The Characterisation of N-Acetyltransferase (NAT) in Mycobacterium tuberculosis

by Carolyn Sholto-Douglas-Vernon

Thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the department of Medical Biochemistry at the University of Stellenbosch

April 2005

Supervisor: Professor P. van Helden

Co-supervisor: Professor T. Victor

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or part submitted it at any university for a degree.

Signature: ..... Date: 14.02.05

SUMMARY

A gene coding for Arylaminie N-acetyltransferase (NAT) has been found in *Mycobacterium tuberculosis*, the casual agent of tuberculosis (TB). N-acetyltransferase acetylates and inactivates isoniazid (INH), which is a front line drug used in TB therapy. A guanine to adenine SNP at base-pair 619 (G619A) has previously been identified in this gene, which results in a glycine to arginine change at amino acid 207 (G207R) (Upton *et al.* 2001). In this study the *nat* gene was further characterised. The frequency of the G619A SNP was analysed in 37 *M. tuberculosis* strain families found in the Western Cape Province of South Africa, and it was found that the G619A SNP is conserved in two strain families (strain family 3 and strain family 28). Further sequence analysis identified a new thymine to cytosine SNP at base-pair 529 (T529C) resulting in a tyrosine to histidine change at amino acid 177 (Y177H). This SNP was found only in isolates from strain family 3. These results imply that these SNPs may be used in epidemiology studies to classify isolates into these strain families.

Using Real Time PCR, the expression of *nat* in *M. bovis* BCG and *M. tuberculosis* (reference strain H37Rv) was determined over a 7 and 28 day growth cycle, respectively. Using 16S rRNA as an endogenous control, the *nat* gene was shown to be expressed early during the growth curve and reach its maximum expression level at approximately mid-log phase. The expression of *nat* was induced in drug susceptible *M. tuberculosis* isolates (reference strain H37Rv and isolate 1430 containing both SNPs) exposed to INH at a concentration of 0.01μg/ml, but minimal change in expression was observed in resistant isolates (isolate 816) exposed to INH at the same concentration. *Mycobacterium bovis* BCG cultures exposed to INH, at a final concentration of 0.28μg/ml, showed an increase in protein production. The increase of *nat* mRNA and NAT protein in *M. tuberculosis* and *M. bovis* BCG, respectively, implies that INH affects the expression of NAT.

The NAT protein was localised to all fractions of the cell in *Mycobacterium smegmatis*, *M. bovis* BCG and *M. tuberculosis*, using the Western blot technique. However, protein fractions from the cell envelope region showed a protein (detected with specific NAT antibodies) that ran at a higher molecular weight (MW). This implies that the cytosolic hydrophilic NAT undergoes some type of post-translational process that may make it hydrophobic, and enable it to pass into the cell envelope region.

These results show for the first time how *nat* is expressed during the entire growth cycle of *M. tuberculosis* and *M. bovis* BCG. It was shown that *nat* is expressed early during the growth cycle of the bacterium reaching maximum expression levels at mid-log phase. These results are in concordance with those obtained using *M. smegmatis nat* mutants, which taken together, show that early expression of *nat* is important for early growth and development of mycobacteria. The results in this study also showed that NAT appeared to be translocated into the cell envelope of the bacterium, implying that NAT may be involved in one of the pathways needed for complete formation of the cell envelope. These results suggest that NAT may be an important target for drug development, as inhibitors of NAT could result in hindered growth and hence spread of the bacterium within its host. Inhibitors may also result in the incomplete development of the cell wall, enabling the host to combat the disease using its own immune system.

OPSOMMING

'n Geen wat kodeer vir Arielamien N-asetieltransferase (NAT), is gevind in *Mycobacterium tuberculosis*, die bacterium wat tuberkulose (TB) veroorsaak. NAT asetileer en inaktiveer isoniasied (INH), wat 'n belangrike antibiotika is in die behandeling van TB. 'n Enkel guanien na adenien nukleotied verandering (SNP) by basispaar 619 (G619A) is vroeër in hierdie geen gevind. Die basispaar verandering gee aanleiding tot 'n glisien na arginien aminosuur verandering by posisie 207 (G207R) in die proteïen (Upton *et al.* 2001). In hierdie studie is die *nat* geen verder gekarakteriseer. Die frekwensie van die G619A SNP is ondersoek in 37 verskillende *M. tuberculosis* stamfamilie wat geisoleer was van TB pasiente in die Weskaap provinsie van Suid-Afrika. Daar is gevind dat die G619A SNP teenwoordig is in twee stamfamilie (stamfamilie 3 en 28). Verdere DNA volgorde bepalings het 'n nuwe timien na sitosien SNP by basisparr 529 (T529C) getoon, wat aanleiding gee tot 'n tirosien na histidien aminosuur verandering by posisie 177 (Y177H) in die proteïen. Hierdie SNP is slegs gevind in isolate van stamfamilie 3. Die resultate dui daarop dat hierdie SNP's gebruik kan word in epidimiologie studies om isolate van hierdie stamfamilies te identifiseer.

Die uitdrukking van *nat* in *M. bovis* BCG en *M. tuberculosis* (die verwysingsstam H37Rv is gebruik) is gedoen met "Real-Time PCR" na groeisiklusse van 7 en 28 dae van die bakterie, onderskeidilik. Die 16S rRNA is as endogene kontrole gebruik om te wys dat die *nat* geen vroeg in die groeikurwe uitgedruk word en dat die maksimum uitdrukking verekry is by die middle eksponensiële van groei. Die uitdrukking van *nat* is geïnduseer deur INH teen 'n konsentrasie van 0.01 µg/ml toe te dien in middlevatbare *M. tuberculosis* isolate (H37Rv en isolaat 1430 wat albei SNP's bevat). 'n Minimale verandering in uitdrukking is gevind in weerstandige isolate (isolaat 816) wat blootgestel is aan INH. *M. bovis* BCG wat blootgestel was aan 'n finale konsentrasie van 0.28 µg/ml het a toename van proteïen produksie getoon. Die toename van *nat* mRNA in *M. tuberculosis* en NAT proteïen in *M. bovis* BCG impliseer dat INH die uitdrukking van NAT affekteer.

n' "Western blot" tegniek is gebruik om te wys dat die NAT proteïen gevind is in alle sel fraksies van *M. smegmatis, M. bovis* BCG en *M. tuberculosis.* Proteïen fraksies van die selomhulsel (selwand, selmembraan en intermembraan spasies) het egter 'n proteïen van 'n hoer moleklêre gewig getoon (gevind deur NAT-spesifieke antiliggaampies te gebruik). Dit mag impliseer dat die

hidrofiliese NAT proteïen in die sitosol 'n na-translasie proses ondergaan wat die proteïen hidrofobies maak sodat dit na die selomhulsel kan beweeg.

Met hierdie resultate word vir die eerste keer gewys hoe die *nat* geen gedurende die hele groei siklus van *M. tuberculosis* en *M. bovis* BCG uitgedruk word. Daar is gewys dat *nat* vroeg in die groeisiklus van die bacterium uitgedruk word en die hoogste vlakke word bereik in die middle eksponensiële groei fase. Hierdie resultate stem ooreen met wat verkry is wanneer *M. smegmatis* *nat*-mutante gebruik is en dit dui daarop dat vroë uitdrukking van *nat* belangrik is vir groei en ontwikkling van mycobakterië. Die resultate wys ook dat NAT betrokke is by een van die paaie wat benodig word vir voledige vorming van die selomhulsel. Resultate van hierdie studie dui daarop dat NAT 'n belangrike teiken vir medikasie ontwikkeling kan wees, want inhibeerders mag veroorsaak dat die selwand onvolledig ontwikkel, wat die gasheer sal help om die bacterium dan met sy imuunsisteem te beveg.

ACKNOWLEDGEMENT

I would like to acknowledge:

- My two supervisors at the University of Stellenbosch, Prof. P. van Helden and Prof. T. Victor, for their patience, understanding and encouragement throughout this project.

- Prof. van Helden for giving me the opportunity to do some of my research at the University of Oxford, Department of Pharmacology.

- Prof. E. Sim for allowing me to spend time within her lab at the University of Oxford, Department of Pharmacolgy.

- Prof. E. Sim for all her help and guidance during my stay in Oxford.

- All my colleagues both at Stellenbosch and Oxford for their continual encouragement and willingness to help whenever the need arose.

ABBREVIATIONS

α	alpha
β	beta
γ	gamma
°C	degrees Celsius
>	greater than
<	less than
µl	microlitre
µg	microgram
µM	microMolar
#	number
%	percentage
3D	three dimensional
π	pi
A_x	absorbance at the wavelength of x nanometres
AcCoA	Acetyl Coenzyme A
ACP	enoyl-Acyl Carrier Protein
ADC	Albumin-Dextrose Complex
AG	Arabinogalactan
aGPL	apolar glycopeptidolipid
AIDs	Acquired Immunodeficiency Syndrome
Arg	Arginine
Asp	Aspartate
Ave	Average
BCG	Bovine Calmette-Guerin
bp	base pair
BSA	Bovine Serum Albumin
cDNA	copy Deoxynucleic Acid
CFU	Colony Forming Unit
C_T	threshold cycle
Cys	Cysteine
ddH_2O	double distilled water

dH_2O	distilled water
DMSO	Dimethyl Sulfoxide
DNA	Deoxynucleic Acid
DNAse	Deoxyribonuclease
dNTPs	Deoxynucleoside Triphosphates
DOT	Directly Observed Therapy
DR	Direct Repeats
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ED	electron dense
EDTA	Ethylenedinitro Tetra acetic Acid
ET	electron transparent
EtOH	alcohol
F1	Fragment 1
F2	Fragment 2
FAS II	Fatty Acid Synthase II
FB	Fractionation Buffer
FISH	Fluorescent *in situ* Hybridisation
g	gram/ gravitational force (see in context)
G207R	Glycine to Arginine change at amino acid 207
G619A	Guanine to Adenine change at nucleotide 619
Gly	Glycine
GPL	glycopeptidolipid
H_2O	water
HCl	Hydrocholoric Acid
His	Histidine
HIV	Human Immunodeficiency Virus
INH	isoniazid
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kas	Keto-Acyl Synthase
KatG	catalase-peroxidase
kb	kilo base
KCl	potassium chloride
kDa	kilo Daltons

LAM	Lipoarabinomannan
LB medium	Lauria Bertani medium
LB	Lysate Buffer
l	litre
LM	Lipomannan
M	Molar
mAGP	mycolyl Arabinogalactan Peptidoglycan
M. bovis	*Mycobacterium bovis*
MCS	Multiple Cloning Sites
MDR TB	Multi-Drug Resistant Tuberculosis
mg	milligram
$MgCl_2$	Magnesium Chloride
MIC	Minimum Inhibitory Concentration
min	minute
MIRU-VNTR	Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats
ml	millilitre
M. leprae	*Mycobacterium leprae*
M. loti	*Mycobacterium loti*
mM	milliMolar
MP	Membrane Proteins
mRNA	messenger Ribonucleic Acid
M. smegmatis	*Mycobacterium smegmatis*
M. smeg	*Mycobacterium smegmatis*
MSNAT	*Mycobacterium smegmatis* N-acetyltransferase
M. tuberculosis	*Mycobacterium tuberculosis*
M. tb	*Mycobacterium tuberculosis*
MW	Molecular Weight
n	number
NaCl	sodium chloride
NADH	reduced Nicotinamide Adenine Dinucleotide
NAD^+	Nicotinamide Adenine Dinucleotide
NaOH	sodium hydroxide
nat	N-acetyltransferase gene

NAT	N-acetyltransferase protein
ng	nanograms
NH_4^+	ammonium ion
nm	nanometre
nM	nanoMolar
nsSNP	nonsynonymous SNPs
NTC	no template control
OADC	Oleic Acid-Albumin-Dextrose-Catalase
OD_{600}	Optical Density at 600 nanometres
ORF	Open Reading Frame
Ori E	Origin of replication for *Escherichia Coli*
pABA	para aminobenzoic acid
pABA-Glu	para aminobenzoyl-glutamate
PAGE	Polyacrylamide Gel Electrophoresis
PAS	para aminosalicyclic acid
P. aeruginosa	*Pseudomonas aeruginosa*
PBS	Phosphate-Buffer Saline (140mM NaCl/ 10mM Na Phosphate pH 7.5)
PCR	Polymerase Chain Reacction
PG	peptidoglycan
PGRS	Polymorphic GC-Rich Sequence
Phe	Phenylalanine
PIMs	Phosphatidylinosityl Mannoside
PM	Plasma Membrane
Prof.	Professor
PZA	Pyrazinamide
Q-PCR	Quantitative Polymerase Chain Reaction
Qty	Quantity
RFLP	Restriction Fragment Length Polymorphism
RIF	Rifampin
Rn	normalised reporter signal
RNA	Ribonucleic Acid
RNase	Ribonuclease
rRNA	ribosomal Ribonucleic Acid
rpm	revolutions per minute

RT	reverse transcription
S. aureus	*Staphylococcus aureus*
SAM	S-adenosylmethionine
sec	second
SDS	Sodium Dodecyl Sulphate
S. hygroscopicus	*Streptomyces hygroscopicus*
Staph. epidermis	*Staphylococcus epidermis*
SNP	Single Nucleotide Polymorphism
STDEV	Standard Deviation
STNAT	*Salmonella typhimurium* N-acetyltransferase
sSNP	synonymous SNP
S. typhimurium	*Salmonella typhimurium*
T529C	Thymine to cytosine change at nucleotide 529
TB	Tuberculosis
tbNAT	*Mycobacterium tuberculosis* N-acetyltransferase protein
tbnat	*Mycobacterium tuberculosis* N-acetyltransferase gene
TBS	Tris-Buffer Saline
TBST	Tris-Buffer Saline plus 0.05% Tween 20
TDM	Trehalose Dimycolates
Thin Layer Chromatography	TLC
Thr	Threonine
TIGR	The Institute for Genomic Research
TMM	Trehalose Monomycolates
Tris	Tromethamine, 2-Amino-2-(hydroxymethyl)-1,3-propanediol
Tyr	Tyrosine
U	Units
US	United States
V	Volts
V. parahaemolyticus	*Vibrio paraheamolyticus*
vs	versus
WHO	World Health Organisation
Y177H	Tyrosine to Histidine change at amino acid 177
ZN	Ziehl Nielson

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1. Overview

Tuberculosis (TB) was declared a global emergency by the World Health Organisation (WHO) in 1993 (WHO 1999). A third of the global population is infected with *Mycobacterium tuberculosis* (WHO 1999), the causal agent of TB (Ellard *et al.* 1972). Of those infected, approximately 10% will develop active TB, however, this figure is rising due to the AIDs epidemic and the increase in drug resistant isolates (WHO 1999;Mukherjee et al. 2004). Around 30% of AIDs related deaths are due to TB (Grange and Zumla 2002). The emergence of multidrug resistant TB (MDR TB) is a threat to TB control (Grange and Zumla 2002;Mukherjee et al. 2004). MDR TB is treatable with a combination of drugs, however, this is very expensive and is not undertaken in many poor nations (Grange and Zumla 2002).

1-Isonicotinyl hydrazide, more commonly known as Isoniazid (INH), is an important component of effective multidrug therapy and prophylaxis for *M. tuberculosis* (Torres *et al.* 2000). In a recent survey of 35 countries, 7.3% of *M. tuberculosis* isolates were shown to be resistant to INH (Pablos-Mendez et al. 1998). An isolate of *M. tuberculosis* is regarded as multidrug-resistant (MDR TB) when it is resistant to at least INH and Rifampin (RIF) (WHO 2004). It has been reported that approximately 90% of RIF resistant isolates are also INH resistant (Ramaswamy and Musser 1998;Watterson et al. 1998;Piatek et al. 2000). This suggests that once INH resistance occurs, resistance to other drugs frequently follows, leading to MDR TB, which complicates treatment and control (Grange and Zumla 2002).

A single mutation in at least one of the following genes; *katG, inhA, kasA, ahpC* and *ndh*, can be related to resistance in many INH resistant *M. tuberculosis* clinical isolates characterised to date (Banerjee et al. 1994;Musser et al. 1996;Kelly et al. 1997;Victor et al. 1997;Mdluli et al. 1998;Lee et al. 2001). However, in approximately 25 to 50% of clinical isolates, INH resistance cannot be accounted for (Ramaswamy and Musser 1998;Slayden and Barry 2000). For example, a recent study on INH resistant isolates in Singapore showed that 37.5% of the isolates had no known causative mutations (Lee *et al.* 1999).

The discovery of the gene for human N-acetyltransferase 2 (NAT2) (Blum *et al.* 1990), which acetylates and hence inactivates INH, has important implications for INH as a therapeutic agent. Different polymorphisms in this gene have functional effects on the enzyme (conferring fast, intermediate or slow acetylator status to individuals) (Parkin *et al.* 1997). Identification of human NAT2 polymorphisms provide a rationale for the different therapeutic doses of INH needed for TB treatment in fast and slow acetylators (Ellard and Gammon 1976;Deguchi et al. 1990;Payton et al. 1999). A homologous gene encoding arylamine N-acetyltransferase (NAT) has been discovered in *M. tuberculosis* (Payton et al. 1999;Upton et al. 2001b), and a G619A single nucleotide polymorphism (SNP) has been identified in the open reading frame (ORF) of this gene, in some *M. tuberculosis* strains (Upton *et al.* 2001b).

Activation of INH in *M. tuberculosis* by catalase-peroxidase (encoded for by *katG*) involves oxidation of the hydrazine moiety, which cannot occur when INH is N-acetylated (Upton *et al.* 2001b). Recombinant NAT from *M. tuberculosis* (tbNAT) has been shown to N-acetylate INH *in vitro*, and when the *M. tuberculosis nat* gene (*tbnat*) is over expressed in *M. smegmatis*, the resistance to INH of the transformed organism increases three fold (Payton *et al.* 1999). These results suggest that *M. tuberculosis* NAT may be involved in INH resistance.

In the work presented here the *M. tuberculosis nat* gene and its protein was further investigated. To understand this bacterium and its complexities one needs to understand the biology of mycobacteria in general, and what makes *M. tuberculosis* such an effective pathogen.

1.2. Mycobacteria

There are over 40 recognised species of mycobacteria (Parish and Stoker 1998), which can affect most species of animals, including rodents, birds, fish and humans (Parish and Stoker 1998). Mycobacteria are gram-positive, rod-shaped bacteria of the Actinomycete family, and therefore are most closely related to the nocardia, corynebacteria, and streptomyces (Clark-Curtiss 1990). Mycobacteria are prototrophic and are able to build most of their required components from basic carbon (glucose or glycerol) and nitrogen (ammonium ion – $NH4^+$ or asparagines) sources. They are also acid-fast bacilli and can easily be identified using the Ziehl-Nielson (ZN) acid-fast stain. Mycobacteria fall naturally and taxonomically into two main groups: slow and fast-growers. The group of mycobacteria that cause most of the major human and animal pathogens are the slow-

growers, whereas the fast growers include non-pathogenic species e.g. *M. smegmatis*. *M. smegmatis* is widely used as a convenient, if imperfect, model organism for the study of mycobacterial pathogens (Parish and Stoker 1998).

Although mycobacteria affect most species of animals, much attention has been devoted to them since they are major human pathogens. TB, caused by *M. tuberculosis*, remains an extremely important infectious cause of mortality in the world, and leprosy, caused *by M. leprae*, still afflicts large numbers of people. Other species can be pathogenic, which has recently become apparent as they are major opportunists in HIV-infected people of developing countries (Parish and Stoker 1998).

The mycobacterial genome contains a high guanine plus cytosine (GC) content, ranging from 58% to 69%, which creates a biased amino-acid composition in proteins (Clark-Curtiss 1990;Cole et al. 1998). Similar to closely related family members (e.g. Actinomycete family) mycobacteria have complex cell wall components, adjuvant activity, cord factor, sulfolipids and iron-chelating compounds (Minnikin 2001;Minnikin et al. 2002). Their most characteristic feature is their complex cell envelope, containing a high percentage of lipids, which include the large-branched mycolic acids (Clark-Curtiss 1990;Minnikin 2001). The envelope makes the bacteria resistant to breakage and relatively impermeable to antibiotics, and is responsible for the acid-fast staining property used to identify the organism (i.e. ZN acid-fast staining) (Clark-Curtiss 1990;Brennan 2003).

1.3. *Mycobacterium tuberculosis*

1.3.1. Brief History

M. tuberculosis is the causal agent of TB, which has been a threat to humans since antiquity. To the ancient Greeks it was known as phthisis (to waste), and to the Western World in the 1700 to 1800s it was known as consumption (to consume). It has been suggested that *M. tuberculosis* and related bacteria (which can also be responsible for this disease) originated from domesticated cattle 10,000 years ago and transferred to humans. However, more recent findings contradict this statement (Gutacker *et al.* 2002). References to tuberculosis-like symptoms and characteristics can be found throughout early writings, such as Hippocrates. TB is also well known for claiming the

lives of many famous individuals, including a number of the Bronte family members (writers), John Keats and Hendry David Thoreau (Sequella 2003). In the 1800s, common treatments for TB included induction of lung collapse (which helped because the bacteria require a high oxygen environment to survive) and isolation of patients in institutions (sanatoria), which were located away from cities and where the air was fresh. In 1882 Robert Koch identified the bacteria that caused TB (Lindsten and Ringerte 2003), however, the era of modern medical treatment only began after 1944, when the antibiotics streptomycin and para aminosalicyclic acid (PAS) were first administered to TB patients.

1.3.2. Infection

TB claims twice as many lives as AIDs and more lives than any other infectious disease (WHO 1999). It is estimated that approximately three million people die from TB yearly, and its prevalence is increasing globally (in both developing and industrialised countries) despite the availability of effective drugs such as INH and RIF (Cole et al. 1998;WHO 1999). Even more alarming is the fact that *M. tuberculosis* strains resistant to multiple drugs have been observed in infected individuals, making treatment extremely difficult (Grange and Zumla 2002;Mukherjee et al. 2004). In addition, *M. bovis* Bacillus Calmette-Guerin (BCG) vaccine, which is widely used in developing countries to immunise infants against TB, is only partially effective (Lawson et al. 2003;Parthasarathy 2003). It is no wonder that the World Health Organisation (WHO) declared TB a global emergency in 1993 (Cole *et al.* 1998). The spread of TB among AIDs patients and the emergence of multidrug resistant strains emphasise a need for new drugs (with shortened therapy) and vaccines to prevent or cure *M. tuberculosis* infection and disease (Bloom and Murray 1992;Young and Cole 1993;Glickman et al. 2001). Unfortunately, no new anti-mycobacterial has been characterised in the past 25 years (Azad *et al.* 1997). New drugs and vaccine candidates will only come with more thorough understanding of the mechanisms of *M. tuberculosis* pathogenesis (Glickman *et al.* 2001).

