Characterisation of gene regulation in mycobacteria.

Laura Dixon

A thesis submitted for the degree of:

Doctor of Philosophy

September 2017

Department of Molecular Biology and Biotechnology

The University of Sheffield

Abstract

This study focuses on the characterisation of gene regulation in mycobacteria in response to various signals within the cell. The regulatory mechanisms explored include two types of protein regulators (cAMP-receptor proteins (CRPs) and CsoR) and an RNA-based method of regulation (a member of the *ydaO*-type riboswitch). The signals concerned are copper and the small nucleotide molecules cyclic AMP and cyclic di-AMP. The study focuses primarily on the regulation of *rpfA*, which encodes a resuscitation promoting factor that is involved in resuscitation of *Mycobacterium tuberculosis* from dormancy. The main findings of the study are as follows. A copper-sensitive repressor protein from *M. tuberculosis,* CsoR, was purified and shown to bind to the *rpfA* gene. The presence of copper caused the release of the DNA:protein complex and kinetic parameters of the interactions were determined. Two mycobacterial CRP proteins, Rv3676 from *Mycobacterium tuberculosis* and Msmeg_6189 from *Mycobacterium smegmatis*, were purified and their interaction with the *rpfA* or *sdh1* promoters were characterised. The effect of cyclic AMP binding on the interaction with DNA was assessed and thermodynamic and kinetic parameters of the interaction between the CRP proteins and cyclic AMP was studied. Comparisons were made with the well characterised *E. coli* CRP protein and the two mycobacterial proteins. Finally, a putative *ydaO*-type riboswitch present in the 5' untranslated region of *rpfA* messenger RNA was shown to play a role in regulation of the gene in a *Mycobacterium marinum* salt stress model. In vitro evidence was gathered to suggest the riboswitch is able to bind cyclic di-AMP and pApA, which favours a shift to a specific structural confirmation. Mutation of G168C/G169C residues in the riboswitch abolished this effect.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Jeff Green, for giving me the opportunity to carry out this PhD. Thank you for all the help and guidance you have given me throughout the project and especially for your support during difficult times. I am truly grateful to have had such an understanding supervisor.

I would also like to thank all members of the Green lab, past and present, for your help with methods, techniques and suggestions for my research. A special thanks to Dr Laura Smith for her help with all things mycobacterial! I would also like to thank Katie, Ben and Peter for all the laughs and good times!

I would like to thank Jenni, whose contributions to my CsoR chapter have been invaluable. I really enjoyed supervising you during your Masters project (during which no Gilson pipettes were harmed..!). Who would have thought we would end up being such close friends after the project.

I would like to also thank my best friends Constantinos and Tom for all the good times, weekend trips and laughs over the years. I would like to thank my parents for their love and support and my brother, Sam. I have the fondest memories of the years that you lived with me in Sheffield during the PhD.

I would like to thank Nam. You have always believed in me. I honestly don't think I could have got this far without you there by my side cheering me on. Thank you.

Finally, I would like to dedicate this thesis to Tanner. You always brought a smile to my face.

Publications

Aung, H.L. in, **Dixon, L.L.**, Smith, L.J., Sweeney, N.P., Robson, J.R., Berney, M., Buxton, R.S., Green, J., and Cook, G.M. (2015). Novel regulatory roles of cAMP receptor proteins in fast-growing environmental mycobacteria. Microbiology 161, 648–661.

List of Abbreviations

AIDS	Acquired immune deficiency syndrome
ASP	Ammonium sulphate precipitation
ATP	Adenosine triphosphate
BLItz	Biolayer interferometry
βME	β-mercaptoethanol
bp	Base pair(s)
CCCP	Carbonyl cyanide m-chlorophenyl-hydrazone
CDS	Coding sequence
CFE	Cell-free extract
CHISAM	Chloroform- isoamyl alcohol 24:1
Ci	Curie (unit)
CIAP	Calf intestinal alkaline phosphatase
Cyclic di-AMP$_{SS}$	Phosphorothioate-modified analog of cyclic di-AMP
Cso	Copper sensitive operon
DNA	Deoxyribonucleic acid
DosR	Dormancy survival regulon
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
EMP	Ethambutol
F-cyclic di-AMP	2'-O-(6-[Fluoresceinyl]aminohexylcarbamoyl)-cyclic di-AMP
gDNA	Genomic DNA
HIV	Human immunodeficiency virus
HPLC	High-performance liquid chromatography
ICP-MS	Inductively coupled plasma mass spectrometry
IF-γ	Interferon-gamma
ILP	In-line probing
INH	Isoniazid
iNOS	Inducible nitric oxide synthase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
IVEGI	*in vivo* expressed genomic island
IVT	*in vitro* transcription
LC	Liquid chromatography
LTBI	Latent tuberculosis infection
MctB	Mycobacterial copper transport protein B
MDR-TB	Multi-drug resistant TB
M. tuberculosis	*Mycobacterium tuberculosis*
mRNA	messenger RNA
MS	Mass spectrometry
MST	Microscale thermophoresis
MTBC	*Mycobacterium tuberculosis* complex
MWCO	Molecular weight cut off
NADH	Reduced nicotinamide adenine dinucleotide
NAG-NAM	N-acetyl glucosamine-N-acetyl muramic acid
nm	Nanometers
NRAMP1	Natural resistance associated membrane protein
PAGE	Polyacrylamide gel electrophoresis
PE protein	Proline-glutamic acid protein

PEI	Polyethyleneimine
PG	Peptidoglycan
PPE protein	Proline-proline-glutamic acid protein
PreQ1	7-aminomethyl-7-deazaguanine
P*rpfA*	Region of *rpfA* promoter containing CRP site
PZA	Pyrazinamide
RicR	Regulated in copper repressor
RIF	Rifampcin
RNA	Ribonucleic acid
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
rpfA$_L$,	185 bp region of *rpfA* gene containing CsoR site
rpfA$_S$,	48 bp region of *rpfA* gene containing CsoR site
rpm	Revolutions per minute
STB	Smooth tubercle bacilli
T4 PNK	T4 polynucleotide kinase
T7 RNAP	T7 RNA polymerase
TB	Tuberculosis
TLC	Thin layer chromatography
T$_m$	Melting temperature or transition midpoint
TNF	Tumour necrosis factor
UTR	Untranslated region
UV	Ultraviolet
WHO	World Health Organisation
WT	Wild-type
XDR-TB	Extensively-drug resistant TB

List of Figures

List of Tables

Table of Contents

1 Introduction

1.1 Introduction to TB

Tuberculosis (TB) is a major cause of death worldwide. The causative agent of TB, *Mycobacterium tuberculosis* has been found to infect humans dating back to 70,000 years ago (Comas et al., 2013). According to the World Health Organisation (WHO), 10.4 million people were infected with TB in 2015 and almost 2 million died as a result (World Health Organization, 2017). TB occurs in all parts of the world but its prevalence is most common in South-East Asia and Africa, collectively accounting for 60% of all cases (World Health Organization, 2017).

There are several factors behind the increased prevalence of TB in these areas. They include: delayed diagnosis and treatment failure; poverty and social instability; and human immunodeficiency virus (HIV) co-infection (Dye, 1999; Tang et al., 2015). TB is able to lie dormant in the host, with the infected person showing no symptoms of the disease (Barry et al., 2009). In fact, one-third of the world's population has latent *M. tuberculosis* infections aiding the delayed diagnosis of the disease. This coupled with treatment failure increases the duration of TB and fuels secondary spread of infection (Dye, 1999). People with HIV are more likely to develop active TB; in fact, TB is the leading cause of death in people with HIV and acquired immune deficiency syndrome (AIDS) (Zumla et al., 2015). In 2015, TB was responsible for 35% of deaths in people with HIV (World Health Organization, 2017).

1.1.1 Treatment of TB and emergence of multi-drug resistance

The effective treatment of TB requires strict adherence to the use of several drugs for a time period of 6 to 9 months. The first-line antibiotics that form the core of the treatment regimen are isoniazid (INH), rifampicin (RIF), ethambutol (EMB) and pyrazinamide (PZA)(CDC, Division of Tuberculosis Elimination, 2016). The antibiotics can have serious side effects, which include ototoxicity, hepatotoxicicty, hyperuricemia and neuropsychiatric manifestations (Gülbay et al., 2006). The therapeutic efficacy of the treatment is deteriorated by patient non-compliance due to the long length of treatment and the associated side effects (Sumartojo, 1993). The emergence of drug-resistant TB goes hand-in-hand with patient non-compliance and over the years, we have seen the rise of multi-drug resistant TB (MDR-TB), defined as infections caused by *M. tuberculosis* that are resistant to at least two of the first-line anti-TB antibiotics, INH and RIF (World Health Organization, 2017). There are also forms of extensively-drug resistant TB (XDR-TB) with additional resistance to second-line medications and fluoroquinolones (World Health Organization, 2017). MDR- and XDR-TB pose an increasingly growing threat to global health and hinder successful treatment of TB, frequently leaving patients without any additional treatment options. In fact, only 52% and 28% of patients with MDR- and XDR-TB

respectively are successfully treated and accounted for almost half a million new cases and 200, 000 deaths in 2014 alone (World Health Organization, 2017). Worryingly, resistance has been shown to arise almost immediately upon introduction of new antibiotics (Cohen et al., 2015; Eldholm et al., 2015).

The WHO has launched five priority actions to address the global MDR-TB crisis. They include:

- Prevention of MDR- and XDR-TB cases as first priority through high quality treatment of drug-susceptible TB.
- Expansion of rapid detection of drug-resistant TB.
- Provision of prompt access to effective treatment of MDR-TB by ensuring adequate supplies of quality medications and scaling up capacity to deliver treatment.
- Implement proper infection control measures and promptly enroll diagnosed patients onto appropriate treatment to minimize disease transmission.
- Strengthen and maintain response to MDR-TB by increased political commitment and effective leadership across multiple governmental sectors and ensure adequate funding for care and research.

The WHO has recently launched a goal of 90% reduction in TB incidence and 95% reduction in TB related death by 2035 (compared with baseline data in 2015) in an effort to terminate the TB pandemic. To achieve this bold goal, as well as tackling the priority actions listed above; it is essential to conduct research into *M. tuberculosis* to help discover novel strategies of TB treatment.

1.1.2 The evolution of *Mycobacterium tuberculosis*

TB is caused by members of the *Mycobacterium tuberculosis* complex (MTBC) comprising of *M. tuberculosis*, the causative agent of TB in humans; *Mycobacterium africanum*, which causes TB-like symptoms in humans in specific regions of Africa; *Mycobacterium bovis*, the causative agent of TB in cattle; and *Mycobacterium microti*, that causes TB in voles (Imaeda, 1985). *M. tuberculosis* most likely emerged as a human pathogen some 70, 000 years ago in Africa and spread throughout the world upon migration of humans (Hershberg et al., 2008; Gutierrez et al., 2005). The ancient *M. tuberculosis* strain oringally emerged from a soil-dwelling mycobacteria (smooth tubercle bacilli, STB)(Supply et al., 2013). Strains of STB have been isolated from immunocompromised patients in East Africa, but are significantly less virulent than *M. tuberculosis* and do not transmit amongst humans (Supply et al., 2013). The unusual life cycle of *M. tuberculosis* involving a long latency period in the host may be due to the coevolution of *M. tuberculosis* with humans since the beginning of our species when human population was much smaller and isolated (reviewed by Eldholm and Balloux, 2016). Activation

of TB in the elderly, and spreading to the next generation, in which it would remain in its latent form would allow *M. tuberculosis* to spread through generations without destroying its small and isolated populations of hunter-gatherer hosts (Gagneux, 2012).

It is plausible that the modern *M. tuberculosis* has evolved towards shorter latency periods and increasing virulence in accordance with advancements in agriculture and civilization, which has resulted in a vast human population of more than 7 billion people in densely populated urban areas (Comas et al., 2013; Wirth et al., 2008). A worrying example of this is the recent global expansion of the Beijing TB lineage, which is associated with augmented progression to active TB compared to other TB strains (de Jong et al., 2008).

1.2   Cell wall of *M. tuberculosis*

M. tubercolosis is slow growing, with a doubling time of 18-24 h at 37 °C, with optimal supply of oxygen and nutrients. The Gram-positive obligate pathogen is an aerobe, and does not replicate under anaerobic conditions. The bacterium has a unique cell wall structure, which forms a highly impermeable barrier to antibiotics and has a major role in virulence (reviewed by Brennan and Nikaido, 1995; Daffé and Draper, 1997; Kieser and Rubin, 2014).

The cell wall of *M. tuberculosis* is a complex structure composed of three distinct layers: the mycolipid; the arabinogalactan; and the peptidoglycan (PG) layers, which are encased by an outermost capsule of non-covalently linked polysaccharides and proteins (reviewed by Kieser and Rubin, 2014) (Figure 1.1). The high density of lipids in the cell wall of *M. tuberculosis* prevents accurate Gram staining and mycobacteria are known as acid-fast as they are stained by acid-fast dyes, for example the Ziehl-Neelsen stain instead (Hett and Rubin, 2008). The thick, waxy lipid layer of acid-fast bacteria retains the carbol fuchsin dye despite attempts at destaining with acid-alcohol and appears red even when the counter stain is applied. In contrast, non-acid fast bacteria pick up the counter-stain and appear blue.

The *M. tuberculosis* inner membrane is surrounded by the PG layer, which is comprised of long polymers with recurring disaccharide N-acetyl glucosamine-N-acetyl muramic acid (NAG-NAM), linked by peptide bridges (reviewed Kieser and Rubin, 2014). Transpeptidases allow crosslinking between PG polymers to occur and *M. tuberculosis* PG is more heavily cross-linked compared to model organisms such as *Escherichia coli* and *Bacillus subtilis* with an unusually high content (~80 %) of 3-3 type non-traditional peptide cross-links (Kumar et al., 2012; Lavollay et al., 2008). Modifications such as the glycolylation of NAM residues potentially mask recognition of *M. tuberculosis* by the innate receptor cells, as in *B. subtilis*. The PG layer is surrounded by a layer of arabinogalactan. Galactan contains recurring units of 6-$_D$-Gal$f$β1-D-
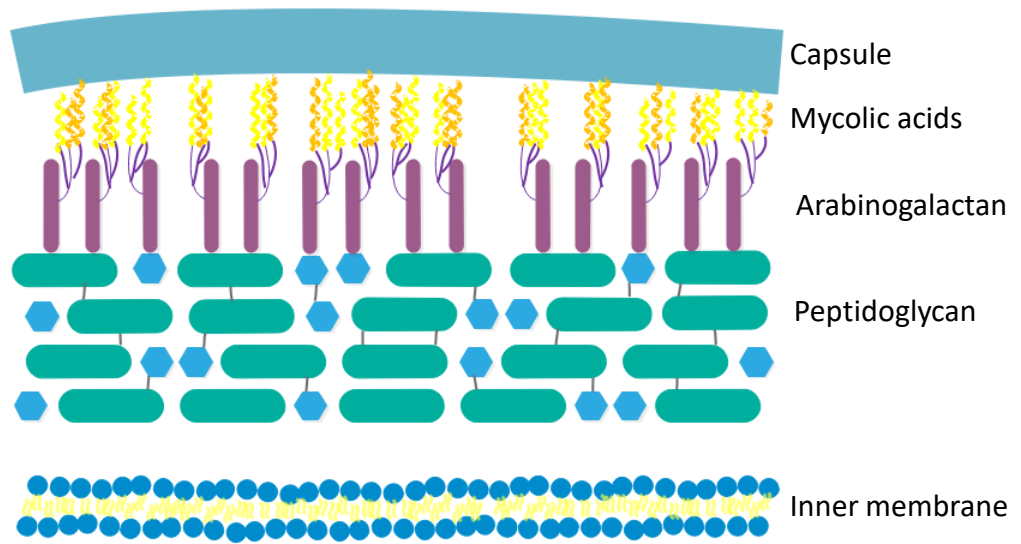
Figure 1.1: Composition of the mycobacterial cell wall. The different portions of a mycobacterial cell wall are shown.

Gal$f$β and can have modifications with long arabinan polymers (Kaur et al., 2009; Makarov et al., 2009). Additionally, arabinan can have modifications such as the addition of non-N-acetylated galactosamine (GalN) or succinyl moieties (Kaur et al., 2009). Such modifications are predominantly found in pathogenic mycobacteria and potentially augment infection (Skovierová et al., 2010).

The presence of mycolic acids ligated to arabinan in the cell wall of *M. tuberculosis* leads to the characteristic thick, waxy lipid coat, which aids its virulence and increases impermeability of the cell wall (reviewed by Kieser and Rubin, 2014). Mycolic acids provide *M. tuberculosis* with properties that resist medical treatment, such as reduced effectiveness of hydrophilic antibiotics. Asselineau and Lederer define mycolic acids as β-hydroxy fatty acids with a long α-alkyl side chain (Asselineau and Lederer, 1950). *M. tuberculosis* produces three different structural classes of mycolic acids: α- mycolic acids, which form the majority (70 %) of mycolic acids found in the organism, methoxy- and keto-mycolic acids, which are the minor components (Barry et al., 2007; Takayama et al., 2005). In mouse models of infection, deletion of the proximal cyclopropane ring of α- or methoxy- and keto-mycolates led to significant attenuation of bacterial growth (Dubnau et al., 2000; Glickman et al., 2000). Furthermore, deletion of keto-mycolates in *M. tuberculosis* restricted its growth in host macrophage (Yuan et al., 1998), thus demonstrating the association between virulence and mycolic acids.

1.3 *M. tuberculosis* infection process

The transmission of *M. tuberculosis* infections occurs via aerosols of highly infectious bacilli. *M. tuberculosis* enters the human body by inhalation of the exhaled bacilli from patients with active TB (Figure 1.2)(Fennelly et al., 2004). The symptoms of TB include coughing, weakness, weight loss, fever and blood in sputum (World Health Organization, 2017). Once inhaled, *M. tuberculosis* enters the alveolar space in the lungs, where it encounters macrophages and is ingested by phagocytosis. In 10-30% of the people exposed to *M. tuberculosis*, the invading bacteria are successfully killed, leading to no infection (Figure 1.2). In 70-90% of the cases however, the bacilli resists the defense mechanisms of the innate immune system and begins replication in alveolar macrophages, dendritic cells and diffusion to nearby endothelial and epithelial cells (Wolf et al., 2008). The ability of *M. tuberculosis* to delay the activation of the adaptive immune system allows itself to reach a critical mass in host cells before attempts to control the infection are made (Russell, 2007; Russell et al., 2010). *M. tuberculosis* can also spread to other tissues and organs via dissemination through the bloodstream and the lymphatics to form secondary lesions (Makarov et al., 2009).
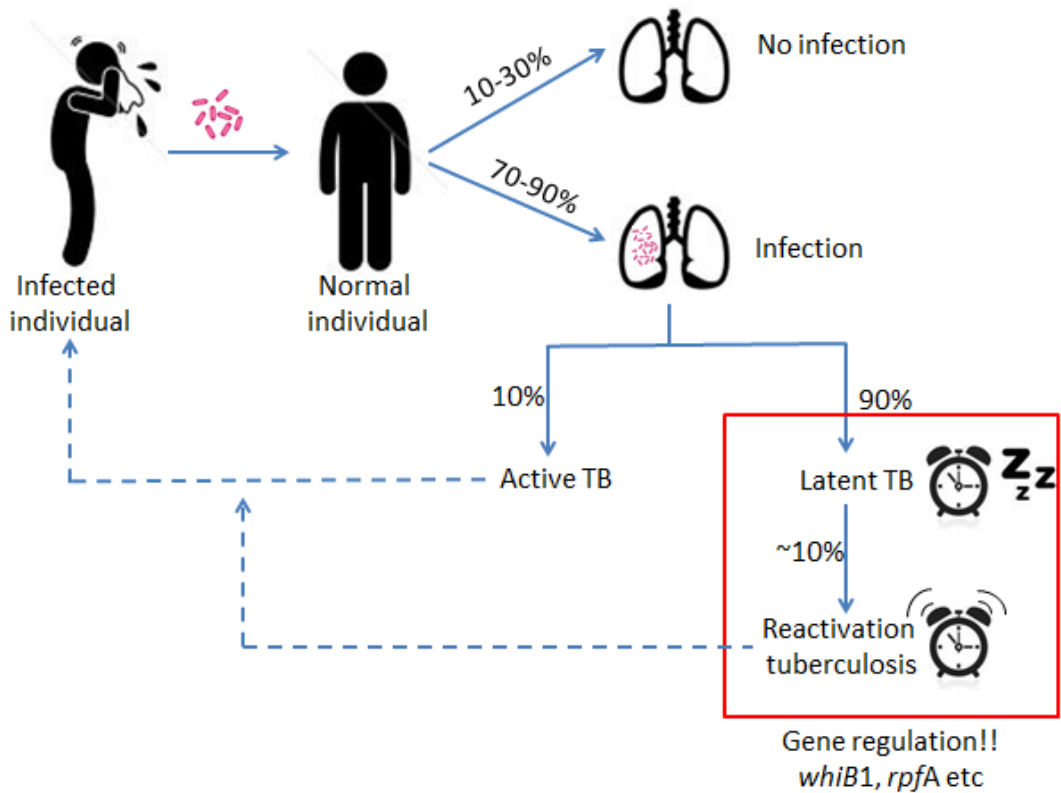
Figure 1.2: Overview of the tuberculosis disease. *M. tuberculosis* is transmitted in air-bourn droplets which are distributed by coughing, sneezing and the spit of infected individuals. From those exposed, 70 to 90% will be infected. Those infected individuals have a 10% chance of having active TB, where they will experience symptoms such as coughing, fatigue, fever, night sweats and weight loss. The other 90% will develop latent TB. The success of *M. tuberculosis* as a pathogen relies on its ability to enter a dormant, non-replicating state from which it can persist and later emerge from when conditions are more favourable. This typically happens when the host has a compromised immune system due to another infection. Understanding the gene regulation and how *M. tuberculosis* is able to emerge from this state may place a key role in elucidating potential future drug targets.

The adaptive immune response then kicks in, resulting in the recruitment of mononuclear cells and T lymphocytes, forming a cellular infiltrate that is known as a granuloma, the hallmark of *M. tuberculosis* infection (Ottenhoff and Kaufmann, 2012)(Figure 1.3). In 90% of the people infected with TB, the infection is contained at this stage, with the bacteria in a non-replicating "dormant" state and latent tuberculosis infection (LTBI) is established (Figure 1.2). People with LTBI show no symptoms of TB and in this way, *M. tuberculosis* maintains a pool of ~2 billion latently infected individuals (World Health Organization, 2017). The success of *M. tuberculosis* as a pathogen relies on its ability to persist in a dormant state and only re-emerging when conditions are more favourable. Dormancy is a state where the bacteria remain viable, but do not divide. However, they remain capable of resuscitation and resuming cell division. There have been numerous studies linking bacterial dormancy to clinical latency (Wayne and Hayes, 1996; Wayne and Sohaskey, 2001). This typically occurs when the host has a compromised immune system, for example as result of co-infection with HIV (Pawlowski et al., 2012). Section 1.7 describes the reactivation process of *M. tuberculosis* in more detail.

Interestingly, new studies are emerging that point towards a dynamic model of TB infection whereby both dormant and actively replicating bacteria are present alongside each other during infection and the balance between the two populations are skewed towards one or the other depending on whether the clinical status of the disease is active or latent (Barry et al., 2009). Primary infections by *M. tuberculosis* is followed by attempts of the adaptive immune response to block bacterial replication. The bacteria are able to resist host defense mechanisms (See Section 1.3.1 for more information) and can persist in a dormant state. A small portion of the dormant bacteria resuscitate, begin replication and act as 'scouts', testing the environmental for conditions appropriate for growth (Figure 1. 3)(Chao and Rubin, 2010). An immunocompetent host would be able to successfully clear actively growing bacteria whilst maintain the dormant population, thus the clinical status of a latent TB infection. An immunocompromised host however would fail to combat the actively growing bacteria, thus giving rise to clinically active TB.

1.3.1 Mycobacterial adaptation and survival strategies in the macrophage

When *M. tuberculosis* enters the body, it firstly comes under attack by alveolar macrophages and neutrophils as part of the body's innate immune response. Once *M. tuberculosis* is engulfed by macrophages through phagocytosis, it is subjected to a harsh intracellular environment, which is highly acidic and with a plethora of reactive oxygen species (ROS) and reactive nitrogen species (RNS)(Cambier et al., 2014). *M. tuberculosis* has a variety of defence mechanisms by which it combats this hostile environment.

Solid granuloma

Necrotic granuloma

Caseous granuloma
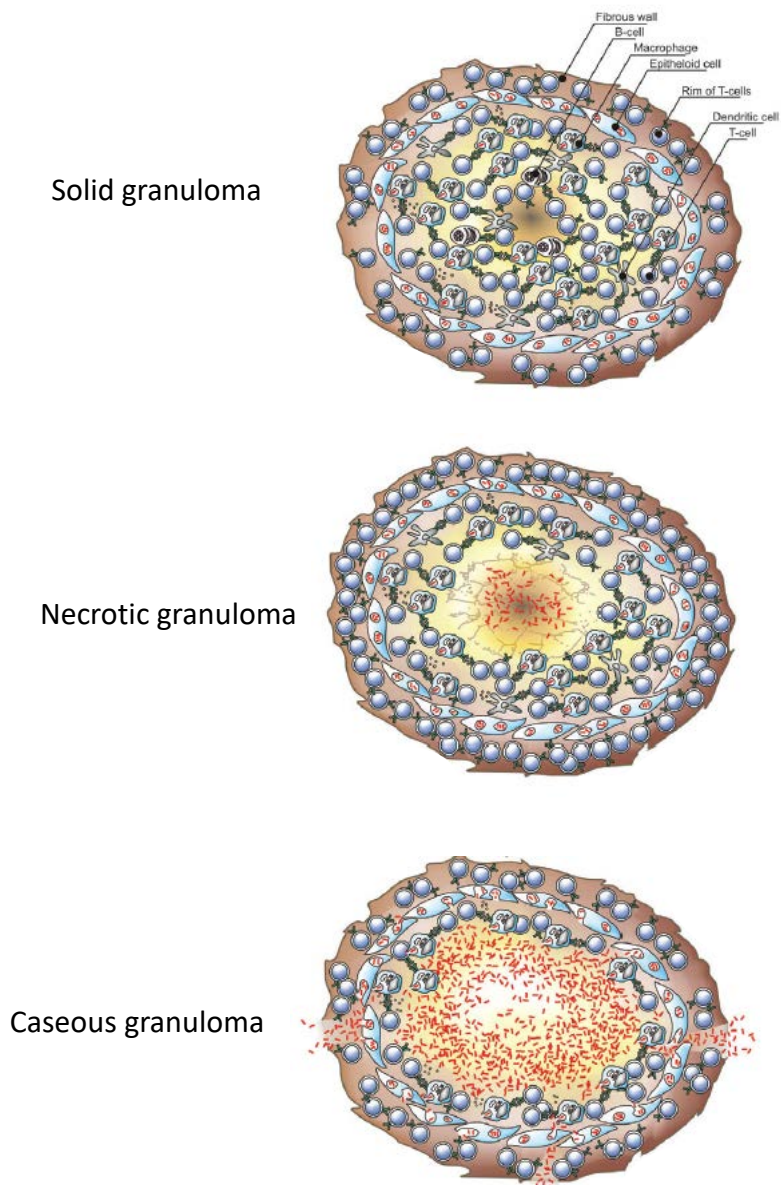
Figure 1.3: Pathology of TB and granuloma structure. Figure adapted from Gengenbacher et al. (2012). One of the major problems in treating TB is there is a huge reservoir of disease – over 2 billion latently infected individuals. Only a small proportion of these directly switch to active TB. Reactivation and reinfection cause ~9 million new cases per annum. Upon inhalation of infected droplets, *M. tuberculosis* reaches the lungs and is phagocytosed by alveolar macrophages. There is an initial pro-inflammatory response which recruits mononuclear cells and T lymphocytes, resulting in granuloma formation. Normal and healthy individuals are able to control the infection at this level, but will have a latent infection and be at risk of reactivation tuberculosis in future when the immune system is compromised. Granulomas maturation (solid, necrotic and caseous) occur at different speeds. Caseating granuloma lose solidarity of its structure due to decay of the centre and accumulation of host cell debris (the caseum). The *M. tuberculosis* cells are released , enter the airway and are coughed out to enable the spread of the disease.

1.3.1.1 Acid stress

M. tuberculosis arrests phagosome maturation at an early stage so as to avoid its fusion with lysosomes, therefore avoiding acidification of the internal environment (Sturgill-Koszycki et al., 1996). In addition, the unique lipid rich cell wall (Figure 1.1) has been shown to act as a permeability barrier against acidic environments when *M. tuberculosis* is exposed to phagolysosomes and acid-sensitive *M. tuberculosis* mutants also show defects in genes involving cell wall biosynthesis (Vandal et al., 2008). Abramovitch et al. showed that the *aprABC* locis found in pathogenic mycobacteria aids resistance against acidic environments in phagosomes (Abramovitch et al., 2011).

1.3.1.2 Defense against oxidative and nitrosative stress

The defense repertoire of the body's immune system against *M. tuberculosis* in activated macrophages also involves ROS, generated by the phagocyte oxidase NOX2; and RNS produced by inducible nitric oxide synthase (iNOS) (Nathan and Shiloh, 2000). ROS such as superoxides and hydroxyl radicals; and RNS such nitric oxide and peroxynitrite can have bactericidal effects by negatively interfering with nucleic acids, proteins, carbohydrates and lipids. *M. tuberculosis* employs strategies of both detoxification and repairment of damage to counteract such stresses (reviewed by Schnappinger et al., 2003).

An example of this is the mycobacterial catalase peroxidase enzyme encoded by *katG*, which detoxifies H_2O_2 to H_2O and O_2. *M. tuberculosis* mutants lacking *katG* were shown to be hypersensitive to H_2O_2 in culture (Ng et al., 2004). Another example is the production of superoxide dismutases, which convert superoxides to H_2O_2 and several studies have highlighted their importance for virulence in many pathogenic bacteria such as *Salmonella typhimurium* (Fang et al., 1999) and *Helicobacter pylori* (Seyler et al., 2001). *M. tuberculosis* contains two such superoxide dismutases encoded by *sodA* and *sodC* (Dussurget et al., 2001; Wu et al., 1998). SodC is copper and zinc based and found in the cell envelope of *M. tuberculosis* (wu al., 1998). It plays an important role in the ability of *M. tuberculosis* to resist oxidative burst in activated macrophages (Piddington et al., 2001). However, *MtbΔsodC* was still found to be virulent in guinea pigs (Dussurget et al., 2001) and the mutant was found to grow in mice up to 60 days post infection (Piddington et al., 2001). The role of the Cu, Zn-based SodC in chronic TB infections has not been elucidated yet.

The second superoxide dismutase in *M. tuberculosis* is encoded by *sodA*, which is iron-based, provides protection against the respiratory burst in naïve macrophages and is found to be essential for growth *in vitro* (Dussurget et al., 2001). Other mycobacterial genes implicated in

detoxification of ROS and RNS are: *mshA*, which is involved in the biosynthesis of mycothiol, a reduced-sulfur-containing molecule that acts as an antioxidant (Vilchèze et al., 2008); *cysH*, which is responsible for sulphur assimilation and is required for the production of thiol-containing antioxidants (Senaratne et al., 2006); and the *ahpC*, *ahpD*, *lpd* and *dalT* genes that encode peroxidasereductases (Gold et al., 2008; Shi and Ehrt, 2006).

1.3.1.3 Metal stress in the macrophage

An important feature of *M. tuberculosis* is its ability to exploit and counteract trafficking of metal cations inside infected macrophages (reviewed by Neyrolles et al., 2015; Rowland and Niederweis, 2012). Essential metals such as iron and manganese are tightly controlled by the host and kept away from the invading *M. tuberculosis* by sequestering it to host proteins, such as ferritin and transferrin, or by pumping them out of the phagosome using the metal cation tranporter NRAMP1 (natural resistance associated membrane protein) (Cellier, 2012). On the other hand, other metal cations like zinc and copper are accumulated in macrophages in response to infections with mycobacteria (Gouzy et al., 2013, 2014). To counteract toxicity by metal overload, *M. tuberculosis* utilizes detoxification systems, metal efflux pumps, sequestration and oxidases (Gold et al., 2008; Rowland and Niederweis, 2012, 2013). The host immune system combines mechanisms of nutritional immunity, i.e. depriving *M. tuberculosis* of essential metals (Fe, Mn) whilst overloading it with toxic levels of Cu and Zn. The sections below discuss the ways in which *M. tuberculosis* manages metal stresses imposed by the immune system in more detail.

Iron

 M. tuberculosis has developed effective iron acquisition mechanisms in order to survive the iron-starved conditions in the host macrophage. This involves the production and secretion of small iron chelators called siderophores, such as mycobactins and carboxymycobactins, which bind free iron with high affinity and transport it into *M. tuberculosis* (Luo et al., 2005; Snow et al., 1970). The majority of the host iron (~70%) is bound in haem (Finch and Huebers, 1982). As a result, many Gram-positive and Gram-negative bacteria have found ways to utilize haem as a major source of iron (Cescau et al., 2007; Wandersman and Stojiljkovic, 2000). Until recently, it was thought that *M. tuberculosis* acquired iron primarily through the production of siderophores however studies have emerged pointing towards a haem acquisition system in mycobacteria (Parish et al., 2005; Raghu et al., 1993; Tullius et al., 2011). Tullius et al. (2011) identified Rv0203, a secreted haem-binding protein, and several membrane encoded by genes *Rv0202c - Rv0207c* that are involved in sequestering haem from the host and transporting it

into *M. tuberculosis*. Furthermore, a haem-degrading protein, MhuD, found in the cytoplasm of *M. tuberculosis* has been characterised (Nambu et al., 2013). Interestingly, in contrast to canonical haem oxygenases, the non-canonical MhuD breaks down haem without the release of CO and it is thought that this unique mechanism of haem degradation may have evolved to avoid the production of signal (CO) for *M. tuberculosis* to switch to a dormant state inappropriately (Nambu et al., 2013).

Zinc

Recent studies on the transcriptional changes that occur in *M. tuberculosis* and human macrophages have shown substantial evidence for heavy metal poisoning during the infection process (Botella et al., 2011; Tailleux et al., 2008). The gene that encodes a putative zinc efflux pump, *ctpC* was upregulated in mycobacteria indicating that the bacilli were exposed to zinc in the phagosome during infection (Botella et al., 2011). When the *ctpC* gene was disrupted, it resulted in the *M. tuberculosis* mutant being hypersensitive to zinc. Indeed, a previous study had shown that heavy metals, for example, zinc and copper (See below for more info on copper) accumulated in phagosomes infected with different mycobacterial species and the accumulation was even greater when inflammatory signals such as tumour necrosis factor (TNF) and interferon-gamma (IF-γ) were used to activate macrophages (Wagner et al., 2005). Zinc can be toxic to bacteria it can displace other cations in essential enzymes, therefore impeding their activity (Stafford et al., 2013). Zinc poisoning is therefore one of the mechanisms macrophages use to destroy *M. tuberculosis* and increasing amounts of studies show that *M. tuberculosis* counteract the stressor zinc with the use of efflux pumps (Botella et al., 2011).

Copper

Recent studies have recognised that high levels of copper are accumulated in mycobacterial phagosomes and as well as ROS, RNS and acid stress, copper is an important addition to the macrophage's arsenal of weapons against *M. tuberculosis* (Adams et al., 1997; Russell et al., 1998; Wagner et al., 2005). The antimicrobial effects of copper are multi-faceted and have been known for some time (Domek et al., 1984). Copper can produce hydroxyl radicals via the Fenton reaction (Wardman and Candeias, 1996), can remove metal co-factors in proteins and destabilise Fe-S clusters (Macomber 2009). On the other hand, copper is an essential micro-nutrient with trace amounts necessary for various cellular functions. It serves as a cofactor in metalloenzymes such as cytochrome *c* oxidases and the superoxide dismutase, SodC (See Section 1.3.1.2 on ROS for more detail). Since copper is both toxic and essential for *M. tuberculosis*, copper metabolism and stress response has to be carefully regulated by the cell.

M. tuberculosis has been shown to be able to respond to changes in copper levels (Liu et al., 2007a). Transcriptome analysis identified a set of at least 30 copper-responsive genes in *M. tuberculosis*, approximately half of which are only induced at toxic levels of copper (Ward et al., 2008). Figure 1.4 shows the known copper homeostasis pathways in *M. tuberculosis*. It is believed that further characterising and elucidating the role of the copper response for survival in the human host may provide novel targets for disease control.

1.4 Copper Homeostasis in *M. tuberculosis*

1.4.1 The *cso* operon

The initial evidence that copper plays an important role in TB infections came from transcriptional analysis of *M. tuberculosis* grown *in vivo* in mice compared to that grown *in vitro* in broth culture (Talaat et al., 2004). Talaat et al. (2004) discovered the first copper responsive operon in *M. tuberculosis* after the transcriptomic studies revealed that several genes encoding putative metal cation transporters were highly induced in *M. tuberculosis* isolated from infected mice versus bacteria grown in broth culture. These genes were located in a unique genomic island called the '*in vivo* expressed genomic island' (IVEGI). Present in the genomic island, they discovered ctpV, a gene encoding a P- type ATPase which may transport copper out of *M. tuberculosis* (Ward et al., 2008). Moreover, analysis of the IVEGI for transcriptional units revealed the presence of an operon, which was highly induced upon treatment with copper, so it was named the copper sensitive operon (*cso*) (Liu et al., 2007).

1.4.1.1 CsoR – the copper-sensitive operon repressor

Similar to the cadmium/lead metal transporter (*cm*t) operon in *M. tuberculosis*, where the first gene encodes CmtR, a lead-sensing transcriptional regulator that controls the entire operon (Cavet et al., 2003), it was found that the first gene in the *cso* operon encodes a copper responsive transcriptional regulator called CsoR (Liu et al., 2007). CsoR, or copper-sensitive operon repressor, was one of the first identified members of a large family of transcriptional repressors and is a Copper(I) responsive transcriptional regulator. The family members are made up of an all α-helical dimer of dimers structure and many are able to sense metal(s) (Higgins and Giedroc, 2014). CsoR-like regulators have been fairly well-characterised and are widespread across the five major classes of eubacteria. Among those characterised include CsoR proteins from *Steptomyces lividans*, *Staphylococcus aureus*, *Listeria monocytogenes*, *Geobacillus thermodenitrificans*, *Thermus thermophilus*. An alignment of these proteins with *M. tuberculosis* CsoR is shown later in Figure 3.1 (located in Chapter 3),
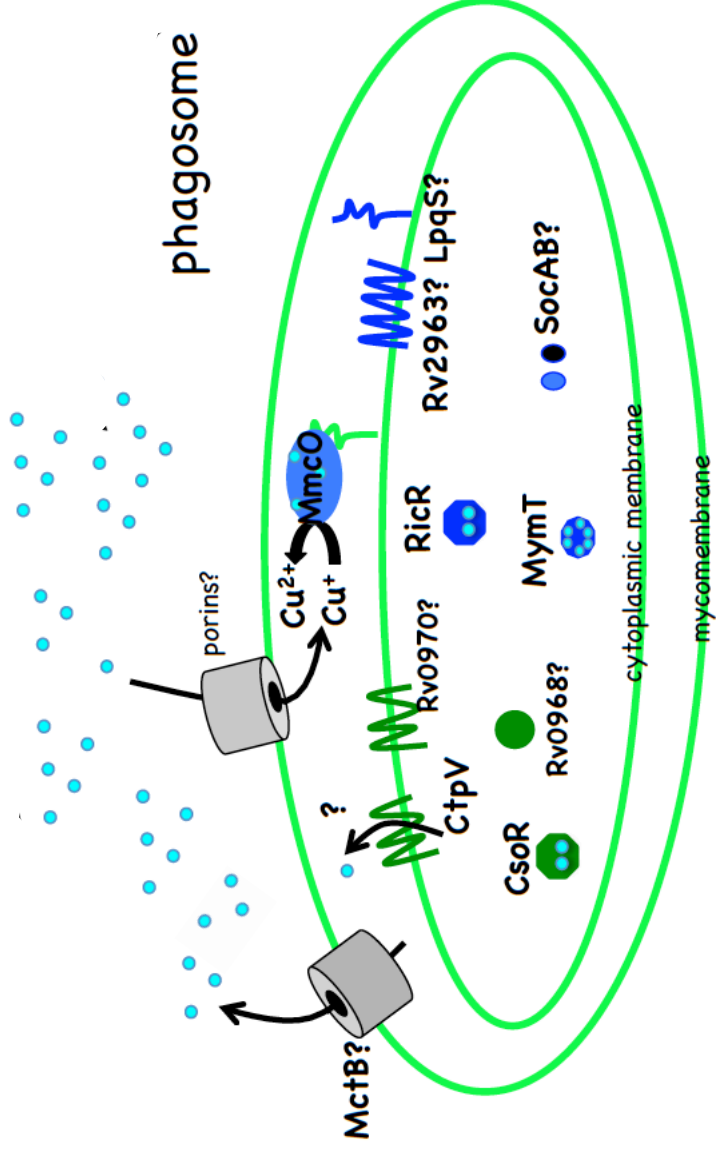
Figure 1.4: Copper in *M. tuberculosis*. Adapted from Darwin (2015). An *M. tuberculosis* bacterium is shown in green. It has at least three independent Cu-responsive pathways. The CsoR regulon (gene products shown in dark green); RicR regulon (gene products in dark blue) and proteins of unknown regulation are shown in grey. Copper ions are shown in light blue.

which also highlights important amino acid residues that will be discussed in greater detail later.

In the apo form when no copper is bound, CsoR binds DNA in the GC-rich, pseudopallindromic site (5' GTAGCCCACCCCCAGTGGGGTGGATAC 3') located in the promoter region of the *cso* operon. This site overlaps with the -10 and -35 regions of the *cso* promoter (Liu et al., 2007). The *cso* operon is comprised of genes encoding CsoR itself, CtpV (a Cu(I) efflux P1-type ATPase) and Rv0968 (a non-essential protein of unknown function). In the presence of Cu(I), CsoR no longer binds the promoter of the cso operon due to the conformational change in the CsoR-DNA complex. It established that the derepression of the *cso* operon observed in *M. tuberculosis* cells treated with Cu was due to the mechanism of CsoR (Ward et al., 2008).

1.4.1.2   Structure of CsoR

CsoR forms a disc-like dimer of dimers that assemble as an overall tetrameric structure (Higgins and Giedroc, 2014). A key structural feature of the dimers is the four-helix bundle (α1, α2, α1', α2'), which is flanked at the C-terminal end by α3/α3', the two helices which form the basis of the tetramer interface.

CsoR proteins contain a signature x-C-H-C motif (where x can be any amino acid) and the copper-binding site is located close to where two monomers bridge to form a homodimer. A ratio of one Cu(I) per monomer binds in a trigonal geometry via interactions with three residues: Cys36 (from α2), Cys65' (from α2') and His61' (from α2') (shown later in Chapter 3, Figure 3.2a). Crystal structures of *M. tuberculosis* CsoR revealed a hydrogen-bonding network between His61', Glu81' of one subunit with Tyr35 of the other (Liu et al., 2007)(Figure 3.2b). This network has been shown to be important in linking copper binding with DNA release for CsoR proteins in *M. tuberculosis* and *B. subtilis* (Liu et al., 2007a; Ma et al., 2009). It was proposed by Liu et al. that the network drives negative allosteric regulation of DNA binding by Cu(I) in *M. tuberculosis* CsoR. These residues are well-conserved among members of the CsoR-like repressor family. The exact details of the DNA-binding mechanism remain elusive at present (Higgins and Giedroc, 2014). However, pulse-chase amidination experiments using Cu(I)-bound and DNA-bound *B. subtilis* CsoR complexes revealed potential roles of C-terminal lysine residues (K96, K97 and K100) and the N-terminus of the α1 helix (Chang et al., 2011; Higgins and Giedroc, 2014). It has been proposed that the CsoR tetramer remodels after Cu(I) binding, which perturbs these residues and results dissociation of CsoR from the DNA site. Only K96 is present in *M. tuberculosis* CsoR.

DNA binding studies suggested that two tetramers of CsoR bind per DNA site in a CsoR tetramer:DNA "sandwich" (Jacobs et al., 2015)(Figure 3.3). The protein is also able to bind other metals, including Ni(II), Zn(II), Co(II), but interactions with other metals are of lower affinity and there is a strong thermodynamic preference for Cu(I) (Ma et al., 2009). Ni(II) binds with a square-planar geometry, Co(II) binds forms a tetrahedral geometry (Higgins and Giedroc, 2014).

1.4.1.3 The role of CsoR in *M. tuberculosis* pathogenesis and host defence

CsoR has multifaceted roles in helping *M. tuberculosis* survive high levels of copper. Firstly, the high affinity of CsoR to Cu(I) means that it can act as a buffer, reducing levels of free copper and therefore preventing its toxic effects to *M. tuberculosis* (Liu et al., 2007). High levels of copper means that CsoR derepresses genes in the *cso* operon, including *ctpV*, encoding for a copper exporter (Ward et al., 2010). Ward et al. (2010) showed that the *M. tuberculosis ctpV* mutant is hyper-sensitive to copper and that mice infected with a *ctpV* mutant were able to survive TB infections longer that mice infected with WT *M. tuberculosis* (Liu et al., 2007). Similarly, in guinea pigs infected with *M. tuberculosis*, lower bacterial counts of *ctpV* mutants were found compared to WT bacteria at 21 days post-infection. However, no defects in the ability of colonisation were found in the *ctpV* mutant. All in all, CtpV is clearly important for resistance of copper toxicity but its role in the virulence of *M. tuberculosis* remains unclear. Interestingly, Marcus et al. (2016) showed that a *csoR* mutant, which does not produce CsoR but which has an intact *cso* operon, had similar survival curves as those of a CtpV overproducing *M. tuberculosis* strain when both were exposed to high levels of copper (Marcus et al., 2016). In both cases, these strains survived better than WT strains in the presence of copper. This suggests that it is the genes within the CsoR regulon as opposed to CsoR itself that are mainly responsible for alleviating the toxic effects of copper stress.

The interesting finding that *csoR* is induced during the enduring hypoxic response of *M. tuberculosis* has pointed towards the role of the *cso* in the dormancy stage of *M. tuberculosis*. Since there is both increased accumulation and enhanced toxicity of copper in hypoxic conditions, an advanced hypoxia-specific copper response would be of important to ensure survival of *M. tuberculosis* (Beswick et al., 1976a; White et al., 2009). Indeed, both hypoxia and increased copper concentrations have been found in granulomas of guinea pigs with TB, highlighting the need for *M. tuberculosis* to combat both stresses concurrently (Wolschendorf et al., 2011). Interestingly, *M. tuberculosis* reduces its need for copper under hypoxic environments by using the copper-independent cytochrome *bd* oxidase instead of the copper-based cytochrome *c* oxidase (Neyrolles et al., 2015).

1.4.2   The MctB protein

The discovery of CsoR was followed by the characterisation of a second locus necessary for copper resistance. The mycobacterial copper transport protein B (MctB) was initially thought to be a mycomembrane protein, functioning as porins to maintain low copper concentrations in *M. tuberculosis* (Song et al., 2008; Wolschendorf et al., 2011). Recent studies have however indicated MctB may actually be anchored to the inner membrane (Rowland and Niederweis, 2012). This has therefore cast uncertainty to whether MctB forms a channel to export copper out of the cell. It could have other mechanisms, yet to be known, in order to maintain copper homeostasis in *M. tuberculosis*.

1.4.3   The RicR regulon

Studies have revealed another copper homeostasis system in *M. tuberculosis* known as the 'regulated in copper repressor' (RicR) regulon (Festa et al., 2011). The RicR regulon includes the following genes: (1) *mmco*, encoding a mycobacterial multicopper oxidase, which oxidises Cu(I) to the less toxic Cu(II) (Rowland and Niederweis, 2013); (2) *mymT*, a gene encoding a Mycobacterium metallothionien, which are small cysteine-rich proteins that bind up to 6 Cu^+ ions and protect against copper toxicity (Gold et al., 2008; Jacob et al., 1998); (3) *lpqs* & *rv2963*, two genes encoding putative membrane proteins (Festa et al., 2011); (4) *socAB*, which encodes two highly basic proteins of unknown function (Festa et al., 2011); and (5) *ricR*, which is a homologue of *csoR* in *M. tuberculosis*. It encodes RicR, which regulates the expression of the above-mentioned genes, as well as itself, in a similar fashion to CsoR. RicR binds to DNA in the absence of Cu^+ and causes repression of genes in the regulon. The binding of Cu^+ to RicR releases it from the DNA and causes derepression of genes (Festa et al., 2011; Gold et al., 2008; Ward et al., 2008).

1.5   Small nucleotide signalling

1.5.1   Cyclic AMP

Cyclic AMP is a universal second messenger, which is utilised by a range of species to regulate a plethora of cellular functions. Cyclic AMP is synthesised from ATP via adenylate cyclase enzymes. Mycobacteria are unique because their genome encodes large numbers of biochemically distinct adenylate cyclases, including 17 in *M. tuberculosis* (Shenoy and Visweswariah, 2006) and 31 in *M. marinum* (McCue et al., 2000). This indicates the molecule is important to mycobacteria. At least 1 of these adenylate cyclase enzymes, Rv0386, is essential for virulence (Agarwal et al., 2009). An overview of the role of cyclic AMP signalling in *M. tuberculosis* is shown in Figure 1.5. The signals in the red boxes stimulate adenylate cyclase expression. Shleeva et al. (2013) showed that increased levels of unsaturated fatty acids, such
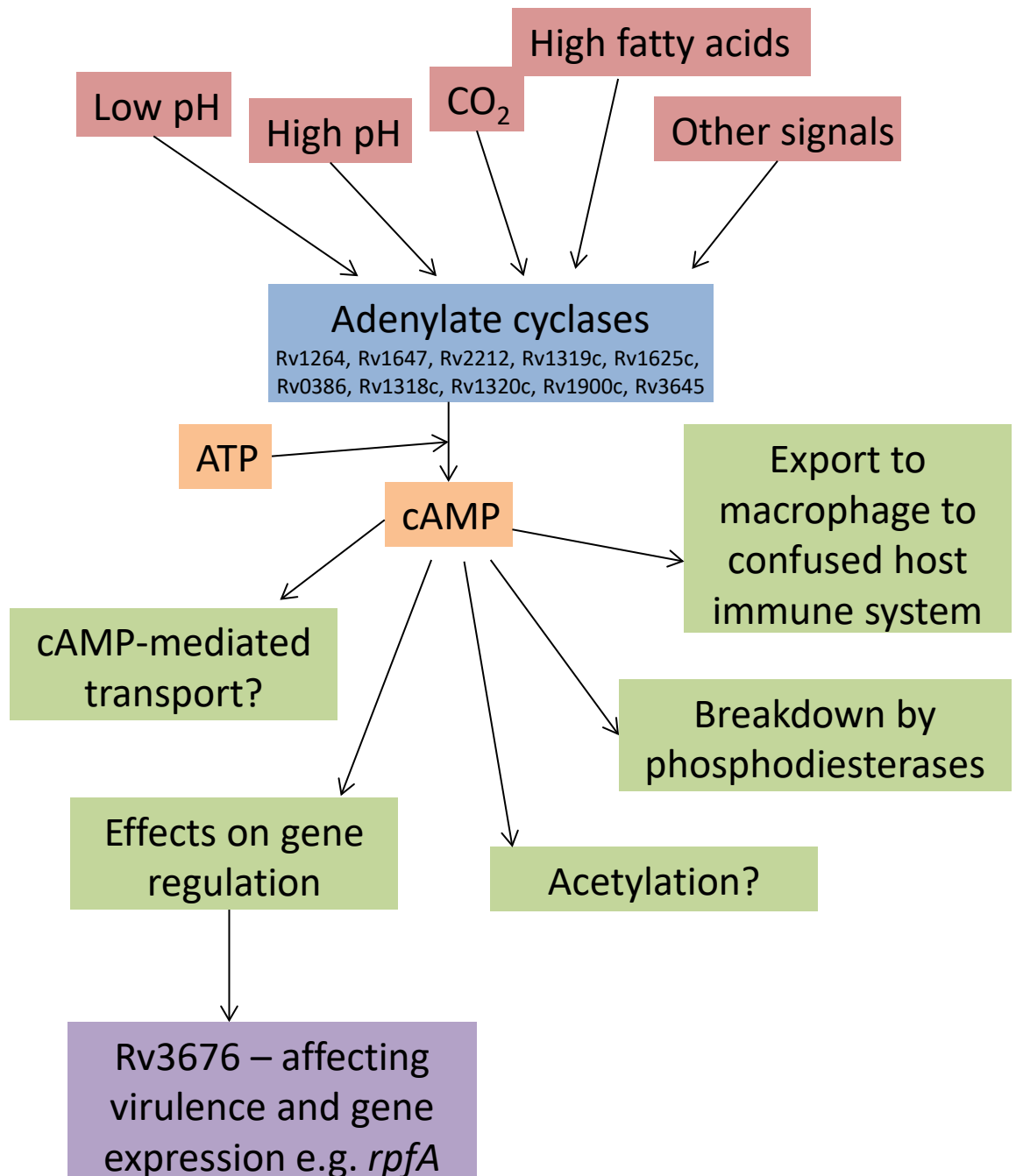
Figure 1.5: Overview of cAMP signalling in *M. tuberculosis*. The adenylate cyclase enzymes are shown in the blue box and different types of signals they respond to are in the red boxes. Cyclic AMP is made by adenylate cyclases from ATP. The cyclic AMP can act as a signal to other parts of the cell (green boxes) , be broken down (by phosphodiesterases) or be exported to the macrophage in attempts to confuse the host immune system and prevent phagosome maturation. The factors in the red boxes are stimuli for the production of adenylate cyclases. For example, Shleeva et al. (2013) provided evidence that high levels of unsaturated fatty acids, such as arachidonic acid, increases production of cyclic AMP by adenylate cyclase in *M. tuberculosis*.

as arachidonic acid results in higher cellular levels of cyclic AMP, which they suggest is due to increases in activity of a fatty acid-activated adenylate cyclase.

1.5.1.1 Confusing host immune system

Many bacterial pathogens use mechanisms that subvert host cell signalling pathways via enzymes capable of synthesising or degrading second messengers, such as cyclic AMP (Baker and Kelly, 2004; Sands and Palmer, 2008). Rv0386 is involved in the delivery of *M. tuberculosis*-derived cyclic AMP to the cytoplasm of the host macrophage. Deletion of the protein and loss of this increased in intramacrophage cyclic AMP leads to decreased immunopathology in animal tissue and reduced survival of *M. tuberculosis* (Agarwal et al., 2009). It has been suggested that this extracellular burst of cyclic AMP by *M. tuberculosis*, alongside other mechanisms, inhibits phagosome maturation which would otherwise be fatal (Agarwal et al., 2009; Axelrod et al., 2008; Fratti et al., 2003).

1.5.1.2 Effect on gene regulation

In addition, cyclic AMP has been shown to be important in virulence and host infection by *M. tuberculosis* via cyclic AMP-associated regulation of virulence genes.

There are several predicted cyclic nucleotide binding proteins in *M. tuberculosis*, the best characterised of which is Rv3676. Rv3676 is a member of the CRP/FNR family of transcription regulators and will be discussed in detail in Chapter 4. It is capable of directly binding both cyclic AMP and DNA targets (Rickman et al., 2005; Stapleton et al., 2010). There is a Rv3676 homolog (Msmeg_6189) present in the *M. smegmatis* genome, which is also discussed in Chapter 4. Other predicted cyclic nucleotide binding proteins include Rv0073, Rv0104, Rv2434c, Rv2564, Rv2565, Rv3239 and Rv3728, which contain different functional domains, including ones associated with transport and esterase activities (Bai et al., 2011; McCue et al., 2000).

Rv3676 is the best characterised cAMP-responsive protein in *M. tuberculosis* and has been implicated in pathogenesis and for *in vitro* growth (Rickman et al., 2005). Microarray analysis revealed differential regulation of 16 genes in the *rv3676* mutant compared to wild-type H37Rv, which included *rpfA* – a resuscitation promoting factor involved in revival from dormancy (Rickman et al., 2005), discussed in detail in Section 1.7 and Chapter 4.

Shleeva et al. (2013) suggests that host mycobacterial lipds play a role during resuscitation of dormant mycobacteria. Resuscitation of dormant *M. tuberculosis* and *M. smegmatis* cells in liquid medium was observed upon addition of free unsaturated fatty acids like as arachidonic
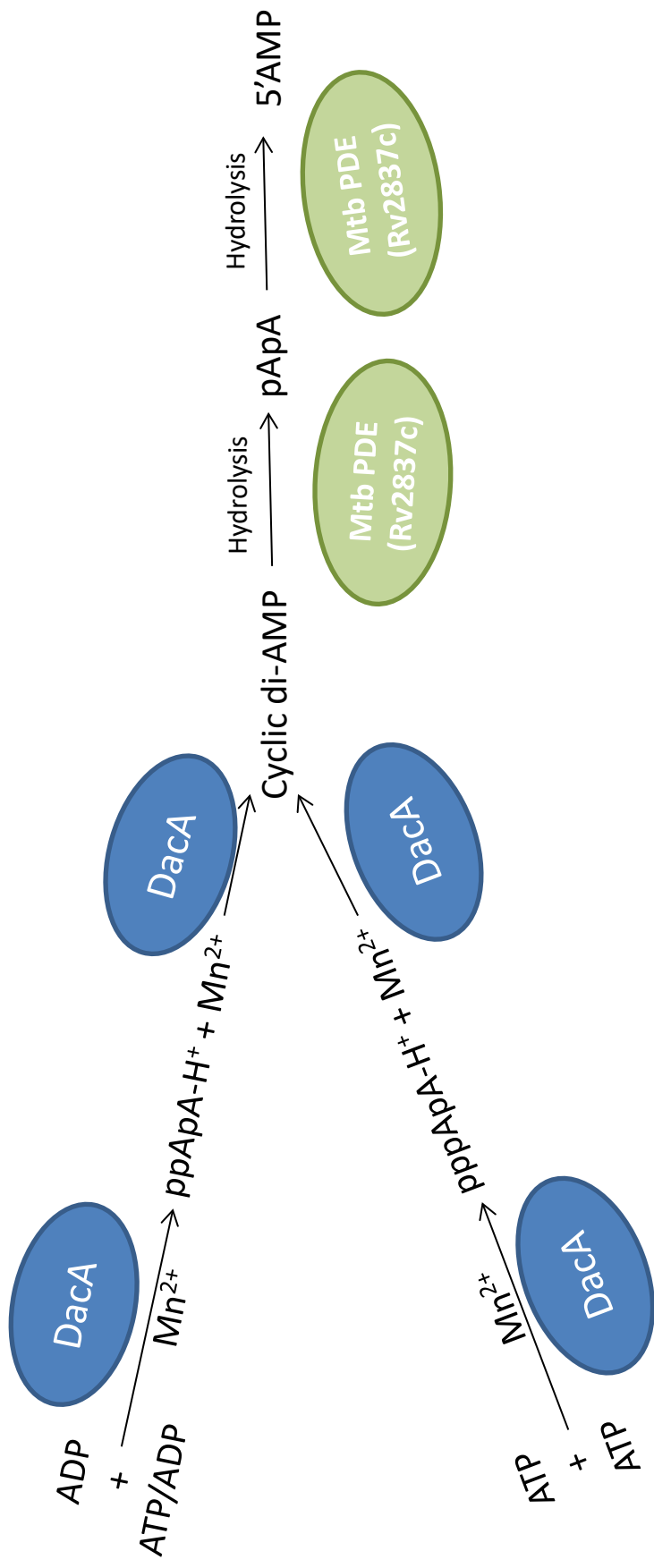
Figure 1.6: Overview of cyclic di-AMP synthesis and degradation in *M. tuberculosis*. DacA is the diadenylyl cyclase that is responsible for synthesising cyclic di-AMP from ATP and/or ADP. The phosphodiesterase Rv2837c is responsible for the breakdown of cyclic AMP to pApA and ultimately 5' AMP. The potential role of cyclic di-AMP and pApA in the regulation of the *rpfA* gene is discussed in Chapter 5.

acid. Addition of the fatty acids caused an increase in cellular cyclic AMP levels in the mycobacteria, and subsequent resuscitation (Shleeva et al., 2013). Shleeva et al. (2013) discovered that a *M. smegmatis* mutant lacking the gene encoding the fatty acid-activated adenylate cylase Msmeg_4279, was not resuscitated by fatty acids, but could be resusciatated with cyclic AMP. Addition of 8-bromo-cyclic AMP, an inhibitor of adenylate cyclase enzymes, also prevented resuscitation by fatty acids in wild-type *M. smegmatis* cells. The study provides support for the importance of cyclic AMP during resuscitation and suggests that the fatty acids cause resuscitation by increasing cyclic AMP, which subsequently results in increased RpfA expression and resuscitation.

1.5.2 Cyclic di-AMP

Cyclic di-AMP is a signalling molecule which has been implicated in the regulation of a wide range of cellular processes across species. It is one of the more recent nucleotide signalling molecules to be identified (Corrigan and Gründling, 2013).

1.5.2.1 Cyclic di-AMP synthesis and degredation

An overview of the synthesis and breakdown of cyclic di-AMP in *M. tuberculosis* is shown in Figure 1.6. Cyclic di-AMP is synthesised from ATP(or ADP) by diadenylate cyclase enzymes, such as DacA in *M. tuberculosis* (Bai et al., 2012) or DisA in *B. subtilis* (Zheng et al., 2013). Cyclic di-AMP is degraded by specific phosphodiesterase enzymes, such as Rv2837c from *M. tuberculosis* into pApA and then 5'AMP.

1.5.2.2 Cellular role

Cyclic di-AMP has thus far been implicated in playing a signalling role in a wide range of bacterial cellular functions. Functions that cyclic di-AMP has been demonstrated to be involved in include DNA integrity and the sensing of DNA damage, cell wall homeostasis, membrane lipid homeostasis, sporulation and associated processes, virulence and ion transport (Reviewed in Corrigan and Gründling, 2013). The ligand has been shown to bind protein target effectors, such as TetR transcription factors, but also regulatory RNAs, such as *ydaO*-type riboswitches (Nelson et al., 2013). The role of cyclic di-AMP in the regulation of ydaO-class ribowitches, in particular the one associated with the *rpfA* gene in *M. tuberculosis* will be explored in Chapter 5.

1.6 Metabolic state during dormancy and reactivation of *M. tuberculosis*

The harsh conditions in the macrophage and granuloma, which consist of nutrient starvation, low oxygen tension, and the presence of carbon monoxide and nitric oxide trigger actively growing *M. tuberculosis* into dormancy (Kumar et al., 2007). Dormant *M. tuberculosis* is able to switch to anaerobic metabolism by down regulating central metabolism and activating alternative metabolic pathways like the triacyl glycerol biosynthesis pathway, and the activation of the glyoxylate cycle (Chao and Rubin, 2010; Daniel et al., 2004). Production of stress proteins and the ultrastructural modifications of the *M. tuberculosis* cell wall exemplify the dormant state and are coordinated by the dormancy survival regulon, DosR (Voskuil et al., 2004). The DosR regulon controls approximately 50 genes. It governs the metabolic shift of *M. tuberculosis* from aerobic to anaerobic functioning; promotes survival of bacteria in hypoxia-induced dormancy and also controls the switch back to active replication upon reintroduction of oxygen in the environment (Leistikow et al., 2010; Rustad et al., 2009).

1.7 The Rpf family

The reactivation of dormant *M. tuberculosis* is achieved by re-establishing the replicative and metabolic activity. The process is known as resuscitation, and was first shown to be induced in *Micrococcous luteus* by resuscitation-promoting factor (Rpf) (Mukamolova et al., 2002; Young et al., 2010). Analysis of genomic sequences of Gram-positive bacteria in the *Actinobacteria* phylum later made it clear that the Rpf from *M. luteus* actually belongs to a large family of Rpf-like proteins and is found in genera of bacteria, including *Corynebacterium*, *Nocardia* and *Mycobacterium* (reviewed by Nikitushkin et al., 2011). Each species of bacteria has between one to five orthologs of *rpf*. *M. tuberculosis* has five *rpf* like genes, named *rpf* A-E (Kaprelyants et al., 2012).

1.7.1 Structure and function of Rpf proteins

Studies performed on the structure of Rpf proteins revealed that all of them possessed a c-type lysozyme-like domain, with glutamate residue in its active site (Cohen-Gonsaud et al., 2004). Alongside the catalytic domain, *M. tuberculosis* contains other intrinsic domains, which are: the Domains of Unknown Function (DUF); the G5 domain with 5 consecutive glycine residues; and domains important in binding certain proteins with the cytoplasmic membrane (reviewed by Nikitushkin et al., 2016). The G5 domain is a vital component for proteins involved in the degradation of cell wall and the formation of biofilms (Bateman et al., 2005). The G5 domain in RpfB has β-strands and assumes a β-TH- β motif structure and is adhesive to mycobacterial PG, with its bended shape proving to be beneficial for steric localisation of RpfB

to PG in cells (Nikitushkin et al., 2015). The DUF domains have recently been demonstrated to have one α-helix surrounded by four β-strands (Nikitushkin et al., 2015). Although the role of the DUF domain is not clear, its spatial organisation is similar to the eukaryotic protein ubiquitin. Ubiquitin interacts with many macromolecules in eukaryotes and the presence of a ubiquitin-like domain in Rpf allows it to interact with other proteins like RipA, an endopeptidase capable of cleaving peptide bonds within the PG peptide chain (Dasgupta et al., 2006; Hett and Rubin, 2008; Kaprelyants et al., 2012; Molle and Kremer, 2010). It acts as a partner to both RpfB and RpfE (Hett et al., 2007). RipA also plays a key role during the final stage of cell division (reviewed by Nikitushkin et al., 2016). Rpf proteins hydrolyse the 1→4 glycosidic bonds of the bacterial PG and together with RipA, lead to the synergistic hydrolysis of the mycobacterial cell wall (Hett et al., 2007). Likewise, addition of both RipA and Rpf led to syngergistic effects on the reactivation of dormant *M. smegmatis* cells (Nikitushkin et al., 2015). The combination of RipA and Rpf have also shown to result in the generation of low molecular weight products like ahydromuropeptides, analogs of which can stimulate the resuscitation of mycobacterial cells (Nikitushkin et al., 2015).

Other mechanisms that could take place under the influence of Rpf, such as the disintregration of mycobacterial aggregations (Nikitushkin et al., 2011) may well precede the ensuing production of muropeptides, stimulating resuscitation. In summary, Rpf proteins can directly stimulate cell division of mycobacteria and also stimulate production of muropeptides capable of inducing early stages of resuscitation (Nikitushkpin et al., 2013).

1.8   Scope of this study

The general aim of this study is to gain further insight into mechanisms of gene regulation in mycobacteria, with particular interest in *rpfA* regulation in *M. tuberculosis*. It is hoped that by contributing to the greater knowledge of gene regulation in the bacteria, potential new drug targets and ideas for therapeutics may arise. The following topics will be investigated in depth:

1) Characterising if CsoR, a copper-sensitive repressor protein, regulates *rpfA* and the effect of copper on the interaction.

2) Characterising two mycobacterial CRP proteins and their interaction with cyclic AMP and how this affects regulation of *rpfA*, *sdh1* and other genes.

3) Characterising the putative ydaO-type riboswitch located in the 5' UTR of *rpfA* messenger RNA.

A full introduction and specific aims are provided at the beginning of each chapter.

2   Materials and methods

2.1   Primers

Primer sequences for the entire study can be found in Table 2.1.

2.2   Growth media and conditions

2.2.1   Growth media

Escherichia coli, *Mycobacterium smegmatis* and *Bacillus subtilis* strains were typically grown in low salt Luria-Bertani (LB) medium, unless otherwise stated.

2.2.1.1   Luria Bertani (low-salt) medium

	LB broth (g/l)	LB agar (g/l)
Yeast extract	5	5
Sodium chloride	5	5
Tryptone	10	10
Agar bacteriological	-	15

The medium was sterilised by autoclaving at 121°C, 15 psi for 15 min.

For LB agar, the medium was cooled in a water bath to 50°C and appropriate antibiotic stock solutions were added. It was subsequently poured into petri dishes (25 ml aliquot per dish) under sterile conditions. Once set, the plates were dried in a sterile hood and were either used immediately or stored at 4°C for up to one month.

2.2.1.2   M9 minimal media

M9 media was used to wash, and subsequently for the induction period, cells in certain protein overproduction trials.

5X M9 salts:

	g/l
$Na_2HPO_4\text{-}7H_2O$	64
KH_2PO_4	15
NaCl	2.5
NH_4Cl	5

The 5X M9 salts were autoclaved at 121°C, 15 psi for 15 min and allowed to cool. To make 1 litre of M9 minimal media, 200 ml of M9 salts; 2 ml of filter-sterilised 1M $MgSO_4$; 20 ml of

Table 2.1: Table of primers used in this thesis

NAME	SEQUENCE (5'->3')	.	F/R
LD59	cgacccgggtccgcgcCGgtccagcgcgaccacg		F
LD60	cgtggtcgcgctggacCGgcgcggacccgggtcg		R
LD61	ccggctgcgccgcaccaGcaacaacctggccgacc		F
LD62	ggtcggccaggttgttgCtggtgcggcgcagccgg		R
LD63	cgggtccgcgcacgtGGagcgcgaccacgatcacc		F
LD64	ggtgatcgtggtcgcgctCCacgtgcgcggacccg		R
LD65	cgacccgggtccgcgcTGgtccagcgcgaccacg		F
LD66	cgtggtcgcgctggacCAgcgcggacccgggtcg		R
LS49	cccccagtcaccgc		F
LS50	ccgccaagaatcgg		R
LD27	gccgatagcacgccgGactccatccagtggg		F
LD28	cccactggatggagtCcggcgtgctatcggc		R
LD29	gccgatagcacgccgTactccatccagtggg		F
LD30	cccactggatggagtAcggcgtgctatcggc		R
LD31	ccgatagcacgccgaTctccatccagtgggc		F
LD32	gcccactggatggagAtcggcgtgctatcgg		R
LD33	ccgatagcacgccgaGctccatccagtgggc		F
LD34	gcccactggatggagCtcggcgtgctatcgg		R
LD35	cggctccttggggtgGagccaacgtcgttcgc		F
LD36	gcgaacgacgttggctCcaccccaaggagccg		R
LD37	cggctccttggggtgTagccaacgtcgttcgc		F
LD38	gcgaacgacgttggctAcaccccaaggagccg		R
LD39	ggctccttggggtgaGgccaacgtcgttcgcc		F
LD40	ggcgaacgacgttggcCtcaccccaaggagcc		R
LD41	ggctccttggggtgaTgccaacgtcgttcgcc		F
LD42	ggcgaacgacgttggcAtcaccccaaggagcc		R
LD43	GGCGAATCCAACCCGGACCCGACAGCTGACC		F
LD44	GGTCAGCTGTCGGGTCCGGGTTGGATTCGCC		R
LD45	GGCGAATCCAACCCGTACCCGACAGCTGACC		F
LD46	GGTCAGCTGTCGGGTACGGGTTGGATTCGCC		R
LD47	CAACCCGAACCCGACGGCTGACCTCGTAGGC		F
LD48	GCCTACGAGGTCAGCCGTCGGGTTCGGGTTG		R
LD49	CAACCCGAACCCGACTGCTGACCTCGTAGGC		F
LD50	GCCTACGAGGTCAGCAGTCGGGTTCGGGTTG		R
LD51	tgtacgaccgcggccCCgtgatcctctccacg		F
LD52	cgtggagaggatcacGGggccgcggtcgtac		R
Bio-LD67	BIO-GTCGCCAAGAATCGCCTTTA		F
LD68	GCCCAGGTGCTTTGAGTGAA		R
Bio-CsoR-F	BIO-CGCGGCCACCGACGGGGAATGGGATCAGGTGGCCCGCTGCGAGTCGGG		F
CsoR-R	CCCGACTCGCAGCGGGCCACCTGATCCCATTCCCCGTCGGTGGCCGCG		R
Bio-LD70	BIO-GTAGCCCACCCCCAGTGGGGTGGGATAC		F
LD71	GTAGCCCACCCCCAGTGGGGTGGGATAC		R
LD70	cgcgcgtgtggaccttccagcgccggcgtga		F
LD71	tcacgccggcgctggaaggtccacacgcgcg		R
LD72	cgcgtgtggaccgtccagcgccggcgtg		F
LD73	cacgccggcgctggacggtccacacgcg		R
Myc657	GAATTCGACGGGATTGTCGTCCGA		F
Myc658	GGATCCCACAATCGCCGGCAAGAA		R

filter-sterilised 20% (w/v) glucose and 100 µl of filter-sterilised 1M CaCl$_2$ were combined. This was adjusted to 1 litre by the addition of sterile distilled water.

2.2.1.3 NZY+ broth

NZY+ broth was used in the immediate recovery of transformants after heat shock during the site-directed mutagenesis protocol (Section 2.3.10).

NZY broth:

	(g/l)
NZ amine (casein hydrolysate)	10
Yeast extract	5
NaCl	5

This was made up to 1 litre with distilled water and the pH was adjusted to pH 7.5 with NaOH. This was sterilised by autoclaving at 121°C, 15 psi for 15 min. The following filter-sterilised supplements were added to the brother prior to use: 12.5 ml of 1M MgCl$_2$; 12.5 ml of 1M MgSO$_4$ and 20 ml of 20% (w/v) glucose.

2.2.2 Antibiotic supplementation of growth media

Antibiotics were added to growth media to allow for the selection of specific plasmids and to minimise contamination. The final concentration of antibiotics used to supplement media is listed below. Stocks of the different antibiotics were made and kept at -20°C for up to one month. Antibiotics stocks were diluted in sterile water, except for tetracycline, which was in 70% ethanol.

