gonorrhoeae* (Huang *et al.*, 2000; Withey and Friedman, 1999). In addition, the tmRNA system may have a regulatory role since some phenotypes generated by mutations in tmRNA can relate to misregulation of individual substrates. For example, *Y. pseudotuberculosis* cells lacking tmRNA display a defect in the secretion of virulence factors possibly resulting from the misregulation of the transcriptional factor VirF that controls synthesis of the secretion apparatus (Okan *et al*., 2006) and tmRNA regulates synthesis of the stress sigma factor RpoS in *E. coli* (Ranquet and Gottesman, 2007). Additionally tmRNA was shown to act as an antisense RNA to regulate expression of CrtM/N and influence the pigment synthesis of *Staph. aureus* (Liu *et al*., 2010).

1.9.1.3 The tmRNA pathway in *M. tuberculosis*

M. tuberculosis contains *ssrA* (Tyagi and Kinger, 1992) and *smpB* genes confirming the presence of tmRNA tagging system in this species. *smpB* is not predicted to be essential whilst there are no data available for *ssrA* (Sassetti *et al.*, 2003). The *M. tuberculosis* SsrA tag is 12 residues long: (A)ADSHQRDYALAA and this sequence is conserved in most mycobacterial species. *smpB* was found to be down-regulated in a SenX3-RegX3 deletion mutant strain; smpB expression levels are therefore predicted to be controlled by this two-component system (Parish *et al.*, 2003). As the tmRNA pathway is absent from eukaryotes it may be a promising drug target. Recently pyrazinamide was found to target a component of trans-translation. Pyrazinamide is a pro-drug, which is currently used in TB chemotherapy, and requires conversion into its active form pyrazinoic acid (POA) by the bacterial pyrazinamidase. POA was found to bind to RpsA, which is involved in tmRNA tagging system, suggesting that pyrazinamide is effective due to its effect on the tmRNA system (Shi *et al.*, 2011) and confirming that the tmRNA pathway is an attractive drug target for novel antibacterial.

1.9.1.4 Alternative ribosome rescue pathways

Several bacteria have some tmRNA-independent rescue pathways to recycle stalled ribosomes. For example in the absence of tmRNA, ArfA (for Alternative Ribosome rescue Factor; former YhdL) is required for the growth of *E. coli* cells and rescues stalled ribosomes (Chadani *et al.*, 2010). ArfA binds to the large

subunit of ribosomes and enhances the hydrolysis of peptidyl-tRNA residing in the stalled ribosome, although the molecular mechanism of how it rescues ribosomes is still unknown. Moreover, ArfA synthesis is regulated by tmRNA tagging and proteolysis so ArfA-mediated ribosome rescue is increased in response to decreased tmRNA activity (Chadani *et al.*, 2011a). The combination of *ssrA* and *arfA* mutations is lethal; however over-expression of the YaeJ protein supports growth of cells lacking *ssrA* and *arfA* (Chadani *et al.*, 2011b). YaeJ was shown to rescue stalled ribosomes independently of SsrA and ArfA demonstrating a third alternative ribosome rescue system.

1.9.2 N-end rule pathway

The N-degron is a second degradation tag present in eukaryotes and prokaryotes. The N-end rule states that the stability of a protein is determined by the identity of its N-terminal residue, with proteins commencing with large residues preferred for degradation (Bachmair *et al.*, 1986; Varshavsky, 1996). *E. coli* destabilising residues are organised into two levels: primary and secondary. The primary destabilising residues are aromatic amino acids (tryptophan, phenylalanine, and tyrosine) and leucine, while secondary destabilising residues are basic amino acids arginine and lysine (Tobias *et al.*, 1991). In bacteria protein synthesis starts with formyl-methionine (fMet) considered to be a stabilising residue, thus the generation of a N-degron requires some post-translational modifications. Several different models have been proposed for the generation of N-degron *in vivo* (Dougan *et al.*, 2010). The first model involves the removal of the N-terminal fMet on the target protein to reveal the N-degron. Alternatively, the N-terminal fMet is untouched but a primary destabilising residue is attached to the N-terminus of the protein. The remaining model involves processing of a pre-N-degron (containing a stabilising N-terminal residue) by an unknown endopeptidase resulting in exposure of a primary destabilising residue or resulting in exposure of a secondary residue onto which a primary residue is attached (Dougan *et al.*, 2010).

The N-end rule pathway has largely been defined using artificial substrates but two natural substrates have been identified in *E. coli*: DpS (DNA protection during starvation) and putrescine-aminotransferase (PATase) (Schmidt *et al.*, 2009). DpS, as the name suggests, protects DNA during starvation and also oxidative stress, whilst PATase is involved in the catabolism of putrescine. Interestingly, DpS is also degraded as a full-length protein by the ClpXP

protease complex which targets the first residues of the N-terminal segment (Schmidt *et al.*, 2009).

The protein ClpS is an essential component of the *E. coli* N-end rule pathway. In *E. coli* ClpS specifically binds to destabilising N-terminal residues and mediates degradation of the selected substrate by the ClpAP complex (Erbse *et al.*, 2006; Schmidt *et al*., 2009). Structurally ClpS is a 'cone-shaped' molecule with a coiled N-terminal extension; the N-degron binding site is located at the base of the cone, while the tip docks to the N-domains of ClpA (Guo *et al.*, 2002; Zeth *et al.*, 2002). ClpS binding to ClpAP prevents recognition and degradation of other types of substrates by ClpAP (Dougan *et al*., 2002). ClpS binds more tightly to ClpA in presence of N-end substrates so it does not prevent degradation of other proteins in absence of such substrates (Roman-Hernandez *et al.*, 2011). *M. tuberculosis* has the adaptor protein ClpS (Rv1331) suggesting the N-end rule pathway is present in mycobacteria. ClpS is predicted non essential for mycobacterial growth (Sassetti *et al*., 2003) but is predicted to be required for survival in primary murine macrophages (Rengarajan *et al.*, 2005).

1.10 *M. smegmatis* and *M. marinum* as model organisms

Despite the recent advancements in genetic tools available for studying *M. tuberculosis*, it remains a difficult organism to work with. Related mycobacteria such as *M. smegmatis* and *M. marinum,* are frequently used as model organisms to study the biology of *M. tuberculosis* (Stinear *et al.*, 2008). *M. smegmatis* was first isolated in 1824 by Lustgarten and in 1885 was isolated from human smegma by Alvarez and Tavel (Alvarez and Tavel, 1885). The genome of *M. smegmatis* is about 1.7 times bigger than that of *M. tuberculosis*. It is a fast-growing non-pathogenic mycobacterium and has a generation time of three to four hours compared to 24 hours for *M. tuberculosis*. It does not infect humans and it does not require ACDP Hazard Group 3 containment. However, because it is saprophytic and not pathogenic, it cannot be used as a model for pathogenesis (Reyrat and Kahn, 2001).

M. marinum can infect more than 150 species of fish and is an occasional zoonotic human pathogen, with transmission to humans occurring through direct skin inoculation. *M. marinum* has more than 85% nucleotide identity with *M. tuberculosis* and the two organisms share virulence determinants (Stinear *et*

al., 2008). *M. marinum* is a slow growing organism with a generation time of ten to 12 hours but its optimal growth temperature is 30°C (Shiloh and DiGiuseppe Champion, 2009).

M. smegmatis and *M. marinum* have a total of 153 and 125 proteases respectively including the cytoplasmic Lon protease, which is absent in *M. tuberculosis* (Rawlings *et al.,* 2010). *M. smegmatis* and *M. marinum* have two *clpP* homologues arranged in tandem similarly to *M. tuberculosis* (Kapopoulou *et al.*, 2011). The ATPase subunits ClpX, ClpC1, and ClpC2 are also present in both organisms, as well as the adaptor protein ClpS and the tmRNA tagging components *ssrA* and SmpB, in both organisms. Thus these two organisms appear to be good models for unravelling the role of the Clp proteins in mycobacteria.

1.11 Aims and objectives

Bacterial species have evolved multiple proteolytic systems, and the genome of *M. tuberculosis* encodes over 140 proteases including the proteasome, which is absent from most bacteria. While most bacteria have a single *clpP* gene, *M. tuberculosis* has two ClpP protease subunits (ClpP1 and ClpP2), with at least one of them essential for growth and activators of ClpP activity have been shown to have activity against *M. tuberculosis* (Ollinger *et al.*, 2011). Although Clp proteases have been extensively studied in organisms such as *E. coli* and *B. subtilis*, little is known regarding their function in mycobacteria and in *M. tuberculosis* in particular.

The aim of this project was to determine why *M. tuberculosis* has two ClpP proteases and why they are essential for mycobacterial growth, although other systems such as the proteasome are present. We hypothesised that ClpP1 and ClpP2 perform different functions, as in other bacteria with multiple ClpP proteases (Fedhila *et al*., 2002), or may be active under different conditions.

The objectives of this project were to determine:

- the mechanism of activation of the ClpP proteases in *M. tuberculosis* and identify the conditions that influence their expression

- the specific interaction of the two ClpP subunits with the other members of the complex

- the effect of the ClpP proteases over-expression on the transcriptome of *M. tuberculosis*

- the specificity of ClpP1 and ClpP2 substrate recognition

- the role of the accessory components ClpS, SmpB and SsrA

2 Materials and methods

Unless otherwise stated all chemical reagents were obtained from Sigma-Aldrich.

2.1 Bacterial growth

2.1.1 Media

E. coli DH5α and *E. coli* BTH101 (Karimova *et al.*, 1998) were grown at 37°C with shaking at 225 rpm in Luria-Bertani (LB) broth or in M63 medium (2 g L^{-1} $(NH_4)_2SO_4$, 13.6 g L^{-1} KH_2PO_4, 0.5 mg L^{-1} $FeSO_4$.7H2O and 1 mg L^{-1} vitamin B1, pH 7.0) for liquid cultures or with Bacto agar (15g L^{-1}) for solid media. Ampicillin was used at 100 mg L^{-1}, kanamycin at 50 mg L^{-1}, isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.5 mM and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at 50 mg L^{-1} where required.

M. smegmatis mc^2155 (Snapper *et al.*, 1990) and *M. smegmatis* ΔHisA (O'Hare *et al.*, 2008) were grown at 37°C with shaking at 160 rpm in Lemco medium (10 g L^{-1} peptone, 5 g L^{-1} Lemco powder, 5 g L^{-1} NaCl) containing 0.05% w/v Tween 80 for liquid cultures or with Bacto agar (15g L^{-1}) for solid media. For Split-Trp growth assays, *M. smegmatis* ΔHisA was grown on 7H9 agar (5 g L^{-1} Difco™ Middlebrook 7H9 powder (Becton Dickinson), 15 g L^{-1} agar) supplemented with 0.2% w/v glucose, 0.2% v/v glycerol and 60 mg L^{-1} histidine at 30°C.

M. marinum strain M (ATCC BAA-535), was obtained from N. Andreu, Imperial College, London and grown at 30°C with shaking at 100 rpm in Middlebrook 7H9 medium supplemented with 10% v/v oleic-albumin-dextrose-catalase (OADC) supplement (Becton Dickinson) and 0.05% w/v Tween80 (7H9-Tw-OADC) or on Middlebrook 7H10 agar (Becton Dickinson) with 10% v/v OADC supplement.

M. tuberculosis H37Rv was grown at 37°C in 7H9-Tw-OADC or in 7H9-Tween supplemented with AD (5% w/v bovine serum albumin fraction V, 2% w/v glucose) or on Difco™ Middlebrook 7H10 agar supplemented with 10% v/v OADC. Cultures were grown without agitation in 50 mL tubes unless otherwise stated. For LacZ reporter system experiments *M. tuberculosis* was grown in 7H9-tw supplemented with 10% v/v BSA and 0.1% w/v succinate with or without 0.1% w/v acetamide.

For growth under hypoxic conditions, *M. tuberculosis* was grown in Dubos medium (Difco, Becton Dickinson) supplemented with Dubos Medium Albumin (Becton Dickinson) and 0.05% w/v Tween®80.

Gentamicin was used at 10 mg L^{-1}, hygromycin B at 100 mg L^{-1}, kanamycin at 20 mg L^{-1}, streptomycin at 20 mg L^{-1} where required.

2.1.2 Growth of mycobacteria

Cultures of *M. smegmatis* and *M. marinum* were prepared as follows: a loopful of cells from a plate was used to inoculate 5 mL of medium and grown overnight; cultures were then inoculated 1/10 in 5 mL medium.

All work involving *M. tuberculosis* was carried out in a Containment Level 3 laboratory equipped for Advisory Committee on Dangerous Pathogens (ACDP) Hazard Group 3 work. To prepare seed cultures of *M. tuberculosis*, a loopful of cells from a plate was used to inoculate 10 mL medium containing 3 mL of 1 mm glass beads (BioSpec Products Inc.) and incubated at 37°C without agitation for 14 d. 1 mL of this culture was used to inoculate 10 mL medium and grown without agitation for 7 d.

To assay promoter activity, a seed culture was used to inoculate 10 mL 7H9-Tw-AD medium to a starting OD_{580} of 0.1. Liquid cultures were grown until an OD_{580} of 0.6-0.8 was reached and cell-free extracts were prepared. For growth profile of promoter activity, 100 mL of medium in a 450 cm^2 roller bottle was inoculated with 3 mL of a seed culture and incubated rolling at 37°C. The OD_{580} of the roller culture was measured and cell-free extracts were prepared every day.

To measure LacZ turnover in cells carrying the acetamidase promoter, a seed culture was used to set up 10 mL medium with an OD_{580} of 0.1. Liquid cultures were grown in 7H9-Tween-BSA with 0.1% w/v sodium succinate with or without 0.1% w/v acetamide until an OD_{600} of 0.6-0.8 was reached and cell-free extracts were prepared. To measure LacZ turnover in cells carrying the tetracycline inducible promoter, a seed culture was used to set up 100 mL roller culture with an O.D of 0.025. Cultures were grown in 7H9-Tw-AD with or without 150 or 300 ng/mL anhydrotetracycline (ATc) for 7 d. Cell-free extracts were prepared from

10 mL of this culture, while 50 mL were spun, washed three times in 7H9 medium and used to inoculate new 100 mL roller culture starting with an O.D of 0.05.

For growth in the Wayne model of hypoxia (Wayne and Hayes, 1996), liquid cultures were grown in DTA in 20 mm x 125 mm round bottom screw cap glass tubes. To create a theoretical starting OD_{580} of 0.004, 170 µL of a culture with an OD_{580} of 0.4 was used to inoculate 17 mL medium. Caps were tightly screwed on and sealed with parafilm. Cultures were grown stirring at 150 rpm with a 3 x 8 mm magnetic stirring bar. A control tube containing methylene blue (1.5 µg mL^{-1}) was used as a visual indication of oxygen depletion. Fading of the methylene blue dye indicated onset of microaerophilic conditions (Non-replicating Phase stage 1) and decolourisation signified onset of anaerobic conditions (Non-replicating Phase stage 2). For reaeration, 3 mL of hypoxic cultures were used to inoculate 100 mL of fresh medium in roller bottles and incubated with rolling at 37°C.

Aerobic growth curves in *M. tuberculosis* were carried out in 4 mL 7H9-Tw-AD medium in 16 x 125 mm round bottom screw cap glass tubes with a starting OD_{580} of 0.02. Cultures were stirred with 3 x 8 mm magnetic stirring bars at 150 rpm.

2.2 Nucleic acid isolation

2.2.1 Extraction and purification of genomic DNA

M. tuberculosis cells were harvested by centrifugation for 10 min at 2700 x *g*, resuspended in 1 mL 10 mM Tris-Cl pH 7.5, and added to 2 mL lysing matrix B tubes (MP Biomedicals) on ice. Cells were disrupted using one 30 s cycle at speed 6.0 using a FP120 FastPrep (Qbiogene). Samples were spun at 16000 x *g* for 4 min and the supernatants were filter-sterilised through a 0.2 µm filter unit. Samples were incubated at 55°C for a minimum of 90 min with 0.2 vol of 5% w/v sodium deoxycholate. An equal volume of phenol:chloroform:isoamyl acohol 25:24:1 was added, the mixture was vortexed for 30 s and spun at 16000 x *g* for 2 min. The top layer was transferred to a new 1.5 mL microcentrifuge tube and the phenol-chloroform extraction was repeated 1-3 times as required. An equal volume of chloroform was added to the samples and the mixture was vortexed for 30 s and spun for 2 min at 16000 x *g*. The top

layer was transferred to a new 1.5 mL tube and ethanol precipitation was carried out by adding 0.1 vol of 3 M sodium acetate buffer pH 5.2, and 2.5 vol of 100% ethanol. Samples were incubated at -80°C for a minimum of 1 h and spun at 16000 x *g* for 15 min. Supernatant was removed and the pellet was re-suspended in 1 mL of ice cold 100% ethanol. Samples were centrifuged for 2 min at 16000 x *g* and the supernatant was removed. The pellet was dried for 5 min at 30°C in a vacuum concentrator and re-suspended in 50 µL SDW. DNA concentration was determined using a ND-1000 spectrophometer (NanoDrop).

2.2.2 Extraction of mycobacterial DNA for colony PCR

One loopful of cells from a plate was added to 1 mL Tris buffer and incubated for 10 min at 100°C. Samples were left to cool for 5 min and filter-sterilised through 0.2 µM filter units.

2.2.3 Extraction and purification of mycobacterial RNA

RNase and DNase-free filtered tips, water and microcentrifuge tubes were used for all RNA work. RNA was extracted as previously described (Rustad *et al.*, 2009b). Briefly, *M. tuberculosis* cells were harvested by centrifugation for 10 min at 2700 x *g*. The pellets were resuspended in 1 mL TRI Reagent (Invitrogen) and added to 2 mL tubes containing lysing matrix B. Cells were disrupted using two 20 s cycles at speed 4 using a FastPrep, placed on ice for 5 min, spun at 16000 x *g* for 4 min and filter-sterilised through a 0.2 µm filter unit. The solutions were added to 2 mL Heavy Phase Lock Gel (5 Prime) snap cap tubes containing 300 µL chloroform:isoamyl alcohol (24:1). Samples were mixed by inverting rapidly for 15 s and placed on ice. Once all samples were transferred, inverting was continued periodically for 2 min. Samples were spun for 5 min at 16000 x *g* and the aqueous layer (~540 µL) was transferred to a 1.5 mL tube containing 270 µL ispopropanol and 270 µL high salt solution (0.8 M sodium citrate, 1.2 M NaCl). Tubes were inverted several times to mix. Samples were left to precipitate overnight at 4°C and centrifuged for 10 min at 16000 x *g* to remove the supernatant. The pellets were then washed with 1 mL 75% ethanol, tubes were inverted and spun for 5 min before decanting the ethanol. Pellets were resuspended in 100 µL RNase-free water.

RNA was purified and DNase digested using the RNeasy Minikit (QIAGEN) protocol as follows. RLT buffer was prepared by adding 10 µL ß-mercaptoethanol per mL of buffer prior to use and 350 µL RLT buffer was

added to the sample and mixed. 265 µL of 95% ethanol was added and mixed by vortexing. The mixture was transferred to an RNeasy spin column, centrifuged for 15 s and transferred to a new 2 mL collection tube. 350 µL of buffer RW1 was added, the column spun for 15 s and the flow through was discarded. 70 µL of buffer RDD was added to a 10 µL aliquot of DNase I stock solution and pipetted directly onto the column membrane. Digestions were carried out at RT for 1 h. 350 µL buffer RW1 was added and the column spun for 15 s before adding 500 µL RPE buffer. The column was spun for 15 s and the flow through was discarded. An additional 500 µL RPE buffer was added and the column, centrifuged for 2 min and the flow through was discarded and the column spun for 1 min to dry completely. The column was transferred to a new 1.5 mL microcentrifuge tube and RNA eluted in 40 µL RNAse-free water. RNA concentrations were measured on a ND-1000 Spectrophometer. To check for RNA degradation, 1 µL of RNA sample was run on a 2% w/v agarose gel.

2.3 Polymerase chain reaction (PCR)

Reactions with GoTaq Green Master Mix (Promega) (containing *Taq* DNA polymerase, buffer, $MgCl_2$ and dNTPs) were carried out in a total volume of 20 µL with 10-100 ng of DNA template, 1 µL of each 10 µM primer, 10 µL 2x Master Mix and SDW. Reactions with *Pfu* high fidelity DNA polymerase (Promega) were carried out in a total volume of 50 µL containing 5 µL 10X buffer, 5 µL DMSO, 1 µL 10 mM dNTPs, 10-100 ng of DNA template, 2.5 µL of each 10 µM primer, 0.5 µL *Pfu* DNA polymerase and SDW. PCRs were carried out on a thermocycler TC-312 (Techne, Duxford, UK). For amplification of inserts, PCR program consisted of: initial denaturation at 94°C for 1 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at optimal temperature for 30 s and extension at 72°C for 1 min/kb. Final extension at 68°C for 5 min was followed by hold at 4°C.

2.4 Limiting dilution RT-PCR

SuperScript II reverse transcriptase (Invitrogen) was used for the preparation of cDNA according to the manufacturer's instructions. Two tubes were prepared for each sample. In a total volume of 12 µL, 1 µg of RNA was mixed with 50 ng random primers (Invitrogen) and 1 µL of 10 mM deoxyribonucleotide triphosphates (dNTPs) mix (Promega). The mixture was heated at 65°C for 5 min and quickly chilled on ice. 4 µL of 5X First Strand Buffer, 2 µL 0.1 M dithiothreitol (DTT) and 1 µL RNasin® (Promega) were added, mixed gently and

incubated at 25°C for 10 min. 1 µL of reverse transcriptase was added to one tube and 1 µL DNAse- and RNAse-free water was added to the negative control tube. Samples were mixed and incubated at 25°C for 10 min, 42°C for 50 min and 70°C for 15 min.

PCR was carried out using undiluted cDNA and serial four-fold dilutions (1:4, 1:16, 1:64, 1:256, and 1:1,026) with GoTaq Green Master mix and primers listed in Appendix 1.

2.5 Cloning PCR products into pSC-A and pSC-B vectors

A-tailed PCR products were cloned into pSC-A while blunt PCR products were cloned into pSC-B using the StrataClone PCR cloning kit (Stratagene). A reaction mixture was set up using 3 µL buffer, 10 ng PCR product and 1 µL Strataclone vector mix. After 1 h of incubation at RT, the reaction was transformed into StrataClone SoloPack competent cells as follows. Cells were thawed on ice, 1 µL of the reaction was added and mixed gently. Cells were incubated on ice for 20 min, heat shocked at 42°C for 45 s and incubated on ice for 2 min. Cells were recovered in 250 µL pre-warmed SOC medium shaking at 225 rpm at 37°C for 1 h. Samples of 100 µL, 50 µL, and 5 µL of cells were plated onto LB agar plates containing the appropriate selection antibiotic plus X-gal and incubated overnight. White or light blue colonies were picked.

2.6 Cloning PCR products into pGEM vector

A-tailed PCR products were cloned into pGEM T-Easy Vector System (Promega). A reaction mixture was set up using 5 µL ligation buffer, 1 µL vector, 1 µL T4 DNA ligase, 3 µL PCR product and SDW in a total volume of 10µL. After 1 h of incubation at RT, the reaction was transformed into subcloning efficency DH5α competent cells (Invitrogen). Samples were plated onto LB plates containing the appropriate selection antibiotic plus X-gal and incubated overnight; white or light blue colonies were picked.

2.7 Blunting of sticky-end DNA fragments

Sticky ends of DNA fragments were converted to blunt ends by mixing 1 µL *Pfu* high fidelity DNA polymerase (Promega), 2 µL of 10X buffer, 0.5 µL 10 mM dNTPs with 100-500 ng DNA and SDW in a 20 µL reaction. The mix was incubated for 25 min at 72°C. Samples were purified using a clean and concentrate kit (Zymo).

2.8 A-tailing of DNA fragments

DNA fragments were A-tailed by mixing 4 µL of 5X buffer with 0.4 µL 1 mM deoxyadenosine triphosphates (dATPs) (Promega), 1 µL *GoTaq* DNA polymerase (Promega) and 100-500 ng DNA in a total volume of 20 µL. The mix was incubated for 25 min at 72°C. Samples were purified using a clean and concentrate kit (Zymo).

2.9 Restriction enzyme digests

For a 20 µL reaction, the digestion mixture contained 200 ng of plasmid DNA, 2 µL of 10X enzyme buffer, 1-2 µL of enzyme and SDW. Digests were incubated at 37°C for 1-2 h. For a 50 µL reaction, the digestion mixture contained 1-5 µg of plasmid DNA, 5 µL of 10X enzyme buffer, 2-3 µL of enzyme and SDW. Where appropriate, 2 µL Alkaline Phosphatase (Promega) was added as well as 5 µL of 10X buffer. Digests were incubated at 37°C overnight.

2.10 DNA extraction from agarose gels

DNA fragments were separated on 0.8% w/v or 1% w/v agarose gels made with UltraPure™ agarose (Invitrogen) in 1X TAE (242 g L^{-1} Tris base, 37.2 g L^{-1} Na_2EDTA and 57.1 mL L^{-1} glacial acetic acid pH 8.3) with 0.5 µg mL^{-1} ethidium bromide. Orange G buffer (1X) was used to load samples (2.5 g L^{-1} Orange G sodium salt, 40% w/v glycerol, 60% w/v TE buffer).

DNA was extracted from agarose gels using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. The DNA fragment was excised using a scalpel and dissolved in 3 vol of buffer QG for 10 min at 50°C. 1 gel vol of isopropanol was added and the mixture was transferred to a QIAquick spin column. The DNA was bound to the column by centrifugation at 16000 x *g* for 1 min followed by a wash with 750 µL buffer PE and a centrifugation at 16000 x *g* for 1 min. DNA was eluted by adding 20 µL of SDW to the column and centrifugation for 1 min.

2.11 Cleaning and concentrating of DNA

Small quantities of DNA (≤5 µg) were cleaned and concentrated using the "Zymo 5" DNA Clean and Concentrator-5 kit (Zymo Research Corporation) while large quantities of DNA (between 5 µg and 25 µg) were cleaned and concentrated using the "Zymo 25" DNA Clean and Concentrator-25 kit

according to the manufacturer's instructions. Briefly, in a 1.5 mL microcentrifuge tube, 2 vol of DNA binding buffer was added to each vol of DNA sample and mixed briefly by vortexing. The mixture was transferred to a Zymo-Spin Column in a collection tube. The tubes were centrifuged at 16000 x *g* for 30 s and the flow-through discarded. 200 µL of wash buffer was added to the column, spun at 16000 x *g* for 30 s, and the wash step was repeated. Small quantities of DNA were eluted by applying in 10 µL of SDW to the column while large quantities of DNA were eluted in 20-35 µL of SDW. DNA concentrations were measured on a ND-1000 spectrophometer.

2.12 Ligation of DNA fragments

Ligations were performed using the Rapid DNA Dephos & Ligation kit (Roche) according to the manufacturer's instructions with a vector:insert ratio of 1:3. Reactions were carried out in a total volume of 20 µL containing 50 ng of linearised vector, a corresponding amount of insert, 2 µL of 5X dilution buffer and SDW and 10 µL of 2X ligation buffer. 1 µL of DNA ligase was added and the ligation reaction was incubated at RT for 1 h before transformation into competent *E. coli* cells.

2.13 Transformation of *E. coli*

Subcloning efficency DH5α competent cells (Invitrogen) were used for cloning. DNA (1-10 ng) was added to 50 µL of cells thawed on ice and mixed gently. Cells were incubated on ice for 30 min, heat shocked for 20 s at 42°C and placed on ice for 2 min. Cells were diluted into pre-warmed 0.9 mL of LB medium and incubated with shaking at 225 rpm at 37°C for 1 h; 100 µL of cells and the pellet were plated onto an LB agar plate containing the appropriate selection antibiotic and grown overnight at 37°C.

2.14 Preparation of plasmid DNA

Small scale plasmid DNA preparations were carried out using the Wizard Plus SV Minipreps DNA Purification Systems kit (Promega) according to the manufacturer's instructions as follows. A 5 mL overnight culture was pelleted by centrifugation and the pellet was resuspended in 250 µL cell resuspension solution. 250 µL of cell lysis solution was added and mixed by inversion; 10 µL of alkaline protease was added, mixed by inversion, and incubated at RT for 5 min. To stop the reaction, 350 µL of neutralisation buffer was added and mixed by inversion. The mix was centrifuged at 16000 x *g* for 10 min and the clear

lysate was transferred into a spin column and spun at 16000 x *g* for 1 min. The column was washed by adding 750 µL of wash solution and spun at 16000 x *g* for 1 min. The wash step was repeated with 250 µL wash solution and the column spun for 2 min. DNA was eluted in 50 µL of SDW and stored at -20°C.

Large scale plasmid DNA preparations were carried out using the HiSpeed Plasmid purification kit (QIAGEN) according to the manufacturer's instructions as follows. A 100 mL overnight culture was pelleted by centrifugation (2700*x g* for 10 min) and the pellet was resuspended in 6 mL of buffer P1 containing RNase A. 6 mL of buffer P2 was added, mixed by inversion and the mixture was incubated at RT for 5 min. To stop the reaction, 6 mL of buffer P3 was added and mixed by inversion. The lysate was transferred into a QIAfilter cartridge, incubated for 10 min at RT and filtered into a HiSpeed tip pre-equilibrated with 4 mL of buffer QBT. The HiSpeed tip was then washed with 20 mL buffer QC and DNA was eluted with 5 mL of buffer QF. 3.5 mL isopropanol were added to precipitate the DNA and the mixture was incubated at RT for 5 min. The solution was filtered through a QIAprecipitor and washed with 2 mL of 70% ethanol. DNA was eluted in 1 mL of SDW and stored at -20°C.

2.15 Site directed mutagenesis

Site directed mutagenesis (SDM) was carried out in 50 µL total volume containing 2.5 units *PfuUltra* Hot Start high fidelity DNA polymerase (Stratagene), 1X buffer, 0.5 mM dNTPs, 125 ng of mutagenic primers and 10 ng plasmid template. The PCR program consisted of: 95°C for 30 s, followed by 18 cycles of 95°C for 30 s, 60°C for 1 min and 1 min/kb at 68°C. Template DNA was degraded using 10 units of *DpnI* at 37°C for 2 h. Samples were purified using a clean and concentrate kit (Zymo) and 2 µL were used to transform Subcloning DH5α competent cells. Recombinant plasmids were isolated and sequence-verified. Primers used for SDM experiments are listed in Appendices 2 and 3.

2.16 Construction of plasmids

For testing promoter activity, the upstream regions of the predicted translational start site were PCR-amplified using the primers listed in Appendix 2. *ScaI* restriction sites were incorporated into the primers where required and the amplified regions were cloned into the PCR cloning vector pSC-A, from which they were excised via *ScaI* digestion. Inserts were cloned into the *ScaI* site of

the mycobacteriophage L5-derived promoter probe vector pSM128 (Dussurget *et al.*, 1999). Successful cloning was confirmed by restriction enzyme digest and sequence verification. Lists of plasmids used in this study are provided in Appendix 4.

Protein tags AADENYA-ASV and AADENYA-LAA, added to the C-terminal end of *lacZ* in pSM128, were provided by Dr D. Schuessler (NIMR). AADENYA-ASV was mutated to AADENYA-GGG by site directed mutagenesis using primers listed in Appendix 13. To construct AADSHQRDYA-LAA protein tag, LacZ was excised with *HindIII*, PCR amplified with a primer harbouring the tag and cloned into pSC-A. The tagged LacZ was cut out of pSC-A and cloned into pSM128. The acetamidase promoter (P_{ami}) (Parish *et al.*, 1997) was excised from pFLAME-3-ace (Blokpoel *et al.*, 2003) with *BamHI* and *EcoRI*, blunted and cloned into the *ScaI* sites of the plasmids containing the lacZ constructs. *M. smegmatis rpsA* promoter with tetO operator (P_{smyc}) was cut out of pCon3 (Dr Paul Carroll, QMUL) with *XbaI*/*BamHI*, blunted and cloned into the *ScaI* sites of the plasmids containing the lacZ constructs.

The tetracycline repressors were extracted from pMC1m and pMC1s (Guo *et al.*, 2007) with *Not*I, blunted and cloned into pSC-B. The constructs were excised with *EcoR*I and cloned into the *EcoR*I sites of pTT1A (Pham *et al.*, 2007). Successful cloning was confirmed by restriction enzyme digest and sequence verification. Primers and plasmids used in this study are provided in Appendix 13.

2.17 Electroporation of mycobacteria

2.17.1 Electroporation of *M. smegmatis* and *M. marinum*

A 5 mL pre-culture was grown overnight and 1 mL was used to inoculate a 100 mL culture grown until an OD_{580} of 0.8-1.0 was reached. The culture was chilled on ice for 90 min before pelleting the cells by centrifugation at 2700 x *g* for 10 min at 4°C. The pellet was resuspended in 20 mL 10% w/v glycerol and washed in 5 mL 10% glycerol. The final pellet was redissolved in 1 mL of 10% glycerol, dispensed in 200 µL aliquots and stored at -80°C. Electrocompetent cells were thawed on ice. To each vial of cells, 1 µg of salt-free DNA (in a volume of ≤5 µL) was added, mixed gently and the sample incubated on ice for 10 min. The cells were transferred into a pre-chilled 2mm electroporation

cuvette (Flowgen Biosciences) and pulsed in an electroporator (Bio-Rad Laboratories Ltd.) set at 2.5 kV, 25 µF and 1000 Ω. Cells were incubated on ice for 10 min, recovered in 5 mL Lemco or 7H9 broth for *M. smegmatis* or *M. marinum* respectively and grown at 37°C with shaking at 100 rpm for 2 h. Serial dilutions were plated onto Lemco or 7H9 agar plates containing the appropriate selection antibiotic. Transformants were picked after 3 to 5 d of growth for *M. smegmatis* and after 7 d of growth for *M. marinum*.

2.17.2 Electroporation of *M. tuberculosis*

An inoculum of 3 mL was used to seed a 100 mL roller bottle culture which was grown for 7 d rolling at 37°C. The culture was treated with 10 mL of 2 M glycine 24 h before harvesting. Half of the culture was pelleted by centrifugation at 2700 x *g* for 10 min. The cells were washed twice in 10% w/v glycerol using 10 and 5 mL respectively and resuspended in 1 mL of 10% w/v glycerol. 1 µg of plasmid DNA (in a volume of ≤ 5 µL) was mixed with 200 µL of cells. Cells were transferred into a 2 mm electroporation cuvette and pulsed in an electroporator set to 2.5 kV, 25 µF and 1000 Ω. Cells were recovered in 10 mL medium and incubated at 37°C for 24 h. Serial dilutions were plated onto 7H10 plates containing the appropriate selection antibiotic. Transformants were picked after 4 weeks of growth.

2.18 Reporter gene assays

2.18.1 Preparation of cell-free extracts

Mycobacterial cells were grown to an OD_{600} of 0.6-0.8, harvested by centrifugation (10 min at 2700 x *g*), washed in 5 mL of 10 mM Tris-Cl (pH 8) and resuspended in 1 mL of 10mM of Tris-Cl. The solution was transferred into a 2 mL lysing matrix B tube (MP Biomedicals). Cells were disrupted using a FastPrep FP120 set at speed 6.0 for 30 s. Samples were spun at 16 000 x *g* for 4 min and the supernatants were recovered. Cell-free extracts from *M. tuberculosis* cells were filter sterilised through a 0.2 µm filter unit.

2.18.2 Determination of protein concentration

Total protein concentration of the samples was determined using the Pierce BCA protein assay kit (Fisher). Standards of bovine serum albumin (BSA) were prepared using 0.9% w/v NaCl and used at 0, 25, 125, 250, 500, 750, 1000, 1500 and 2000 µg mL^{-1}. Reagents A and B were mixed at a ratio of 50:1 and 200 µL of this solution was added to 25 µL of standard or cell-free extract in a

96-well PVC microtitre plate. Samples were incubated at 37°C for 30 min and the OD_{562} was measured. A standard curve was plotted from which protein concentrations of the cell-free extracts were calculated.

2.18.3 ß-galactosidase activity assays

β-galactosidase assays were carried out following the method previously described by Miller (Miller, 1972). Briefly, 100 µL of cell-free extract was added to 900 µL of Z Buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7) and incubated at 37°C for 5 min. 200 µL of 4 mg mL^{-1} ONPG was added and reaction mixtures were incubated at 37°C for a minimum of 60 min. Reactions were stopped with 500 µL of 1 M $NaHCO_3$. The OD_{420} was measured and β-galactosidase activity, expressed in Miller Units (MU), was calculated using the following formula:

$$\text{Units} = \frac{OD_{420} \times 1.7}{\text{time (min)} \times \text{volume of cell-free extract (mL)} \times \text{total protein concentration (mg mL}^{-1}\text{)} \times 0.0045}$$

2.19 Protein-protein interactions assays

2.19.1 Bacterial two hybrid system

Genes encoding the proteins of interest were amplified by PCR using the primer pairs listed in Appendix 8. The PCR products were cloned into pSC-A vector and sequence verified before being subcloned into pKT25 or pUT18 vectors (Karimova *et al.*, 1998). Plasmids are listed in Appendix 10. Different combinations of the resultant constructs were co-transformed into *E. coli* BTH101 and plated onto LB agar plates at 30°C containing X-Gal and appropriate selection antibiotics.

To measure interaction between pairs of hybrid proteins, colonies were inoculated into 4 mL of LB broth with appropriate antibiotics and IPTG. Cultures were diluted to an O.D of 0.5 and 2 mL of M63 medium was added to 0.5 mL of bacterial cultures. Cells were permeabilised with 30 µL of toluene and 32 µL of 0.1% w/v SDS. Tubes were vortexed for 10 s, lightly plugged with cotton and incubated at 37°C with shaking at 225 rpm for 35 min. For the enzymatic reaction, aliquots of 0.1 to 0.5 ml of the permeabilised cells were added to PM2 buffer (70 mM $Na_2HPO_4.12H_20$, 30 mM NaH_2PO_4 H_20, 1 mM $MgSO_4$, 0.2 mM $MnSO_4$, 100 mM β-mercaptoethanol, pH 7.0) to a final volume of 1 mL. Tubes were incubated at 28°C for 5 min before addition of 250 µL of 4 mg mL^{-1} ONPG

(pre-equilibrated to 30°C). Tubes were incubated at 37°C for 30 min. The reaction was stopped by addition of 0.5 ml of 1 M Na_2CO_3. The OD_{420} was measured and β-galactosidase activity, expressed in Miller Units, was calculated using the formula:

$$\text{Units} = \frac{OD_{420}}{\text{time (min)}} \text{ x dilution factor}$$

2.19.2 Split-Trp

Genes encoding the proteins of interest were amplified by PCR using the primer pairs listed in Appendix 9. The PCR products were cloned into pSC-A vector and sequence verified before being subcloned into PL240 and PL242 vectors (O'Hare *et al.*, 2008) between *SpeI* and *HpaI* restriction sites (plasmids are listed in Appendix 11). The resultant constructs were introduced into *M. smegmatis* Δ*hisA* and plated on Lemco agar. Three separate co-transformants were inoculated into 2 mL LB plus 0.05% Tween 80 and incubated at 37°C with shaking at 160 rpm until OD_{600} reached 1.0. 1 mL of culture was harvested by centrifugation at 2700 x *g* for 1 min. The bacteria were washed in 1 mL SDW plus 0.05% Tween 80 and resupended in 1 mL SDW. Drops of 5 µL suspension of undiluted, 1/10 and 1/100 dilutions were spotted onto 7H9 agar plates lacking tryptophan. The same dilutions were spotted onto plates containing tryptophan to check cell numbers. The plates were incubated for 3 weeks at 25°C.

2.20 Construction of deletion mutants of *M. tuberculosis*

Mutant strains were created using recombineering (van Kessel and Hatfull, 2007). The flanking regions and the gene of interests were amplified and cloned into pGEM vector; the gene was removed by inverse PCR with primers carrying *BglII* restriction sites. The hygromycin cassette with flanking *dif* sites was excised from pAL58 (Cascioferro *et al.*, 2010) using *BglII* and cloned in between the two flanking regions. A linear allelic exchange substrate was prepared by digestion with *XbaI* and *HindIII*. pJV53 plasmid DNA (van Kessel and Hatfull, 2007) was electroporated into *M. tuberculosis* cells and cells carrying pJV53 were inoculated into 100 mL roller bottle of 7H9 induction medium (7H9, 10% w/v BSA, Tween 80, 0.2% w/v succinate, kanamycin) and incubated at 37°C until an OD_{580} of 0.4 was reached. 10 mL of 2M glycine and 0.2% w/v of acetamide were added and the culture was incubated rolling for 24

h at 37°C. 100 ng of linearised DNA were elecroporated and cells were recovered in 10 mL medium incubated at 37°C for 2 d before plating. Colonies resistant to kanamycin and hygromycin were screened and validated by Southern blot analysis. Primers and plasmids used in this study are listed in Appendices 14 and 15.

Complementing vectors, carrying *smpB* or *ssrA* with their own promoters, were constructed by amplifying the corresponding gene sequences and 200 bp of upstream sequence from genomic DNA with primers listed in Appendix 14. The products were cloned into pGEM and the Gm-int cassette from pUC-Gm-Int (Mahenthiralingam *et al.*, 1998) was introduced as a *Hind*III fragment. The resulting vector was introduced in *M. tuberculosis smpB*Δ (for SmpB complementing vector) or cells carrying pJV53 (for *SsrA* complementing vector) and transformants were selected on 7H10 plates containing gentamycin.

2.21 Southern blotting

1.5 µg *M. tuberculosis* genomic DNA was digested overnight with a selected restriction enzyme and separated by gel electrophoresis. The DNA gel was incubated for 30 min at RT with slow shaking in denaturing buffer (3M NaCl, 1M NaOH) followed by 30 min incubation with neutralizing buffer (3M NaCl, 1M Tris). DNA was transferred to a Nytran SuPerCharge nylon membrane using a turboblotter (Whatman) following manufacturer's instructions as follows. 20 sheets of dry GB004 blotting paper (thick) were placed in the stack tray and four sheets of dry 3MM Chr blotting paper (thin) were placed on top of stack. One sheet of prewet 3MM Chr blotting paper in transfer buffer (3M NaCl, 8 mM NaOH) and the transfer membrane was then added. The membrane was covered with the agarose gel, making sure there were no air bubbles in between. The top surface of the gel was weted with transfer buffer, and three sheets of 3MM Chr blotting paper, presoaked in transfer buffer, placed on top of the gel. The "buffer tray" of the transfer device was added to the bottom tray and filled with transfer buffer. Transfer was carried out overnight.