Of the third of the world's population that is infected with *M. tuberculosis*, only 10% of infected individuals will manifest active infection (in the absence of immunosupression). Active TB is characterised by coughing, fever, weight loss, night sweats, and shortness of breath (Cole *et al.* 1998). *M. tuberculosis* does not remain as only a pulmonary infection, but can also infect many other organs including the brain, kidneys and heart; however active infection does usually begin in the lungs. A person with active TB will cough out *M. tuberculosis* contained in 'droplet nuclei', which can remain suspended in the air for several hours, making TB a contagious disease (Cole *et*

al. 1998). In most cases the inhaled bacteria are rapidly destroyed by the host immune system. However, it is thought that sometimes *M. tuberculosis* bacteria remain dormant in the lungs, within macrophages lodged in calcified, scar-like structures called tubercles (hence the name of the disease), which are produced as a result of the body's attempt to isolate the area of infection. Such dormant bacteria can persist in the body for years, held in check by the immune system, and may emerge as an active infection under times of stress or immune compromise (Cole *et al.* 1998). It is therefore not surprising that TB is a major problem in the care and treatment of HIV infected individuals. It has been hypothesised that the state of dormancy that occurs in infected tissues may reflect metabolic shutdown resulting from the action of a cell-mediated immune response that contains but does not eradicate the infection (Cole *et al.* 1998). The mechanisms of this dormancy are unknown, but are expected to be genetically programmed and to involve intracellular signalling pathways (Cole *et al.* 1998).

1.3.3. The Cell Envelope

The characteristic features of the tubercle bacillus are slow growth, dormancy, intracellular pathogenesis, genetic homogeneity and a complex cell envelope (Cole *et al.* 1998). *M. tuberculosis* is able to synthesise most of its essential amino acids, vitamins and enzyme co-factors, although some of the biochemical pathways differ from those found in other bacteria (Cole *et al.* 1998). It can also metabolise a variety of carbohydrates, hydrocarbons, alcohols, ketones and carboxylic acids (Ratledge 1982;Wheeler and Ratledge 1994). *M. tuberculosis* has the ability to adapt its metabolism to environmental change, enabling it to compete with the lung for oxygen and to survive in a microaerobic/aerobic environment inside a granuloma (Cole *et al.* 1998).

In 1998 the complete genome sequence of *M. tuberculosis* (H37Rv) was released and it was shown to have 4,411,529 base pairs that constituted approximately 4,000 genes (Cole *et al.* 1998). *M. tuberculosis* (like other mycobacteria) differs from other bacteria in that a very large portion of its coding capacity is devoted to the production of enzymes involved in lipogenesis and lipolysis, and two new families of glycine-rich proteins with a repetitive structure that may represent a source of antigenic variation (Cole *et al.* 1998).

The cell envelope of *M. tuberculosis* is highly specialised and confers resistance to chemical injury, low permeability to antibiotics, resistance to dehydration, and ability to thrive within the hostile environment of the phagolysome within the macrophage (Barry *et al.* 1998). It is based on unusual lipid molecules, ranging from inert waxes to biologically active glycolipids (Minnikin

2001), including mycolic acids, lipoarabinomannan, trehalose dimycolate, and phthiocerol dimycocerosate (Brennan and Nikaido 1995;Daffe and Draper 1998). Most known lipid and polyketide biosynthetic systems are found in mycobacteria, including enzymes usually found in mammals and plants as well as the common bacterial system. There are approximately 250 distinct enzymes involved in fatty acid metabolism in *M. tuberculosis* compared with only 50 in *Escherichia coli* (Cole *et al.* 1998). This highly complex array of distinctive lipids and glycolipids have been intensely scrutinised as a potential effecter in the interaction of *M. tuberculosis* and the human host (Brennan and Nikaido 1995;Daffe and Draper 1998;Barry et al. 1998). These molecules range from simple fatty acids such as palmitate and tuberculostearate, through isoprenoids, to very-long-chain, highly complex molecules such as mycolic acids and the phenolphthiocerol alcohols that esterify with mycocerosic acids to form the scaffold for attachment of the mycosides (Riley and Labedan 1996;Cole et al. 1998).

These unique lipids are important in cell structure and pathogenicity (Minnikin 2001), and therefore play an important role in the survival of the pathogen in the hostile environment within the host (Mehra et al. 1984;Vachula et al. 1989;Chan et al. 1989). A range of specific waxes (phthiocerol dimycocerosates) and glycolipid antigens are exposed on the cell surface of mycobacteria. The antigens range from simple 2,3-diacyltrehaloses to complex glycopeptidolipids and glycosylphenolphthiocerol dimycocerosates, the so called phenolic glycolipids (Minnikin 2001). Surface-exposed lipids have been found to be unique to pathogenic mycobacteria (Minnikin 1982;Thurman and Draper 1989;Daffe and Laneelle 1989) and play important roles in mycobacterial interaction with the host (Daffe and Laneelle 1988;Brennan and Nikaido 1995). Drugs that inhibit the synthesis of cell wall components can be effective antimycobacterial therapy (Young 1994).

1.3.3.1. Structure

Much of the early structural definition of the cell wall of Mycobacteria species was initiated in the 1960s and 1970s (Petit et al. 1969;Adams et al. 1969;Weitzerbin-Falzpan et al. 1970;Lederer et al. 1975). After these initial findings there was a long period of inactivity, but with the recent developments in analytical techniques combined with the definition of the *M. tuberculosis* genome, more information could be obtained, resulting in a better understanding of the mycobacterial cell wall, its lipids and its basic genetics and biosynthesis (Brennan 2003).

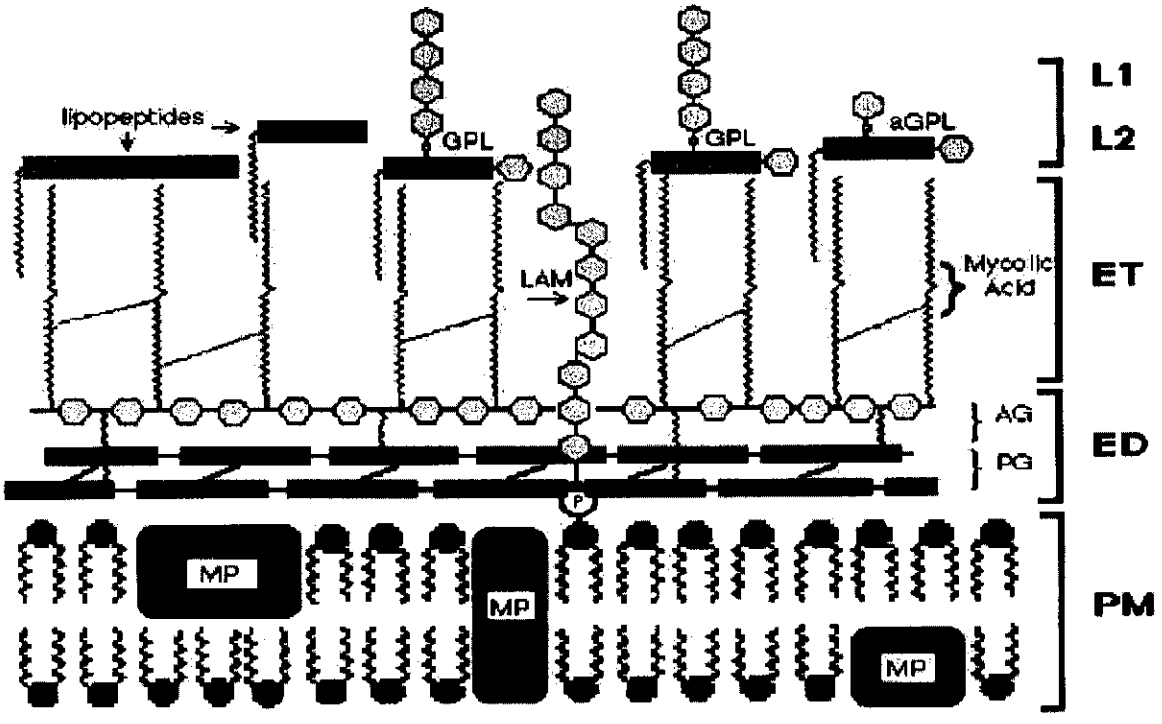


Figure 1.1: Model of mycobacterial cell envelope depicting the plasma membrane (PM), the electron dense (ED), electron transparent (ET), and L1 and L2 layers.

In this model, the plasma membrane (PM) is shown to contain membrane proteins (MP). Also represented are peptidoglycan (PG), arabinogalactan (AG), mycolic acid, lipoarabinomannan (LAM), lipopeptides, glycopeptidolipid (GPL), and apolar glycopeptidolipid (aGPL). This model is not drawn to scale.

This model is an amended version of that published in 'Frontiers in Bioscience' (Barrow 1997).

The cell envelope of mycobacteria has been divided into four major layers (see Figure 1.1). The first layer is the cytoplasmic, or plasma membrane layer (PM). This layer is a permeable lipid bilayer that interacts with proteins enclosing the cytoplasm, and is similar to those found in most bacteria. The second layer is referred to as the electron dense layer (ED) because of its staining properties when observed by transmission electron microscopy. It consists of the peptidoglycan/ or arabinogalactan moieties that form the basic structural component of the cell wall. The third layer is called the electron transparent layer (ET) due to its appearance under the transmission electron microscopy. This layer consists of complex lipids e.g. the mycolic acids. The fourth and outermost layers (L1 and L2) are composed of many different components that are mycobacteria specific. The layers generally appear fibrillar in nature when observed by freeze fracture or negative staining. They are most important with regard to initial host interaction as they contain

(or streptomycin), for the first two months of treatment, and INH and RIF, for at least a further four months (Banerjee et al. 1994;Palandez et al. 2003). Unfortunately, the long time course required for effective treatment, the unpleasant side effects of the drugs, and the foibles of human nature often lead to poor compliance, one of the major factors contributing to rising infection rates. Thus, the World Health Organisation has been pushing for universal adherence to a process known as directly observed therapy (DOT), through which health care workers counsel patients, monitor their progress, and ensure that the swallowing of each dose of medication is observed by health care workers or volunteers (Grange and Zumla 2002).

M. tuberculosis is naturally resistant to many of the available antibiotics due to its highly hydrophobic cell envelope, which acts as a permeable barrier, and its endogenous proteins and enzymes (Brennan and Draper 1994;Cole et al. 1998). These include hydrolytic or drug-modifying enzymes such as β-lactamases and aminoglycosides, acetyl transferases, and many potential drug-efflux systems, such the numerous ABC transporters (Cole *et al.* 1998). Knowledge of these putative resistance mechanisms will promote better use of existing drugs and facilitate the conception of new therapies (Cole *et al.* 1998). The unique lipids may also be used as drug targets, as exemplified by the action of the front-line INH on mycolic acid biosynthesis (Minnikin *et al.* 2002).

1.3.4.1. Isoniazid (INH)

INH was first synthesised by Meyer and Mally in 1912, but its chemotherapeutic value was not discovered until forty years later. Today it is one of the most widely used chemotherapeutic and prophylactic drugs for treatment of TB (Ellard et al. 1972;Ramaswamy et al. 2003). It was first reported to be an effective anti-TB drug in 1952, displaying particular potency against *M. tuberculosis* and *M. bovis* (Bernstein et al. 1952;Banerjee et al. 1994;Quemard et al. 1995). It can be an effective prophylactic anti-tubercular drug on its own, however, it is mostly used in combination with RIF, and PZA (Quemard *et al.* 1995). INH has a simple chemical structure consisting of a hydrazide group attached to a pyridine ring (see Figure 1.2), but its mode of action is very complex (Ramaswamy *et al.* 2003). It has been proposed that INH enters *M. tuberculosis* as a prodrug by passive diffusion and is subsequently activated by catalase-peroxidase, encoded by *katG* (see Figure 1.2), to generate free radicals, which then attack multiple targets in the cells (Bardou et al. 1998;Ramaswamy et al. 2003). Two intracellular enzymatic targets for activated INH have been proposed, both of which are enzymes involved in the biosynthesis of mycolic acids. They are a NADH-dependent enoyl acyl carrier protein (ACP) synthase, encoded by *inhA*,

and a β-keto-acyl ACP synthase, encoded by *kasA* (Banerjee et al. 1994;Mdluli et al. 1998;Ramaswamy et al. 2003).

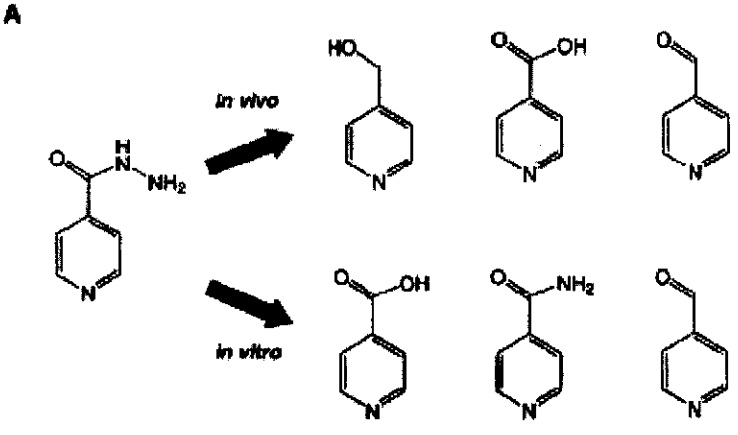


Figure 1:2: Metabolism of INH by whole cells of *M. tuberculosis* gives defined products resulting from KatG-mediated oxidation (in vivo) while purified KatG (with manganese ions) gives a different product profile (in vitro).
This figure is an amended version of that published in 'Microbes and Infection' (Slayden and Barry 2000).

INH resistant strains were isolated almost immediately after this antibiotic was introduced for therapeutic use (Quemard *et al.* 1995). Resistance-associated amino acid substitutions have been identified in the *katG*, *inhA* and *kasA* genes of INH resistant clinical isolates of *M. tuberculosis* (Banerjee et al. 1994;Musser et al. 1996;Mdluli et al. 1998;Ramaswamy et al. 2003). Mutations in the *oxyR-aphC* intergenic region have also been identified in INH resistant isolates (Sreevatsan et al. 1997b;Ramaswamy et al. 2003). Further studies showed that certain promoter mutations of alkylhydroperoxide reductase, encoded by *aphC*, in INH resistant isolates result in overexpression of *aphC* as a compensatory mechanism for the loss of catalase activity due to *katG* mutations (Sherman et al. 1996;Kelly et al. 1997;Ramaswamy et al. 2003). More recently missense mutations have been identified in the *ndh* gene of INH resistant isolates. This gene encodes NADH dehydrogenase, which is a essential respiratory chain enzyme that regulates the $NADH/NAD^+$ ratio in cells (Lee et al. 2001;Ramaswamy et al. 2003). The molecular mechanism by which mutations in *ndh* confer INH resistance in *M. tuberculosis* is poorly understood (Ramaswamy *et al.* 2003). In addition, low-level INH resistance in mycobacteria has been shown to be associated with enhanced expression of NAT, which is believed to inactivate the prodrug INH by acetylating the molecule (Payton *et al.* 1999). However, no INH resistance-associated

mutations have been described in the NAT gene (Payton et al. 1999;Upton et al. 2001b;Ramaswamy et al. 2003).

1.4. Arylamine N-acetyltransferase (NAT)

(This section has been published; Brooke, E., Sholto-Douglas-Vernon, C. and Sim, E. (2003) The role of arylamine N-acetyltransferase in bacteria. In Recent Research developments in bacteriology Transworld Research Network, Kerala, India; and is reproduced here)

NATs are a family of Phase II drug-metabolizing enzymes that have been found in a wide range of prokaryotes and eukaryotes (Upton *et al.* 2001a). The enzyme was first characterised in humans as the factor responsible for the polymorphic inactivation of the anti-tubercular drug INH (Evans *et al.* 1960). The two human isoenzymes are 30-34kDa cytosolic proteins found in a range of human tissues (Grant et al. 1989;Hickman et al. 1998). The acetyl transfer reaction involves the use of acetyl Coenzyme A (AcCoA) as a cofactor and proceeds by a 'ping-pong' bi-bi mechanism (see Figure 1.3) (Riddle and Jencks 1971). NATs can generally acetylate a wide range of arylamines, arylhydroxylamines and aryl hydrazines including environmental carcinogens (Hein *et al.* 1993) and drugs (Weber and Hein 1985). Subsequent work on NATs has focussed on a wide variety of fields: carcinogenesis (Hein *et al.* 1993), developmental studies (Smelt *et al.* 2000) and pharmacogenetics (Weber and Hein 1985) in particular. NAT activity was first described in bacteria in one of the Ames mutagenicity tester strains that had a higher susceptibility to food-based carcinogens (Saito *et al.* 1983). Subsequent advances in recombinant protein technology and genomics has allowed the in depth analysis of NAT structure and function and the discovery of many more bacterial NATs. Current investigations in this area are focussed around the role of NATs in different bacterial families and the effect that bacterial NATs may have on human diseases such as TB and inflammatory bowel disease.

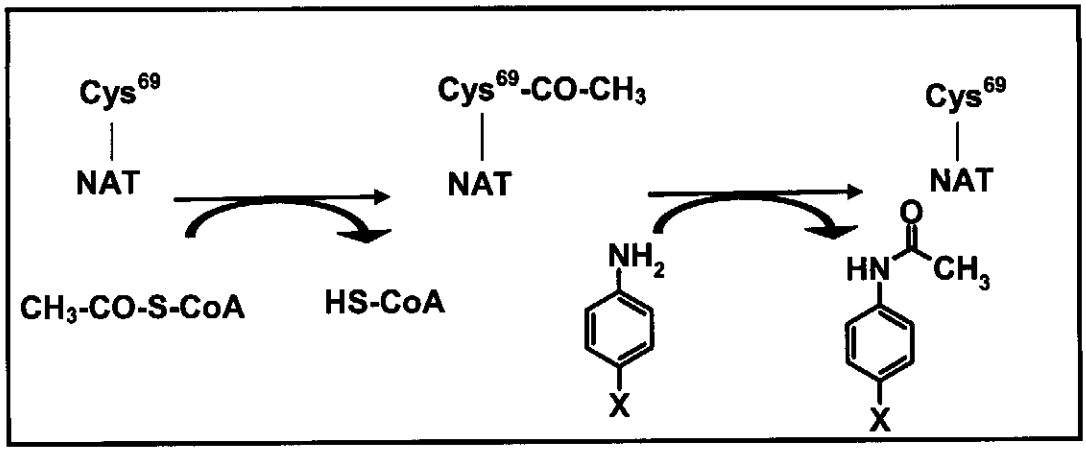


Figure 1.3: Catalytic mechanism of NAT

The acetyl group is transferred by a ping-pong mechanism from the acetyl CoA to the arylamine via a cysteine residue. This figure was published in 'Recent research developments in bacteriology' (Brooke *et al.* 2003b), and is reproduced here.

1.4.1. NAT in *Salmonella typhimurium*

Many environmental arylamines are mutagenic in humans and their mutagenic effects have been related to NAT acetylation status and polymorphism (Fretland *et al.* 2002). The relationship between NAT activity and mutagen susceptibility is complex and depends on the tumour type and location in the body (Agundez et al. 1995;Varzim et al. 2002). It is generally accepted that *N*-acetylation of arylamines leads to the deactivation of the mutagenic properties and the excretion of the compound. However, if the arylamine is first oxidised by cytochrome P450 (e.g. CYP1A2, (Brockmoller *et al.* 1996)) the *N*-hydroxy product formed can be *O*-acetylated and leads to the formation of highly reactive arylnitrenes that form DNA adducts.

The Ames tester strains of *S. typhimurium* have been used for many years as a diagnostic test of the mutagenicity of chemical and biological samples (Mortlemans and Zeiger 2000). The strains possess a mutation in the gene controlling histidine production that greatly inhibits growth without a histidine source. Incubation of the bacteria with a mutagenic compound can lead to a reverse mutation in the mutant *his* gene, allowing the bacteria to grow again without added histidine. The frequency of growing bacteria indicates the mutagenicity of the compound (Ames *et al.* 1973). One particular strain, TA98/1,8-DNP, was resistant to the toxic and mutagenic effects of arylamines (Saito *et al.* 1983). Another strain, YG1024, was more susceptible to these effects (Watanabe *et al.* 1993). Subsequent studies showed that strain TA98/1,8-DNP was deficient in

NAT activity, and YG1024 overexpressed the *nat* gene. Hence, the increased rate of *O*-acetylation by NAT leads to an increased toxicity of these compounds. Recombinant NAT from *S. typhimurium* (STNAT) was produced and showed 25-33% similarity over the 170 amino acids at the N-terminus to the eukaryotic enzymes that had been previously sequenced (Watanabe *et al.* 1992). A key cysteine residue was found (Cys^{69}) that was conserved in all NAT sequences and a mechanism was suggested involving an acetylated cysteine intermediate. This correlated with the inhibition of NAT activity by iodoacetamide and *N*-ethylmaleimide. The C-terminus shows much weaker homology to the eukaryotic proteins. A deletion mutant of the eleven C-terminal amino acids of STNAT possessed poor acetyl transfer activity but could hydrolyse AcCoA in the absence of substrate. Without the entire third domain, the enzyme possessed no acetyl transfer activity, as observed with a truncated human NAT isozyme (Sinclair and Sim 1997), but again the truncated STNAT hydrolysed AcCoA, suggesting a role for the C-terminus in substrate binding and protection of the acetylated enzyme intermediate from hydrolysis (Mushtaq *et al.* 2002).