Antibiotic	Final (µg/ml)	Stock (mg/ml)
Ampicillin	100 – 200	200
Tetracycline	35	35
Kanamycin	20 – 50	20
Streptomycin	50	50

2.2.3 Growth of *Escherichia coli* strains

Escherichia coli was typically cultured in LB broth, supplemented with the appropriate antibiotics, at 37°C with shaking at 250 rpm until the desired OD$_{600}$ (often OD$_{600}$ of 0.6) was reached.

Overnight cultures of *E. coli* were prepared by the inoculation of a single colony in to 5 ml of LB broth, supplemented with the appropriate antibiotics. Inoculated cultures were grown at 37°C with shaking at 250 rpm for ~16 hours. Overnight cultures were either used directly (e.g. for plasmid mini prep) or as an inoculum for larger cultures.

2.2.4   Growth of *Mycobacterium smegmatis* strains

An overnight culture of LB (5 ml) was inoculated with 100 µl of *Mycobacterium smegmatis* Mc^{2}155 glycerol stock. The cultures were grown for 36-48 h at 37°C with 250 rpm shaking.

Aliquots (100 µl) of the liquid culture were plated on to standard LB agar plates and grown at 37°C until large enough single colonies had formed.

2.2.5   Storage of strains

Bacterial strains were stored at 4°C on agar plates for up to four weeks. To maintain strains for longer than four weeks, glycerol stocks were made. Glycerol stocks were made as follows. A 5 ml overnight culture of the strain was grown and subsequently pelleted by centrifugation (6,000 rpm, 10 mins at 4°C). The pellet was resuspended in 1 ml of ice-chilled LB broth, 1 ml sterile ice-chilled 80% glycerol and supplemented with the relevant antibiotic where appropriate. These were stored at -20°C for use at a later date.

2.2.6   Production of electrically competent cells

A 5 ml overnight culture of cells was prepared, typically a BL21 or DH5α strain. The overnight culture was added to 200 ml of LB broth and shaken at 250 rpm at 37°C until an OD_{600} of 0.4 to 0.6 was reached. The culture was split into 50 ml aliquots and chilled on ice for 15 – 20 min. Aliquots were centrifuged at 6,000 rpm for 10 min at 4°C and the supernatants were discarded. The pellets were resuspended in 50 ml of ice-chilled, sterile water and subsequently centrifuged as before. The supernatants were discarded and the pellets were resuspended in 25 ml of ice-chilled sterile water, and were then centrifuged as before. The pellets were resuspended and combined in a total of 2 ml of 10% glycerol (ice-chilled, sterile). A final centrifugation step of 6,000 rpm for 10 min at 4°C was performed, after which the supernatant was discarded. The pellets were resuspended in 200 µl of 10% glycerol (ice-chilled, sterile). This was divided in to 50 µl aliquots and stored at -80°C until required.

2.2.7   Transformation of electrically competent cells

Aliquots (50 µl) of electrically competent cells were removed from storage at -80°C and thawed gently on ice. A volume of 2 to 5 µl of the desired plasmid (typical concentration of 50

to 100 ng/µl) was added to the cell aliquot and pipetted up and down to mix. A control reaction containing dH₂O instead of plasmid was also set up. These reactions were left on ice for 20 to 30 min. The cells were transferred to pre-chilled electroporation cuvettes and were electroporated at 1,800 V. Immediately following electroporation, 1 ml LB broth was added to the cells and mixed thoroughly by pipetting. The cells were transferred to sterile Eppendorf tubes and shaken for 1 hour at 250 rpm and 37°C. A 10 µl sample from each reaction was plated onto ampicillin-containing LB agar plates. The remaining sample was pelleted by centrifugation (13,000 rpm for 10 min). The pellet was resuspended in a small volume (typically 20 µl) of LB broth and was then plated onto ampicillin-containing LB agar. The agar plates were grown at 37°C overnight.

2.3 Nucleic methods

2.3.1 Storage of DNA

DNA was typically stored in nuclease free water at -20°C.

2.3.2 Plasmid purification

Plasmid DNA was purified from overnight cultures (5 ml) using a Qiagen QIAprep spin miniprep kit, following the standard manufacturers protocol. DNA was typically eluted in nuclease free water.

2.3.3 Digestion of DNA using restriction endonucleases

Plasmid DNA was linearised or cut into fragments using the appropriate restriction endonuclease enzyme(s) – see plasmid table in Table 2.1. Cut DNA was subsequently electrophoresed on an agarose gel.

For a single digest, a typical reaction was set up as follows:

1 µl   Relevant buffer* (10X)

8 µl   DNA

1 µl   Enzyme (10 units)

This was incubated in a 37°C water bath for 1 h.

For a double digest, a typical reaction was set up as follows:

2 µl   Relevant compatible buffer* (10X)

16 µl   DNA

1 µl   Enzyme #1

1 µl   Enzyme #2

This was incubated in a 37°C water bath for 2 h.

Relevant compatible buffers were selected using the Promega Restriction Enzyme Tool, located at http://www.promega.com/a/apps/reTool/

2.3.4 Polymerase Chain Reaction (PCR)

PCR was typically carried out as follows:

25 µl 2X Entensor Hi-Fidelity PCR Master Mix (Thermo Scientific)

1 µl Primer #1 (1 µM final)

1 µl Primer #2 (1 µM final)

x µl DNA template*

up to 50 µl with nuclease free water

If genomic DNA was used as a template then 50-150 ng was used. If a simple template, such as a previous PCR product, was being used then ~10 ng of template was used.

Following this samples were placed in a PCR machine and the following program was performed.

Initial denaturation at 94°C for 2 minutes.

35 cycles of:

Step	Temperature	Time
Denaturing	94°C	30 sec
Annealing	56°C	30 sec
Extension	68°C	40 sec

Final extension 68°C for 4 minutes.

For PCR products > 1kb in size, the extension time was increased to 1 min/kb. If the melting temperature of certain primers were significantly below 56°C, the annealing temperature was lowered until a suitable product could be obtained. If this caused other non-specific products to additionally be formed, the desired product was gel extracted and sequenced to check it was the desired product. This was subsequently used as the template DNA for future PCRs.

2.3.5 PCR purification

PCR clean-up was performed on the remainder of the reaction with Qiagen's QIAquick PCR purification kit using the standard manufacturer's protocol.

2.3.6 Radioactive PCR

Radioactive PCR reactions were set up as follows: 22.5 µl PCR mastermix (1.1x); 2 µl genomic template DNA; 0.5 µl primer 1 (100 pmol/µl); 0.5 µl primer 2 (100 pmol/µl) and 1 µl radiolabelled α-[^{32}P]dCTP (2 µCi/µl, 3000 Ci/mmol^{-1} supplied by Perkin Elmer). The PCR mastermix used was Thermo Scientific 1.1X ReddyMix™ PCR Master Mix with 1.5 mM MgCl$_2$. The addition of the radioactive material was performed on a designated radiation bench. Primers were diluted before addition to a concentration of 100 pmole/µl in buffer EB (10 mM Tris-HCl, pH 8.5). PCR was carried out typically using 25 cycles of:

Temperature	Time
96°C	30 sec
58°C	30 sec
72°C	60 sec

This was followed by 4 min at 72°C. PCR clean-up was performed the PCR products with Qiagen's QIAquick PCR purification kit using the standard manufacturer's protocol. Radiolabelled DNA was transferred within the laboratory in lead pots and stored at -20°C in the radioactive freezer.

2.3.7   InstaGene™ extraction of Mycobacterial and *Bacillus subtilis* genomic DNA

This method was used for DNA extraction from *Mycobacterium smegmatis* cells. A colony from a plate (or a pellet from 100 µl mid-log phase liquid culture) was resuspended in 200 µl InstaGene™ Matrix (BioRad) and incubated at 56°C for 30 min to allow the matrix to bind to the components in the preparation that may inhibit PCR. Samples were boiled at 100°C in a heat block for 8 min to kill the *M. smegmatis* and release the genomic DNA. Samples from this reaction mixture (typically 1 – 5 µl) were used in future PCR reactions as genomic DNA.

2.3.8   Agarose gel electrophoresis

Agarose gel electrophoresis was used for the visualisation of DNA fragments. Agarose (1% (w/v)) was added to 1X TAE buffer and dissolved into solution by heating in a microwave (320 W). The gel was cooled to 50°C before the addition of GelRed™ solution (1:10,000 dilution), from Biotium, to the mix before setting. Before loading the DNA, Bioline 6X loading dye was added. Typically, 1 kb Generuler™ ladder (Thermo Scientific) or HyperLadder™ 1 kb (Bioline) was loaded in the first lane of the gel to allow for size calibration of the DNA fragments. Gels placed in a tank and were electrophoresed in 1X TAE buffer at 100 V for 1 hour. They were subsequently visualised using a UVItech photodocumentation system and images were printed.

50X TAE buffer:

Tris base	242 g
Acetic acid (glacial)	57.1 ml
0.5M EDTA (pH 8.0)	100 ml

made up to 1 litre with dH_2O.

2.3.9 Gel extraction from agarose gels

Agarose gels were placed on a UV-transillumitator in a dark room and exposed to UV light (254 nm). The separated DNA fragment of desired size was neatly and closely excised using a scalpel. DNA fragments were purified from the gel matrix using a QIAquick Gel Extraction Kit, supplied by Qiagen, following the standard manufacturer's protocol.

2.3.10 Site-directed mutagenesis

Primers were designed, that would incorporate the desired mutation in the DNA, according to the advice provided in the QuikChange II XL site-directed mutagenesis kit (Agilent). Site-directed mutagenesis experiments were then performed following the standard manufacturer's instructions.

2.3.11 Sequencing

DNA sequencing was carried out by GATC Biotech on a 96 capillary ABI 3730xl DNA analyser system. The DNA to be sequenced and a relevant primer were sent at the following concentrations.

Final concentration

Purified plasmid DNA	40 – 50 ng/µl
Purified PCR product	10 – 40 ng/µl
Primer	2.5 pmol/µl

2.3.12 Determination of nucleic acid concentration

The DNA or RNA concentration in a sample was determined using a Nanodrop ND-1000 (Thermo Scientific). The machine was blanked (using nuclease free water or the relevant buffer) and the absorbance at 260 nm was recorded.

The sequence, type of nucleic acid and absorbance at 260 nm were entered in to the OligoCalc tool from Northwestern. This provided the concentration of nucleic acid present in the sample. The tool can be found at: www.basic.northwestern.edu/biotools/oligocalc.html

Purity of RNA could be assessed using the ratio of absorbance at 260 nm to 230 nm. RNA was deemed to be pure if 260:230 was between 1.98 and 2.

2.3.13 In vitro transcription for generation of RNA

Typically HiScribe™ High Yield Transcription Kit (New England Biolabs) was used for generation of riboswitch RNA by *in vitro* transcription, following the standard manufacturer's instructions.

2.3.14 Denaturing PAGE analysis of RNA

An acrylamide gel (6 to 8%) containing 1X TBE and 8M urea was poured and set. After setting the gel was prerun (200V) in 1X TBE buffer for at least 20 min. Gels were flushed using a syringe prior to loading the RNA sample.

2X gel-loading buffer (18M Urea, 20% (w/v) sucrose), 0.1% (w/v) SDS, 0.05% (w/v) bromophenol blue, 0.05% xylene cyanol, 90 mM Tris-HCl, 90 mM borate and 1 mM EDTA (pH 8.0 at 25°C)) was added to the RNA sample to be analysed by denaturing PAGE. The sample was loaded into the relevant lane(s) of the gel and the gel was run at 200 V until the bromophenol blue had migrated to approximately ¾ of the gel length.

2.3.15 Crush/soak buffer for removal of RNA from denaturing PAGE-gels

After denaturing PAGE has been carried out, the glass plates were separated and the gel was transferred to RNaseZap-treated clingfilm. The clingfilm was placed over a fluorescently coated thin-layer chromatography (TLC) plate (supplied by Applied Biosystems) in a darkened room and UV light (254 nm) was applied using a hand-held lamp. RNA bands were visible as dark shadows in the gel and the location of bromophenol blue and xylene cyanol in the loading dye was also visible. The migration of these dyes from the loading dye was used to check the size of the IVT product was roughly the size expected.

The RNA band of interested was excised from the gel using a scalpel (RNase free). The band was chopped into smaller cubes and transferred to an Eppendorf tube (1.5 ml). Crush/soak buffer (500 µl) was added which covered all the gel cubes. The tube was incubated at room temperature for a minimum of 30 min. The crush/soak buffer was removed using a clean pipette tip and transferred to a new tube, where ethanol precipitation of the RNA was performed.

Crush/soak buffer consisted of: 200 mM NaCl, 10 mM Tris–HCl (pH 7.5 at 25°C), and 1 mM EDTA (pH 8.0 at 25°C).

2.3.16 Phenol/chloroform extraction

Remove the CIAP or other enzymes from nucleic acid preps was achieved by phenol/chloroform extraction as follows. Equal volumes of Tris-buffered phenol and CHISAM (chloroform- isoamyl alcohol 24:1) were combined. Phenol/CHISAM of equal volume to the sample was added and the mixture was vortexed. The phases were separated by brief (5 s) centrifugation and the aqueous phase was transferred to a new microfuge tube. An equal volume of CHISAM was added to the aqueous phase and the mixture was vortexed. The layers were separated by brief (5 s) centrifugation and then the aqueous phase was transferred to a new microfuge tube for subsequent experiments.

2.3.17 Ethanol precipitation of nucleic acids

Ethanol precipitation was performed on nucleic acid samples as follows. 0.1X volume of sodium acetate (3M, pH 5.0) was added to a sample and mixed. For RNA samples, glycogen (final concentration 0.2 mg/ml) was added to the sample to act as a carrier. 2.5 to 3X volume of ice-chilled ethanol was added to samples. Samples were left on ice for a minimum of 30 min. Nucleic acids were pelleted by centrifugation (16,800 xg, 20 min at 4 °C).

2.3.18 Spin column purification of RNA

Riboswitch RNA that had been synthesised by IVT could be purified by spin column purification. MEGAClear™ Transcription Clean-Up Kit by Ambion was used, following the standard manufacturer's instructions.

2.3.19 Radiolabelling of RNA

KinaseMax™ 5' End-Labeling Kit (by Ambion) was used for 5' end-labelling of RNA. The standard manufacturer's protocol was followed.

2.3.20 In-line probing

In-line probing assays were initially performed by the methods outlined in Regulski and Breaker (2008). This was optimised throughout the process and the following adaptions were made. A 'Low salt' ILP buffer was used that contained a 1X final concentration of 50 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$ and 10 mM KCl was used as the buffer for reactions and typically ~1 pmol of radiolabelled riboswitch RNA was incubated with ligand(s) at a specified concentration. For B. subtilis samples were incubated at the specified times, often 36 h for *B. subtilis* and 14 h for *M. tuberculosis* samples. A 2X TBE buffering systems was used during

resolution of the species, compared to the 1X TBE proposed in Regulski and Breaker (2008), as this provided better resolution.

2.3.21 RNA gel shifts

Gel shift reactions were set up that contained riboswitch RNA and the ligand of interest in a final concentration of 1X refolding buffer (10 mM sodium cacodylate (pH 6.8) 10 mM potassium chloride and 10 mM magnesium chloride).

Samples underwent refolding and were incubated for the relevant amount of time before analysis via native PAGE on a gel containing typically 8% acrylamide. 1x Native PAGE buffer = 100 mM Tris/HEPES (pH 7.5); 10mM $MgCl_2$ and 0.1mM EDTA All buffers and tanks were cooled to 4°C before and during running. Gels were run at 125V until the bromophenol blue approached near the bottom of the gel before imaging on a UV transilluminator photodocumentation system.

2.3.22 Microscale thermophoresis with RNA

Reactions were set up with a constant concentration of F-c-di-AMP and differing concentrations of RNA in 1X refolding buffer (10 mM sodium cacodylate (pH 6.8) 10 mM potassium chloride and 10 mM magnesium chloride). Samples underwent refolding and once cooled were transferred to standard MST capillaries. MST was performed on a Monolith NT.115 machine by NanoTemper following the standard guide for experiments. The blue LED was used with power typically at 40% with MST power between 30-70% with the following settings:

Fluo. before: 5 s
MST on: 30 s
Fluo. after: 5 s

Data was analysed using NT analyse software, provided by NanoTemper.

2.3.23 Refolding of riboswitch RNA

RNA samples were refolded by heating to > 70°C for 3 min and slow cooling to room temperature. It could be achieved in a standard thermal cycling machine as follows: 90°C for 3 min,, 70°C for 1 min, 60°C for 1 min, 50°C for 1 min, 45°C for 1 min, 37°C for 1 min and leaving to equilibrate to room temperature.

2.4   Protein methods
2.4.1   SDS-PAGE

SDS-PAGE running buffer:

Glycine	14.5 g
SDS	1 g
Tris	3 g

made up to 1 litre with dH_2O

Coomassie stain and destain:

	Stain (500 ml)	Destain (1 litre)
Coomassie R250	0.575 g	-
Methanol	200 ml	400 ml
Acetic acid	50 ml	100 ml
dH_2O	250 ml	500 ml

Gel drying solution:

Methanol	300 ml
Glycerol	50 ml
dH_2O	650 ml

SDS-loading buffer (2x):

Glycerol	20% (v/v)
Tris-HCl	100 mM
SDS	4% (w/v)
Bromophenol blue	0.2% (w/v)
β-mercaptoethanol	200 mM

Acrylamide gels:

	15% resolving gel	5% stacking gel
30% (w/v) Acrylamide	3.8 ml	1.7 ml
3M Tris-HCl (pH 8.3)	0.95 ml	-
0.5M Tris-HCl (pH 6.8)	-	1.25 ml
dH_2O	2.7 ml	6.95 ml
10% (w/v) SDS	75 µl	100 µl
10% (w/v) APS	100 µl	100 µl
TEMED	20 µl	20 µl

Glass plates were assembled in the spacers and inserted in the gel clamps. The resolving gel (15% acrylamide (w/v)) was poured between the plates leaving a gap with a width slightly over one comb. Isopropanol was poured on top of the gel and the gel was left to set. After setting, the isopropanol was thoroughly rinsed off using dH$_2$O and a stacking gel was made. The stacking gel was poured, the comb was inserted and the gel was left to set, typically for 8-10 min. The gel was either stored or placed in a gel tank in 1X SDS running buffer. SDS-loading buffer was added (in a 1:1 ratio) to protein samples, which were subsequently denatured by boiling at 95°C for 10 min. Typically 5 – 10 µl of Precision Plus Protein™ All Blue Standards (from Bio-Rad) was loaded in the first well to allow for size calibration and the other samples were loaded in the other wells. The gel was electrophoresed at 200 V for 1 hour, approximately until the loading dye reached the base of the gel.

For storage, SDS gels were wrapped in tissue (moistened with dH$_2$O) and wrapped in clingfilm. These were stored at 4°C and used within 5 days.

2.4.2 Protein concentration determination

The concentration of protein samples was calculated using either the Bradford assay, or spectroscopically using the extinction coefficient of the protein of interest.

2.4.2.1 Bradford assay for protein concentration

Protein concentration within samples was calculated using a standard Bradford assay, using Bio-Rad Protein Assay Dye Reagent Concentrate. The concentrate was diluted (1:5) before use. Varying volumes of protein solution, typically 1 µl, 5 µl or 10 µl, was added to 1 ml of diluted Bio-Rad reagent in a cuvette. Parafilm was applied over the cuvette and the cuvette was inverted 3 times to allow for mixing. The A$_{595}$ of the solution was determined using a Unicam HEλIOS spectrophotometer. This value was used to determine the protein concentration using a bovine serum albumin (BSA) standard curve. If the protein concentration was too high to provide an A$_{595}$ reading in a suitable range, a small amount of the original protein sample was diluted (typically 1:10) and the protein concentration was calculated from this.

2.4.2.2 Protein concentration using extinction coefficient

Protein concentrations were calculated using the extinction coefficient of the relevant protein, which was calculated by the ExPASy ProtParam tool.

Concentration of protein (M) = Extinction coefficient/(A$_{280}$ – A$_{320}$)

For CsoR an extinction coefficient of 1,615 (assuming all pairs of Cys form cysteines) or 1,490 (assuming all Cys reduced) were used.

For CRP proteins an extinction coefficient of 12,490 was used.

2.4.3 Concentration and buffer exchange of protein samples

Buffer exchange was performed in order to lower the salt concentration of the buffer the protein was present in. After elution from the AKTA™ prime, the protein was in elution buffer (0.5 M NaCl and 20 mM sodium phosphate, pH 7.4). AKTA™ fractions containing pure protein we combined (max volume of 6 ml) and applied to a Vivaspin 6 column (GE Healthcare) with MWCO of 10,000 Da. If the sample was smaller than the maximum volume (6 ml) it was diluted up to the maximum value before the first centrifugation step. The column was centrifuged for 20 min cycles at 6000 rpm at 7°C, until only a small volume remained in the Vivaspin sample chamber. The filtrate was removed and discarded from the column. Fresh buffer (100 mM NaCl, 20 mM sodium phosphate, pH 7.4) was applied to the column to a total volume of 6 ml. The centrifugation cycles were repeated. The filtrate was discarded and the concentrated, desalted protein sample was recovered.

2.4.4 Overproduction of CRPs

CRP proteins were usually overproduced by inoculating 5 ml overnight cultures of the host stain in a 2 l conical flask containing 500 ml LB broth, which was supplemented with the appropriate antibiotic (typically 200 µg/ml ampicillin). The cultures were grown at 37°C with shaking (250 rpm) under aerobic conditions until OD_{600} of 0.4 – 0.6 were achieved (typically 3 h). Cells were induced with a final concentration of 30 µg/ml IPTG and were placed aerobically at 37°C with shaking (250 rpm) for 2 to 3 h. Cells were subsequently centrifuged at 8,000 rpm at 4°C for 30 min to form pellets. These pellets were either stored overnight by freezing at -20°C or were resuspended in binding buffer (20 mM sodium phosphate and 0.5 M NaCl, pH 7.4) and used immediately to generate cell-free extracts.

2.4.5 Overproduction of CsoR

2.4.5.1 LB version

Initially, CsoR was overproduced in 500 mL LB broth inoculated with 5 mL overnight cultures (1% inoculum) of the BL21 λDE3 host strain containing pGS2628 in a 2 L conical flask. The cultures were supplemented with 200 mg/L ampicillin and grown at 37°C with shaking at 250 rpm under aerobic conditions until OD_{600} of 0.4-0.6 was reached (typically 3-4 h). Expression was induced by addition of IPTG to a final concentration of 120 µg/mL and allowed to continue for 2-3 h at 37°C with shaking at 250 rpm under aerobic conditions. Cultures were subsequently pooled and centrifuged at 8,000 rpm at 4°C for 20 min. The supernatant was discarded and the resulting cell pellets were stored at 20°C.

2.4.5.2 M9 minimal media

CsoR was expressed under copper-deficient conditions to keep the protein in its apo-form. Following growth to OD_{600} 0.4-0.6 in 500 mL LB broth, as above, cultures were transferred to sterile centrifuge pots and were centrifuged at 8,000 rpm at 4°C for 15 min. Cell pellets were washed twice in M9 minimal media (200 ml). Under sterile conditions, cells were resuspended in 500 ml M9 minimal media and were transferred to a sterile 2 L conical flask. Expression of CsoR was induced with IPTG (120 µg/ml) and cells were placed at 25°C with shaking at 250 rpm for 18 h. Cells were collected by centrifugation at 8,000 rpm at 4°C for 20 min, the supernatant was discarded and cells were stored at -20°C and were resuspended in Buffer A.

2.4.6 Production of cell-free extracts

Cells were thawed and resuspended in a total volume of typically 10 to 20 ml of the relevant buffer that would be used in the next purification step. Resuspended cells underwent cell lysis via passage through a French press at 16,000 psi. Cells were passed through the French press either 2 to 3 times. Separation of the soluble cell extract and the insoluble cell pellet was achieved by centrifugation at 15,000 rpm at 4°C for 30 min. Insoluble and soluble fractions were stored at -20°C until required.

2.4.7 Purification of CRPs

Overproduced proteins were purified from the soluble cell extract using their N-terminal histidine tag and nickel affinity chromatography. The purification step was performed on the AKTA™ prime machine (GE Healthcare) using the His-tag purification programme. The standard manufacturer's protocol for the His-tag purification program was followed and the machine was configured according to manufacturer's advice. Cell extract was removed from storage (at -20°C) was centrifuged (6,000 rpm, 4°C for 15 min). The supernatant was passed through a 0.45 µm sterile membrane filter and 2.5 ml to 10 ml was applied to the AKTA™ prime sample loop using a syringe. A 1 ml HiTrap™ chelating column was used for the purification process. Fractions, of volume 1 ml were collected, and a graph of the absorbance at 280 nm was produced. The initial peak on the 280 nm line corresponded to non-specifically binding proteins, so the fractions which corresponded to the second peak (after the addition of the elution buffer) were collected and resolved on SDS-PAGE gel to check for the presence of purified protein. The fractions were stored at 4°C until needed.

Buffers for His-tag purification

Elution buffer (pH 7.4)   Binding buffer (pH 7.4)

20 mM sodium phosphate   20 mM sodium phosphate

0.5 M NaCl   0.5 M NaCl

0.5 M imidazole

Ni-loading element

100 mM NiSO$_4$

2.4.8   DisA overproduction and purification

Overnight cultures (5 ml) of the DisA overexpression strain were set up and used to inoculate (1%) 2L conical flasks containing 500 ml low-salt LB broth. Cultures were supplemented with the relevant antibiotic and grown at 37°C with shaking (250 rpm) until an OD$_{600}$ of ~0.6 was achieved. Cultures were induced with 120 µg/ml IPTG and were placed at 25°C with shaking (250 rpm) for 3 hours. Cells were collected by centrifugation (8,000 rpm, 20 min at 4 °C) and were stored at -20°C until required.

Cells were thawed and resuspended in elution buffer (10 ml). Cell free extract was produced and filtered through a 0.45 µm filter. DisA was purified from the cell free extract using the His-tag purification method on the AKTA Prime with the following buffers. Buffers for His-tag purification

Elution buffer (pH 7.4)   Binding buffer (pH 7.4)

20 mM sodium phosphate   20 mM sodium phosphate

0.5 M NaCl   0.5 M NaCl

0.5 M imidazole

Ni-loading element

100 mM NiSO$_4$

Fractions were analysed by SDS-PAGE, and those containing sufficiently pure DisA were combined and concentrated using a 5 ml Vivaspin column (MWCO 10kDa). The protein was dialysed into 'c-di-AMP synthesis buffer' containing 5% glycerol, 50 mM Tris (pH 7.5), 150 mM NaCl and 10 mM MgCl$_2$.

2.4.9   Purification of CsoR
2.4.9.1   PEI precipitation

Cell free extract was kept on ice and precipitated by with PEI (pH 5.8, final concentration 0.15 % v/v). The sample was subsequently centrifuged at 18,000 rpm, 30 min at 4°C. The supernatant was transferred to a fresh tube for further purification.

2.4.9.2 Ammonium sulphate precipitation

The supernatant from the PEI precipitation was subjected to ammonium sulphate precipitation (ASP). Optimal percentages (40-50%) for an ASP cut of CsoR were ascertained by testing different percentages by the methods in outlined in Burgess (2009). The 40-50 % cut was found to be optimal for CsoR purification. Crystalline ammonium sulphate was ground to a fine powder for more effective mixing. Initially, 291 mg/ml of sample was added slowly whilst mixing. The sample was then incubated for 30 min on ice and subsequently centrifuged at 10000 xg for 10 min at 4°C in a precooled centrifuge. The pellet was discarded and 57 mg powder was mixed into the remaining 0.95 ml supernatant. The sample was incubated on ice for 30 min before being centrifuged at 10000 xg for 10 min at 4°C in a precooled centrifuge. The supernatant was discarded and the pellet was kept and resuspended in Buffer A containing 50 mM NaCl before being purified by affinity chromatography.

2.4.9.3 Affinity chromatography

Overexpressed proteins were purified from the soluble cell free extract, or following salt precipitations using affinity chromatography on the ÄKTA Prime machine (GE Healthcare). The purification step was carried out using the Anion Exchange programme on a 1 ml HiTrap™ Heparin HP column. If using cell free extract, it was passed through a 0.45 μm filterbefore loading. If using precipitants from the ASP cut step, the pellet was resuspended in 5 ml Buffer A with 50 mM NaCl and dialysed overnight in 1 L Buffer A with 50 mM NaCl at 4°C to remove residual ammonium sulphate. The dialysed sample was loaded into the ÄKTA sample loop. Unless otherwise stated elution was achieved with a linear gradient of 50 mM – 750 mM NaCl in Buffer A. Fractions were collected and those corresponding to the relevant peak (280 nm) during the salt gradient were analysed by SDS-PAGE. Fractions containing the relevant samples were concentrated in Vivaspin column (5 ml, 10,000 MWCO), dialysed into Buffer A containing 50 mM NaCl and were stored at 4°C.

Buffer A (pH 5.8) contained 25 mM MES; 3 mM DTT; 1 mM EDTA; 1 mM Benzamidine.

2.4.9.4 Gel filtration standard curve

Gel filtration was performed by loading protein samples on a 1.6x60 HiLoad Superdex200 column. Gel filtration was performed using a flow rate of 1.5 ml/min in 50 mM Tris (pH 8.0) containing 0.5 M NaCl (Figure 3.8c). Fractions (2 ml) were collected and analysed by SDS-PAGE In order to be able to determine the oligometic state of purified protein a calibration curve was created. A set of proteins of known molecular weights were loaded and eluted from the the column. The V_o for the column was 41 ml and the V_t was 116 ml. The K_{av} was calculated for the known proteins and plotted against the $\log_{10}$ of their molecular weights (figure 2.1).

HiLoad Superdex200 calibration curve
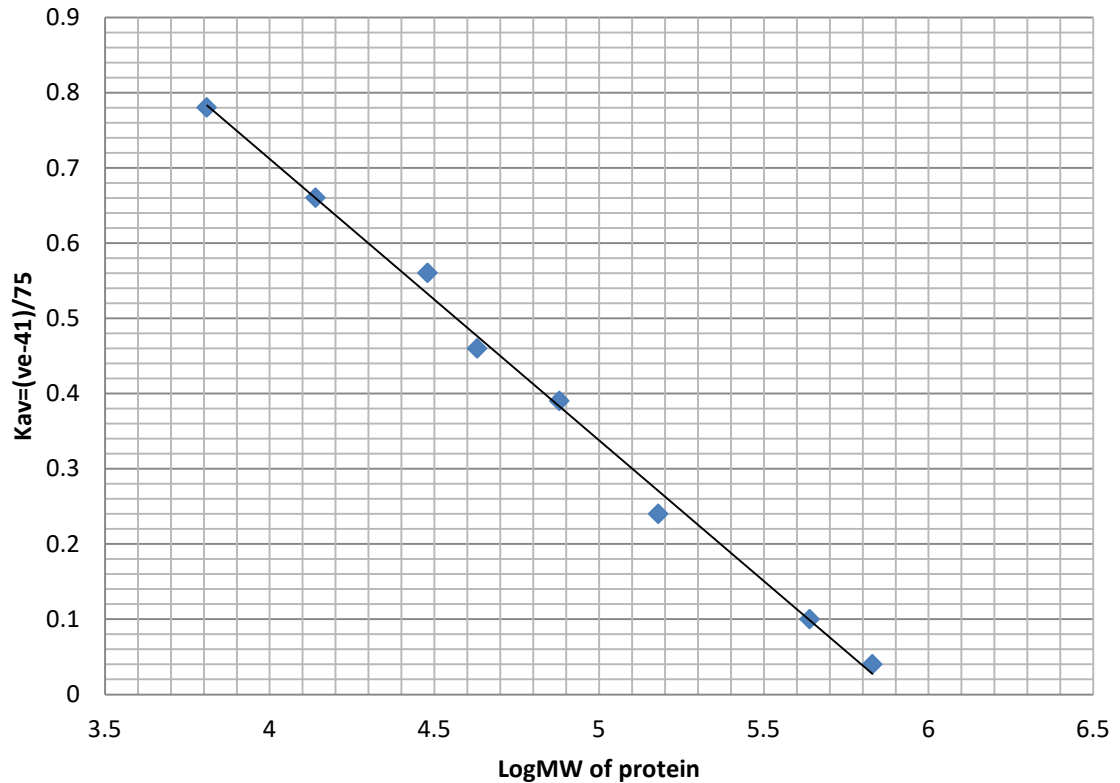
Figure 2.1: Calibration curve of protein standards for gel filtration. The HiLoad Superdex 200 column was calibrated by determining the elution volumes of the proteins of known sizes (aproteinin, ribonuclease, carbonin anhydrase, ovalbumin, conalbumin, aldolase, ferritin and thyroglobulin). V_o= 41 ml, V_t= 116 ml. The K_{av} was calculated for each protein and plotted against the Log_{10} of the molecular weight of the corresponding protein.

2.4.10 Cyclic di-AMP synthesis using purified DisA

Cyclic di-AMP synthesis reactions were carried out with DisA, by the methods outlined in Zheng et al. (2013). ATP stock (25 μM) was from Perkin Elmer and reactions were incubated at 45°C overnight.

2.4.11 Isothermal titration calorimetry

Isothermal titration calorimetry was performed on a Nano ITC by TA Instruments. Purified protein (300 μl) typically at 200-500 μM concentration was loaded in the sample chamber. Ligand was present in the syringe, typically at concentrations 5-10-fold higher than the protein. A set of 25 injections (1.97 μl) were performed per run at 25°C, with typically 240 s between injections, or sufficient time for the trace to return to baseline before the next injection. Ligand(s) were diluted in the dialysis buffer used to dialyse the protein (20 mM sodium phosphate + salt) in order to minimise any effects caused by buffer mismatch.

2.4.12 Differential scanning fluorimetry

Reaction (50 μl) for DSF were set up that contained the protein of interest (100 μM) and the ligand of interest (1 mM) in the protein buffer (typically 20 mM sodium phosphate pH 7.4 with 200 mM NaCl) and 10X SYPRO orange dye (supplied from manufacturers at 5000X). Samples were placed in RT-PCR 96-well plates and a program was ran that increased the temperature of the samples by 1°C/min over a range from 25-100°C. The fluorescence was measured throughout and the first derivative curves were analysed to determine the T_m for the protein with and without the ligand of interest.

2.4.13 Mass spectrometry on protein samples

Protein mass spectrometry was carried out by Dr Simon Thorpe from the Department of Chemistry, CCIAS, the University of Sheffield. The sample was examined by MALDI-ToF spectrometry using a Bruker Reflex III machine.

2.5 DNA:Protein methods
2.5.1 Electrophoretic mobility shift assays (EMSAs)

2.5.1.1 EMSAs with radiolabelled DNA

Several different concentrations of acrylamide and buffer combinations were tested during the project to try to find a suitable method, but typically 10% acrylamide (w/v) and 0.5X TBE buffer gels were used as the optimal method for EMSA experiments.

10% acrylamide, 0.5X TBE buffer gel:

Acrylamide (30% w/v)	3.3 ml
10X TBE	0.5 ml
dH$_2$O	6.1 ml
APS (10% w/v)	100 µl
TEMED	20 µl

10X TBE buffer:

890 mM	Tris-Base
890 mM	Boric acid
20 mM	EDTA (pH 8.0)

To make the acrylamide gels, clean glass plates were placed in plate holder and assembled in to the clamps. The gel was poured between the plates to the top and the comb was inserted. The gel was left to set, typically 10 min, before placing in to a tank containing 0.5X TBE buffer. The gel was pre-run at 30 mA for 20 to 30 min to allow for the removal of any un-polymerised acrylamide ions before the assay samples were loaded, as this could disrupt potential DNA:protein interactions.

EMSAs with CRP proteins were carried out in 1X EMSA buffer containing: 100 mM NaCl, 50 mM Tris (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 1 mM DTT and 0.25 µg/µl BSA, before the addition of radiolabelled DNA (~3 ng) for a further 10 min (total reaction volume of 15 µl). Calf thymus DNA (~1 µg) was used as a non-specific competitor in all assays. After incubation, loading buffer (2 µl of 50% glycerol, 0.25% bromophenol blue) was added to the samples immediately before applying to 6 or 10% polyacrylamide gels buffered with 0.5X TBE. Electrophoresis was carried out at 30 mA for ~30 min with 0.5XTBE running buffer. The gels were then dried and visualized by autoradiography.

2.5.1.2 EMSAs with biotinylated DNA

EMSAs were carried out with biotinylated DNA and different concentrations of the protein and ligand(s) of interest using a LightShift Chemiluninescent EMSA Kit (Thermo Fisher Scientific) following the standard manufacturers protocol.

2.5.2 BLItz analysis of protein:DNA interactions

Bio-layer Interferometry was carried out on the BLItz (by ForteBio) machine in the Medical School, The University of Sheffield. DNA was amplified with 5'-end biotinylated forward primers and diluted to a final concentration of 100 nM. Binding reactions were carried out

using Streptavidin coated probes that were able to immobilise the biotinylated target DNA. Protein samples were diluted to a range of concentrations. The 'Advanced kinetics' program was used with the following steps:

Initial Baseline	60 s	Tube
Loading DNA	240 s	Drop holder
Baseline	60 s	Tube
Association	240 s	Drop holder
Dissociation	120 s	Tube

Association rate (k_a) and dissociation rate (k_d) constants were calculated using the 'Advanced Kinetics' analysis program on BLItz Pro GxP software by ForteBio using the 'global fit' option.

2.5.3 Slot blotting

Slot blotting was performed on reactions that were set up identically to those in Section 2.5.1.1, but scaled up larger. The slot blot was assembled as shown in Figure 4.12. 1X EMSA buffer was used to moisten the blotting paper and Hybond N and C membranes. A vacuum was attached and samples (20 µl) were applied to each well in triplicates. After the sample had passed through the membranes, the wells were washed three times with 0.5 ml 1X EMSA buffer, which was allowed to pass through the membranes. The membranes were air dried and subsequently exposed to x-ray film and developed.

3 Characterising the regulation of *rpfA* by the copper-sensing protein CsoR

3.1 Introduction

3.1.1 Copper and *M. tuberculosis*

The majority of bacteria that make it into the body are killed after phagocytosis. Despite this, the success of *M. tuberculosis* is largely due to its ability to interfere with host signalling pathways and ultimately prevent phagosome maturation (Sturgill-Koszycki et al., 1995). However, the mycobacterial phagosome contains many sources of stress, including limiting levels of oxygen, iron and other nutrients and high levels of reactive nitrogen intermediates (Schnappinger et al., 2003).

Copper is an essential micronutrient for microorganisms. However, it is toxic at high concentrations and hence cellular levels of copper must be tightly regulated. Copper homeostasis is covered in more detail in Introduction Section 1.4. *M. tuberculosis* has been shown to be able to respond to changes in copper levels (Liu et al., 2007b). Transcriptome analysis identified a set of at least 30 copper-responsive genes in *M. tuberculosis*, approximately half of which are only induced at toxic levels of copper (Ward et al., 2008). It is believed that further characterising and elucidating the role of the copper response for survival in the human host may provide novel targets for disease control. Currently, the best characterised mechanism for copper response in *M. tuberculosis* is the copper-sensor protein CsoR.

3.1.2 CsoR – the copper-sensitive operon repressor

CsoR, or copper-sensitive operon repressor, was one of the first identified members of a large family of transcriptional repressors and is a Cu(I) responsive transcriptional regulator. The family members are made up of an all α-helical dimer of dimers structure and many are able to sense metal(s) (Higgins and Giedroc, 2014). CsoR-like regulators have been fairly well-characterised and are widespread across the five major classes of eubacteria. Among those characterised include CsoR proteins from *Steptomyces lividans*, *Staphylococcus aureus*, *Listeria monocytogenes*, *Geobacillus thermodenitrificans*, *Thermus thermophilus*. An alignment of these proteins with *M. tuberculosis* CsoR is shown in Figure 3.1, which also highlights important amino acid residues.

In the apo form when no copper is bound, CsoR can bind to the GC-rich, pseudopallindromic site (5' GTAGCCCACCCCCAGTGGGGTGGATAC 3') located in the promoter region of the *cso* operon. This site overlaps with the -10 and -35 regions of the *cso* promoter (Liu et al., 2007b).

```
S. aureus                  -----MGY-------------SIVRNVNEGVNAMTE-QDNAHHSEQIKTNLKSRLNRIEG
S. lividans                MTTTEAGASAPSPAVDGAVNQTARQAEADGTDIVTDHDRGVHGYHKQKAEHLKRLRRIEG
C. glutamicum              -------------------------------MSNSECHTHGYIEEKQRYLARLKRIEG
L. monocytogenes           -------------------------------MKHDQPIVPRKEDETKLLQNRLRRIEG
B. subtilis                ----------------------------MEKHNEHKTLNHKSSKEKDQITNRLKRIEG
G. thermodenitrificans     ------------MVIEV-----ITVNSNQYQDNNINAKMVPRTEEEIESIVKRLKRIEG
M. tuberculosis            ------------------------------------MSKELT-AKKRAALNRLKTVRG
T. thermophilus            ------------------------------MPHSHLHLDPKVREEARRRLLSAKG
                                                                        **   .*

S. aureus                  QVRAINRMIEE-DVYCDDVLTQIRATRSALNSVAIKLLEQHMKSCIMNKVNQGAQE—EA-
S. lividans                QIRGLQRMVDE-DVYCIDILTQVSASTKALQSFALQLLEEHLRHCVADAALKGGTEIDAK
C. glutamicum              QTRGIHRMIDE-EQYCIDILTQISAVNSALKNVAFGLLDDHLAHCVKEAADLGGDELDAK
L. monocytogenes           QIRGIAQMVED-DRYCTDILVQISAANKALKNVGLQVLEHHTAHCVVDAAKNGEDD---V
B. subtilis                QVRGIQNMVEN-DRYCVDILVQISAVQAAMKNVALHLLEDHAHHCVADAIKSGDGE—QA-
G. thermodenitrificans     QVRGVQKMVED-NRYCIDILVQISAIQAALNKVGVNLLERHVNHCVAKAIREGSGE—ES-
M. tuberculosis            HLDGIVRMLES-DAYCVDVMKQISAVQSSLERANRVMLHNHLETCFSTAVLDGHGQ—AA-
T. thermophilus            HLEGILRMLEDEKVYCVDVLKQLKAVEGALDRVGEMVLRAHLKDHVATAHERGDVE—EI-
                           :  .: .*::. . ** *:: *: *   ::.      :*  *  .       *  :

S. aureus                  MEELLVTFQKLIKD-------------------------
S. lividans                VEEATKAIGRLLRT-------------------------
C. glutamicum              LKEVSDAIARFSKA-------------------------
L. monocytogenes           MEDLLKAIRQFSKT-------------------------
B. subtilis                ISELLDVFKKFTKS-------------------------
G. thermodenitrificans     IRELMDVIKQF----------------------------
M. tuberculosis            IEELIDAVKFTPALTGPHARLGGAAVGESATEEPMPDASNM
T. thermophilus            VEELMEALKYR----------------------------
                           : :   ..
```

Figure 3.1: Alignment of CsoR protein sequences from difference bacterial species. Bacterial species are *S. aureus*, *S. lividans*, *C. glutamicum*, *L. monocytogenes*, *B. subtilis*, *G. thermodenitrificans*, *M. tuberculosis* and *T. thermophilus*. The residues which play an important role in allosteric switching in *M. tuberculosis* CsoR (shown in red) are widely conserved in CsoR proteins present in other species. The residues involved in the binding of Cu(I) (Cys36, His61 and Cys65) are shown in green. The C-terminal tail, which is only present in *M. tuberculosis* CsoR, and the putative phosphorylation target Thr93 is shown in orange.

The *cso* operon is comprised of genes encoding CsoR itself, CtpV (a Cu(I) efflux P1-type ATPase) and Rv0968 (a non-essential protein of unknown function).

3.1.3   Structure of CsoR

CsoR forms a disc-like dimer of dimers that assemble as an overall tetrameric structure (Figure 3.2c) (Higgins and Giedroc, 2014). A key structural feature of the dimers is the four-helix bundle (α1, α2, α1', α2'), which is flanked at the C-terminal end by α3/α3', the two helices which form the basis of the tetramer interface.

CsoR proteins contain a signature x-C-H-C motif (where x can be any amino acid) and the copper-binding site is located close to where two monomers bridge to form a homodimer. A ratio of one Cu(I) per monomer binds in a trigonal geometry via interactions with three residues: Cys36 (from α2), Cys65' (from α2') and His61' (from α2') (Figure 3.2a). Crystal structures of *M. tuberculosis* CsoR revealed a hydrogen-bonding network between His61', Glu81' of one subunit with Tyr35 of the other (Liu et al., 2007b) (Figure 3.2b). This network has been shown to be important in linking copper binding with DNA release for CsoR proteins in *M. tuberculosis* and *B. subtilis* (Liu et al., 2007b; Ma et al., 2009). It was proposed by Liu et al. that the network drives negative allosteric regulation of DNA binding by Cu(I) in *M. tuberculosis* CsoR. These residues are well-conserved among members of the CsoR-like repressor family. The exact details of the DNA-binding mechanism remain elusive at present (Higgins and Giedroc, 2014). However, pulse-chase amidination experiments using Cu(I)-bound and DNA-bound *B. subtilis* CsoR complexes revealed potential roles of C-terminal lysine residues (K96, K97 and K100) and the N-terminus of the α1 helix (Chang et al., 2011; Higgins and Giedroc, 2014). It has been proposed that the CsoR tetramer remodels after Cu(I) binding, which perturbs these residues and results dissociation of CsoR from the DNA site. Only K96 is present in *M. tuberculosis* CsoR.

DNA binding studies suggested that two tetramers of CsoR bind per DNA site in a CsoR tetramer:DNA "sandwich" (Jacobs et al., 2015) (Figure 3.3). The protein is also able to bind other metals, including Ni(II), Zn(II), Co(II), but interactions with other metals are of lower affinity and there is a strong thermodynamic preference for Cu(I) (Ma et al., 2009). Ni(II) binds with a square-planar geometry, Co(II) binds forms a tetrahedral geometry (Higgins and Giedroc, 2014).

3.1.4   CsoR may regulate the *rpfA* gene in *M.tuberculosis*

A recent study aimed to gather ChIP-seq data to provide genome wide information on binding sites for *M. tuberculosis* transcription factors (Cortes et al., 2013). ChIP-seq data indicated that the *rpfA* (*Rv0867c*) gene is regulated by CsoR (also known as Rv0967). As the scope of this
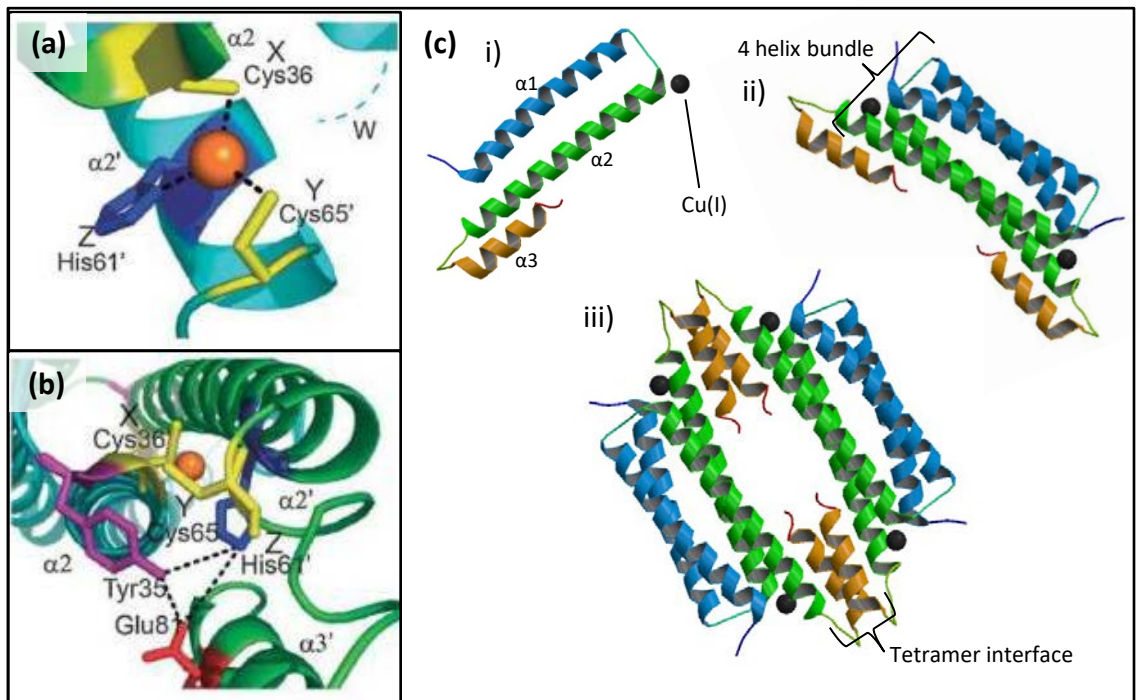
Figure 3.2: Crystal structures of *M. tuberculosis* CsoR with Cu(I) associated. (a) Cu(I) (orange) binds in a trigonal geometry between Cys35 of one monomer and His61' and Cys65' from the other monomer of the homodimer. (b) Suggested hydrogen-bonding network between His61', Glu81' of one subunit with Tyr35 of the other. (c) i) Structure of a *M. tuberculosis* CsoR monomer with Cu(I) associated. ii) Structure of a *M. tuberculosis* CsoR homodimer with 2 Cu(I) associated. iii) Structure of a dimer of dimers of CsoR with 4 Cu(I) associated.

(a) and (b) taken from Liu et al. (2007). (c) taken from 2HH7 structure (PDB).
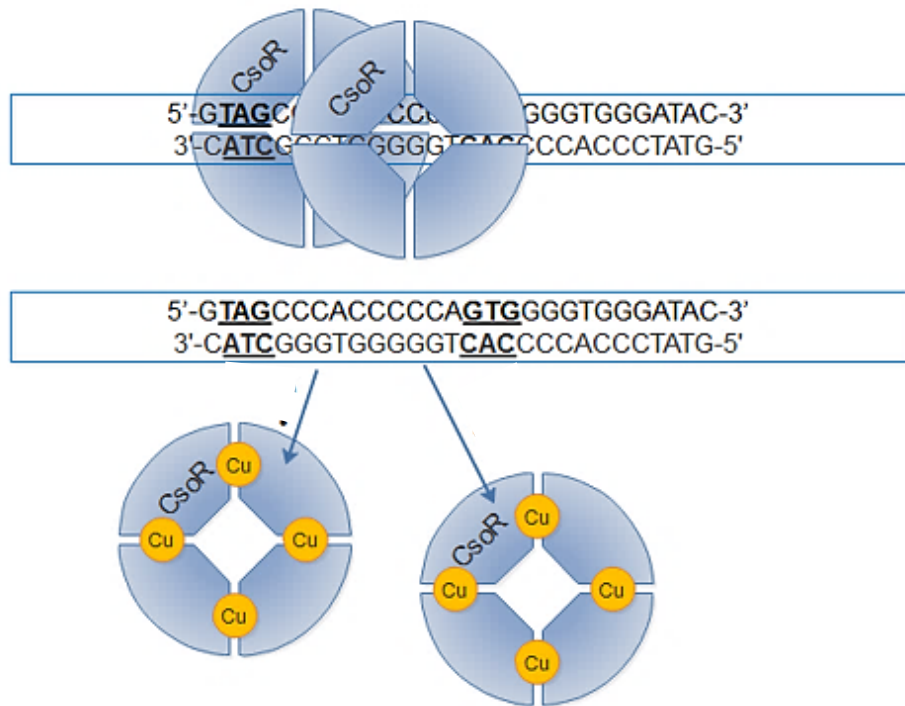
Figure 3.3: Two CsoR tetramers bind per target DNA site. CsoR teramers (blue) bind to target sites in a "sandwich-like" manner, where two tetramers associate per one DNA site. Upon binding of Cu(I) The CsoR tetramers dissociate from the DNA.

thesis centres around the characterisation of different mechanisms of regulation of RpfA, it was decided to further investigate if CsoR can bind to the putative CsoR-site in the *rpfA* promoter and if this interaction is sensitive to copper.

3.1.5   Aims of this chapter

The main aims of this chapter are listed below.

- Initially to produce a map of the different regulatory sites present in the *rpfA* gene, particularly focusing on identifying the putative CsoR binding site sequence.
- To successfully overproduce and purify unlabelled CsoR protein from an *E. coli* host.
- To ascertain using EMSA experiments if the purified CsoR is able to bind to the predicted *rpfA* promoter region.
- To determine if the interaction with DNA is sensitive to Cu(I).
- To obtain data on the binding affinity (K_D) of the interaction and to assess if the rate of dissociation (k_d) alters when Cu(I) is present.
- To investigate the putative role that phosphorylation of the T93 residue of CsoR may play in *rpfA* regulation by generating and purifying phosphomimetic mutants (T93D and T93E).
- To characterise the DNA-binding ability and the effect of copper-binding on DNA-binding for the phosphomimetic mutants.

3.2   Mapping of regulatory sites on the *rpfA* gene

ChIP-Seq data on TBDB (Tuberculosis Database) suggested that a CsoR regulatory site was present in the *rpfA* gene of *M. tuberculosis*. To determine the exact location of the putative CsoR regulatory site, bioinformatics analysis was performed and the sites of all the different regulators of *rpfA* were mapped (Figure 3.4). The CsoR binding sequence was identified to be located 129 bp downstream of the start codon. This site is a GC-rich pseudopallindromic sequence. The MtrA site is located 421 bp upstream of the TSS and was identified using TBDB ChIP-seq data and comparison to the consensus binding site (Brocker et al., 2011). Lsr2 is an iron-responsive protein and the binding site was mapped 362 bps upstream of the TSS based on ChIP-seq data from TBDB. The CRP binding site is located upstream of the TSS and binding to this site has been demonstrated by Rickman et al. (2005) and also during later stages of this study (Section 4.4.1). The region containing the *ydaO*-type riboswitch is shown in green and this regulatory element is discussed in depth in Chapter 5.

~1000 bp

(approx to scale)

5'-CGTTATCCGCCGGCGACA-3'

MtrA site

Lsr2 site

CRP site

5'-TGTGACATTACCCAC-3'

P_1

Riboswitch

P_2

P_3

CsoR site

5'-CGGGGAATGGGATCAGGTGGCCCG-3'

rpfA (Rv0867c)

5'

3'
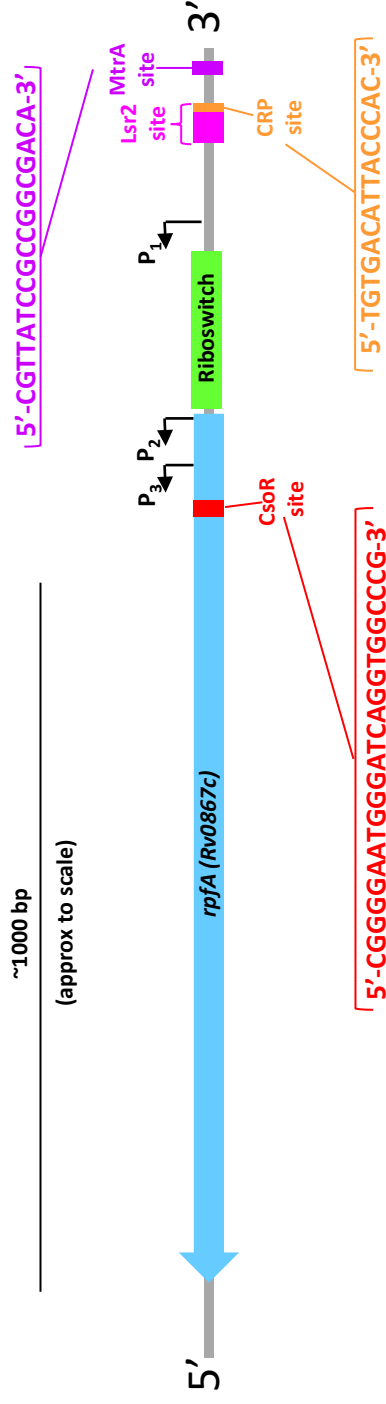
Full coding sequence
Located between co-ordinates 9654312 to 9655535

Transcript start sites (TSS) – Identified in Cortes et al. (2013)

P^1 is located at co-ordinate 965808 and produces mRNA encoding the riboswitch (in 5' UTR) and full coding sequence

P^2 is located at co-ordinate 965535 and produces mRNA encoding the full coding sequence of RpfA

P^3 is located at co-ordinate 965463 and produces mRNA encoding a shortened, cytoplasmic RpfA that lacks the signal sequence for extracellular export via the Sec system.

CsoR site 129 bp downstream of the start codon
MtrA site is 491 bp upstream of the start codon
Lsr2 site is 362 bp upstream of the start codon
CRP site is 434 bp upstream of the start codon
The riboswitch is located from 232 bp to 9 bp upstream of the start codon

Figure 3.4: Map of binding sites and regulatory elements of the *M. tuberculosis rpfA* gene. The binding sites of protein regulators CsoR, CRP, Lsr2 and MtrA are indicated. The location of the putative *ydaO*-type riboswitch is also indicated. This thesis covers the regulatory roles of CRP (Rv3676) in Chapter 4, the riboswitch in Chapter 5 and this chapter (Chapter 3) discusses regulation by CsoR.

3.3 Purification of CsoR

In order to investigate if CsoR interacts with the predicted CsoR site present in the *rpfA* promoter, it was necessary to purify the protein. It was desirable to purify untagged CsoR to avoid altering the protein's native properties. It has been suggested that the C-terminal lysine residues (K96, K97 and K100) and the N-terminus of the α1 helix of the *B. subtilis* CsoR play a role in remodelling of the CsoR tetramer after copper binding and release from the DNA (Chang et al., 2011). For this reason, it was deemed inappropriate to tag the protein close to the N- or C-terminus and risk altering the native properties of the protein, so the untagged protein was overproduced for purification.

3.3.1 Overexpression of CsoR in *E. coli* host

Electrically competent *E. coli* BL21 λDE3 cells were transformed (Method Sections 2.2.6 and 2.2.7) with pGS2528 plasmid (map in Figure 3.5) generated in a previous study. This plasmid consists of a pLATE11 vector (by Thermo Scientific) with the CsoR coding region cloned between the XbaI and SwaI restriction sites. The plasmid has an ampicillin resistance gene and uses a pET expression system, under control of the *lac* promoter and operator. The expression of the cloned gene of interest is driven by a strong bacteriophage T7 promoter, which is recognised by T7 polymerase. The host *E. coli* strain possesses an IPTG-inducible T7 RNAP encoded on the chromosome. In the absence of IPTG, the *lac* operator represses the expression of T7 polymerase and the gene of interest. IPTG is able to displace the Lac repressor and allow T7 polymerase to be expressed, which allows for the expression and production of the CsoR protein.

Initially, overproduction of CsoR was carried out in LB broth by the methods outlined in Method Section 2.4.5.1. After overproduction, cells were collected by centrifugation and were lysed using a French pressure cell (Method Section 2.4.6). To determine if CsoR was overproduced, the soluble cell-free extract (CFE) and the insoluble cell pellet were analysed by SDS-PAGE (Method section 2.4.1) (Figure 3.6a). An abundant ~12.7 kDa polypeptide was present in the CFE lane, suggesting that CsoR has successfully been overproduced under these conditions.

3.3.2 Heparin affinity chromatography purification of CsoR from CFE

Cells were resuspended in 10 mM sodium phosphate (pH 7.0) before being lysed to produce CFE (Method Section 2.4.6). CsoR purification from the cellular proteins was attempted using heparin affinity chromatography. Heparin affinity chromatography is a useful method during the purification of DNA binding proteins such as CsoR. This is because DNA-binding proteins can often bind to heparin columns due to the high negative charge density of heparin
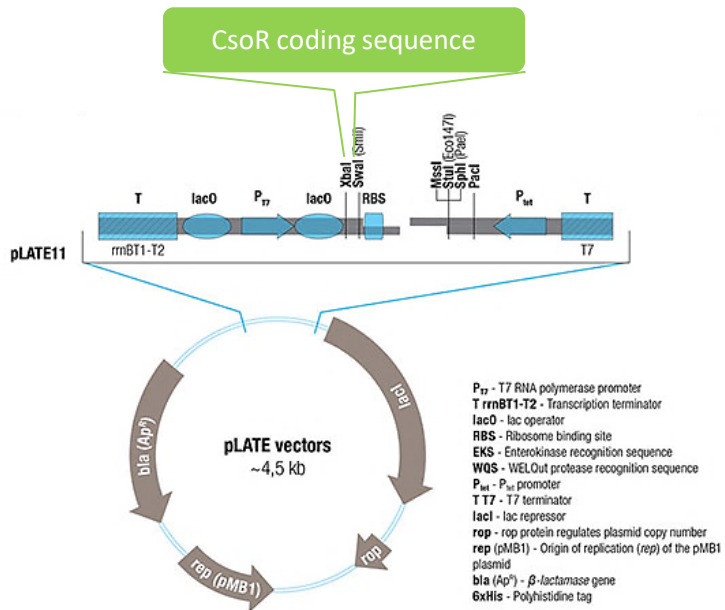
Figure 3.5: Map of pGS2528 plasmid used for the overproduction of CsoR. The plasmid based on the pLATE11 vector (by Thermo Scientific) with the CsoR coding region cloned between XbaI and SwaI cloning sites (shown in green).
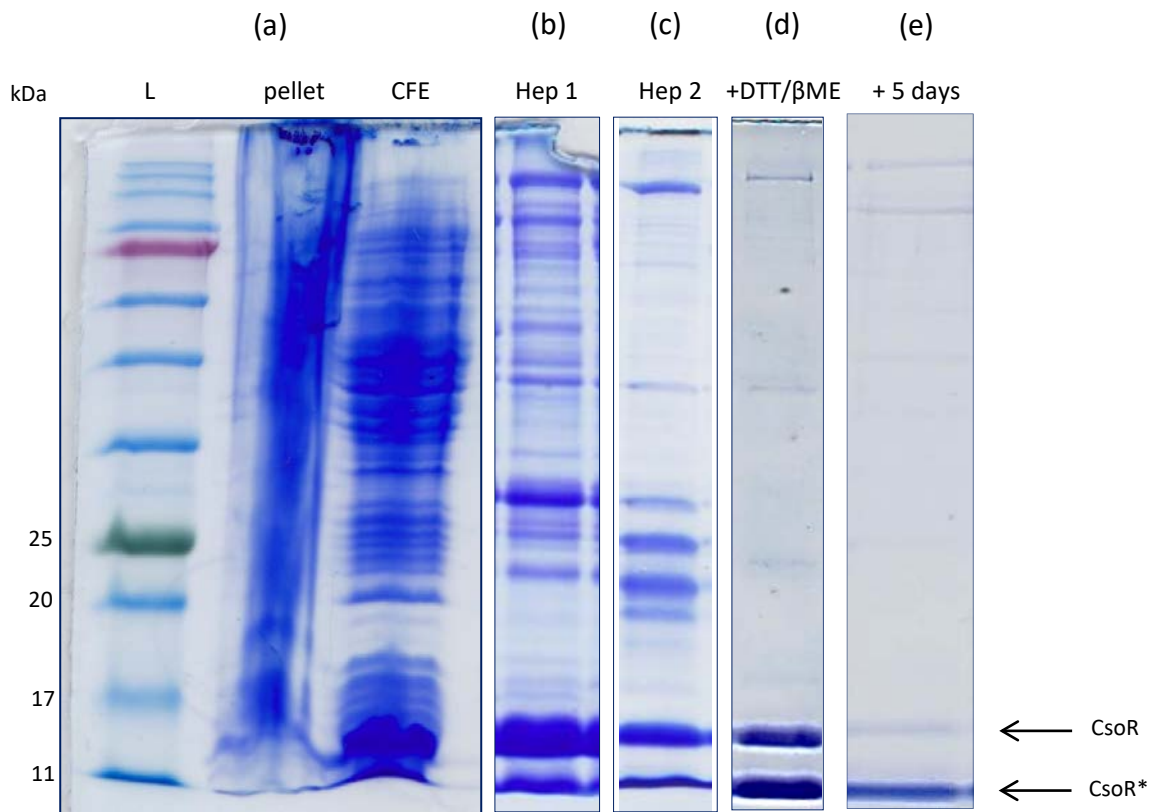
Figure 3.6: Initial purification of CsoR. a) SDS-PAGE analysis of overproduced CsoR. Lane L contains 10 µl BLUEstain™ Protein Ladder (by Gold Biotechnology). Pellet lane contains insoluble cell pellet after lysis (10 µl) and CFE lane contains soluble cell-free extract (10 µl). b) Purity of CsoR after first heparin purification step. Hep 1 contains 10 µl of fraction 19 eluted from a HiTrap heparin column with a 0.2 – 0.8 M salt gradient. c) Purity of CsoR after second heparin purification step. Hep 2 contains 10 µl of fraction 18 eluted from a HiTrap heparin column with a 0.325 – 0.8 M salt gradient. d) Treatment of CsoR with additional reducing agents. +DTT/βME lane contains 10 µl of CsoR sample treated with 100 mM DTT and 1M βME at 95°C for 10 min. e) Degradation of CsoR to CsoR* over time. Sample in +5 days the same sample as in (d) but after 5 days storage. The position of full-length CsoR (~12.7 kDa) and CsoR* (CsoR breakdown product) are indicated. This figure is produced from representative lanes from individual gels that have been scaled.

mimicking the negatively charged DNA backbone and will only be released from the column at higher ionic strengths.

In the first purification step, affinity chromatography was carried out using a HiTrap Heparin 1 ml column on an AKTA Prime machine (GE Healthcare) as detailed in Method Section 2.4.9.3. The cellular proteins were resolved with a linear salt gradient (10 mM sodium phosphate (pH 7.0) with 0.2 to 0.8 M NaCl). Fractions (1 ml) were collected and analysed by SDS-PAGE. Fraction 19 contained the highest purity of CsoR protein compared to other cellular proteins (Figure 3.6b). Although this fraction had higher purity than CFE, it still contained considerable amounts of contaminating proteins and was not suitable for further experiments. CsoR*, corresponding to CsoR after C-terminal tail degradation (discussed in depth in Section 3.3.3), was observed in this fraction, but initially was suspected to just be another unrelated cellular protein.

In order to further purify CsoR, a second heparin affinity chromatography step was performed. Fractions containing the highest purity and amount of CsoR were combined and dialysed against low salt buffer (10 mM sodium phosphate with 50 mM NaCl) and the sample was then applied to the heparin column. The method was the same as for the first heparin purification, but a steeper salt gradient (10 mM sodium phosphate (pH 7.0) with 0.325 to 0.8 M NaCl) was applied during elution. During elution, fractions (1 ml) were collected and analysed by SDS-PAGE. The fraction containing the fewest contaminating proteins and the highest yield of CsoR was fraction 18, which is shown in Figure 3.6c, Lane 'Hep 2'. Considerably fewer contaminating proteins were observed after the second purification step (Lane 'Hep 2'), compared to after just the first purification (Lane 'Hep 1'). CsoR* was present in the sample in similar proportion to CsoR as in the previous purification step.

It was speculated that some, or all, of the protein contaminants present after the second purification step could be multimeric forms of CsoR. Moreover, it was possible that internal disulphide bridges between cysteine residues of CsoR may not have been fully reduced before analysis by SDS-PAGE and may cause incomplete denaturation and migration to suggest a larger size. To test if these species were CsoR, samples were treated with higher concentrations of the reducing agents dithiothreitol (DTT) (final concentration 100 mM) and β-mercaptoethanol (βME) (final concentration 1M) for 10 min at 95°C. The sample was analysed by SDS-PAGE (Figure 3.6d). After treatment with reducing agent many of the previously suspected 'contaminant' species disappeared or were reduced. This suggested that they were different forms of CsoR and that treatment with increased concentrations of reducing agent (1M βME and 100 mM DTT at 95°C for 10 min) is sufficient to reduce any intramolecular and

intermolecular bonds present, resulting in fully denatured monomeric CsoR. The 2X SDS-PAGE loading dye with extra reducing agent was used for all subsequent SDS-PAGE analysis with CsoR protein samples. The smaller polypeptide species (CsoR*) was present at a slightly higher proportion in the protein sample that was treated with reducing agent.

3.3.3 The C-terminal tail region of CsoR is degraded

As mentioned previously in section 3.3.2 a species corresponding to a smaller, <12.7 kDa polypeptide was observed. This species progressively increased in proportion to the full length CsoR during the purification steps and over time. Initially it was a minor component, but after the second heparin purification and treatment of the sample with reducing agents, the smaller polypeptide species (CsoR*) was present at almost equal proportions to CsoR.

It was important to determine what this species was before using the purified CsoR for further studies. It was hypothesised the species would like be either another form of CsoR that had not completely denatured before migration; a degradation product of full length CsoR or another protein that was co-purifying. It was unlikely that it was a form of CsoR that had not fully denatured as samples were analysed by SDS-PAGE with increased reducing agents present actually showed an increase in abundance of this polypeptide. The sample was sent for mass spectroscopy analysis (Method Section 2.4.13) with the aim of identifying the smaller polypeptide (Figure 3.7). Several different sized species were present and their molecular weights were compared to the CsoR amino acid sequence. Analysis showed that the sample contained almost entirely CsoR species, but that the full length CsoR protein was being degraded from the C-terminus. The exact residues between which cleavage was occurring are shown by the red lines in Figure 3.7. SDS-PAGE analysis was performed on the CsoR sample several days after purification, which (Figure 3.6e) revealed that a higher proportion of the protein present had been degraded, as there is an increase in the abundance of CsoR*, compared to Figure 3.6d.

Interestingly, a flexible tail region is located at the C-terminus, which is not present in the CsoR proteins from many other species e.g. *Bacillus subtilis*, *Listeria monocytogenes*, *Staphylococcus aureus* or *Steptomyces lividans* (Dwarakanath et al., 2012). We observed that the loss of the C-terminal tail region of the protein results in loss of DNA-binding ability in EMSA experiments (data not shown) under the same conditions that binding was observed with full length CsoR (Section 3.4.2.1). It has been proposed that CsoR binds to DNA targets as a dimer of dimers and that the C-terminal tail plays a role in the association between the two dimers (Higgins and Giedroc, 2014). This is a logical explanation for why the ability to bind DNA was not observed.

CsoR protein sequence

```
        10          20          30          40          50          60          70
MSKELTAKKR  AALNRLKTVR  GHLDGIVRML  ESDAYCVDVM  KQISAVQSSL  ERANRVMLHN  HLETCFSTAV

        80          90         100         110         119
LDGHGQAAIE  ELIDAVKFTP  ALTGPHARLG  GAAVGESATE  EPMPDASNM*
```
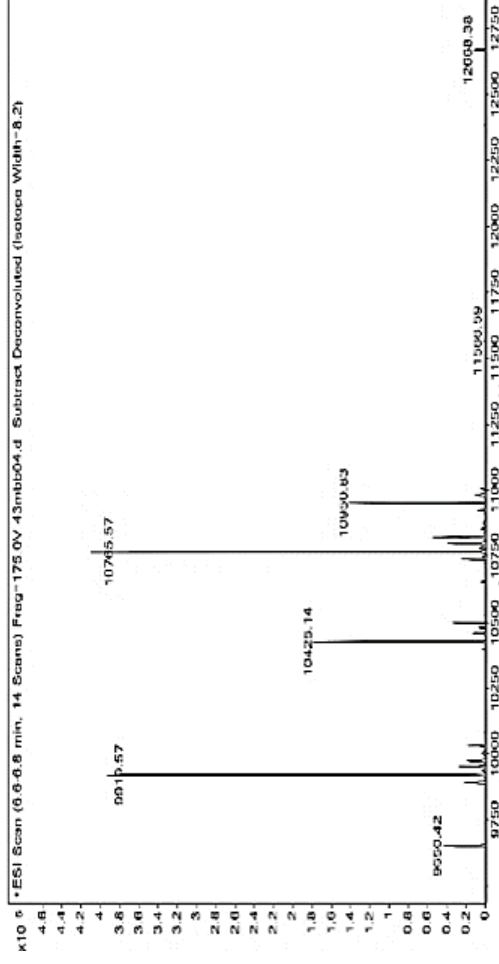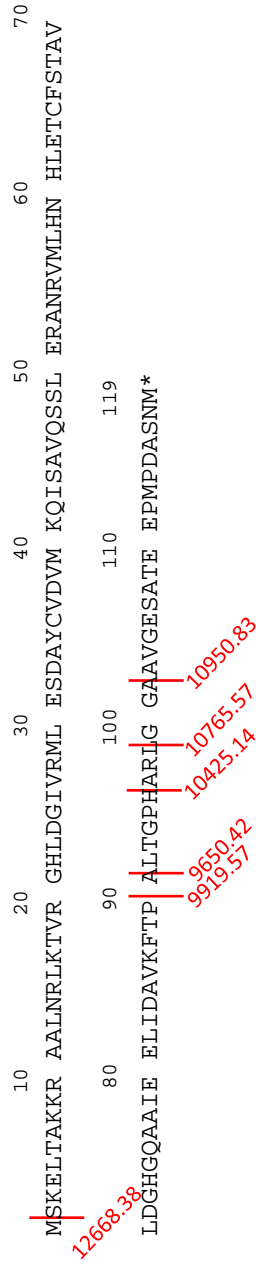
Figure 3.7: Degradation of the C-terminal tail of CsoR. Mass spectroscopy analysis was performed on CsoR after the second heparin affinity chromatography purification step. The sizes of species present were compared to the CsoR amino acid sequence to determine the points where the protein has been cleaved.

In addition to this, the tail region is missing from the published crystal structures of *M. tuberculosis* CsoR (Liu et al., 2007b). Not only is the tail region likely important for dimer interaction and hence DNA binding ability, but one residue (T93) may also act as a target for a member of the Pkn kinase family (unpublished data), which is discussed in Section 3.5.1. CsoR that has lost the C-terminal tail will not be able to be phosphorylated and is not functional for regulatory roles in DNA binding. It is plausible that phosphorylation of T93 may stabilise the tail region, making it less likely to be lost, and hence keeping an active, DNA-binding form of CsoR in the cell.