The membrane was placed in a glass hybridization tube containing 15 mL of AlkPhos Direct hybridisation buffer (GE Healthcare) and incubated rolling at 60°C for 5-6 h. To generate a probe, a DNA fragment was PCR amplified and labeled with AlkPhos Direct system (GE Healthcare) according to the manufacturer's instructions as follows. DNA was diluted to a concentration of 10

ng/μL, heated for 5 min in a boiling water bath and immediately cooled on ice for 5 min. 10 μL of reaction buffer, 2 μL of labelling reagent, 10 μL of cross linker solution were added and the mixture was incubated at 37°C for 30 min. The probe was added to the buffer and hybridised overnight at 60°C. Primary and secondary post hybridization washes were carried out. 50 mL of pre-heated primary wash buffer (2M urea, 0.1% SDS, 50 mM NaOH, 150 mM NaCl, 1mM MgCl2 and 0.2% blocking reagent) was incubated for 30 min with gentle agitation at 60°C. The wash step was repeated before placing the blot in a clean container containing 100 mL of secondary wash buffer (1M Tris, 2 M NaCl, 1M MgCl2) for 5 min at RT with gentle agitation twice. The probe was detected by CDP-Star (GE Healthcare) using 1 mL of detection reagent for 5 min at RT. The blot was wrapped in a detection bag and placed in a film cassette. HyperfilmECL autoradiography film (GE Healthcare) was placed on top of the blot in a dark room, exposed for minimum 1 h and developed.

2.22 Macrophage infections

J774A murine macrophages (derived monocytes from Balb/C) (Rastogi *et al.*, 1989) were grown in complete medium (Dulbecco's Modified Eagle Medium, DMEM) (Invitrogen), supplemented with 10% fetal calf serum (FCS) (Invitrogen) and 2 mM L-Glutamine, at 37°C with 5% CO_2. Activated macrophages were prepared by adding 100 units mL^{-1} interferon gamma (Invitrogen) for 24 h and 100 ng ml^{-1} LPS for 4 h at 37°C. To get a multiplicity of infection of 10, 1 mL of an *M. tuberculosis* culture at O.D=0.05 was spun and resuspended in 10 mL complete medium. 100 μL of the suspension was added to $5x10^5$ monocytes per well of a 24 well plate and incubated at 37°C with 5% CO_2 for 24 h. The extracellular bacteria were removed by washing the macrophages three times with 1 mL PBS. To determine the number of intracellular bacteria at each time point (1, 3 and 7 d), bacteria were harvested using 4 mL of 0.25% SDS and serial dilutions of the lysates were plated on plates and incubated at 37°C for 4 weeks before the colonies were counted.

2.23 Electrophoretic mobility shift assay

The Rv2745c coding sequence was PCR amplified and cloned into pET-28a (Novagen) using primers to encode mutations V111D and A112D at the C-terminus listed in Appendix 5. Protein expression and purification were carried out by protein production facility at Queen Mary University (see Appendix 7 for detailed method).

PCR products were amplified using primers listed in Appendix 6 with biotin modification on one primer. Electrophoretic mobility shift assays (EMSA) were carried out using the LightShift Chemiluminescent EMSA Kit (Pierce) according to the manufacturer's instructions as follows. Binding reactions were performed with 10X binding buffer, 1 µg poly dI:dC, 15 fmol target DNA, protein extract and SDW for 20 min at RT. Samples were electrophoresed in 6% native polyacrylamide gel for 2 h in 0.5X TBE buffer. Reactions were transferred to a nylon membrane (Thermo Scientific) in an electrophoretic transfer unit with 0.5X TBE buffer at 100V for 1 h. The membrane was crosslinked in a UV-light crosslinking instrument using the auto crosslink function. To block the membrane 20 mL of blocking buffer was added and incubated for 15 min with gentle shaking. The buffer was replaced by 20 mL blocking buffer containing 66.7 µL streptavidin-horseradish peroxidase conjugate and incubated for 15 min at RT. The membrane was washed 4 times for 5 min each in wash solution followed by 5 min incubation in 30 mL of substrate equilibration buffer. Excess buffer was removed and the membrane was incubated for 5 min with substrate solution composed of luminol-enhancer solution mixed with stable peroxide solution. The moist membrane was wrapped in plastic wrap and exposed to X-ray film.

2.24 Microarray

RNA was extracted and cDNA was synthesised and labelled using SuperScript Indirect cDNA Labeling System (Invitrogen) following manufacturer's instructions as follows. Two µg total RNA was mixed with 5 µg anchored oligo$(dT)_{20}$ primer, 1 µL of random hexamers and SDW in a total volume of 18 µL. The mixture was heated at 70°C for 5 min and quickly placed on ice; 6 µL of 5X First Strand Buffer, 1.5 µL 0.1 M dithiothreitol (DTT), 1 µL dNTP mix (including amino-modified nucleotides), 1 µL RNaseOUT and 1 µL Superscript III Reverse transcriptase were added and incubated at 46°C overnight. The synthesis reaction was stopped by adding 10 µL 0.5M EDTA and 10µL 1M NaOH and incubated at 65°C for 15 min. 25 µL 1M Tris was added to neutralise the pH. In parallel, 4 µg gDNA were digested overnight with *EcoRI* and cleaned DNA was mixed with 10 µg of random hexamers and distilled water in a total volume of 39 µL. The mixture was mixed, incubated at 100°C for 10 min and chilled on ice for 30 s. Five µL of 10X EcoPol (Klenow) buffer, 2 µL of 5 mM dNTP/aa-UTP labelling mix and 4 µL of DNA Polyl Klenow fragment were

added to the mixture and incubated at 37°C overnight. DNA transcription was stopped by adding 5 µL of 0.5M EDTA.

Unincorporated aa-dUTP and free amines were removed from gDNA and cDNA samples following a modified method from the MinElute PCR purification kit (Qiagen). cDNA was vigorously mixed with 400 µL PB buffer before transferring to MinElute column and spun for 1 min; 750 µL phosphate wash buffer was added to the column and spun for 1 min. The collection tube was emptied and the column centrifuged for an additional 1 min. The column was transferred to a new 1.5 microfuge tube and 30 µL phosphate elution buffer was added. The buffer was eluted by centrifugation and a second elution was repeated with another 30 µL of phosphate elution buffer.

The gDNA and cDNA solutions were dried in a vacuum concentrator until the volume was reduced to 3 µL and 5 µL of 2X coupling buffer was added. 2 µL of DMSO was added to vials of Cy3 or Cy5 Alexa Fluor Reactive dyes (Invitrogen) to resuspend the dyes, vortexed thoroughly and added to the relevant gDNA or cDNA tubes. The mixes were incubated at RT in the dark for 1-2 h. Free dye was removed by purification following a modified method from the MinElute PCR purification kit (Qiagen) as previously. cDNA was eluted by adding 30 µL EB buffer twice. cDNA concentrations as well as dye incorporations were measured on a ND-1000 spectrophometer.

Cy3 and Cy5 probes were mixed together and dry in a speed vac concentrator. Microarrays were provided by the Bacterial Microarray Group at St Georges, University of London (BµG@S). Printed slides were incubated at 42°C with prehybridisation buffer (5X SSC, 0.1% SDS and 1% BSA) for at least 1 h. The buffer was removed and SDW was added and removed until suds could no longer be seen. The slides were placed in a slide holder inside a staining dish containing water and placed on a rotor shaker. The water was changed every 2 min until 2 L total of wash water was used. The staining dish was filled with isopropyl alcohol for 2 min and the slides were centrifuged at 600 x *g* for 10 min.

65 µL of hybridisation buffer (40% formamide, 5X SSC, 0.1% SDS, 0.6 $\mu g.\mu L^{-1}$ salmon sperm DNA) was added to the Cy3/Cy5 probe mixture and vortexed. The mixture was heated at 95°C for 5 min, vortexed and heated again at 95°C

for 5 min. A clean lifterslip was placed over the prehybridised microarray slide and the labelled probe mixture was pipetted onto the lower edge of the lifterslip. Slides were placed in a hybridisation chamber and incubated in a 42°C waterbath overnight. Slides were removed from the chamber and placed in low stringency wash buffer (2X SSC, 0.1% SDS, 0.1 mM DTT) pre-warmed to 55°C. As the cover slips were removed, the slides were placed in a glass slide holder placed into a staining dish containing low stringency wash buffer pre-heated to 55°C and agitated for 5 min. Slides were transferred to a dish with new low stringency buffer and agitated for 5 min; washed twice in medium stringency medium (0.1X SSC, 0.1% SDS, 0.1 mM DTT) for 5 min followed by 2 washes for 5 min in high stringency medium (0.1X SSC, 0.1 mM DTT).

Slides were read with a Genepix scanner and microarray analysis was performed by the Bacterial Microarray Group at St. George's, University of London (BμG@S). Each experiment was analysed separately and also combined as one. GeneSpring v7.3 analysis software was used for statistical analysis to identify any significant differences between the expression levels of individual genes. Analysis was restricted to only genes with 2-fold expression differences. One-way ANOVA was used to select differentially expressed genes with a *p*-value of less than 0.05 using Benjamini and Hochberg false discovery rate correction.

2.25 qRT-PCR

5 μL of DNA or cDNA was added to a 2X mix composed of 25 μL Taqman mastermix (Roche), 9 μL of 5 μM primer mix, 1 μL of 25 μM probe and 10 μL SDW. 25 μL of the reaction mix were aliquoted into a 96 well qRT-PCR plate and centrifuged briefly. Amplification reactions were monitored with the 7500 Real Time PCR System (Applied Biosystems). Three biological replicates were assayed in duplicate. In order to measure relative gene expression levels, standard curves for each primer-probe set were generated using genomic DNA and used to calculate copy number for cDNA reactions. Copy number was corrected by substractiong background from genomic DNA in the samples (no reverse transcription reaction). In order to standardise the samples to ensure that equal amounts of cDNA were used each value was standardized to *sigA* to generate unit-less values. Primer-probes sequences are listed in Appendix 12.

3 Regulation of ClpP1 and ClpP2 expression and activity

3.1 Introduction

Clp proteases have proteolytic cores capable of cleaving a broad range of peptide bonds, and yet their activity is directed against only misfolded or aggregated proteins and specific substrate proteins (Thompson *et al.*, 1994). The active Clp protease is composed of a proteolytic and a regulatory ATPase subunit as the proteolytic subunit is unable to bind and degrade protein without an ATPase control partner. Degradation by the Clp complex is a highly regulated process and this control is critical for the cell viability. For example activation of ClpP activity by ADEPs is lethal in Gram positive bacteria including *M. tuberculosis* (Brotz-Oesterhelt *et al.*, 2005; Ollinger *et al*., 2011).

Prokaryotic gene expression is controlled mainly at the level of initiation of gene transcription (Alberts *et al.*, 2002b). Clp proteases are involved in protein quality control by degrading misfolded or aggregated proteins, thus it is not surprising that in most bacteria expression of ClpP is induced by conditions which lead to protein misfolding, such as heat shock or oxidative stress (Engels *et al.*, 2004; Msadek *et al.*, 1998). In *M. tuberculosis,* the *clpP* genes are highly expressed during aerobic and hypoxic growth and are further up-regulated during reaeration following exposure to anaerobic conditions (Muttucumaru *et al.*, 2004; Sherrid *et al.*, 2010). Additionally, an increase in *clpP1* and *clpP2* expression was observed after different stresses such as oxidative shock in *M. tuberculosis* and after vancomycin treatment in *M. smegmatis* (Barik *et al.*, 2009; Mehra and Kaushal, 2009). The transcriptional regulator ClgR activates ClpP1 and ClpP2 expression in actinobacteria including *M. tuberculosis* (Bellier and Mazodier, 2004; Engels *et al.*, 2004; Sherrid *et al*., 2010). ClgR-mediated up-regulation of the *clp* genes is required for replication in a macrophage model of infection (Estorninho *et al.*, 2010) supporting an important role for the Clp proteases during adverse conditions.

Regulation of proteolysis may also occur at the level of substrate selection. The Clp ATPase subunits recognise and control substrate entry into the proteolytic core for degradation and therefore ensure proteolytic specificity (Hoskins *et al.*, 1998). Different ATPases generally have distinct substrate preferences; for

example, *E. coli* ClpXP complex degrades the stationary-phase sigma factor σ^S and Mu transposase, which are not substrates for ClpAP (Levchenko *et al.*, 1995; Schweder *et al.*, 1996). Conversely, ClpAP, but not ClpXP, degrades MazE and HemA (Engelberg-Kulka and Glaser, 1999; Wang *et al.*, 1999).

M. tuberculosis contains three potential ATPase subunits (ClpC1, ClpC2 and ClpX) but it is not known how many of them associate with ClpP1 or ClpP2 (Fig 3.1 A). ClpX and ClpC1 possess 'LGF' loops (Fig 3.1 B), a determinant for interaction with ClpP (Kim *et al.*, 2001), suggesting there are the only two ATPases interacting with ClpP1 and ClpP2. ClpC2 may not be functional since it does not have ATPase activity (Benaroudj *et al*., 2011). Interaction between ClpP2 and ClpC1 has been demonstrated (Singh *et al.*, 2006) but this does not exclude interaction between ClpP1 and ClpC1; RseA was shown to be degraded by the ClpP2-ClpC1 complex (Barik *et al.*, 2009) further demonstrating an interaction between ClpP2 and ClpC1.

The presence of two ClpP proteases in *M. tuberculosis* as well as their essentiality is intriguing. One could hypothesise that ClpP1 and ClpP2 are present or induced in the cell under different conditions. For example, one ClpP subunit could be constitutively expressed while the other one may be induced in the presence of stress conditions or alternatively, the two ClpP proteins may be induced in response to different environmental conditions. The presence of two ClpP proteins suggests that they may target different substrates or be involved in different cellular pathways. In order to determine the specific role of the two ClpP proteases in *M. tuberculosis*, the first objective of the project was to determine if there was a difference in *clpP1* and *clpP2* gene expression and regulation. ClpP1 and ClpP2 could target different substrates by interacting with different ATPase subunits or accessory subunits; thus the next objective was to determine protein-protein interaction between the two ClpP subunits, the ATPases and the accessory components.

3.2 Results

3.2.1 Co-transcription of *clpP1* and *clpP2*

It seemed likely that *clpP1* and *clpP2* were co-expressed since they are organised in an apparent bicistronic operon (Fig 3.1 A) and the start

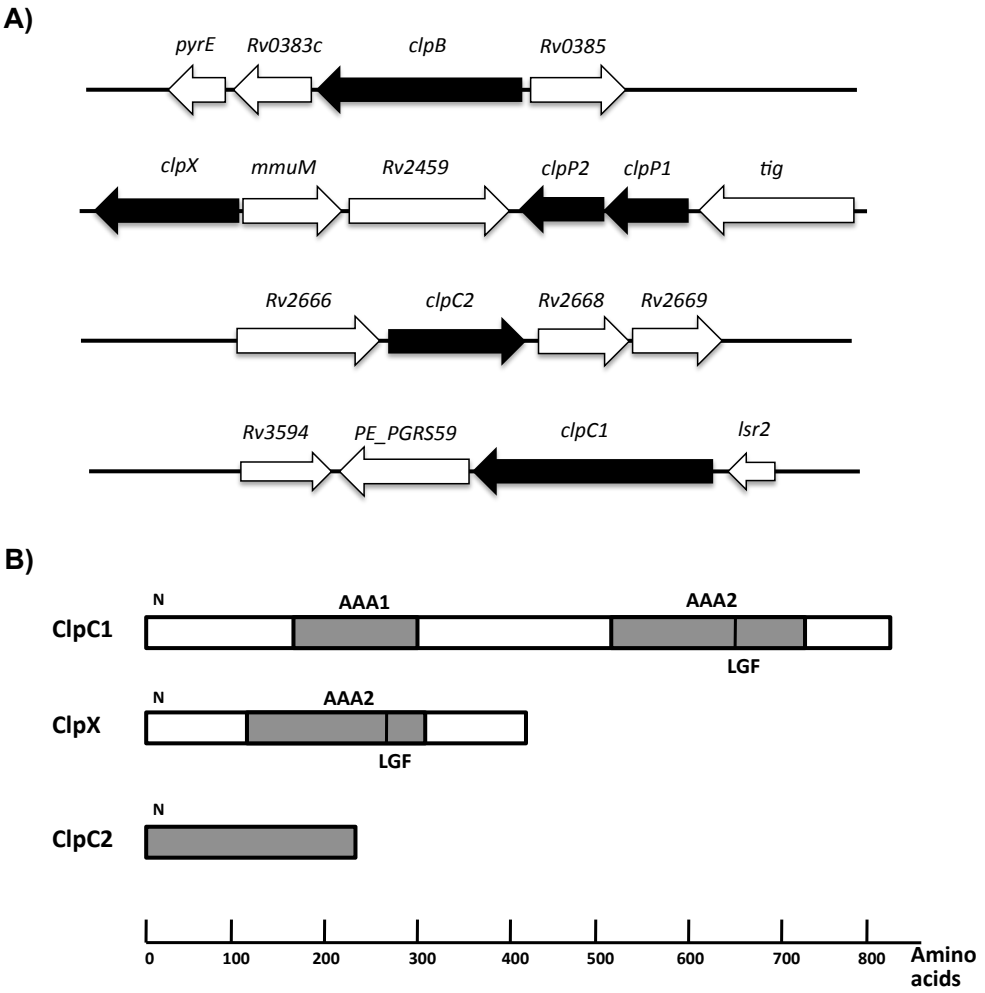

Figure 3.1 ClpP and Clp ATPases present in *M. tuberculosis*

A) Genomic organisation of the ClpP and Clp ATPase subunits in *M. tuberculosis.* The genetic coordinates for the H37Rv strain of *M. tuberculosis* are given.

B) Overall architecture of the Clp ATPases present in *M. tuberculosis*. Class I ATPase proteins (ClpC1) contain two ATPase (AAA) modules while class II (ClpX) contain only one AAA module. The 'LGF' loops sequences, predicted to be necessary for interaction between the Clp ATPases with ClpP, are present in ClpC1 and ClpX.

and stop codons overlap. In order to determine if *clpP1* and *clpP2* are co-transcribed as an operon, semi-quantitative RT-PCR was conducted on RNA extracted from *M. tuberculosis* during late exponentional phase. cDNA was synthesised and primers were designed to amplify regions specific to *clpP1* (258 bp) and *clpP2* (264 bp), or that spanned *clpP1-clpP2* junction (257 bp). To determine the relative amounts of each mRNA species, limiting dilution RT-PCR was used as a semi-quantitative method; *sigA,* whose expression is considered to be constant was used as a control (Manganelli *et al.*, 1999) (Fig 3.2). Products were identified for all three mRNA species, indicating that the two genes are co-transcribed. The relative levels of amplification of *clpP2* were lower compared to levels of *clpP1* mRNA as expected if the gene was at the 3' end of an operon. These data demonstrate that *clpP1* and *clpP2* are transcribed in late exponentional phase and the two *clpP* genes are co-expressed under the control of a single promoter.

3.2.2 Identification of the promoter of the *clpP1P2* operon

In order to identify the promoter of the operon, a 125 bp region (P_{125}) encompassing the first two codons of *M. tuberculosis clpP1*, the intergenic region between *clpP1* and the upstream gene *tig,* and the *tig* stop codon, was cloned into pSM128, a plasmid which integrates in a single copy in the genome and contains a promoterless *lacZ* reporter gene (Dussurget *et al.*, 1999). The corresponding plasmid was electroporated into *M. smegmatis*, *M. marinum,* and *M. tuberculosis* and promoter activity was assayed in aerobic standing cultures. Promoter activity, measured by β-galactosidase activity, was detected in all three organisms confirming that a functional promoter was present in this region (Fig 3.3 B). However, the promoter was much more active in *M. tuberculosis* (106 MU) as compared to *M. marinum* (12 MU) or *M. smegmatis* (15 MU).

To exclude the possibility of another promoter sequence upstream of *clpP2,* the region encompassing the first two codons of *clpP2* and 280 bp upstream of *clpP2,* was cloned upstream of a promoterless *lacZ* reporter gene. Promoter activity was assayed in aerobic standing cultures in *M. tuberculosis*; no activity was detected in this region (Fig 3.3 C) confirming that, at least under aerobic culture conditions, *clpP2* is not independently expressed and *clpP1* and *clpP2* are co-expressed.

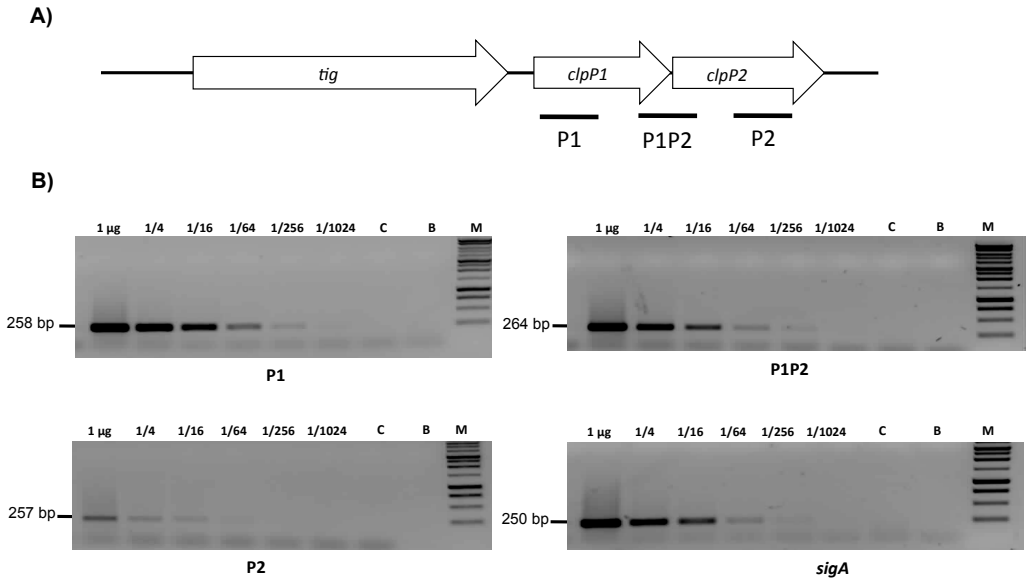

Figure 3.2 Co-transcription of *clpP1* and *clpP2* in *M. tuberculosis*

A) Chromosomal organisation of *clpP1* and *clpP2*. Regions amplified for RT-PCR are marked. B) Limiting dilution RT-PCR. RNA was extracted from *M. tuberculosis* grown to late exponential phase in liquid cultures and cDNA was synthesised form 1 µg of RNA. Serial-four fold dilutions of cDNA were used as a template for PCR using primers specific for *clpP1* (P1), *clpP2* (P2), the spanned *clpP1-clpP2* junction (P1P2) and *sigA*. C: no RT control; B: no template blank, M: markers.

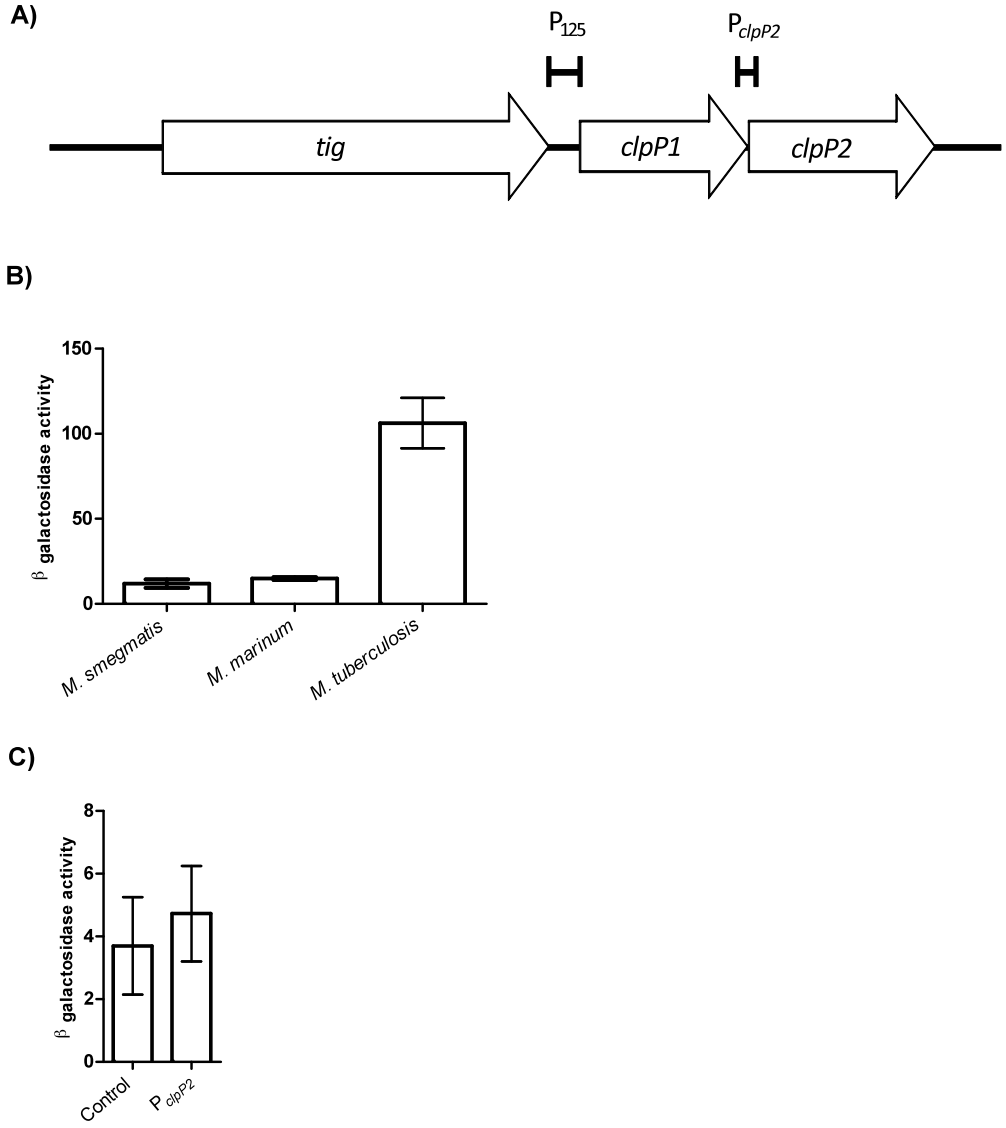

Figure 3.3 Identification of the promoter of the *clpP1P2* operon

A) Upstream regions of *clpP1* or *clpP2* tested for promoter activity. B) P_{125} activity in *M. smegmatis*, *M. marinum* and *M. tuberculosis*. C) P_{clpP2} activity in *M. tuberculosis*. Promoter activity was measured in transformants grown in late exponential phase in standing liquid cultures. Results are the average activity ± standard deviation of three independent transformants assayed in duplicate. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from pSM128 (control vector) was 4 ± 2 MU.

Mycobacterial promoters are very diverse although the -10 element is frequently similar to the *E. coli* consensus sequence (TATAAT) (Gomez and Smith, 2000). Looking at the sequence data it was possible to identify two putative -10 elements in the 125 bp promoter region (Fig 3.4 A). To determine if either of these are the genuine -10 element of the promoter, the TAGTGT hexamer (10A) was mutated to **C**AGTG**G** and the TAGAAG hexamer (10B) was mutated to **CG**GAAG and promoter activity, using *lacZ* reporter gene, was measured in *M. marinum* and *M. tuberculosis* (Fig 3.4 B and C). In *M. marinum* and *M. tuberculosis*, no significant difference in activity was observed when the TAGAAG hexamer was mutated to **CG**GAAG but when the TAGTGT was mutated to **C**AGTG**G** promoter activity was abolished, indicating that the TAGTGT hexamer is the -10 element of the ClpP1P2 operon promoter.

In mycobacteria the -35 region of a promoter is often hard to predict, and it is not uncommon for regulated promoters to completely lack a -35 element (Gomez and Smith, 2000). Two potential -35 elements were identified (35A and 35B in Fig 3.4 A) with similarity to the canonical -35 element of *E. coli* promoters (TTGACA), located 17 and 23 nucleotides upstream of the -10 element respectively. Mutation of each element was conducted and the effect on promoter activity measured in *M. tuberculosis* (Fig. 3.4 D); neither of the mutations reduced promoter activity suggesting that they are not functional -35 elements and that this promoter may lack this region. However, mutation of GTGACC (35A) to **CC**GACC significantly increased promoter activity; a possible explanation for this may be that this region may be a binding site for a regulator as discussed in a later section (3.2.9).

3.2.3 Promoter of the *clpP1P2* operon is weak

The frequency of transcription initiation is dependent on the strength of the promoter i.e. its affinity for the RNA polymerase, and the action of regulatory proteins. In order to determine the intrinsic strength of the ClpP1P2 operon promoter, a short upstream region, encompassing the two first ClpP1 codons and 86 bp upstream of *clpP1* (denoted P_{92}), was cloned upstream of a promoterless *lacZ* and activity was measured in *M. tuberculosis*. The activity was very low (12 MU) demonstrating that the promoter is weak (Fig 3.5). Expression of *clpP1P2* is induced after addition of vancomycin in *M. smegmatis* and after oxidative stress, generated by diamide, in *M. tuberculosis*

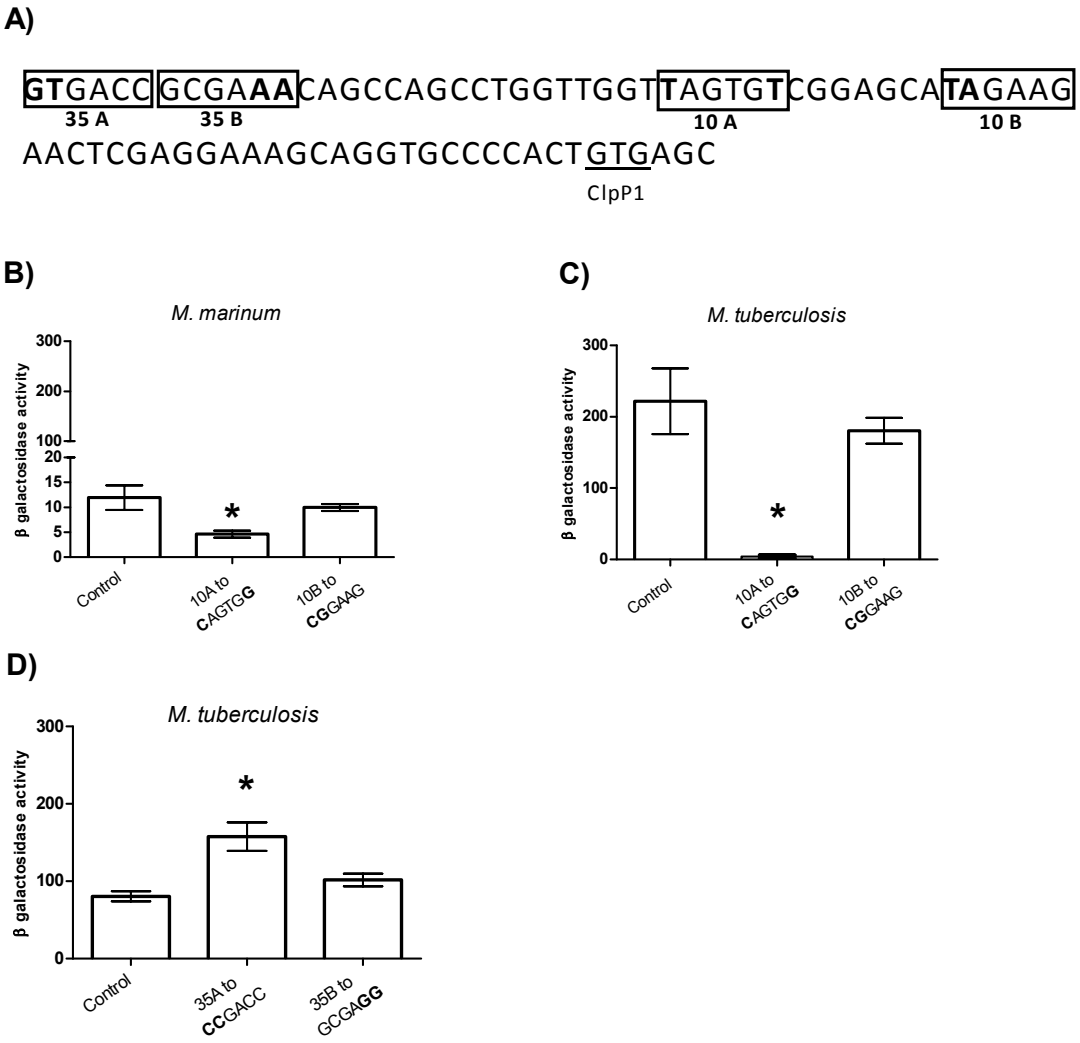

Figure 3.4 Identification of the promoter elements of the *clpP1P2* operon

A) Sequence of the upstream region of *clpP1P2*. Putative -10 elements (10A and 10B) and putative -35 elements (35A and 35B) are boxed. Mutated residues are in bold. ClpP1 start codon is indicated. B and C) Identification of the -10 element in *M. marinum* and in *M. tuberculosis* respectively. The following mutations were made- 10A: TAGTGT mutated to **C**AGTG**G**; 10B: TAGAAG mutated to **CG**GAAG. D) Absence of a -35 element. The following mutations were made- 35A: GTGACC to **CC**GACC; 35B: GCGAAA to GCGA**GG**.

Promoter activity was measured in transformants grown in late exponential phase in standing liquid cultures. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from pSM128 (control vector) was 4 ± 2 MU. A significant difference, measured by the student's t-test (unpaired, two sided), compared to the control vector (pSM128) is marked by an * ($p < 0.05$).

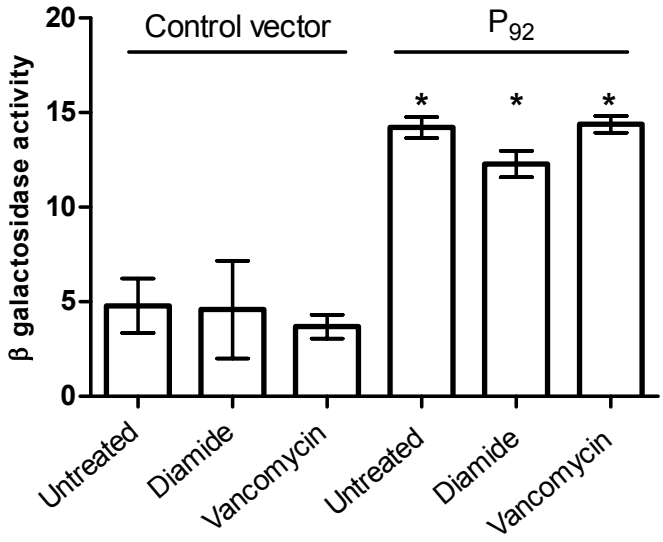

Figure 3.5 Promoter activity of P_{92} in response to diamide and vancomycin

M. tuberculosis transformants harbouring the P_{92} plasmid or the control vector pSM128 were grown to late exponential phase in standing liquid cultures. Stress treatments were 10 mM diamide for 1 h or 6 µg/mL of vancomycin for 90 min. Cell-free extracts were prepared after treatment and results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. A significant difference compared to the control vector pSM128 is marked by an * ($p < 0.05$) using the student's t-test (unpaired, two sided).

(Barik *et al.*, 2009; Mehra and Kaushal, 2009). Activity of P_{92} was tested after these treatments but was not induced (Fig 3.5).

3.2.4 Promoter activity in response to heat and oxidative shocks

Under stress conditions such as high temperature or oxidation, misfolded and aggregated proteins can accumulate and are potentially toxic for the cells. Clp proteases are involved in the degradation of misfolded or aggregated proteins (Kruger *et al.*, 2000; Thomsen *et al*., 2002) and are heat induced in several organisms including *B. subtilis* and *C. glutamicum* (Engels *et al*., 2004; Msadek *et al.*, 1998). Promoter activity of P_{125} in *M. smegmatis* cultures exposed to moderate (42°C) or even severe (50°C) heat shocks were not significantly different from the untreated control (Fig. 3.6 A). Moderate heat shock was tested in *M. tuberculosis,* as well as oxidative shock (diamide) to determine if the published observations regarding increased RNA expression after oxidative stress correlated with increased promoter activity. Promoter activity after heat and oxidative stress was not significantly different from the untreated control (Fig. 3.6 B) indicating that the promoter was not induced by these conditions.

3.2.5 Increased promoter activity of a longer fragment

Since the promoter activity was not increased after oxidative stress despite previous published reports of RNA induction, it was considered that the intergenic region being examined might not contain all the regulatory sites. A longer fragment upstream of *clpP1/clpP2* (278 bp) was cloned upstream of a promoterless *lacZ* reporter gene (denoted P_{278}) and promoter activity was measured in *M. tuberculosis*, *M. smegmatis* and *M. marinum* (Fig. 3.7). In all species the longer fragment had a higher promoter activity than the shorter region (Fig 3.3B and Fig 3.7), suggesting that binding sites for regulatory elements were missing in the shorter region. Again, promoter activity was much lower in *M. smegmatis* (95 MU) and *M. marinum* (207 MU), than in *M. tuberculosis* (865 MU). Plasmid instability was noted in *M. marinum,* since only one of the three transformants tested contained the insert.

The longer fragment (P_{278}) was used in all the future studies as the shorter fragment misses binding sites for regulatory elements. The effect of heat and oxidative stress treatments on promoter activity was first tested in *M. smegmatis* and *M. marinum*. However, once again promoter activity was not induced (Fig 3.7 A and B). As treatments with chlorpromazine, menadione and

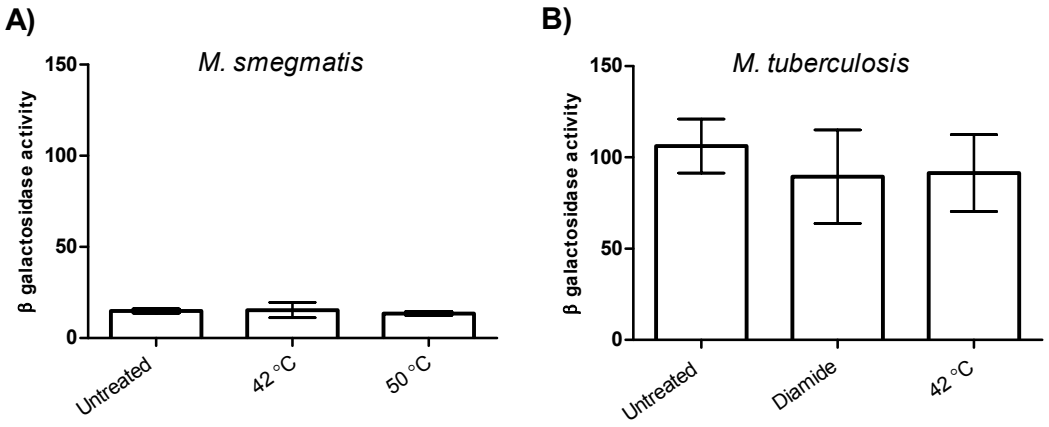

Figure 3.6 Promoter activity in response to heat and oxidative stresses in *M.smegmatis* and *M. tuberculosis*

Promoter activity of P_{125} was measured in transformants grown to late exponential phase in standing liquid cultures. Stress treatments were 10 mM diamide for 1 h, heat shock at 42°C for 1 h, or 50°C for 1 h. Cell-free extracts were prepared after treatment and results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from the control vector pSM128 was 5 ± 2 MU under the different conditions tested.

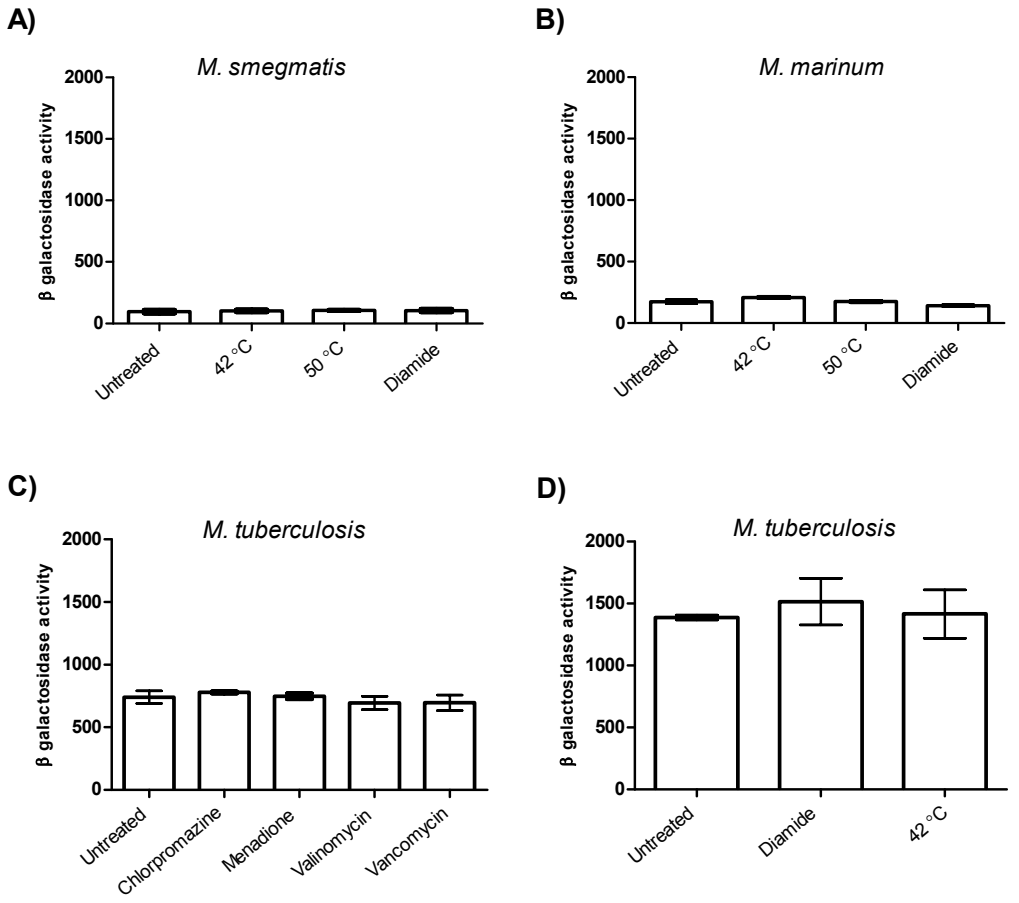

Figure 3.7 Promoter activity of P_{278} in response to stress treatments

Promoter activity of a 278 bp fragment upstream of *clpP1* was measured in A) *M. smegmatis*, B) *M. marinum,* C and D) in *M. tuberculosis* in standing or rolling cultures respectively.

Transformants harbouring P_{278} were grown to late exponential phase in standing liquid cultures unless otherwise stated. Treatments were: 42°C for 1 h, 50°C for 1 h, 10 mM diamide for 1 h, 50 µg/mL of chlorpromazine for 3 h, 10 µg/mL of menadione for 3 h, 10 µg/mL of valinomycin for 3h, 6 µg/mL of vancomycin for 90 min. Cell-free extracts were prepared after treatment and results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from pSM128 (control vector) was 6 ± 3 MU under the different conditions tested.

valinomycin in *M. tuberculosis* (Boshoff *et al.*, 2004) and vancomycin treatment in *M. smegmatis* (Barik *et al.*, 2009) have previously been shown to increase ClpP expression, these conditions were also examined. Promoter activity was tested in *M. tuberculosis* following these treatments but the promoter was not significantly induced or repressed in standing cultures (Fig 3.7 C). To determine if the lack of induction was specific to standing cultures, promoter activity was also measured in aerated rolling cultures before and after heat and oxidative stress treatments in *M. tuberculosis*. Promoter activity was different in rolling cultures (1,380 MU) compared to standing cultures (740 MU) but again was not induced after heat or oxidative stress (Fig 3.7 D).

