Insights into the Mycobacterial Response to Nitrogen Limitation; Characterisation of the GlnR Regulon

Victoria Alice Jenkins

A thesis submitted for the degree of Doctor of Philosophy and Diploma of Imperial College London

MRC Centre for Molecular Bacteriology and Infection
Division of Infectious Diseases
Department of Medicine

Imperial College London
May 2013

ABSTRACT

The ability to sense and initiate a response to situations of nitrogen-limitation is essential for bacterial survival. In extensively investigated organisms, the nitrogen-stress response consists of changes in intracellular metabolite levels, post-translational modification of proteins (such as metabolic enzymes) and a transcriptomic response mediated by a global response regulator. However, in mycobacteria the nitrogen stress response has not been comprehensively investigated. In this study mycobacterial nitrogen limiting conditions were optimised and the mechanism of GlnR activation investigated; *M. smegmatis* GlnR requires a highly conserved aspartate residue (D48), corresponding to a putative phosphorylation site, for function. In addition, a ChIP-seq approach combined with global expression analyses, permitted characterisation of the GlnR mediated global transcriptomic response stimulated during nitrogen.

In *M. smegmatis,* 52 GlnR binding sites were identified, controlling the expression of at least 103 genes in response to nitrogen limitation. The majority of GlnR regulated genes were involved in nitrogen uptake and nitrogen scavenging. A consensus GlnR DNA binding motif was identified and AC-n9-AC DNA residues shown to be essential for GlnR:DNA binding. In *M. tuberculosis* 36 GlnR binding sites were identified in nitrogen limitation, however no consensus GlnR:DNA binding motif could be determined. Initial analysis suggests GlnR may be involved in a general stress response in *M. tuberculosis*, rather than mediating a nitrogen scavenging response as observed in *M. smegmatis*.

This study provides the first global analysis of nitrogen limitation in mycobacteria and identifies GlnR as the main nitrogen response regulator. From this analysis it appears that the role of GlnR is different in *M. tuberculosis* compared to *M. smegmatis*, which may provide key insights into how pathogenic and non-pathogenic species survive nutrient limiting conditions.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1:

CHAPTER 3:

CHAPTER 4:

CHAPTER 5:

CHAPTER 6:

CHAPTER 7:

CHAPTER 8:

LIST OF TABLES

AKNOWLEDGMENTS

Firstly I would like to thank my two supervisors, Dr. Kerstin Williams and Dr. Brian Robertson, without whom this project and my PhD would not have been possible. Kerstin made the project enjoyable, offered me encouragement and pushed me to reach my potential. I am extremely grateful for all the time, effort and patience Kerstin has invested in me. Brian I would like to thank for over seeing the whole project, for his valuable input and for the opportunity to share my work at various conferences, enabling me meet to meet some fascinating people. I would also like to thank them both for their detailed critique of this thesis. I am also grateful to the MRC for providing me with funding to complete both my PhD and MSc projects.

In addition I would like to thank the members of the LoLa consortium. They have provided me with the opportunity to utilise various resources and learn about new techniques. In particular I would like to thank Geraint Barton who carried out the bioinformatics analysis of this project and Dr. Volker Behrends who helped with the initial media development.

I would like to express my gratitude to all members of CMBI3 for their support and help throughout this project. An invaluable experience has been the Thursday lab meetings, which provided a forum to share and troubleshoot ideas. I would also like to thank Michaella who listened to my woes over lunch for the whole 3 years. This lunch group later expanded to include various additional members and I would like to thank them all, in particular Ana and Dan.

My friends have also played an important role in enabling me to complete this PhD. In particular Sarah, without whom I wouldn't have been able to physically complete these past few years. She provided a shoulder to cry on, an ear to listen to all my problems and most importantly support in taking care of my daughter. For her friendship and support towards looking after Isabella I am eternally grateful.

Finally I would like to thank my family. My parents and Elizabeth for helping me realise my dreams and encouraging me every step of the way. To Russell and his family for their support and enabling Isabella and me to take a break when needed and go on holiday. Most importantly this thesis is for my daughter, Isabella. She has given me perspective, making me realise what is important in life. To my beloved daughter Isabella, I would like to express my thanks for always being so amazing and constantly making me smile.

DECLARATION OF ORIGINALITY

I declare that the work presented in this thesis is my own, except where indicated by specific references that are displayed in the text and listed in the bibliography. The work presented in this thesis was completed in accordance to the regulations of Imperial College London.

ABBREVIATIONS

A	Alanine
AES	Allelic exchange substrate
ATP	Adenosine triphosphate
BC	Before Christ
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary DNA
cfu	Colony forming units
ChIP	Chromatin-Immunoprecipitation
ChIP-seq	Chromatin-immunoprecipitation coupled with high-throughput sequencing
D	Aspartic acid
Da	Daltons
DE	Differentially expressed
DIG	Digoxigenin
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DOTS	Directly Observed Therapy Short-course
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electromobility shift assay
FPLC	Fast protein liquid chromatography
g	Grams
GDH	Glutamate dehydrogenase
gDNA	Genomic DNA
GOGAT	Glutamate synthase
GS	Glutamine synthetase
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
Hyg	Hygromycin
IGV	Integrated genome viewer
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITB	Inoue transformation buffer
Kan	Kanamycin
kbp	Kilobase pair
kDa	Kilodaltons
KW	Work carried out by Kerstin Williams
L	Litre
LB	Luria Bertani
MAMA PCR	Mismatch amplification mutation assay polymerase chain reaction
MDR-TB	Multidrug-resistant *M. tuberculosis*
MEME	Multiple EM (Expectation Maximization) for Motif Elicitation
min	Minutes
ml	Millilitre
mM	Millimolar
MSX	Methionine sulfoximine
ng	Nanograms
NMR	Nuclear magnetic resonance

OADC	Oleic acid-albumin-dextrose-catalyse
OD	Optical Density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Q	Capacitance
qRT-PCR	Quantitative Real-Time PCR
R	Resistance
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Seconds
SISSRs	Site Identification from Short Sequence Reads
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
ssDNA	Single stranded DNA
TAE	Tris-Acetate-EDTA buffer
TBE	Tris-Borate EDTA buffer
TBS	Tris-buffered saline
TF	Transcription factor
UV	Ultraviolet
V	Volts
v/v	Volume/volume
VB	Work carried out by Volker Behrends
WHO	World health organisation
WT	Wild type
x g	Times gravity (relative centrifugal force)
XDR-TB	Extensively drug-resistant *M. tuberculosis*
µl	Microlitre

CHAPTER 1: Introduction

1.1 Mycobacteria

Mycobacterial pathogens are known to cause severe human disease, and include the causative agents of leprosy, *Mycobacterium leprae,* and *Mycobacterium tuberculosis*, the etiological agent of tuberculosis. These prominent human diseases, tuberculosis and leprosy, have been documented since antiquity. Furthermore, tuberculosis has been named as the suspected cause of death in some Egyptian mummies circa 3000 BC (204). In the 19th Century tuberculosis, known as the 'white plague', was a leading cause of death in Europe and the United States of America; the estimated mortality rate from tuberculosis was as high as 400 in 100,000 individuals in the USA during 1830 (124). It was in 1873 that a Norwegian doctor Armauer Hansen identified *M. leprae* from leprosy patients, with *M. tuberculosis* discovered nine years later by Koch in 1882 (109, 137). The incidence of leprosy has since declined to 192,246 reported cases, according to official reports received during 2011 from 130 countries (190). However, to-date, in some developing countries tuberculosis is still a leading cause of mortality due to an infectious disease.

Currently, the genus *Mycobacterium* contains more than 120 species, the majority of which are thought to be free living, ubiquitous bacteria (124, 168). Characteristics used to define mycobacteria are the absence of motility, resistance to acid-alcohol following coloration with phenicated fuchsin, a guanine/cytosine (G-C) rich (62-71%) genome and a slightly curved and rod-shaped morphology (124, 148). For convenience, mycobacteria have traditionally been divided into two major divisions based on phenotypic growth differences; rapid growers and slow growers. Rapid growers, such as the saprophytic soil bacteria *Mycobacterium smegmatis,* are defined as organisms that produce isolated visible colonies on nutrient-rich solid media within seven days of inoculation (134, 148). Slow growers, such as *M. tuberculosis,* generate colonies apparent only after seven days or more (134, 148). *M. leprae* is documented as the only mycobacterial species that has not been cultivated *in vitro*, special growth requirements are necessary such as cultivation on the footpad of mice or within the nine-banded armadillo (170). These growth rate divisions of mycobacteria have no formal taxonomic standing, however they are useful clinically in identification schemes. Typically, slow growing mycobacteria cause disease in humans and animals, whilst the fast growers do not, however a few notable exceptions apply (148).

Taxonomically, mycobacteria belong to the genus *Mycobacterium*, which is the single genus within the family of *Mycobacteriaceae*, order *Actinomycetales* (148). *Actinomycetales* comprise of diverse micro-organisms, including *Streptomyces* and *Corynebacteria*. All *Actinomycetales* share a G-C rich genome and are classified, due to phylogenetic analysis, as Gram-positive

organisms (18, 134, 148). Within the order *Corynebacterium*, *Mycobacterium* and *Nocardia* form a monophyletic taxon, termed the CMN group (43). Members of the CMN group are the only micro-organisms that are able to synthesise mycolic acids (43). Mycolic acids are cell wall components composed of long β-hydroxy fatty acids with a shorter α-alkyl side chain. Each molecule contains between 60 and 90 carbon atoms, however the exact number varies between species, in addition to the presence of different functional groups (167). It is the presence of mycolic acids that makes the cell envelope of mycobacteria structurally distinct from that of both Gram-positive and Gram-negative bacteria; their thick waxy cell envelope is impermeable to the Gram-stain. In addition, it is widely believed that the limited permeability of the mycobacterial cell envelope is due to the high mycolic acid content, combined with a variety of other intercalated lipids (167).

1.1.1 Pathology of *M. tuberculosis*

Tuberculosis is predominantly a pulmonary disease, transmitted via aerosolised droplets containing infectious *M. tuberculosis*. Infectious droplets are generated from pulmonary or laryngeal tuberculosis, and dispersed via a cough, sneeze or speaking, and subsequently inhaled by an uninfected person (53, 86). Upon infection, *M. tuberculosis* replicates within a membrane bound phagosome within human alveolar macrophages. A primary complex develops, consisting of a small lesion at the initial site of implantation (53). *M. tuberculosis* replication occurs at local lymph sites within the lung, leading to eventual dissemination of bacilli to remote sites of the body. At this stage of infection approximately 95% of infected individuals mount an immune response capable of controlling, but not eliminating the infection (140).

In an infected person who does not succumb to primary tuberculosis, the bacilli enter a latent state contained within an aggregate of immune cells. This requires the coordinated recruitment of macrophages and lymphocytes, which aggregate to form a granuloma surrounding the infected macrophages (140). The granuloma serves to contain the foci of infection, preventing dissemination of *M. tuberculosis*. For the majority of healthy individuals these immune structures can prevent disease progression indefinitely, however these quiescent bacilli can potentially reactivate at any time during the remainder of the person's life. Generally, about 5% of those infected develop primary tuberculosis within five years of infection, with a further 5-10% subsequently developing post-primary disease (86, 140). This disease ratio shows some regional variation, and it is notably much higher in immunosuppressed individuals or in the presence of human immunodeficiency virus (HIV) (140). For instance, co-infection with HIV

raises the chance of reactivation of *M. tuberculosis* from ~10% within the individual's lifetime to ~50% (38, 140). Reduction in immune competence significantly raises the chances of granuloma disruption and thus progression to active disease.

Failure of granuloma formation, or its breakdown, may release *M. tuberculosis* into the lung tissue allowing growth and replication. The tuberculosis disease is a chronic wasting illness characterised by fever, weight loss and, in the case of pulmonary reactivation, a cough (53). Many of the symptoms of tuberculosis, including tissue destruction, are a result of the host response against *M. tuberculosis*, rather than direct toxicity from the bacilli. Cell-mediated immunity activates macrophages leading to delayed-type hypersensitivity, resulting in caseous necrosis and ultimately killing of *M. tuberculosis* infected macrophages, however this is at the expense and destruction of nearby tissues (53, 86). In the majority of cases (~75%) the infection is restricted to the lungs causing pulmonary tuberculosis (101). A more severe, but less common, outcome of infection is dissemination from the lungs to other vital organs, leading to extra-pulmonary tuberculosis (142).

1.1.2 *M. tuberculosis* Prevalence and Treatment

In 1993 the World Health Organization (WHO) launched a global effort to control tuberculosis infection. Since then, progress towards the WHO's global targets for reduction in tuberculosis cases and deaths continues. The Millennium Development Goal, to halt and reverse the tuberculosis epidemic by 2015, has already been achieved (189). New cases of tuberculosis have fallen in recent years, and fell at a rate of 2.2% between 2010 and 2011 (189). The tuberculosis mortality rate has also decreased by 41% since 1990. Despite this, the global burden of tuberculosis incidence remains high; in 2011, the WHO estimated 8.7 million new cases of tuberculosis worldwide (13% co-infected with HIV) and 1.4 million people died from tuberculosis (189). To-date, tuberculosis still disproportionately affects those in low-income countries, with 22 high-burden countries accounting for over 80% of the world's tuberculosis cases (189).

A major control policy implemented by the WHO was the introduction of the Directly Observed Therapy Short-course (DOTS), which remains central in the effort to control tuberculosis (192). The DOTS treatment recommendation is a four drug cocktail consisting of rifampin, isoniazid, pyrazinamide and ethambutol, taken daily for two months followed by four to six months of two drugs, usually rifampin and isoniazid (194). Poor patient compliance, due to the high drug burden, and long treatment regimens reduces treatment effectiveness and increases the risk of

drug resistant strains emerging. Consequently, DOTS incorporates a comprehensive tuberculosis management programme, a five-element strategy for the control of tuberculosis; political commitment, improved laboratory analysis, direct patient observation during drug administration, a free drug supply for the complete short course of anti-tuberculosis drugs, and a reporting system documenting patient progress. As such, treatment success rates have been maintained at high levels for several years; in 2010 the treatment success rate among newly diagnosed cases was 85% (189). However, the DOTS policy still requires a developed infrastructure with medical expertise and funding, which is lacking in many less economically developed countries with high rates of tuberculosis.

Despite the global DOTS policy, the appearance of multidrug-resistant *M. tuberculosis* (MDR-TB) is an increasing problem. MDR-TB is caused by *M. tuberculosis* strains that are resistant to the most effective anti-tuberculosis drugs, isoniazid and rifampicin (194). MDR-TB results from either infection with organisms that are already drug-resistant or may develop during the course of the patient's treatment. MDR-TB strains have been detected worldwide and are estimated to account globally for approximately 3.7% of new cases and 20% of previously treated cases (189). Notably, MDR-TB is found at a higher incidence in less economically developed countries where DOTS compliance is difficult; in 2008 1.1% of *M. tuberculosis* isolates were multi-drug resistant in the UK compared to approximately 30% in countries with the highest incidence (62, 188). Treatment for MDR-TB is known as DOTS-plus (DOTS plus second line drugs), which relies on less effective, more expensive drugs, with a longer treatment period and more severe side effects (194). Extensively drug-resistant TB (XDR-TB), has also been documented in 84 countries, although with low prevalence (191). XDR-TB is a form of tuberculosis caused by organisms that are resistant to isoniazid and rifampicin as well as any fluoroquinolone and any of the second-line anti-TB injectable drugs (amikacin, kanamycin or capreomycin) (191). XDR-TB is virtually untreatable as few chemotherapeutic options remain.

New classes of anti-tuberculosis drugs are required to combat the emerging drug resistant *M. tuberculosis* strains. In addition, first-line anti-tuberculosis drugs like isoniazid have limited activity against dormant *M. tuberculosis* and consequently latent *M. tuberculosis* remains a major hindrance to effective chemotherapy (78). It is thought a shift in *M. tuberculosis* metabolic pathways, due to the adverse conditions experienced by the bacilli, enables them to survive in a state of dormancy prior to reactivation. Therefore for the design of innovative drugs, a more thorough understanding involving the survival strategy of mycobacteria, linked with its ability to adapt in changing environments is crucial. One such target would be the ability of the bacilli to survive and adapt in a nitrogen-limiting environment. The nitrogen metabolic pathway is of particular interest as the pathways assimilate nitrogen into essential biological macromolecules

such as proteins, nucleic acids and cell wall constituents. The essential nature of this metabolic pathway makes it an attractive system to investigate, potentially providing insight into survival strategies employed by mycobacteria, as well as uncovering novel drug targets.

1.1.3 *M. smegmatis* as a Model Organism

Despite determination of the entire genomic sequence of *M. tuberculosis,* attempts to elucidate biological pathways, that underlie virulence of the bacilli, have been severely hampered. *M. tuberculosis*, which has a generation time of approximately twenty four hours in optimal laboratory conditions, requires two to three weeks to yield colonies on solid medium. In addition, biosafety level 3 containment imposes stringent logistic constraints on cultivation of the pathogen (147). To this end, it is therefore advantageous to study some aspects of mycobacterial biology on fast-growing, non-pathogenic species, such as *M. smegmatis*.

M. smegmatis has been used as a non-pathogenic substitute to study the metabolic and regulatory pathways of *M. tuberculosis.* Trevisan in 1889 first applied the name *smegmatis* to the Smegma Bacillus discovered by Alvarez and Tavel (55). Despite initial cultivation from human smegma, *M. smegmatis* is generally a soil-dwelling, saprophytic, avirulent mycobacterial species (77). The bacterium is characterised as a fast growing mycobacterial strain with a generation time of approximately three to four hours. In rare cases, *M. smegmatis* has been reported to be pathogenic in humans; *M. smegmatis* was isolated from infected orbital tissue in a patient from Thailand and identified the causative agent in a fatal disseminated infection in an immune compromised infant (32, 119). The *M. smegmatis* mc^2 155 strain is most frequently used for research purposes, due to its relative ease of genetic manipulation. It is a high frequency transformation mutant originating from mc^26, a single colony isolate of the *M. smegmatis* reference strain ATCC 607 (182). However, the cause of the enhanced transformability of this strain is unknown. In 2006 the complete genome of *M. smegmatis* mc^2 155 was sequenced by TIGR, adding further to its use as a favourable alternative for the study of mycobacterial molecular processes.

1.2 Nitrogen Availability

1.2.1 Nutrient Limitation in *M. tuberculosis*

During infection the location of latent *M. tuberculosis* remains to be precisely established. The current paradigm is that quiescent *M. tuberculosis* resides within fibrotic granulomatous lesions in the lung where the bacilli have become dormant, in response to hypoxic conditions (22, 183). Hypoxic, non-replicating *M. tuberculosis* has been studied *in vitro* in the Wayne model of persistence; mycobacterial cultures are exposed to gradual oxygen depletion in a sealed system (184). In this *in vitro* model, mycobacteria undergo changes in their energetic and metabolic status (123, 145). The molecular mechanisms involved in the survival of non-replicating hypoxic mycobacteria remains largely undefined, however the DosR/DosT two-component system has been shown to be essential for anaerobic adaptation (84, 161). In addition to hypoxia, *M. tuberculosis* is exposed to many other environmental stresses during infection including nutrient deprivation, altered pH, iron limitation and exposure to reactive nitrogen and oxygen species (13). As such, an interplay among regulatory mechanisms is thought to enable the bacilli to remain in a quiescent state in these hostile conditions.

Some evidence suggests that persistent *M. tuberculosis* in lung lesions experiences nutrient deprivation. Nyka (1974) demonstrated that *M. tuberculosis* cells in lung lesions differ in their morphology and staining properties, when compared with bacilli grown in optimum conditions *in vitro*; they are small spherical cells rather than rods and are chromophobic (not stained with conventional stains and are not acid fast) (110). Similar morphology and staining properties were documented *in vitro* with *M. tuberculosis* cultures starved in distilled water, however the bacilli recovered growth and regained acid fast properties when added to nutrient-rich medium (110). In addition, nutrient starvation of *M. tuberculosis*, induced in a phosphate-buffered saline solution, resulted in a gradual shutdown of respiration to minimal levels; again, the bacilli remained viable and recovered when returned to a nutrient rich medium (88). *M. smegmatis* has also been documented to remain viable after prolonged periods of nutrient deprivation; cultures starved of carbon, nitrogen or phosphorous remained viable for over 650 days (150). As *M. tuberculosis* is deprived of nutrients in a granuloma, it is assumed the bacilli shut down certain metabolic pathways to economise energy consumption. The molecular control of this metabolic shift and reactivation, in regards to nutrient utilisation, is therefore of particular interest for establishment and revival of infection.

1.2.2 Nitrogen Sources

Ammonium

Ammonia supports the fastest growth rate of many bacteria, and is considered the preferred nitrogen source for *E. coli* (127). In solution, the protonated ammonium ion (NH_4^+) is in equilibrium with uncharged ammonia (NH_3). Ammonia is a lipophilic molecule and able to pass bacterial cell membranes by diffusion, which occurs under high ammonia/ammonium concentrations outside the cell. When this concentration decreases, passive diffusion is no longer sufficient and subsequently the bacterial ammonium transport system, Amt, is synthesised (3, 26, 195). Almost all bacteria and archaea encode at least one Amt protein, and frequently the gene encoding the transporter (AmtB) is found in an operon with GlnK, encoding a small signal transduction protein (7, 163). The Amt proteins are high-affinity ammonia transporters that function to scavenge ammonium and recapture ammonium lost from cells by diffusion (195). In *Saccharomyces cerevisiae* and *Rhodobacter capsulatus*, Amt proteins have also been implicated in sensing ammonium in the external medium (89, 200). Once inside the cell, ammonium is assimilated under high concentrations by glutamate dehydrogenase (GDH), and at low ammonium concentrations, glutamine synthetase and glutamate synthase (GS and GOGAT) (Section 1.3).

The *E. coli* AmtB transporter is a high-affinity, low capacity ammonia channel (10–10,000 molecules/s) and provides a paradigm for the Amt protein family (71, 121, 202). Transcription of the *amtB* gene is induced in response to nitrogen limitation and activated by the gene activator protein, NtrC (100, 203). Each subunit of AmtB has a narrow, predominantly hydrophobic, pore containing a number of highly conserved residues that have a role in periplasmic NH_4^+ binding, NH_4^+ deprotonation, and NH_3 translocation (121, 202). Further regulation of the AmtB ammonium channel is provided via protein-protein interactions with GlnK, with direct evidence for the interaction of GlnK and AmtB in *E. coli* provided by crystal structure studies (37, 57, 163). GlnK is a member of the P_{II} protein family and in *E. coli* acts as a sensor of cellular nitrogen status. In response to nitrogen deprivation, GlnK is covalently modified by uridylylation at residue Tyr51 at the apex of the T-loop, with this process reversed during nitrogen sufficiency (8, 68). Therefore in nitrogen sufficient conditions, the un-modified T loop of GlnK inserts into the cytoplasmic exit channel of AmtB, blocking ammonium transport into the cell (121). The *E. coli* AmtB-GlnK complex has a stoichiometry of 1:1, and all molecules of GlnK within the complex are fully deuridylylated (42). Under nitrogen limitation GlnK is uridylylated, resulting in GlnK dissociation from AmtB permitting ammonia influx through the channel (Figure 1.1) (37, 70).

During infection the availability of ammonium for *M. tuberculosis* utilisation is unknown; therefore the role of ammonium as a nitrogen source is unclear. Despite this homologues of AmtB exist in *M. tuberculosis* and *M. smegmatis,* suggesting both bacteria have a requirement for external ammonium uptake (5). In addition, *M. smegmatis* contains two further ammonium transporters AmtA (msmeg_4635) and Amt1 (msmeg_6259) (5). These additional genes, encoding Amt transporters in *M. smegmatis,* may indicate that ammonium utilisation in this species has a greater significance than for the pathogen. In *M. smegmatis* Amt1 and AmtB have been proposed to take up ammonium during nitrogen limitation; this is based on an increase in transcription of *amtB* and *amt1* under nitrogen limiting conditions (3). In addition, in *M. smegmatis* transcriptional regulation of *amtB* and *amt1* is controlled by GlnR, a transcriptional regulator of nitrogen control in mycobacteria (3). However, no biochemical data is available for mycobacteria in the role of AmtB or additional ammonium transporters in the uptake of ammonium.

In actinomycetes the *amtB* gene is encoded in an operon *amtB-glnK-glnD*. The additional gene present *glnD*, functions as an adenylyl transferase in *C. glutamicum*, *S. coelicolor* and *M. tuberculosis* (64, 158, 193). Recently it has been demonstrated that in response to nitrogen limitation GlnD adenylylates GlnK at residue Tyr51 in the T-loop of *M. tuberculosis* (193). The function of GlnK modification in response to nitrogen limitation is unknown; a *glnD* mutant lacking the ability to adenylylate GlnK displayed no impaired growth phenotype compare to the wild type strain during nitrogen limitation (193). However, comparing the structure of *M. tuberculosis* GlnK (P_{II}) (apo- and ATP-bound forms) with the *E. coli* GlnK:AmtB complex, suggests that *M. tuberculosis* GlnK (P_{II}) could form a complex with AmtB, but this has yet to be experimentally verified (144).

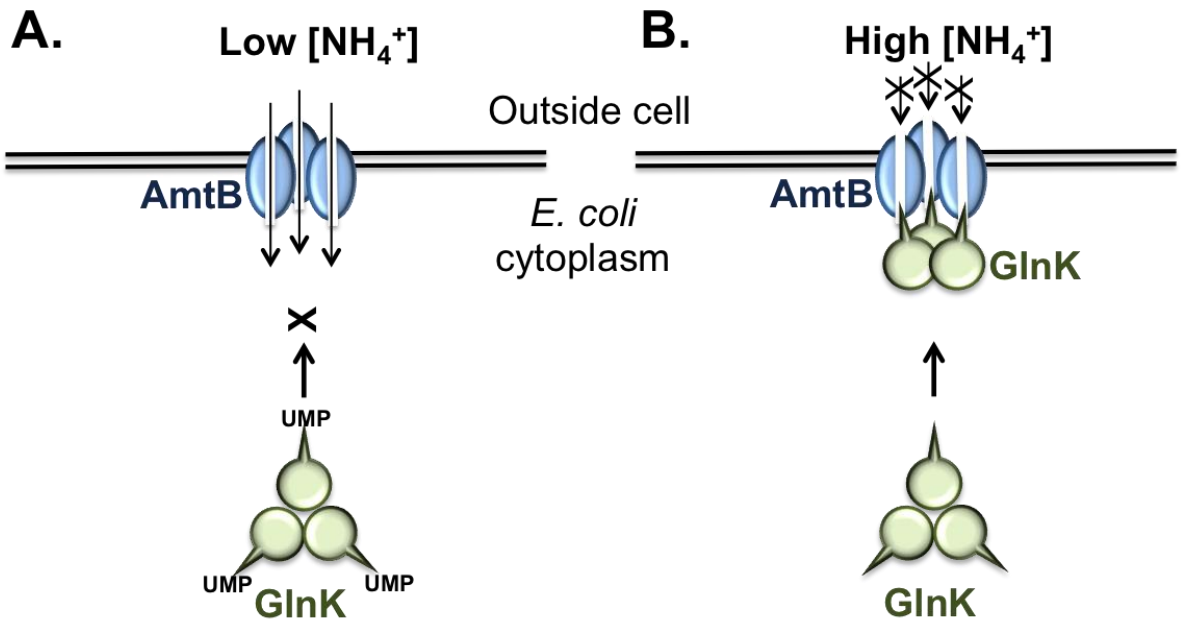


Figure 1.1. Model for the role of GlnK in regulating ammonium uptake via AmtB in *E. coli*.

(A) When the extracellular ammonium concentration is low, GlnK is uridylylated and unable to form a complex with AmtB, permitting ammonium influx through the AmtB channels. (B) An increase in extracellular ammonium leads to deuridylylation of GlnK, which favours the complex formation with AmtB. The process is reversible. Adapted from (121).

Nitrate

M. tuberculosis is phagocytosed by macrophages during infection, which limits nutrient availability to the intracellular pathogen (104). Nitric oxide (NO) is produced by activated macrophages via a nitric oxide synthase, and functions as a tumoricidal and antimicrobial molecule *in vitro* and *in vivo* (21). Nitrate is generated spontaneously from NO within the macrophage, and nitrate may therefore provide a source of nitrogen during growth. Evidence exists for a role of nitrate reductase enzymes during *M. tuberculosis* infection. Historically, *M. tuberculosis* has been differentiated from *Mycobacterium bovis* by the fact that only *M. tuberculosis* can reduce significant amounts of nitrate to nitrite (154). High nitrate reductase activity has also been correlated with increased virulence of some *M. tuberculosis* lineages and the NarGH locus was found actively transcribed in granulomas from the lungs of *M. tuberculosis* patients (46, 54). These results indicate a role for nitrate and mechanisms involved in its reduction during *M. tuberculosis* infection.

Assimilation of nitrate by *M. tuberculosis* as a nitrogen source has been reported previously (40, 63, 95). Genes encoding components for the complete reduction of nitrate to ammonium have been identified in mycobacterial genomes, with the exception of *M. leprae* (see Table 1.1 for *M. tuberculosis* and *M. smegmatis* genes) (5). Before assimilation of nitrogen into biomolecules can occur nitrate must first be reduced to ammonium. This proceeds via reduction of nitrate to nitrite by a nitrate reductase and the subsequent reduction of nitrite to ammonium via a nitrite reductase. The *M. tuberculosis narGHJI* and *narX* display homology with other prokaryotic nitrate reductases (154). However, it has been shown that *narGHJI* alone is responsible for nitrate-reducing activity in culture, the function of *narX* has yet to be identified (154). Encoded in an operon with the inactive nitrate reductase *narX*, is *narK2* (154). In *M. tuberculosis* four genes, *narK1*, *narK2*, *narK3* and *narU*, are homologous to the *E. coli* nitrate/nitrite transporters *narK* and *narU*. NarK2 is a putative nitrate/nitrite transporter. Early work in *E. coli* suggested that NarK was involved in nitrite export, and consequently the homologous NarK2 in *M. tuberculosis* is annotated as a 'nitrite extrusion protein' (133). Subsequent analysis of an *E. coli narK/narU* double mutant indicated that the two proteins could transport nitrate into and nitrite out of the cell (34, 73). The *M. tuberculosis narK2* could complement this *E. coli* double mutant, supporting a role for *narK2* transporter of nitrate into and nitrite out of the cell (154). In *M. tuberculosis* NirBD functions as an assimilatory nitrite reductase, responsible for the reduction of nitrite to ammonia, and was able to support growth on nitrite as a sole source of nitrogen (95). GlnR regulates the expression of *nirBD*, and a *glnR* mutant is unable to grow on either nitrate or nitrite as sole nitrogen sources (95). Therefore, it has been recognised that

NarGHJI and NirBD mediate the assimilatory reduction of nitrate and nitrite respectively in *M. tuberculosis* (Figure 1.2).

It is interesting to note that nitrate utilisation in *M. tuberculosis* can serve one of two functions; it can act as an energy source during anaerobic growth, in addition to a nitrogen source. In an anaerobic environment, many bacteria are able to use nitrate as a final electron acceptor in place of oxygen for the maintenance of a proton motive gradient. Nitrate reductase activity occurs at a low level during aerobic growth of *M. tuberculosis*, but increases significantly upon entry into the microaerobic stage (185). The increase in nitrate reductase activity in hypoxic culture is due not to increased transcription of *narGHJI,* which appears to be constituently expressed, but to increased levels of the transporter *narK2* (145). It has been proposed that nitrate is reduced by a nitrate reductase (NarGHJ) to produce nitrite and this excess nitrite, which is toxic in large amounts, is then excreted by a nitrite extrusion protein (NarK1, NarK2, NarK3) (185). Nitrate reductase does not appear to support anaerobic growth of *M. tuberculosis,* as microaerobic conditions develop the bacilli enter a state of non-replicating persistence. The role for nitrate reductase in *M. tuberculosis* could therefore be in redox balancing, or it may serve a temporary function to provide energy during shift-down to non-replicating persistence.

Enzyme/function	**Gene name**	***M. smegmatis* mc² 155**	***M. tuberculosis* H37Rv**
Assimilatory nitrite reductase	*nirB*	*msmeg_0427*	*Rv0252*
	nirD	*msmeg_0428*	*Rv0253*
Assimilatory nitrate reductase	*narI*	*msmeg_5137*	*Rv1164*
	narJ	*msmeg_5138*	*Rv1163*
	narH	*msmeg_5139*	*Rv1162*
	narG	*msmeg_5140*	*Rv1161*
	narX	n/a	*Rv1736c*
Nitrite/nitrate transporter	*narK*	*msmeg_5141*	n/a
	narK1	n/a	*Rv2329c*
	narK2	n/a	*Rv1737c*
	narK3	*msmeg_0433*	*Rv0261c*
	narU	n/a	*Rv0267*

Table 1.1. Genes involved in the reduction of nitrate to ammonium in *M. tuberculosis* and *M. smegmatis*.

Adapted from (5).

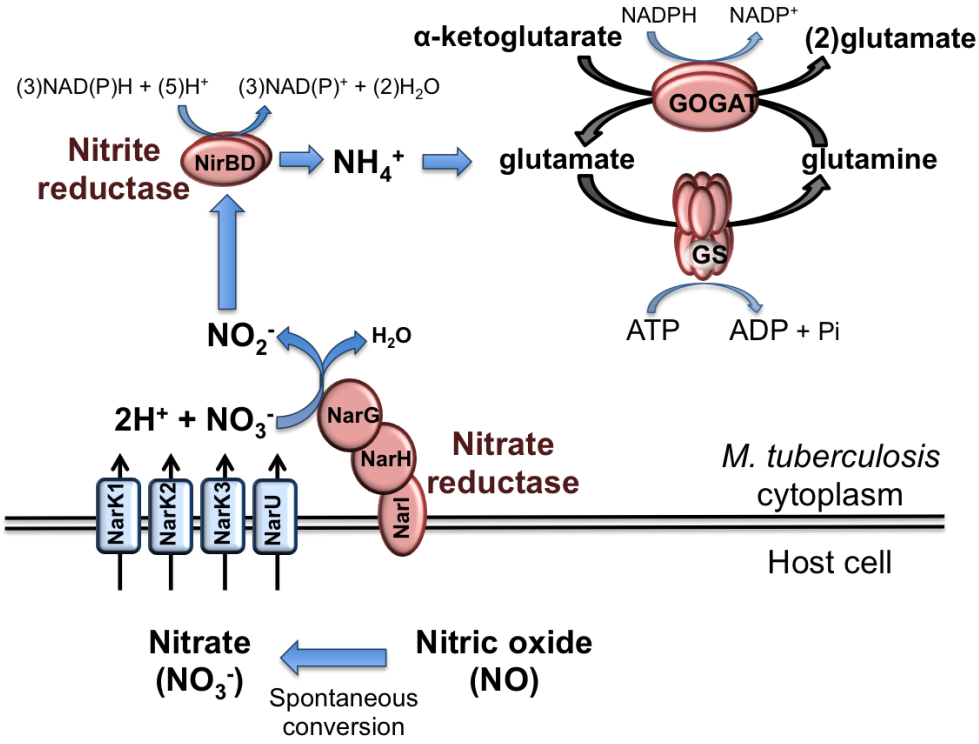


Figure 1.2. Biochemical pathway of nitrate assimilation in mycobacteria via nitrate reductase (NarGHJI), nitrite reductase (NirBD) and GS (GlnA1).

Exogenous nitrate is transported inside the cell through nitrate transporters (NarK2 and potentially NarK1, NarK3 and NarU) and reduced to nitrite by nitrate reductase (NarGHJI), using the quinol pool as electron donor. Nitrite is subsequently reduced to ammonia by nitrite reductase (NirBD) using the NADH pool as electron donor. Ammonium is assimilated into the central metabolite glutamate using the GS-GOGAT pathway.

Amino acids

Many micro-organisms utilise amino acids as an energy and nitrogen source, in addition to biosynthetic purposes. *M. tuberculosis* can utilise a variety amino acids as a nitrogen source, including L-asparagine, L-alanine and L-glycine (91). In addition, studies with *M. smegmatis* demonstrated that both D- and L-isomers of alanine, glutamic acid, and valine were taken up by the bacilli, however the uptake of L-amino acids was greater than that of corresponding D-amino acids (199).

Lyon (1974) noted that L-asparagine, as a nitrogen source, was the preferred amino acid for the growth of mycobacteria (90). Washed cells as well as cell-free preparations of *M. tuberculosis* H37Rv, H37Ra, *M. smegmatis* and *M. bovis* Bacillus Calmette–Guérin (BCG) could deamidate L-asparagine to aspartic acid and ammonia, by the action of an asparaginase enzyme (112). No difference in activity was reported between asparaginase from *M. smegmatis* and *M. tuberculosis* other than the amount present in the extracts (112). Of note no ammonia was formed when D-asparagine was incubated with cell-free extract and D-asparagine inhibited the formation of ammonia from L-asparagine by extracts from H37Ra (112). Analysis of culture filtrates of H37Ra in media containing L-asparagine sole amino acid, displayed an accumulation of extracellular amino acids (aspartic acid, glutamine, alanine, and lysine) (90). These particular amino acids were rapidly utilised after the disappearance of asparagine from the medium. The results reveal a rapid but inefficient metabolism of asparagine and its preferential utilisation in the presence of additional amino acids.

The utilisation of alanine as a nitrogen source has also been documented previously, focusing on the role of an alanine dehydrogenase. Alanine dehydrogenases (Ald) are well-studied enzymes found in a wide range of bacterial species. In mycobacteria, Ald was first identified as an enzyme absent in vaccine strains of *M. bovis* BCG, but present in virulent *M. tuberculosis* (31). Ald catalyses the oxidative deamination of L-alanine to pyruvate and ammonia (catabolic reaction) or, in the reverse direction, the reductive amination of pyruvate to L-alanine (biosynthetic reaction) (45). Studies with the Ald enzyme of *M. tuberculosis* (Rv2790) and *M. smegmatis* (msmeg_2659) suggest that its primary role is the catabolism of alanine for nitrogen utilisation; *ald* null mutants displayed impaired growth with alanine as the sole nitrogen source (45, 51). In addition, microarray analysis of the *M. tuberculosis ald* transcript demonstrated that it was overproduced under nutrient starvation and under hypoxic conditions (20, 143). In *M. smegmatis* increased Ald activity was observed in cells grown under anaerobic conditions, and analysis of an *ald* null mutant demonstrated that Ald is necessary for sustained anaerobic growth (45). The possible role for Ald under anaerobic conditions was suggested to be involved

in NADH recycling, however this requires further investigation. The role of Ald in both nitrogen assimilation and hypoxia, is intriguing, suggesting overlap between mechanisms for the two pathways in mycobacteria.

Urea

Urea has been proposed to be available to *M. tuberculosis* in both its intracellular and extracellular locations within the host (87). Ammonia generated by the action of urease may serve one of two functions during infection; it may contribute to alkalizing the microenvironment of *M. tuberculosis*, preventing phagosome-lysosome fusion/ phagosome acidification and ammonia generated would be available to the bacilli for assimilation of nitrogen into biomolecules (35, 87). The concerted action of urease and GS could therefore serve to scavenge and assimilate environmental nitrogen during infection.

M. tuberculosis expresses a functional urease *ureABCDFG*; Rv1848-Rv1853 (35). Lin *et al* 2012 investigated the effect of a urease-deficient *M. tuberculosis* strain and confirmed the alkalizing effect of the urease activity within the mycobacterium-containing vacuole in resting macrophages. However, this was not detected in the more acidic phagolysosomal compartment of activated macrophages (87). In addition, the urease-mediated alkalizing effect did not confer any growth advantage on *M. tuberculosis* in macrophages, suggesting the alkalizing effect provided by the mycobacterial urease activity is somewhat modest (87).

M. tuberculosis is able to assimilate urea for growth and this ability is urease dependent. A *M. tuberculosis* urease deletion mutant displayed impaired growth *in vitro* when urea was provided as the sole source of nitrogen (87). It was demonstrated that ammonia arising from ureolysis is the actual nitrogen source utilised by *M. tuberculosis* for its *in vitro* growth (87). Therefore, it is proposed that *M. tuberculosis* assimilates urea via its urease activity and that this process generates ammonia, for use as a nitrogen source during growth (Figure 1.3). Despite this, a urease-deficient mutant did not have altered growth phenotype in macrophages in culture medium containing a variety of carbon and nitrogen sources (87). It is possible that other readily available nitrogen sources within the host cell, such as ammonia from the metabolism of nitrogenous compounds and nitrate, could bypass the need for *M. tuberculosis* to metabolise urea. The ability to utilise urea as a nitrogen source may however, be critical at specific sites of infection where other sources of nitrogen are limited, for instance during intestinal tuberculosis infection, urea-containing body fluids such as saliva and tissue exudates could provide a constant source of energy for the bacilli. Thus, the absence of *in vivo* phenotype of a urease-

deficient *M. tuberculosis* mutant potentially reflects the metabolic versatility of *M. tuberculosis* with the ability to adapt to any microenvironment encountered in its host (87).

For *M. smegmatis* two putative urease encoding operons were found. Only msmeg_3622-3627 exhibits homology to the *ure* gene clusters in *M. tuberculosis* and *M. bovis* based on gene identity and arrangement (5). The second (msmeg_1091-1096) exhibits similarity to the urease subunits encoding genes from proteobacteria, with DNA sequence identities between 60 and 70%, and an identical operon arrangement (5). No urease operon or urease-related genes were found in *M. avium* and *M. leprae* (5). Also, *M. smegmatis* is the only mycobacterial species to feature a distinct operon (msmeg_2978-2982) encoding the subunits of a putative urea ABC transporter (5). It has been reported previously in *M. smegmatis*, that the level of urease activity in the crude bacterial extract is 11-fold greater than for *M. tuberculosis* (35). Emphasising, along with the presence of an increased number or urea utilisation genes, the potential importance of urea as a nitrogen source for this species.

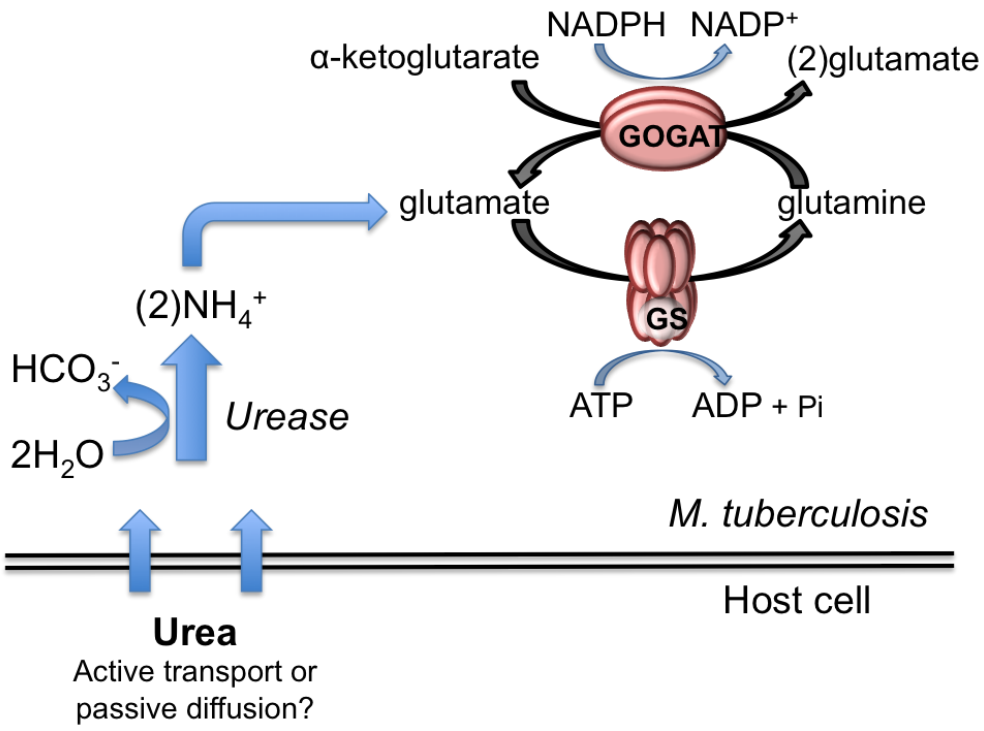


Figure 1.3. Proposed route of urea assimilation in *M. tuberculosis.*

Urea enters *M. tuberculosis* via possible diffusion across the cell membrane. Urea is then converted to ammonium via a urease enzyme. Ammonium can then enter the GS/GOGAT pathway for assimilation into biomolecules. Adapted from (87).

1.2.3 Nitrogen Availability Summary

From the literature it is apparent that the uptake and utilisation of nitrogenous compounds by mycobacteria is not well understood. Uncertainty exists as to the main, essential nutrients utilised by *M. tuberculosis* whilst residing inside the host, which may depend on the location of *M. tuberculosis* inside the human body. This could vary from phagosomes in macrophages and dendritic cells to granulomas and even fat cells (106, 135, 173). As such, there is a need to determine which nutrients are available in different environments and identify the proteins that are employed by *M. tuberculosis* to utilise these. Current data obtained, regarding the utilisation of various nitrogen sources for *M. tuberculosis,* supports the concept that its virulence correlates with metabolic versatility and an ability to utilise a variety of nitrogen sources available in its environment.

1.3 Nitrogen Assimilation Enzymes

Two nitrogen incorporating mechanisms are present in most prokaryotes; the glutamate dehydrogenase (GDH) pathway and the glutamine synthetase/glutamate synthase pathways (GS/GOGAT) (For a review see (4, 58)). Both pathways incorporate nitrogen into glutamate or glutamine, which form the major biosynthetic donors for all other nitrogen containing components in the cell. Glutamine is a source of nitrogen for the synthesis of purines, pyrimidines, asparagine, glucosamine and a variety of amino acids. Conversely, glutamate provides nitrogen for most transaminases, and is responsible for 85% of nitrogenous compounds within a cell (59). GDH is the energetically more favourable nitrogen incorporating pathway, as GS/GOGAT utilises an additional ATP per molecule of ammonium assimilated (58). Consequently, during high nitrogen levels the GDH pathway is active. However, at low nitrogen levels the GDH pathway is not sufficient due to the low affinity of the enzyme for ammonium; instead during nitrogen starvation the higher affinity GS/GOGAT system is active (58).

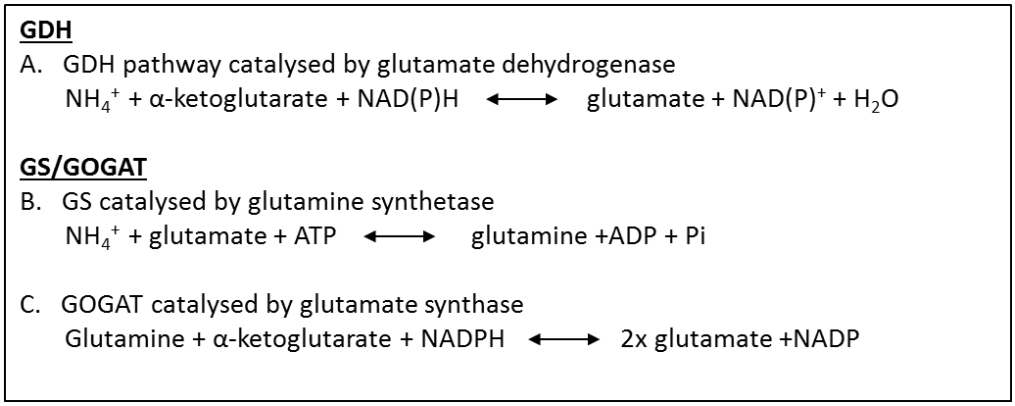


Figure 1.4. Reaction mechanism of enzymes responsible for ammonium assimilation.

Reaction A, catalysed by glutamate dehydrogenase (GDH), reduces NAD(P)H to produce glutamate from α-ketoglutarate and ammonium. Reaction B, glutamine synthetase (GS) utilises ATP in the formation of glutamine from ammonium and glutamate. Reaction C, glutamate synthase (GOGAT) catalyses the formation of glutamate from glutamine and α-ketoglutarate, with the reduction of NADPH. Reactions B and C supply glutamate and glutamine to each respective reaction producing a cyclic pathway.

1.3.1 Nitrogen Assimilation Enzymes: GDH

Glutamate dehydrogenase (GDH) enzymes catalyse the reversible amination of α- ketoglutarate to form glutamate, with the concomitant reduction of NAD(P)H. Furthermore, GDH enzymes serve as metabolic branch enzymes regulating a flux of intermediates, such as α- ketoglutarate, between the Krebs cycle and nitrogen metabolism (24). In prokaryotes, GDH enzymes function with co-factors, $NADP^+$ (EC 1.4.1.4) or NAD^+ (EC 1.4.1.2). *M. smegmatis, M. tuberculosis* and *M. bovis* all encode for a putative NAD^+ GDH enzyme. This NAD^+ GDH protein has been captured from cell free extracts of *M. tuberculosis* and *M. smegmatis,* and its activity detected in *M. smegmatis* (59, 111). Additionally, the *M. smegmatis* genome is thought to encode a second putative NAD^+ GDH, and a single $NADP^+$ specific GDH enzyme (59, 139). The presence of additional GDH genes in *M. smegmatis* suggests that this pathway is, to a greater extent, central to nitrogen utilisation than in other mycobacterial species.

Recently it has been reported that the activity of NAD^+ GDH, in *M. smegmatis* and *M. tuberculosis,* is modulated by a small soluble protein, glycogen accumulation regulator (GarA; msmeg_3647, Rv1827) (111). GarA is highly conserved among actinomycetes; Rv1827 shares 82% identity at the amino acid level with Odh1, the GarA analogue in *C. glutamicum* (107, 178). Native or

unphosphorylated GarA has been demonstrated to interact with NAD^+ GDH of *M. tuberculosis* and *M. smegmatis* from cell extracts (111). Binding leads to a reduction in NAD^+ GDH activity, in *M. smegmatis,* by altering the affinity of the enzyme for its substrate (111). Phosphorylation of GarA, by a serine/threonine protein kinase PknG (Rv0410c, msmeg_0786), prevented binding to NAD^+ GDH (111, 178). The condition under which PknG is stimulated to phosphorylate GarA has not yet been investigated, and it is not clear how this relationship may affect nitrogen metabolism in mycobacteria.

1.3.2 Nitrogen Assimilation Enzymes: GS

Glutamine synthetases (GS; L-glutamate ammonium ligase; EC 3.6.2) are involved in the ATP-dependant synthesis of glutamine from glutamate and ammonium. The genome sequence of *M. tuberculosis* contains four *glnA* gene copies encoding independent GSs; *glnA1, glnA2, glnA3* and *glnA4* (36). Of these *glnA1*, *glnA3* and *glnA4* produce L-glutamine, with *glnA2* synthesising D-glutamine and D-isoglutamine used in cell wall biosynthesis (61). To date only the major GS, *glnA1,* has been demonstrated to be essential in *M. tuberculosis* with deletion mutants auxotrophic for glutamine and growth attenuated in the macrophage and guinea pig model of infection (171, 172). Activity of the other enzymes has been demonstrated by Harth *et al*. 2005, but their essentiality and physiological role have yet to be verified (61). It is therefore considered that GlnA1 is the major GS used for glutamine synthesis in mycobacteria.

In addition to encoding enzymes that catalyse glutamine synthesis, evidence suggests that in *M. tuberculosis* GlnA1 has evolved to perform other specialised functions, not present in non-tuberculosis causing mycobacteria. In pathogenic mycobacterial species it has been established that GlnA1 is exported from the cell to the extracellular milieu (60, 125). Increasing evidence suggests that extracellular GlnA1 is implicated in the production of poly-L-glutamine-glutamate, a polymer found only in pathogenic mycobacterial cell-wall. An *M. bovis* mutant, lacking a functional *glnA1* gene, contained no detectable poly-L-glutamine in its cell wall and showed marked sensitivity to a variety of chemical and physical stresses (30). It has also been speculated that this extracellular GlnA1 activity may modulate phagosome pH, thereby preventing phagosome lysosome fusion (126). In this instance, adaptive evolution may have led to functional promiscuity whereby GlnA1 exerts other functions, whilst maintaining the same active site for the original singular activity.

The genome of *M. smegmatis* encodes for each of the four classes of GS proteins found in *M. tuberculosis* (3). Of these homologues msmeg_4290 has the greatest amino acid identity with

glnA1 of *M. tuberculosis*, encoding for a type 1 ammonium assimilatory enzyme (61). In contrast to the GlnA1 of *M. tuberculosis, M. smegmatis* GlnA1 does not appear to be expressed at such a high level, nor does it appear to be exported to the extracellular milieu (60, 172). Additionally, the *M. smegmatis* genome contains open reading frames encoding GS-like proteins with unknown physiological function (msmeg_1116, msmeg_3827, msmeg_5374 and msmeg_6693) (5). These additional proteins have similarities shared with other saprophytic soil dwelling organisms; msmeg_5374 shares 51% identity with GS-like proteins in the soil dwelling α-proteobacteria *Rhodopseudomonas palustris* (5). This suggests that additional *M. smegmatis* GS-like genes have been acquired through horizontal gene transfer, but their role in nitrogen metabolism has yet to be defined.

Regulation of GS activity is essential, as failure would lead to depletion of intracellular levels of glutamate and ATP. Post-translational modification of GS in *E. coli* is controlled by a regulatory cascade of three proteins GlnD, GlnB (P_{II}) and GlnE (Reviewed; (4, 83, 100, 127)). GlnD is an uridylyl transferase, which modifies the GlnB (P_{II}) protein adding a UMP residue during nitrogen limiting conditions. The status of GlnB (P_{II}) in turn controls the activity of GlnE; GlnB binding to GlnE promotes the adenylylation reaction of GlnE to GS, while GlnB-UMP binding to GlnE promotes the deadenylylation reaction of GlnE to GS (48). Therefore, under conditions of nitrogen excess the GlnE enzyme transfers an adenylyl group (AMP), from ATP, to the tyrosine hydroxyl on GlnA1, converting it into the adenylylated, inactive enzyme (GS-AMP). GS-AMP is more sensitive than active GS to feedback inhibition by the end products of glutamine metabolism (4). The GlnA1 GS enzyme is composed of 12 identical subunits arranged in two superimposed hexagonal rings (52, 60). Activity of GlnA1 GS is directly proportional to the number of subunits that are adenylylated; GS0 is an enzyme that does not carry an AMP moiety, while GS12 carries a moiety on each subunit. GS can therefore display a range of activities depending on the degree of adenylylation.

In contrast to other bacteria, the *glnE* gene is essential in *M. tuberculosis,* with this requirement linked to its adenylylation activity of GlnA1 (28, 116). GS does undergo a change in adenylylation state in response to nitrogen availability as described in *E. coil*, however the mechanism of this modification is unclear (193). Both *M. tuberculosis* and *M. smegmatis* have a single P_{II} homologue (GlnK), contained in an operon *amtB-GlnK-GlnD* (5). Recent, experimental determination of the status of GlnK during nitrogen limitation demonstrated that it is susceptible to adenylylation by GlnD. However, work with a *glnD* knockout mutant displayed no altered adneylylation phenotype of GS, suggesting modification of GS is not linked to the adenylylation state of GlnK (193). This is in agreement with observations made in *S. coelicolor* and *C. glutamicum* (64, 158). Thus the targets for GlnD and GlnK (P_{II}) in actinomycetes are still

to be elucidated, but it appears their function is not analogous to the role in *E. coli* GS modification.

1.3.3 Nitrogen Assimilation Enzymes: GOGAT

GOGAT (Glutamate synthase; glutamine amide-2-oxoglutarate aminotransferase (NADPH), EC 1.4.1.13) synthesizes glutamate during nitrogen-limiting conditions (127, 159). The GOGAT enzyme exists as a heterodimeric protein consisting of a large and small subunit, forming a tetrameric holoezyme (159). Synthesis of glutamate via the GOGAT pathway occurs via a two-step process. Firstly GS is involved in the amidation of endogenous glutamate to glutamine. Glutamine is then fed into the GOGAT pathway and with the concomitant reduction of NADPH, GOGAT catalyses the reductive transfer of the glutamine amide-nitrogen to α-ketoglutarate (2-oxoglutarate). Thereby the GOGAT pathway produces a net synthesis of 2 molecules of L-glutamate per L-glutamine molecule (159).

In silico analysis demonstrated that all mycobacterial genomes contain highly conserved operons encoding the large (*gltB*) and small (*gltD*) subunits of GOGAT (5). In addition the genome of *M. smegmatis* contained several additional copies of *gltB* (msmeg_5594, msmeg_6263, msmeg_6459) and *gltD* (msmeg_6262, msmeg_6458) that are not found in other mycobacteria (5). Again the presence of these additional GOGAT subunits have yet to be further investigated, and as such their role in nitrogen metabolism is unknown.

1.4 Transcriptional Regulation of Genes Involved in Nitrogen Metabolism

Regulation of nitrogen metabolism occurs predominately on two levels; transcriptional regulation of genes implicated in nitrogen metabolism and post-translational control of enzyme activity within the nitrogen assimilatory pathway. Previous work, involving nitrogen regulation in mycobacteria, focused largely on the post-translational control of enzymes such as GS and more recently GlnK (28, 193). Other than this, little information has been established regarding the transcriptional control of nitrogen metabolism in mycobacteria.

1.4.1 Transcriptional Regulation in *E. coli*: The NtrB/C Response

Mechanisms involved in the transcriptional response to nitrogen availability have been extensively investigated in *E. coli*. In *E. coli,* coordination of gene expression for nitrogen assimilation is controlled by the NtrB/NtrC two-component system. The signal relay involves four key components; the uridylyltransferase GlnD, P_{II}-type signal transduction protein GlnB, histidine kinase NtrB and its corresponding response regulator NtrC (for reviews see; (83, 100, 127)). An interplay of these molecules senses the nitrogen status of the cell, and coordinates the transcriptional response accordingly.

In *E. coli* the indicator of nitrogen availability is the intracellular concentration of glutamine and α-ketoglutarate. During nitrogen excess conditions the level of glutamine is high compared to α-ketoglutarate; this ratio changes under nitrogen limitation (76). Under nitrogen excess conditions GlnD binds glutamine, which is accumulated in the cell. This leads to a conformational change of GlnD permitting deuridylylation of GlnB (P_{II}) (75). Unmodified GlnB (P_{II}) binds to NtrB activating its phosphatase activity leading to dephosphorylation and subsequent inactivation of NtrC under nitrogen surplus (75).

During growth in nitrogen limiting conditions the nitrogen-regulated (Ntr) response is stimulated. Nitrogen limitation is detected via a rise in α-ketoglutarate concentration, and binding of ATP and α-ketoglutarate to GlnB (P_{II}). This induces a conformational change resulting in uridylylation of GlnB (P_{II}) by GlnD. Under these conditions the GlnB-UMP dissociates from the cytoplasmic sensor histidine kinase NtrB. This dissociation activates the kinase activity of NtrB, which leads to autophosphorylation at histidine residue 139 (74). Subsequently NtrB phosphorylates its corresponding response regulator NtrC at a conserved residue, aspartate 45, which in turn activates gene transcription from σ^{54} dependent promoters (79, 187). In total,

approximately 100 genes are regulated by NtrB/C during nitrogen limitation, including genes for transcriptional regulation *glnLG* (NtrB/C) and *nac* (Nitrogen assimilatory control protein; an adapter for σ^{70} dependent genes), a P_{II} signal transduction protein *glnK*, and genes involved in ammonium utilisation *glnA* (GS) and *amtB* (ammonium transporter) (127, 203).

Regulation of NtrB/C also occurs at the transcriptional level upon nitrogen starvation (reviewed in (100)). The genes encoding NtrB (*glnL*) and NtrC (*glnG*) are located in an operon *glnALG*, in which *glnA* encodes GS. Transcription of these genes is strictly regulated by the usage of different promoters. The *glnA* gene is expressed from tandem promoters *glnA*p1 and *glnA*p2, whereas the downstream *glnLG* genes are expressed either by read through from the *glnA* promoters or from a separate promoter p*glnL*, located between *glnA* and *glnLG*. Under nitrogen-sufficient conditions, *glnA* is expressed at low levels from *glnA*p1, which is transcribed by the major vegetative RNA polymerase, σ^{70} and expression of *glnLG* occurs primarily from p*glnL* (128). Under nitrogen limiting conditions, the transcriptional regulator NtrC is activated, which inactivates transcription from *glnA*p1 by competitive binding. NtrC binds to its enhancer binding sites upstream of the high affinity *glnA*p2 leading to elevated sigma σ^{54} dependent expression of *glnA* and *glnLG* (128). This mechanism guarantees low level of GS, NtrB and NtrC under nitrogen surplus, as well as increased protein concentrations under nitrogen limitation.

1.4.2 Transcriptional Control in Response to Nitrogen Limitation in Actinomycetes

NtrC/B homologues have not been identified in any actinomycetes genomes, suggesting a different mechanism of transcriptional regulation of genes involved in nitrogen metabolism. *In silico* analysis of the *M. smegmatis* genome found two nitrogen metabolism transcriptional regulators known in other actinomycetes. These regulators were AmtR, sharing 42% identity with AmtR of *C. glutamicum*, and GlnR sharing 60% identity to GlnR of *S. coelicolor*, both at the amino acid level (5). Other mycobacterial species investigated only revealed high identity with their GlnR sequences, *M. tuberculosis* displayed 65% identity with that of *S. coelicolor* (165). Involvement of the second regulator, AmtR, in other mycobacterial species is questionable; *M. tuberculosis* Rv3160c possesses 27.9% homology with AmtR of *C. glutamicum* (58). Despite this, Betts *et al.* detected a reduction in transcription of Rv3160c in a nutrient starvation model (20). This reduction in Rv3160c transcription may suggest that AmtR is in fact an additional regulator of nitrogen metabolism in *M. tuberculosis*, despite its low sequence homology with other actinomycetes.

1.4.3 Transcriptional Regulator GlnR

Transcriptional Regulation in *Streptomyces* by GlnR

GlnR is an OmpR-type transcriptional regulator, first identified in *Streptomyces coelicolor* as a protein able to restore glutamine auxotrophy (197). A variety of studies have been conducted in *Streptomyces* to determine the GlnR regulon, these include transcriptional and proteomic analysis with a *glnR* deletion strain and ChIP-chip analysis. From these data at least 50 genes in *S. coelicolor* and 44 genes in *Streptomyces venezuelae* have been determined to be GlnR regulated in response to nitrogen limitation (120, 164, 165). The highly conserved *amtB-glnK-glnD* operon has been demonstrated to be under GlnR control in both *S. coelicolor* and *S. venezuelae*, activated in nitrogen limitation (47, 120). Further genes under positive GlnR control include *glnA1* (GS), *gltBD* (GOGAT), *nirBD* a nitrate reductase, *nnaR* nitrate/nitrite regulator and *nasA* a periplasmic nitrate reductase (2, 165, 181). Interestingly, GlnR can also act as an inhibitor of transcription, repressing the transcription of the *gdhA* (GDH)*, ureA* and various other uncharacterised ORFs (165). This suggests a role for GlnR in both induction and repression of gene expression during nitrogen limitation.

The phosphate response regulator, PhoP, has been reported to have a direct, negative effect on transcription of *glnR* and its most prominent targets in *S. coelicolor* (131, 132). PhoP competes with GlnR in binding to overlapping sites on the DNA, and can also act as a physical block of transcription, depending on the structure and organisation of the gene's promoter region (132). PhoP has been demonstrated to block transcription of *glnA, glnII* and *amtB,* as well as *glnR*. Paradoxically, transcriptome studies showed no response of the GlnR regulon to phosphate limitation (131). As such, the physiological relevance of this observation, between the control of phosphate and nitrogen metabolism, is still debatable.

Signal transduction and regulation of GlnR are still unknown. Observations based on the conserved phosphorylation domain, analogous to the situation in *E. coli*, suggests involvement of a protein kinase (47, 201). OmpR-like proteins regularly have a conserved aspartate residue in their N-terminal domains, which are sites of specific phosphorylation. In addition conserved serine/threonine and tyrosine residues are also assumed to be involved in phosphotransfer. GlnR of *S. coelicolor* does contain a conserved aspartate residue (D-50) at the potential phosphorylation site, as well as a tyrosine residue corresponding to the OmpR T-83. However, a valine residue (V-95) occupies the position normally allocated to the conserved tyrosine T-83 in OmpR (47). Therefore GlnR may be subject to phosphorylation by a sensor kinase, nevertheless no operon linked sensor kinase has been identified. Thus the corresponding protein kinase and phosphorylation state of GlnR during nitrogen limitation remain unknown.

Role of GlnR in Mycobacteria

Using the GlnR binding motif of *S. coelicolor* putative GlnR binding sites were found in all the available mycobacterial genomes. Three highly conserved *cis* elements were found in *M. smegmatis,* suggesting a similar mechanism of DNA protein interaction (3). These putative binding motifs were located upstream of *amtB, amt1* encoding ammonium transporters and *glnA1* encoding GS (3). As *amtB* is transcribed in an operon, *amtB-glnK-glnD,* these two signal transduction proteins are also assumed to be under GlnR transcriptional control. Analysis of a GlnR deletion mutant confirmed the role of the putative GlnR binding sites; during nitrogen starvation transcripts of *glnA1*, *amtB* and *amt1* showed an increase in WT, yet no increase was observed in the mutant (3). In *M. tuberculosis* GlnR was demonstrated to positively regulate *nirB* a nitrite reductase, and the transcriptional role of GlnR was deemed essential in nitrate/nitrite utilisation (95). In total, seven genes have been identified in the mycobacterial GlnR regulon, much lower than the total number of genes identified in Streptomyces. Interestingly, GlnR transcription itself did not alter under nitrogen starvation conditions suggesting *glnR* transcription is not subject to nitrogen control (3). This conflicts results observed in *S. coelicolor* that suggests transcription of GlnR is dependent on the nitrogen status of the cell (165). GlnR in mycobacteria may therefore be subject to posttranslational modification in response to nitrogen limitation as with other OmpR family members.

Mechanism of OmpR Transcriptional Regulation in *E. coli*

According to sequence homology GlnR is part of the OmpR-family of transcriptional regulators (3, 47). Typically, OmpR family proteins contain a conserved N-terminal phosphorylation domain that controls activity of the C-terminal DNA-binding domain in a phosphorylation dependant manner (23). The mechanism of OmpR activation by EnvZ, the corresponding histidine kinase, in *E. coli* has been well described and represents a two-component His-Asp phosphorelay signal transduction system (reviewed in (156)). In response to changing osmolarity EnvZ is autophosphorylated at histidine residue 243, and subsequently transfers this phosphate group to OmpR at aspartic acid 55 (39, 130). EnvZ has kinase activity to phosphorylate OmpR, and also phosphatase activity toward phosphorylated OmpR (OmpR-P). Thus EnvZ it is able to regulate the level of OmpR-P in the cell depending upon the medium osmolarity. At low medium osmolarity, the phosphatase activity is relatively higher than the kinase activity so that the cellular concentration of OmpR-P is maintained at a lower concentration .

OmpR regulation of two genes, *ompC* and *ompF,* has been extensively investigated. OmpC and OmpF are outer membrane porins; OmpF is mainly present at low osmolarity and possesses a larger pore and a higher flow rate than OmpC, consequently the OmpC porin dominates at high osmolarity. Upstream of the *ompF* and *ompC* promoters are four F sites (F1, F2, F3, and F4) and three C sites (C1, C2, and C3) respectively. Each site consists of 20 base pairs, providing tandem-binding sites for two OmpR-P molecules (19, 66). The hierarchy of OmpR-P binding to these F and C sites was determined to be F1, C1 > F2, F3 > C2 > C3 (19). At low osmolarity OmpR-P cooperatively binds to F1-F2/F1-F2-F3 to activate *ompF* transcription. Under this condition of low osmolarity, only the C1 site is occupied by OmpR-P, which is not sufficient to activate *ompC* transcription. However, high osmolarity leads to a higher amount of OmpR-P in the cells, and at this elevated level OmpR-P also occupies C2 and C3, resulting in *ompC* expression. In addition, this increase in OmpR-P results in binding to the F4 site, a weak OmpR-P-binding site. OmpR-P binding to the F4 site is proposed to form a loop that interacts with OmpR-P molecules binding to F1, F2, and F3, thereby blocking *ompF* transcription (Figure 1.5) (19). Within each of these OmpR binding sites (F1, F2, F3, F4, C1, C2 and C3) there are two 10 bp tandem subunits, a low affinity "a site" and a high affinity "b site" (201). Yoshida et al., 2006 proposed a hierarchic model of DNA binding in a "discontinuous, galloping manner"; OmpR-P first binds with high affinity to the "b site", phosphorylation results in stable protein-protein interactions and leads to subsequent recruitment of additional OmpR-P molecules the "a site" (Figure 1.5)(201).

A discontinuous galloping manner of GlnR:DNA binding has also been proposed in *Streptomyces*. This is based on sequence analysis of consensus binding sites in *Streptomyces*. In *S. coelicolor* a low affinity "a-site' and a higher affinity "b-site" were identified via sequence analysis (165). However in *S. venezuelae* two "a-sites" were identified, rather than a "b-site" (120). Further investigation in *Streptomyces* needs to be conducted to establish whether these conserved sites represent a similar manor of GlnR:DNA interaction as proposed in OmpR. However, these results are interesting, suggesting a conserved mechanism of OmpR-family member DNA interaction.