Protease enzymes could be responsible for the observed degradation or it could be due to the tail region generally being less stable under the storage conditions. In order to prevent degradation of CsoR by proteolytic enzymes various steps were carried out. Acid-washed glassware was used to hold all purification and storage buffers. Buffers were autoclaved (121°C for 15 mins at 15 psi) and once cooled, protease inhibitors were added (final concentration 1 mM benzamidine and 0.1 mM PMSF). In addition to using protease inhibitors, purification steps and storage were altered (Section 3.3.4) to allow for the purification of high purity and stable CsoR. Elements of the purification protocol were based on those used in Liu et al. (2007) and Marcus et al. (2016).

3.3.4 Optimised method for purification of CsoR

CsoR is a copper sensing protein, which can bind one Cu(I) per protein monomer. Copper binding acts as an allosteric switch and results in a confirmation change in the C-terminal region of CsoR, which ultimately results in the release of the CsoR tetramers from the DNA (Figure 3.3) (Chang et al., 2011; Jacobs et al., 2015; Liu et al., 2007b; Ma et al., 2009). Because of this, it is desirable to purify apo-CsoR. In order to minimise the amount of holo-CsoR purified, steps were taken to minimise copper exposure during overproduction and purification. It was desirable to purify holo-CsoR to allow for the investigation of the DNA-binding in the absence of Cu(I) and then treating the protein with Cu(I) to determine the effect. In order to reduce the likelihood of purifying a high proportion of copper-bound CsoR, steps were taken to minimise levels of copper during overexpression of the protein. This included switching from rich LB medium to M9 medium for overexpression steps and taking the precautionary steps during buffer preparation suggested in Liu et al. (2007) and Marcus et al. (2016).

In addition to this, to try to minimise C-terminal degradation, protease inhibitors (final concentration 1 mM benzamidine and 0.1 mM PMSF) were included in all purification buffers and during cell lysis.

The optimised overproduction and purification method for CsoR is as follows. Like before, *E. coli* BL21 λDE3 cells containing pGS2528 were grown to an OD_{600} between 0.4 and 0.6. Cells were pelleted washed twice in M9 minimal medium before being resuspended in M9 minimal medium and being induced with IPTG (120 µg/ml) (Method Section 2.4.5.2). Overproduction was carried out in M9 medium rather than LB media (as before) in order to reduce the available copper during overproduction of CsoR, which should result in a reduction in the amount of holo-CsoR present. All glassware was acid washed (nitric acid) prior to use to prevent copper contamination as suggested in Marcus et al. (2016).

After cell lysis with French pressure cell (Method Section 2.4.6), samples were kept at 4°C in 25 mM Tris (pH 8.0) with 50 mM sodium chloride between all purification steps and once purified protein had been obtained. This was to minimise breakdown of the C-terminal tail region and to stabilise the protein for longer periods. The following purification steps were used for the generation of CsoR from CFEs. Between each stage excess salt was removed from the protein sample by dialysis against "low-salt" buffer containing 50 mM NaCl.

After CFE was produced, the first purification step was polyethyleneimine (PEI) precipitation with 0.15% (v/v) final concentration at pH 5.8 (Method Section 2.4.9.1). The supernatant from the PEI precipitation was subjected to ammonium sulphate precipitation (ASP) (Method Section 2.4.9.2). ASP can be used as a crude purification method to remove cellular proteins that possess different inherent precipitation properties to the protein of interest (CsoR). Different cuts of ASP were tested to establish the % cut that would provide the highest amount of CsoR protein, whilst resolving the most impurities (data not shown). The 40-50% cut was selected as the best option.

The sample was dialysed into 25 mM Tris (pH 8.0) and affinity chromatography using a Heparin-HP column (5 ml, GE Healthcare) on an AKTA Purifier syatem with a flow rate 5 ml/min was performed. Proteins were eluted in 25 mM Tris (pH 8.0) over a 50 ml NaCl gradient (0 to 1 M) (Figure 3.8a). Fractions (2.5 ml) were collected and fractions 13 to 18 contained CsoR were combined (~15 ml). They were diluted to ~45 ml before being applied to a Resource Q column (6 ml, GE Healthcare) and undergoing ion exchange chromatography. This was carried out on an AKTA purifier system with a flow rate 4 ml/min over a 50 ml NaCl gradient (0 to 0.7 M) (Figure 3.8b). Fractions (2.5 ml) were collected and those containing CsoR (fractions 21 to 23) were combined and concentrated using a Vivaspin 3000 MWCO until a volume of <1 ml was achieved (~12.6 mg/ml). This sample was applied to a 1.6x60 HiLoad Superdex200 column. Gel filtration was performed using a flow rate of 1.5 ml/min in 50 mM Tris (pH 8.0) containing 0.5 M NaCl (Figure 3.8c). Fractions (2 ml) were collected and analysed by SDS-PAGE (Figure 3.8d)
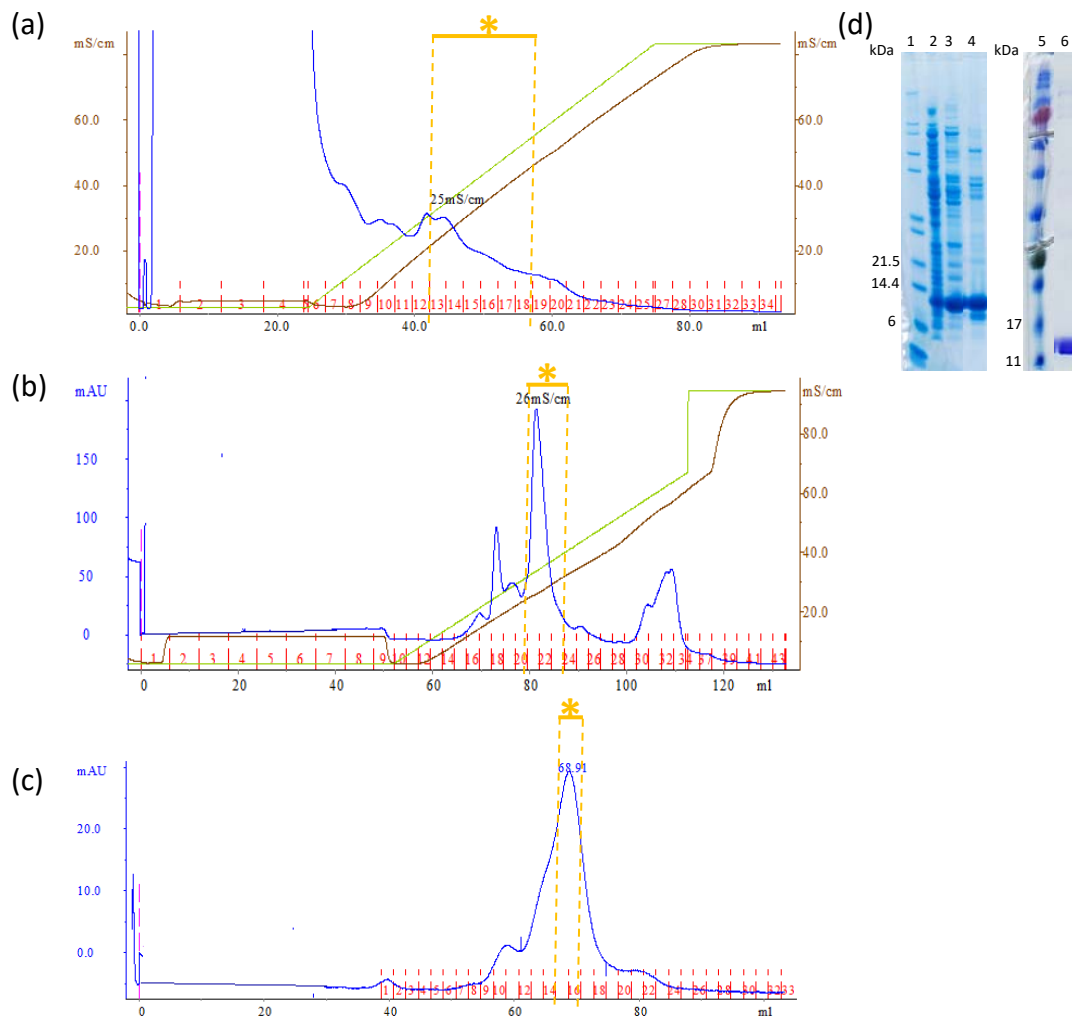
Figure 3.8: Optimised purification of CsoR. a) Absorbance during heparin affinity chromatography. Fractions 13 to 18 (*) were combined for further purification. b) Absorbance (280 nm) during ion exchange chromatography. Fractions 21-23 (*) were combined for further purification. c) Absorbance (280 nm) during gel filtration. Fractions 15 and 16 contained purified CsoR protein. d) SDS-PAGE analysis of samples from the stages of purification of CsoR. All lanes contain 10 µl of the relevant sample. Lane 1 contains Mark12™ Ladder (by Thermo Scientific); Lane 2 contains cell-free extract; Lane 3 contains a sample after heparin affinity chromatography; Lane 4 contains a sample after ion exchange chromatography; Lane 5 contains BLUEstain™ Protein Ladder (by Gold Biotechnology) and lane 6 contains purified CsoR protein after gel filtration.

and sufficiently pure CsoR was obtained in fractions 15 and 16 (Lane 6, Figure 3.8d). A set of protein standards with known molecular weights were run down the column to allow for the oligomeric state of the purified CsoR protein to be determined. The calibration curve is shown in Figure 2.1. The purified CsoR eluted from the gel filtration column at ~67ml (K_{av} = 0.34), indicating that the CsoR was purified as an octomer.

The typical yield from a 1 litre culture was ~0.5 ml of ~200 μM protein. The CsoR sample was estimated to be at least 90% pure and was deemed sufficiently pure for further experiments to be carried out. Storing the protein at 4°C in 50 mM Tris (pH 8.0), 50 mM NaCl in the presence of 3 mM DTT and protease inhibitors provided a stable, functional, full length protein for approximately two weeks. To our knowledge, this is is the first purified full length *M. tuberculosis* CsoR protein.

3.4 Binding studies with wildtype CsoR

3.4.1 Generation of *rpfA* DNA containing the predicted CsoR binding site

It was necessary to produce regions of DNA containing the predicted CsoR site to use in binding experiments with the purified CsoR protein. Two *rpfA* regions were selected for this study, which will be referred to as *rpfA*$_S$ and *rpfA*$_L$, the sequences of which are shown in Figure 3.9. *rpfA*$_L$, is a longer region (185 bp) containing the CsoR site with ~80 bp flanking region either side. *rpfA*$_S$ is a shorter region (48 bp) containing the CsoR site with 12 bp flanking either side of the regulatory site. The shorter promoter region was selected to narrow down the location of the binding site, but the region with longer flanking regions was also tested in case the flanking DNA was important for stabilisation of binding of the two CsoR tetramers to the DNA.

Primers were designed to allow for the amplification of the relevant regions of *M. tuberculosis* genomic DNA (gDNA). For *rpfA*$_L$, the forward primer Bio-LD67 and reverse primer LD68 (sequence in Table 2.1) were designed. The forward primer possessed a 5'-end biotin label, which were necessary for subsequent EMSA and BLItz experiments. PCR was carried out (Method Section 2.3.4) with the primers and genomic DNA. The products were analysed by agarose gel electrophoresis (Method section 2.3.8) to confirm for the presence of the desired 185 bp DNA species (Figure 3.10). A DNA species of the predicted size was present (Figure 3.10, lane 2) indicating the amplification of was successful *rpfA*$_L$. The promoter region was purified from the reaction mixture using a PCR purification kit (Method section 2.3.5) and eluted in nuclease water.

5' – GTCGCCAAGATCGCCTTTACCGGCGCAG
TACTCGGTGGCGGCGGCATCGCCATGGCCGC
TCAGGCGACCGCGGCCACCGA**CGGGGAATG**
GGATCAGGTGGCCCGCTGCGAGTCGGGCGG
CAACTGGTCGATCAACACCGGCAACGGTTAC
CTCGGTGGCTTGCAGTTCACTCAAAGCACCT
GGGC - 3'

Figure 3.9: Region of the _rpfA_ sequence used for CsoR binding experiments. The putative CsoR binding site is indicated in red. The complete _rpfA_$_L$ sequence is shown. The shorter _rpfA_$_S$ is indicated in green. CsoR binding sites are typically psuedopalindromic in sequence. The underlined letters are the palindromic bases within the binding site.
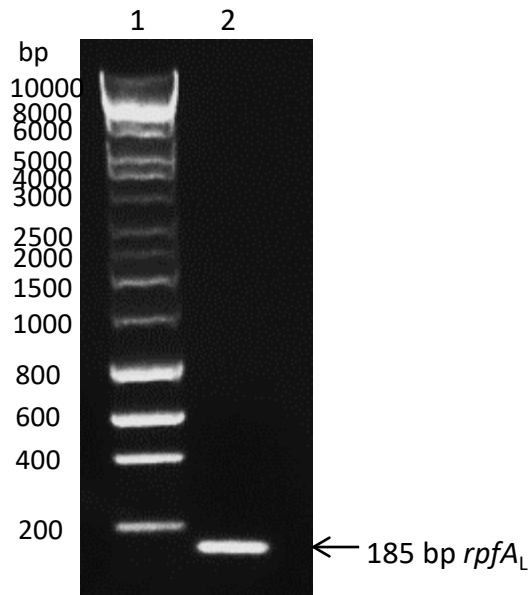
Figure 3.10: Agarose gel electrophoresis analysis of the products of PCR amplification of _rpfA_L. Lane 1 contained 5 µl Hyperladder 1 Kb (Bioline). Lane 2 contained 5 µl of the PCR products of a reaction containing primers Bio-LD67 and LD68 and _M. tuberculosis_ gDNA. The presence of an 185 bp species, corresponding to _rpfA_L is indicated.

To produce $rpfA_S$, oligos Bio-CsoR-F and CsoR-R were synthesised by Eurofins of the complete 48 bp sequence. These were diluted to the relevant concentration, heated to 94°C for 5 minutes and allowed to cool to room temperature to anneal before use in binding experiments.

3.4.2 EMSA analysis of the interaction between CsoR and *rpfA*

3.4.2.1 CsoR binds to the *rpfA* gene and binding is released in the presence of copper

EMSA experiments were carried out using purified CsoR and Biotinylated *rpfA* DNA ($rpfA_L$) to determine if CsoR is able to bind to the putative CsoR regulatory site. Binding reactions containing biotinylated $rpfA_L$ DNA (20 pmol) and differing concentrations of CsoR (1 to 10 μM) were set up in the presence of absence of copper chloride (100 μM), and were incubated and electrophoresed by the methods detailed in Section 2.5.1.2.

The EMSA analysis (Figure 3.11) revealed that CsoR was able to bind to $rpfA_L$ in the absence of copper. Lane 1 shows the migration of free $rpfA_L$ in the absence of the protein. The formation of CsoR:$rpfA_L$ complex can be observed in lanes 3 to 5, where the migration of $rpfA_L$ DNA has been retarded when CsoR is present. The shift becomes more pronounced with increased protein concentration and can first be observed at CsoR concentration above ~2.5 μM (monomeric). When copper was present in reactions (lanes 6 to 9), the DNA migrated similarly to free $rpfA_L$ DNA, suggesting that the $rpfA_L$:CsoR complex is released in the presence of copper.

EMSA experiments were also performed using the shorter promoter fragment $rpfA_S$ (20 pmol) with different concentrations of CsoR (0.01 to 10 μM) (Figure 3.12). The migration of free $rpfA_L$ in the absence of the protein is shown in lane 1. The formation of CsoR:$rpfA_S$ complex can be observed in lane 5, where the migration of $rpfA_L$ DNA has been retarded in the presence of 10 μM CsoR. This point at which the shift occurs is between 1 and 10 μM, which supports the findings of the EMSA experiments with $rpfA_L$ where the shift was first observed at 2.5 μM CsoR.

3.4.2.2 Copper titration shows 0.5X copper is sufficient to release CsoR from the $rpfA_L$:CsoR

EMSA experiments were performed with reactions containing purified CsoR (10 μM), biotinylated $rpfA_L$ DNA (20 pmol) and different proportions of copper, ranging from 0.25X to 10X the CsoR monomer concentration (Figure 3.13). The migration of free DNA is shown in lane 1, where no protein or copper are present. Shifts are observed in lanes 2 to 4, which are
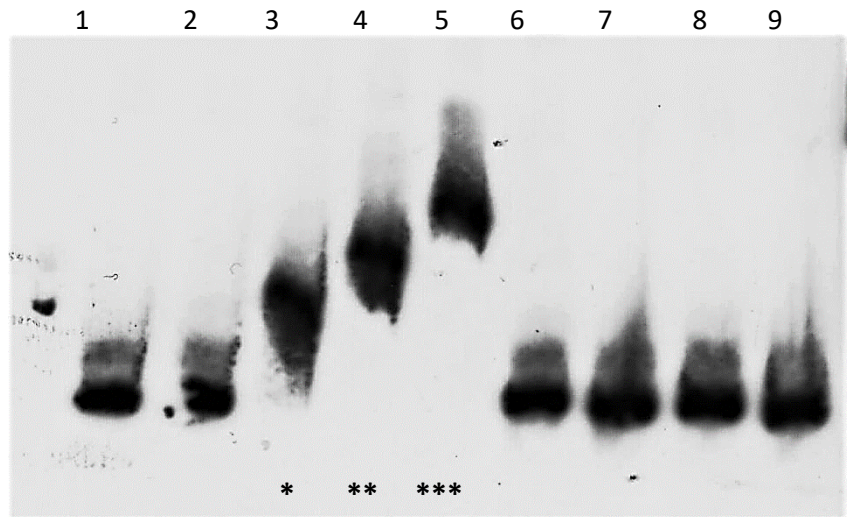
	1	2	3	4	5	6	7	8	9
CsoR conc. (µM)	0	1	2.5	5	10	1	2.5	5	10
Copper (100 µM)	-	-	-	-	-	+	+	+	+

Figure 3.11: EMSA analysis of the interaction between CsoR and *rpfA*$_L$ in the presence or absence of copper. Biotinylated *rpfA*$_L$ (1 µM) was incubated with different concentrations of CsoR protein (1-10 µM) in the presence or absence of copper chloride (100 µM). The protein concentrations provided are for CsoR monomers. This shift is a representative selected from 5 replicates.

	1	2	3	4	5

CsoR conc.   0   10 nM 100 nM 1 µM   10 µM
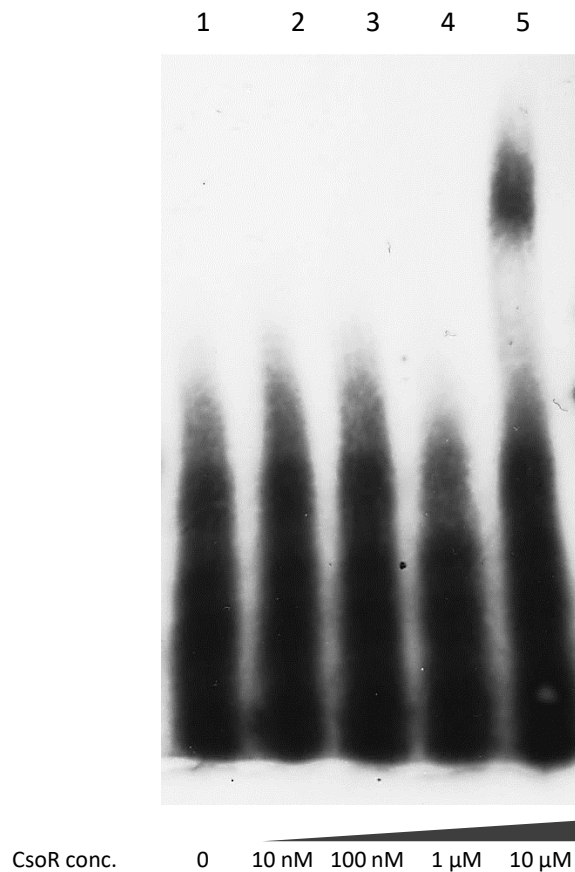
Figure 3.12: EMSA analysis of the interaction between CsoR and *rpfA*$_S$. Biotinylated *rpfA*$_S$ (1 µM) was incubated with different concentrations of CsoR (0.01-10 µM monomeric concentration) in the absence of copper. The migration of free *rpfA*$_S$ (no protein) can be seen in lane 1. A shift in migration of the *rpfA*$_S$ DNA can be observed in lane 5, containing 10 µM CsoR. Representative shift selected from 3 replicates.
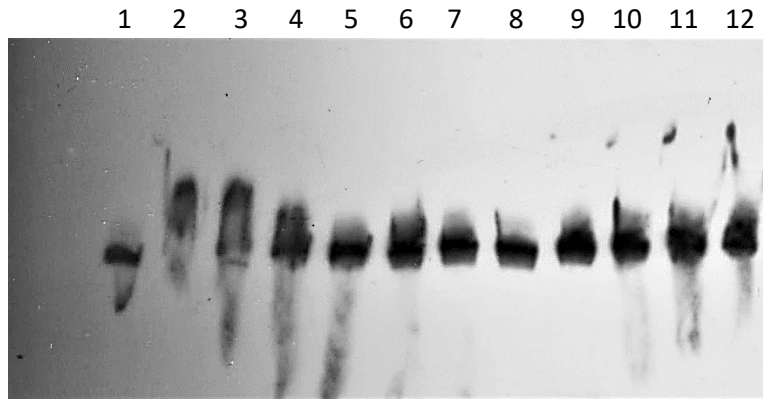
	1	2	3	4	5	6	7	8	9	10	11	12
Copper conc. (µM)	0	0	1	2.5	5	7.5	10	15	25	50	75	100
Molar ratio (copper:protein)	-	-	0.1	0.25	0.5	0.75	1	1.5	2.5	5	7.5	10

Figure 3.13: EMSA analysis of the interaction between CsoR and *rpfA*_L with different copper concentrations. Biotinylated $rpfA_L$ (1 µM) was incubated with CsoR (10 µM) with different concentrations of copper (0-100 µM). These concentrations provide a molar ratio of 0.25X to 10X copper:protein monomer. Lanes 2 to 4 contain CsoR: $rpfA_L$ complex as a shift in migration of the $rpfA_L$ DNA can be observed when compared to free $rpfA_L$ (Lane 1). Shift is a representative of 3 replicates.

due to retarded migration of *rpfA*_L because of binding of CsoR to the CsoR site. Lanes 2 to 4 contain 0.1 to 0.5X molar concentration of copper compared to CsoR monomer. No shift in migration of DNA was observed in lanes 5 to 13, corresponding to 0.75X to 10X molar concentration of copper compared to CsoR. This suggests that 0.75X copper:CsoR monomer is sufficient to release CsoR from the DNA.

Previous literature has indicated each CsoR monomer binds one copper ion. Taking this into account, a 1X molar ratio of copper:CsoR monomer should theoretically be required to completely release CsoR from the promoter DNA. However, only 0.75X copper was required to fully release CsoR from the promoter. This could be explained by some of the purified CsoR protein being purified in the copper-bound form. Efforts were made to minimise the proportion of copper present during overexpression and purification of the protein, which are discussed in Section 3.3. Attempts were made to determine the copper content of the purified protein using a BCS assay (discussed in section 3.7.2), but unfortunately during the duration of this project a copper content was not successfully determined.

3.4.2.3 CsoR does not bind the *Rv1675c* promoter

A region of the *Rv1675c* gene promoter was selected as a control DNA to test for sequence specificity of DNA binding by CsoR. It was suggest by Festa et al. (2011) that CsoR, like RicR, can bind to DNA without sequence specificity in *in vitro* binding experiments. In order to ascertain if the interaction between CsoR and *rpfA* is sequence specific, we tested for binding of purified CsoR to another promoter region from the *Rv1675c* gene.

A 230 bp region of the *Rv1675c* promoter (designated P*Rv1675c* and sequence shown in Figure 3.14a) was amplified by PCR (Method Section 2.3.4). The primers Bio-LD49 and LS50 (sequences in Table 2.1) were used in the reaction with *M. tuberculosis* gDNA as template DNA. The products were analysed by agarose gel electrophoresis (Method Section 2.3.8). The relevant 230 bp product corresponding to P*Rv1675c* was purified by gel extraction (Methods Section 2.3.9), before use in EMSA experiments.

Gel shift experiments were carried out with P*Rv1675c* and different concentrations of purified CsoR (Figure 3.14c). The free migration of P*Rv1675c* can be observed in lane 1, where no protein in present. In the presence of CsoR (1 to 10 µM) under the conditions tested, no alterations in migration of P*Rv1675c* were observed (lanes 2 to 6), revealing that CsoR does not bind to the *Rv1675c* promoter. This supports the notion that CsoR-binding is sequence specific and that it recognises and binds only specific DNA sequences that are closely related to the consensus CsoR site.

(a)

5'-ccgccaagaatcggtttgtcagcgtgctgacagagtgcctggggcgcagaac
gggcttgcgtgttctcttcgatcgactaatttcccattcagggcaagccaacccgca
cgcaatccacgcgcgctgacaattggcagcacgatgacggcatcgcctgctctca
aggaggggccatggcagatcggtcggtgcgcccgctgcggcatctcgttcatgcg
gtgactggggg – 3'

▩ = binding region for LD50

▩ = binding region for Bio-LD49

(b)

| CsoR conc. (µM) | 0 | 1 | 5 | 10 | 5 | 10 |
| Copper (100 µM) | - | - | - | - | + | + |
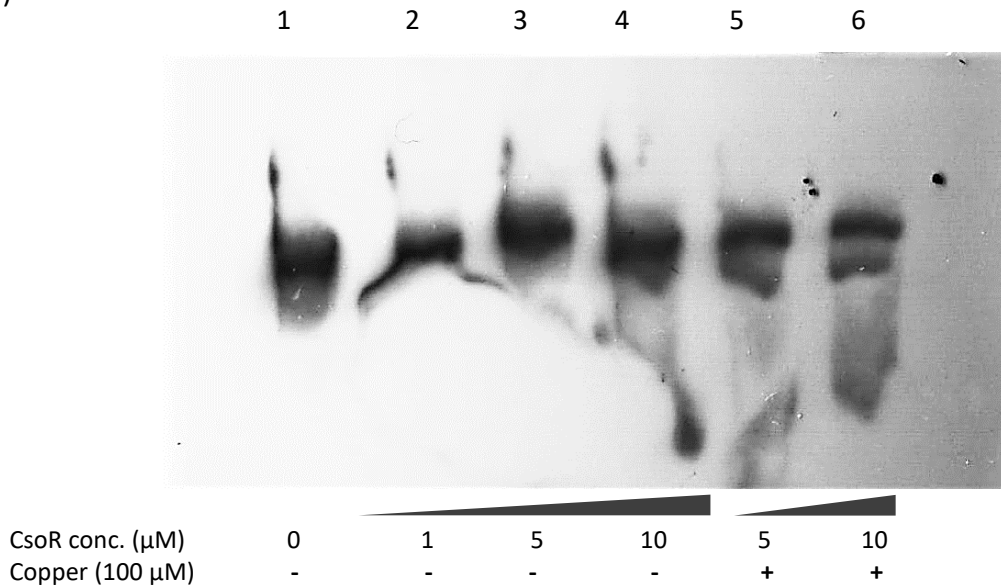
Figure 3.14 CsoR does not bind to the PRv1675c. a) Full length sequence of the region of the *Rv1675c* promoter and the primers used for amplification. This promoter region was used as a negative control to show CsoR-binding to *rpfA* is sequence specific. b) EMSA analysis of reactions containing differing CsoR concentrations and PRv1675c (1 µM) in the presence and absence of copper. No shift in migration has been observed at CsoR concentrations that retarded migration could be observed with *rpfA*, suggesting that CsoR does not bind to PRv1675c. This is representative shift from 3 replicates.
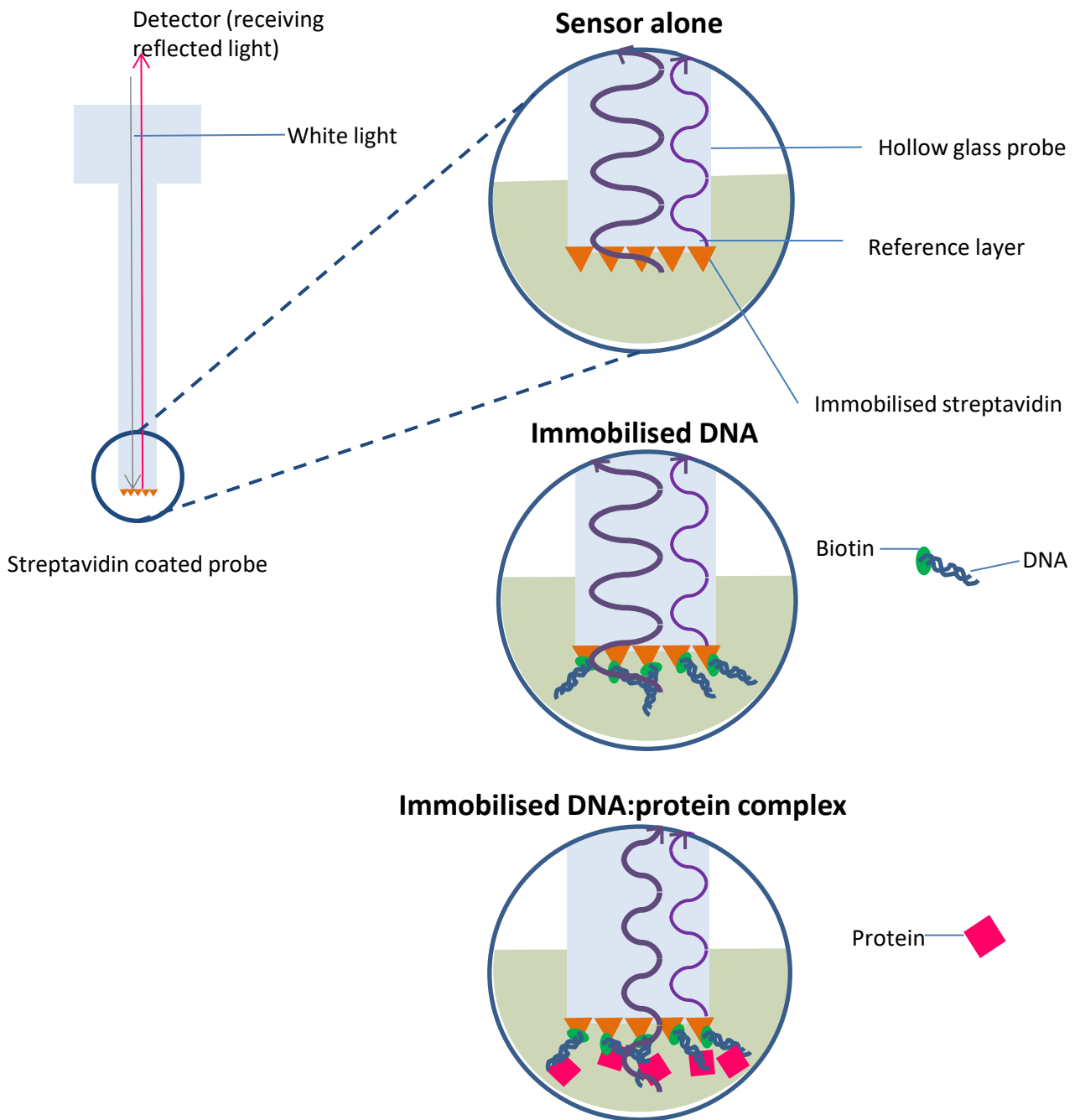
Figure 3.15: Overview of a BLItz kinetic experiment. A beam of white light is passed through the centre of a hollow probe, which has streptavidin immobilised on the tip, just below the reference layer. At each step of the experiment, the interference pattern of the white light reflected from 2 surfaces is recorded (purple lines). The two surfaces are the reference layer and the tip of the biosensor. The light reflected from the tip of the biosensor is dependent on the thickness which is dependent on the number of biomolecules that are bound. It is possible to measure interference patterns of reflected light in real time as 5' end biotinylated target promoter DNA is immobilised to the streptavidin coated biosensor tip. Interference patterns of reflected light can then be measure as the tip is placed in different concentrations of protein. If a protein binds to the DNA, shifts in interference pattern and thickness of the tip of the biosensor will be recorded. From this data, it is possible to determine the kinetics of binding and gain values for the association and dissociation constants.

3.4.3 Biolayer interferometry studies into the binding of CsoR to *rpfA*

Biolayer interferometry is a technique that allows for the real-time, label-free analysis of binding kinetics. It relies on the principle that the light reflected from a biosensor tip to a detector is dependent on the number (and thickness) of biomolecules bound to the tip. A more detailed overview of the technique is shown in Figure 3.15.

3.4.3.1 Binding kinetics of the interaction between CsoR and *rpfA$_L$*

BLItz experiments were carried out with a range of concentrations of purified CsoR from 200 µM to 250 nM and *rpfA$_L$* DNA) by the methods outlined in Section 2.5.2. Three repeats were performed and a representative trace is shown in figure 3.16. An increased thickness when probes containing immobilised DNA were dipped into CsoR samples (comparing section 3 and 4 of Figure 3.16a). The thickness increased was greater in samples with higher protein concentration, indicating that CsoR can bind to *rpfA$_L$*. This supports the findings EMSA experiments in Section 3.4.2.1. BLItz software (by ForteBio) was used to analyse the data to provide binding kinetic constants (Figure 3.16b). CsoR showed an association rate (k_a) of $8.79 \times 10^3 \pm 5.39 \times 10^2$ Ms^{-1} and a dissociation rate (k_d) of $3.62 \times 10^{-3} \pm 1.70 \times 10^{-4}$ S^{-1}. The K$_D$ was calculated to be 411 nM.

3.4.3.2 Binding kinetics of the interaction between CsoR and the *cso* site

BLItz experiments were carried out to assess the binding interaction of CsoR with 5'-end biotinylated consensus *cso* binding sequence (P*cso* 5' GTAGCCCACCCCCAGTGGGGTGGGATAC 3') from Liu et al. (2007). To produce P*cso*, oligos Bio-LD70 and LD71 were synthesised (by Eurofins) and diluted to the relevant concentration. They were heated to 94°C for 5 min and allowed to cool to room temperature to anneal before use in binding experiments. BLItz experiments were performed on a range of CsoR concentrations from 200 µM to 500 nM (Figure 3.17) by the methods outlined in Section 2.5.2. At least three repeats with ten different protein concentrations were gathered and representative traces are shown.

An increase in thickness was observed upon exposing probes containing immobilised DNA to CsoR (comparing section 3 and 4 of Figure 3.17a). A greater increase in thickness was observed at higher CsoR concentrations. This data provides evidence that CsoR can bind to P*cso*, supporting the findings of Liu et al., 2007. BLItz software (by ForteBio) was used to analyse the data to provide binding kinetic constants. CsoR showed an association rate (k_a) of $1.67 \times 10^3 \pm 5.41 \times 10^1$ Ms^{-1} and a dissociation rate (k_d) of $1.64 \times 10^{-3} \pm 7.71 \times 10^{-5}$ s^{-1}. The K$_D$ was calculated to be 985 nM. This suggests that the affinity of CsoR for the *cso* promoter is weaker than for the *rpfA* promoter.
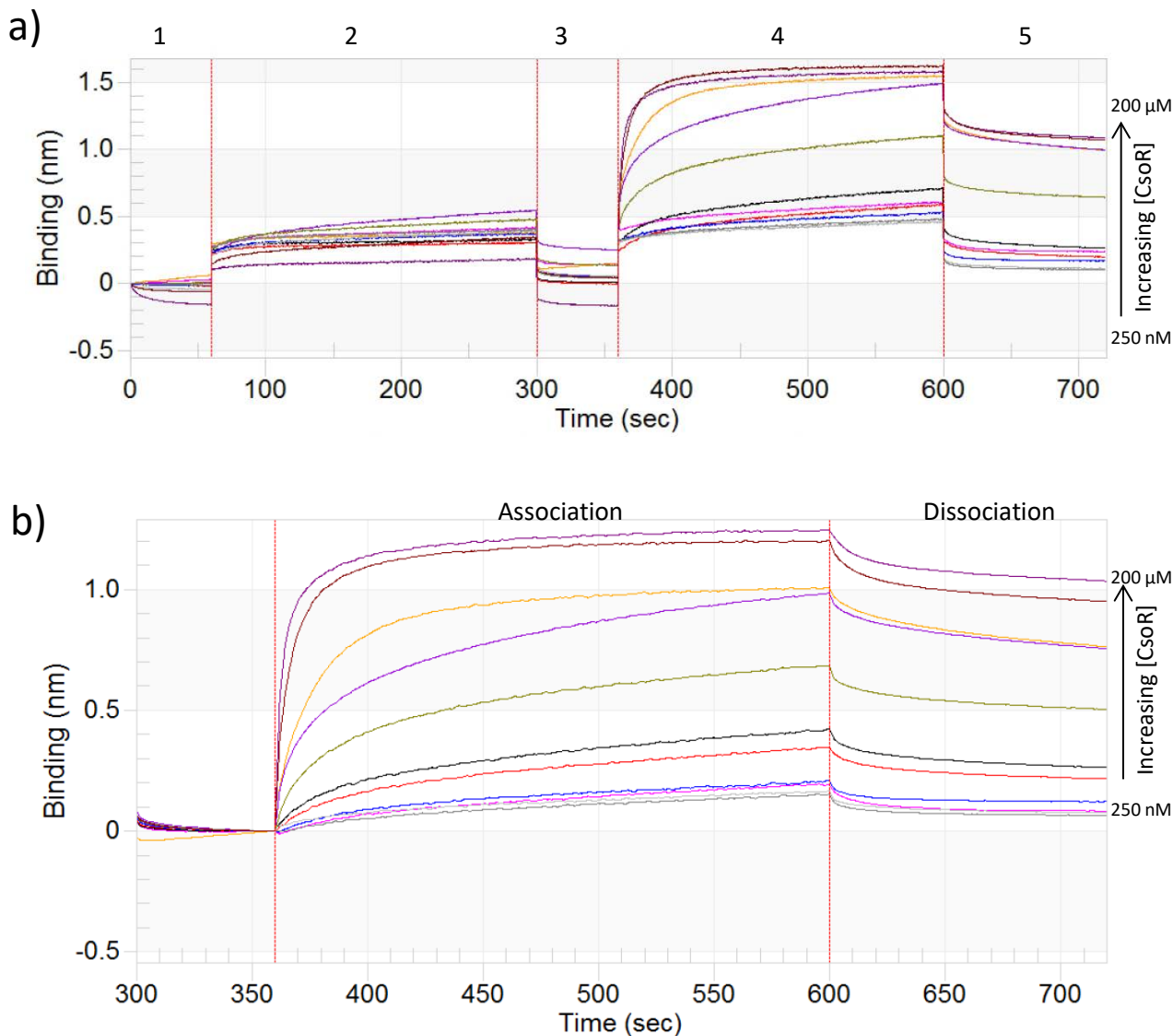
Figure 3.16: BLItz analysis of the interaction between CsoR and *rpfA*$_L$. a) Overall BLItz trace with different concentrations of CsoR (200, 100, 50, 25, 10, 5, 2.5, 1, 0.5 and 0.25 μM). Step 1: wash step. Step 2: association of DNA to probe. Step 3: wash step. Step 4: association of CsoR to DNA. Step 5: dissociation. b) Baseline-adjusted association and dissociation of CsoR with *rpfA*$_L$. Data shown are representative graphs of three independent repeats. These were analysed using BLItz software (by ForteBio) to provide an association rate (k_a) of $8.79 \times 10^3 \pm 5.39 \times 10^2$ Ms^{-1}, a dissociation rate (k_d) of $3.62 \times 10^{-3} \pm 1.7 \times 10^{-4}$ s^{-1}. and a K_D value of 411 nM. Three replicates were performed.
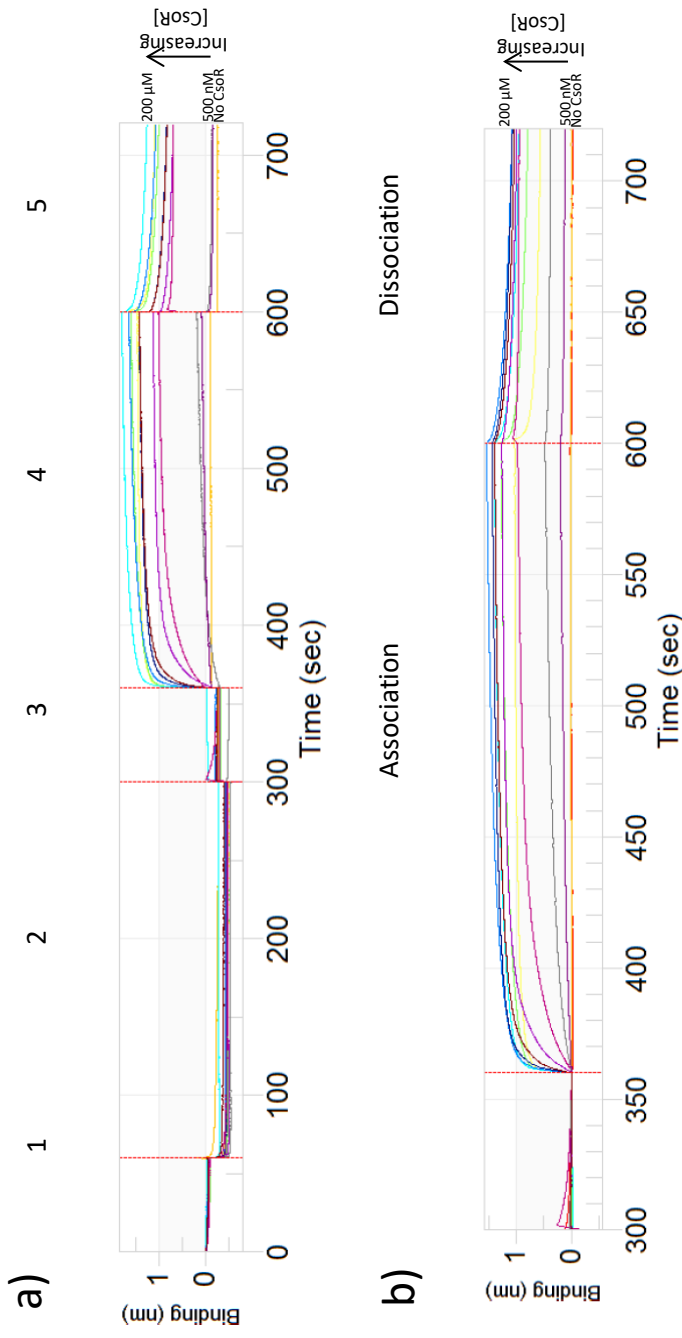
Figure 3.17: BLitz analysis of the interaction between CsoR and Pcso. a) Overall BLitz trace with different concentrations of CsoR (200, 100, 50, 25, 10, 7.5, 5, 2.5, 1, 0.5 and 0.25 µM) and when no protein is present. Step 1: wash step. Step 2: association of Pcso to probe. Step 3: wash step. Step 4: association of CsoR to Pcso. Step 5: dissociation. b) Baseline-adjusted association and dissociation of CsoR with Pcso. These were analysed using BLitz software (by ForteBio) to provide an association rate (k_a) of $1.67 \times 10^3 \pm 54.1$ Ms^{-1}, a dissociation rate (k_d) of $1.64 \times 10^{-3} \pm 7.7 \times 10^{-5}$ s^{-1} and a K_D value of 985 nM. Data shown are representative graphs of three independent repeats.

3.4.3.3 Release of CsoR from DNA in the presence of copper

To test the effect of copper on the k_a, k_d and K_D of the CsoR DNA-binding, BLItz experiments were carried out as in Sections 3.4.3.1 and 3.4.3.2, but the buffer used for the dissociation stage of the experiment was supplemented with copper chloride (100 µM). The dissociation curves of the release of CsoR (100 µM and 25 µM) from P$rpfA_L$ DNA in presence or absence of copper are shown in Figure 3.18. Control curves were produced showing how immobilised DNA (without CsoR protein) responded to the presence (purple line) or absence (orange line) of copper. Another control reaction (green line) showing the effect of copper in the buffer has on a probe without any P$rpfA$L DNA immobilised on it is shown.

Samples containing CsoR in the absence of copper (black and brown line) showed a change in thickness during dissociation of ~0.5 nm. As expected, the sample containing 100 µM CsoR (black line) had a steeper reduction in thickness than the 25 µM CsoR sample. The change in thickness for the CsoR samples was greater in the presence of copper (~1.25 nm, blue and red lines) than when copper was not supplemented in the buffer (~0.5 nm). This supports the notion that copper binding causes the release of CsoR from the DNA complex that was observed during EMSA experiments (section 3.4.2.1) and has been proposed in other studies (Liu et al., 2007b).

Initially the thickness decreased by ~1.25 nm in the presence of copper and the curve was steeper for the sample containing the higher CsoR concentration (100 µM (blue line) compared to 25 µM (red line)). However, after the initial decrease in thickness, the curves begin to show an increase in thickness. This was unexpected and controls were performed where just DNA was present (orange line), DNA and copper were present (purple line) or just copper and no immobilised DNA was present (green line) during the dissociation. These revealed that the presence of copper during dissociation caused an increase in thickness. This could be explained by the positively charged copper ions interacting with the negatively charged DNA or interacting directly with the streptavidin probe. This increase in thickness observed in the presence of copper explains why after an initially steep drop in thickness during dissociation in the presence of copper, after ~10 s there was an increase in thickness, which is suggested to be due to the presence of copper in the reaction. The combination of these two factors complicates the analysis and meant that further deconvolution would be necessary for accurately determining the effect of copper on the rate of dissociation (k_d) and hence K_D of the interaction between CsoR and $rpfA_L$.

From this experiment, it is possible to draw the conclusion that the initial rate of dissociation for CsoR and $rpfA_L$ is increased in the presence of copper. However, the effects of copper on
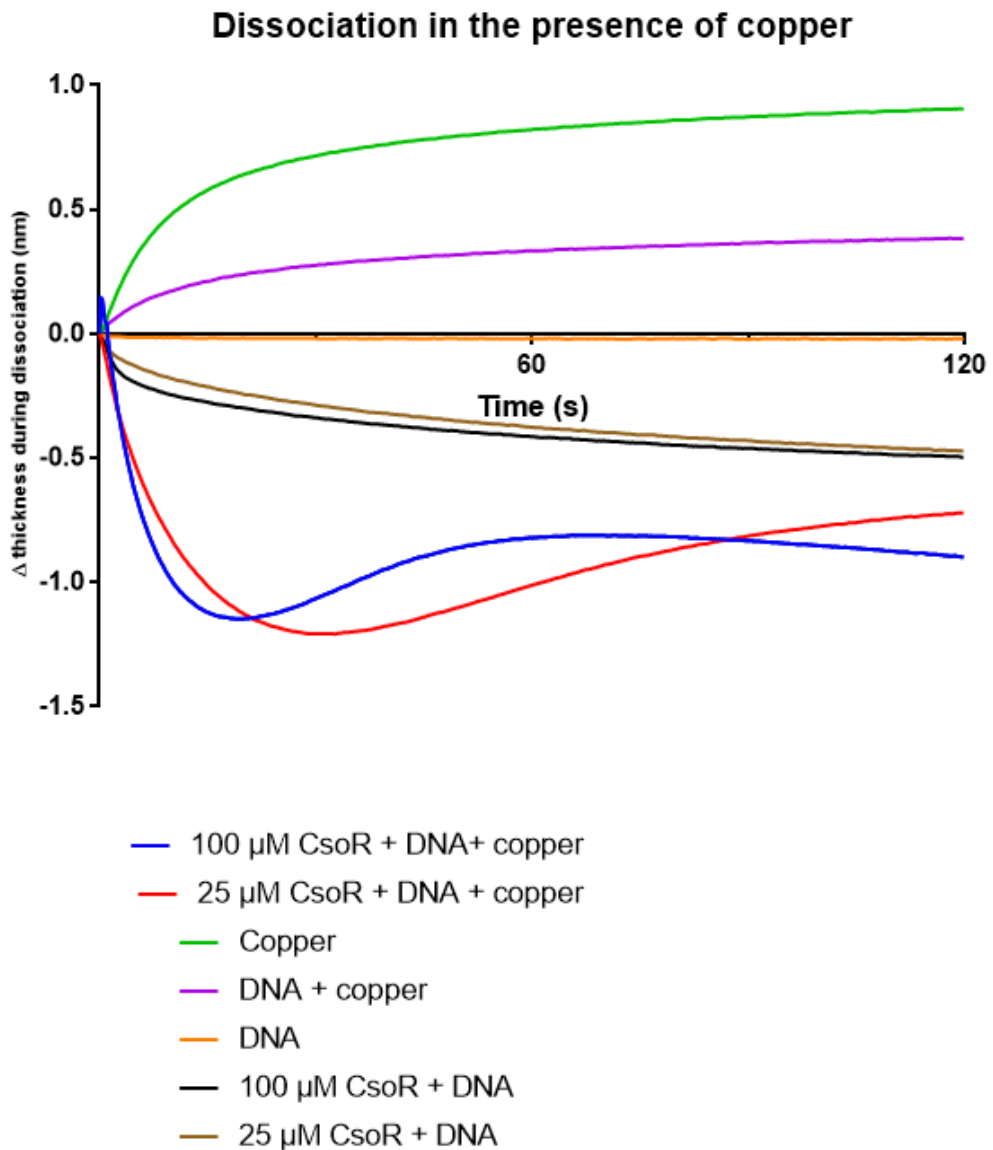
Figure 3.18: Differences in the dissociation CsoR and *rpfA*$_L$ and the effect of copper during dissocation. Data shows the change in binding (thickness in nm) during the dissocation stage of BLItz experiments. *rpfA*$_L$ DNA (100 nM) was immobilised to the streptavidin probe as in previous experiments. Copper was included at a concentration of 100 µM in the buffer used for the dissociation step in indicated samples (blue, red, purple and green lines). CsoR concentrations used are indicated in the key. Data shown are representative graphs of three independent repeats.

the thickness of the probe (without protein or DNA) has meant that it has not been possible to quantifiably establish the effect of copper on the dissociation rates between CsoR with P$rpfA_L$ using this technique.

3.4.3.4 Binding kinetics of the interaction between CsoR and control *rv1675c* promoter DNA

BLItz experiments were carried out in the same methods as outlined in Results Section 3.4.3.1, but instead of *rpfA*$_L$ DNA being used, a 230 base pair region of the *Rv1675c* promoter was used as a control for site specific DNA-binding. DNA was generated like for the EMSA experiments (Section 3.4.2.3) and was diluted to 100 nM for use in BLItz experiments.

BLItz experiments were carried out with a range of concentrations of purified CsoR from 200 μM to 200 nM and unrelated promoter DNA (P*Rv1675c* DNA) containing no CsoR site as a negative control (data not shown). An interaction between the control promoter and CsoR was not observed under the conditions tested and copper had no effect.

3.5 The putative role of phosphorylation of CsoR in *rpfA* regulation

3.5.1 CsoR as a target for phosphorylation

Unpublished data provided by Dr Galina Muklamova at The University of Leicester, revealed that CsoR is a putative substrate for phosphorylation by a member of the PknA to F kinase family. The residue of CsoR that is phosphorylated is T93, which is located at the beginning of the C-terminal tail of the protein region of the protein and is indicated in Figure 3.1. This tail region is not present in CsoR proteins from other species, such as *B. subtilis*, *L. monocytogenes*, *S. aureus* or *S. lividans* (Dwarakanath et al., 2012).

In many biological systems, phosphorylation can act as a switch to activate or deactivate a protein involved in gene regulation and interestingly the C-terminal region of *M. tuberculosis* CsoR was prone to proteolysis (Section 3.3.3). During this study, site-directed mutagenesis was used to substitute the putative phosphorylation target threonine 93 to aspartate (D) or glutamate (E). These mutations were selected as they are phosphomimetic and can act as a model for the phosphorylated CsoR protein. The aim of this part of the study was to ascertain if the phosphorylation state of CsoR affects its stability, activity or binding affinity to DNA targets and hence potentially play a role in the regulation of the *cso* operon, *rpfA* or other CsoR-regulated genes.
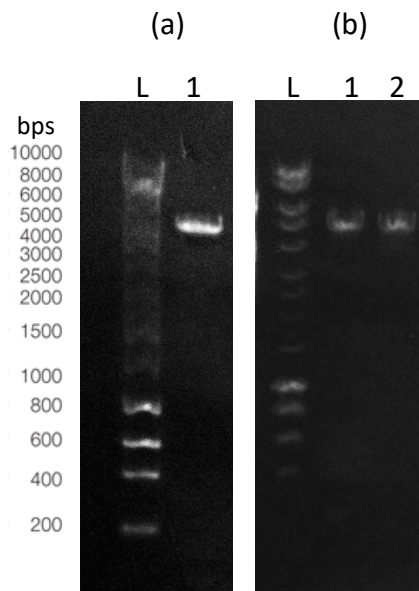
Figure 3.19: Generation of constructs for overexpression of phosphomimetic CsoR proteins (a) Agarose gel electrophoresis analysis of linearised pGS2528 DNA. Lane L contained 5 µl DNA Hyperladder I (Bioline) and lane 2 contained 10 µl of the sample of pGS2528 which had been treated with XbaI for 1 h at 37°C. (b) Agarose gel electrophoresis analysis of linearised pGS2528 after site-directed mutagenesis to introduce either T93D or T93E mutation. Lane L contained 5 µl DNA Hyperladder I (Bioline), lanes 2 and 3 contained 10 µl of linearised pGS2528 T93D and pGS2528 T93E respectively. The plasmids were linearised by treatment with XbaI for 1 h at 37°C.

3.5.2 Generation of T93D and T93E site-directed mutants

Overnight cultures of an E. coli DH5α strain containing the pGS2528 plasmid were set up. The pGS2528 plasmid map is shown in Figure 3.5 and contains the wildtype CsoR protein coding region cloned between the XbaI and SwaI cloning sites. Plasmid mini preps (Method Section 2.3.2) were performed to extract pGS2528 DNA from the cells. Restriction digests using XbaI were performed on the plasmid DNA (Method Section 2.3.3) and the presence of the correct size DNA fragment was verified by agarose gel electrophoresis (Method Section2.3.8) (Figure 3.19a).

Site-directed mutagenesis was performed using pGS2528 as template DNA using the Quikchange XL II site-directed mutagenesis kit (by Agilent). Primers LD72 and LD73 were used for generate the T93D mutant (switching the ACC codon to GAC). Primers LD70 and LD71 were used to generate the T93E mutant (switching the ACC codon to GAA). Sequences of primers are listed in Table 2.1. Single colonies were streaked onto ampicillin-containing agar plates. Plasmid mini preps were used to extract plasmid DNA from overnight cultures. Plasmid DNA was treated with XbaI (Method Section 2.3.3) and was analysed by agarose gel electrophoresis (Method Section 2.3.8) to check for the presence of a DNA species of ~5 kb, corresponding to the T93D pGS2528 and T93E pGS2528 plasmids (Figure 3.19b). Lanes 1 and 2 both contained DNA species of ~5 kb, which is the relevant size for the mutant plasmids. Plasmid DNA was sequenced (by GATC Biotech) to confirm the presence of the desired mutations, but no other undesirable mutations.

3.5.3 Purification of CsoR T93D and T93E

CsoR T93D and T93E proteins were overexpressed in *E. coli* BL21 λDE3 cells containing the T93 D pGS2528 and T93E pGS2528 overproduction plasmids using the same overproduction method as developed for WT CsoR (discussed in section 3.3.4).

The phosphomimetic CsoR proteins were purified by the same methodology that was used for the purification of stable and full length WT CsoR protein. These are discussed in Section 3.3.4 and Method Section 2.4.9. Purification was carried out with the help of Dr Svetlana Sedelnikova. After PEI precipitation, ASP, heparin affinity chromatography, ion exchange chromatography and gel filtration, the purity of the CsoR was assessed via SDS-PAGE (Figure 3.20a). Similarly to the WT CsoR sample, the purity of both CsoR T93D and CsoR T93E exceeded 90%. To confirm that the full length proteins had been produced, samples of purified CsoR T93D and CsoR T93E were sent for mass spectroscopy analysis (Figure 3.20, b&c). The CsoR T93D sample contained one major species of the expected size (12,696.77 Da) as did
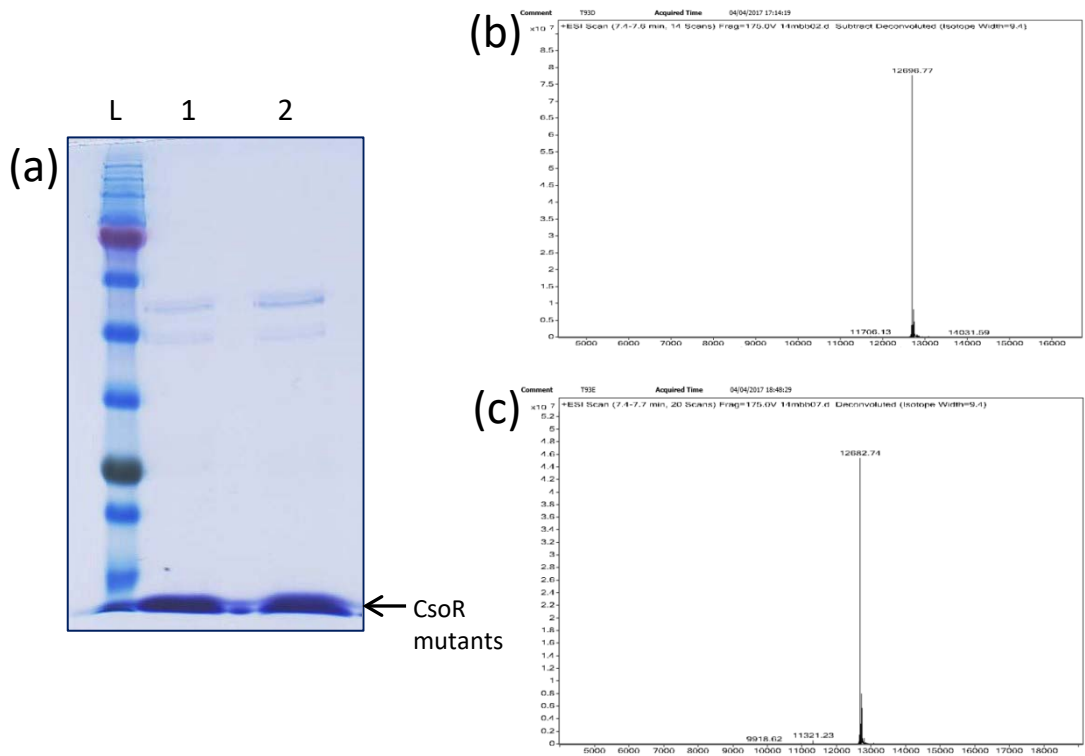
Figure 3.20: Purification of full-length CsoR T93D and CsoR T93E. (a) SDS-PAGE analysis of CsoR T93D and CsoR T93E after purification steps. Lane L contained 10 μl Precision Protein Standards (BioRad). Lane 1 and 2 contained 10 μl sample of purified CsoR T93D and CsoR T93E respectively. The migration of full length mutant CsoR proteins are indicated with an arrow. (b) Mass spec analysis of CsoR T93D. One major species of size 12,696.77 Da is present corresponding to full-length CsoR T93D. (c) Mass spec analysis of CsoR T93E. One major species of size 12,682.74 Da is present corresponding to full-length CsoR T93E.

CsoR(μM)	0	1	2.5	5	10	5	10	0	1	2.5	5	10	5	10
Cu(I)(100μM)	-	-	-	-	-	+	+	-	-	-	-	-	+	+

ΔT93D   ΔT93E
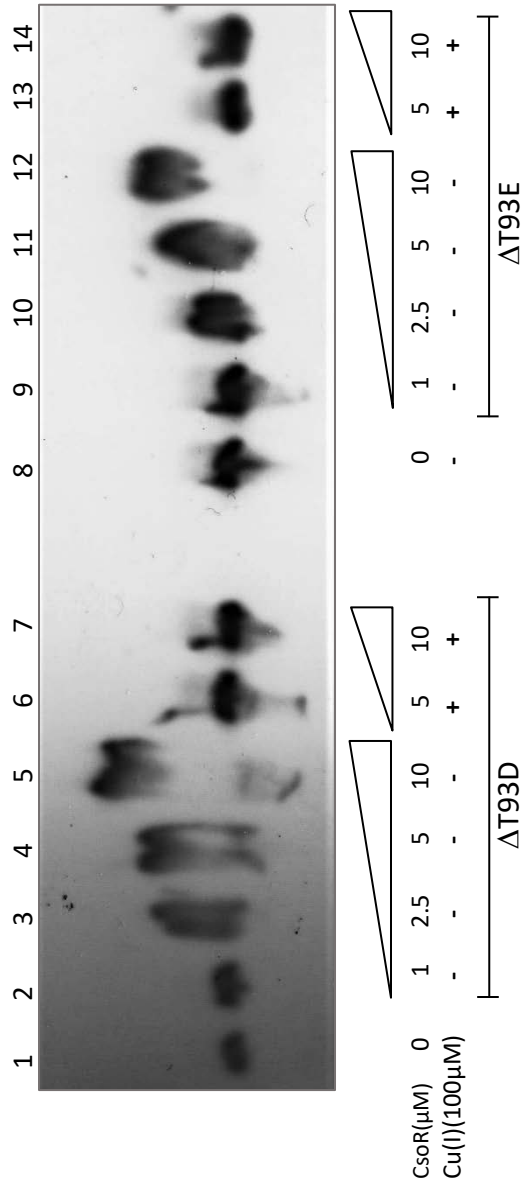
Figure 3.21: EMSA analysis of the interaction between phosphomimetic CsoR mutants (T93D and T93E) and *rpfA*L. Biotinylated *rpfA*L (1 μM) was incubated with different concentrations of CsoR protein (1-10 μM) in the presence or absence of copper chloride (100 μM). The protein concentrations provided are for CsoR monomers. Three replicates were performed.

CsoR T93E (expected size 12,682.74 Da). This indicated the successful purification of full length CsoR T93D and CsoR T93E.

Mutant proteins were stored by the same methods as for WT CsoR, discussed in Section 3.3.4.

3.5.4   EMSA analysis of binding of $rpfA_L$ by phosphomimetic CsoR mutants

To determine if the phosphorylation of CsoR plays a regulatory role by altering the DNA binding ability or affinity of the protein, the ability of the phosphomimetic mutants to bind $rpfA$ DNA ($rpfA_L$) was assessed via EMSA experiments (Figure 3.21). These were carried out under the same conditions used for WT CsoR in Section 3.4.2. When incubated with CsoR T93D (concentrations ≥2.5 µM) there was a shift in the migration of $rpfA_L$ (lanes 3 to 5) compared to the migration when no protein is present (lane 1). When incubated with CsoR T93E (concentrations ≥2.5 µM) there was a shift in the migration of $rpfA_L$ (lanes 10 to 12) compared to when no protein is present. The shifts were observable at mutant CsoR monomer concentrations above 2.5 µM. The alteration in migration of $rpfA_L$ in the presence of CsoR T93D and CsoR T93E suggests that both phosphomimetic CsoR mutants bind the CsoR site present in the $rpfA$ gene. The initial shift for both CsoR T93D and CsoR T93E occurred at the same concentration (2.5 µM) as that of WT CsoR, which implies the binding affinity in a similar range. In order to ascertain a more accurate K_D value, BLItz analysis was performed on the phosphomimetic mutant proteins (shown in Section 3.5.5).

No shift in migration of $rpfA_L$ was observed in the presence of either CsoR T93D or CsoR T93E when copper was included in EMSA reactions (Lanes 6 to 7 and 13 to 14). This suggests that, similarly to the WT CsoR, the proteins are released from the CsoR site when copper binds to the protein. If the phosphomimetic mutations are an accurate representative of native phosphorylated CsoR, then it is likely that if phosphorylation does play a regulatory role, then its method of action is not by affecting the copper sensitivity of the protein. It is plausible that phosphorylation may affect the stability of the C-terminal tail region, which is often lost during purification Initial observations suggested the mutant protein seems to be stable for longer after purification but further experiments would be needed to clarify this.

3.5.5   BLItz analysis with phosphomimetic CsoR mutants

BLItz experiments were carried out with a range of concentrations of mutant CsoR proteins (100 µM to 500 nM) and $rpfA_L$ DNA (Figure 3.22a) using the methods outlined in Section 2.5.2.

An increased thickness when probes containing immobilised DNA were dipped into CsoR T93D and CsoR T93E samples (comparing sections 3 and 4 of the graphs in Figure 3.22a and 3.22b). The thickness increase was greater in samples with higher protein concentrations, indicating

a)

BLItz data of the interaction between ΔT93D CsoR with P*rpfA*$_L$

b)

BLItz data of the interaction between ΔT93E CsoR with P*rpfA*$_L$

c)
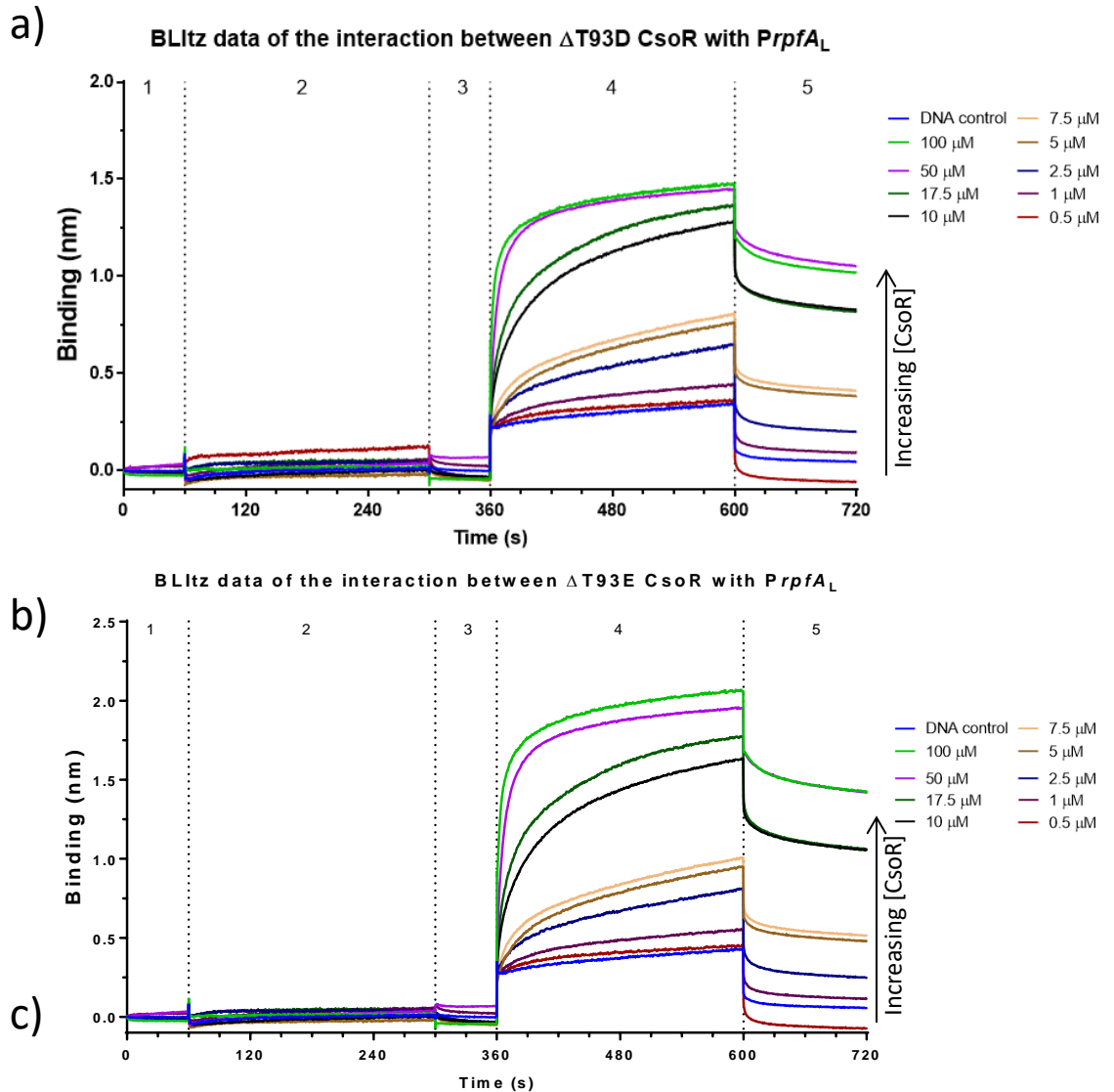
Protein	DNA	k_a (Ms^{-1})	k_d (s^{-1})	K_D (nM)
WT CsoR	*rpfA*$_L$	$8.79 \times 10^3 \pm 5.39 \times 10^2$	$3.62 \times 10^{-3} \pm 1.70 \times 10^{-4}$	411
WT CsoR	*cso* promoter	$1.67 \times 10^3 \pm 5.41 \times 10^1$	$1.64 \times 10^{-3} \pm 7.71 \times 10^{-5}$	985
CsoR-T93D	*rpfA*$_L$	$2.79 \times 10^3 \pm 1.01 \times 10^2$	$1.57 \times 10^{-3} \pm 8.80 \times 10^{-5}$	561
CsoR-T93E	*rpfA*$_L$	$5.37 \times 10^3 \pm 7.73 \times 10^1$	$1.02 \times 10^{-3} \pm 1.77 \times 10^{-5}$	189

Figure 3.22: BLItz analysis of the interaction between the phosphomimetic CsoR mutants and *rpfA*$_L$. a & b) BLItz trace with different concentrations of CsoR T93D (in a) and CsoR T93E (in b). Concentrations of 100, 50, 17.5, 10, 7.5, 5, 2.5, 1 and 0.5 µM were used. The BLItz steps are as follows. Step 1: wash step. Step 2: association of DNA to probe. Step 3: wash step. Step 4: association of CsoR to DNA. Step 5: dissociation. Data shown are representative graphs of two independent repeats for each mutant protein. c) Curves of the association and dissociation were analysed using BLItz software (by ForteBio) to provide the association rates (k_a), dissociation rates (k_d) and K_D values listed in c.

that both CsoR mutants bound to $rpfA_L$. This supports the findings of the EMSA experiments in Section 3.4.2. BLItz software (by ForteBio) was used to analyse the data to provide binding kinetic constants (Figure 3.22c). CsoR T93D showed an association rate (k_a) of $2.79 \times 10^3 \pm 1.01 \times 10^2$ Ms^{-1} and a dissociation rate (k_d) of $1.57 \times 10^{-3} \pm 8.80 \times 10^{-5}$ s^{-1}. The K_D was calculated to be 561 nM. This is a similar K_D value for the interaction between the WT CsoR protein and $rpfA_L$ (411 nM). CsoR T93E showed an association rate (k_a) of $5.37 \times 10^3 \pm 7.73 \times 10^1$ Ms^{-1} and a dissociation rate (k_d) of $1.02 \times 10^{-3} \pm 1.77 \times 10^{-5}$ s^{-1}. The K_D was calculated to be 182 nM. This affinity is at least two-fold tighter than for the interaction between the WT CsoR protein and $rpfA_L$ (411 nM).

3.6 Conclusions

In this chapter untagged *M. tuberculosis* CsoR protein has been successfully overproduced in an *E. coli* host. A protocol for the purification of CsoR from cell-free extract has been developed. Regulatory sites present in the *rpfA* gene have been mapped. EMSAs showed that CsoR is able to bind the predicted CsoR site present in the *rpfA* gene. In the presence of Cu(I) the CsoR is released from the CsoR:DNA complex. It took up to 0.5X Cu(I) per CsoR monomer to cause complete dissociation from the *rpfA* DNA, suggesting that a proportion of the CsoR protein purified may be in the copper-bound form and for this reason it would be useful if further experiments could be performed to allow quantification of copper levels in the purified protein (discussed in Section 3.7.2).

BLItz experiments supported the findings from the EMSA experiments and suggested a K_D value of ~400 nM for the interaction between CsoR and the *rpfA* promoter. Similar experiments using the *cso* promoter showed that purified CsoR protein could bind with a K_D of ~1000 nM.

In order to determine if phosphorylation of CsoR at residue T93 by a Pkn family kinase affected the DNA-binding or copper sensitivity of CsoR, constructs that would allow for the overproduction of phosphomimetic mutants (T93D and T93E) were generated by site-directed mutagenesis. EMSAs and BLItz experiments were carried out on both mutants, which showed both were able to bind to the CsoR site in the *rpfA* gene and that this interaction was released when Cu(I) was added to reactions. A K_D value of 561.3 nM (comparable to WT CsoR) for CsoR T93D and of 182.2 nM (at least two fold tighter than WT CsoR) for CsoR T93E were obtained.

3.7 Discussion and future work

3.7.1 Crystal structures

The current published crystal structure of *M. tuberculosis* CsoR is missing the C-terminal tail region of the protein (Liu et al., 2007b). However, it has been proposed that CsoR binds to DNA targets as a dimer of dimers and that the C-terminal tail plays a role in the association between the two dimers (Higgins and Giedroc, 2014). Additionally, during this study we observed that loss of the C-terminal tail region results in loss of the DNA-binding ability. For this reason, it would be useful to use the purification methods developed in this study to produced enough purified protein for crystal trials to be performed, with the ultimate aim of producing a complete crystal structure for the protein. Determining the structure of this tail region could provide useful insights into the association between CsoR dimers. Similarly, it would also be useful to obtain a crystal structure of the CsoR:DNA complex.