The difference in promoter activity between rolling and standing cultures could indicate that the promoter activity is growth dependent. Promoter activity was measured from rolling aerated cultures over a time course from $O.D_{580}$ 0.15 to $O.D_{580}$ 1.7. The activity was constant from an $O.D_{580}$ of 0.15 to stationary phase at $O.D_{580}$ 1.5 (from 720 to 980 MU) (Fig 3.8). However in late stationary phase, at $O.D_{580}$ of 1.7 to 1.8, promoter activity increased significantly (to 1,050 MU) compared to the activity measured at $O.D_{580}$ 0.15 (Fig 3.8).

3.2.6 The lack of induction is not due to experimental factors

A series of experiments were designed to try to understand why the lack of ClpP induction in this study does not correlate with previous studies (Barik *et al.*, 2009; Boshoff *et al.*, 2004; Mehra and Kaushal, 2009).

Microarray data suggested that streptomycin induces *clpP1* and *clpP2* expression (Boshoff *et al.*, 2004). However, streptomycin was used throughout the culturing to maintain selection of the promoter constructs raising the possibility that the lack of induction observed in these studies may result from the fact that the promoter was already induced. To determine if this was the case, promoter activity was measured in cultures grown with or without antibiotic selection for seven days. There was no significant difference in promoter activity between cultures grown in presence or absence of antibiotic selection and no induction of activity was observed following diamide or vancomycin treatment in either condition (Fig 3.9 A). This shows that antibiotic selection with streptomycin was not inducing *clpP1P2* promoter activity.

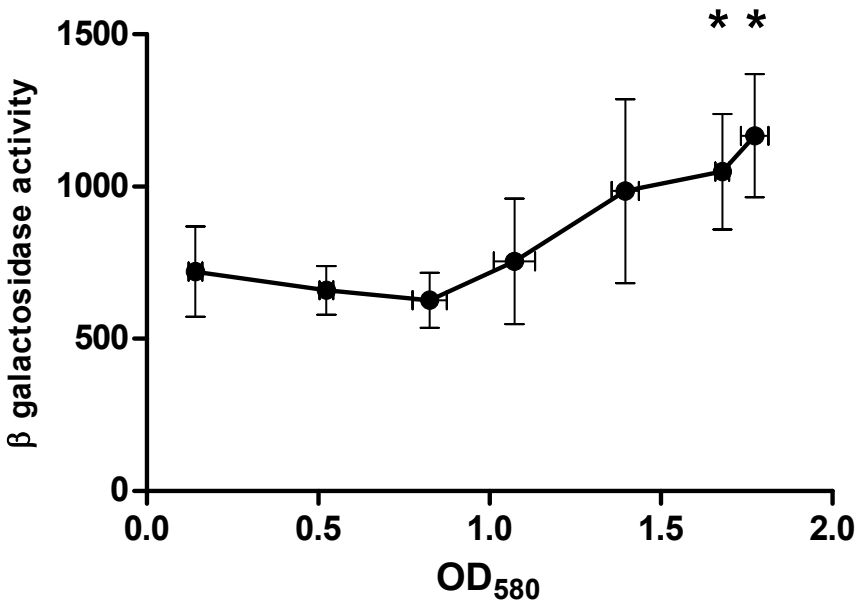

Figure 3.8 Promoter activity during aerobic growth in *M. tuberculosis*

Transformants harbouring P_{278} were grown in aerobic cultures. Cell-free extracts were prepared and ß-galactosidase activity assayed in duplicate. Results are the average activity of three transformants against average OD_{580}. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from pSM128 (control vector) was 8 ± 3 MU. A significant difference, measured by the student's t-test (unpaired, two sided), compared to promoter activity at $O.D_{580}$ 0.15 is marked by an * ($p < 0.05$).

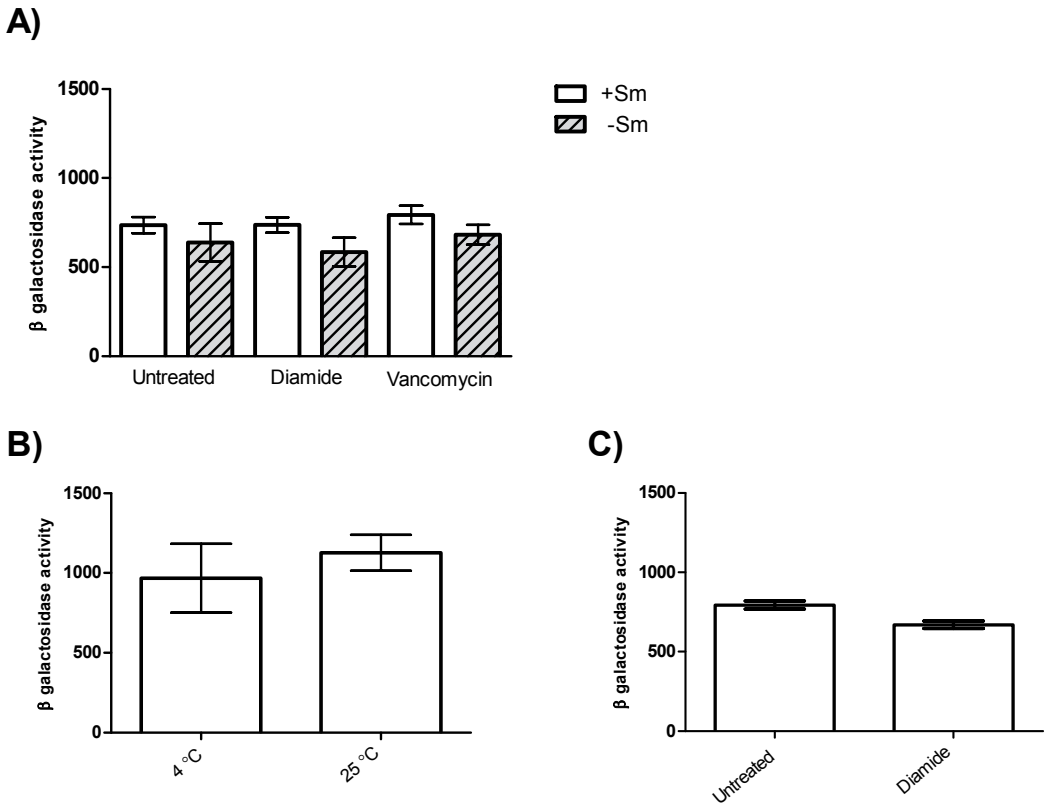

Figure 3.9 Promoter activity was not induced during experimental conditions

A) Promoter activity in response to diamide and vancomycin treatments in presence or absence of streptomycin selection. B) Promoter activity of cultures processed at 4°C or at 25°C. C) Promoter activity in response to diamide treatment in *M. tuberculosis* CDC1551.

M. tuberculosis transformants were grown to late exponential phase in standing liquid cultures. Stress treatments were 10 mM diamide for 1 h or 6 µg/mL of vancomycin for 90 min. Cell-free extracts were prepared after treatment and results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from pSM128 was 6 ± 2 MU under the different conditions tested.

+Sm: presence of streptomycin, -Sm: absence of streptomycin

ClpP1 can be induced by cold shock in cyanobacteria (Porankiewicz *et al.*, 1998). During the making of cell free extracts, liquid cultures were harvested at 4°C, raising the possibility that promoter activity may have been induced during processing. To address this question, promoter activity was measured in cultures harvested at 4°C or at room temperature (25°C) but no significant different in activity was noted excluding cold shock induction (Fig. 3.9 B).

Previously, ClpP induction in response to diamide treatment has been demonstrated in *M. tuberculosis* CDC1551 strain (Mehra and Kaushal, 2009). To determine whether *clpP1P2* induction might be strain-specific, promoter activity was compared between *M. tuberculosis* H37Rv and *M. tuberculosis* CDC1551. The activity of the CDC1551 strain (800 MU) was comparable to that observed in H37Rv and was not induced after diamide treatment (Fig 3.9 C) demonstrating that promoter induction after diamide treatment was not strain-specific.

3.2.7 Promoter activity during hypoxia and reaeration

M. tuberculosis has the ability to survive inside the host environment for decades in a latent state before reactivation. The Wayne model is frequently used to study hypoxia (Wayne and Hayes, 1996), one condition believed to be encountered by the bacteria during latency, and inoculating hypoxic cultures into aerated medium (reaeration) can be used to mimic reactivation of the disease (Sherrid *et al.*, 2010). Promoter activity was measured during adaptation to hypoxia and over 12 weeks of survival in hypoxia (Fig 3.10 A). Promoter expression was similar between aerated and hypoxic cultures (typically around 700-800 MU) for up to 56 days (eight weeks) of hypoxia. Promoter activity was reduced to 350 MU between weeks 8 and 12 (84 days) of exposure to hypoxia. Hypoxic cultures of 84 days were then inoculated into aerated medium, and promoter activities were measured once the cultures reached a minimum $O.D_{580}$ of 0.3. Promoter activity returned to its original activity quickly after reaeration (Fig. 3.10 B).

3.2.8 Mapping of a regulatory sequence

Higher promoter activity was observed for P_{278} compared to P_{125} suggesting that the longer fragment contained at least one regulatory element binding site that is missing in the shorter fragment. In all bacteria containing the

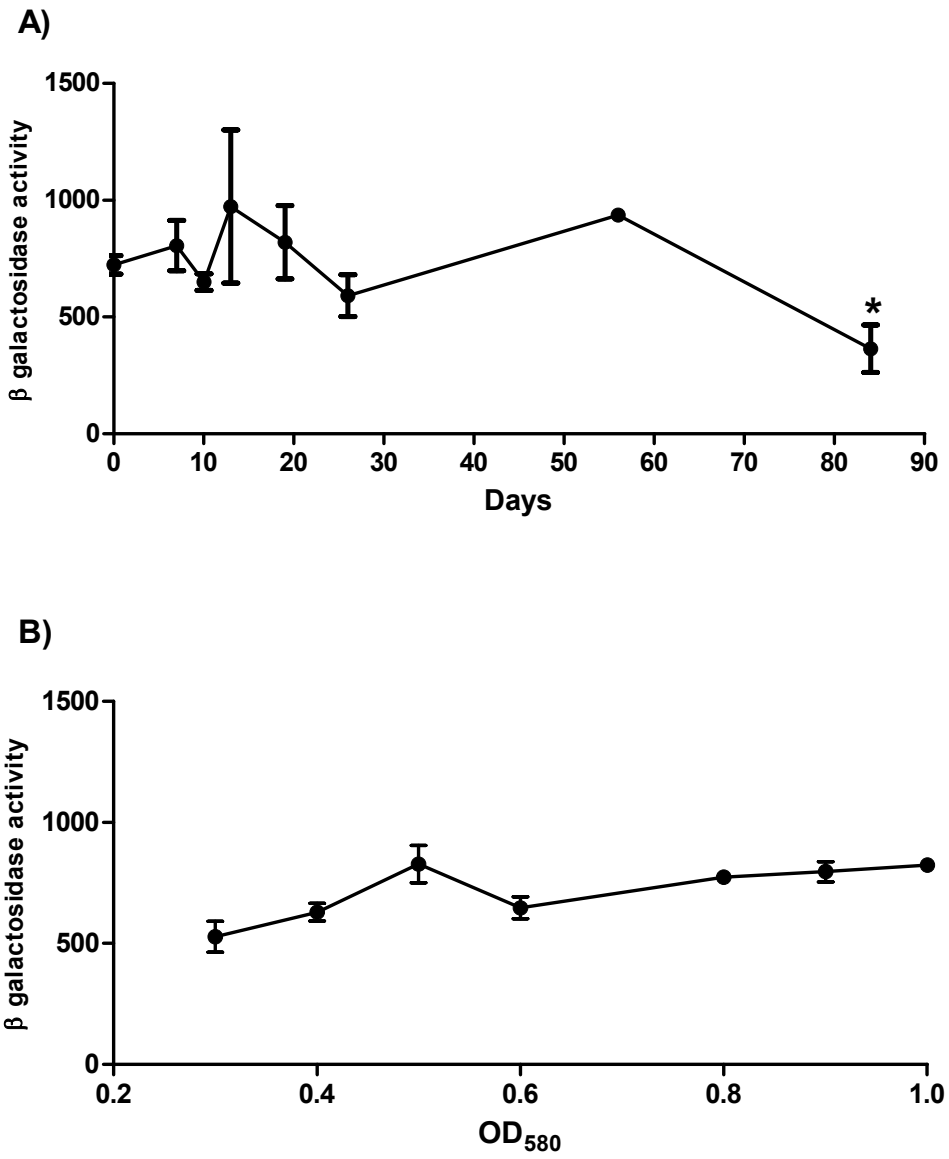

Figure 3.10 Promoter activity in hypoxic cultures and after reaeration

A) Promoter activity in the Wayne model of hypoxia. *M. tuberculosis* liquid cultures were inoculated to a theoretical starting OD_{580} of 0.004 in DTA medium. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. A significant difference compared to activity at day 0 is marked by an * ($p < 0.05$) using the student's t-test (unpaired, two sided).

B) Promoter activity after reaeration. 84 days hypoxic cultures were used to inoculate aerobic rolling cultures. Cell-free extracts were prepared once the cultures reached an $O.D_{580}$ of 0.3 and ß-galactosidase activity in the samples was measured in duplicate. Activity is given in Miller Units- measured as nmol of O-nitrophenol produced per min per mg of protein.

transcriptional regulator ClgR, the palindromic motif CGC-N5-GCG is found in the promoter region of *clpP* (Russo *et al.*, 2009). This motif is found in the region upstream of *clpP1P2* in *M. tuberculosis,* approximately 100 bp away of ClpP1 start codon (Fig 3.11 A). The first half of the motif (CGC) was mutated to 'AAA' and promoter activity was measured using *lacZ* as before. Promoter activity was significantly reduced after mutation in *M. smegmatis* (3.6-fold reduction), *M. marinum* (4-fold reduction) and *M. tuberculosis* (3.4-fold reduction) suggesting the possible binding of a regulator at this site, most likely to be ClgR (Fig 3.11 B).

To determine the nucleotides in this region that are responsible for the binding of a regulatory element, single nucleotide substitutions were created where native A/T were mutated to G bases and C/G were replaced by As and activity was tested (Figure 3.11 C). Twenty bases were mutated in total and numbered 1 to 20. Two bases (A_1G and $C_{20}A$) whose mutation had no effect on promoter activity were considered to be outside the sequence bound by a regulator. All mutations in between these two nucleotides had a significant effect on promoter activity and therefore constitute the binding site, which is thus 18 bp long. Two mutations induced a higher promoter activity (C_5A and $G_{19}A$) while all other mutations induced a significant reduction in promoter activity compared to the unmutated plasmid. This regulatory 18 bp sequence (TGACGCTGTAAGCGAACG) matches the consensus sequence of ClgR binding site in *C. glutamicum* (Engels *et al.*, 2005). As regulation of *clpP1P2* by ClgR was previously demonstrated in *M. tuberculosis* (Sherrid *et al*., 2010) these data strongly suggest that the sequence mapped is the ClgR binding sequence.

To determine if the differences in promoter activity following mutagenesis were due to alteration in binding of ClgR, a protein-DNA binding assay was used. The final two amino acids of ClgR (V111, A112) were mutated to aspartates to stabilise the protein as previously described (Sherrid *et al.*, 2010) and the protein was purified by affinity chromatography. Using this assay (EMSA), a DNA shift demonstrates binding of a selected protein to a DNA sequence. However, no DNA shift was observed when the 278 bp promoter region of *clpP1P2* and purified ClgR (200ng or 500 ng) were used; the addition of $MgCl_2$ or NP-40 were all found to show no difference, as again there was no observable DNA shift (Fig 3.12). Folding of the DNA sequence was considered

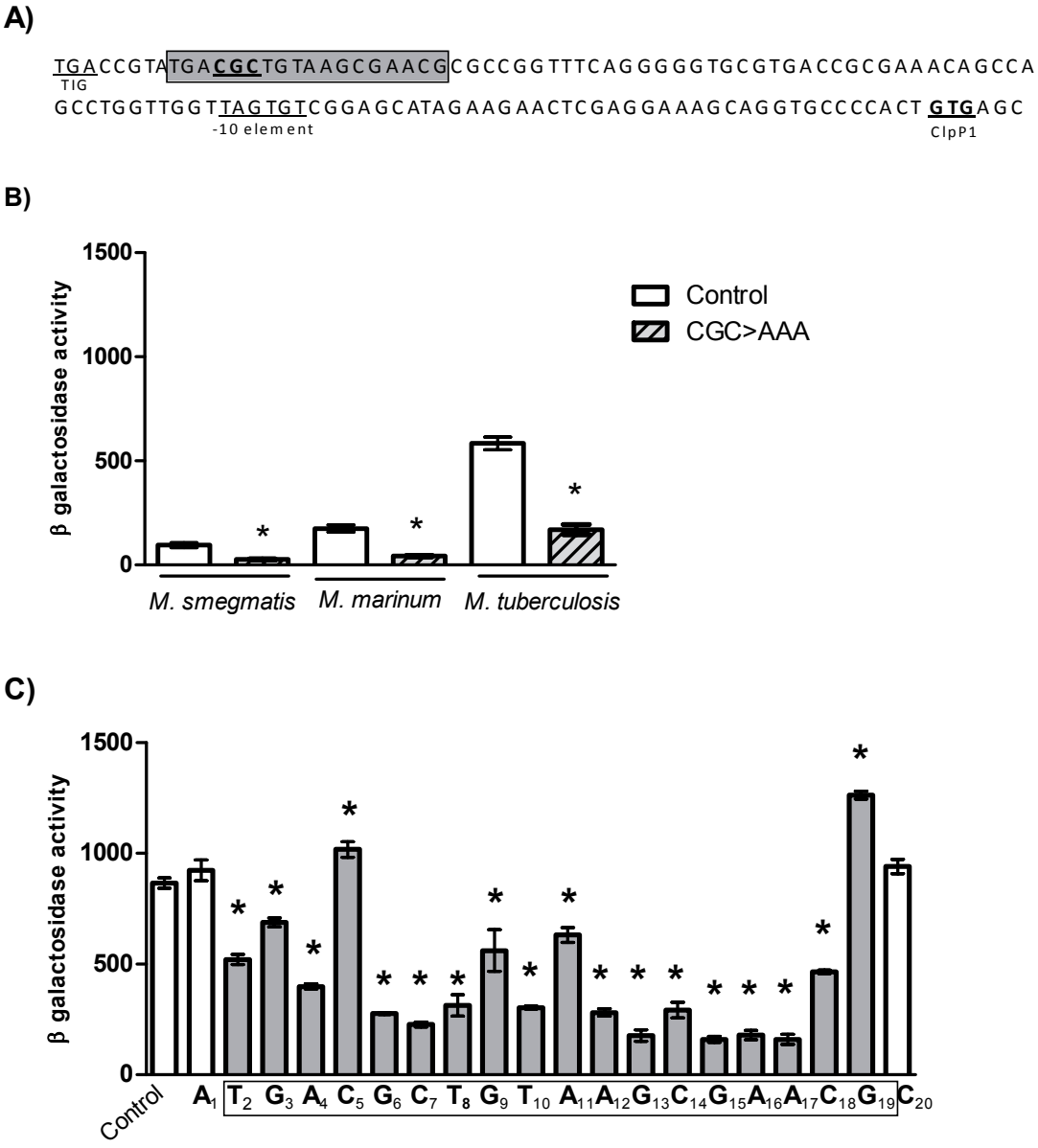

Figure 3.11 Mapping of a regulator binding site in the *clpP1P2* promoter region

A) Sequence of the region upstream of *clpP1P2*. Tig stop codon and ClpP1 start codon are indicated. Residues that constitute the binding site of a regulator are boxed. B) Identification of a putative regulatory site in *M. smegmatis*, *M. marinum* and *M. tuberculosis.* CGC (underlined bold) was mutated to AAA. C) Mapping of a regulator binding site in *M. tuberculosis*. Single nucleotide substitutions in P_{278} were made by SDM. Residues A or T were mutated to G or and residues C or G were mutated to A. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units- measured as nmol of O-nitrophenol produced per min per mg of protein. A significant difference of activity compared to unmutated P_{278} (control) is marked by an * ($p < 0.05$) using the student's t-test (unpaired, two sided).

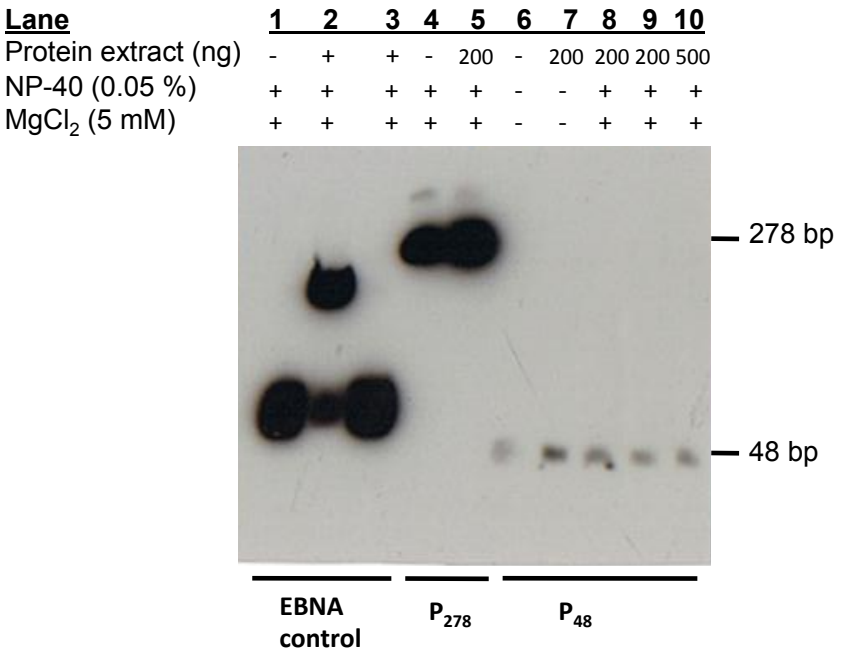

Figure 3.12 No interaction detected between ClgR and *clpP1P2* upstream regions

Binding reactions were performed with P_{278} (Lanes 4 and 5) and a 48 bp sequence (P_{48}) upstream of *clpP1* (Lanes 6 to 10). DNA fragments were PCR amplified with a biotin labeled primer, gel extracted and cleaned. 200 ng of purified ClgR protein extracts was used in lanes 7 to 9 while 500 ng was used in lane 10. Epstein Barr nuclear antigen (EBNA) system (EBNA) provided in the kit was used as a positive control (Lanes 1 to 3). Lane 3 contains excess unlabelled DNA.

to have prevented ClgR binding, thus binding of ClgR to a shorter DNA fragment was tested. This fragment of 48 bp (P_{48}) contained the predicted 18 bp ClgR binding site from the mutagenesis study and 15 bp on each side. However, no shift was observed with this fragment either (Fig 3.12). As expected, a shift was observed for the positive control reaction provided with the kit (Fig 3.12).

3.2.9 Identification of a second regulatory sequence

An increase in promoter activity was noticed when the GTGACC hexamer was mutated to CCGACC (Fig 3.4 D), suggesting that other regulator(s) besides ClgR may bind in the upstream region of *clpP1P2*. The presence of inverted repeats is an indication for possible regulator binding sites and the imperfect complementary inverted repeat sequence 'GTTTCAGGG-N_{58}-CAGGTGCCCC' was found upstream of *clpP1P2* (Fig 3.13 A). The last three nucleotides of the first part of the motif were mutated to A's by SDM and activities measured using *lacZ* in *M. tuberculosis*. Promoter activities of P_{125} and P_{278} were found to be significantly reduced (4.6 fold reduction) compared to the wild-type sequence, suggesting that this motif may be the binding region of a second positive regulator (Fig 3.13 B).

3.2.10 Protein-protein interactions

ClpP1 and ClpP2 are co-expressed and therefore are present in the cell under the same conditions. However, even though ClpP1 and ClpP2 are active under the same conditions they may be performing different functions by interacting with different ATPase subunits. Indeed, ClpP subunits need to associate with an ATPase partner to be able to degrade substrate proteins. The ATPase proteins can enlarge the number of substrates they can recognise by interacting with adaptor proteins such as ClpS, which is involved in recognition of substrates harbouring a N-degron. The presence of ClpP1 and ClpP2 in addition to multiple ATPase subunits in *M. tuberculosis* (ClpC1, ClpC2 and ClpX) suggests that several possible Clp protease complexes could form. ClpP1 and ClpP2 may have different affinities for different ATPases or different adaptors in order to recognise various substrates. Protein-protein interactions between the two ClpP proteins with ATPases and adaptor proteins were investigated using two independent methods. ClpB was also included in the screening to verify that it does not interact with any Clp proteins.

A)

TGACCGTATGACGCTGTAAGCGAACGCGCCG**GTTTCAGGG**GGTGCGTGACCGCGAAACAGCCA

Putative ClgR binding site

GCCTGGTTGGTTAGTGTCGGAGCATAGAAGAACTCGAGGAAAG**CAGGTGCCCC**ACTGTGAGC

-10 element ClpP1 start codon

B)

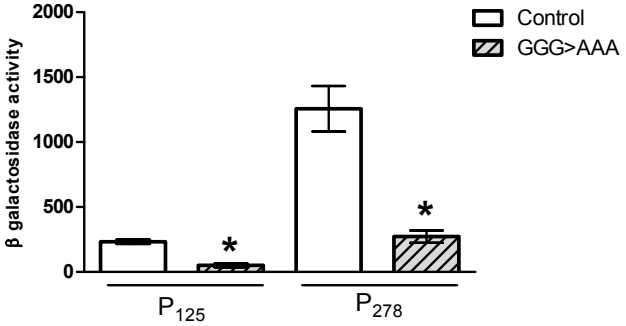

Figure 3.13 Identification of a second regulatory sequence

A) Sequence of the upstream region of *clpP1P2.* Sequence of a second putative regulatory site is indicated in bold. Putative ClgR binding site and -10 element are indicated. The three G residues (underlined bold) were mutated to AAA. B) Effect of GGG mutation on promoter activity. *M. tuberculosis* transformants harbouring the mutated or unmutated plasmid (control) were grown to late exponential phase in standing liquid cultures. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units (MU)- measured as nmol of O-nitrophenol produced per min per mg of protein. The background activity from pSM128 (control vector) ranged from 5 to 8 MU. A significant difference compared to the unmutated plasmid (control) is marked by an * ($p < 0.05$) using the student's t-test (unpaired, two sided).

3.2.10.1 Detection of interactions using the bacterial two-hybrid system

Interactions between the different Clp components proteins were studied with the bacterial two hybrid (BACTH) system (Karimova *et al.*, 1998) (Fig 3.14). This system uses the fast growing organism *E. coli* and interactions between *M. tuberculosis* proteins have been previously demonstrated using this method (Klepp *et al.*, 2009). In this assay, the proteins of interest are fused to two fragments (T25 and T18) of the catalytic domain of *Bordetella pertussis* adenylate cyclase (CyaA) and co-expressed in an adenylate cyclase deficient *E. coli* strain. Interaction between the two hybrid proteins results in functional complementation between the T25 and T18 fragments, leading to cAMP synthesis. cAMP binds to the catabolite gene activator protein CAP leading to transcription of *lacZ*. Thus, bacteria expressing interacting proteins form blue colonies on LB medium in presence of X-Gal whereas cells expressing non-interacting proteins remain white; in addition the strength of interaction can be quantified by measuring β-galactosidase activity.

In order to investigate interactions between the ClpP proteins, ClpP1 and ClpP2, the different ATPases ClpB, ClpC1, ClpC2, ClpX, and the accessory proteins, ClpS and SmpB, the bacterial two-hybrid system was used. DNA fragments encoding ClpP1, ClpP2, ClpC1, ClpS and SmpB were cloned into vectors as fusions to the C-terminal or N-terminal regions of the T25 domains (T25 or NT25) while DNA fragments encoding ClpB, ClpC1, ClpC2 and ClpX were cloned into vectors as fusions to the T18 domains (T18 or T18C) of adenylate cyclase (Karimova *et al.*, 1998). Pairs of recombinant plasmids were co-transformed into the *E. coli* adenylate cyclase deficient strain BTH101 (Karimova *et al.*, 1998). Interaction between the different proteins was determined by colour of the colonies formed on LB agar in the presence of X-Gal (white for non-interacting proteins, blue for interacting proteins) and the results were confirmed by measuring β-galactosidase activity (Table 1 and 2). The GCN4 leucine zipper motifs were used as a positive control for complementation (Karimova *et al.*, 1998). The level of β-galactosidase activity in liquid cultures for bacteria expressing the positive controls was typically 5,000 units per mg of dry weight bacteria while the background levels were around 80 units per mg of dry weight bacteria, confirming a strong interaction between the two control proteins (Table 1).

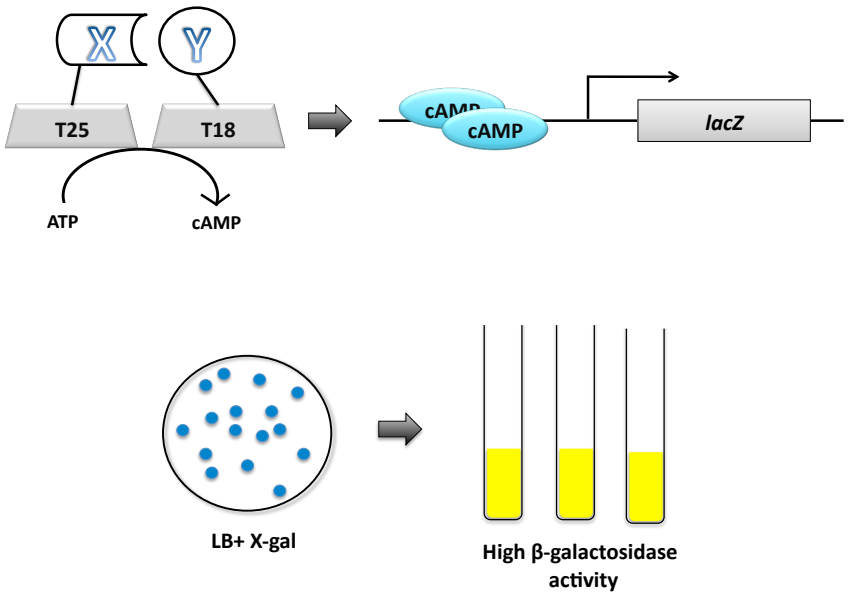

Figure 3.14 Principle of the *E. coli* two hybrid system based on a functional complementation of CyaA fragments

The T25 and T18 fragments correspond to amino acids 1–224 and 225–399 of the CyaA protein. If the two fragments, fused to two interacting proteins (X and Y), are brought into close proximity cAMP is produced. cAMP, bound to the transcriptional activator CAP, is a regulator of the expression of various genes including *lacZ*. Bacteria expressing interacting proteins form blue colonies on LB medium in the presence of X-Gal. Cells expressing non-interacting proteins remain white. The level of β-galactosidase activities, taken as an indicator of the strength of the interaction, can be measured in liquid cultures.

A)

		T25 fusion proteins		
		ClpP1-T25	**ClpP1- NT25**	**T25 zip**
T18 (C) fusion proteins	ClpP1	126 ± 45	64 ±37	
	ClpP2	78 ±7	48 ±9	
	ClpB	97 ±16	46 ±10	
	ClpX	85 ± 4	50 ±10	
	ClpC1	69 ±5	66 ±12	
	ClpC2	85 ±6	45 ±13	
	T18	88 ±8	/	
	T18C	/	51± 16	
	T18Czip	84 ±10	43 ±6	**4 193 ±298**

B)

		T25 fusion proteins		
		ClpP2-T25	**ClpP2- NT25**	**T25 zip**
T18 (C) fusion proteins	ClpP1	99 ±23	102 ±15	
	ClpP2	85 ±5	**346 ±147**	
	ClpB	84 ±9	160 ±34	
	ClpX	104 ±13	97 ±5	
	ClpC1	78 ± 15	65 ±5	
	ClpC2	100 ±5	122 ±38	
	T18	96 ±18	/	
	T18C	/	114 ±12	
	T18Czip	100 ±5	87 ±4	**5 509 ±278**

C)

		T25 fusion proteins		
		ClpC1-T25	**ClpC1- NT25**	**T25 zip**
T18 (C) fusion proteins	ClpP1	53 ±14	74 ±7	
	ClpP2	57 ±16	76 ±10	
	ClpB	89 ±19	78 ±5	
	ClpX	46 ±7	113 ±70	
	ClpC1	48 ±3	95 ±29	
	ClpC2	49 ±4	65 ±16	
	T18	50 ±10	/	
	T18C	/	103 ±29	
	T18Czip	47 ±5	87 ±12	**2 630 ±261**

Table 1 Detection of protein-protein interactions between A) ClpP1, B) ClpP2, and C) ClpC1 with the different Clp subunits

Transformants harbouring T25 (T25 or NT25) and T18 (T18 or T18C) fusion proteins were grown in presence of 0.5 mM of IPTG for 16h. The cells were permeabilised and ß-galactosidase activity was measured. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation and are given in units per mg of dry weight bacteria. One unit of activity corresponds to 1 nmol of ONPG hydrolysed per min at 28°C. Values significantly different from the empty vectors (approximately 80 units/mg of dry weight bacteria) using the student's t-test (unpaired, two sided) are indicated in bold ($p<0.05$). /: Non-tested combinations.

A)

		T25 fusion proteins		
		ClpS-T25	**ClpS- NT25**	**T25 zip**
T18 (C) fusion proteins	ClpP1	80 ±2	94 ±15	
	ClpP2	94 ±7	107 ±9	
	ClpB	82 ±3	90 ±12	
	ClpX	95 ±6	90 ±5	
	ClpC1	66 ±12	56 ±4	
	ClpC2	94 ±6	93 ±5	
	pUT18	94 ±9	/	
	pUT18C	/	108 ±6	
	pUT18Czip	96 ±4	105 ±5	**4 114 ±397**

B)

		T25 fusion proteins		
		SmpB-T25	**SmpB- NT25**	**T25 zip**
T18 (C) fusion proteins	ClpP1	69 ±29	39 ±8	
	ClpP2	46 ±10	35 ±11	
	ClpB	38 ±5	20 ±8	
	ClpX	42 ±6	16 ±11	
	ClpC1	41 ±8	37 ±15	
	ClpC2	57 ±34	27 ±9	
	pUT18	37 ±3	/	
	pUT18C	/	27 ±18	
	pUT18Czip	44 ±6	43 ±6	**4 948 ±1318**

C)

		T25 fusion proteins		
		Cfp10-T25	**Cfp10- NT25**	**T25 zip**
T18 fusion proteins	Esat6- T18	**8,305 ±3,023**	/	
	Esat6- T18C	/	86 ±5	
	T18	73 ±32	/	
	T18C	/	97 ±19	
	T18Czip	87 ±9	80 ±2	**4 789 ±615**

D)

		T25 fusion proteins		
		ClpP-T25	**ClpP-NT25**	**T25 zip**
T18 fusion proteins	ClpX- T18	73 ±5	/	
	ClpX- T18C	/	64 ±4	
	T18	68.6 ±10	/	
	T18C	/	80 ±20	
	18Czip	69.7 ±8	78 ±8	**4 789 ±615**

Table 2 Detection of protein-protein interaction, between A) ClpS and B) SmpB with the different Clp subunits; C) between Esat6 with Cfp10; D) between the *E. coli* ClpP and ClpX proteins.

Transformants harbouring the T25 (T25 or NT25) and T18 (T18 or T18C) fusion proteins were grown in presence of 0.5 mM of IPTG for 16h. The cells were permeabilised and ß-galactosidase activity was measured. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation and are given in units per mg of dry weight bacteria. One unit of activity corresponds to 1 nmol of ONPG hydrolysed per min at 28°C. Values significantly different from the empty vectors (approximately 80 units/mg of dry weight bacteria) using the student's t-test (unpaired, two sided) are indicated in bold ($p<0.05$). /: Non-tested combinations.

Since the ClpP tetradecameric complex is formed by two ClpP heptamers, it was predicted that interaction between ClpP1 and ClpP1 or ClpP2 would be observed. However, no interaction between ClpP1 and ClpP1 or ClpP2 was observed (Table 1A). In addition no interaction between ClpP1 and ClpC1, ClpC2, ClpX, ClpS or SmpB was seen, as all the measured activities were similar from the activity detected for the empty vectors (approximately 80 units) (Table 1A). ClpP2 interacted with itself (346 units), but did not interact with ClpP1, ClpB, ClpC1, ClpC2, ClpX, ClpS or SmpB (Table 1B). ClpP2 self-interaction was seen in one combination only (ClpP2-NT25 + ClpP2-T18C) suggesting that orientation of the fusion protein may influence protein folding. Similar observations were made during interaction analysis of *E. coli* FtsW and FtsI where interaction was seen in one configuration only (FtsI-T18 + FtsW-T25) (Karimova *et al.*, 2005).

Singh *et al*., (2006) previously demonstrated that ClpC1 interacts with ClpP2, thus it was expected to see an interaction between these two proteins. To determine if the lack of interaction could be due to the vector used (T18 or T25) *clpC1* was cloned into the T25 vectors (T25 and NT25) and co-transformed with ClpP1, ClpP2, ClpC1, ClpC2, ClpX, ClpS or SmpB fused to T18 domains (T18 or 18C). All the resulting strains appeared white on LB medium containing X-Gal and no activity was detected demonstrating that there was no interaction between ClpC1 and the other members of the complex in this assay (Table 1C).

ClpS is involved in recognition of substrates harbouring a N-degron, while SmpB participates in the formation of SsrA-tagged proteins. It was expected that interaction would be detected between ClpS and at least one ATPase for the degradation of proteins tagged at their N-terminus; and protein-protein interaction with SmpB was also tested to determine if SmpB had a role in substrate recognition. ClpS and SmpB were cloned into T25 vectors and tested for interaction with the different ATPases and ClpP proteins cloned in the T18 domains. ClpS and SmpB were not found to interact with any of the Clp proteins tested (Table 2 A and B).

To try to understand the lack of interactions described above it was decided to test if the interaction between two mycobacterial proteins could be detected with this system. ESAT-6 (early secreted antigen target 6) and CFP-10 (culture

filtrate protein 10) secreted antigens interact in *M. tuberculosis* and form a tight 1:1 complex (Renshaw *et al.*, 2002). ESAT-6 was cloned into the T18 vectors and CFP-10 was cloned into the T25 vectors. The plasmids were transformed into *E. coli* BTH101 and blue colonies were observed. β-galactosidase activity in liquid cultures was approximately 8,000 units per mg of dry weight bacteria (Table 2C). As previously seen with ClpP2-ClpP2 interaction, the interaction only occurred in one combination (CFP10-T25 + ESAT6-T18). The high level of activity indicates that ESAT-6 and CFP-10 proteins interact strongly and confirms that the system can be used to detect interaction between *M. tuberculosis* proteins. However, these proteins are monomers and do not reflect the complex structure of ClpP and ATPase proteins.

To see if the lack of interactions was due to the complex structure of the Clp proteins, interaction between *E. coli* ClpP and ClpX proteins, known to interact (Grimaud *et al.*, 1998), was tested. *E. coli clpP* was cloned into the T25 vectors while *clpX* was cloned into T18 vectors. *E. coli* BTH101 was co-transformed with the corresponding fusion proteins and plated on LB X-Gal medium. The bacterial colonies that were obtained were white and exhibited a β-galactosidase activity similar to the negative controls (Table 2D). This result shows that the interaction between *E. coli* Clp proteins could not be detected with this method.

3.2.10.2 Assay of interactions using Split-Trp

Many genes from mycobacteria yield folded proteins and active enzymes when expressed in *M. smegmatis*, whereas the same genes yield neither folded proteins or active enzymes in *E. coli* (Garbe *et al.*, 1993; Thangaraj *et al.*, 1990; Zhang *et al.*, 1991). It was therefore decided to use a second two-hybrid system, based on a mycobacterial species. Split-protein sensors are an important tool for studying protein interactions in living cells (Piehler, 2005). The Split-Tryptophan (Split-Trp) method uses a simple growth assay to detect protein-protein interactions (O'Hare *et al*., 2008). In this assay, the proteins of interest are fused to two fragments of *Saccharomyces cerevisiae* Trp1p involved in tryptophan (Trp) biosynthesis, and are co-expressed in a tryptophan-auxotroph *M. smegmatis* strain. Interaction between the two hybrid proteins results in reconstitution of active Trp1p and rescue of tryptophan auxotrophy. Interactions between *M. tuberculosis* proteins have been

previously demonstrated with this method using *M. smegmatis* as a bacterial host (O'Hare *et al*., 2008; Sacco *et al.*, 2007).

clpP1, *clpP2, clpB, clpC1, clpC2, clpX, clpS* and *smpB* genes were cloned into PL240 and PL242 vectors to generate fusion proteins with the N-terminal and C-terminal fragments of *Saccharomyces cerevisiae* Trp1p protein (O'Hare *et al*., 2008). Pairs of plasmids were electroporated into *M. smegmatis* Δ*hisA* and recombinant bacteria were grown in parallel on solid medium lacking or supplemented with tryptophan. If an interaction occurs between the proteins encoded by a pair of plasmids, Trp1p function is restored and the recombinant strain is able to grow in absence of exogenous Trp. Esat6-Cfp10 and C2-C1 plasmid pairs were used as positive controls. As expected, the two positive controls grew on medium lacking tryptophan, demonstrating interaction between the corresponding proteins (Fig 3.15). All strains harbouring Clp proteins grew readily on medium supplemented with tryptophan but could not grow on medium lacking tryptophan demonstrating there were no interactions between the different Clp proteins (Fig 3.15).

3.3 Discussion

The functional significance of the presence of two ClpP proteases in *M. tuberculosis* is currently unknown. *clpP1* and *clpP2* are co-expressed, thus excluding the possibility they are present under different conditions (Fig 3.2). ClpP1 and ClpP2 co-expression suggest a coordinated function for the two proteins, and ClpP1 and ClpP2 heptameric rings have been found to form a ClpP1P2 tetradecamer complex (Akopian *et al.*, 2012). Presence of a mixed complex was reported previously in *Listeria monocytogenes* (Zeiler *et al.*, 2011) and *Synechococcus elongatus* (Stanne *et al.*, 2007). The presence of a mixed complex could allow the possibility of degrading a large subset of protein substrates simultaneously if ClpP1 and ClpP2 recognise different substrates.