1.4.2. Distribution of bacterial NATs

The rapid production of bacterial genomes has been an invaluable resource to many areas of bacteriology (Fitzgerald and Musser 2001;Raczniak et al. 2001;Schoolnik 2002). In the NAT field, the availability of bacterial genomes has shown the diversity of NATs in bacteria and pinpointed key conserved residues found in all NAT sequences. NATs are not found in all bacterial phyla, but have been discovered extensively in three of the eubacterial phyla: proteobacteria, gram-positive bacteria (firmicutes) and actinobacteria. Genome searches have shown *nat*-like genes in over twenty eubacteria (Payton et al. 2001b;Rodrigues-Lima and Dupret 2002). Most of these organisms have exhibited NAT activity and show 30-80% amino acid similarity over the first two domains (see Figure 1.4). Certain areas of homology are observed in all members of the NAT family: the GGxC motif at the active site cysteine (position 66-70), the catalytic histidine (108-111) and aspartate DxG (120-127) motifs, and the FENL motif around position 40 in all NATs with acetylating activity. Similarly, in an activity based study, NAT activity was detected in fifteen strains of bacteria from the gut flora (Delomenie *et al.* 2001). Many of the bacteria that possess NAT activity live in the soil or the gut where high concentrations of potentially toxic organic material are present, reinforcing the suggestion of NAT playing a general role as a protective enzyme for the breakdown of exogenous

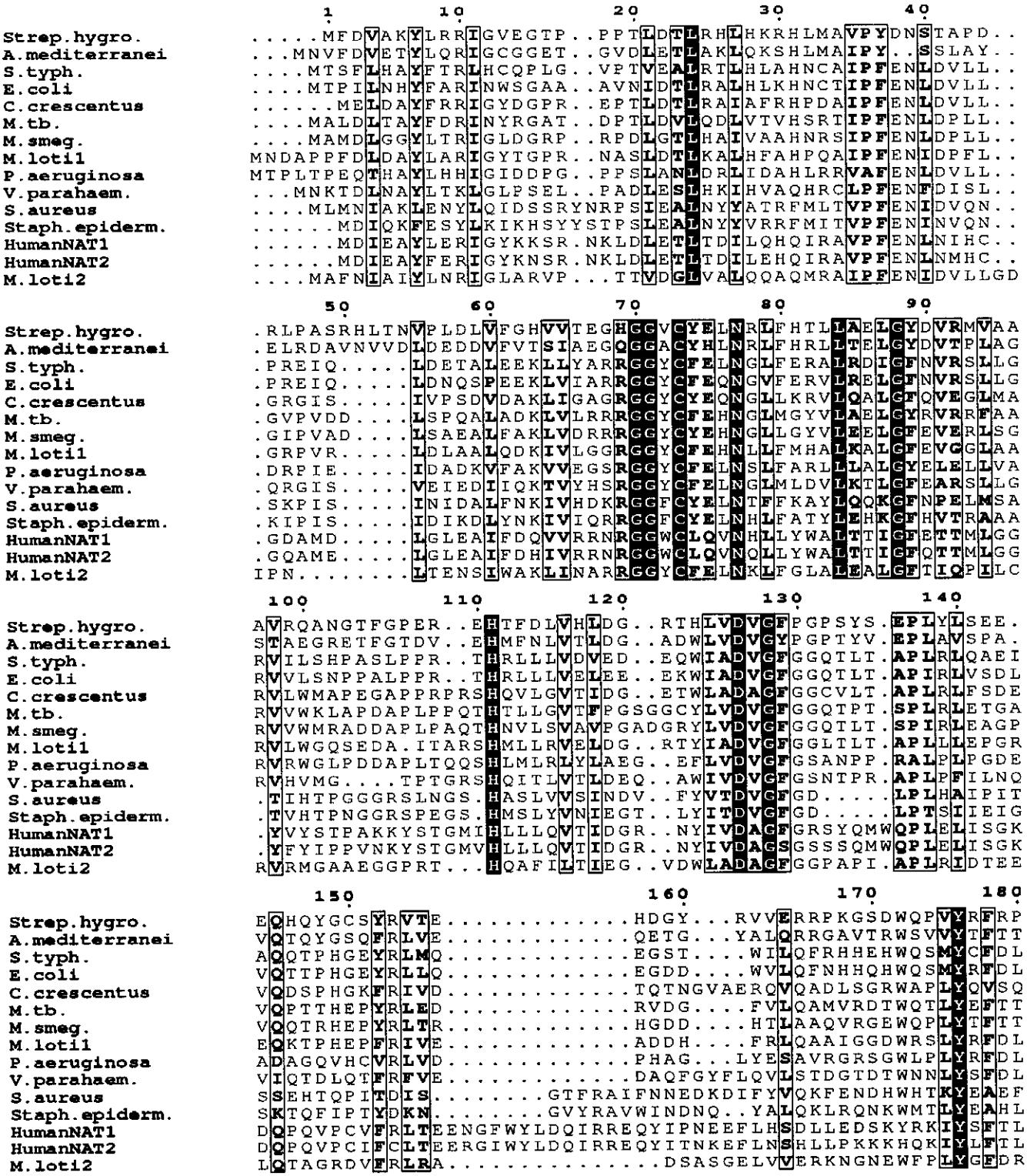


Figure 1.4: Sequence alignment of a selection of bacterial and human NATs

Sequence alignment was performed on ClustalW on http://www.ebi.ac.uk and picture generated with ESPript on http://www.expasy.ch. The first two domains are shown. Abbreviations are: Strep.hygro. – *Streptomyces hygroscopicus*; S.typh. – *Salmonella typhimurium*; M.tb. – *M. tuberculosis*; M.smeg. – *M. smegmatis*; M.loti1 & M.loti2 – *Mesorhizobium loti*; V.parahaem. – *Vibrio parahaemolyticus*; S.aureus – *Staphylococcus aureus*; Staph.epiderm. – *Staphylococcus epidermidis*

This figure is an amended version of that published in 'Recent research developments in bacteriology' (Brooke *et al.* 2003b).

arylamine toxins (Delomenie et al. 2001;Payton et al. 2001b;Brooke et al. 2003a). This is also reinforced by the presence of several genes in the region of the *nat* gene that show homology to enzymes used for the degradation of aromatic compounds such as biphenyl-2,3-diol-1,2-dioxygenase (Payton *et al.* 2001b). In *M. smegmatis*, the *nat* gene is transcribed as part of a larger message suggesting that NAT forms part of an operon in this organism (Payton *et al.* 2001a). It will be very interesting to discover whether NAT occurs as part of an operon in other organisms to suggest an endogenous role for this enzyme.

NATs have been found extensively throughout the eubacterial and the vertebrate kingdoms (Rodrigues-Lima and Dupret 2002). However, no NAT sequences have so far been discovered in any members of the Archaea (Payton *et al.* 2001b) or in non-vertebrates, apart from the urochordate *Ciona intestinalis* (Rodrigues-Lima and Dupret 2002) which possesses considerable similarities to vertebrates in development. In other lower organisms and plants, NATs may not be required due to other arylamine metabolising pathways. Alternatively the organisms may synthesise arylamines themselves making NAT activity detrimental (Deguchi 1992). Studies of the distribution of NATs throughout eukaryotes and prokaryotes have suggested that NATs have spread from a common ancestor (Rodrigues-Lima and Dupret 2002). NAT had been placed as part of a set of 41 genes thought to have passed laterally from bacteria to vertebrates (horizontal gene transfer) (Salzberg *et al.* 2001) but the existence of this phenomenon is not convincingly supported (Stanhope *et al.* 2001).

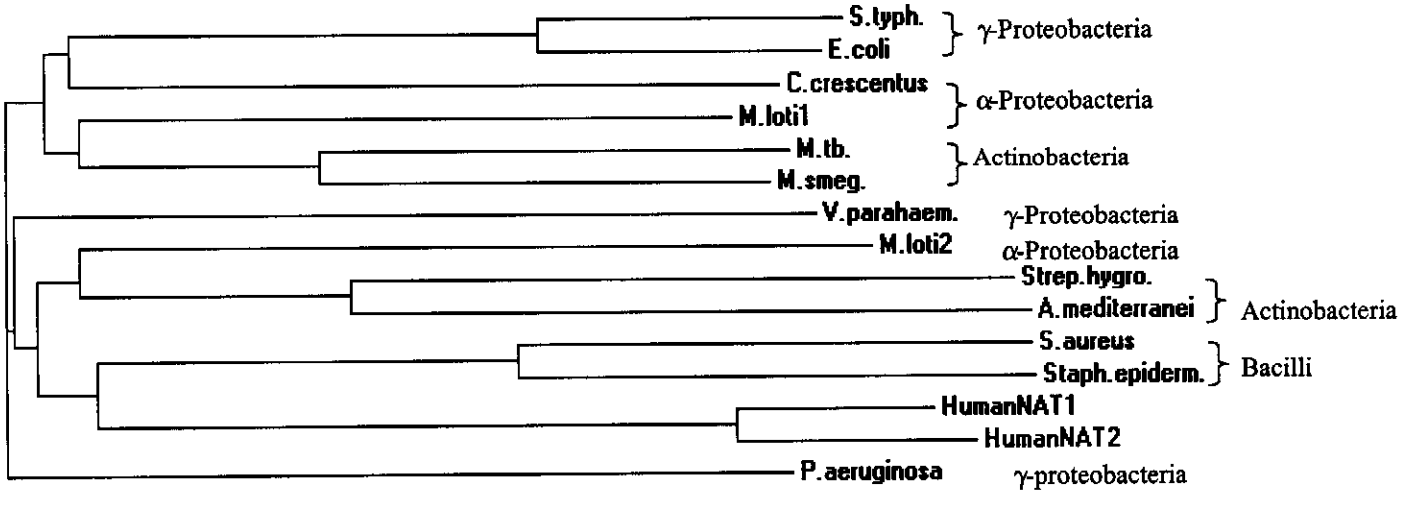


Figure 1.5: Phylogenetic tree of a selection of bacterial and human NATs

Phylogram was created with ClustalW (http://www.ebi.ac.uk). Distances indicate an estimate of the evolutionary time since the most recent common ancestor. Abbreviations are as per Figure 2.

This figure is an amended version of that published in 'Recent research developments in bacteriology' (Brooke et al. 2003b).

Interestingly, the two NAT paralogs in *Mesorhizobium loti* are the first example of NAT paralogs in bacteria (Rodrigues-Lima and Dupret 2002). Although the sequence identity between the paralogs is very low (31%, Figure 1.5), it has been suggested that the paralogs are a result of ancestral gene duplication rather than horizontal gene transfer. Bacteria within the same bacterial class (e.g. γ-Proteobacteria) can have both similar (e.g. *E. coli* and *S. typhimurium*) and diverse (e.g. *P. aeruginosa*) NAT sequences, but differing rates of evolution can account for the differences observed. Those bacteria that do not possess a NAT sequence have probably undergone gene loss, as observed in the *M. leprae* genome compared to the *M. tuberculosis* genome (Cole *et al.* 2001).

1.4.3. Rifamycin Amide Synthase

Several NAT homologues certainly do appear in operons and have very specific endogenous functions. Genome searches showed that there was a NAT homologue in *Amycolaytopsis mediterranei* (Payton *et al.* 2001b), an actinomycete that produces the antibiotic Rifamycin B. The homologue was shown to be the final enzyme in the rifamycin synthetic pathway, RifF. Rifamycin B is formed from an aromatic precursor that is then extended by the polyketide synthase genes *rifA-rifE* (Yu *et al.* 1999). This produces an aromatic amine base with a long polyketide chain, attached to the polyketide synthase by a thioester bond. RifF catalyses the amide bond formation via an intra-molecular acyl transfer reaction. The protein is 40% similar to STNAT and still contains the three key catalytic residues but does not possess the FENL region around position 40. This may open out the active site cleft and allow the larger substrate to access the cysteine residue. This protein has been expressed and purified and does not show any acetyl transfer activity (Pompeo *et al.* 2002b). The catalytic site has been modelled and shows a more open structure, as expected to accommodate the much larger substrate. The NAT-like sequences in *Actinosynnema pretiosum* and *Streptomyces achromogenes* play similar roles in the synthesis of ansamitocin and rubranosol (Payton *et al.* 2001b). The genes from *S. hygroscopicus*, *ShnN* and *GdmF* (strain NRRL 3602) (Rascher *et al.* 2003), are also NAT-like genes in the gene-clusters for ansamycin and geldamycin antibiotics. These are the first examples of NAT-like proteins with an intra-molecular acyl transfer activity and comparisons between these enzymes and the other members of the NAT family could give key insights to the function and mechanism of NATs.

1.4.4. Substrates of bacterial NATs

In humans, there are two NAT genes, NAT1 and NAT2 (Grant *et al.* 1989). These two enzymes exhibit differing yet overlapping substrate specificities (Minchin et al. 1992;Minchin 1995) and tissue distributions (Jenne 1965;Hickman et al. 1998;Smelt et al. 2000). This evidence has suggested differing roles for the two isoenzymes: NAT1 can acetylate p-aminobenzoic acid (pABA) and p-aminobenzoylglutamate (pABA-Glu), a folate catabolite, suggesting a role for NAT1 in the folate cycle. NAT1 may hence play a key role early in development and has been linked with neural tube defects (Mills and Conley 1996;Sim et al. 2000). NAT2 acetylates a broad range of substrates and its tissue distribution suggests that it plays a role in xenobiotic metabolism (Pompeo *et al.* 2002a). Similar analysis in bacteria may help to suggest endogenous roles for NATs in bacterial species.

The one substrate that is rarely acetylated by bacterial NATs is pABA (Delomenie et al. 2001;Brooke et al. 2003a). In bacteria, pABA plays an essential role as a folate precursor (Herrington 1994) so acetylation of pABA by bacterial NATs would be highly destructive to the organism. However, *Pseudomonas aeruginosa* has shown weak pABA acetylating activity in cell lysate (Delomenie *et al.* 2001) and when expressed as a recombinant protein (unpublished data). This is a very interesting discovery and raises several questions as to the folate production cycle in these bacteria. *P. aeruginosa* lives and grows in many toxic environments including growth on organic pollutants (Prescott *et al.* 1999) and, in line with NAT playing a protective role against aromatic toxins, exhibited the highest NAT activities of the series of bacterial cell lysates tested (Delomenie *et al.* 2001). Similarly, a partial NAT sequence has been found in *Sphingomonas aromaticivorans* (Rodrigues-Lima and Dupret 2002), a bacteria noted for its aromatic compound metabolizing capabilities (Shi *et al.* 2001). Further investigation of the enzymes in these bacteria could prove useful in defining and exploiting NAT activity.

Studies with both recombinant NATs and cell lysates have shown the substrate specificities of several bacterial NATs. Although bacterial NATs were originally thought to be homologues of human NAT2, overall bacterial NATs exhibit different substrate specificities to either human isoenzymes. In tests with recombinant NAT from *M. smegmatis* and STNAT both the human NAT1 substrate pABA and the human NAT2 substrate sulfamethazine were acetylated poorly (Brooke *et al.* 2003a). Good substrates of these enzymes include simple lipophilic arylamines such as 2-aminofluorene, 4-hexyloxyaniline and the aromatic hydrazine hydralazine. In several studies, 5-aminosalicylate has been shown to be a good substrate for bacterial NATs (Delomenie et

al. 2001;Brooke et al. 2003a). 5-aminosalicylate is a highly effective treatment for inflammatory bowel disease so the acetylation of this compound by gut flora may have important ramifications in the treatment of this disorder.

1.4.5. Structure of NAT

The first three-dimensional structure of NAT was determined by X-ray diffraction of recombinant STNAT crystals (Sinclair *et al.* 2000). The structure, at 2.8Å resolution, was sufficiently detailed to elucidate the catalytic residues involved in the acetyl transfer reaction. A catalytic triad of Cys^{69}-His^{107}-Asp^{122} was observed that provided the general base catalysis required for the reaction to occur. Previously, a highly conserved arginine residue, Arg^{65}, was thought to play a catalytic role (Watanabe *et al.* 1992). However, correlation with known human NAT polymorphisms suggest that the conserved arginine may well be key to the stability of the NAT structure (Rodrigues-Lima *et al.* 2001). The unique fold showed that the protein fell into three clear domains: an α-helical bundle of 84 residues, a β-sheet barrel of 95 residues and an α/β lid of 98 residues. The catalytic residues are found in the first two domains, and these residues are conserved throughout the NAT family. The triad provides a charge transfer system, with the aspartate pulling a proton from the histidine and the histidine either partially or fully deprotonating the cysteine residue thiol. The thiolate is thus sufficiently active to attack the AcCoA thioester and allow the acetyl transfer to occur.

The structure of STNAT showed a strong similarity to the structures of several members of the cysteine protease superfamily, including the papain family. Cysteine proteases have been the target of several drugs to treat diseases as diverse as malaria (Rosenthal *et al.* 2002) and Alzheimer's disease (Hook *et al.* 2002). NAT and the cysteine proteases are very likely to have evolved from a common precursor and it is thus possible that NAT may have a role in the cell other than the acetyltransferase action, at least in some organisms.

A further 1.7Å structure of NAT from *M. smegmatis* (MSNAT) showed a near identical fold and suggested an oxy-anion hole that would stabilise the transition state of acetyl transfer (Sandy *et al.* 2002). Alignment of eukaryotic and prokaryotic NATs shows an extra 16 amino acid section towards the end of the second domain (see Figure 1.4). This is predicted to form a random loop on the exterior of the protein, and this may explain the difficulties observed in crystallising eukaryotic NATs. A partial model of the first two domains of human NAT2 predicted that the eukaryotic NATs would maintain the same fold (Rodrigues-Lima *et al.* 2001).

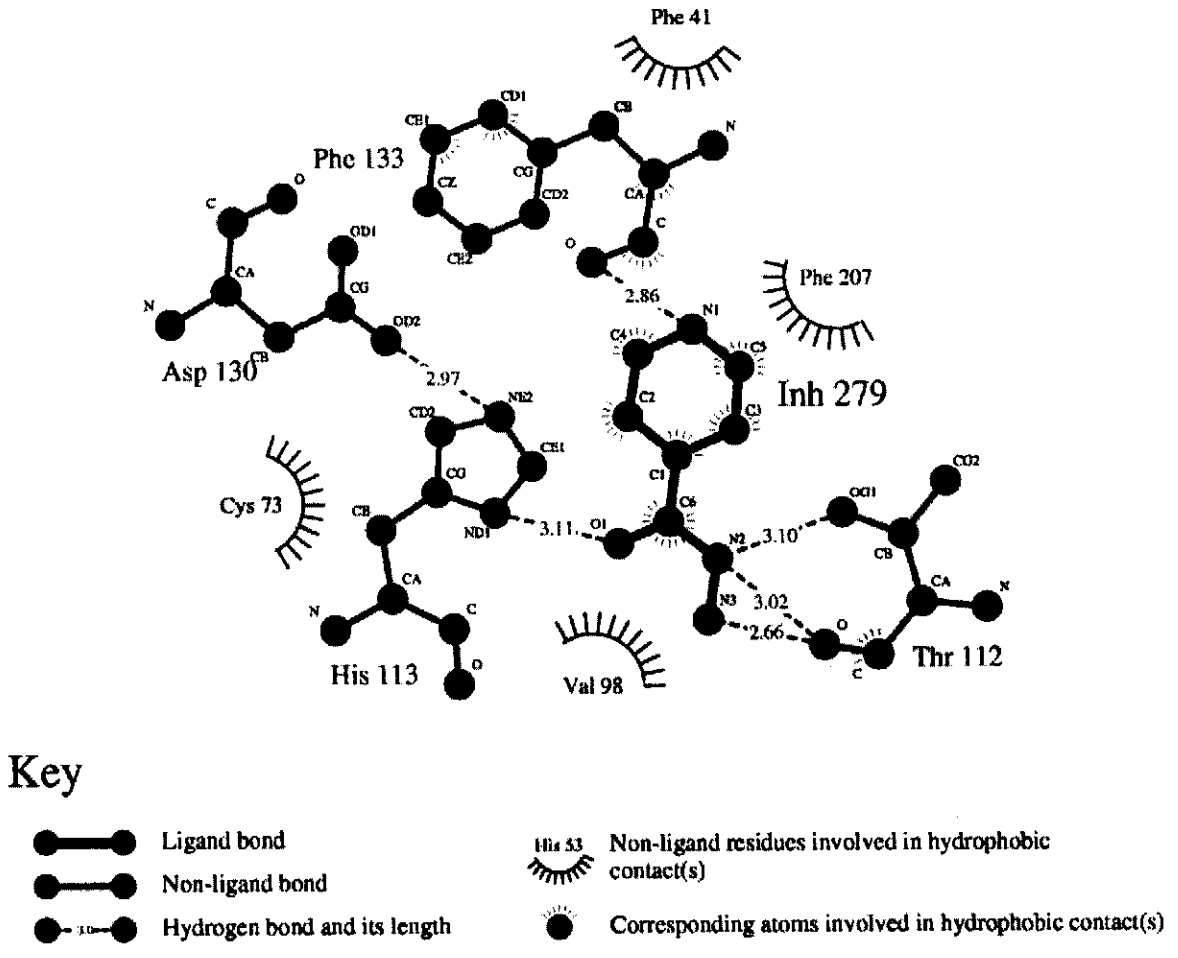


Figure 1.6: Binding of INH to the MSNAT active site

INH was docked into the crystal structure of MSNAT *in silico* using the program AutoDock (Goodsell *et al.* 1996). Figure was created using LigPlot (Wallace *et al.* 1995) to show the ligand protein interactions. Interactions are observed with the catalytic triad (Cys^{73}, His^{113}, Asp^{130}), the three phenylalanine residues and a threonine residue Thr^{112}.

This figure is an amended version of that published in 'Recent research developments in bacteriology' (Brooke *et al.* 2003b).

The binding of substrates to the NAT active site has been investigated *in silico* (Mushtaq et al. 2002;Brooke et al. 2003a). The active site of MSNAT consists of a cleft approximately 11Å long with the active site Cys residue at one end. The area immediately surrounding the cysteine is surrounded by three highly conserved phenylalanine residues and a tryptophan. This area is termed the α-site and will be the area where the arylamine must bind when it removes the acetyl group from the acetylated enzyme intermediate (see Figure 1.6). The large, flat, hydrophobic

residues can form lipophilic and π-stacking interactions with the aryl group of the substrate. Very few specific residues have been pinpointed that account for the varying substrate specificities of NATs and crystal structures containing bound substrate or Coenzyme A derivatives will provide invaluable information as to the catalytic mechanism of this family of enzymes.

1.4.6. NAT in *M. tuberculosis*

One particularly fascinating area of research in the NAT field is the role of NAT in *M. tuberculosis*, the bacterium responsible for TB. The current resurgence of TB, particularly with strains resistant to the front-line drugs, has re-sparked interest in this devastating disease (Raviglione et al. 1995;Ramaswamy and Musser 1998;WHO 1999). The molecular mechanism of action of even the oldest anti-TB drug INH is still only poorly understood (Banerjee et al. 1994;Mdluli et al. 1998;Basso et al. 1998;Marrakchi et al. 2000). The discovery of NAT with INH-acetylating activity in *M. tuberculosis* suggests that NAT may play a role in INH resistance in TB (Payton *et al.* 1999).