3.7.2 Copper content of purified CsoR proteins

The gel filtration indicated that the purified CsoR protein was in an octomeric form, which supports the notion that the CsoR was purified in a copper-free state, because the incorporation of copper has been suggested to cause the octomer to dissociate (Liu et al, 2007).

However, as mentioned in Section 3.4.2.2, we observed that 0.5X Cu(I) per CsoR monomer was sufficient to abolish retarded migration in EMSA experiments. It is the general consensus in the literature that each monomer of CsoR binds 1X Cu(I) (Higgins and Giedroc, 2014). Although steps were taken to minimise levels of copper present during overproduction and purification of CsoR, based on the EMSA results it is likely that a proportion of the purified protein is in the copper-bound form. For this reason, further work to quantify the ratio of copper in the purified samples would be useful. During this study, efforts were made to quantify the copper levels within protein samples via bathocuprione disulphonate (BCS) fluorescence assay that was modified from Rapisarda et al. (2002). A standard curve was produced to be used for quantification of Cu(I) concentrations, but unfortunately due to the high amount of protein required per sample and time constraints, the copper content of purified protein was not successfully determined. Further work should include producing sufficient protein to allow for the BCS assay to determine the copper content. ICP-MS analysis of purified protein samples could be performed as an additional method to determine the copper content of purified CsoR proteins. This would provide greater insight into why we observed 0.5X copper being sufficient to release CsoR from the *rpfA*.

3.7.3 Phosphorylation of CsoR

During this study, phosphomimetic mutants, which mimic the T93 residue after phosphorylation were characterised (Section 3.5). However, it would be interesting to performed experiments to phosphorylate CsoR with the relevant Pkn kinase. This would involve overproduction and purification of Pkn proteins and setting up *in vitro* phosphorylation reactions using [γ-^{32}P] ATP as the substrate. The use of a radiolabel would allow for easy determination of successful phosphorylation of the protein. This would be useful as it would allow confirmation as to whether CsoR is a target for phosphorylation by a Pkn kinase.

DNA binding by the phosphorylated protein could then be characterised by the same methods used during this study and compared to the data collected for the phosphomimetic mutants. The CsoR T93E mutant showed over two-fold tighter binding to *rpfA*$_L$ compared to WT CsoR, but CsoR T93D showed a similar affinity. Performing the experiments with phosphorylated proteins would reveal which phosphomimetic mutant behaved more like natively phosphorylated protein.

3.7.4 Stability of the tail region

The C-terminal tail region of CsoR may be important both in the phosphorylation of CsoR and potentially in stabilisation of the protein. Our preliminary observations during this study suggested that the phosphomimetic proteins did not lose the C-terminal tail region as quickly as the WT protein. Additionally they were still functional two weeks after purification, by which time WT CsoR preparations were no longer functional. Further experiments to confirm this and determine the importance of this tail region would be beneficial. These experiments could include testing the functionality (by EMSA or BLItz) of WT, phosphomimetic mutants and *in vitro* phosphorylated CsoR proteins at different time lengths after purification and analysing the protein samples tested by SDS-PAGE (and/or mass spectrometry). These experiments would determine if the phosphorylation affects the stability of the tail region of CsoR and if this in turn has an effect on the DNA binding ability of the protein.

4 Characterisation of mycobacteria cyclic-AMP receptor proteins (CRPs)

4.1 Introduction

In this chapter the role of mycobacterial cyclic AMP receptor proteins (or CRPs) will be explored, with particular interest in the regulation of *rpfA* by Rv3676 and in comparing the regulatory roles of Rv3676 (from *M. tuberculosis*) and Msmeg_6189 (from *M. smegmatis*) with the well-characterised *E. coli* CRP. It should be noted that the use of 'cyclic AMP' or 'cAMP' is referring to 3'5'-cyclic adenosine monophosphate. If referring to 2'3'-cyclic adenosine monophosphate, the 2'3' is always stated.

4.1.1 *Escherichia coli* CRP

Escherichia coli CRP, encoded by *crp*, is a classic example of bacterial gene regulatory system and is arguably the best characterised protein from the CRP/FNR family of transcription factors. The main role of CRP is controlling the expression of multiple genes with a plethora of biological functions in response to intracellular cyclic AMP concentrations (Gosset et al., 2004). *Escherichia coli* CRP is not essential for growth under standard conditions, but allows the bacterium to exploit a wider range of carbon sources when the preferred source, glucose, is exhausted (Sabourin and Beckwith, 1975)

The *E. coli* genome encodes just one adenylate cyclase (CyaA, encoded by *cyaA*), which is able to synthesis cyclic AMP from ATP (Roy et al., 1983). *Escherichia coli* CRP is a cyclic AMP responsive protein, where binding cyclic AMP enables the protein to interact with promoter DNA that contains the CRP consensus sequence (TGTGANNNNNNTCACA) (Berg and von Hippel, 1988). The protein contains a cyclic nucleotide binding site and a helix-turn-helix DNA-binding domain. Once bound to target promoters, it is able to play a role in RNAP recruitment and promotes transcription of target genes (Busby and Ebright, 1999). Although binding at the vast majority of target promoters results in activation, in some cases binding of CRP causes promoter occlusion and subsequent transcriptional repression (Kolb et al., 1993).

In the apo-CRP dimer, the DNA-binding helices are buried beneath the surface of the protein, rendering them unable to bind to target DNA, as can be seen in Figure 4.1 (taken from Green et al., 2014). cyclic AMP binding to *E. coli* CRP results in allosteric changes in the structure of the protein, during which the DNA recognition helices are exposed on the surface of the protein. Cyclic AMP binding to CRP results in ~1000-fold enhancement in DNA-binding, which is logical as the DNA-binding helices are positioned in an optimal location for interacting with the CRP site in the DNA (Green et al., 2014).

Apo-CRP   cAMP-CRP
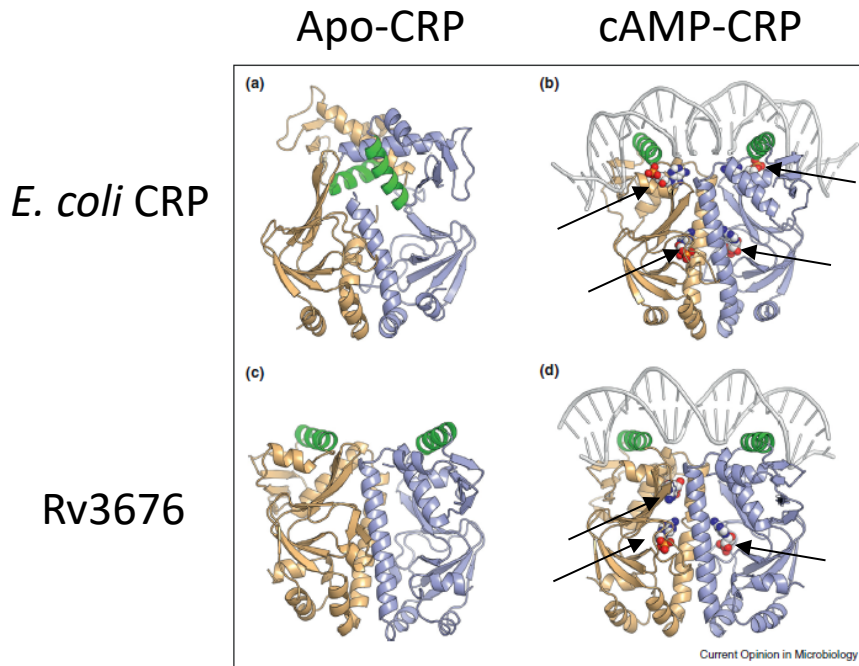
E. coli CRP

Rv3676

Taken from Green et al, 2014

Figure 4.1: X-ray crystal structures of apo- and holo- *E. coli* **and** *M. tuberculosis* **CRP proteins.** (a) *Escherichia coli* apo-CRP (PDB ID 3HIF). (b) *Escherichia coli* cAMP-CRP (PDB ID 2CGP). (c) *Mycobacterium tuberculosis* apo-Rv3676 (PDB ID 3D0S). (d) *Mycobacterium tuberculosis* cAMP-Rv3676 (PDB ID 3MZH). One protomer is shown in brown and the other in blue. The DNA recognition helices are shown in green. The cyclic AMP molecules bound to the CRP structures in (b) and (d) are indicated with arrows.

4.1.2 Rv3676 from *M. tuberculosis*

Sequencing of the *M. tuberculosis* genome revealed it encoded orthologues of previously characterised regulatory genes, which included *Rv3676* (Cole et al., 1998). Rv3676 is a member of the CRP/FNR family of transcription regulators and contains a potential cyclic AMP-binding site and a helix-turn-helix DNA-binding domain. Although Rv3676 is not essential, deletion of *Rv3676* in *M. tuberculosis* H37Rv caused growth defects in both laboratory cultures and bone-marrow derived macrophages and was attenuated in mouse models of infection (Rickman et al., 2005). These effects could be successfully complemented. Rv3676 is required for normal growth and multiplication within mouse models and macrophage, but is not necessary for the persistence stage of growth (Rickman et al., 2005).

Rv3676 is a homo-dimer, made up of two subunits. Each protomer is ~25 kDa and is capable of binding one molecule of cyclic AMP (Stapleton et al., 2010). The presence of a regulatory protein capable of binding cyclic AMP, suggests *M. tuberculosis* is able to respond to intracellular cyclic AMP concentrations and that the ligand plays a role in metabolic processes within the bacterium. One example of this is in response to glucose levels, as decreases in cyclic AMP levels within the pathogen have been observed upon the addition of glucose (Padh and Venkitasubramanian, 1976). The notion that cyclic AMP signalling is important in mycobacteria is supported by the fact that the *M. tuberculosis* genome encodes ≥17 adenylate cyclase enzymes (Cole et al., 1998), compared to just one adenylate cyclase (encoded by *cyaA*) in the *E. coli* genome. At least one of the adenylate cyclase enzymes encoded in the *M. tuberculosis* genome is essential for virulence (Agarwal et al., 2009).

Structural information

Rv3676 has 32% sequence identity with *E. coli* CRP and 53% similarity (alignment in Figure 4.2). The majority of the differences are located at the N-terminal region of the protein, which, in *E. coli* CRP, have been implicated in interacting with RNA polymerase to activate transcription at target promoters via residues H19 and H21 (Busby and Ebright, 1999). The differences in the N-terminal region of the proteins imply that Rv3676 may interact with the mycobacterial RNAP polymerase in a different manner to the *E. coli* system. Rv3676 is not functional *in vivo* in *E. coli* cells and was unable to complement a CRP deletion mutant (Spreadbury et al., 2005). If the two proteins interact with RNAP in different ways it could explain the inability of Rv3676 to complement the CRP deletion mutant in *E. coli*.

Four of the six key residues involved in cyclic AMP binding (Schultz et al., 1991) in *E. coli* CRP are conserved in Rv3676 and two out of the three residues involved in direct DNA-binding (Schultz et al., 1991) are conserved, whilst the third is similar in charge (Figure 4.2). The

```
Rv3676:   1                                      VDEILARAGIFQGVEPSAIAALTKQLQPVD   30
                                                              +
crp:      1                                      MVLGKPQTDPTLEWFLSHCHIHK         23

Rv3676:  31  FPRGHTVFAEGEPGDRLYIIISGKVKIGRRAPDGRENLLTIMGPSDMFGELSIFDPGPRT   90
             +P    T+  +GE  + LY I+ G V +  +  +G+E +L+ +    D  GEL +F+ G
crp:     24  YPSKSTLIHQGEKAETLYYIVKGSVAVLIKDEEGKEMILSYLNQGDFIGELGLFEEGQER   83

Rv3676:  91  SS---ATTITEVRAVSMDRDALRSWIADRPEISEQLLRVLARRLRRTNNNLADLIFTDVP  147
             S+   A T  EV +S +  R  I   P+I +L   +ARRL+ T+  + +L F DV
crp:     84  SAWVRAKTACEVAEISYKK--FRQLIQVNPDILMRLSAQMARRLQVTSEKVGNLAFLDVT  141

Rv3676: 148  GRVAKQLLQLAQR---FGTQEGGALRVTHDLTQEEIAQLVGASRETVNKALADFAHRGWI  204
             GR+A+ LL LA++       +G  +++    T++EI Q+VG SRETV + L     +  I
crp:    142  GRIAQTLLNLAKQPDAMTHPDGMQIKI----TRQEIGQIVGCSRETVGRILKMLEDQNLI  197

Rv3676: 205  RLEGKSVLISDSERLARRAR                                          224
               GK++++
crp:    198  SAHGKTIVVYGTR                                                210
```

Figure 4.2: Sequence alignment of Rv3676 from *M. tuberculosis* and CRP from *E. coli*. Figure taken from Rickman et al. (2005). Residues shown to be involved in the binding of cyclic nucleotides, including cyclic AMP, in *E. coli* CRP are highlighted in grey. Those conserved (4 out of 6) in *M. tuberculosis* Rv3676 are highlighted in grey. Residues R83 (in *E. coli* or T90 in *M. tuberculosis*) and S129 (in *E. coli* or N135 in *M. tuberculosis*) are not conserved and mutants of these are discussed in Results Section 4.6). The amino acids (Arg180, Glu181 and Arg185) of *E. coli* CRP directly involved in DNA recognition are indicated by underlining.

conservation in these residues support the notion that Rv3676 is able to bind cyclic AMP (discussed in Section 4.1.2.1) and recognise and interact with a similar DNA consensus to *E. coli* CRP (Bai et al., 2005) (discussed in Section 4.1.2.2).

4.1.2.1 Cyclic AMP binding

Cyclic AMP binding induces a conformation change in structure of *E. coli* CRP, which allows it to interact with specific target DNA sequences (Kolb et al., 1993; Krakow and Pastan, 1973). Allosteric changes in Rv3676 occur upon cyclic AMP binding. Partial proteolysis experiments, where Rv3676 and DNA were treated with trypsin, revealed reduced proteolysis of Rv3676 and partial protection of Rv3676 DNA-binding functionality in the presence of cyclic AMP. The DNA-binding function of Rv3676 was lost in the absence of cyclic AMP and greater levels of proteolysis were observed (Bai et al., 2005). Other studies have demonstrated that cyclic AMP is able to bind to Rv3676 (Stapleton et al., 2010).

4.1.2.2 Targets of Rv3676

Consensus site and DNA-binding

In *E. coli* CRP, the contacts made between the DNA-binding helix of CRP and specific nucleotides in the DNA is essential for binding (Jansen et al., 1987) and as these DNA-binding residues are largely conserved in Rv3676 (shown in Figure 4.2), it is likely that these contacts are important for Rv3676 binding to promoters. Studies have identified the consensus Rv3676 binding site to be GTGNNANNNNNCACA and the proportion of sites identified with these nucleotides is shown in Figure 4.3 (Bai et al., 2005; Rickman et al., 2005). Substitution of the G in position 4 or the C in position 17 (as labelled in Figure 4.3) leads to loss of Rv3676 binding. The nucleotides at these positions are highly conserved, reflecting their importance (Bai et al., 2005). The consensus sequence for Rv3676 is close to the *E. coli* CRP consensus, and expression of Rv3676 in *E. coli* inhibits the activation of transcription from a CRP-regulated promoter, suggesting that the two proteins are able to compete to bind to the same site (Rickman et al., 2005). Rv3676 can bind to the *E. coli lac* promoter (TTGATN$_6$TCACT), but is unable to bind the FNR-binding site present in the *E. coli ndh* promoter (TTGATN$_4$ACTAA) (Bai et al., 2005).

What target promoters have been identified?

Bai et al. (2005) searched for putative Rv3676 binding sites in the intergenic regions of the *M. tuberculosis* genome. The model identified that 73 promoter regions containing a CRP site, which were associated with 114 genes (Bai et al., 2005), some of which are listed in Table 4.1.
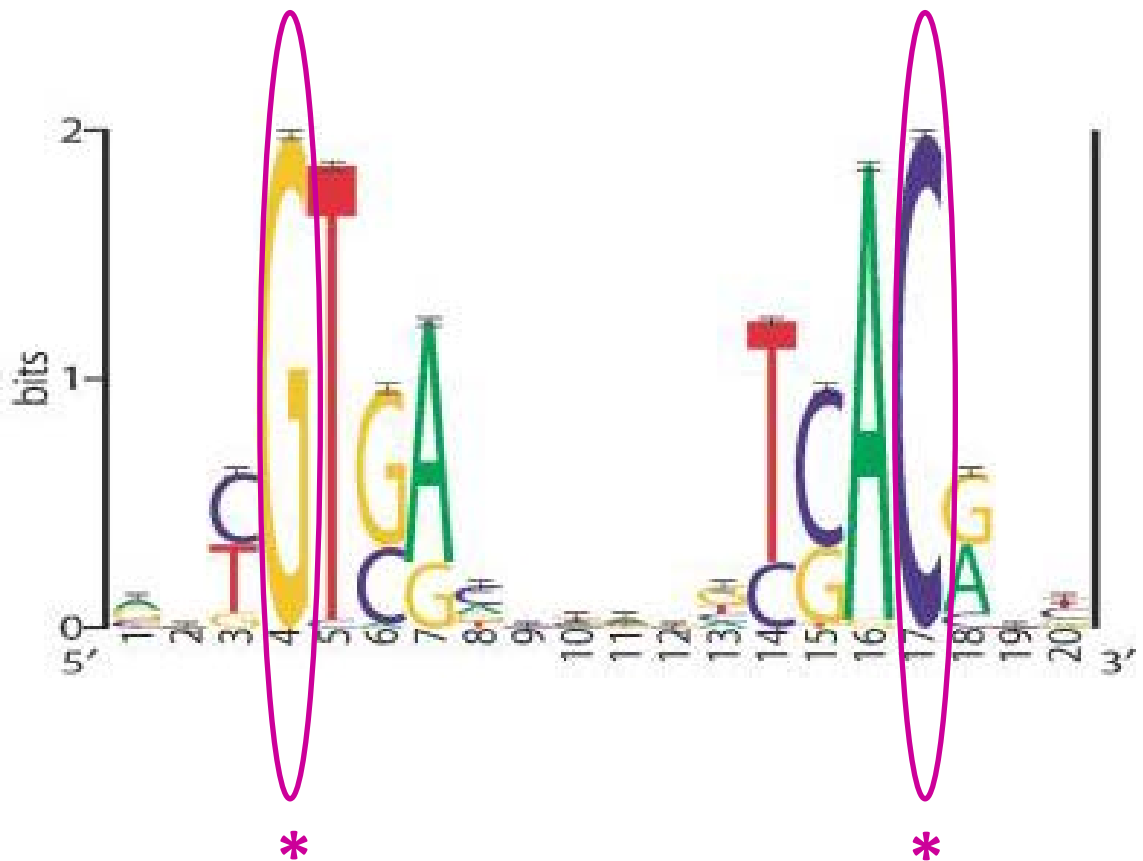
Figure 4.3: The Rv3676 binding site (based on Bai et al., 2005). The putative Rv3676 binding motif based upon 58 predicted sites in the *M. tuberculosis* genome. The size of the letters is proportional to the relative frequency of each base at each position of the motif. Strongly conserved nucleotides at positions 4 and 17 are essential for interactions between Rv3676 and DNA. Single substitutions of G to C at position 4 and C to G at position 17 were sufficient to abolish Rv3676 binding (Bai et al., 2005).

		Table 4.1: Genes regulated by Rv3676		
Gene	**Protein**	**Function**	**Evidence**	**Reference**
Rv0884c	SerC	Phosphoserine aminotransferase	EMSA	Bai et al., 2005
Rv1230c	Rv1230c	membrane protein	EMSA	Bai et al., 2005
rpfA	RpfA	Resuscitation promoting factor A	Identified in	Spreadbury et al., 2005
			Microarray	Rickman et al., 2005
			EMSA	Rickman et al., 2005
frdA			Identified by	Spreadbury et al., 2005
fadD10		Fatty-acid-CoA ligase	Microarray	Rickman et al., 2005
Rv0188	Rv0188	Possible transmembrane protein	Microarray	Rickman et al., 2005
lpqR	LPQR	Conserved lipoprotein	Microarray	Rickman et al., 2005
Rv1387	PPE20	Proline-proline-glutamic acid protein	Microarray	Rickman et al., 2005
Rv1592c	Rv1592c	Unknown function	Microarray	Rickman et al., 2005
cydA	CydA	Cytochrome d ubiquinol oxidase subunit	Microarray	Rickman et al., 2005
ahpC	AhpC	Alkyl hydroperoxide reductase	Microarray	Rickman et al., 2005
lipQ	LipQ	Probably carboxylesterase	Microarray	Rickman et al., 2005
whiB1	WhiB1	Regulatory protein	Microarray	Rickman et al., 2005
			EMSA	Stapleton et al., 2010
Rv3136	Rv3136	Proline-proline-glutamic acid protein	Microarray	Rickman et al., 2005
Rv3613c	Rv3613c	Hypothetical protein	Microarray	Rickman et al., 2005
Rv3614c	Rv3614c	Hypothetical protein	Microarray	Rickman et al., 2005
Rv3615c	Rv3615c	Hypothetical protein	Microarray	Rickman et al., 2005
Rv3616c	Rv3616c	Hypothetical protein	Microarray	Rickman et al., 2005

The candidate genes identified were involved mainly in starvation, hypoxia, dormancy and latency, PE and PPE proteins, cell wall remodelling, essential growth genes or uncharacterised genes (Bai et al., 2005). PE (proline-glutamic acid) and PPE (proline-proline-glutamic acid) family of proteins are present only in virulent mycobacteria and are upregulated during infection. Members of the family have been shown to be involved in cell wall remodelling (Singh et al., 2016).

Microarray analysis was performed which compared WT H37Rv with an *Rv3676* deletion mutant which revealed 16 genes with significantly altered regulation (>2.4-fold) during *in vitro* growth (Rickman et al., 2005). Several of the genes identified overlapped with the findings in Bai et al. (2005) and functions of the products of these genes varied from PPE proteins (*Rv3136* and *Rv1387*); regulatory proteins (*whiB1*); cell wall remodelling and resuscitation (*rpfA*) and many of unknown functions.

Genes with experimental evidence for direct regulation by Rv3676

<u>*rpfA*</u>

Experimentally, *rpfA* has been shown to be directly regulated by Rv3676 by *in vitro* and *in vivo* methods. Beta-galactosidase assays performed with an *rpfA*::*lacZ* promoter fusion construct showed 7.5-fold lower expression in the Rv3676 deletion mutant than in WT H37Rv. Mutating the predicted Rv3676 binding site at positions 4 and 17 lowered expression by 9-fold, suggesting that Rv3676 regulates *rpfA* expression from this site. EMSAs were performed with Rv3676 and *rpfA* promoter DNA containing the putative Rv3676 binding site, which is located ~440 bp upstream of the translation start site (CRP site is shown in Figure 3.4). EMSAs demonstrated binding of Rv3676 to the *rpfA* promoter DNA, and, unlike promoter binding by *E. coli* CRP, this binding was not enhanced in the presence of cyclic AMP or cyclic GMP (Rickman et al., 2005). The product of *rpfA* is RpfA, a resuscitation promoting factor which is linked to reactivation to actively growing cells during latency (discussed further in Introduction Section 1.7). Rv3676 is a regulator of *rpfA*, thus is implicated in the control of *M. tuberculosis* infection and reactivation.

whiB1

The *whiB1* promoter contains two Rv3676 binding sites, that are located at -58.5 (site 1) and -37.5 (site 2) relative to the transcriptions start. Stapleton et al. (2010) provided evidence that site 1 activates transcription of *whiB1* and that occupation of site 2 causes repression of *whiB1* transcription. Site 2 appeared to only be bound by Rv3676 when cyclic AMP was present or when the concentration of Rv3676 was high in the absence of ligand (Stapleton et al., 2010).

Rv3676 is not required for RNAP binding to the DNA and switching from a closed to an open complex, but is necessary for activation of transcription. Transcriptional activation caused by the Rv3676 interacting with site 1 occurs at a point after the formation of open complex of RNAP. At high cyclic AMP concentrations both sites may be bound by Rv3676, which results in repression by preventing the formation of an open complex (Stapleton et al., 2010).

4.1.2.3 Interaction of Rv3676 with promoter DNA causes DNA bending

When specific target DNA is bound by *E. coli* CRP sharp (~90 degrees) bending of the adjacent DNA is induced, which is important for the function of activation of transcription by CRP (Lin and Lee, 2003; Schultz et al., 1991; Wu and Crothers, 1984). EMSAs performed with *rv0884c* promoter DNA regions with the CRP site located in different relative locations on the 174 bp fragment showed lower electrophoretic mobility of the Rv3676:DNA complex when the CRP site was located centrally in the DNA fragment (Bai et al., 2005). Lower electrophoretic mobility is observed in bent DNA when the bend is centrally located in the DNA (Kolb et al., 1983), which supports the notion that DNA is bending upon Rv3676 binding (Bai et al., 2005).

4.1.3 Msmeg_6189 from *M. smegmatis*

Msmeg_6189, encoded by the *Msmeg_6189*, is one of two CRP/FNR family proteins encoded in the *M. smegmatis* genome.

4.1.3.1 Structural information

The four cyclic AMP binding residues (E80, S91, R129 and T134, shown in Figure 4.4) in Rv3676 are conserved in Msmeg_6189. Similarly to Rv3676, the two other residues involved in cyclic nucleotide binding by *E. coli* CRP are not present in Msmeg_6189. This would suggest that the mechanism by which Msmeg_6189 binds cognate ligands may be more similar to the Rv3676 mechanism than the *E. coli* CRP one. *Escherichia coli* CRP is capable of binding other ligands, such as cyclic GMP (Ren et al., 1990). Whereas Rv3676 is believed to bind specifically cyclic AMP (Stapleton et al., 2010 and Results Section 4.4.3.1 of this study). The residues in the putative cyclic AMP-binding pocket of Msmeg_6189 are identical to Rv3676, which implies that Msmeg_6189 may also show specificity for cyclic AMP and not other cyclic nucleotides. One aim of this study was to test this hypothesis.

The two residues (R188 and E189, shown in Figure 4.4) in Rv3676 which directly interact with the CRP site in DNA targets are both conserved in Msmeg_6189. The residue R186 that is involved in direct interaction with DNA in *E. coli* CRP is missing from both Rv3676 and Msmeg_6189. The consensus binding site for Msmeg_6189 has been identified as

(a)

```
MSMEG_6189   MDEILARAGIFQGVEPTAVAALTKQLQPVDFPRGHTVFAEGEPGDRLYIIISGKVKIGRR 60
Rv3676       VDEILARAGIFQGVEPSAIAALTKQLQPVDFPRGHTVFAEGEPGDRLYIIISGKVKIGRR 60
             :***************:*:*****************************************

                                                     R
MSMEG_6189   SPDGRENLLTIMGPSDMFGELSIFDPGPRTSSATTITEVRAVSMDRDALRAWIADRPEIA 120
Rv3676       APDGRENLLTIMGPSDMFGELSIFDPGPRTSSATTITEVRAVSMDRDALRSWIADRPEIS 120
             :*****************.*************.*********************.*******:

                        S
MSMEG_6189   EQLLRVLARRLRRTNNNLADLIFTDVPGRVAKQLLQLAQRFGTQEGGALRVTHDLTQEEI 180
Rv3676       EQLLRVLARRLRRTNNNLADLIFTDVPGRVAKQLLQLAQRFGTQEGGALRVTHDLTQEEI 180
             ********.****.***********************************************

                    *
MSMEG_6189   AQLVGASRETVNKALADFAHRGWIRLEGKSVLISDSERLARRAR 224
Rv3676       AQLVGASRETVNKALADFAHRGWIRLEGKSVLISDSERLARRAR 224
             *******.************************************
```

(b)

Figure 4.4: Alignment of *M. tuberculosis* Rv3676 and *M. smegmatis* Msmeg_6189. (a) sequence alignment of Rv3676 and Msmeg_6189. There is a large degree of homology between the two CRP proteins, and only 6 out of 224 residues differ (shown in purple). Alignment was produced using CLUSTAL 2.1. Residues that form hydrogen bonds with DNA bases in *E. coli* CRP and FNR and are conserved in Msmeg_6189 and Rv3676 are shown in green. The residue R185 in *E. coli* CRP (equivalent to K193 in both mycobacterial CRPs) is involved in direct DNA binding. This is not directly conserved in the mycobacterial CRP proteins, but the residue in this position is similarly charged. Residues equivalent to those involved in the interaction of *E. coli* CRP with cyclic AMP are shown in orange. (b) Crystal structure of Rv3676 (PDB: 3I54) over which the locations of residues which differ in Msmeg_6189 have been highlighted in yellow. The residues that differ between Rv3676 and Msmeg_6189 are not located in the DNA recognition helices or the cyclic AMP binding pocket. Each green sphere represents a molecule of cyclic AMP.

T**G**TGNNNNNNNNCA**C**A (Lee et al., 2014), which is highly similar to the sites identified for *E. coli* CRP and Rv3676. It also contains the G and C residues (shown in bold), which when substituted to other nucleotides prevented interaction with Rv3676 or *E. coli* CRP (Bai et al., 2005).

Although there is no published crystal structure of Msmeg_6189 available at present, but structure prediction models of Msmeg_6189 (A0R5H1) on the SWISS-MODEL repository suggest the structure would theoretically be highly-similar to Rv3676. Thus in terms of the DNA-binding in the absence of cyclic AMP, it is logical that the protein may behave similarly to Rv3676 and only show slight enhancement of DNA-binding in the presence of cyclic AMP.

4.1.3.2   Target promoters

Collaborators in the research group under the supervision of Dr Gregory Cook, University of Otago were interested in the regulatory roles of the CRP proteins present in *M. smegmatis* and the work was published in Aung et al., (2015).

The *M. smegmatis* genome, unlike *M. tuberculosis* which encodes just Rv3676, encodes two CRP proteins - Msmeg_0539 and Msmeg_6189. Msmeg_6189 shows a high degree of homology (~98%) (Figure 4.4) with Rv3676 and for this reason is of interest to us.

Our collaborators made extensive attempts to delete the *msmeg_6189* gene via allelic exchange in order to perform microarray analysis to allow determination of regulatory targets. However, an *msmeg_6189* deletion mutant, unlike *msmeg_0539*, was not readily obtained, suggesting that the gene may be essential for growth under the conditions tested. Instead, Msmeg_6189 was overexpressed on a tetracycline inducible plasmid (pMind) (Blokpoel et al., 2005) and transcriptional profiling analysis was carried out to determine the effects on gene expression (Aung et al., 2015). Tetracycline induction upregulated *msmeg_6189* expression by 6-fold and caused upregulation of 49 (≥1.4-fold) and the downregulation of 9 genes (≤1.4-fold). The consensus sequence for Msmeg_6189 was shown to be TGTGNNNNNNNCACA. The proposed Msmeg_6189 regulated genes spanned a wide-range of biological functions and included *sdh1* (encoding a succinate dehydrogenase), members of the *whiB* family and *rpf* family, which are linked to the resuscitation of dormant cells (Aung et al., 2015; Kana et al., 2008).

The *whiB1* gene is regulated by Rv3676 in *M. tuberculosis* (Stapleton et al., 2010). The *whiB4* and *whiB6* genes of *M. smegmatis* contain the CRP-binding site motif suggesting that they may be directly regulated by Msmeg_6189. The promoter of the *rpfA* gene contains a CRP site is directly regulated by Rv3676 in *M. tuberculosis* (Rickman et al., 2005 and section 4.3.1 of this

study). As well as *rpfA*, a potential CRP-site is present in the *rpfE* promoter in *M. smegmatis*, which suggests that Msmeg_6189 may regulate *rpfE* (Aung et al., 2015).

The study (Aung et al., 2015) also performed microarray analysis on an *Msmeg_0539* deletion mutant. There were eleven genes predicted to be members of both the Msmeg_6189 and Msmeg_0539 regulons. This included *sdh1*, but not *sdh2* which was only present in the Msmeg_0539 regulon. In order to validate the array data, qRT-PCR was performed for *sdh1* and *sdh2* (Figure 4.5), which further confirmed that *sdh1* appeared to be regulated by both *M. smegmatis* CRPs and *sdh2* was only regulated by Msmeg_0539. Our role in the study of these mycobacterial CRPs was to provide *in vitro* evidence of the regulation of *sdh1* by Msmeg_6189 and test whether it was able to bind to *sdh2*.

ahpC

The *ahpC* gene, which encodes alkyl hydroperoxide reductase protein C (AhpC), contains a CRP site in the upstream promoter. AhpC has a catalytic function in the detoxification of organic peroxides, reactive nitrogen intermediates and anti-oxidants (Chen et al., 1998; Hillas et al., 2000; Master et al., 2002), and is upregulated in cells that are exposed to various oxidative stresses (Lee et al., 2014).

Lee et al. (2014) demonstrated that Msmeg_6189 interacts with this site to activate transcription. qRT-PCR showed that expression of *ahpC* in *M. smegmatis* was significantly upregulated (6 to 50-fold depending on compound) after treatment with compounds (e.g. diamide) that are known to cause oxidative stress than in control cells with no stressors. Levels of induction of *ahpC* in the *Msmeg_6189* deletion strain after treatment with oxidising agents were significantly lower than for wild-type *M. smegmatis*, which supports the notion that Msmeg_6189 is having an effect on the expression of *ahpC*. Western blot analysis was performed, which showed that in *M. smegmatis* AhpC was more abundant after treatment with oxidative stress reagents than without and no such difference in abundance was observed in the *Msmeg_6189* mutant when treated (Lee et al., 2014).

EMSAs were performed, which showed that Msmeg_6189 is able to directly interact with the putative CRP site located in the *ahpC* promoter. Interestingly, Lee et al. (2014) only performed these binding assays in the presence of cyclic AMP, so the question of the Msmeg_6189:DNA interaction being dependent on cyclic AMP was not discussed in the study. Substitutions in the C and the G nucleotides (shown in bold in the consensus sequence T**G**TGNNNNNNNNCA**C**A) resulted in loss of binding. Levels of expression of *ahpC* were observed to be proportional to
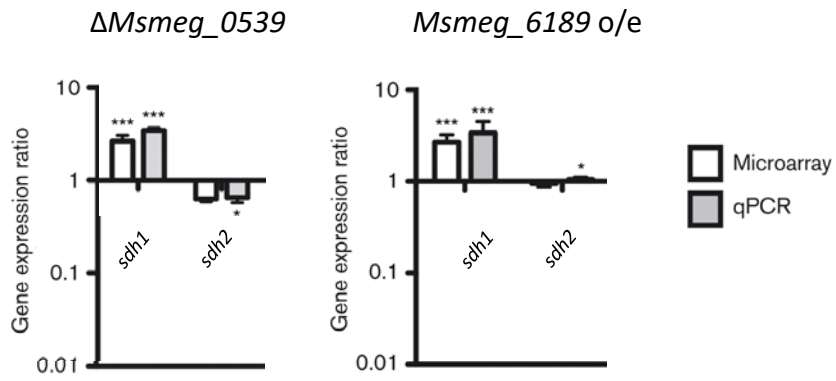
Figure 4.5: Gene expression ratio in response to deletion of *Msmeg_0539* or overexpression of *Msmeg_6189*. This figure is adapted from Aung et al. (2015). Value is the mean±SD for four (microarray) or three (qRT-PCR) independent biological repeats. *P<0.05, **P<0.01 and ***P<0.001.

intracellular cyclic AMP levels and at high levels of cyclic AMP, such as after treatment with oxidative stress inducers, *ahpC* expression levels increased (Lee et al., 2014).

4.2   Aims

- Purify His-Rv3676 and His-Msmeg_6189.
- Gain greater insight into DNA-binding by mycobacterial CRPs. This includes experimentally assaying binding to putative target promoter DNA for both proteins, with particular interest in:
    - The interaction between *rpfA* and Rv3676, and the effect of cyclic-AMP on the interaction.
    - Determining whether Msmeg_6189 can bind to the putative CRP sites identified in *sdh1* by our collaborators and determining the importance of cyclic AMP for the DNA-binding function of Msmeg_6189.
    - Determine whether Msmeg_6189 can bind to the promoter of *sdh2*, which is proposed to be regulated by Msmeg_0539, but not Msmeg_6189.
- Gain greater insight into cyclic AMP binding by mycobacterial CRPs.
    - Determine whether Msmeg_6189 binds cyclic AMP and if the binding pocket is specific for this cyclic nucleotide.
    - Determine whether Rv3676 is able to interact with cyclic GMP or 2'3'-cylic AMP as well as 3'5'-cyclic AMP.
    - Gain thermodynamic parameters for the interactions of the proteins with the ligand(s). Compare the binding affinity for cyclic-AMP for both CRP proteins.
    - Mutate the residues in Rv3676 that are equivalent to the two non-conserved *E. coli* CRP residues that involved in cyclic AMP binding and determine the effects of these mutations on the ability to bind cyclic AMP and the affinity of the interaction.

4.3   Purification of mycobacterial CRPs

The initial aim of the project was to perform experiments to characterise the interaction of wild-type Rv3676 and Msmeg_6189 proteins with target promoters. This included determining how the presence or absence of cyclic-AMP bound to the protein affected the interaction of the CRP proteins with DNA. To carry out experiments to investigate this, it was necessary to purify a sufficient quantity of each protein.

The approach taken was to overproduce a His-tagged version of each CRP in an *Escherichia coli* host and to use nickel affinity chromatography to purify the His-tagged CRP from the other cellular proteins.

4.3.1   Purification of Rv3676

Electrically competent *E. coli* BL21 λDE3 cells with a deletion of the *cyaA* gene (strain JRG5876), were produced and successfully transformed with pGS2132 by the methods in Sections 2.2.6 and 2.2.7. The pGS2132 plasmid was generated in a previous study. This plasmid consists of a pET28a (by Novagen) vector with the Rv3676 protein coding region (675 bp) cloned between the EcoRI and XhoI sites in the multiple cloning site. The vector also contains the *bla* gene for ampicillin resistance incorporated into the ClaI site, which disrupted the kanamycin resistance gene. The vector uses a pET expression system, under control of the *lac* promoter and operator. The expression of the cloned gene of interest is driven by a strong bacteriophage T7 promoter, which is recognised by T7 polymerase. The host *E. coli* strain (BL21 λDE3) possesses an IPTG inducible T7 RNAP encoded on the chromosome. In the absence of IPTG, the Lac repressor inhibits the expression of T7 polymerase and the gene of interest. IPTG is able to release the Lac repressor and allow T7 polymerase to be expressed, which allows for the expression and production of Rv3676.

The *cyaA* gene encodes for a class-1 adenylate cyclase enzyme, whose biological role in the cell is the production of the secondary messenger cyclic adenosine monophosphate (cyclic-AMP). The gene is non-essential for the growth in *E. coli*. In order to be able to purify Rv3676 without any cAMP already bound, a strain with *cyaA* deletion was selected for overproduction, in order to eliminate endogenously produced cAMP, hence lowering the likelihood of the protein being produced in a cAMP-bound state.

4.3.1.1   Overproduction of Rv3676

The presence of pGS2132 in *E. coli* BL21 λDE3 strain was verified by purifying plasmid DNA from an overnight culture and performing a restriction digest with EcoRI and XhoI. The cut plasmid was analysed by agarose gel electrophoresis (Figure 4.6) and species of size ~700 bp and ~6.7 kb were present as expected. The plasmid was sent for sequencing to further verify there were no mutations in the *Rv3676* coding region.

 Overnight cultures of *E. coli* BL21 λDE3 containing pGS2132 were set up and used to inoculate (1% inoculum) conical flasks containing LB medium supplemented containing ampicillin (200 µg/ml). Flasks were grown aerobically with shaking (250 rpm) at 37°C until an OD_{600} between 0.4 and 0.6 was reached. At this point the culture was split in four and protein overproduction was induced with IPTG at either 60 µg/ml or 120 µg/ml. One culture of each IPTG concentration was placed at 25°C and 37°C where they were grown aerobically with shaking (250 rpm) for up to 3 h. The OD_{600} of the final cultures were recorded and samples containing equivalent numbers of cells from each condition and time point were boiled for 10 min at 95°C
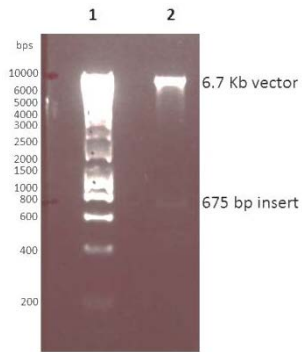
Figure 4.6: Agarose gel electrophoresis of pGS2132 treated with EcoRI and XhoI. Lane 1 contained 5 µl of DNA Hyperladder I (Bioline) and lane 2 contained 5 µl of pGS2132 DNA after digestion with EcoRI and XhoI for 1 h at 37°C.

in the presence of 1X SDS-PAGE loading dye to lyse the cells and denature the proteins. Samples were analysed by SDS-PAGE (Methods Section 2.4.1) (Figure 4.7).

An abundant polypeptide of ~26 kDa was present in all the samples, corresponding to His-Rv3676. There was no significant difference in the relative abundance of His-Rv3676 between cells induced at 37°C or 25°C. After induction with IPTG (120 µg/ml), the relative abundance of His-Rv3676 increased over time (30 min to 180 min) under both temperature conditions. There appeared to be no significant difference between the relative abundance of His-Rv3676 in 180 min post-induction samples induced with the two different IPTG concentrations. The conditions selected for all subsequent large-scale overproduction were 120 µg/ml IPTG for 3 h at 37°C with shaking (250 rpm).

After cells were grown and induced to overproduce His-Rv3676, the cells were collected by centrifugation (8,000 rpm for 20 min at 4°C). Cells were resuspended in (20 mM sodium phosphate (pH 7.4) with 0.5 M NaCl, containing EDTA-free protease inhibitor cocktail mix (Roche)) and lysed using a French pressure cell and cell free extract was produced (Methods Section 2.4.6).

4.3.1.2 Nickel affinity chromatography purification of His-Rv3676

The cell-free extract was filtered through a 0.2 µm filter before being applied to a nickel-charged 1 ml Hi-Trap chelating column (GE Healthcare). The His-tag purification program on the AKTA Prime™ Machine was run and the protein was eluted over an imidazole gradient (0 to 0.5 M) as detailed in Section 2.4.7. Fractions (1 ml) were collected and protein samples (10 µl) were analysed by SDS-PAGE (Figure 4.8). Lanes 7 to 9, corresponding to fractions 12 to 14 contained purified His-Rv3676. Fractions containing purified His-Rv3676 were combined and imidazole was removed typically by dialysis against 20 mM sodium phosphate (pH 7.4) with 0.2 M NaCl. Protein samples were concentrated using VivaSpin columns with a MWCO of 10,000 Da and proteins were stored at room temperature.

4.3.2 Purification of Msmeg_6189

Msmeg_6189, a CRP protein from *M. smegmatis*, was overproduced and purified by the same methods used for Rv3676, which are discussed in Results Section 4.3.1. The only difference being that pGS2229 was the overproduction vector used instead of pGS2132. Plasmid pGS2229 was generated in a previous study and is the identical to pGS2132, except the Msmeg_6189 protein coding region has been cloned between EcoRI and XhoI instead. The Msmeg_6189 coding region present in pGS2229 was verified by sequencing prior to commencing protein purification.
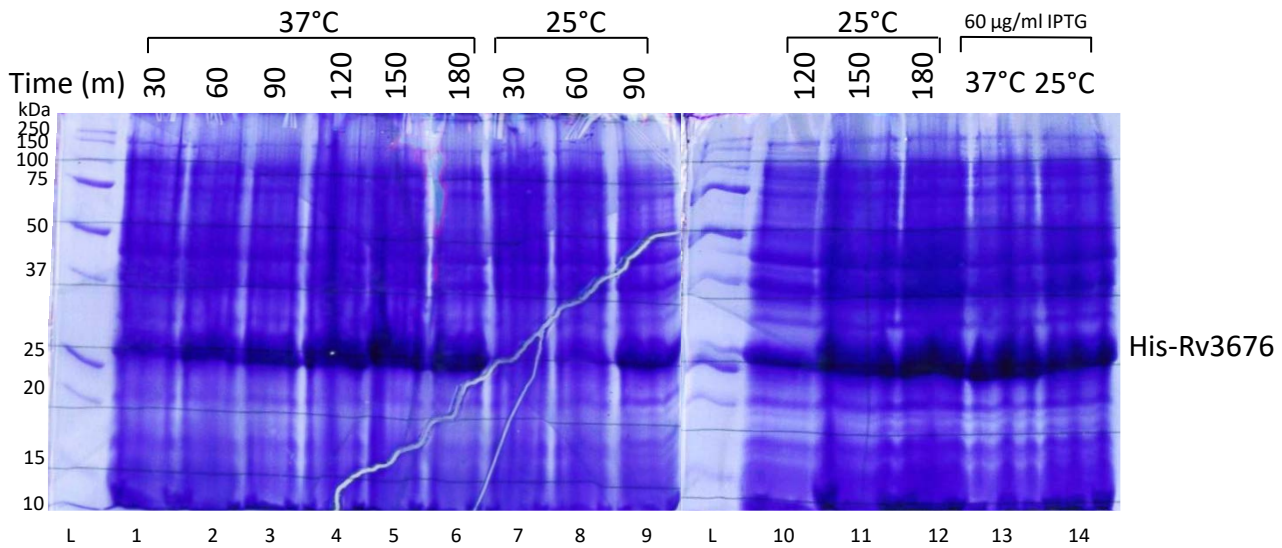
Figure 4.7: SDS-PAGE analysis of trial overproduction of His-Rv3676 under different conditions. All samples were induced with 120 μg/ml IPTG, except for those in lanes 13 and 14, which were induced with 60 μg/ml IPTG and were incubated for 180 min at the relevant temperature.
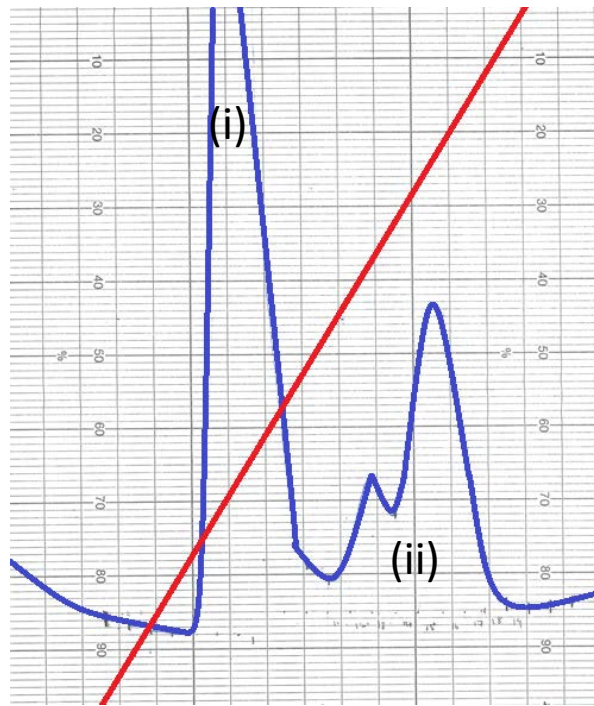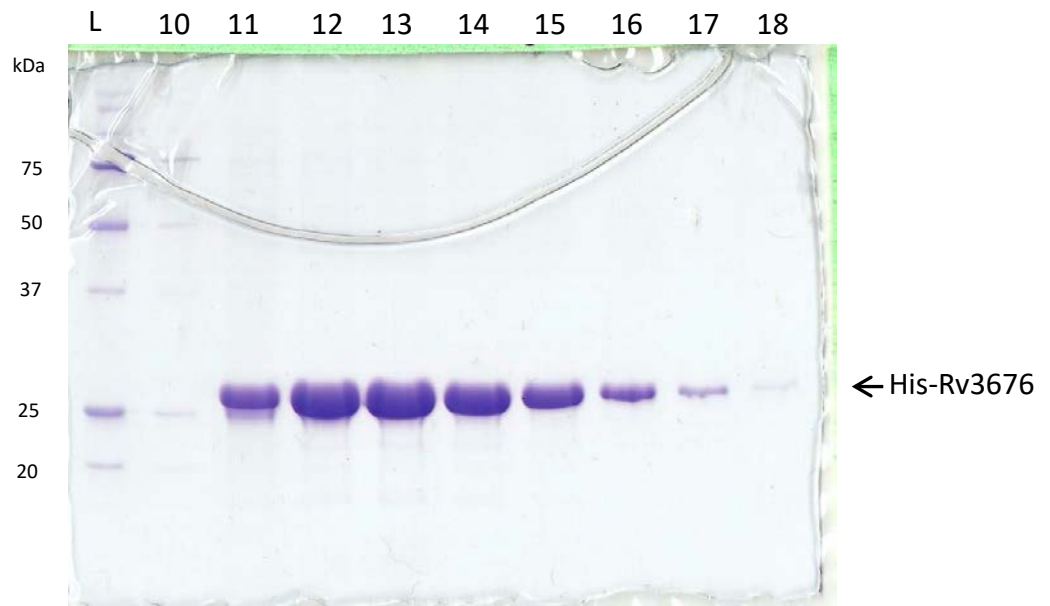
Figure 4.8: Nickel affinity chromatography purification of His-Rv3676. (a) Trace of A_{280} of fractions collected from AKTA prime™ His-tag purification program after the application of an imidazole gradient (0 to 0.5 M). The A_{280} of the eluent is shown by the blue line. The imidazole gradient (0 to 0.5 M) being applied to the column is shown by the red line. The first large peak (i) corresponds to non-specific proteins that contain a level of histidine residues high enough to bind the immobilised nickel in the HiTrap column. The second set of peaks (ii) contain the protein of interest. (b) SDS-PAGE analysis of fractions 10 to 18 from affinity purification of His-Rv3676. Lane L contains Precision Plus Protein™ All Blue Standards (8 µl) from BioRad. Lanes 1 to 9 contain samples (10 µl) of fractions 10 to 18 collected during affinity chromatography combined with SDS-PAGE loading dye (2x). The location of His-Rv3676 is indicated.

Cells were grown up, induced, collected and lysed as detailed in Section 4.3.1.1. Nickel affinity chromatography was performed as detailed in Section 4.3.1.2. SDS-PAGE analysis was performed on fractions from the eluent (Figure 4.9). Lanes 7 to 10 contained purified His-Rv3676. Fractions containing purified His-Rv3676 were combined and imidazole was removed typically by dialysis against 20 mM sodium phosphate (pH 7.4) with 0.2 M NaCl. Protein samples were concentrated using VivaSpin columns with a MWCO of 10,000 Da and proteins were stored at room temperature.

4.4 Binding of CRPs with promoter DNA

4.4.1 Rv3676 interacts with P*rpfA* in the absence of cyclic AMP, but cyclic AMP enhances binding

As covered in the introduction to this chapter (Section 4.1.2.2), it has been demonstrated in numerous studies that cyclic AMP is not required for site-specific DNA-binding (Bai et al., 2005; Rickman et al., 2005; Stapleton et al., 2010). Rickman et al. (2005) discussed how cyclic AMP affects the interaction between Rv3676 and the CRP site located in the promoter of the *rpfA* gene. They proposed that cyclic AMP did not enhance binding to this site. Other studies on other genes, such as *whiB1*, have suggested that cyclic AMP enhances binding to DNA targets (Bai et al., 2005; Stapleton et al., 2010). Due to our interest in the regulation of the *rpfA* gene in *M. tuberculosis*, it was decided to further explore the role of cyclic AMP in the interaction between Rv3676 and the CRP binding site located in *rpfA* promoter.

4.4.1.1 Generation of promoter DNA

In order to perform binding assays with purified Rv3676 it was necessary to produce radiolabelled *rpfA* promoter (P*rpfA*) DNA containing the CRP site (Figure 4.10a). PCRs were performed (Methods Section 2.3.5) with primers Myc657 and Myc658 (sequences in Table 2.1). PCR products were analysed by agarose gel electrophoresis (Figure 4.10b). A species of size ~200 base pairs was observed that corresponded to P*rpfA*. The DNA was purified using a PCR purification kit and subsequently underwent restriction digest with EcoRI (Methods Section 2.3.3). The 5' sticky end was radiolabelled by incubating with Klenow enzyme in the presence of $[\alpha^{32}P]$-dATP (Methods Section 2.5.1.1). End-labelled DNA was purified and quantified before use in binding studies.

4.4.1.2 EMSA analysis of the interaction between Rv3676 and P*rpfA* in the presence and absence of cyclic AMP

EMSAs were performed (Methods Section 2.5.1.1) to investigate the interaction of Rv3676
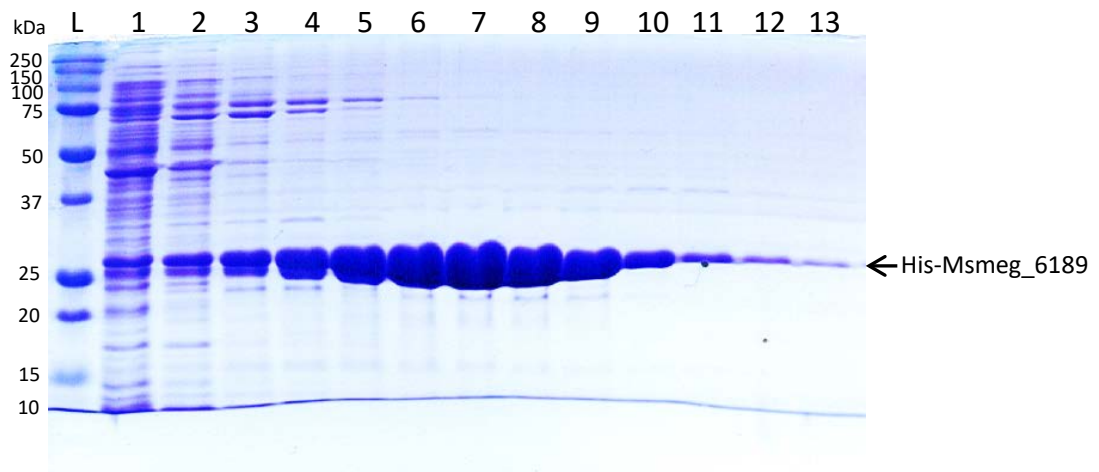
Figure 4.9: SDS-PAGE analysis of fractions after nickel affinity chromatography purification of His-Msmeg_6189. Lane L contains Precision Plus Protein™ All Blue Standards (8 µl) from BioRad. Lanes 1 to 13 contain samples (10 µl) of fractions 6 to 18 collected during affinity chromatography combined with SDS-PAGE loading dye (2x). The location of His-Msmeg_6189 is indicated.

(a)

P*rpfA*

5'-GAATTCGACGGGATTGTCGTCCGAGATGACGCCGTCGCGTTATCCG
CCGGCGACACGATTGACGTACTACCCCCTTTCGCCGGCGGCTAAGGCCA
TGTGACATTACCCACATAACGGAACGATAACGGGGCGGACACGCTCCGA
TTGCGAGGGCTTACTACCTGCGCAAATGCCTATATTTCTTGGCGTTCTT
GCCGGCGATTGTGGGATCC-3'

Myc657 5'-GAATTCGACGGGATTGTCGTCCGA-3'
Myc658 5'-GGATCCCACAATCGCCGGCAAGAA-3'

(b)
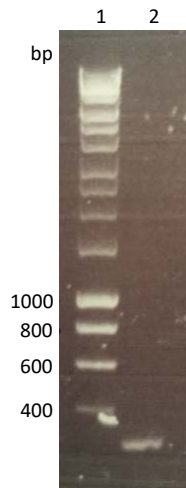
Figure 4.10: The structure of P*rpfA* region of the *rpfA* gene promoter that contains the CRP binding site. (a) Sequence of P*rpfA*. The 212 base pair region of the *rpfA* promoter containing the CRP binding site is shown. The location of CRP binding site is indicated in bold. The EcoRI site is indicated in pink. (b) Agarose gel electrophoresis analysis of P*rpfA* PCR products. Lane 1 contained 5 µl DNA Hyperladder I (Bioline) and lane 2 contained 5 µl PCR product.

with the CRP site located in P*rpfA*. Samples containing varying concentrations of Rv3676 (0 to 10 µM) with 0.5 ng/µl P*rpfA* were set up with or without cyclic AMP (Figure 4.11a). The migration of free, unbound P*rpfA* is shown in lanes 1 and 7, where no protein is present. When no cyclic AMP is present (lanes 1 to 6), a shift corresponding to retarded migration due to interaction between Rv3676 and P*rpfA* is observable at concentrations of Rv3676 from ~5 µM and above. When cyclic AMP is present in reactions (in lanes 7 to 12), the shift in migration is observable from concentrations of Rv3676 of 2.5 µM and above. This suggests that in contrast to the observations of Rickman et al. (2005), binding of Rv3676 at P*rpfA* is somewhat enhanced in the presence of cyclic AMP, in accord with Rv3676 binding at the whiB1 promoter (Stapleton et al., 2010). Thus, it was concluded that Rv3676 can bind P*rpfA* in the absence of cyclic AMP, but higher concentrations are required to produce a complete shift compared to when cyclic AMP is present.

4.4.1.3 Slot blotting analysis of the interaction between Rv3676 and P*rpfA* in the presence and absence of cyclic AMP

Slot blotting is a technique that can be used to gain information on DNA:protein binding. A diagram of the slot blot machine and brief description of the principles behind the technique is shown in Figure 4.12. Reactions were set up in triplicate containing radiolabelled P*rpfA* and different concentrations of Rv3676 in the presence and absence of cyclic AMP (Figure 4.11b). For reactions in both the presence and absence of cyclic AMP the proportion of DNA on the 'Protein-bound' membrane increases as the concentration of Rv3676 increases,. This further consolidates the EMSA evidence that Rv3676 is able to bind the CRP site located in the *rpfA* promoter. It was observed that cyclic AMP enhances binding. At 5 µM Rv3676 when no cyclic AMP is present only ~50% of PrpfA was bound to Rv3676. However, at Rv3676 concentrations of 5 µM in the presence of cyclic AMP (100 µM) at least ~95% of P*rpfA* in in complex with Rv3676. Slot blotting experiments have provided further evidence that cyclic AMP enhances the binding of Rv3676 to the CRP site in the *rpfA* promoter under the conditions tested.

4.4.1.4 BLItz analysis of the interaction between Rv3676 and P*rpfA*

As discussed previously in Section 3.4.3, BLItz allows for the real-time, label-free analysis of binding kinetics. BLItz experiments were carried out with a range of concentrations of purified Rv3676 from 50 µM to 50 nM and biotinylated P*rpfA* DNA (Figure 4.13) by the methods outlined in Section 2.5.2. At least three independent repeats were performed and a representative trace is shown. An increased thickness was observed when probes containing immobilised DNA were dipped into Rv3676 samples. The thickness increased was greater in samples with higher protein concentration, indicating that Rv3676 can bind to P*rpfA* (Figure
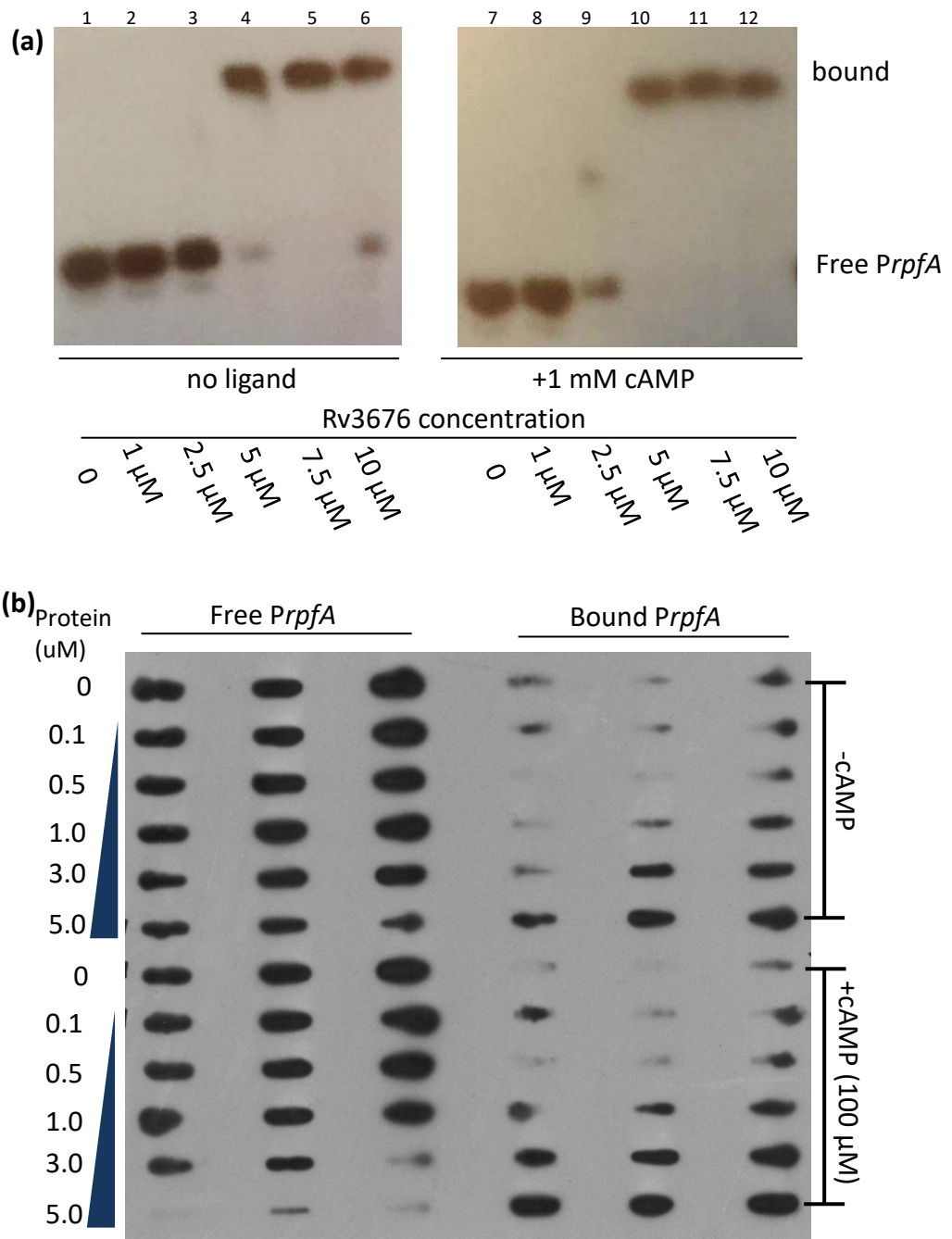
Figure 4.11: The effect of cyclic AMP on the interaction between Rv3676 and PrpfA. (a) EMSA analysis of the interaction between Rv3676 and PrpfA in the presence and absence of cyclic AMP. EMSA was carried out by the methods in section XX. The protein and ligand concentrations are shown. Radiolabelled PrpfA was present at a ~4 pM. (b) Slot blotting analysis of the interaction between Rv3676 and PrpfA in the presence and absence of cyclic AMP. The concentration of Rv3676 and cyclic AMP are indicated. DNA was Radiolabelled PrpfA was present at a ~4 pM. Concentrations of Rv3676 are shown as concentration of monomer. Three replicates were performed.
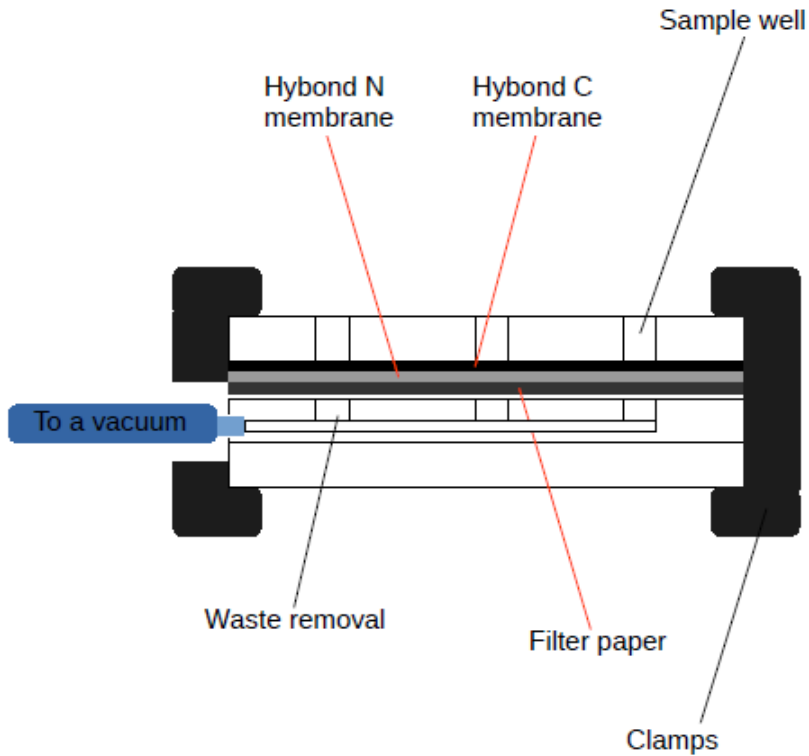
Figure 4.12: The principle behind the slot blot machine for assaying DNA binding. The principle behind the technique is to set up binding reactions containing the DNA and protein of interest in the presence and absence of the ligand of interest. Instead of analysing the reactions by electrophoresis and screen for alterations in migration like is done in an EMSA experiment, samples are loaded onto the slot blot and a vacuum is applied. This sucks the sample through two layers of membrane. The first membrane is able to bind protein and any DNA which is bound to the protein will also be immobilised on the membrane. The free DNA which is not in complex with the protein is able to pass through this first protein membrane and is immobilised on the second membrane which is positively charged.

(a)
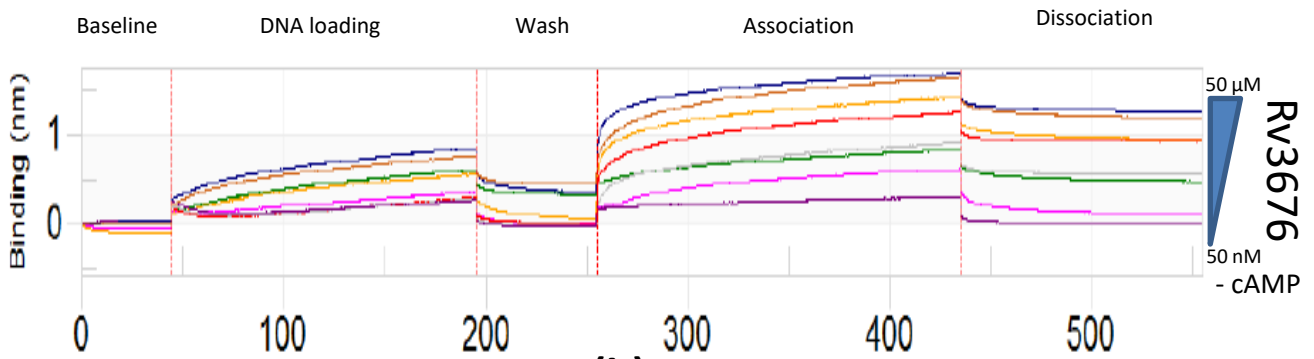
(b)
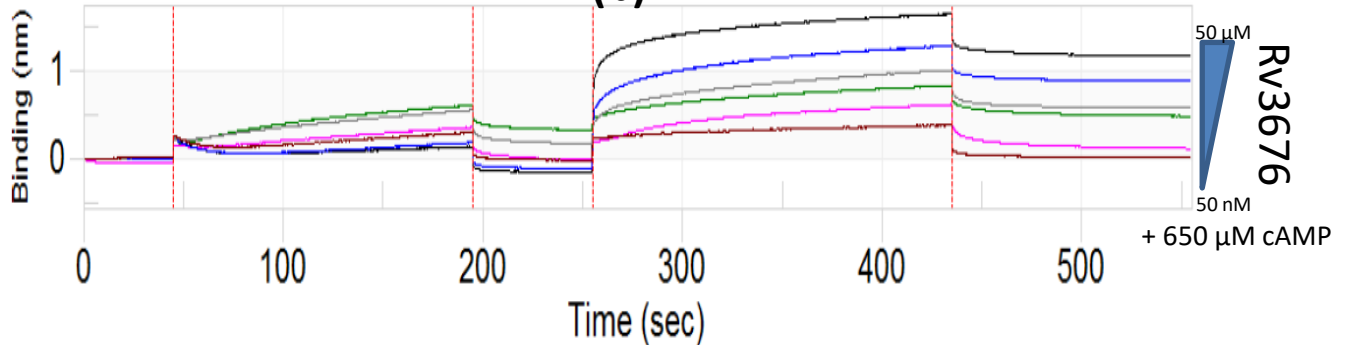
(c)

	- cAMP	+ cAMP (650 µM)
KD (M)	2.65×10^{-6}	2.71×10^{-7}
Ka (Ms^{-1})	$7.20 \times 10^{2} \pm 40.6$	$3.80 \times 10^{3} \pm 190$
Kd (s^{-1})	$1.91 \times 10^{-3} \pm 1.35 \times 10^{-4}$	$1.03 \times 10^{-3} \pm 5.89 \times 10^{-5}$

Figure 4.13: BLItz analysis of the interaction between Rv3676 and P*rpfA* in the presence and absence of cyclic AMP. (a) Overall BLItz trace with different concentrations of Rv3676 (50, 25, 10, 5, 2.5, 1, 0.5, 0.05 µM). (b) Overall BLItz trace with different concentrations of Rv3676 (50, 10, 5, 2.5, 1, 0.05 µM) in the presence of cyclic AMP (650 µM). Step 1: wash step. Step 2: association of DNA to probe. Step 3: wash step. Step 4: association of Rv3676 to DNA. Step 5: dissociation. (c) Rates of association, dissociation and binding affinity after analysis using BLItz software (by ForteBio).

4.13a). This supports the EMSA findings and slot blot experiments in Sections 4.4.1.2 and 4.4.1.3. Similar experiments were performed with Rv3676 samples that had been treated with cyclic AMP and all subsequent buffers contained 650 µM cyclic AMP (Figure 4.13b). Again, an increased thickness was observed when probes containing immobilised DNA were dipped into Rv3676 samples when cyclic AMP was present. The thickness increase was greater in samples with higher protein concentration, indicating that Rv3676 can bind to P*rpfA*.

BLItz software (by ForteBio) was used to analyse the data to provide binding kinetic constants for the interaction between Rv3676 and P*rpfA* in the presence and absence of cyclic AMP (Figure 4.13c). In the absence of cyclic AMP, the interaction had an association rate constant (k_a) of 7.2 x 10^2 1/Ms and a dissociation rate constant (k_d) of 1.9 x10^{-3} 1/s. The K_D was calculated to be ~2.7 µM. In the presence of cyclic AMP, the binding affinity (K_D) for the interaction was approximately 0.27 µM, which is ~10-fold tighter. The tighter binding is mainly due to an increased binding constant (k_a), which increased ~5-fold to 3.8 x 10^3 1/Ms. The dissociation rate constant was similar, ~0.5-fold lower at 1.0 x 10^{-3} 1/s. The binding affinities measured by BLItz are tighter than those suggested by EMSA experiments, particularly for the samples with cyclic AMP. However, they do support the notion that Rv3676 is able to bind to the CRP site located in the *rpfA* promoter and that there is an enhancement of binding (~10-fold) in the presence of cyclic AMP.

4.4.2   Binding of Msmeg_6189 to *sdh1* promoter is dependent on cyclic AMP

Due to the high degree of homology between Rv3676 and Msmeg_6189 and the conservation of the residues involved in cyclic AMP-binding and DNA-binding between the two proteins (as discussed in Section 4.1.3), it was reasonable to theorise that the interaction between Msmeg_6189 and target DNA would be independent of cyclic AMP, with some enhancement in the presence of cyclic AMP. The initial aim was to test this hypothesis and to determine if Msmeg_6189 is able to interact with the putative CRP sites located in the *sdh1* and *sdh2* genes of *M. smegmatis* as discussed in Section 4.1.3.

4.4.2.1   Generation of promoter DNA

In order to determine if purified Msmeg_6189 directly regulated either *sdh1* or *sdh2*, regions of the promoter containing the putative CRP-binding site and some flanking DNA were amplified by PCR (Methods Section 2.3.4). The promoter regions selected were termed P*sdh1* P*sdh2* for the *sdh1* and the *sdh2* gene respectively. The sequences are shown in Figure 4.14a. PCR reactions were set up (Methods section 2.3.4) to amplify the promoter regions (Figure 4.14b). PCR reactions contained dCTP, dGTP and dTTP, but [α^{32}P]-dATP, which allowed for

radiolabelling of the promoter regions during amplification. Radiolabelling was necessary for subsequent EMSA experiments. Primers LD3 and LD4 for P*sdh1* and LD5 and LD6 for P*sdh2* (sequences in Table 2.1). In lane 1, a DNA species of ~160 bp was observed which corresponds to the successful amplification of P*sdh1*. In lane 2 a DNA species of ~180 bp was observed which corresponds to the successful amplification of P*sdh2*. Promoter regions were purified from the components of the PCR mix (Methods Section 2.3.5) and the DNA concentration was determined (Section 2.3.12).

4.4.2.2   Msmeg_6189 binds P*sdh1* but not P*sdh2*

EMSA experiments were performed where radiolabelled promoter DNA was incubated with purified Msmeg_6189 in the presence or absence of cyclic AMP (Figure 4.15a) (Methods Section 2.5.1.1). When both cyclic AMP and Msmeg_6189 are present, the migration of P*sdh1* was retarded, compared to free P*sdh1* in the absence of any Msmeg_6189. This suggests that Msmeg_6189 is able to bind the CRP-binding site present in the *sdh1* promoter. No alteration in the migration of P*sdh1* was observed in the absence of cyclic AMP, which suggests that the interaction between Msmeg_6189 is dependent on cyclic AMP.

Under the conditions tests, no alteration in the migration of P*sdh2* was observed when incubated with Msmeg_6189 in either the presence or the absence of cyclic AMP (Figure 4.15a). This suggests that Msmeg_6189 does not directly regulate *sdh2*. The results of the EMSA experiments corroborate gene expression profiles from qRT-PCR and microarray analyses performed by out collaborators (Figure 4.5).