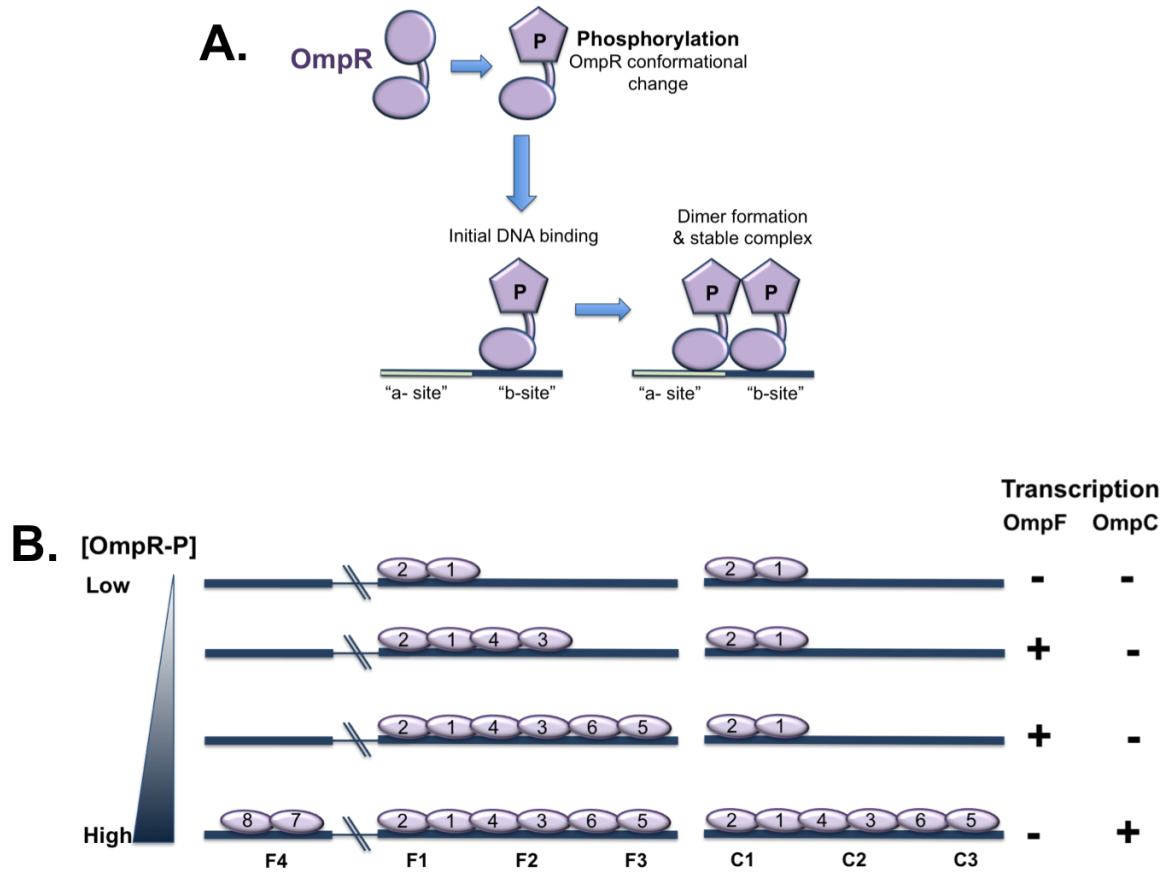


Figure 1.5. Model of OmpR transcriptional activation in *E. coli*.

A. Activation of OmpR and the "galloping" model of OmpR-P:DNA binding. Phosphorylation, via histidine kinase EnvZ, leads to an OmpR conformational change. OmpR-P binds to the high affinity "b site" preferentially. Recruitment of a second OmpR-P to the "a site" results in dimer formation and a stable complex. Only OmpR-P dimers able to completely bind DNA.

B. System of OmpR-P recruitment to binding sites, leading to activation or repression of *ompF* and *ompC*. The oval shape represents an OmpR-P molecule. OmpR-P molecules bind in the order depicted by the numbered molecules. F1, F2, F3, and F4 represent the regulatory region of the *ompF* promoter. C1, C2, and C3 represent the regulatory region of the *ompC* promoter. Adapted from (201).

1.4.2.2 Transcriptional Control in Response to Nitrogen Limitation in *C. glutamicum* by the Response Regulator AmtR

Compared to the extended *S. coelicolor* GlnR regulon, *M. smegmatis* GlnR has a reduced number of target genes (5). Genes that are under control of GlnR in *S. coelicolor,* for example the NADP-dependent glutamate dehydrogenase gene *gdhA*, the urease operon and the nitrite reductase genes, do not contain GlnR *cis* acting elements in *M. smegmatis* (3). AmtR may therefore be an interesting candidate for a second nitrogen regulator, especially since the urease encoding genes in *C. glutamicum* are under AmtR control.

Little analysis has been carried out with the AmtR response regulator in *M. smegmatis*, more information is known about the closely related bacteria *C. glutamicum*. In *C. glutamicum* AmtR governs the nitrogen-starvation-dependent gene expression (69). AmtR blocks transcription of various genes during growth in nitrogen rich-medium. By a combination of bioinformatics, transcriptome and proteome analysis at least 35 genes have been revealed to be directly controlled by the AmtR repressor protein (15, 25). These include genes encoding transporters and enzymes for ammonium assimilation (*amtA*, *amtB*, *glnA*, *gltBD*, *dapD*), creatinine metabolism (*codA*, *crnT*), urea metabolism (*urtABCDE*, *ureABCEFGD*), signal transduction proteins GlnK and GlnD, and a number of biochemically uncharacterised enzymes and transport systems (for review see (27)).

AmtR is a TetR family member, sharing homology within their C-terminal DNA binding domain, binding in a helix-turn-helix manner (122). The TetR family encompasses a number of repressor proteins involved in adaptation to environmental changes. For family members whose function has been characterised, regulation of DNA binding is influenced by binding of a small inducer molecule to the non-conserved, N-terminal domain. Binding of the effecter molecule produces a conformational change in the conserved DNA binding region, resulting in a release of the repressor from its operator allowing DNA transcription (122).

In contrast to most TetR-type regulators, the dissociation of AmtR from its target promoters in *C. glutamicum* is not triggered by the binding of a low-molecular weight ligand, but by a complex formation with the P_{II}-type signal transduction protein GlnK (15). GlnK is a tetrameric, P_{II} type signal-transduction protein, expressed in an operon with *amtB*. GlnK, under nitrogen excess conditions, is found sequestered to the cytoplasmic membrane by AmtB (Figure 1.6 A). During nitrogen limitation, an unknown sensory kinase detects the nitrogen status of the cell and activates GlnD to modify GlnK. GlnD then adds an AMP moiety, at tyrosyl residue 51, to GlnK. GlnK is subsequently released from AmtB, relocating GlnK-AMP to the cytoplasm (158). GlnK-AMP then interacts with AmtR releasing it from its target DNA, removing its repressor activity

(Figure 1.6 B) (15). GlnD is a bi-functional protein capable of catalysing adenylylation and de-adenylylation of GlnK. Interestingly truncated C-terminal GlnD is still active for GlnK modification, however lacked the ability to demodify GlnK (158). Of note, this kind of GlnK modification is observed in mycobacteria and *S. coelicolor*, yet the targets of GlnK are currently unknown (64, 193).

The signal to which AmtR responds to fluctuating nitrogen levels in *C. glutamicum* is unknown. *C. glutamicum* accumulates large intracellular pools of amino acids such as glutamate, which respond slowly to changes in nitrogen availability, making them unlikely signals for cellular nitrogen status (108). Intracellular glutamine concentrations have been shown to vary in response to nitrogen availability, however the cellular transcriptional response of *gltB*, a gene demonstrated to be nitrogen regulated, did not alter (103). In addition, protein sequence analysis of GlnD in *C. glutamicum* indicated that the protein does not have a ligand binding domain, as found in *E. coli* which binds glutamine, implying that GlnD does not sense the nitrogen status of the cell (166). Supporting this hypothesis, transcription of *glnD* is not constitutive, but varies in response to nitrogen availability, suggesting that GlnD would provide a poor sensory mechanism to nitrogen availability (108). Muller *et al* (2006) demonstrated that intracellular α-ketoglutarate and ammonium levels altered rapidly in response to changes in nitrogen availability, which corresponded to GlnK adenylylation and the transcriptional response of *gltB* (103). Thus, both ammonium and α-ketoglutarate may play a role as indicators of cellular nitrogen status; however, further experimental evidence regarding the nitrogen sensors is required.

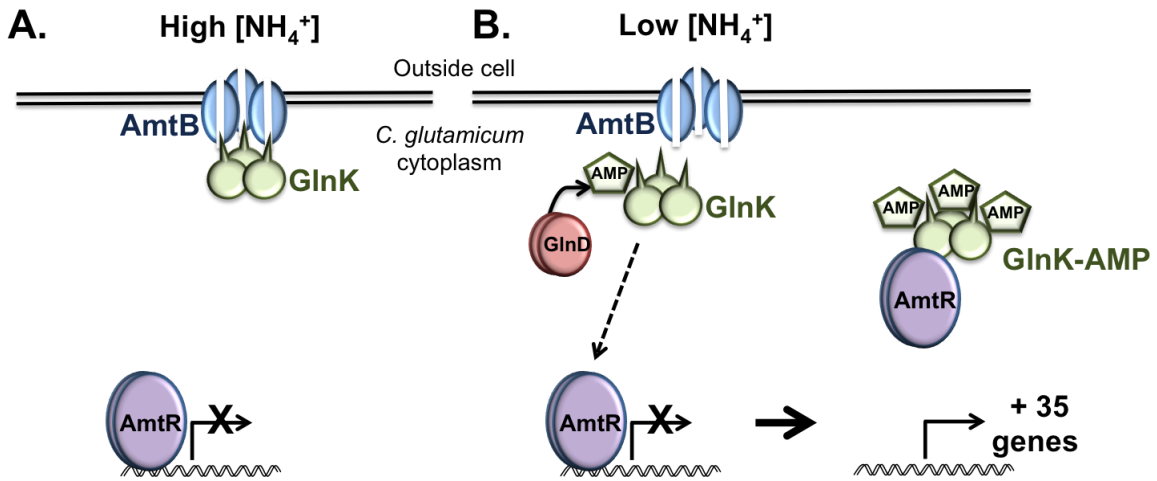


Figure 1.6. Mechanism of AmtR regulation in *C. glutamicum*.

(A) High external nitrogen level AmtR binds to the promoter region of nitrogen-regulated genes, preventing transcription. GlnK is unmodified and bound to AmtB. (B) Low external nitrogen is detected via an unknown mechanism. GlnD is activated and adenylylates GlnK releasing it from AmtB. GlnK-AMP relocates to the cytoplasm where it binds AmtR. Binding of GlnK releases the repressor activity of AmtR allowing transcription of nitrogen-regulated genes.

1.5 Aims of this Study

A comprehensive understanding of nitrogen assimilation and control of these pathways is lacking for mycobacteria. There are many fundamental questions regarding the post-translational modifications and interactions between proteins of the nitrogen regulation cascade. For instance, the signal of nitrogen limitation and regulation of GlnR activity is unknown. In addition only 7 genes have been shown to be part of the mycobacterial GlnR regulon, approximately 50 genes are GlnR regulated in *Streptomyces*, suggesting more unidentified genes are subject to GlnR regulation in mycobacteria.

Therefore, the overall aim of this study is to increase our understanding of the transcriptional response to nitrogen limitation in mycobacteria, focusing primarily on the transcriptional regulator GlnR under nitrogen limiting conditions. As nitrogen-limiting conditions are an important aspect of this study I will first develop a defined medium that stimulates a nitrogen stress response in mycobacteria. These conditions will then be used for the subsequent analysis. To examine the effect of the conserved phosphorylation site of GlnR on transcriptional activation, I will create an *in vivo* point mutant; mutating the conserved aspartic acid-48 to an alanine. This will allow me to determine if the conserved residue Asp-48 is important for GlnR transcriptional activation. To determine the GlnR regulon in *M. smegmatis* I will use chromatin immunoprecipitation coupled with next generation sequencing (ChIP-seq) to locate GlnR-binding sites in the *M. smegmatis* genome. Aligning this data to a microarray of a GlnR deletion strain I will be able to identify genes directly and indirectly regulated by GlnR. Finally using the methodology from the *M. smegmatis* study, I will analyse GlnR binding in the pathogenic species *M. tuberculosis* via Chip-seq analysis. Data obtained from the *M. smegmatis* and *M. tuberculosis* GlnR binding analysis will enable direct comparison of the responses of pathogenic *verses* saprophytic organisms to nitrogen limiting conditions.

CHAPTER 2: Materials and Methods

2.1 Bacterial Strains and Culture Conditions

Plasmids	Description and features	Antibiotic resistance	References
pET28b+	Cloning and expression of recombinant proteins in *E. coli* with a thrombin cleavable N-terminal His_6-tag and optional C-terminal His_6-Tag. Expression of target genes is under control of the bacteriophage T7 promoter.	Kanamycin (50 μg/ml)	Novogen
pYUB854	For the generation of homologous recombination substrates containing a Hyg^R cassette flanked by γδ–res sites for removal of the hygromycin cassette for unmarked deletions. Two multiple cloning sites flank the Hyg^R cassette.	Hygromycin: *E. coli* (200 μg/ml) mycobacteria (50 μg/ml)	(12)
pJV128	Used for the generation of chromosomal point mutations in mycobacteria. pJV128 is an extrachromosomally-replicating plasmid that contains phage encoded Che9c recombineering gene *60* cloned downstream of an acetamidase expression cassette. Contains oriE, oriM, Kan^R, HygSamber and *sacB* cassettes.	Kanamycin (50 μg/ml)	(175)
pJV126	Used for the generation of chromosomal gene replacement mutant in mycobacteria. pJV126 is an extrachromosomally-replicating plasmid that contains phage encoding Che9c recombineering genes *60* and *61* downstream of an acetamidase expression cassette. Contains OriE, oriM, Kan^R and *sacB* cassettes.	Kanamycin (50 μg/ml)	(177)
pMV306	Single copy mycobacterial shuttle plasmid used for the generation of gene complementation strains. Contains *attP* and *int* for the integration of the plasmid at the *attB* site on Mycobacterial chromosomes. Also contains *aph* for kanamycin resistance.	Kanamycin (50 μg/ml)	(157)
pCR2.1 TOPO	Part of the TOPO cloning system for the amplification and sequencing of PCR products. TOPO vectors are provided linearised with topoisomeraseI covalently bound to each 3' phosphate. This enables the vector to readily ligate DNA sequences with compatible ends. PCR products with an A overhang are suitable for TOPO cloning.	Kanamycin (50 μg/ml)	Invitrogen

Table 2.1. Plasmids used in this study.

Cell Strains	Description and features	Genotype	Antibiotic resistance	References
Mycobacterium smegmatis mc^2 155	Wild type strain (ATCC 700084)			(153)
Mycobacterium smegmatis mc^2 155 Δ*glnR*	Wild type *M. smegmatis* with a gene replacement, a hygromycin cassette, in the msmeg_5784 *glnR* gene rendering the gene inactive.	*glnR*$^-$, Hyg^R	Hygromycin (50 µg/ml)	This study
Mycobacterium smegmatis mc^2 155 Δ*glnR*::*glnR*	*M. smegmatis glnR* deletion strain with a *glnR* replacement gene under activation of its own promoter in the integrating vector pMV306.	*glnR*$^-$, *aph*, Hyg^R ::*glnR*	Hygromycin (50 µg/ml) and Kanamycin (50 µg/ml)	This study
Mycobacterium smegmatis mc^2 155 GlnR_D48A	Wild type *M. smegmatis* with a GlnR (msmeg_5784) chromosomal point mutation at amino acid residue 48, nucleotide change results in an aspartic acid to alanine substitution.	*glnR*^{D48A}		This study
Mycobacterium smegmatis mc^2 155 GlnR_D48A::*glnR*	*M. smegmatis* GlnR_D48A strain with a *glnR* replacement gene under activation of its own promoter in the integrating vector pMV306.	*glnR*^{D48A} , *aph* ::*glnR*	Kanamycin (50 µg/ml)	This study
Mycobacterium tuberculosis H37Rv	Wild type strain (ATCC 25618)			
E. coli DH5α	An *E. coli* K12 derived transformable strain. This strain is *endA1* (does not produce Endonuclease I) therefore avoids non-specific digestion resulting in high quality plasmid DNA. DH5α also lacks the alpha portion of the lacZ gene and is thus suitable for blue-white screening.	F$^-$, endA1, glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG, Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–		Promega

Cell Strains	Description and features	Genotype	Antibiotic resistance	References
E. coli BL21(DE3)pLysS	BL21(DE3)pLysS allows high-efficiency protein expression of target genes under the control of a T7 promoter. BL21 (DE3) is lysogenic for λ-DE3 which contains the T7 bacteriophage gene I, encoding a T7 RNA polymerase. Gene I is under control of the *lac*UV5 promoter, inducible by addition of IPTG. BL21(DE3)pLysS carries a plasmid encoding a T7 lysozyme. T7 lysozyme lowers the background expression of target genes under the T7 promoter, but does not interfere with expression levels following induction with IPTG.	F–, *omp*T, *hsd*S_B (r_B–, m_B–), *dcm*, *gal*, λ(DE3), pLysS, Cm^r.	Chloramphenicol (34 µg/ml)	Promega
E. coli HB101	HB101 is a hybrid K12 x B transformable strain. The strain is *rec*A13 negative, which minimizes undesirable recombination events and aids insert stability. Additionally, the strain carries the *hsd*S20(r_B^- m_B^-) restriction minus genotype that allows better representation when cloning methylated DNA and prevents cleavage of cloned DNA by endogenous restriction enzymes. HB101 competent cells are suitable for high efficiency sub-cloning of DNA and vectors that do not require α-complementation for blue/white screening.	F^- mcrB mrr hsdS20(r_B^- m_B^-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm^R) glnV44 λ^-	Streptomycin (50 µg/ml)	Sigma

Table 2.2. Bacterial strains used in this study.

2.1.2 Bacterial Growth Conditions

M. smegmatis mc²155 and *M. tuberculosis* H37Rv were grown aerobically at 37°C with shaking at 180 rpm and 100 rpm respectively, in modified Sauton's medium (0.05% (w/v) KH_2PO_4, 0.05% (w/v) $MgSO_4$, 0.2% (w/v) citric acid, 0.005% (w/v) ferric ammonium citrate, 0.2% (v/v) glycerol, 0.4% (w/v) asparagine, 0.0001% (v/v) $ZnSO_4$, 0.015% (v/v) Tyloxapol). For nitrogen free medium, asparagine was removed and 0.005% (w/v) ferric citrate replaced ferric ammonium citrate. Different nitrogen sources were added to the nitrogen free medium in stated quantities. Stocks of 100 mM glutamine (99.99% pure; Sigma), 100 mM potassium nitrate (Sigma) 100 mM ammonium chloride (99.99% pure; Sigma) and 100 mM ammonium sulphate (99.99% pure; Sigma) were made in solution using the nitrogen free medium and filter sterilised through a 0.2 μm filter. All medium was stored at 4°C and nitrogen stocks were used within 24 hours of preparation.

When specified *M. smegmatis* mc² 155 and *M. tuberculosis* H37Rv was cultured in Middlebrook 7H9 broth (Difco) supplemented with 0.2% (v/v) glycerol, 10% (v/v) oleic acid-albumin-dextrose-catalyse (OADC; Becton Dickinson) and 0.05% (v/v) Tween 80 (Sigma), shaking at 180 rpm (*M. smegmatis*) or 100 rpm (*M. tuberculosis*), aerobically at 37°C. For growth on solid medium, *M. smegmatis* was grown aerobically at 37°C on Middlebrook 7H11 agar (Difco) supplemented with 0.5% (v/v) glycerol and 10% (v/v) OADC.

For growth analysis in nitrogen limiting and nitrogen excess medium, a 24 hour *M. smegmatis* mc² 155 culture or 7 day *M. tuberculosis* H37Rv culture was washed twice by centrifugation at 4000 x *g* in nitrogen free Sauton's medium. The pellet was resuspended in half original volume of Sauton's nitrogen free medium to produce the inoculum. The cell suspension was added to 30 ml Sauton's nitrogen free medium supplemented with ammonium sulphate (*M. smegmatis*) or ammonium chloride (*M. tuberculosis*) at 1 mM (nitrogen limiting) or 30 mM (nitrogen excess), to a starting OD_{600} of 0.08 (Biochrom spectrophotometer). Cultures were grown at 37°C, 180 rpm for *M. smegmatis* and 100 rpm for *M. tuberculosis*. OD_{600} and cfu/ml readings were taken during growth at periodic intervals, cfu counts were conducted as described by (102). Ammonium ions in the culture medium during growth were monitored using an Ammonium AquaQuant kit (Merck) (Section 2.9.1).

E. coli strains were cultured in Luria-Bertani (LB) broth (Miller), aerobically at 37°C, unless otherwise stated, shaking at 180rpm. For growth on solid medium *E. coli* strains were grown aerobically on LB agar (Miller) at 37°C.

Where required, antibiotics were added to the standard growth medium as described in Table 2.1 and Table 2.2.

2.2 Molecular Cloning

2.2.1 Preparation of *M. smegmatis* Genomic DNA

Purification of *M. smegmatis* DNA was carried out as described by Parish *et al.* (113). Standard 7H9 growth medium (10 ml) was inoculated with *M. smegmatis,* from frozen stocks, and grown aerobically until the culture reached late log to early stationary phase ($OD_{600} \sim 1.5$).

Bacterial cells were harvested by centrifugation at 4,500 x *g* for 15 min at room temperature and the supernatant discarded. Lysis of the cells was carried out in 450 µl GTE solution (25 mM Tris-HCL (pH8.0), 10 mM EDTA, 50 mM glucose) and 50 µl of 10 mg/ml lysozyme (Sigma) solution, made up in Tris-HCL buffer (25 mM Tris-HCL (pH8.5)). This suspension was transferred to a 2 ml microcentrifuge tube and incubated overnight at 37°C, shaking at 180 rpm.

Following incubation, cell lysis was stopped with the addition of a final concentration of 2% (w/v) sodium dodecyl sulphate (SDS; Gibco), and mixed gently by inversion. Proteinase K (Sigma) 1 mg/ml final concentration, was added to the suspension, mixed gently, and incubated for 30 min at 55°C. Next, 1.25 M NaCl solution was added and mixed gently by inversion. A 160 µl volume of preheated (65°C) 10% (w/v) cetyltrimethylammonium bromide (CTAB) solution (10 g Cetrimide (Sigma) dissolved in saline solution (4.1 g NaCl dissolved in 90 ml dH_2O)) was added, mixed gently by inversion and the mixture was incubated at 65°C for 10 min.

A 1 ml aliquot of chloroform-isoamyl alcohol (24:1) was added, the solution mixed, and centrifuged at 16,100 x *g* for 5 min at room temperature. The aqueous layer was aspirated and transferred to a fresh 2 ml microcentrifuge tube. An equal volume of chloroform-isoamyl alcohol (24:1) was added, mixed and centrifuged at 16,100 x *g* for 5 min at room temperature. The aqueous layer was aspirated and transferred to a fresh 1.5 ml microcentrifuge tube. To precipitate the DNA isopropanol was added at 0.7 x the aliquot volume, and mixed gently by inversion until the DNA precipitated out of solution. The solution was incubated for a further 5 min at room temperature then centrifuged at 16,100 x *g* for 10 min at room temperature. The supernatant was aspirated and the pellet washed with the addition of 1 ml 70% ethanol, and centrifuged at 16,100 x *g* for 5 min at room temperature. The supernatant was discarded and the DNA pellet allowed to air dry for 15 min. DNA was re-suspended in 50 µl of dH_2O and stored at 4°C overnight to allow the pellet to fully dissolve. The DNA solution was stored at -20°C.

2.2.2 Polymerase Chain Reaction (PCR) for Amplification of DNA Fragments from Mycobacterial Chromosomal DNA

To amplify target DNA sequences BioMix complete master mix (Bioline) was used. The master mix contained required dNTPs, Taq DNA polymerase and produced a 2 mM $MgCl_2$ final concentration. Template DNA (50 ng) was added to a PCR reaction mixture consisting of 20 µl BioMix (2x), 1 µM of each primer and 5% (v/v) dimethyl sulfoxide (DMSO; Sigma), to a final volume of 40 µl with the addition of dH_2O. PCR was carried out in a thermocyler, T3000 (Biometra). Typical PCR conditions are stated below however the annealing temperature was optimised for each primer pair (Appendix 1). Five µl of the product was analysed by agarose gel electrophoresis; the remaining PCR product was purified using a QIAquick spin column (Qiagen), or used immediately without purification for TOPO cloning.

95°C	5 min		Denaturing
95°C	30 sec	30 cycles	Denaturing
55°C	30 sec		Primer annealing*
72°C	1 min/kb		Extension
72°C	8 min		
4°C	∞		Hold

*This step was optimised for each reaction

2.2.3 Colony PCR for Amplification of Desired Insert from *E. coli* Plasmid DNA

Colonies were screen for the presence of a desired insert by amplification of the region of interest with the PCR or vector sequencing primers. A fresh colony was picked from an LB agar plate and used to inoculate a PCR solution containing 5 µl BioMix (2x), 1 µM of each primer and 5% (v/v) DMSO, to a final volume of 10 µl with the addition of dH_2O. Cycler conditions were used as above. The full 10 µl solution was run on an agarose gel to size the DNA fragments. Bacterial colonies, containing plasmids with the correct size insert, were grown up for mini preparation of the plasmid DNA.

2.2.4 Gel Electrophoresis of DNA

DNA was visualised by gel electrophoresis using 0.8% (w/v) agarose (DNA > 200 bp) or 2% (w/v) agarose (DNA < 200bp). The gel was prepared by dissolving electrophoresis grade agarose in a 1x Tris-Acetate-EDTA buffer (TAE) (Fisher). SYBR Safe solution (Invitrogen) was added to the molten agarose according to manufacturer's instructions, before pouring into the mould and allowed to set. The DNA samples were prepared for electrophoresis by adding 1 µl of 10 x Blue Juice gel loading buffer (Invitrogen) for every 10 µl of DNA sample. A 1 kbp DNA ladder (Fermentas) or 100 bp DNA ladder (Biolabs) was run alongside the DNA samples, to size the separated DNA fragments. Electrophoresis was carried out in 1x TAE buffer at 100 volts for 50 minutes. The DNA bound SYBR safe dye was visualised by exposing the gel to ultraviolet light using Gel Doc™ EZ Imager (BioRad).

2.2.5 DNA Purification

DNA was purified from PCR and restriction enzyme digest reactions, in preparation for downstream applications, using a QIAquick PCR purification kit (Qiagen). Five volumes of Buffer PBI (supplied with the kit) was mixed with the sample and loaded onto a QIAquick spin column, in a 2 ml collection tube. DNA was bound to the column by centrifugation at 16,200 x *g* for 60 seconds. The flow-through was discarded and the column was washed by addition of 0.75 ml of buffer PE (supplied) supplemented with 80% ethanol (v/v) (Sigma), and centrifuged again at 16,200 x *g* for 60 seconds. The flow-through was discarded once again, and any residual buffer was removed by additional centrifugation at 16,200 x *g* for 60 seconds. The QIAquick spin column was then transferred to a clean 1.5 ml Eppendorf tube. DNA was eluted by addition of 30 µl of dH_2O to the centre of the QIAquick membrane and centrifuged at 16,200 x *g* for 60 seconds. All samples were stored at -20°C.

2.2.6 Cloning PCR-amplified Target Genes into TOPO pCR 2.1 Vector

Target genes, amplified by PCR, were cloned into TOPO pCR-2.1 vector (Invitrogen) for DNA sequencing prior to cloning into an expression vector. The reaction was carried out according to the manufacturer's instructions using the supplied reagents. Fresh PCR product (1 µl ~ 10 ng) was ligated into the pCR 2.1 vector (2 µl ~ 25 ng/µl) with the addition of 1 µl of 10 x ligation buffer, 1 µl of T4 ligase in a total volume of 10 µl with sterile water (supplied). The sample was

incubated at 14 °C for 16 hours, before being transformed in DH5α *E. coli* cells as described in Section 2.2.10. Transformants were selected on Kanamycin (50 µg/ml) LB agar plates, supplemented with 50 µg/ml 5-Bromo-4-Chloro-3-indolyl β-D-galactopyranoside (Sigma). Colonies displaying a white phenotype were selected for further analysis.

2.2.7 Restriction Enzyme Digestion

Restriction Enzymes, bovine serum albumin (BSA) and buffers were all sourced from New England Biolabs.

Analytical digests of ligated vector/ PCR product were carried out using appropriate restriction enzymes. Five µl of plasmid DNA, obtained by mini-preparation, was incubated with 0.5 µl of restriction enzyme(s), in 1 x appropriate buffer and 10 x BSA diluted (when recommended by the manufacturer). A total volume of 10 µl was made up by the addition of dH_2O, and samples incubated at 37°C for 1 hour. The digests were then separated by electrophoresis.

Prior to ligation of PCR products into an expression vector, both were digested with appropriate enzymes to produce cohesive ends. PCR fragments (30 µl), obtained by PCR amplification or expression vector (approximately 600 ng), were incubated with 0.5 µl of restriction enzyme(s) in the presence of 1 x reaction buffer and 10 x BSA (when recommended by the manufacturer). The final reaction volume of 40 µl was made up by the addition of dH_2O. The reactions were incubated at 37°C for 3 hours. Double digests were performed with enzymes compatible in a single buffer. Digested PCR products were purified using a QIAquick spin column (Qiagen), vector digests were purified by gel extraction followed by the QIAquick gel extraction kit.

2.2.8 DNA Purification from Agarose Gels

Restriction digest products were extracted and purified from agarose gels using the QIAquick gel extraction kit (Qiagen). The gels, stained with SYBR Safe (Invitrogen), were viewed under UV light in a darkroom, and the desired DNA bands were excised from the gel using a scalpel. The excised gel slices were dissolved in 3 volumes of Buffer QG (supplied) and incubated at 60°C for 10 minutes. The resulting solutions were applied onto QIAquick spin column, placed in 2 ml collection tube, and the Qiagen spin column protocol described in Section 2.2.5 followed.

2.2.9 DNA Ligation of Plasmid Vector and DNA Insert

All ligations were carried out on vector/insert DNA with cohesive ends, produced by prior enzymatic cleavage. Purified vector and insert were combined at a molar ratio of approximately 1:3, this was based on quantitative comparison of the purified products using a NanoDrop 1000 spectrophotometer. The digested vector and DNA insert were added to 1 µl T4 DNA ligase buffer (supplied at a 10 x concentration) (NewEngland BioLabs) and 0.5 µl of T4 DNA ligase (supplied at 5 units/ml) (NewEngland BioLabs), in a total volume of 10 µl made up with dH_2O. Reaction mixtures were incubated at RT for 1 hour. Ligations were then transformed into selected bacterial strains.

2.2.10 Transformation of *E. coli* with Plasmid DNA

Competent BL21(DE3)pLysS, HB101 and DH5α strains of *E. coli* were used for recombinant protein expression or plasmid amplification. Aliquots (50 µl) of the appropriate bacterial cells were prepared for transformation by incubation with 5 µl of the ligation mixture, or 50 ng plasmid DNA, on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds by placing in a pre-heated water bath and immediately followed by incubation on ice for 2.5 min. Two hundred and fifty µl SOC medium (Invitrogen) was added to the transformed culture and incubated at 37°C for 1 hour with moderate shaking (180 rpm). Cell aliquots (20 µl and 200 µl) of transformed bacteria were spread on LB agar plates supplemented with appropriate antibiotic(s) and incubated at 37°C until colonies were present, usually 12-24 hours.

2.2.11 Plasmid DNA Mini-Preparations from Small-Scale Cultures of Transformed Bacteria

Transformed bacteria were screened for the presence of the required plasmid and insert by amplification of a single bacterial colony and extraction of plasmid DNA. Individual colonies were selected and grown in 5 ml LB broth, supplemented with appropriate antibiotic(s), and shaken (180 rpm) at 37°C overnight. Bacteria were harvested by centrifugation at 4,000 x *g* for 10 minutes.

DNA extraction was carried out using a Mini-Prep Kit (Qiagen). The supernatant was discarded and the bacterial pellet re-suspended in 250 µl of Resuspension Solution (containing 100 µg/ml RNase A). This was followed by addition of 250 µl of Lysis Solution (supplied), and the contents

inverted 4-6 times. Neutralisation Solution (350 µl; supplied) was then added to the cell lysate and the tubes were inverted 4-6 times prior to centrifugation at 16,100 x *g* for 10 minutes. To isolate plasmid DNA the supernatants were loaded onto QIAquick columns, in a 2 ml collection tubes, and centrifuged at 16,100 x *g* for 1 minute. Five hundred µl Wash Solution (supplied) was applied to the column and centrifuged for a further 60 seconds. The flow through was discarded and the wash step repeated. Removal of any residual wash solution was carried out by centrifuging the empty column at 16,100 x *g* for a further 1 minute. The spin columns were then transferred to a sterile 1.5 ml Eppendorf tubes; DNA was eluted by addition of 30 µl dH_2O to the centre of the QIAquick membrane and centrifugation at 16,100 x *g* for 2 minutes.

Recovered plasmid DNA was analysed by restriction enzyme digestion, followed by agarose gel electrophoresis and visualized with SYBR Safe and UV illumination. Plasmids containing correctly sized fragments were submitted for DNA sequencing analysis at Medical Research Council genomics core facility (Imperial College Hammersmith campus, UK).

2.2.12 Sequencing of Plasmid DNA

DNA sequences were determined using the Imperial College Core Sequencing Service at the MRC genomics laboratory. Sequencing reactions contained 3.2 pmoles of primer, 150 - 300 ng of plasmid DNA per 3 kb or 500 - 600 ng for > 3kb plasmid DNA in 10 µl total dH_2O. Samples were cycle sequenced using BigDye v 3.1 (Applied Biosystems) as follows:

94°C	1 min	
94°C	10 sec	30 cycles
55°C	15 sec	
60°C	4 min	
4°C	∞	

Products were purified by EDTA-ethanol precipitation, resuspended in highly deionised formamide and run on 3730xl DNA Analyser (Applied Biosystems). All sequences were then analysed using Geneious genome viewer.

2.2.13 Quantification of DNA Concentration

DNA concentrations were determined using a Nanodrop ND 1000 Spectrophotometer (NanoDrop Technologies, Inc.). For this, 1.5 µl of DNA sample was placed on the nanodrop stage and the sample was analysed spectrophotometrically giving concentration (ng/µl) and purity (260 nm/280 nm ratio). The blank used in each case was the buffer in which the DNA was resuspended.

2.2.14 Large-Scale Preparation of Plasmid DNA from Bacterial Cultures (Midi Prep)

For large-scale preparation of plasmid DNA midi preparations were performed. Bacteria were transformed with the desired plasmid and grown on LB agar over night at 37°C. LB broth (100 ml), supplemented with appropriate antibiotic(s), was inoculated with a single colony and the culture incubated overnight, shaking 180 rpm at 37°C.

Plasmid DNA was isolated from the bacterial culture using the DNA Midi-prep Kit (Qiagen). The bacteria were harvested by centrifugation at 4,500 x *g* for 15 minutes at 4°C and supernatants discarded. Four ml Buffer P1 (containing 100 µg/ml RNase A; supplied) was used to re-suspend the pellet, followed by the addition of 4 ml Buffer P2 (supplied) to lyse the cells; the mixture then incubated for 5 minutes at room temperature. Termination of the lysis was conducted by addition of 4 ml chilled Buffer P3 (supplied), and the mixed thoroughly by vigorously inverting 4-6 times. The cell lysate was transferred to the barrel of a QIAfilter Cartridge and incubated at RT for 10 min. Buffer QBT (4 ml) (supplied) was applied to a QIAGEN-tip 100, to equilibrate the column, and the column allowed to empty by gravity flow. Cell lysate was then filtered through the QIAfilter and applied to the equilibrated QIAGEN-tip. The lysate was allowed to enter the resin of the tip via gravity flow. Column washes were carried out by applying 2 x 10 ml buffer QC (supplied). DNA was eluted from the column with the addition of 5 ml Buffer QF (supplied). Eluted DNA was precipitated by the addition of 0.7 volumes of isopropanol and pelleted by centrifugation 4,500 x *g* for 1 hour at 4°C. The supernatant was carefully removed and 5 ml 70% ethanol was used to wash the DNA pellet, followed by centrifugation at 4,500 x *g* for 1 hour. Ethanol was removed and the DNA air-dried before dissolving in 50 µl dH_2O deionised water. Concentrations of resulting plasmid DNA were determined by NanoDrop 1000 spectrophotometer.

2.2.15 One-step Preparation of Competent *E. coli* Cells: Chung Method

Cells for protein expression, *E. coli* BL21(DE3)pLysS, were made chemically competent using the Chung method (33). From frozen stocks, bacteria were streaked out on LB agar plates, and grown at 37°C overnight. One colony was selected and used to inoculate 10 ml LB broth. The inoculated culture was grown overnight at 37°C at 180 rpm, until stationary phase reached. Fifty ml of LB broth was inoculated with 0.5-1 ml of the overnight culture and grown at 37°C 180 rpm until cells reached early exponential phase (OD_{600} ~0.3–0.4). Cells were then harvested by gentle centrifugation at 1000 x *g,* 4°C, for 10 minutes. The supernatant was discarded and the pellet re-suspended in 1/10 original culture volume, ice cold, TSS solution (LB broth, 10% (w/v) polyethylene glycol (PEG; Sigma), 5% (v/v) DMSO, 50 mM $MgSO_4$, pH 6.5). Cells were divided into 0.1 ml aliquots, flash frozen in dry-ice and 100% ethanol and stored at -80°C.

2.2.16 Preparation of Ultra-competent *E. coli* Cells: Inoue's Method

Cells for DNA cloning, DH5α and HB101 strains of *E. coli*, were made ultra-competent using the method describe by Inoue *et al.* (67). From frozen stocks, bacteria were streaked onto LB agar plates and incubated aerobically overnight at 37°C. Ten to twelve colonies were used to inoculate 250 ml SOB (Difco) and cultures incubated at 18°C at 180 rpm. When optical density reached OD_{600} ~ 0.6, cells were transferred to 50 ml Falcon tubes and incubated on ice for 10 minutes. Cells were harvested by centrifuging at 2500 x *g* for 10 minutes at 4°C. Pellets were re-suspended in 1/3 original volume in ice-cold Inoue Transformation Buffer (ITB; 55 mM $MnCl_2.4H_2O$, 15 mM $CaCl_2.2H_2O$, 250 mM KCl, 0.5 M PIPES pH6.7). A further incubation on ice was carried out for 10 min before centrifugation at 2500 x *g* at 4°C for 10 min. The pellets were then re-suspended in 1/12 original volume of ice cold ITB with the addition of 7.5% (v/v) DMSO. The solution was mixed gently before incubation on ice for 10 min. Cells were divided into 0.2 ml aliquots, flash frozen in dry-ice and 100% ethanol and stored at -80°C.

2.3 Protein Expression and Purification

2.3.1 Growth of Cells for Protein Expression

Plasmids were transformed into BL21 (DE3) or BL21 pLysS strains of *E. coli* for protein expression. Transformants were selected by growth on LB agar plates, supplemented with appropriate antibiotic(s), and incubated at 37°C overnight. A small scale culture (100 ml) LB broth, plus appropriate antibiotic(s), was inoculated with a loop full of colonies and grown aerobically shaking at 37°C overnight. A 2 L conical flask containing 800 ml of LB, plus appropriate antibiotics, was inoculated with 50 ml of the overnight culture and grown shaking at 37°C. When the OD_{600} ~ 0.4-0.5 the culture was temperature shifted to 20°C in a H_2O bath shaking at 145 rpm. Expression was induced with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma). The cultures were grown for a further 3 hours, shaking at 20°C. Cells were harvested by centrifugation at 4, 000 x *g* at 4°C and the pellet frozen at -20°C overnight or until required.

To test for protein induction, samples (1 ml) were taken before and after addition of IPTG. These samples were centrifuged at 16, 100 x *g* and the pellet re-suspended in 100 µl H_2O. Ten µl of these samples were analysed by SDS-PAGE using a 4-12% Bis-Tris gel (Invitrogen).

2.3.2 Protein Purification: Cell Lysis

Soluble protein fractions were obtained by disruption of the *E. coli* cell wall by probe sonication. Bacterial cell pellets were thawed on ice before being re-suspended in 30 ml of lysis buffer (Phosphate buffered saline (PBS; Sigma) containing: 3 complete mini EDTA-free protease inhibitor tablets (Roche), 100 µg/ml lysozyme, 85.5 units deoxyribonuclease I (Invitrogen)). Bacterial suspensions were placed on ice and lysed by sonication (30 second on/off pulses for 15 minutes at 11 amplitude microns). Cellular debris was removed by centrifugation of the lysate at 17, 000 x *g*, 4°C, for 30 minutes. The supernatant (soluble fraction) was then transferred to a chilled tube and kept on ice until affinity purification. Affinity purification was carried out directly after cell lysis. A 10 µl sample of the soluble fraction was removed analysed by SDS-PAGE using a 4-12% Bis-Tris gel.

2.3.4 Protein Purification: Nickel Affinity Chromatography

Proteins were purified after expression by affinity purification on an AKTA Purifier FPLC system (GE Healthcare). Buffers A (25 mM $Na_2H_2PO_4$ (pH7), 0.5 M NaCl, 5% glycerol), and Buffer B (Buffer A + 1 M imidazole (Sigma)), were used. A 5 ml HiTrap Ni affinity column (GE healthcare) was prepared for purification by washing with 15 ml 0.1 M EDTA, to remove any residual compounds. The column was equilibrated to Buffer A (15 ml), before washing with 15 ml dH_2O. The column was charged by loading 5 ml of 0.1 M $NiCl_2$, before a repeated wash with H_2O. Three column washes of 100% Buffer A 15 ml, 100% Buffer B 15 ml and 100% Buffer A 20 ml were carried out before loading of the bacterial lysate. Clarified bacteria lysate was loaded onto a superloop and injected onto the 5 ml Ni column. Non-specific binding of proteins to the column was removed with a 9% buffer B column wash. The hexa-histidine-tagged protein complex was eluted from the Ni column in a gradient of 1- 100% buffer B, over 80 ml. Fractions (0.5 ml) were collected and 10 µl analysed on a 4-12% Tri-Bis gel, according to the chromatogram UV reading.

Purified protein fractions were dialysed into storage buffer C (10 mM Tris-HCl pH 8, 50 mM NaCl, 20% (v/v) glycerol, 0.1 mM EDTA) for antibody production or storage buffer D (10 mM Tris-HCl pH 8, 50 mM NaCl, 5% (v/v) glycerol) for gel shift assays, overnight at 4°C using dialysis cassettes (Thermo Scientific) with 20 kDa molecular weight cut off. All proteins were split into aliquots and frozen at -20°C.

2.4 Analysis of Protein Samples

2.4.1 BCA Determination of Protein Concentration

Protein concentration was determined using the BCA protein assay kit (Pierce). Concentrations of the neat protein sample or a 1/10 dilution in PBS were analysed. To 50 µl of sample, 1 ml of Working Reagent (supplied; 50:1 parts BCA reagent A to BCA reagent B) was applied, mixed and incubated at 37°C for 1 hour. A blank standard replicate was also incubated for 1 hour at 37°C, containing 50 µl PBS and 1 ml Working Reagent. Absorbance of the samples at 562 nm was analysed, using water as the blank. Subtraction of the blank standard replicate from the sample readings, gave a corrected absorbance value. This corrected absorbance value was compared to a standardised curve with known protein concentrations to determine the sample concentration.

2.4.2 SDS Polyacrylamide Gel Electrophoresis (PAGE): NuPAGE Novex 4-12% Bis-Tris Gels (Invitrogen)

Protein samples were analysed by SDS-PAGE using NuPAGE Novex 4-12% Bis-Tris gel according to manufacturer's instructions. Samples were prepared by the addition of 2.5 µl of NuPAGE LDS Sample Buffer (4x), 1 µl NuPAGE Reducing agent (10x), and water to a final concentration of 10 µl. The prepared samples were then heated at 95°C for 10 minutes, before being centrifuged at 16, 000 x *g* for 2 min to settle the contents. An XCell *SureLock* system (Invitrogen) was arranged to hold the pre-cast 4-12% Bis-Tris gel. The inner chamber was filled with 200 ml 1 x MES buffer (Invitrogen) containing 500 µl NuPAGE Antioxidant (Invitrogen), with 600 ml of 1 x MES only filling the outer chamber. Wells were flushed before loading of the sample. A SeeBlue pre-stained protein ladder (Invitrogen) was run adjacent to the samples to size the resolved products. Gels were run for 35 min at 200 V.

Bands were visualised using SimplyBlue Safe Stain (Invitrogen). Gels were removed from their cast and washed in dH_2O 3 times for 5 minutes. After washing, 20 ml of SimplyBlue Safe Stain was added to cover the gel and incubated with gentle agitation for 1 hour. De-staining was carried out by rinsing the gel in dH_2O, before covering the gel in dH_2O and incubating with gentle agitation for 1 hour; the bands were then visualised. For clearer gels, water was replaced and the gel incubated for a further 1 hour. De-stained gels were then scanned directly using a flatbed scanner (Epson) or dried using the Invitrogen system.

2.4.3 Western Blot

Protein samples were detected using specific antibodies raised against a tag or recombinant protein. The Invitrogen XCell *SureLock* blot module was used to perform the transfer. Proteins samples were separated on 4-12% Bis-Tris gels as described in Section 2.4.2. Immediately following electrophoresis the blot module was assembled, with all components of the blot sandwich pre-soaked in 1 x NuPAGE transfer buffer (Invitrogen). The transfer sandwich composed of, from cathode to anode, blotting pads, filter paper, 4-12% Bis-Tris gel, Hybond-C extra (Amersham) nitrocellulose transfer membrane, filter paper, blotting pads. The blot module was assembled with 1 x transfer buffer covering the transfer sandwich in the inner section, and H_2O filled the outside chamber. Proteins were electro-blotted onto the membrane for 1 hour at constant 30 V. The membrane was removed and blocked overnight at 4°C in a high protein block solution (5% (w/v) milk powder in PBS).

The unbound block solution was removed from the membrane by two washes in PBS/Tween (PBS, 0.05% (v/v) Tween-20 (Sigma)), followed by one wash in PBS only. Blocked membranes were incubated with primary antibody, amount and type as specified, diluted in 20 ml block solution, for 1 hour, gently rocking, at room temperature. Unbound antibody was removed by washing as stated previously. Incubation with a secondary antibody, polyclonal HRP conjugated swine anti-rabbit (DakoCytomation), diluted 1 in 10,000 with block, followed for 1 hour at room temperature. Washing of the membrane occurred as stated earlier.

Super Signal West Femto (Thermo Scientific) was used to detect the HRP conjugated antibody. Equal volumes of the stable peroxide solution and the enhancer solution were mixed before applying to the drained membrane. The membrane and solutions were incubated for 5 min before chemiluminescence detected via a LAS-3000 Fuji imager.

2.4.4 Affinity Purification of GlnR Polyclonal Antibody

The polyclonal serum was subject to affinity purification against the recombinant *M. smegmatis* His-GlnR or *M. tuberculosis* His-GlnR. Purified protein (50 µg) was separated via SDS PAGE and transferred to a nitrocellulose membrane. The membrane was stained with Ponceau S (Sigma) and rinsed with ddH_2O to visualise the bands. A membrane slice was extracted surrounding His-GlnR, before incubating the membrane for 1 hr at RT in Block (PBS with 5% milk powder). Following this 5 ml serum, diluted in 25 ml Block, was incubated with the membrane over night at 4°C with gentle agitation. The membrane was washed 4 times in PBS before elution of the

antibody. Antibody elution commenced with 2 min incubation with 4 ml 100 mM glycine pH 2.7, before transfer of the solution to 300 µl of 1.5 M Tris-HCl pH 8.8. Purified antibody was dialysed against PBS before storage at -20°C.

2.4.5 Preparation of *M. smegmatis* Cell Lysates

M. smegmatis cell lysates were prepared from 30 ml samples. Cells were harvested by centrifugation at 4000 x *g* and re-suspended in 0.5 ml PBS buffer (Sigma) with EDTA- free complete protease inhibitor cocktail. The solution was transferred to tubes containing Zirconium beads (MP Biomedicals) and the cells lysed in a ribolyser Fastprep FP120 (Thermo *Savant*) for two cycles of 30 seconds at 6 m/s. Cell debris was removed by centrifugation at 16,100 x *g* for 10 min and supernatant removed and stored in aliquots at -20°C. Lysate concentrations were quantified using BCA assay (Pierce).

2.5 Generation of *M. smegmatis* Mutants: Recombineering Method

2.5.1 Preparation of Electrocompetent *M. smegmatis* Cells

Fifty ml of 7H9 medium was inoculated with *M. smegmatis* mc^2 155 from frozen stocks and grown at 37°C at 180 rpm until OD_{600} ~1.8. Cell cultures were incubated on ice for 1 hour before harvesting by centrifuging at 1,500 x *g*, 4°C for 10 minutes. Pellets were washed with 50 ml ice cold 10% (v/v) glycerol. Cells were harvested by re-centrifuging at 1,500 x *g,* 4°C for 10 min. The wash step was repeated a further 3 times. After the final spin cells were re-suspended in 1/10 original volume of 10% ice-cold glycerol, and split into 200 µl aliquots, before flash freezing in dry ice and ethanol and storage at -80°C.

2.5.2 Transformation of Electrocompetent *M. smegmatis* Cells

For the uptake of DNA into *M. smegmatis*, electrocompetent cells were transformed using a bench top electroporator, Gene Pulser (BioRad). Competent *M. smegmatis* cells were thawed on ice before the addition of 200 ng of DNA. Cells were mixed by gentle tapping and incubated on ice for 20 min. The bench top electroporator was set at: resistance (R) 1000 Ω, capacitance (Q) 25 µF, voltage (V) 2.5 kV. Samples were transferred to a 2 mm cuvette and tapped to remove any air bubbles. The cuvette was attached to the electroporator and the voltage applied. A time constant of between 15-23 indicated a successfully applied charge across the cuvette. One ml of 7H9 medium was added directly to the cuvette, before transferring the sample to a 1.5 ml eppendorf tube. Cells were incubated at 37°C for 3 hours, before plating onto selective 7H11 plates.

2.5.3 Preparation of Recombineering Strain of Electrocompetent *M. smegmatis* Cells

M. smegmatis cells, already containing the desired recombineering plasmid via electroporation, were induced for protein expression and made electrocompetent. Five ml of 7H9 medium, supplemented with appropriate antibiotic(s), were inoculated with recombineering plasmid containing cells, and grown at 37°C to log phase. Modified 7H9 medium (Middlebrook 7H9 broth supplemented with; 0.05% (v/v) Tween, 0.2% (w/v) succinate, antibiotic) was inoculated to an OD_{600} ~0.025 and incubated at 37°C, 180 rpm. When cells reached an OD_{600}~ 0.4 they were induced for protein expression with the addition of acetamide to a final concentration of

0.2%, and incubated as before for 3 hours. Following protein induction, the culture was placed on ice for 1 hour before harvesting of cells by centrifuging at 4,000 x *g*, 4°C for 10 min. The pellet was washed with ½ original volume ice-cold 10% glycerol and cells harvested as before. Two further wash steps were carried out in ¼ original volume ice-cold 10% glycerol, and harvested as before. The pellet was finally re-suspended in 1/20 original volume 10% glycerol, split into 200 µl aliquots and flash frozen in dry ice and 100% ethanol, before storage at -80°C.

2.5.4 Gene Replacement Mutant: Allelic Exchange Substrate (AES)

Generation of gene replacement mutants was carried out as described by van Kessell *et al.* (175, 176). An allelic exchange substrate (AES) was generated by amplification of regions of homology flanking the target gene. Homologous regions (~500-1000 bp) were amplified by PCR. PCR products were directly cloned into the pCR2.1-TOPO vector system (Section 2.2.6) and inserts sequenced. Flanks were excised from the vector with appropriate restriction enzymes and purified with gel extraction (Sections 2.27 and 2.2.8 respectively).

Directional cloning of homologous regions of DNA was achieved by insertion and transformation of one flank, before the second. The target vector, pYUB854 containing a Hyg^R cassette separating two multiple cloning sites, was digested for a single insertion using appropriate enzymes and purified with gel extraction. Concentration comparison of the target vector and digested PCR product was carried out with a NanoDrop 1000 spectrophotometer, before ligation at a concentration of 1:3 vector:insert (Section 2.2.9). HB101 *E. coli* cells were transformed with the plasmid/insert ligation and positive transformants selected for (Section 2.2.10). Positive colonies were screened for the presence of the correct sized insertion by mini preparation and subsequent restriction digestion (Section 2.2.11). Resulting plasmid, obtained via mini preparation, was then subjected to digestion using restriction enzymes for the alternate cloning site, with ligation and positive transformant selection repeated as before. Directionality of the cloned products was checked by sequence analysis.

The AES was obtained by digestion of the upstream-Hyg^R-downstream sequence from the plasmid backbone via restriction digestion, followed by gel extraction. Resulting AESs were checked for purity and size on a 0.8% agarose gel and quantified using a NanoDrop 1000 spectrophotometer.

2.5.5 Gene Replacement Mutant: Recombineering

Gene replacement mutants were generated by transformation of the linear AES in to *M. smegmatis* cells. AES DNA was prepared by digesting the pYUB85-*glnR* construct with *Afl*I and *Spe*I. Linear AES DNA (200 ng) was used to transform 200 µl of *M. smegmatis* cells containing the pJV126 recombineering plasmid (a kind gift from Graham Hatfull). Putative null mutants were selected on 7H11 agar containing hygromycin (50 mg ml^{-1}) and kanamycin (50 mg ml^{-1}). Confirmation of gene deletion was carried out by PCR on gDNA using primers outside the upstream and downstream flanking regions in combination with hygromycin cassette specific primers. PCR products would only be obtained with insertion of the hygromycin cassette by recombination onto the chromosome at the correct location. Further confirmation of *glnR* deletion phenotype was provided by Western analysis using a custom made GlnR polyclonal antibody (Eurogentec, Belgium).

2.5.6 GlnR Chromosomal Point Mutation and MAMA PCR Screen

The point mutation was generated using *M. smegmatis* containing the pJV128 recombineering plasmid (a kind gift from Graham Hatfull). Cells were co-transformed with 100 ng of two ssDNA oligonucleotides containing the base pair changes for the required *glnR* D48A point mutation and containing the required base pair changes to convert the *hygS* cassette contained within the pJV128 vector from hygromycin sensitive to hygromycin resistant. This hygromycin resistance repair method was used to select colonies that had undergone positive recombination. A mismatch amplification mutation assay (MAMA PCR) screen using primer pairs MAMA_PCR_F&R was performed to identify *glnR* containing the desired point mutation (29, 160). The rationale behind MAMA PCR is that a single nucleotide mismatch at the 3′ extremity of the annealed reverse primer renders *Taq* polymerase unable to extend the primer. Therefore, the absence of the specific PCR product reveals a deviation from the desired DNA sequence. MAMA PCR conditions were 95°C 5 min, 39 cycles of 95°C 15 sec, 32°C 1 min with final extension time of 72°C 7 min. Recombineering plasmids were removed from the mutant strains via negative *sacB* selection (117).

2.6 RNA Analysis

2.6.1 RNA Isolation from *M. smegmatis* Whole Cell Extracts

RNA was extracted from *M. smegmatis* cells grown as specified. Cells were initially mixed with an equal volume of GTC solution (5 M guanidine thiocyanate, 0.5% N-lauryl sarcosine, 0.1 M β-mercaptoethanol, 0.5% Tween-80, 10 mM Tris.HCl pH7.5) and cell pellets recovered by centrifugation at 4,000 x *g* for 10 minutes. The supernatant was discarded and the pellet re-suspended in 1 ml Trizol (Invitrogen) and stored at -80°C until required. Thawed cells were homogenised with ribolyser Fastprep FP120 (Thermo *Savant*) at 6 $msec^{-1}$ for 30 sec x 2, with cells incubated on ice for 5 minutes between runs. Cell debris was removed by centrifugation at 16,100 x *g* for 10 min at 4°C. The supernatant was transferred to a 1.5 ml eppendorf tube and mixed with 600 µl chloroform, followed by centrifugation at 4°C 16,100 x *g* for 5 minutes. The aqueous phase was recovered and chloroform extraction repeated. Finally, the RNA was precipitated with 100% isopropanol and RNA was recovered via centrifugation at 16,100 x *g* for 10 min at 4°C. Supernatant was removed and the pellet washed with 70% ethanol. Following this the RNA was purified using the RNeasy kit (Qiagen) according to manufacturer's instructions. Residual DNA was removed from the sample with a TURBO DNA-*free* (Ambion) treatment following the manufacturer's instructions. RNA was recovered and RNA quality and quantity was determined by OD 260/280 and 260/230, gel electrophoresis and bio-analyzer analysis. RNAsecure (Ambion) (1 µl) was added to protect the sample and stored at -20°C.

2.6.2 cDNA Preparation

To determine gene expression levels cDNA was amplified from RNA using the SuperScript III first strand synthesis super mix (Invitrogen). To 100 ng RNA the following were added as per the manufacturer's instructions; 10 µl of 2x RT reaction mix, 2 µl RT enzyme mix, DEPC-treated water to a final volume of 20 µl. Samples were incubated at 25°C for 10 min followed by 50°c for 30 min and to terminate the reaction 85°C for 5 min. RNaseH was added to the cooled sample and incubated at 37°C for 20min. cDNA was stored at -20 until required.

2.6.3 Quantitative Real-Time PCR (qRT-PCR)

To determine gene expression levels, qRT-PCR was performed on cDNA. qRT-PCR reactions were carried out in a final volume of 10 µl (1 µl of cDNA, 5 µl of TaqMan PCR master mix (Applied Biosystems), 0.5 µl of the appropriate TaqMan probe (Applied Biosystems)). Amplification was performed on an Applied Biosystems 7500 Real-Time System (conditions 50°C 5 min, 95°C 10 min, and 40 cycles of 95°C 15 sec, 60°C 1 min). Amplification efficiency for each TaqMan probe was determined to be 100%, based on amplification of serial dilutions of template cDNA, with the slope of linear regression used to determine efficiency. Real-time analysis was performed on RNA from three independent cultures and quantification of *sigA* expression served as an internal control. The threshold value C_T was converted to gene expression, an arbitrary unit in relation to *sigA* (gene expression = $2^{-\chi}$, where $\chi = C_T$ sample gene – C_T sample *sigA*). All data displayed reflects the mean of triplicate experiments, with error bars indicating the standard deviation. Statistical comparison of means was performed with a Student's *t*-test, a *P* value of ≤ 0.05 was considered significant.

2.6.4 Preparation of Labelled cDNA from Total RNA for Microarray Analysis (carried out at BUGS @ St. George's Hospital)

Labelled cDNA was prepared from 1 µg total RNA using Cy3-dCTP (GE Healthcare) and SuperScript II reverse transcriptase with random hexamer primers (Life Technologies – Invitrogen division). Agilent One Color Spike-In controls were labelled together with the RNA samples according to manufacturer's instructions. Labelled cDNA was purified by Qiagen MinElute column, combined with 10 x CGH blocking agent and 2 x Hi-RPM hybridisation buffer (Agilent) and heated at 95°C for 5 minutes prior to loading onto microarray slides which were incubated overnight in an Agilent rotating oven at 65°C, 20 rpm. After hybridization, slides were washed for 5 minutes at room temperature with CGH Wash Buffer 1 (Agilent) and 1 minute at 37°C with CGH Wash buffer 2 (Agilent) and scanned immediately, using an Agilent High Resolution Microarray Scanner, at 2 µm resolution, 100% PMT. Scanned images were quantified using Feature Extraction software v 10.7.3.1.

2.6.5 *M. smegmatis* Microarray Design

The microarray was constructed by determining all unique genes from the 6887 chromosomal predicted coding sequences of *M. smegmatis* strain mc^2 155, downloaded from Ensembl Bacteria

Release 5 (http://bacteria.ensembl.org/). Multiple optimal hybridisation 60-mer oligonucleotide sequences were designed (Oxford Gene Technologies), from which a minimal non-redundant subset of oligonucleotides were selected with target coverage of three 60-mers per gene. Arrays were manufactured on the Inkjet in-situ synthesized platform (Agilent) using the 8 x 60 k format. The full array design is available in BμG@Sbase (BμG@Sbase: A-BUGS-40) and also in ArrayExpress (ArrayExpress: A-BUGS-40).

2.6.6 Statistical Analyses of Differential Gene Expression (Conducted in collaboration with Geraint Barton at CISBIO)

Statistical analyses of the gene expression data was carried out using the statistical analysis software environment R together with packages available as part of the Bioconductor project (http://www.bioconductor.org). Data generated from the Agilent Feature Extraction software for each sample was imported into R. Replicate probes were mean summarised and quantile normalised using the preprocess Core R package. The limma R package (152) was used to compute empirical Bayes moderated *t*-statistics to identify differentially expressed gene between time points. Generated p-values were corrected for multiple testing using the Benjamini and Hochberg False Discovery Rate. A corrected p-value cut-off of less than 0.01 was used to determine significant differential expression. Fully annotated microarray data has been deposited in BμG@Sbase (accession number E-BUGS-143; http://bugs.sgul.ac.uk/E-BUGS-143) and also in ArrayExpress (accession number E-BUGS-143).

2.7 Chromatin-Immunoprecipitation

2.7.1 Cell Preparation and Cross-linking

M. smegmatis or *M. tuberculosis* (3 x 60 ml cultures) were grown as specified in nitrogen limiting and excess medium before cross-linking with the addition of formaldehyde (Sigma) (final concentration 1% (v/v)). Cross-linking proceeded for 20 min with continued agitation at 37°C, before glycine addition (final concentration 125 mM) and incubation for 5 min at 37°C. Cells were harvested by centrifugation at 4,000 x *g* and washed twice with TBS. The pellet was frozen at -80°C until required.

For DNA fragmentation the pellet was re-suspended in 8 ml immunoprecipitation (IP) buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) Na deoxycholate, 0.1% (w/v) SDS) supplemented with EDTA- free complete protease inhibitor cocktail (Roche), before sonication at 100% amplitude in 30 sec pulses for 10 min (Misonix Ultrasonic Processor S4000). Debris was removed by centrifugation at 4,000 x *g* and the supernatant recovered. Samples were stored on ice at 4°C and 100 µl sample taken to confirm sonication conditions. To the 100 µl sample DNA was precipitated with sodium acetate/ ethanol, before analysis on a 2% agarose gel. Sonication was deemed complete when DNA fragments were between 100-200 bp and no visible genomic DNA present. Once sonication was confirmed a further 100 µl sample was taken and stored at -20°C, this sample was subject to protein degradation as the rest of the sonicated extract and used as the control, 'input' sample. The rest of the sample was subject to immunoprecipitation.

2.7.2 Immunoprecipitation and Elution of DNA

To the sonicated extract 200 µl of our purified rabbit anti-GlnR specific polyclonal antibody was added and incubated overnight on a rotating wheel at 4°C. Sheep anti-rabbit IgG Dynal beads (Invitrogen) were prepared by washing in 2x 500 µl PBS and 2x 500 µl IP buffer, before saturating the beads overnight at 4°C in 1 ml blocking solution (IP buffer, EDTA-free protease inhibitor tablet, 1 mg/ml BSA). Following saturation of the beads the blocking solution was removed and sonicated sample plus antibody incubated with the beads for 3 hours at 4°C on a rotating wheel. To harvest the bead-antibody-DNA complex a magnet was used (Invitrogen). The complex was then subject to a series of washing steps; 2x 500 µl IP buffer, 500 µl IP buffer plus 500 mM NaCl, 500 µl wash II (10 mM Tris pH 8, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet-P40, 0.5% (w/v) Na deoxycholate), 500 µl TE buffer (50 mM Tris, 10 mM EDTA pH 7.5). Elution

of DNA was performed by addition of 100 µl elution buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% (w/v) SDS) and incubation at 65°C with rocking for 40 min. Beads were separated by magnetism and the supernatant harvested. Elucidate was diluted 2-fold in nuclease free H_2O (Qiagen), followed by protein degradation with the addition of 4 mg/ml Pronase and incubated at 42°C for 2 hours and 65°C for 6 hours. DNA was subsequently purified using the Qiagen MiniElute kit and DNA quantified using the HS dsDNA Qubit (Invitrogen).

2.7.3 Library Preparation for Next Generation Sequencing

DNA was prepared for next generation high throughput sequencing using the Illumina ChIP-seq DNA sample prep kit according to the manufacturer's protocol, with the addition of a second gel extraction step after PCR amplification to remove excess primer dimers. DNA size and purity was confirmed via HS DNA Bioanalyser (Agilent) and sequencing conducted on an Illumina HiSeq2000 sequencer (MRC Clinical Sciences Centre, Hammersmith).

2.7.4 Site Identification from Short Sequence Reads (SISSRs) (Conducted in collaboration with Geraint Barton at CISBIO)

GlnR binding regions were identified using SISSRs as described in (105). An example call of the SISSRs script is listed below:

perl sissrs.pl -i VJ_V1.bed -o V1all_p0.005.bsites -m 1 -s 6988159 -p 0.005 -b VJ_V3.reallySorted.bed
-i = input file in a BED file format (http://www.ensembl.org/info/website/upload/bed.html)
-o = output results file
-b = background (input control) file in bed format
-m = fraction of the genome mappable by reads

2.7.5 GlnR DNA Binding Consensus Sequence Generated by MEME (Conducted in collaboration with Geraint Barton at CISBIO)

First, a fasta file was created of the 100 bases on either side of each peak, with sequences retrieved from the NCBI GenBank | CP000480 | *Mycobacterium smegmatis* str. MC^2 155 complete genome. (http://www.metalife.com/Genbank/118168627). This was performed in

R using the BSgenome library. The fasta file was then imported into motif based analysis online tool MEME (http://meme.nbcr.net/meme/) (10).

2.7.6 COG Functional Classification (Conducted in collaboration with Geraint Barton at CISBIO)

Genes for each of the *M. smegmatis* main functional roles were downloaded from JCVI (http://cmr.jcvi.org/). The Model Based Gene Set Analysis (MGSA)(14) R package was used for the functional enrichment analysis of the differentially expressed genes against the functional role categories.

2.8 DNA Mobility Shift Assay

2.8.1 DIG 3' Labelling of DNA for Mobility Shift Assay

DNA was labelled using a DIG Oligonucleotide 3' End Labelling Kit (Roche) according to manufacturer's instructions. DNA of the region of interest was amplified using PCR and product size checked via agarose gel electrophoresis, before purification by gel extraction. DNA (100 ng) was made up to a volume of 10 µl with the addition of dH_2O. The following components were added on ice directly to the tube containing the DNA; 4 µl of 5 x labelling buffer (final concentration 1 x; supplied), 4 µl $CoCl_2$ solution (final concentration 5 mM; supplied), 1 µl DIG ddUTP (final concentration 0.05 mM; supplied) and 1 µl terminal transferase (20 Units/ µl; supplied). Following addition, the components were mixed and centrifuged briefly, before incubation at 37°C for 15 min. After incubation the mixture was placed on ice before the addition of 2 µl of 0.2 M EDTA (pH 8.0) to quench the reaction. Three µl of dH_2O was then added to produce a final volume of 25 µl.

Calculation of labeling efficiency was carried out by comparison of spot intensity with the control pre-labelled DNA (Supplied). Spots (1 µl) were placed on a nylon membrane ranging in concentration from neat to 1/1000 in 10-fold dilutions. The Membrane was developed as described in Section 2.8.3. Comparison of spot intensity between labelled DNA and the control pre-labelled DNA at various concentrations permitted calculation into the amount of labelled DNA present.

2.8.2 GlnR:DNA Binding Reaction

The DNA binding reaction was carried out by the addition, on ice, of; 0.4 ng of labelled DNA, specified amounts of protein of interest, 1 µl binding buffer (250 mM Hepes (pH 7.9), 1500 mM NaCl, 250 mM $MgCl_2$) made up to a total of 10 µl with the addition of dH_2O. The contents were mixed briefly and incubated at 37°C for 15 min. After incubation the mixture was returned to ice before the addition of 2.5 µl of TBE high density loading buffer (Invitrogen). Samples were immediately loaded onto a pre-electrophoresed gel (Section 2.8.3).

2.8.3 DNA Retardation Gel Running Conditions

Pre-cast 6% DNA retardation gels (Invitrogen) were used to resolve DNA from the binding reactions. The Invitrogen XCell *SureLock* system was used to run the gel, assembled as per the manufacturer's instructions. TBE running buffer (0.5 x) (Invitrogen) was used to fill the inside and outside chambers of the gel system. The gel was pre-run for 5 min at 100 V, before the wells were flushed and samples loaded. Running of the gel was carried out at 100 V ~60 min, until the loading dye had migrated 2/3 into the gel.

2.8.4 DNA Transfer from DNA Retardation Gel to Nylon Membrane and Membrane Development

The Invitrogen XCell *SureLock* Blot module was used to transfer the labelled DNA to a nylon membrane. Immediately following electrophoresis the blot module was assembled, with all components of the blot sandwich pre-soaked in 0.5 x TBE running buffer. The transfer sandwich composed of, from cathode to anode, blotting pads, filter paper, 6% DNA retardation gel, Hybond-N membrane (Amersham), filter paper, blotting pads. The blot module was assembled with 0.5 x TBE running buffer covering the transfer sandwich in the inner section, and H_2O filled the outside chamber. DNA was electro-blotted onto the membrane for 1 hour at constant 30 V. The membrane was removed and DNA cross-linked to the membrane with a UV stratalinker (Stratagene). The membrane was either developed immediately following UV cross linking or stored overnight at 4°C.