INH is a front line pro-drug used in TB treatment. INH is oxidised inside the cell by a catalase peroxidase (KatG) (Zhang and Young 1993;Zhang et al. 1996). This oxidation step is vital for the activation of INH and hence the effectiveness of INH as a therapeutic drug in TB treatment (Zhang and Young 1993). In its activated form INH disrupts the pathway of mycolic acid biosynthesis critical to the maintenance of the mycobacterial cell wall (Takayama *et al.* 1972). *N*-acetyl-INH is therapeutically inactive both *in vitro* and *in vivo* (Bernstein *et al.* 1952). It has been shown that heterologous expression of *M. tubercolosis* NAT in *M. smegmatis* results in a three-fold increase in resistance to INH (Payton *et al.* 1999) and a *M. smegmatis nat* knockout mutant gave a slight increase in INH sensitivity (Payton *et al.* 2001a). It can thus be suggested that *M. tuberculosis* NAT could also be involved in INH resistance (Upton *et al.* 2001b). A similar set of experiments with an NADH-specific enoyl-ACP reductase resulted in a 20-fold increase in INH resistance. This enzyme is involved in the fatty acid synthase II (FAS II) system of mycobacteria (Larsen *et al.* 2002). Mutations in this gene have been associated with INH resistance (Kelly *et al.* 1997).

A G619A single nucleotide polymorphism (SNP) of the *nat* open reading frame (ORF) was identified in *M. tuberculosis* clinical isolates resulting in G to R change at amino acid 207 (Upton *et al.* 2001b). This substitution results in a lower structural stability or solubility and a 10 times lower affinity of the enzyme for INH. The modelled structures of *M. tuberculosis* 207G NAT and 207R NAT suggest that this mutation causes the substrate-binding Phe^{204} residue to be moved, and

hence alters the INH acetylation activity (Kawamura *et al.* 2003). SNPs are rare in the slow-growing mycobacteria so often will have arisen through evolutionary pressure such as INH therapy (Sreevatsan *et al.* 1997a).

NAT may not only play a role in INH metabolism in mycobacteria but also possess an endogenous function separate to the acetylation of exogenous arylamines. The growth of the *M. smegmatis nat* knockout strain is considerably delayed when compared with the wild-type, due to an extended lag phase. This suggests that the presence of NAT is required during early growth (Payton *et al.* 2001a). Further investigation of mycobacterial *nat* knockout strains may elucidate what function this enzyme plays within the cell itself.

AIMS OF THE STUDY

The main aim of this study was to further characterise the *nat* gene in *M. tuberculosis*. To achieve this overall aim it was broken down into more specific aims, which are described below:

1. To determine the expression of *nat* over the growth cycle of *M. tuberculosis* and *M. bovis*.
2. To analyse the previously identified G619A SNP further, and to search the *nat* gene and its flanking regions for further mutations/polymorphisms.
3. To analyse the effects INH may have on the expression of the *nat* gene in growing cultures of *M. tuberculosis* and *M. bovis* BCG.
4. To locate the NAT protein in different protein fractions.

In addition to the steps described above, work was done to develop a suicide delivery vector construct containing the *nat* gene, with the T529C and G610A SNPs (described in this thesis), and its flanking regions. Two vectors were used, the p2NIL and pGoal19 vectors. These vectors would be used in future to transform the *M. bovis* BCG strain. Changes in phenotypic characteristics in the transformed isolate would be monitored and compared to that of the wild-type strain.

The characterisation of *M. tuberculosis nat* may enable other researchers to make informed decisions on whether this gene could be used as a candidate target for new TB drugs. It may also answer questions concerning whether this gene plays a role in INH resistance. However, the *nat* gene is thought to be part of a 5 gene operon (*nat*; Rv3567c; Rv3568c; Rv3569c and Rv3570c) and to answer this question fully, other genes in this operon would also need to be characterised. Monitoring the dispersion of the SNPs in different strain families may identify the SNPs as a rapid test for the characterisation of specific *M. tuberculosis* strains, i.e. if the SNP is specific to certain strain families.

CHAPTER 2

EXPRESSION OF NAT DURING THE GROWTH CYCLE OF *M. tuberculosis* AND *M. bovis* BCG

The aim of this chapter was to monitor the expression of the *nat* gene in *M. tuberculosis* and *M. bovis* BCG using Quantitative-PCR (Q-PCR). This technique quantified the relative mRNA levels of the *nat* gene at different stages of the growth cycle of the bacterium.

2.1. Introduction

NAT in *M. tuberculosis* may play a role in INH resistance (Upton *et al.* 2001b), and should be further investigated to understand this gene and its role within the bacterium. As a model for *M. tuberculosis*, the non-pathogenic *M. smegmatis* has been much studied (Parish and Stoker 1998). The *nat* gene was inactivated (i.e. the gene has been knocked-out) in *M. smegmatis* and it has been found that the knock-out isolate had an extended lag phase during the growth cycle when compared to the wild-type strain. The characteristic wild-type growth cycle was regained in the knock-out isolate when it was complemented with an expression vector containing the wild-type *nat* gene (Payton *et al.* 2001a). This suggests that the presence of NAT may be required during early growth of the bacterium (Payton *et al.* 2001a).

To characterise the *M. tuberculosis nat* gene further, the expression of *nat* was monitored over the growth cycle of the bacterium. There are a number of ways to analyse the expression of genes. These include Western blots (Payton *et al.* 2001b), RNA blotting techniques (Hansen *et al.* 2001), microarrays (Eisen and Brown 1999), RNase protection assay (McCaughan-Carucci 2003) and Q-PCR.

The Western blot technique can be used to measure the expression of a gene, by measuring the protein levels expressed by the gene of interest. An antibody specific for the protein of the studied gene is required, and protein levels are determined by the intensity of bands hybridised to the specific antibody (Payton *et al.* 2001b). RNA blotting techniques include Northern blotting, slot or dot blotting, and are widely recognised techniques used to quantitate or identify gene expression (Hansen *et al.* 2001). These techniques measure the mRNA levels by the intensity of bands

hybridised to specific probes (Hansen *et al.* 2001). Microarrays are based on similar principles to that of the RNA blotting techniques, except many more genes can be analysed at one time (Smith *et al.* 2002). The negative aspect of all these techniques is that a large amount of starting RNA is required (Smith *et al.* 2002). RNase protection is a newly developed assay to monitor gene expression (McCaughan-Carucci 2003). A labelled probe (antisense to mRNA sequence) is added to a sample of RNA and then RNase is also added. All single stranded RNA will be degraded but not the RNA annealed to the probe. The sample is loaded and run on a polyacrylamide gel, and the mRNA levels are measured as the intensity of the band hybridised to the probe (McCaughan-Carucci 2003).

In this study, Q-PCR, using the ABI PRISM 7700 Sequence Detection System, was used to quantify the expression of *nat* (normalised to 16S rRNA) over the growth cycle of the reference *M. tuberculosis* strain H37Rv (Steenken 1934; Allen 1969) and the *M. bovis* BCG strain ATCC 35734 (Behr 2002). The concept and methodology of this technique is discussed below.

2.1.1. Quantitative PCR (Q-PCR)

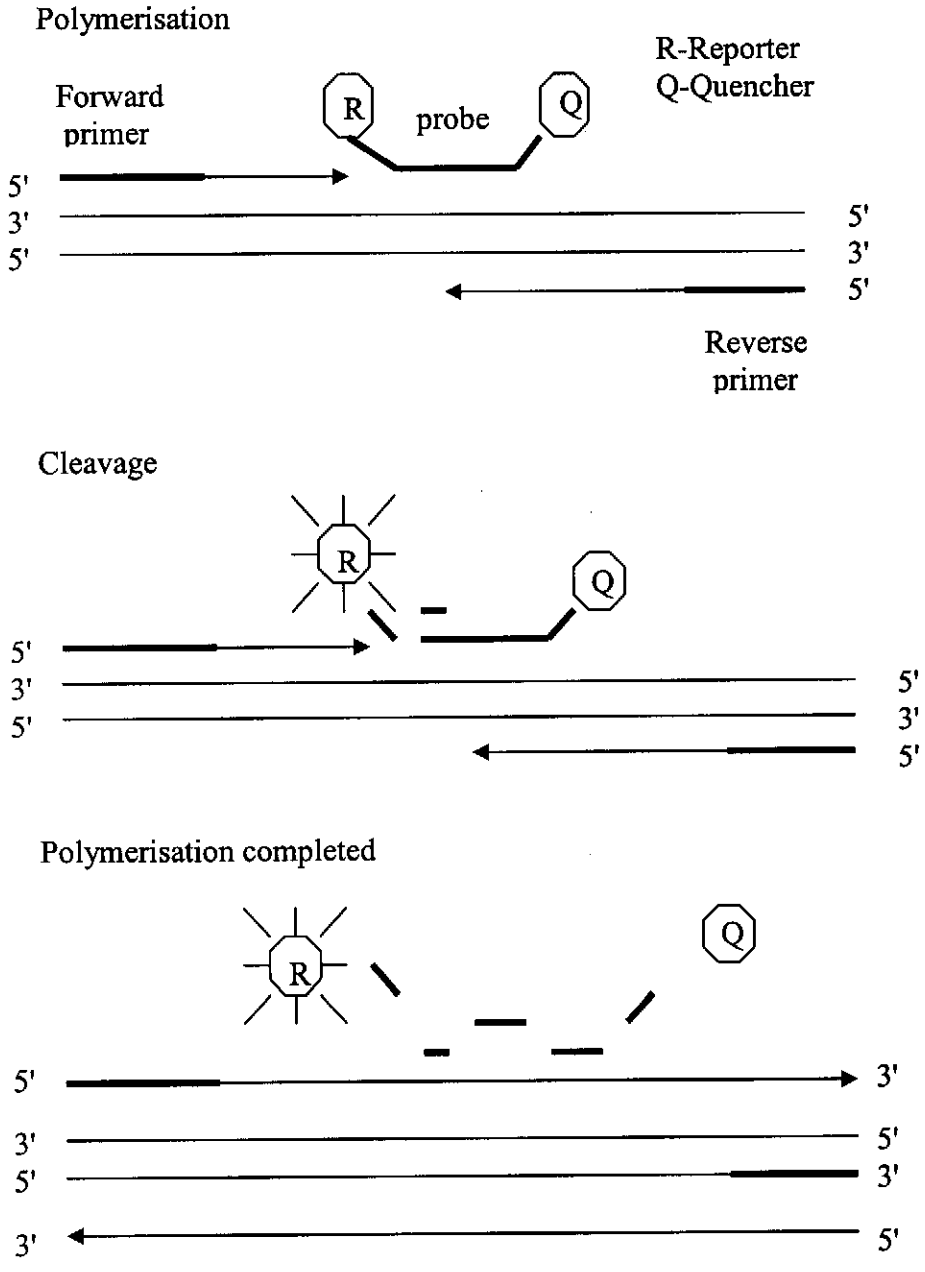
Q-PCR system (also known as Real-Time PCR) is capable of detecting PCR products as they accumulate during PCR and so enable the accurate and reproducible quantitation of DNA and RNA over a wide dynamic range (Applied Biosystems, Services Literature). This technique has been used successfully for studying gene expression, for example, Q-PCR was successfully used to monitor cytochrome P450 gene induction in human hepatocytes (Bowen *et al.* 1999).

This concept of analysis of PCR kinetics was first demonstrated in 1992 and 1993, whereby PCR products could be detected as they accumulated (Higuchi *et al.* 1992; Higuchi *et al.* 1993). This 'real-time' system monitored fluorescence of amplified products, which included the intercalator ethidium bromide, with each amplification reaction. During active PCR, the sample is subjected to ultraviolet light and a computer-controlled cooled CCD camera detects the fluorescence. The increasing amounts of double-stranded DNA produced during amplification bind to the ethidium bromide, resulting in an increase in fluorescence. The increase in fluorescence versus the cycle number is measured, producing amplification plots that provide a more complete picture of the PCR process than assaying product accumulation after a fixed number of cycles ((Higuchi *et al.* 1992; Higuchi *et al.* 1993) and Applied Biosystems, Services Literature).

This form of Q-PCR was later improved by probe based PCR product detection. The probe-based detection method minimised errors caused by the detection of non-specific products that generated a signal when using the intercalator-based method. This alternative method, the 5' nuclease assay (Gelfand *et al.* 1993; Holland *et al.* 1991), enables the detection of specific products of amplification (Applied Biosystems, Services Literature). It had previously been demonstrated that cleavage of a target probe during PCR, by the 5' nuclease activity of *Taq* DNA polymerase, could be used to detect amplification of target-specific product (Holland *et al.* 1991). The probes were 5' end-labelled with ^{32}P and 3' blocked so that the probe cannot prime. The probe annealed to its target sequence and was cleaved by the 5' nuclease activity of *Taq* DNA polymerase when the enzyme extends from an upstream primer into the region of the probe. This ensured that the cleavage of the probe only occurred if the target sequence was being amplified. The cleavage of the probe was measured using thin layer chromatography (TLC) to separate cleavage fragments from intact probe ((Holland *et al.* 1991) and Applied Biosystems, Services Literature).

The development of fluorogenic probes (Lee *et al.* 1993) enabled the PCR reaction to be monitored without the need to separate the cleaved probes using TLC. In this case the probe was an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An intact probe does not emit fluorescence due to the close proximity of the reporter dye and the quencher. Forster Resonance energy transfer through space causes this reduction of fluorescence. Adequate quenching can be achieved for probes with the reporter at the 5' end and the quencher at the 3' end (Livak *et al.* 1995) and Applied Biosystems, Services Literature).

The explanation of Q-PCR is illustrated in Figure 2.1. The fluorogenic probe anneals to the target sequence downstream from one of the primer sites and is cleaved by the 5' nuclease activity of *Taq* DNA polymerase as this primer is extended. This cleavage separates the reporter dye from its quencher, thereby increasing the reporter dye signal. The cleaved probe is also removed from the target strand, and therefore does not interfere with the overall PCR process as primer extension can continue past the position of the probe. Additional reporter dye molecules are cleaved from their respective probes with each cycle, resulting in an increase in fluorescence intensity proportional to the amount of amplicon produced. That is, the starting copy number of the nucleic acid target determines the initial and significant increase in fluorescence. The ABI PRISM 7700 Sequence Detection System detects this fluorescence during thermal cycling (Applied Biosystems, Services Literature).



Two fluorescent dyes, a reporter (R) and a quencher (Q), are attached to the probes. When both dyes are attached to the probe, reporter dye emission is quenched

During each extension cycle the Taq DNA polymerase cleaves the reporter dye from the probe.

Once separated from its quencher the, the reporter dye emits its characteristic fluorescence.

Figure 2.1: Stepwise representation of the forklike-structure-dependent, polymerisation-associated, 5' to 3' nuclease activity of Taq DNA polymerase acting on a fluorogenic probe during one extension phase of PCR (figure is an amended version of that found in Applied Biosystems, Services Literature).

The parameter C_T (threshold cycle) is defined as the fractional cycle number at which the reporter fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The threshold is usually set at the beginning to middle of the exponential phase of the PCR amplification before reagents can become limiting. Finally the parameter Rn is the normalised reporter signal (Applied Biosystems, Services Literature).

2.1.1.1. Data Analysis

Before Q-PCR can be used to analyse cDNA samples, the primers need to be optimised. The purpose of this procedure is to determine the minimum primer concentrations giving the maximum change in Rn and minimum non-specific amplification. To do this, various primer concentrations are tested on the same template sample and the primer concentration generating the desired results is used in all following Q-PCR reactions (Applied Biosystems, Services Literature). Relative quantitation with data from the ABI PRISM 7700 Sequence detection system (using version 1.6 software) can be performed using one of two methods:

a) The Standard Curve Method: where a set of standard samples of known concentration is run on each reaction plate. For absolute quantitation, plasmid DNA or *in vitro* transcribed RNA is used to prepare the standards. Concentration is measured by A_{260} and converted to the number of copies using the molecular weight of the DNA or RNA. Where absolute values are not needed, the relative standard curve method is used. It is much easier to prepare standard curves for relative quantitation, because quantity is expressed relative to a reference sample, usually referred to as the calibrator. For all experimental samples, target quantity is determined from the standard curve and divided by target quantity of the calibrator. Thus, the calibrator becomes the 1X sample, and all other quantities are expressed as an n-fold difference relative to the calibrator. The calibrator can be one of the experimental samples, usually the sample expected to have the lowest amount of targeted template. All that is required of the standards is their relative dilutions i.e. the unit value is not needed because the sample quantity is divided by the calibrator quantity. Any RNA or DNA sample can be used to make up the standards. If a DNA standard curve is used for relative quantitation of RNA then it must be assumed that the reverse transcription efficiency of the target is the same in all samples, but exact value of this efficiency need not be known (Applied Biosystems, Services Literature).

In most cases the quantitation of each sample including the calibrator needs to be normalised to an endogenous control (in the work presented in this chapter 16S rRNA was used). Standard curves are therefore needed for both the target and the endogenous reference. The normalised target value is the target amount divided by the endogenous reference amount (Applied Biosystems, Services Literature). The data analysed in these experiments was subjected to a further normalisation step where the mRNA levels were estimated as a cellular value (Hansen *et al.* 2001).

b) Comparative C_T Method: this method is similar to the standard curve method except that it uses arithmetic formulas to achieve the same results for relative quantitation (Applied Biosystems, Services Literature). Once the primers have been validated, it is possible to eliminate the use of standard curves for relative quantitation. This validation experiment demonstrates that efficiencies of target and reference are approximately equal. The absolute value of the slope of the log input amount (DNA/RNA added to the reaction) compared to the difference in C_T value between the target and endogenous reference values, should be <0.1. Once this is proven, one can use arithmetic formulas to achieve the same result as for relative quantitation i.e. eliminates the use of standard curves with every reaction (Applied Biosystems, Services Literature).

2.2. Materials and Methods

2.2.1. Bacterial Strains

2.2.1.1. *M. tuberculosis*

For the purpose of the present study, a collection of DNA samples together with a corresponding database containing clinical and molecular information from *M. tuberculosis* clinical isolates, was available. The isolates were obtained from TB patients residing in the Western Cape Province of South Africa. The isolates were grouped into 37 strain families and 230 unique isolates. The grouping into families was done by determining the number of times the insertion element IS*6110* (McAdam *et al.* 1990; Thierry *et al.* 1990) appears within the genome of the isolate. A family of strains is defined as strains which are > 65% related to each other in terms of RFLP IS*6110* DNA banding patterns. Unique isolates contain unique IS*6110* patterns which do not fall into any of the families in the database (Warren *et al.* 2000). All the isolates selected from the database had been genotyped by others (R. Warren and M. Richardson, Department of Medical Biochemistry, University of Stellenbosch). All known information about the strains used in the work presented here is provided in Appendix 2, Table 2.1. As a reference strain, H37Rv was used. A family tree, of the majority of the strain families used in this study, is presented in Appendix 4, Figure 4.1.

2.2.1.2. *M. bovis* BCG (Bovine Calmette-Guerin)

The *M. bovis* BCG Pasteur strain ATCC 35734 was used in this study (Behr 2002).

2.2.2. Buffers and Reagent Mixes

Unless stated otherwise all reagents used in this study were from Sigma-Aldrich.

2.2.2.1. ADC and OADC

	ADC	OADC
Oleic Acid	-	0.25g
BSA (fraction V)	25g	25g
Glucose	10g	10g
Catalase	0.75ml	1ml
NaCl	-	4.25g
dH_2O to make final volume	500ml	500ml

2.2.2.2. 7H9 medium

Consists of 2.35g Middlebrook 7H9 powder (DIFCO), 2ml glycerol, 500µl Tween80 and 450ml dH_2O. The mixture was autoclaved at 121°C for 15 minutes only. After the medium had cooled to 45°C, ADC (see section 2.2.2.1) was added using a sterilised filter to a final volume of 10% (i.e.50ml to 450ml of medium).

2.2.2.3. 7H10/7H11 agar

Consists of 9.5g 7H11 (7H10 for *M. bovis* BCG) powder (DIFCO), 2.5ml glycerol and 450ml dH_2O. Medium was mixed thoroughly and autoclaved at 121°C for 15 minutes only. Immediately before plates were poured, OADC (see section 2.2.2.1) was added to final a volume of 10% (i.e. 10ml to 90ml medium).

2.2.3. Growth of Cultures

An aliquot of 100µl, from a glycerol stock of the reference strain *M. tuberculosis* H37Rv, was added to 10ml of 7H9 medium (see section 2.2.2.2) and left stirring (using magnetic stirrers) at 37°C. For *M. bovis* BCG, 1ml of a glycerol stock was added to 100ml of 7H9 medium and cultures were placed at 37°C in a rolling incubator. Once the growing cultures reached an absorbance (see section 2.2.4) at 600nm of 0.8, an aliquot of 100µl of the *M. tuberculosis* culture was added to nine labelled Falcon tubes containing 10mls of 7H9 medium; and for *M. bovis* BCG, 1ml of culture was added to 100ml of 7H9 medium. Eight of the *M. tuberculosis* containing Falcon tubes were left stirring at 37°C and the ninth culture was used immediately in subsequent reactions (as day 0). An aliquot of *M. bovis* BCG cells was removed from growing culture at

various time points (viz. 1, 2, 3, 4, 5, 6, 7 days) for RNA extraction (1 X 10^9 cells – volume adjusted accordingly (see section 2.2.7)) and absorbance readings (see section 2.2.4).

At various time points of the growing *M. tuberculosis* cultures (viz. 0, 1, 2, 4, 7, 14, 21 and 28 days) absorbance readings at 600nm (see section 2.2.4), protein extractions (see section 2.2.6) and total RNA extractions (see section 2.2.7) were done and CFUs (see section 2.2.5) were counted.

To test that the cultures were pure *M. tuberculosis* or *M. bovis* BCG, and not contaminated with other bacteria, the ZN stain (Kubica and Kent 1985; Isenberg 1992) was preformed and 5µl of growing culture was plated on blood agar plates. The blood agar plates were analysed for bacterial growth after two days. No growth indicated pure mycobacterium growth.

2.2.4. Absorbance readings of growing mycobacteria cultures

The spectrophotometer was set to a wavelength of 600nm. The machine was zeroed using the sterile 7H9 growth medium (see section 2.2.2.2) as the blank. An aliquot of 1ml of the growing bacterial culture was removed aseptically and placed in a 1ml cuvette and the absorbance was read.

2.2.5. Colony Forming Units (CFUs)

For each CFU count, five 7H10/ or 7H11 (depending on Mycobacterium sp.) medium (see section 2.2.2.3) plates were poured. An aliquot of 1ml was removed from a growing Mycobacterium culture of known absorbance at 600nm and was subjected to four 10X serial dilutions i.e. 100µl from the removed 1ml was added to 900µl of 7H9 medium (see section 2.2.2.2) and this was repeated four times. Each 7H10/7H11 plate was divided into four sections and onto each section five drops of 5µl of one of the dilutions was added and left to dry. This was repeated for each of the four sections. Each plate corresponds to a dilution (i.e. 1X, 10X, 100X, 1,000X and 10,000X). The plates were stored at 37°C and colonies were counted weekly for up to 3 weeks. The number of colonies was counted on the dilution plate where the colonies were most clear and an average of the number of colonies per µl was calculated. This value was adjusted according to the dilution factor.