4.4.2.3   Binding of Msmeg_6189 at *sdh1* is dependent on cyclic AMP

Further experiments were performed to consolidate the role of cyclic AMP in the regulation of *sdh1* by Msmeg_6189. EMSA experiments were performed where reactions contained different concentrations of cyclic AMP (0 to 133 µM) incubated with Msmeg_6189 (3.6 µM) and P*sdh1* DNA (Figure 4.15b). P*sdh1* incubated with Msmeg_6189 in the absence of cyclic AMP migrated the same distance as free P*sdh1* which suggested that Msmeg_6189 is not able to interact with the CRP site located in P*sdh1* when cyclic AMP is not present. Msmeg_6189 caused a shift in mobility of P*sdh1* at concentrations of cyclic AMP above 6.7 µM and above (Lanes 4 to 8, Figure 4.15b), with a complete shift occurring at 133 µM (Lane 8) (Aung et al., 2015).

EMSAs were also performed with P*sdh1* and increasing concentrations of Msmeg_6189 in the presence of cyclic AMP to validate that the specific binding of Msmeg_6189 at P*sdh1* is cAMP-

(a)

P*sdh1*
5'ggcgcgacgcctcgacaccgccttgccggaagggcaccgggaatgcgcaccgcgc
accacgcccctcgctgacgtgatgcgcgtcacgttcggatcgtgatgtggctaccat
cggttgtcgtaa**tgtgattcacgacacc**cttcagcccgaggaggtcac 3'

P*sdh2*
5'agcgcactcggcacagagacctagaccata**tgagacattctccaca**ttgcatccg
gtcaccgtccggccgtcgtcgaccagcgcggccgcaccgaccggaaagcccgaatag
ggcgcatatgcgtgccgagagacctcgattgccttgctgcgcagggcattccaattc
acgtttgcgttc 3'

(b)
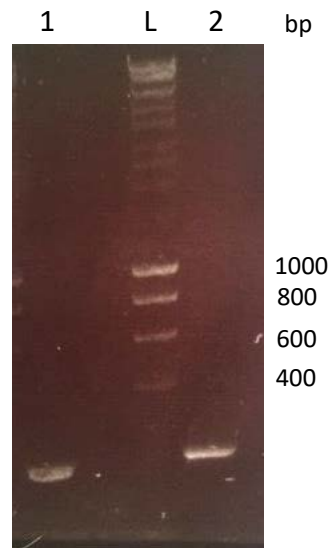
Figure 4.14: Sequences and generation of P*sdh1* and P*sdh2*. (a) Sequences of promoter regions P*sdh1* and P*sdh2*. P*sdh1* is a 160 bp region from the promoter of the *sdh1* gene. P*sdh2* is a 181 bp region from the promoter of the *sdh2* gene. The location of the putative CRP binding sites are shown in **purple**. (b) Agarose gel electrophoresis of P*sdh1* and P*sdh2* PCR products. PCR reactions were set up using primers LD3 and LD4 for P*sdh1* and LD5 and LD6 for P*sdh2*. Lane L contained 5 µl DNA Hyperladder I (Bioline); lane 1 contained 5 µl of P*sdh1* PCR product and lane 2 contained 5 µl of P*sdh2* PCR product.
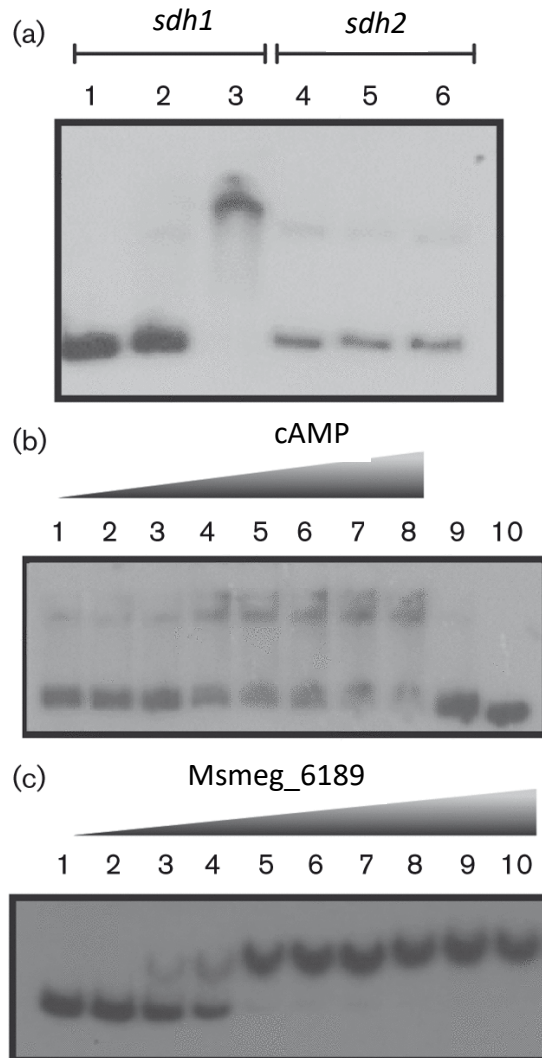
Figure 4.15: Assessing the interaction between Msmeg_6189 and *sdh1* or *sdh2* promoter DNA. Published in Aung et al. (2015). (a) EMSAs of Msmeg_6189 with radiolabelled *M. smegmatis sdh1* (lanes 1–3) and *sdh2* (lanes 4–6) promoter DNA. Promoter DNA was incubated without Msmeg_6189 (lanes 1 and 4), with Msmeg_6189 (5.5 µM; lanes 2 and 5) and with Msmeg_6189 plus cAMP (5.5 µM Msmeg_6189, 0.2 mM cAMP; lanes 3 and 6). (b) EMSAs with the *sdh1* promoter incubated with Msmeg_6189 (3.6 µM) in the presence of increasing concentrations of cAMP (lanes 1–8: 0.66, 1.33, 3.33, 6.7, 13.3, 33, 67, 133, 0 µM). The reaction in lane 9 contained no cAMP. The reaction in lane 10 lacks Msmeg_6189. (c) Binding of Msmeg_6189 at the wild-type *sdh1* promoter. Promoter DNA was incubated at 25 °C for 5 min with Crp2 in the presence of 2 mM cAMP. Lane 1, no protein; lanes 2–10, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5 and 1 µM Msmeg_6189, respectively. Three independent replicates were performed.

dependent (Figure 4.15c). Free migration of P*sdh1* was observed in lane 1. In the presence of 1 nM Msmeg_6189 and 2 mM cyclic AMP, no alterations in the migration of P*sdh1* were observed. A partial shift of migration of P*sdh1* was observed at Msmeg_6189 concentrations between 5 and 10 nM when cyclic AMP is present (lanes 3 and 4). A complete shift in migration of P*sdh1* DNA was observed at concentrations of Msmeg_6189 above 25 nM (lanes 5 to 10) confirming previous findings that Msmeg_6189 binding to the *sdh1* promoter is dependent on cyclic AMP under the conditions tested.

4.5 Characterisation of the interaction between mycobacterial CRPs with cyclic AMP

4.5.1 Msmeg_6189 and Rv3676 are able to bind 3'5' cyclic AMP

4.5.1.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) analysis was performed to provide insight into the thermodynamic parameters of the interaction between cyclic AMP and the mycobacterial CRPs Rv3676 and Msmeg_6189. Proteins were overproduced and purified as discussed in Section 4.3 and ITC experiments were performed where cyclic AMP was titrated into Rv3676 or Msmeg_6189 (Methods Section 2.4.11). Representative traces for the binding of cyclic AMP to Rv3676 and Msmeg_6189 are shown in Figure 4.16 alongside the thermodynamic parameters of the interaction. Both Rv3676 and Msmeg_6189 can interact with 2 molecules of cyclic AMP (1 per monomer) and the reaction is endothermic. The dissociation constants of the interaction with cyclic AMP are in the micromolar range for both proteins (~300 µM for Rv3676 and ~350 µM for Msmeg_6189). This is relatively weak binding compared to the affinities observed for other CRP proteins. For example, K_d values have been measured at ~13-16 µM for the interaction between *E. coli* CRP and cyclic AMP (Yu et al., 2012). It is also weaker than the 60 µM that has previously been suggested for the interaction of Rv3676 and cyclic AMP (Stapleton et al., 2010). These differences will be discussed further in Section 6.1.

4.5.2 Cyclic AMP binding affects the thermal stability of Rv3676 and Msmeg_6189

Differential scanning fluorimetry (DSF) experiments were performed with purified CRP proteins to determine if cyclic AMP-binding affected the thermal stability. The purpose of this was not only as another potential method to support previous evidence for cyclic AMP-binding, but also to compare any potential differences between Msmeg_6189 with Rv3676. The thermal stability of a protein is dependent on its current conformational state and three dimensional structure. Msmeg_6189 showed cyclic-AMP dependent binding to P*sdh1* (Aung et al., 2015), whereas the DNA-binding ability of Rv3676 is only slightly enhanced in the presence of cyclic

Figure 4.16: ITC analysis for mycobacterial CRP proteins titrated with 3'5'-cyclic AMP. The traces belonging to Rv3676 and Msmeg_6189 are indicated. Proteins samples (500 µM) and were subjected to 25 injections (1.97 µl) of 3'5'-cyclic AMP (7.5 mM) and data was fitted to an independent binding site model to provide thermodynamic parameters of the interaction.
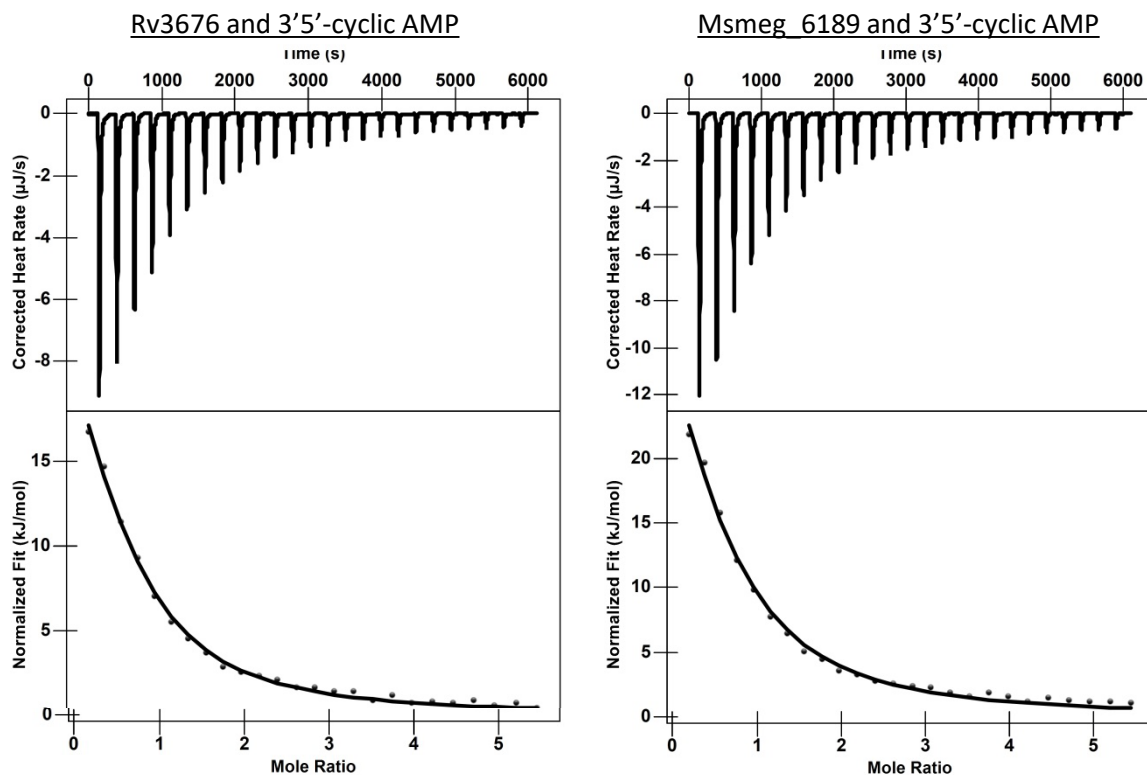
Parameter	Rv3676	Msmeg_6189
K_a (M^{-1})	$3.438 \times 10^3 \pm 475.1$	$2.902 \times 10^3 \pm 380.8$
K_d (M)	2.908×10^{-4}	3.446×10^{-4}
n (per monomer CRP)	0.962 ± 0.072	0.936 ± 0.099
ΔH (kJ/mol)	36.63 ± 5.21	54.34 ± 7.6

Figure 4.16: ITC analysis for mycobacterial CRP proteins titrated with 3'5'-cyclic AMP. The traces belonging to Rv3676 and Msmeg_6189 are indicated. Proteins samples (500 µM) and were subjected to 25 injections (1.97 µl) of 3'5'-cyclic AMP (7.5 mM) and data was fitted to an independent binding site model to provide thermodynamic parameters of the interaction.

AMP (Rickman et al., 2005; Stapleton et al., 2010). The proteins have a high degree of homology (~98%) and exhibit similar affinities for cyclic AMP (section 4.5.1.1) so it is plausible that the cyclic AMP-dependence could be due to a conformational difference in the two proteins in the holo or apo-form.

In the apo-form of *E. coli* CRP, the DNA-binding helices are buried below the surface of the protein (Figure 4.1, green helices). Cyclic AMP binding causes a conformational change, resulting in exposure of the DNA-binding helices to the surface of the protein and thus allowing them to interact with the target DNA. In apo-Rv3676, the DNA-binding helices are already presented on the surface, which explains why the protein is able to interact with target promoters in the absence of cyclic AMP. In holo-Rv3676, the orientation of the DNA-binding helices in comparisons with the main body of the protein is shifted slightly, which could explain why the DNA-binding affinity is mildly enhanced in the ligand-bound form.

Although the residues which differ between the two mycobacterial CRPs discussed in this study (Msmeg_6189 and Rv3676) are not those proposed to be directly interact with the cyclic AMP (E80, S91, R129 and T134), there was a difference in the dependence on cyclic AMP for DNA-binding. Thus it would be interesting to compare the relative thermostability of the proteins in the apo- and holo- forms.

DSF (Methods Section 2.4.12) is a technique which can be used to identify the thermostability or melting point (T_m) of a protein by the use of a fluorescence dye (SYPRO orange). It is based on the method discussed in Niesen et al. (2007). SYPRO orange has an affinity for hydrophobic regions of the protein. The technique involves gradually heating the protein of interest in the presence of SYPRO orange. As the protein unfolds, hydrophobic regions are exposed and Sypro Orange is able to bind, which can be detected by changes in fluorescence. A fitting process is performed, which allows the transition midpoint or 'melting temperature' to be calculated. Comparisons between the T_m of a protein in the presence and absence of a small ligand provide insights into if ligand-binding affects the thermal stability of the protein under the conditions tested.

DSF assays were performed on Rv3676 and Msmeg_6189 (100 µM) in the presence and absence of cyclic AMP (1 mM) (Figure 4.17a and b). The T_m for both Rv3676 and Msmeg_6189 is 64°C in the absence of cyclic AMP and the thermostability of the proteins increase by an identical amount to 68.2°C in the presence of cyclic AMP. This suggests that conformational changes resulting from the binding of cyclic AMP stabilise the proteins. Lysozyme was used as a control and was found to have a T_m of 70.2°C in both the presence and absence of cyclic AMP. This value is consistent with published findings (Kato et al., 1992).
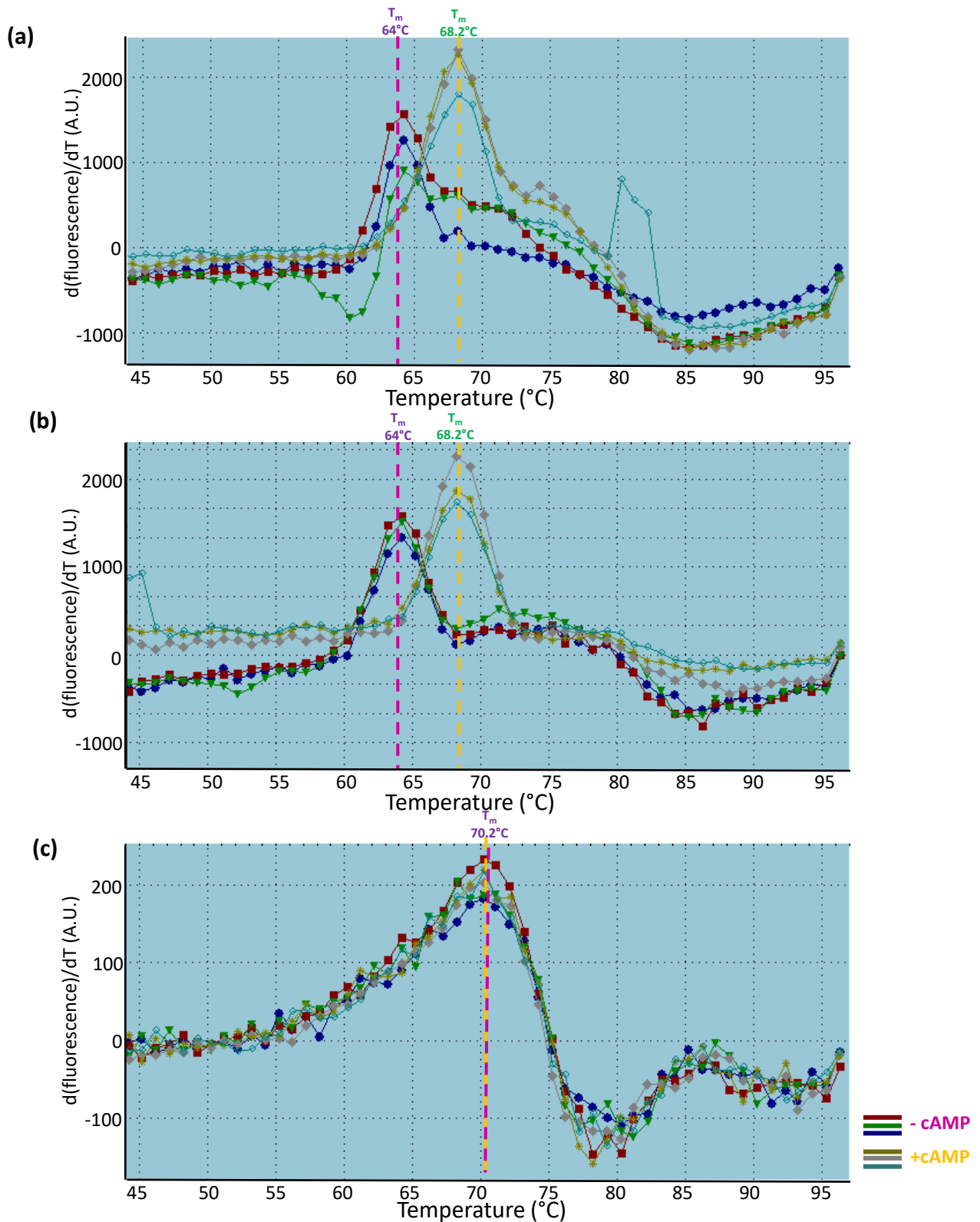
Figure 4.17: DSF analysis of the melting temperatures of mycobacterial CRP proteins in the presence and absence of cyclic AMP. DSF analysis was performed on proteins (100 µM) in the presence or absence of cyclic AMP (1 mM final). Changes in fluorescence were monitored and the first derivative curves were plotted, at which the minimum corresponds T_m (°C) of the proteins at that condition. (a) The first derivative curves for Rv3676 samples. (b) The first derivative curve for Msmeg_6189 samples. (c) The first derivative curves for Lysozyme control samples. The T_m for both Rv3676 and Msmeg_6189 is 64°C in the absence of cyclic AMP and in the presence of cyclic AMP is 68.2°C. Lysozyme has a T_m of 70.2°C in the presence and absence of cyclic AMP.

4.5.3   Ligand specificity studies

4.5.3.1   The cyclic nucleotide binding site in Rv3676 is specific for 3'5'-cyclic AMP

Escherichia coli CRP is able to bind other cyclic nucleotides, such as cyclic GMP in the ligand binding site (Takahashi et al., 1980). Bai et al. (2005) suggested that the cyclic AMP binding site present in Rv3676 is specific for cyclic AMP. Partial proteolysis experiments with trypsin-treated Rv3676 and DNA in the presence and absence of ligand showed that presence of cyclic AMP had a protective effect during proteolysis. This protective effect was not observed when AMP was present (Bai et al., 2005).

In order to verify the suggestions by Bai et al. (2005), ITC experiments were performed (Methods Section 2.4.11) to screen for interactions between Rv3676 and cGMP or 2'5'-cyclic AMP (Figure 4.18). The interaction between Rv3676 and cyclic AMP was used as a positive control to verify that the prep of purified Rv3676 contained a functional cyclic nucleotide binding pocket. Cyclic AMP was shown to bind endothermically (blue traces, Figure 4.18) with the parameters discussed in Section 4.5.1.1. There was no observable interaction between cyclic GMP (red line) or 2'3'-cyclic AMP (green line) and purified Rv3676. This supports the notion that the cyclic AMP binding site of Rv3676 is specific for 3'5'-cyclic AMP and differs from the *E. coli* CRP paradigm.

4.5.3.2   The cyclic nucleotide binding site in Msmeg_6189 is specific for 3'5'-cyclic AMP

In order to investigate the specificity of the Msmeg_6189 cyclic nucleotide binding site, ITC experiments were performed (Methods Section 2.4.11) to screen for interactions between Msmeg_6189 and cGMP and 2'3'-cyclic AMP. The interaction between Msmeg_6189 and cyclic AMP was used as a positive control to verify that the prep of purified Msmeg_6189 contained a functional cyclic nucleotide binding pocket. Cyclic AMP was shown to bind endothermically (blue traces, Figure 4.18) with the parameters discussed in Section 4.5.1.1. There was no observable interaction between cyclic GMP (red line) or 2'3'-cyclic AMP (green line) and purified Msmeg_6189. This supports the notion that the cyclic AMP binding site of Msmeg_6189 is specific for 3'5'-cyclic AMP. This is the the same as the observations for Rv3676.

4.6   Mutational analysis of cyclic AMP binding site

As discussed previously, Rv3676 only possesses 4 of the 6 cyclic nucleotide binding residues present in *E. coli* CRP (residues shown in the alignment in Figure 4.2). Collaborators in the Buxton laboratory carried out growth experiments (unpublished) comparing the growth rate of
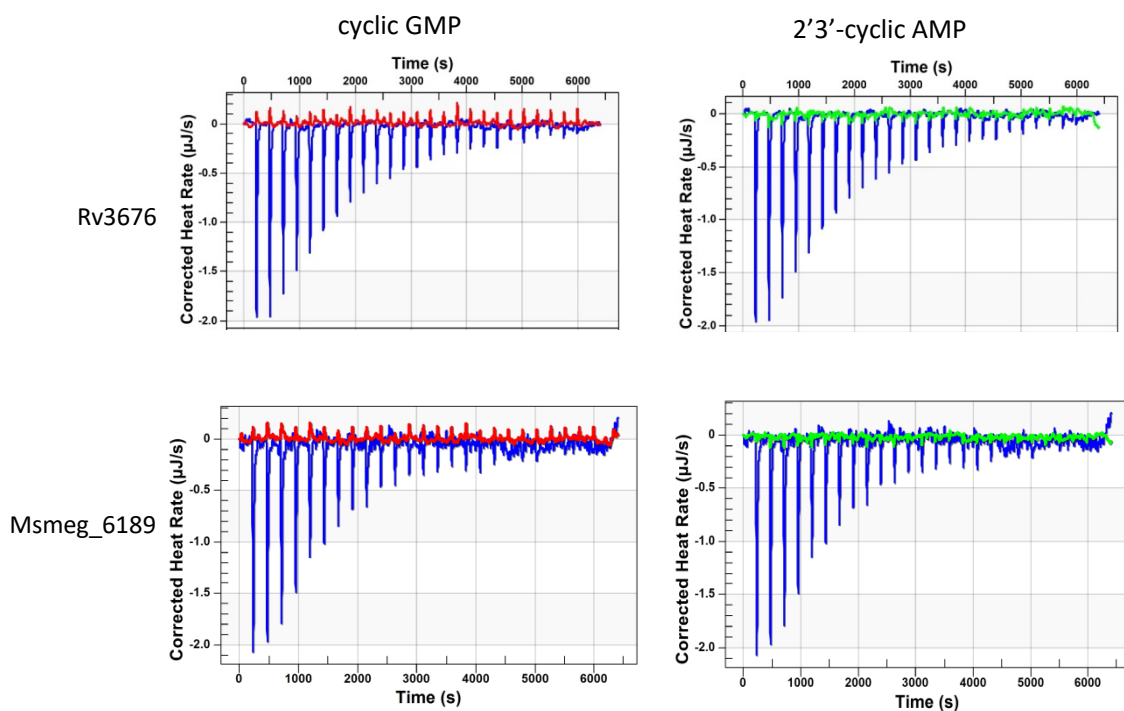
Figure 4.18: ITC traces for mycobacterial CRP proteins titrated with different cyclic nucleotides. The traces belonging to Rv3676 and Msmeg_6189 are indicated. Proteins samples (200 μM) and were subjected to 25 injections (1.97 μl) of ligands (2 mM). The **red** lines corresponds to cyclic GMP. The **green** lines correspond to 2'3'-cyclic AMP. The **blue** lines correspond to 3'5'-cyclic AMP.

wild type *Mycobacterium tuberculosis* H37Rv with an Rv3676 deletion mutant (Δ*crp*). The Δ*crp* mutant had a significantly lower rate of growth than wild-type H37Rv (Figure 4.19). Complementation of the Δ*crp* strain with a copy of the WT *Rv3676* gene restored the growth rate to one comparable with wild-type (WT) H37Rv. Constructs with site-directed mutations in the in the cyclic-AMP binding pockets of Rv3676 were produced (S91W, S91A, N135W and N135A) and were introduced into the *crp* deletion strain and growth data were collected. In addition to this DNA-binding mutant (N67M+N137K and R188A+E189A) versions of Rv3676 were introduced into the Δ*crp* strain.

The growth rate of both *crp* deletion strains that were complemented with a DNA-binding mutant version of Rv3676 (Δ*crp*/R188A+E189A or Δ*crp*/ N67M+N137K), were significantly lower than that of WT H37Rv or the Δ*crp*/WT *crp* complement strain. This suggests that the DNA-binding function of Rv3676 is important in growth for *M. tuberculosis*. However, complementation of the *crp* deletion strain with *crp* coding for mutations in the cyclic-AMP binding site (Δ*crp*/S91W, Δ*crp*/S91A, Δ*crp*/N135W and Δ*crp*/N135A) showed similar or only slightly reduced growth rates compared with WT H37Rv and the Δ*crp*/WT complement strain. This implies that cyclic AMP is not essential for Rv3676 function *in vivo* and requires experimental evidence that these mutants actually have reduced or diminished ability to bind cyclic AMP.

Additionally, it would be interesting to investigate the effect of mutating the two residues involved in cyclic AMP binding in *E. coli* CRP that are missing in Rv3676. Does the resulting protein bind cyclic-AMP with a higher affinity than WT Rv3676? Does the protein show more dependence on cyclic-AMP for DNA-binding activity?

Aims:

- Use site-directed mutagenesis and purify the cyclic AMP binding site mutants (S91W and N135W) and perform ITC and EMSA experiments to verify that they either cannot bind cyclic AMP or have reduced ability for binding cyclic AMP. This will complement the *in vivo* data.
- Investigate what happens if the two additional residues that are involved in cyclic AMP binding in *E. coli* CRP, but differ in Rv3676, are introduced into Rv3676 (i.e. T90R and N135S). Will this create a cyclic-AMP "hyper"-binder or have any downstream effects on cyclic AMP dependence of Rv3676?

4.6.1 Generation of site-directed cyclic-AMP binding mutants

Primers were designed to incorporate the desired mutations into the Rv3676 coding region
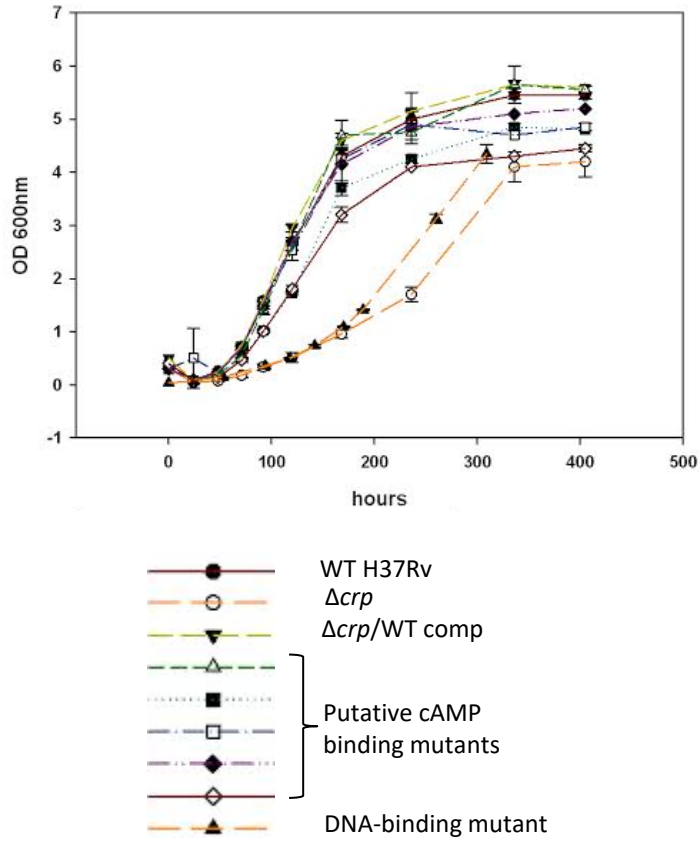
Figure 4.19: Growth of various CRP mutants in comparison to wild-type H37Rv. The growth of various strains of *M. tuberculosis* H37Rv in 7H9 media. Error bars show mean ± SEM of three biological replicates.

located in pGS2132 template DNA. Site-directed mutagenesis was then performed by the methods outlined in Section 2.3.10 using the following primer combinations for each mutation: LD59 and LD60 for T90R; LD61 and LD62 for N135S; LD63 and LD64 for S91W and LD65 and LD66 for N135S. The sequences of all primers can be found in Table 2.1. After the thermal cycling stage of the protocol was completed, samples were digested with EcoRI and analysed on an agarose gel (Figure 4.20a) to confirm the presence of plasmid DNA. Plasmid DNA was observed of the expected size (~7.4 kb) in the lanes corresponding to all 4 of the mutants. The DNA was used in transformation reactions (Methods Section 2.2.7) and a single colony for each mutant construct was restreaked. A single colony was used to inoculate an overnight culture and plasmid DNA was purified and sent for sequencing. Once it was confirmed that the new plasmids contained the desired mutation and no other mutations within the Rv3676 coding region, they were used to transform *E. coli* BL21 λDE3 Δ*cya* cells in order to create a strain for overproduction of the protein for purification.

4.6.2 Purification of cyclic AMP binding site mutants

Mutant proteins were overproduced and purified by the same methods used for Rv3676, which are discussed in Results Section 4.3.1. The only difference being that the overproduction vector used was a derivative of pGS2132 containing the relevant mutation (discussed in Section 4.6.1).

Cells were grown up, induced, collected and lysed as detailed in Section 4.3.1.1. Nickel affinity chromatography was performed as detailed in Section 4.3.1.2. SDS-PAGE analysis was performed on fractions from the eluent (Figure 4.20b). In (i) an abundant polypeptide of size ~26 kDa is present in lanes 9 and 10, corresponding to purified Rv3676 T90R. In (ii), an abundant polypeptide of size ~26 kDa is present in lanes 5 to 9, corresponding to purified Rv3676 N135S. In (iii) an abundant polypeptide of size ~26 kDa is present in lanes 8 to 13, corresponding to purified Rv3676 S91W. In (iv), an abundant polypeptide of size ~26 kDa is present in lanes 9 to 13, corresponding to purified Rv3676 N135W.

Fractions containing purified mutant Rv3676 protein were combined and imidazole was removed typically by dialysis against 20 mM sodium phosphate (pH 7.4) with 0.2 M NaCl. Protein samples were concentrated using VivaSpin columns with a MWCO of 10,000 Da and proteins were stored at room temperature. Samples were sent for mass spectrometry analysis, which confirmed that the full length mutant protein was purified.

4.6.3 ITC analysis of cyclic AMP binding in Rv3676 mutant

Isothermal titration calorimetry was performed on purified cyclic AMP binding mutant versions of Rv3676 in order to ascertain if they could bind cyclic AMP and if so with what affinity.
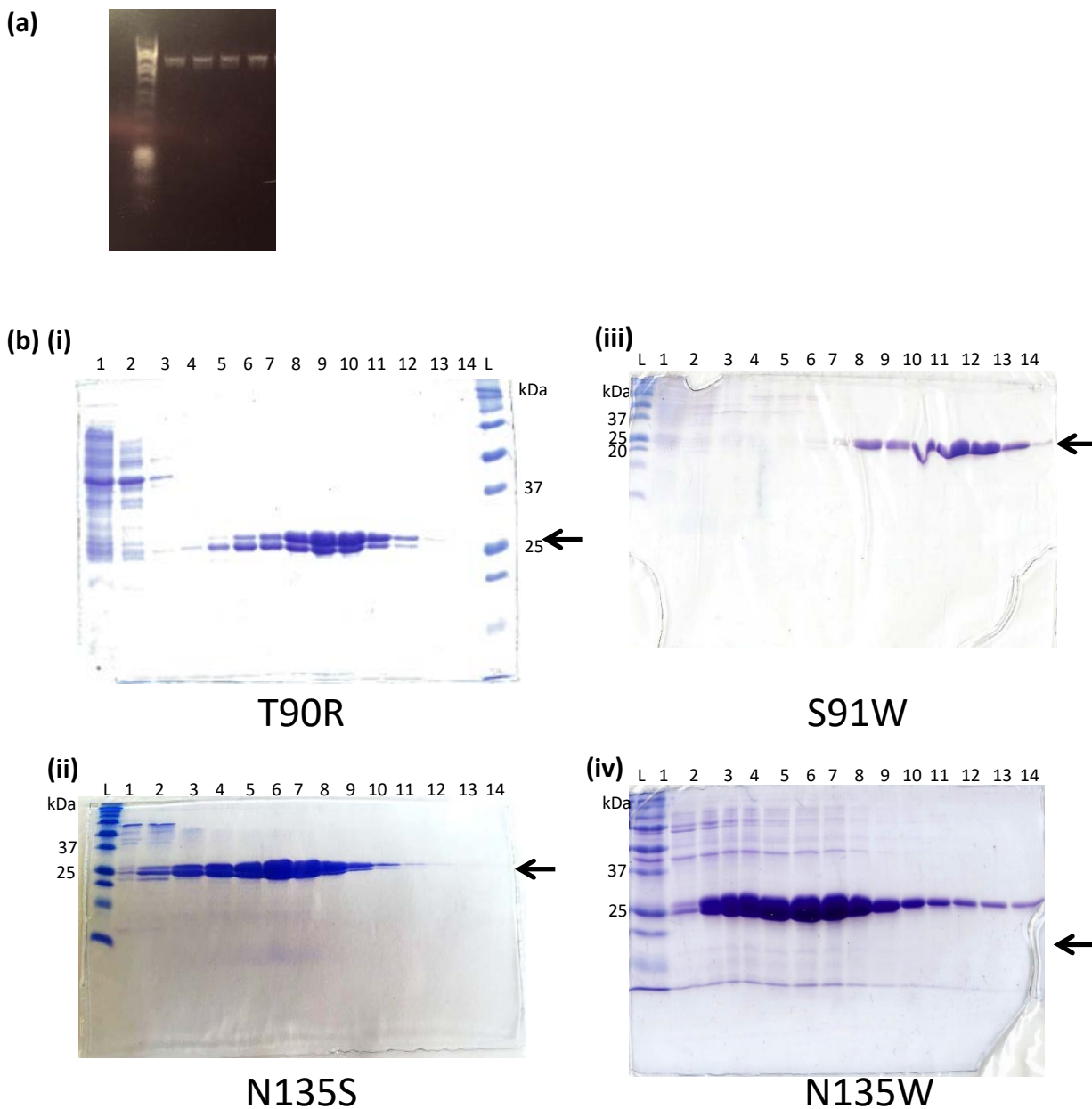
Figure 4.20: Generation and purification of cyclic AMP binding pocket mutants. (a) Agarose gel electrophoresis analysis of site-directed mutagenesis products. Lane L contained 5 µl DNA Hyperladder I (Bioline). Lanes 1 to 4 contained 5 µl of T90R, N135S, S91W and N135W thermal cycling products after digestion with EcoRI for 1 h at 37°C. (b) SDS-PAGE analysis of fractions (10 µl) after nickel affinity chromatography purification of the following cyclic AMP binding mutant Rv3676 proteins: i) T90R, ii) N135S, iii) S91W and N135W. Lanes L contained Precision Plus All Blue protein standards (BioRad).

4.6.3.1 Rv3676 T90R is not able to bind cyclic AMP

In order to determine the effect that substituting T90 with R in Rv3676 had on the proteins cyclic AMP binding ability, ITC experiments were performed (Methods Section 2.4.11) using 200 µM Rv3676 T90R protein with 25 injections of 1 mM cyclic AMP. Our predictions were that the protein would be able to bind cyclic AMP, as it still contained the four residues required for interaction with cyclic AMP, but may have increased affinity for cyclic AMP due to the presence of R90, which is involved in cyclic AMP binding in *E. coli* CRP. Interestingly we observed no binding of cyclic AMP to Rv3676 T90R (Figure 4.21a). The interaction between Rv3676 and cyclic AMP was used as a positive control. This result was surprising and in order to investigate if the DNA-binding function of the protein had been affected, BLItz or EMSA analysis should be performed. Preliminary BLItz experiments indicated that T90R was able to bind to P*rpfA* with similar affinity to WT Rv3676 (data not shown). However, further repeats are necessary to confirm this finding.

4.6.3.2 Rv3676 N135S is able to bind cyclic AMP with greater affinity than WT Rv3676

In order to determine the effect that substituting N135 with S in Rv3676 had on the proteins cyclic AMP binding ability, ITC experiments were performed (Methods Section 2.4.11) using 200 µM Rv3676 N135S protein with 25 injections of 1 mM cyclic AMP. Our predictions were that the protein would be able to bind cyclic AMP, as it still contained the four residues required for interaction with cyclic AMP, but may have increased affinity for cyclic AMP due to the presence of S135, which is involved in cyclic AMP binding in *E. coli* CRP. Rv3676 N135S was able to bind cyclic AMP (Figure 4.21b) and fitting to an independent binding model provided a K_D of ~45 µM. This affinity is ~6.5-fold tighter than the interaction between cyclic AMP and WT Rv3676. In E. coli CRP, which contains an S at the position equivalent to 135 in Rv3676, the affinity for cyclic AMP is much tighter (~13-16 µM) than it is for Rv3676. This result suggests that this residue plays an important role in the interaction with cyclic AMP.

The next step of the project would involve characterising if any changes to the DNA-binding ability of the Rv3676, or dependence on cyclic AMP for DNA-binding, have occurred in this mutant. This could be achieved by EMSA experiments and quantified by BLItz. Unfortunately, due to time constraints this work has not yet been carried out.

4.6.4 Conclusions and further work

In this Section (4.6), a set of four predicted cyclic AMP binding mutants have been successfully generated using site-directed mutagenesis. Two of which were involved in introducing residues

(a)

(b)
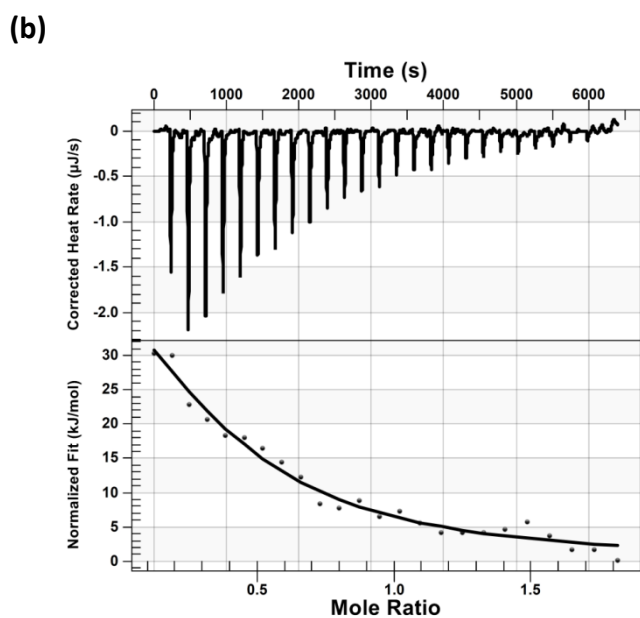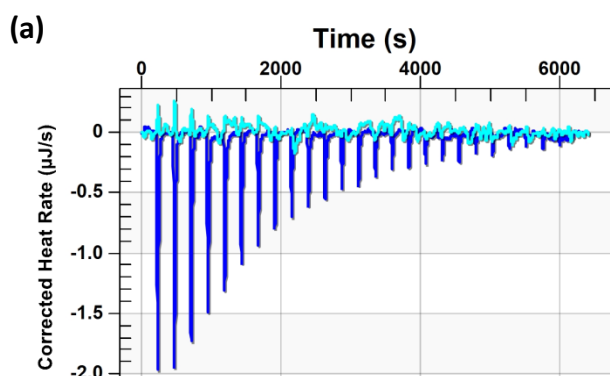
Parameter	Rv3676 N135S
K_a (M^{-1})	2.322 x 10^4 ± 1.86 x 10^3
K_d (M)	4.307 x 10^{-5}
n (per monomer CRP)	0.922 ± 0.103

Figure 4.21: ITC analysis for Rv3676 cyclic AMP binding site mutant proteins titrated with 3'5'-cyclic AMP. (a) Trace belonging to Rv3676 T90R is shown in light blue, which is overlaid on a trace of wild-type Rv3676. Proteins samples (200 µM) and were subjected to 25 injections (1.97 µl) of 3'5'-cyclic AMP (2 mM). (b) Trace belonging to Rv3676 N135S (200 µM) was titrated with 25 injections (1.97 µl) of 3'5'-cyclic AMP (1 mM).Data was fitted to an independent binding site model to provide thermodynamic parameters of the interaction.

to Rv3676 which are present and important in cyclic AMP binding in *E. coli* CRP, but differ in Rv3676. This resulted in the generation of one protein (T90R) which was unable to bind cyclic AMP under the conditions tested. Our preliminary BLItz experiments indicated that the DNA-binding ability of Rv3676 T90R remained unaffected. The second protein (N135S) showed increased affinity for cyclic AMP compared to WT Rv3676. Further work should include characterising the DNA-binding ability of Rv3676 N135S, which can be achieved by EMSA experiments or BLItz using P*rpfA*.

In addition to this, two other mutants (N135W and S91W), believed to abolish or significantly reduce cyclic AMP binding by steric hindrance, have been successfully overproduced and purified. ITC experiments should be performed on these proteins to establish their ability to interact with cyclic AMP and the DNA-binding functions should be tested by EMSA and/or BLItz with P*rpfA*. These results will provide more insight into the growth curve data (Figure 4.19).

4.7 Discussion and further work

4.7.1 Conclusions

During this study, two mycobacterial CRP proteins, Rv3676 and Msmeg_6189, were successfully overproduced and purified from an *E. coli* host. Rv3676 was shown to bind to the CRP site located in the promoter of the *rpfA* gene by slot blots, BLItz and EMSA experiments. This interaction did not depend on cyclic AMP, but a slight enhancement of DNA-binding (up to ~10-fold) was observed in the presence of cyclic AMP. ITC experiments provided evidence that Rv3676 can bind cyclic AMP with relatively weak affinity (~300 µM), which could be due to the intracellular concentration of cyclic AMP being as high as 4 mM (Padh and Venkitasubramanian, 1977). Rv3676 was not able to bind cyclic GMP or 2'3'- cyclic AMP under the conditions tested. cAMP-binding mutants were created by SDM and successfully overproduced and purified. ITC experiments revealed that T90R was not able to bind cyclic AMP, however N135S was able to bind cyclic AMP with increased affinity compared to the WT protein.

Msmeg_6189 was shown to interact with the CRP site located in the promoter of the *sdh1* gene, but not with *sdh2*. The interaction with P*sdh1* was dependent on cyclic AMP being present, which was surprising as the important residues in cAMP and DNA-binding are conserved with the Rv3676 protein. ITC experiments verified Msmeg_6189 binds cyclic AMP with a fairly weak affinity (~350 µM). Msmeg_6189 was not able to bind cyclic GMP or 2'3'-cyclic AMP under the conditions tested.

4.7.2 Controversy over the cyclic AMP dependence for DNA-binding by Msmeg_6189

During the course of this work, Sharma et al. (2014) published a study comparing the two CRP proteins present in *M. smegmatis* (Msmeg_6189 and Msmeg_0539). The paper suggested that Msmeg_6189 can bind to a CRP site located in the intergenic region between the *Msmeg_3780* and *Msmeg_3781* genes. The CRP site is located approximately ~25 bp upstream of *Msmeg_3781* transcript start site and ~200 bp upstream of the *Msmeg_3780* transcript start site (Sharma et al., 2014). Interestingly, neither of these genes were identified as having significantly altered expression during transcriptomic analysis performed with the Δ*Msmeg_0539* or the *Msmeg_6189* overexpression strains (Aung et al., 2015). Sharma et al. (2014) observed binding of Msmeg_6189 to this intergenic region in the absence of cyclic AMP and propose that the protein does not require cyclic AMP for DNA-binding. This is contrary to what has been observed in this study with the *sdh1* promoter.

Lee et al. (2014) observed binding of Msmeg_6189 to the CRP site located in the promoter of the *ahpC* gene. Interestingly, these binding assays were only performed in the presence of cyclic AMP, so the question of the Msmeg_6189:DNA interaction being dependent on cyclic AMP was not addressed in the study.

In order to clarify this issue it would be useful to perform further experiments with both DNA regions. Initially repeating the EMSA protocol used in this study with the *Msmeg_3780* and *Msmeg_3781* intergenic region to ascertain if the observations concur with the Sharma et al. (2014) study. This may help if it is a property of the buffers or purification steps that is affecting the properties of the protein. The protein used in this study is tagged with a His-tag, which is removable with thrombin. EMSA experiments performed with Rv3676 (data not shown) after the His-tag had been removed showed no difference in binding, so for the experiments performed with Msmeg_6189 the tag was left to minimise the number of steps to subject the protein to before use. The Sharma et al. (2014) study purified Msmeg_6189 using a GST-tag, which was subsequently removed. Performing the EMSA experiments with P*sdh1* again after removal of the His-tag would be a simple experiment to perform and could provide insight on whether the tag is affecting the dependence on cyclic AMP. However, previous work with the closely related Rv3676 suggests the tag does not affect binding.

 It would be useful to perform promoter::*lacZ* reporter fusion assays with different promoter regions, including *Msmeg_3780* and *Msmeg_3781*intragenic region, *sdh1*, *ahpC* and others identified in the microarray analysis. These should be performed in a cyclic AMP-free background (where no adenylate cyclases are present), a background of "normal" intracellular

cyclic MAP concentration and perhaps also in a strain overexpressing *cya* which should have elevated cyclic AMP concentrations (which may be more representative of a mycobacterial cell). Due to the *M. smegmatis* genome encoding multiple adenylate cyclases, the experiment would be more complicated to perform. *Escherichia coli* Δ*cya* mutant is available in which the gene encoding the only adenylate cyclase in the *E. coli* genome (*cyaA*) is deleted. This would be an ideal cyclic AMP-free background to test the reporter constructs in, compared with WT *E. coli* cell and a WT *E. coli* cell containing an inducible *cya* overexpression vector. Beta-galactosidase assays could then be performed and the levels of expression resulting from the different promoters in the different intracellular cyclic AMP concentrations can then be compared to look for differences which will provide insight into whether the levels of *lacZ* expression are dependent on the cyclic AMP. Experiments should be performed to measure the intracellular levels of cyclic AMP in the different strains. There are numerous enzyme-linked immunosorbent assay (ELISA) kits available for ascertaining the levels of cyclic AMP in cell lysates.

Our collaborators were not able to obtain a *Msmeg_6189* deletion mutant and suspected that the gene may be essential for growth in *M. smegmatis* (Aung et al., 2015) and other published studies have had issues deleting the gene (Sharma et al., 2014). Thus, transcriptomic experiments to identify putative genes regulated by Msmeg_6189 were performed using a strain overexpressing the protein (Aung et al., 2015). *Msmeg_6189* is not an essential gene and a study has since been published that demonstrated impaired growth rate of an *Msmeg_6189* deletion mutant and suggested the strain reached stationary phase at a lower cell density compared to WT *M. smegmatis* cells (Lee et al., 2014).

Discussions over Msmeg_6189 and Rv3676

It is interesting how the cyclic AMP binding affinity of Rv3676 and Msmeg_6189 are similar; the proteins are 98% homologous, including the key residues involved in cyclic AMP and DNA interactions; yet the effect of cyclic AMP on the DNA-binding properties of the proteins are so different.

Further investigation into the residues which differ between between the two mycobacterial CRPs would be beneficial and may provide insights into this profound difference. Performing site directed mutagenesis of Rv3676 towards Msmeg_6189 sequence (i.e. substitution of residues S17T, I19V, A61S, S111A and S120A) and combinations of double, triple and quadruple mutants and purification of these proteins will allow their importance to investigated. Purified proteins can be screened using EMSAs to determine if any of the mutants revert from cyclic AMP independence for binding promoters (such as *sdh1* or *rpfA*) to

requiring cyclic AMP for binding the promoter region. This work has already commenced, whereby constructs allowing the overproduction of the single amino acid substitution proteins have been created using SDM (Method Section 2.3.10) on the pGS2132 plasmid. Mutant plasmids were sequenced to confirm incorporation of the desired mutations and no others and successful plasmid DNA was used to transform *E. coli* BL21 λDE3. The proteins have successfully been overproduced and can be purified (Figure 4.22).

In addition to investigating the effect of the differences in the amino acid sequence of Rv3676 and Msmeg_189, it would also be worthwhile further investigating if these produce any subtle differences in the structure of the cyclic AMP binding sites, despite the cyclic-AMP binding residues themselves being conserved between the two proteins. It will also be worth considering if the specific sequence of the CRP binding site and the wider context of the promoter have an effect.

Interaction of Rv3676 and *sdh1*

The direct binding of CRP to *sdh1* (Rv0249c–Rv0247c) in *M. tuberculosis* has not yet been studied, but Rv3676 is computationally predicted to bind this promoter region (Bai et al., 2005; Krawczyk et al., 2009). Further EMSA experiments should be performed with purified Rv3676 and *M. tuberculosis* P*sdh1* DNA to confirm if it can bind this gene. It would also be interesting to determine if cyclic AMP is necessary and/or enhances DNA-binding.

It would also be useful to determine if Rv3676 is able to bind the P*sdh1* (from *M. smegmatis*) and what effect cyclic AMP has on the interaction. These results could be directly compared to EMSAs performed with Msmeg_6189 and may provide further insight onto if the exact DNA sequence of the CRP site, and potentially surrounding DNA, may have an effect on cyclic AMP dependence, or if this difference is solely related to slight differences in the strutures of the two proteins. For similar reasons, EMSAs could be performed with Msmeg_6189 and P*rpfA* to determine if the protein behaves in a cyclic AMP dependent manner with other promoters.

Crystal structures of Msmeg_6189

In order to gain further structural information to provide insights into the cyclic AMP dependence observed with Msmeg_6189, it would be valuable to generate crystal structures of apo- and holo-Msmeg_6189 in complex with cyclic AMP, and additionally when in complex with DNA. It would be of particular interest to overlay the structures with the Rv3676 counterparts and look for possibly causes for the differences in cyclic AMP dependence between the two proteins, paying particular attention to the orientation and exposure of the DNA-binding helices.
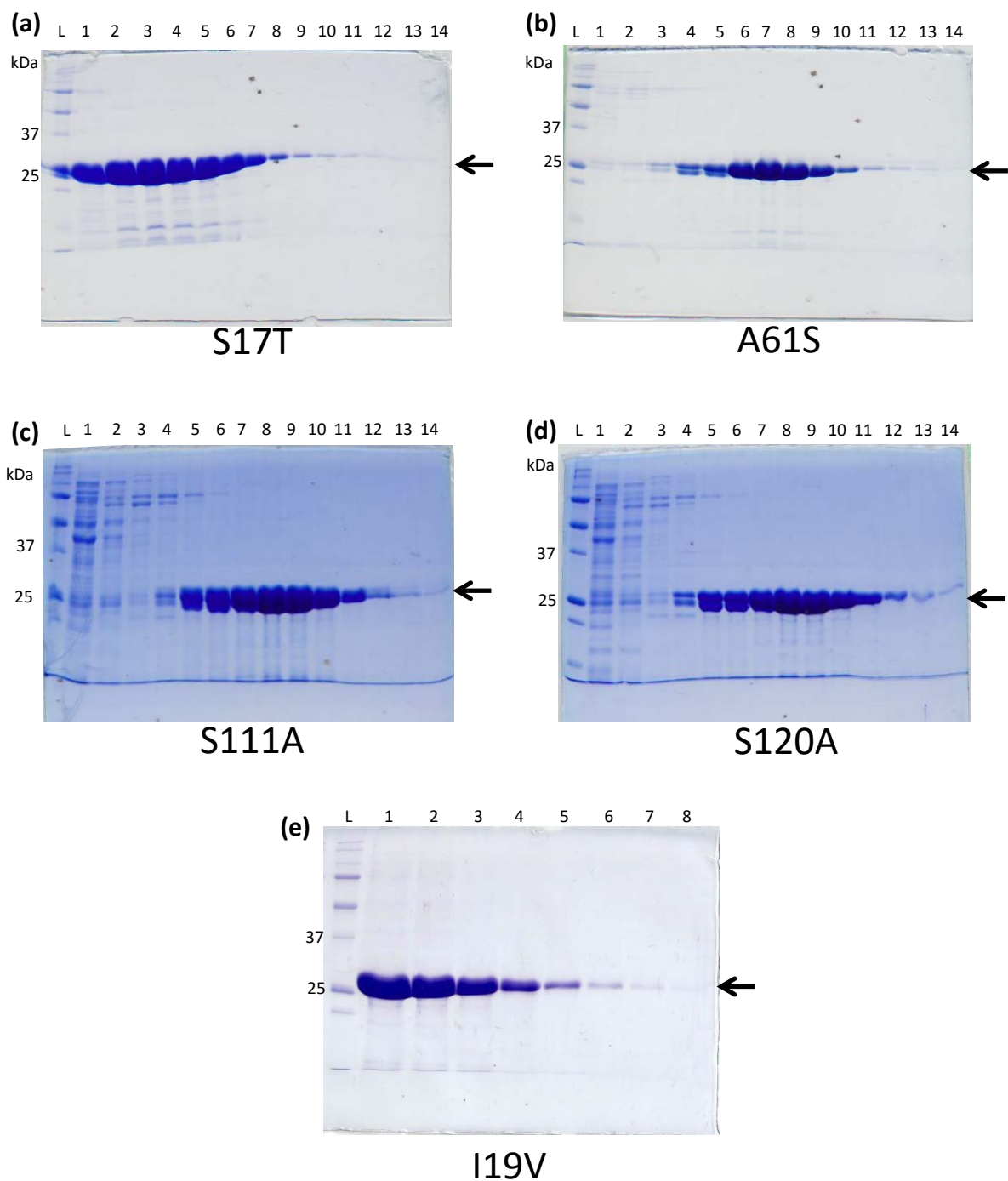
Figure 4.22: SDS-PAGE analysis of mutant Rv3676 proteins after purification by nickel affinity chromatography. Lane L contained 5 µl Precision Plus All Blue protein standards (BioRad). Lanes 1 to 14 contained fractions (10 µl) after nickel affinity chromatography purification of the indicated Rv3676 mutant protein.

COPASI modelling to compare *M. tuberculosis* Rv3676 with *E. coli* CRP

It would of interest to produce a model to allow for comparison of cyclic AMP-independent Rv3676 with the well-characterised cyclic AMP-dependent *E. coli* CRP paradigm. This model would take the interaction overview shown in Figure 4.23, for which data for the rate values for both Rv3676 and *E. coli* CRP have been generated in this work or in other published studies. The user would be able to input different in intracellular cyclic-AMP concentrations for the two different proteins, which would allow for modelling of outputs (i.e. changes in transcription and chromosome structure resulting from interactions between CRP and DNA). This would be useful as it would allow simulation of the effects a mycobacterial intracellular cyclic AMP concentration (i.e. as high at 4 mM) would have on the *E. coli* system and also what *E. coli* intracellular levels of cyclic AMP (i.e. moderate) would have on the *M. tuberculosis* system of regulation.

Comparison of CRP-based regulation of rpfA in *S. coelicolor* and *M. tuberculosis*

Loss of *rpfA* in *S. coelicolor* results in reduced ability of the bacteria to germinate from spores (Haiser et al., 2009) and tight gene regulation is necessary. St-Onge et al. (2015) propose that *rpfA* regulation in *S. coelicolor* is complex and involves at least three levels of control. Two of regulators are also involved in *rpfA* regulation in *M. tuberculosis,* as discussed throughout this thesis. The first of the common regulatory methods is regulation by Crp and cyclic AMP. The second common regulatory element is the cyclic di-AMP responsive riboswitch. The third level of control is proposed to be a growth stage dependent proteolysis of RpfA in response to ppGpp.

Loss of *rpfA* function in *S. coelicolor* results in delays in germination initiation (Haiser et al., 2009). CRP has been shown to bind to and activate transcription of *rpfA* in *S. coelicolor*. Transcriptomic analysis revealed that *rpfA* expression was upregulated 28-fold when *crp* was induced in a Δ*crp* mutant (St-Onge et al., 2015). Transcriptomic analysis of *rpfA* expression in *M. tuberculosis* showed a >6.4-fold increase in expression of *rpfA* in the wild-type strain compared to the Δ*rv3676* mutant (Rickman et al., 2005). qRT-PCR analysis was performed which showed *rpfA* transcript levels increased 2-fold in wild-type *S. coelicolor* strains that were overexpressing crp in comparison to overexpressing *crp* in a Δcrp mutant strain (St-Onge et al., 2015). In M. tuberculosis, qRT-PCR showed a 3.8-fold increase in gene expression in the wildtype strain compared with the Δ*rv3676.*

Chromatin immunoprecipitation experiments were performed with anti-CRP to isolate DNA that had bound to *S. coelicolor* Crp with the aim of determining if Crp directly or indirectly

Input
C

P'D0   P*C2D0

P'D1   P*C2D1

Outputs   P' ⇌ P ⇌ PC ⇌ PC2 ⇌ P*C2 ⇌   Outputs
P'D2   R9-12 R1   R2   R3   R4   R5-8   P*C2D2

P'D3   P*C2D3
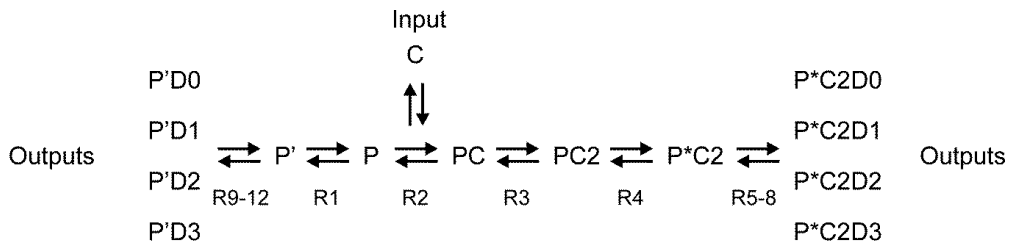
C - cAMP
P' - CRP
P - CRP capable of cAMP binding
P*C2 - CRP-cAMP complex capable of site-specific DNA-binding
D0 - high-affinity CRP sites
D1 - medium affinity CRP sites
D2 - low affinity CRP sites
D3 - non-specific DNA

Figure 4.23: Overview of a potential model to simulate the response of CRP to changes in the intracellular concentration of cAMP. Where R values are the rate constants. The CRP homodimer can exist in several different forms as indicated by the key and the description of the model in the text. CRP emerges from the ribosome in a form (P′) capable of low affinity DNA binding (P′D0-3; reactions R9-12). P′ is capable of conversion to a form (P, in R1) competent for binding cAMP (C). P binds two molecules of C in reactions R2 and R3 to form species PC2. PC2 undergoes a conformational change that permits site-specific high affinity DNA-binding in reaction R4. The DNA-binding competent CRP-(cAMP)$_2$ species (P*C2) binds at CRP-binding sites in target promoters (P*C2D0-D2) and to non-specific sites across the chromosome (P*C2D3) in reactions R5-8. Changes in transcription and chromosome structure (OUTPUTS) result from interactions between CRP and DNA.

regulates the expression of *rpfA*. qPCR analysis revealed significant levels of *rpfA* promoter DNA was enriched, suggesting that Crp directly binds the *rpfA* promoter. St-Onge further identified the putative Crp binding sequence (**TGAGA**gatagg**TCTCA**) ~300 nt upstream of the *rpfA* translation start site (St-Onge et al., 2015), which is similar to the binding site present in the rpfA gene promoter in *Mycobacterium tuberculosis* (Rickman et al., 2005).

5 Investigating the putative *ydaO*-type riboswitch present in the *rpfA* gene of mycobacteria

5.1 Introduction

5.1.1 What are riboswitches?

Riboswitches are complex folded RNA domains, which are typically located in the non-coding 5' untranslated region (UTR) of messenger RNAs (mRNAs). They act as receptors for specific ligands and alter gene expression in response to ligand binding (Mandal and Breaker, 2004). Riboswitches consist of two main components – an aptamer and an expression platform. The aptamer domain is formed of a precise three-dimensional structure, which is able to selectively bind a target ligand. This induces structural changes, which result in differences in the folding pattern of the adjacent expression platform. Ultimately this is translated in to changes in gene expression.

There are eight main classes of riboswitches that have been identified (Barrick et al., 2004), of which five have had their cognate ligand identified. These ligands include Mg^{2+} (Dann et al., 2007), glycine (Mandal et al., 2004), $PreQ_1$ (Roth et al., 2007), glucosamine-6-phosphate (Winkler et al., 2004) and cyclic di-AMP (Nelson et al., 2013). The remaining 3 classes of riboswitch currently remain as 'orphan' riboswitches, where the ligand is waiting to be discovered. The different classes have numerous members, that are based on individual consensus sequences and the representatives are widespread across different bacterial species. This supports the theory that they play an important role in the control of cellular processes.

5.1.2 The *ydaO* family of riboswitches

In this project, I explored what ligand or ligands are capable of binding to the *rpfA* riboswitch in mycobacteria. The *rpfA* riboswitches have a high degree of sequence homology to the *ydaO*-type riboswitch, which was first identified in *B. subtilis* (Block et al., 2010). There are over 3,000 representatives of the *ydaO* class of riboswitch that can be found across a plethora of species. They are involved in the regulation of a diverse range of genes, including amino acid transporters, acetyltransferases and those involved in cell wall metabolism, e.g. *rpfA* (Watson and Fedor, 2012). Of particular relevance to this project is the *ydaO*-class riboswitch in the *ydaO* gene of *Bacillus subtilis* (Nelson et al., 2013; Watson and Fedor, 2012) and the putative *ydaO*-class riboswitches present in the *rpfA* genes of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium smegmatis* and *Mycobacterium marinum*.
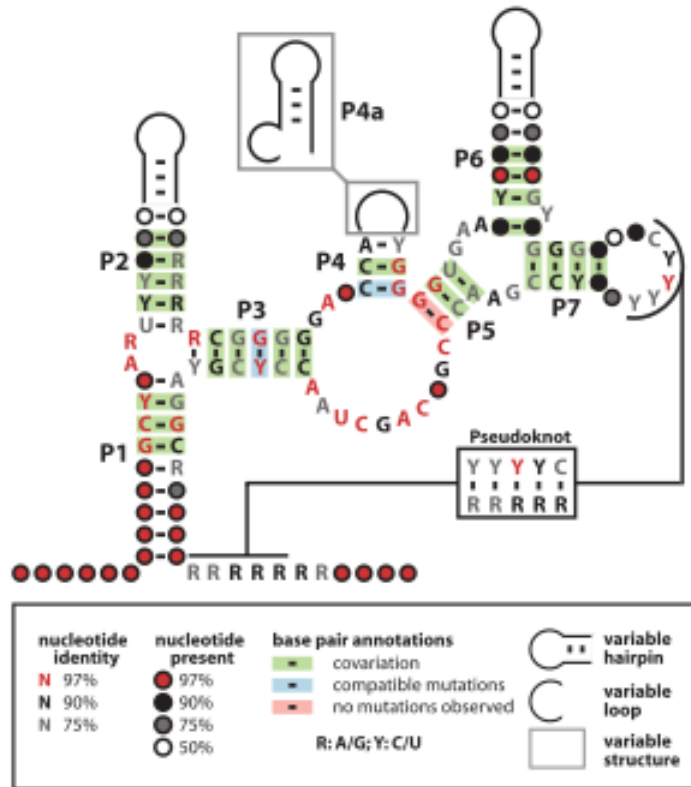
Figure 5.1: shows the consensus sequence for the *ydaO*-class of riboswitches. This figure is taken from Block *et. al*, 2010. There are 3,012 identified representatives of this riboswitch class. Nucleotides that are present in greater than 97% (black); 90% (grey) and 75% (red) of representatives are shaded. Predicted substructures are labelled P1 through to P7 and the predicted psuedoknot is also labelled. In certain family members where no transcription terminator is found, a putative psuedoknot is formed between the loop of P7 and the 3' end of the RNA, where the Shine Dalgano sequence is located. Green shading is present where there is phylogenetic evidence for base pairing.

5.1.2.1 Structural elements of the *ydaO* consensus RNA motif

A consensus sequence for the *ydaO*-class, which is based on comparison of 580 family members is shown in Figure 5.1 (structure taken from Block *et. al*, 2010). The main structural elements present in the consensus sequence are substructures P1 to P7 (base-paired hairpins) and a psueoknot structure (Nelson et al., 2013). In some members, including the *B. subtilis* *ydaO* riboswitch, a transcription terminator element is present directly 3' of, but not overlapping, the aptamer region. In species containing these terminators, gene regulation by the riboswitch is at a transcriptional level (Block et al., 2010). In *B. subtilis*, the *ydaO* riboswitch acts as an "off" switch, whereby ligand binding results in repression of gene expression, which has been further evidenced in reporter constructs containing wild-type and mutated riboswitches (discussed in more detail below) (Block et al., 2010). However, for the majority of members, including the *Streptomyces coelicolor* riboswitch responsible for the regulation of various cell wall enzymes (St-Onge et al., 2015), the gene regulation is believed to be at the translational level. The currently accepted model is that a pseudoknot structure is formed between P7 and the 3' end of the riboswitch, where the Shine-Dalgano sequence is located, resulting in occlusion of the ribosome binding site (Haiser et al., 2009). In bacteria and archaea, the Shine-Dalgano sequence plays a vital role in ribosome recruitment to the messenger RNA and hence support the notion that the majority of *ydaO*-class riboswitches regulate genes at the translational level (Block et al., 2010).

5.1.2.2 Evidence for regulatory role of the *B. subtilis ydaO* riboswitch

A transcriptional reporter construct was created containing the *B. subtilis ydaO* riboswitch fused to the *lacZ* gene. In order to explore the importance of certain substructural elements on the riboswitch function, mutant versions of the reporter construct were generated. This included M1 with altered pairing in the P4a stem and M3 with altered pairing in the P6 stem. Interestingly, the effect of M1 mutation, when compared to wild-type, was only minor and hence is not deemed essential for riboswitch function. The P4a substructure is not present in all family members, and, when it is present, is not highly conserved like other substructural elements. In addition, it is located further from the conserved core region of the structure. Taken together, these observations support the theory that P4(a) is not essential for functionality. In contrast, the M3 mutation showed significantly altered expression levels, which could be returned to wild-type levels in a 'complement' construct (M4), which reinstated base-pairing within the P6 stem. It is believed that P6 has an important structural role in functionality of the *ydaO* riboswitch class (Block et al., 2010). Mutations disrupting the base-pairing in P3 and P7 regions also resulted in significant alterations in gene expression.

However, successful 'complementation' of these constructs to wild-type expression levels was not achieved. Nevertheless, the high-degree sequence identity of these regions within the family, support their potential importance in functionality of the riboswitch (Block et al., 2010). Overall the study provides convincing evidence that sequence and structural changes within the *ydaO* motif can result in dysregulation of gene expression and confirm that the motif has a regulatory role.

The majority of the *ydaO* motifs identified are present in Gram-positive bacteria, so perhaps it is unsurprising that many of the genes associated with the motif are involved in cell wall metabolism. This includes genes encoding enzymes responsible for the degradation of polysaccharides and peptide linkers within the peptidoglycan in the cell wall. The motif is associated with transporters of compatible solutes, ions (especially potassium) and amino acids (Block et al., 2010) and other genes linked to the cells response to osmotic shock (including *rpfA*, whose link to osmotic shock is discussed in section 5.1.7).

5.1.3   What is the ligand?

Bioinformatics can be useful in the identification of the ligand for a riboswitch, for example this was pivotal in ligand identification for the molybdenum cofactor (MoCo) riboswitch (Regulski et al., 2008). Although bioinformatic analysis of the *ydaO* motif strongly suggests it functions a riboswitch, the genetic distribution provided little insight in to the likely ligand. After work with their wild-type reporter construct, Block *et al.*, 2010 proposed that the cognate ligand of the *ydaO* riboswitch is likely both present abundantly in rich medium and is also produced within host cells when grown in both rich media and minimal media.

A dual selection screen was developed to allow screening of transposon mutants, whilst assessing riboswitch reporter activity. It was hoped that mutants with alterations in intracellular ligand concentrations would be isolated and, by determining the mutation present, it may provide useful insight into potential candidate ligands e.g. if the mutation directly disrupted the specific biosynthetic pathway involved in the ligand synthesis (Block et al., 2010). The genes identified in the screen include *menH*, which encodes a methylase involved in menaquinone production, and *ndh*, which encodes a major NADH dehydrogenase. As both these mutations will effect energy efflux and the citric acid cycle in the cell, it was suggested that altered regulation of the riboswitch in the mutant strains could be attributed as a secondary effect of mutation of such a gene and not necessarily due to significant alterations to concentrations of the cognate (Block et al., 2010). In addition to this, vast numbers of changes to the cell metabolome of the mutants were identified, which did not provide much help in narrowing down the potential ligands.

There has been some controversy over the identity of the natural ligand for the *ydaO* riboswitch in *B. subtilis*. In 2011, Watson and Fedor published a study providing evidence for the role of ATP as a ligand (Watson and Fedor, 2012), which is covered in more detail in section 5.1.4. However in 2013, the Breaker group published work evidencing that cyclic di-AMP could actually be the natural ligand of the riboswitch, which is also able to bind other structurally-related compounds containing AMP moieties (Nelson et al., 2013). This is covered in more detail in section 5.1.5. Previously, it has been demonstrated that individual riboswitches are capable of interacting with multiple metabolites with structural similarities (Watson and Fedor, 2011). This means it is entirely feasible that the *ydaO* riboswitch could be regulated by a wide range of small adenosine-related nucleotides.

These studies were used as a starting place for identifying the ligand(s) of the *rpfA* riboswitches present in mycobacteria. It was decided that initially experiments would test for interactions with cyclic di-AMP, ATP and different forms of yeast extract. Yeast extract provides a complex and vast mix of metabolites to screen for potential candidate ligands. If a successful interaction was detected in one of the types of yeast extract tested, then further purification steps could be performed to further separate metabolites from the yeast extract into fractions. Fractions with a metabolite that showed an interaction with the RNA could be further purified, each time enriching for the candidate ligand and reducing down the number of other metabolites within the mixture. Eventually it was hoped that mass spectrometry or other analytical techniques could help to identify the ligand.

5.1.4 Adenosine triphosphate (ATP) binds the *ydaO* riboswitch

Research published by Watson and Fedor, suggested that adenosine triphosphate binds to the *ydaO* riboswitch motif from *B. subtilis*. Equilibrium dialysis experiments showed the binding of ATP to the riboswitch motif as a fairly weak interaction with a K_d value of 0.6 ± 0.1 mM. In-line probing (ILP) experiments provided further evidence that the *ydaO* riboswitch can bind ATP and the same K_d value of 0.6 ± 0.1 mM was found. Mutations of bases A69, A70, A82 or a double mutation of C76/U74 abolished ATP-binding (Watson and Fedor, 2012).

Typically the binding affinity between a riboswitch and its ligand is tight, often in the picomolar to low nanomolar range. This millimolar affinity between ATP and the *ydaO* riboswitch is unusual. However, it is also plausible taking into consideration that ATP is a much more common cellular metabolite and is present at much higher concentrations, than other less common riboswitch ligands would be. Intracellular ATP concentrations in *B. subtilis* have been recorded at 2.0 ± 0.5 mM (Watson and Fedor, 2012). The intracellular concentration of cyclic di-GMP, the ligand for the Vc2 riboswitch in *Vibrio cholerae*, has been reported to typically be

between 0.9 and 5.6 µM in *V. cholerae* cells grown in common laboratory media (Koestler and Waters, 2013). This is significantly lower than the intracellular concentration of ATP in bacterial cells. The dissociation constant (K_D) of the interaction between the *Vibrio cholerae* 110 Vc2 riboswitch and cyclic di-GMP is reported to 1 nM (Sudarsan et al., 2008).

The *ydaO* riboswitch of *B. subtilis* is predicted to act as an 'off switch', and ligand binding is believed to cause a termination of transcription (Block et al., 2010). A tight interaction between ATP and the *ydaO* riboswitch when intracellular concentrations of ATP are so high would mean that majority of riboswitches present within a cell would be in the holo-form, terminating transcription of the *ydaO* gene. With a tight affinity, it would take much larger drop in intracellular ATP concentrations, arguably to significantly disruptive concentrations, to see significant changes in expression levels.