All components for the development of the membrane were source for the DIG wash and Block buffer set (Roche). Stocks of 1 x buffer concentrations were prepared and autoclaved before use. The membrane was rinsed briefly for 1 minute in 1 x washing buffer before incubation in blocking solution for 30 min. Anti-DIG-AP antibody solution (Anti-DIG-AP antibody diluted 1:10,000 with 1x Blocking solution) 30 ml was applied to the membrane and incubated at RT with gentle agitation for 30 minutes. Antibody solution was subsequently drained from the membrane before addition of Wash buffer (30 ml) and incubated at RT for 15 min with gentle agitation. The wash step was repeated with fresh wash buffer for 15 min. Wash buffer was then drained from the membrane before it was equilibrated in 20 ml detection buffer for 5 minutes. The membrane was placed on a transparent film and 1 ml CSPD working solution (supplied) applied to cover the membrane. A transparent film was placed on top to spread the solution evenly over the membrane. This was then incubated at RT for 5 min before excess liquid

removed and the film containing the membrane incubated at 37°C for 10 min. The resulting bands were visualised via a LAS-3000 Fuji imager.

2.9 Analytical techniques

2.9.1 Aquaquant for Quantification of NH_4 Media Concentration

To detect the ammonium concentration in cell free medium Aquaquant (Merk) NH_4 estimations were performed. Fifty µl of culture was centrifuged at 16, 000 x *g* for 2 minutes. A 10 µl sample of the supernatant was removed and transferred to a new eppendorf. Samples were diluted 50-fold with the addition of 450 µl H_20. A further 325 µl H_2O was added to the solution, in addition to 125 µl NH4-1B (supplied). One micro spoon (supplied) of NH4-2B was dissolved in 1 ml H_2O and 50 µl of this suspension add to the sample. The sample was mixed and incubated at RT for 5 min. Two µl of NH4-3B was added to the solution, mixed, and incubated further for 7 min at RT. The colour change was detected via OD readings of the samples at 690 nm. These readings were compared to a standard curve of known ammonium concentrations.

CHAPTER 3: Optimisation of Nitrogen Limiting Conditions for Mycobacteria

3.1 Aim

To optimise a defined mycobacterial growth medium that supports growth of *M. smegmatis* and *M. tuberculosis* and, using this defined medium, determine nitrogen limiting conditions in *M. smegmatis*.

3.2 Introduction

To date, limited reports are available which investigate mycobacterial growth during nitrogen limitation (3, 28, 58-61, 116, 126). Discrepancies exist regarding the nitrogen limiting conditions used, the nitrogen source, and the justification behind its use (Table 3.1). This inconsistency is confounded by the lack of supporting experimental evidence that during mycobacterial growth the medium used in these studies did limit nitrogen availability. As the aim of this project was to investigate the response of *M. smegmatis* to nitrogen limitation, it was important to first optimise and define our nitrogen limiting conditions.

For the growth of mycobacteria a variety of complex and defined mediums are already developed (113). Most commonly used is the commercially sourced Middlebrook medium range. Commercially manufactured growth mediums offer less batch variation; however the manipulation of individual ingredients is arduous, in particular for this project, which requires the addition and removal of individual nitrogen sources. Sauton's medium is a defined mycobacterial growth medium produced in house (113), which permits the manipulation of individual ingredients medium. In addition, Sauton's medium has been used in liquid form since 1912, and permits large scale growth of *M. tuberculosis* (113). With the addition of Tyloxapol and Zinc, it permits homogeneous growth of mycobacteria at comparable rate to that obtained in the widely used Middlebrook 7H9 medium (Data not shown). As such, Sauton's medium was an appealing choice and it was subsequently chosen for use in this study.

Organism	Medium	Nitrogen source	Conc [a]	Comments	Ref
M. tuberculosis	TSA	Ammonium sulphate	0.3 mM/ 30 mM	Contains OADC (potential nitrogen source). No differential growth phenotype noted between 0.3 mM and 30 mM	(28, 116, 126)
M. tuberculosis	7H9	Ammonium sulphate	3.8 mM/ 38 mM	3.8 mM ammonium sulphate is not limiting nitrogen[b]	(60, 61)
M. smegmatis	Kirchner's	Ammonium sulphate	3 mM / 60 mM	60 mM ammonium sulphate causes a notable pH shift[b] 3 mM not limiting nitrogen[c]	(59)
M. smegmatis	7H9	MSX starvation	200 µM	Non-specific MSX[d] affects may influence data	(3)
M. smegmatis	7H9 (lacking ammonium sulphate, glutamic acid and iron ammonium sulphate)	No additional nitrogen source	N/A	Contains OADC (nitrogen source from bovine albumin V fraction and catalase)	(3)

Table 3.1. Previously published nitrogen-limiting mycobacterial growth medium.

[a] Represents concentration of nitrogen source used or MSX added. First number represents nitrogen limiting conditions followed by nitrogen excess

[b] Observed result in this study

[c] Observed result Figure 3.4

[d] MSX methionine sulfoximine. Blocks glutamine synthetase activity.

3.3 Results

3.3.1 Modification of Sauton's Medium for Optimal Growth of *M. smegmatis* and *M. tuberculosis*

In order to study the nitrogen stress response in both *M. smegmatis* (this project) and *M. tuberculosis* (BBSRC LoLa project, Kerstin Williams), an important consideration was the development an identical defined culture medium that permitted growth of both mycobacterial species. Initial growth studies with *M. tuberculosis* revealed that the concentration of glycerol in the original Sauton's medium (6%) was detrimental to *M. tuberculosis* growth (KW, Figure 3.1). In the standard Middlebrook 7H9 medium the glycerol concentration is 0.2% (v/v), making this a logical choice for the glycerol content in Sauton's medium. Reduction of the glycerol content to 0.2% enhanced the growth of *M. tuberculosis* and did not significantly alter the *M. smegmatis* growth rate (Figure 3.1). NMR analysis of the culture medium during *M. smegmatis* growth in 0.2% glycerol also confirmed that the decreased glycerol concentration was not a limiting factor to growth (VB, Figure 3.2). Consequently, 0.2% (v/v) glycerol was used for all subsequent mycobacterial studies.

Sauton's medium contains Tyloxapol (0.025%), a non-ionic detergent to prevent bacterial clumping; however these detergents can present a problem with some downstream analysis applications, for example LC-MS analysis required for metabolomics for the LoLa project. As such, the minimum level of detergent that could be used without compromising bacterial growth was investigated. Based on studies carried out on *M. tuberculosis* (Kerstin Williams) the lowest level of detergent that did not affect growth based on OD_{600}, and gave a homogenous suspension, was 0.015% Tyloxapol, reduced from the original 0.025% (Figure 3.3 A). Similar studies in *M. smegmatis* confirmed that 0.015% Tyloxapol levels did not affect growth (data not shown) and produced a homogenous cell suspension (Figure 3.3 B). Tyloxapol (0.015%) was therefore used in all subsequent mycobacterial studies. The optimised defined Sauton's medium is displayed in Table 3.2; this medium was used for all subsequent growth curves (Figure 3.4 - Figure 3.8).

Original Sauton's Medium Recipe (113)		Modified Sauton's Medium Recipe (This study)	
KH_2PO_4	0.05 % (w/v)	KH_2PO_4	0.05 % (w/v)
$MgSO_4$	0.05 % (w/v)	$MgSO_4$	0.05 % (w/v)
Citric acid	0.2 % (w/v)	Citric acid	0.2 % (w/v)
* Ferric ammonium citrate	0.005% (w/v)	* Ferric ammonium citrate	0.005% (w/v)
Glycerol	**6 % (v/v)**	**Glycerol**	**0.2 % (v/v)**
* Asparagine	0.4 % (w/v)	* Asparagine	0.4 % (w/v)
$ZnSO_4$	0.0001 % (v/v)	$ZnSO_4$	0.0001 % (v/v)
Tyloxapol	**0.025 % (v/v)**	**Tyloxapol**	**0.015 % (v/v)**

Table 3.2. Sauton's Medium Recipe; Original and Modified.

The left-hand column displays the original Sauton's medium recipe as described in (113). On the right, the modified Sauton's recipe, with the changes highlighted in bold.

* For nitrogen limiting medium asparagine is removed and ferric citrate used instead of ferric ammonium citrate.

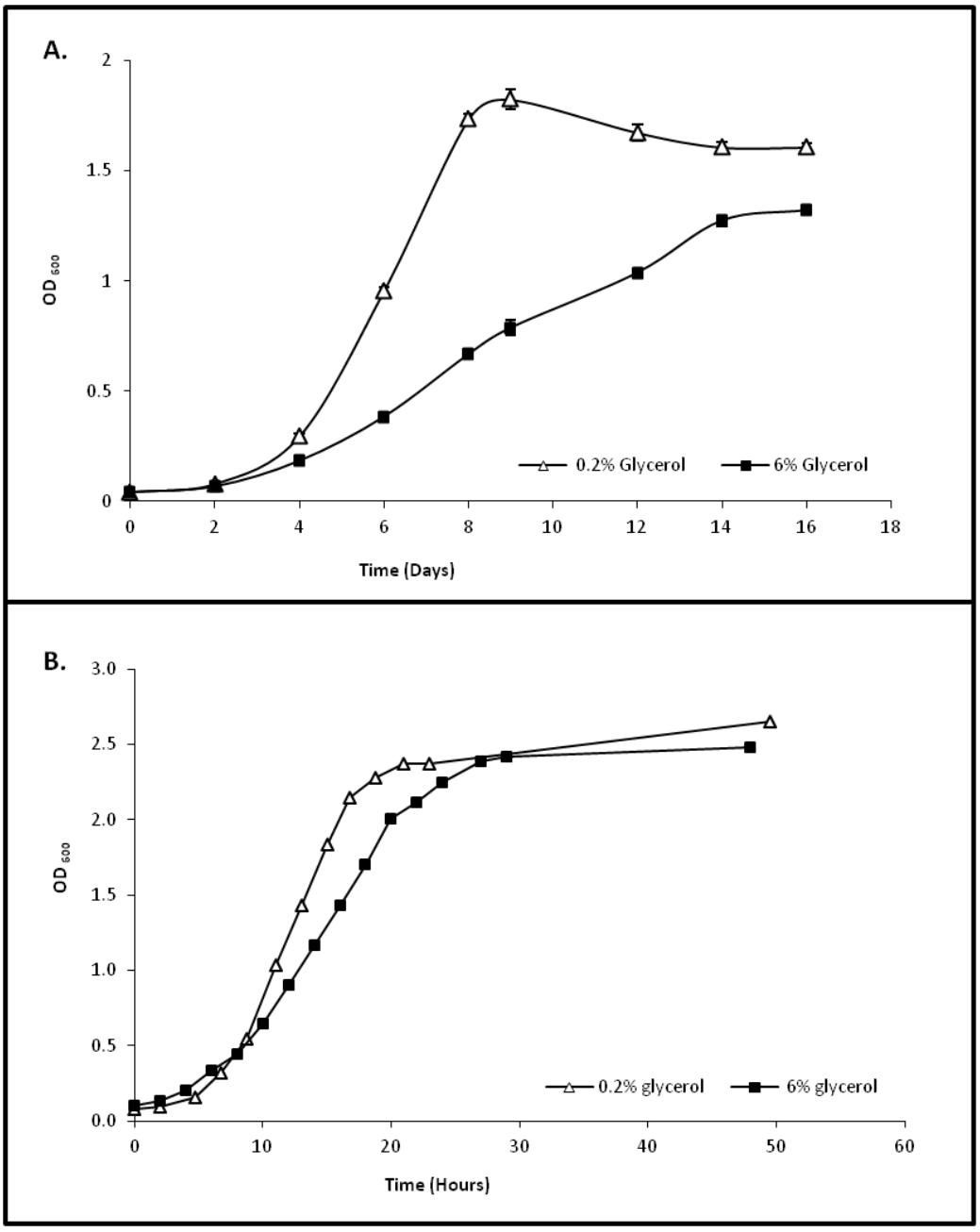


Figure 3.1. Effect of glycerol concentration on the growth of (A) *M. tuberculosis* (KW) and (B) *M. smegmatis.*

M. tuberculosis and *M. smegmatis* were grown in Sauton's modified medium (0.025% (v/v) Tyloxapol) with the addition of 0.2% (open triangles) or 6% (closed squares) (v/v) glycerol. Growth was measured by OD_{600nm}.

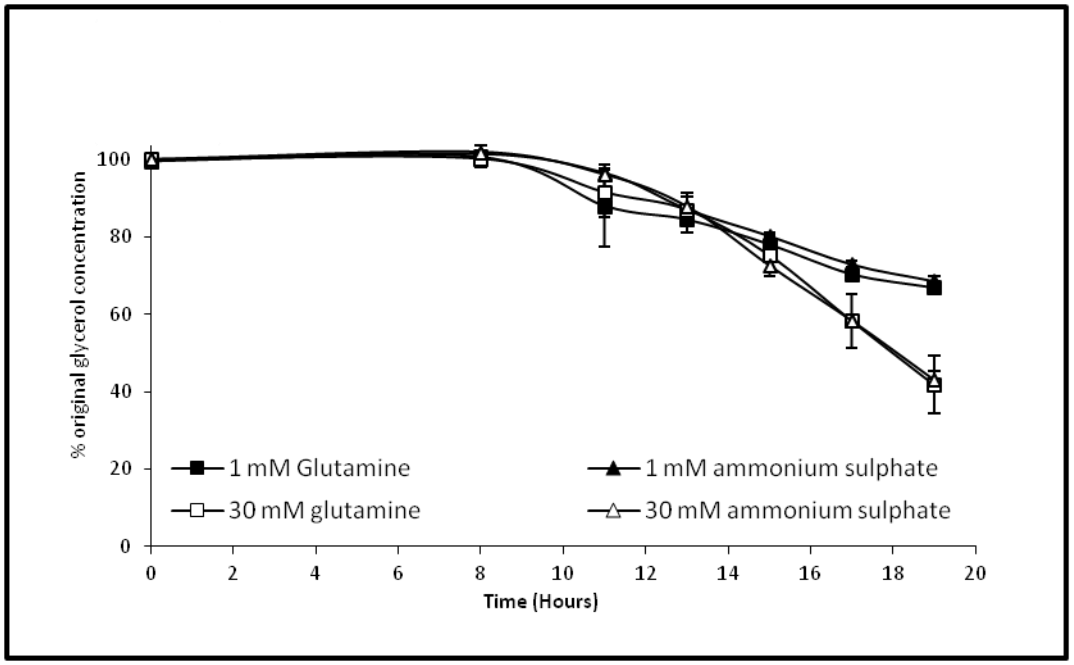


Figure 3.2. NMR analysis of glycerol concentration in modified Sauton's medium (0.2% v/v glycerol) during *M. smegmatis* growth (VB).

M. smegmatis was grown in Sauton's modified medium with the addition of 0.2% (v/v) glycerol and either 30 mM (open shapes) or 1 mM (closed shapes) ammonium sulphate (triangles) or glutamine (squares). Growth, measured by $OD_{600nm,}$ is displayed in Figure 3.1. Culture supernatant was subject to NMR analysis and glycerol intensity plotted as a percentage of original concentration.

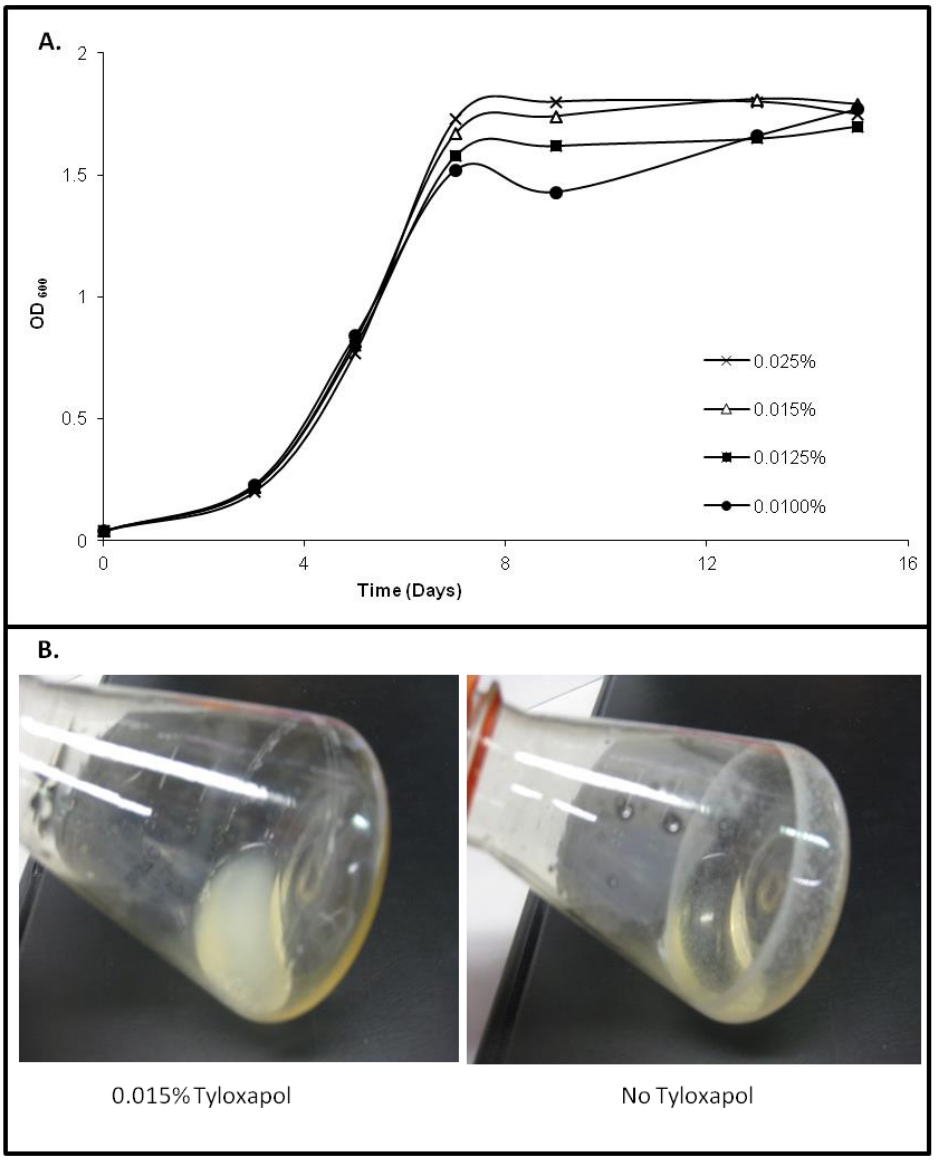


Figure 3.3. Effect of Tyloxapol concentration on the growth of (A) *M. tuberculosis* and (B) *M. smegmatis* (KW).

(A) *M. tuberculosis* was grown in Sauton's modified medium (0.2% (v/v) glycerol) with a range of Tyloxapol concentrations. Growth was measured by OD_{600nm}. (B) Image displaying growth of *M. smegmatis* in medium with 0.015% Tyloxapol or without Tyloxapol addition. Cells appeared to adhere to the side of the flask in medium containing no Tyloxapol.

3.3.2 Optimisation of Nitrogen Limiting Conditions for *M. smegmatis*

Nitrogen limiting medium was optimised by analysing the growth of *M. smegmatis* in a range of glutamine and ammonium sulphate concentrations. As the preferred nitrogen source for mycobacteria is unknown, two different nitrogen sources were chosen for initial study. *M. smegmatis* was grown as described in Section 2.1.2. Essentially, nitrogen free Sauton's medium, was supplemented with the addition of a single nitrogen source at various concentrations. Bacterial growth was monitored over a 24 hour period by OD_{600nm} and cfu/ml. Growth of *M. smegmatis* was proportional to the concentration of nitrogen source contained within the medium. Minimal growth was observed in medium containing no nitrogen source (Figure 3.4).

For nitrogen limiting medium, a nitrogen concentration had to be determined that would stimulate an observable nitrogen-stress response by OD_{600}, yet yield enough cells for downstream analysis. At very low nitrogen levels (0.3 mM), although growth rate of *M. smegmatis* was greatly reduced in comparison to the non-nitrogen limiting medium, the density of cells (OD_{600nm} ~0.4 corresponding to a cfu/ml of 5 $x10^7$) was quite low. Potentially, this low cell number may have proved problematic when obtaining sufficient cells for downstream analyses, such as RNA extraction for microarray analysis. However, a concentration of 1 mM nitrogen source produced a reduction in growth rate observed by OD_{600nm} and a cfu/ml of 5 $x10^8$ at the point of nitrogen limitation (Figure 3.5 and Figure 3.6). This was deemed more suitable for the downstream applications required in this study.

To confirm our optimised medium was nitrogen limiting, the concentration of nitrogen present in the supernatant during *M. smegmatis* growth, was determined. Aquaquant analysis was performed to determine the concentration of ammonium ions in the medium. Samples were centrifuged to pellet the cells and the supernatant retained for further analysis. The first stage of the Aquaquant reaction adjusts the pH to 13, which converts the equilibrium of ammonium to ammonia. Ammonia then reacts with a chlorinating agent to form a monochoramine, which in turn reacts with thymol to form a blue indophenol derivative producing a detectable colour change. The ammonium concentration was then calculated by OD_{690nm} comparison to a standard curve of known ammonium concentrations.

Aquaquant analysis revealed that no ammonium remained in the culture medium after 11 hours *M. smegmatis* growth, in medium containing 1 mM ammonium sulphate (Figure 3.7). This correlated with a decrease in *M. smegmatis* growth rate from this time point onwards (Figure 3.6). NMR analysis was conducted on culture supernatants collected from bacteria grown in 1 mM glutamine and readings indicated that glutamine was depleted from at 11 hours (VB; data not shown), which again correlated with a decrease in growth rate (Figure 3.5). This suggests

that nitrogen depletion leads to the reduction in growth rate seen in our nitrogen limiting medium when compared to our nitrogen excess medium.

A non-limiting, nitrogen excess medium was also established. Initial results indicated that 30 mM or 3 mM of both glutamine and ammonium sulphate produced similar growth dynamics to the original un-modified Sauton's medium (Figure 3.4). Analysis using Aquaquant to determine the ammonium concentration left in the medium at a given point indicated that in the 3 mM cultures ammonium levels were close to depletion by 24 hours (Data not shown). In contrast, Aquaquant of the high (30 mM ammonium sulphate) supernatant showed nitrogen did not run-out (Figure 3.7) and NMR analysis showed no run-out of glutamine (Data not shown; VB). In addition, analysis of the medium's pH showed that for concentrations greater than 30 mM ammonium sulphate there was a notable shift in the starting pH of the medium (ammonium sulphate 1 mM: pH 7.4, 30 mM: pH 7.2 and 60 mM: pH 7.1). Therefore, 1 mM nitrogen source was used for our nitrogen limiting medium and 30 mM for our nitrogen excess medium.

In general, both glutamine and ammonium sulphate gave similar growth rates and phenotypes in excess and nitrogen limiting conditions (Figure 3.5 and Figure 3.6). As the phenotypes were similar it was assumed that the responses generated, as a consequence of nitrogen limitation, would be analogous. With Aquaquant detection permitting a fast and relatively simple measure of the rate of ammonium depletion, ammonium sulphate was consequently chosen for use from this point on as the nitrogen source in Sauton's medium.

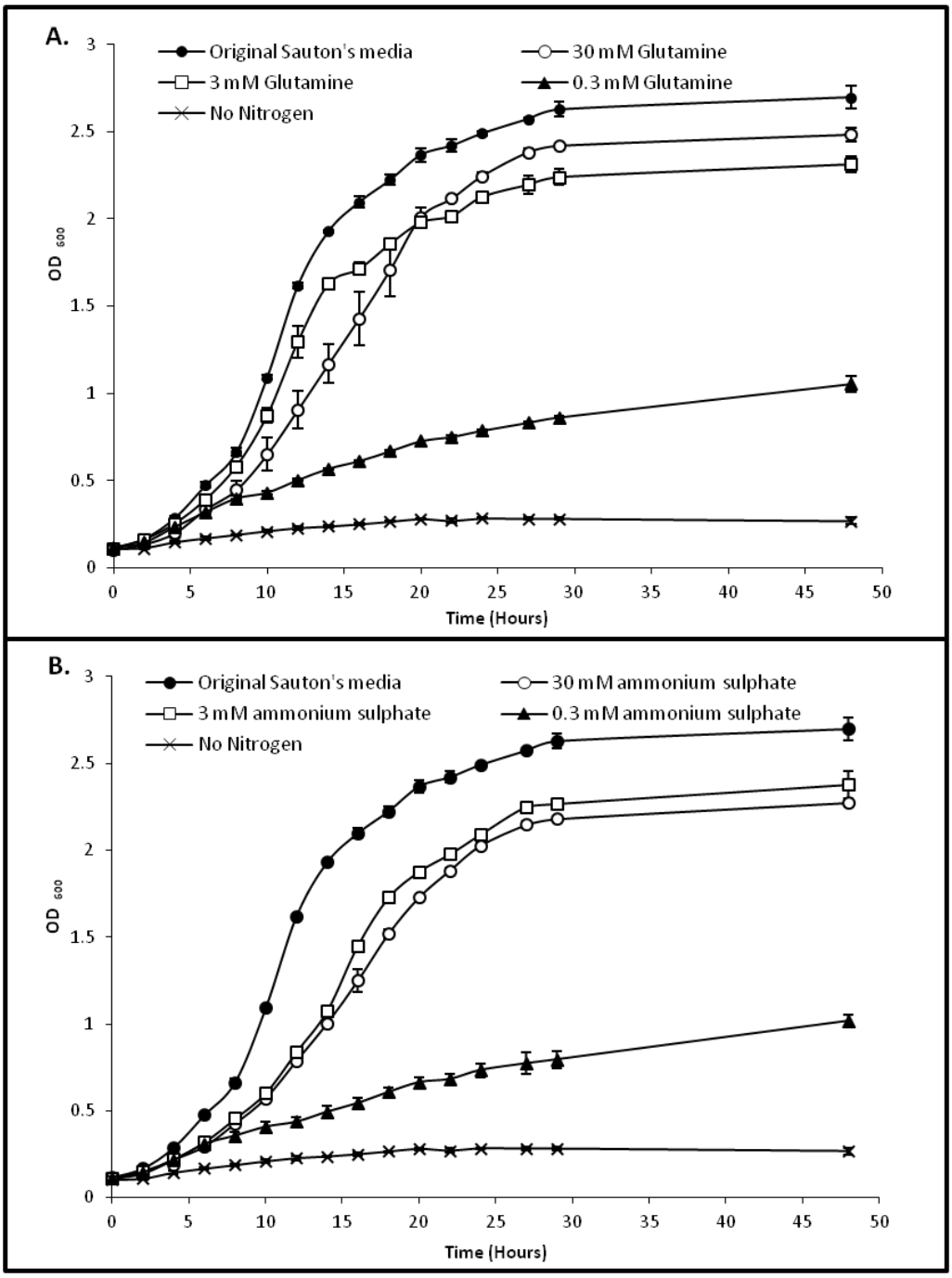


Figure 3.4. Growth of *M. smegmatis* in Sauton's modified medium with nitrogen sources at various concentrations.

M. smegmatis was grown in Sauton's modified medium with either (A) glutamine or (B) ammonium sulphate at concentrations of 30 mM (open circles), 3 mM (open squares) or 0.3 mM (closed triangles). Sauton's modified medium with no nitrogen addition (X) and Original Sauton's medium (closed circles), with asparagine and ferric ammonium citrate as the nitrogen source included as controls. Growth was measured by monitoring OD_{600nm}, data represents the average (±SD) of three independent experiments.

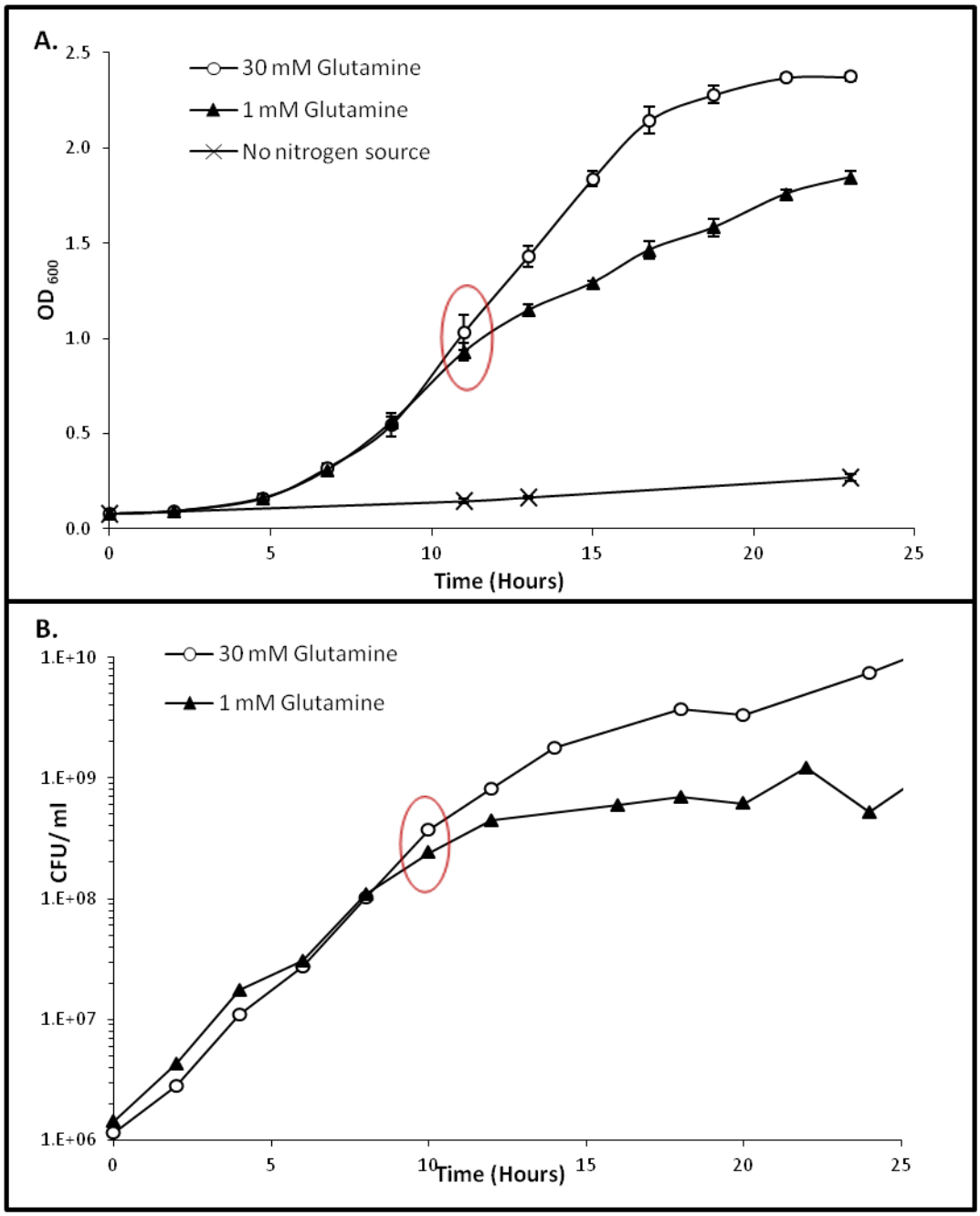


Figure 3.5. Growth of *M. smegmatis* in Sauton's modified medium with 1 mM or 30 mM glutamine.

M. smegmatis was grown in Sauton's minimal medium with the addition of either 30 mM (open circles) or 1 mM (closed triangles) glutamine. Growth was monitored by (A) OD_{600nm} or (B) cfu/ml. Nitrogen run out in the (1 mM) glutamine containing medium was determined to be at 11 hours by NMR analysis (highlighted by the red oval). Data represents the average (±SD) of three independent experiments.

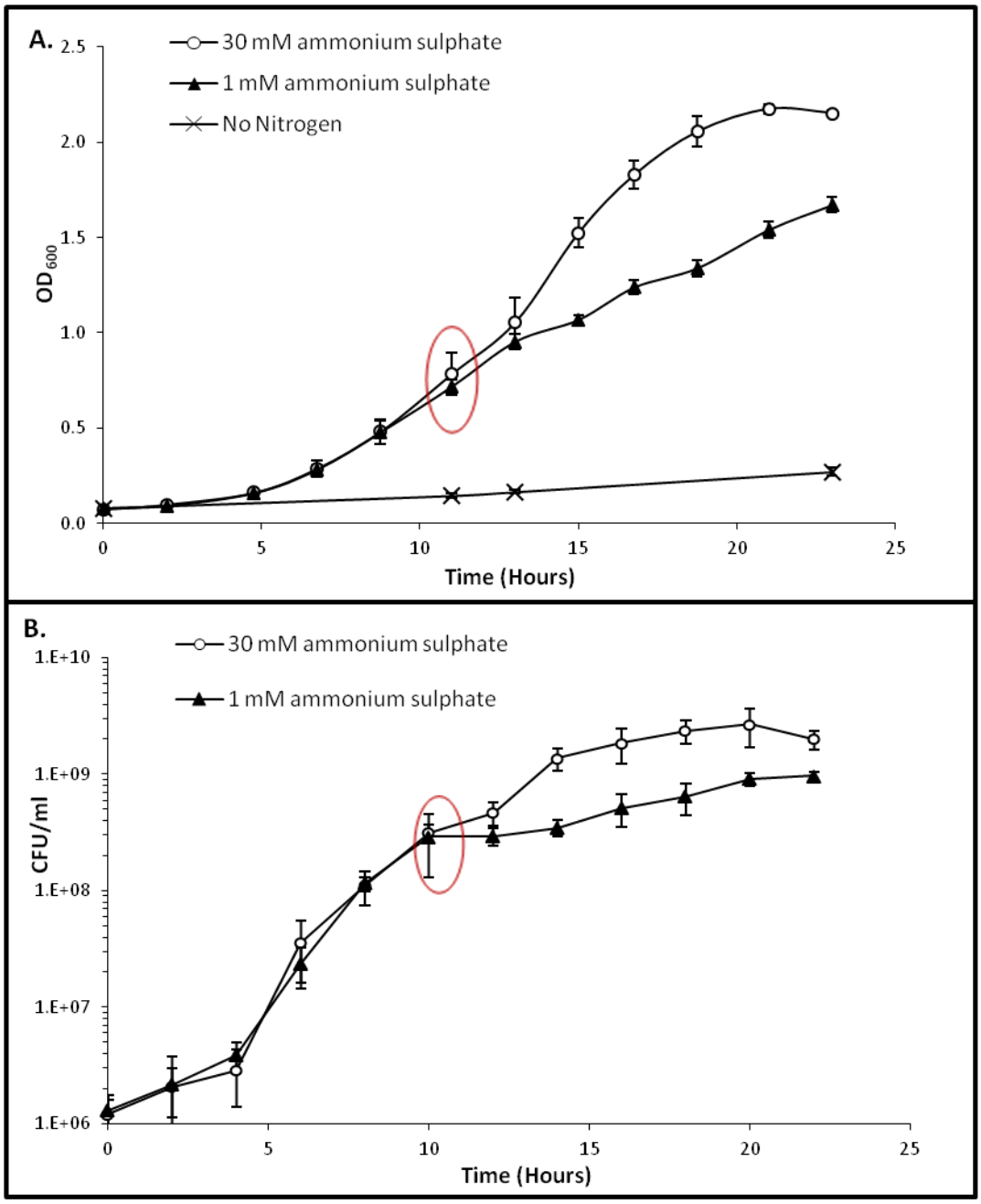


Figure 3.6. Growth of *M. smegmatis* in Sauton's modified medium with 1 mM or 30 mM ammonium sulphate.

M. smegmatis was grown in Sauton's minimal medium with either 30 mM (open circles) or 1 mM (closed triangles) ammonium sulphate. Growth was monitored by (A) OD_{600} or (B) cfu/ml. Nitrogen run out in the (1 mM) nitrogen containing medium was determined by Aquaquant analysis to be at 11 hours (highlighted by a red oval) (Figure 3.7). Data represents the average (±SD) of three independent experiments.

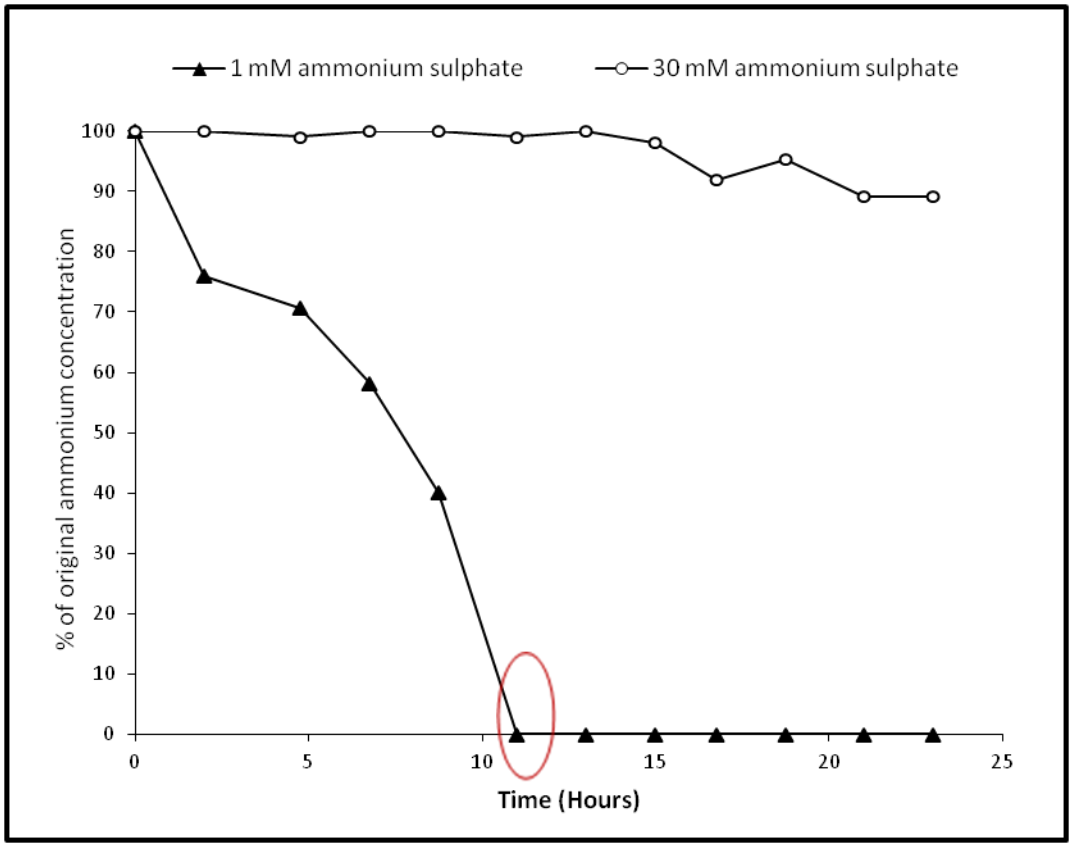


Figure 3.7. Aquaquant analysis of ammonium concentration in Sauton's modified medium during *M. smegmatis* growth.

Aquaquant analysis was conducted on the supernatants taken during *M. smegmatis* growth (Figure 3.6) in Sauton's modified medium containing low (closed triangles) 1 mM ammonium sulphate or (open circles) 30 mM ammonium sulphate. The graph displays the Aquaquant analysis readings as a percentage of the starting ammonium concentration. Depletion of ammonium from the 1 mM medium occurred at 11 hours, correlating with the reduction in growth rate observed in Figure 3.6.

3.3.3 Spiking of *M. smegmatis* Cultures in Nitrogen Excess and Limiting Medium with Nitrogen Source

In order to confirm conditions were limiting for only nitrogen and not any other substance, cultures were spiked with the addition of an exogenous nitrogen source at the time of nitrogen limitation. Cells were grown for 12 hours in Sauton's medium, containing 1 mM or 30 mM ammonium sulphate. It was assumed, that if nitrogen limitation was limiting growth of *M. smegmatis*, nitrogen addition would promote growth. Both nitrogen limiting and excess ammonium sulphate containing medium was spiked with the addition of 10 mM ammonium sulphate, made up in water; water alone was added as a no nitrogen-spike control. Growth was monitored by measuring OD_{600nm} and cfu/ml, with Aquaquant readings taken to confirm nitrogen run out in the low nitrogen sample.

As expected, growth was not affected by the addition of 10 mM ammonium sulphate to the nitrogen excess (30 mM) cultures (Figure 3.8). Addition of water alone did not alter the growth dynamics of the control sample and Aquaquant readings confirmed the presence of nitrogen in these samples at the point of nitrogen addition.

In both the control and the spiked nitrogen limiting cultures (1 mM ammonium) the nitrogen had run out by 12 hours (Figure 3.8). With the addition of 10 mM ammonium sulphate to the culture at 12 hours, growth rate increased compared to the water-spiked control (Figure 3.8 A); cfu/ml readings confirmed this observation (Figure 3.8 B). Growth of the nitrogen-spiked culture was at a similar rate to the nitrogen excess, non-limiting, medium and the final OD_{600nm} was similar. This observation provides definitive evidence that 1 mM nitrogen is limiting growth rate in our optimised medium. Therefore, for all subsequent analysis nitrogen excess medium contained 30 mM nitrogen and nitrogen limiting medium contained 1 mM.

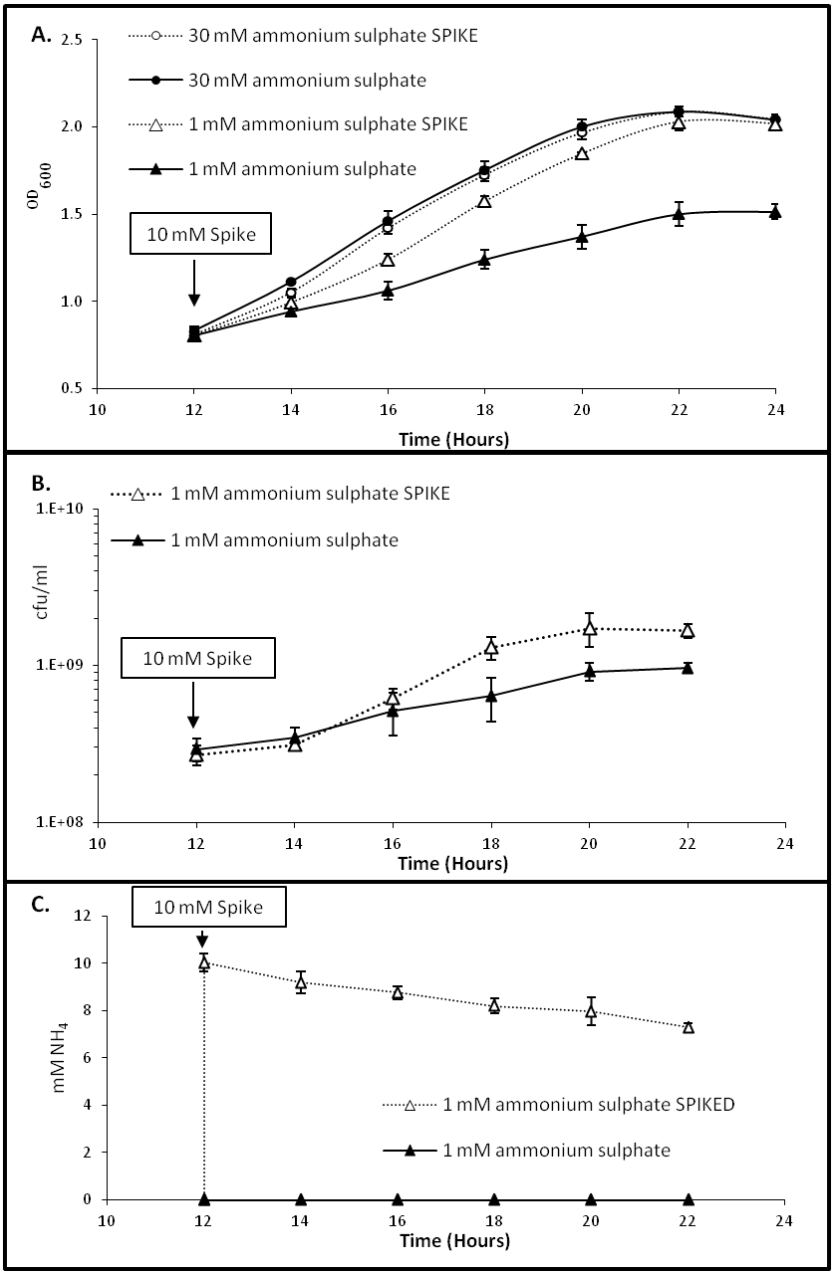


Figure 3.8. Growth of *M. smegmatis* following addition of exogenous (10 mM) ammonium sulphate after nitrogen depletion.

M. smegmatis was grown in Sauton's minimal medium with 1 mM (triangles) or 30 mM (circles) ammonium sulphate for 12 hours. At 12 hours 10 mM ammonium sulphate was added to the cultures (spike, open shapes), or water alone (control, coloured shapes). Growth was measured with change in (A) OD_{600nm} and (B) cfu/ml. (C) Aquaquant analysis on the culture supernatants confirming nitrogen run out at 12 hours for the low cultures (closed triangles) and ammonium addition (10 mM) at 12 hours to the spiked sample (open triangles). Data represents the average (±SD) of three independent experiments.

3.4 Discussion

The overall aim of this project was to investigate mycobacterial adaptation to nitrogen-limitation, therefore it was a fundamental prerequisite that our defined liquid medium was limiting for nitrogen. To date, a few reports are available that investigate mycobacterial growth during nitrogen limitation (3, 28, 58-61, 116, 126). However, inconsistency between the nitrogen limiting conditions used, and the reasons behind their use, made reliance on these studies difficult.

Mycobacterial nitrogen-limiting conditions exist in the literature, however each study fails to convincingly justify the medium choice and nitrogen concentrations used (Table 3.1) (28, 58-61, 116, 126). Pashley *et al.* developed a *M. tuberculosis* liquid medium, TSA, supplemented with 0.1 mM and 30 mM ammonium sulphate for nitrogen limiting and excess conditions (116). However, no difference in the growth dynamics of *M. tuberculosis* were observed between the two conditions (116). One possibility for the lack of variation is that OADC, a supplement in TSA, contains protein sources (Bovine Albumin V fraction and catalase), and possible degradation may lead to the availability of an alternative nitrogen source. Harth *et al.* used 7H9 medium, supplemented with ammonium sulphate (3.8 mM limiting and 38 mM excess) (60, 61). While Harper *et al.* used Kirchner's minimal medium in which asparagine was replaced with ammonium sulphate concentrations of 3 mM and 60 mM (59). We noted that the pH of our medium, containing greater than 30 mM ammonium sulphate, decreased from pH 7.4 to 7.1, possibly influencing the growth dynamics seen for the medium containing 60 mM ammonium sulphate. In addition, Aquaquant analysis during our investigations indicated that 3 mM ammonium sulphate was only depleted from *M. smegmatis* cultures when the cells had reached stationary phase, after 24 hours growth. As the samples in these studies were taken at mid-log it queries whether this medium did limit nitrogen availability and whether their conclusions with respect to nitrogen-limitation are valid (59-61).

An alternative approach to investigate *M. smegmatis* growth under nitrogen-limitation is to induce nitrogen-starvation with the addition of methionine sulfoximine (MSX) (3, 60). MSX addition to growing cells has been demonstrated to block glutamine synthetase (GS) activity, preventing glutamine metabolism via this pathway (60, 96). A concern using this approach arises; the nitrogen starvation response is not necessarily being investigated, but rather the effect of a non-functional GS. Transcriptional data may then only reflect the product, glutamine, starvation in the cell rather than complete nitrogen starvation, as other nitrogen sources are contained in the growth medium. In addition, non-specific effects of MSX may result in inaccurate representation of cellular responses. Consequently, this approach was determined to

be unsuitable for use in this study. Thus we concluded, after examination of the published medium, for confidence in our results we first needed to optimise our own nitrogen-limiting conditions.

For the growth of mycobacteria a variety of complex and defined mediums are already developed (113). Most commonly used is the commercially sourced Middlebrook range. Commercially manufactured growth mediums offer less batch variation, however the manipulation of individual ingredients is laborious. Defined mycobacterial growth medium produced in house, such as Sauton's, permits the manipulation of individual ingredients. Sauton's medium has been used as a mycobacterial liquid growth medium since 1912, and supports large scale growth of *M. tuberculosis* (113). As such Sauton's medium was assessed for its suitability for use in this study.

Despite the established use of Sauton's as a mycobacterial liquid growth medium, problems were noted with initial growth analysis. The high glycerol content of the medium proved to be detrimental to *M. tuberculosis* H37Rv growth. Middlebrook 7H9 contains 0.2% (v/v) glycerol, and this concentration was used in Sauton's medium. Reduction of the glycerol content in Sauton's medium to 0.2% (v/v) enhanced the growth of *M. tuberculosis* and did not significantly alter the *M. smegmatis* growth rate. NMR analysis of the culture-supernatant collected during *M. smegmatis* growth in 0.2% glycerol also confirmed that the decreased glycerol concentration was not a limiting factor to growth. Further modification, to reduce the level of Tyloxapol (a non-ionic detergent) in the medium, was examined. In the absence of a detergent the hydrophobic nature of the mycobacterial cell-surface results in clumping of bacilli and reduced bacterial growth (162). A non-ionic detergent prevents clumping, providing a homogenous solution, resulting in a reproducible and increased growth rate (162). Despite their advantages, detergents can be problematic for some analysis applications, such as LC-MS metabolomic profiling required for the LoLa project (Kerstin Williams and Volker Behrends). As such, the minimum level of Tyloxapol that could be used without compromising bacterial growth was determined to be 0.015%, reduced from 0.025%. These alterations, reduction in glycerol (0.2% v/v) and Tyloxapol (0.015% v/v) concentrations, gave us the basis for our standard liquid growth medium, permitting further investigation into the effects of nitrogen availability on mycobacterial growth.

Once the composition of Sauton's medium had been optimised, *M. smegmatis* growth in various concentrations of nitrogen sources was examined. As the preferred nitrogen source and signal for nitrogen-limitation in mycobacteria is unknown two nitrogen containing compounds, ammonium sulphate and glutamine, were investigated. Initially, growth analysis in 0.3 mM

ammonium sulphate and glutamine, indicated that at the point of nitrogen-limitation cell density was low, problematic for downstream applications such as ChIP-seq and microarray analysis. A nitrogen concentration of 3 mM was also examined, as this concentration had been used previously for nitrogen-limitation (59-61). However, at 3 mM a growth phenotype similar to 30 mM control and the original Sauton's medium was observed, suggesting that 3 mM nitrogen was not limiting. Finally, *M. smegmatis* grown in a nitrogen concentration of 1 mM produced a clear phenotype when compared to 30 mM and the original Sauton's medium. Quantification of nitrogen remaining in the medium confirmed a correlation between reduced growth rate and nitrogen depletion. In addition, at the point the growth rate started to reduce under nitrogen limitation (1 mM glutamine or 1mM ammonium sulphate) cell density was approximately 5 $x10^8$ cfu/ml, a level considered more suitable for downstream applications. To further confirm 1 mM ammonium sulphate was limiting *M. smegmatis* for nitrogen only, exogenous ammonium sulphate was added back to the medium at the point of nitrogen run-out. Spiking of the 1 mM medium with exogenous nitrogen increased growth rate, when compared to the non-spiked control, to a level comparable to the 30 mM medium. This observation provides further evidence that 1 mM nitrogen is limiting growth rate and nitrogen addition after run-out can restore growth. Consequently 1 mM nitrogen source, glutamine or ammonium sulphate, was chosen for our nitrogen limiting medium.

A nitrogen excess medium, which did not limit the growth rate of mycobacteria, was also required for comparison in this study. Early observations had demonstrated that a concentration of 3 mM for glutamine and ammonium sulphate produced a similar phenotype to the 30 mM nitrogen source and the original Sauton's medium. However, Aquaquant analysis on the 3 mM ammonium sulphate sample indicated that nitrogen was depleted after 24 hours growth when the cells had entered stationary phase, unsuitable for a nitrogen-excess medium. Growth rate similar to the original Sauton's medium was also noted for 30 mM ammonium sulphate and glutamine concentrations. Aquaquant and NMR analysis confirmed that for the 30 mM concentrations the nitrogen contained in the medium had not been depleted during growth. Spiking of the medium at 12 hours, the point of nitrogen run out for the limiting medium, did not affect the growth rate of *M. smegmatis*, indicating that nitrogen was not a limiting factor to growth in Sauton's medium containing 30 mM ammonium sulphate. Consequently our nitrogen excess medium was determined to be 30 mM ammonium sulphate and 30 mM glutamine.

Due to insufficient evidence regarding the merits of published nitrogen-limiting conditions, we began by establishing and characterising a new medium to limit nitrogen availability for mycobacteria with two nitrogen sources. However, the growth phenotypes for ammonium sulphate and glutamine were comparable, and due to the relative ease of quantifying the

amount of ammonium ions in the medium, ammonium sulphate was chosen as the nitrogen source for all subsequent analysis. Nitrogen limiting and excess conditions were therefore established to be 1 mM and 30 mM ammonium sulphate.

CHAPTER 4: Aspartate 48 is Essential for the GlnR-Mediated Transcriptional Response to Nitrogen Limitation in *Mycobacterium smegmatis*

4.1 Aim

To investigate the role of the putative GlnR phosphorylation site, Asp-48, with regards to *M smegmatis* GlnR functionality during nitrogen limitation. Primarily focusing on the role of GlnR Asp-48 with regards to the transcriptional response of GlnR regulated genes during nitrogen limitation.

4.2 Introduction

GlnR is thought to be the global transcriptional regulator in mycobacteria in response to nitrogen limitation, however the activation mechanism and post-translational modifications of GlnR are unknown. Transcript levels of *glnR* do not significantly alter during nitrogen limitation, suggesting *glnR* transcription is not regulated in response to nitrogen availability, but rather GlnR activity is subject to an alternate control mechanism (3). Bioinformatic analysis of *M. smegmatis* GlnR places the protein within the OmpR family of two-component response regulators (3). Typically, OmpR-type response regulators are transcriptional activators, phosphorylated by a sensor kinase in response to extracellular stimuli (80). A prominent feature of the OmpR family is a highly conserved aspartate residue, which undergoes phosphorylation by a sensor kinase. Known phosphorylation sites for related proteins include OmpR Asp-55, phosphorylated by OmpZ, PhoB Asp-53 phosphorylated by PhoR, and CheY Asp-57 phosphorylated by CheA (39, 41, 205). Sequence alignments of *M. smegmatis* GlnR to other OmpR-family response regulators indicates the presence of a corresponding conserved residue, Asp-48, suggesting that GlnR may undergo phosphorylation during nitrogen limitation (Figure 4.1) (3). However, phosphorylation of GlnR has yet to be confirmed, possibly due to the labile nature of the phospho-aspartate bond making the detection of this modification by conventional methods problematic.

Consequently, the importance of this putative phosphorylation site, with regard to the functionality of GlnR in response to nitrogen limiting conditions, was investigated. In this study, a recombineering approach was applied to create a chromosomal point mutation in *M. smegmatis,* changing the GlnR Asp-48 residue to alanine, and the effects studied.

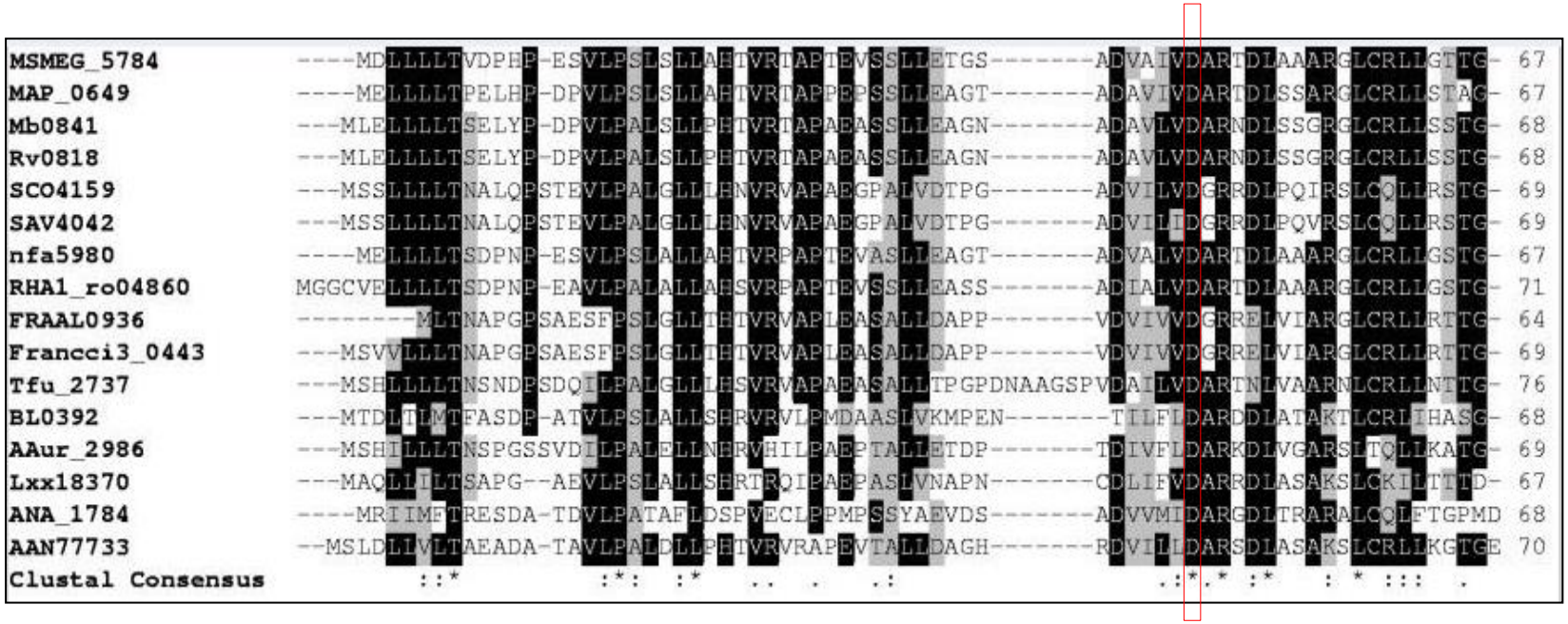


Figure 4.1. Sequence alignment of the N-terminal domain of GlnR-type proteins in various Actinomycetes (adapted from (3)).

A black background indicates identical amino acid residues, and a gray background indicates similar amino acids. *M. smegmatis* GlnR contains a putative phosphorylation pocket, located at amino acid residues 43 to 53, with the conserved putative phosphorylation site (GlnR D48) indicated with a red rectangle.

Abbreviations used for GlnR homologs with annotation numbers: MSMEG, *Mycobacterium smegmatis*; MAP, *Mycobacterium avium* subsp. *paratuberculosis*; Mb, *Mycobacterium bovis*; Rv, *Mycobacterium tuberculosis* H37rv; SCO, *Streptomyces coelicolor*; SAV, *Streptomyces avermitilis*; nfa, *Nocardia farcinica*; RHA1, *Rhodococcus* sp. strain RHA1; FRAAL, *Frankia alni* ACN14a; Francci3, *Frankia* sp. strain Cci3; Tfu, *Thermobifida fusca*; BL, *Bifidobacterium longum*; AAur, *Arthrobacter aurescens*; Lxx, *Leifsonia xylii*; ANA, *Actinomyces naeslundii*; AAN77733, *Amycolatopsis mediterranei*.

4.3 Results

4.3.1 Construction of GlnR_D48A Mutant

M. smegmatis cells containing the recombineering vector pJV128 were used to generate a chromosomal point mutation. The pJV128 plasmid contains a mycobacterial phage gene product, Che9c 61, which facilitates ssDNA recombination (174). In addition, the plasmid pJV128 contains a hygromycin resistance gene with two adjacent nonsense mutations inactivating its function (*hygS*). Colonies that have undergone recombination can therefore be selected by co-electroporation of a 100 base oligonucleotide to correct the *hygS* mutation, producing hygromycin resistant colonies. As such, a chromosomal *glnR* aspartate 48 to alanine (GlnR_D48A) substitution was generated using this approach. Two single-stranded oligonucleotides one containing the GlnR_D48A point mutation, the other to repair the *hygS* cassette, were co-transformed into *M. smegmatis*_ pJV128 cells.

A mismatch amplification mutation assay (MAMA) screen was performed to selectively amplify the *glnR* point mutation (Section 2.5.6). The principal of MAMA PCR is that a single nucleotide mismatch at the 3′ extremity of the annealed reverse primer renders *Taq* polymerase unable to extend the primer (29, 160). Therefore, the absence of the specific PCR product reveals a deviation from the desired DNA sequence. As we were searching for a two base pair change within the msmeg_5784 gene, we incorporated these changes into the 3' extremity of the reverse primer. As such a PCR product would be produced if the D48A mutation were present, while the wild type produces no product. Over 100 colonies were screened to find a positive result for the GlnR_D48A substitution, and Figure 4.2 displays a positive PCR product for the chromosomal point mutation (Figure 4.2 A). Genome sequencing of the *glnR* gene was subsequently performed on the GlnR_D48A mutant to confirm the desired chromosomal point mutation and that no other changes were present.

Counter selection of the *sacB* gene on pJV128 allowed removal of the plasmid and was confirmed via PCR. The pJV128 plasmid contained a *sacB* casette that confers sucrose sensitivity, allowing selection on sucrose for colonies that no longer contain pJV128 (117, 118). Serial dilutions of the mutant were plated on 2% sucrose and replica plated on medium supplemented with hygromycin and kanamycin. Colonies which displayed sucrose resistance and antibiotic sensitivity, had lost plasmid pJV128. PCR was then performed as described on a sucrose resistant colony with primers designed to amplify the hygromycin cassette present on the pJV128 plasmid. Figure 4.2 B indicates the colony had lost pJV128. As a result we obtained an unmarked chromosomal point mutation.

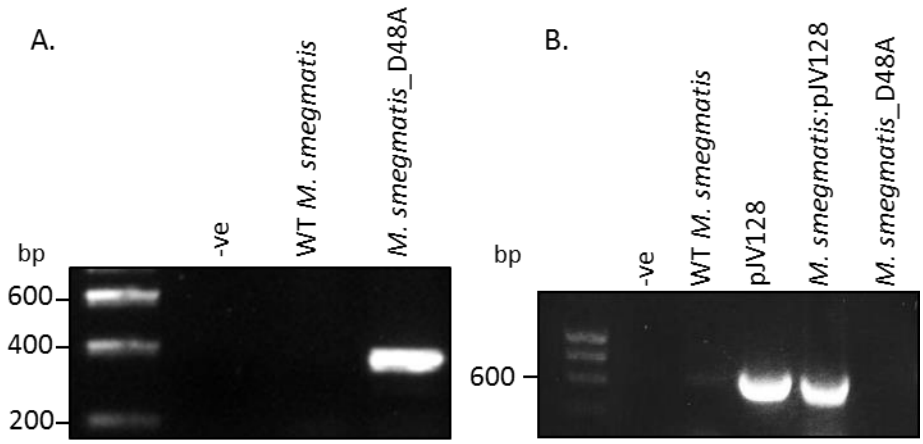


Figure 4.2. MAMA PCR of *M. smegmatis* GlnR_D48A mutant and PCR to confirm loss of plasmid pJV128 after sucrose selection.

(A) MAMA PCR on *M. smegmatis* WT and GlnR_D48A genomic DNA. GlnR_D48A displays a 350bp fragment indicating required incorporation of the point mutation onto the chromosome in this strain. (B) PCR on the hygromycin cassette on the pJV128 vector. PCR was performed on mini preparations from selected strains. Strains containing the hygromycin cassette show product amplification at 600 bp. Lack of amplification in the *M. smegmatis* GlnR_D48A strain indicates plasmid loss.

4.3.2 Construction of the *glnR* Deletion Strain

In order to study GlnR mediated transcriptional activation during nitrogen limitation, a *M. smegmatis glnR* deletion strain was required for comparison. For the generation of the *glnR* deletion mutant, the recombineering approach was again utilised (175, 176). *M. smegmatis* strains containing the pJV126 plasmid were used. This plasmid again contained a *sacB* cassette for selective removal of pJV126, and also phage genes encoding Che9c 60 and 61 enabling homologous recombination of linear dsDNA (118, 175, 176). As such an allelic exchange substrate (AES) was generated as described in Section 2.5.4, permitting homologous recombination of 800 bp regions flanking the *glnR* gene. A hygromycin cassette was inserted between the flanking regions, ultimately creating a marked deletion strain. Linear AES dsDNA was transformed into *M. smegmatis*_pJV126 and putative *glnR* deletion colonies were then selected on hygromycin plates.

Putative null mutants were confirmed by PCR with oligonucleotides specific for the hygromycin cassette and a site outside the *glnR* flanking regions used to construct the mutant. As such, PCR products would only be obtained if the hygromycin cassette had inserted into the correct location on the chromosome. PCR products of the expected size (approximately 1.5 kb) were obtained for the GlnR deletion mutant; no products were obtained for the wild type strain (Figure 4.3 A). Additional confirmation that the *glnR* gene had been disrupted was provided by Western blot, using a polyclonal GlnR specific antibody (Section 2.4.3), showing the absence of the corresponding 27.9 kDa GlnR protein in the mutant strain compared to the wild type (Figure 4.3 B).

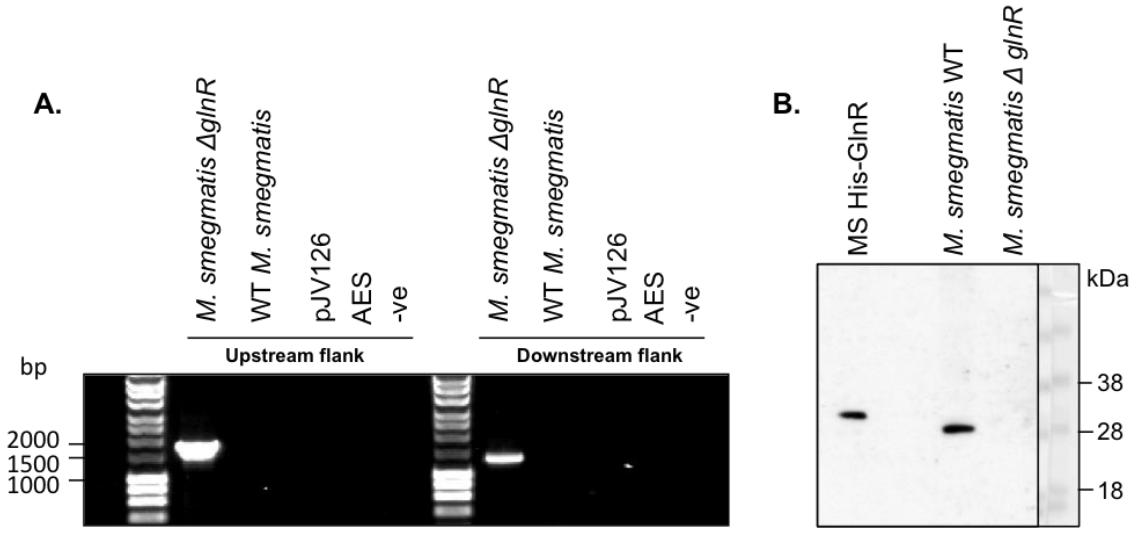


Figure 4.3. Confirmation of a *M. smegmatis glnR* deletion strain.

(A) PCR confirming the construction of a *glnR* deletion strain. Amplification of 1.5 kbp fragments indicate inclusion of the hygromycin cassette at the site of the *glnR* gene. (B) Western blot analysis of *M. smegmatis* strains incubated with an affinity purified polyclonal anti-GlnR antibody. Lane 1: 12.5ng of recombinant *M. smegmatis* His-GlnR protein, Lane 2: *M. smegmatis* wild type cell extracts (20 µg) and Lane 3: *M. smegmatis* GlnR deletion strain cell extracts (20 µg). Lane 2 displays a band at 28 kDa corresponding to the native GlnR protein, which is absent in the mutant.

4.3.3 Generation of GlnR_D48A and *ΔglnR* Complementation Strains

For complementation of the *M. smegmatis glnR* mutants, plasmid pMV306 was chosen (157). Plasmid pMV306 is an integrating vector, integrating at the *attB* site on the mycobacterial chromosome, providing plasmid stability for long-term growth analysis (157). The *glnR* gene was cloned into the vector under its own predicted promoter; an 80 bp region up-stream of the *glnR* gene was incorporated to allow inclusion of promoter elements. Cloning was performed as described and the construct sequence was confirmed via plasmid sequencing. Mutant *M. smegmatis* strains were transformed with the plasmid and *glnR* expression from the plasmid was confirmed via qRT-PCR (Figure 4.4).

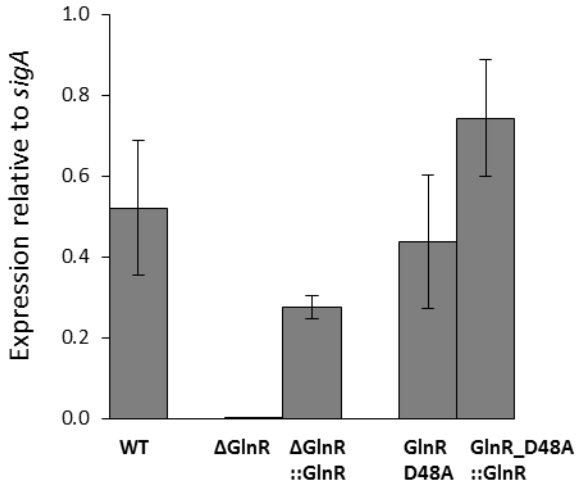


Figure 4.4. qRT-PCR to show *glnR* expression levels during nitrogen limitation.