2.2.6. Protein Extraction

At each specified day, total protein was extracted from the growing culture. Duplicate 1ml culture samples were centrifuged in a benchtop centrifuge (13,800 X g) for 5min, and the tubes were

rotated through 180° and centrifuged again for 5min to produce compact cell pellets. The pellets were washed with 1ml of phosphate-buffered saline (PBS), pH 7.0 (without resupending the cells), and were centrifuged as described above. Pellets were then resuspended in 100µl 1M NaOH and the sealed tubes were placed in an incubator set at 100°C for 10min. The samples were neutralized by adding 20µl of 5M HCl and the volumes were adjusted to 1ml by adding 880µl of PBS, pH 7.0. Samples were then centrifuged for 30min, and 800µl of each supernatant was removed for protein determination. For each sample, the absorbance was measured at 230 and 260nm, and the protein concentration (µg/ml) was determined (Meyers *et al.* 1998).

2.2.7. RNA extraction

For *M. tuberculosis*, RNA was extracted from 6.3ml growing cultures of which absorbance readings were known. For *M. bovis* BCG, RNA from approximately 1 X 10^9 cells (volume was adjusted according to absorbance readings) was extracted from the growing cultures. By CFU analysis (see section 2.2.5) it was calculated that a growing *M. bovis* BCG culture with an absorbance at 600nm of 1 is equivalent to 1 X 10^9 cells/ml. RNA was prepared using the RNA Mini-Prep kit (Qiagen). Changes made to the protocol stated in the Qiagen manual are described below. Cells resuspended in 350µl RLT buffer (Qiagen) (containing 3.5µl ß- mercaptoethanol) were either ribolysed twice at speed 6 for 45 sec (for *M. tuberculosis*) or sonicated in a cold room on ice at 7 microns for 15 sec (for *M. bovis* BCG). The lysed cells were centrifuged, in a benchtop centrifuge, for 2 min at 13,000rpm at 4°C. The supernatant (approximately 350µl) was removed and added to 350µl 100% EtOH and mixed by inverting 3-5 times. The 700µl was added onto a Qiagen column and centrifuged for 20 sec at 10,000rpm. For the DNase step 70µl RDD buffer (Qiagen) was added to 10 µl DNase (Qiagen). The flow-through was removed from the column, 350µl RW1 reagent (Qiagen) was added and the column was centrifuged at 10,000rpm for 20sec. The flow-through was removed; 80µl of the recently prepared DNase solution was added directly onto the column and left at room temperature for 15 min. Finally 350µl RW1 (Qiagen) solution was added and extraction was continued as stated in the Qiagen manual.

Extracted RNA was quantified using absorbance readings at 260nm and 280nm. To test for degradation of RNA, an aliquot of RNA was run on a 1% agarose gel. Loaded RNA sample consisted of 6.5µl RNA, 3.5µl formaldehyde, 10µl formamide and 5µl gel loading buffer. Samples were run at 80 volts for 45 min. All 1.5ml tubes used were certified RNase free (Eppendorf).

2.2.8. Extra DNase Step

To a purified RNA sample, 1U DNase (Promega) per μg RNA, 1X DNase buffer (Promega) and 10U RNase inhibitor (Promega) per μg RNA was added. Finally, nuclease free dH_20 was added to make up required volume. The reaction mix was heated for 30 min at 37°C and the reaction was stopped by adding 1X stop solution (Promega) or heating for 10 min at 65°C. For the experiments outlined here, usually 2μg RNA was subjected to this DNase step and then the reaction mix was halved and each half was used in either the RT+ or RT- reaction (see sections 2.2.9).

2.2.9. Reverse Transcription (cDNA production)

For reverse transcription (RT), the following reagents were added: 12.5U AMV reverse transcriptase (Roche), 1X AMV RT buffer (Roche), 10μM random hexamers (Applied Biosystem), 0.2μg/ul RNase/DNase free BSA (Roche), 1U/μl RNase inhibitor (Ambion), 5μl RNA extraction and nuclease free H_2O (Promega) to make up a final volume of 25μl. The reaction mix was heated at 42°C for 1 hour, then 95°C for 5 min and finally held at 30°C (Roche). For every RT reaction (RT+), a RT minus (RT-) reaction was also done, where the reaction mix lacked reverse transcriptase and RNase inhibitor.

2.2.10. Quantitative PCR (Q-PCR)

All solutions and dilutions were in nuclease free water or were obtained from the manufacturer.

2.2.10.1. Primers and Probes

Table 2.1: The primers and probes used in the Real Time PCR reactions for the *tbnat* gene

Gene	Probe name	[a]Fluorescent Probe sequence	Primer name	Forward primer sequence	Primer name	Reverse primer sequence
tbnat	Taqnat1 prb	5'-/56-FAM/ATCGCGGCGC TACCGATC CA/3*BHQ*-1/-3'	Taqnat1F	5'-GCGTACTTCG ATCGCATCAA-3'	Taqnat1R	5'-TCCTGCAGA ACATCCAGGG-3'
16S rRNA	rRNA prb	5'/5TET/CGGGTTCTCTCGG ATTGACGG T/3*BHQ*-1/-3'	rRNA F	5'-GGCCTTCGGG TTGTAAACCT-3'	rRNA R	5'-CGTAGTTGG CCGGTGCTT-3'

The primers and probes used were designed using the computer program Primer Express™ (Applied Biosytems) and synthesised by Integrated DNA Technologies (www.idtdna.com). The 5' Phosphoramidite conjugated [a]fluorescent dye is underlined (6-FAM and TET). The 3' quencher (italics) used is a black hole quencher (BHQ™-1).

2.2.10.2. Q-PCR Mix (remains constant throughout all experiments):

1X master mix (Applied Biosytem), 5µl RNA, probe to a final concentration of 200nM, primers to concentration of 300nM each (this value had previously been worked out (see section 2.2.10.4) and made up to the final volume 25µl with nuclease free water containing 50ng/µl of yeast RNA (Prepare by diluting 5mg/ml of yeast RNA (Ambion) 1:100 in nuclease free H_2O).

2.2.10.3. PCR Cycle

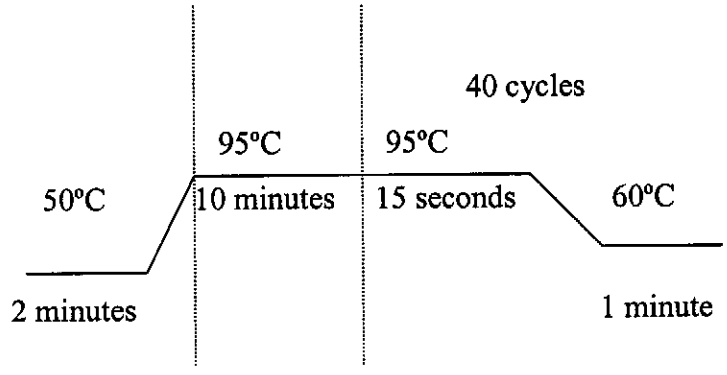


Figure 2.2: Thermal conditions of the Q-PCR cycle

The initial 50°C allows time for contaminating amplicons generated by previous Q-PCR to be degraded. UNG is an Uracil N-glycosylase enzyme that is a constituent of the master mix (Applied Biosystems), and has the ability to degrade DNA in which uradine (dUTP) has been incorporated instead of the usual thymine (dTTP). The deoxynucleoside triphosphates used in the Q-PCR reactions are adenine, guanine, cytosine and uradine (which replaces thymine). The 10 min at 95°C is for the activation of the *Taq* polymerase. These thermal conditions are recommended by the manual and were always used.

2.2.10.4. Primers and Probe Optimisation

To work out the optimal working concentration of the primers and probe, a series of primer concentrations were used in the Q-PCR reaction where the template cDNA remained constant. The cDNA was transcribed from RNA extracted from *M. tuberculosis* cells growing at mid log phase. The probe concentration was usually kept at 200nM. A working stock solution of 5µM was made for both primers and probe.

Table 2.2: Q-PCR Reaction Mix with different primer concentration

Primer Concentration F/R	Q-PCR Mix (25X)	Probe (200nM)	Forward Primer	Reverse Primer	Yeast RNA (50ng/µl)	cDNA
50nM/50nM	55.4 µl	8 µl	2 µl	2 µl	127.6 µl	5 µl
50nM/300nM	55.4 µl	8 µl	2 µl	12 µl	121.6 µl	5 µl
50nM/900nM	55.4 µl	8 µl	2 µl	36 µl	93.6 µl	5 µl
300nM/50nM	55.4 µl	8 µl	12 µl	2 µl	117.6 µl	5 µl
300nM/300nM	55.4 µl	8 µl	12 µl	12 µl	107.6 µl	5 µl
300nM/900nM	55.4 µl	8 µl	12 µl	36 µl	83.6 µl	5 µl
900nM/50nM	55.4 µl	8 µl	36 µl	2 µl	93.6 µl	5 µl
900nM/300nM	55.4 µl	8 µl	36 µl	12 µl	83.6 µl	5 µl
900nM/900nM	55.4 µl	8 µl	36 µl	36 µl	59.6 µl	5 µl

This reaction mix was used for both the *nat* and 16s rRNA primer sets. For primers see Table 2.1. F and R refer to forward and reverse primers respectively. The cDNA used was transcribed from *M. tuberculosis* RNA.

2.2.10.5. Standards for the Standard Curve Method

Dilutions of known amounts of *M. tuberculosis* (H37Rv) DNA or cDNA template were needed when the Standard curve method was being used for Q-PCR. When only relative amounts (i.e. not exact) of mRNA were needed, then dilutions of DNA can be used instead of cDNA. DNA of H37Rv was used as the standards in the subsequent reactions.

The gDNA from H37Rv strain of *M. tuberculosis* used as a template was diluted with nuclease free water by ten times serial dilutions, such that the stock solution could be stored at -20°C in 20µl aliquots to allow a single aliquot to be used for each series of reactions.

2.2.10.6. Primer Validation

When using the Comparative Method (Applied Biosystems, Services Literature) to analyse the data, a once-off primer validation test is required. This demonstrates that the efficiencies of detecting levels of mRNA of the target gene and the endogenous control (16s rRNA) are approximately equal. The absolute value of the slope of the log input amount (DNA/RNA added to the reaction) vs. the difference in C_T value between the target and reference values, should be <0.1.

Once this is proven, one can use arithmetic formulas to achieve the same result as for relative quantitation i.e. eliminates the use of standard curves with every reaction (Applied Biosystems, Services Literature).

cDNA was reverse transcribed from 1μg of RNA (see sections 2.2.8 and 2.2.9). Three sets of dilutions were made from the initial RT reaction using yeast RNA (50ng/μl in nuclease free water). The concentrations of cDNA should be 10/ 1.0/ 0.5 and 0.2ng/μl. Each dilution of cDNA used in the Q-PCR reactions (see section 2.2.10.2) was done in triplicate.

Table 2.3: cDNA Dilutions

cDNA (ng/μl)	Volume of cDNA (μl)	Volume of Yeast RNA (μl)	Final concentration of cDNA (ng/μl)
10	2	18	1
1	10	10	0.5
0.5	8	12	0.2
0.2	10	10	0.1

The starting cDNA concentration was an estimate and assumed that the RT efficiency was 100% (Applied Biosystems 2004). This was used as the first concentration in the dilution series. The cDNA was diluted with nuclease free water containing yeast RNA at a concentration of 50ng/μl. The yeast RNA has not been transcribed to cDNA and therefore does not interfere with the reaction. The final concentration of yeast RNA does vary per reaction but this has an insignificant effect on the final results (Applied Biosystems, Services Literature).

The data obtained from the above experiment should show that the primer set used in Q-PCR produces the same results for the same sample even when diluted. Precise concentrations are not needed and, therefore, for convenience sake the cDNA concentration is determined by assuming 100% RT efficiency (Applied Biosystems 2004).

2.2.10.7. Data Analysis

RNA was extracted from cells at each specified time point (see section 2.2.7) and converted to cDNA by reverse transcription (see sections 2.2.8 and 2.2.9). The cDNA was used in subsequent Q-PCR reactions (see section 2.2.10.2) run on an ABI PRISM 7700 Sequence Detection System (Applied Biosystem). The Standard Curve method was used to analyze the data generated from

the *M. tuberculosis* cDNA and, therefore, on each Q-PCR plate (a plate supplied by Applied Biosystem containing 96 wells) serial dilutions of known concentration of template DNA was used for both sets of primers. The following equation (Hansen *et al.* 2001) was used to calculate the *nat* expression levels at each time interval: mRNA/cell = (mRNA/total RNA) X (total RNA/cell). This equation is used to give a normalised *nat* mRNA level. The following modifications were made: In Hansen *et al.* (2001), the 16S rRNA value per cell was quantified using Fluorescent *in situ* Hybridisation (FISH) and the total 16S rRNA content per total RNA was determined by slot blotting. In this study Q-PCR was used to quantify the amount of 16S rRNA. In order to determine the rate of 16s RNA to cell number, the values were divided by one of three variables:

a) number of CFUs per ml of culture (see section 2.2.5)
b) protein concentration of cells in culture (see section 2.2.6)
c) absorbance at 600nm of culture (see section 2.2.4)

All three of these methods give an estimation of the total number of bacterial cells in the growing culture. The quantified 16S rRNA was also divided by the total RNA concentration to give a 16S rRNA content per total RNA. The total RNA concentration was determined using the absorbance readings at 260nm and 280nm. The cellular total RNA is estimated by dividing the 16S rRNA value per cell value by the 16S rRNA content per total RNA (total RNA/cell) (Hansen *et al.* 2001).

The comparative method was used to analyze the data generated for the *M. bovis* BCG cDNA. This method was used as more samples can be quantified on one plate. This method is therefore less time consuming and more cost efficient. The method followed was that explained in the Applied Biosystems, Services Literature.

2.2.10.8. Verification of the sensitivity of Q-PCR

The *nat* gene was amplified in a PCR reaction using the primers *nat5* and *nat3* (see section 3.2.1) and an annealing temperature of 56°C. The PCR product was diluted 10X and this was used as the first dilution to spike a sample of cDNA. From this dilution three 10X serial dilutions were made and labelled 10X, 100X and 1000X. The original solution used for the serial dilutions was labelled 1X.

Each dilution was then used to spike 3µl of a cDNA sample with 3µl of diluted PCR product; this combined sample under goes Q-PCR in triplicate (see section 2.2.10.2).

2.3. Results

2.3.1. Optimal Primer Concentration

Before Q-PCR can be applied to samples, the primers to be used in the reaction need to be optimised. This identifies the lowest primer concentration that produces early and strong fluorescence during the PCR cycle.

An example of the results for the Taqnat1 primers is presented in Figure 2.3. Each coloured line in the graph represents the fluorescence produced during each PCR cycle, where the reaction mix contained a specific primer concentration. The primer concentration (300nM/300nM) producing the most efficient results are shown here.

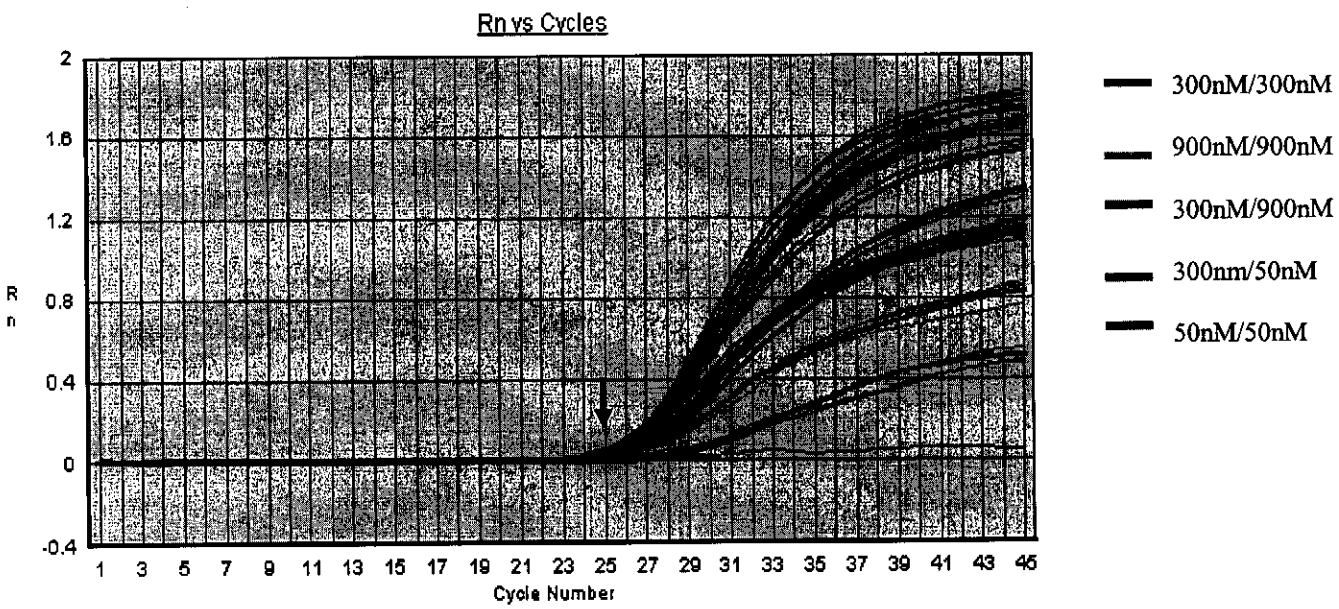


Figure 2.3: The effect of primer concentration on the efficiency of early amplification
Results generated by ABIPRISM 7700 Sequence Detection System for each Taqnat1 primer set concentration. Rn is the amount of fluorescence detected and cycle number is the cycle of the PCR reaction at which fluorescence is detected. Each coloured line shows the fluorescence detected (e.g. the efficiency of the reaction) for each PCR reaction with a different primer concentration. Four primer concentrations have been selected and are shown on the right of the figure. Arrow indicates the C_T value.

The optimal primer concentration is that which has a low C_T value but a relatively high Rn value (see Figure 2.3). If more than one primer concentration fits these specifications then the lowest

primer concentration is chosen (more cost effective). For both the 16S rRNA and Taqnat1 primers a concentration of 300nM was chosen for both the forward and reverse primer (see Figure 2.3).

2.3.2. Primer Validation

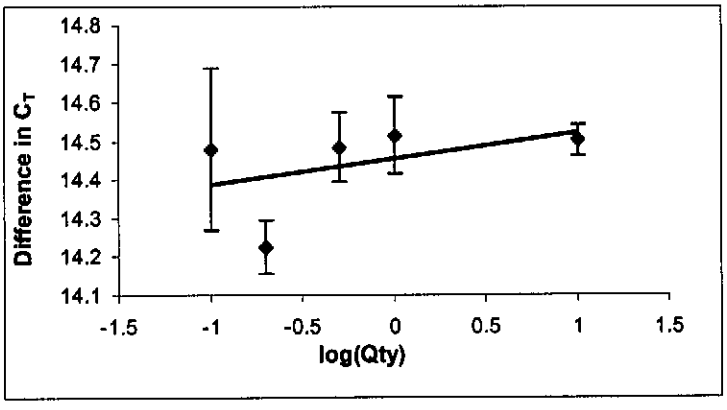
When the comparative method is being used, the primers need to be validated. Once the primers are validated they may be used to analyse samples without standard concentrations of gDNA being present with each reaction plate. This method saves on costs, as fewer Q-PCR reactions are needed to generate results for each sample. The results obtained are not as accurate as the Standard curve method, but because only relative values are needed it was decided to use this method to analyse the expression of *nat* in *M. bovis* BCG.

A control sample of cDNA was diluted and each dilution was subjected to Q-PCR. The difference in the 16S rRNA and *nat* mRNA C_T value was calculated for each dilution. These values were plotted and the linear regression of the values was worked out.

Table 2.4: The C_T values for both the Taqnat1 primer/probe and 16S rRNA primer/probe sets and the difference between the averages for each cDNA concentration.

Qty of cDNA (ng/µl)	Log(Qty)	Ave *nat* C_T	STDEV	Ave *rRNA* C_T	STDEV	Difference Ave *nat* – Ave rRNA	STDEV
10	1.00	31.10	+/- 0.14	16.60	+/- 0.09	14.51	+/- 0.04
1	0.00	32.52	+/- 0.13	18.00	+/- 0.27	14.52	+/- 0.10
0.5	-0.30	33.55	+/- 0.18	19.06	+/- 0.06	14.49	+/- 0.09
0.2	-0.70	34.71	+/- 0.13	20.48	+/- 0.04	14.23	+/- 0.07
0.1	-1.00	35.99	+/- 0.38	21.51	+/- 0.08	14.48	+/- 0.21

The quantity (Qty) of cDNA transcribed from *M. bovis* RNA was known, and measured as a concentration of cDNA. Results obtained were done in triplicate. Average (Ave) C_T values and their corresponding standard deviations (STDEV) were calculated. The 'difference' is the average C_T value of *nat* subtracted from the average C_T of 16S rRNA. The cDNA concentration was worked out from the initial amount of RNA added to the RT reaction. The RT efficiency was estimated as 100% (Applied Biosystems 2004).



Linear regression equation:
$y = 0.069x + 14.456$

Figure 2.4: Graph of the linear regression of the difference between the C_T values of the two primer/probe sets. Data used to produce this figure are found in Table 2.4.

The slope (0.069) of the primer validation tests was < 0.1 (see Figure 2.4), therefore, proving that the primer sets are valid (i.e. their efficiency is comparable) and can be used in subsequent Q-PCR reactions where the Comparative Method will be used. This experiment shows that results obtained are not dependent on the concentration of cDNA, implying that varying cDNA concentrations will not affect the final outcome of the results.

2.3.3. Verification of the sensitivity of Q-PCR

The accuracy and sensitivity of the methods used to produce the results presented in this chapter were tested. This was achieved by spiking a sample of cDNA with different serial dilutions of *nat* template. The relative increase of template generated using Q-PCR compared to the dilution used to spike the cDNA sample was tested, i.e. if 10X more *nat* template was added to the cDNA sample, did Q-PCR generate a result of 10-fold increase in template?