Watson and Fedor developed a *ydaO*-reporter β-galactosidase assay. Results suggested that, under conditions with intracellular ATP concentration above the K_d value of 0.6 mM provided by equilibrium dialysis and in-line probing, LacZ expression was low (Watson and Fedor, 2012). This fits with the hypothesis that this riboswitch acts as an 'off switch' and transcriptional terminator (Block et al., 2010). The reverse was demonstrated where when intracellular ATP concentrations were reduced to 0.4 mM, by the usage of an ATP synthesis inhibitor (CCCP), LacZ activity was significantly increased (Watson and Fedor, 2012).

However, despite the evidence presented in Watson and Fedor, 2012, there is on-going controversy over whether ATP is indeed the primary ligand for the *ydaO*-class. Firstly, other studies have provided evidence that the affinity of the binding between the *ydaO* riboswitches and ATP is too weak to be of biological relevance (Block et al., 2010). It is lower than that of other riboswitches and their ligands, which typically have affinities in the nanomolar or even picomolar range (Smith et al., 2010).

The mutations identified in the dual-selection screen (discussed in the previous section 5.1.3) were indeed mutations that would affect energy metabolism and the production of ATP, which could fit with the theory that ATP could act as a ligand for the riboswitch. However, metabolomics revealed that the concentrations of a wide plethora of different metabolites changed in these mutants, so these data alone are not sufficient evidence to support the theory that ATP is the primary ligand.

Additionally, since the identification of ATP as the putative ligand, others have shown tight binding between the riboswitch and a metabolite present in yeast extract. Mass spectrometry analysis of purified yeast extract samples did not identify the metabolite responsible, but
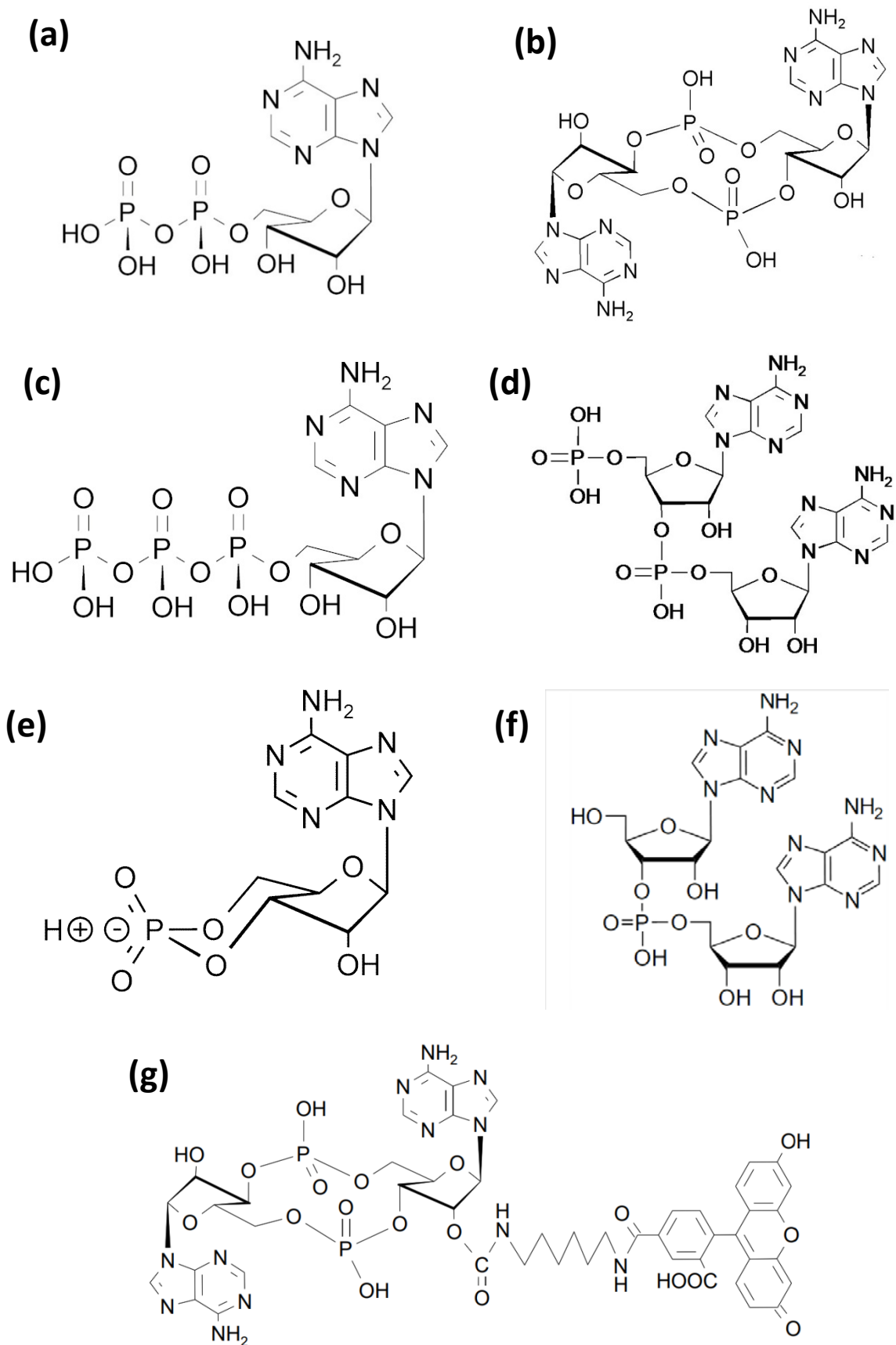
Figure 5.2: shows chemical structures of important ligands. (a) adenosine diphosphate; (b) cyclic di-AMP; (c) adenosine triphosphate; (d) pApA; (e) 3' 5'-cyclic AMP; (f) ApA and (g) 2'-Fluo-AHC-cyclic di-AMP.

indicated that an AMP-moiety was present in the compound. Further experiments have demonstrated binding of cyclic di-AMP, with almost a million-fold higher affinity than ATP (Nelson et al., 2013).

Logically if ATP was the primary ligand, because of its widespread use as an energy 'currency' within the cell and higher intracellular concentration (2.0 ± 0.5 mM (Watson and Fedor, 2012)), it would be expected for the K_D value and binding affinity to be reasonably weak, or otherwise it is likely that the riboswitch would only ever be found in the ligand-bound form and thus always forming the transcription terminator structure, resulting in no gene expression.

5.1.5   Cyclic di-AMP binds the *B. subtilis ydaO* riboswitch

Due to the low affinity of interaction between ATP and the *ydaO* motif, it was suspected that this was not the primary ligand. Nelson et al., developed a strategy to identify the primary ligand that involved starting with complex mixture of chemically diverse molecules and screen for binding interactions by in-line probing experiments and/or equilibrium dialysis. Yeast extract was selected as it is a rich mixture of cellular metabolites and there was believed to be a high chance the natural ligand may be present in the mix. Alterations in the spontaneous cleavage patterns of the RNA motif in the presence of the yeast extract suggested that the ligand of interest, or a closely related molecule, was present within the mix (Nelson et al., 2013).

Further purification and concentration methods were performed on the extract, such as reverse-phase liquid chromatography (LC) and high-performance LC (HPLC). Between each purification step, the ligand-containing fractions were selected by assessing the pattern of spontaneous cleavage of the RNA after incubation with the fractions. After several rounds of purification, the active fraction was analysed by mass spectrometry (MS). MS analysis revealed that the major compound present had a fragmentation pattern and mass that was suggestive of possessing an AMP moiety (Nelson et al., 2013). Both 5'-AMP, ATP and a handful of derivatives had previously been tested in ILP experiments and were ruled out as being the primary ligand (Block et al., 2010). This led to speculation that perhaps AMP could be a breakdown product of the primary ligand and work began to screen other small compounds that contained an AMP moiety within their structure.

Incubation with nanomolar concentrations of cyclic di-AMP, or micromolar concentrations of pApA or ApA (structures shown in Figure 5.2) caused significant differences to appear in the breakdown pattern of the *ydaO* RNAs, which were identical to those patterns produced after incubation with purified yeast extract. The changes that showed reduced cleavage are caused by ligand-induced stabilisation of substructural elements and these primarily occurred in

structural regions that had a higher degree of conservation between *ydaO*-motif family members. Certain areas showed increases in spontaneous cleavage or others showed constant cleavage. These regions were typically located in the non- or less-conserved loop regions within the structure. The binding of cyclic di-AMP has been observed in the *ydaO* representatives from *B. subtilis* (*ydaO* and *yuaA*), *Syntrophus acidictrophicus, Nostoc punctiforme* and *Clostridium acetobutylicum* and the findings support the hypothesis that the *ydaO* riboswitch family respond to cyclic di-AMP as their primary ligand (Nelson et al., 2013).

An estimated K_D value of 700 pM was found during ILP experiments with radiolabelled *ydaO B. subtilis* RNA incubated with RNA, however due to the steep nature of the curve and proximity to the lower concentration of radio-labelled RNA required for detection, it was suggested that the true K_D for the interaction could actually be lower (<100 pM) (Nelson et al., 2013). The K_D value for the interaction with pApA was estimated, by ILP, to be 300 nM. This suggests the riboswitch is able to discriminate between the two structurally similar molecules by over 1,000-fold. ILP was also used to provide a K_D value of a phosphorothioate-modified analog of cyclic di-AMP (cyclic di-AMP$_{SS}$), which was found to be ~7 nM (Nelson et al., 2013). It has been suggested that cyclic di-AMP$_{SS}$ could have a future use as a *ydaO*-type riboswitch inhibitor, due to its tight affinity, decreased degradation by phosphodiesterases and increased ability to enter cells via the membrane (Nelson et al., 2013; Shanahan et al., 2013).

Interestingly, the K_D for the interaction of ATP with *ydaO* from *B. subtilis* is estimated to be ~10 nM (Watson and Fedor, 2012), which is significantly weaker than the K_D of interaction with cyclic di-AMP (Nelson et al., 2013). It has not been possible to detect binding between ATP and the *ydaO* representatives from *Syntrophus acidictrophicus*, *Clostridium acetobutylicum* and *Nostoc punctiforme,* even at concentrations of ATP as high as 1 mM (Nelson et al., 2013)*.* It is not advisable to use ligand concentrations above 1 mM in ILP experiments due to the emergence of non-specific interactions making data interpretation difficult (Regulski and Breaker, 2008).

In addition to this, the ATP-binding ability of *ydaO* from *B. subtilis* could be abolished when Mg^{2+} concentrations were reduced to those that more accurately represented conditions within the bacterial cytosol (10 mM or lower). The affinity of the interaction with cyclic di-AMP was reduced to ~10 nM (Nelson et al., 2013). This finding suggests that the weak interaction with ATP is likely only an artefact of conditions present *in vitro* and is unlikely to be of biological significance or relevance.

Mutational studies revealed that alterations in bases that are important for the formation of the pseudoknot significantly reduced the K_D (~20 nM) of the interaction between the *ydaO* riboswitch and cyclic-di-AMP (Nelson et al., 2013).

Transcription termination experiments were performed using a derivative of the *yuaA* RNA of *B. subtilis*, which showed that concentrations up to 3 mM of ATP did not increase the termination of transcription below background level. Conversely, less than micromolar concentrations of cyclic di-AMP were sufficient for a significant rise in transcription termination (Nelson et al., 2013). This is further evidence that it is likely that cyclic di-AMP is the primary, biological ligand of the *ydaO*-motif. It supports the hypothesis that this particular *yuaA* RNA is functioning as an 'off' switch and exerts control at a transcriptional level (Nelson et al., 2013).

In vivo work using the *ydaO* promoter fused to *lacZ* were carried out in wild-type *B. subtilis*, and also in a *disA* knockout strain. *disA* encodes DisA, one of the three proteins present in the *B. subtilis* genome that is capable of synthesising cyclic di-AMP, and the knockout was used to provide an environment with lower than WT concentrations of cyclic di-AMP. In the knockout strains the activity of *lacZ* was higher than that of the control strain. This supports the theory that there is riboswitch-mediated gene regulation and that cyclic di-AMP levels within a cell can produced an effect on gene expression from the *ydaO* promoter. Mutant riboswitch constructs were also assessed using the reporter system, and perhaps unsurprisingly, it was the mutant with disruption to the predicted terminator stem structure (M6) that resulted in the largest increase in *lacZ* expression (Nelson et al., 2013). Although, the evidence indicates that *ydaO* gene expression is controlled in response to cyclic di-AMP via a terminator stem, the exacting mechanism of how ligand binding causes the formation of the terminator is unknown. However, in-line probing of the *B. subtilis ydaO* riboswitch suggests that the majority of the RNA structure is pre-folded before the riboswitch senses its ligand (Block et al., 2010).

5.1.6 The *rpfA* gene of several mycobacterial species contain a putative *ydaO*-type riboswitch

Members of the *ydaO*-type family of riboswitches are present in a wide range of bacteria and are associated with a wide variety of genes. Of particular interest to this study is the putative *ydaO*-type riboswitch that is associated with the *rpfA* gene of *Mycobacterium tuberculosis*. As mentioned in more detail in Introduction Section 1.7, the *rpfA* gene encodes resuscitation promoting factor A (RpfA). RpfA, and other members of the Rpf family (Rpfs A to E) play an important role in the resuscitation of cells from the dormant, non-replicating state to growth during an active infection (Kana and Mizrahi, 2010; Kana et al., 2008). Cyclic di-AMP had not
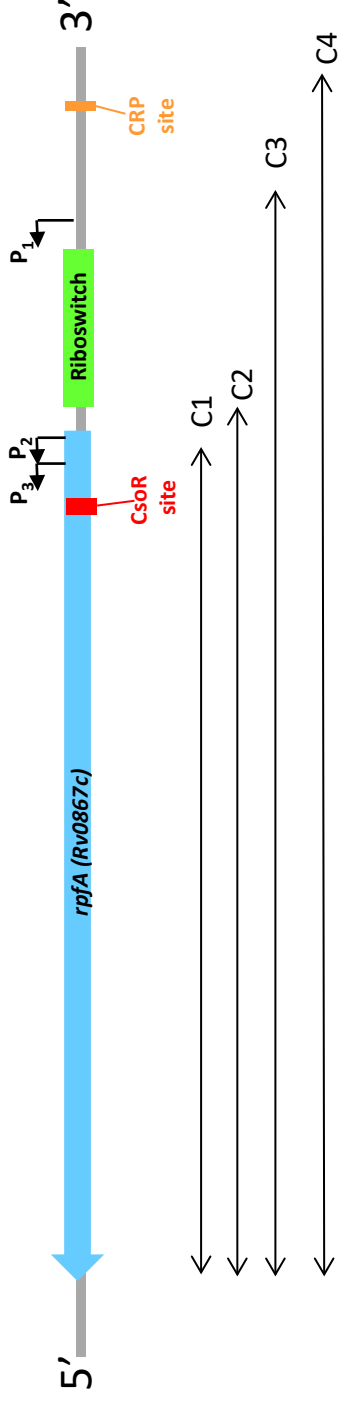
Figure 5.3: Schematic representation of *rpfA*, *ydaO* and the promoter regions (red boxes) upstream and internal to *rpfA*. Construct C1 amplified the region from the intragenic promoter to the end of the gene (965486-964315), C2 amplified from the promoter region before the transcriptional start site (965555-964315), construct C3 amplified from the promoter before the riboswitch (965832-964315) and construct C4 amplified the entire intragenic region (965985-964315). Constructs C1 to C4 were transformed in to the Δ*rpfA* and the phenotype was assessed to determine if the riboswitch has an essential regulatory role

previously been implicated in playing an important signalling role during the resuscitation process. Due to our interest in the regulation of *rpfA* and recent identification of a putative *ydaO*-type riboswitch being present in the 5' UTR of the *rpfA* mRNA, it was of interest to us to investigate and gain more information.

The central aim of this project was to confirm if the putative riboswitch from *M. tuberculosis* played a role in the regulation of *rpfA* and if so, to determine the signal which the riboswitch responds to. The project combined expertise from two different research groups. The majority of the mycobacterial *in vivo* work was carried out by our collaborators under the supervision of Dr Galina Muklamova at The University of Leicester and is outlined in section 5.1.7. The *in vitro* binding studies will form the basis of this chapter and were carried out at The University of Sheffield.

5.1.7 *In vivo* evidence that the *M. tuberculosis rpfA* riboswitch is functional (unpublished)

Our collaborators used an *rpfA* deletion mutant in *M. marinum,* which was created by Dr Nino Iakobachvili. After osmotic stress (2 M NaCl), the Δ*rpfA* strain had consistently lower colony forming units (cfus) and slower growth on agar plates (over a week for colonies to appear) compared to the wild-type stain. This phenotype was exploited to study the importance of the different regulatory portions of the *rpfA* 5' UTR in gene expression. In particular, it has been useful in confirming a regulatory role of the predicted riboswitch by complementing the Δ*rpfA* strain with different *rpfA* constructs (shown in Figure 5.3). These include constructs with the riboswitch present and with the riboswitch deleted.

Ideally, these experiments would have been performed in a *Mtb* Δ*rpfA* mutant, but at the time of writing this thesis, a *M. tuberculosis* Δ*rpfA* with the complete *rpfA* region deleted (including riboswitch) that was able to be fully complemented was not available. For this reason, the experiments were carried out using the *M. marinum* Δ*rpfA* mutant. This strain is closely related to *M. tuberculosis* and also possessed a riboswitch within its *rpfA* gene.

The Δ*rpfA* strain had increased sensitivity to osmotic stress (2M NaCl) and upon regrowth on agar plates the morphology of colonies altered to smaller colonies. Several constructs were made containing different sections the *M. tuberculosis rpfA* promoter regions which were then used to complement the *M. marinum* mutant (data not shown). The long construct C4 (containing the CRP binding site through to the end of the *rpfA* coding region) complemented the osmotic sensitivity of the *M. marinum* Δ*rpfA*. C3 lacked the CRP-binding site region and complemented the osmotic sensitivity of the *M. marinum rpfA* mutant. Constructs C2 and C1,

both lacking the riboswitch, did not complement. This suggests the riboswitch could be important and functional and may play a role beyond repression.

5.1.8 Aims of this chapter

As mentioned in section 5.1.6, the primary aim of our project was to ascertain if the *ydaO*-type riboswitch present in the *rpfA* gene of *Mycobacterium tuberculosis* was indeed functional (covered in section 5.1.7) and if so, attempt to determine the natural ligand of the riboswitch. This would initially centre around determining if cyclic di-AMP (suggested in Nelson *et al.*, 2013) or ATP (suggested in Watson *et al.*, 2012) could bind to the riboswitch and with what affinity.

Initially, it was decided to perform in-line probing (ILP) experiments to achieve this aim, but depending on the results, other techniques, such as microscale thermophoresis or gel shift asays, would be employed. Upon determining the ligand-binding capabilities of the *Mtb rpfA* riboswitch, attempts would be made to determine binding parameters, such as K_D, of the interaction between the riboswitch and the relevant ligand(s).

In addition to this, bioinformatics approaches would be utilised in order to identify riboswitches present within the *rpfA* gene of other mycobacterial species. Of particular interest were *Mycobacterium smegmatis* (a soil-dwelling, 'fast'-growing, model mycobacteria); *Mycobacterium marinum* (a fish pathogen related to *M. tuberculosis*) and *Mycobacterium bovis* (closely related to *M. tuberculosis* and the causative agent of TB in badgers). Once identified, the sequences would be compared to that of the *M. tuberculosis* riboswitch and structural predictions would be made. It was an aim to characterise the ligand-binding capabilities of these riboswitches for different ligands, which would be selected based upon our findings with the *M. tuberculosis* riboswitch.

The final aims of the project included performing mutational studies with the *M. tuberculosis rpfA* riboswitch. A selection of mutations would be designed with the aim of disrupting important substructural elements of the riboswitch. The ligand-binding ability of these riboswitches would subsequently be assessed by methods developed for the wild-type riboswitch. Mutations that caused significant reduction in, or abolished, ligand-binding would be screened *in vivo* using the reporter system developed for the wild-type riboswitch.

Overall, this project aims to provide confirm the functionality and provide ligand binding data to give insight into several previously uncharacterised members of the *ydaO*-type riboswitch. It could also provide useful insights in the debate over what the natural ligand for the family actually is.

(a)

Name	Sequence (5' to 3')	F/R	*rpfA* riboswitch in organism
LD20	TAATACGACTCACTATAGGGGCCGATAGCACGCCGAACTC	F	*M. tuberculosis*
LD21	AATTCCTCTCGGTACACGCC	R	*M. tuberculosis* and *M. marinum*
LD24	TAATACGACTCACTATAGGGCGCCGAGCTCCATCCGGTAG	F	*M. marinum*
LD22	TAATACGACTCACTATAGGGGCCGGTAGCGCCCGAGCTCCA	F	*M. smegmatis*
LD23	TCCTCTCGTATGCGCCCACG	R	*M. smegmatis*
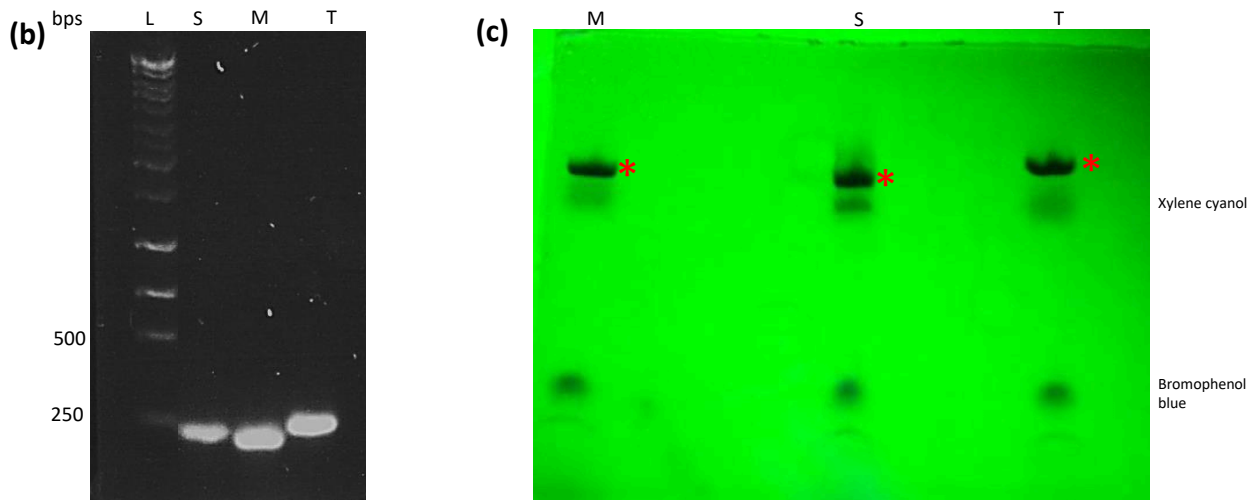
Figure 5.4: shows stages required to synthesise mycobacterial riboswitch RNA.
(a) Primers designed to generate DNA templates for subsequent RNA synthesis experiments. F/R category indicates if the primer is a forward (F) or reverse (R) primer. All sequences are in 5' to 3' orientation. (b) Agarose gel electrophoresis analysis of PCR products of mycobacterial *rpfA* riboswitch DNA templates. Lane L corresponds to 5 µl Hyperladder 1kb Plus . Lanes S, M and T correspond to samples of the PCR products for genomic DNA from *M. smegmatis*, *M. marinum* and *M. tuberculosis* respectively. A total of 10 µl was loaded per well (8 µl DNA + 2 µl loading dye (5X)). (c) Denaturing PAGE analysis of products from RNA synthesis. Products from *in vitro* RNA transcription reactions using *M. marinum* (M); *M. smegmatis* (S) or *M. tuberculosis* (T) *rpfA* riboswitch template DNA were analysed by denaturing PAGE.

In summary, we aim to:

- Ascertain if the putative *rpfA* riboswitch plays a functional role in the regulation of *rpfA* in *Mycobacterium tuberculosis*.

- Determine the primary ligand of this riboswitch and generate convincing evidence of binding.

- Generate and characterise the ability of mutant riboswitches to bind the ligand(s). Test any "non-binder" mutants *in vivo*.

- Identify riboswitches present in other mycobacterial species using bioinformatics and characterise ligand-binding abilities of these.

5.2 Generation of mycobacterial riboswitch RNA

In order to carry out experiments to explore the putative ligand(s) of the mycobacterial riboswitches, it was necessary for full length riboswitch RNA to be generated and purified.

5.2.1 Generation of DNA templates to be utilised in RNA synthesis reactions

Initially, it was necessary to produce DNA templates suitable for subsequent use in *in vitro* transcription (IVT) reactions to produce full RNA species. These template DNAs contained the T7 RNA polymerase recognition site (TAATACGACTCACTATA), followed by three G nucleotides, followed by the full length riboswitch sequence to be transcribed by RNA polymerase. The three G nucleotides promoted optimal transcription yields of small transcripts (Regulski and Breaker, 2008).

Forward primers were designed that incorporated the T7 RNAP recognition site, triple G and 20 bp from the beginning of the riboswitch sequence (5'-TAATACGACTCACTATA-GGG-20 bp of riboswitch-3'). Reverse primers were designed that contained 20 bp of the 3' end of the riboswitch. To amplify the DNA templates required for IVT reactions to produce full length riboswitch RNA, PCR was carried out using the relevant primers (listed in Figure 5.4a) and methods detailed in Section 2.3.4. Genomic DNA, extracted by methods in Section 2.3.7, from the relevant organisms was used as template DNA. After thermal cycling was complete, products were analysed by agarose gel electrophoresis (Method Section 2.3.9) to confirm the presence of a species of the correct size (Figure 5.4b).

The length of the putative riboswitches in the different mycobacterial species were as follows: 224 bp for *M. smegmatis*, 223 bp for *M. marinum* and 243 bp for *M. tuberculosis*. The DNA templates were expected to be 20 bp longer than this, due to the incorporation of the 5' end region containing the T7 recognition site. In lane S, the presence of a species at the expected size of 244 bp indicated the successful amplification of the *M. smegmatis* riboswitch DNA from

Full length *Mtb rpfA* riboswitch RNA

Xylene cyanol

Bromophenol blue
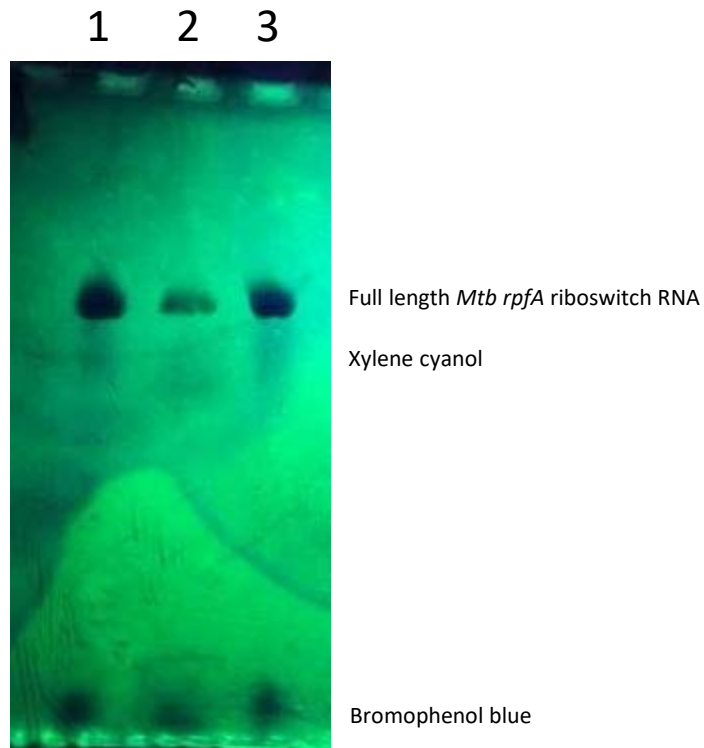
Figure 5.5: Denaturing PAGE analysis of *M. tuberculosis rpfA* riboswitch products from RNA synthesis after different methods of purification. Lane 1 contains a sample of RNA taken directly from the IVT reaction mix with no purification. Lane 2 contains equivalent "amount" of RNA that has been purified by denaturing PAGE, followed by elution in crush/soak buffer and ethanol precipitation. Lane 3 contains the equivalent amount of RNA purified by MegaClear™ column (Ambion). The location of the full length RNA and loading dyes are labelled.

genomic DNA using primers LD22 and LD23. The presence of a ~243 bp product in lane M indicated the successful amplification of the *M. marinum* riboswitch DNA from genomic DNA using primers LD24 and LD21. The presence of a ~263 bp product in lane T indicated the successful amplification of the *M. smegmatis* riboswitch DNA from genomic DNA using primers LD20 and LD21.

DNA was purified (Method Section 2.3.5) and eluted in nuclease free water. The concentration was determined spectrophotometrically (Method Section 2.3.12). If the concentration was not high enough for use in the relevant *in vitro* transcription kit, DNA was concentrated.

5.2.2   Synthesis of mycobacterial riboswitch RNA

RNA synthesis (or IVT) reactions were carried out by the method listed in Section 2.3.13. Products were analysed using denaturing polyacrylamide gel electrophoresis (PAGE) (Method Section 2.3.14). Visualisation of RNA species present in the gel was carried out on thin layer chromatography (TLC) plates. Figure 5.4c shows denaturing PAGE analysis of IVT reaction. The full length products (indicated with *) for all mycobacterial species were successfully synthesised and can be seen as a dark shadow on the TLC plates upon exposure to UV light (254 nm). The locations of bromophenol blue and xylene cyanol in the loading dye are indicated. There is no smearing present in the lanes, which indicates there has been no degradation of the RNA after synthesis.

5.2.3   Purification of riboswitch RNA

In the initial experiments generating full length riboswitch RNA, purification from the IVT reactions was achieved by a PAGE-based method.

5.2.3.1   Successful purification of riboswitch RNA by PAGE

Initially the DNA template was degraded in the samples by treatment with DNase I. The RNA product was then purified by denaturing PAGE (Method Section 2.3.14). The gel was exposed to UV light (254 nm) on a TLC plate and the band of interest was excised with a razor blade, removing as little surrounding gel as possible. Gel cubes were covered by crush/soak buffer and incubated at room temperature for at least one hour. The buffer was removed from the gel and RNA was precipitated (Method Section 2.3.17) and after washing, the pellet was resuspended in nuclease free water.

However, the yield of purified RNA produced by PAGE was significantly lower (Figure 5.5, lane 2) when compared to an equivalent unpurified RNA from the IVT reactions (Figure 5.5, lane 1). There were other caveats of the PAGE method of purification, which are listed below:
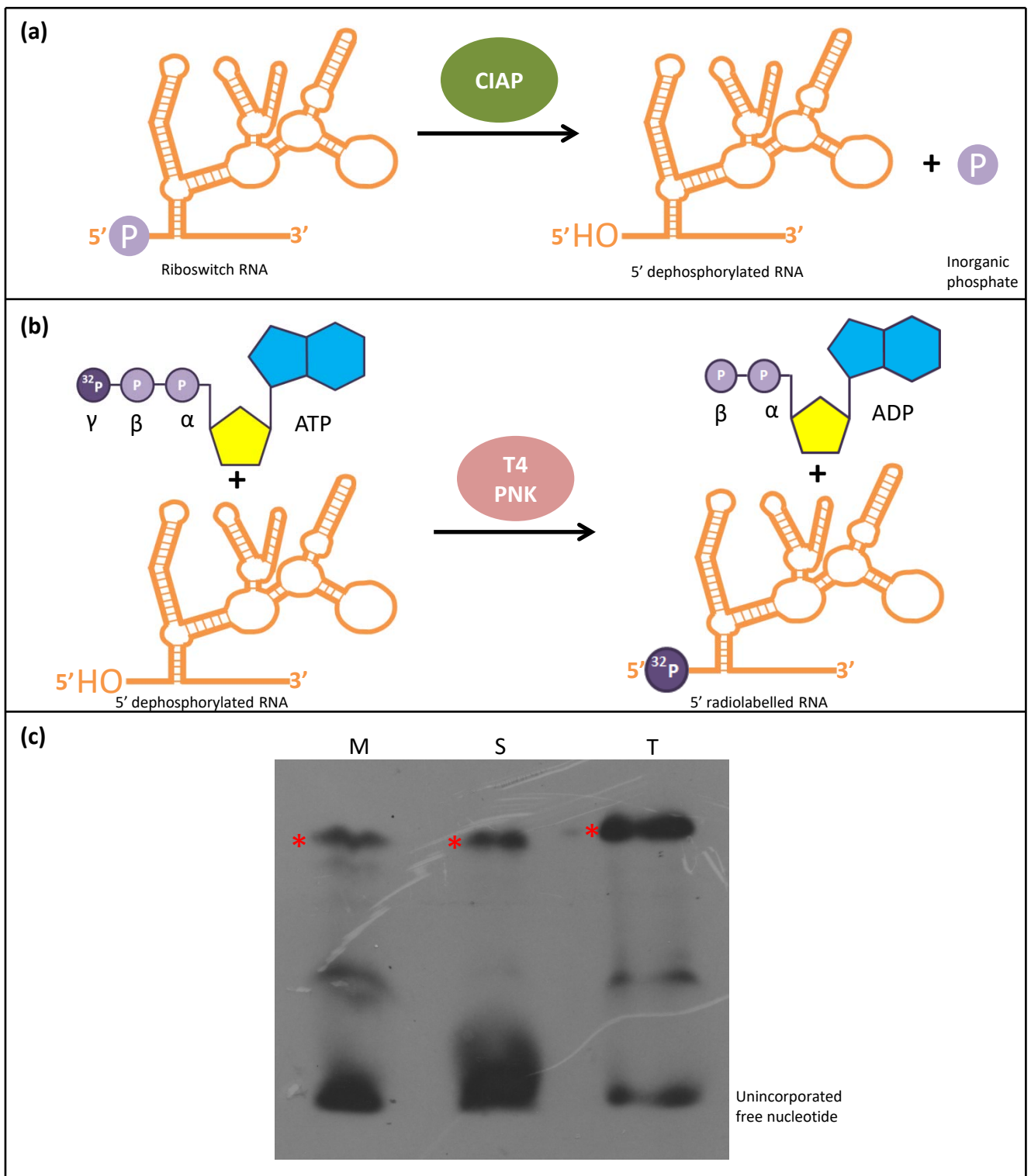
Figure 5.6: outline of the process 5' end labelling of riboswitch RNA. (a) shows the dephosphorylation of riboswitch RNA by CIAP. This reaction involves the removal of the 5' phosphate group by CIAP, resulting in a 5' hydroxyl group on the RNA molecule and an inorganic phosphate molecule . (b) shows the phosphorylation of desphophorylated RNA by T4 PNK. The phosphate groups of the γ^{32}P-[ATP] substrate are labelled in order of their distance from the adenosine (α, β and γ). The T4 PNK is catalyses the transfer of the γ^{32}P from the ATP, to the 5' hydroxyl group of the RNA. The end products are ADP and a 5' radiolabelled RNA. (c) Denaturing PAGE analysis of products from end-labelling reactions. Lanes contain *rpfA* riboswitch RNA end-labelling reactions from *M. marinum* (M), *M. smegmatis* (S) and *M. tuberculosis* (T). The successfully 5' 32P end-labelled species of full length is indicated with *. A band corresponding to unincorporated [γ-^{32}P] ATP that has not been used during the phosphorylation reaction is indicated at the bottom of the gel.

a) There are more steps where RNase contamination could potentially occur.

b) There is requirement for careful RNAse-free preparation of all the gel constituents and running buffer.

c) All gel tank and apparatus must be treated with RNase ZAP or a 10% SDS, followed by ethanol wash to avoid degradation of the samples.

d) The method is more time consuming than the subsequent spin column purification method (Section 2.3.18).

Although the PAGE purification successfully purified RNA and was the sole method used at the beginning of the project, it was replaced later in the project by the more efficient spin column purification (Section 5.2.3.2).

5.2.3.2 Successful purification of riboswitch RNA by spin column

An alternative method for the purification of riboswitch RNA from synthesis reactions used a spin column. Firstly, the DNA template in the samples was degraded by treatment with DNase I. Next, RNA was purified by using a MEGAclear™ spin column by the methods outlined in Section 2.3.18. A good yield of RNA was successfully purified using the spin column method (Figure 5.5, lane 3) when compared to an equivalent amount of unpurified RNA from the IVT reactions (Figure 5.5, lane 1). This method provided a better yield of purified RNA than the previously used PAGE purification method (Figure 5.5, lane 2) and all subsequent RNA purification steps were performed this way.

5.2.4 ^{32}P end-labelling of riboswitch RNA

In order to perform some further experiments, including in-line probing experiments and 'gel shift' assays, it was necessary to produced 5' ^{32}P end-labelled riboswitch RNA. To begin with the 5' phosphate group of the purified RNA needed to be removed (Figure 5.6a). This was achieved by treatment with calf intestinal alkaline phosphatase (CIAP) by methods detailed in Section 2.3.19. After the dephosphorylation reaction had occurred, it was necessary to remove the CIAP from the samples, as carryover would inhibit subsequent end-labelling reactions by T4 polynucleotide kinase (T4 PNK). There was limited success with subsequent kinase reactions when the phosphatase removal reagent supplied in the KinaseMaxTM kit (data not shown) was utilised for this purpose. Instead, successful removal of the CIAP was achieved by phenol/chloroform extraction, detailed in methods Section 2.3.16.

Glycogen (100 µg/ml) was added to the samples, which acted as a carrier molecule during the precipitation of RNA species by ethanol precipitation (Section 2.3.17). RNA was resuspended in nuclease free water at a relevant concentration for kinase reactions to be carried out. T4 PNK

is an enzyme that is capable of catalysing the transfer of a phosphate group from the γ position of an ATP molecule to the 5'-hydroxyl terminus of a dephosphorylated nucleic acid molecule (Figure 5.6b). Kinase end-labelling reactions were set up as outlined in Section 2.3.19 using [γ-^{32}P] ATP as a substrate. The 5' ^{32}P end-labelled RNA was purified by denaturing PAGE on gels containing 8% acrylamide (Methods Section 2.3.14). The gel was placed on acetate with Stratagene markers attached, to allow for correct orientation and alignment, and exposed to x-ray film and developed. A typical film is shown in Figure 5.6c. Successfully labelled, full length *rpfA* riboswitch RNA migrated slowest and is visible as a dark defined band at the top of the gel (marked with *) for all desired mycobacterial species. Free, unincorporated [γ-^{32}P] ATP migrated faster and its location is market near the lower portion of the gel.

There was a small proportion of species present in the gel, which migrated through the gel at a faster speed than the full length riboswitch RNA. These are likely to correspond to smaller fragments of the riboswitch, which has partially broken down. RNA, particularly single stranded RNA, is inherently unstable, in comparison to DNA. The increased lability of RNA is attributed to several factors, including the presence of 2' hydroxyl (-OH) group, the presence of uracil bases in RNA instead of the thymine bases present in DNA, and other factors (Soukup et al., 2001).

This property of instability is troublesome during the processing and purification of specific RNA species, such as in these experiments. However, the instability can also be beneficial. In Section 5.3.1, details on how this inherent instability can be exploited experimentally, to help elucidate information on ligand binding via in-line probing experiments, will be discussed. In many organisms RNA plays an important role as a messenger i.e. a short-lived intermediate that is able to communicate information contained in genes to the rest of the cell. The DNA in genes is stable. Often, there is a requirement for genes to be expressed in short bursts and thus the less stable property of RNA is extremely useful in the degradation of mRNAs and other functions.

Due to the presence of these labelled, smaller species, it was not possible to purify the labelled riboswitch RNA by the spin column purification method detailed in Section 5.2.3.2. Using the Stratagene markers, the gel was aligned with the x-ray film and the band of interest (corresponding to the full length riboswitch RNA) was excised with a razor blade, whilst removing as little of the surrounding gel as possible. Gel cubes were covered by crush/soak buffer and incubated at room temperature for at least one hour. The buffer was removed from the gel and RNA was precipitated and, after washing, the pellet was resuspended in nuclease
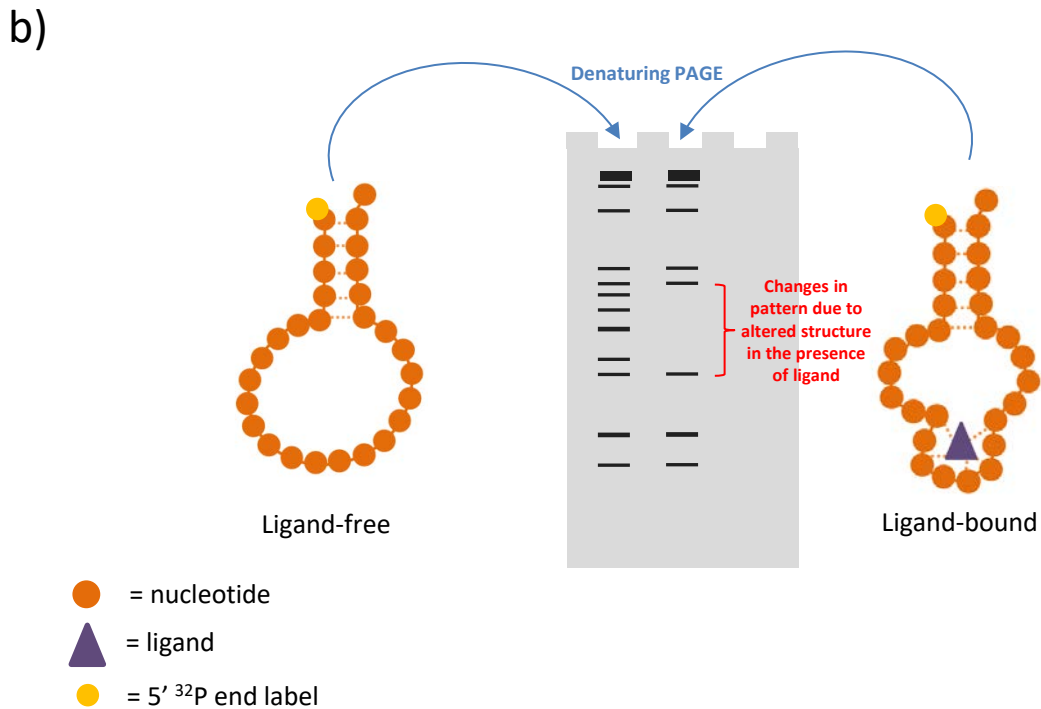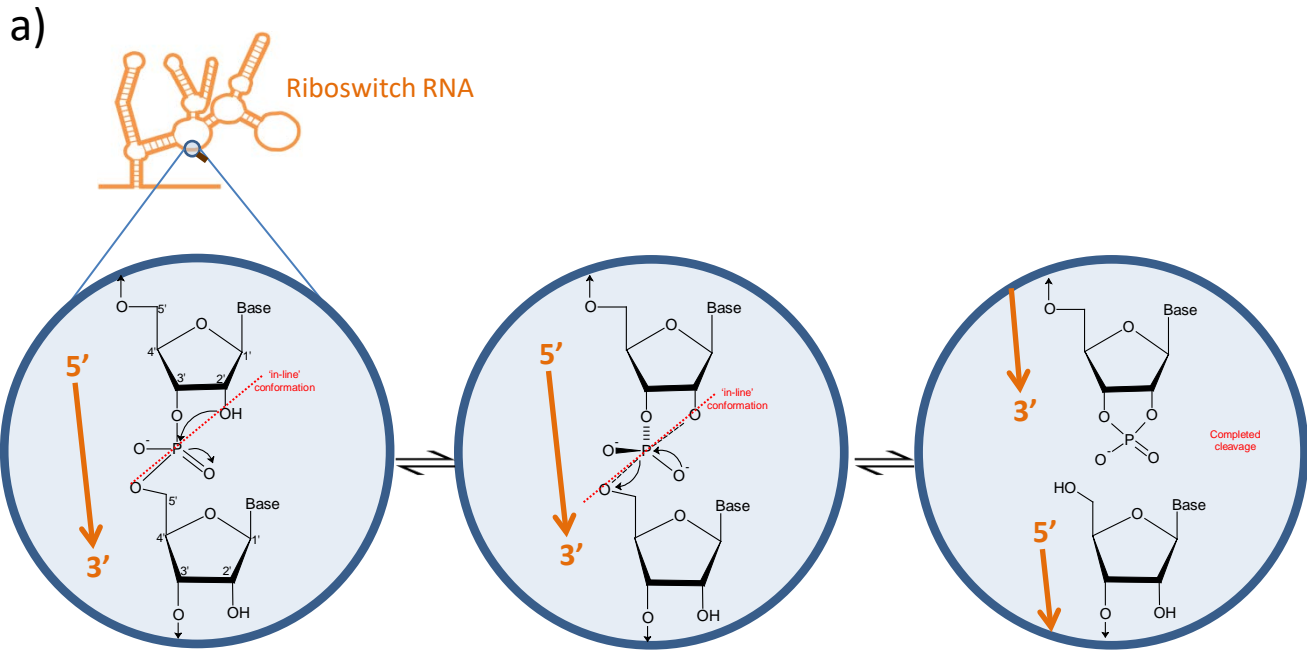
Figure 5.7: The principles behind in-line probing experiments.

a) Shows the mechanism of 'in-line' RNA backbone cleavage between two nucleotides. The direction of the RNA is indicated in orange (5'→3'). Cleavage can occur when the 2' oxygen, the phosphorus centre, and the 5' oxygen atom from the adjacent nucleotide enter a linear conformation (known as 'in-line'). Once in the 'in-line' conformation (shown by the red dotted line), the 2' oxygen executes an efficient nucleophilic attack on the adjacent phosphorus centre, causing intramolecular displacement of the 5' oxygen and internucleotide cleavage. b) Schematic of an in-line probing experiment. Example structures of a 5' ^{32}P end labelled RNA in ligand free and ligand bound form are shown in orange. Riboswitches are incubated in the presence and absence of ligand to allow for partial break down. Next they are analysed by denaturing PAGE. In the ligand-bound form, certain internucleotide linkages are held in position and are less flexible. This leads to a decrease in linkages entering the in-line conformation, hence less cleavage. When the gel is exposed to x-ray film and developed the differences in the pattern of break down in the two different samples can be detected (shown in red bracket). If a ligand doesn't interact, there should be no alteration in structure and hence no differences in the species visible on the x-ray film.

free water. Radiolabelled riboswitch RNA was stored at -20°C and used quickly for in-line probing experiments (Section 2.3.20).

5.3 Optimisation and use of in-line probing as a method to elucidate cognate ligands of riboswitches

5.3.1 In-line probing (ILP) as a method for elucidating ligands of riboswitches

ILP is a technique that which can be useful in the analysis of riboswitches or other RNA molecules, particularly to gain information about their secondary structure and their ability to bind small metabolites. It exploits the innate instability of RNA molecules and their tendency to spontaneously and non-enzymatically break down. Spontaneous and efficient cleavage of RNA between two nucleotides can occur by 'in-line' nucleophilic attack of the 2' oxygen atom on the nearby phosphorus centre (Soukup and Breaker, 1999). This mechanism is shown in detail in Figure 5.7a.

The likelihood and speed of cleavage of an internucleotide phosphodiester bond occurring is highly dependent on the local structure surrounding it (Li and Breaker, 1999; Soukup and Breaker, 1999; Soukup et al., 2001). Often, phosphodiester linkages present in highly structured regions, such as hairpins or base-paired helixes, are less susceptible to cleavage in this manner due to the 2' oxygen atom, phosphorus and adjacent 5' oxygen atom not being held 'in-line' in the structure. In addition, in these ordered structures, the atoms are less likely to enter this conformation (Regulski and Breaker, 2008).

On the other hand, free and unstructured regions of RNA are not restricted in such manner. Thus are more likely to randomly enter the 'in-line' conformation, allowing for nucleophilic attack and higher rates of cleavage of the linkage (Mandal and Breaker, 2004). It is also worth noting that in certain cases, internucleotide linkages within specific tertiary or secondary structures can have increased rates of non-enzymatic cleavage. This occurs when the structure itself holds the atoms in the relevant configuration for the nucleophilic attack to occur (Soukup and Breaker, 1999).

The core principle of in-line probing experiments is described in Figure 5.7b. Structural changes occur within riboswitches when a metabolite binds to the aptamer region. The rate of cleavage of nucleotide linkages is highly dependent on the surrounding structure of an RNA. If 5' ^{32}P end-labelled riboswitch RNA is refolded and incubated and absence of candidate ligands, it would be possible to analyse the products of break down by denaturing PAGE and determine if there are any alterations in the bands present with and without the ligand. If a ligand binds under those conditions, alterations in the species produced are expected.
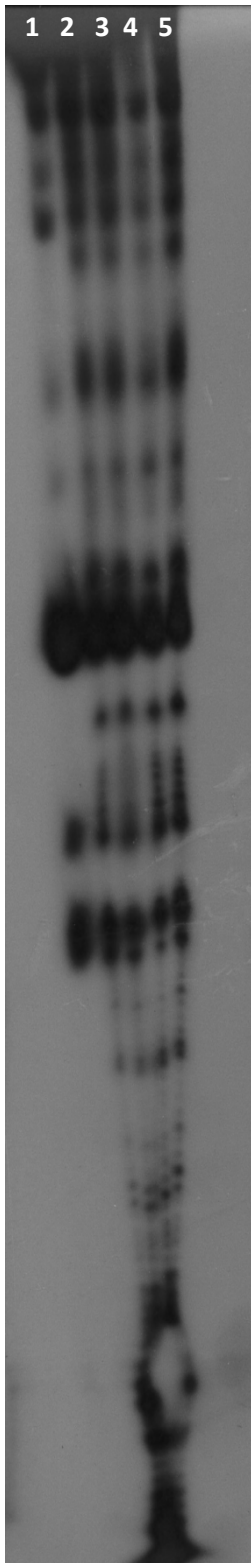
Figure 5.8: shows denaturing PAGE analysis of in-line probing reaction products after 4 day incubation with 1 mM ligand. All lanes contain *M. tuberculosis rpfA* riboswitch RNA (~1 pmol). Lane 1 contained labelled RNA that had not undergone incubation at room temperature in 1X ILP buffer. Samples in lanes 2 to 5 were incubated at room temperature for a period of 4 days. Lane 2 contained no ligand. Lanes 3, 4 and 5 contained 1 mM ATP, c-di-AMP and cAMP respectively.

With this knowledge, coupled with the successful 5' ^{32}P end labelling of the mycobacterial *rpfA* riboswitches (shown in Section 5.2.4), it was decided to utilise this approach to screen for putative ligands in yeast extract. Yeast extract provides a large pool of cellular metabolites, which could act as ligands. If alterations in structure were observed in the presence of yeast extract, further purification steps could be carried out to enrich for the relevant metabolite. It was also decided to assess if the *rpfA* riboswitches could bind ATP or cyclic di-AMP, which have both been demonstrated as able to bind to the closely related *ydaO* riboswitch from *B. subtilis* (Nelson et al., 2013; Watson and Fedor, 2012).

5.3.2 Optimisation of in-line probing methodology

Initially, ILP reactions were performed using the buffers and methods in Regulski and Breaker in an RNase-free environment and with RNase-free reagents. Ligands were used at a maximum final concentration of 1 mM. This concentration was deemed to be sufficient to ascertain if there are any structural changes in the RNA in response to the ligand, but not so high to result in non-specific interactions.

Although the initial method tested was based upon an existing published method by the Breaker group (Regulski and Breaker, 2008), there were numerous issues that arose during the experiment, which required troubleshooting. Overviews of the problems encountered are detailed in the following sections.

5.3.2.1 Pinning during electrophoresis

Initially, ILP samples were set up according to the methods in Regulski and Breaker (2008). Reactions contained radiolabelled *M. tuberculosis rpfA* riboswitch RNA (~1 pmol) and, if relevant, a ligand at a final concentration of 1 mM. The reactions were prepared in 1X ILP buffer that contained a final concentration of 50 mM Tris-HCl (pH 8.3), 40 mM MgCl$_2$ and 100 mM KCl. Samples were analysed on a 10% acrylamide (19:1) denaturing PAGE gel (40 cm x 0.75 mm) as outlined in. A 1x TBE buffering system was used during electrophoresis at 45 W. The gel was electophoresed until the bromophenol blue dye had migrated nearly the full length of the gel. The gel was fixed, transferred to blotting paper and dried, and was then exposed to x-ray film for the required length of time to produce a sufficient exposure and was developed (Figure 5.8).

Pinning is an artefact of electrophoresis, which can be seen in Figure 5.8 where the samples in lanes 1 to 5 have migrated towards each other in a pinpoint during the process of electrophoresis. Pinning can be caused by the concentration of salt(s) in the samples being too

(a) 1 2 3 4 5 **(b)** 1 2 3 4
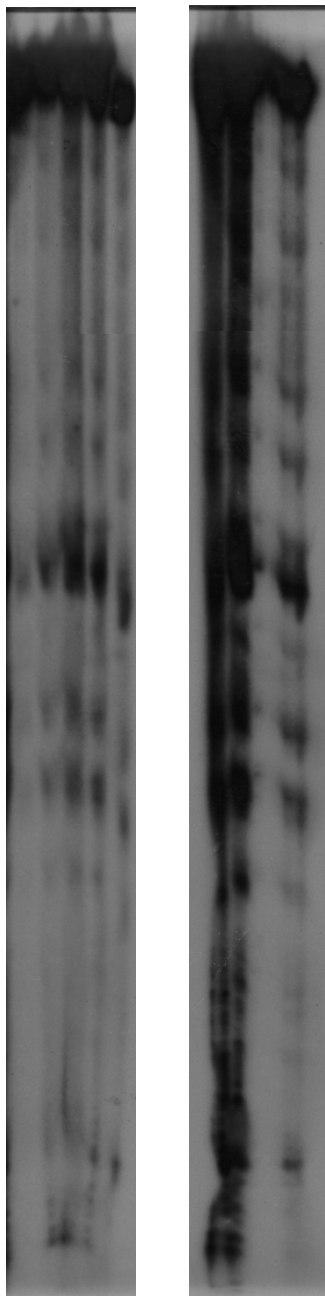
Figure 5.9: shows the comparison of denaturing PAGE analysis of *M. tuberculosis rpfA* riboswitch in-line probing products after ethanol precipitation or incubation in lower salt ILP buffer. (a) shows ILP reactions containing RNA (~2 pmol) with relevant ligands (1 mM) in 1X 'normal' ILP buffer, that were subjected to ethanol precipitation and resuspended in water prior to loading. (b) shows ILP reactions containing RNA (~1 pmol) with relevant ligands (1 mM) that were set up in 1X low salt ILP buffer. Samples in (a) and (b) were incubated for 4 days at room temperature.

In (a), lane 1 contained riboswitch RNA that had been stored at -20°C and had not undergone incubation at room temperature. Lane 2 contained incubated RNA with no ligand. Lanes 3, 4 and 5 contained RNA incubated with 1 mM ATP, c-di-AMP or cyclic AMP respectively. In (b), lane 1 contained incubated RNA with no ligand. Lanes 3, 4 and 5 contained RNA incubated with 1 mM ATP, c-di-AMP or cyclic AMP respectively.

high during electrophoresis, as high salt can affect the conductivity of samples and result in atypical migration.

In order to rectify the issue of pinning, two approaches were trialled to reduce the salt content of reactions after the incubation step, but prior to the loading stage. The first was to further purify reactions using Dyex 2.0 (Qiagen) columns. The other was to perform ethanol precipitation on the samples after incubation and resuspend the pellets in nuclease-free water. Both methods were trialled and the results for ethanol precipitation are shown in Figure 5.9a. The DyEx 2.0 column resulted in a large loss of RNA (data not shown). The samples purified by ethanol precipitation (Figure 5.9a, lanes 1 to 5) showed some reduction in pinning during electrophoresis (compare with Figure 5.8, lanes 1 to 5). However, ethanol precipitation appeared to give an inefficient recovery of RNA, particularly of the small sized species. Despite a higher concentration of RNA (~2 pmol) originally being present in the reactions from Figure 5.9a, when compared to lanes in Figure 5.8, in which lanes contained ~1 pmol, there is far less signal on the film after equivalent exposure times. This indicates that there is a significant loss of RNA during the purification process, which is particularly noticeable in the smaller sized species. It was concluded that this method of reducing pinning was not suitable for further experiments.

As both of the methods attempted resulted in loss of RNA, it was decided to try to reduce the salt concentration within the 1X ILP buffer in the hope of reducing the pinning, but maintaining sufficient levels of RNA in samples. However, it would be important to not reduce it so much, especially the Mg^{2+}, that it may affect the refolding and stability of the RNA. A new 'Low salt' ILP 1X was developed, which contained of 50 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$ and 10 mM KCl. Samples were prepared as before, but with the altered low salt ILP buffer and were analysed by denaturing PAGE on a sequencing gel, see Figure 5.9b. The samples that were incubated in low salt ILP buffer (Figure 5.9b, lanes 1 to 4) had reduced pinning compared to those samples incubated in normal ILP buffer (Figure 5.8). For subsequent ILP experiments the low salt ILP buffer was utilised.

5.3.2.2 Smearing during resolution

Initially 1X TBE buffer with 10% polyacrylamide was for the separation of RNA products during ILP experiments. However, smearing occurred during electrophoresis and many species were not clearly distinguishable from one another. In order to solve this issue several different buffering systems and acrylamide percentages were trialled. After testing several combinations, 2X TBE with 10% acrylamide was shown to improve the resolution of species with the gel matrix the most (data not shown).
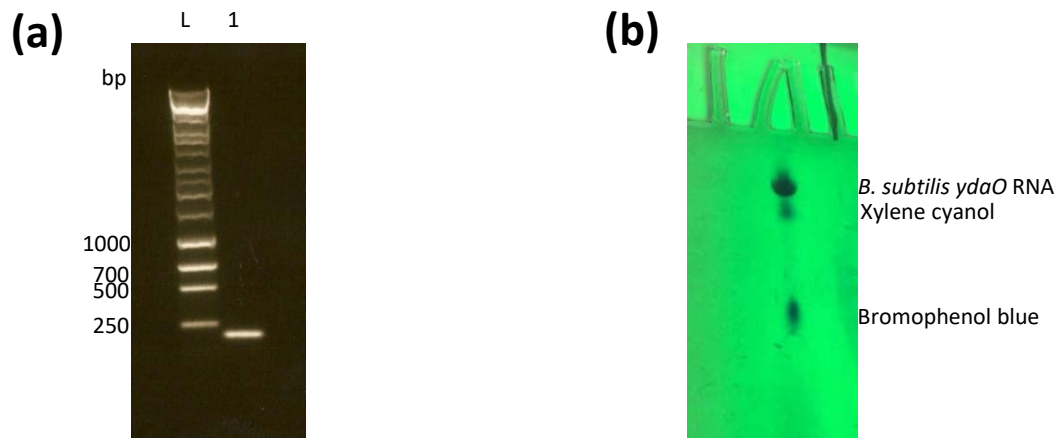
Figure 5.10: Generation of *B. subtilis ydaO* riboswitch RNA. (a) Agarose gel electrophoresis of *ydaO* template DNA. Lane 1 contained 5 µl of Hyperladder 1 kb Plus (Bioline) and lane 2 contained 5 µl of PCR product. (b) Denaturing PAGE analysis of products of transcription from *ydaO* DNA template.

5.3.2.3 Use of higher activity RNA

At the early stages of the project the end-labelling reactions of 5' dephosphorylated RNAs were performed using [γ-^{32}P] ATP with a specific activity of 3000 Ci/mmol. This was sufficient to allow low concentrations of total RNA (~1 pmol) to be detected during ILP. However, the exposure time for such gels was typically in the magnitude of several days. In order to increase productivity by quickening the exposure time, isotope with an increased specific activity 6000 Ci/mmol was utilised.

5.3.2.4 Time for incubation

Different incubation periods were tested to ascertain the optimal time for sufficient breakdown of the RNA to occur, without the majority of the RNA species being completely degraded. We observed that at seven day incubation the majority of RNA species were completely degraded (data not shown). Incubation periods of 14 h to 96 h were optimal, with shorter incubation periods not providing a sufficient level of spontaneous cleavage of the RNA (data not shown). In subsequent experiments, incubation for 14 h of *M. tuberculosis rfpA* riboswitch RNA was used.

5.3.2.5 Use of *B. subtilis ydaO* as a control

The initial ILP experiments with *M. tuberculosis rpfA* riboswitch RNA did not clearly demonstrate if alterations in spontaneous breakdown products occurred with the range of ligands tested. It was decided to perform the optimised ILP methodology we had developed with the *B. subtilis ydaO* riboswitch as a control. The theory behind this was that according to Nelson et al. (2013) we should expect to observe alterations in the spontaneous breakdown pattern of this particular RNA in the presence of c-di-AMP and yeast extract, providing a positive control.

The *B. subtilis ydaO* riboswitch RNA was synthesised by the same methods used for the mycobacterial *rpfA* riboswitch RNA, which are outlined in sections 5.2.1 to 5.2.4. Initially, a DNA template, containing the T7 RNAP recognition site at the 5' end, was synthesised via PCR using primers ydaO1 and ydaO2 (sequences published in Nelson et al., 2013) as the forwards and reverse primers. Genomic DNA was extracted from *B. subtilis* (Method Section 2.3.7) and used as template DNA for the PCR reaction (Method Section 2.3.4). The products of the PCR reaction were analysed by agarose gel electrophoresis and are shown in Figure 5.10a. A product corresponding to the full length *ydaO* DNA template of size 185 bp can be seen in lane 2. This product was purified using a PCR purification kit (Method Section 2.3.5) and used as a template for an IVT transcription reaction.
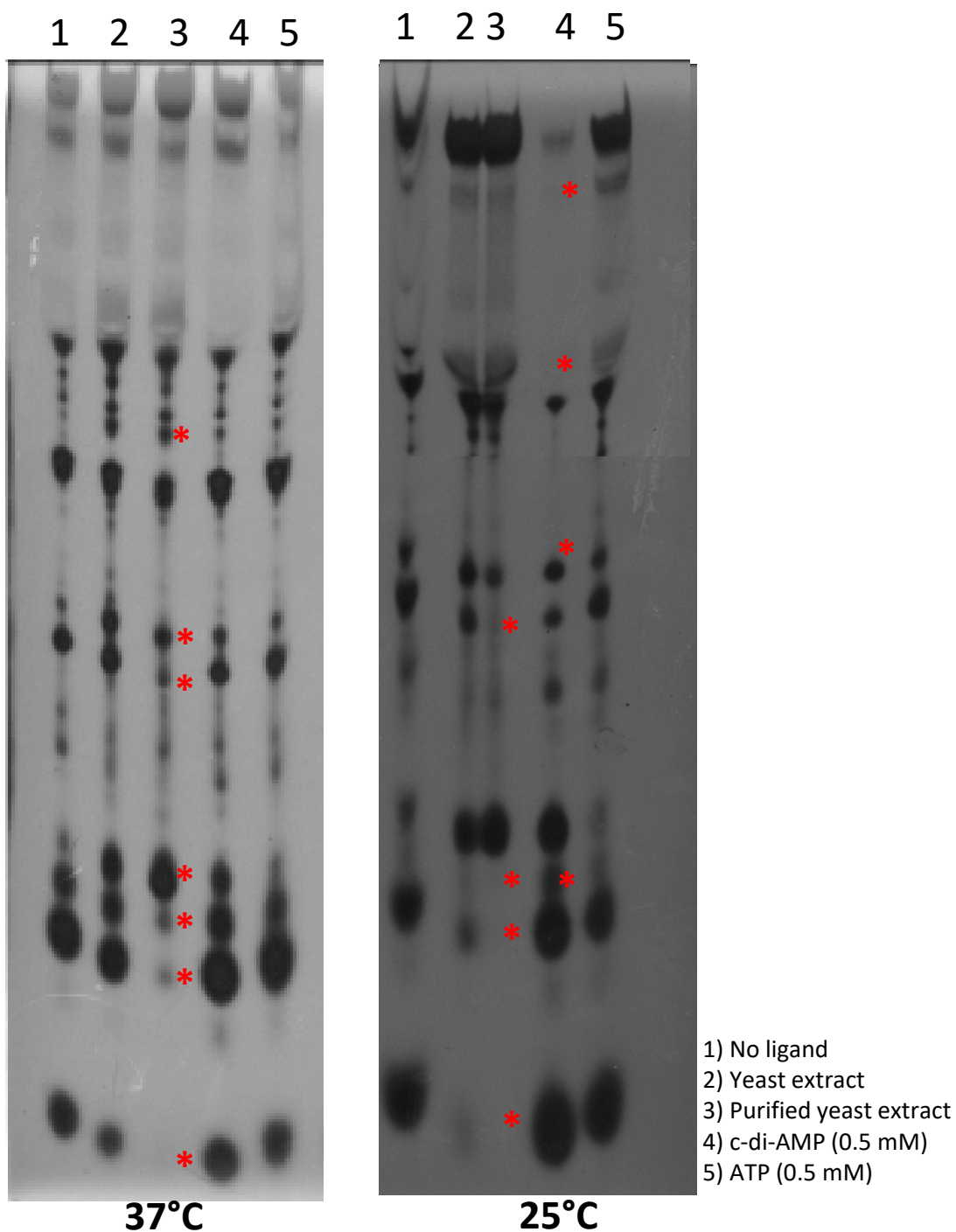
Figure 5.11: In-line probing analysis of *B. subtilis ydaO* riboswitch RNA in the presence of ligands incubated at either 25°C or 37°C. Samples were incubated with the specified ligand and were refolded. They were subsequently incubated for 36 h at either 25°C or 37°C and analysed by denaturing PAGE. RNA species that were observed to increase or decrease in intensity after incubation with a ligand are marked with *. These are representative gels selected from 4 repeats.

1   **2**   **3**   **4**   **5**   **1**   **2**   **3**   **4**   **5**

1) No ligand
2) Yeast extract
3) Purified yeast extract
4) c-di-AMP (0.5 mM)
5) ATP (0.5 mM)

37°C   **25°C**
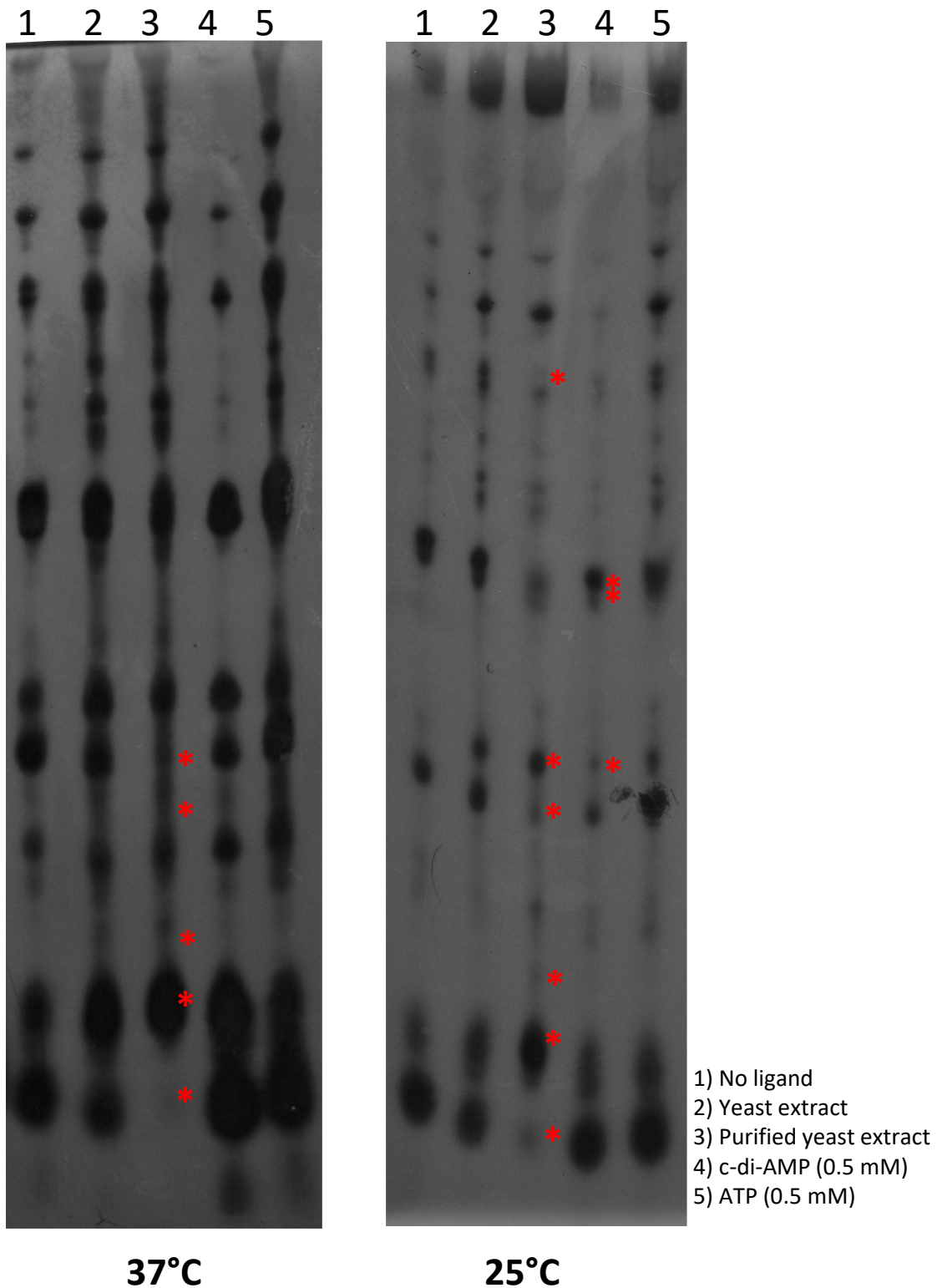
Figure 5.12: In-line probing analysis of *M. tuberculosis rpfA* riboswitch RNA in the presence of ligands incubated at either 25°C or 37°C. Samples were incubated with the specified ligand and0 were refolded . They were subsequently incubated for 14 h at either 25°C or 37°C and analysed by denaturing PAGE. RNA species that were observed to increase or decrease in intensity after incubation with a ligand are marked with *. These are representative gels selected from 4 repeats.
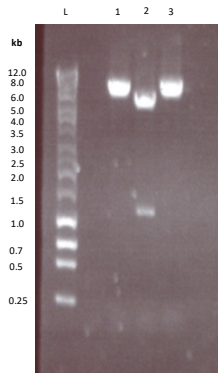
Figure 5.13: Agarose gel electrophoresis of digest product of the pET28b-disA$_{BT}$-His overpression plasmid after treatment with NcoI and/or XhoI. Lane (L) contains 5 µl of DNA Hyperladder 1kb Plus (Bioline) to be used to determine the size of species present in the digest products. Lanes 1 to 3 each contain 10 µl of restriction digest products with 2 µl of DNA loading dye (6X). Lane 1 contains XhoI digested plasmid; Lane 2 contains XhoI/NcoI digested plasmid and Lane 3 contains NcoI digested plasmid.

IVT reactions for the *B. subtilis* riboswitch sample were set up in the same manner as for the mycobacterial species riboswitches, which is outlined in detail in Results Section 5.2. The IVT reaction products were analysed by denaturing PAGE, the results of which are shown in Figure 5.10b. After exposure to UV (254 nm) on a TLC plate, one clean band corresponding to the full length (165 bp) *ydaO* riboswitch could be observed. RNA was purified from the IVT reaction components by column purification.

[32]P-5' end labelling was carried out for *ydaO* RNA as it was for the mycobacterial RNA, which is outlined in Results Section 5.2.4. End-labelling was verified by analysing the labelling reaction by denaturing PAGE and exposing the electrophoresed gel to a film (not shown). The band corresponding to the full length RNA was excised and the RNA was purified by the elution in to crush soak buffer, followed by ethanol precipitation.

5.3.3   In-line probing analysis of *B. subtilis ydaO* riboswitch RNA

After *B. subtilis ydaO* RNA was successfully synthesised and end-labelled with ^{32}P, ILP experiments were carried out. RNA (~1 pmol) was incubated with different ligands (yeast extract, 'purified' yeast extract, cyclic di-AMP (0.5 mM) or ATP (0.5 mM) or with no ligand present. Samples were refolded (Method Section 2.3.20) and incubated for 36 hours at either 25°C or 37°C. Samples were analysed on a 10% acrylamide (19:1), 2X TBE gel by denaturing PAGE gel (40 cm x 0.75 mm). The resulting gel was exposed to x-ray film (Figure 5.11).

Differences were observed (marked with *) in the breakdown of the *ydaO* riboswitch RNA in the presence of cyclic di-AMP (Figure 5.12, lane 4), suggesting that cyclic di-AMP interacts with the riboswitch resulting in conformational changes, which result in ligand-induced structural stabilisation of regions corresponding to species decreasing in prevalence or ligand-induced structural instability in regions corresponding to species increasing in prevalence.

5.3.4   In-line probing analysis of *M. tuberculosis rpfA* riboswitch

ILP experiments were carried out on ^{32}P end-labelled *M. tuberculosis rpfA* riboswitch RNA. RNA (~1 pmol) was incubated with different ligands (yeast extract, 'purified' yeast extract, cyclic di-AMP (0.5 mM) or ATP (0.5 mM) or with no ligand present. Samples were refolded (Method Section 2.3.23) and incubated for 14 h at either 25°C or 37°C. Samples were analysed on a 10% acrylamide (19:1), 2X TBE gel by denaturing PAGE gel (40 cm x 0.75 mm). The resulting gel was exposed to x-ray film (Figure 5.12).

Differences were observed (marked with *) in the breakdown of the *rpfA* riboswitch RNA in the presence of cyclic di-AMP (Figure 5.12, lane 4), suggesting that cyclic di-AMP interacts with the riboswitch resulting in conformational changes, which result in ligand-induced structural

stabilisation of regions corresponding to species decreasing in prevalence or ligand-induced structural instability in regions corresponding to species increasing in prevalence.