Figure represents the average expression of *glnR* from three independent experiments relative to *sigA*. Expression of *glnR* is restored in the *glnR* deletion strain by complementation of the *glnR* gene on the pMV306 plasmid.

4.3.4 GlnR Mutants Exhibit Reduced Growth Rates During Nitrogen Limitation

Growth kinetics of the *M. smegmatis* wild type, D48A and *ΔglnR* strains grown in the optimised nitrogen limiting and nitrogen excess conditions were analysed. Growth dynamics of each strain were monitored by OD_{600} and CFU/ml over 24 hours. Aquaquant readings were also taken to determine ammonium levels in the supernatant.

All *M. smegmatis* strains grew similarly in nitrogen excess conditions (Figure 4.5). However, during nitrogen limitation the *glnR* deletion and GlnR_D48A mutants exhibited a reduced growth rate when compared to the wild type (Figure 4.6). Growth rate could be restored by complementing both mutants by reintroduction of the *glnR* gene on the pMV306 vector (Figure 4.6). Another interesting observation was the increased levels of ammonium in the media, suggesting reduced uptake of ammonium by both mutants (Figure 4.6). Again ammonium uptake levels could be restored to wild type levels by reintroduction of a functional *glnR* gene. As such it is apparent that both mutants exhibit a reduced growth phenotype during nitrogen limitation, with both the *glnR* deletion strain and GlnR_D48A mutant having very similar phenotype to each other, suggesting that the D48A residue is important for wild type growth rate in nitrogen limiting media. As such the transcriptomic effect of the D48A and *ΔglnR* mutants were investigated.

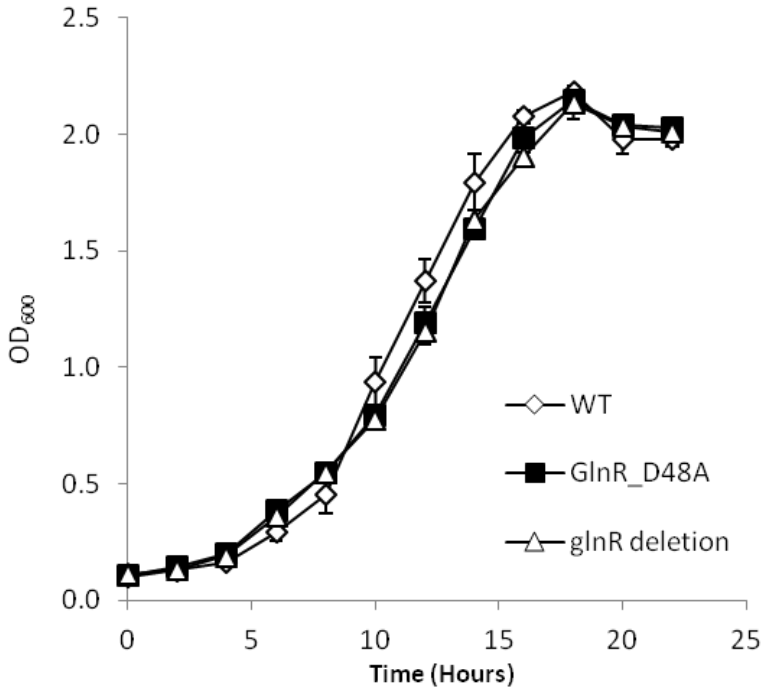


Figure 4.5. Growth analysis of wild type *M. smegmatis* and *glnR* mutants in nitrogen excess medium (30 mM ammonium sulphate).

Growth analysis monitored by OD_{600} of *M. smegmatis* wild type (open diamonds), GlnR_D48A (closed squares) and GlnR deletion (open triangles). Data represents the average (±SD) of three independent experiments.

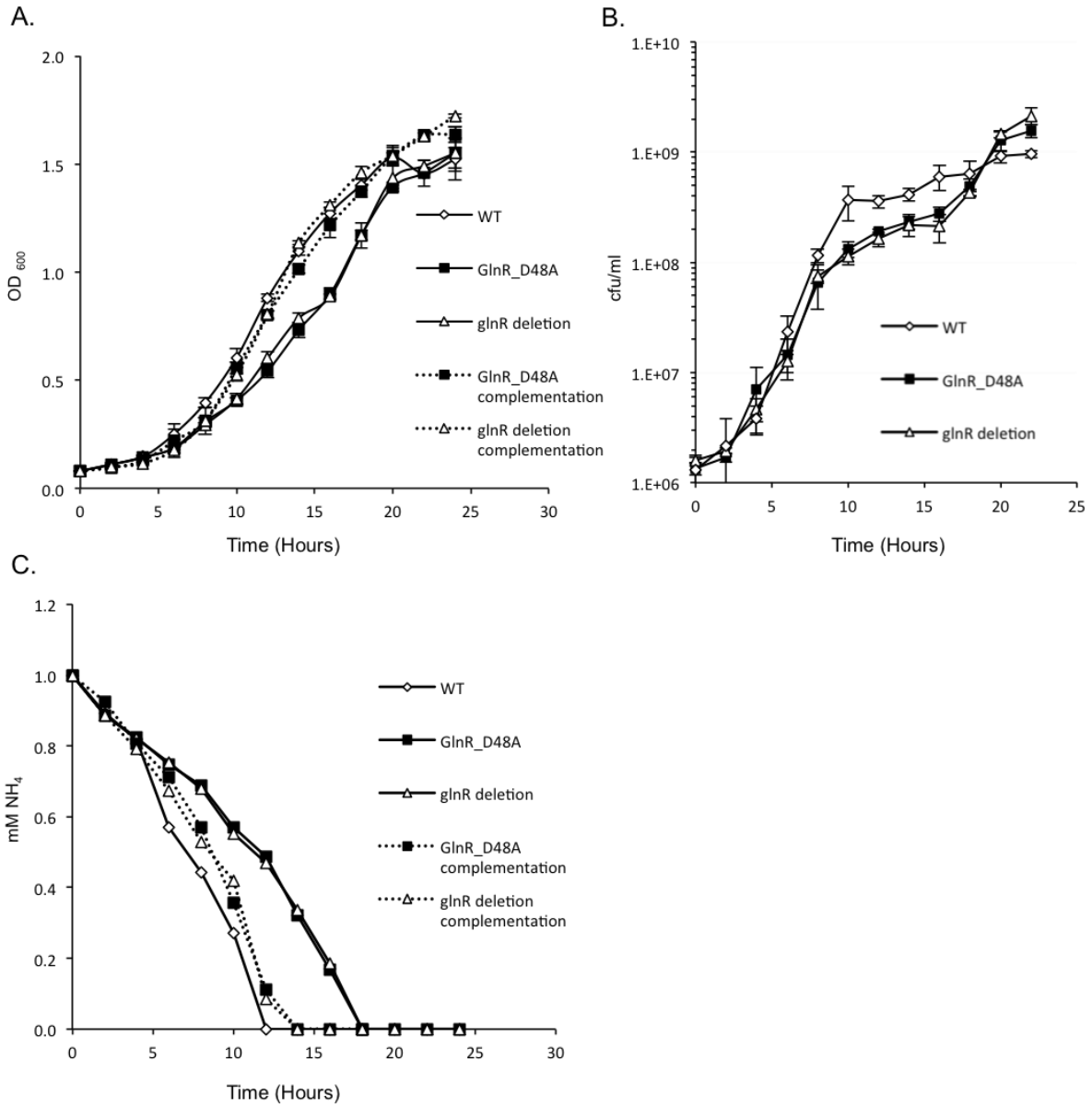


Figure 4.6. Growth analysis of wild type *M. smegmatis* and *glnR* mutants in nitrogen limiting medium (1 mM ammonium sulphate).

(A & B) Growth analysis monitored by OD_{600} and CFU/ml respectively. (C) External ammonium concentration during growth in nitrogen limitation.

M. smegmatis wild type (open diamonds), GlnR_D48A (closed squares), GlnR deletion (open triangles) strains grown in nitrogen limiting conditions. Complementation strains indicated with a hashed line. Data represents the average (±SD) of three independent experiments.

4.3.5 Transcriptomic Response to Nitrogen Limitation is Abolished in GlnR Mutants

Due to the reduction in growth rate of the GlnR_D48A strain the transcriptomic response during nitrogen limitation was investigated. *M. smegmatis* GlnR has previously been shown to control the transcription of five nitrogen metabolism genes during nitrogen limitation. As such three of these genes *glnA1* (msmeg_4290), *amtB* (msmeg_2425) and *amt1* (msmeg_6259) were chosen for initial analysis (3). Four other genes were also chosen for investigation due to their proposed role in nitrogen metabolism in mycobacteria (5). These were *amtA* (msmeg_4635), *nirB* (msmeg_0427), *gltD* (msmeg_3226), and *glnE* (msmeg_4293). Expression of *glnR* (msmeg_5784) was also analysed, to confirm that any transcriptional difference observed was a direct result of the point mutation and not reduced GlnR expression levels. Wild type and mutant strains were grown in nitrogen limiting or nitrogen excess conditions for 13 hours. Expression values for each gene analysed were compared to the housekeeping gene *sigA* (msmeg_2758) whose expression did not alter in the conditions tested (data not shown).

Genes previously shown to be under GlnR control, *glnA1*, *amtB* and *amt1,* were all highly up regulated in the wild type during nitrogen limitation when compared to their expression in nitrogen excess conditions (Figure 4.7, Table 4.1). Expression of *glnA1* was induced in wild type from nitrogen excess to nitrogen limiting conditions by approximately 13-fold, *amtB* 153-fold and *amt1* 219-fold (Table 4.1), confirming that the conditions used were stimulating a nitrogen stress response in *M. smegmatis.* However, there was no induction of these genes in both GlnR mutant strains grown under nitrogen limitation (Figure 4.7, Table 4.1). To account for the fact that the GlnR mutants deplete the external ammonium at a slower rate than the wild type (Figure 4.6), and therefore may not initiate a stress response at 13 hours, qRT-PCR was repeated using RNA samples taken at 19 hours, when external nitrogen was no longer detectable. However, there was also no induction of *glnA1*, *amtB* or *amt1* gene expression in either mutant strain at this later time point (data not shown).

To exclude the possibility that the GlnR_D48A mutation inhibited *glnR* expression, leading to the observed null phenotype of this strain, transcriptomic analysis of *glnR* levels were performed. No significant change in *glnR* expression was observed under nitrogen limiting conditions for either the wild type or GlnR_D48A mutant (Figure 4.7, Table 4.1), confirming previous observations that *M. smegmatis glnR* expression levels are not subject to transcriptional regulation during nitrogen limitation (3). As expected there was no detectable *glnR* expression in the GlnR deletion strain (Figure 4.7, Table 4.1).

Transcriptional control of other genes proposed to be involved in mycobacterial nitrogen metabolism, not shown previously to be under GlnR control, was subsequently investigated.

Fold change represents the change in gene expression from nitrogen excess to nitrogen limiting conditions, with positive values representing genes that are significantly upregulated upon nitrogen limiting conditions. Figure 4.8 and Table 4.1 displays that *amtA, gltD* and *nirB* were up-regulated in the wild type strain in response to nitrogen limitation at 13 hours, compared to nitrogen excess, while *glnE* expression was down regulated. During nitrogen limitation, *amtA* was induced approximately 337-fold, *nirB* 103-fold, and *gltD* 8-fold; *glnE* was down regulated 3-fold. Again, no significant change in the expression levels of these genes was observed in the GlnR mutants at either 13 hours (Figure 4.8, Table 4.1) or 19 hours (data not shown). Indicating that a functional GlnR is necessary to induce expression of these genes upon nitrogen limitation.

	Wild Type			**GlnR_D48A**			***glnR* Deletion**		
	Fold change[a]	SD	*P*	Fold change	SD	*P*	Fold change	SD	*P*
amt1	**218.7**	**± 34.1**	**< 0.01**	0.56	± 0.05	0.02	0.64	± 0.21	0.22
amtB	**152.9**	**± 68.2**	**< 0.01**	1.0	± 0.07	0.54	0.91	± 0.14	0.32
glnA1	**13.2**	**± 0.86**	**< 0.01**	1.0	± 0.14	0.95	0.93	± 0.26	0.76
amtA	**336.5**	**± 101.2**	**< 0.01**	1.2	± 0.3	0.29	1.3	±0.31	0.35
nirB	**102.6**	**± 30.0**	**< 0.01**	ND[b]	ND	ND	ND	ND	ND
gltD	**7.5**	**± 1.2**	**< 0.01**	1.7	± 0.12	0.02	2.6	± 0.56	0.03
glnE	**-3.2**	**± 0.42**	**< 0.01**	-1.4	± 0.19	0.07	-1.4	± 0.15	0.05
glnR	2.5	± 0.05	0.28	1.6	± 0.54	0.19	ND	ND	ND

[a] Average fold change of nitrogen excess vs nitrogen limitation. Data normalised to *sigA* and represents three independent samples.
[b] Expression not detected

Table 4.1. Relative changes in gene expression in wild type and GlnR mutants during nitrogen limitation.

Positive fold change values represent an up-regulation of gene expression during nitrogen limitation (1 mM ammonium sulphate) compared to nitrogen excess (30 mM ammonium sulphate); negative fold change values indicate a down-regulation in gene expression during nitrogen limitation compared to nitrogen excess. Significant changes (P<0.01) in gene expression are shown in bold.

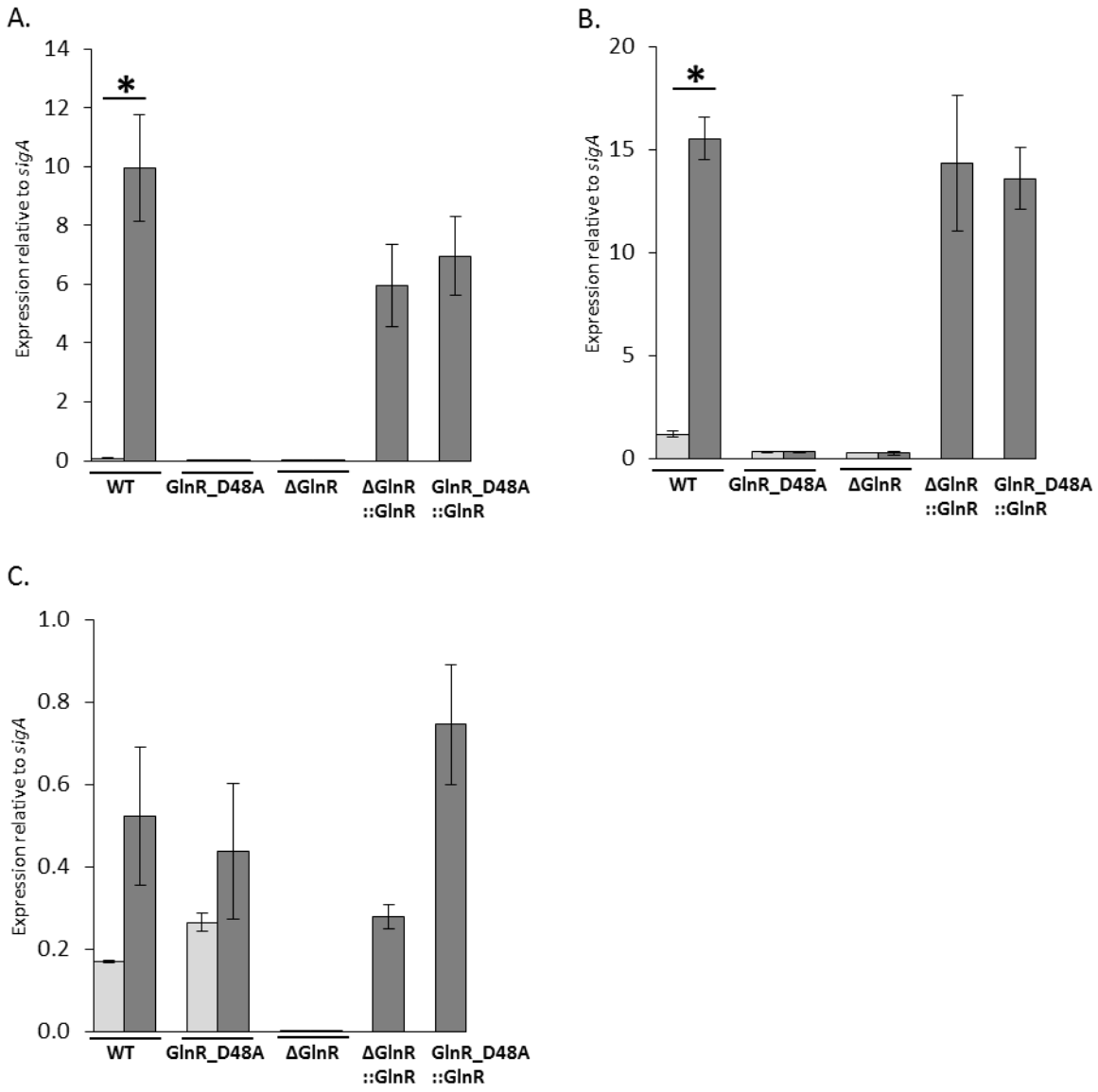


Figure 4.7. qRT-PCR analysis of (A) *amtB*, (B) *glnA1* and (C) *glnR* expression in response to nitrogen availability.

Cells were grown in nitrogen excess (30 mM ammonium sulphate) (light bars) or nitrogen limiting (1 mM ammonium sulphate) (dark bars) conditions for 13 hours. Complemented mutant strains were only examined under nitrogen limiting condition. Expression level of mRNA is given as expression normalised to *sigA*. Data presented is the mean expression value from three independent biological samples. Error bars represent standard deviation with statistically significant mRNA expression values ($P<0.01$) between nitrogen excess and limiting conditions denoted by *.

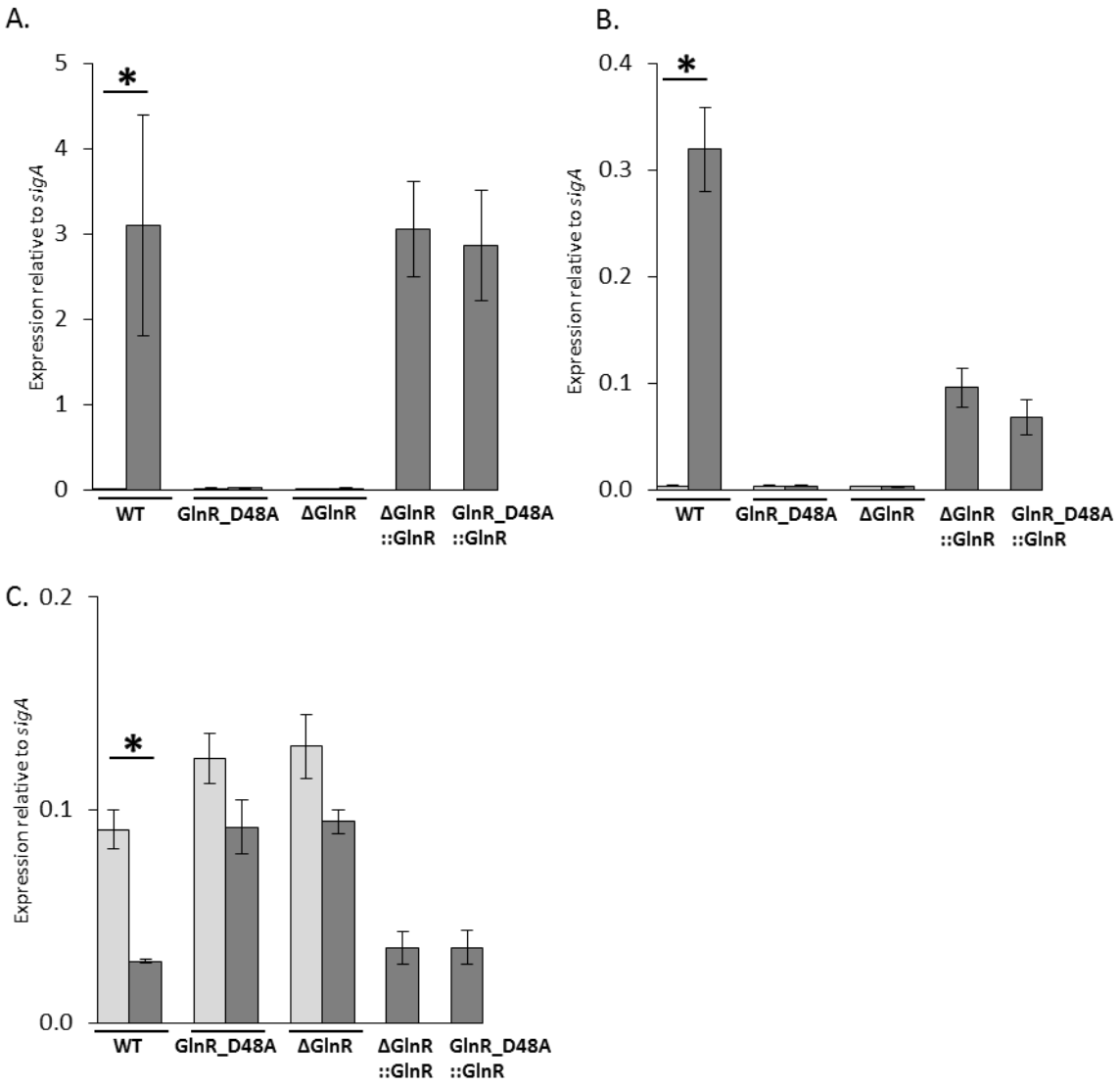


Figure 4.8. qRT-PCR analysis of (A) *amtA* (B) *nirB* and (C) *glnE*, expression in response to nitrogen availability.

Cells were grown in nitrogen excess (30 mM ammonium sulphate) (light bars) or nitrogen limiting (1 mM ammonium sulphate) (dark bars) conditions for 13 hours. Complemented mutant strains were only examined under nitrogen limiting condition. Expression level of mRNA is given as expression normalised to *sigA*. Data presented is the mean expression value from three independent biological samples. Error bars represent standard deviation with statistically significant mRNA expression values ($P<0.01$) between nitrogen excess and limiting conditions denoted by *****.

4.3.6 GlnR Mutants Fail to Grow in Nitrate as Sole Nitrogen Source

To further our analysis of the growth phenotype of the *glnR* mutants, we examined growth in potassium nitrate as the sole nitrogen source. Potassium nitrate was chosen as a nitrogen source as the *nirB* gene (msmeg_0427) was shown to be GlnR regulated in this study. NirB is a nitrite reductase responsible for the conversation of nitrite to ammonium, the second step in nitrate assimilation. It had been previously demonstrated that NirB is essential for *M. tuberculosis* growth in nitrate as a sole nitrogen source, and mutants lacking *glnR* and *nirB* were unable to assimilate nitrate (95). Consequently both mutants and wild type strains were tested for a growth phenotype in nitrate.

Cells were grown as previously described in 10 mM potassium nitrate and monitored via OD_{600}. The phenotype in 10 mM potassium nitrate was far more striking than that seen in ammonium sulphate; both mutants failed to grow in the medium when compared to the wild type strain (Figure 4.9). Restoration of growth could be achieved via the complementation of the mutants with *glnR*. Suggesting a functional GlnR is required for growth with nitrate as the sole nitrogen source.

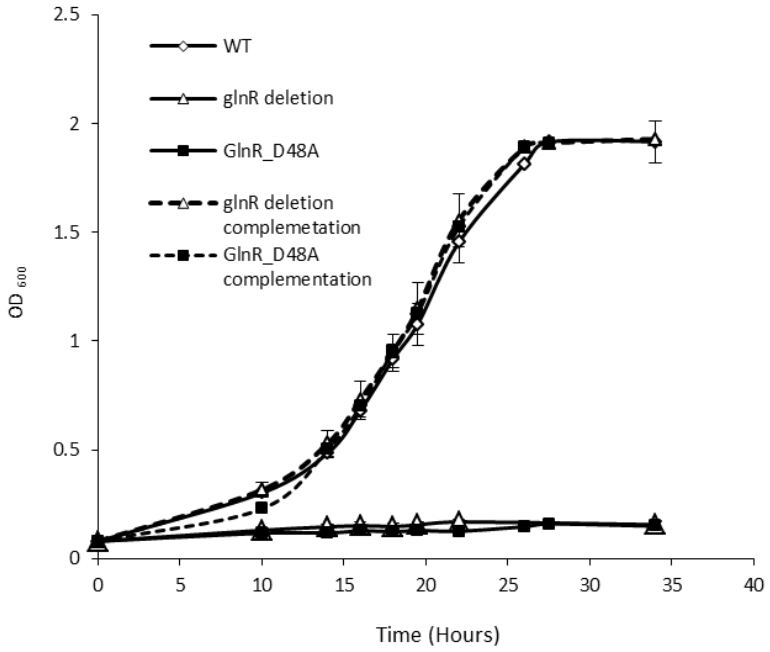


Figure 4.9. Growth analysis of wild type *M. smegmatis* and *glnR* mutants in potassium nitrate.

Growth monitored by OD_{600} of *M. smegmatis* wild type (open diamonds), GlnR_D48A (closed squares), GlnR deletion (open triangles) strains grown in nitrate as sole nitrogen source. Complementation strains indicated with a hashed line. Data represents the average (±SD) of three independent experiments.

4.4 Discussion

GlnR belongs to the OmpR family of two-component transcriptional regulators. OmpR-type regulators are typically regulated in a phosphorylation dependant manner at a conserved aspartate residue. Although GlnR is an orphan regulator, lacking a corresponding histidine kinase adjacent to its chromosomal location, GlnR still contains a conserved phosphorylation site (3). As such, the presence of the conserved phosphorylation site in GlnR (Asp_48) and the lack of regulation of *glnR* at the transcriptional level during nitrogen limitation, prompted the investigation into the role of aspartate 48 in mediating the transcriptomic response to nitrogen limitation (3).

Conventional methods to detect phosphorylation are problematic when investigating phosphorylation of aspartate residues. The labile nature of the phospho-aspartate bond results in the loss of the phosphate during sample processing for applications such as mass spectrometry (9). Previous methods to identify phosphorylation have used *in vitro* phosphorylation assays with the corresponding kinase (98); however as the kinase is unknown for GlnR these assays could not be preformed. Other attempts to establish the phosphorylation state of GlnR were attempted with PhosTag and IEF protein separation, however these were unsuccessful (Kerstin Williams unpublished data). As such the role of the Asp-48 residue was explored by investigating the function of GlnR_D48A in nitrogen limitation.

To investigate the role of the conserved aspartate (D48) for GlnR in *M. smegmatis,* an *in vivo* point mutation was created. Based on structural similarity, the most conservative amino acid substitution for an aspartate residue would be asparagine. However, as reported by Wolanin *et al*., (2003) asparagine can spontaneously deaminate, regenerating an aspartate residue (196). Consequently, several investigations with related response regulators have used the substitution of an alanine for an aspartate residue, creating an inactive protein. Notable examples include an *E. coli* PhoB_D53A mutant which was unable to undergo phosphorylation *in vitro* (205), while chemotaxis studies with a CheY_D57A mutation rendered an inactive phenotype; no *in vivo* phosphorylation was observed and the flagella machinery was not activated (41). As such, the aspartate 48 to alanine chromosomal substitution was made in *M. smegmatis* by a recombineering method described by van Kessel (174-176).

In addition a *glnR* deletion strain was constructed, where the *glnR* gene was replace with a *hygR* cassette. This provided a negative control to compare any affect observed for the GlnR_D48A mutant; previously a *glnR* deletion mutant had failed to up regulate genes implicated in nitrogen metabolism (3). Therefore, a hygromycin marked deletion strain was generated to compare the

effect of the point mutation on GlnR activity, with respect to transcription activation and growth on nitrate.

Initially the growth of the *glnR* deletion and GlnR_D48A strains were compared to wild type when grown in nitrogen limiting and excess conditions. The *M. smegmatis* strains grew similarly in nitrogen excess conditions (Figure 4.5) while the *glnR* and GlnR_D48A mutants exhibited a reduced growth rate when compared to the wild type under nitrogen limiting conditions (Figure 4.6). Despite this, no major growth defect was noted for either mutant strain; this is intriguing, suggesting that the *M. smegmatis* GlnR-mediated transcriptomic response is not essential for growth during nitrogen limitation in ammonium. Another interesting observation was the reduced uptake of ammonium from the medium by both mutants. Two ammonium transporters (AmtB and Amt1) are regulated by GlnR during nitrogen limitation (3). The inability of the GlnR mutant strains to induce expression of ammonium transporters could lead to a reduction in ammonium uptake in these mutants and reduced growth rate. As such the transcriptomic effect of the GlnR_D48A mutant was probed further, to analyse if the point mutation was responsible for the growth phenotype observed.

Previous studies in *M. smegmatis* have implicated GlnR in the expression of five nitrogen metabolism genes during nitrogen limitation. These are a glutamine synthetase (GS) enzyme *glnA1* (msmeg_4290), responsible for glutamine synthesis during nitrogen limitation, and the ammonium transporters *amtB* (msmeg_2425) and *amt1* (msmeg_6259) (3). In addition two further genes, *glnD* and *glnK*, contained in an operon with *amtB,* have been demonstrated to be GlnR regulated (3). As such, three known GlnR regulated genes, *glnA1, amtB* and *amt1* were selected for analysis in this study. Four other genes were also chosen for investigation due to their proposed role in nitrogen metabolism in mycobacteria (5). These were *amtA* (msmeg_4635) an ammonium transporter*, nirB* (msmeg_0427) a nitrite reductase enzyme, *gltD* (msmeg_3226) a glutamate synthase enzyme involved in glutamate synthesis during nitrogen limitation, and *glnE* (msmeg_4293) a bi-functional adenylyl-transferase thought to modulate GS enzymatic activity in response to nitrogen availability. Expression levels of *glnR* (msmeg_5784) were also monitored.

Genes previously shown to be under GlnR control, *glnA1*, *amtB* and *amt1,* were all highly up regulated in the wild type during nitrogen limitation when compared to their expression levels in nitrogen excess conditions (Figure 4.7). This provided additional confirmation that a nitrogen stress response was being induced in our optimised conditions. However, no induction of these genes was noted when either *glnR* mutant was grown under nitrogen limitation (Figure 4.7, Table 4.1), suggesting that GlnR controls the transcriptional response of these genes and the

D48 residue is important for GlnR function with respect to the transcriptional response to nitrogen limitation. Transcriptional regulation of other genes proposed to be involved in nitrogen metabolism, not shown previously to be under GlnR control, was also investigated. The data displayed that *amtA, gltD* and *nirB* were up-regulated in the wild type strain in response to nitrogen limitation at 13 hours, compared to nitrogen excess, while *glnE* expression was down regulated (Table 4.1). Again, no significant change in the expression levels of these genes was observed during nitrogen limitation in either of the GlnR mutants (Figure 4.8, Table 4.1). To exclude the possibility that the GlnR_D48A mutation inhibited *glnR* expression, leading to the observed null phenotype of this strain, transcriptomic analysis of *glnR* was performed. No significant change in *glnR* expression was observed under nitrogen limiting conditions between the wild type and the GlnR_D48A mutant (Figure 4.7, Table 4.1), confirming previous observations that *M. smegmatis glnR* expression levels are not subject to transcriptional regulation during nitrogen limitation (3). This transcriptomic data indicates that the D48A residue is essential for GlnR mediated transcriptomic response to nitrogen limitation.

Finally growth of the two *glnR* mutants was assessed in potassium nitrate. In this study the nitrate reductase *nirB* gene was demonstrated to be GlnR regulated; both mutant strains failed to up regulate *nirB* expression, when compare to the wild type, under nitrogen limiting conditions. Failure to up regulate the *nirB* gene would prevent conversion of nitrate into ammonium and, as such, the GS or GDH enzymes would not assimilate the nitrogen source. As predicted both mutant strains failed to grow in potassium nitrate, while the wild type strain grew as expected (Figure 4.9). The growth phenotype could be restored by reintroduction of a functional *glnR* gene. Consequently, it can be concluded that GlnR is required for nitrate assimilation in *M. smegmatis*.

In summary, this study demonstrates that the proposed phosphorylation site of GlnR (D48) is essential for the GlnR-mediated transcriptional response to nitrogen limitation in mycobacteria. It has also been shown that GlnR mediates the transcriptional response of at least 9 genes; *amtB-glnK-glnD*, *glnA1*, *amt1* demonstrated by (3) and confirmed in this study, *amtA*, *nirB*, *gltD* and *glnE*. Genes in operons include *nirB-nirD* and *gltD-gltB* extending the total number of expected GlnR regulated genes to 11. The GlnR transcriptional response is not essential for growth during nitrogen limitation in 1 mM ammonium sulphate, however growth is reduced when compare to wild type. GlnR and the GlnR residue D48 are essential however for growth in nitrate as the soul nitrogen source. The effect of the mutated GlnR D48 residue indicates that this residue is important for GlnR function, suggesting that GlnR activation occurs at this residue. However, as GlnR activation was not directly demonstrated the effects of the GlnR_D48A mutation on protein structure, structural changes in the phosphorylation pocket

and effects on possible dimer formation cannot be dismissed. In other OmpR family regulators phosphorylation at this conserved aspartate residue leads to dimerization, resulting in a stable DNA-Protein complex, thus leading to transcriptional activation. The lack of GlnR mediated transcriptional response in the GlnR_D48A mutant suggests that this form of activation may occur in mycobacteria.

CHAPTER 5: Optimisation and Validation of Mycobacterial ChIP-seq Conditions using *Mycobacterium smegmatis*

5.1 Aim

To develop a GlnR-specific antibody and an optimised protocol to permit the genome-wide identification of GlnR-binding sites in *M. smegmatis* by ChIP-seq.

5.2 Introduction

DNA-binding proteins play a crucial role in many major cellular processes, such as the regulation of gene transcription. Therefore, identification of the genomic location of these proteins, and the specific DNA sequences to which the proteins preferentially bind, is of particular interest to understand the mechanism of global transcriptional response to environmental stress conditions. The genomic locations of bound transcription factors, such as GlnR, have been predicted using *in silico* DNA sequence analysis (3). However, additional functional assays are necessary to identify and confirm these bio-informatically predicted protein:DNA interactions. As such, chromatin immunoprecipitation coupled with short-tag sequencing (ChIP–seq) has become a standard technique to identify the genomic location of DNA-binding proteins (49, 114).

The basic steps of the ChIP–seq assay have been reviewed in (114) and are depicted in Figure 5.1. Given the diversity of cell types, conditions and DNA binding proteins assayed it is impossible to comprehensively define common guidelines that are appropriate for all studies. As such methods for each step of the ChIP-seq analysis should be validated for each organism and DNA binding protein investigated (49).

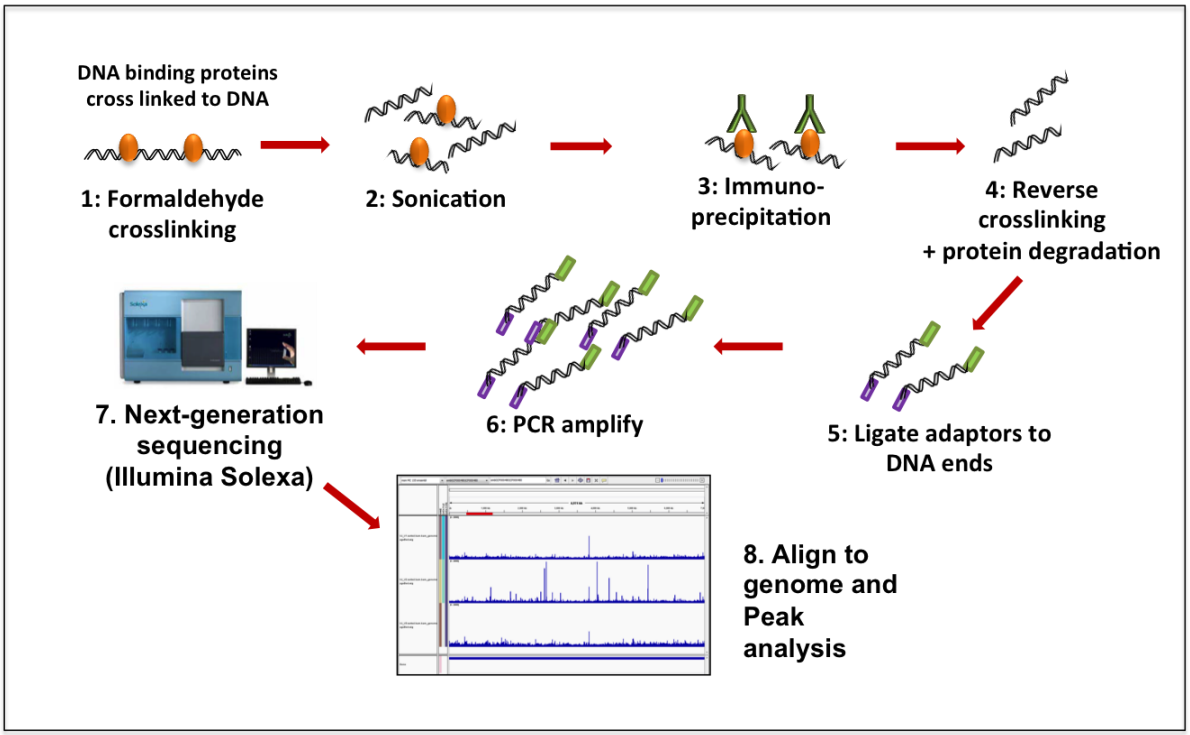


Figure 5.1. Diagrammatic representation of the steps involved in ChIP-seq analysis.

1: At a pre-determined time point formaldehyde is added to the bacterial cultures to reversibly cross-link all proteins bound to DNA. 2: Cells are lysed and DNA is fragmented by sonication. 3: The protein bound DNA is immunoprecipitated using an antibody against the protein of interest. 4: Cross-linking is reversed and proteins are degraded. 5-7 represents steps from the Illumina ChIP-seq library preparation kit. 5: DNA fragments are ligated to specific adaptor sequences. 6: PCR amplification of DNA using primers specific to the adaptor sequences. 7: Amplified DNA is sequenced on the Illumina Solexa platform using next-generation sequencing. 8: Alignment of DNA sequences to the genome of interest, permitting identification of DNA regions enriched (peaks) by the immunoprecipitation.

5.3 Results

5.3.1 Sonication

Sonication conditions for *M. smegmatis* cells cross-linked with formaldehyde were optimised to ensure sufficient cell lysis and DNA fragmentation to the optimum size of 200-400 bp. Bacteria were grown in nitrogen limiting and excess conditions, and cross-linked with formaldehyde as described (Section 2.7.1). The cells were then stored at -80°C overnight before sonication. A range of sonication times and amplitudes were tested, Figure 5.2 A and B displays DNA samples after 5 minutes of sonication (30 sec on 30 sec off, 100% amplitude). Incomplete fragmentation is observed with DNA fragments larger than 400 bp in both samples (Figure 5.2 A, B). Increasing the sonication time to 10 minutes (30 sec on 30 sec off, 100% amplitude) resulted in complete fragmentation of the *M. smegmatis* genomic DNA (Figure 5.2 C).

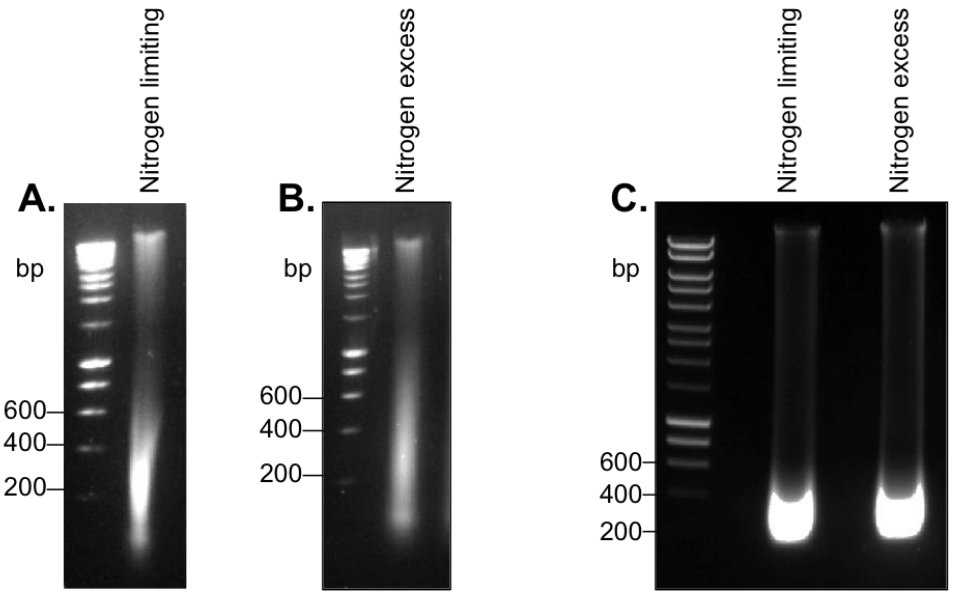


Figure 5.2. Sonication of *M. smegmatis* cells for ChIP-seq analysis.

M. smegmatis cells were grown in nitrogen limiting (1 mM ammonium sulphate) or nitrogen excess (30 mM ammonium sulphate) for 13 hours, before the addition of formaldehyde. Cells were harvested by centrifugation and subject to sonication, either 5 minutes in total (30 sec on 30 sec off, 100% amplitude, panel A and B), or 10 minutes in total (30 sec on 30 sec off, 100% amplitude, panel C).

5.3.2 GlnR Polyclonal Antibody Production

Specific antibodies are required for the immunoprecipitation step of ChIP (Figure 5.1), and *M. tuberculosis* GlnR protein (Rv0818) was chosen as the antigen for this purpose. *M. tuberculosis* and *M. smegmatis* GlnR proteins share high amino acid identity (73%), therefore it was predicted that the antibody generated against Rv0818 would also recognise the *M. smegmatis* GlnR protein, and be suitable for ChIP-seq studies in both organisms (5).

Cloning of glnR into pET28b+

The GlnR coding sequence was cloned into the pET28b+ vector to create a fusion with a hexahistidine tag (His-tag) at the N-terminus, allowing purification by nickel affinity chromatography. Located between the N-terminal His-tag and the coding sequence of the desired gene is a thrombin cleavage site, allowing removal of the His-tag if required. The vector also contained a ribosome-binding site for the T7 RNA polymerase, permitting expression in the BL21(DE3)pLysS strain of *E. coli* cells. The BL21(DE3)pLysS strain of *E. coli* is lysogenic for λ-DE3 which contains the bacteriophage *gene I*, encoding a T7 RNA polymerase. *Gene I* is under control of the *lacUV5* promoter, inducible with the addition of IPTG. In addition BL21(DE3)pLysS *E. coli* contain a plasmid encoding a T7 lysozyme which lowers background expression of target genes under the T7 promoter before induction with IPTG.

Rv0818 was amplified by PCR from *M. tuberculosis* genomic DNA then ligated into the linearised pET28b+ vector. PCR primers were designed to incorporate a *Nde*I restriction site at the 5' end of the gene, resulting in fusion to the N-terminal His-tag when expressed and a stop codon at the 3' end of the gene. The addition of a stop codon prevented the translation of the optional C-terminal His-tag contained within the vector. PCR amplification was carried out as described in Section 2.2.2 and specific PCR products purified, before restriction endonuclease digestion and ligation into the vector (Sections 2.2.7 and 2.2.9). Sequencing of the inserted DNA fragment into pET28b+ was carried out as described, to confirm 100% identity to the native gene and that protein translation was in frame with the His tag.

GlnR Protein Expression and Purification

BL21(DE3)pLysS *E. coli* cells were used to express the His-GlnR protein, and transformed and grown as described in Section 2.2.10 and 2.3.1. His-GlnR expression in *E. coli* appeared insoluble when cultures were grown at 37°C (Data not shown). Solubility of the expressed protein was improved by reducing the temperature after induction with 1 mM IPTG to 20°C, when cultures reached mid-log phase (OD_{600}~0.45).

Recombinant His-GlnR was purified from bacterial lysates by capture on a nickel affinity column, using FPLC to wash and elute His-tagged proteins. Initially the eluted extracts contained a high proportion of a higher molecular weight contaminant (Figure 5.3). In order to fully saturate the column with the His-GlnR the amount of lysate loaded was doubled. The column wash was also increased from 3% to 9% buffer B, increasing the imidazole concentration of the wash buffer to remove non-specific binding, before elution with a gradient of imidazole. Eluted fractions were measured at OD_{280} and readings displayed on a chromatogram (Figure 5.4 A). Fractions were separated on SDS PAGE gel to check the purity of the protein product (Figures 5.4 B), then selected fractions were pooled and dialysed into a storage buffer containing 20% glycerol before storage at -20°C (Section 2.2.4). Protein concentration was determined by BCA analysis described in Section 2.4.1. Ten µl of the protein was separated on an SDS PAGE gel to check purity and then a Western blot probed with an anti-His antibody, to confirm the product was the desired recombinant His-tagged protein (Figure 5.5). The purified His-GlnR protein was subsequently sent for antibody production at Eurogentec, Germany.

The *M. smegmatis* GlnR protein was purified and used to test the specific reactivity of the polyclonal antibody. This was carried out using the same cloning procedure described for *M. tuberculosis* and the same protein expression and purification conditions. Figure 5.6 displays the fractions from the FPLC elution and the final pooled GlnR product.

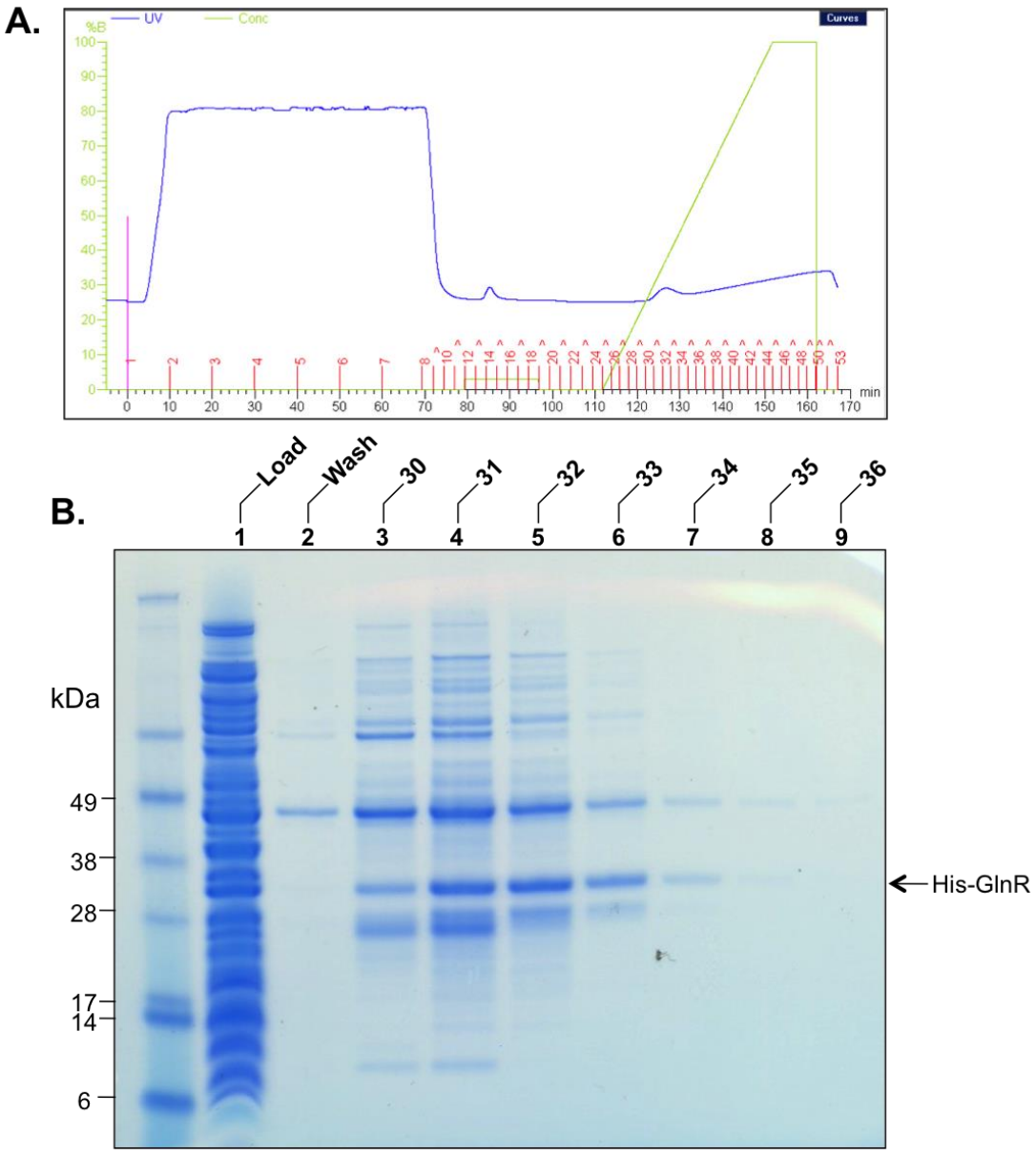


Figure 5.3. High molecular weight contaminants were observed in the initial purification of *M. tuberculosis* His-GlnR protein.

(A) Chromatogram of His-GlnR purification on a FPLC. Absorbance (blue line) is read at 280 nm, the green line represents percentage of Buffer B, this is indicative of imidazole concentration. (B) SDS PAGE gel of fractions taken from the protein purification displayed in the chromatogram. Lane 1 displays 10 µl of the soluble fraction of the cell lysate before purification. Lane 2 contains the wash fraction on the FPLC at 3% buffer B (fraction 15). Lanes 3-9 show 10 µl samples of the eluted fractions 30-36 selected based on absorbance at 280nm on the FPLC.

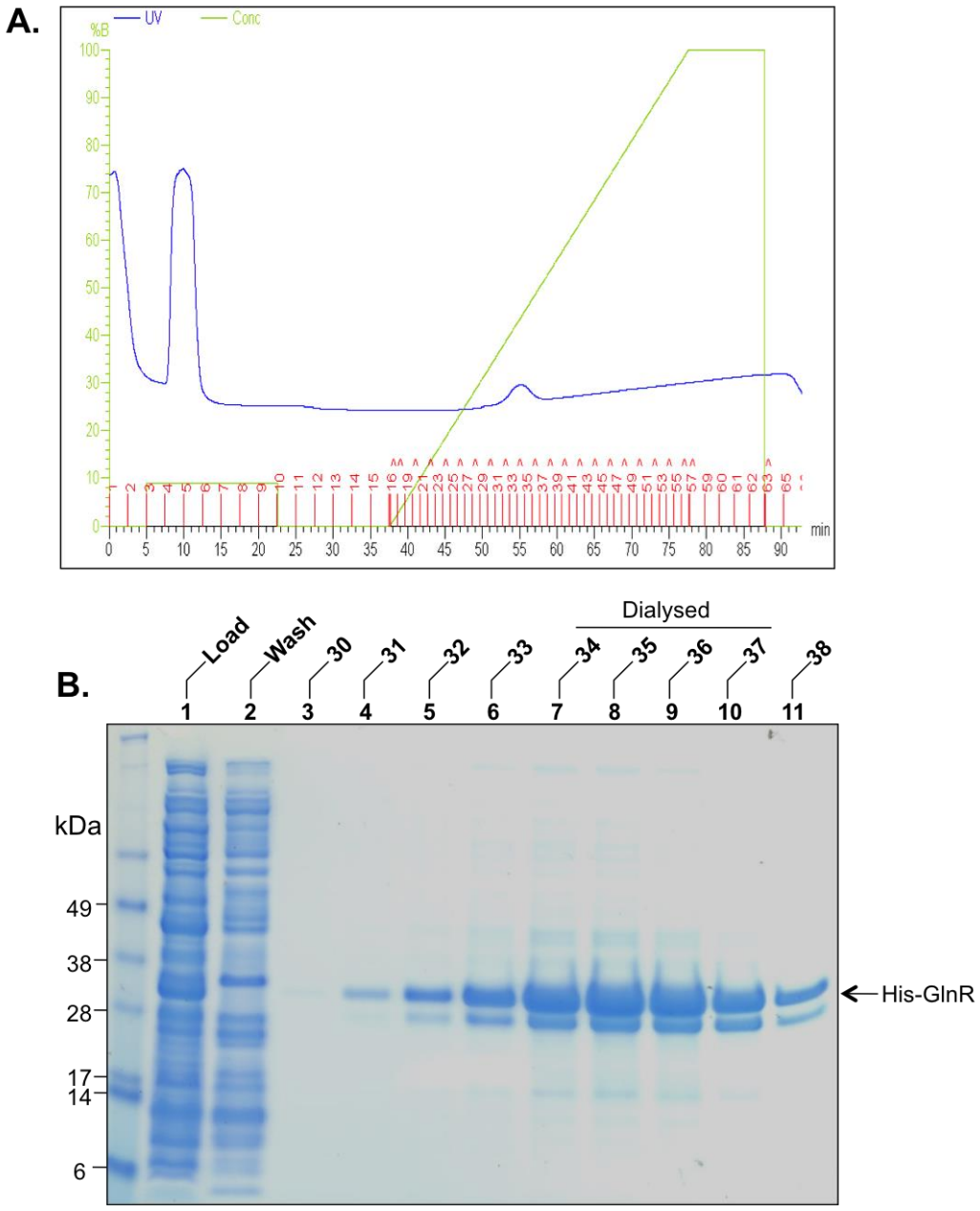


Figure 5.4. *M. tuberculosis* His-GlnR protein purification with increased imidazole wash step.

(A) Chromatogram of His-GlnR purification by FPLC. Absorbance (blue line) is read at 280 nm, the green line represents percentage of Buffer B, this is indicative of imidazole concentration. (B) SDS PAGE gel of fractions taken from the protein purification displayed in the chromatogram. Lane 1 displays 10 µl of the soluble fraction of the cell lysate before purification. Lane 2 exhibits the wash fraction on the FPLC at 9% buffer B (fraction 5). Lanes 3-11 show 10 µl samples of the eluted fractions 30-38 analysed selected based on absorbance at 280nm. Fractions 34-37 were pooled and dialysed for future use (Figure 5.5).

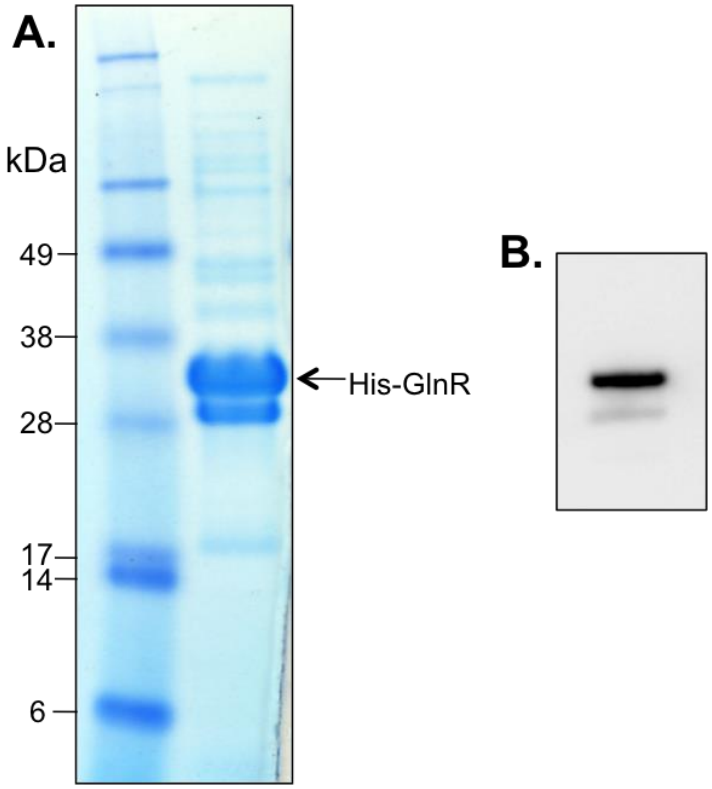


Figure 5.5. Confirmation of *M. tuberculosis* His-GlnR protein purification.

(A) SDS PAGE displaying 10 µl sample of the purified *M. tuberculosis* His-GlnR protein from pooled and dialysed fractions 34-37 in Figure 5.4. (B) Western blot of the purified *M tuberculosis* His-GlnR protein (0.54 µg) reacted with an anti-His tag antibody, and visualised by chemiluminescence.

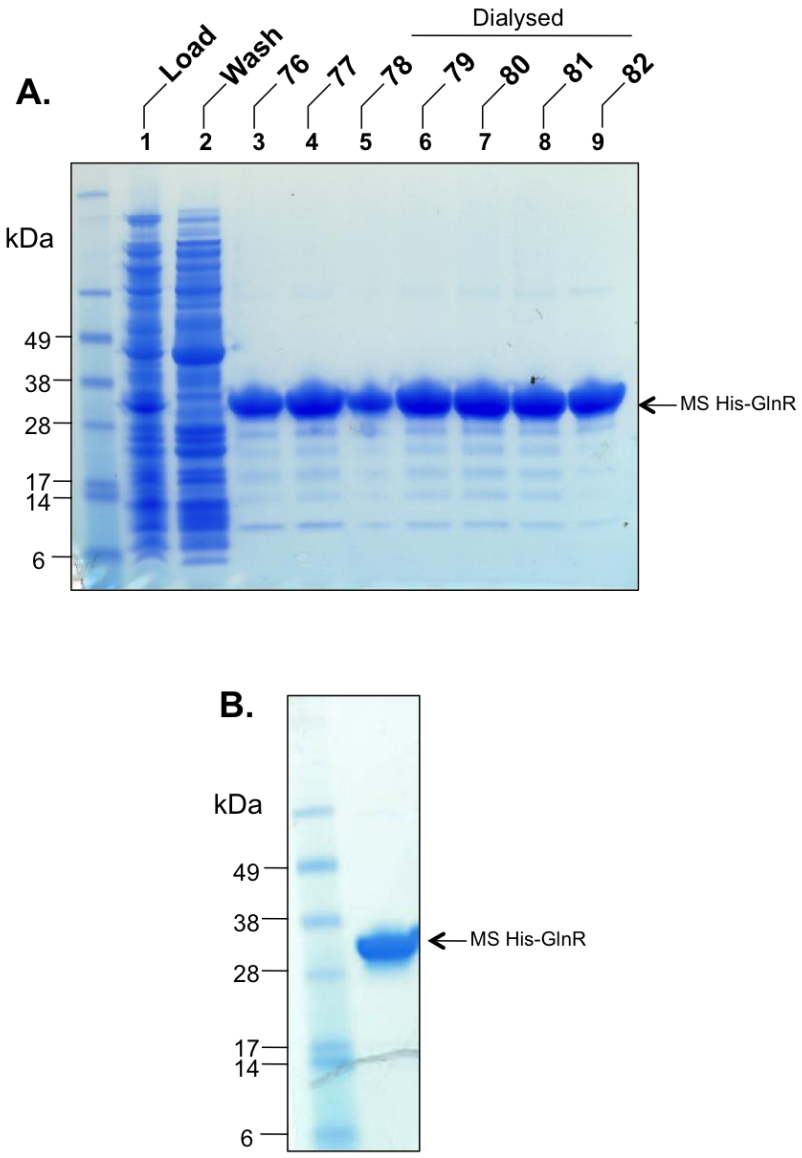


Figure 5.6. *M. smegmatis* His-GlnR protein purification.

(A) SDS PAGE gel of fractions taken from the protein purification. Lane 1 displays 10 µl of the soluble fraction of the cell lysate before purification. Lane 2 exhibits the wash fraction on the FPLC at 9% buffer B. Lanes 3-9 show 10 µl samples of the eluted fractions 76-82 analysed selected on the basis of absorbance at 280nm. Fractions 78-82 were pooled and dialysed for future use. (B) SDS PAGE displaying 10 µl sample of the purified *M. smegmatis* His-GlnR protein from pooled and dialysed fractions 78-82.

GlnR Antibody Specificity

Western blot analysis confirmed the specific reactivity of the antibodies produced against the recombinant proteins *in vitro* and *in vivo*. Purified recombinant *M. smegmatis* His-GlnR (250 ng) and *M. smegmatis* cell lysates (20 µg) from wild type and a *glnR* deletion strain (Chapter 4), were tested. The polyclonal GlnR antibody recognised the *M. smegmatis* protein specifically, but cross reactivity was seen in both cell lysates (Figure 5.7 A). As such the polyclonal antibody was affinity purified against the *M. smegmatis* His-GlnR protein as described (Section 2.4.4). Ranges of antibody concentrations were tested to give the optimum detection signal for GlnR in the cell lysates; a 1 in 50 dilution was subsequently used (Data not shown). The purified polyclonal GlnR antibody specifically detected the His-GlnR purified protein and the GlnR protein in wild type *M. smegmatis* cell lysate, represented by a band at 28 kDa that was absent in the *glnR* deletion strain (Figure 5.7 B). Further analysis was conducted on the *M. smegmatis* cell lysates using the same conditions to be used in ChIP-seq analysis. In both nitrogen limiting and excess conditions a single band representing GlnR was present, which was absent in the *glnR* deletion strains (Figure 5.7 C). Only single bands were obtained suggesting that the anti-GlnR reaction is specific, with very little background cross-reactivity, confirming the suitability of this purified GlnR antibody for ChIP-seq analysis.

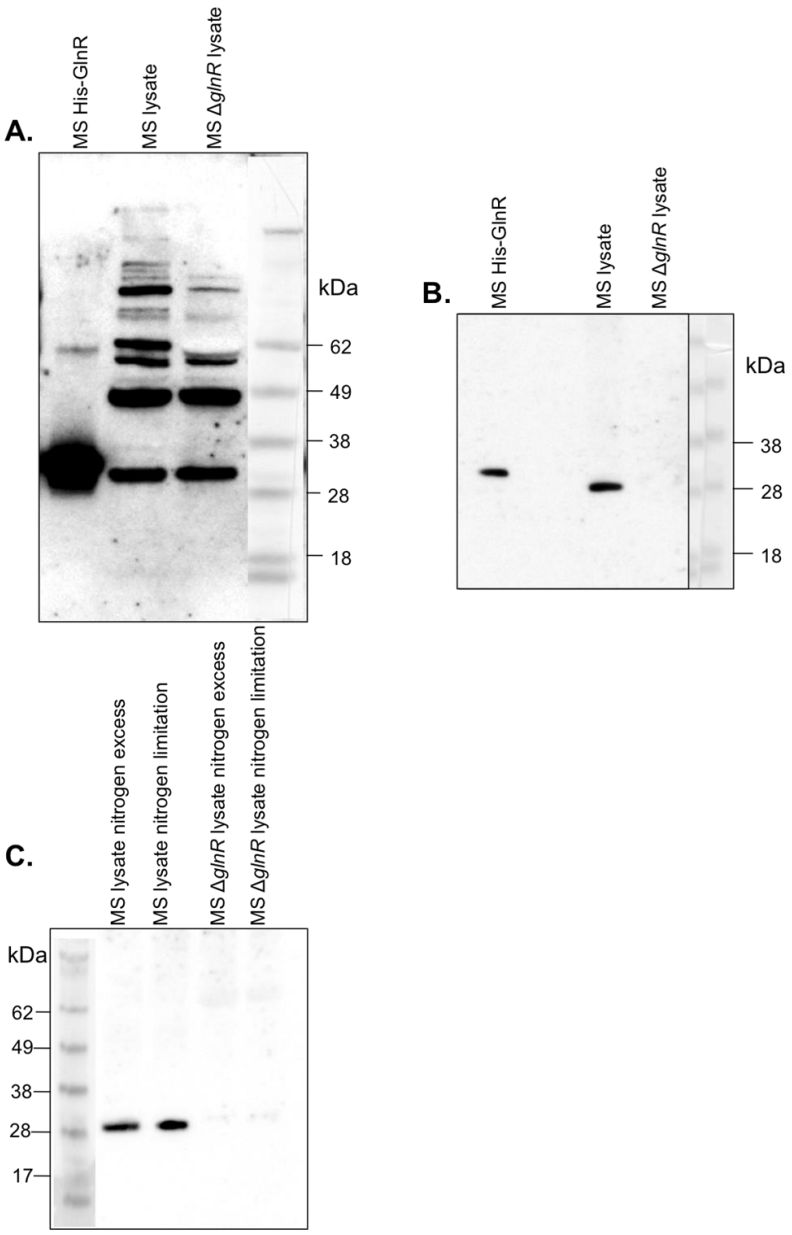


Figure 5.7. Western blot analysis of *M. smegmatis* cell lysates with the polyclonal anti-GlnR antibody.

(A) Western blot analysis with the polyclonal anti-GlnR antibody 1:1000 dilution (not purified) against *M. smegmatis* His-GlnR (250 ng) and *M. smegmatis* wild type and *glnR* deletion lysates (20 µg). (B) Western blot analysis using the purified anti-GlnR antibody at 1:50 dilution against *M. smegmatis* His-GlnR (12.5 ng) and *M. smegmatis* wild type and *glnR* deletion lysates (20 µg). (C) Western blot analysis using the purified anti-GlnR antibody at 1:50 dilution of *M. smegmatis* cell lysates under the conditions used for ChIP-seq. Cell lysates (20 µg) were taken at 13 hours.

5.3.3 Immunoprecipitation of *M. smegmatis* Cross-linked and Sonicated DNA with anti-GlnR Antibody

The immunoprecipitation step of *M. smegmatis* fragmented DNA with the purified GlnR antibody was optimised. Samples from nitrogen limiting cultures at 13 hours were processed for ChIP-seq up to the point of immunoprecipitation (Section 2.7.1). Samples were then divided and immunoprecipitation proceeded with a range of volumes of the neat, purified anti-GlnR antibody. A rate limiting PCR was conducted on the *glnA1* promoter region in order to determine which concentration of antibody precipitated the most fragmented DNA. GlnR had been previously shown to bind the *glnA1* promoter region via EMSA analysis (3). The purpose of the rate limiting PCR was to determine which concentration of antibody precipitated the most fragmented DNA without overloading the beads used for precipitation of the complex. Overloading of the beads would result in saturation of the beads with antibody only. Figure 5.8 displays amplification of DNA precipitated with 20 µl and 100 µl of neat purified antibody, with 100 µl clearly showing greater enrichment of the DNA fragment. Since the band intensity from the 100 µl sample and the input control (10 ng of total sonicated DNA) were similar, this volume of antibody was chosen for ChIP-seq analysis.

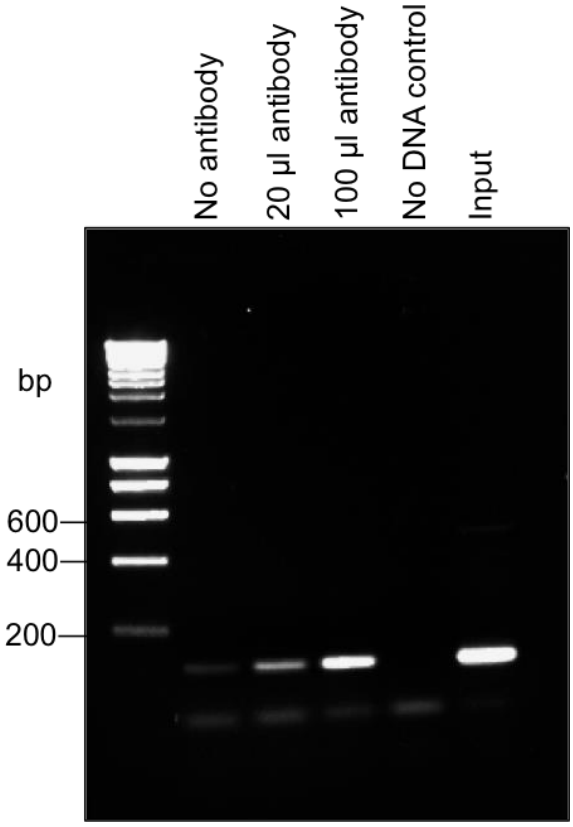


Figure 5.8. Rate-limiting PCR of *glnA1* promoter region from DNA precipitated with a range of anti-GlnR volumes.

Rate limiting PCR involved 23 cycles amplification of 180 bp promoter region of *glnA1*, using 0.3 ng of template DNA. No antibody represents a sample which did not have the GlnR antibody present for immunoprecipitation of DNA, 10 µl and 100 µl indicates the volume of neat purified antibody added to the *M. smegmatis* cell lysate. Input represents 10 ng of total sonicated *M. smegmatis* DNA before immunoprecipitation.

5.3.4 Illumina Next Generation ChIP-seq Library Preparation

The GlnR ChIP DNA library was prepared for sequencing on the Illumina Solexa platform as per the manufacturer's instructions, with the addition of a second gel extraction step (Section 2.7.3). Briefly, 10 ng of ChIP DNA or Input sample was processed; the DNA overhangs were converted into phosphorylated blunt ends, before addition of A bases to the 3' end of the DNA fragments. The 3' A overhangs were subsequently utilised to allow annealing of the adaptor fragments, which contained a 3' T base overhang. To remove excess adaptor dimers, and to select only DNA in the region of 200-400 bp, the sample was gel extracted. The next step involved PCR amplification, using primers specific to the adaptor sequences. The DNA library was subsequently examined on a bioanalyser to confirm DNA size and purity. During this analysis it became apparent that a large volume of primer dimers remained in the sample after the PCR amplification step. Consequently an additional gel extraction step was integrated into the protocol to remove this primer dimer contaminant (Figure 5.9 A). This resulted in a DNA library that was free from contaminants with DNA in the optimum size range for sequencing (Figure 5.9 B).