Table 2.5: Q-PCR results where cDNA has been spiked with dilutions of PCR amplified *nat*

CDNA Sample	Ave *nat* C_T	Ave *rRNA* C_T	Ave *nat* – Ave rRNA	STDEV	Difference from cDNA	Relative to cDNA
cDNA	33.67	20.75	12.91	+/- 1.01	0.00	1.00
cDNA 1000X	16.37	22.59	-6.22	+/- 0.15	-19.13	5.74E+05
cDNA 100X	12.68	22.56	-9.88	+/- 0.20	-22.79	7.27E+06
cDNA 10X	9.24	22.23	-12.99	+/- 0.31	-25.90	6.28E+07
cDNA 1X	6.49	21.64	-15.15	+/- 0.27	-28.06	2.80E+08

Each Q-PCR reaction consisted of 3µl cDNA and 3µl of diluted *nat* PCR product (1X to 1,000X dilution), where cDNA was not spiked 3µl ddH_2O was added. The relative increase is correlated to the C_T value of cDNA (relative to cDNA) not spiked with *nat* PCR product (i.e. cDNA not spiked with DNA had a relative increase of 1). For all equations please see Applied Biosystem's Services Literature.

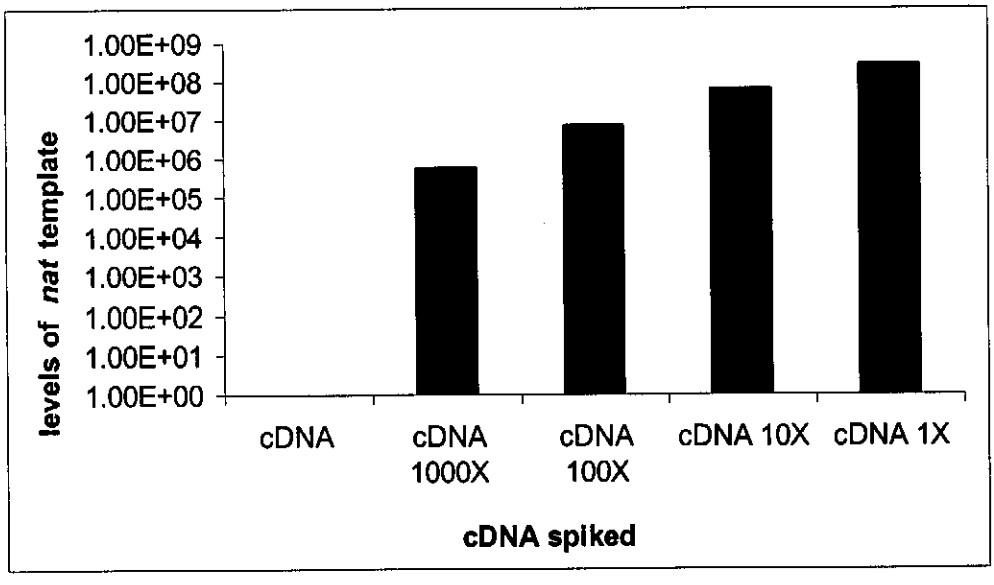


Figure 2.5: Q-PCR can distinguish accurately between 10-fold dilutions
Graph portraying the relative increase as cDNA is spiked with varying dilutions of amplified *nat*. Data used to produce this figure are found in Table 2.8.

It is evident from Figure 2.5 that Q-PCR using the comparative method distinguished between each 10X serial dilution i.e. each reaction spiked with 10X more PCR product was shown to have approximately a 10-fold increase in *nat* template using Q-PCR.

2.3.4. Standard Curve Method – *M. tuberculosis* Data

To generate results for the expression of *nat* in *M. tuberculosis,* the standard curve method was used. The copy number of a sample of DNA was worked out and serial dilutions of this sample were made i.e. the amount (in ng) of *nat* and 16S *rRNA* template was known. These DNA dilutions, in effect a standard curve, were used in duplicate on each reaction plate containing samples with unknown amounts of template.

The regression equation was calculated for the standards (difference in C_T vs. log(Qty)). For the data to be accepted the slope of the equation must be approximately –3, anything greater than –4 is rejected. Similarly the correlation co-efficient of the equation should be greater than 0.95 (see Figure 2.6).

Table 2.6: Results of the set of DNA standards used in Q-PCR reactions for the quantification of *nat* mRNA during the growth cycle of *M. tuberculosis* H37Rv

			nat probe		16S *rRNA* probe	
Sample	copy #	log copy #	C_T 1	C_T 2	C_T 1	C_T 2
std 1	1.00E+00	0	34.60	26.82	36.42	28.28
std 1	1.00E+00	0	34.60	27.14	35.13	28.19
std 2	1.00E+01	1	31.55	23.73	32.13	24.77
std 2	1.00E+01	1	31.43	23.86	32.04	24.64
std 3	1.00E+02	2	27.25	20.40	27.84	21.58
std 3	1.00E+02	2	27.30	20.51	28.03	21.43
std 4	1.00E+03	3	23.32	17.21	24.30	18.02
std 4	1.00E+03	3	23.64	17.29	24.21	18.00
std 5	1.00E+04	4	20.09	14.96	20.34	14.92
std 5	1.00E+04	4	20.00	15.08	20.33	14.87
std 6	1.00E+05	5	16.04	14.00	16.27	12.77
std 6	1.00E+05	5	16.12	14.10	16.18	12.71

C_T 1 and C_T 2 are the average cell cycle values of the DNA standards (with known copy number (copy#)) used in two different Q-PCR reactions on separate plates.

a) b)

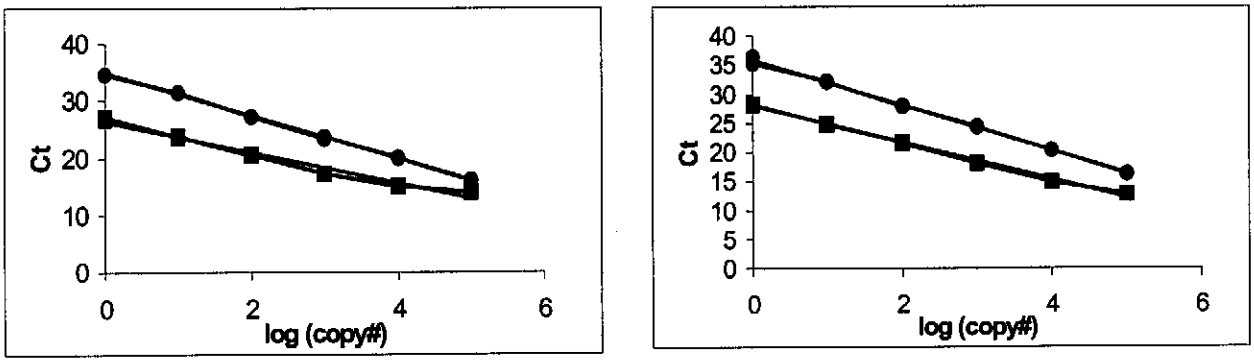


Figure 2.6: The linear regression of both sets of primers applied to the DNA standards of two separate Q-PCR reactions on different plates (a, b).

The C_T value for each standard was plotted against the log of the number of copies of template present (log(copy#)). Data used to produce these figures are found in Table 2.6.

Linear regression equations:

a) *nat* DNA (Circles): y = -3.733x + 34.823, R2 = 0.9989 and
16S rRNA DNA (Squares): y = -3.9051x + 35.865, R2 = 0.9981

b) *nat* DNA (Circles): y = -3.2535x + 27.001, R2 = 0.9993 and
16S rRNA DNA (Squares): y = -3.1543x + 27.901, R2 = 0.9952

To calculate the amounts of template in the unknown RNA samples, the values calculated needed to be normalised. This was achieved by dividing the calculated template amounts by the estimated cell number (calculated as cell volume) from which the RNA was extracted, i.e. the value of mRNA per cell volume. The total cell volume was estimated using absorbance readings and total protein concentration. All cultures showed no bacterial contamination using the ZN stain and the blood agar plates.

Table 2.7: mRNA levels of *nat* at different growth stages of *M. tuberculosis*

a)

	Absorbance at 600nm				Relative mRNA levels of *nat*		Relative to day 28			
Day	Trial 1	Trial 2	Average	STDEV	Trial 1	Trial 2	Trial 1	Trial 2	Average	STDEV
0	0.00	0.04	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.04	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.07	0.04	0.05	0.00	1.76E+02	0.00	0.11	0.05	0.08
4	0.25	0.20	0.23	0.03	2.70E+03	2.81E+03	0.52	1.74	1.13	0.86
7	0.58	0.92	0.75	0.24	1.83E+05	1.17E+05	34.93	72.26	53.59	26.40
14	1.29	1.25	1.27	0.03	3.03E+04	4.88E+04	5.80	30.16	17.98	17.23
21	1.04	1.26	1.15	0.15	5.02E+04	1.76E+02	9.60	9.94	9.77	0.24
28	1.38	1.32	1.35	0.05	5.23E+03	1.62E+03	1.00	1.00	1.00	0.00

b)

Protein Concentration (μg/ml)					Relative mRNA levels of *nat*		Relative to day 28			
Day	Trial 1	Trial 2	Average	STDEV	Trial 1	Trial 2	Trial 1	Trial 2	Average	STDEV
0	17.48	19.67	18.57	1.55	0.000	0.000	0.00	0.00	0.00	0.00
1	10.37	8.43	9.40	1.37	0.000	0.000	0.00	0.00	0.00	0.00
2	5.88	11.35	8.61	3.87	0.000	0.000	0.00	0.10	0.05	0.07
4	48.22	40.93	44.58	5.15	0.002	0.002	0.56	1.33	0.94	0.54
7	106.25	122.36	114.31	11.39	0.142	0.126	39.88	84.08	61.98	31.26
14	166.99	148.35	157.67	13.18	0.034	0.059	9.46	39.26	24.36	21.07
21	201.67	150.40	176.04	36.25	0.037	0.019	10.45	12.90	11.68	1.73
28	291.46	203.66	247.56	62.09	0.004	0.001	1.00	1.00	1.00	0.00

This table shows the relative *nat* levels relative to absorbance readings at 600nm (a) and protein concentration (b). For a complete print out of all the results used to obtain these results see APPENDIX 1 section 1.1 and 1.2.

For every sample tested, the respective RT minus (RT-) reaction also undergoes Q-PCR to ensure that the RNA does not have DNA contamination. To ensure that the reagents are not contaminated a 'no template' control (NTC) is also done for each reaction mix. The Q-PCR reaction of the RT- and the NTC should not produce C_T values less than 37, otherwise DNA contamination present is significant. To ensure C_T values greater than 37 the RNA was subjected to a final DNase step just before the RT reaction. All the controls used in the reactions to obtain the data presented here, had C_T values greater than 37.

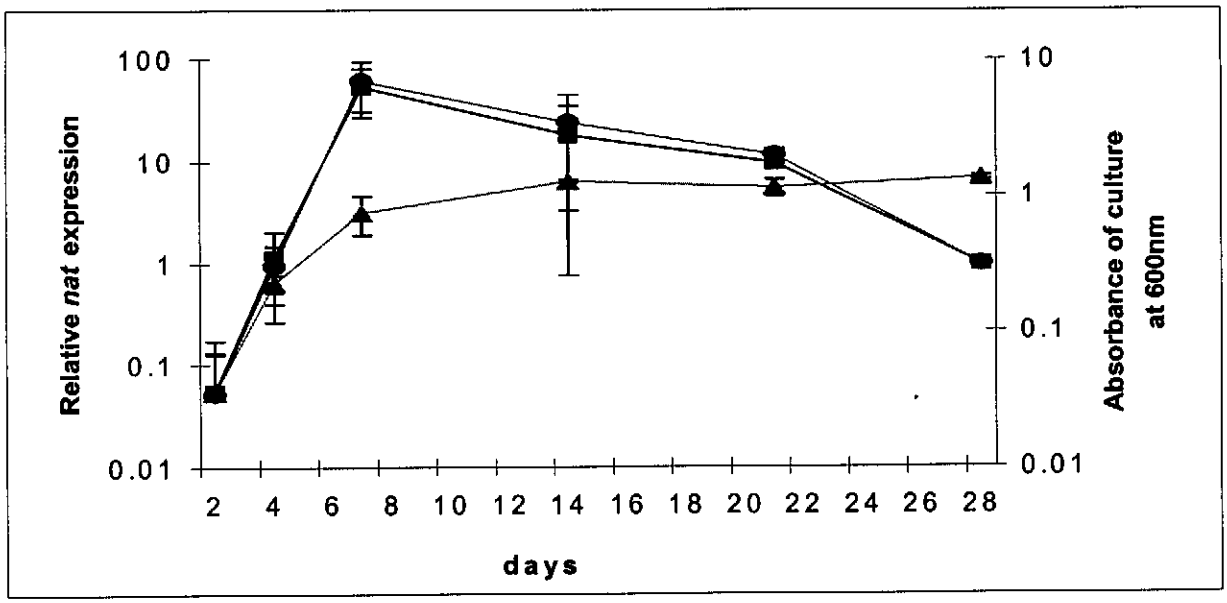


Figure 2.7: The expression of *nat* during the growth of strain *M. tuberculosis* H37Rv.

This figure portrays the results shown in Table 2.7. Triangles indicate the average standard growth curve, squares indicate relative *nat* expression measured against absorbance readings (A_{600}) and circles indicate relative *nat* expression measured against protein concentration (µg/ml).

Figure 2.7 presents the average of two standard growth curves (over a period of 28 days) of the *M. tuberculosis* strain H37Rv, using absorbance readings at 600nm and protein concentrations, and the relative increase of *nat* expression to *nat* expression on day 28 (the expression level of *nat* on day 28 is therefore 1). The levels of *nat* are calculated from Q-PCR data and expressed as mRNA concentration correlated to cell number (*nat* mRNA/cell). Cell number is estimated using the absorbance readings at 600nm and protein concentrations. The CFU results were inaccurate due to clumping and were discarded. For the Q-PCR mRNA values to have any significance, the values need to be normalised. The use of 16S rRNA on its own was not sufficient for a normalisation step as the 16S rRNA levels were slightly affected by the stage the cells were in during the growth cycle (Hansen *et al.* 2001; Gonzalez-y-Merchand *et al.* 1998; Vandecasteele *et al.* 2001; Cangelosi and Brabant 1997). The following equation: mRNA/cell = (mRNA/total RNA) X (total RNA/cell) (Hansen *et al.* 2001)was adapted to give more realistic expression levels of *nat*. An estimated cellular total RNA (total RNA/cell) was calculated by dividing the amount of 16S rRNA per cell number by the amount of 16S rRNA per total amount of RNA.

2.3.5. Comparative Method – *M. bovis* BCG data

The comparative method was used to generate results for the expression of *nat* during the growth cycle of *M. bovis* BCG. No DNA standards are required with each reaction plated, providing the primers have been previously optimised.

Table 2.8: mRNA levels of *nat* at different growth stages of *M. bovis* BCG

Day	OD (600nm)	Relative *nat* expression	STDEV
1	0.08	1.46	0.37
2	0.13	2.01	0.30
3	0.35	2.33	0.08
4	0.89	1.84	0.08
5	1.40	1.56	0.15
6	1.85	1.00	0.31
7	1.88	/	/

The results show the relative increase of *nat* mRNA compared to the ratio of the C_T values generated for *nat* and 16S rRNA cDNA for day 6 (i.e. relative increase on day 6 is 1). The absorbance readings at 600nm (OD) are shown for each day that RNA was extracted. The standard deviation values (STDEV) were derived from 3 replicates of the experiment. For full results see ADDENDIX 1 section 1.3.

Figure 2.8 illustrates the similarity between the expression of *nat* in *M. bovis* BCG with that of *M. tuberculosis* (Figure 2.7), even though two different methods were used to calculate the data. *M. bovis* BCG has an approximate 7 day growth cycle and *M. tuberculosis* a 28 day growth cycle. In both cases *nat* reached its highest expression levels at approximately mid-log phase of the growth cycle of these mycobacteria.

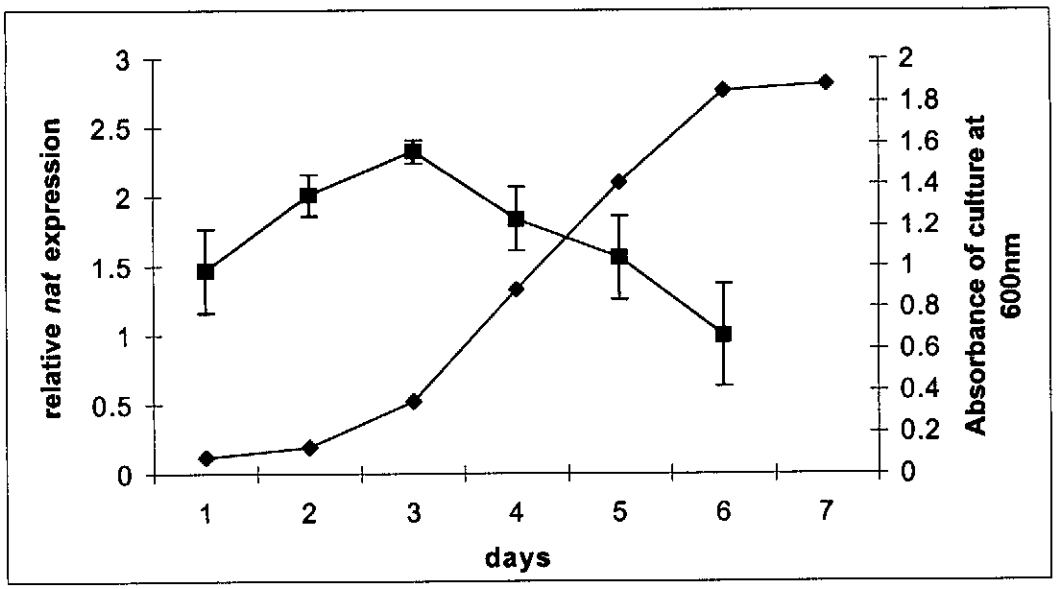


Figure 2.8: The expression of *nat* during the growth cycle of M. *bovis* BCG

The relative increases of *nat* (squares) expression during the growth cycle of *M. bovis* BCG together with optical density readings at 600nm (diamonds). Data used to produce this figure are in Table 2.7.

2.4. Discussion

Targeted disruption of *nat* in *M. smegmatis* resulted in a delayed entry of the cells into exponential phase of growth, extending the length of the lag phase (Payton *et al.* 2001a). Alternatively, overexpression of NAT in *M. smegmatis* decreases the lag phase of the growth cycle of *M. smegmatis* (Bhakta *et al.* 2003a). Preliminary experiments with the non-pathogenic mycobacterium *M. bovis* BCG, in which the *nat* gene has been deleted, appear to indicate that NAT is important in cell growth (Bhakta *et al.* 2003b). This suggests that the presence of NAT is required during early growth. If NAT in *M. tuberculosis* is involved in the growth and development of the bacterium, then NAT could become a potential candidate for drug development. An inhibitor of NAT could result in controlling the TB infection in patients by hindering the growth, and therefore spread of the bacterium within these patients. NAT may be involved in the development of the complex cell envelope, and therefore without a functional NAT alterations may occur within the cell envelope, enabling the host's own immune system to successfully fight the infection.

The expression of *nat* in strain H37Rv of *M. tuberculosis* was monitored over 28 days. The results obtained (see Figure 2.7) show that the gene is expressed during the initial stages of cell growth, with the maximum expression levels being obtained at mid-log phase that occurs around day 7 (OD_{600} = 0.75). As the bacterium reaches stationary phase, the expression of *nat* drops. Increasing expression early in the growth cycle could indicate that NAT is important for growth. In addition to these results, the expression of *nat* was also monitored in *M. bovis* BCG over 7 days. The expression pattern of *nat* in this bacterium showed a similar pattern to that of *M. tuberculosis* (see Figure 2.7 and Figure 2.8). The expression of *nat* in *M. bovis* BCG was found to occur early in the growth cycle of the bacterium, reaching maximum levels during early to mid-exponential phase of the growth cycle.

To obtain these results (*M. bovis* BCG), a different method was used to analyse the Q-PCR data, that being the comparative method. For *M. tuberculosis* the standard curve method was used to analyse the results, however, similar findings were achieved. This suggests that the results obtained and methods used are valid, thereby reiterating that *nat* is expressed early in the growth cycle of the bacterium and reaching maximum levels during mid-exponential phase of the growth cycle.

Methods used previously to monitor *nat* expression (Payton *et al.* 2001a; Payton *et al.* 2001b; Upton *et al.* 2001b; Payton *et al.* 1999) were not sensitive enough to detect levels of expression at the early phase of the bacterial growth. These techniques included northern and western blots. The negative aspect of these techniques is that large amounts of starting product (i.e. growing culture) are needed for detection of either the mRNA or protein (Smith *et al.* 2002). Cultures during the early stages of growth do not always produce enough target to allow detection using these techniques. In this study Q-PCR was shown to be highly sensitive and has been able to detect expression very early in the growth cycle of the bacterium. The accuracy of this technique was also verified by spiking cDNA with different dilution of *nat* PCR product. When ten times mores target sequence was added to the reactions, an approximate ten-fold increase was seen in the quantitative values generated (see Figure 2.5).

Finally, similar to our conclusions that *nat* is expressed early in the growth cycle reaching maximum levels at mid-exponential phase of the growth cycle, is the work done on *Salmonella typhimurium* (*S. typhimurium*) (Payton *et al.* 2001b). Here, the western blot technique was used to visualise proteins over a 4-hour period after an initial 1 in 10 inoculation. The expression of

NAT in *S. typhimurium* was shown to increase with time, reaching maximum levels by late exponential/ early stationary phase (Payton *et al.* 2001b). This implies that NAT protein increases during the growth cycle of the bacterium reaching maximum levels towards the end of the exponential phase of the growth cycle. These results correspond with the results reported in this chapter. However, the work here has enabled the expression of *nat* to be detected very early in the growth cycle and then accurately monitored throughout the entire growth cycle of the bacterium. These results show for the first time how the *nat* gene is expressed during the entire growth cycle of *M. tuberculosis* and *M. bovis* BCG.

These results are in concordance with those obtained using *M. smegmatis nat* mutants, which taken together, show that early expression of *nat* is important for early growth of mycobacteria. If this proves correct then NAT may be a good candidate for drug development.

CHAPTER 3

ANALYSIS OF *M. tuberculosis nat* FOR POLYMORPHISMS

The aim of this chapter was to analyse the *nat* gene for polymorphisms. The *nat* gene and its flanking regions were screened for polymorphisms by sequencing. The previously identified G619A Single Nucleotide Polymorphism (SNP) was further investigated to determine how widespread it is amongst the available strain families, and to determine whether isolates with the SNP showed a change in their INH susceptibility levels when compared to the reference strain H37Rv.

3.1. Introduction

A G619A SNP has been identified in the *nat* gene of *M. tuberculosis* (Upton et al. 2001). This SNP confers a change from Gly to Arg at amino acid 207, and was identified in approximately 20% of a series of genetically distinct *M. tuberculosis* clinical isolates (Upton et al. 2001). Low-level changes in INH susceptibility were detected in isolates containing this SNP and the recombinant protein demonstrated a 10-fold decrease in affinity for INH (Upton et al. 2001).