The promoter of the *clpP1/clpP2* operon was identified. Interestingly its activity was significantly higher in *M. tuberculosis* compared to the model organisms *M. smegmatis* or *M. marinum,* which are often used as genetic hosts for *M. tuberculosis* studies. A possible explanation for this may be the absence of regulatory control elements in the model organisms. The -10 sequence was identified (Fig 3.4 C) and found to match the consensus sequence for the binding of the sigma A factor; the principal sigma factor of *M. tuberculosis*

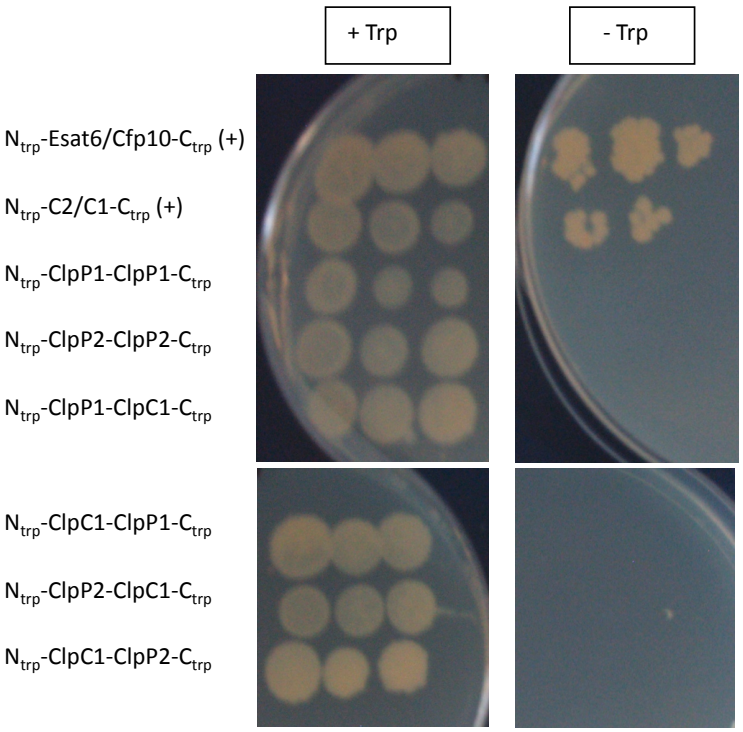

Figure 3.15 Split-Trp growth assay in *M. smegmatis* Δ*hisA*

Pairs of proteins were fused to N_{trp} and C_{trp} and co-expressed in *M. smegmatis* Δ*hisA* grown at 25°C on 7H9 agar in presence or absence of tryptophan. Undiluted and serial dilutions 1/10 and 1/100 were plated (from left to right).
-Trp: Medium lacking tryptophan; + Trp: Medium containing tryptophan, (+): Positive control of interaction

(Gomez *et al.*, 1998). No -35 element was identified (Fig 3.4 D), however this is not uncommon in mycobacteria. Several promoters that do not require a -35 region for activity have been identified but they typically contain a TGN motif immediately upstream of the -10 region described as an extended -10 element (Bashyam *et al.*, 1996). Since the *clpP1P2* promoter does not have this TGN motif and does not possess a -35 element, it is possible that a regulator makes contact with a subunit of the RNA polymerase to facilitate the polymerase binding.

The -10 element sequence of the *clpP1P2* operon (TAGTGT) has three nucleotides in common with the *E. coli* consensus sequence (TATAAT), however the intrinsic promoter activity is weak (Fig 3.5). High promoter activity is therefore dependent on the binding of regulator (s). A regulatory region was mapped: the 18 bp sequence identified forms an imperfect palindromic sequence (Fig 3.11) and matches the ClgR binding site in *C. glutamicum* (Engels *et al*., 2005). This regulatory region is present in P_{125} and P_{278} but binding of the regulator may be reduced in the shorter region due to presence of a terminator in the pSM128 plasmid close to the cloning site. Binding of ClgR in the upstream region of *clpP1P2* was previously demonstrated but could not be repeated in this study (Sherrid *et al*., 2010). The main difference with the previous study was the use of biotin labelled DNA while cy3 labelling was used previously. The presence of a binding site for a second regulator was suggested (Fig 3.13), however two positive regulators of ClpP1P2 activity would be surprising; thus it may be possible that the introduced mutations induced a conformational change in the sequence allowing a stronger binding of ClgR.

A knockdown strain of ClpP1P2 showed reduced growth (Carroll *et al.*, 2011) suggesting that a high promoter activity is necessary for optimal growth conditions. Promoter activity was slightly increased in late stationary phase (after an OD_{580} of 1.7) (Fig 3.8) suggesting that an increase in the Clp protease activity is necessary to degrade the accumulation of misfolded proteins emerging during stationary phase (Kwiatkowska *et al.*, 2008). The importance of the Clp proteases in stationary phase was previously demonstrated as *E. coli* and *B. subtilis clpP* mutants have a loss of viability during stationary phase survival (Msadek *et al*., 1998; Weichart *et al*., 2003).

Promoter activity during hypoxia was high and constant for up to eight weeks (Fig 3.10 A). Promoter activities are generally reduced during hypoxia as metabolic activity is greatly reduced, thus the high level of the ClpP1P2 promoter activity demonstrates an essential role for the Clp protease in this condition and suggests an important activity of the Clp proteases during infection. This confirms a previous report of high *clpP* expression in a hypoxic environment (Muttucumaru *et al.*, 2004) thus this illustrates the importance of the Clp protease in conditions where oxygen is limited, such as within a macrophage and this may explain why the Δ*clgR* strain was not able to replicate during macrophage infection (Estorninho *et al*., 2010). In addition, a ClpP1P2 knockdown strain displayed a reduction of virulence in macrophages further confirming their importance during infection (Carroll *et al*., 2011). Promoter activity was reduced between eight to 12 weeks of hypoxia but returned to its original activity quickly after reaeration (Fig 3.10 B) confirming previous reports of *clpP* induction during reaeration (Sherrid *et al.*, 2010). Reaeration is used to mimic reactivation of the disease and the quick return of promoter activity after hypoxia may suggest a role for the Clp proteases during reactivation of the disease. Targeting Clp activity, either directly or via ClgR, could therefore be a novel and attractive approach to prevent *M. tuberculosis* survival during infection and avoid reactivation of the disease.

The role of proteases is particularly vital during stresses that increase the occurrence of damaged proteins. *clpP* induction under heat shock conditions is variable among actinomycetes: while *clpP1P2* expression is induced upon severe heat stress in *C. glutamicum* (Engels *et al*., 2004), none of the *clpP* genes are heat induced in *S. lividans* (Bellier and Mazodier, 2004). Promoter activity was not increased at 42°C in *M. tuberculosis* (Fig 3.7) consistent with microarray studies which did not detect any transcript increase for either *clpP1* or *clpP2* following heat shock (45°C for 30 min) (Stewart *et al.*, 2002). Altogether this suggests that ClpP1 and ClpP2 are not involved in the heat shock response in *M. tuberculosis*.

The lack of induction after stress treatments demonstrates that the ClpP proteases are constitutively expressed. Since *clpP1P2* induction was previously demonstrated at the mRNA level after diamide or vancomycin treatment (Barik *et al*., 2009; Mehra and Kaushal, 2009), and mRNA levels are affected by both

transcription and degradation, expression of the two ClpP may be controlled by mRNA degradation.

ClpP1 and ClpP2 not being differentially expressed transcriptionally suggests that they may be regulated post-transcriptionally. To determine if regulation of activity was occurring at the level of substrate selection, interactions between all of the Clp proteins were examined. Two methods were used: the *E. coli* bacterial two-hybrid system and the Split-Trp systems. In the bacterial two-hybrid system, ClpP2 was found to interact with itself but not with any other members of the Clp system (Table 1) and no interactions were detected with the Split-Trp method (Fig 3.15). This was surprising since ClpX and ClpC1 both possess 'LGL' loops, a determinant for interaction, and interaction of ClpC1 with ClpP2 was previously demonstrated (Singh *et al.*, 2006). One plausible explanation for the lack of interaction in these systems may be that the structure of the Clp complex, made of two heptameric rings binding to a hexameric ATPase, is too complex to be detected with these methods as suggested by the lack of interaction between *E. coli* ClpP and ClpX proteins tested. For example, the adenylate cyclase fragments might be trapped inside the barrel structure of the ClpP proteins or the hybrid proteins might not fold properly.

However Singh *et al*. (2006) demonstrated interaction between ClpC1 and ClpP2 using the mycobacterial protein fragment complementation (M-PFC), a two-hybrid system based in *M. smegmatis* such as the Split-Trp. The size of the protein mDHFP used for M-PFC is about 21 kDa while the fusion protein Trp1P used for the Split-Trp studies is 25 kDA so the size of the fusion protein does not explain the difference of results obtained. The reason for the detection of interaction with this system but not with the BACTH or Split-Trp is still unknown.

The Clp proteins were expressed from multicopy plasmids and therefore their expression levels were elevated. Clp over-expression could be toxic to the cell and the fusion proteins might aggregate into inclusion bodies. In addition, toxicity of over-expression of ClpP2 has been demonstrated in *M. tuberculosis* partially supporting this hypothesis (Ollinger *et al.*, 2011).

It is also possible that the *E. coli* or *M. smegmatis* endogenous Clp proteins were interacting with the recombinant proteins thus preventing detection of

interaction in the assays. The use of ClpP deletion strains may be an alternative for future studies to counteract this issue. It is important to note that when ClpP1 and ClpP2 were previously expressed in *E. coli,* the proteins were enzymatically inactive (Akopian *et al.*, 2012; Benaroudj *et al.*, 2011; Ingvarsson *et al.*, 2007). Proteolytic activity was detected only when the two ClpP proteins were present together in presence of small activating molecules (Akopian *et al.*, 2012). This demonstrates that ClpP1 and ClpP2 form a mixed complex in order to become active, thus it is possible that association of ClpP1 and ClpP2 with each other induce a conformational change that enable them to bind to the ATPases, which would explain the lack of interaction observed in the assays.

To conclude, data gained from this study showed that there is no difference in regulation of transcription of *clpP1* and *clpP2* as the two genes are co-expressed. The promoter region of *clpP1P2* was identified and the binding site for a positive regulator, most likely to be ClgR was mapped. The results generated also suggest that *clpP1* and *clpP2* are constitutively expressed and the levels of expression remain high during hypoxia indicating the Clp proteases are important for bacterial survival inside the host environment. Different interactions between the various Clp proteins were expected but were not demonstrated.

4 Substrate specificities of ClpP1 and ClpP2

4.1 Introduction

Bacterial proteins are usually targeted for degradation when they harbour a sequence tag. One such degradation tag, encoded by *ssrA*, is added to the C-terminus of a nascent polypeptide in a trans-translation reaction (Keiler *et al.*, 1996). SsrA-tagged proteins are subsequently degraded by proteases including the Clp proteases (Gottesman *et al.*, 1998). Proteolytic degradation directed by tmRNA can be exploited experimentally to produce proteins with altered half-lives. The *E. coli ssrA* tag has the sequence AANDENYALAA (Keiler *et al.*, 1996) but alternative degradation tags that vary in the final three amino acids (e.g. AAV, ASV) alter protein stability (Andersen *et al.*, 1998). The C-terminal end of the tag (LAA) is recognised by ClpX and degradation is enhanced by the adaptor protein SspB, which recognises the AANDENY portion of the tag (Flynn *et al*., 2001; Levchenko *et al*., 2000). The *M. tuberculosis* SsrA tag sequence has been identified and consists of a ten amino acids portion (AADSHQRDYA) and a terminal LAA sequence. SsrA-tagged GFP proteins are degraded in *M. smegmatis* (Blokpoel *et al.*, 2003; Triccas *et al.*, 2002), but a direct role of the Clp proteases in the degradation of SsrA-tagged proteins in *M. tuberculosis* has not been demonstrated and *M. tuberculosis* does not have a SspB homolog (Kim *et al.*, 2010). Since *M. tuberculosis* has two ClpP, it is possible that only one of them is involved in the degradation of SsrA-tagged proteins or that the two ClpP recognise different tag sequences.

One of the most common methods to investigate gene function is to create a deletion mutant. Since essential genes cannot be deleted, over-expression in combination with microarray profiling provides an alternative approach for investigating their role (Stewart *et al.*, 2004). Microarrays are commonly used for analysis or to identify global patterns of gene expression and offer the possibility to monitor the expression level of thousands of genes in parallel.

Expression of genes of interest can be inducible. Conditional gene expression systems generally consist of an inducible promoter that can be turned on and off in a controlled manner, so gene expression can be up or down-regulated. The acetamidase system was one of the earliest inducible promoter systems used in mycobacteria. It is based on the induction of the acetamidase gene of

M. smegmatis by acetamide through a complex mechanism involving several promoters, two positive regulators of gene expression (AmiC and AmiD) and one repressor (AmiA) are involved (Draper, 1967; Mahenthiralingam *et al.*, 1993; Narayanan *et al.*, 2000; Parish *et al.*, 1997; Parish and Stoker, 1997; Roberts *et al.*, 2003). The system has been used for conditional over-expression of various genes of interest (Brown and Parish, 2006; Brown *et al.*, 2010; Kang *et al.*, 2005; Manabe *et al.*, 1999; Parish *et al.*, 1997; Park *et al.*, 2008). In addition, several tetracycline-inducible promoter systems have also been used to express foreign and native genes in both fast and slow-growing mycobacteria (Blokpoel *et al.*, 2005; Carroll *et al.*, 2005; Ehrt *et al.*, 2005; Guo *et al.*, 2007). They are based on regulation of expression from a promoter through a tetracycline-responsive repressor (TetR). The repressors and the promoters used to express TetR vary between the systems, but the general principle of tetracycline-inducible gene expression is the same. In the absence of inducer, TetR binds to the operator region of the promoter, blocks transcription, and promoter activity is switched off. In the presence of tetracycline or anhydrotetracycline, a conformational change in the regulator prevents binding and the promoter is available for RNA polymerase to bind and start gene transcription (Klotzsche *et al.*, 2009).

B. thuringiensis has two ClpP proteins involved in different cellular pathways: ClpP1 is essential for normal cell division at low temperature whereas ClpP2 is required for motility and sporulation (Fedhila *et al.*, 2002). Thus one could hypothesise that the two ClpP proteins of *M. tuberculosis* also have different substrate specificities. *clpP1* is essential for mycobacterial growth and *clpP2* is predicted to be essential (Ollinger *et al.*, 2011; Sassetti *et al.*, 2003), their deletion is therefore not possible. To determine if ClpP1 and ClpP2 have different regulatory functions, one objective of the project was to analyse the effect of ClpP1 and ClpP2 over-expression on the transcriptome. It may be possible that ClpP1 and ClpP2 exhibit different proteolytic activities and have a separate range of protein substrates; one could be involved in general protein turnover while the other one may be involved in the degradation of specific substrate proteins and therefore ClpP1 or ClpP2 over-expression could have distinct effects on the transcriptome. One of the main substrates of the Clp proteases is the SsrA-tagged proteins. To determine which of the Clp proteases degrade SsrA-tagged proteins in *M. tuberculosis*, SsrA-tagged LacZ was used as a model protein, and to determine if ClpP1 and ClpP2 have different

substrate specificities their affinity with different SsrA tag sequences was measured.

4.2 Results

4.2.1 Effect of ClpP1 and ClpP2 over-expression on the mycobacterial transcriptome

Clp proteolysis contributes to a variety of functions in many bacteria, varying their levels may therefore have a serious effect on the transcriptome of the cell. Microarrays were used to analyse the global transcriptional response of *M. tuberculosis* to over-expression of both *clpP1* and *clpP2*. *M. tuberculosis* ClpP1-ClpP2 over-expressing strain (pOPPY7), under the control of the constitutive *hsp60* promoter, was obtained from Dr J. Ollinger (Ollinger *et al*., 2011). The over-expressing strain and a strain carrying the empty vector (pSMT3) were grown to late exponential phase in rolling cultures and total RNA was extracted (Rustad *et al.*, 2009b). cDNA was synthesised, labelled with fluorescent dye, and hybridised against labelled *M. tuberculosis* genomic DNA; to avoid variation coming from dye incorporation the dyes were swapped. Genes whose expression differed by at least two-fold were identified using one-way ANOVA (Table 3). It is to note that the expression data was of low intensity, thus explaining the low number of differentially expressed genes and limiting their reliability.

As expected*, clpP1* (5.1-fold) and *clpP2* (5.7-fold) were up-regulated and, in addition, six genes were found to be up-regulated in the ClpP1-ClpP2 over-expressing strain. These genes are responsible for various functions: Rv2205c encodes a hypothetical protein, Rv2526 encodes an antitoxin, DevB is involved in the pentose phosphate pathway, HemD is involved in the biosynthesis of siroheme and cobalamin; Wbbl2 is possibly involved in cell wall arabinogalactan linker formation and FadD13 is involved in lipid degradation. This suggests that the Clp proteases are involved in regulation of multiple cellular pathways.

Quantitative real-time PCR (qRT-PCR) was used to validate the results obtained from the microarray study (Fig 4.1). As Rv2526 antitoxin forms a complex with toxin Rv2527, the associated toxin was added to the qRT-PCR analysis to determine the effect of the over-expression of ClpP1 and ClpP2 on the whole operon. RNA levels were measured in strains over-expressing ClpP1

ORF	Gene name	Fold change	p value	Product
Rv2460c	*clpP2*	5.7	0.002	Clp protease proteolytic subunit 2
Rv2461c	*clpP1*	5.1	0.016	Clp protease proteolytic subunit 1
Rv2205c		2.1	0.044	Hypothetical protein
Rv1445c	*devB*	2.1	0.044	6-phosphogluconolactonase
Rv0511	*hemD*	2.0	0.044	uroporphyrin-III C-methyltransferase
Rv1525	*wbbL2*	2.1	0.044	Rhamnosyl transferase
Rv2526		2.3	0.044	Antitoxin
Rv3089	*fadD13*	2.5	0.016	Fatty-acid-CoA ligase

Table 3 Differentially expressed genes in the ClpP1-ClpP2 over-expressing strain

Total RNA was extracted from three *M. tuberculosis* liquid cultures, grown until late exponential phase, carrying pSMT3 (empty vector) or the plasmid over-expressing *clpP1* and *clpP2* (pOPPY7). cDNA was synthesised and labelled with Cy3 or Cy5 dyes and hybridised against genomic DNA. Genepix software was used for image reading and analysis was performed using GeneSpring v7.3 software. Genes which were more than two-fold up-regulated in the over-expressing strain with a p value of less than 0.05 were identified by ANOVA.

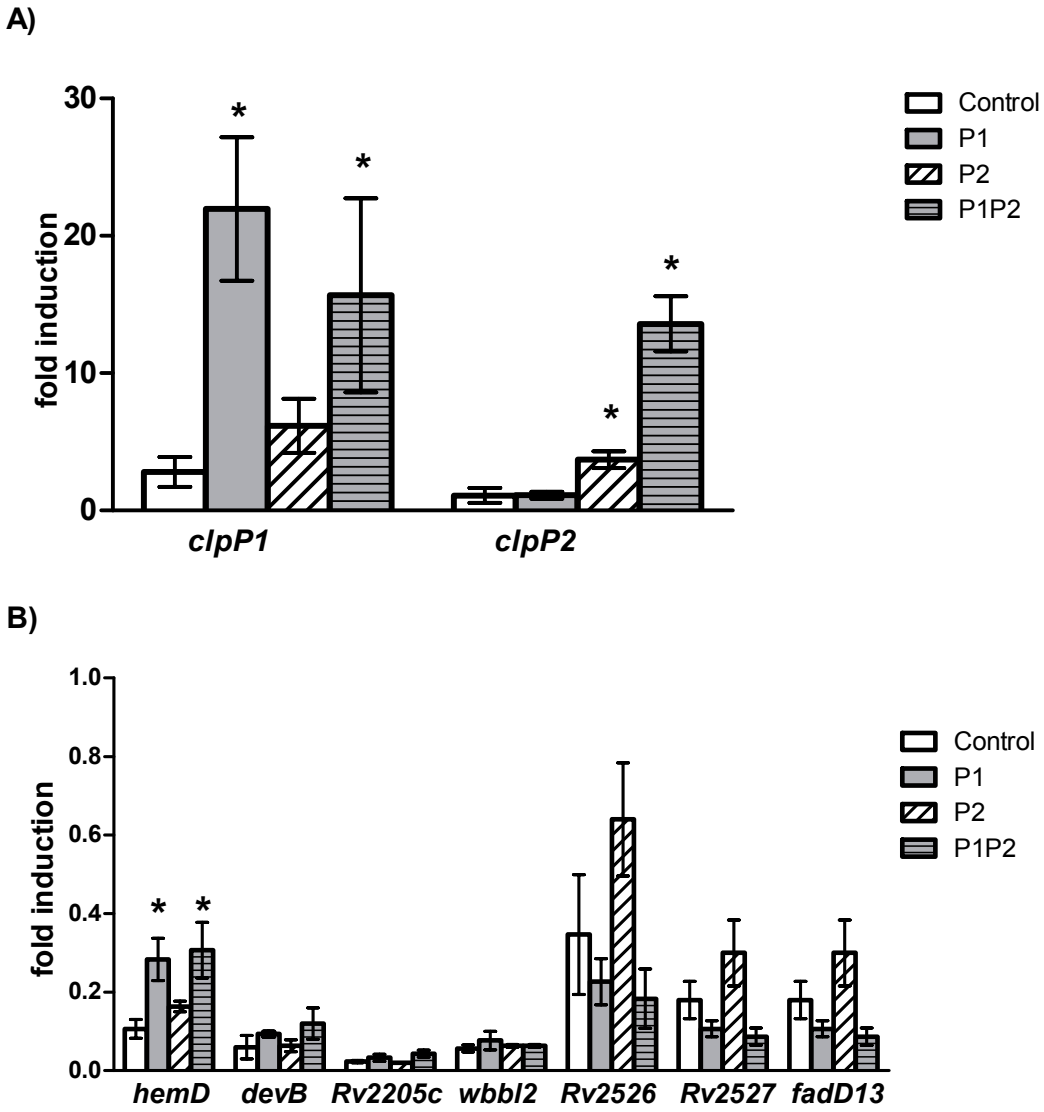

Figure 4.1 mRNA levels of target genes in strains over-expressing ClpP1 and/or ClpP2

Total RNA was extracted from three independent *M. tuberculosis* liquid cultures grown until late exponential phase and levels of mRNA were measured using qRT-PCR. A) Levels of mRNA of *clpP1* and *clpP2*. B) Levels of mRNA of target genes. The amount of mRNA is given as an arbitrary value standardised to *sigA* expression values. The mean ± standard deviation of three biological samples assayed in duplicate is given. A significant difference, measured by Student's t-test (unpaired, two sided), compared to the control is marked by an * ($p < 0.05$).

Strains are- Control: empty vector (pSMT3); P1: over-expressing ClpP1; P2: over-expressing ClpP2; P1P2: over-expressing ClpP1 and ClpP2.

and ClpP2 together or independently in order to differentiate specific effect of the over-expression of each protease. The results generated were normalised to *sigA,* whose expression is considered to be constant (Manganelli *et al.*, 1999).

As expected, *clpP1* and *clpP2* were both over-expressed in the double over-expressing strain (5.6-fold increase in *clpP1* and 12-fold increase in *clpP2*). A 7.8-fold increase of *clpP1* expression was measured in the ClpP1 over-expressing strain and a 3.4-fold increase was measured for *clpP2* expression in the ClpP2 over-expressing strain (Fig 4.1 A).

HemD expression was significantly increased in the ClpP1 and the ClpP1-ClpP2 over-expressing strains, but no significant difference of expression was noted in the ClpP2 over-expressing strain. There was no significant difference in the expression levels of *devB*, *Rv2205c*, *wbbl2*, *Rv2526*, *Rv2527* and *fadD13* between the control wild-type and all the over-expressing ClpP strains (Fig 4.1 B). The expression levels of the genes were very low compared to *sigA* showing that the selected genes are expressed at low levels. Overall, the microarray experiment did not show an important effect of ClpP1 and ClpP2 over-expression on the transcriptome.

4.2.2 Reporter system of ClpP1 and ClpP2 proteolytic activity

In order to assay ClpP1 and ClpP2 enzymatic activities, a LacZ based reporter incorporating an SsrA tag was developed. AANDENYA-LAA and AANDENYA-ASV protein tags were added to the C-terminal end of LacZ (referred as LacZ-LAA and LacZ-ASV respectively) and expressed under the control of the inducible acetamidase promoter (P_{ami}) from *M. smegmatis* (Parish *et al.*, 1997). These tags have previously been used to target protein for degradation in *M. smegmatis* and differ in the half-life of the tagged protein; GFP-LAA has a half-life of 165 minutes while GFP-ASV has a half-life of 110 minutes (Blokpoel *et al*., 2003).

Plasmids harbouring untagged or tagged LacZ were co-transformed in *M. tuberculosis* harbouring *clpP1*, *clpP2* or *clpP1P2* over-expressing plasmids and steady state levels of LacZ proteins were measured. Over-expression of ClpP1 or ClpP2 was used to determine if SsrA-tagged LacZ was degraded by one particular ClpP or by both of them. Liquid cultures of *M. tuberculosis*

transformants were grown under induced (presence of acetamide) or non-induced conditions, cell-free extracts were prepared and LacZ activity was determined using the β-galactosidase assay. The steady state level reflects a balance between protein synthesis and protein degradation, consequently if degradation of LacZ is performed by the ClpP proteases, a strain over-expressing the protease would have a higher protein turnover rate than the wild-type and therefore the steady state level of a substrate protein would be reduced. A reduction of the steady state level of LacZ will be visible by a reduction of its activity.

Under non-induced conditions untagged LacZ, LacZ-ASV, and LacZ-LAA were expressed at a similar level (approximately 2,000 Miller units) in the wild-type and the different over-expressing strains (Fig 4.2). This suggests that protein degradation by the Clp proteases was not occurring for untagged and SsrA-tagged LacZs when they were produced at low level.

There was a significant difference in LacZ activity between cells grown in induced conditions (acetamide) compared to cells grown in non-induced conditions for untagged (4.1-fold induction), ASV-tagged (4.8-fold induction) and LAA-tagged LacZ (2-fold induction), confirming that induction of the acetamidase promoter in *M. tuberculosis* was occurring and resulting in increase in LacZ levels (Fig 4.2).

No difference in the steady state levels of untagged LacZ activity was observed when ClpP1 was over-expressed (Fig 4.2 A). However, there was a significant decrease in the steady state levels of untagged LacZ when ClpP2 or ClpP1-ClpP2 were over-expressed (2.3-fold and 1.6-fold reduction respectively). This indicates that the rate of degradation of untagged LacZ was not increased when ClpP1 was over-expressed but the rate of degradation was increased when ClpP2 was over-expressed suggesting that ClpP2 degrades untagged LacZ but ClpP1 does not.

Incorporation of the ASV tag changed the dynamics of protein turnover. There was a significant decrease in steady state levels of LacZ-ASV when either ClpP1 or ClpP2 were over-expressed independently or together (1.3-fold reduction in all over-expressing strains) (Fig 4.2 B). This demonstrates that over-expression of either ClpP protease increased the degradation rate of

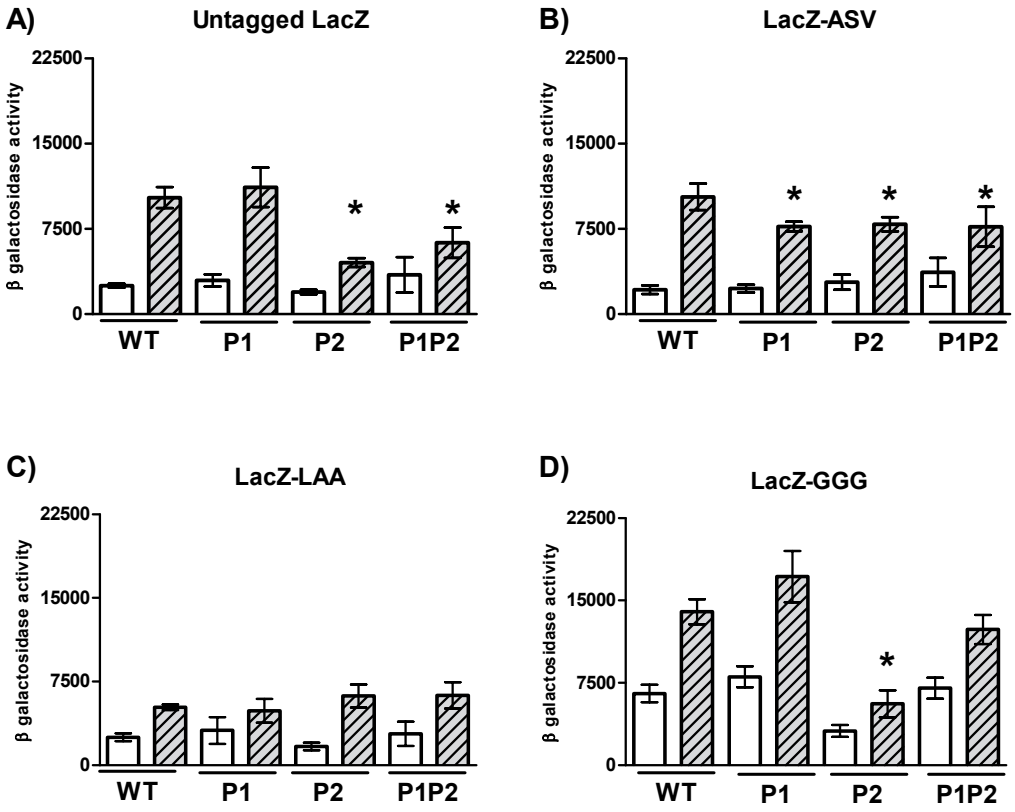

Figure 4.2 Activity of LacZ variants in over expressing ClpP strains

M. tuberculosis transformants were grown to late exponential phase in standing liquid cultures in presence of succinate +/- acetamide (0.1 % w/v) and cell-free extracts were prepared. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units- measured as nmol of O-nitrophenol produced per min per mg of protein. A significant difference measured by the student's t-test (unpaired, two sided) compared to the induced LacZ level in the WT strain is marked by an * ($p < 0.05$).

Empty bars: uninduced conditions (succinate); Grey striped bars: induced conditions (succinate + acetamide). Strains are- WT: wild-type; P1: over-expressing ClpP1; P2: over-expressing ClpP2; P1P2: over-expressing ClpP1 and ClpP2.

LacZ-ASV. This shows that this tag sequence directs protein degradation by both ClpP1 and ClpP2 in *M. tuberculosis*.

Similarly to *E. coli*, the last three residues of the *M. tuberculosis* SsrA tag sequence are LAA. Tagging LacZ with LAA had a profound effect on protein activity. A reduction in steady state quantities of LacZ-LAA was seen in all strains, even in the WT as compared to untagged LacZ (Fig 4.2 C); LacZ-LAA activity was about 5,200 Miller units in induced conditions in all strains while untagged and LacZ-ASV activity was about 10,000 Miller units in the WT strain. This demonstrates that LacZ-LAA is degraded quickly in the cell, thus the tag ending with LAA is more efficient at directing protein degradation compared with the tag ending with ASV. Over-expression of ClpP1 and ClpP2 did not have an effect on steady state levels of LacZ-LAA as the quantities were similar in all strains confirming that degradation does not occur when LacZ levels are low.

The *E. coli* ClpXP complex recognises the last three amino acids of the *ssrA* tag (AANDENYALAA) (Flynn *et al.*, 2001; Karzai *et al.*, 2000). To determine if the last three amino acids of the tag were also the determinant for recognition in mycobacterial degradation, the last three residues were mutated to GGG. Interestingly, the uninduced level of LacZ-GGG (7,400 MU) was higher compared to untagged and LacZ-ASV and LacZ-LAA (around 2,000 MU) (Fig 4.2 D). The steady state levels of LacZ-GGG were similar between the wild-type and over-expressing ClpP1 strains demonstrating that the protein was not recognised by ClpP1 for degradation; this indicates that the last three residues are the determinant for degradation by ClpP1. LacZ-GGG steady state levels were significantly reduced when ClpP2 was over-expressed (2.5-fold reduction) confirming that ClpP2 degrades proteins in the absence of a C-terminal degradation signal. However when both *clpP1* and *clpP2* were over-expressed the steady state levels were not significantly different from the wild-type.

4.2.2.1 Steady state levels of LacZ-ASV were not reduced after stress treatments

Since LacZ-ASV appeared to be a good indicator of ClpP1 and ClpP2 activity, LacZ-ASV steady state levels were measured after heat shock (42°C), oxidative shock (addition of diamide), or vancomycin treatment in order to determine if these conditions increased protein turnover (Fig 4.3). *M.*

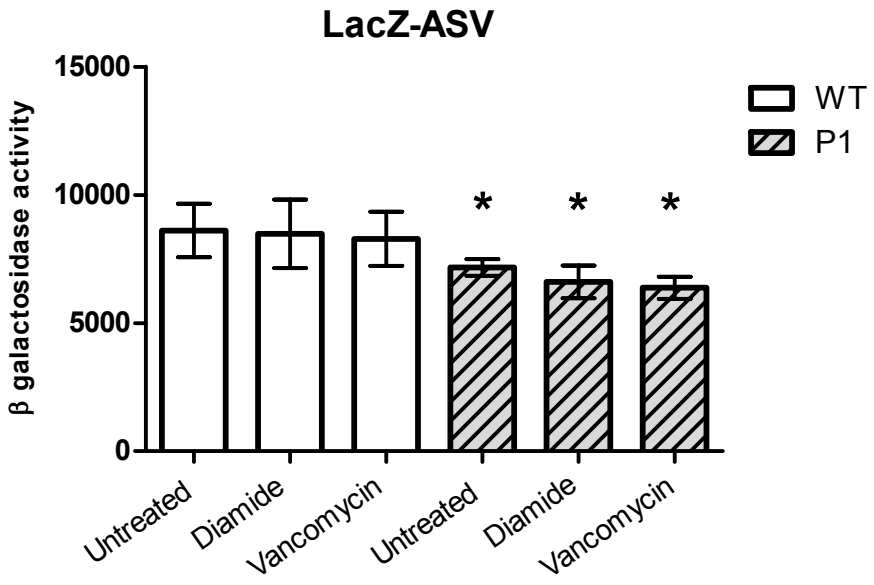

Figure 4.3 LacZ-ASV activity after heat and oxidative shocks

Three *M. tuberculosis* transformants carrying LacZ-ASV were grown to late exponential phase in standing liquid cultures in presence of acetamide (0.1 % w/v) and cell-free extracts were prepared. Treatments were 10 mM diamide for 1 h or 6 µg/mL of vancomycin for 90 min. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units- measured as nmol of O-nitrophenol produced per min per mg of protein. A significant difference from untreated WT is marked by an * ($p < 0.05$).

WT: wild type; P1: over-expressing ClpP1.

tuberculosis transformants carrying LacZ-ASV were grown in presence of acetamide to induce synthesis and steady state levels of LacZ were measured after stress treatments. There was no significant difference in LacZ-ASV steady state levels between untreated and stressed cultures suggesting that protein turnover was not increased in these conditions. LacZ-ASV levels were reduced when ClpP1 was over-expressed, as observed previously. However an increase in LacZ-ASV degradation in the ClpP1 over-expressing strain was not observed between untreated and stressed cultures suggesting that degradation by ClpP1 was not induced by these conditions, supporting evidence from the promoter data on the lack of induction of the Clp proteases.

4.2.2.2 Degradation kinetics

To verify that ClpP2 was degrading untagged LacZ and to measure the kinetics of degradation, untagged LacZ levels were measured in WT and over-expressing ClpP2 strains. P_{ami} is always expressed at a low level, even in the absence of acetamide (Parish *et al.*, 1997; Roberts *et al.*, 2003), thus this system is not suitable to follow protein turnover over time where a strict on/off switch is required. Therefore untagged LacZ was expressed under the control of an anhydrotetracycline (ATc) inducible promoter ($P_{smyc\text{-}tetO}$) (Ehrt *et al.*, 2005). Transformants were grown to late exponential phase in rolling cultures in presence of ATc to induce LacZ expression. Cultures were then washed and inoculated into new rolling cultures without ATc to follow LacZ turnover. In presence of 150 ng/mL ATc the level of LacZ was around 238 MU and in presence of 300 ng/mL ATc the level of LacZ was around 362 MU, confirming increased levels of LacZ in presence of increasing concentrations of ATc (Fig 4.4). Protein stability following the removal of ATc was measured, the level of LacZ was reduced from 238 MU to 30 and from 362 MU to 60 MU after two days, thus removal of ATc lead to a rapid decrease in untagged LacZ levels in the wild- type strain. Over-expression of ClpP2 was found to have no effect on the turnover rates observed (Fig 4.4).

4.3 Discussion

To determine if ClpP1 and ClpP2 were targeting different substrate proteins, the effect of their over-expression on the transcriptome was analysed. Antitoxins have previously been shown to be degraded by Clp protease complexes in *E. coli* and *Staph. aureus* (Aizenman *et al*., 1996; Donegan *et al*., 2009; Lehnherr and Yarmolinsky, 1995) so it was anticipated that this may also

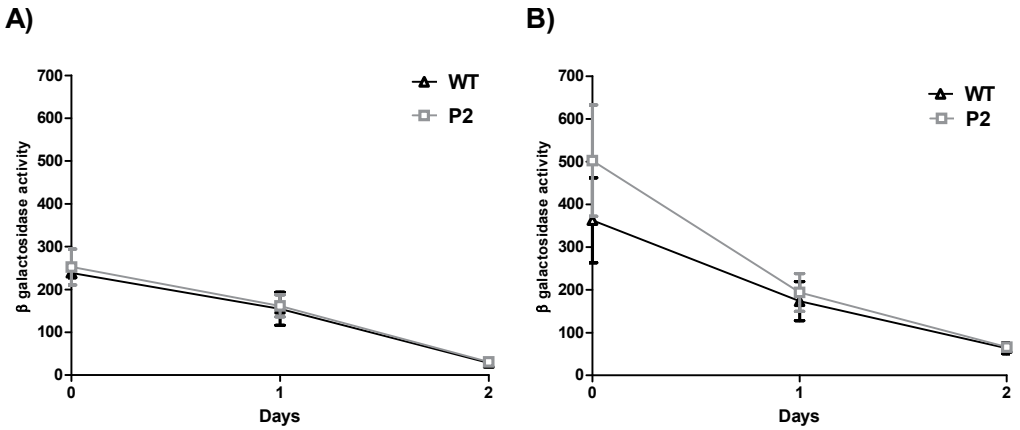

Figure 4.4 Untagged LacZ turnover in WT and ClpP2 over-expressing strains

M. tuberculosis transformants carrying untagged LacZ were grown to late exponential phase in rolling cultures in the presence of 150 ng/mL (A) or 300 ng/mL of ATc (B) and cell free extracts were prepared from 10 mL. Cultures were washed three times to remove ATc and new rollers were inoculated, cell-free extracts were prepared after 1 and 2 days of incubation. Results are the average activity of three independent transformants assayed in duplicate ± standard deviation. Activity is given in Miller Units- measured as nmol of O-nitrophenol produced per min per mg of protein.
WT: wild-type, P2: over-expressing ClpP2.

be the case in *M. tuberculosis*. Microarray analysis showed that six genes were up-regulated when ClpP1 and ClpP2 were over-expressed including antitoxin Rv2526 but over-expression of the toxin or antitoxin was not confirmed when the corresponding RNA levels were measured by qPCR (Fig 4.1). Despite the use of a strong constitutive promoter (P_{hsp60}), there was only a 3-fold up-regulation of *clpP2* expression in the ClpP2 over-expressing strain; toxicity of over-expression of ClpP2 has been reported for mycobacterial cells suggesting that the plasmid used may be unstable (Ollinger *et al*., 2011). Over-expression of ClpP1 resulted in increased *hemD* expression while over-expression of ClpP2 had no detectable effect on transcription of the genes measured (Fig 4.1). HemD is involved in the biosynthesis of siroheme and cobalamin (vitamin B12). Over-expression in HemD expression suggests that ClpP1 may degrade a negative transcriptional regulator of HemD. Interestingly, a potential ClgR binding site was identified upstream of *hemD* (Sherrid *et al*., 2010). ClgR is a positive transcriptional regulator, which is induced and degraded by the Clp proteases (Bellier and Mazodier, 2004; Engels *et al*., 2004). Degradation of ClgR in *S. lividans* is dependent on C-terminal sequence (AA) (Bellier *et al*., 2006), which has similarity to the *M. tuberculosis* ClgR C-terminus (AVA), suggesting that ClgR may be degraded by ClpP1 and/or ClpP2 in *M. tuberculosis*. ClpP1 and ClpP2 may be competing for ATPase binding so when ClpP1 is over-expressed it occupies most of the ATPases reducing ClpP2 activity. If ClgR is a ClpP2 substrate, a reduction in ClpP2 activity may result in ClgR over-expression which in turn activates *hemD* expression.

In order to determine if the Clp proteases were involved in the degradation of SsrA-tagged proteins in *M. tuberculosis* and if a difference in tag recognition was seen between the two ClpP proteins, unstable LacZ variants were used, which had different degradation tags. The first eight amino acids of the tags were similar (AANDENYA) and the last three residues differed (LAA, ASV, GGG). No LacZ degradation was observed in the non-induced conditions suggesting that degradation occurs only when protein concentration reaches a threshold level. When ClpP1 and ClpP2 were over-expressed individually or together, LacZ-ASV steady levels were significantly reduced, demonstrating that both ClpP1 and ClpP2 degrade LacZ-ASV (Fig 4.2 B). However degradation was not increased during heat or oxidative stresses (Fig 4.3). This demonstrates that the *M. tuberculosis* Clp proteases are involved in the degradation of proteins harbouring a C-terminus tag and degradation is not

increased after stress. When the ASV residues were mutated to GGG, ClpP1 degradation was not observed, demonstrating that the final three amino acids are the recognition signal for degradation by ClpP1 (Fig 4.2 D). *M. tuberculosis* SsrA tag sequence ends with LAA and LacZ-LAA was degraded faster than LacZ-ASV confirming the importance of the last three residues for proteolytic degradation (Fig 4.2 C).

Untagged LacZ and LacZ-GGG were not targeted by ClpP1 but were degraded by ClpP2 (Fig 4.2 A). Protein degradation of untagged LacZ in the cell was confirmed using a different promoter ($P_{smyc\text{-}tetO}$); degradation was occurring quickly as difference in protein levels was observed after one day (Fig 4.4). No difference in protein turnover was visible after one or two days when ClpP2 was over-expressed probably due to the fast turnover. One strategy to follow protein degradation within hours would be to use a fluorescent protein such as GFP for example. The difference in substrate degradation by ClpP1 and ClpP2 suggests that ClpP1 might be involved in the degradation of specific substrates while ClpP2 might be responsible for a general and central housekeeping function. The apparent lack of sequence determinant for ClpP2 degradation may explain the toxicity when ClpP2 is over-expressed, which may result in degradation of proteins necessary for growth for example.

In *E. coli,* ClpAP, ClpXP and Tsp protease complex preferentially target proteins which contain small and uncharged residues (Ala, Cys, Ser, Thr, and Val) in the last three positions of their sequence (Keiler *et al.*, 1996; Keiler and Sauer, 1996). A search of the *M. tuberculosis* genome for proteins whose last two residues consisted of any combination of the residues Ala, Cys, Ile, Leu, Ser, Thr, and Val and whose third-to-last residue was not Asp, Gly, Glu, or His, (found to be stabilising residues) revealed that 10% of the proteins encoded by essential genes may be susceptible to C-terminal proteolysis (Chang *et al.*, 2008). This high number of potentially short-lived proteins indicates that ClpP1 and ClpP2 may be involved in the regulation of a large range of pathways; most proteins identified in the study were hypothetical reflecting a lack of knowledge regarding these short-lived proteins. Functions of these short-lived proteins may be unravelled by studying regulatory proteolysis by the Clp proteases.

Protein degradation by the Clp proteases is dependent on ClpP proteins binding to an ATPase subunit. An increase in protein degradation of LacZ was

detected when the ClpP subunits were over-expressed suggesting that the number of ClpP subunits is the limiting factor for formation of the complex in the cell. The excess of Clp ATPase subunits in the cell is probably due to the fact that besides participating in protein degradation they also exhibit chaperone activity for refolding misfolded proteins (Wawrzynow *et al.*, 1996).