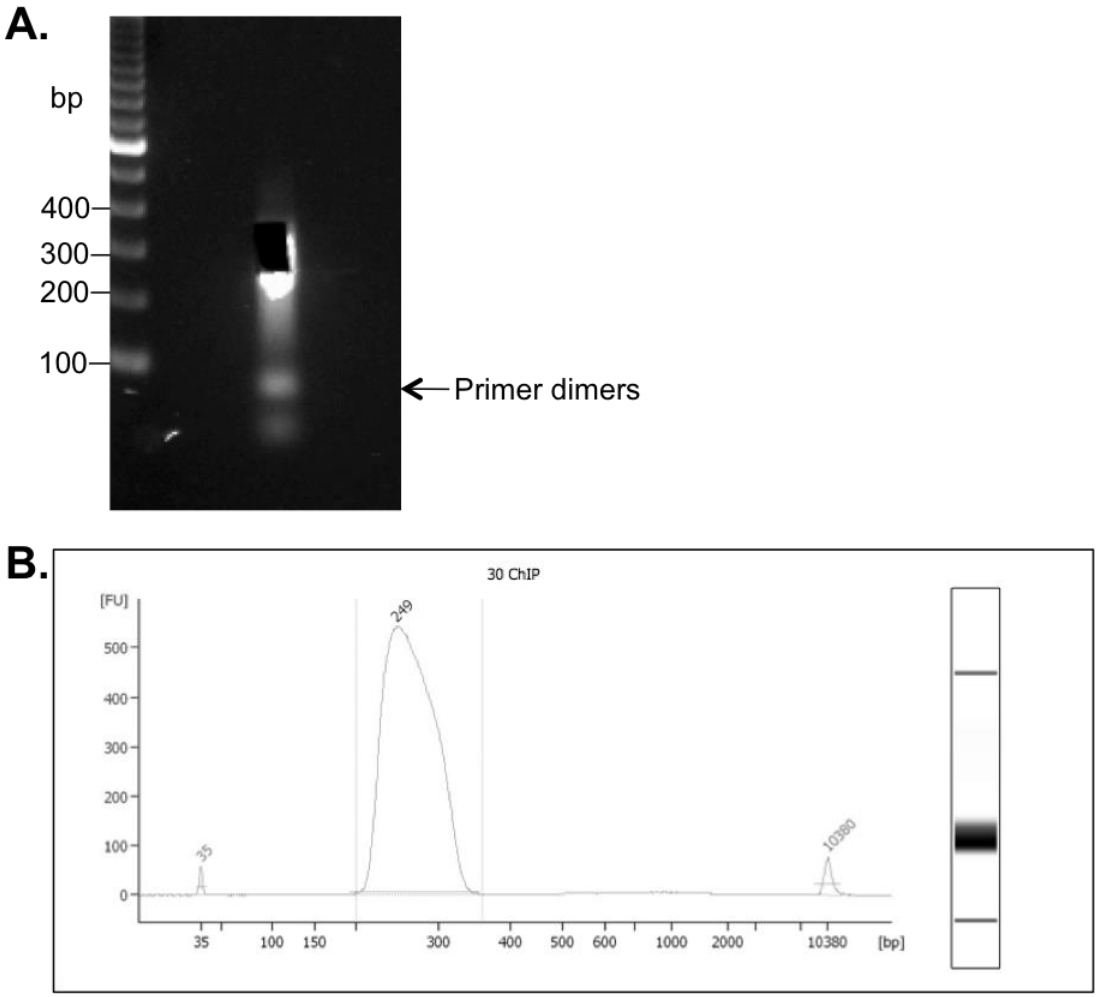


Figure 5.9. (A) Second gel extraction step to remove primer dimer contaminants and (B) the subsequent bioanalyser reading to confirm purity of DNA.

(A) A 2% agarose gel showing the gel slice extracted from the PCR amplified DNA fragments between 200-400 bp. Primer dimer contaminants are highlighted. (B) The bioanalyser reading of the purified gel extracted sample. The two peaks at 35 and 10,380 are DNA markers, the large peak between 200 and 350 bp represents the DNA library sample.

5.4 Conclusion

A validated protocol was optimised for the identification of genome-wide GlnR binding sites in *M. smegmatis*, using ChIP-seq analysis (Section 2.7). Optimisation of the sonication conditions ensured all *M. smegmatis* genomic DNA was sheared to fragments of between 200-400 bp. A polyclonal anti-GlnR antibody was generated against the *M. tuberculosis* His-GlnR recombinant protein, and after affinity purification against *M. smegmatis* His-GlnR, the antibody specifically recognised GlnR in *M. smegmatis* cell lysates. Analysis of the DNA fragments co-immunoprecipitated with GlnR gave an optimum quantity of antibody to precipitate the GlnR:DNA complexes. Finally the addition of a second gel extraction step after PCR amplification of the GlnR DNA ChIP library meant that samples were in the correct size range for next-generation sequencing and contaminant free. This optimised protocol was used in Chapter 6 to identify GlnR binding sites across the genome during nitrogen limitation.

CHAPTER 6: Genome Wide Analysis of the GlnR Regulon During Nitrogen Stress in *Mycobacterium smegmatis*

6.1 Aim

To determine the entire GlnR regulon in *M. smegmatis* during nitrogen limitation, using ChIP-seq to determine *in vivo* GlnR binding sites, and transcriptomic data analysis of wild type and *glnR* deletion strains to characterise direct and indirect GlnR regulated transcripts.

6.2 Introduction

In *Streptomyces,* GlnR acts as a global transcriptional regulator for genes encoding proteins related to nitrogen uptake, metabolism and regulation (198). In *S. coelicolor* a global proteomic analysis of a *glnR* deletion strain compared to wild type and transcriptome data showed 50 genes to be GlnR regulated in response to nitrogen limitation (164, 165). Recent analysis in *S. venezuelae* combined ChIP-CHIP with gene expression microarray data, and 44 genes were demonstrated to be GlnR regulated (120), indicating GlnR as a global transcriptional regular of genes involved in nitrogen metabolism in *Streptomyces.*

GlnR of *M. smegmatis* shares 60% identity at the amino acid level with GlnR of *Streptomyces* (3). Using the GlnR consensus binding motifs from *S. coelicolor*, putative GlnR binding sites were found in all the available mycobacterial genomes. Three highly conserved *cis* elements were found in *M. smegmatis* upstream of *glnA1* and *amt1* genes, and the *amtB-glnK-glnD* operon (3). In this study GlnR was also shown to regulate the expression of *amtA*, *glnE*, *nirB/D* (nitrite reductase) and *gltB/D* in *M. smegmatis* in response to nitrogen stress. This expanded the *M. smegmatis* GlnR regulon to 11 genes. However, given the number of nitrogen metabolism-related genes in the *M. smegmatis* genome, it is likely that many other GlnR-regulated genes exist.

As such, to gain insight into the regulatory role of GlnR during nitrogen-limitation in *M. smegmatis*, a global *in vivo* approach was applied. Chromatin Immunoprecipitation (ChIP) coupled with high-throughput sequencing (ChIP-seq), permitted identification of *in vivo* GlnR:DNA interactions. GlnR regulated transcripts were identified by combining this with WT and *glnR* mutant genome-wide expression profiles during nitrogen limitation. Using these techniques direct and indirect regulated GlnR genes could be identified.

6.3 Results

6.3.1 Global GlnR Regulated Gene Expression in Nitrogen Limitation

In order to identify all the GlnR regulated transcripts during nitrogen limitation, global expression profiles of *M. smegmatis* wild type and *glnR* deletion strains grown in nitrogen limiting conditions were analysed by microarray. *M. smegmatis* wild type and *glnR* deletion strains were harvested one hour after nitrogen run-out. Total RNA was extracted and cDNA hybridised to the *M. smegmatis* microarray as described (Section 2.6.4). Data was normalised (Section 2.6.6) and genes showing greater than 2 fold difference in expression between the wild type and *glnR* deletion strain, with an FDR corrected *p* value of <0.01, were considered significant. The microarray confirmed earlier studies; previously identified GlnR-regulated genes were controlled by GlnR under nitrogen stress (Table 6.1).

In total 392 genes were significantly up regulated and 291 genes were significantly down regulated in wild type compared to the *glnR* deletion strain under nitrogen limitation. Fully annotated microarray data is available at BμG@Sbase (accession number E-BUGS-143; http://bugs.sgul.ac.uk/E-BUGS-143) and also ArrayExpress (accession number E-BUGS-143). Intriguingly, several response regulators were up regulated only in the wild type strain, suggesting that these are GlnR-activated response regulators. This indicates that GlnR has both a direct and indirect effect on the transcriptional response of *M. smegmatis* to nitrogen stress.

Gene ID	Name	Fold change WT vs Δ*glnR* this study	Reference GlnR regulated in previous studies
msmeg_6259	*amt1*	255.9	(3)
msmeg_4635	*amtA*	102.0	(72)
msmeg_2425	*amtB*	98.8	(3)
msmeg_4290	*glnA1*	20.0	(3)
msmeg_0427	*nirB*	76.4	(72)
msmeg_4293	*glnE*	-3.1	(72)
msmeg_2426	*glnK*	29.8	(3)
msmeg_2427	*glnD*	29.1	(3)

Table 6.1. Differential gene expression between *M. smegmatis* wild type vs Δ*glnR* during nitrogen limitation (this study).

Data obtained from the microarray analysis in this study of genes previously shown to be GlnR regulated.

6.3.2 Global GlnR Binding Regions in Nitrogen Limitation

In order to determine which 681 genes identified by microarray are directly regulated by GlnR, rather than indirectly through GlnR activation of other response regulators, a ChIP-seq approach was applied. ChIP-seq enabled the identification of *in vivo* GlnR binding sites during nitrogen limitation. Cells were grown in 1 mM (limiting) or 30 mM (excess) ammonium sulphate and were cross-linked one hour after external ammonium depletion in the limiting cultures; nitrogen excess samples were cross-linked at the same time point. After cross-linking, cells were lysed and the DNA sheared by sonication. Immunoprecipitation of GlnR-bound DNA fragments was carried out using a purified anti-GlnR polyclonal antibody as described in Chapter 5.

To confirm enrichment of the GlnR binding regions during nitrogen limitation in the immunoprecipitated DNA fragments, a rate limiting PCR was performed. Two known GlnR regulated genes were chosen for analysis; the glutamine synthetase (*glnA1*) and nitrite reductase (*nirB*) promoter regions (Figure 6.1 A and B). A gene not thought to be GlnR regulated (msmeg_3224) was included as a negative control (Figure 6.1 C). The rate limiting PCR confirmed that DNA enrichment was seen in the immonoprecipitated sample for nitrogen limiting conditions, compared with the nitrogen excess sample, and no enrichment was seen for the negative control in either condition.

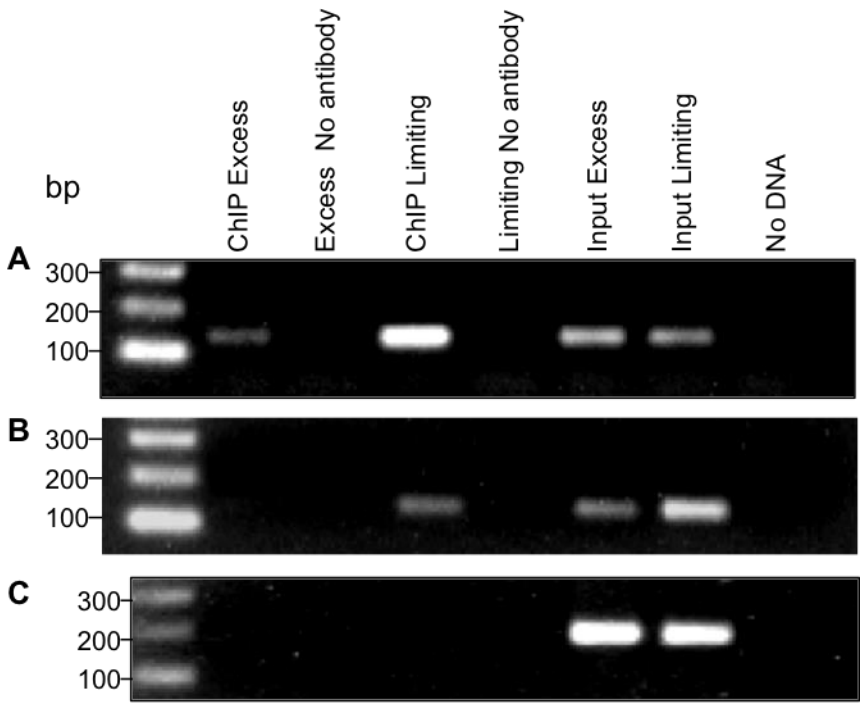


Figure 6.1. Rate limiting PCR confirmed enrichment of GlnR immunoprecipitated DNA.

(A) Promoter region of *glnA1.* (B) Promoter region of *nirB.* (C) Promoter region of msmeg_3224, negative control.

Rate limiting PCR involved 23 cycles of 200 bp promoter regions, using 0.3 ng of template DNA. ChIP DNA represents DNA that has been immunoprecipitated with a GlnR specific antibody during nitrogen excess or nitrogen limiting conditions. Input DNA represents the total amount of DNA that was subject to immunoprecipitation.

Immunoprecipitated DNA was prepared for sequencing using the Illumina ChIP-seq library kit as described in Section 2.7.3. High-throughput next generation sequencing of the DNA libraries, using the Illumina HiSeq2000, generated approximately 160 million reads per sample which were mapped to the *M. smegmatis* genome using Bowtie (81). GlnR binding regions were identified using the peak calling algorithm SISSRs (Site Identification for Short Sequence Reads) (105). GlnR binding sites were defined as regions showing greater than 5-fold enrichment in the immunoprecipitated sample compared to the input control with p value of < 0.005. This identified 53 GlnR binding sites in nitrogen limitation and 5 GlnR binding sites in nitrogen excess (Figure 6.2, Table 6.2 and Table 6.3 respectively). However all binding sites identified in nitrogen excess conditions were also observed in nitrogen limiting conditions, but with much lower peak intensity value. For example a GlnR binding site identified in both conditions upstream of *glnA1* (msmeg_4290) had a peak intensity of 6.3 in nitrogen excess and 184.7 in nitrogen limitation (Table 6.3). All GlnR binding sites (except peak number 52) were located in intergenic regions, close to promoter regions of genes.

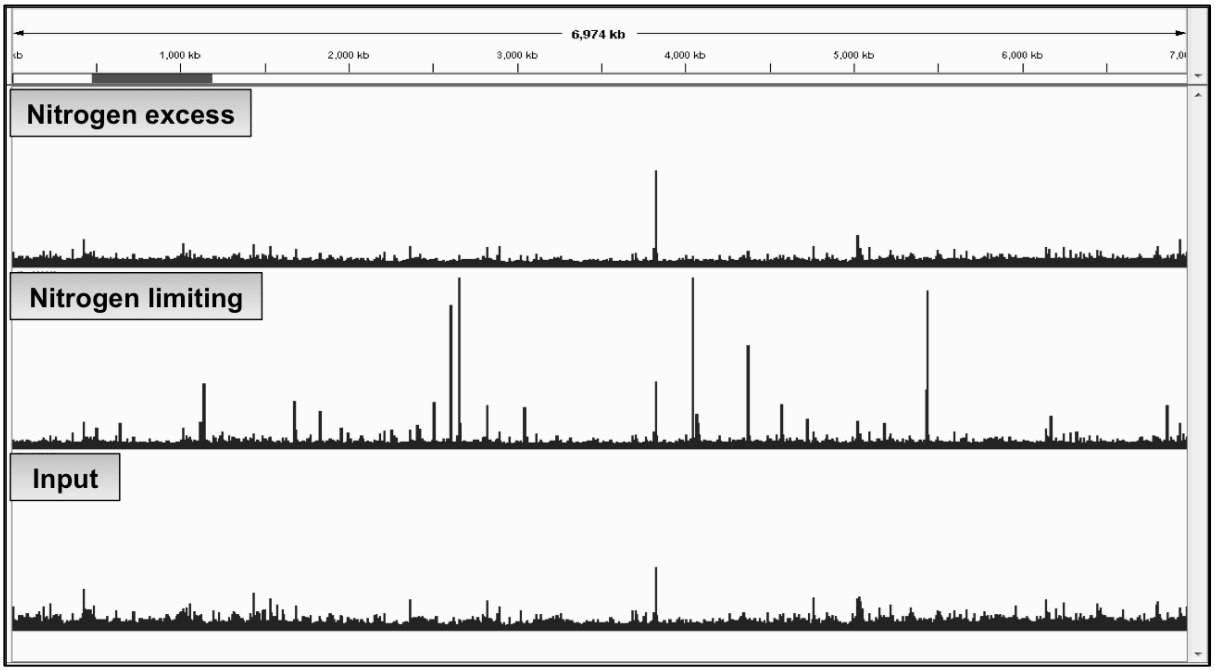


Figure 6.2. Whole genome view of GlnR binding sites identified by ChIP-seq in *M. smegmatis.*

Whole genome view of GlnR binding sites identified by ChIP-seq in *M. smegmatis* displayed in IGV. GlnR binding sites are represented by peaks indicating DNA enrichment in the ChIP samples. Upper track displays nitrogen excess (30 mM ammonium sulphate), middle track nitrogen limiting (1 mM ammonium sulphate), bottom track is the input control (total DNA without immunoprecipitation).

Peak[a]	Chr[b] start	Chr[b] end	Peak Intensity[c]	D. reg[d]	Gene Fold Change[e]	Gene ID	Gene Description
1	501431	501471	8.4		76.4	MSMEG_0427*	*nirB* Nitrite reductase, large subunit
2	508651	508691	42.9		18.3	MSMEG_0432*	*nnaR* Transcriptional regulator
3	510091	510131	8.4		24.6	MSMEG_0433	*narK3* Nitrate extrusion protein
4	647871	647911	27.1		263.4	MSMEG_0572*	Putative uncharacterised protein
5	864391	864431	6.1	5L	23.0	MSMEG_0780*	Phosphotransferase enzyme family protein
				5R	8.4	MSMEG_0781	Amino acid permease
6	1121631	1121671	54.3		6.3	MSMEG_1052	Amino acid carrier protein
7	1142851	1142891	6.5	7L	-3.8	MSMEG_1078	Hydrolase
				7R	3.4	MSMEG_1079*	Putative uncharacterised protein
8	1146711	1146751	71.9		277.4	MSMEG_1082	Putative response regulator, LuxR family
9	1238491	1238531	19.3	9L	10.7	MSMEG_1177	Cytosine/purines/uracil/thiamine/allantoin permease
				9R	3.5	MSMEG_1178	Transcriptional regulator
10	1385631	1385671	6.48	10L	2.4	MSMEG_1292*	Dehydrogenase protein
				10R	4.2	MSMEG_1293*	Xanthine/uracil permeases family protein
11	1684231	1684271	64.6		2.8	MSMEG_1597	Transcription factor WhiB
12	1832291	1832331	46.5		-13.2	MSMEG_1738	Probable conserved transmembrane protein
13	1965171	1965211	19.3		No DE	MSMEG_1886	Fatty acid desaturase
14	2000471	2000511	10.9		No DE	MSMEG_1919	Transcription factor WhiB
15	2070111	2070151	9.9		120.7	MSMEG_1987*	Putative uncharacterised protein
16	2081471	2081511	19.5		-2.1	MSMEG_1999	Putative uncharacterised protein
17	2260871	2260911	39.9	17L	2.3	MSMEG_2183	Conserved hypothetical protein
				17R	38.8	MSMEG_2184*	Amino acid permease
18	2414891	2414931	67.9		10.1	MSMEG_2332	Amino acid carrier protein
19	2508191	2508231	101.5		98.8	MSMEG_2425*	*amtB* Ammonium transporter
20	2592931	2592971	18.6		-4.1	MSMEG_2506*	Carboxyvinyl-carboxyphosphonate phosphorylmutase
21	2608351	2608391	171.1		165.9	MSMEG_2522*	Efflux ABC transporter, permease protein
22	2612531	2612571	331.2		782.4	MSMEG_2526	Copper amine oxidase
23	2655531	2655571	56.3		50.8	MSMEG_2570*	Xanthine/uracil permease
24	3048291	3048331	105.9		583.8	MSMEG_2982*	Putative periplasmic binding protein
25	3206851	3206891	8.7	25L	No DE	MSMEG_3131	Polypeptide: AMP-binding protein
				25R	No DE	MSMEG_3132	Polypeptide: DNA-binding protein
26	3237471	3237511	6.5		No DE	MSMEG_3166	Enzyme: beta-lactamase
27	3471571	3471611	8.2		228.0	MSMEG_3400*	Glutamyl-tRNA(Gln) amidotransferase subunit A
28	4043191	4043231	22.8		2.1	MSMEG_3975	Putative regulatory protein, PucR family
29	4069251	4069291	58.9		9.7	MSMEG_3995	N-carbomoyl-L-amino acid amidohydrolase
30	4070051	4070091	13.2	30L	8.3	MSMEG_3996	*hydA* Dihydropyrimidinase
				30R	6.5	MSMEG_3997	Regulatory protein, PucR family
31	4082411	4082451	77.2		49.2	MSMEG_4008*	Oxidoreductase, 2OG-Fe(II) oxygenase family protein
32	4136531	4136571	7.4		No DE	MSMEG_4063	Polypeptide: amidohydrolase
33	4290471	4290511	8.0		115.7	MSMEG_4206	Molybdopterin oxidoreductase
34	4374791	4374831	184.7		20.0	MSMEG_4290	*glnA* Glutamine synthetase
35	4381891	4381931	49.8		12.6	MSMEG_4294	*glnA* Glutamine synthetase, type I
36	4580191	4580231	384.4		103.3	MSMEG_4501	Sodium:dicarboxylate symporter
37	4722511	4722551	17.1		102.0	MSMEG_4635*	*amtA* Ammonium transporter
38	4726751	4726791	63.6		57.3	MSMEG_4639*	Putative uncharacterised protein
39	4729431	4729471	11.1		No DE	MSMEG_4643	Resuscitation-promoting factor
40	4729931	4729971	34.4		No DE	MSMEG_4643	Resuscitation-promoting factor
41	5183411	5183451	57.5		27.1	MSMEG_5084*	Glycosyl transferase, group 2 family protein
42	5440611	5440651	233.9		14.9	MSMEG_5358	Acetamidase/Formamidase family protein
43	5442051	5442091	27.2		29.1	MSMEG_5360*	Formate/nitrate transporter

44	5651011	5651051	18.6		No DE	MSMEG_5561	HPP family protein
45	5840591	5840631	11.6		4.1	MSMEG_5765	*glbN* Globin
46	6177591	6177631	31.6		24.8	MSMEG_6116	Putative allantoicase
47	6323551	6323591	23.7		255.9	MSMEG_6259	*amt1* Ammonium transporter
48	6714291	6714331	16.3		8.1	MSMEG_6660	Permease, cytosine/purine/uracil/thiamine/allantoin
49	6747051	6747091	9.9	49L	No DE	MSMEG_6695	Cytochrome P450
				49R	No DE	MSMEG_6697	IS1096, tnpA protein
50	6782771	6782811	17.7		128.3	MSMEG_6735*	Amino acid permease, putative
51	6865371	6865411	199.7		385.3	MSMEG_6816	Molybdopterin oxidoreductase
52	*6867931*	*6867971*	*12.7*		*Inside*	*MSMEG_6817*	
53	6930751	6930791	10.8		5.8	MSMEG_6881	Transcriptional regulator, GntR family

Table 6.2. GlnR binding sites identified in *M. smegmatis* during nitrogen limitation and corresponding gene expression levels of WT vs Δ*glnR* during nitrogen limitation.

[a]Ordered by chromosome position, the table displays enriched binding regions, peaks, with the [b]coordinates on the *M. smegmatis* genome. [C]Peak intensity was calculated using SISSRs, based on the number of sequenced tags at each site vs the input control sample. [d]D.reg indicates the direction of the gene in relation to GlnR binding (L=left and R=right), where GlnR may have a divergent role. [e]Gene fold change represent the fold change value between *M. smegmatis* WT vs Δ*glnR* strain during nitrogen limitation. Genes in operons are denoted by *. Genes highlighted in grey displayed no DE (differential expression) on the microarray; these genes are potential GlnR regulated genes, but not included in our analysis.

Peak[a]	Chr[b] start	Chr[b] end	Peak Intensity Excess[c]	Peak Intensity Limiting[c]	Gene ID	Gene Description
1	1832291	1832331	6.7	46.5	MSMEG_1738	Probable conserved transmembrane protein
2	2508171	2508211	5.38	101.46	MSMEG_2425*	*amtB* Ammonium transporter
3	4374771	4374811	6.27	184.71	MSMEG_4290	*glnA* Glutamine synthetase
4	4381891	4381931	6.94	49.84	MSMEG_4294	*glnA* Glutamine synthetase, type I
5	5651011	5651051	5.51	18.6	MSMEG_5561	HPP family protein

Table 6.3. GlnR binding sites identified in *M. smegmatis* during nitrogen excess.

[a]Ordered by chromosome position, the table displays enriched binding regions, peaks, with the [b]coordinates on the *M. smegmatis* genome. [c]Peak intensity of each peak was calculated using SISSRs, based on the number of sequenced tags at each site vs the input control sample. Genes in operons are denoted by *. Peak intensity during nitrogen limitation is included in the table for direct comparison at the binding sites.

Identification of three previously known GlnR binding sites, upstream of *amt1*, *amtB* and *glnA1*, confirmed that the ChIP-seq approach had successfully identified specific GlnR binding regions (Figure 6.3) (3). Further validation of the ChIP-seq results was provided by performing electromobility shift assays (EMSAs) using purified recombinant His-GlnR protein and four novel GlnR DNA binding regions. DNA sequences of 200 bp representing promoter regions of peak 19 (*amtB*), included as a positive control, peak 17 (msmeg_2184), peak 21 (msmeg_2522), peak 22 (msmeg_2526), and peak 42 (msmeg_5358), were assayed for specific GlnR binding. All putative binding sites analysed bound GlnR specifically with a protein-concentration dependent shift (Figure 6.4). The promoter region of msmeg_3224, a region not identified as a GlnR binding site in this study and included as a negative control, displayed no GlnR binding (Figure 6.4). No difference was observed in GlnR binding to the different DNA regions assayed, despite notable differences in fold peak enrichment. For example the GlnR binding site of msmeg_2526, with a peak fold enrichment of 331.2, displayed similar protein:DNA binding (determined by EMSA) to the GlnR protein, as did msmeg_2184 which had a peak fold enrichment of 39.9 (Figure 6.4).

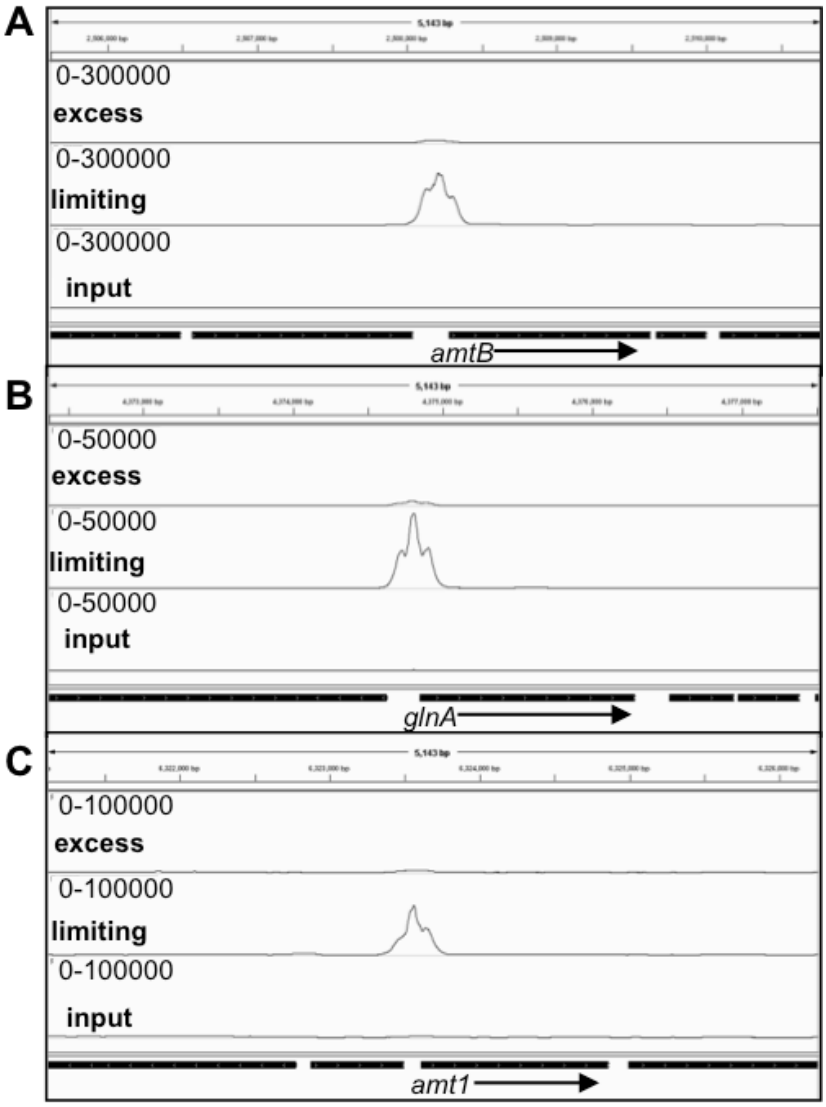


Figure 6.3. ChIP-seq confirmed GlnR binding during nitrogen limitation upstream of (A) *amtB* (B) *glnA1* (C) *amt1*.

Binding data was visualised using the Integrated Genome Viewer (IGV). Upper track in each panel indicates ChIP-seq data from nitrogen excess conditions (30 mM ammonium sulphate), middle track ChIP-seq data from nitrogen limiting conditions (1 mM ammonium sulphate) and the total DNA input in aligned at the bottom track. The black bars at the bottom signify gene transcripts, location of known GlnR regulated genes are labelled accordingly.

(A) GlnR binding upstream of *amtB* during nitrogen limitation. (B) GlnR binding upstream of *glnA1* during nitrogen limitation. (C) GlnR binding upstream of *amt1* during nitrogen limitation.

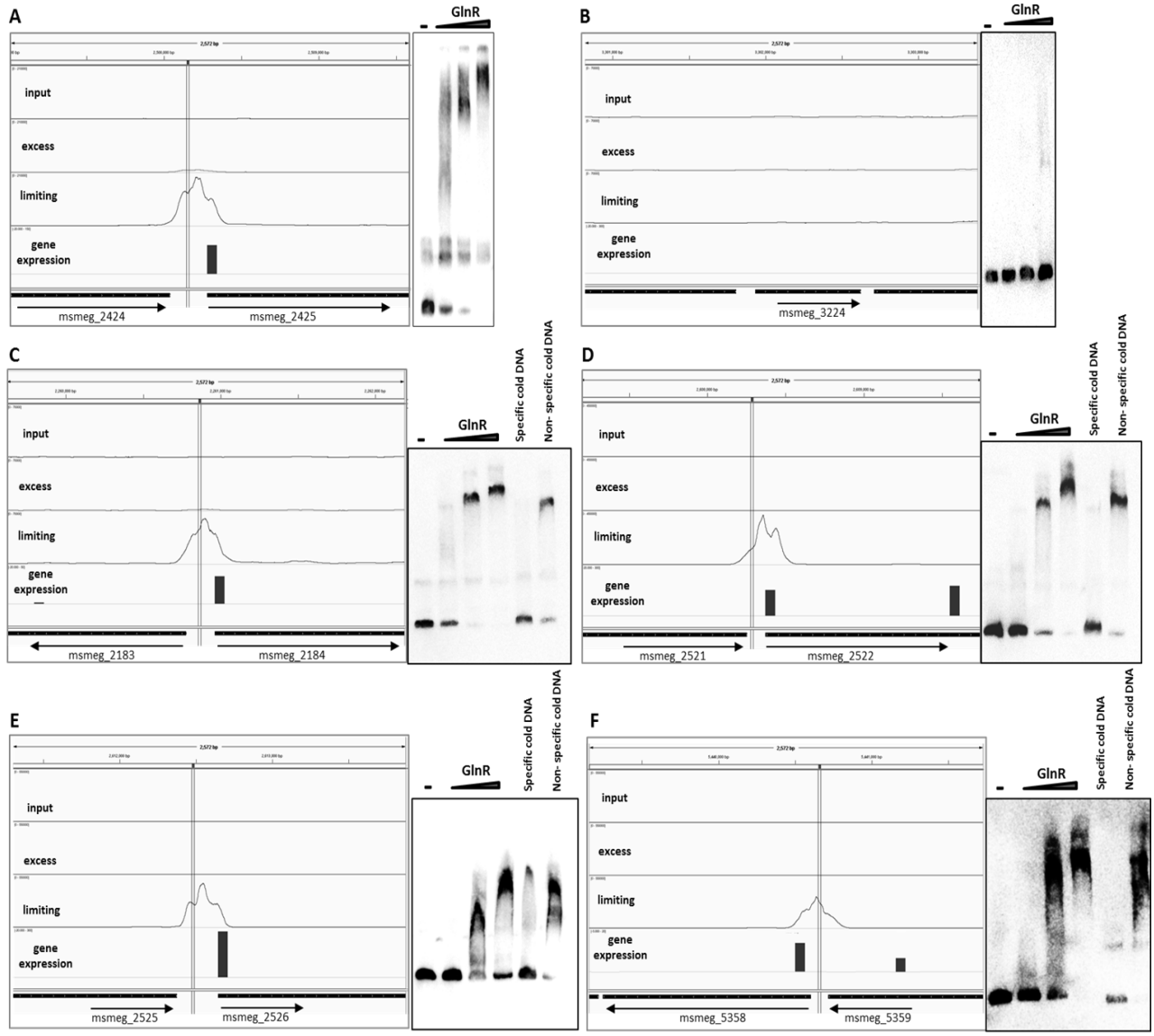


Figure 6.4. Confirmation of GlnR binding to DNA by EMSA.

EMSAs were carried out with increasing amounts of His-GlnR 0-0.9 µg with labelled 200 bp DNA corresponding to the promoter regions of the genes. Specific and non-specific cold competitor DNA was added at 1000 x excess to the labelled probe. GlnR ChIP-seq data is visualised in IGV, with the top track representing the total DNA input, with nitrogen excess (30 mM ammonium sulphate) and limiting (1 mM ammonium sulphate) conditions in the second and third track respectively. Gene expression is indicated by vertical black bars on the bottom track at the start site of each gene; bar height is representative of fold change in gene expression (WT vs *glnR* deletion strain). The location of the GlnR consensus binding site, determine via MEME, is indicated by parallel vertical lines through the peak.

(A) Peak 9 msmeg_ 2425 (*amtB*). (B) Negative control msmeg_3224. (C) Peak 17 msmeg_2184. (D) Peak 21 msmeg_2522. (E) Peak 22 msmeg_2526. (F) Peak 42 msmeg_5358.

6.3.3 Determination of the GlnR Regulon in Nitrogen Limitation

In order to determine the regulon, the GlnR binding sites and transcripts regulated by GlnR under nitrogen limitation were combined. Array expression data was mapped onto the ChIP-seq GlnR binding site data using IGV, such that GlnR binding and GlnR regulated gene expression could be visualised (Figure 6.5). Specific binding sites could also be visualised by zooming into the region of interest (Figure 6.5 B). Figure 6.5 displays GlnR bound adjacently to differentially expressed genes for the *amtB* operon. From the 53 peaks identified by ChIP-seq analysis, 44 binding sites were associated with differential gene expression of 103 genes (Table 6.4). Genes were classified into operons according to annotation on Biocyc Data base collection (http://biocyc.org/MSME246196/NEW-IMAGE?type=ORGANISM&object=TAX-1763).

Of the 53 GlnR binding sites identified, 8 of the associated downstream genes did not show any evidence of differential expression in nitrogen limitation, suggesting that these genes are either not GlnR-regulated under these conditions, or are expressed at a different time point to that taken in this study. In addition 2 peaks were upstream of genes not represented on the microarray. These genes were analysed for differentia expression by qRT-PCR. Msmeg_2332 was differentially expressed in the wild type compared to *glnR* deletion strain (up regulated 10.1 fold) and deemed to be in GlnR regulated, but the other, msmeg_6697 was not differentially expressed.

Further analysis was conducted on the 9 genes that had adjacent GlnR binding sites but did not exhibit significant GlnR mediated gene expression. Rate limiting PCR was carried out for all 9 binding sites on independent ChIP samples, to confirm that the amplification observed was not a result of sample processing during library preparation, and all the binding sites, except peak 52, displayed enrichment (Figure 6.6 and Figure 6.7). Analysing the IGV view of peak 52 this was determined to be a miscall by SISSRs (Figure 6.7). In addition, GlnR binding to peak 13 was confirmed by EMSA (Figure 6.8). This suggests that these 8 peaks do represent GlnR binding sites in nitrogen limitation and therefore the 8 genes downstream of these binding sites are categorised as putative GlnR regulated genes (Table 6.2).

Interestingly, as well as the 96 genes up-regulated by GlnR during nitrogen limitation, 7 genes were down regulated. Down regulated genes represent 6.8 % of all genes under GlnR control, indicating that GlnR can both activate and repress transcription, with the main role as an activator. Six GlnR binding sites are associated with divergently transcribed genes under nitrogen limitation, suggesting GlnR may act in a bidirectional manner at some binding sites.

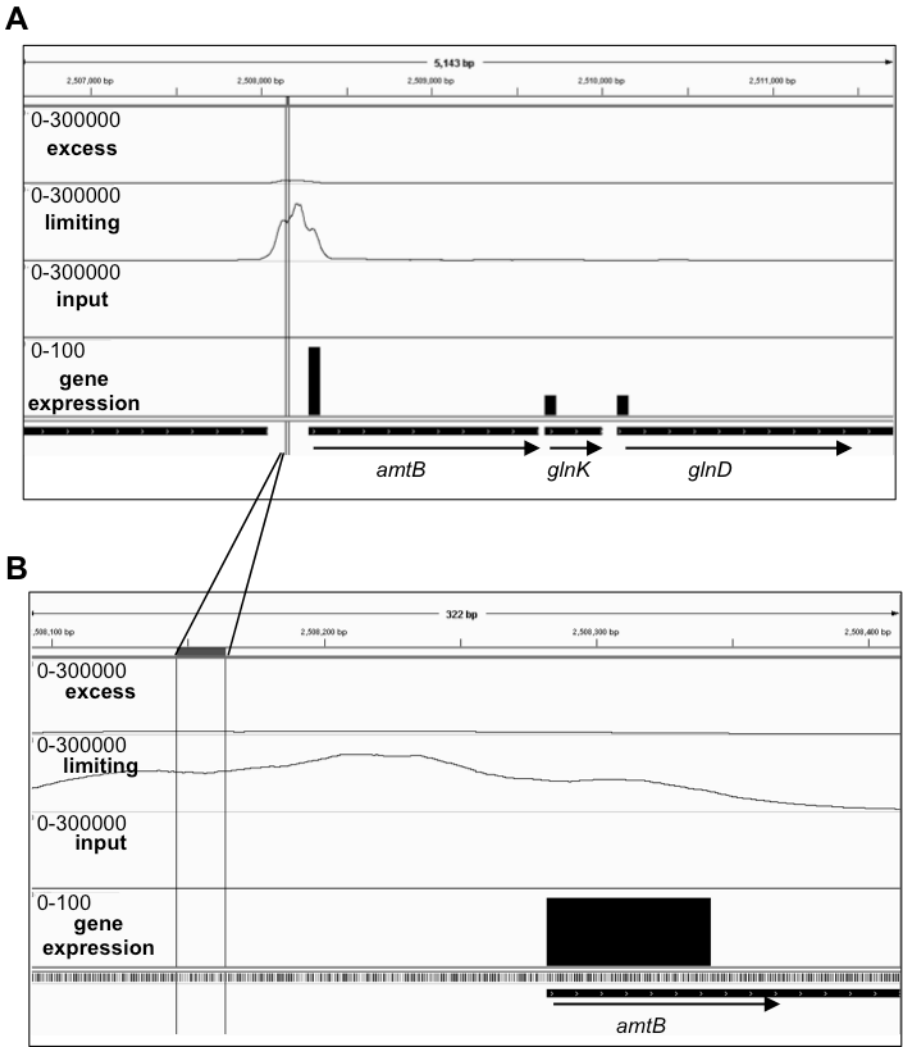


Figure 6.5. GlnR ChIP-seq binding data for the *amtB-glnK-glnD* operon.

Binding data was visualised using IGV. Excess track indicates ChIP-seq data from nitrogen excess conditions, the limiting track ChIP-seq data from nitrogen limiting conditions and the total DNA input in aligned in the bottom input track. The black vertical bars represent gene expression during nitrogen limitation; fold change of the WT vs *glnR* mutant. Scale for gene expression is 0-100 fold with the height of the bars representing the fold change value. The location of the GlnR consensus binding site, determine via MEME, is indicated by parallel vertical lines through the peak.

(A) GlnR binding and gene expression data for the *amtB-glnK-glnD* operon window size 5243 bp. (B) Zoomed in image to 322 bp of the GlnR consensus sequence in relation to gene start site.

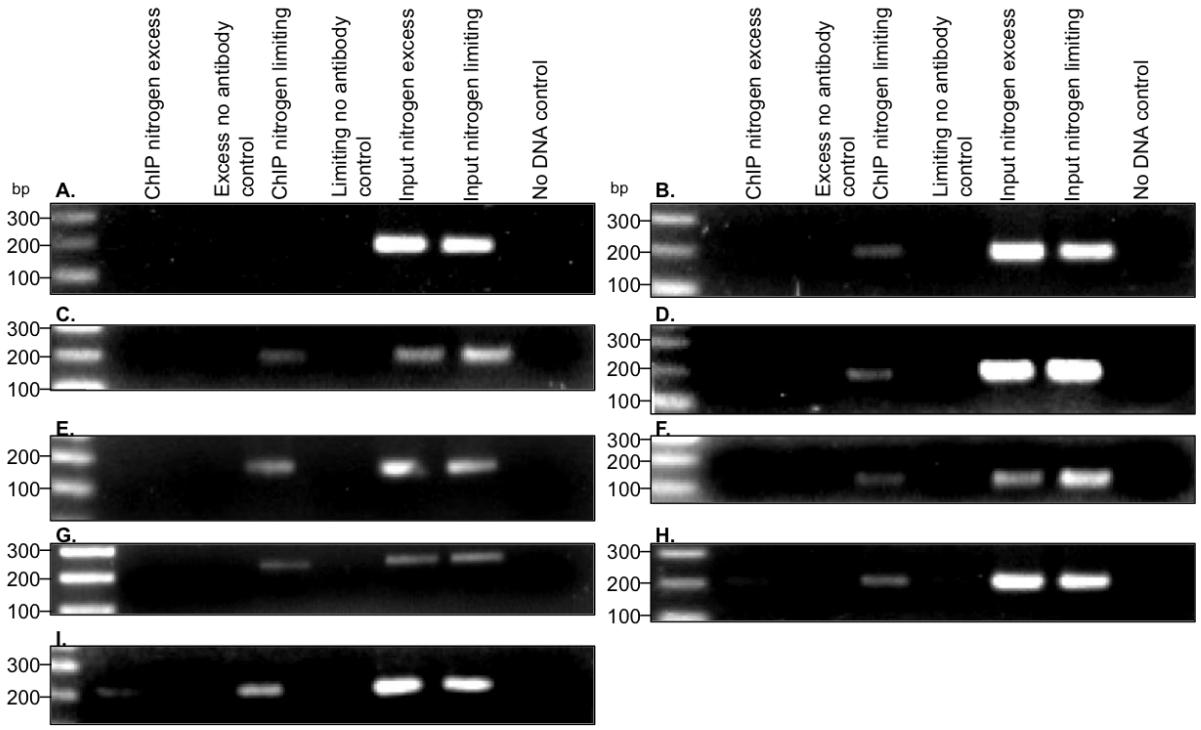


Figure 6.6. Rate limiting PCR indicating enrichment of the 8 GlnR immunoprecipitated DNA for genes that showed no significant DE during nitrogen limitation.

Rate limiting PCR involved 23 cycles amplification of 200 bp promoter regions, using 0.3 ng of template DNA. ChIP DNA represents DNA that has been immunoprecipitated with a GlnR specific antibody during nitrogen excess or nitrogen limiting conditions. Input DNA represents the total amount of DNA that was subject to immunoprecipitation.

A. Promoter of msmeg_3224 negative control
B. Promoter region representing peak 13
C. Promoter region representing peak 14
D. Promoter region representing peak 26
E. Promoter region representing peak 32
F. Promoter region representing peak 39
G. Promoter region representing peak 40
H. Promoter region representing peak 44
I. Promoter region representing peak 49

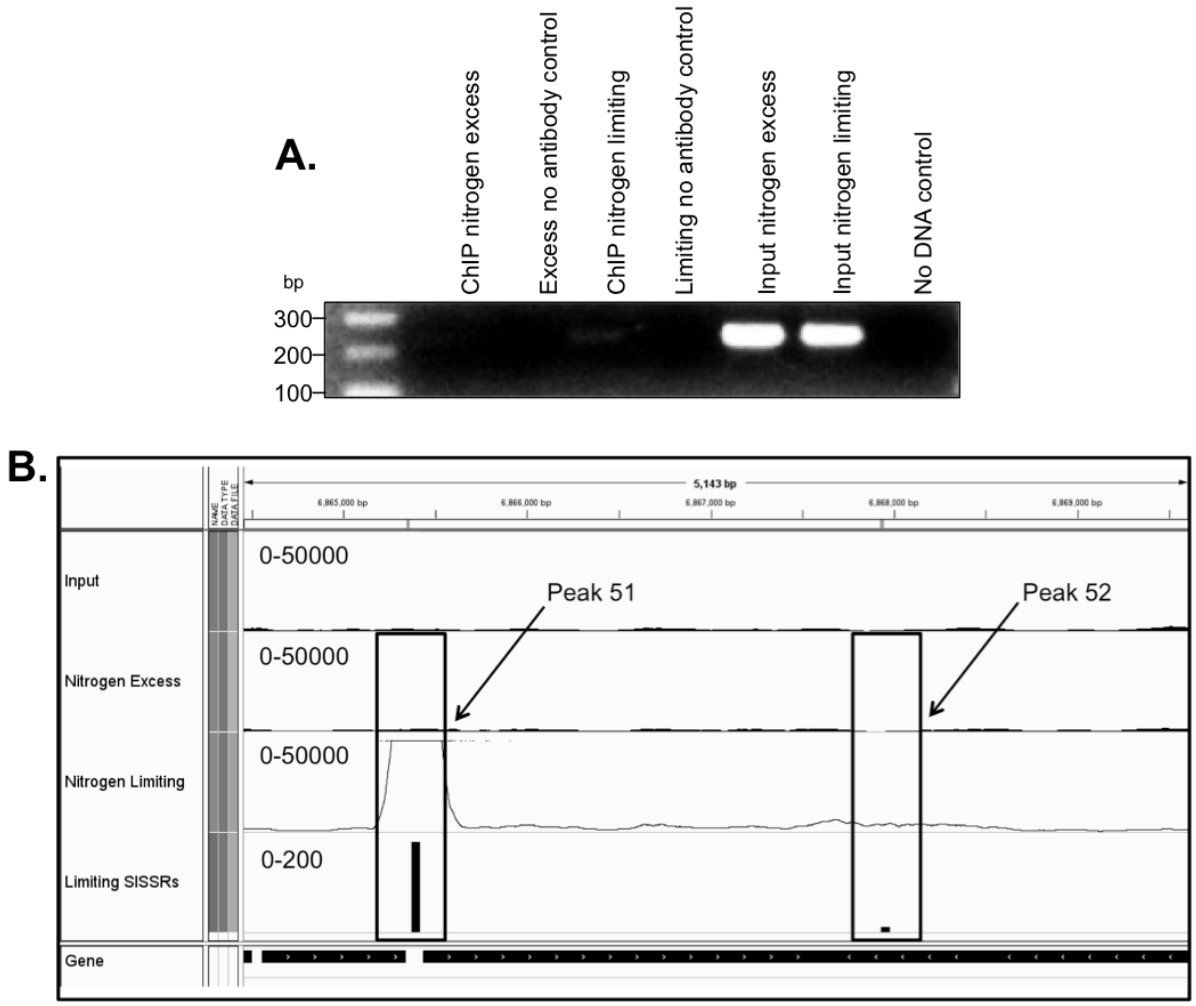


Figure 6.7. Rate limiting PCR of peak 52 showing no enrichment and IGV view of peak 52.

(A) Rate limiting PCR of promoter region of peak 52 involved 23 cycles amplification of 200 bp promoter regions, using 0.3 ng of template DNA. ChIP DNA represents DNA that has been immunoprecipitated with a GlnR specific antibody during nitrogen excess or nitrogen limiting conditions. Input DNA represents the total amount of DNA that was subject to immunoprecipitation. No enrichment is seen in nitrogen limiting conditions. (B) Binding data was visualised using IGV. Upper track indicates ChIP-seq data from the Input sample representing the total DNA, middle track is nitrogen excess conditions and then ChIP-seq data from nitrogen limiting conditions. Aligned to the bottom track is the SISSRs value for the peaks highlighted by the vertical black bars.

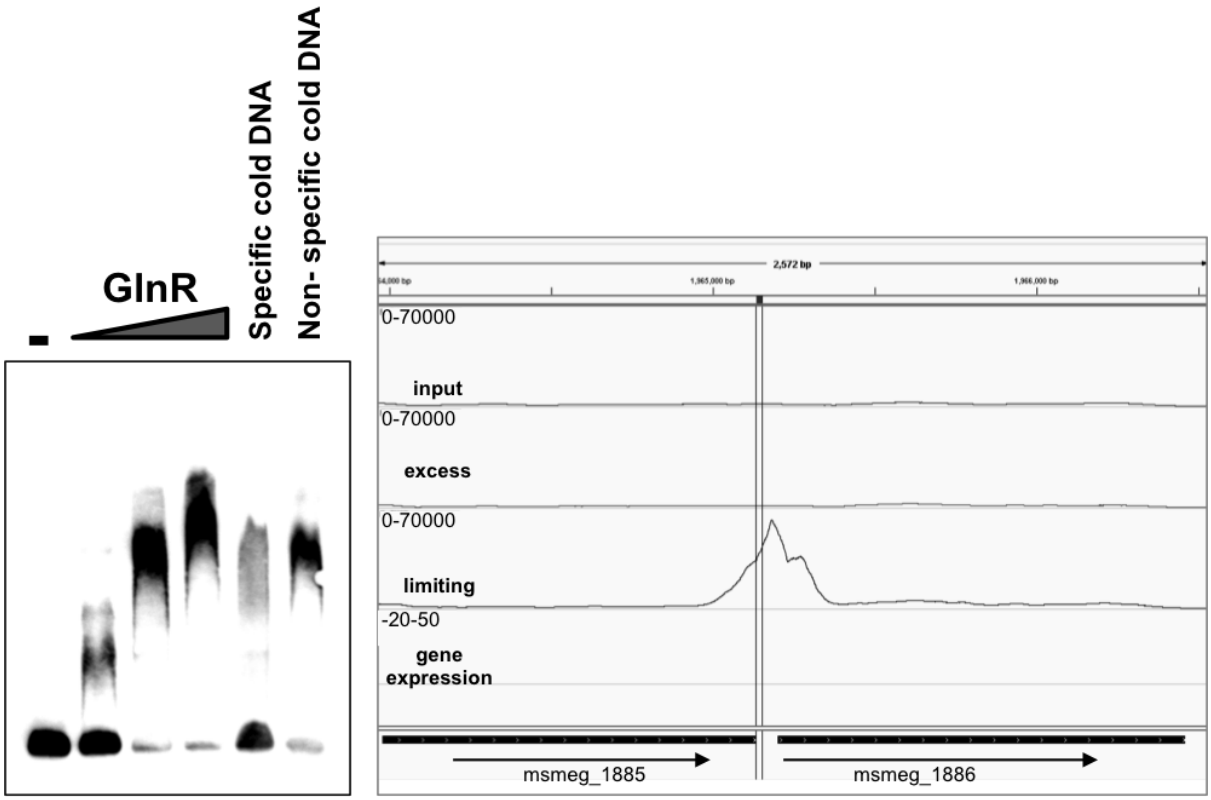


Figure 6.8. EMSA of GlnR binding to 200 bp region, peak 13 alongside ChIP-seq data.

Increasing amounts His-GlnR (0-0.9 μg) were incubated with 200 bp around peak 13. Specific and non-specific cold competitor DNA was added at 1000x excess to the labelled probe. Binding data was visualised using IGV. Upper track indicates ChIP-seq data from the Input sample representing the total DNA, middle track is nitrogen excess conditions and the ChIP-seq data from nitrogen limiting conditions aligned at the bottom track. The black bars at the bottom signifying genes are labelled accordingly. The location of the consensus binding site is indicated by the vertical line through the peak.

P.[a]	Peak intensity[b]	D[c]	Fold change WT vs Δ*glnR* Gene 1	Gene 1 ID	Fold change WT vs Δ*glnR* Gene 2	Gene 2 ID	Fold change WT vs ΔglnR Gene 3	Gene 3 ID	Fold change WT vs ΔglnR Gene 4	Gene 4 ID	Fold change WT vs ΔglnR Gene 5	Gene 5 ID	Fold change WT vs ΔglnR Gene 6	Gene 6 ID	Fold change WT vs ΔglnR Gene 7	Gene 7 ID	Fold change WT vs ΔglnR Gene 8	Gene 8 ID
1	8.41		76.4	MSMEG_0427	13.0	MSMEG_0428												
2	42.9		18.3	MSMEG_0432	17.3	MSMEG_0431												
3	8.4		24.6	MSMEG_0433														
4	27.07		263.4	MSMEG_0572	109.6	MSMEG_0571	27.6	MSMEG_0570	37.3	MSMEG_0569	30.0	MSMEG_0568	26.3	MSMEG_0567	6.8	MSMEG_0566	4.4	MSMEG_0565
5	6.13	L	23.0	MSMEG_0780	14.6	MSMEG_0779	4.3	MSMEG_0778										
		R	8.4	MSMEG_0781														
6	54.34		6.3	MSMEG_1052														
7	6.47	L	-3.8	MSMEG_1078														
		R	3.4	MSMEG_1079	4.5	MSMEG_1080	2.5	MSMEG_1081										
8	71.93		277.4	MSMEG_1082														
9	19.34	L	10.7	MSMEG_1177														
		R	3.5	MSMEG_1178														
10	6.48	L	2.4	MSMEG_1292	3.3	MSMEG_1291												
		R	4.2	MSMEG_1293	3.1	MSMEG_1294	3.6	MSMEG_1295	3.5	MSMEG_1296	3.7	MSMEG_1297	4.2	MSMEG_1298				
11	64.64		2.8	MSMEG_1597														
12	46.51		-13.2	MSMEG_1738														
13	19.25		No DE															
14	10.94		No DE															

P.[a]	Peak intensity[b]	D[c]	Fold change WT vs Δ*glnR* Gene 1	Gene 1 ID	Fold change WT vs Δ*glnR* Gene 2	Gene 2 ID	Fold change WT vs ΔglnR Gene 3	Gene 3 ID	Fold change WT vs ΔglnR Gene 4	Gene 4 ID	Fold change WT vs ΔglnR Gene 5	Gene 5 ID	Fold change WT vs ΔglnR Gene 6	Gene 6 ID	Fold change WT vs ΔglnR Gene 7	Gene 7 ID	Fold change WT vs ΔglnR Gene 8	Gene 8 ID
15	9.94		120.7	MSMEG_1987	42.4	MSMEG_1988	19.8	MSMEG_1989	32.9	MSMEG_1990								
16	19.5		-2.1	MSMEG_1999														
17	39.9	L	2.3	MSMEG_2183														
		R	38.8	MSMEG_2184	16.7	MSMEG_2185	28.8	MSMEG_2186	6.8	MSMEG_2187	3.0	MSMEG_2189						
18	67.87		10.1	MSMEG_2332														
19	101.46		98.8	MSMEG_2425	29.8	MSMEG_2426	29.1	MSMEG_2427										
20	18.6		-4.1	MSMEG_2506	-3.8	MSMEG_2505	-2.4	MSMEG_2504	-2.5	MSMEG_2503								
21	171.12		165.9	MSMEG_2522	191.5	MSMEG_2523	499.1	MSMEG_2524										
22	331.2		782.4	MSMEG_2526														
23	56.3		50.8	MSMEG_2570	3.5	MSMEG_2569												
24	105.92		583.8	MSMEG_2982	130.1	MSMEG_2981	276.1	MSMEG_2980	141.6	MSMEG_2979	77.4	MSMEG_2978						
25	8.66		No DE															
26	6.46		No DE															
27	8.19		228	MSMEG_3400	57.1	MSMEG_3401	76.6	MSMEG_3402	36.5	MSMEG_3403								
28	22.78		2.1	MSMEG_3975														
29	58.91		9.7	MSMEG_3995														
30	13.19	L	8.3	MSMEG_3996														
		R	6.5	MSMEG_3997														

P.[a]	Peak intensity[b]	D[c]	Fold change WT vs Δ*glnR* Gene 1	Gene 1 ID	Fold change WT vs Δ*glnR* Gene 2	Gene 2 ID	Fold change WT vs ΔglnR Gene 3	Gene 3 ID	Fold change WT vs ΔglnR Gene 4	Gene 4 ID	Fold change WT vs ΔglnR Gene 5	Gene 5 ID	Fold change WT vs ΔglnR Gene 6	Gene 6 ID	Fold change WT vs ΔglnR Gene 7	Gene 7 ID	Fold change WT vs ΔglnR Gene 8	Gene 8 ID
31	77.17		49.2	MSMEG_4008	6.6	MSMEG_4009	21.2	MSMEG_4010	18.4	MSMEG_4011	15.9	MSMEG_4012	10.5	MSMEG_4013	5.6	MSMEG_4014		
32	7.38		No DE															
33	7.98		115.7	MSMEG_4206														
34	184.71		20.0	MSMEG_4290														
35	49.84		12.6	MSMEG_4294														
36	384.39		103.3	MSMEG_4501														
37	17.09		102.0	MSMEG_4635	57.6	MSMEG_4636												
38	63.56		57.3	MSMEG_4639	191.5	MSMEG_4638	157.9	MSMEG_4637										
39	11.12		No DE															
40	34.38		No DE															
41	57.47		27.1	MSMEG_5084	5.2	MSMEG_5083	2.9	MSMEG_5082										
42	233.93		14.9	MSMEG_5358														
43	27.24		29.1	MSMEG_5360	7.0	MSMEG_5359												
44	18.6		No DE															
45	11.62		4.1	MSMEG_5765														
46	31.59		24.8	MSMEG_6116														
47	23.65		255.9	MSMEG_6259														
48	16.34		8.1	MSMEG_6660														
49	9.9		No DE															

P.[a]	Peak intensity[b]	D[c]	Fold change WT vs Δ*glnR* Gene 1	Gene 1 ID	Fold change WT vs Δ*glnR* Gene 2	Gene 2 ID	Fold change WT vs ΔglnR Gene 3	Gene 3 ID	Fold change WT vs ΔglnR Gene 4	Gene 4 ID	Fold change WT vs ΔglnR Gene 5	Gene 5 ID	Fold change WT vs ΔglnR Gene 6	Gene 6 ID	Fold change WT vs ΔglnR Gene 7	Gene 7 ID	Fold change WT vs ΔglnR Gene 8	Gene 8 ID
50	17.72		128.25	MSMEG_6735	30.68	MSMEG_6734	5.20	MSMEG_6733										
51	199.66		385.33	MSMEG_6816														
52	12.65		No DE															
53	10.78		5.78	MSMEG_6881														

Table 6.4. List of GlnR binding sites during nitrogen limitation and genes regulated by GlnR listed in their operons.

[a]P. indicates peak number corresponding to GlnR binding sites listed in Table 6.2. [b]Peak intensity is calculated with SISSRs and represents the fold change in number of tags sequenced at the region comparing nitrogen limiting conditions to the input control sample. [c]D represents the direction from GlnR binding of gene transcription. Gene operons are based on predicted operons according to Biocyc Data base collection (http://biocyc.org/MSME246196/NEW-IMAGE?type=ORGANISM&object=TAX-1763). Gene 1 represents the gene in closest proximity to the GlnR binding site, followed by gene 2 etc. Expression values represent the fold change between *M. smegmatis* WT vs Δ*glnR* during nitrogen limitation and are shaded in the table.

The 103 GlnR regulated genes were categorised using Clusters of Orthologous Groups of Proteins (COG) functional classifications (Section 2.7.6). Most genes annotated are involved in nitrogen uptake and metabolism, as well as a large number of genes with unknown function (Table 6.5). Nitrogen transport genes represented the largest group upregulated during nitrogen limitation with 26 genes in total. In addition genes that encoding enzymes predicted to be involved in the release of ammonia from cellular sources, such as amine oxidase and hydrolases, are listed in Table 6.6. These genes are of particular interest as they suggest a nitrogen scavenging response from alternate nitrogen sources. The largest category was proteins of unknown function (FUN) and although these need confirming experimentally, it seems likely that these genes encode proteins involved in nitrogen metabolic processes.

COG functional Classification	Number of genes
Amino Acid	4
Cell Envelope	3
Cellular Process	1
Central Intermediary Metabolism	16
DNA Metabolism	1
Energy Metabolism	11
FUN	29
Protein Fate	1
Protein Synthesis	2
Regulatory Functions	9
Transport And Binding Proteins	26

Table 6.5. Functional classification of genes in the GlnR regulon.

Gene ID	Name	Product	COG Classification	EC number	Reaction
MSMEG_0427	*nirB*	Nitrite reductase [NAD(P)H], large subunit	Central Intermediary	E 1.7.1.4	NO_2^- + 3 NAD(P)H + 5 H^+<=> NH_4^+ + 3 NAD(P)$^+$ + 2 H_2O
MSMEG_0428	*nirD*	Nitrite reductase [NAD(P)H] small subunit	Central Intermediary	E 1.7.1.4	NO_2^- + 3 NAD(P)H + 5 H^+<=> NH_4^+ + 3 NAD(P)$^+$ + 2 H_2O
MSMEG_0566		Aliphatic amidase	Energy Metabolism		
MSMEG_0567		Selenophosphate synthetase/ N-acetyltransferase activity	Central Intermediary		
MSMEG_0571		Hydrolase, carbon-nitrogen family protein	NA		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_0779		Short-chain dehydrogenase/reductase SDR	Central Intermediary		
MSMEG_1078		Hydrolase	Central Intermediary		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_1080		Large subunit of N,N-dimethylformamidase	Protein Synthesis	E 3.5.1.56	N,N-dimethylformamide + H_2O = dimethylamine + formate
MSMEG_1294		Allantoicase	Energy Metabolism	E 3.5.3.4	allantoate + H_2O <=> (S)-ureidoglycolate + urea
MSMEG_1294		Allantoicase	Energy Metabolism		allantoate + H_2O <=> (S)-ureidoglycolate + urea
MSMEG_1295		Transthyretin	NA		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_1296		Uricase	Central Intermediary	E 1.7.3.3	Urate + O_2 + H_2O <=> 5-hydroxyisourate + H_2O_2
MSMEG_1297		Hydroxydechloroatrazine ethylaminohydrolase	Energy Metabolism	E 3.5.99.3	4-(ethylamino)-2-hydroxy-6-(isopropylamino)-1,3,5-triazine + H_2O <=> N-isopropylammelide + ethylamine
MSMEG_1298		Guanine deaminase	Central Intermediary	E 3.5.4.3	Guanine + H_2O <=> xanthine + NH_3
MSMEG_1989		Phenoxybenzoate dioxygenase beta subunit	Energy Metabolism		
MSMEG_2186		Putative urea carboxylase-associated protein	NA		Urea carboxylase/allophanate hydrolase pathway
MSMEG_2187		Urea amidolyase	Energy Metabolism	E 3.5.1.54 & E 6.3.4.6	ATP + urea + HCO_3 <=> ADP + phosphate + urea-1-carboxylate & Urea-1-carboxylate + H_2O <=> 2 CO_2 + 2 NH_3
MSMEG_2189	*atzF*	Allophanate hydrolase	Central Intermediary	E 3.5.1.54	Urea-1-carboxylate + H_2O <=> 2 CO_2 + 2 NH_3
MSMEG_2506		Carboxyvinyl-carboxyphosphonate phosphorylmutase	Central Intermediary		
MSMEG_2526		Copper amine oxidase	Central Intermediary	E 1.4.3.6	Amine oxidase converts primary amines to aldehydes with release of ammonia and H_2O_2
MSMEG_3403		Formamidase	Energy Metabolism	E 3.5.1.49	Formamide + H_2O <=> formate + NH_3
MSMEG_3995		N-carbamoyl-L-amino acid amidohydrolase	Central Intermediary		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_3996	*hydA*	Dihydropyrimidinase/ primidine base catabolism	Energy Metabolism	E 3.5.2.2	5,6-dihydrouracil + H_2O <=> 3-ureidopropanoate
MSMEG_4009		Vanillate O-demethylase oxidoreductase	Energy Metabolism		
MSMEG_4012	*hydA*	Dihydropyrimidinase/ primidine base catabolism	Energy Metabolism	E 3.5.2.2	5,6-dihydrouracil + H_2O <=> 3-ureidopropanoate
MSMEG_4013		Oxidoreductase	Central Intermediary		
MSMEG_4014		N-carbamoyl-L-amino acid amidohydrolase	Central Intermediary		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_4638		Vanillate O-demethylase oxidoreductase	Energy Metabolism		
MSMEG_5358		Acetamidase/Formamidase family protein	NA		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_5359	*cynS*	Cyanate hydratase	Cellular Process	E 4.2.1.104	Cyanate + HCO_3 + 2 H <=> NH_3 + 2 CO_2
MSMEG_6116		Putative Allantoicase	Central Intermediary		Catalyses the conversion of OHCU into S(+)-allantoin; it is the third step of the conversion of uric acid (a purine derivative) to allantoin
MSMEG_6733		Hydrolase, carbon-nitrogen family protein	Central Intermediary		Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
MSMEG_6734		Dibenzothiophene desulfurization enzyme A	Energy Metabolism		
MSMEG_6816		Molybdopterin oxidoreductase	Central Intermediary		

Table 6.6. GlnR regulated genes include many enzymes that catalyse the release of nitrogen from various sources.

6.3.4 Identification and Analysis of the *M. smegmatis* GlnR DNA Binding Motif

To identify a consensus GlnR DNA binding motif, a 200 bp nucleotide sequence corresponding to the centre of the enriched peak was extracted using the R package Biostrings and submitted to the motif discovery tool Multiple EM (Expectation Maximization) for Motif Elicitation (MEME) (11). A consensus DNA binding motif was identified that was present once in all 52 GlnR binding sites identified by ChIP-seq (E value of 6.5E-30) (Figure 6.9). No direct correlation between the specific GlnR binding DNA sequence, its genomic location and the levels of gene expression was observed (Table 6.7). Key residues required for specific GlnR binding were identified by mutational analysis of the highly conserved AC-n_9-AC and AT-n_9-AC sequences. Figure 6.10 shows that the AC bases at position 5 & 6 and 16 & 17 in the MEME sequence (Figure 6.9) are critical for GlnR binding, since reduced GlnR:DNA binding is observed when these residues are mutated to G. The spacing of 9 nucleotides between these key AC residues at positions 5 & 6 and 16 & 17 was also investigated. Figure 6.10 displays that the spacing between these two adenosine residues was essential for strong GlnR:DNA binding.

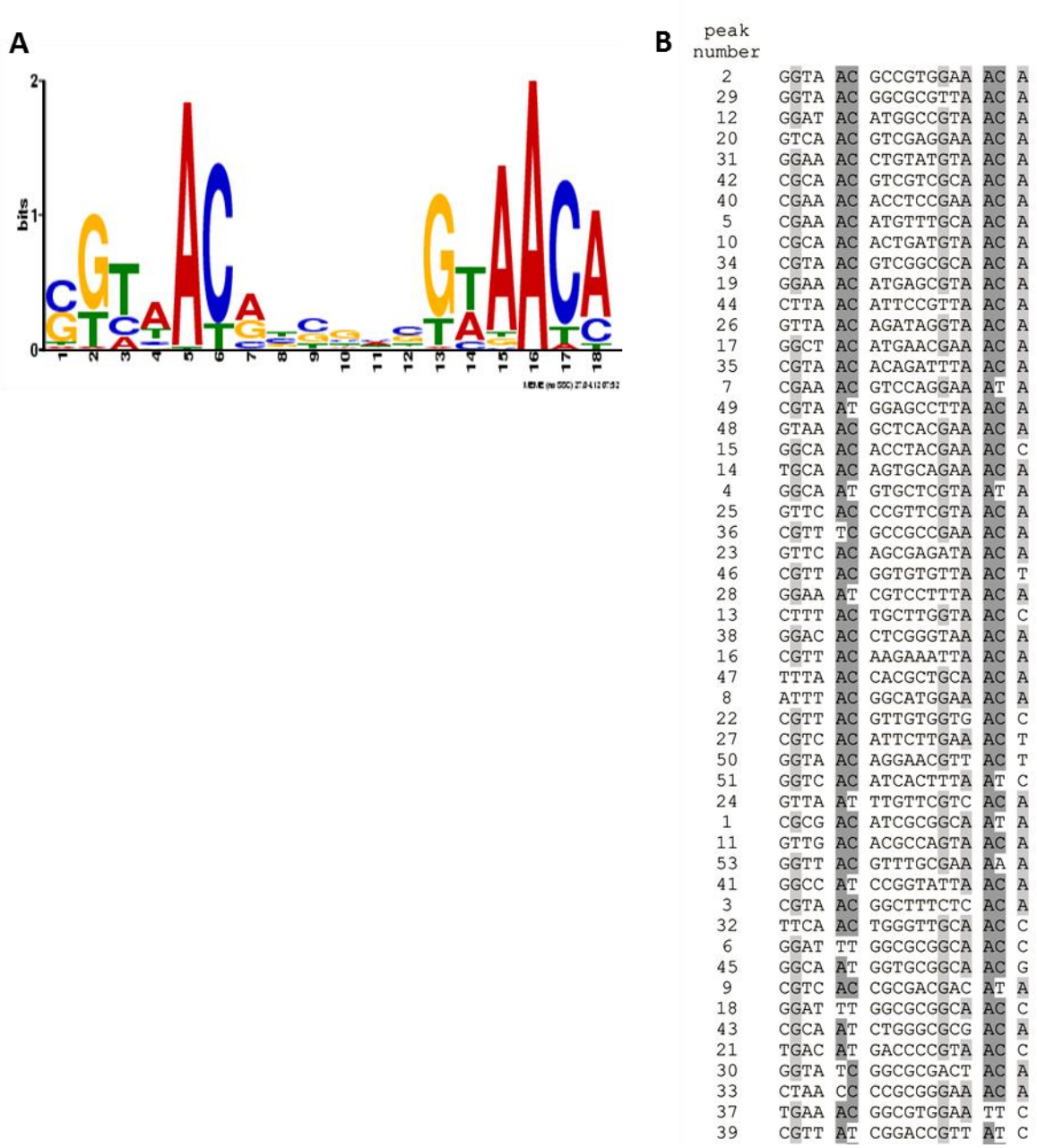


Figure 6.9. MEME-derived consensus sequence from GlnR binding regions identified via ChIP-seq.