M. smegmatis NAT is 60% homologous to *M. tuberculosis* NAT, it acetylates INH (Payton *et al.* 1999) and there is a Gly at the amino acid 207 as in the wild-type *M. tuberculosis* NAT. Using site directed mutagenesis; a *M. smegmatis* Arg 207 NAT mutant was constructed. A 3D structure from crystals was obtained for both the wild-type and the mutant form of NAT, and it was clearly evident that changes in the active site environment did occur due to the 207 Arg mutation (Kawamura *et al.* 2003). It was also shown that the effect of the mutation has implications for changes in protein solubility, folding or stability (Kawamura *et al.* 2003).

To further understand the implications of this SNP, it was screened for in 37 different strain families of *M. tuberculosis*. This was to identify whether the SNP was randomly dispersed or restricted to certain strain families. The INH minimum inhibitory concentration (MIC) of the isolates with the SNP compared to those without the SNP was analysed to determine whether the SNP affects INH susceptibility. The open reading frame (ORF), upstream and downstream regions of different *M. tuberculosis* strains were sequenced to ascertain whether other polymorphisms (genomic) are associated with the *nat* gene.

Identifying new polymorphisms and the frequency with which they occur in the different strain families may give some information as to whether the polymorphisms are random events, or whether they result in the development of new strain families. The INH MICs will give an indication whether or not the SNPs result in changes in susceptibility to INH and predispose certain strains to development of drug resistance.

3.2. Materials and Methods

Unless otherwise stated, all reagents are from Sigma-Aldrich.

3.2.1. Basic PCR

3.2.1.1. PCR Mix

The PCR mix consisted of 1X concentration buffer (Promega, supplied as 10X), 1.5mM $MgCl_2$ (Promega), 0.2μM forward primer, 0.2μM reverse primer, 0.1U/μl Taq polymerase (Promega), 0.8mM dNTPs (Promega), approximately 1ng/μl genomic DNA and the final volume (20/100μl) was made up with double distilled water (ddH_2O).

3.2.1.2. PCR Cycle

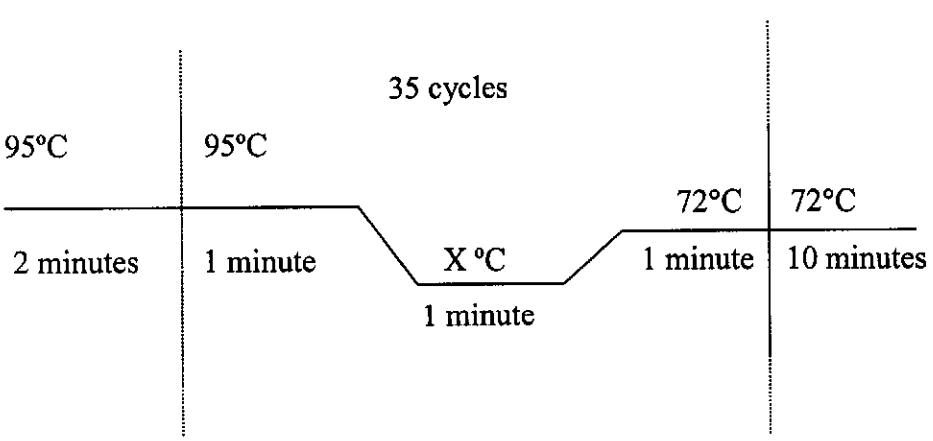


Figure 3.1: Thermal conditions of the PCR cycle.

X – annealing temperature that is primer specific.

Table 3.1: Primers used to amplify and sequence genes associated with INH resistance, and the ORF and flanking regions of the *nat* gene.

Gene	Primers	Sequence of Primers	Annealing Temperature	Size (bp)
KatG (1st frag)	RTB 51	CAGAAACCACCACCGGAGCC	59°C	945bp
	RTB 39	GCTGGTGATCGCGTCCTTAC		
(2nd frag)	RTB 59	TGGCCGCGGCGGTCGACATT	66°C	804bp
	RTB 36	TCGGGGTCGTTGACCTCCCA		
(3rd frag)	RTB 57	CCGACGATGCTGGCCACTGA	64°C	997bp
	RTB 34	GACCTCGACAAGCGCCCGCA		
inhA promoter	inhA P5	CGCAGCCAGGGCCTCGCTG	60°C	246 bp
	inhA P3	CTCCGGTAACCAGGACTGA		
inhA gene	inhA 51	CGGGCAACAAGCTCGACGGG	64°C	169 bp
	inhA 31	GGGTTCATGATCGGCAGGAG		
ahpC promoter	ahpC51	GCTGATTGTCCGAGAGCATCG	60°C	701bp
	ahpC31	GGTCGCGTAGGCAGTGCCCC		
kasA gene	kasA 51	ATTGAGTCGGAGAACCCCGA	56°C	1389 bp
	kasA 31	CCTTCCATATCGGTCCGACT		
ndh gene	ndh3s	GACAGATCGCCGAGCTGGC	60°C	372bp
	ndh3as	GTGCTGCCCGACCTGTCCA		
Upstream *nat* gene	TBnat3a	CCTGCAGAACATCCAGGGTT	60°C	870bp
	TBnat3b	CTAGGCTCACCGATCATCGA		
ORF *nat* gene	*nat5*	GACGAGGTCAAATGGCAAC	60°C	928bp
	nat3	GGGGTTCGTTTGTTCGGATA		
Downstream *nat* gene	Tbnat1	TGGAATCTGAGGTGTGGGAG	60°C	1122bp
	Tbnat2	TATCCGAACAAACGAACCCC		

The positions of the *nat* primers are shown in Figure 3.2. The amplified product size of each primer set is represented as the number of base pairs (bp). Primers *nat* 5 and *nat* 3 were taken from the literature (Upton et al. 2001). The *KatG* gene is large and was divided into three sections, named as first fragment (1st frag), second fragment (2nd frag) and third fragment (3rd frag).

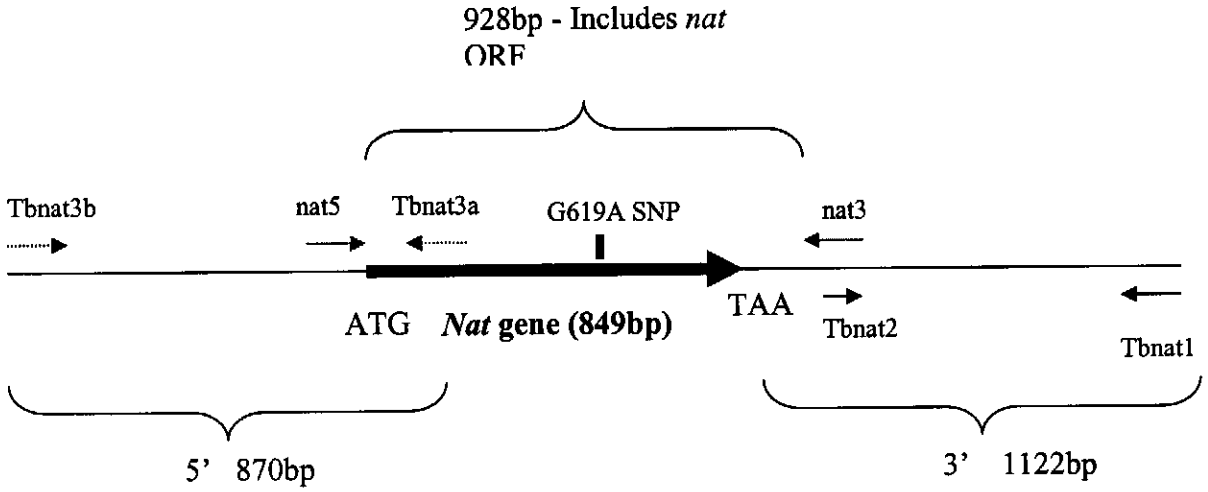


Figure 3.2: Schematic diagram of the positions of primers used to analyse the ORF, upstream (5' end) and downstream (3' end) regions of the *M. tuberculosis nat* gene.

→ the 849bp *M. tuberculosis nat* gene includes the start codon (ATG – 5' end), but not the stop codon (TAA – 3' end).

⇢ ⇠ Position of Tbnat3b and Tbnat3a primers respectively. These primers amplify a product of 870bp, that includes 791bp upstream the 5' end and 79bp within the ORF (downstream the 5' end) of *nat*.

→ ← Position of *nat5* and *nat3* primers respectively. These primers amplify the ORF of *nat* plus 20 bases upstream the start codon and 56bp downstream the stop codon the total amplicon being 928bp.

→ ← Position of Tbnat2 and Tbnat1 primers respectively. These primers amplify 1122bp downstream of *nat* (3' end).

*nat*3 and Tbnat2 are complementary to each other and lie 56bp downstream (3' end) from the stop codon.

3.2.2. Restriction Digest

The restriction digest reaction mix consists of 1X restriction buffer (supplied as 10X concentration), 10U of enzyme per mg DNA and ddH_2O to a final volume (20/50µl). The incubation temperature and duration was enzyme dependent. For the Bsg1 enzyme (NE Biolabs), S-adenosylmethionine (SAM) (NE Biolabs) was added to a final concentration (80µM) of 1X (supplied as 400X).

3.2.3. Analysis of the G619A NAT and T529C SNPs

PCR-RFLP analysis, using the restriction enzymes BsmA1 (NE Biolabs) and Bsg1 (NE Biolabs), was used to screen *M. tuberculosis* DNA for the G619A and T529C SNPs, respectively. A total of 468 representative clinical isolates obtained from the *M. tuberculosis* collection were used to screen for the G619A SNP, and 250 clinical isolates were used to screen for the T529C SNP (identified in this study, described later). These isolates represented all 37 available strain families as well as 10 unique isolates (isolates that could not be classified into a family) (Warren *et al.* 2000). For more information on the strain families used in this study please see section 2.2.1 and Appendix 4.

The primers *nat* 5 and *nat* 3 were used in the PCR reaction (see section 3.2.1 and Table 3.1). The amplified PCR products were restricted (see section 3.2.2) with either BsmA1 or Bsg1 and incubated at 55°C or 37°C, respectively, for 3 hours. The restricted product was run on a 1% agarose gel at 100V for 1 hour.

3.2.4. Sequence Analysis of *M. tuberculosis* isolates

Each gene (*katG, inhA, kasA, ahpC* and *ndh*) known to have mutations associated with INH resistance (Ramaswamy and Musser 1998;Lee et al. 2001) was sequenced, in the INH resistant isolates, to identify mutations that may be responsible for the resistance in these isolates. The isolates screened were: isolate 131 from family 2, isolate 404 from family 28, isolate 766 from family 18 and isolate 816 from family 133. These isolates all showed resistance to INH but mutations associated to their resistance were unknown. The PCR conditions and primers used for the amplification and sequencing of these genes can be viewed in section 3.2.1 and Table 3.1. The ORF and immediate upstream and downstream regions of the *nat* gene of five isolates were amplified and sequenced (for primers see Table 3.1 and Figure 3.2). The five isolates consisted of two INH susceptible isolates (H37Rv and isolate 208 from family 11), two INH resistant isolates (isolate 131 from family 2 and isolate 816 from family 133) and one isolate from family 3 (isolate 1430). The isolates found in family 3 have the G619A SNP, but unlike isolates in family 28, the *nat* gene had not been screened for mutations by sequence analysis. The INH resistant isolates were selected, as their MICs for INH were high and they did not belong to family 3 or family 28. For more information on the isolates and strain families used in this study please see section 2.2.1 and Appendix 4.

DNA sequencing was carried out using an ABI 3100 instrument, using Bigdye chemistry (Applied Biosystems). Sequence alignments to the *nat* sequence of *M. tuberculosis* H37Rv (Rv3566c, (Cole *et al.* 1998)) were done using DNAMAN Version 4.0 (Lynnon BioSoft. Copyright© 1994-1997).

3.2.5. Verification of newly identified SNP

The newly identified T529C SNP (see section 3.3.2) was subjected to restriction digest analysis using a bioinformative program, DNAssist, to identify an extra restriction site.

The *nat* gene of the *M. tuberculosis* isolate 1430 was amplified by PCR using the *nat* 5 and *nat* 3 primers (see sections 3.2.1). Isolate 1430 was used, as this was the isolate in which the T529C SNP was first identified. The PCR product was restricted with the enzyme Bsg1 and left at 37°C for three hours. The restricted fragment was run on a 1% agarose gel.

3.2.6. Minimum Inhibitory Concentration (MICs) of selected strains

Clinical isolates (see section 2.2.1 and Appendix 2, Table 2.1) were selected from the available database according to their INH-resistance status and *nat* G619A mutation status. Four INH resistant isolates (isolate 131 from family 2, isolate 404 from family 28, isolate 766 from family 18 and isolate 816 from family 133), four isolates from family 3 (isolates 929, 1394, 1398 and 1430), six isolates from family 28 (isolates 83, 404, 526, 704, 1077 and 1595), one INH susceptible isolate from family 11(isolate 208) and the reference strain H37Rv, were selected for this study. Isolates from family 3 and family 28 contain the *nat* G619A SNP. Later in this study, it was determined that family 3 had an additional SNP, which resulted in a T529C change (see section 3.3.2). These isolates were grown on BACTEC medium and the Minimum Inhibitory Concentrations (MICs) of INH were determined using concentrations of 0.005, 0.0125, 0.025, 0.05, 0.1, 0.5, 1, 5 and 10μg/ml INH, as described in the BACTEC manual. For more information on the isolates and strain families used in this study please see section 2.2.1 and Appendix 4.

3.2.7. Phylogenetic relationship of isolates in family 3, family 28 and unique isolate 1936

Two IS*6110* insertion loci (ISO480.8 and ISO480.11), cloned from South African clinical isolates, appeared to be shared by isolates from strain family 3 and strain family 28 (see Figure 3.6). The IS*6110* insertion loci had been previously characterised (Sampson *et al.* 1999).

Primers were designed to amplify the insert as well as approximately 20 bases downstream the 3′ end of the insertion sequence.

Table 3.2: Primers used to amplify the insertion loci, ISLO480.8 and ISLO480.11.

Locus name	Forward Primer (Xho1)	Reverse Primer	Tm	Size (bp)
ISLO480.8	5' TTCAACCATCG CCGCCTCTAC 3'	5' GCGAAAGCCCCGGCC GCGGCTGGATGAAC 3'	62°C	± 100
ISLO480.11	5' TTCAACCATCG CCGCCTCTAC 3'	5' TCACTTGAGGGGGTGT TGAGGTTGTGAAC 3'	62°C	± 100

Forward primer was supplied by R. Warren.

Primers were used to amplify the insertion sites ISO480.8 and ISO480.11 using the same PCR conditions as explained in section 3.2.1.

3.2.8. Percentage of resistant isolates in each strain family

Using information available from the database of isolates comprising the 37 strain families (see section 2.2.1), the percentage of INH resistant isolates within each family was calculated. All isolates making up the database had been tested for drug sensitivity and this information was made available for this study.

3.3. RESULTS

3.3.1. The previously identified G619A SNP in *M. tuberculosis nat* is restricted to two specific strain families

Using PCR-RFLP analysis, 468 isolates representing 37 strain families and 10 unique isolates were screened for the *nat* G619A SNP. These isolates all originated from the Western Cape in South Africa. Analysis of these results will indicate whether this SNP is randomly dispersed throughout the strain families, or whether it was a once-off occurrence that is specific to certain strain families.

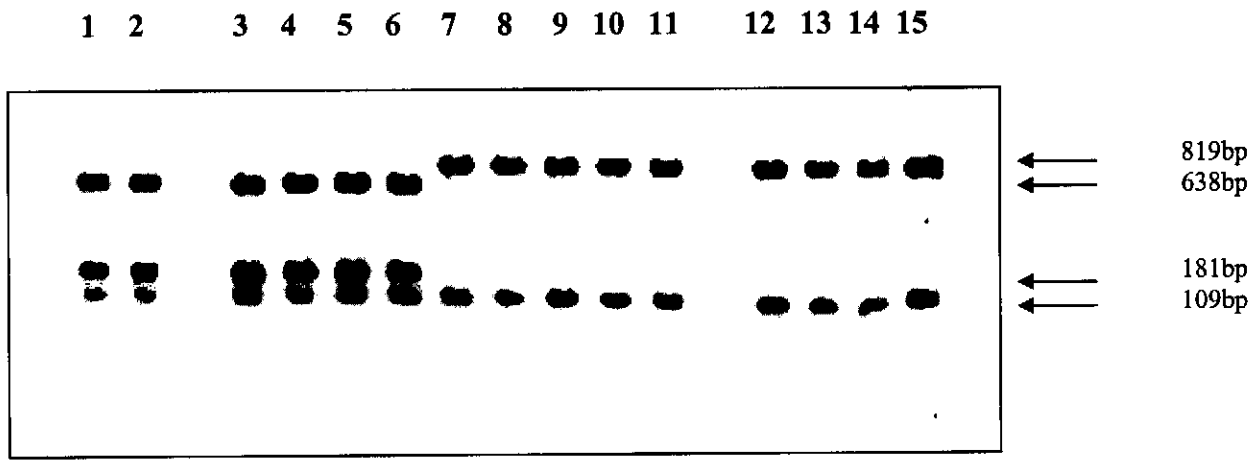


Figure 3.3: PCR-based DNA analysis to identify the *nat* G619A SNP in different strains of *M. tuberculosis.*

PCR-RFLP analysis of isolates representative of three strain families and one unique isolate with the *nat* G619A SNP and others without.

Family 3 (lane 1 – isolate 594; lane 2 – isolate 1430); Family 28 (lane 3 – isolate 44; lane 4 – isolate 136; lane 5 – isolate 509); unique isolate (lane 6 – isolate 1936) and Family 11 (lane 7 – isolate 9, lane 8 – isolate 26; lane 9 – isolate 39; lane 10 – isolate 40; lane 11 – isolate 43; lane 12 – isolate 72; lane 13 – isolate 74; lane 14 – isolate 96; lane 15 – isolate 163). The amplified products were restricted with BsmA1. An extra BsmA1 restriction site (thereby producing an extra band) is present only in samples from strain family 3, 28 and the unique isolate 1936. The number of base pairs (bp) for each fragment is indicated with arrows. Restricted fragments were run with a 1kb DNA ladder (GibcoBRL).

The SNP previously identified introduces an additional restriction site for BsmA1, which enabled numerous isolates to be tested for this SNP using PCR-RFLP. As shown in Figure 3.3, this SNP was found to occur in all isolates of strain family 28 (n = 187), family 3 (n = 6) and in 1 unique isolate (isolate 1936) (Warren *et al.* 2000). No isolates from the other families or unique isolates of *M. tuberculosis* had this SNP.

3.3.2. Sequence Analysis of *M tuberculosis nat*

The resistant isolates 131, 404 and 766 showed mutations in the *katG* gene at codon 315 AGC. For precise nucleotide changes see Appendix 2, Table 2.1. Isolate 816, which has a high INH MIC (>10μg/ml), did not have any of the previously identified mutations in the *katG* gene, however, a SNP in the promoter region of the gene *aphC* (-9 G-A) was identified (see Appendix 2, Table 2.1).

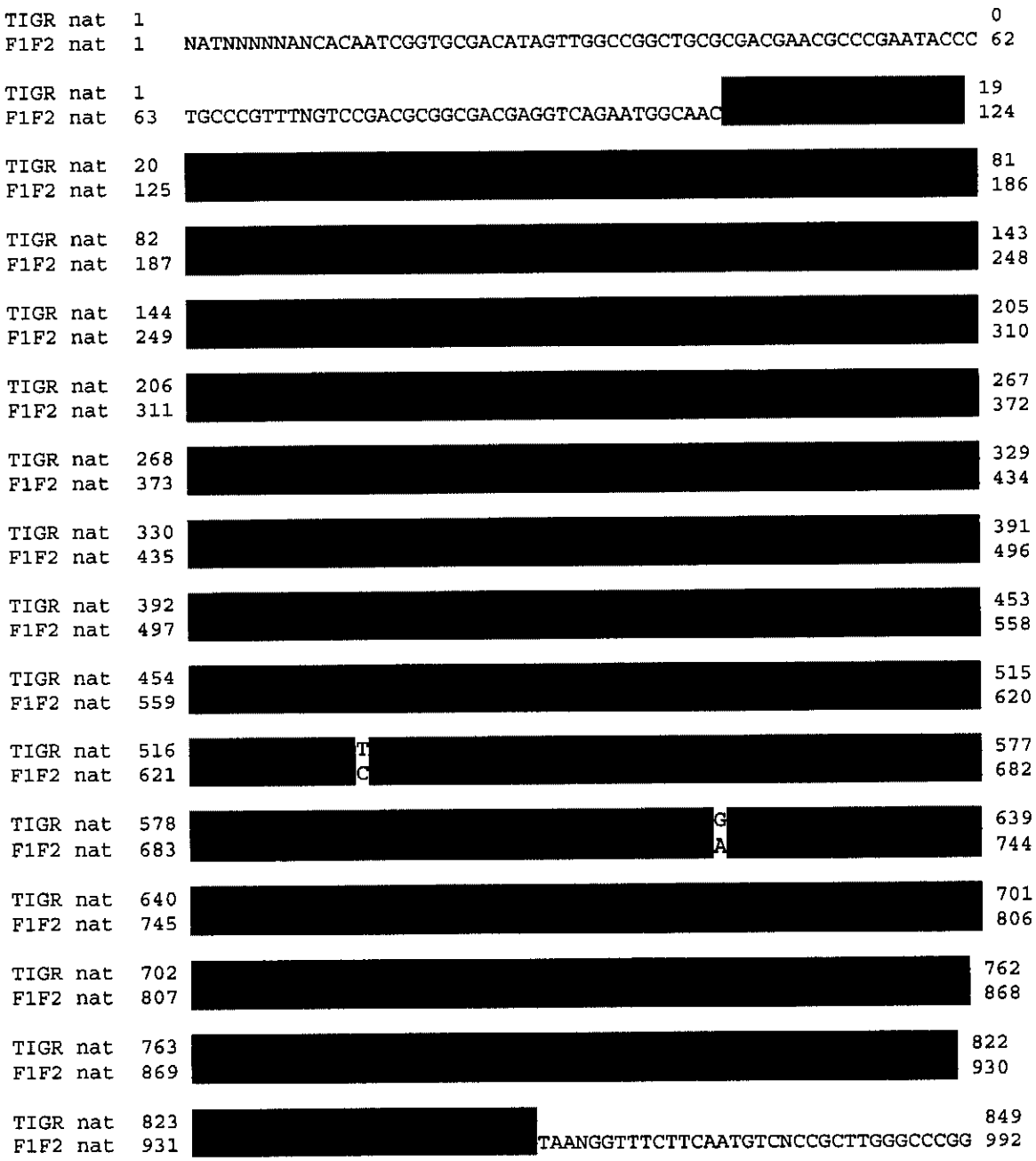


Figure 3.4: Sequence alignment of the ORF of the *nat* gene in isolate 1430, and the TIGR *nat* sequence.