Differences were also observed (marked with *) in the presence of purified yeast extract, suggesting that one of the metabolites present is binding to the riboswitch. We propose that it is unlikely to cyclic di-AMP because, firstly, the breakdown pattern is different to that observed in the lane containing cyclic di-AMP, and secondly the wider literature suggest so. This is discussed further in section 5.3.4.

In the samples containing cyclic di-AMP (Figure 5.12, lanes 4), there is 'overall' a greater amount of breakdown and faster degradation of the RNA in comparison to incubation with no ligand (Figure 5.12, lanes 1) and this is true for the riboswitch RNA from both species. For the *rpfA* riboswitch this effect is noticeable at both temperatures, but in particular a large increase in smaller RNA species can be observed in the samples incubated at 37°C. For the ydaO riboswitch the effect of accelerated breakdown is more observable at 25°C. This suggests that in the presence of elevated levels of cellular cyclic di-AMP where the proportion of the riboswitches present will be in the cyclic di-AMP bound conformation(s) and hence translation is terminated, there is an increased amount of degradation. This could act as a feedback mechanism, so that once the ligand has bound and translation has been inhibited, the RNA will breakdown at a faster rate, which would allow for recycling of the nucleotides in the cell and the ligand could go on to other riboswitches.

In the presence of cyclic di-AMP (Figure 5.12, lane 4) fewer change in the breakdown pattern are observed at 25°C compared to 37°C for the *M. tuberculosis rpfA* riboswitch RNA. *M. tuberculosis* is a human pathogen and the optimal growth temperature is 37°C. For the *B. subtilis ydaO* riboswitch RNA in the presence of cyclic di-AMP more differences in the breakdown pattern are observed at 25°C than at 37°C. Interestingly, *B. subtilis* is naturally found in soil and vegetation and the optimal temperature for growth is 25 to 35°C according to the Entrez Genome Project (NCBI). It is logical that it would be beneficial for more changes in the breakdown of the riboswitch to be observed at temperatures that are physiologically relevant to the host species.

Interestingly, the optimal incubation time to produce an ideal amount of RNA breakdown for ILP was much lower for the *rpfA* riboswitch (14 h) than for the *ydaO* riboswitch (36 h). The *rpfA* RNA was noticeably trickier to work with in terms of generating full-length labelled RNA and the stability. This indicates that the structure of the *rpfA* riboswitch is inherently less stable, or more prone to spontaneous cleavage, than the *ydaO* riboswitch.

5.3.5 Does yeast extract contain cyclic di-AMP?

Performing literature searches provided no evidence that the *Saccharomyces cerevisiae* genome encodes any DisA homologs or diadenylate cyclases. Diadenylate cyclases are the enzymes responsible for the synthesis of cyclic di-AMP and more detail can be found in Introduction Section 1.5.2.1. It is unlikely that it is cyclic di-AMP that is causing the changes in breakdown observed for the *ydaO* and *rfpA* riboswitch samples.

A compound present in purified yeast extract is binding to the riboswitch and is responsible for the differences observed during ILP experiments. If it is not cyclic di-AMP that is responsible then what is? Data published in the supplementary material of Nelson et al. (2013) likewise found differences during ILP experiments with the *ydaO* riboswitch when incubated with yeast extract. They further purified the yeast extract by HPLC and other methods and selected for and enriched the target ligand. Mass spectroscopy on the purified ligand did not reveal cyclic di-AMP as the ligand, and only indicated the ligand likely contained an AMP moiety. There are a whole range of AMP-containing metabolites present in the cell, some of which have been implicated in binding to the riboswitch pApA and ApA (Nelson et al., 2013).

Many riboswitches are capable of binding more than one ligand, but these ligands often are closely related in structure (Bocobza and Aharoni, 2014; Furukawa et al., 2015). It is therefore, not unreasonable to suggest that another structurally related ligand, which has not previously been identified, could bind. Additionally, Watson and Fedor (2013) indicated ATP binds to the riboswitch, albeit with low affinity, so it is plausible due to the commonplace nature of ATP in cells that it is ATP causing the effect. Although we render this unlikely as no effect was observed when the riboswitch was incubated solely with ATP.

The probable lack of cyclic di-AMP within purified cell extract does not mean that the riboswitches cannot bind cyclic di-AMP, but only that another, potentially related (e.g. pApA discussed further in Section 5.4.4.3), metabolite might cause alterations in the stability of the RNA during in-line probing. Only minor differences were observed with c-di-AMP. However, at the point where these experiments were carried out the sodium cacodylate refolding buffer and program used later on for 'gel shifts' and the microscale thermophoresis (MST) experimental procedure had not been developed. If time had permitted, it would have been interesting to return to the in-line probing experiments (for all species) utilising the sodium cacodylate buffer.

5.4 'Gel shift' experiments with riboswitches

Experimental design for an 'electrophoretic mobility shift assay' style of experimental approach was undertaken to explore if cyclic di-AMP bound to the mycobacterial *rpfA* riboswitches. Work by the Gründling group (Campeotto et al., 2015) demonstrated that it is possible to generate ^{32}P radiolabelled cyclic di-AMP utilising purified *Bacillus* versions of DisA, a diadenylate cyclase enzyme. We theorised that incubating and refolding riboswitch RNA in the presence of the labelled ligand and analysing samples by native polyacrylamide gel electrophoresis, it would be possible to monitor for changes in the mobility/migration of the cyclic di-AMP. Retarded migration of the labelled ligand in the presence of riboswitch RNA would indicate a binding interaction.

5.4.1 Purification of *Bacillus thuringiensis* diadenylate cyclase, DisA

ANG3043, an *Escherichia coli* XL1 Blue strain containing DisA overexpression construct (pET28b-disA$_{BT}$-His; KanR) was kindly provided by Angelika Gründling of Imperial College London. The pET28b-disA$_{BT}$-His vector allowed for the overexpression of a C-terminally His-tagged version of the *Bacillus thuringiensis* DisA protein in *E. coli* BL21 λDE3 cells. The pET28b-disA$_{BT}$-His plasmid was used to transform electrically competent *E. coli* BL21 λDE3 cells. Transformants were selected on kanamycin-containing LB agar plates and a single colony was isolated on a streak plate. One colony from the streak plate was restreaked and an overnight culture was set up. The plasmid DNA was purified from the cells in the overnight culture and restriction digests (Method Section 2.3.3) was set up with NcoI, XhoI and a double digest. These were incubated for 2 h at 37°C. The cut plasmid DNA was analysed by agarose gel electrophoresis (Figure 5.13). Species are present in lanes 1 and 3, of ~6.3 kb in size, which correspond to linearised pET28b-disA$_{BT}$-His plasmid. In the double digest two species, corresponding to the inserted DisA coding sequence (1,071 bp) and the remaining pET28a plasmid (~5.3 kb) are present. This indicated successful transformation of the host overexpression strain with the DisA overexpression construct. The insert was sequenced (data not shown) using primers 5'NcoI-DisA-Bt and 3'XhoI-DisA-Bt (sequences provided in Campeotto et al., 2015) to verify that the coding sequence of the gene did not contain any mutations.

Next, DisA was overproduced in 3 litres of LB broth by the method outlined in Section 2.4.8, using a final concentration of 120 μg/ml IPTG. Cells were collected by centrifugation at 6,000 rpm for 20 min and resuspended in 20 mM sodium phosphate buffer (pH 7.4). Cell-free extract was prepared using a French pressure cell, followed by centrifugation and filtration of the sample. Nickel-affinity chromatography was performed to purify His-tagged DisA protein from
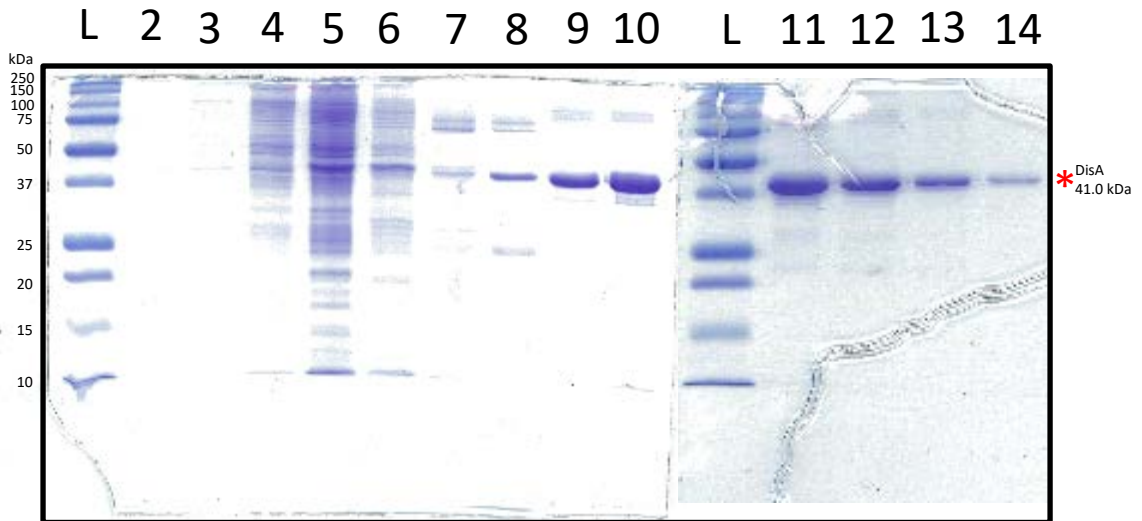
Figure 5.14: showing SDS-PAGE analysis of fractions eluted from a HiTrap column during the purification of DisA by nickel-affinity chromatography. Lanes L each contain 10 µl of Precision Protein Standards (BioRad), the sizes of which are located on the left. The numbers 2 to 14 correspond to the 1 ml elution fraction the protein sample is from. Each protein lane contains a total of 5 µl protein sample with 5 µl of 2X SDS-PAGE loading dye. The location of purified DisA protein (~41 kDa) is indicated (*). Samples 12 to 14 were deemed to contain protein of a sufficient purity for subsequent experiments.
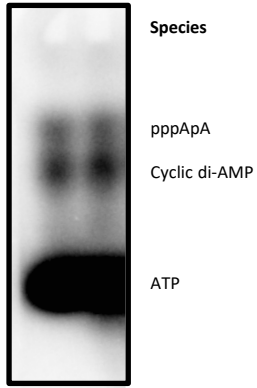
Figure 5.15: denaturing PAGE analysis of cyclic di-AMP synthesis reaction products. Both lanes contain samples (20 µl) of the reaction products after incubation of purified *B. thuringiensis* DisA with α^{32}P-[ATP] mixed with denaturing PAGE loading dye. Three species are visible, which likely correspond to free ATP, pppApA and cyclic di-AMP.

other cellular proteins. This was achieved using an AKTA Prime machine and the methods outlined in Section 2.4.8. Sodium phosphate (pH 7.4) was used as binding buffer to allow proteins to bind to the column, and an increasing gradient of NaCl (up to 1M) was applied to elute proteins in an affinity-based manner. Fractions (1 ml) were collected during purification and were analysed by SDS-PAGE (Method Section 2.4.1) shown in Figure 5.14.

Fractions 12 to 14 were selected as they contained successfully purified DisA, which can be seen in Figure 5.14 as a stained band of polypeptides corresponding to a size of 41,025 Da. DisA was concentrated using a VivaSpin 10 column (5 ml) and dialysed into cyclic di-AMP synthesis buffer (50 mM Tris (pH 7.5), 5% glycerol, 150 mM NaCl and 10 mM $MgCl_2$). It was important to have sufficient concentration of Mg^{2+} in the synthesis buffer, as it is a cofactor and is important for activity of DisA. Bradford assays were used to determine the protein concentration (Method Section 2.4.2.1), which was found to be ~20 μM.

5.4.2   Synthesis of c-di-AMP from α^{32}P-[ATP] by DisA

The main cellular role of DisA is to scan chromosomes for DNA damage, and upon detecting damage, temporarily block the initiation of sporulation (Bejerano-Sagie et al., 2006). However, DisA also possesses diadenylate cyclase activity and generate cyclic di-AMP using ATP as a substrate (Witte et al., 2008). Cyclic di-AMP is a signalling molecule that has been shown to play a role in the regulation of sporulation. Hence, DisA is able to couple DNA integrity to sporulation due to its dual functionality.

2 ATP $\rightarrow$ 2 diphosphate + cyclic di-3',5'-adenylate
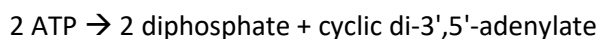
Cyclic di-AMP synthesis reactions were set up (Method Section 2.4.10) containing 10 μM protein with a final concentration of 55 nM α^{32}P-[ATP]. Reactions were incubated at 45°C for 16 hours and quenched. Figure 5.15 shows denaturing PAGE analysis of the products of the reaction. Three distinct species are visible, which we believe correspond to free, unreacted ATP, pppApA and cyclic di-AMP. Cyclic di-AMP synthesis is a two-step reaction (scheme shown in Figure 1.6), whereby initially two molecules of ATP join together and form pppApA. pppApA is then converted into cyclic di-AMP via a metal adduct transition state (Manikandan et al., 2014).

 The yield of cyclic di-AMP synthesised in these reactions was significantly less than anticipated from following the methods in Zheng et al. (2013). Ideally further purification by HPLC would be performed after this prep to isolate purified cyclic di-AMP from the mixture. However, performing HPLC with radioactive samples was not an available option in our laboratory. Instead, a further purification method was attempted that used thin layer chromatography

plates to separate species over a solvent gradient (Bai et al., 2012). However, the method resulted in significant loss of yield of cyclic AMP and the final purified ligand was not sufficient for the experiments.

It was decided to perform trial 'gel shift' experiments with the 'unpurified' cyclic di-AMP regardless to assess for binding to any of the reaction components. It was necessary to remove DisA protein from the reaction, before including using the mixture in 'gel shift' reactions. This was achieved by denaturing and precipitating out the protein.

5.4.3 'Gel shifts' with radiolabelled cyclic di-AMP

A method for riboswitch RNA 'gel shifts' was developed, which was based on previous studies that had successful performed such experiments with riboswitches (Meehan et al., 2016; Smith et al., 2009). The method is detailed in Section 2.3.21. *M. tuberculosis rpfA* and *B. subtilis ydaO* RNA were synthesised and purified as discussed in Results Section 5.2.1 to 5.2.3. Reactions were set up containing differing concentrations of RNA (10 pM to 100 nM) in 1X Refolding buffer (10 mM sodium cacodylate, 10 mM magnesium chloride and 10 mM potassium chloride) in the presence or absence of the radiolabelled cyclic di-AMP-containing mixture. Samples underwent a refolding program involving heating followed by slow cooling, and then incubation at room temperature for 12 h. Samples were analysed by native PAGE using an 8% acrylamide gel (Method Section 2.3.21).

No changes were observed in the presence of the radiolabelled cyclic di-AMP mixture. Some precipitate was also observed in the wells of the gel. It was hypothesised that this could potentially be the RNA oligomerising. In order to ascertain whether this precipitate contained some or all of the riboswitch RNA and hence could account for the absence of alterations in the migration of the cyclic di-AMP (particularly of the *ydaO* riboswitch), staining of the gel with GelRed (1:4,000) was attempted. Examination on a UV transilluminator showed that lanes containing the highest RNA concentrations (10 nM and 100 nM) had two distinct bands of RNA within the gel matrix. No RNA was observed in the lanes corresponding to the lower concentration of RNA – this is due to the limit of detection and sensitivity of GelRed. This confirmed that the RNA was entering the gel.

The ability to visualise the unlabelled RNA at relatively low concentrations inspired the idea to perform label free 'gel shift' experiments. This provided the advantage of being able to work non-radioactively, but more importantly, it would allow us to be able to track the migration of the RNA molecules, allowing us to screen a variety of compounds and not just be limited to our own generation of ^{32}P-cyclic di-AMP mixture. It would allow us to test commercially produced
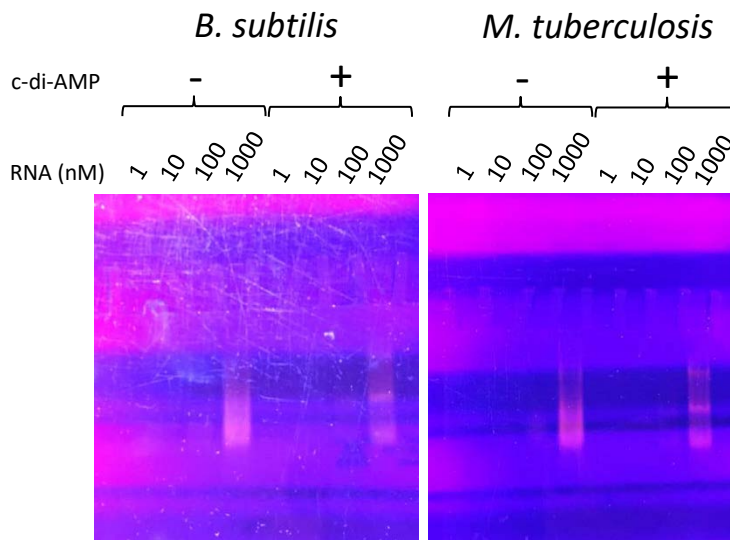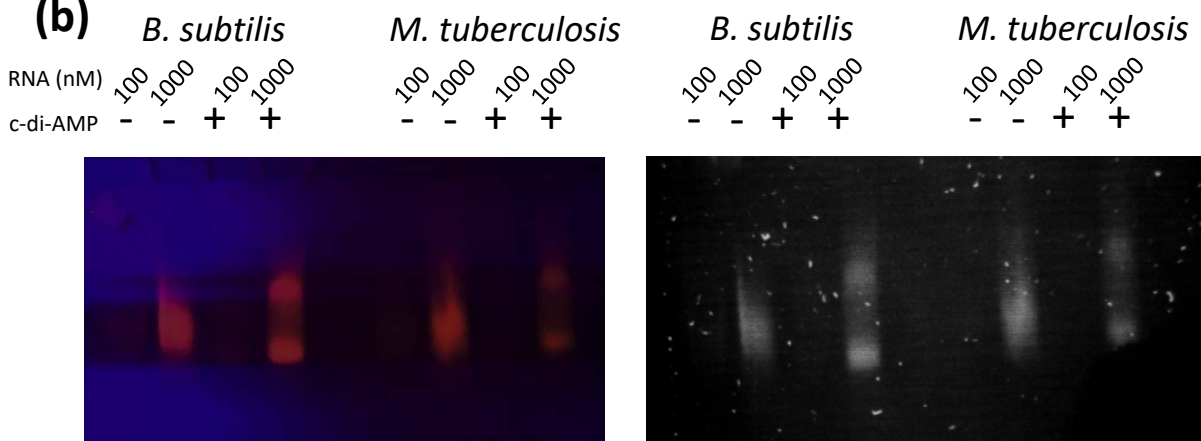
Figure 5.16: EMSA experiments with *B. subtilis* and *M. tuberculosis* riboswitch RNA in the presence and absence of cyclic di-AMP. Reactions containing varying amounts of riboswitch RNA (1 to 1000 nM) were refolded and incubated in the absence or presence of cyclic di-AMP (100 µM). Incubation was at 37°C for (a) 2 hours or (b) 18 hours post-refolding. Samples were analysed by native PAGE. The species from which the riboswitch belongs to is labelled. Three independent replicates were performed.

cyclic di-AMP, as we could not be certain that we did not observe alterations in the migration of cyclic di-AMP because of the mixture of products present (and high proportion of free ATP), if it was because conditions were not amenable for binding or if it was because cyclic di-AMP does not bind to the riboswitches. The major caveat of the technique would be that much higher concentrations of RNA would be required in order to successfully visualise the migration.

EDTA is a chelator of minerals and metals, including magnesium. The use of EDTA (5 mM final concentration) during the quenching of cyclic di-AMP synthesis reactions and subsequent carryover into 'gel shift' reactions could potentially affect the interaction between the RNA and ligand. EDTA will chelate Mg2+, which is often important in the refolding and stability of RNA structures. In unlabelled 'gel shift' experiments, commercially synthesised and purified cyclic di-AMP was used, which eliminated the potential for EDTA to affect interactions.

5.4.4   Unlabelled 'gel shift' experiments

5.4.4.1   *B. subtilis ydaO* and *M. tuberculosis rpfA* riboswitch RNAs interact with cyclic di-AMP

Reactions (30 μl) were set up in 1X refolding buffer (Method Section 2.3.21) that contained varying concentrations of riboswitch RNA (1 – 1000 nM), in the presence or absence of cyclic di-AMP (100 μM). Samples underwent the refolding program (Section 2.3.23) and were incubated for 2 h at room temperature. Aliquots (10 μl) of samples were combined with 2 μl of loading dye and were resolved by native PAGE. However, an increase in magnesium chloride concentration to 50 mM and a reduction in EDTA concentration to 100 mM in the running buffer was implemented, in the hope that it may enhance the stability of the folded RNA or RNA-ligand complex. The gel tank apparatus was kept cool by being placed in an ice box and having ice blocks within the tank during electrophoresis . After electrophoresis, the gels were transferred into 1X native PAGE buffer containing GelRed (1:4000 dilution) for 20 min with agitation. Gels were examined on a transilliuminator.

Figure 5.16a shows differences in the species present when both the *B. subtilis* and *M. tuberculosis* riboswitches were incubated in the presence of cyclic di-AMP. As expected, the GelRed stain was not sensitive enough to detect the lower concentrations of RNA (1 nM and 10 nM). RNA species in the 100 nM samples could be seen faintly, but clearest results were gained from the samples containing 1000 nM RNA. After 2 h of incubation in the presence of cyclic di-AMP, both the *B. subtilis* and *M. tuberculosis* riboswitch RNA samples produced three distinct bands on the native PAGE gel (1000 nM RNA + cyclic di-AMP), compared to a broad smeared species when no cyclic di-AMP was included in reactions (1000 nM RNA - cyclic di-

AMP). Native PAGE allows for RNA to run depending on its three-dimensional structure, compared to denaturing PAGE that unfolds the RNA allowing migration in proportion to its size. The additional species in the presence of cyclic di-AMP suggested that a proportion of the RNA had entered a different structural conformation, likely due to binding of cyclic di-AMP to the aptamer.

The same experiment was repeated for the 100 nM and 1000 nM RNA samples, but this time samples were incubated for 18 h post-refolding (Figure 5.16b). After 18 h of incubation in the presence of cyclic di-AMP, both the *B. subtilis* and *M. tuberculosis* RNA samples produced two distinct bands (1000 nM RNA +cyclic di-AMP in Figure 5.16b), rather than the three observed after only 2 h incubation post-refolding (1000 nM +cyclic di-AMP in Figure 5.16a). Only one (smeary) species was present in the absence of cyclic di-AMP for both the 18 hours and 2 hours post-refolding samples (Figure 5.16a 1000 nM -cyclic di-AMP compared to figure 5.16b 1000 nM -cyclic di-AMP). The presence of new additional species in the presence of cyclic di-AMP, compared to when no ligand is present, suggests that a proportion of the RNA has entered a different structural conformation, likely due to binding of cyclic di-AMP to the aptamer. The fact that after 2 h incubation three species are present, but after 18 h incubation only two were detected could be explained by the formation of some kind of intermediate structure during the cyclic di-AMP binding process and the riboswitch having a slow approach to equilibrium. Other studies have observed increased binding between ligand and riboswitches with longer incubation times (12 h or more) and suggested a slow approach to equilibrium and complete refolding of the RNA (Meehan et al., 2016; Smith et al., 2009).

The broad, smeared RNA species present in the absence of ligand suggests that many alternative conformations of the RNA may exist and the population of RNA is distributed in over a range of these. When the RNA is refolded in the presence of cyclic di-AMP the bands are more defined and this suggests that cyclic-di-AMP binding pushes the equilibrium and distribution of RNA conformations present in the population. The broad, smeared species could additionally be caused by RNA that is still 'refolding' and is between conformations, or has only partially folded. The presence of cyclic di-AMP may stabilise the refolded structures, which is why a more defined spread of conformations are observed.

5.4.4.2 Complete binding of cyclic di-AMP can be observed after 5 h at 37°C

In order to investigate the amount of time required for complete binding and refolding i.e. for only two species to be present, rather than three, a time course EMSA experiment was performed. Duplicate reactions were set up containing RNA (1000 nM) with or without cyclic di-AMP (100 µM) as before, but were staggered by 1 h intervals. Samples were refolded and

2 3 4 5 6 7 8 9 10 11 12   hours
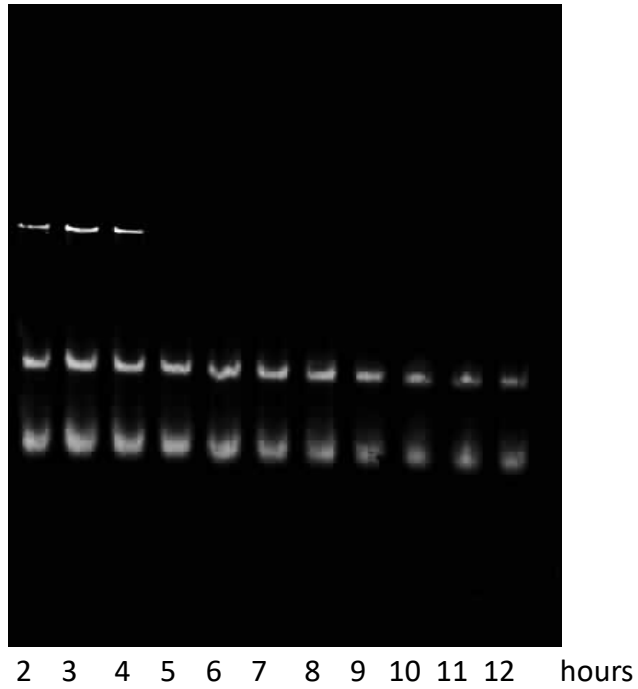
Figure 5.17: Time course EMSA with *M. tuberculosis rpfA* riboswitch RNA in the presence of cyclic di-AMP. All lanes contain 1 µM RNA that was refolded in the presence of 100 µM cyclic di-AMP and incubated at 37°C for the number of hours shown below the relevant well. Two replicates were performed.

(a)
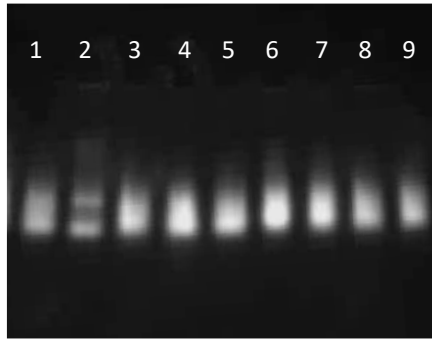
M. tuberculosis rpfA riboswitch

1 = no ligand   6= 3'AMP
2 = c-di-AMP   7= 5' AMP
3= c-di-GMP   8 = ATP
4 = 3'5'cAMP   9 = ADP
5 = 3'5'cGMP

(b)
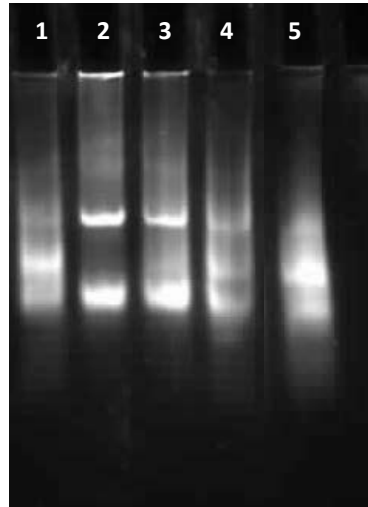
M. tuberculosis rpfA riboswitch

1 = no ligand
2 = c-di-AMP
3= pApA
4 = ApA
5 = c-di-GMP

(c)
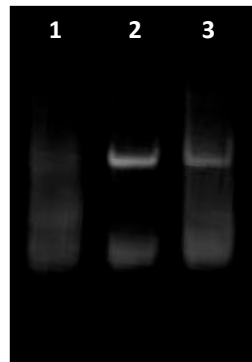
B. subtilis ydaO riboswitch

1 = no ligand
2 = c-di-AMP
3= pApA

Figure 5.18: EMSA screening various different ligands for their interaction with the *M. tuberculosis rpfA* riboswitch. RNA (1 µM) was incubated with different ligands (100 µM) and refolded. They were incubated at 37 °C for 45 minutes and were analysed by native PAGE. Gel was stained using Sybr Gold nucleic acid stain (1:10,000). The ligand the RNA was incubated with is indicated in the key. The species the riboswitch RNA is from is labelled. Three independent replicates were performed.

incubated at 37°C or room temperature for different lengths of time between 2 to 12 h. We hypothesised that incubation at 37°C may increase rate of binding, but may however affect the stability of the secondary structure of the RNA. The reactions were analysed by native PAGE, with the same modifications as above. The gel was stained with GelRed (1:4000 dilution) with agitation before visualisation using a gel imager (Figure 5.17).

The experiment showed that, similarly to the 2 h incubation at room temperature in the presence of cyclic di-AMP (1000 nM RNA +cyclic di-AMP, Figure 5.17), the lanes corresponding to 2, 3 and 4 h post-refolding incubation at 37°C (Figure 5.17) contain three specific species, which likely correspond to three different structural forms of the RNA and support our previous finding that cyclic di-AMP is interacting with the *M. tuberculosis rpfA* riboswitch. In samples from 5 h, only two species were present, suggesting that at 37°C the binding and refolding is able to reach completion within a 5 h timeframe.

5.4.4.3   Assessing the ability of other related ligands to bind to the *M. tuberculosis rpfA* riboswitch

Other ligands were screened for their ability to interact with the *rpfA* riboswitch of *M. tuberculosis*. Samples containing RNA (1 µM) and various ligands with some form of similarity to cyclic di-AMP (100 µM) were set up and incubated as in Section 2.3.21. Analysis by native PAGE was performed and samples were stained with Sybr Gold (1:10000 dilution) (Figure 5.18). Sybr Gold was used as a stain from this point on as it is more sensitive for the detection of nucleic acids, in comparison to GelRed, particularly at lower concentrations.

M. tuberculosis rpfA **riboswitch RNAs show altered migration after incubation with cyclic di-AMP or pApA**

When incubated without any ligand (Figure 5.18a, lane 1) one large, broad band can be observed, which we propose is due to the population of RNA being in a wide range of conformations. When incubated with cyclic di-AMP (Figure 5.18a, lane 2), which is able to bind to the *M. tuberculosis rpfA* riboswitch, the equilibriums between the different conformations shift and 2 to 3 main conformations are favoured and selected for. Lanes 3 to 9 (Figure 5.18a) had one broad band of species with different conformations, suggesting that the *M. tuberculosis rpfA* riboswitch did not interact with cyclic di-GMP, 3'5'-cyclic AMP, 3'5'-cyclic GMP, 3'-AMP, 5'-AMP, ATP or ADP.

The *M. tuberculosis rpfA* riboswitch is also able to bind pApA, as the observed migration of the RNA in the presence of pApA (Figure 5.18b, lane 3) or cyclic AMP (Figure 5.18b, lane 2) were similar, and the equilibrium of RNA conformations was shifted towards 2 specific
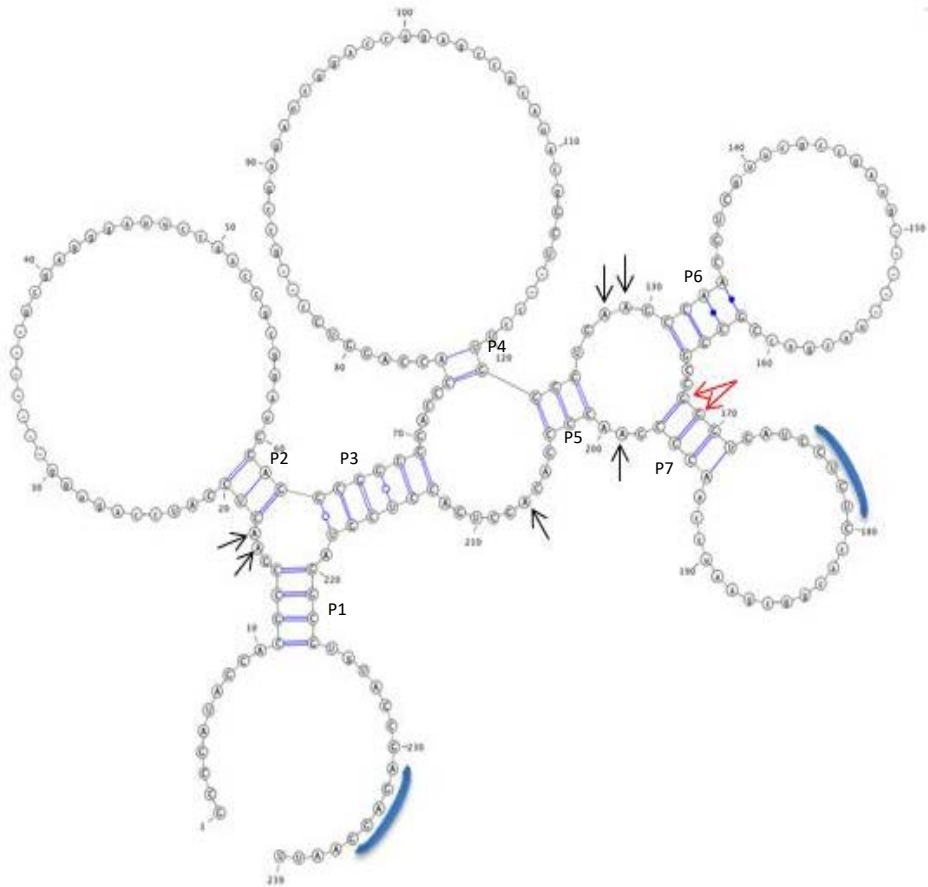
Figure 5.19: Model of the *M. tuberculosis rpfA* riboswitch. The *rpfA* riboswitch sequence was aligned to a cda-e model of the *ydaO* riboswitch and displayed using VARNA (Darty *et al*., 2009). The structural elements of the riboswitch are labelled P1-P7, the regions that interact to form a pseudoknot are indicated by filled arcs, bases that were targeted in mutagenesis studies because they interact with cyclic di-AMP and are indicated by black arrows, bases targeted to disrupt the P7 motif are indicated by red arrows,.

conformations. This differed from when the RNA was incubated in the absence of ligand (Figure 19b, lane 1), where a wider range of conformations were present and the species were located in a broader band, with a smaller proportion of the RNAs preferring 3 different conformations. ApA and cyclic di-GMP were also tested, but the RNA species migrated in the same way as when no ligand was present, suggesting that ApA and cyclic di-GMP do not interact with the *M. tuberculosis rpfA* riboswitch (Figure 19b, lanes 4 and 5).

B. subtilis ydaO riboswitch RNAs show altered migration after incubation with cyclic di-AMP or pApA

The *B. subtilis ydaO* riboswitch is able to bind pApA and cyclic di-AMP as the migration observed after incubation with these ligands was different to the migration observed after incubation with no ligand present (Figure 5.18c). The observed migration of the RNA in the presence of pApA (Figure 5.18c, lane 3) or cyclic AMP (Figure 5.18c, lane 2) were similar, and the equilibrium of RNA conformations was shifted towards 2 specific conformations. This differed from when the RNA was incubated in the absence of ligand (Figure 19c, lane 1), where a wider range of conformations were present and the species were located in a broader band, with a slight proportion of the RNAs preferring 3 different conformations.

5.5 Effect of mutation on the ligand-binding ability of the *rpfA* riboswitch of *Mycobacterium tuberculosis*

After gaining evidence to suggest that cyclic di-AMP is able to interact with the *M. tuberculosis rpfA* riboswitch (shown in section 5.4.4), we decided to see if we could mutate the riboswitch sequence to abolish the ability of the RNA to interact with cyclic di-AMP. A non-binding mutant would be of interest to this study and could be tested using the *in vivo* osmotic stress screen (discussed in Section 5.1.7) developed by our collaborators at the University of Leicester. This would provide greater insight into the alterations in salt tolerance and osmotic stress observed when the riboswitch portion of the *rpfA* mRNA is present or absent by providing information on whether the ability of the riboswitch to bind to the ligand (cyclic di-AMP) has an effect on the regulation of the gene.

5.5.1   Model of the *rpfA* riboswitch structure and selecting bases to mutate

A model of the *M. tuberculosis rpfA* riboswitch was generated by aligning its sequence to a covariance model of the *B. subtilis ydaO* riboswitch (Figure 5.19) and using the program VARNA (Darty et al., 2009). Bases associated with direct or indirect interaction with cyclic di-AMP were conserved (black arrows in Figure 5.19). The *B. subtilis ydaO* riboswitch binds two molecules of cyclic di-AMP per riboswitch. Adenosine residues in positions 10 and 64

Position in *M. tuberculosis rpfA*	Position in *B. subtilis ydaO*	Primers	Plasmid name
A16G	A9G	LD27 + LD28	pGS2581
A16U	A9U	LD29 + LD30	pGS2582
A17U	A10U	LD31 + LD32	pGS2584
A17G	A10G	LD33 + LD34	pGS2585
A128G	A64G	LD35 + LD36	pGS2586
A129G	A65G	LD39 + LD40	pGS2587
A129U	A65U	LD41 + LD42	pGS2588
A199G	A92G	LD43 + LD44	pGS2589
A199U	A92U	LD45 + LD46	pGS2590
A207G	A100G	LD47 + LD48	pGS2591
A207U	A207U	LD49 + LD50	pGS2592
G168C/G169C	n/a	LD51 + LD52	pGS2593

Table 5.1: Point mutations in the *M. tuberculosis rpfA* riboswitch selected for analysis for disruption of the interaction of the riboswitch with cyclic di-AMP. The first column shows the position of the base(s) to be substituted in *M. tuberculosis* RNA. The second column shows the corresponding position of the equivalent base in the *B. subtilis ydaO* riboswitch. The primers which were used to introduce the relevant mutations to pGS2580 are named in the final column. The final column provides the name of the newly generated construct.

(a)

(b)

← Full length RNA

← Bromophenol blue
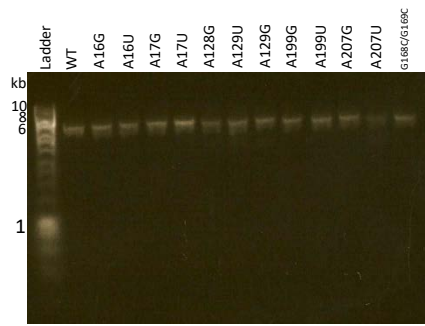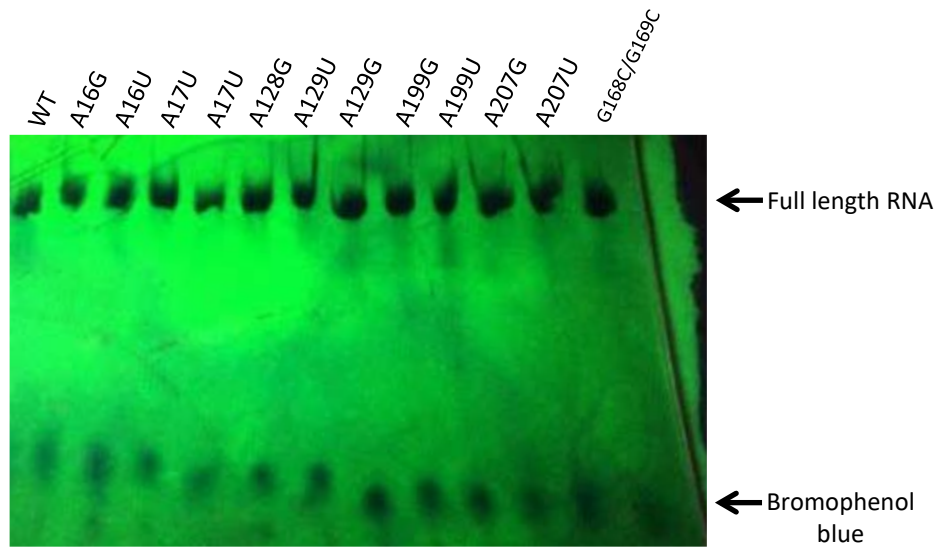
Figure 5.20: Generation of *M. tuberculosis rpfA* mutant RNAs. (a) Agarose gel electrophoresis of linearised mutant *rpfA* riboswitch-containing plasmids. Lanes 1-12 contained 5 µl EcoRI treated A16G, A16U, A17U, A17G, A128G, A129G, A129U, A199G, A199U, A207G, A207U and G168C/G169C mutant plasmids respectively. Lane 1 contained 5 µl DNA hyperladder 1kb plus (Bioline) (b) Denaturing PAGE analysis of mutant *rpfA* riboswitch RNAs. Lanes 1 to 12 contained 5 µl mutant *rpfA* riboswitch RNA generated by *in vitro* transcription.

(equivalent to 17 and 128 in *M. tuberculosis rpfA*) stack on cyclic-di-AMP ligands and adjacent residues adenosine 9 and 65 (equivalent to 16 and 129 in *M. tuberculosis rpfA*) are also involved in base stacking. Similarly, bases contributing to the pseudoknot are conserved (filled arcs in Figure 5.19).

Studies performed on the *B. subtilis ydaO* riboswitch RNA sequence by Jones and Ferré-D'Amaré confirmed the ability of cyclic di-AMP to interact with the RNA by isothermal titration calorimetry (ITC) experiments (Jones and Ferre-D'Amare, 2014). The K_D of the interaction was 53.1±30 nM. ITC performed with RNAs that had the bases that interact by stacking with cyclic di-AMP mutated (listed in Table 5.1, column 2) showed binding was either completely abolished or significantly reduced in affinity. It was decided to mutate the equivalent bases (Table 5.1, column 1) in the *M. tuberculosis rpfA* riboswitch and determine the effect on cyclic di-AMP binding by EMSA experiments. The equivalent base positions in *M. tuberculosis* are shown the first column of Table 5.1.

In addition to the mutations based on disrupting base stacking, a double mutant (G168C/G169C) that would disrupt the predicted pseudoknot of the *M. tuberculosis rpfA* riboswitch was designed in order to investigate the effect of pseudoknot disruption on the ability of the riboswitch to interact with cyclic di-AMP.

5.5.2 Generating mutant riboswitch constructs

5.5.2.1 Site-directed mutagenesis to produce mutant DNA templates

The mutant riboswitch constructs (listed in Table 5.1) were generated by site-directed mutagenesis (Methods Section 2.3.10) using pGS2580 as template DNA. pGS2580 is a pMV306 vector containing an insertion of the *M. tuberculosis* riboswitch and full length *rpfA* CDS, cloned between the EcoRI and KpnI cloning sites. This plasmid was provided by our collaborator, Dr Galina Muklamova of the University of Leicester. Overnight cultures were set up from single colonies recovered from the plate of successful transformants, determined by kanamycin resistance, for each mutation. Plasmid DNA was isolated by mini-prep and plasmid DNA was sequenced. Plasmids containing the desired mutation and no other mutations within the riboswitch coding region were transformed in to *E. coli* DH5α cells for long term storage as a glycerol stock. Samples of all the mutant plasmids after digestion with EcoRI were analysed by agarose gel electrophoresis (method section), which is shown in Figure 5.20a. All sample lanes contain a single DNA species of ~5.7 kb, which corresponds to pGS2580, pGS2581, pGS2582, pGS2584, pGS2585, pGS2586, pGS2587, pGS2588, pGS2589, pGS2590, pGS2591, pGS2592 and pGS2593 respectively.
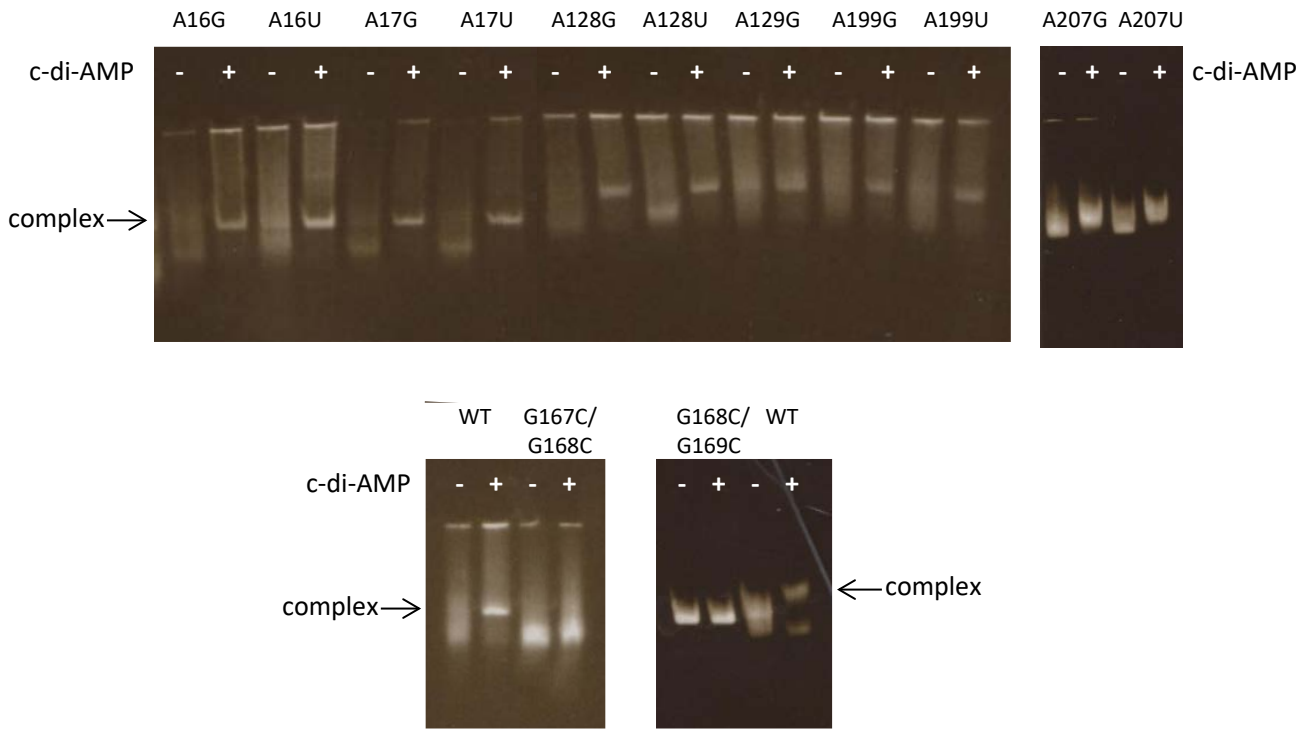
Figure 5.21: Native PAGE analysis of mutant *M. tuberculosis rpfA* riboswitch RNA in the presence and absence of cyclic di-AMP. Reactions contained 1000 nM of the mutant RNA. The exact mutation in lanes is labelled above. In samples containing cyclic di-AMP, the ligand was at a concentration of 100 μM.

5.5.2.2 Generation of mutant riboswitch RNA

The successfully mutated pGS2580 plasmids (pGS2581 to pGS2593, from section 3.5.2.1) were used as templates for PCR reactions with primers LD20 and LD21 to generate DNA templates for the *in vitro* transcription reactions to generate RNA. This was performed by the same methodologies as for the wild-type riboswitch that is detailed in section 5.2.1. This produced DNA templates containing the mutated riboswitch sequence with upstream T7 RNAP recognition site and triple G (GGG) to promote a robust yield of transcription product. PCR was used for the successful amplification of the mutant *rpfA* riboswitch DNA templates that could be used in in vitro transcription reactions to generate mutant riboswitch RNA (data not shown).

RNA was synthesised for all mutant riboswitches in the same manner as for the wild-type RNA (detailed in section 5.2.2). Successful generation of the desired full-length RNA products was shown by denaturing PAGE analysis in Figure 5.20b. Mutant RNAs were purified from the IVT reaction components by spin column purification using the same methodology as for the wild-type RNA (outlined in section 5.2.3.2). RNA was resuspended in nuclease-free water and stored at -20°C until it was used in 'gel shift' experiments.

5.5.3 Determining the cyclic di-AMP binding ability of the *M. tuberculosis* riboswitch mutants

In order to investigate if the mutated riboswitches could bind cyclic di-AMP, or if the mutation has sufficiently altered the secondary structure of the RNA to alter cyclic di-AMP binding ability, unlabelled gel shift experiments were performed by the same methods used for the wild-type (section 5.4.4). Mutant riboswitch RNA (1000 nM) was refolded in the presence or absence of cyclic di-AMP (100 µM) using the refolding program detailed in the methods section 2.3.23. Reactions were subsequently incubated at room temperature for 5 h and analysed by native PAGE and the resulting gel was imaged on a UV transilluminator, see Figure 5.21.

In the presence of cyclic di-AMP, the majority of the mutant riboswitch RNAs, including A16G, A16U, A17G, A17U, A128G, A128U, A129G, A199G, A199U, A207G and A207U, show altered migration under native PAGE conditions in comparison to when no cyclic di-AMP is present (comparing + with - lanes in Figure 5.21 for the relevant mutation). Typically, when no ligand is present (- lanes), the major band observed corresponds to a fast migrating conformation of RNA. For most mutants there is also a second slow migrating conformation of RNA present and there is smearing between the bands, which could be indicative of some structures that are not fully folded or are somewhere between the two main conformations. When the RNA has
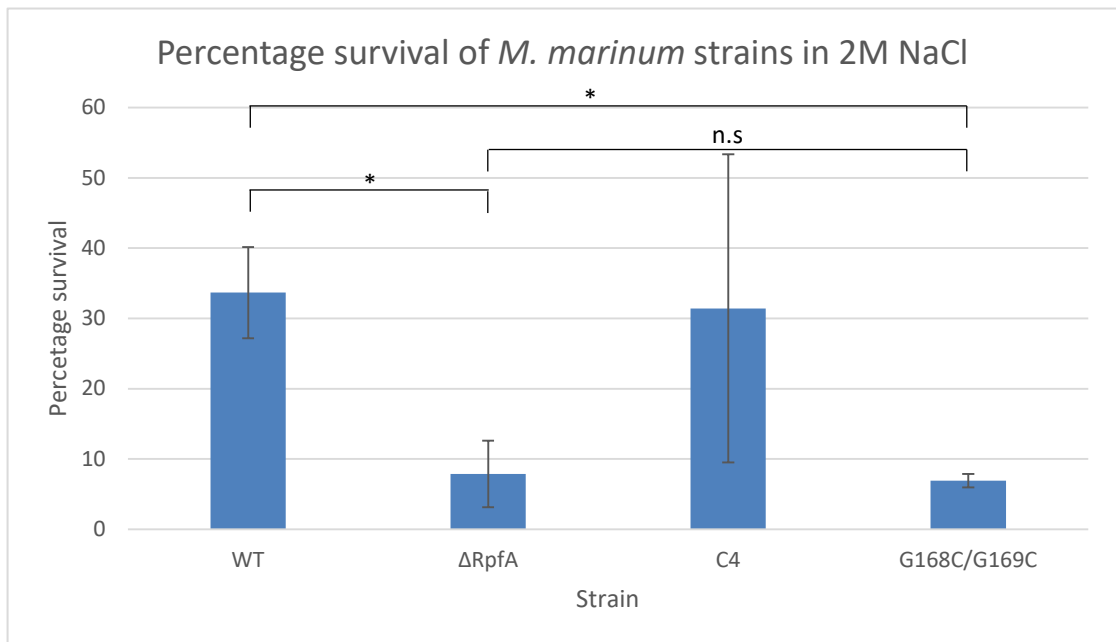
Figure 5.22: Percentage survival of *M. marinum* strains after osmotic stress with 2M NaCl. Strains tested included wild-type *M. marinum* (WT); *rpfA* deletion mutant (Δ*rpfA*) of *M. marinum*; Δ*rpfA* complemented with C4 construct (Figure 5.3) comprising the *M. tuberculosis* riboswtich and Δ*rpfA* complemented with the C4 construct with G168C/G169C mutation in the *M. tubuerculosis rpfA* riboswitch. Error bars show mean ± SD of three biological replicates. *p<0.05, n.s = not statistically significant (paired *t* test).

been refolded and incubated in the presence of cyclic di-AMP, there is a shift of the major conformation of RNA seen, from the fastest migrating species to a slightly slower migrating conformation. We theorise, that this slightly slower migrating species corresponds to a cyclic di-AMP-bound conformation of the riboswitch, which possesses different structural characteristics to the unbound form. These mutants behave in the same manner as wild-type (noted by WT on Figure 5.21), whereby a shift in migration pattern of this type is also seen. Our conclusion is that the A16G, A16U, A17G, A17U, A128G, A128U, A129G, A199G, A199U, A207G and A207U mutants are all capable of binding cyclic di-AMP.

However, the G168C/G169C mutant consistently showed no changes in the conformation of RNA present between when RNA was incubated in the presence or absence of cyclic di-AMP. The major species present in both reactions is the faster migrating RNA conformation. This provides evidence that the cyclic di-AMP binding ability of the *rpfA* riboswitch of *M. tuberculosis* can be abolished by mutating the 168 and 169 position bases. We believe this mutation is able to disrupt the pseudoknot region of the riboswitch, which could account for the diminished ability to bind cyclic di-AMP. The pG2593 (ΔG168C/G169C) vector was sent to our collaborators and was tested in the *M. marinum* salt tolerance model (discussed in Section 5.1.7) to test the effect of abolished cyclic di-AMP binding on the regulation of the *rpfA* gene *in vivo* (Figure 5.22). As mentioned in Section 5.1.7, the Δ*rpfA* deletion strain of *M. marinum* had lower percentage survival rate than WT *M. marinum* after osmotic shock. Construct (C4), containing a section the *rpfA* gene including the coding region and the *M. tuberculosis* riboswitch sequence, was sufficient to complement the percentage survival of the Δ*rpfA* mutant after osmotic stress to levels comparable to WT. A mutated C4 construct containing G168C/G169C substitutions in the riboswitch portion showed similar percentage survival to the Δ*rpfA* mutant and was unable to complement percentage survival levels to those of WT (Figure 5.22). This result supports the EMSA findings (Section 5.5.3, Figure 5.21) that cyclic di-AMP (or any putative ligand) is unable to bind to the G168C/G169C *rpfA* riboswitch mutant, and hence the riboswitch is no longer functional *in vivo* for the regulation of *rpfA*.

5.6 Microscale thermophoresis (MST) to determine kinetics of cyclic di-AMP binding by the *ydaO/rpfA* riboswitch

5.6.1 What is microscale thermophoresis?

Microscale thermophoresis (MST) is a technique of determining binding parameters of an interaction, which exploits the principle that molecules thermophorese either positively (hot to cold) or negatively (cold to hot) when a temperature gradient is applied to them. The thermophoresis of a molecule is dependent on various factors including charge, size and

hydration shell of a molecule. When a molecule interacts with a ligand it alters how the molecule thermophoresis, due to alterations in one or more of the aforementioned variables. By fluorescently labelling one of the interaction partners and applying steep temperature gradients, it is possible to detect changes in thermophoresis using a MST machine. Reactions are set up in capillaries containing a constant concentration of the fluorescently labelled partner with a range of concentrations of the unlabelled partner. Light of a relevant wavelength to excite the fluorophore is applied and the fluorescence recorded. A laser is then used to apply a steep temperature gradient to the capillary and the changes in fluorescence are recorded over time. The laser is subsequently switched off and the back-diffusion is measured. Data from the curves of different concentrations of unlabelled partner can be analysed to provide thermodynamic information on the interaction.

Two approaches could be used to probe the interaction between the *M. tuberculosis rpfA* riboswitch and cyclic di-AMP. The first would involve 5' end labelling the RNA with fluorescein and adding a constant concentration of labelled RNA with a range of concentrations of cyclic di-AMP. This approach is discussed in Section 5.6.2. Towards the later stages of the project, Biolog released a fluorescently labelled version of cyclic di-AMP (2'-O-(6-[Fluoresceinyl]aminohexylcarbamoyl)-cyclic di-AMP or F-cyclic di-AMP), which allowed us to attempt the reverse approach: using labelled cyclic di-AMP and unlabelled RNA. Results from this approach are detailed in Section 5.6.3. The main benefit of the latter approach is that the analysis of the data gathered from the MST machine assumes a 1:1 ratio of fluorophore:ligand. However, in reality, 5'-end labelling of RNA resulted in labelling efficiencies much lower than this (typically 66-89%). The ability to use commercially produced F-cyclic di-AMP removed this issue.

5.6.2 MST of 5'-end fluorescent labelled *M. tuberculosis rpfA* riboswitch RNA and cyclic di-AMP

Riboswitch RNA was synthesised by the same methods as for the in-line probing experiments (detailed in section 5.2), but the 5'-end was fluorescently-labelled with fluorescein by the methods. This involved removal of the 5'-phosphate with CIAP, followed by treatment with T4 PNK in the presence of with PγSH. This added a reactive sulphydryl group to the 5'-end of the RNA, which could be reacted with fluorescein maleimade, resulting in 5'-end fluorescently labelled RNA. The labelling efficiency was determined to be between 66-89% was achieved. RNA was subsequently refolded by heating and slowly cooling. Reactions containing serial dilutions of c-di-AMP were set up and applied to the capillaries. The MST machine was set to 37 °C and a variety of blue laser (excitation) powers and MST powers (gradients) were tested. The reaction buffer was modified by the addition of Tween (0.1%) to improve the quality of

data collected. No alterations in thermophoresis were detected in the presence of c-di-AMP (data not shown).

When these experiments were carried out, the refolding buffer (containing sodium cacodylate) and the refolding protocol used in the successful 'gel shift' experiments had not been developed yet. During these experiments the refolding of the RNA was carried out in the absence of ligand. It is plausible that for interactions to be observed the RNA must be refolded in the presence of the ligand, as within a cell it is likely that the RNA will be folding as it is being transcribed in the presence of the ligand and that the ligand may stabilise substructural elements during folding. The addition of ligand after refolding could account for why no interaction was be observed. It is also interesting to consider the importance of the sodium cacodylate buffer, as this buffer has been widely used across many families of riboswitches for binding studies (Delfosse et al., 2009; Gilbert et al., 2008; Guedich et al., 2016; Jenkins et al., 2011; Luo et al., 2014; Spitale et al., 2009). It would be useful to repeat these MST experiments with the use of the refolding buffer and performing the refolding in the presence of the ligand to determine if an interaction could be observed under these conditions.

5.6.3 MST of riboswitch RNAs and F-cyclic di-AMP

As no interaction was initially observed between the *M. tuberculosis rpfA* riboswitch and cyclic di-AMP using MST, the technique was placed on hold. This was until Biolog released a fluorescently labelled version of cyclic di-AMP (structure shown in Figure 5.2g). It was decided to use MST to screen for any interaction between unlabelled RNA and F-cyclic di-AMP under the conditions under which binding had been observed between cyclic di-AMP and both *B. subtilis ydaO* RNA and *M. tuberculosis rpfA* RNA in 'gel shift' experiments. If the technique was successful, it would provide binding affinities (K_D) and other useful information on the interaction. Although this wouldn't be a K_D value for the interaction with cyclic di-AMP itself, it is a closely related compound and could still give indications of the tightness of binding.

5.6.3.1 Interaction between F-cyclic di-AMP and *B. subtilis ydaO* riboswitch RNA

MST was performed on *B. subtilis ydaO* riboswitch RNA and F-cyclic di-AMP by the methods outlined in Section 2.3.22. Reactions containing RNA (200 nM) with a range of F-cyclic di-AMP concentrations (0.6 nM to 20 µM) were set up in 1X refolding buffer. Samples underwent refolding and were transferred to Standard capillaries (by Nanotemper) and MST was carried out (Figure 5.23). Changes in the thermophoresis of F-cyclic di-AMP were observed that were dependent on the concentration of RNA present. At higher RNA concentrations there is more positive thermophoresis than at lower RNA concentrations, which can be explained by the formation of RNA:F-cyclic di-AMP complex that positively thermphoreses. Analysis was

Figure 5.23: Microscale thermophoresis for the interaction between B. subtilis *ydaO* riboswitch RNA and F-c-di-AMP. MST experiments were performed with 50 nM of F-c-di-AMP incubated with *B. subtilis ydaO* riboswitch RNA (0.6 nM to 20000 nM). The concentration on the axis is referring to nM. Nanotemper Analysis software revealed a dissociation constant of 555 nM ± 47.9 for the the interaction.

performed using NanoTemper Analysis software and provided a K_D value of 555.5± 47.9 nM for the interaction.

Cyclic di-AMP is small (molecular weight 329.2) in comparison to the fluorescent labelling region (molecular weight 829.7). The F-cyclic di-AMP binds to the *ydaO* riboswitch with an affinity lower (~800-fold) than has been demonstrated for cyclic di-AMP in other studies (Nelson et al., 2013). The presence of the fluorescent label in F-cyclic di-AMP would likely decrease the ability of the molecule to enter the ligand-binding site as effectively when compared to the natural cyclic di-AMP. In-line probing experiments provided a K_D value of ~1 nM for the interaction between cyclic di-AMP and the *ydaO* riboswitch (Nelson et al., 2013). For pApA, a structurally close relative of cyclic di-AMP, the interaction with *ydaO* riboswitch showed binding with a K_D value of ~300 nM (Nelson et al., 2013). The interaction observed during MST experiments of the interaction between the riboswitch and F-cyclic-di-AMP was weaker. However this is to be expected as the fluorescein moiety will likely affect the ability of cyclic-di-AMP to bind to the RNA. It would be interesting to return to the original approach of labelling the RNA (discussed in Section 5.6.2) and retry with these new buffer and refolding conditions to see if an interaction can be detected by MST.

5.6.3.2 Interaction between F-cyclic di-AMP and *M. tuberculosis* rpfA riboswitch RNA

MST was performed on *M. tuberculosis rpfA* riboswitch RNA and F-cyclic di-AMP by the methods outlined in Section 2.3.22. Reactions containing RNA (0.6 to 20000 nM) with F-cyclic di-AMP (50 nM) were set up in 1X refolding buffer. Samples underwent refolding and were transferred to Standard MST capillaries (Figure 5.24). Changes in the thermophoresis of F-c-di-AMP were observed in a concentration-dependent manner in the presence of *rpfA* riboswitch RNA, suggesting that an interaction is taking place between the molecules. However, at the time of writing, insufficient repeats had been performed to accurately be able to determine a K_D value for the interaction.

5.7 Conclusions of this chapter

During this project we have succeeded in the generation and purification of DNA templates that can be used in IVT reactions for the synthesis of riboswitch RNAs from *B. subtilis* (*ydaO*), *M. tuberculosis* (*rpfA*), *M. smegmatis* (*rpfA*) and *M. marinum* (*rpfA*). In addition to this we have developed methods to allow for the synthesis and purification of these RNAs and for the 5'-^{32}P end radiolabelling of the purified RNAs.
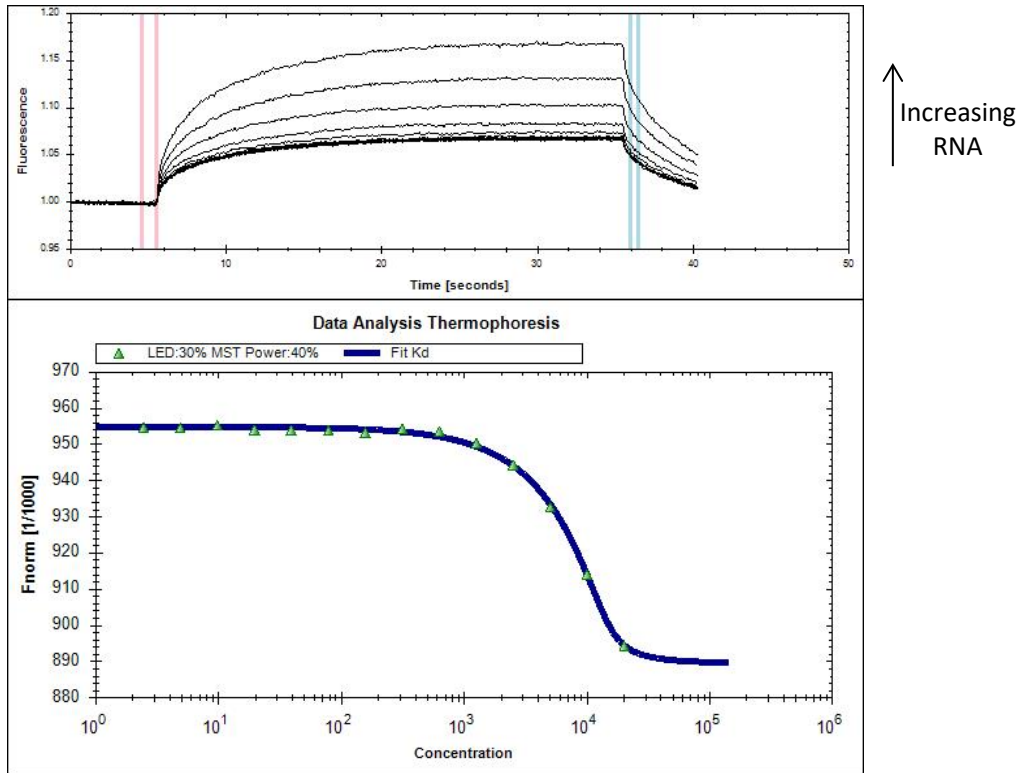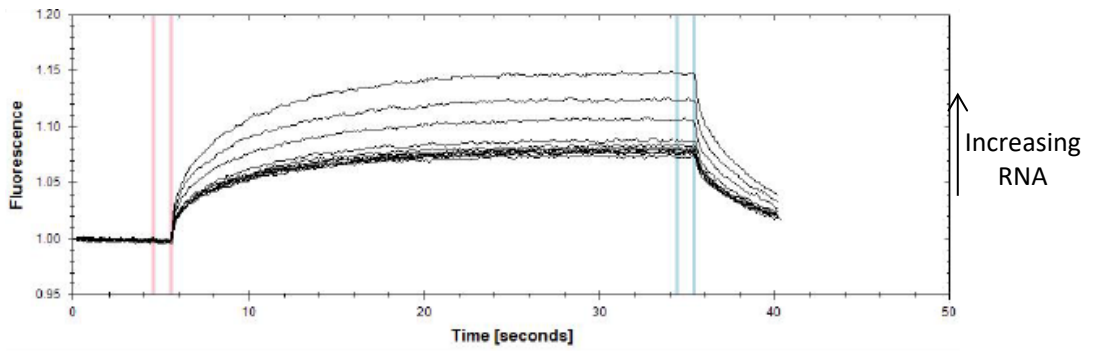
Figure 5.24: Microscale thermophoresis for the interaction between *M. tuberculosis rpfA* riboswitch RNA and F-c-di-AMP. MST experiments were performed with 50 nM of F-c-di-AMP incubated with *M. tuberculosis rpfA* riboswitch RNA (0.6 nM to 20000 nM).

We have extensively optimised different technical issues and developed a method that can be used for in-line probing analysis of riboswitches. This method has been used to analyse the *rpfA* riboswitch from *M. tuberculosis* and the *ydaO* riboswitch from *B. subtilis* with yeast extract, C18 purified yeast extract, cyclic di-AMP and ATP. Clear differences in the breakdown pattern of both riboswitches were observed in the presence of purified yeast extract, suggesting that a metabolite present in yeast extract interacts with both the *B. subtilis* and *M. tuberculosis* riboswitches. Literature suggests that cyclic di-AMP is not produced in yeast, suggesting that another molecule, possibly containing an AMP moiety (reference Nelson), could bind the riboswitches and be responsible for the alterations in breakdown we observed. Differences in spontaneous breakdown of the riboswitches can also be seen when they are incubated with cyclic di-AMP, suggesting both riboswitches are able to bind cyclic di-AMP. This methodology can now be used to assess the ability of the *rpfA* riboswitches of *M. marinum* and *M. smegmatis* to bind cyclic di-AMP and other ligands.

We have successfully purified the adenylate cyclase enzyme DisA (*B. thuringiensis*) in an *E. coli* overexpression strain, but were unsuccessful in synthesising yields of radiolabelled cyclic di-AMP comparable to those published in Zheng et al. (2013).

EMSAs were used to test a panel of potential ligands with the *rpfA* riboswitch. The migration patterns of *M. tuberculosis rpfA* and *B. subtilis ydaO* riboswitches were clearly different in the presence of cyclic di-AMP and linear pApA. Without the addition of a ligand, or in the presence of non-interacting ligands, the *rpfA* RNA appeared a 'smear' punctuated by more intensely stained regions, suggesting the presence of multiple conformations, some of which were more abundant than others. The species with intermediate mobility was the one most intensely stained. Addition of cyclic di-AMP or pApA resolved these multiple conformers such that the species of intermediate mobility was severely decreased in intensity relative to a fast and, predominantly, a slow migrating species. This suggests that the cyclic di-AMP-*rpfA* riboswitch complex is thermodynamically more stable than alternative structures and therefore the conformer(s) capable of interaction with cyclic di-AMP are selected from the assortment of *rpfA* riboswitch structures. We have demonstrated in these unlabelled 'gel shift' experiments that cyclic di-AMP and pApA bind to the *ydaO* riboswitch from *B. subtilis* and the *rpfA* riboswitch from *M. tuberculosis*. We have found no evidence that either of the riboswitches are able to bind cyclic di-GMP, 3'5'-cyclic AMP, 3'5'-cyclic GMP, 3'-AMP, 5'-AMP, ATP or ADP using the 'gel shift' approach under the conditions tested.

Microscale thermophoresis experiments revealed both the *B. subtilis* and *M. tuberculosis* riboswitches interacted with F-cyclic di-AMP with a K_D values of 555.5±7.6 nM for *B. subtilis*

ydaO. It would be useful to perform MST with the G168C/G169C mutant and F-cyclic di-AMP. 'Gel shifts' suggested this mutant is unable to bind cyclic di-AMP and MST data could further evidence this finding.

Twelve site-directed mutants of the *rpfA* riboswitch from *M. tuberculosis* were successfully generated and their ability to bind cyclic di-AMP was screened using 'gel shift' experiments. A16G, A16U, A17G, A17U, A128G, A128U, A129U, A199U, A199G, A207U and A207G were still able to bind cyclic di-AMP. Interestingly, A129G in the absence of cyclic di-AMP was able to migrate in a similar manner to when cyclic di-AMP was present. A double mutant, G168C/G169C, which is predicted to disrupt the psuedoknot structure of the riboswitch, is unable to bind cyclic di-AMP and our collaborators are tested this construct in the osmotic stress screen mentioned in section 5.1.7 and found it to behave similarly to the Δ*rpfA* strain, suggesting it is unable to bind its cognate ligand *in vivo*.

5.8 Discussion and further work

5.8.1 Further characterisation of the riboswitch mutants

Further characterisation of the G168C/G169C mutant riboswitch, which is unable to bind cyclic AMP (Section 5.5.3), should be carried out. It would be useful to assess if the G168C/G169 mutant is able to bind any other ligand(s) – in particular, performing gel shift experiments with pApA would be informative. Gel shift experiments could also be performed with the other site-directed mutants that were created in this study in order to assess if they are able to interact with pApA, as was observed for the wild-type *rpfA* riboswitch. After performing gel shift experiments, microscale thermophoresis could be performed with all mutants to determine the K_d for their interaction with F-c-di-AMP. If any of the mutants appear to have significantly increased affinity for the F-c-di-AMP, these mutants may be of interest for further structural studies as they may provide insights into how alterations in the sequence affect the structure and what structural features are important for binding ligands.

5.8.2 Consolidation of K_D values

Further microscale thermophoresis experiments should be performed with the *M. tuberculosis rpfA* riboswitch and F-cyclic-di-AMP in order to produce enough independent repeats to accurately determine a K_D value for the interaction. This can be used to directly compare the affinity of interaction between F-cyclic-di-AMP and the *M. tuberculosis* with the *B. subtilis* riboswitch.

All quantitative attempts to determine the K_D for the interaction between cyclic di-AMP and *ydaO* or *rpfA* riboswitches involved either labelling the RNA with fluorescein or using a

fluorescent cyclic-di-AMP analogue. This means K_D data generated will likely be affected by the presence of the label – particularly when the cyclic-di-AMP is labelled, as the fluorescent label has a larger molecular weight than the ligand itself. In order to address this issue, a label-free system could be employed. Potentially isothermal titration calorimetry could be performed using the methods outlined in Ren and Patel (2014) as a starting point. Performing ITC would provide additional support for K_D values gained during MST experiments and also information on the stoichiometry of the interaction and if it is an endothermic or exothermic.

The major caveats of using ITC would be the requirement of large quantities of RNA (micromolar concentrations) in order to perform a single run. Producing sufficient quantities for several repeats may be a limiting factor. Additionally, interactions were observed only when the RNA was refolded in the presence of the ligand. This refolding would not be possible in the ITC system. The BLITz system for characterising interactions could also be trialled because biotinylated RNAs are widely available for commercial purchase. However, this would still not allow for the refolding of the RNA in the presence of the ligand because the RNA would have to be immobilised on the tip of the sensor before exposure to the ligand.

This would provide additional support for KD values gained in MST experiments and also information on the stoichiometry of the interaction and if it an endothermic or exothermic binding event.

5.8.3   Screening of other mycobacterial *rpfA* riboswitches

As discussed in Section 5.1.6, other mycobacteria have putative riboswitches located in the 5' UTR of their messenger RNA. This includes *M. smegmatis*, *M. marinum* and *M. bovis*. The sequence of the *M. bovis* riboswitch is identical to that of the *M. tuberculosis* riboswitch, so our *in vitro* observations of interactions would be identical. However, there has been no published studies providing evidence over whether the *M. bovis* riboswitch plays a regulatory role *in vivo*. Hence, it would be useful to perform experiments that would determine is the riboswitch is functional. Mutants with different regulatory features of the gene absent, similar to the constructs used for *M. tuberculosis* in Figure 5.3, could be generated. These mutants could then be screened to ascertain if there was any osmotic stress-related phenotype that could be exploited in order to ascertain a functional role of the riboswitch. If not, qRT-PCR could be performed under various conditions in order to determine if the mutants lacking the riboswitch have significantly altered levels of expression of *rpfA*. In addition, these experiments should be performed with *M. smegmatis* and *M. marinum*.