(A) MEME consensus of GlnR binding regions generated from 200 bp regions from the centre of each peak. (B) Alignment of MEME-generated motifs in different peaks. Conserved residues are highlighted in grey.

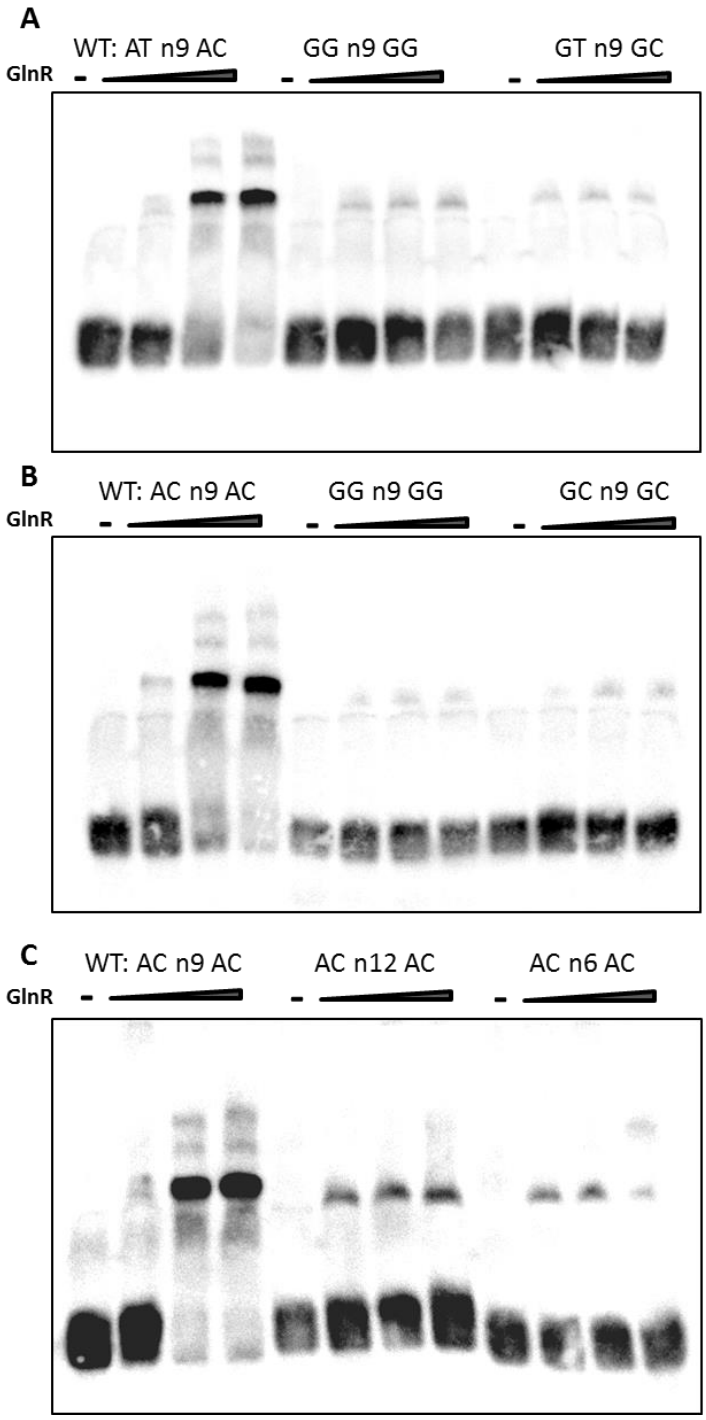


Figure 6.10. EMSA of GlnR binding to 30 bp consensus region with diminished GlnR binding observed upon mutation of highly conserved residues.

EMSAs were carried out with 30 bp annealed complementary oligonucleotides, incubated with increasing amounts of His-GlnR (0-0.9 μg).

(A) WT represents consensus site of peak 24, with the conserved AT n9 Ac residues mutated.
(B) WT represents consensus site of peak 2, with the conserved AC n9 AC residues mutated.
(C) WT represents consensus site of peak 2, with the distance between AC n9 AC altered to increase the spacing by 3 nucleotides and decrease the spacing by 3 nucleotides.

Strand[a]	*p*-value[b]	DNA sequence adjacent motif[c]	GlnR binding sequence[d]	DNA sequence adjacent motif[c]	Peak number[e]	Peak intensity[f]	Direction of gene[g]	Gene expression level WT[h]	Gene ID
-	1.70E-08	GCAATCGCGG	GGTA AC GCCGTGGAA AC A	GAGCCTGCCT	2	42.9	Left	28884	MSMEG_0432
-	1.70E-08	GCAATGTGCC	GGTA AC GGCGCGTTA AC A	ATGATGCACG	29	58.91	Right	12062	MSMEG_3995
-	3.49E-08	TGTCAGCATG	GGAT AC ATGGCCGTA AC A	CCCCCGAAAC	12	46.51	Left	10627	MSMEG_1738
+	1.46E-07	TGGGGTCCCA	GTCA AC GTCGAGGAA AC A	CGTCTGTCAA	20	18.6	Left	501	MSMEG_2506
-	1.75E-07	GCCTTACATC	GGAA AC CTGTATGTA AC A	GACGATCACA	31	77.17	Right	13394	MSMEG_4008
-	1.75E-07	CGAACGCAAC	CGCA AC GTCGTCGCA AC A	TGGCGTGGGT	42	233.93	Left	139956	MSMEG_5358
+	2.09E-07	ATACCGCCCG	CGAA AC ACCTCCGAA AC A	TCAACGGAGC	40	34.38	No DE		
+	2.49E-07	CTCCCAGCTT	CGAA AC ATGTTTGCA AC A	AACCCGCATA	5	6.13	Left	12287	MSMEG_0780
+	2.49E-07	AGCAATGCTC	CGCA AC ACTGATGTA AC A	ACGGCGCGGC	10	6.48	Left	9872	MSMEG_1292
					10		Right	71152	MSMEG_1293
+	2.49E-07	AAACCGATTG	CGTA AC GTCGGCGCA AC A	TCGGGTTGAC	34	184.71	Right	270298	MSMEG_4290
+	2.95E-07	TTTGTCGGCG	GGAA AC ATGAGCGTA AC A	GTGATCGGGA	19	101.46	Right	247467	MSMEG_2425
-	3.49E-07	CGATAACCGG	CTTA AC ATTCCGTTA AC A	TCGTTGGGGC	44	18.6	No DE		
+	4.08E-07	CCGCGCCTTG	GTTA AC AGATAGGTA AC A	CATCGAACAT	26	6.46	No DE		
+	1.00E-06	GCTTACCGAC	GGCT AC ATGAACGAA AC A	TTCGGGTGAC	17	39.94	Right	3778	MSMEG_2184
-	1.00E-06	CTTGCACCCC	CGTA AC ACAGATTTA AC A	GCCGGTGCAT	35	49.84	Left	58365	MSMEG_4294
-	1.51E-06	TACGTCGTTC	CGAA AC GTCCAGGAA AT A	CTGGCGCCCG	7	6.47	Left	4383	MSMEG_1078
					7		Right	2548	MSMEG_1079
-	1.96E-06	TTCTAACAGG	CGTA AT GGAGCCTTA AC A	AAAACGCCGA	49	9.9	No DE		
-	2.23E-06	GACATGTTCT	GTAA AC GCTCACGAA AC A	TTTGCCGTGT	48	16.34	Right	43208	MSMEG_6660
+	3.22E-06	TTTCACCTCA	GGCA AC ACCTACGAA AC C	GTTCATCGCG	15	9.94	Right	69987	MSMEG_1987
-	3.63E-06	AAACATCAGA	TGCA AC AGTGCAGAA AC A	TTTGTGTGCA	14	10.94	No DE		
+	4.08E-06	ATCTCACAGC	GGCA AT GTGCTCGTA AT A	AGTGCAGCAT	4	27.07	Left	47270	MSMEG_5765
-	4.58E-06	GCGAAACGGC	GTTC AC CCGTTCGTA AC A	CGATCTACCC	25	8.66	No DE		
-	6.42E-06	ATATTTCGCG	CGTT TC GCCGCCGAA AC A	TCGCGAACAC	36	384.39	Left	57599	MSMEG_4501
-	7.17E-06	TTTTCGCCGA	GTTC AC AGCGAGATA AC A	CGCGTTGGAT	23	56.3	Left	11315	MSMEG_2570
-	7.17E-06	TGAACAGCGG	CGTT AC GGTGTGTTA AC T	GCGGGCTAAA	46	31.59	Right	110010	MSMEG_6116
-	7.17E-06	TCAATCAGCC	GGAA AT CGTCCTTTA AC A	CGTTTGTGAC	28	22.78	Right	805	MSMEG_3975
+	1.21E-05	GATGTTTAGA	CTTT AC TGCTTGGTA AC C	TACGGAGCCG	13	19.25	No DE		
+	1.21E-05	GGCCCGTTCG	GGAC AC CTCGGGTAA AC A	CGTATCGCCG	38	63.56	Left	31748	MSMEG_4639
-	1.48E-05	GAAAAATTTG	CGTT AC AAGAAATTA AC A	AGCACGATTG	16	19.49	Left	26128	MSMEG_1999
+	1.48E-05	CGGTTTGAGT	TTTA AC CACGCTGCA AC A	CTTGGCGACC	47	23.65	Right	78758	MSMEG_6259
+	1.81E-05	CCGTGGCGTG	ATTT AC GGCATGGAA AC A	GGCTCTGAAC	8	71.93	Right	116062	MSMEG_1082
-	1.81E-05	CACGGCCCCG	CGTT AC GTTGTGGTG AC C	TGACGCAACG	22	331.2	Right	149912	MSMEG_2526
+	2.19E-05	CGACTGATGA	CGTC AC ATTCTTGAA AC T	TCACGACAAC	27	8.19	Right	22922	MSMEG_3400
-	2.89E-05	TGTAGCGAGC	GGTA AC AGGAACGTT AC T	GTGGCGGCCA	50	17.72	Left	26371	MSMEG_6735
+	2.89E-05	CCACCCCTGA	GGTC AC ATCACTTTA AT C	TCGACGCAAT	51	199.66	Right	136949	MSMEG_6816

Strand[a]	*p*-value[b]	DNA sequence adjacent motif[c]	GlnR binding sequence[d]	DNA sequence adjacent motif[c]	Peak number[e]	Peak intensity[f]	Direction of gene[g]	Gene expression level WT[h]	Gene ID
-	3.47E-05	TCATGTCGAG	GTTA AT TTGTTCGTC AC A	CACAGACATT	24	105.92	Left	123674	MSMEG_2982
+	3.79E-05	CTTCACACAG	CGCG AC ATCGCGGCA AT A	TCGGGTTCTT	1	8.41	Right	19569	MSMEG_0427
+	3.79E-05	CTGGCAGTTA	GTTG AC ACGCCAGTA AC A	ATCGCGGCAT	11	64.64	Left	176450	MSMEG_1597
+	3.79E-05	TCAGTCTGTG	GGTT AC GTTTGCGAA AA A	TTTCTGTTGC	53	10.78	Left	7160	MSMEG_6881
+	6.86E-05	CATCAGGACC	GGCC AT CCGGTATTA AC A	AGATCTTTTA	41	57.47	Left	91914	MSMEG_5084
-	8.73E-05	CTGGTGACCG	CGTA AC GGCTTTCTC AC A	GGTGCCCGCC	3	8.4	Left	6511	MSMEG_0433
+	1.02E-04	AGCCCCTGCC	TTCA AC TGGGTTGCA AC C	GTCGAAATGT	32	7.38	No DE		
-	1.10E-04	CTTCGTGTGA	GGAT TT GGCGCGGCA AC C	GTGCAAGATC	6	54.34	Left	148877	MSMEG_1052
-	1.10E-04	GGTACGCCTG	GGCA AT GGTGCGGCA AC G	GGGTTCTCCG	45	11.62	Left	47270	MSMEG_5765
-	1.10E-04	ACTTGTCACC	CGTC AC CGCGACGAC AT A	CCGAGTCCGT	9	19.49	Left	21830	MSMEG_1177
					9		Right	14890	MSMEG_1178
-	1.10E-04	CTTCGTGTGA	GGAT TT GGCGCGGCA AC C	GTGCAAGATC	18	67.87	Left	No array	MSMEG_2332
-	1.59E-04	GTCGCCGAAC	CGCA AT CTGGGCGCG AC A	CGGTCTAGGG	43	27.24	Left	9158	MSMEG_5360
+	2.11E-04	GCGTGGCGCG	TGAC AT GACCCCGTA AC C	CGGGCCGTGT	21	171.12	Right	116597	MSMEG_2522
-	2.42E-04	AACGGACCTT	GGTA TC GGCGCGACT AC A	GGCCAAACTG	30	13.19	Left	6204	MSMEG_3996
							Right	54547	MSMEG_3997
-	2.95E-04	TTTCCTGACG	CTAA CC CCGCGGGAA AC A	TCCCGGAACA	33	7.98	Left	16530	MSMEG_4206
+	2.95E-04	GTCGAGGTCG	TGAA AC GGCGTGGAA TT C	ATGCTTCGTT	37	17.09	Right	136346	MSMEG_4635
-	4.05E-04	CTCATCGAGT	CGTT AT CGGACCGTT AT C	TGCCCGGGAT	39	11.12	No DE		

Table 6.7. MEME-derived GlnR binding sites with corresponding ChIP-seq peak intensity and gene expression for WT in nitrogen limiting conditions.

GlnR binding sites are listed in order of their [b]P value, with sequences with greatest similarity to the MEME-derived consensus sequence listed at the top. [a]Strand represents the DNA strand from which the sequence was derived. [d]GlnR binding sequence derived from MEME is listed in coloured font with the highly conserved AC residues underlined. [c]Nucleotides surrounding this consensus sequence are displayed in grey and labelled DNA sequence adjacent to motif. [e]Peak number represents the GlnR binding site identified during nitrogen limitation. [f]Peak intensity is calculated with SISSRs and represents the fold change in number of tags sequenced at this region comparing nitrogen limiting conditions to the input control sample. [g]Direction of gene is the location of the gene with respect to the peak location. [h]Expression value indicates the normalised expression level of the gene from the WT *M. smegmatis* microarray during nitrogen limitation.

6.4 Discussion

The OmpR-type response regulator GlnR is the global transcriptional regulator during nitrogen stress in *Streptomyces*. In *Streptomyces* GlnR is thought to regulate around 50 genes during times of nitrogen deficiency, including those involved in the uptake and assimilation of nitrogen (120, 164, 165). GlnR is predicted to be the global transcriptional regulator in mycobacteria during times of nitrogen limitation, yet to date this has not been fully investigated (3). Therefore, to investigate the role of GlnR in the adaption of *M. smegmatis* to nitrogen stress, a transcriptomic and ChIP-seq approach was applied to determine the entire GlnR nitrogen regulon.

Initially, the transcriptomic response was determined by comparing the global expression profile of *M. smegmatis* wild type to a *glnR* deletion mutant when both were exposed to nitrogen limitation. In total 680 genes were differentially expressed during nitrogen limitation in the wild type strain compared to the mutant; 392 genes were significantly upregulated and 291 down regulated in the wild type strain compared to the *glnR* mutant, indicating that GlnR acts as both an activator and a repressor. However, these GlnR-dependant effects may be direct or indirect. For instance, many response regulators were also upregulated in the wild type strain and therefore genes under control of these regulators (discussed later) could be incorrectly included in the GlnR regulon if relying on transcriptomic data alone. Consequently ChIP-seq was used to differentiate genes directly and indirectly controlled by GlnR.

Using a GlnR specific antibody to identify where GlnR binds to DNA during nitrogen limitation, we identified 53 GlnR binding sites in *M. smegmatis*. Of these 53 binding sites, 44 sites were upstream of transcripts identified as GlnR-controlled in the microarray. These genes, including genes predicted to be in corresponding operons, were therefore included in the GlnR regulon (Table 6.4). In addition to the 44 sites directly controlling GlnR transcripts, two of the GlnR binding sites identified in this study were located upstream of genes not represented on the microarray and therefore these genes were analysed for differential expression by qRT-PCR. Msmeg_2332, encoding an amino acid carrier protein, was shown to be under GlnR control whereas the other gene (msmeg_6697) was not.

In total, 9 binding sites identified by ChIP-seq were not associated with GlnR specific differential gene expression during nitrogen limitation. Therefore, these binding sites were investigated further to determine if they were true GlnR binding sites during

nitrogen limitation. Rate limiting PCR confirmed enrichment of 8 binding sites in independent ChIP samples to those sent for sequencing. Peak 52 did not show enrichment via rate limiting PCR and a closer visual examination of the region identified in IGV, lead to a conclusion that this was a miscall by SISSRs as no clear peak was observed. EMSA studies further confirmed specific binding of GlnR to one binding site (peak 13). These 8 sites were therefore deemed true GlnR binding sites; the genes downstream of these genes were assumed to be under GlnR control, but may be expressed at a different time point to the one taken for the global expression profiling or requires additional transcription factors not yet recruited for expression.

Genes Involved in Nitrogen Metabolism

The main nitrogen metabolism pathway during nitrogen limitation utilises the glutamine synthetase enzyme GS. The *glnA1* gene encoding a type-I glutamine synthetase (GS) in *M. tuberculosis* has been found to be essential, and has been implicated in virulence as well as nitrogen metabolism (171). In *Streptomyces* the type-I GS in this species (*glnA*), has been demonstrated to be GlnR regulated, indicating the importance of GlnR in GS regulation (120, 165). In this study both *glnA1* and a second glutamine synthetase *glnA2* were upregulated in the *M. smegmatis* WT compared to the *glnR* deletion mutant, and thus under GlnR control. Located between the two genes on the genome is *glnE*, which been implicated in post-translational modification of GS during nitrogen limitation. Whilst *glnE* does not contain a GlnR binding site up stream of its predicted transcriptional start site, it was down regulated in the wild type array when compared to the *glnR* mutant. GlnE is thought to modulate GS activity via adenylylation, inactivating GS activity. Down regulation of *glnE* would result in less GS adenylylation and consequently a more active GS under conditions of nitrogen stress. According to the data presented here, GlnR appears to be involved in the indirect regulation of GlnE at the transcriptional level, suggesting it has a pivotal role in GS transcriptional regulation during nitrogen stress.

Other notable nitrogen metabolism genes under GlnR control are nitrate reductases. Nitrate is an important molecule for bacteria not only under nitrogen stress but also during times of hypoxia. *M. tuberculosis* is predicted to experience hypoxic conditions within macrophages during infection. As such, during hypoxic conditions the terminal electron acceptor would shift from oxygen to nitrate. Regulation of genes involved in nitrate metabolism are therefore important, not only to provide an alternate nitrogen

source that will provide nitrogen compounds for metabolism, but also other areas of energy production. As such we identified GlnR dependant regulation of *nirB* and *nirD*. A nitrate transcriptional regulator (*nnaR*) was also determined to be in the GlnR regulon (discussed later) suggesting the fine tuning of genes involved in nitrate metabolism. Again these results are analogous to *Streptomyces*, with nitrate reductase genes and the nitrate transcriptional regulator under GlnR control (2).

As the external nitrogen source runs out, genes are upregulated that are involved in processes that utilise other cellular compounds as nitrogen sources. Therefore in times of austerity cells appear to switch to using alternative nitrogen sources. Genes that encode enzymes that enable the release of ammonium from other sources (Table 6.6), includes uricase, urea amidolase, hydrolases and amine oxidases. Whilst the pathways and COG classification for these genes may vary, what is apparent is that they all result in the release of ammonium from a variety of cellular or environmental sources. This release of ammonium can then be targeted to the nitrogen assimilation pathways in the cell, providing a valuable source of nitrogen.

Transport and binding proteins represent another major category of genes under GlnR control during nitrogen limitation; 27 genes in total. In addition to the ammonium transporters, *amtA*, *amtB* and *amt1*, several other genes (such as, msmeg_0781, msmeg_1052, msmeg_2522 and msmeg_2524) are predicted to encode ABC transport systems for substrates such as amino acids and small peptides. These observations suggest that the transport of various other nitrogenous compounds into the cell from the environment is important. *M. smegmatis* is predicted to encode many more transporters than other mycobacterial species, probably due to the soil environment in which it has evolved. An operon (msmeg_2978-2982) encoding genes similar to the *C. glutamicum* urease transport operon was shown to be under positive GlnR regulation in this study (5). *M. smegmatis* is the only mycobacterial species to feature such a urea ABC transporter, emphasising the importance of urea as a nitrogen source for this species.

From these results it is clear that the major response of *M. smegmatis* to nitrogen limitation is to scavenge all available nitrogen sources from the environment. Genes involved in scavenging ammonium (*amt1, amtA, amtB, glnA1* and *glnA2*), urea (transporters; msmeg_2978–2982, urea amidolyase, uricase, urea carboxylase-associated protein), nitrate (*nirB, nirD*), guanine deaminase (msmeg_1298) are all upregulated during nitrogen limitation and controlled by GlnR. This would make sense from an evolutionary perspective, as a soil dwelling organism *M. smegmatis* would

encounter numerous nitrogen sources in the environment, and it appears that *M. smegmatis* has evolved to be able take up and assimilate these sources. In addition, in the genus *M. smegmatis* is the species with the most genes annotated as being involved in nitrogen metabolism, showing its importance for survival in this species.

Regulatory Proteins

Nine regulatory proteins are included in the GlnR regulon and are upregulated during nitrogen limitation; this includes seven transcription factors (TF). In *Streptomyces* it was noted that out of 70 genes differentially expressed in the *glnR* deletion mutant, eight were putative TF, suggesting GlnR acts as a global regulator in both species (120). These TF may play a role in fine-tuning nitrogen utilisation. One such example is msmeg_0432 (upregulated 18.3 fold in WT vs Δ*glnR*), which shares 56% identity with SCO2958 (*nnar*) from *S. coelicolor*. Recently, *nnar* in *S. coelicolor* was reported to be GlnR regulated and bound upstream of the nitrate reductases, suggesting NnaR may play a further regulatory role in nitrate assimilation (2). Another notable GlnR regulated transcription factor is WhiB3, which in *M. tuberculosis* is an effector molecule controlling several aspects of virulence (136). *M. tuberculosis* WhiB3 senses fluctuations in the intracellular redox environment associated with O_2 depletion and the metabolic switchover to the favoured *in vivo* carbon source, fatty acids (149). GlnR may, therefore, regulate additional regulatory molecules with roles in other aspects of metabolism and survival.

In addition to transcriptional regulatory proteins, GlnR regulates proteins involved in post-translational modifications. Two proteins, GlnD and GlnK, have been well described in their role during nitrogen metabolism. They form part of the highly conserved *amtB-glnK-glnD* operon in *Actinomycetes*, encoding an ammonium transporter (AmtB), P_{II} signalling protein (GlnK) and an adenylyl transferase (GlnD). In *C. glutamicum*, it is thought that ammonium levels are detected by GlnD which adenylylates the GlnK (PII) protein. Upon adenylylation by GlnD, GlnK dissociates from AmtB porin channel permitting an increase in ammonium influx. It is unclear if this pathway is conserved in mycobacteria. However, it indicates the importance of a cascade of responses during nitrogen limitation, not only at the transcriptional level, but also the regulation of proteins involved in post-translational modification.

Dual Function

GlnR can act as a dual transcriptional regulator, both activating and repressing gene expression in nitrogen limitation, although only 6.8% of GlnR-regulated genes were repressed in *M. smegmatis*. GlnR repression of gene transcription was also documented using proteomics and qRT-PCR of selected genes in a *glnR* deletion mutant of *S. coelicolor* (164, 165). This dual repressor/activator function has been described for other OmpR-like regulators. In *E. coli* OmpR controls the transcription of *ompF* and *ompC*, two genes encoding porin proteins. Under conditions of high osmolarity, OmpR represses *ompF* and activates *ompC* transcription (201). It has been demonstrated in *E. coli* that OmpR-P bind to promoters in a hierarchical model; the OmpF promoter regions are F1>F2>F3>F4. Binding of OmpR to F1 and F2 activates transcription of OmpF, however an increase in OmpR-P results in binding to the F4 site, a weak OmpR-P-binding site. OmpR-P binding to the F4 site is proposed to form a loop that interacts with OmpR-P molecules binding to F1, F2, and F3, thereby blocking *ompF* transcription (201). The mechanism by which GlnR mutually activates and represses transcription is unclear, however it may involve conformational changes in DNA topology.

Of the genes repressed by GlnR in *M. smegmatis*, msmeg_1738 was down regulated to the greatest extent (-13.15). Msmeg_1738 is annotated as a probable conserved membrane protein, predicted to be part of the DosR regulon in *M. tuberculosis* (179). This suggests that the decrease in growth rate under nitrogen limitation stimulates other pathways, such as factors associated with dormancy. In addition, subunits of 3-isopro-pylmalate dehydratase, involved in amino acid biosynthesis, a hydrolase (msmeg_1078) and carboxyvinyl-carboxyphosphonate phosphorylmutase (msmeg_2506) were again down regulated. These changes indicate a shift in the biosynthesis pathways utilised by *M. smegmatis* under nitrogen limitation, possibly due to the observed reduction in growth rate.

GlnR DNA Binding

In order to identify a consensus GlnR DNA binding motif, the DNA sequence 200 bp from the centre of the 52 identified GlnR binding sites, were analysed using the MEME algorithm (Figure 6.9). A consensus sequence of $Gn_2AC\text{-}n_6GnAACA$ was determined and found to be present once in all GlnR binding sites identified in this study. Amon *et al.* (2008) previous *in silico* study only reported 3 binding sites in *M. smegmatis*, however

they used a 44 bp region defined by earlier *Streptomyces* studies (3). Later studies in *Streptomyces* by Tiffert and Pullan further defined the GlnR DNA binding motif to 16 bp, which is similar to the mycobacterial GlnR binding motif identified in this study (120, 165). Using the *M. smegmatis* GlnR DNA binding motif generated by MEME we attempted to search the *M. smegmatis* genome using FIMO (Find Individual Motif Occurrences) for additional putative GlnR binding sites. This generated over 2000 hits, possibly due to the small consensus sequence, and as such was not deemed useful for further analysis.

Tiffert *et al.* proposed the existence of two separate GlnR binding sites within the binding motif in *S. coelicolor* (165). The two binding sites have been termed an "a site" (gTnAc) and a highly conserved "b site" (GaAAc), located 6 bp apart, in which the highly conserved "b site" has a higher affinity for GlnR than the "a site". However, the motif identified in *S. venezuelae*, GTnAC-n_6-GTnAC does not provide a distinction between an "a" and "b site", but rather two copies of an "a site" (120). In *M. smegmatis* the GlnR binding motif contains an "a site" (Gn_2AC) separated by 6 bp followed by a "b site" (GnAAC). Whilst the "b site" sequence is highly similar to the *S. coelicolor* "b site" sequence, the "a site" is more variable. Pullan *et al.* suggested that conservation of the "b site" might be indicative of strong GlnR regulation. Yet combining the ChIP-seq binding data with transcriptomic data in this study, it is apparent some of the most up-regulated genes contain a highly conserved "b site" (msmeg_4501 and msmeg_5358), whilst others (msmeg_6816, msmeg_2982 and msmeg_2526) do not possess a highly conserved "b site", yet the genes are expressed at similar levels (Table 6.7). This indicates that the conserved "b site" is not solely an indication of the effect of GlnR regulation. Other factors such as binding of additional transcription factors or sigma factor usage may also be a contributing factor to gene expression.

While there is some discrepancy between the "a" and "b sites" for each species, what is apparent is the conserved region Ac-n_9-Ac present in the motif search for *M. smegmatis* and *Streptomyces*. Further investigations were made into the requirements for these key residues and the distance between residues in the GlnR DNA binding motif. Mutating the highly conserved A residues ablated specific GlnR binding, indicating its importance for GlnR-DNA interactions (Figure 6.10). Furthermore the distance between these conserved residues was shown to be crucial for specific GlnR binding. The separation of these residues, 9 bp, represents one turn of the major groove of the DNA helix, which may be an important factor for determining the GlnR DNA interaction.

Comparison of the Streptomyces Regulon

Homologues of GlnR regulated genes in *M. smegmatis* are also responsive to changes in nitrogen limitation in *Streptomyces*. A ChIP-seq approach in *S. venezuelae* identified 36 GlnR binding sites compared to the 52 identified here in *M. smegmatis*. Targets found in both species include *glnA, amtB* and *nnaR* (uroporphyrinogen-III synthase). However, some clear differences were noted. Fewer transporter genes were GlnR regulated in *S. venezuelae* when compared to *M. smegmatis,* maybe reflecting difference in environmental nitrogenous compounds, or reflective of the experimental procedures used. In the *Streptomyces* study genes in operons failed to be included in the regulon, possibly accounting for the reduced number of GlnR regulated genes in this species (120). Other notable differences include a glutamate synthetase gene Sven_1677 which contained a GlnR binding site in *S. venezuelae*, whilst glutamate synthetases in *M. smegmatis* (msmeg_6263 and msmeg_6262) do not contain a GlnR-binding site upstream of the genes, even though they are up-regulated (20 and 5 fold respectively) in WT compared to the *glnR* mutant. Again the same was seen for *ureA* in *S. venezuelae* deemed to be GlnR regulated via ChIP-ChIP analysis, and whilst upregulated 2-fold in our array, msmeg_3627 was not under direct GlnR regulation. This indicates that some differences exist between the regulatory roles of GlnR in these species, however in both GlnR has a primary role in regulation of genes required during nitrogen limitation.

Conclusion

In summary, it has been confirmed that GlnR is the global nitrogen response regulator in *M. smegmatis.* The GlnR regulon has considerably increased in size from 11 to at least 103 genes; the majority of which are involved in scavenging and assimilating alternative environmental nitrogen sources when the cell experiences nitrogen starvation.

CHAPTER 7: ChIP-seq Analysis of Global GlnR DNA Binding Sites in *Mycobacterium tuberculosis* During Nitrogen Stress

7.1 Aim

To establish GlnR binding sites in *M. tuberculosis* during nitrogen excess and limiting growth conditions, allowing identification of potentially GlnR regulated genes. To compare these GlnR binding sites with previously identified GlnR regulated genes in *M. smegmatis* to evaluate the role of GlnR in both organisms.

7.2 Introduction

In this study GlnR has been demonstrated to play a pivotal role in regulation of nitrogen metabolism genes during nitrogen limitation in *M. smegmatis,* therefore we sought to investigate the role of GlnR in *M. tuberculosis*. Previous analysis of the *M. tuberculosis* genome identified a GlnR protein that shares 73% identity with GlnR of *M. smegmatis* (5). In addition, 8 GlnR regulated nitrogen metabolism genes in *M. smegmatis* share homology to genes present in the *M. tuberculosis* genome (Table 7.1).

Despite this, the two mycobacterial species vary significantly with respect to the environmental niche in which they reside. *M. smegmatis* contains many more nitrogen metabolism genes, when compared to the genome of *M. tuberculosis* (5). For instance three ammonium transporters are found in the genome of *M. smegmatis* (*amtA, amtB* and *amt1*) whereas only *amtB* is located in the genome of *M. tuberculosis* (5). *M. smegmatis* also contains a greater number of genes for assimilation of various other nitrogen sources, such as an additional urease transport operon (msmeg_2978-2982) and hydrolases releasing ammonium from other nitrogenous compounds ((5), this study). Conversely, *M. tuberculosis* is a pathogenic organism that resides inside nutrient poor conditions of the host, such as macrophages. In this environment it is expected that *M. tuberculosis* experiences a hostile environment, with a limit on nitrogen sources available and/or competition with the host.

During infection *M. tuberculosis* is phagocytosed by macrophages, which is thought to limit nutrient availability to the intracellular pathogen, however the exact nitrogen source available to the bacilli is unknown. In infected tissue nitrate is available; nitrate is generated spontaneously from nitric oxide (NO), the product of nitric oxide synthase. For utilisation as a nitrogen source nitrate must be converted by *M. tuberculosis* to ammonium before entering the GS/GOGAT pathway. Genes encoding enzymes to assimilate nitrate to nitrite then to ammonium are present in the *M. tuberculosis* genome

and were demonstrated by Malm *et al.* to be functional (95). This would put a greater emphasis on genes involved in nitrate uptake and conversion to ammonium for *M. tuberculosis* survival within macrophages. The *M. tuberculosis* genome also contains a conserved ammonium transporter *amtB* suggesting that ammonium transport is important for survival (5). Other interesting nitrogen metabolism operons include genes encoding a urease enzyme, responsible for the conversion of urea to ammonium (87). Indicating that the *M. tuberculosis* genome encodes a variety of genes for the utilisation of various nitrogen sources, that may potentially be GlnR regulated. Consequently we sought to identify all GlnR binding sites in the *M. tuberculosis* genome.

Protein	Putative Function	*M. smegmatis*	*M. tuberculosis*	% Identity
NirB	Assimilatory nitrate reductase	msmeg_0427	Rv0252	76
NirD	Assimilatory nitrate reductase	msmeg_0428	Rv0253	74
Nark3	Nitrite/nitrate transporter	msmeg_0433	RV0261	62
GlnK	PII Protein (signal transduction)	msmeg_2426	Rv2919c	88
GlnD	Post-translational modification of GlnK	msmeg_2427	Rv2918c	58
AmtB	Ammonium uptake	msmeg_2425	Rv2920c	48
GlnA1	GS Ammonium assimilation	msmeg_4290	Rv2220	84
GlnA2	GS Ammonium assimilation	msmeg_4294	Rv2222	88

Table 7.1. Percentage identity at the amino acid level of GlnR regulated nitrogen metabolism genes (from this study) in *M. smegmatis* to homologues in *M. tuberculosis.* (Adapted from (5)).

7.3 Results

7.3.1 Global GlnR Binding Regions in Nitrogen Limitation

To identify putative GlnR regulated genes in the *M. tuberculosis* genome we applied our optimised ChIP-seq approach. Cells were grown with 1 mM (limiting) or 30 mM (excess) ammonium chloride as the nitrogen source. Due to previous growth analysis of *M. tuberculosis* conducted by Dr. K. Williams, the equivalent ammonium source utilised in the *M. smegmatis* study could not be investigated. *M. tuberculosis* failed to grow as expected when ammonium sulphate was used at the soul nitrogen source; growth in 1 mM ammonium sulphate was at a faster rate than 30 mM ammonium sulphate, suggesting high ammonium sulphate levels were impeding growth (Figure 7.1 KW). However, *M. tuberculosis* displayed the preferential growth dynamics when grown in ammonium chloride; after ammonium run-out in the 1 mM culture growth rate slowed, whist in the 30 mM culture growth rate was maintained (Figure 7.1 KW). Ammonium chloride was subsequently used in this study, with *M. tuberculosis* growth monitored via OD_{600} and external ammonium concentration examined by Aquaquant analysis as before.

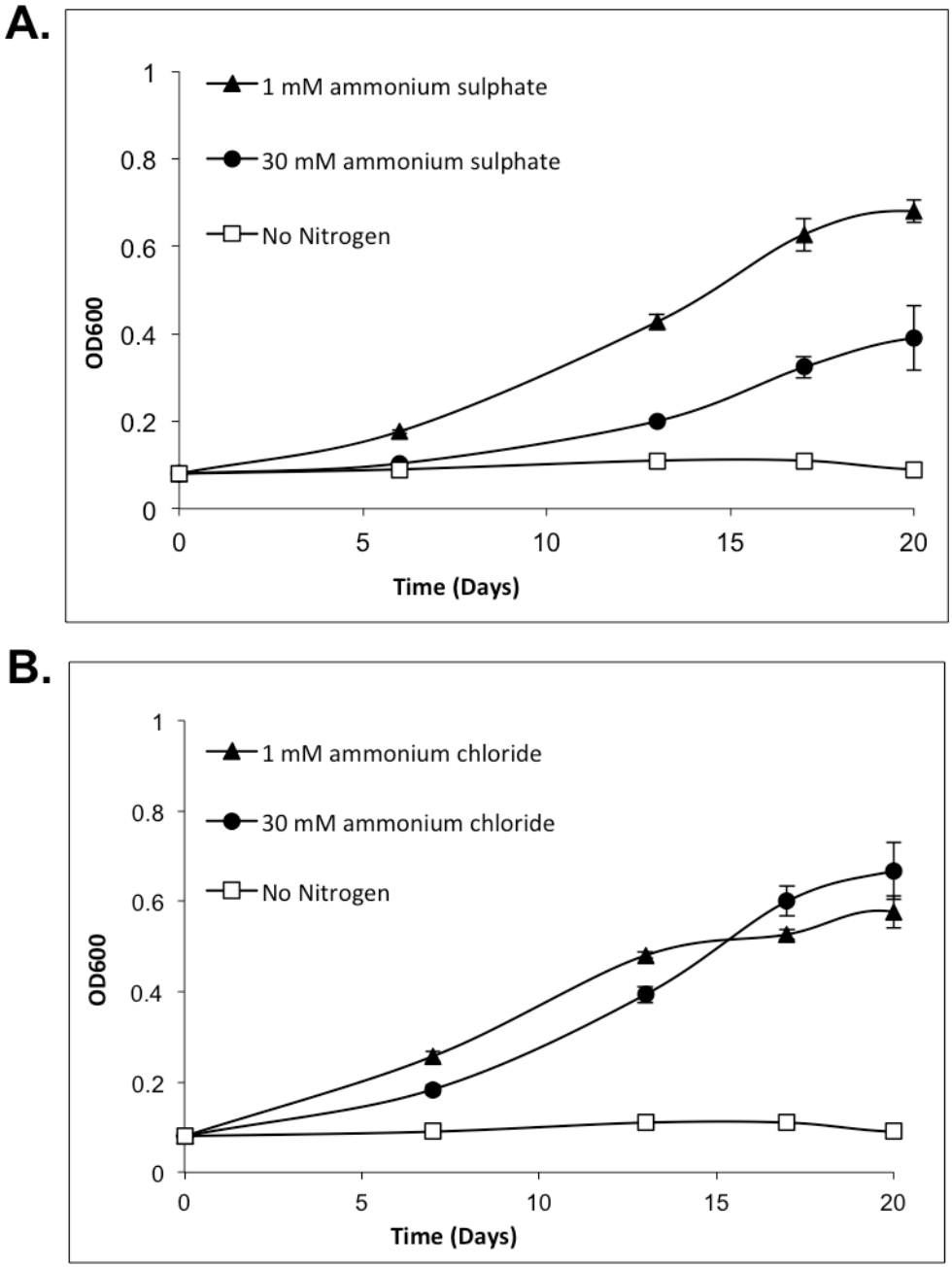


Figure 7.1. Growth of *M. tuberculosis* in (A) ammonium sulphate and (B) ammonium chloride (KW).

M. tuberculosis was grown in Sauton's minimal medium with either (A) ammonium sulphate or (B) ammonium chloride at concentrations of 1 mM (closed triangle), 30 mM (closed circle) or no nitrogen (open square). Growth was monitored by OD_{600}. Nitrogen run out in this experiment for the 1 mM ammonium chloride medium was determined by Aquaquant analysis to be at day 8. Data represents the average (±SD) of three independent experiments.

For ChIP *M. tuberculosis* cells were cross-linked one day after ammonium had run out in the limiting cultures, with nitrogen excess samples cross-linked at the same time point. Nitrogen run out in the 1 mM medium for this analysis corresponded to day 11, this varied from Figure 7.1 as 60 ml cultures were used rather than 30 ml to generate enough cells for ChIP analysis. After cross-linking, cells were lysed and the DNA sheared by sonication. *M. tuberculosis* cells required three rounds of sonication, compared to the one round used previously for *M. smegmatis*. The sonication conditions were optimised until all genomic DNA was sheared to between 200 and 500 bp fragments in size (Figure 7.2). The polyclonal GlnR antibody was purified against the *M. tuberculosis* His-GlnR protein as described in Section 2.4.3 and a Western blot conducted on *M. tuberculosis* whole cells lysate to confirm specific reactivity of the purified anti-GlnR antibody with GlnR (Figure 7.3). Immunoprecipitation of GlnR-bound DNA fragments was carried out using purified anti-GlnR (*M. tuberculosis*) polyclonal antibody, as described (Section 2.7.2).

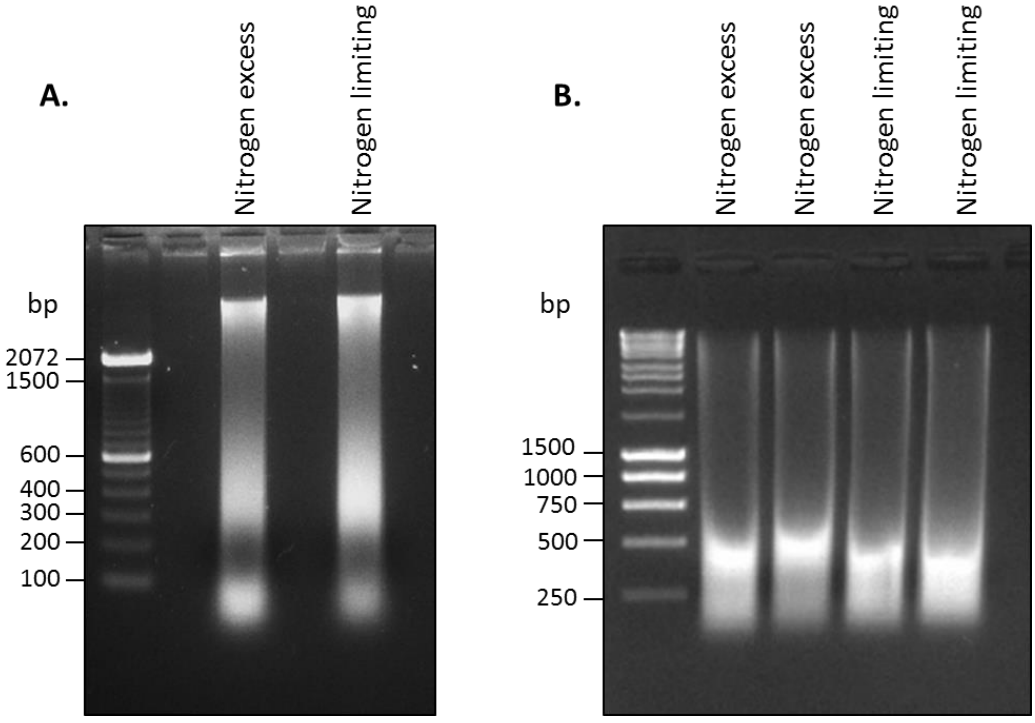


Figure 7.2. Sonication of *M. tuberculosis* DNA for ChIP-seq.

Two per cent agarose gel of *M. tuberculosis* DNA with (A) one round of sonication, as used for *M. smegmatis,* and (B) three rounds of sonication. Genomic DNA can still be visualised in (A), while in (B) all DNA is sheared to between 200 and 500 bp.

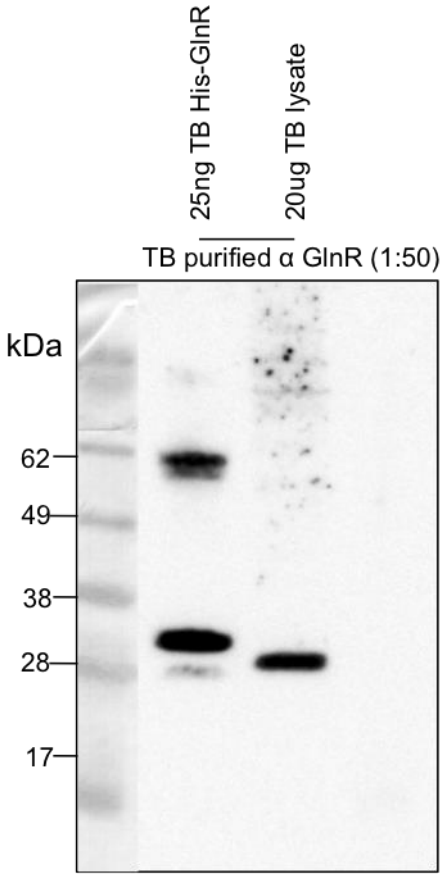


Figure 7.3 Western blot of *M. tuberculosis* whole cell lysate with *M. tuberculosis* purified anti-GlnR antibody.

Western blot demonstrating the specificity of the purified *M. tuberculosis* anti-GlnR antibody, with 25 ng of *M. tuberculosis* His-GlnR recombinant protein and 20 µg of *M. tuberculosis* whole cell lysate. Antibody was used at a 1:50 dilution. GlnR is visualised at 28 kDa (slightly higher for His-GlnR due to His tag). At 62 kDa a dimer of His-GlnR is visible.

To confirm that GlnR binding regions were enriched in the DNA fragments immunoprecipitated during nitrogen limitation compared to nitrogen excess, a rate limiting PCR was performed. One known GlnR regulated *M. tuberculosis* gene, the nitrite reductase (*nirB;* Rv0252) promoter region, was chosen for analysis. A gene not thought to be GlnR regulated (Rv1360) was included as a negative control (Figure 7.4). The *nirB* promoter region was enriched in nitrogen limiting conditions when compared to excess, indicating that GlnR was binding to the known site in our conditions.

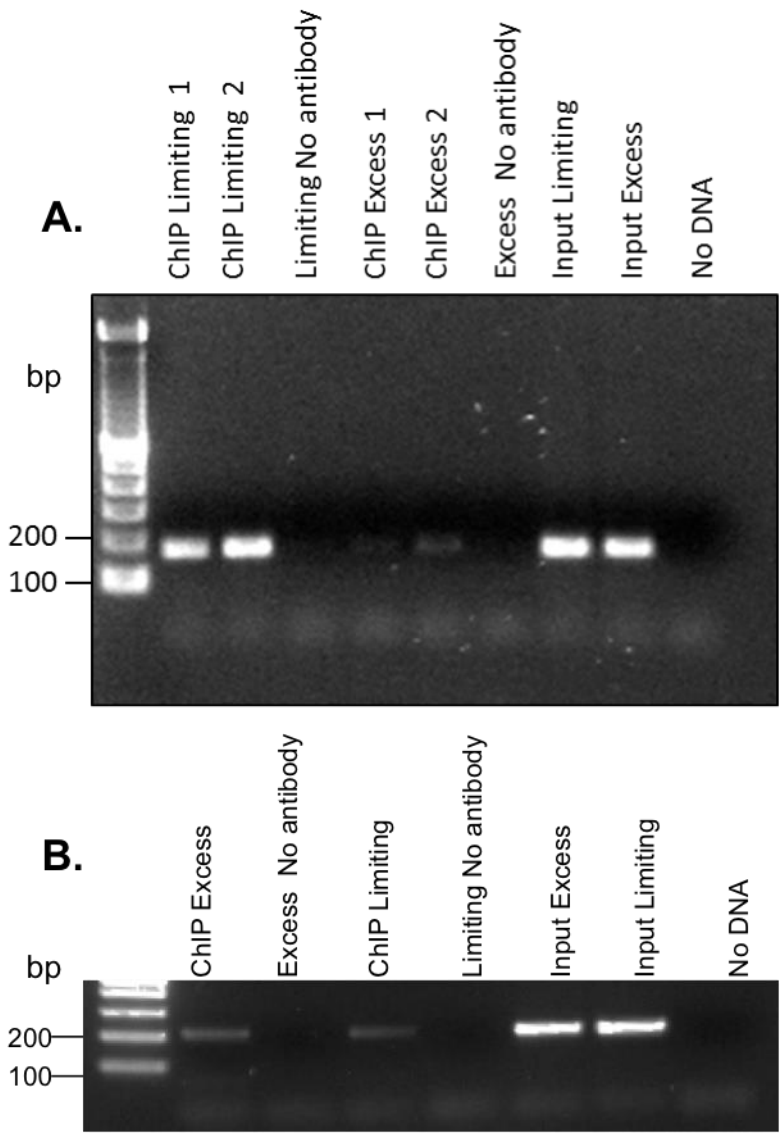


Figure 7.4. Rate limiting PCR indicating enrichment of GlnR immunoprecipitated DNA in *M. tuberculosis*.

Rate limiting PCR involved 23 cycles amplification of 200 bp promoter regions, using 0.3 ng of template DNA. ChIP represents DNA that has been immunoprecipitated with a GlnR specific antibody during nitrogen excess or nitrogen limiting conditions. Input represents the total amount of DNA that was subject to immunoprecipitation.

(A) Promoter region of *nirB* (Rv0252). (B) Promoter region of Rv1360 (negative control)

Immunoprecipitated DNA was prepared for sequencing using the Illumina ChIP-seq library kit according to the manufacturer's instructions. Sequencing of the DNA libraries using the Illumina HiSeq2000 generated approximately 144 million reads per sample which were mapped to the *M. tuberculosis* genome using Bowtie (81). GlnR binding regions were identified using the peak calling algorithm SISSRs (105). GlnR binding sites were defined as regions showing greater than 5-fold enrichment in the immunoprecipitated sample compared to the input control with p value of < 0.005. This analysis identified 36 GlnR binding sites during nitrogen limitation including 2 also observed in nitrogen excess (Figure 7.5, Tables 7.2 and 7.3). These binding regions and genes adjacent to the peaks are listed in Table 7.2. Eight of these GlnR binding sites were in similar regions to those described in the GlnR regulon for *M. smegmatis* (Table 7.2 highlighted in grey).

Unlike *M. smegmatis*, where all GlnR binding sites were identified in intergenic regions, in *M. tuberculosis* six GlnR binding sites were located within genes. Two examples, peaks 10 and 11, are displayed in Figure 7.6. Binding sites observed within gene transcription units are noted for peak 5 inside *umaA,* peak 10 inside *narG,* peak 11 binding inside Rv1173, peak 27 inside *moaA1,* peak 33 inside Rv3528 and peak 34 inside Rv3533. The role of GlnR binding within predicted genes is intriguing and further analysis of gene expression may indicate how GlnR is acting on these genes.

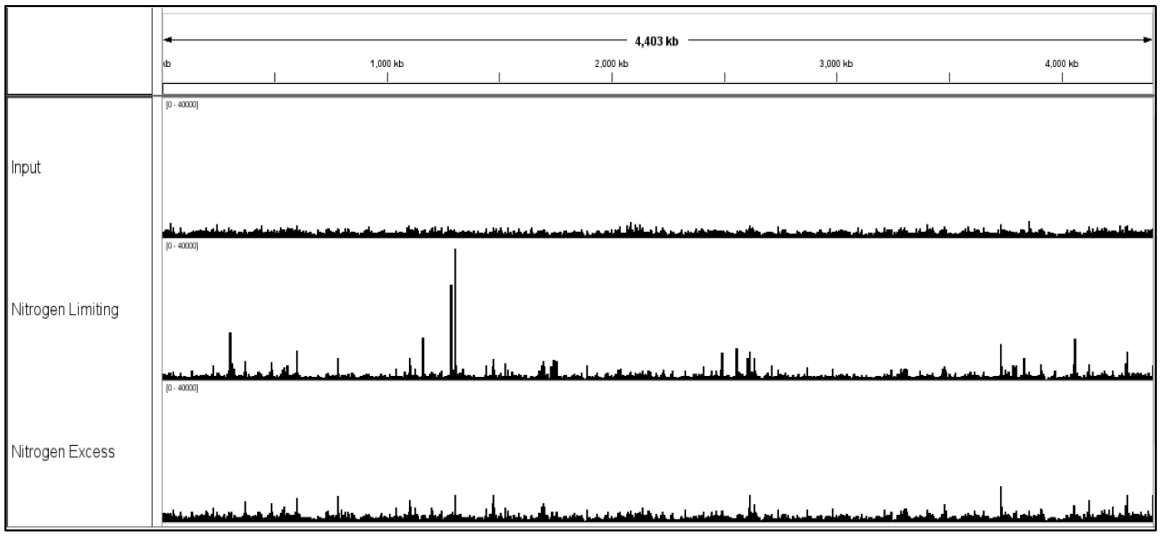
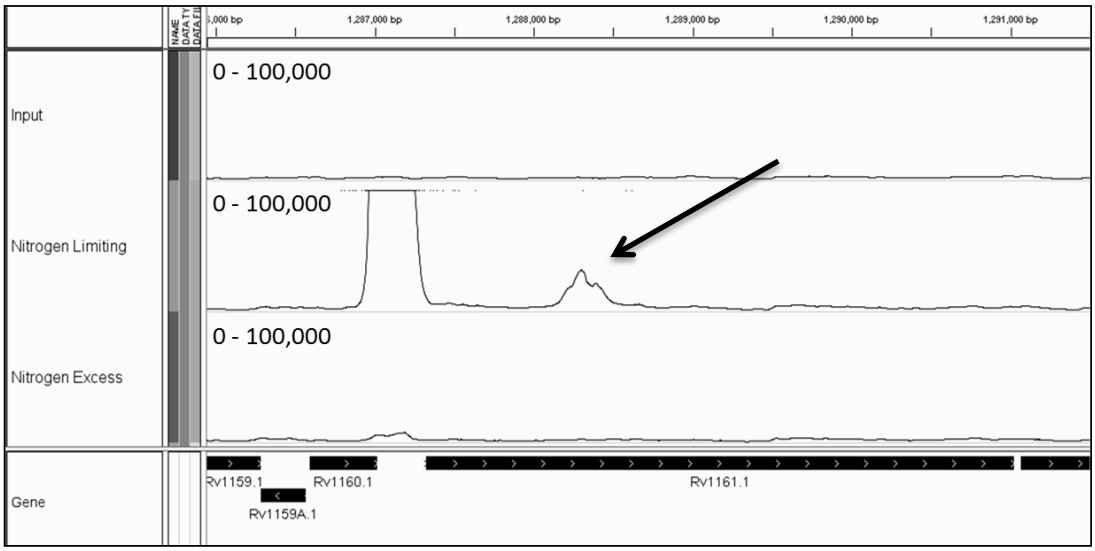


Figure 7.5. Whole genome view of GlnR binding sites in *M. tuberculosis* identified via ChIP-seq.

Whole genome view of GlnR binding sites identified by ChIP-seq in *M. tuberculosis* displayed in IGV. Upper track displays the input control (total DNA without immunoprecipitation), middle track nitrogen limiting (1 mM ammonium sulphate), bottom track is nitrogen excess (30 mM ammonium sulphate).

A.



B.

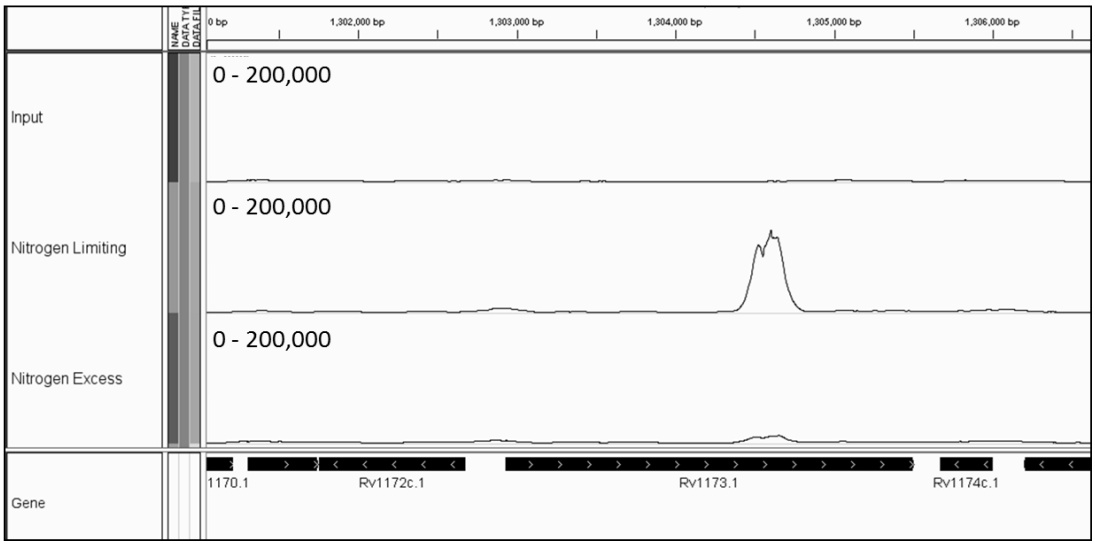


Figure 7.6. Two examples of GlnR binding sites identified within gene coding sequences in *M. tuberculosis*.

(A) Peak 10 within Rv1161 (highlighted by arrow). (B) Peak 11 within Rv1173.

The upper track represents the input control sample of total genomic DNA. The middle and bottom tracks represent sequencing of DNA immunoprecipitated from the nitrogen limiting and excess samples respectively. Horizontal black lines represent the gene transcription units.

Peak[a]	Coordinates[b]		Peak Intensity[c]	RIGHT GENE[d]				LEFT GENE[d]			
				Gene ID	Gene Name	Function	*M. smegmatis* homologue	Gene ID	Gene Name	Function	*M. smegmatis* homologue
1	302811	302851	54.53	Rv0251c	*hsp*	Heat shock protein	MSMEG_0424	Rv0252	*nirB*	Nitrate reductase	MSMEG_0427
2	312651	312691	25.91	Rv0260c	*nnaR*	Transcriptional regulatory protein	MSMEG_0432				
3	314191	314231	11.84	Rv0261c	*narK3*	Excretion of nitrate	MSMEG_0433				
4	559811	559851	11.21	Rv0469	*umaA*	Possible mycolic acid synthase	MSMEG_0913				
5	560371	560411	7.77	Inside Rv0469	*umaA*	Possible mycolic acid synthase	MSMEG_0913				
6	1077911	1077951	5.4	Rv0966c	*FUN*		MSMEG_5505	Rv0967	*csoR*	Copper-sensitive operon repressor of the CSO operon	N/A
7	1163531	1163571	50.83	Rv1040c	*PE8*	PE family protein	N/A				
8	1214391	1214431	5.08					Rv1088	*PE9*	PE family protein	N/A
9	1287091	1287131	60.55					Rv1161	*narG* *	Nitrate reduction	MSMEG_5140
10	1288291	1288331	7.14	inside Rv1161	*narG*	Nitrate reduction	MSMEG_5140				

Peak[a]	Coordinates[b]		Peak Intensity[c]	RIGHT GENE[d]				LEFT GENE[d]			
				Gene ID	Gene Name	Function	*M. smegmatis* homologue	Gene ID	Gene Name	Function	*M. smegmatis* homologue
11	1304571	1304611	34.76	Rv1172c	*PE12*	PE family protein	N/A	Inside Rv1173	*fbiC*	Probable F420 biosynthesis protein FBIC	MSMEG_5126
12	1471651	1471691	5.4	NOT A	PEAK	RIBOSOMAL	ISSUE				
13	1561351	1561391	12.06					Rv1386	*PE15*	PE family protein	MSMEG_0618
14	1728391	1728431	10.62	Rv1527c	*pks5*	Probable polyketide synthase	MSMEG_4727				
15	1728911	1728951	12.52					Rv1529	*fadD24*	Lipid degradation	MSMEG_4731
16	1735631	1735671	6.45	Rv1535		FUN					
17	1744831	1744871	30.01	Rv1542c	*glbN*	Oxygen transport	MSMEG_5765	Rv1543		Possible fatty acyl-coA reductase	N/A
18	1753431	1753471	30.04	Rv1548	*PPE21*	PPE family protein	N/A	Rv1549	*fadD11.1*	Possible lipid degradation	N/A
19	2029771	2029811	5.38	Rv1791	*PE19*	PE family protein	N/A				
20	2487551	2487591	9.56	Rv2219A		Conserved membrane protein	MSMEG_4288	Rv2220	*glnA1*	Glutamine synthase	MSMEG_4290
21	2493731	2493771	41.8	Rv2222c	*glnA2*	Glutamine synthase	MSMEG_4294				

Peak[a]	Coordinates[b]		Peak Intensity[c]	RIGHT GENE[d]				LEFT GENE[d]			
				Gene ID	Gene Name	Function	*M. smegmatis* homologue	Gene ID	Gene Name	Function	*M. smegmatis* homologue
22	2553111	2553151	47.26					Rv2281	*pitB*	Phosphate transport (putative)	N/A
23	2563071	2563111	9.15					Rv2291	*sseB*	Probable thiosulfate sulfurtransferase	MSMEG_3238
24	2603491	2603531	42.33	Rv2329c	*narK1*	Nitrate excretion	N/A				
25	2752931	2752971	12.79					Rv2425c		FUN	MSMEG_4582
26	3079111	3079151	6.27	Rv2769c	*PE27*	PE family protein	N/A				
27	3477871	3477911	6.29	inside Rv3109	*moaA1*	Probable molybdenum cofactor biosynthesis protein	N/A	Rv3110	*moaB1*	Probable molybdenum biosynthesis	N/A
28	3478231	3478271	5.47					Rv3110	*moaB1*	Probable molybdenum biosynthesis	N/A
29	3595491	3595531	6.07					Rv3219	*whiB1*	Transcriptional regulatory protein	MSMEG_1919
30	3784831	3784871	14.79	Rv3370c	*dnaE2*	Probable DNA polymerase III	MSMEG_1633	Rv3371		Possible triacylglycerol synthase	NOT in smeg

				RIGHT GENE[d]				LEFT GENE[d]			
Peak[a]	**Coordinates[b]**		**Peak Intensity[c]**	**Gene ID**	**Gene Name**	**Function**	***M. smegmatis* homologue**	**Gene ID**	**Gene Name**	**Function**	***M. smegmatis* homologue**
31	3799991	3800031	15.45	Rv3385c	*vapB46*	Possible antitoxin VAPB46	N/A	Rv3386		Possible transposase	NOT in smeg
32	3834631	3834671	26.47	Rv3415c		FUN	MSMEG_1598	Rv3416	*whiB3*	Transcriptional regulatory protein	MSMEG_1597
33	3964871	3964911	5.15	inside Rv3528c		FUN	N/A				
34	3971191	3971231	8.71	inside Rv3533c	*PPE62*	PPE family protein	N/A				
35	4060611	4060651	8.02	Rv3620c	*esxW*	Putative ESAT-6 like protein ESXW	N/A				
36	4062351	4062391	48.07	Rv3622c	*PE32*	PE family protein	N/A	Rv3623	*lpqG*	Probable conserved lipoprotein	MSMEG_6109

Table 7.2. GlnR binding sites and adjacent genes identified in *M. tuberculosis* during nitrogen limitation.

[a]Table shows the enriched peak binding regions, with the [b]coordinates on the *M. tuberculosis* genome. [c]Peak intensity corresponds to fold enrichment of each peak calculated using SISSRs, and is based on the number of sequenced tags at each site vs the input control sample. [d]Right gene/left gene indicates the direction of the gene in relation to GlnR binding. Genes highlighted in grey displayed GlnR regulated gene expression in *M. smegmatis*. The hashed line through Rv3219 indicates a corresponding GlnR binding site in *M. smegmatis*, but no change in gene expression of the WT vs the *glnR* deletion mutant.

				RIGHT GENE				LEFT GENE			
Peak[a]	**Coordinates[b]**		**Peak Intensity[c]**	**Gene ID**	**Gene Name**	**Function**	***M. smegmatis* homologue**	**Gene ID**	**Gene Name**	**Function**	***M. smegmatis* homologue**
1	1753431	1753471	8.54	Rv1548	*PPE21*	PPE family protein	N/A	Rv1549	*fadD11.1*	Possible lipid degradation	N/A
2	2493751	2493791	6.96	Rv2222c	*glnA2*	Glutamine synthase	MSMEG_4294				

Table 7.3. GlnR binding site identified in *M. tuberculosis* during nitrogen excess.

[a]Table shows the enriched peak binding regions, with the [b]coordinates on the *M. tuberculosis* genome. [c]Peak intensity corresponds to fold enrichment of each peak calculated using SISSRs, and is based on the number of sequenced tags at each site vs the input control sample. Right gene/left gene indicates the direction of the gene in relation to GlnR binding.

7.3.2 Confirmation of GlnR Binding Sites

Identification of the GlnR binding site for the previously known *nirB* promoter region in our analysis confirmed that ChIP-seq had identified specific GlnR binding sequences (Figure 7.7). Further validation was provided by performing electromobility shift assays (EMSA). Two hundred base pair DNA regions, representing the binding regions from 4 novel binding sites, were incubated with purified *M. tuberculosis* GlnR protein. Using the same conditions as in the *M. smegmatis* EMSAs (this study), peak 18 was confirmed as a GlnR binding site, whilst the negative control Rv1360 displayed no GlnR:DNA binding (Figure 7.8).

Problems with the migration of protein-DNA complexes made identification of the 3 other binding sites via EMSA difficult (Figure 7.9). The GlnR:DNA complex appeared to be aggregating in the well rather than migrating through the gel. Reducing the percentage of acrylamide within the gel from 6% to 4%, permitted migration of the protein-DNA complex for peak 13 into the gel (Figure 7.8). Confirming peak 13 as a specific GlnR binding site. However, the two additional peaks (17 and 20) could not be confirmed by EMSA using the *M. tuberculosis* GlnR protein due to the aggregation issue.

As the DNA binding region of the *M. smegmatis* and *M. tuberculosis* GlnR proteins are similar in sequence and predicted structure (Figure 7.10), binding analysis of these 2 *M. tuberculosis* DNA binding regions was conducted with the *M. smegmatis* purified GlnR protein. Peak 18 region bound specifically to the *M. smegmatis* GlnR protein (data not shown), as observed with the *M. tuberculosis* GlnR protein. Peaks 13, 17 and 20 also bound specifically with the *M. smegmatis* His-GlnR protein (Figure 7.11). Again no GlnR binding was observed for the negative control Rv1360 (Figure 7.11). This confirms that GlnR binds specifically to these promoter regions with a protein-concentration dependent shift.

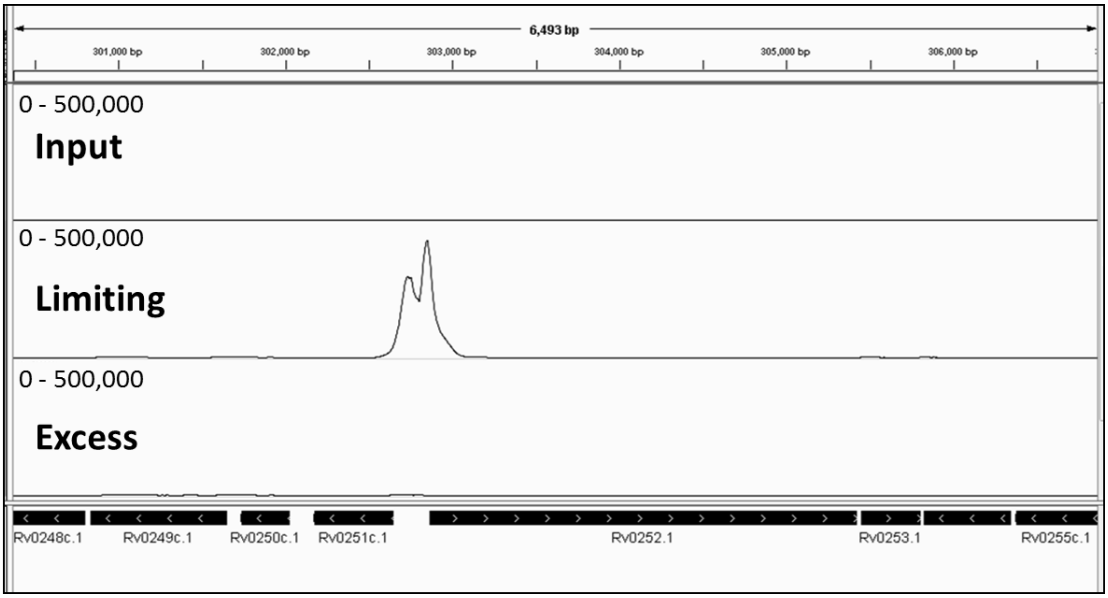


Figure 7.7. Confirmation that *M. tuberculosis* ChIP-seq data identified known GlnR binding site, the *nirB* promoter region.

Binding data for known GlnR-regulated gene *nirB* was visualised using IGV. Upper track indicates total DNA input with the middle and bottom tracks corresponding to ChIP-seq data from nitrogen limiting and excess conditions respectively. The black bars at the bottom signify gene transcripts.