Accumulation of incorrectly folded SsrA-tagged proteins may be toxic for the cells. In *E. coli,* Clp, Lon, FtsH, and Tsp proteases are involved in the degradation of tagged proteins (Choy *et al.*, 2007; Gottesman *et al.*, 1998; Herman *et al.*, 1998; Spiers *et al.*, 2002). Since *M. tuberculosis* does not have Lon or Tsp homologs, FtsH and the Clp proteases may be the only proteases involved in the degradation of SsrA-tagged proteins. Both ClpAP and ClpXP complexes can recognise SsrA-tagged proteins in *E. coli* suggesting that different ATPases can be involved in their degradation although ClpXP is the major complex that degrades SsrA-tagged proteins in the cell (Gottesman *et al.,* 1998). The last three residues of the tag sequence, identified as the recognition motif for degradation by ClpP1, are recognised by ClpX in *E. coli* suggesting that ClpX is involved in the degradation of SsrA-tagged proteins in *M. tuberculosis.* Given that ClpP1 and ClpP2 can degrade SsrA-tagged proteins it is likely that they both interact with ClpX. ClpP2 is also able to degrade untagged protein demonstrating that, in addition to ClpX, ClpP2 can interact with another ATPase.

ClpX is able to interact with SsrA-tagged proteins and deliver them for degradation to ClpP. However the adaptor protein SspB enhances degradation of SsrA-tagged proteins in *E. coli;* SspB recognises the AANDENY portion of the tag which is different from the *M. tuberculosis* sequence (Flynn *et al.*, 2001; Levchenko *et al*., 2000). *M. tuberculosis* does not have a SspB homolog (Kim *et al.*, 2010) and ClpP1 and ClpP2 proteases recognise *E. coli* tag sequences suggesting that an accessory protein may not be necessary for degradation in *M. tuberculosis*.

To conclude, data gained from this study showed that both ClpP1 and ClpP2 are involved in degradation of SsrA-tagged proteins. ClpP2 is also involved in the degradation of untagged proteins demonstrating that ClpP1 and ClpP2 have different substrate specificities. ClpP1 may degrade SsrA-tagged proteins only, while ClpP2 may have a more general role in proteolysis. Substrate

recognition by ClpP1 is dependent on the final three amino acids of the tag sequence whilst ClpP2 does not appear to have any particular recognition signal.

5 Accessory components for Clp activity

5.1 Introduction

Interplay between the Clp ATPases, adaptor proteins and degradation signals ensure substrate recognition for degradation by the Clp proteases. Degrons can be encoded in the protein sequence or added post-transcriptionally and are located near the C- or N-terminus of the sequence. ClpS is involved in recognition of substrates harbouring a N-degron, while tmRNA and SmpB contribute in the addition of the SsrA tag at a protein C-terminus.

The adaptor protein ClpS directly interacts with destabilising N-terminal residues in *E. coli* and transfers substrate proteins to the ClpAP complex for degradation (Erbse *et al*., 2006; Schmidt *et al.*, 2009). In *E. coli* the loss of ClpS does not have phenotypic consequences but a *clpS* mutant in *Pseudomonas aeruginosa* was found to have increase sensitivity to β-lactams and exhibited defects in swarming motility and biofilm formation (Fernandez *et al.*, 2011; Schmidt *et al*., 2009). The presence of ClpS in *M. tuberculosis* indicates that this pathway is conserved in this species, but to date no mycobacterial substrates have been identified. ClpS is not predicted to be essential for mycobacterial growth (Sassetti *et al*., 2003) but is predicted to be required for survival in primary murine macrophages (Rengarajan *et al*., 2005).

Ribosomes can become stalled in the presence of a cluster of rare codons or in mRNA lacking a stop codon for example (Keiler *et al.,* 1996). The accumulation of stalled ribosomes can stop protein synthesis and prevent bacterial growth. tmRNA (encoded by *ssrA*) and its associated protein SmpB are employed in a trans-translation process to rescue stalled ribosomes and add a C-terminal tag to the incomplete nascent protein that will serve as a signal for proteolysis (Himeno *et al.*, 1997; Keiler *et al.*, 1996). Clp proteases are involved in the degradation of SsrA-tagged proteins (Gottesman *et al.*, 1998). tmRNA levels in *M. smegmatis* are high (Andini and Nash, 2011) suggesting a high rate of trans-translation in mycobacteria. Besides its role in ribosome rescue and signalling for protein clearance, tmRNA can play a regulatory role in gene expression (Ranquet and Gottesman, 2007) and acts as an antisense RNA to regulate expression of CrtM/N in *Staph. aureus* (Liu *et al*., 2010).

tmRNA is not essential for viability of many bacteria such as *E. coli* or *B. subtilis*, but it is important for cell growth and resistance to adverse conditions (Abo *et al.*, 2002; Komine *et al.*, 1994; Yang and Glover, 2009). *smpB* mutants have the same phenotype as *ssrA* mutants in *E. coli* indicating that SmpB is essential for trans-translation (Karzai *et al.*, 1999). SmpB is responsible for a variety of actions: it binds to tmRNA, protects it from degradation, enhances its aminoacylation efficiency and is required for stable association of tmRNA with the stalled ribosomes (Barends *et al.*, 2001; Hallier *et al.*, 2006; Karzai *et al.*, 1999). In *M. tuberculosis smpB* is not predicted to be essential whilst no data are available for *ssrA* (Sassetti *et al.*, 2003).

Clp proteases are involved in degradation of N and C-terminal tagged proteins which are recognised or generated by *ssrA*, *smpB* and *clpS*. To determine the importance and investigate the role of these three genes, construction of deletion mutants was attempted.

5.2 Results

The construction of *clpS*, *smpB,* and *ssrA* unmarked deletion mutants was attempted through recombineering (Fig 5.1). Allelic exchange by homologous recombination allows specific genes to be targeted for mutagenesis. Allelic exchange substrates (AES) were constructed to contain approximately 500 bp of DNA flanking each targeted gene, surrounding a hygromycin resistance cassette which possesses terminal *dif* sites (Cascioferro *et al.*, 2010) (Fig 5.1 B). Linear AES were electroporated in *M. tuberculosis* cells expressing the Che9 phage recombinases *gp60* and *gp61* to facilitate DNA recombination, conferring kanamycin resistance (van Kessel and Hatfull, 2007). The endogenous recombinases XerC and XerD recognise and resolve the *dif* sites so the resistance cassette is quickly excised from the chromosome. Following excision, one *dif* site of 28 bp is left at the chromosomal site. The strategy of replacing a gene of interest by a *dif* site allows the construction of an unmarked (no antibiotic marker) and in-frame deletion mutant.

5.2.1 Construction of a SsrA deletion mutant

An attempt to construct a *ssrA* deletion mutant in *M. tuberculosis* was made. The AES, containing a hygromycin resistance cassette flanked by *dif* sites, 500 bp of the *ssrA* flanking regions, and 48 bp of the gene, was constructed and

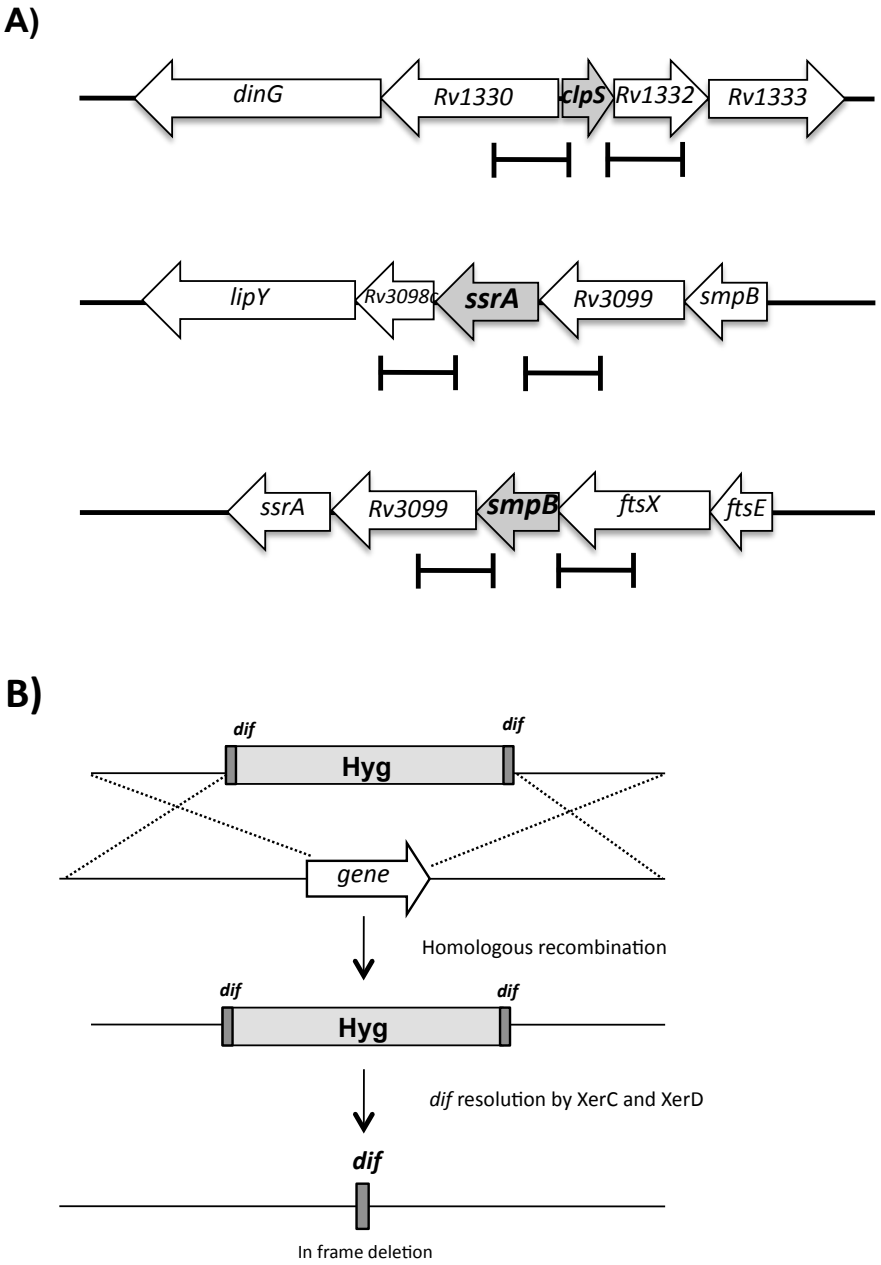

Figure 5.1 Method for construction of the mutant strains

A) Chromosomal organisation of *clpS*, *ssrA* and *smpB* in *M. tuberculosis.* Regions used for construction of the allelic exchange substrate are indicated. B) Method for construction of mutant strains. First, a linear AES recombines with the chromosomal region of the gene of interest by homologous recombination. The hygromycin (Hyg) cassette is incorporated into the chromosome and is then excised due to recombination between the *dif* sites. One *dif* site is left at the chromosomal locus.

electroporated into *M. tuberculosis* cells encoding phage recombinases; recombinants were then selected on hygromycin and kanamycin containing plates. To verify that the *ssrA* AES inserted into the correct chromosomal locus, 48 hygromycin resistant recombinants were isolated, DNA was extracted and Southern blotting was carried out. All the transformants screened contained the *ssrA* chromosomal band (1.8 kb) but also contained other bands at 2.4, 2.8 or 4 kb (Fig 5.2 A and B). The presence of the chromosomal *ssrA* copy and the presence of a second fragment of varying size suggest illegitimate recombination between the AES and the chromosomal DNA. Since none of the transformants lost the *ssrA* chromosomal copy, it is possible that *ssrA* is essential in *M. tuberculosis*.

To determine if *ssrA* is essential, a merodiploid strain was constructed by inserting an integrating vector, containing *ssrA* gene sequence and around 200 bp of the upstream region, into the chromosome at the mycobacteriophage L5 *attB* site (Lee *et al.*, 1991). The SsrA AES was then electroporated into the merodiploid strain and recombinants were selected on hygromycin, kanamycin and gentamicin (complementing vector marker). DNA was extracted from 16 transformants and Southern blotting was carried out. Twelve out of the 16 recombinants screened had the expected hybridisation pattern for a *ssrA* chromosomal replacement: the deletion band (1.2 kb) and the DNA cassette band (2.9 kb), while four recombinants (number 7, 8, 11 and 14) had the *ssrA* wild-type copy (1.8 kb band) (Fig 5.2 C). An additional band of approximately 4 kb indicates the presence of the complementing vector (Fig 5.2 C). Thus the chromosomal copy of *ssrA* could be deleted when a second functional copy was provided and it was not possible to do so in absence of a second copy (Fig 5.2 B); this demonstrates that *ssrA* is essential in *M. tuberculosis*.

tmRNA has two functions: rescue of stalled ribosomes and marking of the associated nascent polypeptides for clearance (Withey and Friedman, 1999). To determine which function is essential in *M. tuberculosis*, a second merodiploid strain was constructed where the integrated copy encoded a SsrA variant in which the last two codons of the tag reading frame (AA) were changed to DD (*ssrA*-DD). This variant enables the stalled ribosomes to be rescued but the resulting tagged proteins are not recognised by the proteolytic machinery and therefore accumulate in the cell

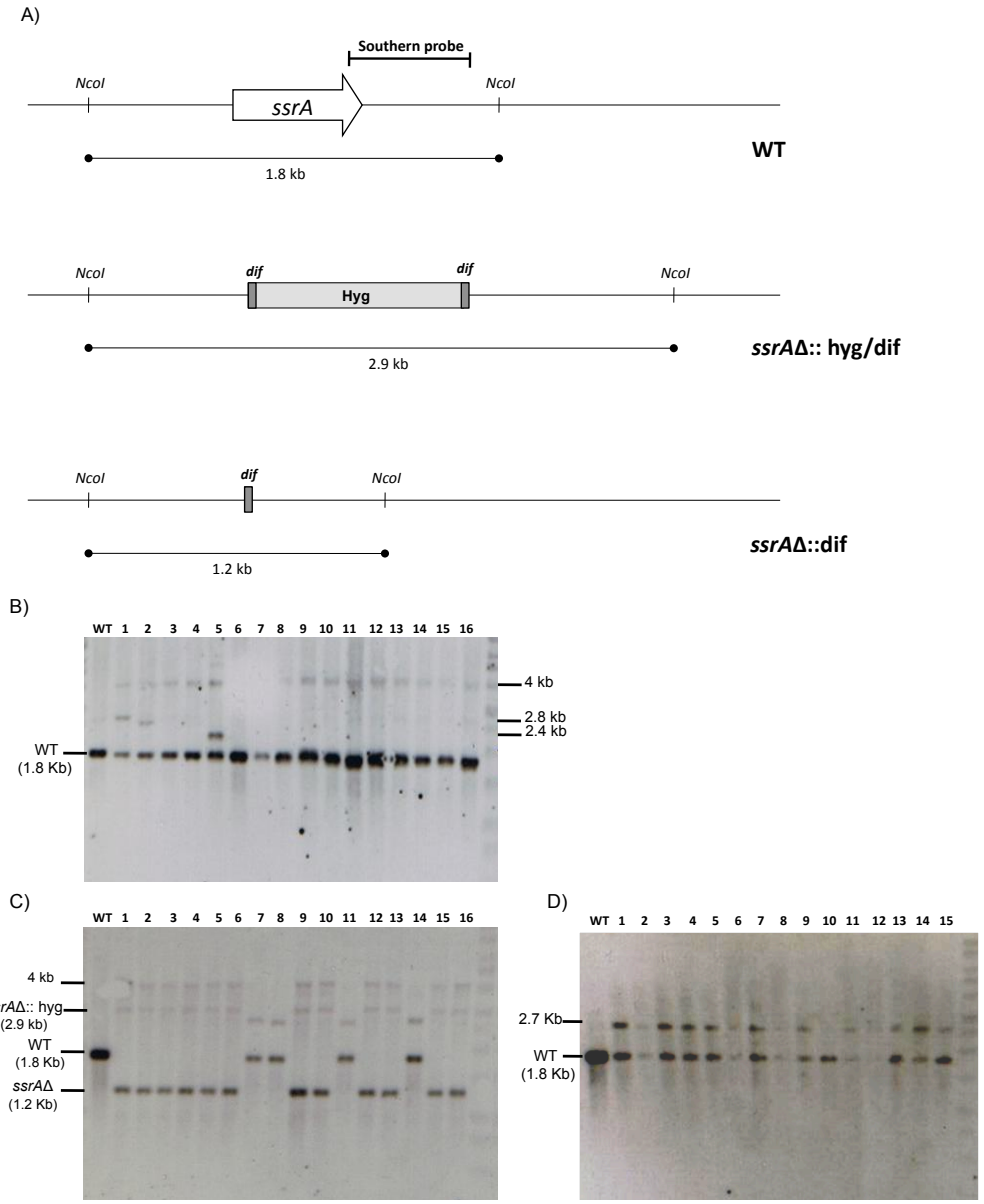

Figure 5.2 Demonstration of the essentiality of *ssrA* in *M. tuberculosis*

A) Expected sizes for Southern analysis. Region used as the probe for Southern blotting is indicated. B) Southern analysis of 16 *ssrA* recombinants. SsrA::hyg AES was electroporated in *M. tuberculosis* and 16 SsrA recombinants were isolated in presence of hygromycin, DNA was extracted and Southern blotting was carried out. Genomic DNA from wild-type (WT) H37Rv and *ssrA* recombinants was digested with *NcoI* and hybridised with the *ssrA* probe. C) Southern analysis of 16 recombinants in SsrA merodiploid strains. D) Southern analysis of 16 recombinants in SsrA-DD merodiploid strains.

(Gottesman *et al.*, 1998; Herman *et al.*, 1998; Keiler *et al.*, 1996). The SsrA AES was electroporated into this merodiploid strain and recombinants were selected on hygromycin, kanamycin and gentamicin. In this background, all the 16 recombinants screened had the WT band and an additional band at approximately 2.7 kb (Fig 5.2 D). This indicates that SsrA-DD is unable to functionally compensate for the loss of *ssrA*. The additional band (~2.7 kb) indicates where the AES inserted into the chromosome, the signal of the bands was faint on this blot which explains why the band of the complementing vector is absent.

5.2.2 Construction of a SmpB deletion mutant

To construct a SmpB deletion mutant the SmpB AES was constructed with a hygromycin resistance cassette flanked by *dif* sites, 500 bp of the *smpB* flanking regions and 47 bp of the gene. The AES was electroporated into *M. tuberculosis* cells encoding *gp60* and *gp61,* and recombinants were selected on hygromycin and kanamycin. Eight recombinants were isolated and further grown on plates containing hygromycin to maintain presence of the resistance hygromycin cassette or on plates without antibiotic selection to allow excision of the hygromycin cassette. DNA was extracted and Southern blot analysis was carried out to determine the site of insertion of the AES. Out of the eight recombinants, seven of them had the WT copy of the gene while one recombinant (number 2) had the expected hybridisation pattern for a deletion strain: the deletion band (1.5 kb) and the DNA cassette band (3.1 kb) (Fig 5.3 A and B). When the same recombinant was grown without antibiotic selection the proportion of cells that lost the resistance cassette was higher showing a rapid excision of the hygromycin cassette, indicated by a stronger deletion band (1.5 kb) on the Southern blot (Fig 5.3 B). Recombinant number 2 was further grown in liquid cultures without antibiotic selection and serial dilutions were plated on no antibiotic plates to allow *dif* excision. Transformants were then isolated in presence or absence of antibiotics in order to identify a sensitive recombinant that lost the hygromycin selection marker. DNA was extracted from one hygromycin sensitive transformant and analysed by Southern blot analysis. The expected genotype of the deletion mutant was confirmed; the deletion band of 1.5 kb was present while the 1.8 kb WT and 3.1 kb DNA cassette bands were lost (Fig 5.3 C). The mutant strain conserved 45 bp of the *smpB* original sequence while 436 nucleotides of the gene were deleted. Thus, an unmarked

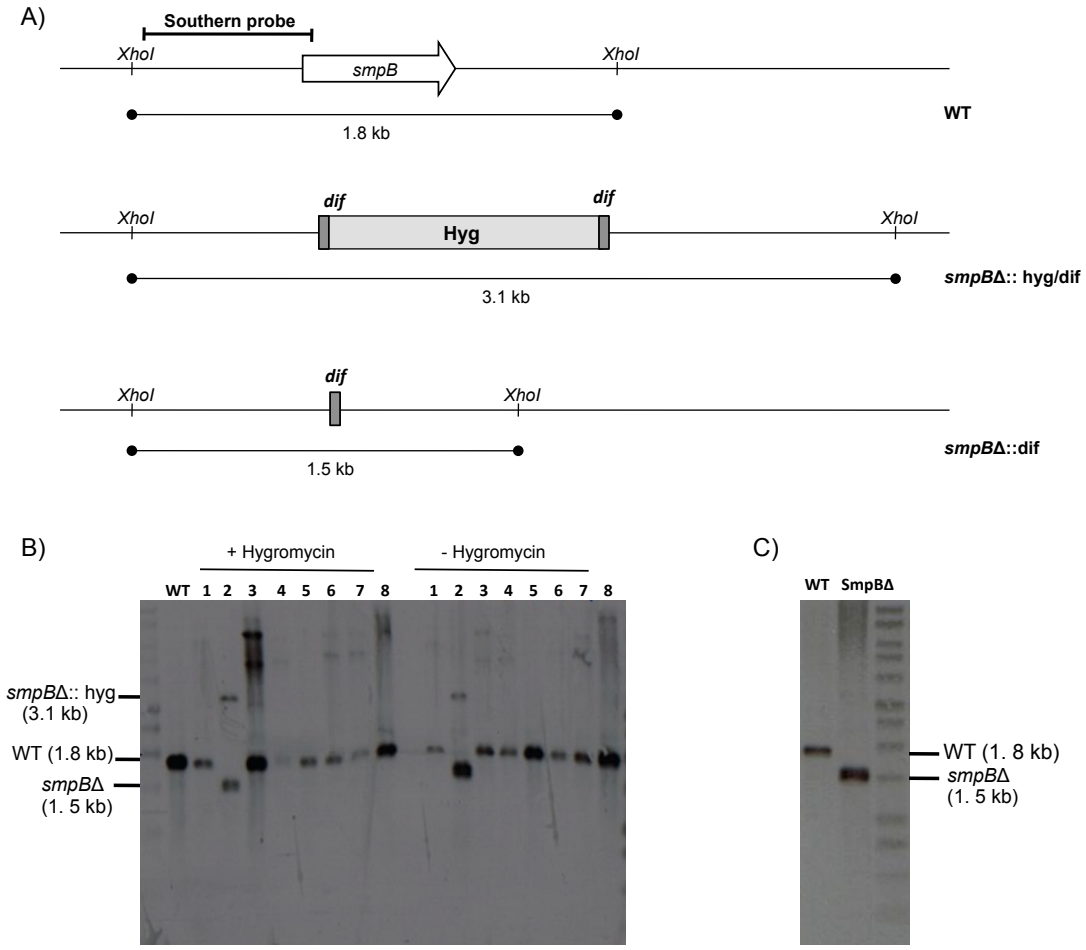

Figure 5.3 Construction of a mutant strain of SmpB

A) Expected sizes for Southern analysis. Region used as the probe for Southern blotting is indicated. B) Southern analysis of 16 *smpB* recombinants. SmpB::hyg AES was electroporated in *M. tuberculosis* and eight SmpB recombinants were grown in presence (+) or absence (-) of hygromycin, DNA was extracted and Southern blotting was carried out. Genomic DNA from wild-type (WT) H37Rv and *smpB* recombinants was digested with *XhoI* and hybridised with the *smpB* probe. C) Southern analysis of genomic DNA from WT and *smpB*Δ strains.

smpB deletion mutant was obtained demonstrating that SmpB is not essential for viability of *M. tuberculosis in vitro*.

A complementing vector was created containing the *smpB* sequence and its predicted promoter region (200 bp of the upstream region) and inserted into the chromosome at the mycobacteriophage L5 *attB* site (Lee *et al.*, 1991) in the SmpB mutant strain to create a complemented strain.

5.2.2.1 Phenotypic analysis of the SmpB mutant strain

A deletion mutant of SmpB was obtained but the loss of SmpB could have serious physiological consequences. To determine if loss of *smpB* had any effect on the growth of *M. tuberculosis*, growth of WT and the SmpB mutant strain (*smpB*Δ) was monitored in aerobic conditions. The deletion strain did not show any defect in growth demonstrating that SmpB is not required for aerobic growth in *M. tuberculosis* (Fig 5.4 A).

The SmpB-SsrA system is thought to play an important role in tolerance to sub-lethal concentrations of translation-specific antibiotics (Abo *et al.*, 2002), most likely by rescuing stalled ribosomes and targeting the associated protein fragments for proteolytic degradation. Thus growth of the WT, mutant, and complemented strain (smpB C'), was assayed in the presence of sub-lethal concentrations of antibiotics targeting ribosome function such as chloramphenicol and erythromycin. Both antibiotics target the 50S ribosomal subunit preventing protein synthesis.

Growth of the *smpB*Δ mutant was compromised in the presence of both chloramphenicol and erythromycin. *smpB*Δ mutant cells were able to grow in presence of 1 µg/mL of chloramphenicol but were unable to grow in 2.5 µg/mL, while the WT cells were able to grow with up to 5 µg/mL of chloramphenicol (Fig 5.4). There was a partial restoration of WT growth in the complemented strain as this strain was able to grow in presence of 2.5 µg/mL of chloramphenicol but not in 5 µg/mL. The growth of the *smpB*Δ strain was significantly reduced in the presence of 2.5, 5 and 10 µg/mL of erythromycin compared to WT strain (Fig 5.5). Once again, growth pattern of the complemented strain displayed an intermediate phenotype between the results observed for WT and mutant strains (Fig 5.5).

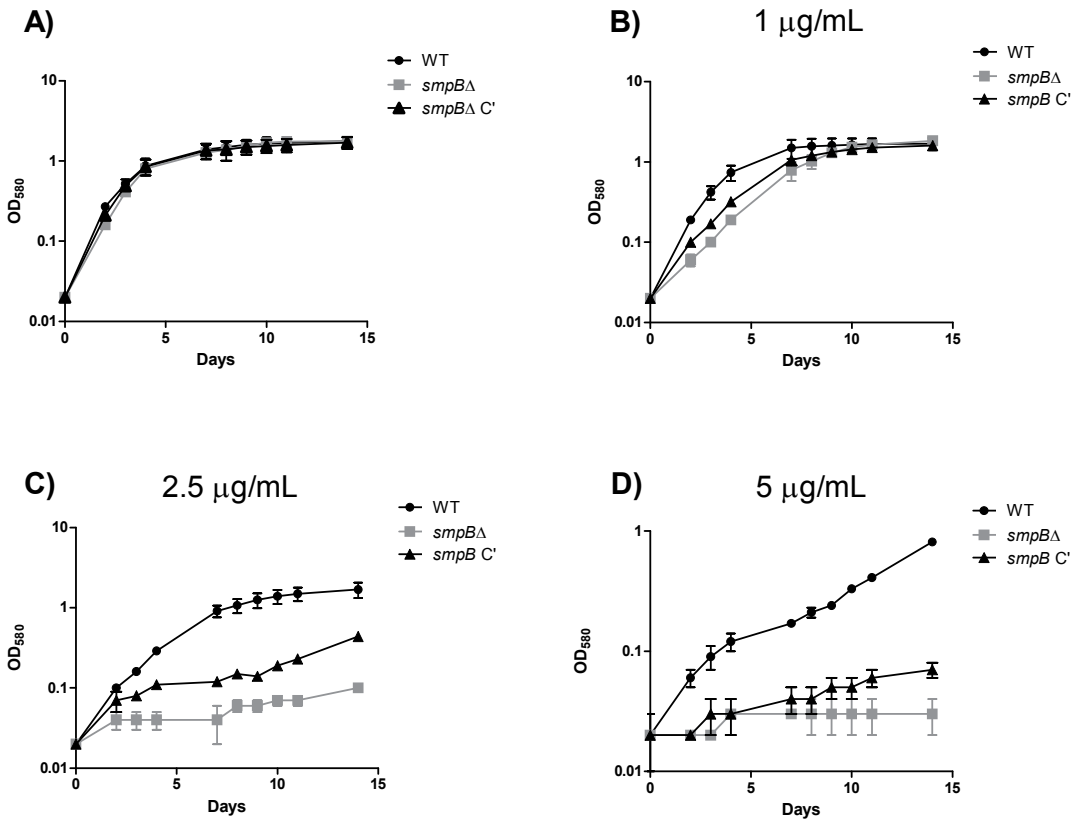

Figure 5.4 Deletion of SmpB leads to chloramphenicol sensitivity in *M. tuberculosis*

M. tuberculosis WT, *smpB*Δ and *smpB* C' strains were cultured in aerobic conditions in 7H9-Tw-AD medium in 16 mm diameter glass tubes and stirred at 120 rpm for 14 days. Results are the mean OD_{580} of three biological replicates ± standard deviation.

A) No antibiotics; B) 1 μg/mL of chloramphenicol; C) 2.5 μg/mL of chloramphenicol D) 5 μg/mL of chloramphenicol.

WT: wild-type; *smpB*Δ: SmpB deletion mutant; *smpB* C': complemented SmpB

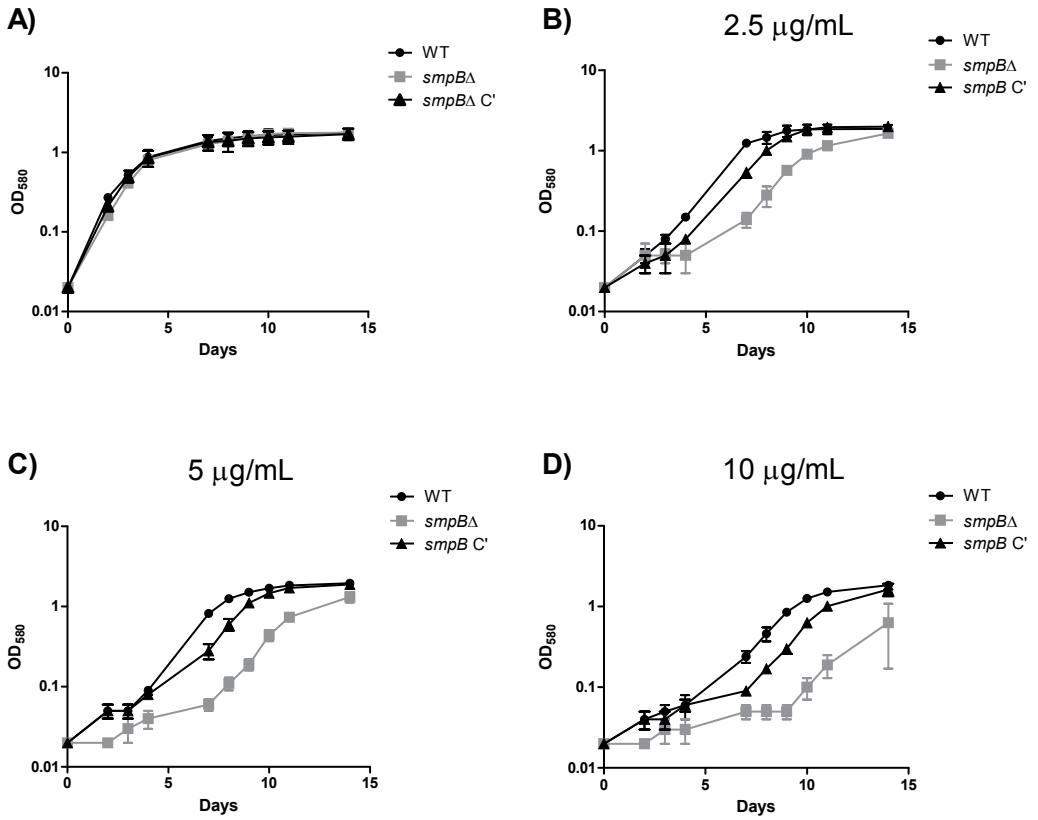

Figure 5.5 Deletion of SmpB leads to erythromycin sensitivity in *M. tuberculosis*

M. tuberculosis WT, *smpB*Δ and *smpB* C' strains were cultured in aerobic conditions in 7H9-Tw-AD medium in 16 mm diameter glass tubes and stirred at 120 rpm for 14 days. Results are the mean OD_{580} of three biological replicates ± standard deviation.

A) No antibiotics; B) 2.5 μg/mL of erythromycin; C) 5 μg/mL of erythromycin at D) 10 μg/mL of erythromycin.

WT: wild-type; *smpB*Δ: SmpB deletion mutant; *smpB* C': complemented SmpB

To determine if the *smpB*Δ growth defect was specific to the presence of translation inhibitors, the effect of rifampicin, an antibiotic that interferes with transcription and RNA elongation, was tested. There was no significant difference in rifampicin susceptibility between WT and *smpB*Δ strains for up to 4 ng/mL of rifampicin (Fig 5.6) and none of the strains were able to grow in 5 ng/mL. This shows that deletion of SmpB does not lead to rifampicin sensititvity, thus the growth defect of the mutant strain was specific for translation inhibitors.

Recent work has suggested that pyrazinamide targets RpsA which is involved in trans-translation (Shi *et al.*, 2011). Since SmpB is also involved in tmRNA activity, the effect of sub-lethal concentrations of pyrazinamide was tested on growth of the WT and the *smpB*Δ mutant. Since pyrazinamide is only effective at acidic pH, growth was assayed at pH 5.5. *smpB*Δ cells were able to grow at pH 5.5; thus SmpB is not necessary for growth under these conditions (Fig 5.7). The minimal inhibitory concentration (MIC) of PZA for *M. tuberculosis* varies from 6.25 to 50 µg/ml at pH 5.5 (Heifets, 2002; Mc Dermott and Tompsett, 1954; Stottmeier *et al.*, 1967). Growth of WT and *smpB*Δ were similarly affected with concentrations of up to 25 µg/mL of pyrazinamide (Fig 5.7) showing that sensitivity to pyrazinamide does not increase in the absence of SmpB.

The SmpB-SsrA system plays a crucial role in bacterial survival in hostile environments, it may therefore be essential inside a macrophage where the bacteria are subjected to adverse concentrations, e.g. limited concentrations of iron and magnesium, changes in pH, exposure to toxic nitrogen and oxygen species (Appelberg, 2006). For example, it was reported that *Salmonella smpB*Δ or *Yersinia pseudotuberculosis smpB-ssrA*Δ strains showed a defect in survival within macrophages (Baumler *et al.*, 1994; Okan *et al*., 2006). Therefore the ability of *M. tuberculosis smpB*Δ to survive and replicate in murine macrophages was assessed.

Resting macrophages were used to mimic initial or latent infection, and activated macrophages were used to model an active immune response. In resting macrophages, mutant and complemented strains both replicated and grew to the same level (Fig 5.8 A); in activated macrophages, both mutant and complemented strains showed similar survival (Fig 5.8 B). These results are

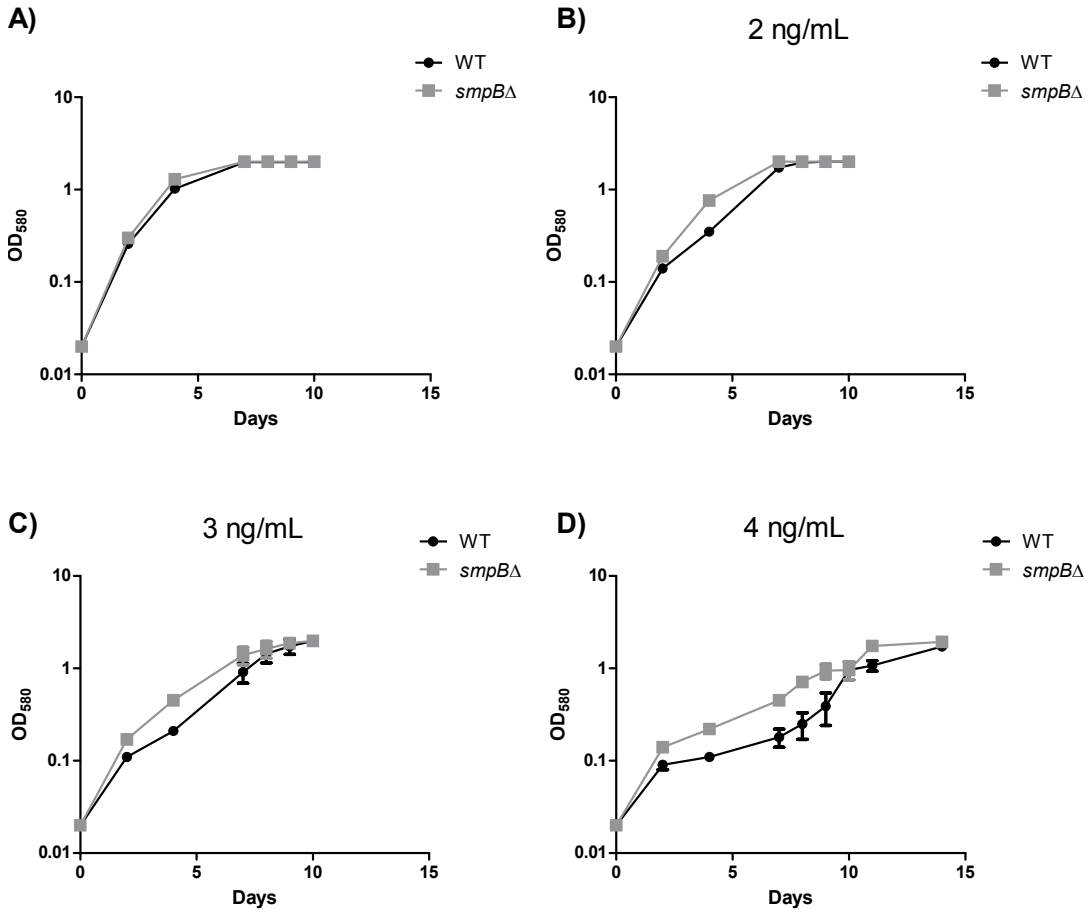

Figure 5.6 Deletion of SmpB does not lead to rifampicin sensitivity in *M. tuberculosis*

M. tuberculosis WT and the *smpB*Δ mutant strain were cultured in aerobic conditions in 7H9-Tw-AD medium in 16 mm diameter glass tubes and stirred at 120 rpm for 14 days. Results are the mean OD_{580} of three biological replicates ± standard deviation.

A) No antibiotics; B) 2 ng/mL of rifampicin; C) 3 ng/mL of rifampicin; D) 4 ng/mL of rifampicin.

WT: wild-type; *smpB*Δ: SmpB deletion mutant; *smpB* C': complemented SmpB

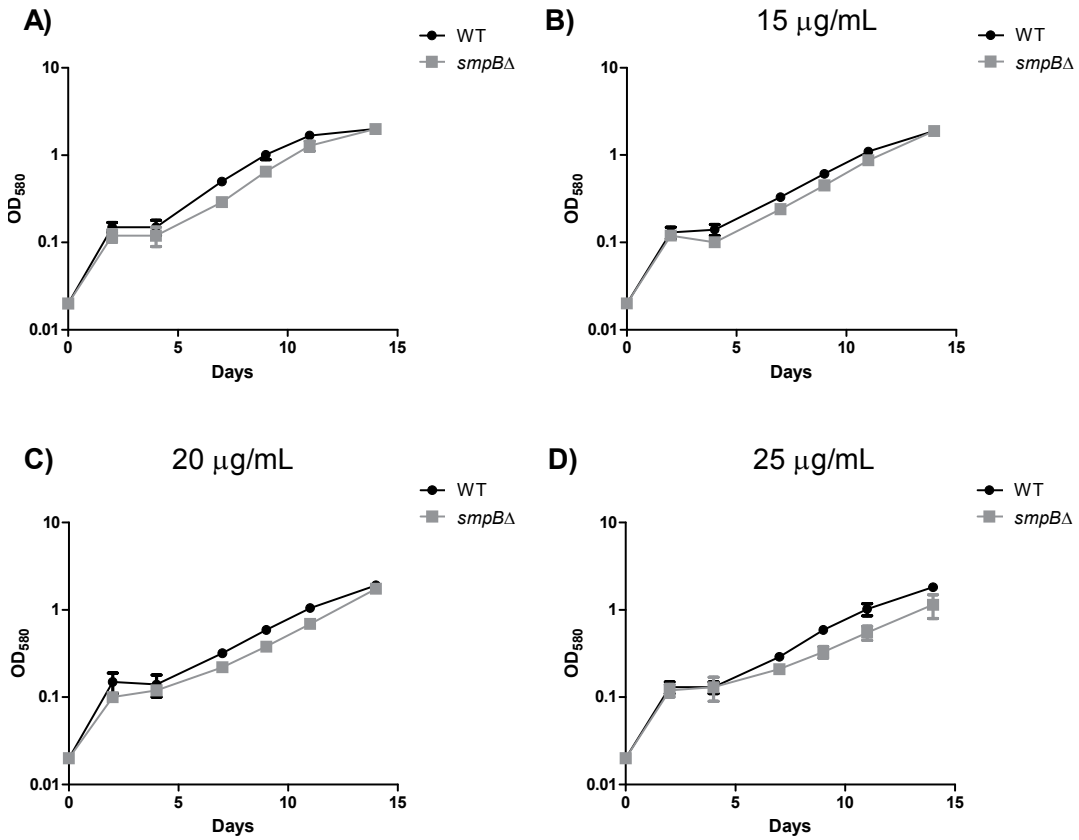

Figure 5.7 Deletion of SmpB does not lead to pyrazinamide sensitivity in *M. tuberculosis*

M. tuberculosis WT and the *smpB*Δ strain were cultured in aerobic conditions in 7H9-Tw-AD medium in 16 mm diameter glass tubes and stirred at 120 rpm for 14 days. Results are the mean OD_{580} of three biological replicates ± standard deviation.

A) No antibiotics; B) 15 μg/mL of pyrazinamide, C) 20 μg/mL of pyrazinamide; D) 25 μg/mL of pyrazinamide.