During this study, the *M. smegmatis* and *M. marinum rpfA* riboswitch RNAs were successfully synthesised by in vitro transcription and were successfully end-labelled (Section 5.2.3). These

RNAs should be screened for ligand binding via similar methods as used for the *M. tuberculosis rpfA* riboswitch, including in-line probing, gel shift and microscale thermophoresis. Predicted structures can be produced using M-fold and comparisons between the structural differences between the different *rpfA* riboswitches and their ability to bind and affinity for different ligands should be compared. This will allow for comparison of which structural elements and bases might be crucial in ligand-binding. The differences in sequence between the different *rpfA* riboswitches may be due to their different requirements for responding to cyclic-di-AMP or other small nucleotides in survival of the species. *M. tuberculosis* and *M. bovis* are both mammalian pathogens who need to respond to different host conditions, whereas *M. smegmatis* is primarily a soil-dwelling bacterium. The levels of different signalling molecules that the different bacteria will experience in their native habitats is different and hence could explain the subtle differences in the predicted structures of their riboswitches.

5.8.4   Crystal structures

Producing crystal structures of the different forms of the *rpfA* riboswitches from the various species would be helpful and would confirm the accuracy of the predicted structures produced in this study. Other riboswitches, including the *ydaO* riboswitch from the thermophilic bacteria *Thermoanaerobacter pseudethanolicus* and *Thermovirga lienii* have had crystal structures solved (Gao and Serganov, 2014). These methods could be trialled and optimised to allow for crystals of the mycobacterial *rpfA* riboswitches to be produced and hopefully solved. Crystals will provide greater insight into where cyclic AMP and/or pApA and other ligands interact with the riboswitch and what conformational changes occur upon ligand binding. It has been suggested that members of this family of riboswitch are stabilised by intermolecular base stacking between the cyclic-di-AMP and specific bases of the riboswitch. Crystal structures would allow greater insight in to whether this is the case for the mycobacterial members.

6 Discussion

The central theme of this study is the characterisation of the regulation of *rpfA*. The *rpfA* gene is important for the entry and exit from dormancy (Haiser et al., 2009; Sexton et al., 2015). Dormancy is crucial as it allows mycobacteria to survive during extended periods of suboptimal conditions, for *M. tuberculosis* within a host with a healthy immune system. Dormancy is defined by low levels of cellular metabolism and a lack of replication (Dworkin and Shah, 2010; Rittershaus et al., 2013). The strategy employed by mycobacteria is to form a 'cyst-like' granuloma structure within their host, which requires remodelling of the cell wall during the entry phase and also during resuscitation from dormancy and commencing active growth (Anuchin et al., 2009; Bradley and Ritzi, 1968; Dewi Puspita et al., 2013; Glauert, 1961; Kaprelyants et al., 2012; Mukamolova et al., 1995).

Resuscitation promoting factors or Rpfs (discussed further in Introduction Section 1.7) are a family of secreted proteins with muralytic function and peptidoglycan cleaving activity (Kana et al., 2008). The muralytic activities of RpfA can be extremely damaging for a cell if produced at an inappropriate time during the life of the cell. The tightly-controlled regulation of expression is vital to maintain the appropriate integrity of the cell wall during each stage of growth. This explains why the regulation of *rpfA* is complex and multi-levelled with a variety of regulatory control mechanisms at play. This includes a riboswitch encoded in the 5' UTR of *rpfA* messenger RNA (covered in Chapter 5); a binding site for Rv3676, a cyclic-AMP binding CRP protein (covered in Chapter 4); a binding site for CsoR, a copper-sensitive repressor protein (covered in Chapter 3); Lsr2 and MtrA binding sites (identified by ChIP-seq data (Cortes et al., 2013). Taking into account the experimental data gathered throughout this study, alongside relevant published data, an updated overall scheme for the regulation of *rpfA* is shown in Figure 6.1. This discussion will cover different elements of this "bigger picture" in more detail.
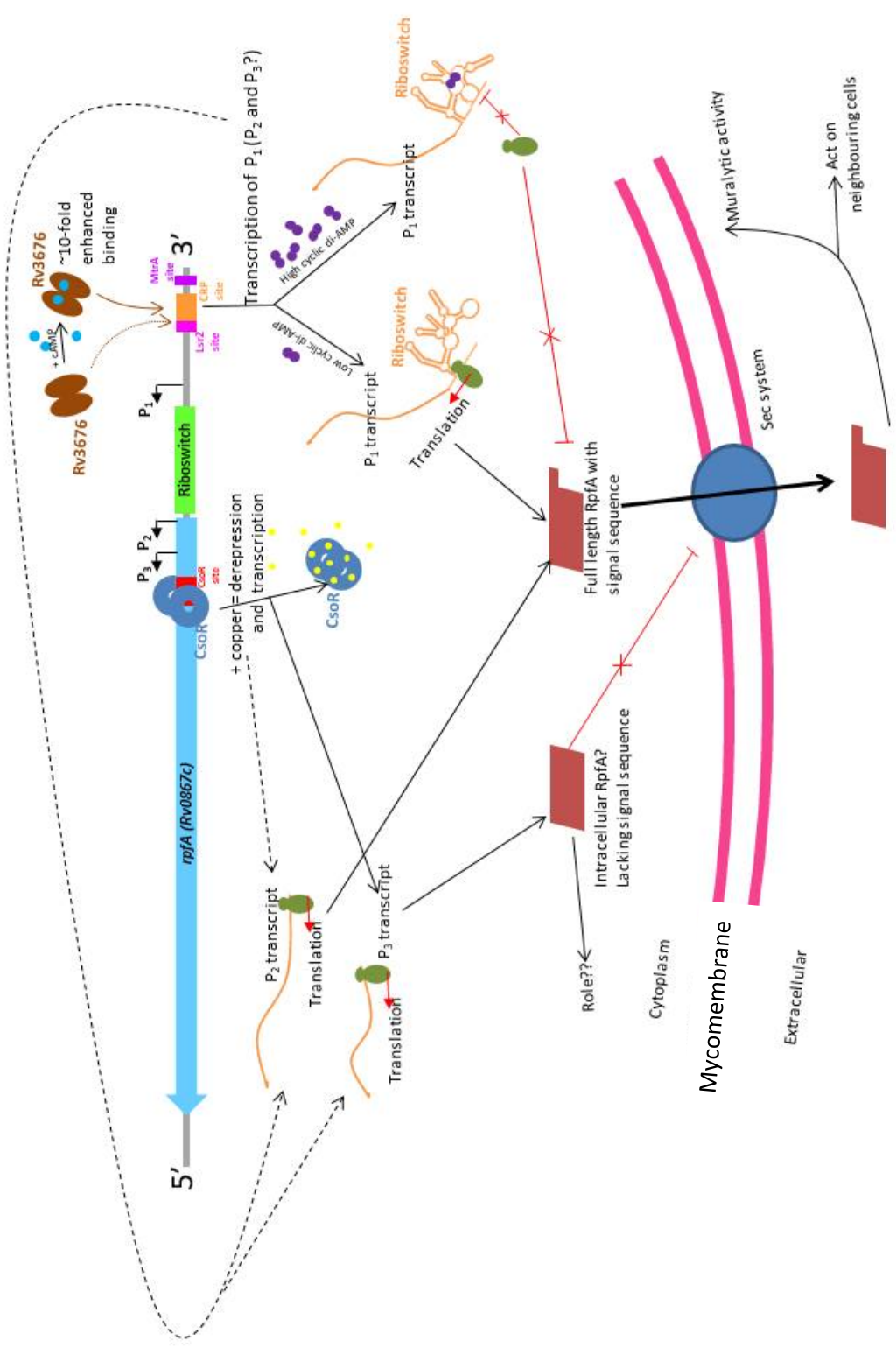
Figure 6.1: An updated overview of the regulation of *rpfA* in *M. tuberculosis*

Figure 6.1: An overview of the complex regulation of the *rpfA* gene in *M. tuberculosis*. The regulation of *rpfA* is complex and there are a plethora of regulatory mechanisms and signals involved. This figure presents a current overview of the complex regulation of *rpfA* in *M. tuberculosis,* taking into account the data gathered during this study and other published data. There are still numerous questions that need to be addressed, however the study has gone someway to increase the overall knowledge into the mechanisms by which *rpfA* is regulated. Firstly, Rv3676 (shown in brown) is able to bind to the CRP site located upstream of all the transcript starts ($P_{1 \text{ to } 3}$). The presence of cyclic AMP enhances DNA-binding approximately 10-fold, primarily via an increase in the on rate constant. Rv3676 acts as a transcriptional repressor and we suggest that binding of Rv3676 plays a role in the regulation of generating *rpfA* transcripts possibly from all three promoters (P1, P2 and P3). P1 is the full length transcript, which contains the riboswitch in the 5' UTR of the messenger RNA. We propose that this riboswitch provides a further level of regulation at a post-transcriptional level. We suggest that cyclic di-AMP and potentially pApA act as ligands for the aptamer of the riboswitch and binding of these ligands causes a shift towards specific favoured structural conformations. Based on predicted structures and the lack of intrinsic terminators present in the *rpfA* riboswitch, we propose ligand binding promotes interaction between the pseudoknot and the RBS, preventing translation of *rpfA*. The CsoR site is located within the coding region of *rpfA* and downstream of all 3 transcript start sites. CsoR is able to bind to this site and act as a copper-responsive repressor which, due to its location is most likely involved in the regulation of transcript P3, but could repress transcription from P1 and P2 also. Transcription from P3 would result in translation of RpfA that lacks the N-terminal region (including the signal sequence that permits Sec-dependent secretion) of the full-length protein. There is evidence suggesting that a shortened cytoplasmic version of RpfA lacking the N-terminal signal sequence exists (unpublished). This is an interesting concept, which needs to be explored further experimentally (discussed later in Section 6.3). In addition to the three regulatory mechanisms outlined in Chapters 3, 4, and 5, there are other regulators proposed to be involved in the regulation of *rpfA*. This includes Lsr2, a nucleoid-associated transcriptional regulator that is involved in adaption to changing oxygen levels and virulence in *M. tuberculosis* (Bartek et al., 2014). ChIP-seq data revealed a putative Lsr2 binding site located upstream of the *rpfA* transcript starts (Cortes et al., 2013). Lsr2 was found to represses a large cohort of genes induced in *M. tuberculosis* during *in vitro* models of latency, including genes implicated in persister formation (Gordon, 2012). The *rpf* genes are downregulated in the Δ*lsr2* mutant with *rpfA* and *rpfB* expression levels ≤2-fold that of WT, providing further evidence that *rpfA* is regulated by Lsr2 (Gordon, 2012).

Further investigation into the different transcription start sites (P_1, P_2 and P_3) would also be useful. *In vitro* transcription assays could be performed with a mycobacterial RNAP, the different regulatory proteins and/or ligands as a way to determine which transcripts are regulated by which proteins and ligands. This would provide further evidence to validate the proposed overview of *rpfA* regulation shown in Figure 6.1.

6.1 Rv3676 regulation and the broader cyclic AMP picture

Gel shift experiments suggested that the interaction between Rv3676 and the CRP site present in *rpfA* was independent of cyclic AMP (Rickman et al., 2005). In 2010, Stapleton et al. performed studies characterising the interaction between the *whiB1* gene and Rv3676 and suggested that binding by Rv3676 to the promoter was enhanced by cyclic AMP by ~5-fold. During this study we have generated kinetic and thermodynamic data that demonstrates cyclic AMP enhances Rv3676 DNA-binding ~10-fold and the majority of this enhancement occurs due to increase in the on rate constant. In *E. coli* CRP, cyclic AMP-binding enhances DNA-binding by as much as ~10^6-fold. The crystal structures of apo- and holo-CRP reveal that when no cyclic AMP is bound the DNA-binding helices are buried beneath the proteins surface, and are unable to interact with the DNA. Upon cAMP-binding, *E. coli* CRP undergoes a significant conformational change, which exposes the DNA-binding helices and allows them to interact with the consensus sequence. We propose that the 10-fold enhancement in DNA-binding observed in holo-Rv3676 compared to apo-Rv3676 is due to realignment of the DNA-binding helices rather than a larger conformational change, such as that observed in *E. coli* CRP.

Despite having a similar structure to Rv3676, Msmeg_6189 appeared to behave more like *E. coli* CRP in terms of the dependence on cyclic AMP for DNA-binding, but more like Rv3676 in terms of the thermodynamics and kinetics of the interaction with cyclic AMP. Experiments with the *sdh1* promoter revealed that cyclic AMP was necessary for Msmeg_6189 to interact. Msmeg_6189 had a similar affinity for cyclic AMP to Rv3676. Other experiments need to be performed with the trajectory mutants (changing the Rv3676 sequence towards the Msmeg_6189) in an attempt to determine exactly which residue(s) may be important for this cyclic AMP dependence and the mechanism by which it is achieved. These mutants should first be screened by EMSA to assess their dependence on cyclic AMP, and if dependence on cyclic AMP observed in a mutant then further characterisation of the structural and thermodynamic parameters would be of interest. Various parameters of the proteins are compared in Table 6.1.

Table 6.1: Comparison of CRP proteins from different species

	Rv3676	*E. coli* CRP	Msmeg_6189
K_D for cAMP	160 to 300 µM	13-16 µM	~300 µM
Interaction type	Independent binding	Cooperative binding	Independent binding
cAMP required for specific DNA binding	No, but slight enhancement (~2 to 10-fold and primarily due to increase in the on rate constrant)	Yes (>1000-fold)	Yes
Intracellular [cAMP]	High	Moderate	High
Bind cGMP?	No	Yes - with lower affinity	No

6.2 COPASI modelling to compare *M. tuberculosis* Rv3676 with *E. coli* CRP

A model to allow for comparison of cyclic AMP-independent Rv3676 with the well-characterised cyclic AMP-dependent *E. coli* CRP paradigm may provide greater insight on the bigger picture of cyclic AMP signalling. An overview of the model is shown in Figure 4.23. The data for the rate values for both Rv3676 and *E. coli* CRP would be based on data generated during this study combined with values in other published studies. The model would allow a user to input different intracellular cyclic-AMP concentrations for model cells containing the two different CRP proteins and would model the outputs in response (i.e. changes in transcription and proportion of protein in each reaction). This would be a useful tool to aid understanding the range of cAMP concentrations within which Rv3676 and CRP can operate effectively as switches at target genes with different binding affinities for the regulator. Such a model should generate experimentally testable hypotheses.

Interestingly Rv3676 is not functional as an activator in *E. coli.* Rv3676 function was tested in an *E. coli lacZ* reporter system and the protein was unable to activate the CRP-dependent promoter (CC) (Spreadbury et al., 2005). This might be due to the possibly because the mycobacterial protein was unable to interact productively with the *E. coli* RNA polymerase, but could also be affected by the comparatively low levels of cyclic AMP present intracellular in *E. coli* compared to mycobacteria. It would be interesting to test if Msmeg_6189 is functional in *E. coli* by similar methods to those outlined in Spreadbury et al. (2005).

6.3 Mode of action of *M. tuberculosis rpfA* riboswitch

The model of the *M. tuberculosis rpfA* riboswitch was generated by aligning its sequence to a covariance model of the *B. subtilis ydaO* riboswitch (Figure 5.19). Bases associated with direct or indirect interaction with cyclic di-AMP were conserved (black arrows) and similarly, bases contributing to the pseudoknot are conserved (filled arcs). Single base substitutions (A16G, A16U, A17G, A17U, A128G, A128U, A129G, A199G, A199U, A207G and A207U; black arrows) based on this gel shift analysis did not prevent the cyclic di-AMP-mediated change in migration observed with the wild-type riboswitch RNA (Section 5.4.4.1). In contrast, single base replacements of bases that directly interact with cyclic di-AMP in the *ydaO* riboswitch, equivalent to positions 16, 128 and 207 in the *rpfA* riboswitch, abolished cyclic di-AMP binding (Jones and Ferré-D'Amaré 2014). However, cyclic di-AMP failed to resolve the structural diversity of the *rpfA* riboswitch when G168 and G169 (red arrows in Figure 5.19) were replaced by C (Section 5.19). This indicates that disrupting the P7 helix impairs the ability of cyclic di-AMP to stabilize the *M. tuberculosis rpfA* pseudoknot conformation. This is response is also

different to the *ydaO* riboswtich where deletion of the residues contributing to the pseudoknot did not change the affinity for cyclic di-AMP (Jones and Ferré-D'Amaré 2014). Thus, it appears that the *ydaO* and *rpfA* riboswitches interact with cyclic di-AMP differently and that this could be related to their different modes of action; the *ydaO* riboswitch operates at the level of transcription, whereas the *rpfA* riboswitch likely operates at the level of translation.

6.4   ATP and *ydaO* class riboswitches

Interestingly, the K_D for the interaction of ATP with *ydaO* from *B. subtilis* is estimated to be ~10 nM (Watson and Fedor, 2012), which is significantly weaker than the K_D of interaction with cyclic di-AMP (Nelson et al., 2013). It has not been possible to detect binding between ATP and the *ydaO* representatives from *Syntrophus acidictrophicus*, *Clostridium acetobutylicum* and *Nostoc punctiforme,* even at concentrations of ATP as high as 1 mM (Nelson et al., 2013)*.* During this study, no interaction was observed between ATP and the *rpfA* riboswitch of *M. tuberculosis*. It is possible that the *B. subtilis ydaO* is an exception, and it is not typical for members of the *ydaO* class of riboswitch to interact with ATP. It is plausible that the *B. subtilis* riboswitch may respond to two distinct signals, ATP and cyclic di-AMP, and regulate *ydaO* in response to these different signalling pathways.

If this were the case, the lower affinity of the riboswitch for ATP that was observed in Watson and Fedor (2013) would be physiologically plausible and logical, as ATP is widespread throughout the cell and functions as the major cellular energy 'currency'. If the affinity for ATP was as tight as determined for c-di-AMP (picomolar range) then the vast majority of the riboswitches present in a *B. subtilis* cell would be in the ligand-bound form and there would be little scope or sensitivity to changes in regulation from changes in ATP levels. It would be interesting to investigate this interplay further and to make efforts to determine the structural effects caused by ATP-binding and to determine if high cellular ATP concentrations affect the interaction between cyclic di-AMP and the riboswitch. This could be achieved by performing gel shift experiments with combinations of radio-labelled ATP, unlabelled cyclic di-AMP and biotinylated RNA and determining migration of the different elements.

6.5   Riboswitch in the *rpfA* gene of *Streptomyces coelicolor*

During the writing of this thesis, a study was published by St-Onge and Elliot (2017) identifying a cyclic di-AMP responsive riboswitch in the 5' UTR of *rpfA* messenger RNA in *Steptomyces coelicolor*. Both streptomycetes and mycobacteria belong to the actinomycetes genus of bacteria and for both groups the *rpfA* gene plays a role in entry to and exit from dormancy (Block et al., 2010; St-Onge et al., 2017). Dormamcy in *Mycobacterium tuberculosis* involves

forming a 'cyst-like' granuloma structure within the host. In contrast, streptomycetes species form spores during dormancy. Despite the differences in the physical dormant states of the two groups, they both require remodelling of the cell wall during the entry phase and resuscitation phase of dormancy, which is when active growth resumes (Sleeva et al., 2011).

Similar to *M. tuberculosis*, the regulation of *rpfA* is tightly-controlled in *S. coelicolor*. Alike to our predictions based on experimental observations during the osmotic stress recovery experiments (Sections 5.1.7 and Figure 5.22), it is suggested that the *rpfA* riboswitch in *S. coelicolor* negatively impacts the *rpfA* expression in response to the second messenger cyclic di-AMP. The non-coding RNA regulation of the *rpfA* gene in *S. coelicolor* is more complex than in *M. tuberculosis* due to the presence of a second antisense small RNA (sRNA), Scr3097, downstream of the stop codon in the flanking region of the mRNA as well as the upstream riboswitch (St-Onge and Elliot, 2017).

Both the *S. coelicolor* and mycobacterial *rpfA* riboswitches lack the intrinsic transcription terminator present in the *ydaO* riboswitches of *B. subtilis* and *Thermoanaerobacter tengcongensis* (St-Onge and Elliot, 2017) and there is no obvious expression platform. The exact mechanism by which the *S. coelicolor rpfA* riboswitch switches between the 'on' and 'off' state remains unknown (St-Onge and Elliot, 2017).

The main findings proposed by the study included the notion that, in *S. coelicolor*, riboswitch-dependent attenuation of *rpfA* expression is inversely proportional to intracellular cyclic di-AMP levels. Experiments were performed using the riboswitch sequence fused to a luciferase reporter system and monitoring luciferase during different stages of growth alongside determining intracellular cyclic di-AMP throughout (where possible). It would be useful to perform such experiments with *M. tuberculosis* and the during the different stages of growth and resuscitation and the various constructs shown in Figure 5.3. Samples could be taken at different phases to allow for determination of the cyclic-di-AMP concentration and also determination of the reporter luciferase could be monitored by methods based on those in St-Onge and Elliot (2017). This would provide insight into the *in vivo* relationship between cyclic di-AMP levels and gene expression resulting from regulation by the *M. tuberculosis* riboswitch. Western blotting and qRT-PCR could be performed to provide quantitative determination of transcript and protein levels.

St-Onge and Elliot (2017) demonstrated, via in-line probing experiments, that cyclic di-AMP binds to the *S. coelicolor rpfA* riboswitch and induces structural changes resulting in stabilisation of some sections of the RNA (P1, P1-P2 link, P2 bulge, P3 stem, the region bridging P3 and P5 and the L7 loop) and increased cleavage of some regions. They propose that cyclic

di-AMP promotes general transcript degradation and that the riboswitch may regulate transcript turnover and stability via recruitment of specific RNases to the messenger RNA in response to cyclic di-AMP binding. '*In-vitro* stability assays' were performed, where *rpfA* transcripts were incubated with cell-free lysate (source of RNases) in the presence or absence of cyclic di-AMP and levels of transcript degradation were compared. It is proposed that greater levels of degradation occurred in samples containing cyclic di-AMP. However, this result is not convincing and arguably there is a greater total quantity of RNA loaded in the lane containing cyclic di-AMP. A loading control would remove this doubt.

It may be interesting to perform these experiments with the *M. tuberculosis* riboswitch, but based on the structural model (Figure 5.19), we propose that the mechanism of action of the riboswitch is through via alterations in the interaction between the pseudoknot and the RBS that prevents translation when cyclic di-AMP is bound. The paper also suggests that cyclic di-AMP may have an unexplored effect on the stability of the *rpfA* transcript in a riboswitch-independent manner.

6.6 CsoR

It has been suggested that CsoR plays more than one role in the cell – not just as a transcriptional repressor, but it plays a buffering role in preventing damage from free copper (Liu et al., 2007b). Marcus et al. (2016) suggests that the genes within the *cso* regulon are directly involved in mitigating the harmful effects of copper stress. Survival of ∆*csoR* at high copper (500 µM) are comparable to *M. tuberculosis* strain overexpressing the copper exporter *ctpV* (Ward et al., 2010). A transcriptional profile of a ∆*csoR* strain showed signs of stress, even at relatively low levels of copper, supporting the notion that CsoR is important for maintaining copper homeostasis (Marcus et al., 2016).

The location of the CsoR binding site is downstream of the P1, P2 and P3 TSS. There are no other suitable transcript start sites in adjacent genes that are in positions that could be regulated by CsoR. It is plausible that CsoR is involved in the regulation of production of a shorter, truncated version of RpfA, which is lacking the N-terminal region that contains a signal sequence which allows export from the cytoplasm via the Sec system. Generating evidence to confirm the existence of cytoplasmic RpfA is important. This could be achieved by introducing a C-terminal fluorescent tag in the *rpfA* coding region and determining the location of the RpfA proteins, whether they are in the supernatant or in the cell. Another approach could be to use RpfA antibodies and western blotting to identify whether significant levels of RpfA are present in the cytoplasm.

The *cso* operon includes the *csoR* gene, which plays a role in *M. tuberculosis* dormancy and induction of *cso* operon is observed during the enduring hypoxic response (Rustad et al., 2008). This hypoxia-specific increase in copper response may be physiologically important for more than one reason. Firstly, hypoxia leads to an increase in copper uptake by macrophages (White et al., 2009) and secondly copper is is more toxic under hypoxic conditions (Beswick et al., 1976b). During dormancy when in the granuloma, *M. tuberculosis* is exposed to high copper concentrations from the host. This means that CsoR will likely be in the copper-bound form and released from the site in the *rpfA* gene, causing depression. At this point it would not be desirable for *M. tuberculosis* to resuscitate as conditions are not favourable. This is when the theory about the second 'cytoplasmic' version of RpfA is especially interesting. If CsoR is involved in the regulation of a cytoplasmic version of the protein only, then high copper levels during extended hypoxia will only lead to increases in the cytoplasmic RpfA, rather than the extracellular one. The function of this cytoplasmic protein has not been confirmed, but neither of the main functions carried out by 'normal' extracellular RpfA (muralytic activity on the cell wall and signalling to nearby cells to resuscitate) would likely be able to be performed by this protein. Further study into the exact role and biological relevance of this cytoplasmic protein would provide further insight into why it is important that *rpfA* is able to be regulated in response to copper levels.

6.7   Interplay between the regulatory methods

As it has already been suggested, there is plethora of different mechanism and signals involved in the regulation of *rpfA*. Extensive work has been done during this study and by others on characterising the various independent elements. However little research has been done to try to characterise the interplay between the different regulatory mechanisms and to ascertain under what cellular conditions

The role of CsoR in regulating *rpfA* was explored in Chapter 3 and the role of Rv3676 in the regulation of *rpfA* was explored in Chapter 4 of this study. It could be interesting to in assess if binding of one regulatory protein to its site in the *rpfA* gene has any effect on the interaction of the other regulatory protein to its site. Performing PCR using primers Myc657 (Forward primer for amplification of region containing CRP site) and LD68 (reverse primer for amplification of region containing the CsoR site) would allow for amplification of a region of *rpfA* containing both regulatory sites. Binding experiments could be performed to look at the effect of binding of both proteins in the presence of the other and different concentrations of copper and cyclic AMP. In vitro transcription experiments could be performed with

mycobacterial RNAP and different concentrations of the regulatory proteins, copper and/or cyclic AMP and the effects on transcription of the *rpfA* transcripts could be monitored.

6.8   Concluding remarks

The regulation of the *rpfA* gene involves a several different signalling molecules and different regulatory mechanisms, both RNA-based and protein-based and at a transcriptional and post-transcriptional level. This study focused mainly on regulation by the *ydaO*-type riboswitch present in the 5' UTR of the *rpfA* mRNA, Rv3676 and CsoR and the summaries of the findings can be found in Sections 3.6, 4.7.1 and 5.7 respectively.

7   References

Abramovitch, R.B., Rohde, K.H., Hsu, F.F., and Russell, D.G. (2011). aprABC: A *Mycobacterium tuberculosis* complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. Mol. Microbiol. 80, 678–694.

Adams, L.B., Dinauer, M.C., Morgenstern, D.E., and Krahenbuhl, J.L. (1997). Comparison of the roles of reactive oxygen and nitrogen intermediates in the host response to *Mycobacterium tuberculosis* using transgenic mice. Tuber. Lung Dis. 78, 237–246.

Agarwal, N., Lamichhane, G., Gupta, R., Nolan, S., and Bishai, W.R. (2009). Cyclic AMP intoxication of macrophages by a *Mycobacterium tuberculosis* adenylate cyclase. Nature 460, 98–102.

Anuchin, A.M., Mulyukin, A.L., Suzina, N.E., Duda, V.I., El-Registan, G.I., and Kaprelyants, A.S. (2009). Dormant forms of *Mycobacterium smegmatis* with distinct morphology. Microbiology 155, 1071–1079.

Asselineau, J., and Lederer, E. (1950). Structure of the Mycolic Acids of Mycobacteria. Nature 166, 782–783.

Aung, H.L. in, Dixon, L.L., Smith, L.J., Sweeney, N.P., Robson, J.R., Berney, M., Buxton, R.S., Green, J., and Cook, G.M. (2015). Novel regulatory roles of cAMP receptor proteins in fast-growing environmental mycobacteria. Microbiology 161, 648–661.

Axelrod, S., Oschkinat, H., Enders, J., Schlegel, B., Brinkmann, V., Kaufmann, S.H.E., Haas, A., and Schaible, U.E. (2008). Delay of phagosome maturation by a mycobacterial lipid is reversed by nitric oxide. Cell. Microbiol. 10, 1530–1545

Bai, G., McCue, L.A., McDonough, K. a, and Acteriol, J.B. (2005). Characterization of Mycobacterium tuberculosis Rv3676 (CRP Mt), a Cyclic AMP Receptor Protein-Like DNA Binding Protein. Society 187, 7795–7804.

Bai, G., Knapp, G.S., and McDonough, K.A. (2011). Cyclic AMP signalling in mycobacteria: Redirecting the conversation with a common currency. Cell. Microbiol. 13, 349–358.

Bai, Y., Yang, J., Zhou, X., Ding, X., Eisele, L.E., and Bai, G. (2012). *Mycobacterium tuberculosis* Rv3586 (DacA) is a diadenylate cyclase that converts ATP or ADP into c-di-AMP. PLoS One 7, e35206.

Baker, D.A., and Kelly, J.M. (2004). Structure, function and evolution of microbial adenylyl and guanylyl cyclases. Mol. Microbiol. 52, 1229–1242.

Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., et al. (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc. Natl. Acad. Sci. 101, 6421–6426.

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

Bartek, I.L., Woolhiser, L.K., Baughn, A.D., Basaraba, R.J., Jacobs, W.R., Lenaerts, A.J., and Voskuil, M.I. (2014). *Mycobacterium tuberculosis* Lsr2 is a global transcriptional regulator required for adaptation to changing oxygen levels and virulence. MBio 5, e01106-14.

Bateman, A., Holden, M.T.G., and Yeats, C. (2005). The G5 domain: A potential N-acetylglucosamine recognition domain involved in biofilm formation. Bioinformatics 21, 1301–1303.

Bejerano-Sagie, M., Oppenheimer-Shaanan, Y., Berlatzky, I., Rouvinski, A., Meyerovich, M., and Ben-Yehuda, S. (2006). A checkpoint protein that scans the chromosome for damage at the start of sporulation in *Bacillus subtilis*. Cell 125, 679–690.

Berg, O.G., and von Hippel, P.H. (1988). Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. J. Mol. Biol. 200, 709–723.

Beswick, P.H., Hall, G.H., Hook, A.J., Little, K., McBrien, D.C., and Lott, K.A. (1976a). Copper toxicity: evidence for the conversion of cupric to cuprous copper in vivo under anaerobic conditions. Chem. Biol. Interact. 14, 347–356.

Block, K.F., Hammond, M.C., and Breaker, R.R. (2010). Evidence for widespread gene control function by the *ydaO* riboswitch candidate. J. Bacteriol. 192, 3983–3989.

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

Bocobza, S.E., and Aharoni, A. (2014). Small molecules that interact with RNA: Riboswitch-based gene control and its involvement in metabolic regulation in plants and algae. Plant J. 79, 693–703.

Botella, H., Peyron, P., Levillain, F., Poincloux, R., Poquet, Y., Brandli, I., Wang, C., Tailleux, L., Tilleul, S., Charrire, G.M., et al. (2011). Mycobacterial P 1-Type ATPases mediate resistance to Zinc poisoning in human macrophages. Cell Host Microbe 10, 248–259.

Bradley, S.G., and Ritzi, D. (1968). Composition and ultrastructure of *Streptomyces venezuelae*. J. Bacteriol. 95, 2358–2364.

Brennan, P.J., and Nikaido, H. (1995). The Envelope of Mycobacteria. Annu. Rev. Biochem. 64, 29–63.

Brocker, M., Mack, C., and Bott, M. (2011). Target Genes, Consensus Binding Site, and Role of Phosphorylation for the Response Regulator MtrA of Corynebacterium glutamicum. J. Bacteriol. 193, 1237–1249.

Burgess, R.R. (2009). Chapter 20 Protein Precipitation Techniques. Methods Enzymol. 463, 331–342.

Busby, S., and Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. 293, 199–213.

Cambier, C.J., Falkow, S., and Ramakrishnan, L. (2014). Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. Cell 159, 1497–1509.

Campeotto, I., Zhang, Y., Mladenov, M.G., Freemont, P.S., and Gründling, A. (2015). Complex structure and biochemical characterization of the *Staphylococcus aureus* cyclic diadenylate monophosphate (c-di-AMP)-binding protein PstA, the founding member of a new signal transduction protein family. J. Biol. Chem. 290, 2888–2901.

Cavet, J.S., Graham, A.I., Meng, W., and Robinson, N.J. (2003). A cadmium-lead-sensing ArsR-SmtB repressor with novel sensory sites. Complementary metal discrimination by NmtR and CmtR in a common cytosol. J. Biol. Chem. 278, 44560–44566.

CDC (2016). CDC | TB | Treatment | Treatment for TB Disease.

Cellier, M.F.M. (2012). Nutritional immunity: Homology modelling of nramp metal import. Adv. Exp. Med. Biol. 946, 335–351.

Cescau, S., Cwerman, H., Létoffé, S., Delepelaire, P., Wandersman, C., and Biville, F. (2007). Heme acquisition by hemophores. In BioMetals, pp. 603–613.

Chang, F.-M.J., Lauber, M.A., Running, W.E., Reilly, J.P., and Giedroc, D.P. (2011). Ratiometric Pulse–Chase Amidination Mass Spectrometry as a Probe of Biomolecular Complex Formation. Anal. Chem. 83, 9092–9099.

Chao, M.C., and Rubin, E.J. (2010). Letting Sleeping Dogs Lie: Does Dormancy Play a Role in Tuberculosis? Annu. Rev. Microbiol. 64, 293–311.

Chen, L., Xie, Q.W., and Nathan, C. (1998). Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. Mol. Cell 1, 795–805.

Clifton E. Barry, Dean C. Crick, and Michael R. McNeil (2007). Targeting the Formation of the Cell Wall Core of *M. tuberculosis*. Infect. Disord. - Drug Targets 7, 182–202.

Cohen, K.A., Abeel, T., Manson McGuire, A., Desjardins, C.A., Munsamy, V., Shea, T.P., Walker, B.J., Bantubani, N., Almeida, D. V., Alvarado, L., et al. (2015). Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of *Mycobacterium tuberculosis* Isolates from KwaZulu-Natal. PLoS Med. 12, e1001880.

Cohen-Gonsaud, M., Keep, N.H., Davies, A.P., Ward, J., Henderson, B., and Labesse, G. (2004). Resuscitation-promoting factors possess a lysozyme-like domain. Trends Biochem. Sci. 29, 7–10.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V, Eiglmeier, K., Gas, S., Barry, C.E., et al. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393, 537–544.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. 45, 1176–1182.

Corrigan, R.M., and Gründling, A. (2013). Cyclic di-AMP: another second messenger enters the fray. Nat. Rev. Microbiol. 11, 513–524.

Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., and Young, D.B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. Cell Rep. 5, 1121–1131.

Daffé, M., and Draper, P. (1997). The Envelope Layers of Mycobacteria with Reference to their Pathogenicity. Advances In Microbial Physiology 39, 131-202.

Daniel, J., Deb, C., Dubey, V.S., Sirakova, T.D., Abomoelak, B., Morbidoni, H.R., and Kolattukudy, P.E. (2004). Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in *Mycobacterium tuberculosis* as it goes into a dormancy-like state in culture. J. Bacteriol. 186, 5017–5030.

Dann, C.E., Wakeman, C.A., Sieling, C.L., Baker, S.C., Irnov, I., and Winkler, W.C. (2007). Structure and Mechanism of a Metal-Sensing Regulatory RNA. Cell 130, 878–892.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics 25, 1974–1975.

Dasgupta, A., Datta, P., Kundu, M., and Basu, J. (2006). The serine/threonine kinase PknB of *Mycobacterium tuberculosis* phosphorylates PBPA, a penicillin-binding protein required for cell division. Microbiology 152, 493–504.

Delfosse, V., Bouchard, P., Bonneau, E., Dagenais, P., Lemay, J.F., Lafontaine, D.A., and Legault, P. (2009). Riboswitch structure: An internal residue mimicking the purine ligand. Nucleic Acids Res. 38, 2057–2068.

Dewi Puspita, I., Uehara, M., Katayama, T., Kikuchi, Y., Kitagawa, W., Kamagata, Y., Asano, K., Nakatsu, C.H., and Tanaka, M. (2013). Resuscitation promoting factor (Rpf) from *Tomitella biformata* AHU 1821(T) promotes growth and resuscitates non-dividing cells. Microbes Environ. 28, 58–64.

Domek, M.J., LeChevallier, M.W., Cameron, S.C., and McFeters, G.A. (1984). Evidence for the role of copper in the injury process of coliform bacteria in drinking water. Appl. Environ. Microbiol. 48, 289–293.

Dubnau, E., Chan, J., Raynaud, C., Mohan, V.P., Lanéelle, M.A., Yu, K., Quémard, A., Smith, I., and Daffé, M. (2000). Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. Mol. Microbiol. 36, 630–637.

Dussurget, O., Stewart, G., Neyrolles, O., Pescher, P., Young, D., and Marchal, G. (2001). Role of *Mycobacterium tuberculosis* copper-zinc superoxide dismutase. Infect. Immun. 69, 529–533.

Dwarakanath, S., Chaplin, A.K., Hough, M.A., Rigali, S., Vijgenboom, E., and Worrall, J.A.R. (2012). Response to copper stress in Streptomyces lividans extends beyond genes under direct control of a copper-sensitive operon repressor protein (CsoR). J. Biol. Chem. 287, 17833–17847.

Dworkin, J., and Shah, I.M. (2010). Exit from dormancy in microbial organisms. Nat. Rev. Microbiol. 8, 890–896.

Dye, C. (1999). Global Burden of Tuberculosis - Estimated Incidence, Prevalence, and Mortality by Country. JAMA 282, 677.

Eldholm, V., and Balloux, F. (2016). Antimicrobial Resistance in *Mycobacterium tuberculosis*: The Odd One Out. Trends Microbiol. 24, 637–648.

Eldholm, V., Monteserin, J., Rieux, A., Lopez, B., Sobkowiak, B., Ritacco, V., and Balloux, F. (2015). Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. Nat. Commun. 6, 7119.

Fang, F.C., DeGroote, M.A., Foster, J.W., Bäumler, A.J., Ochsner, U., Testerman, T., Bearson, S., Giárd, J.C., Xu, Y., Campbell, G., et al. (1999). Virulent *Salmonella typhimurium* has two periplasmic Cu, Zn-superoxide dismutases. Proc. Natl. Acad. Sci. U. S. A. 96, 7502–7507.

Fennelly, K.P., Martyny, J.W., Fulton, K.E., Orme, I.M., Cave, D.M., and Heifets, L.B. (2004). Cough-generated Aerosols of *Mycobacterium tuberculosis*. Am. J. Respir. Crit. Care Med. 169, 604–609.

Festa, R.A., Jones, M.B., Butler-Wu, S., Sinsimer, D., Gerads, R., Bishai, W.R., Peterson, S.N., and Darwin, K.H. (2011). A novel copper-responsive regulon in *Mycobacterium tuberculosis*. Mol. Microbiol. 79, 133–148.

Finch, C.A., and Huebers, H. (1982). Perspectives in Iron Metabolism. N. Engl. J. Med. 306, 1520–1528.

Fratti, R.A., Chua, J., Vergne, I., and Deretic, V. (2003). *Mycobacterium tuberculosis* glycosylated phosphatidylinositol causes phagosome maturation arrest. Proc. Natl. Acad. Sci. U. S. A. 100, 5437–5442.

Furukawa, K., Ramesh, A., Zhou, Z., Weinberg, Z., Vallery, T., Winkler, W.C., and Breaker, R.R. (2015). Bacterial Riboswitches Cooperatively Bind Ni2+ or Co2+ Ions and Control Expression of Heavy Metal Transporters. Mol. Cell 57, 1088–1098.

Gagneux, S. (2012). Host-pathogen coevolution in human tuberculosis. Philos. Trans. R. Soc. B Biol. Sci. 367, 850–859.

Gao, A., and Serganov, A. (2014). Structural insights into recognition of c-di-AMP by the *ydaO* riboswitch. Nat. Chem. Biol. 10, 787–792.

Gengenbacher, M., and Kaufmann, S.H. (2012). *Mycobacterium tuberculosis*: success through dormancy. FEMS Microbiol. Rev. 36, 514–532.

Gilbert, S.D., Rambo, R.P., Van Tyne, D., and Batey, R.T. (2008). Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat. Struct. Mol. Biol. 15, 177–182.

Glauert, A.M. (1961). The fine structure of *Streptomyces violaceoruber* (*S. coelicolor*): III. The Walls of the Mycelium and Spores. J. Cell Biol. 10, 505–516.

Glickman, M.S., Cox, J.S., and Jacobs, W.R. (2000). A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of *Mycobacterium tuberculosis*. Mol. Cell 5, 717–727.

Gold, B., Deng, H., Bryk, R., Vargas, D., Eliezer, D., Roberts, J., Jiang, X., and Nathan, C. (2008). Identification of a copper-binding metallothionein in pathogenic mycobacteria. Nat. Chem. Biol. 4, 609–616.

Gordon, R.G.B. (2012). Lsr2: An H-NS functional analog and global regulator of *Mycobacterium tuberculosis*. University of Toronto.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A., and Saier, M.H. (2004). Transcriptome Analysis of Crp-Dependent Catabolite Control of Gene Expression in *Escherichia coli*. J. Bacteriol. 186, 3516–3524.

Gouzy, A., Larrouy-Maumus, G., Wu, T.-D., Peixoto, A., Levillain, F., Lugo-Villarino, G., Gerquin-Kern, J.-L., de Carvalho, L.P.S., Poquet, Y., and Neyrolles, O. (2013). *Mycobacterium tuberculosis* nitrogen assimilation and host colonization require aspartate. Nat. Chem. Biol. 9, 674–676.

Gouzy, A., Larrouy-Maumus, G., Bottai, D., Levillain, F., Dumas, A., Wallach, J.B., Caire-Brandli, I., de Chastellier, C., Wu, T. Di, Poincloux, R., et al. (2014). *Mycobacterium tuberculosis* Exploits Asparagine to Assimilate Nitrogen and Resist Acid Stress during Infection. PLoS Pathog. 10, e1003928.

Green, J., Stapleton, M.R., Smith, L.J., Artymiuk, P.J., Kahramanoglou, C., Hunt, D.M., and Buxton, R.S. (2014). Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: Adaptation for different ecological niches. Curr. Opin. Microbiol. 18, 1–7.

Guedich, S., Puffer-Enders, B., Baltzinger, M., Hoffmann, G., Da Veiga, C., Jossinet, F., Thore, S., Bec, G., Ennifar, E., Burnouf, D., et al. (2016). Quantitative and predictive model of kinetic regulation by *E. coli* TPP riboswitches. RNA Biol. 13, 373–390.

Gülbay, B.E., Gürkan, Ö.U., Yildiz, Ö.A., Önen, Z.P., Erkekol, F.O., Baççioğlu, A., and Acican, T. (2006). Side effects due to primary antituberculosis drugs during the initial phase of therapy in 1149 hospitalized patients for tuberculosis. Respir. Med. 100, 1834–1842.

Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P., and Vincent, V. (2005). Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 1, 0055–0061.

Haiser, H.J., Yousef, M.R., and Elliot, M.A. (2009). Cell Wall Hydrolases Affect Germination, Vegetative Growth, and Sporulation in *Streptomyces coelicolor*. J. Bacteriol. 191, 6501–6512.

Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W., et al. (2008). High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. 6, 2658–2671.

Hett, E.C., and Rubin, E.J. (2008). Bacterial Growth and Cell Division: a Mycobacterial Perspective. Microbiol. Mol. Biol. Rev. 72, 126–156.

Hett, E.C., Chao, M.C., Steyn, A.J., Fortune, S.M., Deng, L.L., and Rubin, E.J. (2007). A partner for the resuscitation-promoting factors of *Mycobacterium tuberculosis*. Mol. Microbiol. 66, 658–668.

Higgins, K.A., and Giedroc, D. (2014). Insights into Protein Allostery in the CsoR/RcnR Family of Transcriptional Repressors. Chem. Lett. 43, 20–25.

Hillas, P.J., Soto Del Alba, F., Oyarzabal, J., Wilks, A., and Ortiz De Montellano, P.R. (2000). The AhpC and AhpD antioxidant defense system of *Mycobacterium tuberculosis*. J. Biol. Chem. 275, 18801–18809.

Imaeda, T. (1985). Deoxyribonucleic Acid Relatedness Among Selected Strains of. Int. J. Syst. Bacteriol. 35, 147–150.

Jacob, C., Maret, W., and Vallee, B.L. (1998). Ebselen, a selenium-containing redox drug, releases zinc from metallothionein. Biochem. Biophys. Res. Commun. 248, 569–573.

Jacobs, A.D., Chang, F.-M.J., Morrison, L., Dilger, J.M., Wysocki, V.H., Clemmer, D.E., and Giedroc, D.P. (2015). Resolution of Stepwise Cooperativities of Copper Binding by the Homotetrameric Copper-Sensitive Operon Repressor (CsoR): Impact on Structure and Stability. Angew. Chemie Int. Ed. 54, 12795–12799.

Jansen, C., Gronenborn, A.M., and Clore, G.M. (1987). The binding of the cyclic AMP receptor protein to synthetic DNA sites containing permutations in the consensus sequence TGTGA. Biochem. J 246, 227–232.

Jenkins, J.L., Krucinska, J., McCarty, R.M., Bandarian, V., and Wedekind, J.E. (2011). Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. J. Biol. Chem. 286, 24626–24637.

Jones, C.P., and Ferre-D'Amare, A.R. (2014). Crystal structure of a c-di-AMP riboswitch reveals an internally pseudo-dimeric RNA. EMBO J. 33, 2692–2703.

de Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M., Jackson-Sillah, D.J., Fox, A., DeRiemer, K., Gagneux, S., et al. (2008). Progression to Active Tuberculosis, but Not Transmission, Varies by *Mycobacterium tuberculosis* Lineage in The Gambia. J. Infect. Dis. 198, 1037–1043.

Kana, B.D., and Mizrahi, V. (2010). Resuscitation-promoting factors as lytic enzymes for bacterial growth and signaling. FEMS Immunol. Med. Microbiol. 58, 39–50.

Kana, B.D., Gordhan, B.G., Downing, K.J., Sung, N., Vostroktunova, G., Machowski, E.E., Tsenova, L., Young, M., Kaprelyants, A., Kaplan, G., et al. (2008). The resuscitation-promoting factors of *Mycobacterium tuberculosis* are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro. Mol. Microbiol. 67, 672–684.

Kaprelyants, A.S., Mukamolova, G. V., Ruggiero, A., Makarov, V.A., Demina, G.R., Shleeva, M.O., Potapov, V.D., and Shramko, P.A. (2012). Resuscitation-promoting factors (Rpf): in search of inhibitors. Protein Pept. Lett. 19, 1026–1034.

Kato, A., Tanimoto, S., Muraki, Y., Kobayashi, K., and Kumagai, I. (1992). Structural and Functional Properties of Hen Egg-white Lysozyme Deamidated by Protein Engineering. Biosci. Biotechnol. Biochem. 56, 1424–1428.

Kaur, D., Guerin, M.E., Škovierová, H., Brennan, P.J., and Jackson, M. (2009). Chapter 2 Biogenesis of the Cell Wall and Other Glycoconjugates of *Mycobacterium tuberculosis*. Adv. Appl. Microbiol. 69, 23–78.

Kieser, K.J., and Rubin, E.J. (2014). How sisters grow apart: mycobacterial growth and division. Nat. Rev. Microbiol. 12, 550–562.

Koestler, B.J., and Waters, C.M. (2013). Exploring environmental control of cyclic di-GMP signaling in *Vibrio cholerae* by using the ex vivo lysate cyclic di-GMP assay (TELCA). Appl. Environ. Microbiol. 79, 5233–5241.

Kolb, A., Spassky, A., Chapon, C., Blazy, B., and Buc, H. (1983). On the different binding affinities of CRP at the *lac*, *gal* and *malt* promoter regions. Nucleic Acids Res. 11, 7833–7852.

Kolb, A., Busby, S., Buc, I.I., Garges, S., and Adhya, S. (1993). Transcriptional Regulation by cAMP and its Receptor Protein. Annu. Rev. Biochem. 62, 749–797.

Krakow, J.S., and Pastan, I. (1973). Cyclic adenosine monophosphate receptor: loss of cAMP-dependent DNA binding activity after proteolysis in the presence of cyclic adenosine monophosphate. Proc. Natl. Acad. Sci. U. S. A. 70, 2529–2533.

Krawczyk, J., Kohl, T.A., Goesmann, A., Rn Kalinowski, J., and Baumbach, J. (2009). From *Corynebacterium glutamicum* to *Mycobacterium tuberculosis*—towards transfers of gene regulatory networks and integrated data analyses with MycoRegNet. Nucleic Acids Res. 37.

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

Kumar, P., Arora, K., Lloyd, J.R., Lee, I.Y., Nair, V., Fischer, E., Boshoff, H.I.M., and Barry, C.E. (2012). Meropenem inhibits D,D-carboxypeptidase activity in *Mycobacterium tuberculosis*. Mol. Microbiol. 86, 367–381.

Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., Blanot, D., Gutmann, L., and Mainardi, J.-L. (2008). The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L,D-transpeptidation. J. Bacteriol. 190, 4360–4366.

Lee, H.N., Lee, N.O., Han, S.J., Ko, I.J., and Oh, J. Il (2014). Regulation of the *ahpC* gene encoding alkyl hydroperoxide reductase in *Mycobacterium smegmatis*. PLoS One 9.

Leistikow, R.L., Morton, R.A., Bartek, I.L., Frimpong, I., Wagner, K., and Voskuil, M.I. (2010). The *Mycobacterium tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. J. Bacteriol. 192, 1662–1670.

Li, Y., and Breaker, R.R. (1999). Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. J. Am. Chem. Soc. 121, 5364–5372.

Lin, S.-H., and Lee, J.C. (2003). Determinants of DNA bending in the DNA-cyclic AMP receptor protein complexes in *Escherichia coli*. Biochemistry 42, 4809–4818.

Liu, T., Ramesh, A., Ma, Z., Ward, S.K., Zhang, L., George, G.N., Talaat, A.M., Sacchettini, J.C., and Giedroc, D.P. (2007a). CsoR is a novel *Mycobacterium tuberculosis* copper-sensing transcriptional regulator. Nat Chem Biol 3, 60–68.

Luo, M., Fadeev, E.A., and Groves, J.T. (2005). Mycobactin-mediated iron acquisition within macrophages. Nat. Chem. Biol. 1, 149–153.

Luo, Y., Chen, B., Zhou, J., Sintim, H.O., and Dayie, T.K. (2014). E88, a new cyclic-di-GMP class I riboswitch aptamer from *Clostridium tetani*, has a similar fold to the prototypical class I riboswitch, *Vc*2, but differentially binds to c-di-GMP analogs. Mol. Biosyst. 10, 384–390.

Ma, Z., Cowart, D.M., Scott, R.A., and Giedroc, D.P. (2009). Molecular insights into the metal selectivity of the Copper(I)-sensing repressor CsoR from *Bacillus subtilis*. Biochemistry 48, 3325–3334.

Makarov, V., Manina, G., Mikusova, K., Mollmann, U., Ryabova, O., Saint-Joanis, B., Dhar, N., Pasca, M.R., Buroni, S., Lucarelli, A.P., et al. (2009). Benzothiazinones Kill *Mycobacterium tuberculosis* by Blocking Arabinan Synthesis. Science 324, 801–804.

Mandal, M., and Breaker, R.R. (2004). Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol. 5, 451–463.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. (2004). A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. Science 80, 306.

Manikandan, K., Sabareesh, V., Singh, N., Saigal, K., Mechold, U., and Sinha, K.M. (2014). Two-step synthesis and hydrolysis of cyclic di-AMP in *Mycobacterium tuberculosis*. PLoS One 9, e86096.

Marcus, S.A., Sidiropoulos, S.W., Steinberg, H., and Talaat, A.M. (2016). CsoR is essential for maintaining copper homeostasis in *Mycobacterium tuberculosis*. PLoS One 11, 1–20.

Master, S.S., Springer, B., Sander, P., Boettger, E.C., Deretic, V., and Timmins, G.S. (2002). Oxidative stress response genes in *Mycobacterium tuberculosis*: role of ahpC in resistance to peroxynitrite and stage-specific survival in macrophages. Microbiology 1742, 3139–3144.

McCue, L.A., McDonough, K.A., and Lawrence, C.E. (2000). Functional classification of cNMP-binding proteins and nucleotide cyclases with implications for novel regulatory pathways in *Mycobacterium tuberculosis*. Genome Res. 10, 204–219.

Meehan, R.E., Torgerson, C.D., Gaffney, B.L., Jones, R.A., and Strobel, S.A. (2016). Nuclease-Resistant c-di-AMP Derivatives That Differentially Recognize RNA and Protein Receptors. Biochemistry 55, 837–849.

Molle, V., and Kremer, L. (2010). Division and cell envelope regulation by Ser/Thr phosphorylation: *Mycobacterium* shows the way. Mol. Microbiol. 75, 1064–1077.

Mukamolova, G. V, Turapov, O.A., Kazarian, K., Telkov, M., Kaprelyants, A.S., Kell, D.B., and Young, M. (2002). The rpf gene of *Micrococcus luteus* encodes an essential secreted growth factor. Mol. Microbiol. 46, 611–621.

Mukamolova, G. V., Yanopolskaya, N.D., Votyakova, T. V., Popov, V.I., Kaprelyants, A.S., and Kell, D.B. (1995). Biochemical changes accompanying the long-term starvation of *Micrococcus luteus* cells in spent growth medium. Arch. Microbiol. 163, 373–379.

Nambu, S., Matsui, T., Goulding, C.W., Takahashi, S., and Ikeda-Saito, M. (2013). A new way to degrade heme: The *Mycobacterium tuberculosis* enzyme MhuD catalyzes heme degradation without generating CO. J. Biol. Chem. 288, 10101–10109.

Nathan, C., and Shiloh, M.U. (2000). Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Pnas 97, 8841–8848.

Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., and Breaker, R.R. (2013). Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat. Chem. Biol. 9, 834–839.

Neyrolles, O., Wolschendorf, F., Mitra, A., and Niederweis, M. (2015). Mycobacteria, metals, and the macrophage. Immunol. Rev. 264, 249–263.

Ng, V.H., Cox, J.S., Sousa, A.O., MacMicking, J.D., and McKinney, J.D. (2004). Role of KatG catalase-peroxidase in mycobacterial pathogenisis: Countering the phagocyte oxidative burst. Mol. Microbiol. 52, 1291–1302.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. 2, 2212–2221.

Nikitushkin, V.D., Demina, G.R., and Kaprelyants, A.S. (2011). Effect of secreted Rpf protein on intracellular contacts in *Micrococcus luteus* and *Mycobacterium smegmatis* cultures. Microbiology 80, 143–149.

Nikitushkin, V.D., Demina, G.R., Shleeva, M.O., Guryanova, S. V., Ruggiero, A., Berisio, R., and Kaprelyants, A.S. (2015). A product of RpfB and RipA joint enzymatic action promotes the resuscitation of dormant mycobacteria. FEBS J. 282, 2500–2511.

Ottenhoff, T.H.M., and Kaufmann, S.H.E. (2012). Vaccines against tuberculosis: Where are we and where do we need to go? PLoS Pathog. 8, 273–279.

Padh, H., and Venkitasubramanian, T.A. (1976). Adenosine 3',5'-monophosphate in *Mycobacterium phlei* and *Mycobacterium tuberculosis* H37Ra. Microbios 16, 183–189.

Padh, H., and Venkitasubramanian, T.A. (1977). Adenosine 3', 5'-monophosphate in mycobacteria. Life Sci. 20, 1273–1280.

Parish, T., Schaeffer, M., Roberts, G., and Duncan, K. (2005). HemZ is essential for heme biosynthesis in *Mycobacterium tuberculosis*. Tuberculosis 85, 197–204.

Pawlowski, A., Jansson, M., Sköld, M., Rottenberg, M.E., and Källenius, G. (2012). Tuberculosis and HIV co-infection. PLoS Pathog. 8, e1002464.

Piddington, D.L., Fang, F.C., Laessig, T., Cooper, A.M., Orme, I.M., and Buchmeier, N.A. (2001). Cu, Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. Infect. Immun. 69, 4980–4987.

Raghu, B., Sarma, C.R., and Venkatesan, P. (1993). Effect of hemoglobin on the growth of mycobacteria and the production of siderophores. Indian J Pathol Microbiol 36, 376–382.

Rapisarda, V.A., Volentini, S.I., Farías, R.N., and Massa, E.M. (2002). Quenching of bathocuproine disulfonate fluorescence by Cu(I) as a basis for copper quantification. Anal. Biochem. 307, 105–109.

Regulski, E.E., and Breaker, R.R. (2008). In-Line Probing Analysis of Riboswitches. Methods Mol. Biol. 419, 53–67.

Regulski, E.E., Moy, R.H., Weinberg, Z., Barrick, J.E., Yao, Z., Ruzzo, W.L., and Breaker, R.R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. Mol. Microbiol. 68, 918–932.

Ren, A., and Patel, D.J. (2014). c-di-AMP binds the *ydaO* riboswitch in two pseudo-symmetry-related pockets. Nat. Chem. Biol. 10, 780–786.

Ren, Y.L., Garges, S., Adhya, S., and Krakow, J.S. (1990). Characterization of the binding of cAMP and cGMP to the CRP598 mutant of the *E. coli* cAMP receptor protein. Nucleic Acids Res. 18, 5127–5132.

Rickman, L., Scott, C., Hunt, D.M., Hutchinson, T., Menéndez, M.C., Whalan, R., Hinds, J., Colston, M.J., Green, J., and Buxton, R.S. (2005). A member of the cAMP receptor protein family of transcription regulators in *Mycobacterium tuberculosis* is required for virulence in mice and controls transcription of the *rpfA* gene coding for a resuscitation promoting factor. Mol. Microbiol. 56, 1274–1286.

Rittershaus, E.S.C., Baek, S.-H., and Sassetti, C.M. (2013). The normalcy of dormancy: common themes in microbial quiescence. Cell Host Microbe 13, 643–651.

Roth, A., Winkler, W.C., Regulski, E.E., Lee, B.W.K., Lim, J., Jona, I., Barrick, J.E., Ritwik, A., Kim, J.N., Welz, R., et al. (2007). A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat. Struct. Mol. Biol. 14, 308–317.

Rowland, J.L., and Niederweis, M. (2012). Resistance mechanisms of *Mycobacterium tuberculosis* against phagosomal copper overload. Tuberculosis 92, 202–210.

Rowland, J.L., and Niederweis, M. (2013). A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. J. Bacteriol. 195, 3724–3733.

Roy, A., Danchin, A., Joseph, E., and Ullmann, A. (1983). Two functional domains in adenylate cyclase of *Escherichia coli*. J. Mol. Biol. 165, 197–202.

Russell, D.G. (2007). Who puts the tubercle in tuberculosis? Nat. Rev. Microbiol. 5, 39–47.

Russell, D.G., Schaible, U.E., Sturgill-Koszycki, S., and Schlesinger, P.H. (1998). Cytokine Activation Leads to Acidification and Increases Maturation of *Mycobacterium avium*-Containing Phagosomes in Murine Macrophages. J Immunol Ref. 160, 1290–1296.

Russell, D.G., Barry, C.E., and Flynn, J.L. (2010). Tuberculosis: What We Don't Know Can, and Does, Hurt Us. Science 328, 852–856.

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. (2009). Hypoxia: A window into *Mycobacterium tuberculosis* latency. Cell. Microbiol. 11, 1151–1159.

Sabourin, D., and Beckwith, J. (1975). Deletion of the *Escherichia coli crip* gene. J. Bact. 122, 338–340.

Sands, W.A., and Palmer, T.M. (2008). Regulating gene transcription in response to cyclic AMP elevation. Cell. Signal. 20, 460–466.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., et al. (2003). Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. J. Exp. Med. J. Exp. Med 96931200, 693–704.

Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science 253, 1001–1007.

Senaratne, R.H., De Silva, A.D., Williams, S.J., Mougous, J.D., Reader, J.R., Zhang, T., Chan, S., Sidders, B., Lee, D.H., Chan, J., et al. (2006). 5'-Adenosinephosphosulphate reductase (CysH) protects *Mycobacterium tuberculosis* against free radicals during chronic infection phase in mice. Mol. Microbiol. 59, 1744–1753.

Sexton, D.L., St-Onge, R.J., Haiser, H.J., Yousef, M.R., Brady, L., Gao, C., Leonard, J., and Elliot, M.A. (2015). Resuscitation-promoting factors are cell wall-lytic enzymes with important roles in the germination and growth of *Streptomyces coelicolor*. J. Bacteriol. 197, 848–860.

Seyler R.W., J., Olson, J.W., and Maier, R.J. (2001). Superoxide dismutase-deficient mutants of *Helicobacter pylori* are hypersensitive to oxidative stress and defective in host colonization. Infect. Immun. 69, 4034–4040.

Shanahan, C.A., Gaffney, B.L., Jones, R.A., and Strobel, S.A. (2013). Identification of c-di-GMP derivatives resistant to an EAL domain phosphodiesterase. Biochemistry 52, 365–377.

Sharma, R., Zaveri, A., Gopalakrishnapai, J., Thiruneelakantan, S., Varshney, U., and Visweswariah, S.S. (2014). Paralogous cAMP receptor proteins in *Mycobacterium smegmatis* show biochemical and functional divergence. Biochemistry 53, 7765–7776.

Shenoy, A.R., and Visweswariah, S.S. (2006). New messages from old messengers: cAMP and mycobacteria. Trends Microbiol. 14, 543–550.

Shi, S., and Ehrt, S. (2006). Dihydrolipoamide acyltransferase is critical for *Mycobacterium tuberculosis* pathogenesis. Infect. Immun. 74, 56–63.

Shleeva, M.O., Kudykina, Y.K., Vostroknutova, G.N., Suzina N.E., Mulyukin, A.L., and Kaprelyants, A.S. (2011). Dormant ovoid cells of *Mycobacterium tuberculosis* are formed in response to gradual external acidification. Tuberculosis. 91, 146–154.

Shleeva, M., Goncharenko, A., Kudykina, Y., Young, D., Young, M., Kaprelyants, A. (2013). Cyclic AMP-dependent resuscitation of dormant Mycobacteria by exogenous free fatty acids. PLoS One. 23, 1-12.

Singh, P., Rao, R.N., Reddy, J.R.C., Prasad, R., Kotturu, S.K., Ghosh, S., and Mukhopadhyay, S. (2016). PE11, a PE/PPE family protein of *Mycobacterium tuberculosis* is involved in cell wall remodeling and virulence. Sci. Rep. 6, 21624.

Skovierová, H., Larrouy-Maumus, G., Pham, H., Belanová, M., Barilone, N., Dasgupta, A., Mikusová, K., Gicquel, B., Gilleron, M., Brennan, P.J., et al. (2010). Biosynthetic origin of the galactosamine substituent of Arabinogalactan in *Mycobacterium tuberculosis*. J. Biol. Chem. 285, 41348–41355.

Smith, A.M., Fuchs, R.T., Grundy, F.J., and Henkin, T.M. (2010). Riboswitch RNAs: Regulation of gene expression by direct monitoring of a physiological signal. RNA Biol. 7, 104–110.

Smith, K.D., Lipchock, S. V, Ames, T.D., Wang, J., Breaker, R.R., and Strobel, S.A. (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol. Biol. 16, 1218–1223.

Snow, G.A., Zinecard, D., Water, S., Acid, H., Water, S., Acid, H., Water, S., Water, S., Ringer, L., Not, D.O., et al. (1970). Mycobactins: iron-chelating growth factors from mycobacteria. Bacteriol. Rev. 34, 99–125.

Song, H., Sandie, R., Wang, Y., Andrade-Navarro, M.A., and Niederweis, M. (2008). Identification of outer membrane proteins of *Mycobacterium tuberculosis*. Tuberculosis 88, 526–544.

Soukup, G.A., and Breaker, R.R. (1999). Relationship between internucleotide linkage geometry and the stability of RNA. RNA 5, 1308–1325.

Soukup, G.A., DeRose, E.C., Koizumi, M., and Breaker, R.R. (2001). Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. RNA 7, 524–536.

Spitale, R.C., Torelli, A.T., Krucinska, J., Bandarian, V., and Wedekind, J.E. (2009). The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. J. Biol. Chem. 284, 11012–11016.

Spreadbury, C.L., Pallen, M.J., Overton, T., Behr, M.A., Mostowy, S., Spiro, S., Busby, S.J.W., and Cole, J.A. (2005). Point mutations in the DNA- and cNMP-binding domains of the homologue of the cAMP receptor protein (CRP) in *Mycobacterium bovis* BCG: Implications for the inactivation of a global regulator and strain attenuation. Microbiology 151, 547–556.

St-Onge, R.J., Haiser, H.J., Yousef, M.R., Sherwood, E., Tschowri, N. and Mahmoud, A.B. (2015). Nucleotide second messenger-mediated regulation of a muralytic enzyme in *Streptomyces*. Mol. Micro. 96, 779–795.

St-Onge, R.J., and Elliot, M.A. (2017). Regulation of a muralytic enzyme-encoding gene by two non-coding RNAs. RNA Biol. 14, 1592–1605.

Stafford, S.L., Bokil, N.J., Achard, M.E.S., Kapetanovic, R., Schembri, M.A., McEwan, A.G., and Sweet, M.J. (2013). Metal ions in macrophage antimicrobial pathways: emerging roles for zinc and copper. Biosci. Rep. 33.

Stapleton, M., Haq, I., Hunt, D.M., Arnvig, K.B., Artymiuk, P.J., Buxton, R.S., and Green, J. (2010). *Mycobacterium tuberculosis* cAMP receptor protein (Rv3676) differs from the *Escherichia coli* paradigm in its cAMP binding and DNA binding properties and transcription activation properties. J. Biol. Chem. 285, 7016–7027.

Sturgill-Koszycki, S., Schlesinger, P.H., Chakraborty, P., Pryce, L., Collins, H.L., Fok, A.K., Allen, R.D., Gluck, S.L., Heuser, J., and Russell, D.G. (1995). Lack of acidification in *Mycobacterium* phagosomes produced by exclusion of the vesicular proton-ATPase. Science 263, 678–681.

Sturgill-Koszycki, S., Schaible, U.E., and Russell, D.G. (1996). Mycobacterium-containing phagosomes are accessible to early endosomes and reflect a transitional state in normal phagosome biogenesis. EMBO J. 15, 6960–6968.

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science 321, 411–413.

Sumartojo, E. (1993). When Tuberculosis Treatment Fails: A Social Behavioral Account of Patient Adherence. Am. Rev. Respir. Dis. 147, 1311–1320.

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., Majlessi, L., Criscuolo, A., Tap, J., Pawlik, A., et al. (2013). Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nat. Genet. 45, 172–179.

Tailleux, L., Waddell, S.J., Pelizzola, M., Mortellaro, A., Withers, M., Tanne, A., Castagnoli, P.R., Gicquel, B., Stoker, N.G., Butcher, P.D., et al. (2008). Probing Host Pathogen Cross-Talk by Transcriptional Profiling of Both *Mycobacterium tuberculosis* and Infected Human Dendritic Cells and Macrophages. PLoS One 3, e1403.

Takahashi, M., Blazy, B., and Baudras, A. (1980). An equilibrium study of the cooperative binding of adenosine cyclic 3',5'-monophosphate and guanosine cyclic 3',5'-monophosphate to the adenosine cyclic 3',5'-monophosphate receptor protein from *Escherichia coli*. Biochemistry 19, 5124–5130.

Takayama, K., Wang, C., and Besra, G.S. (2005). Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin. Microbiol. Rev. 18, 81–101.

Talaat, A.M., Lyons, R., Howard, S.T., and Johnston, S.A. (2004). The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc. Natl. Acad. Sci. U. S. A. 101, 4602–4607.

Tang, Q., Luo, Y., Zheng, C., Yin, K., Ali, M.K., Li, X., and He, J. (2015). Functional Analysis of a c-di-AMP-specific Phosphodiesterase MsPDE from *Mycobacterium smegmatis*. Int. J. Biol. Sci. 11, 813–824.

Tullius, M. V, Harmston, C.A., Owens, C.P., Chim, N., Morse, R.P., McMath, L.M., Iniguez, A., Kimmey, J.M., Sawaya, M.R., Whitelegge, J.P., et al. (2011). Discovery and characterization of a unique mycobacterial heme acquisition system. Proc. Natl. Acad. Sci. U. S. A. 108, 5051–5056.

Vandal, O.H., Pierini, L.M., Schnappinger, D., Nathan, C.F., and Ehrt, S. (2008). A membrane protein preserves intrabacterial pH in intraphagosomal *Mycobacterium tuberculosis*. Nat. Med. 14, 849–854.

Vilchèze, C., Av-Gay, Y., Attarian, R., Liu, Z., Hazbón, M.H., Colangeli, R., Chen, B., Liu, W., Alland, D., Sacchettini, J.C., et al. (2008). Mycothiol biosynthesis is essential for ethionamide susceptibility in *Mycobacterium tuberculosis*. Mol. Microbiol. 69, 1316–1329.

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 84, 218–227.

Wagner, D., Maser, J., Lai, B., Cai, Z.H., Barry, C.E., Bentrup, K.H.Z., Russell, D.G., and Bermudez, L.E. (2005). Elemental analysis of *Mycobacterium avium*-, *Mycobacterium tuberculosis*-, and *Mycobacterium smegmatis*-containing phagosomes indicates pathogen-induced microenvironments within the host cell's endosomal system. J. Immunol. 174, 1491–1500.

Wandersman, C., and Stojiljkovic, I. (2000). Bacterial heme sources: The role of heme, hemoprotein receptors and hemophores. Curr. Opin. Microbiol. 3, 215–220.

Ward, S.K., Hoye, E.A., and Talaat, A.M. (2008). The global responses of *Mycobacterium tuberculosis* to physiological levels of copper. J. Bacteriol. 190, 2939–2946.

Ward, S.K., Abomoelak, B., Hoye, E.A., Steinberg, H., and Talaat, A.M. (2010). CtpV: a putative copper exporter required for full virulence of *Mycobacterium tuberculosis*. Mol. Microbiol. 77, 1096–1110.

Wardman, P., and Candeias, L.P. (1996). Fenton chemistry: an introduction. Radiat. Res. 145, 523–531.

Watson, P.Y., and Fedor, M.J. (2011). The *glmS* riboswitch integrates signals from activating and inhibitory metabolites *in vivo*. Nat. Publ. Gr. 18, 3–5.

Watson, P.Y., and Fedor, M.J. (2012). The *ydaO* motif is an ATP-sensing riboswitch in *Bacillus subtilis*. Nat. Chem. Biol. 8, 963–965.

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

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

White, C., Kambe, T., Fulcher, Y.G., Sachdev, S.W., Bush, A.I., Fritsche, K., Lee, J., Quinn, T.P., and Petris, M.J. (2009). Copper transport into the secretory pathway is regulated by oxygen in macrophages. J. Cell Sci. 122, 1315–1321.

Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., and Breaker, R.R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428, 281–286.

Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., Van Soolingen, D., Rüsch-Gerdes, S., Locht, C., Brisse, S., et al. (2008). Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog. 4, e1000160.

Witte, G., Hartung, S., Büttner, K., and Hopfner, K.-P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. Mol. Cell 30, 167–178.

Wolf, A.J., Desvignes, L., Linas, B., Banaiee, N., Tamura, T., Takatsu, K., and Ernst, J.D. (2008). Initiation of the adaptive immune response to *Mycobacterium tuberculosis* depends on antigen production in the local lymph node, not the lungs. J. Exp. Med. 205, 105–115.

Wolschendorf, F., Ackart, D., Shrestha, T.B., Hascall-Dove, L., Nolan, S., Lamichhane, G., Wang, Y., Bossmann, S.H., Basaraba, R.J., and Niederweis, M. (2011). Copper resistance is essential for virulence of *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. 108, 1621–1626.

World Health Organization (2017). WHO | Tuberculosis (World Health Organization).

Wu, H.M., and Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. Nature 308, 509–513.

Wu, C. h H., Tsai-Wu, J.J., Huang, Y.T., Lin, C.Y., Lioua, G.G., and Lee, F.J.S. (1998). Identification and subcellular localization of a novel Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis*. FEBS Lett. 439, 192–196.

Young, M., Artsatbanov, V., Beller, H.R., Chandra, G., Chater, K.F., Dover, L.G., Goh, E.B., Kahan, T., Kaprelyants, A.S., Kyrpides, N., et al. (2010). Genome sequence of the fleming strain of *Micrococcus luteus*, a simple free-living actinobacterium. J. Bacteriol. 192, 841–860.

Yu, S., Maillard, R.A., Gribenko, A. V, and Lee, J.C. (2012). The N-terminal Capping Propensities of the D-helix Modulate the Allosteric Activation of the *Escherichia coli* cAMP Receptor Protein.

Yuan, Y., Zhu, Y.Q., Crane, D.D., and Barry, C.E. (1998). The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis*. Mol. Microbiol. 29, 1449–1458.

Zheng, C., Wang, J., Luo, Y., Fu, Y., Su, J., and He, J. (2013). Highly efficient enzymatic preparation of c-di-AMP using the diadenylate cyclase DisA from *Bacillus thuringiensis*. Enzyme Microb. Technol. 52, 319–324.

Zumla, A., George, A., Sharma, V., Hon, R., Herbert, N., Oxley, A., Oliver, M., and Masham, B. (2015). The WHO 2014 Global tuberculosis report — further to go. Lancet Glob. Heal. 3, 10–12.