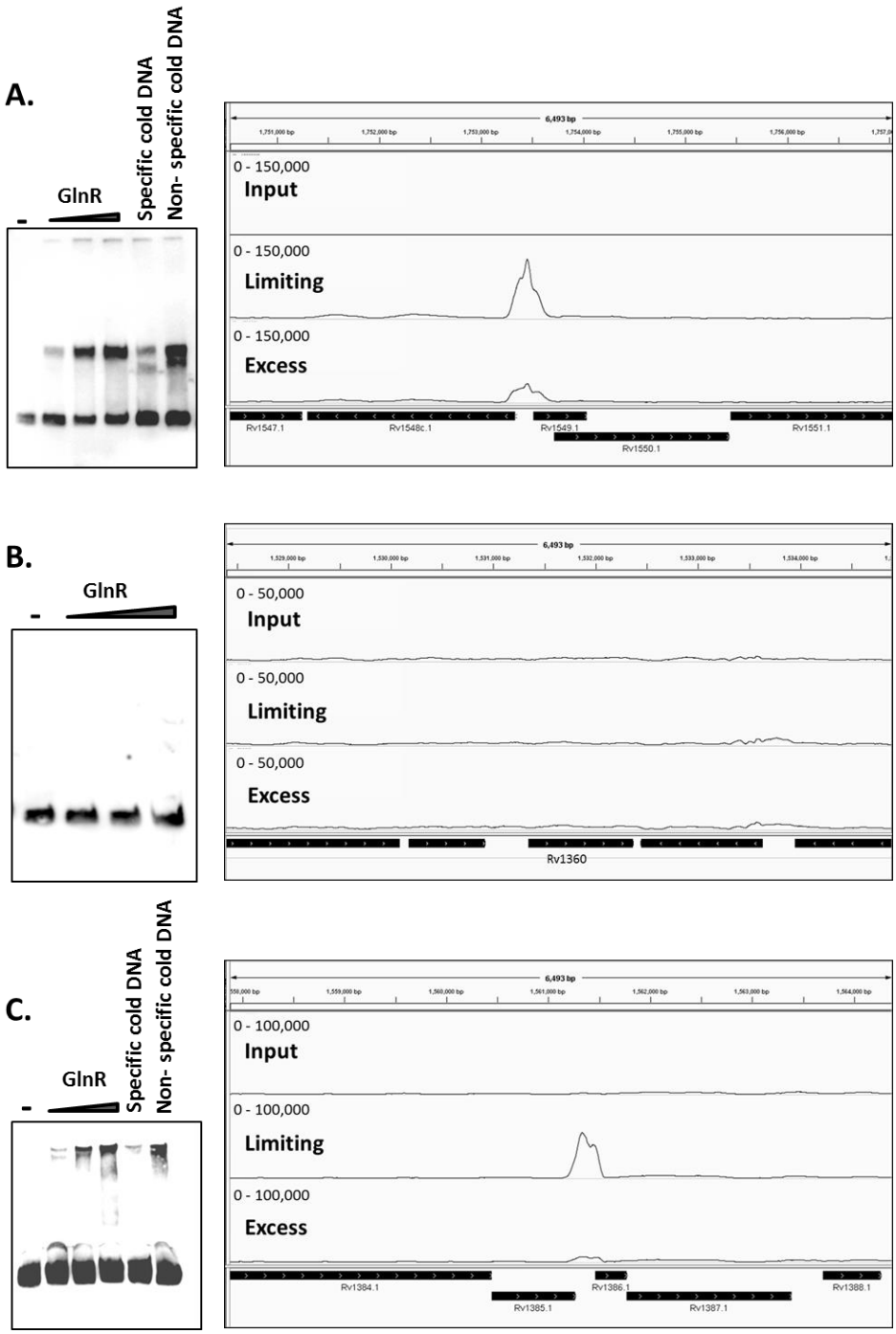


Figure 7.8. EMSA of GlnR binding to 200 bp DNA region and corresponding IGV sequence alignment of: (A) Peak 18, (B) Rv1360 negative control, and (C) Peak 13.

M. tuberculosis His-GlnR was incubated in increasing amounts, 0-0.9 µg, with 200 bp region of labelled DNA. Specific and non-specific competitor cold DNA was added at 1000x excess to labelled probe. Binding data was visualised using IGV. Upper track indicates total DNA sequenced with middle and bottom track displaying ChIP-seq data from nitrogen limiting (1 mM) and nitrogen excess (30 mM) conditions respectively. The black bars at the bottom signifying genes are labelled accordingly.

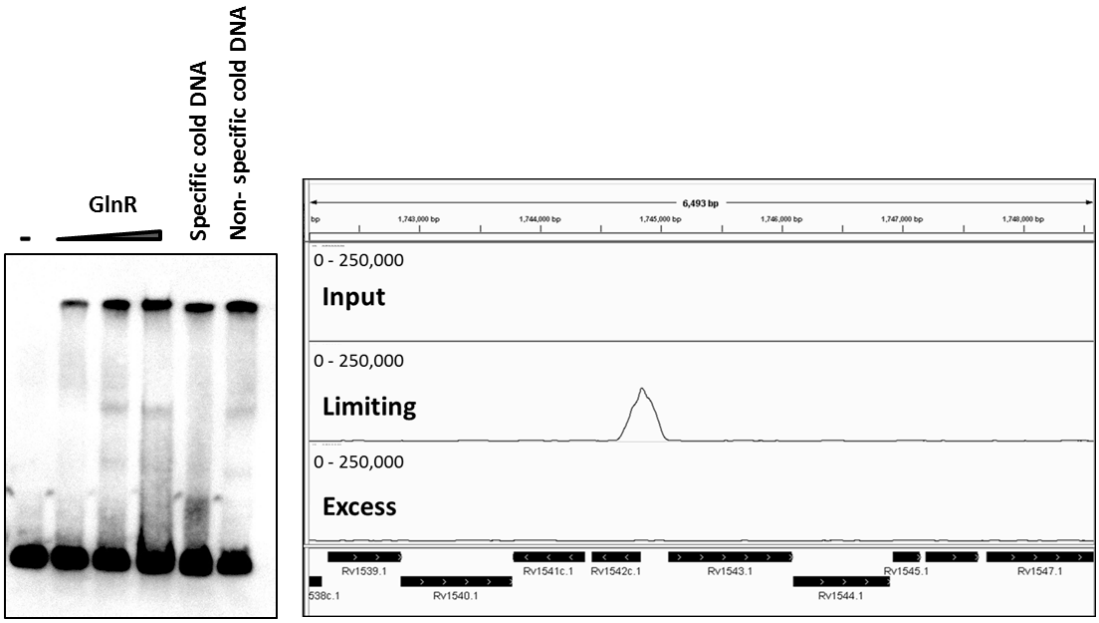


Figure 7.9. Purified *M. tuberculosis* His-GlnR:DNA complexes do not migrate during EMSA analysis.

M. tuberculosis His-GlnR was incubated in increasing amounts, 0-0.9 µg, with 200 bp region of labelled DNA. Specific and non-specific competitor cold DNA was added at 1000x excess to labelled probe. Binding data was visualised using IGV. Upper track indicates total DNA sequenced with middle and bottom track displaying ChIP-seq data from nitrogen limiting (1 mM) and nitrogen excess (30 mM) conditions respectively. The black bars at the bottom signifying genes are labelled accordingly.

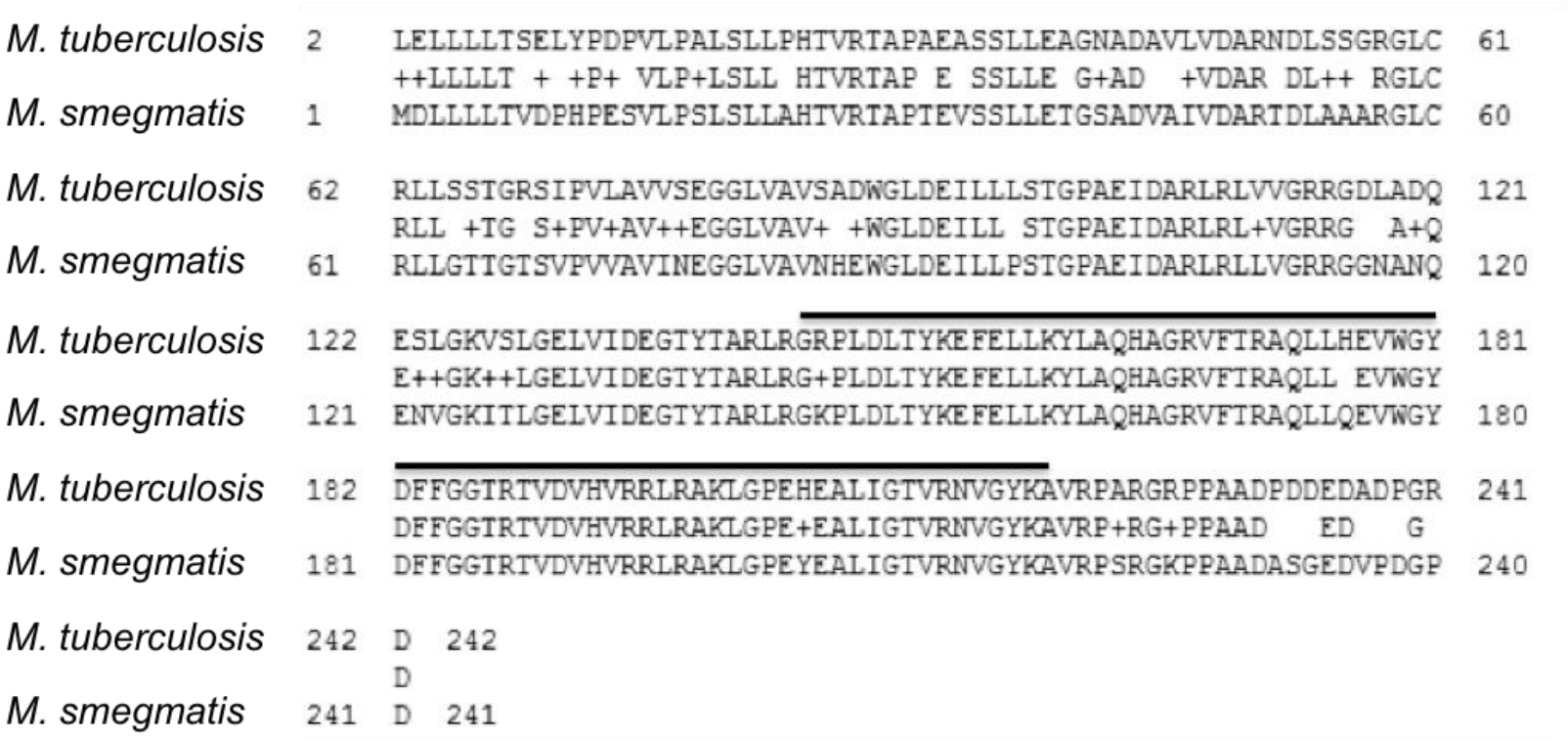


Figure 7.10. Sequence alignment of *M. tuberculosis* and *M. smegmatis* GlnR proteins.

The C-terminal DNA binding domain (helix turn helix) is located between amino acid residues 144 to 218 (indicated by a bar above the sequence) (adapted from (5)). The two DNA binding domains for *M. smegmatis* and *M. tuberculosis* are highly similar, suggesting conservation of DNA:protein interaction.

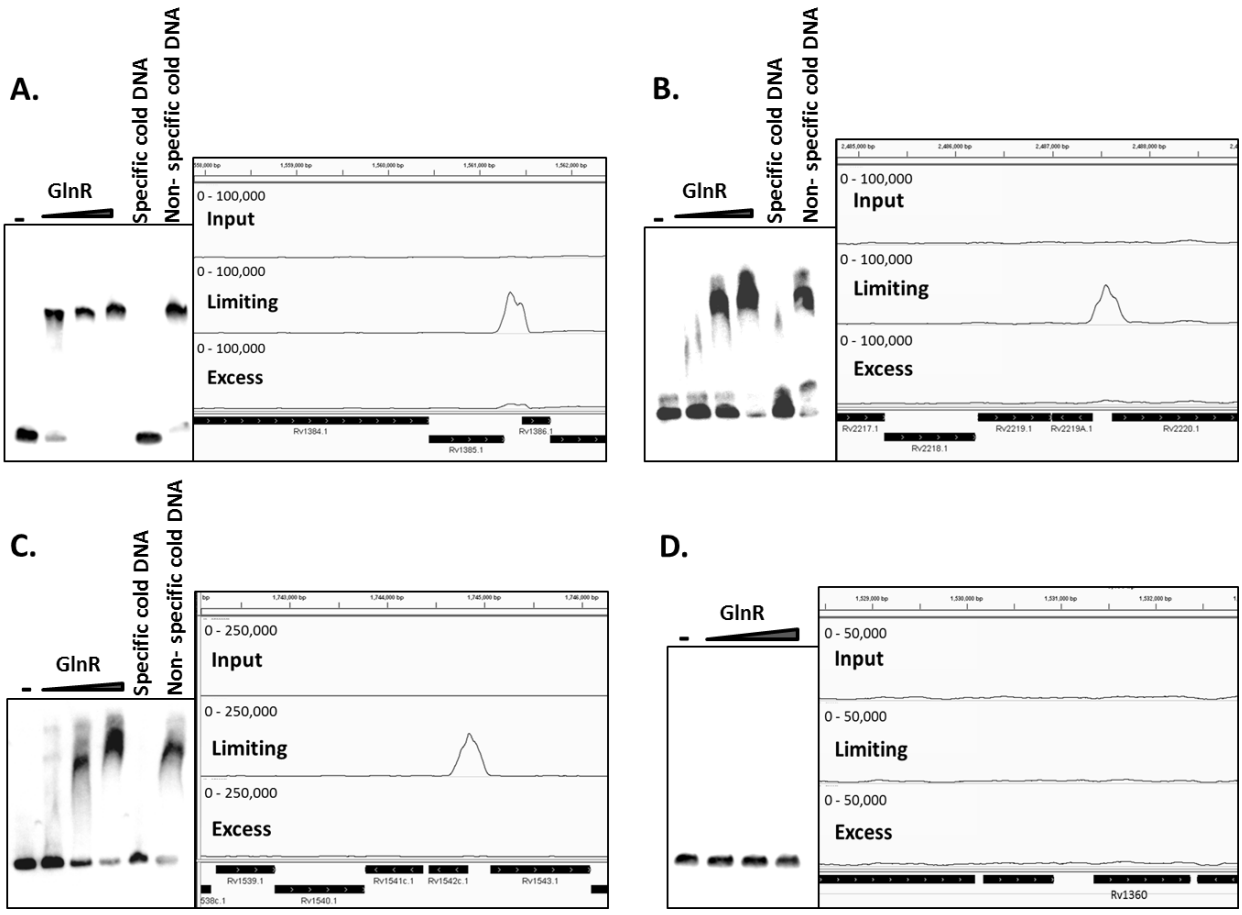


Figure 7.11. *M. smegmatis* GlnR binds to *M. tuberculosis* DNA sequences.

M. smegmatis His-GlnR was incubated in increasing amounts, 0-0.9 µg, with 200 bp region of labelled DNA. Specific and non-specific competitor cold DNA was added at 1000x excess to labelled probe. Binding data was visualised in IGV. Upper track indicates total DNA sequenced with middle and bottom track displaying ChIP-seq data from nitrogen limiting (1 mM) and nitrogen excess (30 mM) conditions respectively. The black bars at the bottom signifying genes are labelled accordingly.

(A) Peak 13 (B) Peak 17 (C) Peak 20 (D) Rv1360 (negative control).

7.3.3 Identification and Analysis of the *M. tuberculosis* GlnR DNA Binding Motif

To identify potential GlnR DNA binding motifs within the nucleotide sequences obtained by ChIP-seq, 200 bp regions centred around the peak identified by SSISRs were submitted to MEME. This generated a consensus motif sequence displayed in Figure 7.12 with an E-value $7.3x10^{-3}$. However, this consensus motif generated was only found in 18 of the 36 GlnR binding sites identified via ChIP-seq, with 3 sequences located in intergenic regions; these peaks and the neighbouring genes are listed in Table 7.4.

In order to ascertain a more comprehensive motif search, sequences of 50 bp representing the GlnR DNA binding site was submitted to MEME. This approach identified 6 different consensus sequences (Figure 7.13). Motif 1 was identified to be the most common motif, located in 10 peaks, with motif 2 located in 9 peaks. Motifs 3, 4, 5 and 6 were all located in two peaks each. Interestingly peak 33 was the only region to contain two binding motifs, 2 and 6. The motif number and genes these controlled are listed in Table 7.5.

As the consensus DNA-binding motif varied between GlnR binding sites in *M. tuberculosis* we sought further verification that the peaks identified via ChIP-seq were due to enrichment of GlnR immunoprecipitated DNA and not due to sequencing error. As such, 4 peaks were chosen for further analysis; peaks 2, 10, 11 and 23. Peak 2 and peak 23 were chosen as no consensus DNA binding motif was detected in these regions. To confirm that GlnR binding sites detected inside genes were true binding sites, peak 10 and 11 were chosen for analysis. Peak 10 bound inside *narG* and had a consensus motif from the second 50 bp MEME search of motif 5, with peak 11 binding inside *fbiC* and containing motif 1 from the 50 bp MEME search. Rate limiting PCR was conducted on independent immunoprecipitated DNA in nitrogen excess and limiting conditions. Figure 7.14 displays that enrichment of DNA is seen in all samples tested in nitrogen limitation, compared to nitrogen excess for the peak regions tested. This confirms that the Peaks identified via ChIP-seq are due to GlnR binding and enrichment.

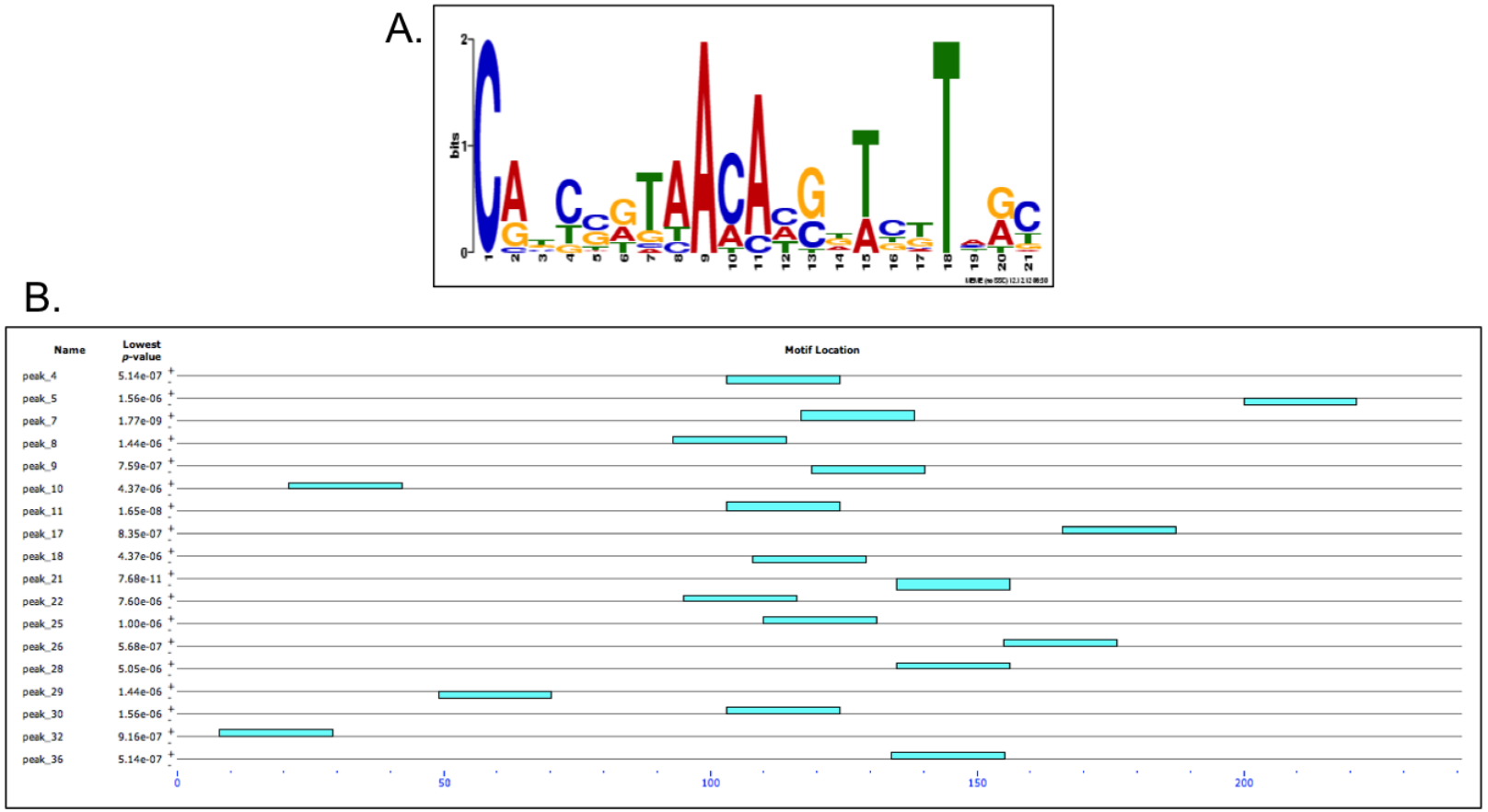


Figure 7.12. MEME-derived GlnR consensus binding sequence from ChIP-seq data.

(A) MEME consensus of GlnR binding regions generated from 200 bp regions surrounding the peak.

(B) Location of MEME generated motif (blue bars) in relation to DNA strand.

			Left Gene		Right Gene	
Peak[a]	**p-value[b]**	**Peak intensity[c]**	**Gene ID**	**Gene Name**	**Gene ID**	**Gene Name**
4	5.1E-07	11.21	Rv0469	*umaA*		
5	1.6E-06	7.77	Inside Rv0469	*umaA*		
7	1.8E-09	50.83	Rv1040c	*PE8*		
8	1.4E-06	5.08			Rv1088	*PE9*
9	7.6E-07	60.55			Rv1161	*narG*
10	4.4E-06	7.14	Inside Rv1161	*narG*		
11	1.7E-08	34.76	Rv1172c	*PE12*	Inside Rv1173	*fbiC*
17	8.4E-07	30.01	Rv1542c	*glbN*	Rv1543	
18	4.4E-06	30.04	Rv1548	*PPE21*	Rv1549	*fadD11.1*
21	7.7E-11	41.8	Rv2222c	*glnA2*		
22	7.6E-06	47.26			Rv2281	*pitB*
25	1.0E-06	12.79			Rv2425c	
26	5.7E-07	6.27	Rv2769c	*PE27*		
28	5.1E-06	5.47			Rv3110	*moaB1*
29	1.4E-06	6.07	Rv3219	*whiB1*		
30	1.6E-06	14.79	Rv3370c	*dnaE2*	RV3371	
32	9.2E-07	26.47	Rv3415c		Rv3416	*whiB3*
36	5.1E-07	48.07	Rv3622c	*PE32*	Rv3623	*IpqG*

Table 7.4. Peaks containing the GlnR DNA binding motif generated by MEME.

All [a]peaks identified via ChIP-seq containing the MEME motif generated in Figure 7.11 are displayed in the table. [b]P value is generated via MEME and relates to the similarity of each site with the MEME derived sequence. [c]Peak intensity corresponds to fold enrichment of each peak calculated using SISSRs, and is based on the number of sequenced tags at each site vs the input control sample. Right gene/left gene indicates the direction of the gene in relation to GlnR binding. Genes highlighted in grey displayed GlnR regulated gene expression in *M. smegmatis.*

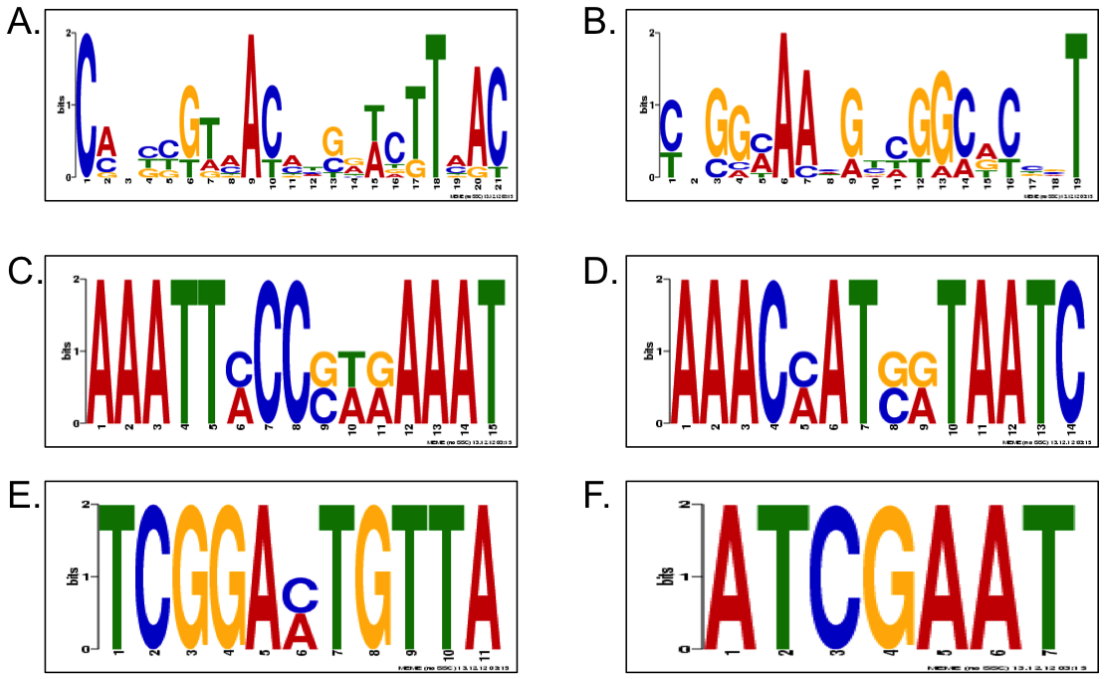


Figure 7.13. Alternate MEME-derived GlnR consensus binding sequences from ChIP-seq data.

MEME consensus of GlnR binding regions generated from 50 bp regions surrounding the peak.

(A) Motif 1 (B) Motif 2 (C) Motif 3 (D) Motif 4 (E) Motif 5 (F) Motif 6.

Peak[a]	p-value[b]	Motif[c]	Peak intensity[d]	Left Gene		Right Gene	
				Gene ID	*Gene Name*	Gene ID	*Gene Name*
1	3.4E-05	3	54.53	Rv0251c	*hsp*	Rv0252	*nirB*
3	9.3E-03	2	11.84	Rv0261c	*narK3*		
4	7.6E-05	1	11.21	Rv0469	*umaA*		
7	7.0E-03	1	50.83	Rv1040c	*PE8*		
8	6.7E-05	4	5.08			Rv1088	*PE9*
9	1.5E-04	4	60.55			Rv1161	*narG*
10	3.7E-04	5	7.14	Inside Rv1161	*narG*		
11	2.8E-06	1	34.76	Rv1172c	*PE12*	Inside Rv1173	*fbiC*
13	3.0E-03	1	12.06			Rv1386	*PE15*
14	4.7E-02	2	10.62	Rv1527c	*pks5*		
15	1.0E-04	1	12.52			Rv1529	*fadD24*
17	3.9E-03	2	30.01	Rv1542c	*glbN*	Rv1543	
18	8.4E-04	1	30.04	Rv1548	*PPE21*	RV1549	*fadD11.1*
19	4.9E-03	1	5.38	Rv1791	*PE19*		
20	1.2E-03	2	9.56	Rv2219A		RV2220	*glnA1*
21	6.1E-03	2	41.8	Rv2222c	*glnA2*		
24	1.5E-02	2	42.33	Rv2329c	*narK1*		
25	6.7E-04	1	12.79			Rv2425c	*FUN*
27	1.6E-03	1	6.29	Inside Rv3109	*moaA1*	Rv3110	*moaB1*
28	1.0E-04	5	5.47			Rv3110	*moaB1*
30	6.7E-04	1	14.79	Rv3370c	*dnaE2*	Rv3371	
31	4.6E-03	2	15.45	Rv3385c	*vapB46*	Rv3386	
33	3.5E-04	6 & 2	5.15	Inside Rv3528c	*FUN*		
34	1.3E-02	2	8.71	Inside Rv3533c	*PPE62*		
35	1.3E-01	6	8.02	Rv3620c	*esxW*		

36	8.2E-05	3	48.07	Rv3622c	*PE32*	Rv3623	*IpqG*

Table 7.5. Peaks containing the various GlnR DNA binding motifs generated by MEME.

All [a]peaks identified via ChIP-seq containing the [c]different MEME motifs generated in Figure 7.13 are displayed in the table. [b]P value is generated via MEME and relates to the similarity of each site with the MEME derived sequence. [d]Peak intensity corresponds to fold enrichment of each peak calculated using SISSRs, and is based on the number of sequenced tags at each site vs the input control sample. Right gene/left gene indicates the direction of the gene in relation to GlnR binding. Genes highlighted in grey displayed GlnR regulated gene expression in *M. smegmatis.*

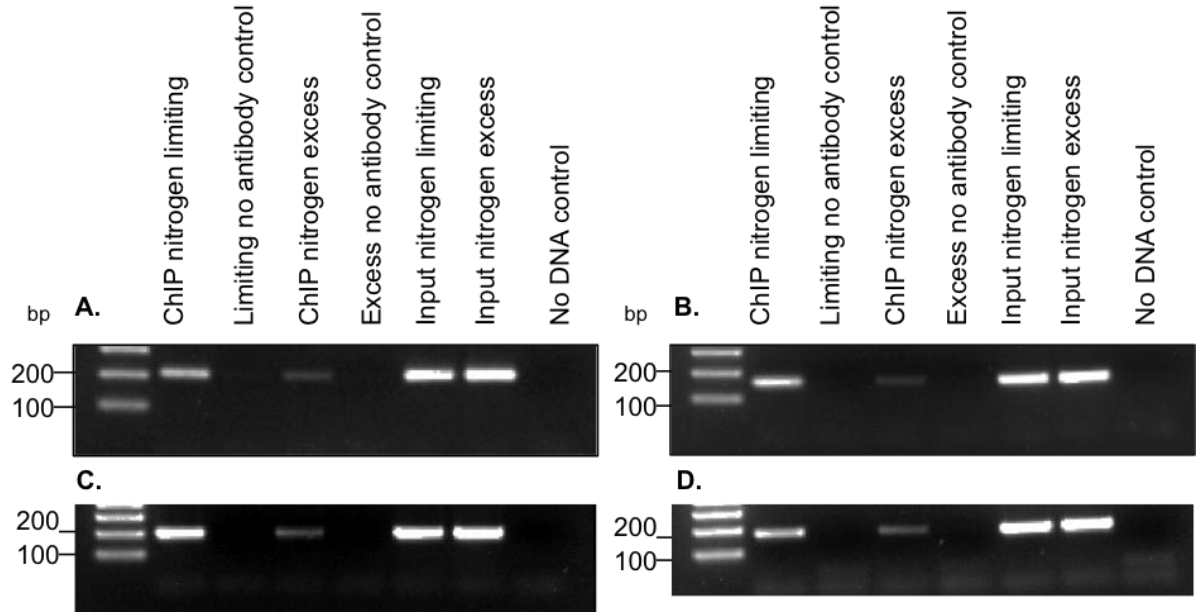


Figure 7.14. Rate limiting PCR on *M. tuberculosis* GlnR immunoprecipitated DNA.

Rate limiting PCR involved 23 cycles of 200 bp promoter regions, using 0.3 ng of template DNA. (A) Peak 2 (B) Peak 10 (C) Peak 11 (D) Peak 23

7.4 Discussion

The aim of this study was to identify GlnR binding sites in *M. tuberculosis* during nitrogen limitation and to compare any conserved GlnR-regulated nitrogen pathways in *M. smegmatis* and *M. tuberculosis*. In addition the identification of novel or unique GlnR binding regions in *M. tuberculosis* could highlight novel mechanisms associated with the establishment of infection in the host.

Unlike *M. smegmatis*, where GlnR binding sites were only identified in intergenic regions, in *M. tuberculosis* six GlnR binding sites were located within genes. The role of GlnR binding within gene coding sequences is intriguing. Three of these sites, namely peak 5 inside *umaA*, peak 10 inside *narG* and peak 27 *moaA1* contained additional GlnR binding sites adjacent to these peaks in gene promoter regions, which may indicate multiple binding of GlnR regulating gene transcription. Peak 11 binds inside Rv1173, however this binding site may regulate a gene adjacent to this location Rv1172, and this could be confirmed via microarray analysis of a *glnR* deletion strain. Peak 33 and peak 34 again were located within gene transcription units and it is not clear from the binding data alone whether GlnR binding influences activation or repression of transcription. As such further analysis of gene expression may indicate how GlnR is acting on these genes.

M. tuberculosis GlnR binding sites were identified upstream of genes also identified in the *M. smegmatis* as GlnR regulon. These include genes involved in nitrite metabolism, *nirB*, *nnaR* and *narK3* (discussed later), in addition to *glnA1* and *glnA2* encoding type-I glutamine synthetases. Interestingly GlnN a globular protein involved in O_2 transport, and the transcriptional regulators WhiB1 and WhiB3 also contain GlnR binding sites upstream of the gene start sites in both species. Of note, *whiB1* in *M. smegmatis* has a GlnR binding site yet no differential expression was observed compared to a *glnR* deletion mutant during nitrogen limitation. The regulation of other transcriptional regulators is intriguing, and extends the possible number of genes and processes controlled by GlnR and potentially links different regulons.

Nitrogen metabolism genes *glnA1* and *glnA2,* encode type-I glutamine synthetases, and contain GlnR binding sites in both *M. smegmatis* and *M. tuberculosis*. GlnR also regulates *glnA* in *Streptomyces*, indicating some conservation of GlnR binding within the Actinomycetes (120, 165). GlnR regulation of GlnA1 in *M. tuberculosis* is interesting, as GlnA1 appears to play a role in virulence as well as nitrogen metabolism. In contrast to *M. smegmatis*, GlnA1 of *M. tuberculosis* is secreted extracellularly and is thought to play an important role in the biosynthesis of the cell wall polymer poly L-glutamate/glutamine (60, 171). In addition, in human mononuclear phagocytic cells (THP-1) infected with *M. tuberculosis,* the bacilli displayed

reduce survival when treated with L-methionine-S-sulfoximine, an inhibitor of the glutamine synthase enzyme (171). As such the effect of GlnR binding on the transcription of *glnA1* in *M. tuberculosis* is intriguing, highlighting a conserved binding site between mycobacterial species, but with a potentially adapted role in pathogenicity.

The majority of genes containing GlnR binding sites in *M. tuberculosis* were different to those described in the *M. smegmatis* GlnR regulon (Figure 7.15). Genes with known homologues in *M. smegmatis,* that are not GlnR regulated in this species, contain a GlnR binding site in *M. tuberculosis,* for instance *narG, psk* and *umaA*. These are interesting as the proteins may serve different roles in the two species during nitrogen limitation since they have a different control mechanism. In addition genes containing no known homologues in *M. smegmatis* contained GlnR binding sites, for example the PE/PPE and ESAT-6 family proteins. This may indicate a novel role for GlnR in the pathogenic species.

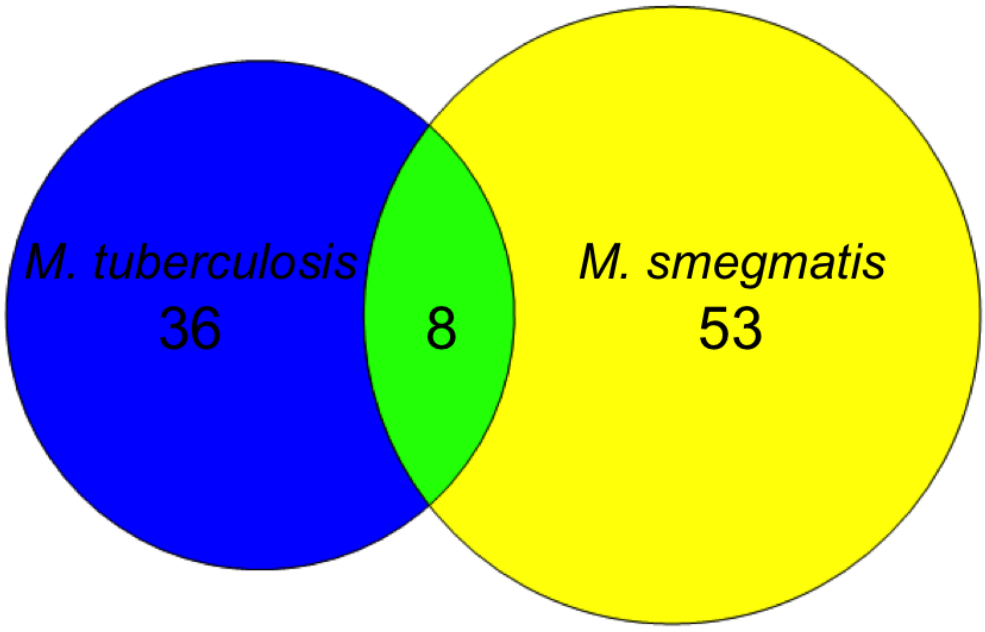


Figure 7.15. Venn diagram of GlnR binding sites in *M. smegmatis* and *M. tuberculosis* displaying the common and unique GlnR binding sites for each species.

A Venn diagram representing binding sites for *M. smegmatis* (yellow) and *M. tuberculosis* (blue). Common GlnR binding sites shared by the two organisms are indicated in green.

PE/PPE Proteins

An interesting category of genes are those that are under GlnR regulation in *M. tuberculosis* without direct homologues in *M. smegmatis*. These include a variety of PPE and PE family proteins. These were designated the *pe* and *ppe* genes, after highly conserved Proline-Glutamate and Proline-Proline-Glutamate residues near the start of their encoded proteins (138). It has been speculated that these proteins play a role in evasion of host immune responses, possibly via antigenic variation. Although a comprehensive understanding of their function has yet to be established, emerging data increasingly supports a role for the PE/PPE proteins at multiple levels of infection. It is known, however, that these genes are unique to mycobacteria and are particularly abundant in pathogenic mycobacteria, such as *M. tuberculosis* (138).

It is probable that GlnR regulates at least six PE and PPE proteins (peaks 7, 8, 13, 19, 26 and 34). However this number may increase; the direction of GlnR action at peaks 11, 18 and 36 is uncertain as genes are divergent from GlnR DNA binding sites and therefore GlnR could regulate additional *pe/ppe* genes. The high number of PE and PPE genes under GlnR control may not be surprising, analysis of the *M. tuberculosis* H37Rv genome revealed PE and PPE proteins comprise almost 10% of the genome coding capacity (36). In addition, regulation of a subset of PE and PPE proteins has been described during characterisation of other regulatory proteins (138, 180). In one example, disruption of the PhoPR two-component regulator resulted in altered expression of at least 14 *pe*/*ppe* genes (180). Of note all *pe*/*ppe* genes identified in the PhoPR study were different to those identified in this analysis. This could indicate that GlnR, as well as other two-component systems, provides regulation of these virulence factors in *M. tuberculosis.*

The GlnR regulated PE/PPE proteins identified in this study all have unknown functions. Table 7.6 summarises known characteristics of each protein. Cell-mediated immunity is important in the control of *M. tuberculosis* infection, so the role of PE and PPE proteins in eliciting this are of particular interest. At least 20 PE/PPE proteins have been reported to elicit CD4 and/or CD8 responses. One potentially GlnR regulated protein PE19 (Rv1791) has been reported previously to elicit a T-cell response (115). Other interesting characteristics are their cellular localisation, PE15 (Rv1386) is cell-membrane-associated and has been identified in exported fractions (93, 94). Comparing transcriptome data on the *pe*/*ppe* genes and aligning this to GlnR binding would provide a valuable way to investigate the regulation of these PE/PPE proteins and how nutrient availability, in this instance nitrogen, effects their expression.

PE/ PPE No[a]	Gene ID	Peak[b]	Peak intensity[c]	pe/ ppe pair[d]	Esx cluster[e]	Sub-group[f]	Surface associated	Exported	T cell response
PE8	Rv1040c	7	50.83	Yes	Yes	IV			
PE9	Rv1088	8	5.08	?	?	V			
PE12	Rv1172c	11	34.76	?	?	V			
PE15	Rv1386	13	12.06	Yes	No	II	Yes (94)	Yes (93)	
PE19	Rv1791	19	5.38	Yes	Yes	IV			Yes (115)
PE27	Rv2769c	26	6.27	Yes	No	IV			
PE32	Rv3622c	36	48.07	Yes	Yes	IV			
PPE21	Rv1548	18	30.04	No	No	V			
PPE62	RV3533	34	8.71	No	No	V			

Table 7.6. Putative GlnR regulated PE and PPE proteins identified in this study.

[a]PE/PPE name. [b]GlnR peak number identified in this study. [c]Peak intensity corresponds to fold enrichment of each peak calculated using SISSRs, and is based on the number of sequenced tags at each site vs the input control sample. [d]PE/PPE protein pair refers to genome localisation of the gene and corresponding pair location adjacent on the genome. [e]ESX cluster refers to PE/PPE protein pair situated within an ESX region. [f]Sub-group is determined by (50).

Table adapted from (138).

Nitrate/ Nitrite Metabolism

Six GlnR binding sites corresponded to promoter regions for genes involved in nitrate/nitrite metabolism and uptake. It has been demonstrated that *M. tuberculosis* can use nitrate as a nitrogen source, which may be vital for bacterial survival during infection (95). *M. tuberculosis* is phagocytosed by macrophages during infection, which could limit nutrient availability to the intracellular pathogen. However, nitrate is available in infected tissue, generated spontaneously from nitric oxide (NO), the product of nitric oxide synthase. For utilisation as a nitrogen source nitrate must be converted by *M. tuberculosis* to ammonium before entering the GS/GOGAT pathway. Nitrate is converted to ammonium via a two-step process; the first step involves reduction of nitrate (NO_3^-) to nitrite (NO_2^-) by NarGHJI, which is succeeded by a second step, nitrite reduction to ammonia (NH_4^+) by NirBD (Figure 7.16). Consequently regulation of the nitrate/nitrite genes involved in this process are of particular interest for *M. tuberculosis* survival.

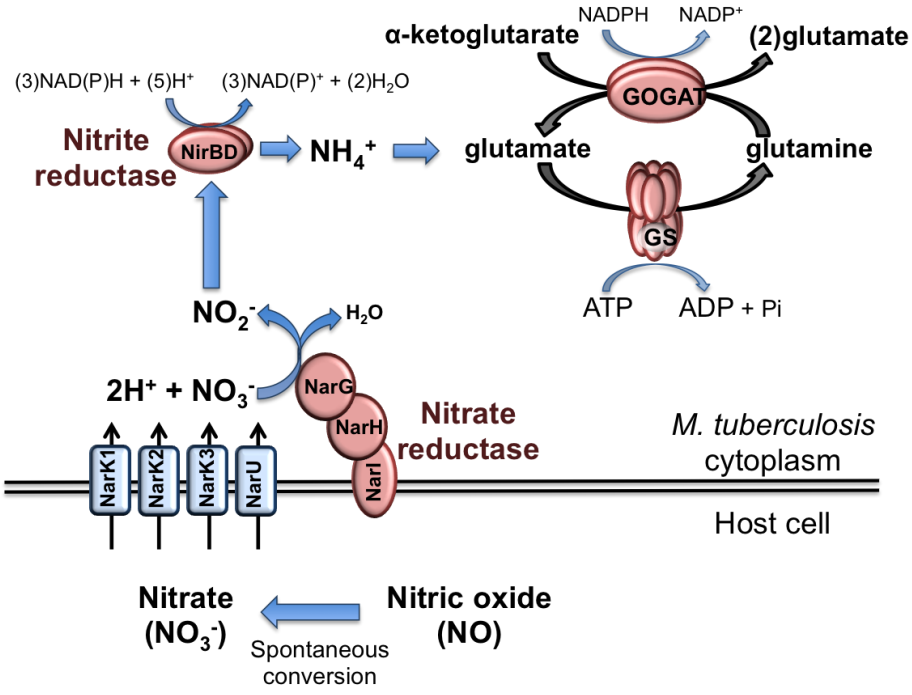


Figure 7.16. Conversion of nitrate to ammonium and subsequent entry into the GS/GOGAT pathway in *M. tuberculosis*.

Nitrate enters *M. tuberculosis,* possibly via the transporters NarK1-3 and NarU, and is converted to nitrate by NarG/H/I. Subsequent conversation of nitrite to ammonium proceeds via NirB/D before ammonium assimilation into glutamate and glutamine via the GS/GOGAT pathway.

GlnR has been demonstrated to regulate the nitrite reductase genes *nirB/D* in both *M. smegmatis* (this study) and *M. tuberculosis* (95). The promoter region of the *nirB* gene in *M. smegmatis* (this study) and *M. tuberculosis* (Rv0252; (95)) both contain a GlnR binding site and regulation of the *nirB/D* operon is disrupted in a *glnR* deletion mutant for both species ((95), this study). In this study it was confirmed that GlnR binds to the *nirB* promoter region in *M. tuberculosis*. In addition GlnR binding sites were identified upstream of a putative transcriptional regulator NnaR (Rv0260c), and a nitrite extrusion protein NarK3 (Rv0261c). All of these were identified in the *M. smegmatis* GlnR regulon (this study). An additional nitrite extrusion protein NarK1 (Rv2329c) also possessed a GlnR binding site, which is not present in *M. smegmatis*. This confirms previous studies that GlnR regulates genes involved in nitrite metabolism in both *M. smegmatis* and *M. tuberculosis*.

In *M. tuberculosis* the nitrate reductase genes also contained a GlnR binding site. Homologues of these genes are present in *M. smegmatis* (msmeg_5140-5137), but they were not GlnR regulated. *NarG* (Rv1161) contained a GlnR binding site within its promoter region and also within the gene itself. *NarG* is found in an operon with *narH*, *narJ* and *narI* (Rv1161-1164) in addition to Rv1165 (*typA;* possible GTP-binding translation elongation factor TYPA). Studies on *M. tuberculosis* NarGHJI found the enzyme to be functionally similar to that of *E. coli*; *M. tuberculosis* NarGHJI complemented a *narGHJI*-defective strain of *E. coli* supporting anaerobic growth, suggesting a role for this protein in *M. tuberculosis* in anaerobic metabolism (154). However, it was also reported that *M. tuberculosis*, unlike *E. coli*, constitutively expressed *narGHJI* independent of oxygen levels. Interestingly *M. tuberculosis* NarGHJI mediates reduction of nitrate under both anaerobic and aerobic conditions (155, 186), indicating that *M. tuberculosis* provides the first example of a *narGHJI*-encoded nitrate reductase that mediates assimilation of nitrate under aerobic conditions. One explanation for this observation is that in *M. tuberculosis* GlnR controls the expression of *narGHJI* and *nirBD* in response to nitrogen levels in the cell, rather than oxygen levels, suggesting a novel control mechanism of nitrate metabolism.

Molybdenum Cofactor Biosynthesis (moa operon)

Two GlnR binding sites were identified for genes encoding proteins involved in molybdenum cofactor biosynthesis. One binding site was inside *moaA1* (Rv3109) and the second in the promoter region of *moaB1* (Rv3110). The molybdenum cofactor is part of the active site of all molybdenum enzymes. There are several molybdenum enzymes in mycobacteria exerting important physiological functions, such as dormancy regulation and metabolism of energy

sources and nitrogen sources (Nitrate reductase NarGHJI being one example) (146). Pterin-based Mo cofactor is the most common cofactor of the molybdenum enzymes in mycobacteria (146).

Pterin-based Mo cofactor is synthesised by a conserved pathway. In *E. coli* more than 15 genes are involved in cofactor biosynthesis, and the function these genes is well understood (6). One locus involved in biosynthesis of Pterin-based Mo cofactor is *moa*. Whilst homologues for *moa* exist in *M. tuberculosis*, its activity has not been characterised biochemically. *moaA1* is probably part of an operon in *M. tuberculosis* due to the location of the *moaABCD* genes, which are organised in a cluster similar to that described in *E. coli*. In *E. coli* the ModE protein mediates transcriptional regulation of *moaABCD* (6). ModE has also been implicated in the regulation of *narG* in *E. coli*, however ModE does not bind directly to the *narG* promoter region, but binds to the *narK*-*narXL* intragenic region (141). In mycobacterial genomes no homologue of *modE* was identified. In this study possible GlnR regulation of *narG* is described, which in *E. coli* is ModE regulated. This evidence, and the lack of *modE* in mycobacterial genomes, may suggest that GlnR is involved in transcriptional regulation of the *moa* operon as well as other pterin-based Mo dependent enzymes in mycobacteria.

Genes Involved in Lipid Metabolism

ChIP-seq identified many GlnR binding sites upstream of genes involved in lipid metabolism. Sequence analysis of the *M. tuberculosis* genome showed that the bacillus encodes a wide range of proteins involved in lipid metabolism (36). In *E. coli* approximately 50 enzymes are involved in lipid metabolism, whereas the *M. tuberculosis* genome contains at least 250 genes, highlighting the importance of lipid metabolism in *M. tuberculosis* (36).

Two members of the FadD family of lipid degradation enzymes (*fadD24* (Rv1529) and *fadD11.1* (Rv1549)), contain GlnR binding sites, which may represent a switch in times of nitrogen poor conditions to an alternate carbon source. *In vivo*-grown mycobacteria are thought to be largely lipolytic, rather than lipogenic, because of the variety and quantity of lipids available within mammalian cells and the tubercle (36). The mycobacterial FadD proteins show homology to acyl-CoA synthetases that convert free fatty acids into acyl-coenzyme A (CoA) thioesters, the first step in fatty acid degradation (169). As such GlnR may regulate genes involved in scavenging alternate carbon sources, as well as a role in nitrogen metabolism.

A GlnR binding site was identified upstream of *pks5* (Rv1527c) a probable polyketide synthase. *M. smegmatis* contains a homologue MSMEG_4727, however it is not GlnR regulated. The role of

pks5 in *M. tuberculosis* has yet to be investigated, although studies in *M. smegmatis* described *pks5* as essential in the synthesis of lipooligosaccharides (44). Lipooligosaccharides are highly antigenic glycolipids, but their precise role in mycobacterial virulence is still a matter of debate (17, 85). Despite this, recent work demonstrated that lipooligosaccharides play a role in sliding motility, biofilm formation, and infection of murine macrophage-like cells by *M. marinum* (129). As such, lipooligosaccharide synthesis and its regulation in *M. tuberculosis* could provide an interesting study into the role of these compounds during infection.

A gene potentially involved in mycolic acid synthesis *umaA* (Rv0469) also contains a GlnR binding site. Mycolic acids are a major constituent of the lipid-rich envelope of mycobacteria and form an outer barrier with extremely low permeability that may explain their intrinsic resistance to many antibiotics. Structural analyses established the function of msmeg_0913, the *M. smegmatis* homologue of *umaA*, as an enzyme that adds a methyl branch to the vicinal position of both a *cis* double bond and cyclopropyl group, to yield *trans* mycolic acid homologues (82). However msmeg_0913 is not under GlnR regulation, which may indicate different roles for the *umaA* gene in each species. *M. tuberculosis* CDC1551, which carries a natural frame shift mutation in *umaA*, contains the same mycolate phenotype as *M. tuberculosis* H37Rv (82). Despite this, McAdam *et al.* revealed that Tn disruption of *M. tuberculosis umaA* resulted in hypervirulence in SCID mice (99). This suggests that the *umaA* gene is functional in H37Rv. It may therefore depend on environmental factors whether *umaA* has a role in *M. tuberculosis* and GlnR may regulate this during stress conditions. Consequently investigation into the effect of GlnR on lipid composition may determine whether control of *umaA* affects the cell wall composition during times of nitrogen deficiency.

ESAT-6 Like Proteins Rv3620c and Rv3219

A GlnR binding site was identified up stream of an ESAT-6 family member *esxW* (Rv3620c), suggesting GlnR controlled transcriptional regulation. Contained in an operon with Rv3620c is *esxV* (Rv3619), indicating GlnR regulation of both proteins. Rv3619c and Rv3620c are members of the Esx family of virulence factors, with *in silico* analysis predicting their presence in *M. leprae*, *M. avium* and *M. marinum* (97). Rv3619:Rv3620 interact in a 1:1 heteromeric structure and are secretory, antigenic proteins (92). One of the major functions associated with ESAT-6 family members is cytolytic activity. Cytolysis of host cells has been demonstrated by analysis of ESAT-6 protein Rv3875 (65). In *M. marinum* the ESAT-6 protein was demonstrated to induce pore formation, enabling the bacterium to escape from the vacuole to the host cell cytosol (151). However, to-date no studies have been performed on other ESAT-6 paralogs to determine

whether other members have similar membrane-destabilising functions. Analysis of Rv3619:Rv3620 function in pore formation may highlight a role for GlnR during *M. tuberculosis* infection.

In addition, ESAT-6 family proteins have been demonstrated to be potent T-cell antigens (1). As such, the antigenic properties of Rv3620c and Rv3619c have been utilised as part of a fusion protein (ID93), in efforts to boost immunity in BCG vaccinated individuals. The fusion protein ID93 incorporates PPE_MPTR42 (PPE family protein), Rv1813 (latency antigen), Rv3620c and Rv3619c (EsX family virulence factors). Recent studies demonstrated that ID93/GLA-SE vaccination elicits protection against *M. tuberculosis* in both mice and guinea pigs (138). Furthermore, ID93 is immunogenic in cynomolgous macaques and elicits polyfunctional CD4 and CD8 responses in peripheral blood mononuclear cells from BCG-vaccinated humans (138). As such the antigenic properties of Rv3620c and Rv3619c are of particular interest and their regulation and expression during infection may provide insight into the immunological response to *M. tuberculosis* infection.

Two GlnR binding Sites Located in Region of Deletion 9

M. tuberculosis shares over 99.9% identity at the DNA level with the other members of the tubercle complex, which includes *M. bovis*. Although highly related, distinct differences exist between host range, virulence in humans and physiological characteristics. As such genomic regions unique to *M. tuberculosis* are of particular interest to understand the pathogenicity of this species with respect to infection in humans.

Comparative genome studies identified 16 regions absent in *M. bovis* with respect to *M. tuberculosis* (16). One region of deletion (RD), classified as RD9 by Behr *et al.* encompasses seven ORFs (Rv3617 to Rv3623) (of note this region is alternatively denoted as RD8 by Gordon *et al*.). RD9 is a stretch of 5516 bp absent in all virulent *M. bovis* strains and the *M. bovis* BCG vaccination strain (16, 56). Within this region two GlnR binding sites were identified in *M. tuberculosis*. One GlnR binding site was located up stream of an ESAT-6 family member Rv3620c, which is contained in an operon with Rv3619c. The second GlnR binding site was positioned upstream of Rv3622c, which is contained in an operon with Rv3621c. ORFs Rv3621c and Rv3622c encode PPE and PE family proteins respectively. The presence of multiple GlnR binding sites in RD9 suggests a novel mechanism of gene regulation for this region in *M. tuberculosis,* potentially indicating a role for GlnR in the pathogenicity of *M. tuberculosis.*

Consensus DNA Binding Motif

To determine a GlnR consensus binding sequence in *M. tuberculosis* 200 bp regions centred from the peaks were submitted to MEME. This generated a motif found in only 18 of the 36 peaks, in contrast to the *M. smegmatis* GlnR DNA motif that was present in all 52 peaks. Consequently, a further investigation was conducted to identify other potential motif sequences present in the *M. tuberculosis* peaks. A MEME search was repeated with 50 bp of the peak sequence identified via ChIP-seq, which generated 6 different motifs. Interestingly genes known to be GlnR regulated in *M. smegmatis* contained GlnR consensus motif 2, with the exception of *nirB* which contained motif 3 (Figure 7.13). Neither motif 2 or 3 are similar to the *M. smegmatis* consensus motif so the results here are intriguing, suggesting other factors play a role in GlnR binding specificity in *M. tuberculosis.*

In contrast to *M. smegmatis* and *Streptomyces,* where the GlnR consensus sequence contained the same AC-n_9-AC spacing (Figure 7.17), the *M. tuberculosis* GlnR motif did not contain a clear AC-n_9-AC pattern and varied from the two previously identified binding regions (Figure 7.17). The C-terminal DNA binding domains of the GlnR proteins for *M. smegmatis* and *M. tuberculosis* are highly similar, therefore it is surprising that the consensus binding motifs differ (Figure 7.10, Figure 7.17). In addition, gel shift assays with peak regions identified from *M. tuberculosis* using recombinant His-GlnR protein from *M. smegmatis,* suggested that some conservation in C-terminal DNA binding domain of GlnR existed. As the *M. smegmatis* GlnR protein binds to the consensus sites in the *M. tuberculosis* peaks, it suggests that GlnR can recognise additional motif sequences that may be utilised by *M. tuberculosis*. It may be the difference in motif sequence in the pathogenic *M. tuberculosis* has switched the function of GlnR regulation from one of nitrogen metabolism to one associated with survival within host cells.

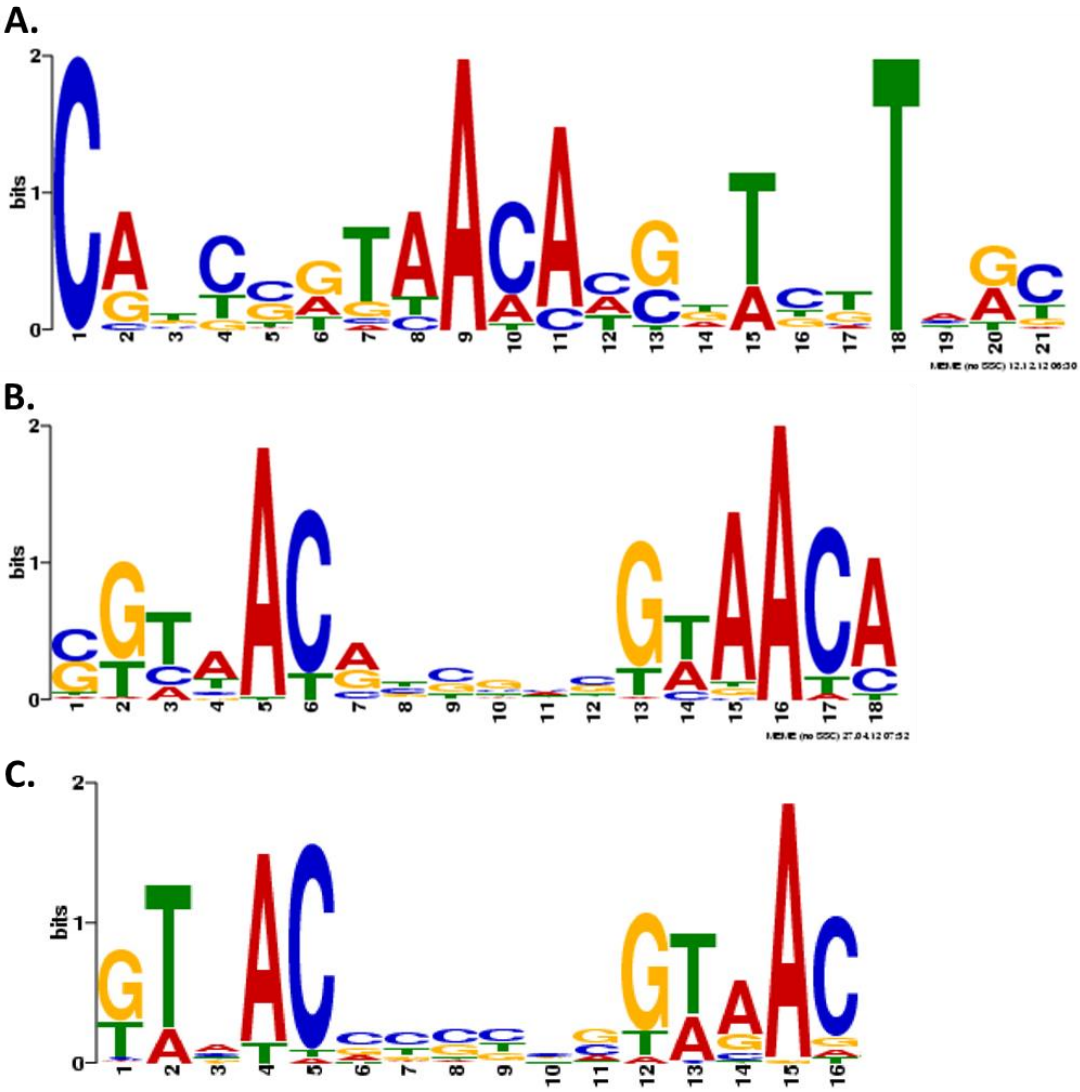


Figure 7.17. Comparison of MEME generated GlnR DNA binding consensus motif sequences.

(A) *M. tuberculosis* (this study from 200 bp MEME search)

(B) *M. smegmatis* (this study)

(C) *S. venezuelae* (120)

CHAPTER 8: FINAL DISCUSSION

8.1 Discussion

Nitrogen is an important bacterial cellular constituent and is required for growth, as such the ability to sense and initiate a response to situations of nitrogen-limitation is essential for bacterial survival. Compared to other extensively investigated organisms such as *E. coli*, information regarding the mechanisms underlying the regulation and assimilation of nitrogen in mycobacteria is limited. The nitrogen stress response has been studied in related Actinomycetes namely *C. glutamicum* and *S. coelicolor*, however the nitrogen stress response in these species differs to *E. coli* and to each other. It is clear that each organism has their own nitrogen-stress response and direct studies in mycobacteria are required.

In *M. smegmatis* the global transcriptional response to nitrogen-limitation is thought to be regulated by GlnR, based on homology to the global nitrogen response regulator GlnR in *Streptomyces*. However, in *M. smegmatis* only seven nitrogen metabolism genes have been reported to be under its direct control (3). For *M. tuberculosis* only one gene, *nirB*, has been demonstrated to be regulated by GlnR (95). In *S. coelicolor* GlnR mediates the expression of at least 50 genes in response to nitrogen limitation (120, 164, 165). Therefore it is likely that many unidentified genes are GlnR regulated, or another response regulator controls the nitrogen stress response in mycobacteria. Consequently, the aim of this study was to investigate the role of GlnR in the nitrogen-stress response and characterise the GlnR regulon in *M. smegmatis*. In addition, initial identification and characterisation of GlnR binding sites in the *M. tuberculosis* genome in nitrogen limitation were performed and the role of GlnR in the pathogenic *versus* saprophytic species discussed.

Development of Mycobacterial Nitrogen Limiting Growth Medium

Few publications are available that study nitrogen limitation in mycobacteria. These investigations vary in the concentration of nitrogen source used as "limiting", often without experimental confirmation of nitrogen limitation or explanation as to why these concentrations were chosen (3, 28, 58-61, 116, 126). Therefore, prior to our comprehensive global analyses using ChIP-seq and expression arrays, it was imperative we optimised our nitrogen limiting conditions. We chose the defined mycobacterial medium Sauton's as the basis of a liquid growth medium which, although would be subject to slight batch to batch variation, permitted the manipulation of nitrogenous components (113). For *M. smegmatis*, we determined the nitrogen limiting conditions to be 1 mM ammonium sulphate and nitrogen excess to be 30 mM ammonium sulphate. Applying these optimised conditions to *M. smegmatis*, a reduced growth

rate was observed in nitrogen limitation after 11 hours of growth, concurrent with external ammonium depletion. To confirm that this growth rate was solely due to nitrogen limitation, ammonium was added back to the limited cultures at this time point and a growth rate similar to nitrogen excess was observed (Figure 3.9). Expression levels of genes known to be induced in nitrogen limitation (*amt1, glnK, glnA1* and *amtB*) were obtained by qRT-PCR for further confirmation that *M. smegmatis* was limited for nitrogen; these genes were induced in our nitrogen limiting but not nitrogen excess conditions (Table 4.1). Therefore, unless stated otherwise, these conditions were applied throughout this study and can be used for any future nitrogen limitation mycobacterial studies.

Investigation into the Mechanism of GlnR Activation

The signal for low nitrogen and how this is translated into a GlnR mediated transcriptomic response is unknown. It was reported previously, and confirmed in this study, that *glnR* expression is not auto-regulated and *glnR* transcripts do not increase during nitrogen limitation (3)(Figure 4.7). GlnR belongs to the OmpR-family of transcriptional regulators that are typically activated by phosphorylation at a conserved aspartate residue. GlnR contains this conserved putative phosphorylation site Asp48, and therefore it is a reasonable assumption that GlnR is activated by phosphorylation during nitrogen limitation (3). However, to date this has not been reported possibly due to the labile nature of the phospho-aspartate bond, making direct investigation difficult. Therefore, to investigate the role the putative phosphorylation site further I generated an *in vivo* GlnR aspartate to alanine mutation, and analysed the transcriptomic response to nitrogen limitation. The GlnR aspartate to alanine mutation resulted in loss of transcriptional activation of known GlnR regulated genes under nitrogen limitation (Table 4.1), indicating that this residue is essential for GlnR function. Although this suggests that that GlnR is activated by modification at Asp48, potentially by phosphorylation, we cannot rule out conformational changes of GlnR induced by this amino acid change. Furthermore, phosphorylation of GlnR was not observed in this study using several approaches (Phos-Tag, IEF and radiolabelling: unpublished data). Consequently, we cannot rule out another mechanism of activation such as modification by protein:protein interaction with P_{II} proteins, as in *E. coli* and *C. glutamicum.*

Deciphering the GlnR Regulon in M. smegmatis

To gain a global insight into the genes regulated by GlnR under nitrogen limitation, a ChIP-seq approach combined with a global expression analysis of WT and *glnR* deletion strain in nitrogen limitation was applied. ChIP-seq permitted identification of *in vivo* GlnR binding sites during nitrogen limitation, and combining this with the microarray data, allowed genes that were directly regulated by GlnR binding to be identified. *In vivo* GlnR genome binding sites under nitrogen limitation were examined and 52 binding sites confirmed. Combined with microarray analysis, 103 genes were assigned as being directly GlnR regulated during nitrogen limitation; 7 genes were down regulated and 96 genes up regulated, indicating that GlnR is the main nitrogen response regulator in *M. smegmatis* and that GlnR acts as a dual activator and repressor.

Under nitrogen limiting conditions the majority of GlnR regulated genes are involved in a nitrogen scavenging response. This is suggested by up regulation of 27 genes for various transporters including amino acids (msmeg_0781, msmeg_1052, msmeg_2184, msmeg_2332, msmeg_2522-2524, msmeg_6735), nucleobases (msmeg_1177, msmeg_1293, msmeg_2570, msmeg_3402, msmeg_4011, msmeg_6660), urea (msmeg_2978-2982), nitrate (*narK3*) and ammonium (*amt1*, *amtA*, *amtB*). GlnR also up regulated 34 genes encoding enzymes for the conversation and break down of compounds to release ammonium. These include nitrite reductase (*nirBD*), guanine deaminase, hydrolases and amine oxidases (Table 6.6). The breakdown of nitrogen containing compounds would lead to a release of ammonium that could then subsequently enter the GS/GOGAT and GDH pathways. An overview of the nitrogen scavenging response and possible assimilation processes are depicted in Figure 8.1.

Eight genes did not show differential gene expression in our microarray analysis despite GlnR binding. This may indicate a role for another, as yet unidentified transcription factor enabling gene regulation at these sites. Binding of additional transcription factors to GlnR regulated genes has been demonstrated in *S. coelicolor* with PhoP binding to promoter regions of *amtB* and *glnA* (132). However, the relevance of this dual GlnR/PhoP binding has yet to be demonstrated. There also appears to be a multi-level control of gene expression during nitrogen limitation, with GlnR regulating 7 other transcription factors. This is not surprising as metabolic pathways do not operate in isolation, which means that there may be cross-talk of regulatory proteins between different pathways. In addition, 29 GlnR regulated genes were classified with function unknown, and further analysis of these may expand our knowledge of the mechanism of nitrogen scavenging in *M. smegmatis*. From this study it is clear however that the majority of GlnR regulated genes in *M. smegmatis* are involved in scavenging, breakdown and assimilation of nitrogen (Figure 8.1).