WT: wild-type; *smpB*Δ: SmpB deletion mutant; *smpB* C': complemented SmpB

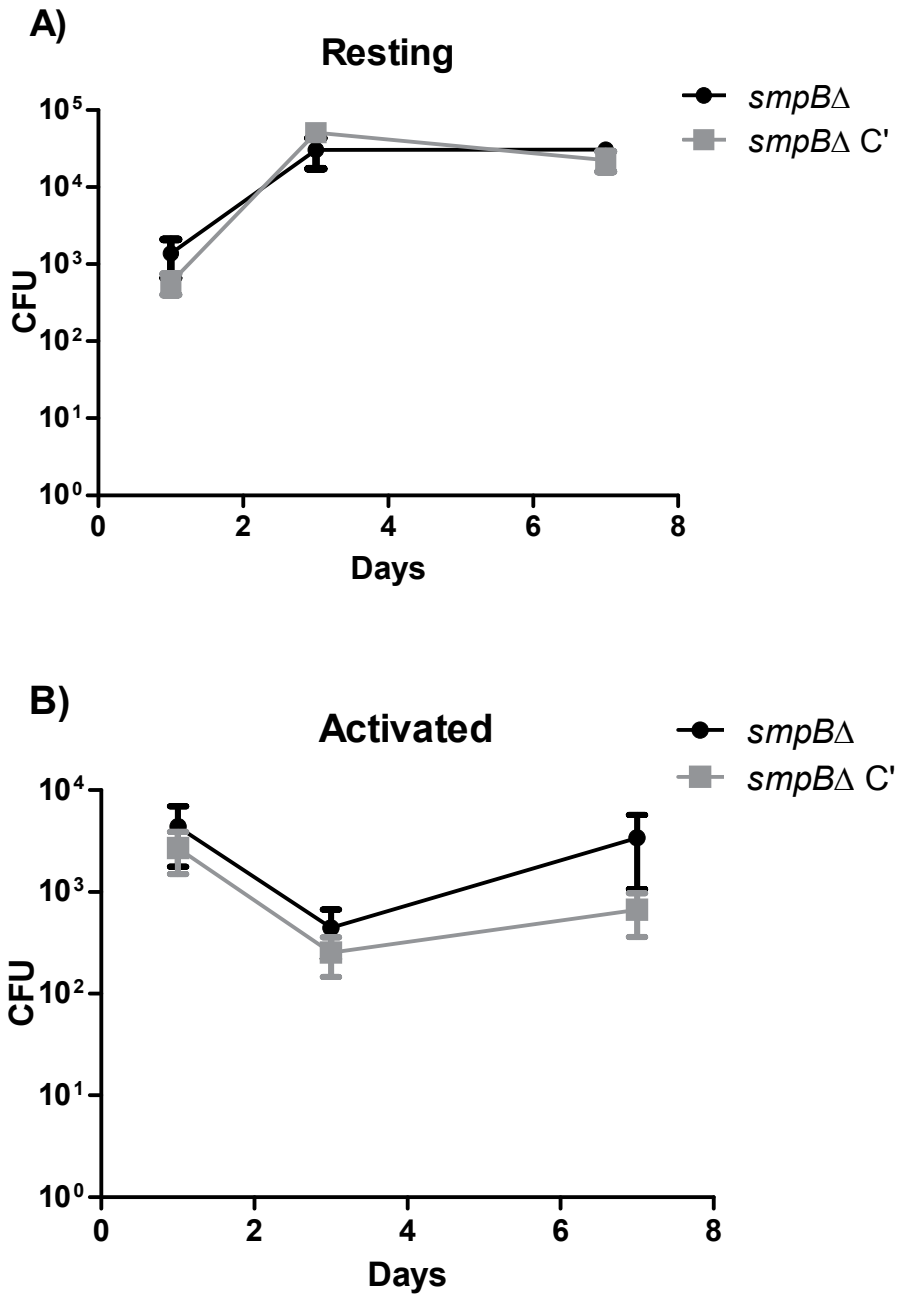

Figure 5.8 Growth of *smpB*Δ and complemented strains in macrophages

J774A murine macrophages were activated by the addition of 100 units/ml of IFN-γ 24 h prior to infection. Macrophages were infected with *smpB*Δ and *smpB*Δ C' strains with a multiplicity of infection of 1:10. Bacteria were harvested and CFU were determined. Results are the mean and standard deviations from three independent samples. The inoculum was $4x10^5$ CFU for *smpB*Δ and $6x10^5$ CFU for the *smpB*Δ C'.

*smpB*Δ: SmpB deletion mutant; *smpB* C': complemented SmpB; CFU: Colony forming units.

similar to what would be expected for the WT strain but in this experiment the CFU of the starting inoculum for the WT was significantly lower than the inoculum of the other strains, and therefore the data was not included. These data suggest that SmpB is not involved in the ability of *M. tuberculosis* cells to survive and replicate into macrophages.

5.2.3 Construction of a ClpS deletion mutant

An attempt to construct a *clpS* deletion mutant in *M. tuberculosis* was made using the same strategy as previously described for *ssrA* and *smpB*. The AES, containing a hygromycin resistance cassette flanked by *dif* sites, 500 bp of the *clpS* flanking regions and 50 bp of the gene sequence, was electroporated into *M. tuberculosis*. DNA was extracted from 32 transformants in total (from two different electroporations) selected on hygromycin and kanamycin containing plates and Southern blotting was carried out to confirm the correct recombination of the AES into the chromosome (Fig 5.9). All the tested transformants had the WT allele of *clpS* (indicated by the presence of a 3 kb band, Fig 5.9 A and B) with an additional band of ~9 kb (Fig 5.9 B). The membrane was re-probed with a probe designed to anneal to the hygromycin cassette. The wild-type band was no longer visible but the faint 9 kb band was present (Fig 5.9 C). This suggests that the AES did not integrate into the *clpS* locus but into another region of the chromosome by illegitimate recombination which occurs at high frequency in *M. tuberculosis* (Kalpana *et al.,* 1991). No deletion strains were obtained possibly due to the fact that ClpS is essential or due to the presence of another chromosomal region similar to the AES where it could recombine.

5.3 Discussion

tmRNA is thought to play a critical role in bacterial physiology but *ssrA* essentiality varies among bacterial species. In *E. coli,* where the tmRNA is not essential, other rescue pathways (ArfA and YaeJ) are present to recycle stalled ribosomes (Chadani *et al*., 2011b; Chadani *et al.*, 2010). In *M. tuberculosis, ssrA* could only be deleted from the chromosome when a second functional copy was provided demonstrating that *ssrA* is essential and that it is the only pathway responsible for ribosome rescue in *M. tuberculosis* (Fig 5.2).

A SsrA-DD variant rescues stalled ribosomes but addition of the peptide tag to the nascent polypeptide does not lead to degradation (Keiler *et al*., 1996). In *M.*

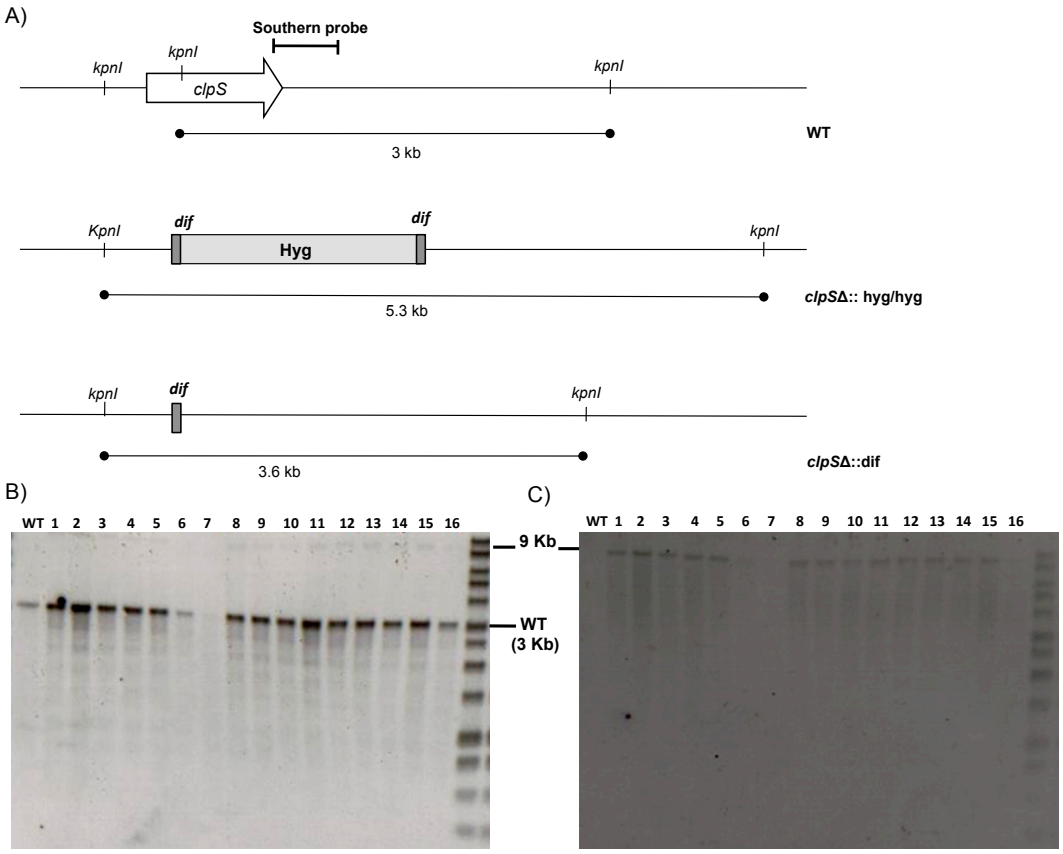

Figure 5.9 Analysis of ClpS recombinants

A) Expected sizes for Southern analysis. Region used as the probe for Southern blotting is indicated. B) Southern analysis of 15 ClpS recombinants. ClpS::hyg AES was electroporated in *M. tuberculosis* and 15 ClpS recombinants were isolated in presence of hygromycin, DNA was extracted and Southern blotting was carried out. Genomic DNA from wild-type (WT) H37Rv and ClpS recombinants was digested with *Kpn*I and hybridised with the *clpS* probe (B) and hybridised with the hygromycin cassette (C).
WT: wild-type.

tuberculosis it was possible to delete *ssrA* from the chromosome when a WT copy was provided but the ssrA-DD allele could not substitute for WT *ssrA* (Fig 5.2 C and D). This suggests that both ribosome rescue and clearance of truncated polypeptides functions are essential in *M. tuberculosis.* This is unusual as clearance of these polypeptides was not essential in *N. gonorrhoeae* where tmRNA is essential (Huang *et al*., 2000) and phenotypes in *E. coli* and *B. subtilis* were complemented by SsrA-DD (Muto *et al.*, 2000; Withey and Friedman, 1999). However, motility in *Y. pseudotuberculosis* and DNA replication control in *C*. *crescentus* are not complemented by the SsrA-DD variant, suggesting that tagging with the wild-type sequence is important in these species (Keiler and Shapiro, 2003; Okan *et al.*, 2006) similarly to the results obtained for *M. tuberculosis*.

Besides its function as tmRNA, SsrA was also shown to act as an antisense RNA to regulate expression of CrtM/N and influence the pigment synthesis of *Staph. aureus* (Liu *et al*., 2010) and this function is independent of SmpB. It is not known if SsrA has a regulatory role in *M. tuberculosis* at this time but if *ssrA* acts as an antisense RNA it may be involved in the reduction of translation of a protein, such as toxin, whose elevated expression is toxic for the cell; this could explain the essentiality of *ssrA* in *M. tuberculosis*.

SmpB is usually essential for tmRNA activity and *smpB* mutant cells have the same phenotype as *ssrA* mutants (Karzai *et al*., 1999). However, a *smpB* deletion mutant was viable in *M. tuberculosis.* Growth was not impaired under normal conditions or during a macrophage model of infection suggesting that SmpB may not be essential *in vivo* (Fig 5.8). In the presence of sub-lethal concentrations of chloramphenicol and erythromycin growth was inhibited but not in presence of rifampicin, indicating that the defect is specific to translation inhibitors (Fig 5.4, 5.5 and 5.6). Erythromycin was previously shown to increase the synthesis of tmRNA in *M. smegmatis* by inducing the *ssrA* promoter, which confirms that tmRNA activity is very important in response to ribosome inhibitors in mycobacteria (Andini and Nash, 2011). Pyrazinamide did not affect growth of SmpB mutant cells (Fig 5.7) confirming reports that this antibiotic is specific to RseA and does not target other components of the tmRNA system (Shi *et al*., 2011).

tmRNA stability depends on SmpB (Hallier *et al.*, 2004; Hong *et al*., 2005) so it is likely that in the absence of SmpB the level of tmRNA in the cell is reduced. The low level of tmRNA in the SmpB mutant may be sufficient to clear stalled ribosomes occurring in normal growth conditions. However tmRNA may be overwhelmed in conditions of stress when the number of stalled ribosomes increases and explains the growth defect observed in presence of antibiotics targeting the ribosome function.

The phenotype observed for *smpB*Δ in the presence of antibiotics targeting ribosomes was not completely restored in the complemented strain (Fig 5.4 and 5.5). This could be due to the fact that the upstream region cloned did not contain the full promoter region or not the true promoter if the gene is part of an operon for example.

ClpS is essential for the degradation of proteins harbouring a N-degron in *E. coli* (Schmidt *et al.*, 2009). The construction of a *clpS* deletion mutant would have helped to determine if Clp proteases participate in the N-end rule pathway in *M. tuberculosis.* The presence of two ClpP proteases allows the possibility that only one of them is involved in clearance of proteins harbouring a N-degron. However construction of a *clpS* mutant was not achieved due to technical difficulties (AES inserting into another locus due to illegitimate recombination) or due to the fact that ClpS is essential. Introducing the ClpS AES into a merodiploid strain would establish if ClpS is essential in *M. tuberculosis.*

To conclude, data gained from this study showed that SmpB is not essential in *M. tuberculosis* but deletion of *smpB* does have profound phenotypic consequences in presence of antibiotics targeting ribosomes. In contrast, the tmRNA mutagenesis study showed that *ssrA* is essential for *in vitro* growth of *M. tuberculosis* and its two functions, ribosome rescue and clearance of truncated polypeptides, are both important, which is unusual for bacterial cells. Thus, targeting tmRNA activity could be an attractive approach for new drugs.

6 General discussion

Clp proteases are involved in protein quality control by degrading misfolded or aggregated proteins, and contribute to the maintenance of cellular functions during normal and stress conditions. Clp are two component proteases composed of a proteolytic component associated with ATPase subunits and Clp ATPases can interact with accessory proteins for targeting additional substrates. The number of Clp proteins and composition of the complexes, as well as adaptor proteins that regulate them, vary between organisms. Deregulation of ClpP function is an effective way to induce bacterial death and could potentially be used in the treatment for tuberculosis (Brotz-Oesterhelt *et al*., 2005; Ollinger *et al*., 2011).

Although the majority of bacteria contain only one ClpP *M. tuberculosis* has two ClpP homologues, with at least one of them known to be essential for mycobacterial growth (Carroll *et al.*, 2011; Ollinger *et al.*, 2011). The overall aim of this work was to determine why two ClpP proteins are present and essential in this pathogenic organism. The main findings were that there is no difference in transcriptional regulation of *clpP1* and *clpP2* as the two subunits are co-expressed. The levels of expression were high and remained constant up to eight weeks of hypoxia. Using two different two-hybrid systems, I was unable to determine any interaction between ClpP1, ClpP2 and the different ATPase subunits ClpX, ClpC1 and ClpC2. A difference in proteolytic activity was found, which suggests that the two ClpP subunits actively target different substrates. The role of the accessory components was investigated: although construction of a ClpS deletion mutant was not achieved, tmRNA was found to be essential, and SmpB was not required for aerobic growth but was important for growth in presence of translation inhibitors.

The functional significance of having two ClpP subunits is not clear. Data gained from this study showed that ClpP1 and ClpP2 proteins are not expressed under different conditions, therefore excluding the possibility they may respond to different stress conditions (Chapter 3; 3.2.1). Instead they form a mixed tetradecameric complex (Akopian *et al.*, 2012) as also is the case in *Listeria monocytogenes* (Zeiler *et al.*, 2011) and *Synechococcus elongatus* (Stanne *et al.*, 2007). The *M. tuberculosis* tetradecameric complex is made of one ClpP1 and one ClpP2 ring, as opposed to the cyanobacterium

Synechococcus ClpP complex where each ring contains a mixture of ClpP3 and ClpPR subunits (Akopian *et al.*, 2012; Andersson *et al.*, 2009). The formation of a mixed complex may be beneficial if ClpP1 and ClpP2 interact with different ATPase subunits targeting different substrates, in order to allow degradation of varied substrates simultaneously.

The level of expression of *clpP1* and *clpP2* is high and ClpP1 and ClpP2 are constitutively expressed as demonstrated by the lack of induction after different stress treatments that induce protein misfolding (Chapter 3, 3.2.5). The constitutive expression of the two subunits associated with reports of ClpP1 and ClpP2 induction at the mRNA level, after diamide treatment for example (Mehra and Kaushal, 2009), suggests that ClpP1 and ClpP2 are regulated post-transcriptionally by RNA degradation. However, further regulation by protein degradation or post- translational modification cannot be excluded.

M. tuberculosis can survive within macrophages, believed to be deficient in oxygen and nutrient supply, and cause latent TB infection (Wayne and Sohaskey, 2001). Following changes in immunocompetency, due to HIV for example, the bacilli can be reactivated and cause active TB. Latent infection and reactivation are major problems for TB eradication but remain poorly understood. Oxygen depletion is commonly studied *in vitro* using the Wayne model (Wayne and Hayes, 1996). ClpP1 and ClpP2 were highly expressed for up to eight weeks of hypoxia, promoter activity was reduced between week 8 and 12 of hypoxia, and re-induced quickly following reaeration (Chapter 3, 3.2.7). The high level of promoter activity during hypoxia is unusual, as many genes are typically down-regulated, and may suggest that the Clp proteases are involved in degradation of proteins that become toxic after damage during infection or in the infection environment. Additionally, it has been shown that the Clp proteases regulate expression of stress response factors in numerous bacteria, such as sigma B in *B. subtilis* (Reeves *et al.*, 2007), so it is could be expected that they regulate factors which may be involved in *M. tuberculosis* colonisation, such as survival or virulence factors. Clp proteases were previously reported to be induced during reaeration, which is used to mimic reactivation of the disease (Sherrid *et al.*, 2010). Clp proteases are likely to control regulation of transcription factors necessary for the bacteria to reactivate and induce active TB, further work is however needed to identify

these factors. Targeting the Clp proteases may therefore be a successful approach to treat latent TB and to prevent reactivation of the disease.

The binding site for a transcriptional positive regulator of *clpP1P2* was mapped by SDM (Chapter 3, 3.2.8), this is likely to be the binding site for ClgR, which also positively regulates the ClpP proteins in *S. lividans* and *C. glutamicum* (Bellier and Mazodier, 2004; Engels *et al*., 2004). ClgR binding in the upstream region of *clpP1P2* and activation of *clpP1P2* transcription was previously demonstrated in *M. tuberculosis* (Mehra and Kaushal, 2009; Sherrid *et al.*, 2010) but the binding site had not been previously mapped. The sequence determined in this study is a perfect match to the ClgR consensus sequence in *C. glutamicum* suggesting a conserved mode of action for ClgR in actinobacteria (Engels *et al.*, 2005). ClgR binding is required for full promoter activity of ClpP1P2 expression, so targeting its activity may be an alternative approach to reduce Clp activity. A mutant of ClgR was found to be viable but had limited survival into macrophages (Estorninho *et al.*, 2010). A knockdown strain of ClpP1-P2 also had reduced growth and virulence in macrophages, confirming the importance of ClpP1 and ClpP2 for growth and infection (Carroll *et al.*, 2011)

Ribosomes decode information from messenger RNA but can stall in the absence of a stop codon on a defective mRNA. An accumulation of stalled ribosomes results in a shortage in the number of active ribosomes, leading to an arrest in protein translation and eventually to cell death. The tmRNA system, encoded by *ssrA*, has a dual role in rescuing the stalled ribosomes and tagging the incomplete nascent polypeptides (Huang *et al.*, 2000; Withey and Friedman, 1999). The presence of *ssrA* in *M. tuberculosis* was previously detected but the pathway has not as yet been characterised (Tyagi and Kinger, 1992). It was only possible to delete *ssrA* in the presence of an additional copy of the gene, thus *ssrA* is essential for *M. tuberculosis* growth (Chapter 5, 5.2.1). *E. coli,* whose tmRNA tagging system is non essential, has alternative rescue pathways (Chadani *et al.*, 2010; Chadani *et al.*, 2011b); as *ssrA* was found to be essential it indicates that *M. tuberculosis* only has this single pathway for ribosome rescue and tagging of nascent proteins; tmRNA is therefore an attractive target for further research work.

The ribosome rescue function of tmRNA is believed to be physiologically more important than its other function, which is to destabilise the nascent incomplete protein as a SsrA-DD variant can complement phenotypes associated with the loss of SsrA in most organisms (Huang *et al.*, 2000; Muto *et al.*, 2000; Withey and Friedman, 1999). Ribosome rescue and clearance of SsrA-tagged proteins are both essential in *M. tuberculosis* (Chapter 5, 5.2.1), which is unusual as only the ribosome rescue function has been previously found to be essential in *N. gonorrhoeae* where tmRNA is essential (Huang *et al*., 2000). The accumulation of SsrA-tagged proteins may be toxic for the cell, possibly due to an inappropriate cellular activity or their accumulation may congest the proteolysis pathways; this would emphasize the importance of the proteolysis function in *M. tuberculosis*.

SmpB is required for tmRNA stability and interaction with the ribosomes (Karzai *et al.*, 2000), but interestingly SmpB was not found to be essential in *M. tuberculosis* (Chapter 5, 5.2.2). A deletion mutant was more sensitive to antibiotics which target ribosomes but was not affected for survival in macrophages. *E. coli* tmRNA half life exceeds the doubling time of the bacterium (Hallier *et al.*, 2004) but tmRNA stability depends on the presence of sufficient SmpB (Hallier *et al*., 2004; Hanawa-Suetsugu *et al.*, 2002). It is likely that in absence of SmpB the number of SsrA molecules present in the cell are sufficient to clear stalled ribosomes occurring in normal growth conditions but they may be overwhelmed when the number of stalled ribosomes increases in conditions of stress. tmRNA can act as an antisense RNA in absence of SmpB, so the essentiality of tmRNA may be due to this function which would allow the possibility of having a second putative ribosome rescue system. SmpB function seems therefore only required during conditions that block the ribosome. Characterisation of the mutant may however be useful to dissect the tmRNA pathway in *M. tuberculosis*.

Using a reporter of activity both ClpP1 and ClpP2 were found to degrade SsrA-tagged proteins in *M. tuberculosis* (Chapter 4, 4.2.2). Even though the Clp proteases are involved in degradation of SsrA-tagged proteins in many organisms, a direct role was not previously shown in *M. tuberculosis.* Accumulation of SsrA-tagged proteins may be toxic in *M. tuberculosis* cells, as the SsrA-DD strain could not be created, thus clearance of these proteins by the Clp proteases is a required function for growth and survival. The SspB

adaptor protein promotes degradation in *E. coli* but there are no homologs in *M. tuberculosis.* In *S. pneumoniae* SsrA-tagged proteins are degraded without adaptor protein (Ahlawat and Morrison, 2009); it is not known if an adaptor protein is necessary for degradation in *M. tuberculosis.*

ClpP proteins cannot degrade proteins on their own but form complexes with Clp ATPases, which regulate substrate entry into the proteolytic compartment. ClpX is the Clp ATPase usually involved in the degradation of SsrA-tagged proteins (Ahlawat and Morrison, 2009; Gottesman *et al.*, 1998), and as ClpP1 and ClpP2 degrade these proteins, both ClpP subunits are likely to interact with ClpX. Mutation of the last three residues of the SsrA tag prevented recognition by ClpP1, demonstrating that these residues were the determinant for identification, similarly to the residues recognised by ClpXP in *E. coli* (Flynn *et al*., 2001) and further suggesting that ClpP1 interacts with ClpX. Additionally to SsrA-tagged LacZ, ClpP2 was shown to degrade untagged LacZ, and LacZ-GGG, demonstrating that ClpP2 was recognising LacZ, present in excess, independently of the tag sequence. ClpP2 is therefore involved in a general and central housekeeping function as opposed to ClpP1, which may be more specific to a particular set of substrates such as proteins harbouring the SsrA degradation tag.

The difference in substrate recognition between ClpP1 and ClpP2 may be due to their association with different ATPase subunits, suggesting that besides ClpX, ClpP2 could interact with an additional ATPase (s). There are three potential Clp ATPases in *M. tuberculosis*; but only ClpX and ClpC1 contain the 'LGL' loop, an essential motif for interaction (Kim *et al*., 2001), leaving the possibility that ClpP2 interacts with ClpC1 and ClpX, whilst ClpP1 interacts with ClpX only. *M. tuberculosis* ClpP2 interaction with ClpC1 has been previously demonstrated, and RseA was shown to be targeted by the ClpC1P2 complex, both confirming the formation of this complex in the cell and also its ability to degrade specific untagged proteins (Barik *et al*., 2009; Singh *et al*., 2006). Altogether the data were used to make a model of protein degradation by the Clp proteases in *M. tuberculosis* (Fig 6.1). It was not possible to verify this model by protein-protein interaction studies between the two ClpP and different ATPase subunits, despite using two different methods (Chapter 3, 3.2.10). The structure of the Clp complex, made of two heptameric rings binding to

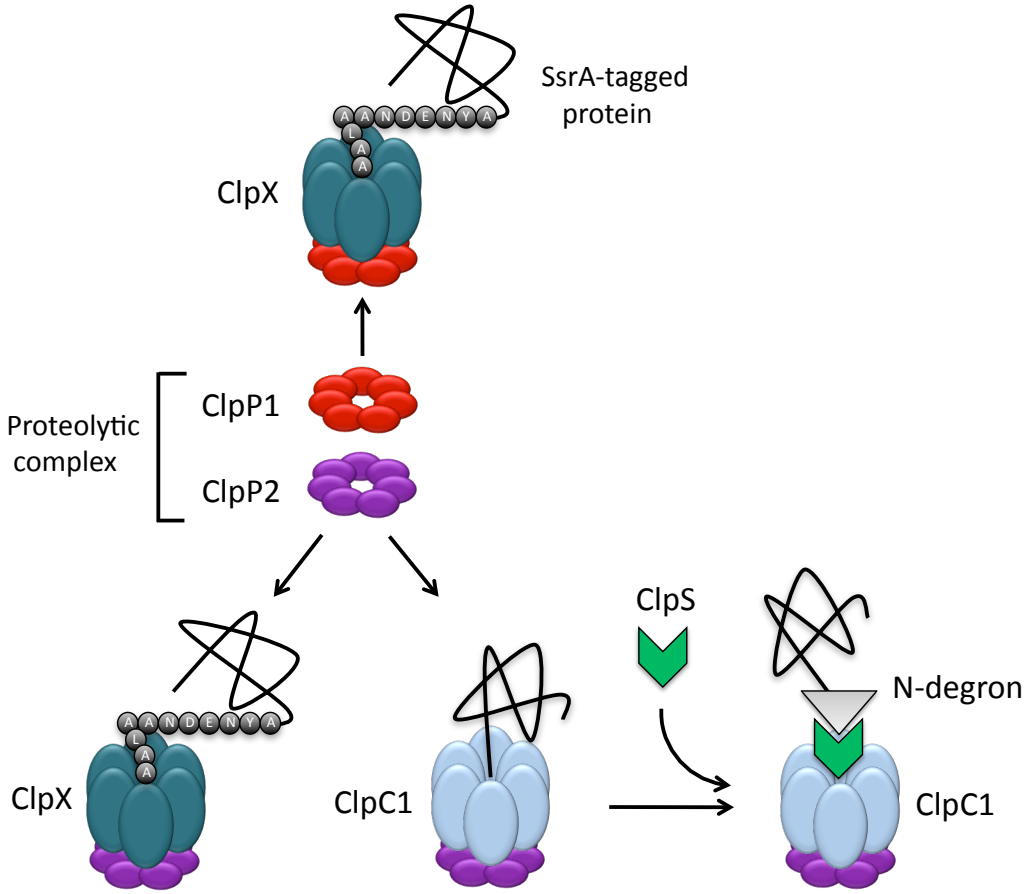

Figure 6.1 Model of ClpP1 and ClpP2 activities in *M. tuberculosis*

ClpP1 and ClpP2 form a mixed tetradecameric complex. ClpP1 and ClpP2 degrade SsrA-tagged proteins by interacting with ClpX. In addition, ClpP2 can degrade specific untagged proteins by interacting with ClpC1. ClpC1 can further recognise proteins harbouring a N-degron by interacting with ClpS. Overall ClpP1 and ClpP2 have distinct functional roles, proteolysis by ClpP1 is specific to the particular set of substrates recognised by ClpX only while ClpP2 is involved in the degradation of proteins recognised by both ClpX and ClpC1.

hexameric ATPase, may be too complex to be detected with the technologies used.

Adaptor proteins enhance substrate recognition of the Clp ATPases. In bacteria the ClpS adaptor protein binds N-end rule substrates and delivers them for degradation by the ClpAP protease complex (Erbse *et al.*, 2006). ClpS normally associates with ClpA but since ClpA is absent in Gram positive organisms ClpS may co-operate with ClpC (Dougan *et al.*, 2002). No N-end rule substrates have been identified in *M. tuberculosis* to date, and construction of a deletion mutant was not successful (Chapter 5, 5.2.3). If ClpS interacts with ClpC1 for clearance of proteins harbouring destabilising residues in their N-terminus; ClpP2 is likely to be the ClpP subunit involved in their degradation; these substrates would be specific to ClpP2 given that ClpP1 does not interact with ClpC1 (Fig 6.1).

Further work on Clp proteases could include new protein-protein interaction studies to verify the model described above. Protein affinity chromatography or pull down assays could be used, or immunoprecipitation if ClpP1 or ClpP2 specific antibodies are available. To fully elucidate ClpP1 and ClpP2 function it is essential to discover the cellular protein substrates targeted by each, preferentially under a range of different conditions such as exponential and stationary phase growth, or during hypoxia and starvation. Protein studies, such as 2D gel electrophoresis followed by mass spectrophotometry, on the ClpP1 and ClpP2 over-expressing strains under the control of a tightly inducible expression system, could be used for the identification of cellular substrates. Once identified, sequence comparison between the different substrates would allow the determination of the recognition signals used by the different Clp complexes in *M*. *tuberculosis*.

Further work on the tmRNA system could be done by replacing the native promoter for an inducible expression system, which would allow the regulation of expression in different environmental conditions and therefore determine *ssrA* requirement and function; and it would be interesting to determine if *ssrA* acts as an antisense RNA in *M. tuberculosis,* as it is the case in *Staph. aureus* (Liu *et al.*, 2010). It would be important to see if another protein can substitute SmpB activity in the binding of tmRNA to stalled ribosomes and determine if stability of tmRNA is dependent on SmpB in *M*. *tuberculosis*. In order to confirm

the role of the Clp proteases in degradation of SsrA-tagged proteins an unstable fluorescent protein could be used, and more studies are also needed to determine if an accessory protein is involved in the recognition of SsrA-tagged proteins in *M. tuberculosis*. More importantly, the identification of compounds that inhibit the tmRNA pathway would determine if this pathway could be a used as a target for the development of new TB drugs.

The identification and characterisation of a N-end rule pathway remains to be examined in mycobacteria. The creation of a ClpS deletion mutant could be carried out in *M. tuberculosis* if ClpS is not essential, as predicted by TraSH data (Sassetti *et al.*, 2003), and over-expression studies could give more insights into ClpS function. Microarray or 2D gel electrophoresis followed by mass spectrophotometry on mutant and over-expressing ClpS strains would be possible ways to identify substrates of the potential mycobacterial N-end rule pathway. Since ClpS is conserved in *M. smegmatis* and *M. marinum,* similar experiments could be carried out in the model organisms. The discovery of substrates would determine the physiological role of the N-end rule pathway in *M. tuberculosis.* The discovery of substrates and accessory components would help to establish a clear N-end rule in *M. tuberculosis* and will determine the destabilising residues. Protein-protein interaction studies between ClpS and the different Clp ATPases would determine which ClpP-ATPase complex-(es) are involved in the degradation of potential N-end substrates. Regulation of proteolysis of a specific protein can be used to determine its function and proteins sequences are commonly modified by altering their C-terminal sequences but modifying the N-terminal sequence may become a novel approach to target protein for degradation.

To conclude, the collective data show that ClpP1 and ClpP2 have distinct functional roles as they have a different proteolytic profile; ClpP1 seems to be involved in degradation of SsrA-tagged proteins only while ClpP2 seems to be involved in general proteolysis. The difference in activity between the two subunits is not due to a difference in transcription as they are co-expressed, but is likely to come from their interaction with different ATPase subunits, which in turn may be interacting with different accessory proteins. Protein tagging through tmRNA is essential for mycobacterial growth and clearance by the Clp proteins of these SsrA-tagged proteins may explain the essentiality of the Clp proteases. SmpB is not essential for growth but mutant cells exhibit phenotypic

consequences. Proteolysis is an important activity in replicating and possibly resting cells so Clp protease activators or inhibitors may become new actors in the fight against tuberculosis.

7 Bibliography

Abo, T., Ueda, K., Sunohara, T., Ogawa, K., and Aiba, H. (2002). SsrA-mediated protein tagging in the presence of miscoding drugs and its physiological role in *Escherichia coli*. Genes Cells *7*, 629-638.

Ahlawat, S., and Morrison, D.A. (2009). ClpXP degrades SsrA-tagged proteins in *Streptococcus pneumoniae*. J Bacteriol *191*, 2894-2898.

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996). An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. Proc Natl Acad Sci USA *93*, 6059-6063.

Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N., and Mekalanos, J.J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proc Natl Acad Sci U S A *99*, 966-971.

Akopian, T., Kandror, O., Raju, R.M., Unnikrishnan, M., Rubin, E.J., and Goldberg, A.L. (2012). The active ClpP protease from *M. tuberculosis* is a complex composed of a heptameric ClpP1 and a ClpP2 ring. EMBO J.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002a). How cells read the genome: from DNA to protein. In *Molecular biology of the cell*. 4th edn. Garland Science, New York. Chapter 6.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002b). Postranscriptional controls. In *Molecular biology of the cell*. 4th edn. Garland Science, New York. Chapter 7.

Alvarez, E., and Tavel, E. (1885). Recherches sur le bacille Lustgarten. Archives de Physologie Normale et Pathologique *6*, 303-321.

Aly, S., Wagner, K., Keller, C., Malm, S., Malzan, A., Brandau, S., Bange, F.C., and Ehlers, S. (2006). Oxygen status of lung granulomas in *Mycobacterium tuberculosis*-infected mice. J Pathol *210*, 298-305.

Andersen, J.B., Sternberg, C., Poulsen, L.K., Bjorn, S.P., Givskov, M., and Molin, S. (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. Appl Environ Microbiol *64*, 2240-2246.

Andersson, F.I., Tryggvesson, A., Sharon, M., Diemand, A.V., Classen, M., Best, C., Schmidt, R., Schelin, J., Stanne, T.M., Bukau, B., Robinson, C.V., Witt, S., Mogk, A., and Clarke, A.K. (2009). Structure and function of a novel type of ATP-dependent Clp protease. J Biol Chem *284*, 13519-13532.

Andini, N., and Nash, K.A. (2011). Expression of tmRNA in mycobacteria is increased by antimicrobial agents that target the ribosome. FEMS Microbiol Lett *322*, 172-179.

Antas, P.R., and Castello-Branco, L.R. (2008). New vaccines against tuberculosis: lessons learned from BCG immunisation in Brazil. Trans R Soc Trop Med Hyg *102*, 628-630.

Ayele, W.Y., Neill, S.D., Zinsstag, J., Weiss, M.G., and Pavlik, I. (2004). Bovine tuberculosis: an old disease but a new threat to Africa. Int J Tuberc Lung Dis *8*, 924-937.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). *In vivo* half-life of a protein is a function of its amino-terminal residue. Science *234*, 179-186.

Baker, T.A., and Sauer, R.T. (2006). ATP-dependent proteases of bacteria: recognition logic and operating principles. Trends Biochem Sci *31*, 647-653.

Barends, S., Karzai, A.W., Sauer, R.T., Wower, J., and Kraal, B. (2001). Simultaneous and functional binding of SmpB and EF-Tu-TP to the alanyl acceptor arm of tmRNA. J Mol Biol *314*, 9-21.

Barik, S., Sureka, K., Mukherjee, P., Basu, J., and Kundu, M. (2009). RseA, the SigE specific anti-sigma factor of *Mycobacterium tuberculosis*, is inactivated by phosphorylation-dependent ClpC1P2 proteolysis. Mol Microbiol *75*, 592-606.

Bashyam, M.D., Kaushal, D., Dasgupta, S.K., and Tyagi, A.K. (1996). A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. J Bacteriol *178*, 4847-4853.

Baumler, A.J., Kusters, J.G., Stojiljkovic, I., and Heffron, F. (1994). *Salmonella typhimurium* loci involved in survival within macrophages. Infect Immun *62*, 1623-1630.

Bellier, A., and Mazodier, P. (2004). ClgR, a novel regulator of *clp* and *lon* expression in *Streptomyces*. J Bacteriol *186*, 3238-3248.

Bellier, A., Gominet, M., and Mazodier, P. (2006). Post-translational control of the *Streptomyces lividans* ClgR regulon by ClpP. Microbiology *152*, 1021-1027.

Benaroudj, N., Raynal, B., Miot, M., and Ortiz-Lombardia, M. (2011). Assembly and proteolytic processing of mycobacterial ClpP1 and ClpP2. BMC Biochem *12*, 61.

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

Beuron, F., Maurizi, M.R., Belnap, D.M., Kocsis, E., Booy, F.P., Kessel, M., and Steven, A.C. (1998). At sixes and sevens: characterization of the symmetry mismatch of the ClpAP chaperone-assisted protease. J Struct Biol *123*, 248-259.

Bewley, M.C., Graziano, V., Griffin, K., and Flanagan, J.M. (2006). The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. J Struct Biol *153*, 113-128.

Blokpoel, M.C., O'Toole, R., Smeulders, M.J., and Williams, H.D. (2003). Development and application of unstable GFP variants to kinetic studies of mycobacterial gene expression. J Microbiol Methods *54*, 203-211.

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

Bouley, D.M., Ghori, N., Mercer, K.L., Falkow, S., and Ramakrishnan, L. (2001). Dynamic nature of host-pathogen interactions in *Mycobacterium marinum* granulomas. Infect Immun *69*, 7820-7831.

Brennan, P.J., and Nikaido, H. (1995). The envelope of mycobacteria. Annu Rev Biochem *64*, 29-63.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., and Cole, S.T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A *99*, 3684-3689.

Brotz-Oesterhelt, H., Beyer, D., Kroll, H.P., Endermann, R., Ladel, C., Schroeder, W., Hinzen, B., Raddatz, S., Paulsen, H., Henninger, K., Bandow, J.E., Sahl, H.G., and Labischinski, H. (2005). Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. Nat Med *11*, 1082-1087.

Brown, A.C., and Parish, T. (2006). Instability of the acetamide-inducible expression vector pJAM2 in *Mycobacterium tuberculosis*. Plasmid *55*, 81-86.

Brown, A.C., Eberl, M., Crick, D.C., Jomaa, H., and Parish, T. (2010). The nonmevalonate pathway of isoprenoid biosynthesis in *Mycobacterium tuberculosis* is essential and transcriptionally regulated by Dxs. J Bacteriol *192*, 2424-2433.

Brown, C.A., Draper, P., and Hart, P.D. (1969). Mycobacteria and lysosomes: a paradox. Nature *221*, 658-660.

Carroll, P., Muttucumaru, D.G., and Parish, T. (2005). Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. Appl Environ Microbiol *71*, 3077-3084.

Carroll, P., Faray-Kele, M.C., and Parish, T. (2011). Identifying vulnerable pathways in *Mycobacterium tuberculosis* by using a knockdown approach. Appl Environ Microbiol *77*, 5040-5043.

Cascioferro, A., Boldrin, F., Serafini, A., Provvedi, R., Palu, G., and Manganelli, R. (2010). Xer site-specific recombination, an efficient tool to introduce unmarked deletions into mycobacteria. Appl Environ Microbiol *76*, 5312-5316.

Chadani, Y., Ono, K., Ozawa, S., Takahashi, Y., Takai, K., Nanamiya, H., Tozawa, Y., Kutsukake, K., and Abo, T. (2010). Ribosome rescue by *Escherichia coli* ArfA (YhdL) in the absence of trans-translation system. Mol Microbiol *78*, 796-808.

Chadani, Y., Matsumoto, E., Aso, H., Wada, T., Kutsukake, K., Sutou, S., and Abo, T. (2011a). trans-translation-mediated tight regulation of the expression of the alternative ribosome-rescue factor ArfA in *Escherichia coli*. Genes Genet Syst *86*, 151-163.

Chadani, Y., Ono, K., Kutsukake, K., and Abo, T. (2011b). *Escherichia coli* YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. Mol Microbiol *80*, 772-785.

Chang, Y., Wesenberg, G.E., Bingman, C.A., and Fox, B.G. (2008). *In vivo* inactivation of the mycobacterial integral membrane stearoyl coenzyme A desaturase DesA3 by a C-terminus-specific degradation process. J Bacteriol *190*, 6686-6696.

Chastanet, A., and Msadek, T. (2003). ClpP of *Streptococcus salivarius* is a novel member of the dually regulated class of stress response genes in gram-positive bacteria. J Bacteriol *185*, 683-687.

Choi, K.H., and Licht, S. (2005). Control of peptide product sizes by the energy-dependent protease ClpAP. Biochemistry *44*, 13921-13931.

Choy, J.S., Aung, L.L., and Karzai, A.W. (2007). Lon protease degrades transfer-messenger RNA-tagged proteins. J Bacteriol *189*, 6564-6571.

Clark-Curtiss, J.E., Jacobs, W.R., Docherty, M.A., Ritchie, L.R., and Curtiss, R., 3rd (1985). Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. J Bacteriol *161*, 1093-1102.

Clemens, D.L., and Horwitz, M.A. (1995). Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited. J Exp Med *181*, 257-270.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537-544.

Comstock, G.W. (1999). How much isoniazid is needed for prevention of tuberculosis among immunocompetent adults? Int J Tuberc Lung Dis *3*, 847-850.

Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem *65*, 801-847.

Cunningham, A.F., and Spreadbury, C.L. (1998). Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. J Bacteriol *180*, 801-808.

Dannenberg, A.M., Jr. (1993). Immunopathogenesis of pulmonary tuberculosis. Hosp Pract (Off Ed) *28*, 51-58.

Darwin, K.H., Ehrt, S., Gutierrez-Ramos, J.C., Weich, N., and Nathan, C.F. (2003). The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. Science *302*, 1963-1966.

Darwin, K.H., Lin, G., Chen, Z., Li, H., and Nathan, C.F. (2005). Characterization of a *Mycobacterium tuberculosis* proteasomal ATPase homologue. Mol Microbiol *55*, 561-571.

Dasgupta, N., Kapur, V., Singh, K.K., Das, T.K., Sachdeva, S., Jyothisri, K., and Tyagi, J.S. (2000). Characterization of a two-component system, devR-devS, of *Mycobacterium tuberculosis*. Tuber Lung Dis *80*, 141-159.

de Crecy-Lagard, V., Servant-Moisson, P., Viala, J., Grandvalet, C., and Mazodier, P. (1999). Alteration of the synthesis of the Clp ATP-dependent protease affects morphological and physiological differentiation in *Streptomyces*. Mol Microbiol *32*, 505-517.

Derre, I., Rapoport, G., and Msadek, T. (1999). CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. Mol Microbiol *31*, 117-131.

Dharmadhikari, A.S., and Nardell, E.A. (2008). What animal models teach humans about tuberculosis. Am J Respir Cell Mol Biol *39*, 503-508.

Diel, R., Nienhaus, A., and Schaberg, T. (2005). Cost-effectiveness of isoniazid chemoprevention in close contacts. Eur Respir J *26*, 465-473.

Dolin, P.J., Raviglione, M.C., and Kochi, A. (1994). Global tuberculosis incidence and mortality during 1990-2000. Bull World Health Organ *72*, 213-220.

Donegan, N.P., Thompson, E.T., Fu, Z., and Cheung, A.L. (2009). Proteolytic regulation of toxin-antitoxin systems by ClpPC in *Staphylococcus aureus*. J Bacteriol *192*, 1416-1422.

Donoghue, H.D. (2009). Human tuberculosis--an ancient disease, as elucidated by ancient microbial biomolecules. Microbes Infect *11*, 1156-1162.

Dougan, D.A., Mogk, A., and Bukau, B. (2002). Protein folding and degradation in bacteria: to degrade or not to degrade? That is the question. Cell Mol Life Sci *59*, 1607-1616.

Dougan, D.A., Truscott, K.N., and Zeth, K. (2010). The bacterial N-end rule pathway: expect the unexpected. Mol Microbiol *76*, 545-558.

Draper, P. (1967). The aliphatic acylamide amidohydrolase of *Mycobacterium smegmatis*: its inducible nature and relation to acyl-transfer to hydroxylamine. J Gen Microbiol *46*, 111-123.

Dussurget, O., Timm, J., Gomez, M., Gold, B., Yu, S., Sabol, S.Z., Holmes, R.K., Jacobs, W.R., Jr., and Smith, I. (1999). Transcriptional control of the iron-responsive fxbA gene by the mycobacterial regulator IdeR. J Bacteriol *181*, 3402-3408.

Dutta, N.K., Mehra, S., Didier, P.J., Roy, C.J., Doyle, L.A., Alvarez, X., Ratterree, M., Be, N.A., Lamichhane, G., Jain, S.K., Lacey, M.R., Lackner, A.A., and Kaushal, D. (2010). Genetic requirements for the survival of tubercle bacilli in primates. J Infect Dis *201*, 1743-1752.

Dziedzic, R., Kiran, M., Plocinski, P., Ziolkiewicz, M., Brzostek, A., Moomey, M., Vadrevu, I.S., Dziadek, J., Madiraju, M., and Rajagopalan, M. (2010). *Mycobacterium tuberculosis* ClpX interacts with FtsZ and interferes with FtsZ assembly. PLoS One *5*, e11058.

Ehlers, S. (2009). Lazy, dynamic or minimally recrudescent? On the elusive nature and location of the mycobacterium responsible for latent tuberculosis. Infection *37*, 87-95.

Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W., and Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res *33*, e21.

Engelberg-Kulka, H., and Glaser, G. (1999). Addiction modules and programmed cell death and antideath in bacterial cultures. Annu Rev Microbiol *53*, 43-70.

Engels, S., Schweitzer, J.E., Ludwig, C., Bott, M., and Schaffer, S. (2004). *clpC* and *clpP1P2* gene expression in *Corynebacterium glutamicum* is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor sigmaH. Mol Microbiol *52*, 285-302.

Engels, S., Ludwig, C., Schweitzer, J.E., Mack, C., Bott, M., and Schaffer, S. (2005). The transcriptional activator ClgR controls transcription of genes involved in proteolysis and DNA repair in *Corynebacterium glutamicum*. Mol Microbiol *57*, 576-591.

Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D.A., and Bukau, B. (2006). ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. Nature *439*, 753-756.

Estorninho, M., Smith, H., Thole, J., Harders-Westerveen, J., Kierzek, A., Butler, R.E., Neyrolles, O., and Stewart, G.R. (2010). ClgR regulation of chaperone and protease systems is essential for *Mycobacterium tuberculosis* parasitism of the macrophage. Microbiology *156*, 3445-3455.

Farrell, C.M., Grossman, A.D., and Sauer, R.T. (2005). Cytoplasmic degradation of ssrA-tagged proteins. Mol Microbiol *57*, 1750-1761.

Fedhila, S., Msadek, T., Nel, P., and Lereclus, D. (2002). Distinct *clpP* genes control specific adaptive responses in *Bacillus thuringiensis*. J Bacteriol *184*, 5554-5562.

Felden, B., and Gillet, R. (2011). SmpB as the handyman of tmRNA during trans-translation. RNA Biol *8*, 440-449.

Fernandez, L., Breidenstein, E.B., Song, D., and Hancock, R.E. (2011). Role of Intracellular Proteases in the Antibiotic Resistance, Motility, and Biofilm Formation of *Pseudomonas aeruginosa*. Antimicrob Agents Chemother *56*, 1128-1132.

Festa, R.A., McAllister, F., Pearce, M.J., Mintseris, J., Burns, K.E., Gygi, S.P., and Darwin, K.H. (2010). Prokaryotic ubiquitin-like protein (Pup) proteome of Mycobacterium tuberculosis [corrected]. PLoS One *5*, e8589.

Fine, P.E. (1995). Variation in protection by BCG: implications of and for heterologous immunity. Lancet *346*, 1339-1345.

Flanagan, J.M., Wall, J.S., Capel, M.S., Schneider, D.K., and Shanklin, J. (1995). Scanning transmission electron microscopy and small-angle scattering provide evidence that native *Escherichia coli* ClpP is a tetradecamer with an axial pore. Biochemistry *34*, 10910-10917.

Flynn, J.M., Levchenko, I., Seidel, M., Wickner, S.H., Sauer, R.T., and Baker, T.A. (2001). Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. Proc Natl Acad Sci U S A *98*, 10584-10589.

Flynn, J.M., Neher, S.B., Kim, Y.I., Sauer, R.T., and Baker, T.A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. Mol Cell *11*, 671-683.

Frees, D., and Ingmer, H. (1999). ClpP participates in the degradation of misfolded protein in *Lactococcus lactis*. Mol Microbiol *31*, 79-87.

Frees, D., Qazi, S.N., Hill, P.J., and Ingmer, H. (2003). Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. Mol Microbiol *48*, 1565-1578.

Frees, D., Chastanet, A., Qazi, S., Sorensen, K., Hill, P., Msadek, T., and Ingmer, H. (2004). Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in *Staphylococcus aureus*. Mol Microbiol *54*, 1445-1462.

Frees, D., Savijoki, K., Varmanen, P., and Ingmer, H. (2007). Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. Mol Microbiol *63*, 1285-1295.

Gaillot, O., Pellegrini, E., Bregenholt, S., Nair, S., and Berche, P. (2000). The ClpP serine protease is essential for the intracellular parasitism and virulence of *Listeria monocytogenes*. Mol Microbiol *35*, 1286-1294.

Gandhi, N.R., Moll, A., Sturm, A.W., Pawinski, R., Govender, T., Lalloo, U., Zeller, K., Andrews, J., and Friedland, G. (2006). Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet *368*, 1575-1580.

Gandotra, S., Schnappinger, D., Monteleone, M., Hillen, W., and Ehrt, S. (2007). *In vivo* gene silencing identifies the *Mycobacterium tuberculosis* proteasome as essential for the bacteria to persist in mice. Nat Med *13*, 1515-1520.

Garbe, T., Harris, D., Vordermeier, M., Lathigra, R., Ivanyi, J., and Young, D. (1993). Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. Infect Immun *61*, 260-267.

Gaudin, C., Zhou, X., Williams, K.P., and Felden, B. (2002). Two-piece tmRNA in cyanobacteria and its structural analysis. Nucleic Acids Res *30*, 2018-2024.

Geiger, S.R., Bottcher, T., Sieber, S.A., and Cramer, P. (2011). A conformational switch underlies ClpP protease function. Angew Chem Int Ed Engl *50*, 5749-5752.

Gomes, M.S., Paul, S., Moreira, A.L., Appelberg, R., Rabinovitch, M., and Kaplan, G. (1999). Survival of *Mycobacterium avium* and *Mycobacterium tuberculosis* in acidified vacuoles of murine macrophages. Infect Immun *67*, 3199-3206.

Gomez, M., Doukhan, L., Nair, G., and Smith, I. (1998). *sigA* is an essential gene in *Mycobacterium smegmatis*. Mol Microbiol *29*, 617-628.

Gomez, M., and Smith, I. (2000). Determinants of mycobacterial gene expression. In Hattfull G, Jacobs WR, editors. *Molecular genetics of mycobacteria.* ASM Press, Washington DC. Chapter 7.

Gominet, M., Seghezzi, N., and Mazodier, P. (2011). Acyl depsipeptide (ADEP) resistance in *Streptomyces*. Microbiology *157*, 2226-2234.

Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Mattick, J.S., Dalrymple, B., Kuramitsu, H., Shiroza, T., and Foster, T. (1990). Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. Proc Natl Acad Sci U S A *87*, 3513-3517.

Gottesman, S., Clark, W.P., de Crecy-Lagard, V., and Maurizi, M.R. (1993). ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. Sequence and *in vivo* activities. J Biol Chem *268*, 22618-22626.

Gottesman, S. (1996). Proteases and their targets in *Escherichia coli*. Annu Rev Genet *30*, 465-506.

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R.T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev *12*, 1338-1347.

Grange, J.M. (1992). The mystery of the mycobacterial 'persistor'. Tuber Lung Dis *73*, 249-251.

Gribun, A., Kimber, M.S., Ching, R., Sprangers, R., Fiebig, K.M., and Houry, W.A. (2005). The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. J Biol Chem *280*, 16185-16196.

Grimaud, R., Kessel, M., Beuron, F., Steven, A.C., and Maurizi, M.R. (1998). Enzymatic and structural similarities between the *Escherichia coli* ATP-dependent proteases, ClpXP and ClpAP. J Biol Chem *273*, 12476-12481.

Guo, F., Esser, L., Singh, S.K., Maurizi, M.R., and Xia, D. (2002). Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. J Biol Chem *277*, 46753-46762.

Guo, X.V., Monteleone, M., Klotzsche, M., Kamionka, A., Hillen, W., Braunstein, M., Ehrt, S., and Schnappinger, D. (2007). Silencing *Mycobacterium smegmatis* by using tetracycline repressors. J Bacteriol *189*, 4614-4623.

Gupta, U.D., and Katoch, V.M. (2005). Animal models of tuberculosis. Tuberculosis (Edinb) *85*, 277-293.

Hallier, M., Ivanova, N., Rametti, A., Pavlov, M., Ehrenberg, M., and Felden, B. (2004). Pre-binding of small protein B to a stalled ribosome triggers trans-translation. J Biol Chem *279*, 25978-25985.

Hallier, M., Desreac, J., and Felden, B. (2006). Small protein B interacts with the large and the small subunits of a stalled ribosome during trans-translation. Nucleic Acids Res *34*, 1935-1943.

Hanawa-Suetsugu, K., Takagi, M., Inokuchi, H., Himeno, H., and Muto, A. (2002). SmpB functions in various steps of trans-translation. Nucleic Acids Res *30*, 1620-1629.

Hanson, P.I., and Whiteheart, S.W. (2005). AAA+ proteins: have engine, will work. Nat Rev Mol Cell Biol *6*, 519-529.

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science *295*, 1852-1858.

Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science *301*, 1496-1499.

Heifets, L. (2002). Susceptibility testing of *Mycobacterium tuberculosis* to pyrazinamide. J Med Microbiol *51*, 11-12.

Herman, C., Thevenet, D., Bouloc, P., Walker, G.C., and D'Ari, R. (1998). Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). Genes Dev *12*, 1348-1355.

Hett, E.C., and Rubin, E.J. (2008). Bacterial growth and cell division: a mycobacterial perspective. Microbiol Mol Biol Rev *72*, 126-156.

Himeno, H., Nameki, N., Tadaki, T., Sato, M., Hanawa, K., Fukushima, M., Ishii, M., Ushida, C., and Muto, A. (1997). *Escherichia coli* tmRNA (10Sa RNA) in trans-translation. Nucleic Acids Symp Ser, 185-186.

Hinnerwisch, J., Fenton, W.A., Furtak, K.J., Farr, G.W., and Horwich, A.L. (2005). Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. Cell *121*, 1029-1041.

Hong, S.J., Tran, Q.A., and Keiler, K.C. (2005). Cell cycle-regulated degradation of tmRNA is controlled by RNase R and SmpB. Mol Microbiol *57*, 565-575.

Hoskins, J.R., Pak, M., Maurizi, M.R., and Wickner, S. (1998). The role of the ClpA chaperone in proteolysis by ClpAP. Proc Natl Acad Sci U S A *95*, 12135-12140.

HPA (2010). Tuberculosis in the UK: Annual report on tuberculosis surveillance in the UK 2010 (London, Health Protection Agency).

Hu, G., Lin, G., Wang, M., Dick, L., Xu, R.M., Nathan, C., and Li, H. (2006). Structure of the *Mycobacterium tuberculosis* proteasome and mechanism of inhibition by a peptidyl boronate. Mol Microbiol *59*, 1417-1428.

Hu, Y., Butcher, P.D., Sole, K., Mitchison, D.A., and Coates, A.R. (1998). Protein synthesis is shutdown in dormant *Mycobacterium tuberculosis* and is reversed by oxygen or heat shock. FEMS Microbiol Lett *158*, 139-145.

Hu, Y., Mangan, J.A., Dhillon, J., Sole, K.M., Mitchison, D.A., Butcher, P.D., and Coates, A.R. (2000). Detection of mRNA transcripts and active transcription in persistent *Mycobacterium tuberculosis* induced by exposure to rifampin or pyrazinamide. J Bacteriol *182*, 6358-6365.

Hu, Y., and Coates, A.R.M. (2003). Tuberculosis. In Dormancy and Low-Growth States in Microbial Disease (New York, Cambridge University Press).

Huang, C., Wolfgang, M.C., Withey, J., Koomey, M., and Friedman, D.I. (2000). Charged tmRNA but not tmRNA-mediated proteolysis is essential for *Neisseria gonorrhoeae* viability. EMBO J *19*, 1098-1107.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., and Venter, J.C. (1999). Global transposon mutagenesis and a minimal Mycoplasma genome. Science *286*, 2165-2169.

Ingvarsson, H., Mate, M.J., Hogbom, M., Portnoi, D., Benaroudj, N., Alzari, P.M., Ortiz-Lombardia, M., and Unge, T. (2007). Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of *Mycobacterium tuberculosis* ClpP1. Acta Crystallogr D Biol Crystallogr *63*, 249-259.

Inobe, T., and Matouschek, A. (2008). Protein targeting to ATP-dependent proteases. Curr Opin Struct Biol *18*, 43-51.

Jain, S.K., Hernandez-Abanto, S.M., Cheng, Q.J., Singh, P., Ly, L.H., Klinkenberg, L.G., Morrison, N.E., Converse, P.J., Nuermberger, E., Grosset, J., McMurray, D.N., Karakousis, P.C., Lamichhane, G., and Bishai, W.R. (2007). Accelerated detection of *Mycobacterium tuberculosis* genes essential for bacterial survival in guinea pigs, compared with mice. J Infect Dis *195*, 1634-1642.

Jenal, U., and Hengge-Aronis, R. (2003). Regulation by proteolysis in bacterial cells. Curr Opin Microbiol *6*, 163-172.

Jennings, L.D., Bohon, J., Chance, M.R., and Licht, S. (2008). The ClpP N-terminus coordinates substrate access with protease active site reactivity. Biochemistry *47*, 11031-11040.

Julio, S.M., Heithoff, D.M., and Mahan, M.J. (2000). ssrA (tmRNA) plays a role in *Salmonella enterica* serovar Typhimurium pathogenesis. J Bacteriol *182*, 1558-1563.

Kallenius, G., Koivula, T., Ghebremichael, S., Hoffner, S.E., Norberg, R., Svensson, E., Dias, F., Marklund, B.I., and Svenson, S.B. (1999). Evolution and

clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau. J Clin Microbiol *37*, 3872-3878.

Kalpana, G.V., Bloom, B.R., and Jacobs, W.R., Jr. (1991). Insertional mutagenesis and illegitimate recombination in mycobacteria. Proc Natl Acad Sci U S A *88*, 5433-5437.

Kang, C.M., Abbott, D.W., Park, S.T., Dascher, C.C., Cantley, L.C., and Husson, R.N. (2005). The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. Genes Dev *19*, 1692-1704.

Kapopoulou, A., Lew, J.M., and Cole, S.T. (2011). The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis (Edinb) *91*, 8-13.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A *95*, 5752-5756.

Karimova, G., Dautin, N., and Ladant, D. (2005). Interaction network among *Escherichia coli* membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. J Bacteriol *187*, 2233-2243.

Karzai, A.W., Susskind, M.M., and Sauer, R.T. (1999). SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). EMBO J *18*, 3793-3799.

Karzai, A.W., Roche, E.D., and Sauer, R.T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. Nat Struct Biol *7*, 449-455.

Kaufmann, S.H. (2000). Robert Koch's highs and lows in the search for a remedy for tuberculosis. Nature Medicine Special Web Focus: Tuberculosis

Kaufmann, S.H. (2010). Future vaccination strategies against tuberculosis: thinking outside the box. Immunity *33*, 567-577.

Keiler, K.C., and Sauer, R.T. (1996). Sequence determinants of C-terminal substrate recognition by the Tsp protease. J Biol Chem *271*, 2589-2593.

Keiler, K.C., Waller, P.R., and Sauer, R.T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science *271*, 990-993.

Keiler, K.C., Shapiro, L., and Williams, K.P. (2000). tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in *Caulobacter*. Proc Natl Acad Sci U S A *97*, 7778-7783.

Keiler, K.C., and Shapiro, L. (2003). tmRNA is required for correct timing of DNA replication in *Caulobacter crescentus*. J Bacteriol *185*, 573-580.

Keiler, K.C. (2008). Biology of trans-translation. Annu Rev Microbiol *62*, 133-151.

Kessel, M., Maurizi, M.R., Kim, B., Kocsis, E., Trus, B.L., Singh, S.K., and Steven, A.C. (1995). Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. J Mol Biol *250*, 587-594.

Kim, D.Y., and Kim, K.K. (2003). Crystal structure of ClpX molecular chaperone from *Helicobacter pylori*. J Biol Chem *278*, 50664-50670.

Kim, K.I., Cheong, G.W., Park, S.C., Ha, J.S., Woo, K.M., Choi, S.J., and Chung, C.H. (2000a). Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in Escherichia coli. J Mol Biol *303*, 655-666.

Kim, Y.I., Burton, R.E., Burton, B.M., Sauer, R.T., and Baker, T.A. (2000b). Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. Mol Cell *5*, 639-648.

Kim, Y.I., Levchenko, I., Fraczkowska, K., Woodruff, R.V., Sauer, R.T., and Baker, T.A. (2001). Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. Nat Struct Biol *8*, 230-233.

Kimber, M.S., Yu, A.Y., Borg, M., Leung, E., Chan, H.S., and Houry, W.A. (2010). Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations. Structure *18*, 798-808.

Kirstein, J., Schlothauer, T., Dougan, D.A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B., and Turgay, K. (2006). Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. EMBO J *25*, 1481-1491.

Kirstein, J., Hoffmann, A., Lilie, H., Schmidt, R., Rubsamen-Waigmann, H., Brotz-Oesterhelt, H., Mogk, A., and Turgay, K. (2009). The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. EMBO Mol Med *1*, 37-49.

Klotzsche, M., Ehrt, S., and Schnappinger, D. (2009). Improved tetracycline repressors for gene silencing in mycobacteria. Nucleic Acids Res *37*, 1778-1788.

Knipfer, N., and Shrader, T.E. (1997). Inactivation of the 20S proteasome in *Mycobacterium smegmatis*. Mol Microbiol *25*, 375-383.

Kock, H., Gerth, U., and Hecker, M. (2004). The ClpP peptidase is the major determinant of bulk protein turnover in *Bacillus subtilis*. J Bacteriol *186*, 5856-5864.

Kojetin, D.J., McLaughlin, P.D., Thompson, R.J., Dubnau, D., Prepiak, P., Rance, M., and Cavanagh, J. (2009). Structural and motional contributions of the *Bacillus subtilis* ClpC N-domain to adaptor protein interactions. J Mol Biol *387*, 639-652.

Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K., and Inokuchi, H. (1994). A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. Proc Natl Acad Sci U S A *91*, 9223-9227.

Kress, W., Maglica, Z., and Weber-Ban, E. (2009). Clp chaperone-proteases: structure and function. Res Microbiol *160*, 618-628.

Kroh, H.E., and Simon, L.D. (1990). The ClpP component of Clp protease is the sigma 32-dependent heat shock protein F21.5. J Bacteriol *172*, 6026-6034.

Kruger, E., Witt, E., Ohlmeier, S., Hanschke, R., and Hecker, M. (2000). The clp proteases of *Bacillus subtilis* are directly involved in degradation of misfolded proteins. J Bacteriol *182*, 3259-3265.

Kruger, E., Zuhlke, D., Witt, E., Ludwig, H., and Hecker, M. (2001). Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. EMBO J *20*, 852-863.

Kumar, A., Toledo, J.C., Patel, R.P., Lancaster, J.R., Jr., and Steyn, A.J. (2007). *Mycobacterium tuberculosis* DosS is a redox sensor and DosT is a hypoxia sensor. Proc Natl Acad Sci U S A *104*, 11568-11573.

Kwiatkowska, J., Matuszewska, E., Kuczynska-Wisnik, D., and Laskowska, E. (2008). Aggregation of *Escherichia coli* proteins during stationary phase depends on glucose and oxygen availability. Res Microbiol *159*, 651-657.

Kwon, H.Y., Kim, S.W., Choi, M.H., Ogunniyi, A.D., Paton, J.C., Park, S.H., Pyo, S.N., and Rhee, D.K. (2003). Effect of heat shock and mutations in ClpL and ClpP on virulence gene expression in *Streptococcus pneumoniae*. Infect Immun *71*, 3757-3765.

Lamichhane, G., Zignol, M., Blades, N.J., Geiman, D.E., Dougherty, A., Grosset, J., Broman, K.W., and Bishai, W.R. (2003). A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A *100*, 7213-7218.

Lamichhane, G., Tyagi, S., and Bishai, W.R. (2005). Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. Infect Immun *73*, 2533-2540.

Lamichhane, G., Raghunand, T.R., Morrison, N.E., Woolwine, S.C., Tyagi, S., Kandavelou, K., and Bishai, W.R. (2006). Deletion of a *Mycobacterium tuberculosis* proteasomal ATPase homologue gene produces a slow-growing strain that persists in host tissues. J Infect Dis *194*, 1233-1240.

Laskowska, E., Kuczynska-Wisnik, D., Skorko-Glonek, J., and Taylor, A. (1996). Degradation by proteases Lon, Clp and HtrA, of *Escherichia coli* proteins aggregated *in vivo* by heat shock; HtrA protease action *in vivo* and *in vitro*. Mol Microbiol *22*, 555-571.

Lee, B.G., Park, E.Y., Lee, K.E., Jeon, H., Sung, K.H., Paulsen, H., Rubsamen-Schaeff, H., Brotz-Oesterhelt, H., and Song, H.K. (2010). Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. Nat Struct Mol Biol *17*, 471-478.

Lee, M.H., Pascopella, L., Jacobs, W.R., Jr., and Hatfull, G.F. (1991). Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guerin. Proc Natl Acad Sci U S A *88*, 3111-3115.

Lehnherr, H., and Yarmolinsky, M.B. (1995). Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli*. Proc Natl Acad Sci U S A *92*, 3274-3277.

Lemos, J.A., and Burne, R.A. (2002). Regulation and Physiological Significance of ClpC and ClpP in *Streptococcus mutans*. J Bacteriol *184*, 6357-6366.

Levchenko, I., Luo, L., and Baker, T.A. (1995). Disassembly of the Mu transposase tetramer by the ClpX chaperone. Genes Dev *9*, 2399-2408.

Levchenko, I., Seidel, M., Sauer, R.T., and Baker, T.A. (2000). A specificity-enhancing factor for the ClpXP degradation machine. Science *289*, 2354-2356.

Li, D.H., Chung, Y.S., Gloyd, M., Joseph, E., Ghirlando, R., Wright, G.D., Cheng, Y.Q., Maurizi, M.R., Guarne, A., and Ortega, J. (2010). Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: A model for the ClpX/ClpA-bound state of ClpP. Chem Biol *17*, 959-969.

Li, Z., Zhao, X., Zhou, C., Gu, B., and Frankel, F.R. (2006). A truncated *Bacillus subtilis dal* gene with a 3' *ssrA* gene tag regulates the growth and virulence of racemase-deficient *Listeria monocytogenes*. Microbiology *152*, 3091-3102.

Licht, S., and Lee, I. (2008). Resolving individual steps in the operation of ATP-dependent proteolytic molecular machines: from conformational changes to substrate translocation and processivity. Biochemistry *47*, 3595-3605.

Lies, M., and Maurizi, M.R. (2008). Turnover of endogenous SsrA-tagged proteins mediated by ATP-dependent proteases in *Escherichia coli*. J Biol Chem *283*, 22918-22929.

Liu, J., Tran, V., Leung, A.S., Alexander, D.C., and Zhu, B. (2009). BCG vaccines: their mechanisms of attenuation and impact on safety and protective efficacy. Hum Vaccin *5*, 70-78.

Mahenthiralingam, E., Draper, P., Davis, E.O., and Colston, M.J. (1993). Cloning and sequencing of the gene which encodes the highly inducible acetamidase of *Mycobacterium smegmatis*. J Gen Microbiol *139*, 575-583.

Mahenthiralingam, E., Marklund, B.I., Brooks, L.A., Smith, D.A., Bancroft, G.J., and Stokes, R.W. (1998). Site-directed mutagenesis of the 19-kilodalton lipoprotein antigen reveals No essential role for the protein in the growth and virulence of *Mycobacterium intracellulare*. Infect Immun *66*, 3626-3634.

Manabe, Y.C., Chen, J.M., Ko, C.G., Chen, P., and Bishai, W.R. (1999). Conditional sigma factor expression, using the inducible acetamidase promoter, reveals that the *Mycobacterium tuberculosis* sigF gene modulates expression of the 16-kilodalton alpha-crystallin homologue. J Bacteriol *181*, 7629-7633.

Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F.R., and Smith, I. (1999). Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. Mol Microbiol *31*, 715-724.

Marks, G.B., Bai, J., Simpson, S.E., Sullivan, E.A., and Stewart, G.J. (2000). Incidence of tuberculosis among a cohort of tuberculin-positive refugees in

Australia: reappraising the estimates of risk. Am J Respir Crit Care Med *162*, 1851-1854.

Martin, A., Baker, T.A., and Sauer, R.T. (2007). Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. Mol Cell *27*, 41-52.

Martin, A., Baker, T.A., and Sauer, R.T. (2008). Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. Nat Struct Mol Biol *15*, 139-145.

Maurizi, M.R. (1992). Proteases and protein degradation in *Escherichia coli*. Experientia *48*, 178-201.

Mc Dermott, W., and Tompsett, R. (1954). Activation of pyrazinamide and nicotinamide in acidic environments in vitro. Am Rev Tuberc *70*, 748-754.

McMurray, D.N. (1996). Mycobacteria and Nocardia. In: Baron S., editor. *Medical Microbiology* 4th ed. Galveston, Texas. Chapter 33.

Mehra, S., and Kaushal, D. (2009). Functional genomics reveals extended roles of the *Mycobacterium tuberculosis* stress response factor sigmaH. J Bacteriol *191*, 3965-3980.

Michel, A., Agerer, F., Hauck, C.R., Herrmann, M., Ullrich, J., Hacker, J., and Ohlsen, K. (2006). Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. J Bacteriol *188*, 5783-5796.

Migliori, G.B., De Iaco, G., Besozzi, G., Centis, R., and Cirillo, D.M. (2007). First tuberculosis cases in Italy resistant to all tested drugs. Euro Surveill *12*, E070517 070511.

Miller, J.H. (1972). Experiments in Molecular Genetics. (Cold Spring Harbor, Cold Spring Harbor Press).

Miltgen, J., Morillon, M., Koeck, J.L., Varnerot, A., Briant, J.F., Nguyen, G., Verrot, D., Bonnet, D., and Vincent, V. (2002). Two cases of pulmonary tuberculosis caused by *Mycobacterium tuberculosis* subsp canetti. Emerg Infect Dis *8*, 1350-1352.

Moore, S.D., McGinness, K.E., and Sauer, R.T. (2003). Structural biology. A glimpse into tmRNA-mediated ribosome rescue. Science *300*, 72-73.

Moore, S.D., and Sauer, R.T. (2005). Ribosome rescue: tmRNA tagging activity and capacity in *Escherichia coli*. Mol Microbiol *58*, 456-466.

Motohashi, K., Watanabe, Y., Yohda, M., and Yoshida, M. (1999). Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones. Proc Natl Acad Sci U S A *96*, 7184-7189.

Msadek, T., Dartois, V., Kunst, F., Herbaud, M.L., Denizot, F., and Rapoport, G. (1998). ClpP of *Bacillus subtilis* is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation. Mol Microbiol *27*, 899-914.

Muto, A., Fujihara, A., Ito, K.I., Matsuno, J., Ushida, C., and Himeno, H. (2000). Requirement of transfer-messenger RNA for the growth of *Bacillus subtilis* under stresses. Genes Cells *5*, 627-635.

Muttucumaru, D.G., Roberts, G., Hinds, J., Stabler, R.A., and Parish, T. (2004). Gene expression profile of *Mycobacterium tuberculosis* in a non-replicating state. Tuberculosis (Edinb) *84*, 239-246.

Nakano, S., Zheng, G., Nakano, M.M., and Zuber, P. (2002). Multiple pathways of Spx (YjbD) proteolysis in *Bacillus subtilis*. J Bacteriol *184*, 3664-3670.

Narayanan, S., Selvakumar, S., Aarati, R., Vasan, S.K., and Narayanan, P.R. (2000). Transcriptional analysis of inducible acetamidase gene of *Mycobacterium smegmatis*. FEMS Microbiol Lett *192*, 263-268.

Nonin-Lecomte, S., Germain-Amiot, N., Gillet, R., Hallier, M., Ponchon, L., Dardel, F., and Felden, B. (2009). Ribosome hijacking: a role for small protein B during trans-translation. EMBO Rep *10*, 160-165.

O'Hare, H., Juillerat, A., Dianiskova, P., and Johnsson, K. (2008). A split-protein sensor for studying protein-protein interaction in mycobacteria. J Microbiol Methods *73*, 79-84.

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

Ocepek, M., Pate, M., Zolnir-Dovc, M., and Poljak, M. (2005). Transmission of *Mycobacterium tuberculosis* from human to cattle. J Clin Microbiol *43*, 3555-3557.

Ogura, T., and Wilkinson, A.J. (2001). AAA+ superfamily ATPases: common structure--diverse function. Genes Cells *6*, 575-597.

Oh, B.K., and Apirion, D. (1991). 10Sa RNA, a small stable RNA of *Escherichia coli*, is functional. Mol Gen Genet *229*, 52-56.

Ohno, H., Zhu, G., Mohan, V.P., Chu, D., Kohno, S., Jacobs, W.R., Jr., and Chan, J. (2003). The effects of reactive nitrogen intermediates on gene expression in *Mycobacterium tuberculosis*. Cell Microbiol *5*, 637-648.

Okan, N.A., Bliska, J.B., and Karzai, A.W. (2006). A Role for the SmpB-SsrA system in *Yersinia pseudotuberculosis* pathogenesis. PLoS Pathog *2*, e6.

Ollinger, J., O'Malley, T., Kesicki, E.A., Odingo, J., and Parish, T. (2011). Validation of the essential ClpP protease in *Mycobacterium tuberculosis* as a novel drug target. J Bacteriol *194*, 663-668.

Pan, Q., Garsin, D.A., and Losick, R. (2001). Self-reinforcing activation of a cell-specific transcription factor by proteolysis of an anti-sigma factor in *B. subtilis*. Mol Cell *8*, 873-883.

Parida, S.K., and Kaufmann, S.H. (2010). Novel tuberculosis vaccines on the horizon. Curr Opin Immunol *22*, 374-384.

Parish, T., and Stoker, N.G. (1997). Development and use of a conditional antisense mutagenesis system in mycobacteria. FEMS Microbiol Lett *154*, 151-157.

Parish, T., Mahenthiralingam, E., Draper, P., Davis, E.O., and Colston, M.J. (1997). Regulation of the inducible acetamidase gene of *Mycobacterium smegmatis*. Microbiology *143 (Pt 7)*, 2267-2276.

Parish, T., Smith, D.A., Roberts, G., Betts, J., and Stoker, N.G. (2003). The senX3-regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology *149*, 1423-1435.

Park, H.D., Guinn, K.M., Harrell, M.I., Liao, R., Voskuil, M.I., Tompa, M., Schoolnik, G.K., and Sherman, D.R. (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. Mol Microbiol *48*, 833-843.

Park, S.T., Kang, C.M., and Husson, R.N. (2008). Regulation of the SigH stress response regulon by an essential protein kinase in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A *105*, 13105-13110.

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

Pearce, M.J., Arora, P., Festa, R.A., Butler-Wu, S.M., Gokhale, R.S., and Darwin, K.H. (2006). Identification of substrates of the *Mycobacterium tuberculosis* proteasome. EMBO J *25*, 5423-5432.

Pederson, K.J., Carlson, S., and Pierson, D.E. (1997). The ClpP protein, a subunit of the Clp protease, modulates ail gene expression in *Yersinia enterocolitica*. Mol Microbiol *26*, 99-107.

Pham, T.T., Jacobs-Sera, D., Pedulla, M.L., Hendrix, R.W., and Hatfull, G.F. (2007). Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology *153*, 2711-2723.

Pickart, C.M., and Cohen, R.E. (2004). Proteasomes and their kin: proteases in the machine age. Nat Rev Mol Cell Biol *5*, 177-187.

Piehler, J. (2005). New methodologies for measuring protein interactions *in vivo* and *in vitro*. Curr Opin Struct Biol *15*, 4-14.

Porankiewicz, J., Schelin, J., and Clarke, A.K. (1998). The ATP-dependent Clp protease is essential for acclimation to UV-B and low temperature in the cyanobacterium *Synechococcus*. Mol Microbiol *29*, 275-283.

Porankiewicz, J., Wang, J., and Clarke, A.K. (1999). New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. Mol Microbiol *32*, 449-458.

Primm, T.P., Lucero, C.A., and Falkinham, J.O., 3rd (2004). Health impacts of environmental mycobacteria. Clin Microbiol Rev *17*, 98-106.

Ranquet, C., and Gottesman, S. (2007). Translational regulation of the *Escherichia coli* stress factor RpoS: a role for SsrA and Lon. J Bacteriol *189*, 4872-4879.

Rastogi, N., Blom-Potar, M.C., and David, H.L. (1989). Comparative intracellular growth of difficult-to-grow and other mycobacteria in a macrophage cell line. Acta Leprol *7 Suppl 1*, 156-159.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. Nucleic Acids Res *38*, D227-233.

Reeves, A., Gerth, U., Volker, U., and Haldenwang, W.G. (2007). ClpP modulates the activity of the *Bacillus subtilis* stress response transcription factor, sigmaB. J Bacteriol *189*, 6168-6175.

Rengarajan, J., Bloom, B.R., and Rubin, E.J. (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. Proc Natl Acad Sci U S A *102*, 8327-8332.

Reyrat, J.M., and Kahn, D. (2001). *Mycobacterium smegmatis*: an absurd model for tuberculosis? Trends Microbiol *9*, 472-474.

Roberts, G., Muttucumaru, D.G., and Parish, T. (2003). Control of the acetamidase gene of *Mycobacterium smegmatis* by multiple regulators. FEMS Microbiol Lett *221*, 131-136.

Robertson, G.T., Ng, W.L., Foley, J., Gilmour, R., and Winkler, M.E. (2002). Global transcriptional analysis of clpP mutations of type 2 *Streptococcus pneumoniae* and their effects on physiology and virulence. J Bacteriol *184*, 3508-3520.

Roche, E.D., and Sauer, R.T. (1999). SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. EMBO J *18*, 4579-4589.

Roman-Hernandez, G., Hou, J.Y., Grant, R.A., Sauer, R.T., and Baker, T.A. (2011). The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. Mol Cell *43*, 217-228.

Russell, D.G. (2001). *Mycobacterium tuberculosis*: here today, and here tomorrow. Nat Rev Mol Cell Biol *2*, 569-577.

Russo, S., Schweitzer, J.E., Polen, T., Bott, M., and Pohl, E. (2009). Crystal structure of the caseinolytic protease gene regulator, a transcriptional activator in actinomycetes. J Biol Chem *284*, 5208-5216.

Rustad, T.R., Harrell, M.I., Liao, R., and Sherman, D.R. (2008). The enduring hypoxic response of *Mycobacterium tuberculosis*. PLoS One *3*, e1502.

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

Rustad, T.R., Roberts, D.M., Liao, R.P., and Sherman, D.R. (2009b). Isolation of mycobacterial RNA. In: Parish T, Brown AC, editors. *Mycobacteria Protocols*, 2nd ed. Humana Press, Totowa. p. 13-21

Sacco, E., Covarrubias, A.S., O'Hare, H.M., Carroll, P., Eynard, N., Jones, T.A., Parish, T., Daffe, M., Backbro, K., and Quemard, A. (2007). The missing piece of the type II fatty acid synthase system from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A *104*, 14628-14633.

Sass, P., Josten, M., Famulla, K., Schiffer, G., Sahl, H.G., Hamoen, L., and Brotz-Oesterhelt, H. (2011). Antibiotic acyldepsipeptides activate ClpP peptidase to degrade the cell division protein FtsZ. Proc Natl Acad Sci U S A *108*, 17474-17479.

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

Sassetti, C.M., and Rubin, E.J. (2003). Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A *100*, 12989-12994.

Saunders, B.M., and Cooper, A.M. (2000). Restraining mycobacteria: role of granulomas in mycobacterial infections. Immunol Cell Biol *78*, 334-341.

Savijoki, K., Ingmer, H., Frees, D., Vogensen, F.K., Palva, A., and Varmanen, P. (2003). Heat and DNA damage induction of the LexA-like regulator HdiR from *Lactococcus lactis* is mediated by RecA and ClpP. Mol Microbiol *50*, 609-621.

Schaible, U.E., Sturgill-Koszycki, S., Schlesinger, P.H., and Russell, D.G. (1998). Cytokine activation leads to acidification and increases maturation of *Mycobacterium avium*-containing phagosomes in murine macrophages. J Immunol *160*, 1290-1296.

Schelin, J., Lindmark, F., and Clarke, A.K. (2002). The *clpP* multigene family for the ATP-dependent Clp protease in the cyanobacterium *Synechococcus*. Microbiology *148*, 2255-2265.

Schirmer, E.C., Glover, J.R., Singer, M.A., and Lindquist, S. (1996). HSP100/Clp proteins: a common mechanism explains diverse functions. Trends Biochem Sci *21*, 289-296.

Schmidt, R., Zahn, R., Bukau, B., and Mogk, A. (2009). ClpS is the recognition component for *Escherichia coli* substrates of the N-end rule degradation pathway. Mol Microbiol *72*, 506-517.

Schmitt, E.K., Riwanto, M., Sambandamurthy, V., Roggo, S., Miault, C., Zwingelstein, C., Krastel, P., Noble, C., Beer, D., Rao, S.P., Au, M., Niyomrattanakit, P., Lim, V., Zheng, J., Jeffery, D., Pethe, K., and Camacho, L.R. (2011). The natural product cyclomarin kills *Mycobacterium tuberculosis* by targeting the ClpC1 subunit of the caseinolytic protease. Angew Chem Int Ed Engl *50*, 5889-5891.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., and Schoolnik, G.K. (2003). Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. J Exp Med *198*, 693-704.

Schweder, T., Lee, K.H., Lomovskaya, O., and Matin, A. (1996). Regulation of *Escherichia coli* starvation sigma factor (sigma s) by ClpXP protease. J Bacteriol *178*, 470-476.

Sherman, D.R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M.I., and Schoolnik, G.K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. Proc Natl Acad Sci U S A *98*, 7534-7539.

Sherrid, A.M., Rustad, T.R., Cangelosi, G.A., and Sherman, D.R. (2010). Characterization of a Clp protease gene regulator and the reaeration response in *Mycobacterium tuberculosis*. PLoS One *5*, e11622.

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

Shiloh, M.U., and DiGiuseppe Champion, P.A. (2009). To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? Curr Opin Microbiol *13*, 86-92.

Shimizu, Y., and Ueda, T. (2002). The role of SmpB protein in trans-translation. FEBS Lett *514*, 74-77.

Shleeva, M.O., Bagramyan, K., Telkov, M.V., Mukamolova, G.V., Young, M., Kell, D.B., and Kaprelyants, A.S. (2002). Formation and resuscitation of "non-culturable" cells of *Rhodococcus rhodochrous* and *Mycobacterium tuberculosis* in prolonged stationary phase. Microbiology *148*, 1581-1591.

Showalter, H.D., and Denny, W.A. (2008). A roadmap for drug discovery and its translation to small molecule agents in clinical development for tuberculosis treatment. Tuberculosis (Edinb) *88 Suppl 1*, S3-17.

Singh, A., Mai, D., Kumar, A., and Steyn, A.J. (2006). Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. Proc Natl Acad Sci U S A *103*, 11346-11351.

Singh, S.K., and Maurizi, M.R. (1994). Mutational analysis demonstrates different functional roles for the two ATP-binding sites in ClpAP protease from *Escherichia coli*. J Biol Chem *269*, 29537-29545.

Smith, C.K., Baker, T.A., and Sauer, R.T. (1999). Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. Proc Natl Acad Sci U S A *96*, 6678-6682.

Smith, I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clin Microbiol Rev *16*, 463-496.

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

Snider, J., Thibault, G., and Houry, W.A. (2008). The AAA+ superfamily of functionally diverse proteins. Genome Biol *9*, 216.

Socha, A.M., Tan, N.Y., LaPlante, K.L., and Sello, J.K. (2010). Diversity-oriented synthesis of cyclic acyldepsipeptides leads to the discovery of a potent antibacterial agent. Bioorg Med Chem *18*, 7193-7202.

Spiers, A., Lamb, H.K., Cocklin, S., Wheeler, K.A., Budworth, J., Dodds, A.L., Pallen, M.J., Maskell, D.J., Charles, I.G., and Hawkins, A.R. (2002). PDZ domains facilitate binding of high temperature requirement protease A (HtrA) and tail-specific protease (Tsp) to heterologous substrates through recognition of the small stable RNA A (ssrA)-encoded peptide. J Biol Chem *277*, 39443-39449.

Sprangers, R., Gribun, A., Hwang, P.M., Houry, W.A., and Kay, L.E. (2005). Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. Proc Natl Acad Sci U S A *102*, 16678-16683.

Stahl, D.A., and Urbance, J.W. (1990). The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. J Bacteriol *172*, 116-124.

Stanne, T.M., Pojidaeva, E., Andersson, F.I., and Clarke, A.K. (2007). Distinctive types of ATP-dependent Clp proteases in cyanobacteria. J Biol Chem *282*, 14394-14402.

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

Stewart, G.R., Robertson, B.D., and Young, D.B. (2004). Analysis of the function of mycobacterial DnaJ proteins by overexpression and microarray profiling. Tuberculosis (Edinb) *84*, 180-187.

Stinear, T.P., Seemann, T., Harrison, P.F., Jenkin, G.A., Davies, J.K., Johnson, P.D., Abdellah, Z., Arrowsmith, C., Chillingworth, T., Churcher, C., Clarke, K., Cronin, A., Davis, P., Goodhead, I., Holroyd, N., Jagels, K., Lord, A., Moule, S., Mungall, K., Norbertczak, H., Quail, M.A., Rabbinowitsch, E., Walker, D., White, B., Whitehead, S., Small, P.L., Brosch, R., Ramakrishnan, L., Fischbach, M.A., Parkhill, J., and Cole, S.T. (2008). Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. Genome Res *18*, 729-741.