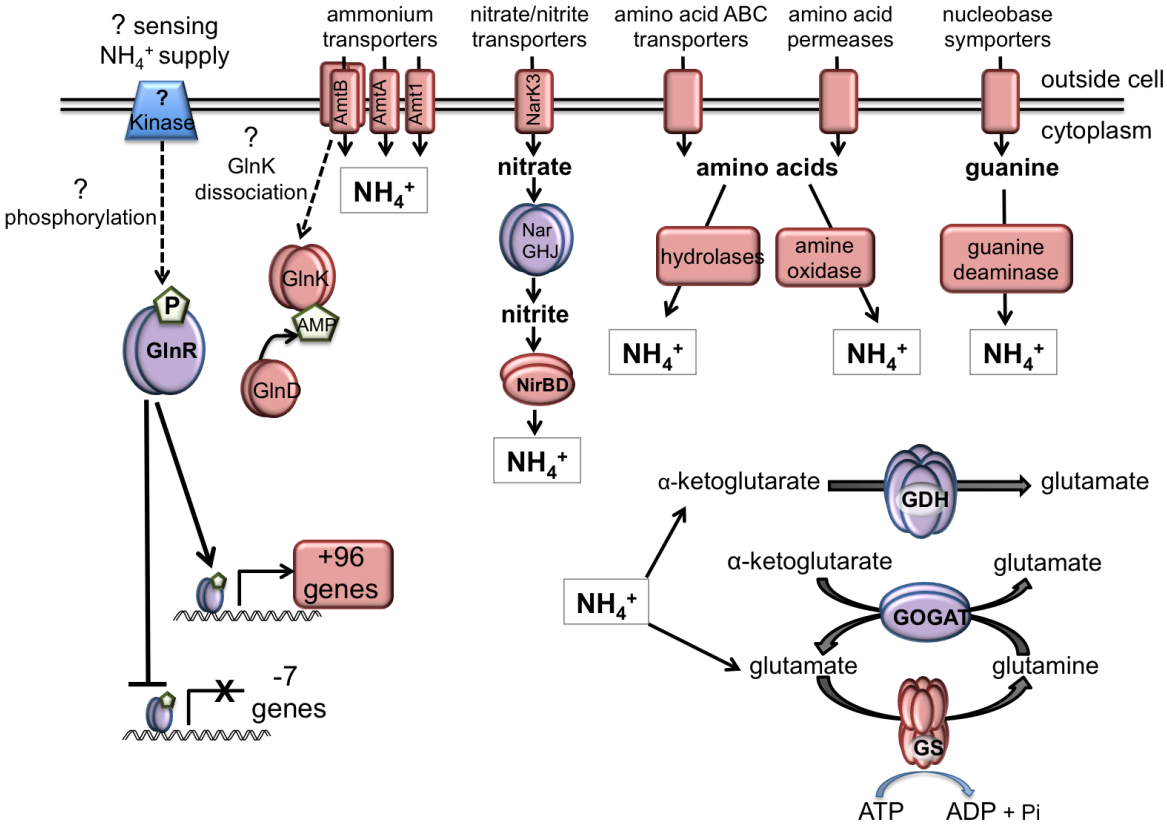


Figure 8.1. Proposed model for the GlnR-mediated nitrogen scavenging response in *M. smegmatis*.

In response to nitrogen limitation GlnR is activated by an unknown mechanism; potentially phosphorylated by an unknown sensor kinase that could serve to detect the level of extracellular NH_4 (unknown kinase depicted in blue). In nitrogen limitation, GlnR binds to 52 sites and up-regulates the transcription of 96 genes and the repression of 7 genes (up-regulated GlnR genes are depicted in red; genes not GlnR regulated are depicted in purple). GlnR activation leads to increased expression of genes encoding numerous nitrogen transporters including AmtB, AmtA, Amt1 (ammonium transporters), NarK3 (nitrate/nitrite transporter), amino acid transporters/ permeases and nucleobase symporters. Genes that encode enzymes to break down nitrogen containing molecules into ammonium are also up-regulated by GlnR, including NirBD (nitrite reductase), guanine deaminase and various amino acid hydrolases and oxidases. Ammonium can subsequently enter the GOGAT/GS pathway or GDH pathway for assimilation into glutamate and glutamine and other biomolecules.

GlnR Binding Sites in M. tuberculosis

ChIP-seq methodology optimised in *M. smegmatis* was then applied to analyse the *in vivo* binding sites of GlnR in *M tuberculosis*, providing a preliminary insight into the role of GlnR in pathogenic species. In addition, this permitted direct comparison of the role of GlnR in the pathogenic and saprophytic species. *M. smegmatis* is often used as a model organism for *M. tuberculosis*, however the regulatory mechanisms with respect to GlnR in the nitrogen stress response are unknown.

In *M. tuberculosis* 36 GlnR binding sites were identified in nitrogen limitation. As for *M. smegmatis* a selection of binding sites were confirmed via EMSA and rate limiting PCR analysis. However, the expression array of WT vs *glnR* mutant in nitrogen limitation, to compare binding to GlnR binding to gene expression levels, is still in progress and could not be used for analysis in this study. Initial ChIP-seq analysis provided interesting data, suggesting that GlnR performs a different role in *M. tuberculosis* under nitrogen limiting conditions. In *M. smegmatis* the majority of GlnR regulated genes (where known) were involved in nitrogen metabolism, however in *M. tuberculosis* the only known nitrogen metabolism genes to contain GlnR binding sites upstream were *glnA1*, *glnA2* (both GS) and genes involved in nitrate/ nitrite metabolism. Common binding sites to both species were identified upstream of *glnA1*, *glnA2*, *nirB, nnaR, narK3, whiB1, whiB3,* and *glbN,* while other binding sites were unique to *M. tuberculosis*. For an overview and direct comparison of nitrogen metabolism genes putatively regulated by GlnR in *M. tuberculosis* see Figure 8.2. An interesting observation is that GlnR does not regulate a nitrogen scavenging response in *M. tuberculosis,* but does regulate nitrate metabolism in both species. Potentially highlighting the importance of nitrate as a nitrogen source for *M. tuberculosis in vivo* as GlnR has retained this regulatory function, whereas regulation of other nitrogen metabolism genes has been lost, presumably due to reductive evolution.

In *M. tuberculosis* initial analysis suggests GlnR may have adapted to a more a global role, with a general stress response. Genes potentially regulated by GlnR include those involved in fatty acid metabolism, PE and PPE proteins with potential roles in virulence, and two binding sites located in the RD8 region. The latter region is particularly interesting as it may provide a novel mechanism of regulation involving GlnR that is unique to the pathogenic strain. This work requires further validation before genes can be formally assigned to the GlnR regulon in *M. tuberculosis*. However, the results from ChIP-seq provide an interesting initial insight into the regulatory pathways involving GlnR in *M. tuberculosis*, whereby GlnR in *M. smegmatis* and *M. tuberculosis* appears to have different roles during nitrogen limitation.

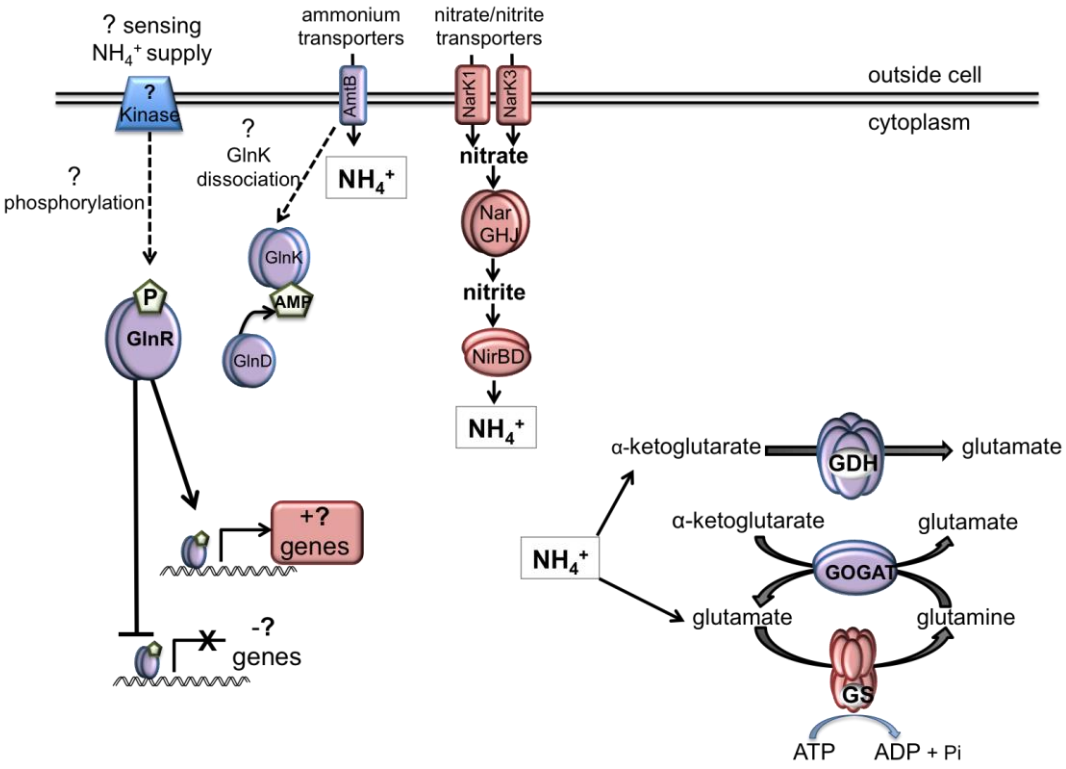


Figure 8.2. GlnR regulated nitrogen metabolism genes in *M. tuberculosis.*

In response to nitrogen limitation, GlnR is activated by an unknown mechanism, potentially phosphorylated by an unknown sensor kinase that could detect the level of extracellular NH_4 (unknown kinase depicted in blue). Activation of GlnR leads to binding to 36 regions on the *M. tuberculosis* genome, the transcriptional activation and repression of genes is unknown. For proteins involved in nitrogen metabolism GlnR binding (depicted in red) occurs up stream of NarK1, NarK3 (nitrate/nitrite transporters), NarGHJ (nitrate reductase), NirBD (nitrite reductase), and GS (*glnA1* and *glnA2*).

Comparison of GlnR DNA Binding Consensus Sequences

Combining the GlnR binding site DNA sequences a consensus DNA binding motif for *M. smegmatis*, but not *M. tuberculosis* was identified. Using MEME, a GlnR consensus motif of $Gn_2AC\text{-}n_6GnAACA$ was identified in all 52 binding sites and was highly similar in sequence to the GlnR binding motif identified in *Streptomyces*. What is strikingly apparent from these motifs is the conservation of the $AC\text{-}n_9\text{-}AC$ sequence in GlnR binding motifs in *M. smegmatis* and *Streptomyces*. The AC residues and the spacing between them was shown to be essential for DNA binding by a mutagenesis approach. These key residues are spaced 9 bp apart, representing one turn of the major groove of the DNA helix, and it would be interesting to see if this mechanism of DNA interaction and the importance of these key residues is conserved in *Streptomyces.*

A consensus GlnR DNA binding sequence for *M. tuberculosis* could not be identified. The "putative" binding motif for GlnR in *M. tuberculosis* was $Cn_7AnAn_3Tn_2T$ and only identified in 18 of the 36 GlnR binding sites. Deviations from the consensus binding sequence have been noted in *Streptomyces,* where nine of the thirty-six GlnR binding regions identified by ChIP-CHIP did not contain the "consensus" GlnR DNA binding motif (120). An additional *Streptomyces* study investigated the consensus GlnR binding motif upstream of *nasA* by DNase digestion, confirming GlnR DNA binding to a "non-consensus" site; GlnR bound to two distantly separated copies of a $GTAAC\text{-}n_{18}\text{-}GTAAC$ motif (181). Pullan *et al.* suggested local changes in DNA topology and/or the co-operative interactions of multiple transcriptional factors reduce the requirement for a consensus sequence and facilitate GlnR binding to these "non-consensus" sites (120); a similar situation may be occurring in *M. tuberculosis*. Confirmation and characterisation of the GlnR binding sites in *M. tuberculosis* are in progress and it is hoped that upon analysis of the expression array of WT vs *glnR* mutant, GlnR binding sites can be further investigated. Understanding the requirements for GlnR:DNA binding in *M. tuberculosis* in comparison to *M. smegmatis* may identify novel GlnR control mechanisms in pathogenic species, where GlnR may serve a different regulatory function.

8.2 Future Work

This study provides a global understanding of the GlnR mediated transcriptional response to nitrogen limitation in mycobacteria. However, this project has also raised several interesting questions regarding the control of nitrogen metabolism in mycobacteria:

- What is the mechanism of GlnR activation and the signal to which GlnR responds to in nitrogen limitation? Although it has been demonstrated that GlnR Asp-48 residue is essential for the GlnR mediated transcriptomic response to nitrogen limitation, whether this residue is modified under nitrogen limiting conditions is unknown. Determination of the post-translational status of GlnR during nitrogen limitation is a key area for future work. *In vitro* phosphorylation assays, or immunoprecipitaion of GlnR from whole cell lysates followed by sensitive detection analysis such as LC-MS could be used. The corresponding kinase, if GlnR is phosphorylated, would also be an interesting area of study; generating a library of kinase mutants may permit identification of the corresponding kinase. As the transcription of GlnR does not increase during nitrogen limitation, the cellular localisation of GlnR during nitrogen excess conditions is of particular interest, and may provide a mechanism of GlnR activation. Cell fractionation would permit detection of GlnR at the cell membrane or cytoplasm and this location may alter depending on the nitrogen status of the media and the cell.

- To confirm the role of GlnR in nitrogen limitation in *M. tuberculosis* the GlnR binding sites identified via ChIP-seq need to be aligned to RNA microarray data from WT versus *glnR* deletion strain grown under nitrogen limiting conditions. Assigning these genes would confirm whether the response to nitrogen limitation in *M. tuberculosis* is a general stress response, or a specific up regulation of genes in the nitrogen metabolism pathway.

- Investigation into the role of GlnR in *in vitro* and *in vivo* models of *M. tuberculosis* infection. Analysing an *M. tuberculosis glnR* deletion mutant in *in vitro* stress tests (such as the Wayne model of dormancy) and *in vivo* infection (in particular macrophage infection models). This would identify whether the GlnR transcriptomic response impacts on *M. tuberculosis* survival during infection.

- Several other transcriptional regulators are potentially involved in the *M. smegmatis* transcriptomic response during nitrogen limitation. The involvement of AmtR, PhoP and GlnR mediated transcription factors such are NnaR could be investigated by a similar combined approach (ChIP-seq and micro array analysis).

- Chemostat *versus* batch nitrogen limitation experiments. As we investigated the transcriptional response to nitrogen limitation in batch culture, metabolic changes due to the observed reduction in growth rate between the nitrogen limiting and excess conditions will be included in our 'nitrogen stress' response. To delineate the effect of growth rate from the effect of nitrogen limitation the ChIP-seq and microarray analysis should be repeated using a chemostat where growth rate could be accurately controlled.

REFERENCES

1. **Alderson, M. R., T. Bement, C. H. Day, L. Zhu, D. Molesh, Y. A. Skeiky, R. Coler, D. M. Lewinsohn, S. G. Reed, and D. C. Dillon.** 2000. Expression cloning of an immunodominant family of *Mycobacterium tuberculosis* antigens using human CD4(+) T cells. Journal of Experimental Medicine **191:**551-560.
2. **Amin, R., J. Reuther, A. Bera, W. Wohlleben, and Y. Mast.** 2012. A novel GlnR target gene, *nnaR,* is involved in nitrate/nitrite assimilation in *Streptomyces coelicolor*. Microbiology **158:**1172-1182.
3. **Amon, J., T. Brau, A. Grimrath, E. Hanssler, K. Hasselt, M. Holler, N. Jessberger, L. Ott, J. Szokol, F. Titgemeyer, and A. Burkovski.** 2008. Nitrogen control in *Mycobacterium smegmatis*: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-type regulator GlnR. Journal Bacteriology **190:**7108-7116.
4. **Amon, J., F. Titgemeyer, and A. Burkovski.** 2010. Common patterns - unique features: nitrogen metabolism and regulation in Gram-positive bacteria. FEMS microbiology reviews **34:**588-605.
5. **Amon, J., F. Titgemeyer, and A. Burkovski.** 2009. A genomic view on nitrogen metabolism and nitrogen control in mycobacteria. Journal of molecular microbiology and biotechnology **17:**20-29.
6. **Anderson, L. A., E. McNairn, T. Lubke, R. N. Pau, and D. H. Boxer.** 2000. ModE-dependent molybdate regulation of the molybdenum cofactor operon moa in *Escherichia coli*. Journal of bacteriology **182:**7035-7043.
7. **Arcondeguy, T., R. Jack, and M. Merrick.** 2001. P(II) signal transduction proteins, pivotal players in microbial nitrogen control. Microbiology and molecular biology reviews : MMBR **65:**80-105.
8. **Atkinson, M. R., and A. J. Ninfa.** 1999. Characterization of the GlnK protein of *Escherichia coli*. Molecular microbiology **32:**301-313.
9. **Attwood, P. V., P. G. Besant, and M. J. Piggott.** 2011. Focus on phosphoaspartate and phosphoglutamate. Amino acids **40:**1035-1051.
10. **Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble.** 2009. MEME SUITE: tools for motif discovery and searching. Nucleic acids research **37:**W202-208.
11. **Bailey, T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:**28-36.
12. **Bardarov, S., S. Bardarov Jr, Jr., M. S. Pavelka Jr, Jr., V. Sambandamurthy, M. Larsen, J. Tufariello, J. Chan, G. Hatfull, and W. R. Jacobs Jr, Jr.** 2002. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis*, *M. bovis* BCG and *M. smegmatis*. Microbiology **148:**3007-3017.
13. **Barry, C. E., 3rd, H. I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R. J. Wilkinson, and D. Young.** 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nature Reviews Microbiology **7:**845-855.
14. **Bauer, S., P. N. Robinson, and J. Gagneur.** 2011. Model-based gene set analysis for Bioconductor. Bioinformatics **27:**1882-1883.
15. **Beckers, G., J. Strosser, U. Hildebrandt, J. Kalinowski, M. Farwick, R. Kramer, and A. Burkovski.** 2005. Regulation of AmtR-controlled gene expression in *Corynebacterium glutamicum*: mechanism and characterization of the AmtR regulon. Molecular microbiology **58:**580-595.
16. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small.** 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284:**1520-1523.

17. **Belisle, J. T., and P. J. Brennan.** 1989. Chemical basis of rough and smooth variation in mycobacteria. Journal of bacteriology **171:**3465-3470.
18. **Bergey, D. H., G. M. Garrity, D. R. Boone, and R. W. Castenholz.** 2001. Bergey's manual of systematic bacteriology/ Vol 1, The archaea and the deeply branching and phototrophic bacteria. Springer, London.
19. **Bergstrom, L. C., L. Qin, S. L. Harlocker, L. A. Egger, and M. Inouye.** 1998. Hierarchical and co-operative binding of OmpR to a fusion construct containing the *ompC* and *ompF* upstream regulatory sequences of *Escherichia coli*. Genes to cells : devoted to molecular & cellular mechanisms **3:**777-788.
20. **Betts, J. C., P. T. Lukey, L. C. Robb, R. A. McAdam, and K. Duncan.** 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Molecular microbiology **43:**717-731.
21. **Bogdan, C.** 2001. Nitric oxide and the immune response. Nature immunology **2:**907-916.
22. **Boshoff, H. I., and C. E. Barry, 3rd.** 2005. Tuberculosis - metabolism and respiration in the absence of growth. Nature Reviews Microbiology **3:**70-80.
23. **Bourret, R. B., J. F. Hess, and M. I. Simon.** 1990. Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY. Proc Natl Acad Sci U S A **87:**41-45.
24. **Britton, K. L., P. J. Baker, D. W. Rice, and T. J. Stillman.** 1992. Structural relationship between the hexameric and tetrameric family of glutamate dehydrogenases. European Journal of Biochemistry **209:**851-859.
25. **Buchinger, S., J. Strosser, N. Rehm, E. Hanssler, S. Hans, B. Bathe, D. Schomburg, R. Kramer, and A. Burkovski.** 2009. A combination of metabolome and transcriptome analyses reveals new targets of the *Corynebacterium glutamicum* nitrogen regulator AmtR. Journal Biotechnology **140:**68-74.
26. **Burkovski, A.** 2003. I do it my way: Regulation of ammonium uptake and ammonium assimilation in *Corynebacterium glutamicum*. Arch Microbiol **179:**83-88.
27. **Burkovski, A.** 2007. Nitrogen control in *Corynebacterium glutamicum*: proteins, mechanisms, signals. Journal Microbiology Biotechnology **17:**187-194.
28. **Carroll, P., C. A. Pashley, and T. Parish.** 2008. Functional analysis of GlnE, an essential adenylyl transferase in *Mycobacterium tuberculosis*. Journal of bacteriology **190:**4894-4902.
29. **Cha, R. S., H. Zarbl, P. Keohavong, and W. G. Thilly.** 1992. Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. PCR Methods Appl **2:**14-20.
30. **Chandra, H., S. F. Basir, M. Gupta, and N. Banerjee.** 2010. Glutamine synthetase encoded by *glnA-1* is necessary for cell wall resistance and pathogenicity of *Mycobacterium bovis*. Microbiology.
31. **Chen, J. M., D. C. Alexander, M. A. Behr, and J. Liu.** 2003. *Mycobacterium bovis* BCG vaccines exhibit defects in alanine and serine catabolism. Infection and immunity **71:**708-716.
32. **Chierakul N, M. P., Nana A, Jearanaisilavong J, Sriumpai S, Bovornkitti S.** 1993. *Mycobacterium smegmatis* Infection in a Thai Woman. Journal of Infection and Disease Antimicrobial Agents **10:**25-28.
33. **Chung, C. T., and R. H. Miller.** 1993. Preparation and storage of competent *Escherichia coli* cells. Methods in Enzymology **218:**621-627.
34. **Clegg, S., F. Yu, L. Griffiths, and J. A. Cole.** 2002. The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. Molecular microbiology **44:**143-155.
35. **Clemens, D. L., B. Y. Lee, and M. A. Horwitz.** 1995. Purification, characterization, and genetic analysis of *Mycobacterium tuberculosis* urease, a potentially critical determinant of host-pathogen interaction. Journal of bacteriology **177:**5644-5652.
36. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T.**

Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537-544.

37. **Conroy, M. J., A. Durand, D. Lupo, X. D. Li, P. A. Bullough, F. K. Winkler, and M. Merrick.** 2007. The crystal structure of the *Escherichia coli* AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. Proc Natl Acad Sci U S A **104:**1213-1218.
38. **Corbett, E. L.** 2003. The growing burden of tuberculosis - Global trends and interactions with the HIV epidemic. Archives of internal medicine **163:**1009-1021.
39. **Delgado, J., S. Forst, S. Harlocker, and M. Inouye.** 1993. Identification of a phosphorylation site and functional analysis of conserved aspartic acid residues of OmpR, a transcriptional activator for ompF and ompC in *Escherichia coli*. Molecular microbiology **10:**1037-1047.
40. **Deturk, W. E., and F. Bernheim.** 1958. Effects of ammonia, methylamine, and hydroxylamine on the adaptive assimilation of nitrite and nitrate by a *Mycobacterium*. Journal of bacteriology **75:**691-696.
41. **Drake, S. K., R. B. Bourret, L. A. Luck, M. I. Simon, and J. J. Falke.** 1993. Activation of the phosphosignaling protein CheY. I. Analysis of the phosphorylated conformation by 19F NMR and protein engineering. The Journal of biological chemistry **268:**13081-13088.
42. **Durand, A., and M. Merrick.** 2006. *In vitro* analysis of the *Escherichia coli* AmtB-GlnK complex reveals a stoichiometric interaction and sensitivity to ATP and 2-oxoglutarate. The Journal of biological chemistry **281:**29558-29567.
43. **Embley, T. M., and E. Stackebrandt.** 1994. The molecular phylogeny and systematics of the actinomycetes. Annual Reviews Microbiology **48:**257-289.
44. **Etienne, G., W. Malaga, F. Laval, A. Lemassu, C. Guilhot, and M. Daffe.** 2009. Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. Journal of bacteriology **191:**2613-2621.
45. **Feng, Z., N. E. Caceres, G. Sarath, and R. G. Barletta.** 2002. *Mycobacterium smegmatis* L-alanine dehydrogenase (Ald) is required for proficient utilization of alanine as a sole nitrogen source and sustained anaerobic growth. Journal of bacteriology **184:**5001-5010.
46. **Fenhalls, G., L. Stevens, L. Moses, J. Bezuidenhout, J. C. Betts, P. Helden Pv, P. T. Lukey, and K. Duncan.** 2002. *In situ* detection of *Mycobacterium tuberculosis* transcripts in human lung granulomas reveals differential gene expression in necrotic lesions. Infection and immunity **70:**6330-6338.
47. **Fink, D., N. Weissschuh, J. Reuther, W. Wohlleben, and A. Engels.** 2002. Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2). Molecular microbiology **46:**331-347.
48. **Forchhammer, K.** 2004. Global carbon/nitrogen control by PII signal transduction in cyanobacteria: from signals to targets. FEMS microbiology reviews **28:**319-333.
49. **Furey, T. S.** 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nature Reviews Genetics **13:**840-852.
50. **Gey van Pittius, N. C., S. L. Sampson, H. Lee, Y. Kim, P. D. van Helden, and R. M. Warren.** 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC evolutionary biology **6:**95.
51. **Giffin, M. M., L. Modesti, R. W. Raab, L. G. Wayne, and C. D. Sohaskey.** 2012. *ald* of *Mycobacterium tuberculosis* encodes both the alanine dehydrogenase and the putative glycine dehydrogenase. Journal of bacteriology **194:**1045-1054.

52. **Gill, H. S., G. M. Pfluegl, and D. Eisenberg.** 2002. Multicopy crystallographic refinement of a relaxed glutamine synthetase from *Mycobacterium tuberculosis* highlights flexible loops in the enzymatic mechanism and its regulation. Biochemistry **41:**9863-9872.
53. **Glickman, M. S., and W. R. Jacobs, Jr.** 2001. Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline. Cell **104:**477-485.
54. **Goh, K. S., N. Rastogi, M. Berchel, R. C. Huard, and C. Sola.** 2005. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (*narGHJI*) operon promoter. Journal of clinical microbiology **43:**4010-4014.
55. **Gordon, R. E., and M. M. Smith.** 1953. Rapidly growing, acid fast bacteria. I. Species' descriptions of *Mycobacterium phlei* Lehmann and Neumann and *Mycobacterium smegmatis* (Trevisan) Lehmann and Neumann. Journal of bacteriology **66:**41-48.
56. **Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Molecular microbiology **32:**643-655.
57. **Gruswitz, F., J. O'Connell, 3rd, and R. M. Stroud.** 2007. Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GlnK, at 1.96 A. Proc Natl Acad Sci U S A **104:**42-47.
58. **Harper, C., D. Hayward, I. Wiid, and P. van Helden.** 2008. Regulation of nitrogen metabolism in *Mycobacterium tuberculosis*: a comparison with mechanisms in *Corynebacterium glutamicum* and *Streptomyces coelicolor*. IUBMB Life **60:**643-650.
59. **Harper, C. J., D. Hayward, M. Kidd, I. Wiid, and P. van Helden.** 2010. Glutamate dehydrogenase and glutamine synthetase are regulated in response to nitrogen availability in *Myocbacterium smegmatis*. BMC microbiology **10:**138.
60. **Harth, G., D. L. Clemens, and M. A. Horwitz.** 1994. Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity. Proc Natl Acad Sci U S A **91:**9342-9346.
61. **Harth, G., S. Maslesa-Galic, M. V. Tullius, and M. A. Horwitz.** 2005. All four *Mycobacterium tuberculosis glnA* genes encode glutamine synthetase activities but only GlnA1 is abundantly expressed and essential for bacterial homeostasis. Molecular microbiology **58:**1157-1172.
62. **Health Protection Agency Centre for, I.** 2009. Tuberculosis in the UK: Annual report on tuberculosis surveillance in the UK 2009. Health Protection Agency Centre for Infections.
63. **Hedgecock, L. W., and R. L. Costello.** 1962. Utilization of nitrate by pathogenic and saprophytic mycobacteria. Journal Bacteriology **84:**195-205.
64. **Hesketh, A., D. Fink, B. Gust, H. U. Rexer, B. Scheel, K. Chater, W. Wohlleben, and A. Engels.** 2002. The GlnD and GlnK homologues of *Streptomyces coelicolor* A3(2) are functionally dissimilar to their nitrogen regulatory system counterparts from enteric bacteria. Molecular microbiology **46:**319-330.
65. **Hsu, T., S. M. Hingley-Wilson, B. Chen, M. Chen, A. Z. Dai, P. M. Morin, C. B. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R. G. Russell, S. C. Derrick, F. M. Collins, S. L. Morris, C. H. King, and W. R. Jacobs, Jr.** 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A **100:**12420-12425.
66. **Huang, K. J., C. Y. Lan, and M. M. Igo.** 1997. Phosphorylation stimulates the cooperative DNA-binding properties of the transcription factor OmpR. Proc Natl Acad Sci U S A **94:**2828-2832.
67. **Inoue, H., H. Nojima, and H. Okayama.** 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene **96:**23-28.
68. **Jaggi, R., W. Ybarlucea, E. Cheah, P. D. Carr, K. J. Edwards, D. L. Ollis, and S. G. Vasudevan.** 1996. The role of the T-loop of the signal transducing protein PII from *Escherichia coli*. FEBS letters **391:**223-228.

69. **Jakoby, M., L. Nolden, J. Meier-Wagner, R. Kramer, and A. Burkovski.** 2000. AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*. Molecular microbiology **37:**964-977.
70. **Javelle, A., E. Severi, J. Thornton, and M. Merrick.** 2004. Ammonium sensing in *Escherichia coli*. Role of the ammonium transporter AmtB and AmtB-GlnK complex formation. The Journal of biological chemistry **279:**8530-8538.
71. **Javelle, A., G. Thomas, A. M. Marini, R. Kramer, and M. Merrick.** 2005. *In vivo* functional characterization of the *Escherichia coli* ammonium channel AmtB: evidence for metabolic coupling of AmtB to glutamine synthetase. Biochemical Journal **390:**215-222.
72. **Jenkins, V. A., B. D. Robertson, and K. J. Williams.** 2012. Aspartate D48 is essential for the GlnR-mediated transcriptional response to nitrogen limitation in *Mycobacterium smegmatis*. FEMS Microbiology Letters.
73. **Jia, W., and J. A. Cole.** 2005. Nitrate and nitrite transport in *Escherichia coli*. Biochemical Society transactions **33:**159-161.
74. **Jiang, P., and A. J. Ninfa.** 2009. Alpha-ketoglutarate controls the ability of the *Escherichia coli* PII signal transduction protein to regulate the activities of NRII (NrB but does not control the binding of PII to NRII). Biochemistry **48:**11514-11521.
75. **Jiang, P., J. A. Peliska, and A. J. Ninfa.** 1998. Reconstitution of the signal-transduction bicyclic cascade responsible for the regulation of Ntr gene transcription in *Escherichia coli*. Biochemistry **37:**12795-12801.
76. **Jiang, P., J. A. Peliska, and A. J. Ninfa.** 1998. The regulation of *Escherichia coli* glutamine synthetase revisited: role of 2-ketoglutarate in the regulation of glutamine synthetase adenylylation state. Biochemistry **37:**12802-12810.
77. **Kamala, T., C. N. Paramasivan, D. Herbert, P. Venkatesan, and R. Prabhakar.** 1994. Isolation and Identification of Environmental Mycobacteria in the *Mycobacterium bovis* BCG Trial Area of South India. Applied and environmental microbiology **60:**2180-2183.
78. **Karakousis, P. C., E. P. Williams, and W. R. Bishai.** 2008. Altered expression of isoniazid-regulated genes in drug-treated dormant Mycobacterium tuberculosis. The Journal of antimicrobial chemotherapy **61:**323-331.
79. **Keener, J., and S. Kustu.** 1988. Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NtrB and NtrC of enteric bacteria: roles of the conserved amino-terminal domain of NtrC. Proc Natl Acad Sci U S A **85:**4976-4980.
80. **Kenney, L. J.** 2002. Structure/function relationships in OmpR and other winged-helix transcription factors. Current opinion in microbiology **5:**135-141.
81. **Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10:**R25.
82. **Laval, F., R. Haites, F. Movahedzadeh, A. Lemassu, C. Y. Wong, N. Stoker, H. Billman-Jacobe, and M. Daffe.** 2008. Investigating the function of the putative mycolic acid methyltransferase UmaA: divergence between the *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* proteins. The Journal of biological chemistry **283:**1419-1427.
83. **Leigh, J. A., and J. A. Dodsworth.** 2007. Nitrogen regulation in bacteria and archaea. Annual review of microbiology **61:**349-377.
84. **Leistikow, R. L., R. A. Morton, I. L. Bartek, I. Frimpong, K. Wagner, and M. I. Voskuil.** 2010. The *Mycobacterium tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. Journal Bacteriology **192:**1662-1670.
85. **Lemassu, A., V. V. Levy-Frebault, M. A. Laneelle, and M. Daffe.** 1992. Lack of correlation between colony morphology and lipooligosaccharide content in the *Mycobacterium tuberculosis* complex. Journal of general microbiology **138:**1535-1541.
86. **Leung, A. N.** 1999. Pulmonary tuberculosis: the essentials. Radiology **210:**307-322.

87. **Lin, W., V. Mathys, E. L. Ang, V. H. Koh, J. M. Martinez Gomez, M. L. Ang, S. Z. Zainul Rahim, M. P. Tan, K. Pethe, and S. Alonso.** 2012. Urease activity represents an alternative pathway for *Mycobacterium tuberculosis* nitrogen metabolism. Infection and immunity **80:**2771-2779.
88. **Loebel, R. O., E. Shorr, and H. B. Richardson.** 1933. The Influence of Adverse Conditions upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli. Journal of bacteriology **26:**167-200.
89. **Lorenz, M. C., and J. Heitman.** 1998. The MEP2 ammonium permease regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. The EMBO journal **17:**1236-1247.
90. **Lyon, R. H., W. H. Hall, and C. Costas-Martinez.** 1974. Effect of L-asparagine on growth of *Mycobacterium tuberculosis* and on utilization of other amino acids. Journal of bacteriology **117:**151-156.
91. **Lyon, R. H., W. H. Hall, and C. Costas-Martinez.** 1970. Utilization of Amino Acids During Growth of *Mycobacterium tuberculosis* in Rotary Cultures. Infection and immunity **1:**513-520.
92. **Mahmood, A., S. Srivastava, S. Tripathi, M. A. Ansari, M. Owais, and A. Arora.** 2011. Molecular characterization of secretory proteins Rv3619c and Rv3620c from *Mycobacterium tuberculosis* H37Rv. The FEBS journal **278:**341-353.
93. **Malen, H., F. S. Berven, K. E. Fladmark, and H. G. Wiker.** 2007. Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. Proteomics **7:**1702-1718.
94. **Malen, H., S. Pathak, T. Softeland, G. A. de Souza, and H. G. Wiker.** 2010. Definition of novel cell envelope associated proteins in Triton X-114 extracts of *Mycobacterium tuberculosis* H37Rv. BMC microbiology **10:**132.
95. **Malm, S., Y. Tiffert, J. Micklinghoff, S. Schultze, I. Joost, I. Weber, S. Horst, B. Ackermann, M. Schmidt, W. Wohlleben, S. Ehlers, R. Geffers, J. Reuther, and F. C. Bange.** 2009. The roles of the nitrate reductase NarGHJI, the nitrite reductase NirBD and the response regulator GlnR in nitrate assimilation of *Mycobacterium tuberculosis*. Microbiology **155:**1332-1339.
96. **Manning, J. M., S. Moore, W. B. Rowe, and A. Meister.** 1969. Identification of L-methionine S-sulfoximine as the diastereoisomer of L-methionine SR-sulfoximine that inhibits glutamine synthetase. Biochemistry **8:**2681-2685.
97. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch.** 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology **150:**483-496.
98. **Mattison, K., and L. J. Kenney.** 2002. Phosphorylation alters the interaction of the response regulator OmpR with its sensor kinase EnvZ. The Journal of biological chemistry **277:**11143-11148.
99. **McAdam, R. A., S. Quan, D. A. Smith, S. Bardarov, J. C. Betts, F. C. Cook, E. U. Hooker, A. P. Lewis, P. Woollard, M. J. Everett, P. T. Lukey, G. J. Bancroft, W. R. Jacobs Jr, Jr., and K. Duncan.** 2002. Characterization of a *Mycobacterium tuberculosis* H37Rv transposon library reveals insertions in 351 ORFs and mutants with altered virulence. Microbiology **148:**2975-2986.
100. **Merrick, M. J., and R. A. Edwards.** 1995. Nitrogen control in bacteria. Microbiological Reviews **59:**604-622.
101. **Meya, D. B.** 2007. The TB pandemic: an old problem seeking new solutions. Journal of Internal Medicine **261:**309-329.
102. **Miles, A. A., S. S. Misra, and J. O. Irwin.** 1938. The estimation of the bactericidal power of the blood. J Hyg (Lond) **38:**732-749.
103. **Muller, T., J. Strosser, S. Buchinger, L. Nolden, A. Wirtz, R. Kramer, and A. Burkovski.** 2006. Mutation-induced metabolite pool alterations in *Corynebacterium*

glutamicum: towards the identification of nitrogen control signals. Journal Biotechnology **126:**440-453.
104. **Munoz-Elias, E. J., and J. D. McKinney.** 2006. Carbon metabolism of intracellular bacteria. Cell Microbiology **8:**10-22.
105. **Narlikar, L., and R. Jothi.** 2012. ChIP-Seq data analysis: identification of protein-DNA binding sites with SISSRs peak-finder. Methods Molecular Biology **802:**305-322.
106. **Neyrolles, O., R. Hernandez-Pando, F. Pietri-Rouxel, P. Fornes, L. Tailleux, J. A. Barrios Payan, E. Pivert, Y. Bordat, D. Aguilar, M. C. Prevost, C. Petit, and B. Gicquel.** 2006. Is adipose tissue a place for *Mycobacterium tuberculosis* persistence? PloS one **1:**e43.
107. **Niebisch, A., A. Kabus, C. Schultz, B. Weil, and M. Bott.** 2006. Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. The Journal of biological chemistry **281:**12300-12307.
108. **Nolden, L., C. E. Ngouoto-Nkili, A. K. Bendt, R. Kramer, and A. Burkovski.** 2001. Sensing nitrogen limitation in *Corynebacterium glutamicum*: the role of *glnK* and *glnD*. Molecular microbiology **42:**1281-1295.
109. **Nyfors, A.** 1997. Armauer Hansen (1841-1912): The life of the discoverer of the aetiology of leprosy. Journal of the European Academy of Dermatology and Venereology **9:**135-135.
110. **Nyka, W.** 1974. Studies on the effect of starvation on Mycobacteria. Infection and immunity **9:**843-850.
111. **O'Hare, H. M., R. Duran, C. Cervenansky, M. Bellinzoni, A. M. Wehenkel, O. Pritsch, G. Obal, J. Baumgartner, J. Vialaret, K. Johnsson, and P. M. Alzari.** 2008. Regulation of glutamate metabolism by protein kinases in mycobacteria. Molecular microbiology **70:**1408-1423.
112. **Ott, J. L.** 1960. Asparaginase from mycobacteria. Journal of bacteriology **80:**355-361.
113. **Parish, T., and N. G. Stoker.** 1998. Mycobacteria protocols. Humana Press, Totowa, N.J.
114. **Park, P. J.** 2009. ChIP-seq: advantages and challenges of a maturing technology. Nature Reviews Genetics **10:**669-680.
115. **Parra, M., N. Cadieux, T. Pickett, V. Dheenadhayalan, and M. J. Brennan.** 2006. A PE protein expressed by *Mycobacterium avium* is an effective T-cell immunogen. Infection and immunity **74:**786-789.
116. **Pashley, C. A., A. C. Brown, D. Robertson, and T. Parish.** 2006. Identification of the *Mycobacterium tuberculosis* GlnE promoter and its response to nitrogen availability. Microbiology **152:**2727-2734.
117. **Pelicic, V., J. M. Reyrat, and B. Gicquel.** 1996. Expression of the *Bacillus subtilis sacB* gene confers sucrose sensitivity on mycobacteria. Journal of bacteriology **178:**1197-1199.
118. **Pelicic, V., J. M. Reyrat, and B. Gicquel.** 1996. Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors. Molecular microbiology **20:**919-925.
119. **Pierre-Audigier, C., E. Jouanguy, S. Lamhamedi, F. Altare, J. Rauzier, V. Vincent, D. Canioni, J. F. Emile, A. Fischer, S. Blanche, J. L. Gaillard, and J. L. Casanova.** 1997. Fatal disseminated *Mycobacterium smegmatis* infection in a child with inherited interferon gamma receptor deficiency. Clinical Infectious Diseases **24:**982-984.
120. **Pullan, S. T., G. Chandra, M. J. Bibb, and M. Merrick.** 2011. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. BMC Genomics **12:**175.
121. **Radchenko, M. V., J. Thornton, and M. Merrick.** 2010. Control of AmtB-GlnK complex formation by intracellular levels of ATP, ADP, and 2-oxoglutarate. The Journal of biological chemistry **285:**31037-31045.
122. **Ramos, J. L., M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan, and R. Tobes.** 2005. The TetR family of

transcriptional repressors. Microbiology and molecular biology reviews : MMBR **69:**326-356.
123. **Rao, S. P., S. Alonso, L. Rand, T. Dick, and K. Pethe.** 2008. The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A **105:**11945-11950.
124. **Rastogi, N., E. Legrand, and C. Sola.** 2001. The mycobacteria: an introduction to nomenclature and pathogenesis. Revue scientifique et technique **20:**21-54.
125. **Raynaud, C., G. Etienne, P. Peyron, M. A. Laneelle, and M. Daffe.** 1998. Extracellular enzyme activities potentially involved in the pathogenicity of *Mycobacterium tuberculosis*. Microbiology **144 (Pt 2):**577-587.
126. **Read, R., C. A. Pashley, D. Smith, and T. Parish.** 2007. The role of GlnD in ammonia assimilation in *Mycobacterium tuberculosis*. Tuberculosis **87:**384-390.
127. **Reitzer, L.** 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. Annual Reviews Microbiology **57:**155-176.
128. **Reitzer, L. J., and B. Magasanik.** 1985. Expression of *glnA* in *Escherichia coli* is regulated at tandem promoters. Proc Natl Acad Sci U S A **82:**1979-1983.
129. **Ren, H., L. G. Dover, S. T. Islam, D. C. Alexander, J. M. Chen, G. S. Besra, and J. Liu.** 2007. Identification of the lipooligosaccharide biosynthetic gene cluster from *Mycobacterium marinum*. Molecular microbiology **63:**1345-1359.
130. **Roberts, D. L., D. W. Bennett, and S. A. Forst.** 1994. Identification of the site of phosphorylation on the osmosensor, EnvZ, of *Escherichia coli*. The Journal of biological chemistry **269:**8728-8733.
131. **Rodriguez-Garcia, A., C. Barreiro, F. Santos-Beneit, A. Sola-Landa, and J. F. Martin.** 2007. Genome-wide transcriptomic and proteomic analysis of the primary response to phosphate limitation in *Streptomyces coelicolor* M145 and in a *phoP* mutant. Proteomics **7:**2410-2429.
132. **Rodriguez-Garcia, A., A. Sola-Landa, K. Apel, F. Santos-Beneit, and J. F. Martin.** 2009. Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of *glnR, glnA, glnII* and *amtB* expression by the response regulator PhoP. Nucleic acids research **37:**3230-3242.
133. **Rowe, J. J., T. Ubbink-Kok, D. Molenaar, W. N. Konings, and A. J. Driessen.** 1994. NarK is a nitrite-extrusion system involved in anaerobic nitrate respiration by *Escherichia coli*. Molecular microbiology **12:**579-586.
134. **Runyon, E. H., W. G. Lawrence, G. P. Kubica, R. E. Buchanan, and N. E. Gibbons.** 1974. Mycobacteriaceae, p. 681-701, Bergey's Manual of Determinative Bacteriology, vol. 8. The Williams and Wilkins Company.
135. **Russell, D. G.** 2003. Phagosomes, fatty acids and tuberculosis. Nature cell biology **5:**776-778.
136. **Saini, V., A. Farhana, and A. J. Steyn.** 2012. *Mycobacterium tuberculosis* WhiB3: a novel iron-sulfur cluster protein that regulates redox homeostasis and virulence. Antioxidants & redox signaling **16:**687-697.
137. **Sakula, A.** 1983. Robert koch: centenary of the discovery of the tubercle bacillus, 1882. The Canadian veterinary journal. La revue veterinaire canadienne **24:**127-131.
138. **Sampson, S. L.** 2011. Mycobacterial PE/PPE proteins at the host-pathogen interface. Clinical and Developmental Immunology **2011:**497203.
139. **Sarada, K. V., N. A. Rao, and T. A. Venkitasubramanian.** 1980. Isolation and characterisation of glutamate dehydrogenase from *Mycobacterium smegmatis* CDC 46. Biochim Biophys Acta **615:**299-308.
140. **Saunders, B. M., and W. J. Britton.** 2007. Life and death in the granuloma: immunopathology of tuberculosis. Immunology and Cell Biology **85:**103-111.
141. **Self, W. T., A. M. Grunden, A. Hasona, and K. T. Shanmugam.** 1999. Transcriptional regulation of molybdoenzyme synthesis in *Escherichia coli* in response to molybdenum: ModE-molybdate, a repressor of the *modABCD* (molybdate transport) operon is a

secondary transcriptional activator for the *hyc* and *nar* operons. Microbiology **145 (Pt 1):**41-55.
142. **Sharma, S. K., A. Mohan, A. Sharma, and D. K. Mitra.** 2005. Miliary tuberculosis: new insights into an old disease. The Lancet Infectious Diseases **5:**415-430.
143. **Sherman, D. R., M. Voskuil, D. Schnappinger, R. Liao, M. I. Harrell, and G. K. Schoolnik.** 2001. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. Proc Natl Acad Sci U S A **98:**7534-7539.
144. **Shetty, N. D., M. C. Reddy, S. K. Palaninathan, J. L. Owen, and J. C. Sacchettini.** 2010. Crystal structures of the apo and ATP bound *Mycobacterium tuberculosis* nitrogen regulatory PII protein. Protein Science **19:**1513-1524.
145. **Shi, L., C. D. Sohaskey, B. D. Kana, S. Dawes, R. J. North, V. Mizrahi, and M. L. Gennaro.** 2005. Changes in energy metabolism of *Mycobacterium tuberculosis* in mouse lung and under *in vitro* conditions affecting aerobic respiration. Proc Natl Acad Sci U S A **102:**15629-15634.
146. **Shi, T., and J. Xie.** 2011. Molybdenum enzymes and molybdenum cofactor in mycobacteria. Journal of cellular biochemistry **112:**2721-2728.
147. **Shiloh, M. U., and P. A. DiGiuseppe Champion.** 2010. To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? Current opinion in microbiology **13:**86-92.
148. **Shinnick, T. M., and R. C. Good.** 1994. Mycobacterial taxonomy. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology **13:**884-901.
149. **Singh, A., D. K. Crossman, D. Mai, L. Guidry, M. I. Voskuil, M. B. Renfrow, and A. J. Steyn.** 2009. *Mycobacterium tuberculosis* WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. PLoS pathogens **5:**e1000545.
150. **Smeulders, M. J., J. Keer, R. A. Speight, and H. D. Williams.** 1999. Adaptation of *Mycobacterium smegmatis* to Stationary Phase. Journal of bacteriology **181:**270-283.
151. **Smith, J., J. Manoranjan, M. Pan, A. Bohsali, J. Xu, J. Liu, K. L. McDonald, A. Szyk, N. LaRonde-LeBlanc, and L. Y. Gao.** 2008. Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. Infection and immunity **76:**5478-5487.
152. **Smyth, G. K.** 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:**Article3.
153. **Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs, Jr.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Molecular microbiology **4:**1911-1919.
154. **Sohaskey, C. D., and L. G. Wayne.** 2003. Role of *narK2X* and *narGHJI* in hypoxic upregulation of nitrate reduction by *Mycobacterium tuberculosis*. Journal of bacteriology **185:**7247-7256.
155. **Stermann, M., L. Sedlacek, S. Maass, and F. C. Bange.** 2004. A promoter mutation causes differential nitrate reductase activity of *Mycobacterium tuberculosis* and *Mycobacterium bovis*. Journal of bacteriology **186:**2856-2861.
156. **Stock, A. M., V. L. Robinson, and P. N. Goudreau.** 2000. Two-component signal transduction. Annual review of biochemistry **69:**183-215.
157. **Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, and et al.** 1991. New use of BCG for recombinant vaccines. Nature **351:**456-460.
158. **Strosser, J., A. Ludke, S. Schaffer, R. Kramer, and A. Burkovski.** 2004. Regulation of GlnK activity: modification, membrane sequestration and proteolysis as regulatory principles in the network of nitrogen control in *Corynebacterium glutamicum*. Molecular microbiology **54:**132-147.

159. **Suzuki, A., and D. B. Knaff.** 2005. Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. Photosynthesis research **83:**191-217.
160. **Swaminathan, S., H. M. Ellis, L. S. Waters, D. Yu, E. C. Lee, D. L. Court, and S. K. Sharan.** 2001. Rapid engineering of bacterial artificial chromosomes using oligonucleotides. Genesis **29:**14-21.
161. **Tan, M. P., P. Sequeira, W. W. Lin, W. Y. Phong, P. Cliff, S. H. Ng, B. H. Lee, L. Camacho, D. Schnappinger, S. Ehrt, T. Dick, K. Pethe, and S. Alonso.** 2010. Nitrate respiration protects hypoxic *Mycobacterium tuberculosis* against acid- and reactive nitrogen species stresses. PloS one **5:**e13356.
162. **Tang, Y. J., W. Shui, S. Myers, X. Feng, C. Bertozzi, and J. D. Keasling.** 2009. Central metabolism in *Mycobacterium smegmatis* during the transition from O2-rich to O2-poor conditions as studied by isotopomer-assisted metabolite analysis. Biotechnology letters **31:**1233-1240.
163. **Thomas, G., G. Coutts, and M. Merrick.** 2000. The *glnKamtB* operon. A conserved gene pair in prokaryotes. Trends in Genetics **16:**11-14.
164. **Tiffert, Y., M. Franz-Wachtel, C. Fladerer, A. Nordheim, J. Reuther, W. Wohlleben, and Y. Mast.** 2011. Proteomic analysis of the GlnR-mediated response to nitrogen limitation in *Streptomyces coelicolor* M145. Applied Microbiology and Biotechnology **89:**1149-1159.
165. **Tiffert, Y., P. Supra, R. Wurm, W. Wohlleben, R. Wagner, and J. Reuther.** 2008. The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes. Molecular microbiology **67:**861-880.
166. **Tondervik, A., H. R. Torgersen, H. K. Botnmark, and A. R. Strom.** 2006. Transposon mutations in the 5' end of *glnD*, the gene for a nitrogen regulatory sensor, that suppress the osmosensitive phenotype caused by otsBA lesions in *Escherichia coli*. Journal of bacteriology **188:**4218-4226.
167. **Tortoli, E.** 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clinical Microbiology Reviews **16:**319.
168. **Tortoli, E.** 2006. The new mycobacteria: an update. FEMS immunology and medical microbiology **48:**159-178.
169. **Trivedi, O. A., P. Arora, V. Sridharan, R. Tickoo, D. Mohanty, and R. S. Gokhale.** 2004. Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature **428:**441-445.
170. **Truman, R., and P. E. Fine.** 2010. 'Environmental' sources of *Mycobacterium leprae*: issues and evidence. Leprosy review **81:**89-95.
171. **Tullius, M. V., G. Harth, and M. A. Horwitz.** 2003. Glutamine synthetase GlnA1 is essential for growth of *Mycobacterium tuberculosis* in human THP-1 macrophages and guinea pigs. Infection and immunity **71:**3927-3936.
172. **Tullius, M. V., G. Harth, and M. A. Horwitz.** 2001. High extracellular levels of *Mycobacterium tuberculosis* glutamine synthetase and superoxide dismutase in actively growing cultures are due to high expression and extracellular stability rather than to a protein-specific export mechanism. Infection and immunity **69:**6348-6363.
173. **Ulrichs, T., and S. H. Kaufmann.** 2006. New insights into the function of granulomas in human tuberculosis. Journal of Pathology **208:**261-269.
174. **van Kessel, J. C., and G. F. Hatfull.** 2008. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. Molecular microbiology **67:**1094-1107.
175. **van Kessel, J. C., and G. F. Hatfull.** 2008. Mycobacterial recombineering. Methods in molecular biology (Clifton, N.J **435:**203-215.
176. **van Kessel, J. C., and G. F. Hatfull.** 2007. Recombineering in *Mycobacterium tuberculosis*. Nature methods **4:**147-152.

177. **van Kessel, J. C., L. J. Marinelli, and G. F. Hatfull.** 2008. Recombineering mycobacteria and their phages. Nature Reviews Microbiology **6:**851-857.
178. **Villarino, A., R. Duran, A. Wehenkel, P. Fernandez, P. England, P. Brodin, S. T. Cole, U. Zimny-Arndt, P. R. Jungblut, C. Cervenansky, and P. M. Alzari.** 2005. Proteomic identification of *M. tuberculosis* protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. Journal of molecular biology **350:**953-963.
179. **Voskuil, M. I., D. Schnappinger, K. C. Visconti, M. I. Harrell, G. M. Dolganov, D. R. Sherman, and G. K. Schoolnik.** 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. Journal of Experimental Medicine **198:**705-713.
180. **Walters, S. B., E. Dubnau, I. Kolesnikova, F. Laval, M. Daffe, and I. Smith.** 2006. The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Molecular microbiology **60:**312-330.
181. **Wang, J., and G. P. Zhao.** 2009. GlnR positively regulates *nasA* transcription in *Streptomyces coelicolor*. Biochem Biophys Res Commun **386:**77-81.
182. **Warner, D. F., G. Etienne, X. M. Wang, L. G. Matsoso, S. S. Dawes, K. Soetaert, N. G. Stoker, J. Content, and V. Mizrahi.** 2006. A derivative of *Mycobacterium smegmatis* mc(2)155 that lacks the duplicated chromosomal region. Tuberculosis **86:**438-444.
183. **Wayne, L. G.** 1994. Dormancy of *Mycobacterium tuberculosis* and latency of disease. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology **13:**908-914.
184. **Wayne, L. G., and L. G. Hayes.** 1996. An *in vitro* model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. Infection and immunity **64:**2062-2069.
185. **Wayne, L. G., and L. G. Hayes.** 1998. Nitrate reduction as a marker for hypoxic shiftdown of *Mycobacterium tuberculosis*. Tuberculosis and Lung Disease **79:**127-132.
186. **Weber, I., C. Fritz, S. Ruttkowski, A. Kreft, and F. C. Bange.** 2000. Anaerobic nitrate reductase (*narGHJI*) activity of *Mycobacterium bovis* BCG *in vitro* and its contribution to virulence in immunodeficient mice. Molecular microbiology **35:**1017-1025.
187. **Weiss, V., G. Kramer, T. Dunnebier, and A. Flotho.** 2002. Mechanism of regulation of the bifunctional histidine kinase NtrB in *Escherichia coli*. Journal of molecular microbiology and biotechnology **4:**229-233.
188. **WHO.** 2009. Global Tuberculosis Control A short update to the 2009 report. World Health Organization.
189. **WHO.** 2012. Global Tuberculosis Report 2012.
190. **WHO.** 2012. Leprosy Fact sheet N°101.
191. **WHO.** 2012. Multidrug-resistant tuberculosis (MDR-TB) 2012 Update.
192. **WHO.** 2006. The Stop TB Strategy. World Health Organization.
193. **Williams, K. J., M. H. Bennett, G. R. Barton, V. A. Jenkins, and B. D. Robertson.** 2013. Adenylylation of mycobacterial Glnk (PII) protein is induced by nitrogen limitation. Tuberculosis.
194. **Williams, K. J., and K. Duncan.** 2007. Current Strategies for Identifying and Validating Targets for New Treatment-Shortening Drugs for TB. Current Molecular Medicine **7:**297-307.
195. **Wirén, N. M., M.** 2004. Regulation and function of ammonium carriers in bacteria, fungi and plants. Trends in Current Genetics**:**95–120.
196. **Wolanin, P. M., D. J. Webre, and J. B. Stock.** 2003. Mechanism of phosphatase activity in the chemotaxis response regulator CheY. Biochemistry **42:**14075-14082.
197. **Wray, L. V., Jr., M. R. Atkinson, and S. H. Fisher.** 1991. Identification and cloning of the *glnR* locus, which is required for transcription of the *glnA* gene in *Streptomyces coelicolor* A3(2). Journal of bacteriology **173:**7351-7360.
198. **Wray, L. V., Jr., and S. H. Fisher.** 1993. The *Streptomyces coelicolor glnR* gene encodes a protein similar to other bacterial response regulators. Gene **130:**145-150.

199. **Yabu, K.** 1970. Amino acid transport in *Mycobacterium smegmatis*. Journal of bacteriology **102:**6-13.
200. **Yakunin, A. F., and P. C. Hallenbeck.** 2002. AmtB is necessary for NH(4)(+)-induced nitrogenase switch-off and ADP-ribosylation in *Rhodobacter capsulatus*. Journal of bacteriology **184:**4081-4088.
201. **Yoshida, T., L. Qin, L. A. Egger, and M. Inouye.** 2006. Transcription regulation of *ompF* and *ompC* by a single transcription factor, OmpR. The Journal of biological chemistry **281:**17114-17123.
202. **Zheng, L., D. Kostrewa, S. Berneche, F. K. Winkler, and X. D. Li.** 2004. The mechanism of ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. Proc Natl Acad Sci U S A **101:**17090-17095.
203. **Zimmer, D. P., E. Soupene, H. L. Lee, V. F. Wendisch, A. B. Khodursky, B. J. Peter, R. A. Bender, and S. Kustu.** 2000. Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. Proc Natl Acad Sci U S A **97:**14674-14679.
204. **Zink, A., C. J. Haas, U. Reischl, U. Szeimies, and A. G. Nerlich.** 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. Journal of medical microbiology **50:**355-366.
205. **Zundel, C. J., D. C. Capener, and W. R. McCleary.** 1998. Analysis of the conserved acidic residues in the regulatory domain of PhoB. FEBS letters **441:**242-246.

APPENDIX

Name	Application	Sequence (5' – 3')
SEQUENCING		
T7F	Sequencing in pET28b_T7_Promoter_F	TAATACGACTCACTATAGGG
T7R	Sequencing in pET28b_T7_Terminator_R	GCTAGTTATTGCTCAGCGG
M13R	Sequencing in pCR2.1- TOPO	CAGGAAACAGCTATGAC
M13F	Sequencing in pCR2.1-TOPO	TAAAACGACGGCCAG
CHAPTER 4		
GlnR_Down_R	Directional cloning of *glnR* upstream region into pYUB854	GACTAT AAGCTT GTGACGACGTAGATG
GlnR_Down_F	"	GACTAT ACTAGT GTAACCGAGGCCACG
GlnR_Up_R	Directional cloning of *glnR* downstream region into	GACTAT TCTAGA TAGCAGTAGATCCAAC
GlnR_Up_F	pYUB854	GACTAT CTTAAG TGCCGCCATCGATGAGC
GlnR_Down_Seq	Confirmation of *glnR* deletion – downstream region	GACTATACTAGTACAGGCGCGAGGCGTCAAC
Hyg_out_1	"	GCATGCAAGCTCAGGATGTC
GlnR_Up_Seq	Confirmation of *glnR* deletion – upstream region	AGCTTACCCAATGACCCTCG
Hyg_out_2	"	TTCGAGGTGTTCGAGGAGAC
HygS_Repair	Converts Hyg^S into Hyg^R	GCCAGCGGCTCCCAGAATTCCTGGTCGTTCCGCAGGCTCGCGTAGG AATCTCCGAATCAATACGGTCGAGAAGTAACAGGGATTCTTGTGTC ACAGCGG
GlnR_Point_mut	GlnR aspartic acid 48 to alanine (change in **bold**)	GTAGTGCCGACGTCGCGATCGTCG**CG**GCTCGCACAGATCTGGCCGC CGCG
MAMA_PCR_F	Screening for GlnR_D48 A mutation	TAGTGCCGACGTCGCGATCGTCG**CG**
MAMA_PCR_R	"	GAGGTACTTGAGGAGCTCGAATTCCTTG
GlnR_D48A_SeqF	Confirmation of the GlnR_D48A mutation	CTACATGTCAGTCATGAATC
GlnR_D48A_SeqR	"	TTCCCGCAGCACTTGGTC
GlnR_reg_F	Amplification of 1200bp region containing *glnR*	ACATTGTTGCCCACGAGAC
GlnR_reg_R	"	GAGGTTGAGGTATCCGAC

CHAPTER 5&6

MS_His-GlnR_F	Cloning *M. smegmatis glnR* with N terminal His_6 tag into pET28b	GACTATCATATGTTGGATCTACTGCTACTG
MS_His-GlnR_R	"	GACTATCTCGAGTCACTGACTGGTCAACCG
TB_His-GlnR_F	Cloning *M. tuberculosis glnR* with N terminal His_6 tag into pET28b	GACTATCATATGTTGTTGGAGTTATTACTG
TB_His-GlnR_R	"	GACTATCTCGAGTCACTGACTGCGCAACGG
Peak 9F	Amplification of MS peak 9 for EMSA	GAGTGTTTGCGGGGCGTTAC
Peak 9R	"	TTTGTGTGAACCTCCTTGG
Peak 17F	Amplification of MS peak 17 for EMSA	CGTCGATGTGGCGCTGCAC
Peak 17R	"	GCGCTGCTGGTCATGG
Peak 21F	Amplification of MS peak 21 for EMSA	AGCTTGCCTACGAGCTCG
Peak 22R	"	AATGAGGGATGCTGCGAG
Peak 22F	Amplification of MS peak 22 for EMSA	CTACCGGACACACAACG
Peak 22R	"	AACGGTGTGCTTCCTCC
Peak 42F	Amplification of MS peak 42 for EMSA	CATGAGCGCCATCAACTTC
Peak 42R	"	GACGCGTCCATTCGGTTGTC
3224 F	Amplification of msmeg_3224 upstream region for EMSA and	GCCTGTTGCAGTTGATCG
3224 R	rate limiting PCR	GTACGGGTCGCGCACCTTGTC
Peak 1F	Amplification of MS peak 1 for rate limiting PCR	TTGTGGCCTGACTGTGGTCC
Peak 1R		AGGCTAAGAACCCGATATTG
Peak 34F	Amplification of MS peak 34 for rate limiting PCR	ATAGGCGCGTGCGGATGTC AACCCGATGTTGCGCCGAC
Peak 13F	Amplification of MS peak 13 for EMSA and rate limiting PCR	AAGCCGGATCCAGACGTG
Peak 13R	"	GCTCGATACCCAGGTTCTC
Peak 14F	Amplification of MS peak 14 for rate limiting PCR	ACTCGACAGGCGATCGGAAG
Peak 14R	"	GAAACAGCGTTTCTTAC
Peak 26F	Amplification of MS peak 26 for rate limiting PCR	GTCACGGCAAGGGTGGAC

Peak 26R	“	GTTGTGACCGGACACAC
Peak 32F	Amplification of MS peak 32 for rate limiting PCR	CGACAAGAGAAATGGCCGAG
Peak 32R	“	AAGGCAAGAGTGCGAATGAC
Peak 39F	Amplification of MS peak 39 for rate limiting PCR	AGATAACGGTCCGATAAC
Peak 39R	“	TTGCCGTCTACCTGCATG
Peak 40F	Amplification of MS peak 40 for rate limiting PCR	CAACAAACCCCGTGGTCAG
Peak 40R	“	GAATTTATCGTTTCGAG
Peak 44F	Amplification of MS peak 44 for rate limiting PCR	GATGTCTCGGCATCGAGCAAC
Peak 44R	“	CGATAACCGGTGTCGATC
Peak 49F	Amplification of MS peak 49 for rate limiting PCR	TCAGTGCTACCTCCAAG
Peak 49R	“	CGACGACCTCTACACC
Peak 2WT F	WT oligonucleotides for MS binding site analysis of peak 2 EMSA	GCAATCGCGGGGTAACGCCGTGGAAACAGAGCCTGCCT
Peak 2WT R	“	AGGCAGGCTCTGTTTCCACGGCGTTACCCCGCGATTGC
Peak 2GG F	AC to GG mutation for MS binding site analysis of peak 2 EMSA	GCAATCGCGGGGTA**GG**GCCGTGGAA**GG**AGAGCCTGCCT
Peak 2GG F	“	AGGCAGGCTCT**CC**TTCCACGGC**CC**TACCCCGCGATTGC
Peak 2G F	A to G mutation for MS binding site analysis of peak 2 EMSA	GCAATCGCGGGGTA**G**CGCCGTGGAA**G**CAGAGCCTGCCT
Peak 2G R	“	AGGCAGGCTCTG**C**TTCCACGGCG**C**TACCCCGCGATTGC
Peak 24WT F	WT oligonucleotides for MS binding site analysis of peak 24 EMSA	TCATGTCGAGGTTAATTTGTTCGTCACACACAGACATT
Peak 24WT R	“	AATGTCTGTGTGTGACGAACAAATTAACCTCGACATGA
Peak 24GG F	AT to GG mutation for MS binding site analysis of peak 24 EMSA	TCATGTCGAGGTTA**GG**TTGTTCGTC**GG**ACACAGACATT
Peak 24GG R	“	AATGTCTGTGT**CC**GACGAACAA**CC**TAACCTCGACATGA
Peak 24G F	A to G mutation for MS binding site analysis of peak 24 EMSA	TCATGTCGAGGTTA**G**TTTGTTCGTC**G**CACACAGACATT
Peak 24G R	“	AATGTCTGTGTG**C**GACGAACAAA**C**TAACCTCGACATGA
Peak 2long F	Oligonucleotides increasing ACn9AC distance to ACn12AC	GCAATCGCGGGGTAAC**GCCGTCCCGGAA**ACAGAGCCTGCCT

Peak 2long R	"	AGGCAGGCTCTGT**TTCCGGGACGGC**GTTACCCCGCGATTGC
Peak 2short F	Oligonucleotides decreasing ACn9AC distance to ACn6AC	GCAATCGCGGGGTAAC**GTGGAA**ACAGAGCCTGCCT
Peak 2short R	"	AGGCAGGCTCTGT**TTCCAC**GTTACCCCGCGATTGC
CHAPTER 7		
TB_peak18F	Amplification of TB peak 18 for EMSA	ACCATCCCGTCAGCCGGCCACAC
TB_peak18R	"	GTACGTCCACAATCGAAGGA
TB_peak13R	Amplification of TB peak 13 for EMSA	GCTAAATCCCACCAGCATG
TB_peak13R	"	CACAGACTCCATCTGTTG
TB_1360R	Amplification of Rv1360 upstream region for EMSA	ACTCCCTGCGGCAAGGTG
TB_1360R	"	GACATACGTGGATGTGCTG
TB_peak17F	Amplification of TB peak 17 for EMSA	GATCTTGTCGTAGATGCTG
TB_peak17R	"	CATGAGCTGATGAATGGAGT
TB_peak20F	Amplification of TB peak 20 for EMSA	GATATTGCCCGTCAGTC
TB_peak20R	"	TTCGGCATGCCACCGGTTAC
TB_nirB_F	Amplification of TB peak 1 for rate limiting PCR	CTTCGTTGTGAGTTAGC
TB_nirB_R	"	ATCGCCGAATGTGACGCAC
TB_peak2F	Amplification of TB peak 2 for rate limiting PCR	CGAAGCAATGCGCACAG
TB_peak2R	"	TGGCCTACGTCTAGCG
TB_peak10F	Amplification of TB peak 10 for rate limiting PCR	GACAACACCAAGTTCGC
TB_peak10R	"	ACGGCAGGTCGGTGTAGC
TB_peak11F	Amplification of TB peak 11 for rate limiting PCR	GCTTGCCACCGCCGAC
TB_peak11R	"	ACCGACAGCGAGTAGGC
TB_peak23F	Amplification of TB peak 23 for rate limiting PCR	TCGAAGCGACCAGGCAG
TB_peak23R	"	ACCTCCGTGTTGCCTGC

Appendix 1. Primer sequences used in this study.

Underlined sequences are restriction sites used for cloning. Mutated residues are highlighted in bold.

PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

Jenkins, V.A., Barton, G.R., Robertson, B.D., and Williams, K.J. (2013). Genome Wide Analysis of the Complete GlnR Nitrogen-response Regulon in *Mycobacterium smegmatis.* BMC Genomics: 14: 301.

Williams, K.J, Bryant, W.A., **Jenkins, V.A.,** Barton, G.R., Witney, A.A., Pinney, J.W., and Robertson, B.D. (2013). Deciphering the Response of *Mycobacterium smegmatis* to Nitrogen Stress Using Bipartite Active Modules. Submitted BMC Genomics.

Willams, K. J., Bennett, M., Barton, G., **Jenkins, V. A**., and Robertson, B. D. (2013). Adenylylation of Mycobacterial Glnk (PII) Protein is Induced by Nitrogen Limitation. Tuberculosis: 93(2):198-206.

Jenkins, V.A., Robertson, B.D., and Williams, K.J. (2012). Aspartate D48 is Essential for the GlnR Mediated Transcriptional Response to Nitrogen Limitation in *Mycobacterium smegmatis*. FEMS Microbiology Letters: 330(1):38-45.

Behrends, V., Williams, K.J., **Jenkins, V.A.**, Robertson, B.D., Bundy, J.G. (2012). Free Glucosylglycerate is a Novel Marker of Nitrogen Stress in *Mycobacterium smegmatis*. Journal of Proteome Research: 11(7):3888-96

ORAL PRESENTATIONS

Acid Fast Club Meeting. Brighton and Hove Medical School, Brighton, UK. (2012). Genome Wide Analysis of the GlnR Regulon in *M. smegmatis. *Prize for best talk**

Departmental Work in Progress Talk. Imperial College London, UK. (2012). Genome Wide Analysis of the GlnR Regulon in *M. smegmatis.*

POSTER PRESENTATIONS

EMBO Tuberculosis Conference: Institut Pasteur, Paris, France. (2012). Genome Wide Analysis of the GlnR Regulon in *M. smegmatis.*

Department of Medicine Young Scientist Day: Imperial College London, Hammersmith Campus. (2012). Genome Wide Analysis of the GlnR Regulon in *M. smegmatis.*

Keystone Symposia: Mycobacteria: Physiology, Metabolism and Pathogenesis- Back to the Basics (J4): Vancouver, Canada. (2011). Analysis of AmtR binding in *M. smegmatis.*