Stottmeier, K.D., Beam, R.E., and Kubica, G.P. (1967). Determination of drug susceptibility of mycobacteria to pyrazinamide in 7H10 agar. Am Rev Respir Dis *96*, 1072-1075.

Striebel, F., Hunkeler, M., Summer, H., and Weber-Ban, E. (2010). The mycobacterial Mpa-proteasome unfolds and degrades pupylated substrates by engaging Pup's N-terminus. EMBO J *29*, 1262-1271.

Strosser, J., Ludke, A., Schaffer, S., Kramer, R., and Burkovski, A. (2004). Regulation of GlnK activity: modification, membrane sequestration and proteolysis as regulatory principles in the network of nitrogen control in *Corynebacterium glutamicum*. Mol Microbiol *54*, 132-147.

Stryer, L., Berg, J.M., and Tymoczko, J.L. (2002). Biochemistry, 5th edn (New York).

Sun, Z., and Zhang, Y. (1999). Spent culture supernatant of *Mycobacterium tuberculosis* H37Ra improves viability of aged cultures of this strain and allows small inocula to initiate growth. J Bacteriol *181*, 7626-7628.

Swaminathan, S., Ramachandran, R., Baskaran, G., Paramasivan, C.N., Ramanathan, U., Venkatesan, P., Prabhakar, R., and Datta, M. (2000). Risk of development of tuberculosis in HIV-infected patients. Int J Tuberc Lung Dis *4*, 839-844.

Taylor, G.M., Young, D.B., and Mays, S.A. (2005). Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. J Clin Microbiol *43*, 2236-2240.

Thangaraj, H.S., Lamb, F.I., Davis, E.O., Jenner, P.J., Jeyakumar, L.H., and Colston, M.J. (1990). Identification, sequencing, and expression of *Mycobacterium leprae* superoxide dismutase, a major antigen. Infect Immun *58*, 1937-1942.

Thibonnier, M., Thiberge, J.M., and De Reuse, H. (2008). Trans-translation in *Helicobacter pylori*: essentiality of ribosome rescue and requirement of protein tagging for stress resistance and competence. PLoS One *3*, e3810.

Thompson, M.W., Singh, S.K., and Maurizi, M.R. (1994). Processive degradation of proteins by the ATP-dependent Clp protease from *Escherichia coli*. Requirement for the multiple array of active sites in ClpP but not ATP hydrolysis. J Biol Chem *269*, 18209-18215.

Thomsen, L.E., Olsen, J.E., Foster, J.W., and Ingmer, H. (2002). ClpP is involved in the stress response and degradation of misfolded proteins in *Salmonella enterica* serovar Typhimurium. Microbiology *148*, 2727-2733.

Tobias, J.W., Shrader, T.E., Rocap, G., and Varshavsky, A. (1991). The N-end rule in bacteria. Science *254*, 1374-1377.

Tortoli, E. (2006). The new mycobacteria: an update. FEMS Immunol Med Microbiol *48*, 159-178.

Tripathi, L.P., and Sowdhamini, R. (2008). Genome-wide survey of prokaryotic serine proteases: analysis of distribution and domain architectures of five serine protease families in prokaryotes. BMC Genomics *9*, 549.

Tyagi, J.S., and Kinger, A.K. (1992). Identification of the 10Sa RNA structural gene of *Mycobacterium tuberculosis*. Nucleic Acids Res *20*, 138.

van Kessel, J.C., and Hatfull, G.F. (2007). Recombineering in *Mycobacterium tuberculosis*. Nat Methods *4*, 147-152.

Varmanen, P., Ingmer, H., and Vogensen, F.K. (2000). ctsR of *Lactococcus lactis* encodes a negative regulator of clp gene expression. Microbiology *146 (Pt 6)*, 1447-1455.

Varshavsky, A. (1991). Naming a targeting signal. Cell *64*, 13-15.

Varshavsky, A. (1996). The N-end rule: functions, mysteries, uses. Proc Natl Acad Sci U S A *93*, 12142-12149.

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

Via, L.E., Lin, P.L., Ray, S.M., Carrillo, J., Allen, S.S., Eum, S.Y., Taylor, K., Klein, E., Manjunatha, U., Gonzales, J., Lee, E.G., Park, S.K., Raleigh, J.A., Cho, S.N., McMurray, D.N., Flynn, J.L., and Barry, C.E., 3rd (2008). Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. Infect Immun *76*, 2333-2340.

Viala, J., Rapoport, G., and Mazodier, P. (2000). The clpP multigenic family in *Streptomyces lividans*: conditional expression of the clpP3 clpP4 operon is controlled by PopR, a novel transcriptional activator. Mol Microbiol *38*, 602-612.

Viala, J., and Mazodier, P. (2003). The ATPase ClpX is conditionally involved in the morphological differentiation of *Streptomyces lividans*. Mol Genet Genomics *268*, 563-569.

Volker, U., Engelmann, S., Maul, B., Riethdorf, S., Volker, A., Schmid, R., Mach, H., and Hecker, M. (1994). Analysis of the induction of general stress proteins of *Bacillus subtilis*. Microbiology *140 (Pt 4)*, 741-752.

Voskuil, M.I., Visconti, K.C., and Schoolnik, G.K. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis (Edinb) *84*, 218-227.

Wang, C., Li, M., Dong, D., Wang, J., Ren, J., Otto, M., and Gao, Q. (2007). Role of ClpP in biofilm formation and virulence of *Staphylococcus epidermidis*. Microbes Infect *9*, 1376-1383.

Wang, J., Hartling, J.A., and Flanagan, J.M. (1997). The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. Cell *91*, 447-456.

Wang, L., Elliott, M., and Elliott, T. (1999). Conditional stability of the HemA protein (glutamyl-tRNA reductase) regulates heme biosynthesis in *Salmonella typhimurium*. J Bacteriol *181*, 1211-1219.

Wang, T., Li, H., Lin, G., Tang, C., Li, D., Nathan, C., and Darwin, K.H. (2009). Structural insights on the *Mycobacterium tuberculosis* proteasomal ATPase Mpa. Structure *17*, 1377-1385.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M. and Losick, R. (2004). Gene regulation in prokaryotes. In *Molecular biology of the gene,* 5th edition. Pearson Benjamin Cunnings, San Francisco. Chapter 16.

Wawrzynow, A., Banecki, B., and Zylicz, M. (1996). The Clp ATPases define a novel class of molecular chaperones. Mol Microbiol *21*, 895-899.

Wayne, L.G. (1977). Synchronised replication of *Mycobacterium tuberculosis*. Infect Immun *17*, 528-530.

Wayne, L.G., and Lin, K.Y. (1982). Glyoxylate metabolism and adaptation of *Mycobacterium tuberculosis* to survival under anaerobic conditions. Infect Immun *37*, 1042-1049.

Wayne, L.G. (1994). Dormancy of *Mycobacterium tuberculosis* and latency of disease. Eur J Clin Microbiol Infect Dis *13*, 908-914.

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

Wayne, L.G., and Sohaskey, C.D. (2001). Nonreplicating persistence of *Mycobacterium tuberculosis*. Annu Rev Microbiol *55*, 139-163.

Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., Zentgraf, H., Weber-Ban, E.U., Dougan, D.A., Tsai, F.T., Mogk, A., and Bukau, B. (2004). Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. Cell *119*, 653-665.

Weichart, D., Querfurth, N., Dreger, M., and Hengge-Aronis, R. (2003). Global role for ClpP-containing proteases in stationary-phase adaptation of *Escherichia coli*. J Bacteriol *185*, 115-125.

Wells, A.Q., and Oxon, D.M. (1937). Tuberculosis in wild voles. Lancet *1*, 1221.

WHO (2011). Global tuberculosis control. Available at: http://www.who.int/tb/publications/global_report/en/index.html (Accessed 12th December 2011).

Wickner, S., and Maurizi, M.R. (1999). Here's the hook: similar substrate binding sites in the chaperone domains of Clp and Lon. Proc Natl Acad Sci U S A *96*, 8318-8320.

Wiegert, T., and Schumann, W. (2001). SsrA-mediated tagging in *Bacillus subtilis*. J Bacteriol *183*, 3885-3889.

Williams, K.P., and Bartel, D.P. (1996). Phylogenetic analysis of tmRNA secondary structure. RNA *2*, 1306-1310.

Williams, K.P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res *30*, 866-875.

Withey, J., and Friedman, D. (1999). Analysis of the role of trans-translation in the requirement of tmRNA for lambdaimmP22 growth in *Escherichia coli*. J Bacteriol *181*, 2148-2157.

Wladyka, B., and Pustelny, K. (2008). Regulation of bacterial protease activity. Cell Mol Biol Lett *13*, 212-229.

Woo, K.M., Chung, W.J., Ha, D.B., Goldberg, A.L., and Chung, C.H. (1989). Protease Ti from *Escherichia coli* requires ATP hydrolysis for protein breakdown but not for hydrolysis of small peptides. J Biol Chem *264*, 2088-2091.

Wower, J., Kirillov, S.V., Wower, I.K., Guven, S., Hixson, S.S., and Zimmermann, R.A. (2000). Transit of tRNA through the *Escherichia coli* ribosome. Cross-linking of the 3' end of tRNA to specific nucleotides of the 23 S ribosomal RNA at the A, P, and E sites. J Biol Chem *275*, 37887-37894.

Yang, C., and Glover, J.R. (2009). The SmpB-tmRNA tagging system plays important roles in *Streptomyces coelicolor* growth and development. PLoS One *4*, e4459.

Yu, A.Y., and Houry, W.A. (2007). ClpP: a distinctive family of cylindrical energy-dependent serine proteases. FEBS Lett *581*, 3749-3757.

Zeiler, E., Braun, N., Bottcher, T., Kastenmuller, A., Weinkauf, S., and Sieber, S.A. (2011). Vibralactone as a Tool to Study the Activity and Structure of the ClpP1P2 Complex from *Listeria monocytogenes*. Angew Chem Int Ed Engl *50*, 11001-11004.

Zeth, K., Ravelli, R.B., Paal, K., Cusack, S., Bukau, B., and Dougan, D.A. (2002). Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. Nat Struct Biol *9*, 906-911.

Zhang, X., Shaw, A., Bates, P.A., Newman, R.H., Gowen, B., Orlova, E., Gorman, M.A., Kondo, H., Dokurno, P., Lally, J., Leonard, G., Meyer, H., van Heel, M., and Freemont, P.S. (2000). Structure of the AAA ATPase p97. Mol Cell *6*, 1473-1484.

Zhang, Y., Lathigra, R., Garbe, T., Catty, D., and Young, D. (1991). Genetic analysis of superoxide dismutase, the 23 kilodalton antigen of *Mycobacterium tuberculosis*. Mol Microbiol *5*, 381-391.

Zhang, Y., Yang, Y., Woods, A., Cotter, R.J., and Sun, Z. (2001). Resuscitation of dormant *Mycobacterium tuberculosis* by phospholipids or specific peptides. Biochem Biophys Res Commun *284*, 542-547.

Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H., and Nerlich, A.G. (2003). Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol *41*, 359-367.

Zolkiewski, M. (1999). ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from *Escherichia coli*. J Biol Chem *274*, 28083-28086.

8 Appendices

Appendix 1. Primers used for RT-PCR in this study

Primer name	Sequence (5' to 3')	Purpose
sigAF	CAGGACACTACGACCAGCAC	Amplification of 5' end of *sigA* (250 bp)
sigAR	CTCACGTTCGAGGTCTTCGT	
P1F	AGCCAAGTGACTGACATG	Amplification of 5' end of *clpP1* (258 bp)
P1R	GTCACAGGGCGCCAGCAC	
P1P2F	GATCAAGAAAGAAATGTT	Amplification of *clpP1/clpP2* junction (264 bp)
P1P2R	GTATGGATTGGACTCCTT	
P2F	AAGGAGTCCAATCCATAC	Amplification of 3' end of *clpP2* (257 bp)
P2R	GCAGCCGCCGAGGCGGCC	

Appendix 2. Primers used for identification and characterisation of the *clpP1P2* promoter

Name	Sequence (5' to 3')	Purpose
PclpF1	GAGTACTTTCAGGGGGTGCGTGAC	Amplification of 92 bp upstream of *clpP1P2*
PclpF2	AGTACTTGACCGTATGACGCTGTA	Amplification of 125 bp upstream of *clpP1P2*
PclpF3	AGTACTCAGGGCCGCAGTGGAGGC	Amplification of 278 bp upstream of *clpP1P2*
PclpR	AGTACTGCTCACAGTGGGGCACCT	Amplification of upstream regions of *clpP1P2*
SDM10AF	GCCAGCCTGGTTGGTCAGTGGCGGAGCATAGAAG	SDM on putative -10 element from TAGTGT to CAGTGG
SDM10AR	CTTCTATGCTCCGCCACTGACCAACCAGGCTGGC	SDM on putative -10 element from TAGTGT to CAGTGG
SDM10BF	GTTAGTGTCGGAGCACGGAAGAACTCGAGG	SDM on putative -10 element from TAGAAG to CGGAAG
SDM10BR	CCTCGAGTTCTTCCGTGCTCCGACACTAAC	SDM on putative -10 element from TAGAAG to CGGAAG
SDM35AF	GGTTTCAGGGGGTGCCCGACCGCGAAACAGCCAGCC	SDM on putative -35 element from GTGACC to CCGACC
SDM35AR	GGCTGGCTGTTTCGCGGTCGGGCACCCCCTGAAACC	SDM on putative -35 element from GTGACC to CCGACC
SDM35BF	GGGGTGCGTGACCGCGAGGCAGCCAGCCTGGTTGG	SDM on putative -35 element from GCGAAA to GCGAGG
SDM35BR	CCAACCAGGCTGGCTGCCTCGCGGTCACGCACCCC	SDM on putative -35 element from GCGAAA to GCGAGG
PclpP2F	GCAGTACTGTACCAAGGGCAAGCGCTAC	Amplification of 280 bp upstream of *clpP1P2*
Pclp2R	GCAGTACTATTCACTGTGCTTCTCCATTG	Amplification of 280 bp upstream of *clpP1P2*
RegF	CAACGTGACCGTATGAAAATGTAAGCGAACGCGCC	SDM of putative regulatory region from CGC to AAA in clpP1P2 promoter
RegR	GGCGCGTTCGCTTACATTTTCATACGGTCACGTTG	
Reg21F	CGCGCCGGTTTCAGGATATGCGTGACCGCGAAAC	SDM of putative regulatory region from GGG to AAA in clpP1P2 promoter
Reg21R	GTTTCGCGGTCACGCATATCCTGAAACCGGCGCG	

Appendix 3. Primers used for SDM in *clpP1P2* promoter regulatory region

Name	Sequence	Mutation
Reg1F	GACAACGTGACCGTGTGACGCTGTAAGC	A_1G
Reg1R	GTTACAGCGTCACACGGTCACGTTGTC	
Reg2F	CAACGTGACCGTAGGACGCTGTAAGCG	T_2G
Reg2R	CGCTTACAGCGTCCTACGGTCACGTTG	
Reg3F	GACAACGTGACCGTATAACGCTGTAAGCGAAC	G_3A
Reg3R	GTTCGCTTACAGCGTTATACGGTCACGTTGTC	
Reg4F	CAACGTGACCGTATGGCGCTGTAAGCGAACG	A_4G
Reg4R	CGTTCGCTTACAGCGCCATACGGTCACGTTG	
Reg5F	CGTGACCGTATGAAGCTGTAAGCGAACG	C_5A
Reg5R	CGTTCGCTTACAGCTTCATACGGTCACG	
Reg6F	CGTGACCGTATGACACTGTAAGCGAACGC	G_6A
Reg6R	GCGTTCGCTTACAGTGTCATACGGTCACG	
Reg7F	GTGACCGTATGACGATGTAAGCGAACGCG	C_7A
Reg7R	CGCGTTCGCTTACATCGTCATACGGTCAC	
Reg8F	GTGACCGTATGACGCGGTAAGCGAACGCGCC	T_8G
Reg8R	GGCGCGTTCGCTTACCGCGTCATACGGTCAC	
Reg9F	GACCGTATGACGCTATAAGCGAACGCGCC	G_9A
Reg9R	GGCGCGTTCGCTTATAGCGTCATACGGTC	
Reg10F	GACCGTATGACGCTGGAAGCGAACGCGCCGG	$T_{10}G$
Reg10R	CCGGCGCGTTCGCTTCCAGCGTCATACGGTC	
Reg11F	CCGTATGACGCTGTGAGCGAACGCGCCGG	$A_{11}G$
Reg11R	CCGGCGCGTTCGCTCACAGCGTCATACGG	
Reg12F	GTATGACGCTGTAGGCGAACGCGCCGG	$A_{12}G$
Reg12R	CCGGCGCGTTCGCCTACAGCGTCATAC	
Reg13F	CCGTATGACGCTGTAAACGAACGCGCCGGTTTC	$G_{13}A$
Reg13R	GAAACCGGCGCGTTCGTTTACAGCGTCATACGG	
Reg14F	GTATGACGCTGTAAGAGAACGCGCCGGTTTC	$C_{14}A$
Reg14R	GAAACCGGCGCGTTCTCTTACAGCGTCATAC	
Reg15F	GTATGACGCTGTAAGCAAACGCGCCGGTTTCAG	$G_{15}A$
Reg15R	CTGAAACCGGCGCGTTTGCTTACAGCGTCATAC	
Reg16F	GACGCTGTAAGCGGACGCGCCGGTTTCAG	$A_{16}G$
Reg16R	CTGAAACCGGCGCGTCCGCTTACAGCGTC	
Reg17F	GACGCTGTAAGCGAGCGCGCCGGTTTCAG	$A_{17}G$
Reg17R	CTGAAACCGGCGCGCTCGCTTACAGCGTC	
Reg18F	GACGCTGTAAGCGAAAGCGCCGGTTTCAGGG	$C_{18}A$
Reg18R	CCCTGAAACCGGCGCTTTCGCTTACAGCGTC	
Reg19F	GCTGTAAGCGAACACGCCGGTTTCAGG	$G_{19}A$
Reg19R	CCTGAAACCGGCGTGTTCGCTTACAGC	
Reg20F	CTGTAAGCGAACGAGCCGGTTTCAGGG	$C_{20}A$
Reg20R	CCCTGAAACCGGCTCGTTCGCTTACAG	

Appendix 4. Plasmids and vectors used for identification and characterisation of the *clpP1P2* promoter

Plasmid name	Description	Source
pSC-A	PCR cloning vector, Amp^R	Stratagene
pSC-B	PCR cloning vector, Amp^R	Stratagene
pSM128	LacZ promoter probe vector, integrating, Sm^R	Dussurget *et al*., 1999
pPclp1	125 bp upstream of *clpP1P2* in pSM128	This study
pPclp3	pPclp1 with putative -10 element mutated from TAGTGT to **C**AGTG**G**	This study
pPclp4	pPclp1 with putative -10 element mutated from TAGAAG to **CG**GAAG	This study
pPclp5	pPclp1 with putative -35 element mutated from GTGACC to **CC**GACC	This study
pPclp6	278 bp upstream of *clpP1P2* in pSM128	This study
pPclp7	pPclp6 with CGC motif mutated to AAA	This study
pPclp8	pPclp1 with putative -35 element mutated from GCGAAA to GCGA**GG**	This study
pPclp32	pPclp6 with SDM A_1G	This study
pPclp31	pPclp6 with SDM T_2G	This study
pPclp9	pPclp6 with SDM G_3A	This study
pPclp10	pPclp6 with SDM A_4G	This study
pPclp11	pPclp6 with SDM C_5A	This study
pPclp12	pPclp6 with SDM G_6A	This study
pPclp13	pPclp6 with SDM C_7A	This study
pPclp14	pPclp6 with SDM T_8G	This study
pPclp15	pPclp6 with SDM G_9A	This study
pPclp16	pPclp6 with SDM $T_{10}G$	This study
pPclp17	pPclp6 with SDM $A_{11}G$	This study
pPclp18	pPclp6 with SDM $A_{12}G$	This study
pPclp19	pPclp6 with SDM $G_{13}A$	This study
pPclp20	pPclp6 with SDM $C_{14}A$	This study
pPclp21	pPclp6 with SDM $G_{15}A$	This study
pPclp22	pPclp6 with SDM $A_{16}G$	This study

pPclp23	pPclp6 with SDM $A_{17}G$	This study
pPclp24	pPclp6 with SDM $C_{18}A$	This study
pPclp27	pPclp6 with SDM $G_{19}A$	This study
pPclp28	pPclp6 with SDM $C_{20}A$	This study
pPclp25	92 bp upstream of *clpP1P2* in pSC-A	This study
pPclp26	92 bp upstream of *clpP1P2* in pSM128	This study
pPclp33	pPclp6 with GGG motif mutated to AAA	This study
pPclp34	pPclp1 with GGG motif mutated to AAA	This study
pPclp35	pPclp26 with GGG motif mutated to AAA	This study
pPclp36	280 bp upstream of *clpP2* in pSC-B	This study
pPclp37	280 bp upstream of *clpP2* in pSM128	This study

Appendix 5. Primers and plasmids for ClgR purification

Target gene	Primer name	Sequence (5' to 3')
clgR	ClgRDD-F	GCCATATGGCGGCTTTGGTGCGTGAG
	ClgRDD-R	ATGGATCCTTAGTCATCCGCCAGCGACACCACCGGC

Plasmid name	Characteristics	Source
pET-28a	Expression vector, N-terminal His tag, Kan^R	Novagen
pPclp38	ClgR-DD in pGEM, Amp^R	This study
pPclp39	ClgR-DD in pET28-A, Kan^R	This study

Appendix 6. EMSA primers and products

Primer name	Sequence (5' to 3')	Purpose
P278F	CAGGGCCGCAGTGGAGGC	Amplification of 278 bp upstream of *clpP1*
P278R	GCTCACAGTGGGGCACCT [biotin]	
P48F	CGACAACGTGACCGT	Amplification of 48 bp regulatory region upstream of *clpP1*
P48R	CCCTGAAACCGGC [biotin]	

Product name	Characteristics	Source
P_{278}	278 bp upstream of *clpP1*, biotin labelled	This study
P_{48}	48 bp regulatory region upstream of *clpP1*, biotin labeled	This study

Appendix 7. Method for ClgR purification

ClgR-DD was purified by Dr Ruth Rose at the Queen Mary University protein purification facility. ClgR-DD plasmid was transformed in BL21 (DE3) Codon plus competent cells. Cells were grown in one litre LB medium at 37 ºC to an OD_{600} of approximately 0.9, induced with 1mM IPTG, and grown overnight at 18 ºC. Culture cells were harvested by centrifugation at 3,315 x *g*, re-suspended in 50 mL buffer (20 mM Tris pH 7.0, 500 mM NaCl, 10 mM imidazole) and lysed by two passages through an Emusiflex (**lys**); cell debris (**pel**) were removed by centrifugation at 35,000 x *g* for 20 min. Soluble fraction (**s/n**) was incubated with 1 ml of Ni-NTA sepharose (Sigma) for 1 h at 4 ºC on a rotator. The contents were poured into a column and unbound material removed under gravity (**f/t**). The sepharose was washed (**wash**) in 20 mL buffer (20 mM Tris pH 7.0, 500 mM NaCl, 10 mM imidazole). The protein was eluted in elution buffer (20 mM Tris pH 7.0, 500 mM NaCl, 500 mM imidazole) into 1.5 ml eppendorfs. The protein was buffer exchanged in storage buffer (20 mM Tris pH 7.4, 10 % glycerol) using a PD10 column (GE Healthcare), eluted into 3.5 ml storage buffer, and stored at -20 ºC. Fractions that were not exchanged were kept in elution buffer at 4 ºC.

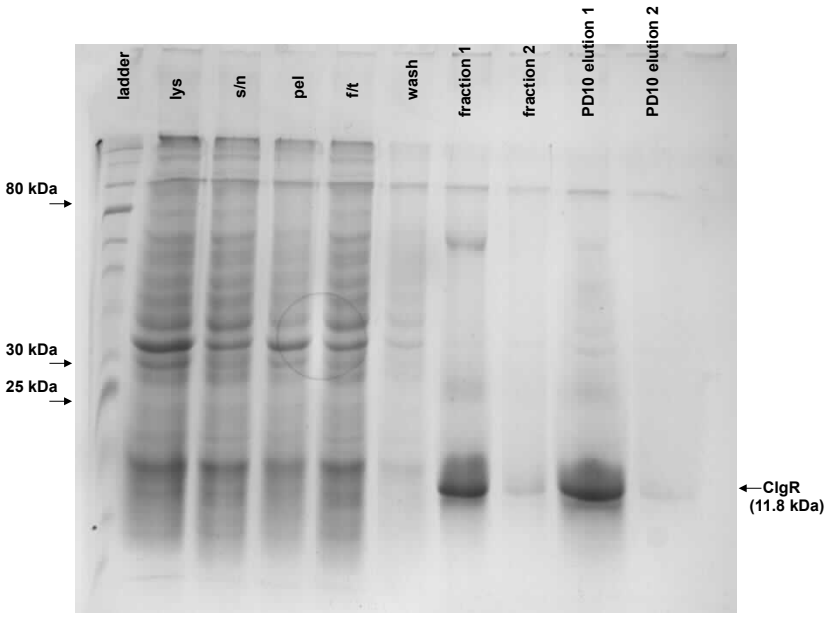

Figure 8.1 SDS gel of purified ClgR-DD

The samples were boiled at 100 ºC for 10 min. 10 µl of sample was loaded on to a 15 % SDS gel run at 200 V for 1 h and stained in coomassie stain.
lys, s/n, pel, f/t, wash = 40 µl sample buffer + 10 µl sample (see text for description)
Elution fractions = 20 µl sample + 20 µl sample buffer

Appendix 8. Primers used for cloning into the bacterial two hybrid system vectors in this study

Target gene	Primer name	Sequence (5' to 3')
clpP1	clpP1F	GTCTAGATGAGCCAAGTGACTGACATG
	clpP1R	GGAGCTCCTGTGCTTCTCCATTGAC
clpP2	clpP2F	GTCTAGAGAATTCCCAAAATTCTCAGA
	clpP2R	CGGTACCGTTTGCGCGGAGAGCTT
clpB	clpBF	ACTCTAGAGGACTCGTTTAACCCGACG
	clpBR	CTCGGTACCGCCCAGGATCAGCGAGTC
clpC1	clpC1F	CCTCTAGAGAACGATTTACCGACCGTGCCC
	clpC1R	CGAGCTCCGCGGCCGGCTCCGGGCCGCCCG
	clpC1R2	AAGCTTAAACCGACGTCAGCAGAGTCTATTGTCACA
clpC2	clpC2F	ACTCTAGAGCCGGAGCCCACACCCACCGCC
	clpC2R	CTCGGTACCGCCGGCATCGGTTGCGCCGGC
clpX	clpXF	GTCGACTCTAGAGGCGCGCATAGGAGACGGT
	clpXR	CTCGGTACCCGCGCTCTTGTCGCGGCG
clpS	clpSF	ACTCTAGAGGCTGTTGTGTCAGCGCCC
	clpSR	GGATCCCCGGTCCTGCTGCATCGT
smpB	smpBF	CTCTAGAGTCCAAGTCGTCGCGTGGC
	smpBR	GGATCCGGTCATGCCCTTAGCGCG
esat6	esat6F	GACTCTAGAGACAGAGCAGCAGTGGAAT
	Esat6R	GGGTACCTGCGAACATCCCAGTGAC
cfp10	cfp10F	GACTCTAGAGGCAGAGATGAAGACCGAT
	cfp10R	GGGTACCGAAGCCCATTTGCGAGGA
E. coli clpX	EcClpXF	GACTCTAGAGACAGATAAACGCAAAGATGGC
	EcClpXR	GGGTACCTTCACCAGATGCCTGTTGCGC
E. coli clpP	EcclpPF	GACTCTAGAGTCATACAGCGGCGAACGAGAT
	EcclpPR	GGGTACCATTACGATGGGTCAGAATCGA

Appendix 9. Primers used for cloning in the Split-Trp vectors

Target gene	Name	Primer Sequences (5' to 3')
clpP1	spltP1F	GCACTAGTGTGAGCCAAGTGACTG AC
	spltP1R	GCGTTAACCTGTGCTTCTCCATTGAC
clpP2	spltP2F	GCACTAGTGTGAATTCCCAAAATTCT
	spltP2R	GCGTTAACGGCGGTTTGCGCGGAGAG
clpB	spltBF	GCACTAGTGTGGACTCGTTTAACCCG
	spltBR	GCGTTAACGCCCAGGATCAGCGAGTC
clpC1	spltC1F	GCACTAGTATGTTCGAACGATTTA CC
	spltC1R	GCGTTAACCCGCGCGGCCGGCTCCGG
	clpC1R2	AAGCTTAAACCGACGTCAGCAGAGTCTATTGTCACA
clpC2	spltC2F	GACTAGTATGCCGGAGCCCACACCCAC CGCCTACCCC
	spltC2R	GGTTAACGCCGGCATCGGTTGCGCCGGCATCGGTT GC
clpX	spltclpXF	GCACTAGTATGGCGCGCATAGGAGAC
	spltXR	GCGTTAACCGCGCTCTTGTCGCGGCGCTC
clpS	spltSF	GACTAGTGTGTCCAAGTCGTCGCGT
	spltSR	GGTTAACGGTCATGCCCTTAGCGCG
smpB	spltSmpBF	GACTAGTATGGCTGTTGTGTCAGCG
	spltSmpBF	GGTTAACCCGGTCCTGCTGCATCGT

Appendix 10. Vectors and plasmids used in the bacterial two-hybrid system

Name	Characteristics	Source
pUT18	T18 fragment *B. pertussis* adenylate cyclase subdomain with a multicloning sequence at the 5' end, Amp^R	Karimova *et al*., 1998
pUT18C	T18 fragment *B. pertussis* adenylate cyclase subdomain with a multicloning sequence at the 3' end, Amp^R	Karimova *et al*., 1998
pKT25	T25 fragment of *B. pertussis* adenylate cyclase subdomain with a multicloning sequence at the 3' end, Kan^R	Karimova *et al*., 1998
pKNT25	T25 fragment of *B. pertussis* adenylate cyclase subdomain with a multicloning sequence at the 5' end, Kan^R	Karimova *et al*., 1998
pUT18C zip	GCN4 leucine zipper motifs cloned into pUT18C, Amp^R	Karimova *et al*., 1998
pKNT25 zip	GCN4 leucine zipper motifs cloned into pKNT25, Kan^R	Karimova *et al*., 1998
pClp1	ClpP1 in pSC-A, Amp^R	This study
pClp2	ClpP1 in pKT25, Kan^R	This study
pClp3	ClpP1 in pKNT25, Kan^R	This study
pClp4	ClpP1' in pSC-A, AmpR	This study
pClp5	ClpP1 in pUT18C, AmpR	This study
pClp6	ClpP1 in pUT18, Amp^R	This study
pClp7	ClpP2 in pSC-A, Amp^R	This study
pClp8	ClpP2 in pKT25, Kan^R	This study
pClp9	ClpP2 in pKNT25, Kan^R	This study
pClp10	ClpP2' in pSC-A, Amp^R	This study
pClp11	ClpP2 in pUT18C, Amp^R	This study
pClp12	ClpP2 in pUT18, Amp^R	This study
pClp13	ClpB in pSC-A, Amp^R	This study
pClp14	ClpB in pUT18C, Amp^R	This study
pClp15	clpB in pUT18, Amp^R	This study
pClp16	ClpX in pSC-A, Amp^R	This study
pClp17	ClpX in pUT18C, Amp^R	This study
pClp18	ClpX in pUT18, Amp^R	This study
pClp19	ClpC1 in pSC-A, Amp^R	This study
pClp20	ClpC1 in pUT18C, Amp^R	This study
pClp21	ClpC1 in pUT18, Amp^R	This study
pClp22	ClpC2 in pSC-A, Amp^R	This study
pClp23	ClpC2 in pUT18C, Amp^R	This study
pClp24	ClpC2 in pUT18, AmpR	This study
pClp25	ClpS in pSC-A, Amp^R	This study
pClp26	ClpS in pKT25, Kan^R	This study

pClp27	ClpS in pKNT25, Kan^R	This study
pClp28	SmpB/pSC-A, Amp^R	This study
pClp29	SmpB in pKT25, Kan^R	This study
pClp30	SmpB in pKNT25, Kan^R	This study
pClp31	ClpC1 in pKT25, Kan^R	This study
pClp32	ClpC1 in pKNT25, Kan^R	This study
pClp44	Esat6 in pSC-A, Amp^R	This study
pClp33	Esat6 in pUT18C, Amp^R	This study
pClp34	Esat6 in pUT18, Amp^R	This study
pClp35	Cfp10 in pSC-A, Amp^R	This study
pClp36	Cfp10 in pKT25, Kan^R	This study
pClp37	Cfp10 in pKNT25, Kan^R	This study
pClp38	*E. coli* ClpP in pSC-A, Amp^R	This study
pClp39	*E. coli* ClpX in pSC-A, Amp^R	This study
pClp40	*E. coli* ClpP in pKT25, Kan^R	This study
pClp41	*E. coli* ClpP in pKNT25, Kan^R	This study
pClp42	*E. coli* ClpX in pUT18, Amp^R	This study
pClp43	*E. coli* ClpX in pUT18C, Amp^R	This study

Appendix 11. Vectors and plasmids used in Split-Trp

Plasmid name	Characteristics	Source
PL240	Ntrp fusions of Trp1p, Gm^R	O'Hare *et al*., 2008
PL242	Ctrp fusions of Trp1p, $Amp^R Hyg^R$	O'Hare *et al*., 2008
PL186	Esat6 in PL240, Gm^R	O'Hare *et al*., 2008
PL339	Cfp-10 into PL242, Amp^R, Hyg^R	O'Hare *et al*., 2008
PL425	C2 into PL240, Gm^R	O'Hare *et al*., 2008
PL426	C1 in PL242, AmpR, Hyg^R	O'Hare *et al*., 2008
pSplt1	ClpP1 in pSC-A, Amp^R	This study
pSplt2	ClpP2 in pSC-A, Amp^R	This study
pSplt3	ClpB in pSC-A, Amp^R	This study
pSplt4	ClpC1 in pSC-A, Amp^R	This study
pSplt5	ClpC2 in pSC-A, Amp^R	This study
pSplt6	ClpX in pSC-A, Amp^R	This study
pSplt7	ClpP1 in PL240, Gm^R	This study
pSplt8	ClpP2 in PL240, Gm^R	This study
pSplt9	ClpB in PL240, Gm^R	This study
pSplt10	ClpC1 in PL240, Gm^R	This study
pSplt11	ClpC2 in PL240, Gm^R	This study
pSplt12	ClpX in PL240, Gm^R	This study
pSplt13	ClpP1 in PL242, Amp^R, Hyg^R	This study
pSplt14	ClpP2 in PL242, Amp^R, Hyg^R	This study
pSplt15	ClpB in PL242, Amp^R, Hyg^R	This study
pSplt16	ClpC1 in PL242, Amp^R, Hyg^R	This study
pSplt17	ClpC2 in PL242, Amp^R, Hyg^R	This study
pSplt18	ClpX in PL242, Amp^R, Hyg^R	This study
pSplt19	SmpB in pSC-A, Amp^R	This study
pSplt20	SmpB in PL240, Gm^R	This study
pSplt21	SmpB in PL242, Amp^R, Hyg^R	This study
pSplt22	ClpS in pSC-A, Amp^R	This study
pSplt23	ClpS in PL240, Gm^R	This study
pSplt24	ClpS in PL242, Amp^R, Hyg^R	This study

Appendix 12. qRT-PCR primer and probe sequences used in this study

Target gene	Name	Sequence (5' to 3')
sigA	SigA probe	CCTCCGGTGATTTC
	SigA-F	CCGATGACGACGAGGAGATC
	SigA-R	GGCCTCCGACTCGTCTTCA
clpP1	ClpP1 probe	TGTTCCGGCTCAACG
	ClpP1-qF	TCGCCGTGATCAAGAAAGAA
	ClpP1-qR	GGCTGGCCGGTGAATTC
clpP2	ClpP2 probe	TGGCGATCTACGACACC
	ClpP2-qF	GCGATATCACCATGTACATCAACTC
	ClpP2-qR	GGCCCGCACGTATTGC
hemD	HemD probe	ACGCCGTCGCACGCACCCA
	HemD-qF	GGATGCGGTAATCAGCGAGGTG
	HemD-qR	CCGGGCACGATTTCGATGTG
fadD13	FadD13 probe	TGGCGCCGTGCCGGCGAT
	FadD13-qF	ATCTCGATGCCGCAGTTCGAT
	FadD13-qR	ACCTGGCGCATGAAGTTGA
Rv2205c	Rv2205c probe	CCGGGCTCGGCAACCGCGT
	Rv2205-qF	GTTTGCGGGCCACTGAACAC
	Rv2205c-qR	CAAGCCTGCGCACACTCCAA
devB	DevB probe	TGAGCAACGCCCGCCGGGC
	DevB-qF	CGAACGCTACGTTCCCGAAGAC
	DevB-qR	GGTGCACCTGGTTCGATGGA
wbbl2	Wbbl2 probe	TCGCGGCTACCGGCCGACG
	Wbbl2-qF	CTCCCACGCAATGACGCTCAAC
	Wbbl2-qR	CAAACATCGGCCCGACCATTCC
Rv2526	Rv2526 probe	CGGGCAGCCCAGGCCGTCA
	Rv2526-qF	GAGGACCACGATTGAGCTGGAC
	Rv2526-qR	GACCGTCGCTCGCAATGTTTCC
Rv2527	Rv2527 probe	ACGCCGCCAGCCGGCATCG
	Rv2527-qF	ACCACCTGGATTCTGGACAAGAG
	Rv2527-qR	ACAGCCATTCAAGTTCGCCGAT

Appendix 13. Primers and plasmids used for reporter system experiments

Primer Name	Sequence	Purpose
GGG -F	CGACGAAAACTACGCTGGAGGAGGATA ATAATAACCGGGC	AANDENYA-ASV tag mutated to AANDENYA-GGG
GGG-R	GCCCGGTTATTATTATCCTCCTCCAGCG TAGTTTTCGTCG	

Plasmid name	Description	Source
pFLAME-3-ace	GFP promoter probe vector (LAA), P_{ami}, kan^R	Blokpoel *et al*., 2003
pCon3	$P_{smyctetO}$ promoter in pGEM, Amp^R	Dr P. Carroll, QMUL
pMC1m	WT TetR, $P_{smyc\text{-}tetO}$	Guo *et al*., 2007
pMC1s	Reverse TetR, $P_{smyctetO}$	Guo *et al*., 2007
pTT1A	Tweety derivative, Amp^R, Kan^R	Pham *et al*., 2007
pDS176	pSM128 with AANDENYA-LAA tag on LacZ	Dr D. Schuessler, NIMR
pDS178	pSM128 with AANDENYA-ASV tag on LacZ, P_{ami} promoter, Sm^R	Dr D. Schuessler, NIMR
pDS179	pSM128, P_{ami} promoter, Sm^R	Dr D. Schuessler, NIMR
pAMIL1	pDS176, P_{ami} promoter, Sm^R	This study
pRep1	pDS178 with AANDENYA-ASV tag mutated to AANDENYA-GGG	This study
pRep7	Untagged LacZ in pSM128, $P_{smyc\text{-}tetO}$ promoter, Sm^R	This study
pRep9	WT TetR in pGEM	This study
pRep10	Reverse TetR in pGEM	This study
pRep11	WT TetR in pTT1A, Kan^R	This study
pRep12	Reverse TetR in pTT1A, Kan^R	This study
pOPPY1	*clpP1* in pSMT3, hsp60 promoter, Hyg^R	Ollinger *et al*., 2011
pOPPY2	*clpP2* in pSMT3, hsp60 promoter, Hyg^R	Ollinger *et al*., 2011
pOPPY7	*clpP1clpP2* operon in pSMT3, hsp60 promoter, Hyg^R	Ollinger *et al*., 2011

Appendix 14. Primers used for recombineering in this study

Target gene	Primer name	Sequence (5' to 3')
clpS	ClpS-UF	GCAAGCTTCGTTTCGAGCAGGACGCATT
	ClpS-UR	GCAGATCTTTGGCGGGCGCTGACACAAC
	ClpS-DF	GCAGATCTGAAGTCGACGTGTCCAAGCTG
	ClpS-DR	GCTCTAGACCGATCTCAAGCATCACT
smpB	SmpB UF	GCGGCGCCCAGCACAAGGACTTTGACA
	SmpB-UR	GCAGATCTCCGCCACGCGACGACTTG
	SmpB-DF	GCAGATCTCGGCGCGCTAAGGGCATGA
	SmpB-DR	GCTCTAGATGTCCTGTTCGTGCATCCAG
ssrA	SsrA-UF	GCAAGCTTAGTGCTGGGTGCGTCATCT
	SsrA-UR	GCAGATCTGCCATTAACTTCCCGGT
	SsrA-DF	GCAGATCTGCCGATCGTCCGGATCTG
	SsrA-DR	GCTCTAGAAACACTCTCCGACGCAAAAC
smpB	SmpB-compF	CATCGCCGTAGCGGGCT
	SmpB-compR	GCAAGCTTTCAGGTCATGCCCTTAGC
ssrA	ssrA-compF	CATCCGCGTTTCCGTGGAC
	ssrA-compR	GCAAGCTTGATCAGATCCGGACGATCG
ssrA	ssrADD-F	GCGACTACGCTCTCGATGACTAAGCGACGGCTAG
	ssrADD-R	CTAGCCGTCGCTTAGTCATCGAGAGCGTAGTCGC

Appendix 15. Plasmids used for recombineering

Plasmid name	Characteristics	Source
pAL58	Dif-Hyg-dif construct in pCR-Blunt II-TOPO, Kan^R, Hyg^R	Cascioferro *et al*., 2010
pUC-Gm-Int	*Hin*dIII-excisable Gm^r-L5 integrase cassette in pUC18	Mahenthiralingam *et al*., 1998
pJV53	Chec9 genes gp60_61 under control of the acetamidase promoter in pLAM12, Kan^R	van Kessel and Hatfull, 2007
pStar7	Dif-Hyg-dif cassette flanked with *ssrA* upstream and downstream regions in pGEM, Amp^R, Hyg^R	This study
pStar9	Dif-Hyg-dif cassette flanked with *smpB* upstream and downstream regions in pGEM, Amp^R, Hyg^R	This study
pStar17	ClpS upstream and downstream regions in pGEM, Amp^R	This study
pStar18	Dif-Hyg-dif cassette flanked with *clpS* upstream and downstream regions in pGEM, Amp^R, Hyg^R	This study
pStar10	ssrA and 201 bp of upstream region in pGEM, Amp^R	This study
pStar11	smpB and 198 bp of upstream region in pGEM, Amp^R	This study
pStar12	clpS and 198 bp of upstream region in pGEM, Amp^R	This study
pStar13	pStar10 with Gm cassette inserted, Amp^R, Gm^R	This study
pStar14	pStar10 with Gm cassette inserted, Amp^R, Gm^R	This study
pStar15	pStar10 with Gm cassette inserted, Amp^R, Gm^R	This study
pStar16	pStar10 with the last two residues of ssrA (AA) mutated to DD	This study