The T529C (Y177H amino acid change) and G619A (G207R amino acid change) SNPs are evident as bases that are not highlighted.

To search for possible nucleotide changes in the immediate upstream, downstream or ORF regions of the *M. tuberculosis nat* gene, five clinical isolates were selected for extensive sequence analysis of these regions (Table 3.1 and Figure 3.2). These isolates were selected

according to their INH susceptibility status and mutation status (see Appendix 2, Table 2.1). Two INH susceptible isolates (H37Rv and isolate 208), two INH resistant isolates (isolates 131 and 816) and one isolate from family 3 (isolate 1430 containing G619A SNP) were selected for analysis. No polymorphisms were identified within 791bp upstream of the 5' end or within 1122bp downstream of the 3'end of the *nat* gene. However, a T529C SNP was identified in the *nat* ORF of isolate 1430 of family 3 (see Figure 3.4). This SNP resulted in an Y177H amino acid change and could be detected using the restriction enzyme Bsg1. This enzyme restricted the *nat* gene only when this T529C SNP was present, producing two DNA fragments of 546bp and 382bp (see Figure 3.5).

3.3.3. The newly identified T529C SNP in *M. tuberculosis nat* is restricted to one strain family

The newly identified SNP introduces a restriction site for Bsg1, which enabled numerous isolates to be tested for this SNP using PCR-RFLP. Using this technique the T529C SNP was verified in isolate 1430 from family 3; a total of 250 isolates representing 37 strain families and 10 unique isolates, from the database in the Western Cape of South Africa, were tested for this SNP. As shown in Figure 3.5 this SNP was found to occur in all isolates of Family 3 (n = 6) (Warren *et al.* 2000). No isolates from the other families or unique isolates of *M. tuberculosis* had this SNP.

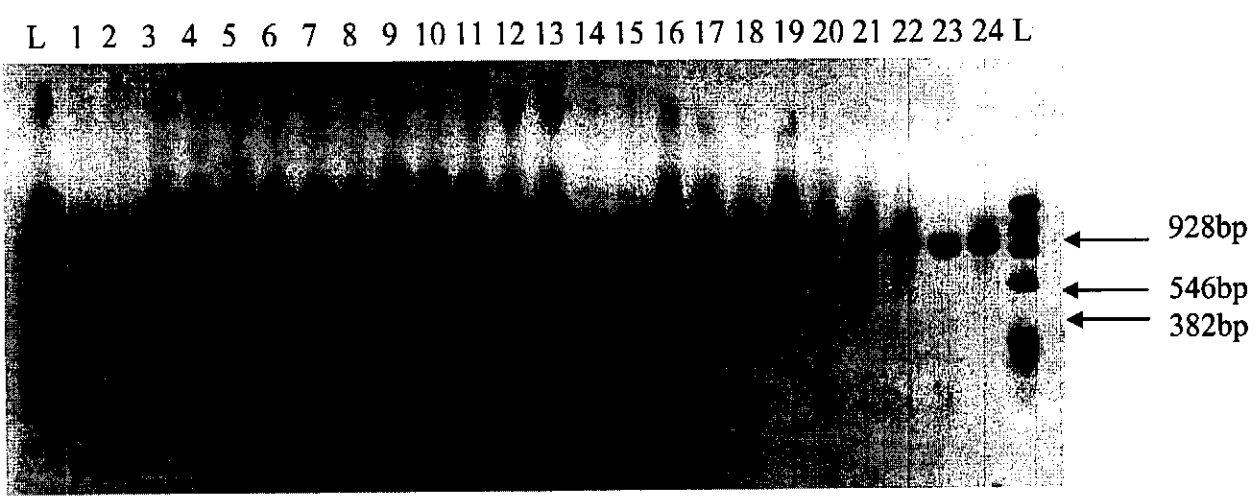


Figure 3.5: PCR-based DNA analysis to identify the *nat* T529C SNP in different isolates of *M. tuberculosis.*

Showing PCR-RFLP analysis of isolates representative of three families and three unique isolates to detect the *nat* T529C SNP. Reference strain H37Rv uncut (1); H37Rv cut (2); Family 3 (lane 3 – isolate

100; lane 4 – isolate 594; lane 5 – isolate 929; lane 6 – isolate 1394; lane 7 – isolate 1398; lane 8 – isolate 1430); Family 28 (lane 9 – isolate 44; lane 10 – isolate 136; lane 11 – isolate 509; lane 12 – isolate 973); unique isolates (lane 13 – isolate 75; lane 14 – isolate 1867; lane 15 – isolate 1936) and Family 11 (lane 16 – isolate 9, lane 17 – isolate 36; lane 18 – isolate 39; lane 19 – isolate 40; lane 20 – isolate 43; lane 21 – isolate 72; lane 22 – isolate 74; lane 23 – isolate 96; lane 24 – isolate 915). The amplified products were restricted with Bsg1. The Bsg1 restriction site is only present in samples from family 3. The number of base pairs (bp) for each fragment is indicated with arrows. Restricted fragments were run with a 1kb DNA ladder (GibcoBRL).

3.3.4. INH Resistance levels in selected strains

Various *M. tuberculosis* isolates, corresponding to the different strain families, INH resistance status and *nat* mutation status, were selected for MIC analysis. The results (See Appendix 2, Table 2.1) show that isolate 816 does not have the *katG* nor the *nat* T529C/or G619A SNPs, but is still resistant to INH at a concentration greater than 10μg/ml (MIC = >10μg/ml). This isolate does, however, have a SNP in the *aphC* promoter region (see Appendix 2, Table 1.2), but it is still possible that mutations in additional locus or loci may be present resulting in this high INH MIC. Analysis of isolates with the *nat* G619A SNP and both the *nat* T529C and G619A SNPs, compared to those without (e.g. reference strain H37Rv), showed an increase in the MIC of approximately 0.02 to 0.05μg/ml of INH (MIC = 0.02-0.05μg/ml). This change in the MIC may not be significant and therefore may suggest that the *nat* SNPs in the structural gene does not play a major role in INH resistance in these strains. No difference in the INH MIC was evident between isolates from strain family 3 and strain family 28, indicating that two SNPs together in one isolate did not affect the INH MIC, when compared to an isolate with just the G619A SNP.

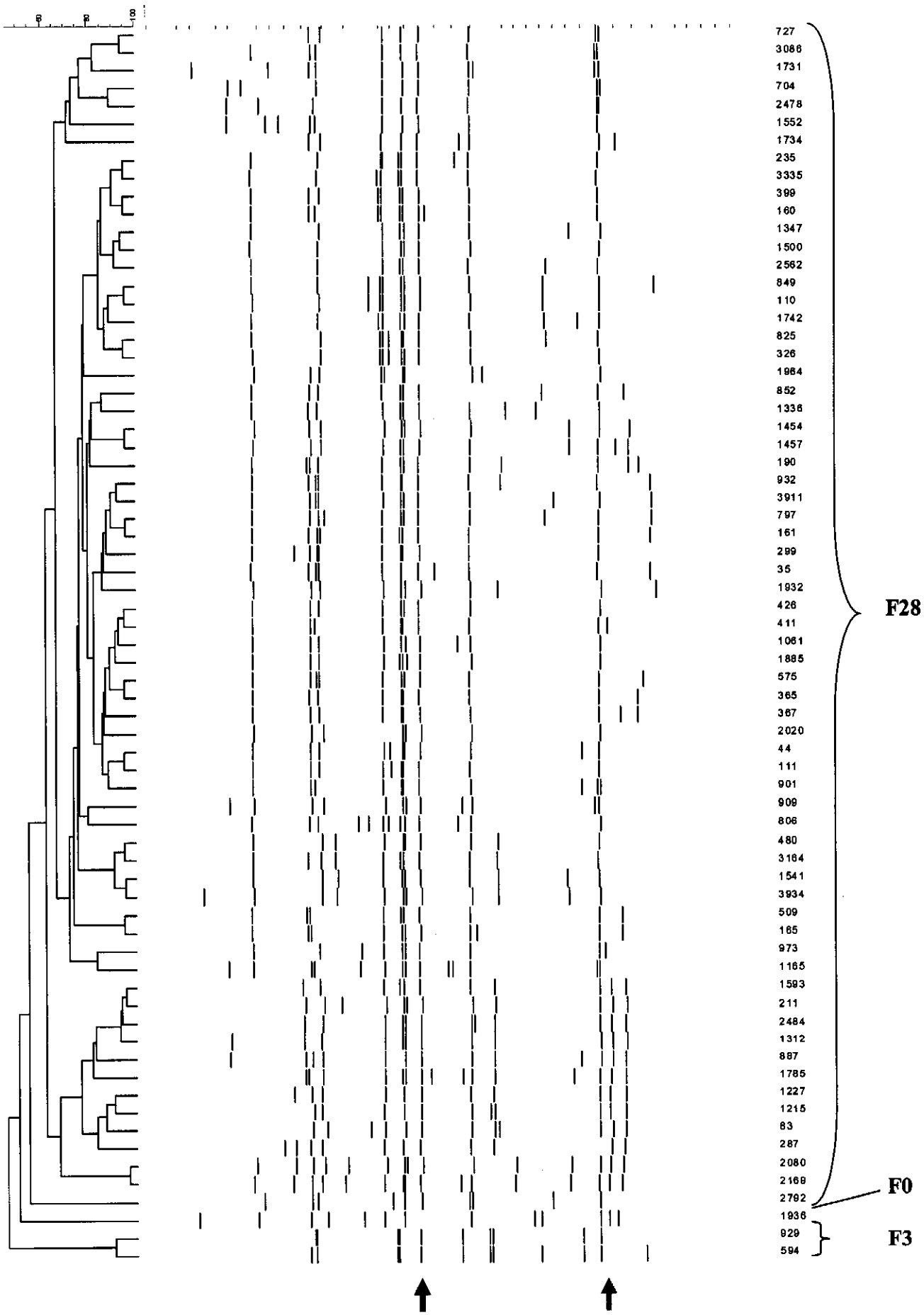


Figure 3.6: A dendogram of isolates representing the strains found in strain family 3 and strain family 28, as well as the unique isolate (F0).

The numbers on the right of the dendogram represent isolate numbers, and the numbers in bold are the strain family to which the isolates belong. ↑ - indicates the insertion site referred to in this study as ISLO480.8 and ↑ - indicates the insertion site referred in this study as to as ISLO480.11. These insertion sites appear to be common between isolates in strain family 3 and isolates in strain family 28.

3.3.5. Phylogenetic relationship of strains in family 3, family 28 and unique isolate 1936

From Figure 3.6, isolates from strain family 3 and strain family 28 seem to share two common insertion loci, indicated by arrows. To confirm the existence of these shared loci, the loci were amplified using the PCR technique.

L - 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

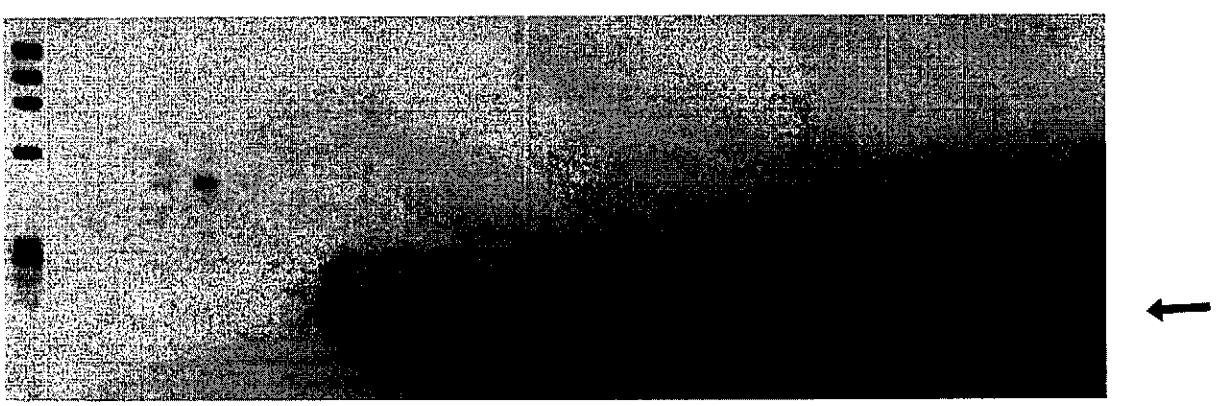


Figure 3.7: PCR-based amplified products of an insertion at locus ISLO480.8

Lane L - 1kb DNA ladder (GibcoBRL); '-' – negative control; Lane 1 – unique isolate 1936; Family 3 (lane 2 – isolate 100; lane 3 – isolate 594; lane 4 – isolate 1398; lane 5 – isolate 1430); Family 28 (lane 6 – isolate 509; lane 7 – isolate 524; lane 8 – isolate 813; lane 9 – isolate 849; lane 10 – isolate 887; lane 11 – isolate 973; lane 12 – isolate 1049; lane 13 – isolate 1077; lane 14 – isolate 1188; lane 15 – isolate 1202; lane 16 – isolate 1215, lane 17 – isolate 1225; lane 18 – isolate 1602; lane 19 – isolate 1773; lane 20 – isolate 1833; lane 21 – isolate 1860; lane 22 – isolate 1868; lane 23 – isolate 2020). Arrow indicates amplified product.

Figure 3.7 shows that isolates from strain family 3 did not yield a DNA product when using primers specific for the ISLO480.8 locus. These results imply that isolates from strain family 28 do have the ISLO480.8 locus, whereas isolates from strain family 3 do not. This particular insert is therefore not shared by these two strain families, or alternatively minor difference make amplification impossible with these primers.

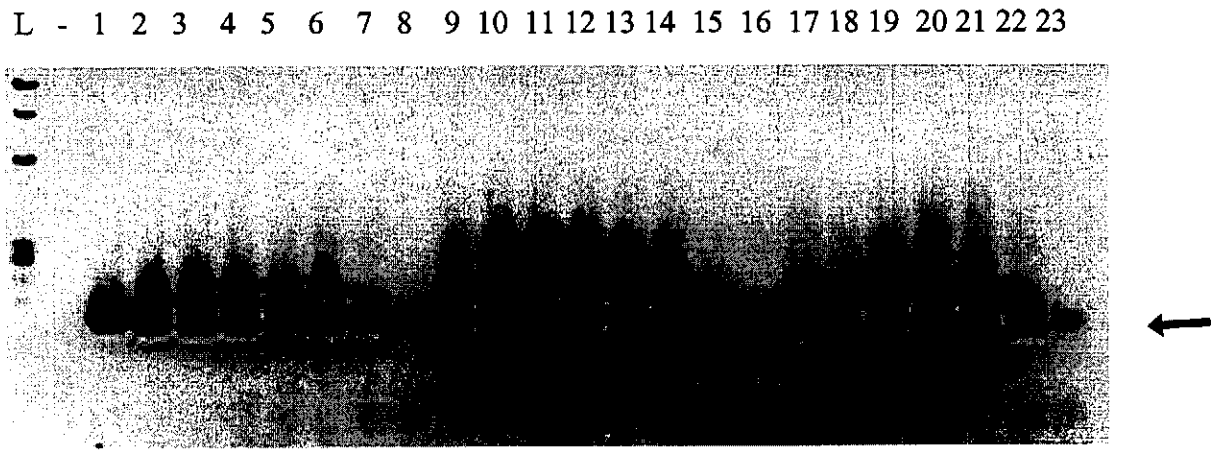


Figure 3.8: PCR-based amplified products of an insertion at locus ISLO480.11.

Lane L - 1kb DNA ladder (GibcoBRL); '-' – negative control; Lane 1 – unique isolate 1936; Family 3 (lane 2 – isolate 100; lane 3 – isolate 594; lane 4 – isolate 1398; lane 5 – isolate 1430); Family 28 (lane 6 – isolate 509; lane 7 – isolate 524; lane 8 – isolate 813; lane 9 – isolate 849; lane 10 – isolate 887; lane 11 – isolate 973; lane 12 – isolate 1049; lane 13 – isolate 1077; lane 14 – isolate 1188; lane 15 – isolate 1202; lane 16 – isolate 1215, lane 17 – isolate 1225; lane 18 – isolate 1602; lane 19 – isolate 1773; lane 20 – isolate 1833; lane 21 – isolate 1860; lane 22 – isolate 1868; lane 23 – isolate 2020). Arrow indicates amplified product.

Isolates from both strain family 3 and strain family 28, as well as the unique isolate 1936, produced an amplified product when using primers specific for the insertion locus ISLO480.11 (Figure 3.8). This implies that locus ISLO480.11 is shared by isolate 1936, and isolates in strain family 3 and strain family 28. This may indicate that there is a distant relationship between these isolates.

3.3.6. Percentage of resistant strains in each strain family

The isolates making up the 37 strain families have been previously tested for INH resistance and this information was available for this study. Using this information, the percentage of INH resistant isolates within the strain families was calculated.

Table 3.3: The percentage of INH resistant isolates within each strain family.

Strain family	F	T	Grand Total	% INH resis
1	13	0	13	0
2	82	16	98	16
3	6	0	6	0
4	67	0	67	0
5	1	2	3	67
6	27	0	27	0
7	31	3	34	9
8	6	0	6	0
9	43	1	44	2
10	8	0	8	0
11	524	57	581	10
12	7	0	7	0
13	37	1	38	3
14	71	8	79	10
15	24	0	24	0
16	33	0	33	0
17	17	0	17	0
18	57	0	57	0
19	7	6	13	46
20	12	6	18	33
21	39	4	43	9
22	5	0	5	0
23	18	27	45	60
24	35	0	35	0
25	33	0	33	0
26	17	12	29	41
27	5	0	5	0
28	187	45	232	19
29	190	436	626	70
30	6	0	6	0
31	2	0	2	0
32	3	0	3	0
Grand Total	1612	624	2236	28

INH susceptible (F) and INH resistant (T) isolates, within each strain family, is depicted in this Table. Accurate data could not be obtained for the remaining five strain families, and these results were omitted. A strain is regarded as INH resistant when it has an INH MIC value greater than 0.1ug/ml (see section 3.2.6).

Strain family 3 has no INH resistant isolates and only 19% of isolates from strain family 28 are INH resistant. This percentage of INH resistance in strain family 28 is fairly low when compared to a number of other strain families, for example 70% of the isolates in strain family 29 are INH resistant (Table 3.3).

3.4. Discussion

DNA fingerprinting has been used to classify epidemiologically related cases of TB. The transposable element IS*6110* is most frequently and widely used. The fingerprint obtained using this transposable element varies in copy number and location in the gene. It is generally considered stable enough to be used as a marker of epidemiologic links among TB cases (Murray and Alland 2002). The *IS6110* is a good marker for identifying different isolates within an infected area, as it detects recent changes in evolving markers (Warren *et al.* 2002). However, recent findings have found that there is no simple relationship between IS*6110* copy number and phylogenetic lineage (Gutacker *et al.* 2002). Hence, IS*6110* copy number alone is not phylogenetically informative (Gutacker *et al.* 2002). To determine linkage between strains, slow evolving markers should be looked at (Warren *et al.* 2002), these include Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR) analysis (Supply *et al.* 1997), Polymorphic GC-Rich Sequence (PGRS) regions (Chaves et al. 1996;Warren et al. 1996) and certain SNPs (Gutacker *et al.* 2002).

Polymorphic genes are good candidates for virulence and immune determinants (Fleischmann *et al.* 2002). Proteins that interact directly with the host are more polymorphic, and can be used as markers for phylogenetic and evolutionary studies (Fleischmann *et al.* 2002). These studies have been hindered by the scarcity of known genetic markers (Fleischmann *et al.* 2002). There are two classes of single nucleotide substitutions, synonymous and nonsynonymous SNPs. Nonsynonymous SNPs (nsSNPS) result in an amino acid change within the protein and could result in evolutionary selection, while synonymous SNPs (sSNPS) do not result in an amino acid change within the protein and are most probably evolutionary neutral (Gutacker *et al.* 2002). The sSNPS are mostly functionally neutral and can therefore be used as targets for large-scale molecular population genetic studies, examining evolutionary relationships among bacterial strains (Gutacker *et al.* 2002). However, recently two nsSNPs located in codon 493 of the *katG*

gene and codon 95 of the *gyrA* gene have prompted the classification of *M. tuberculosis* strains into three principal groups (group1, group 2 and group3) (Gutacker *et al.* 2002).

A G619A SNP of the *nat* ORF was identified in *M. tuberculosis* clinical isolates resulting in a Gly to Arg change at amino acid 207 (Upton et al. 2001). This substitution results in a lower structural stability or solubility and a 10 times lower affinity of the enzyme for INH. The modelled structures of *M. smegmatis* 207G NAT and 207R NAT suggest that this mutation causes the substrate-binding Phe^{204} residue to be moved, and hence alters the isoniazid acetylation activity (Kawamura *et al.* 2003).

The *M. tuberculosis nat* gene and flanking regions were screened for other mutations, and a T529C SNP was found. This SNP resulted in a Tyr to His change at amino acid 177, and created a Bsg1 restriction site. Fifteen NAT homologues show tyrosine at amino acid 177 (see section 1.4.2), this amino acid at this location is highly conserved. This is the change of a tyrosine residue with an aromatic ring and a hydroxyl group for a histidine that has a larger aromatic ring structure but also has a potential H-bonding group in the nitrogen within the heterocyclic aromatic ring. It is very possible that this amino acid change could result in a conformational change of the NAT protein and affect its function within the bacterium. Representative clinical isolates of 37 strain families and 10 unique isolates were tested for both SNPs. It was found that the G619A SNP was restricted to all isolates of strain family 3, strain family 28 and to one unique isolate; and that the T529C SNP was restricted only to all isolates in strain family 3. The G619A SNP constitutes 20% and the T529C SNP constitutes 2.4% of the clinical isolates tested. These nonsynonymous SNPs (Gutacker *et al.* 2002) may be used instead of IS*6110* profiling to identify these specific groups of clinical isolates i.e. they may be used as a quick method to identify these strains in new cases of TB.

The SNPs do not seem to be frequent random events, and it can be assumed that the strains containing these SNPs were derived from a common progenitor. SNPs are rare in the slow-growing mycobacteria so often will have arisen through evolutionary pressure such as isoniazid therapy (Sreevatsan et al. 1997). However, strains from family 3 and family 28 do not appear to be closely related according to their IS*6110* DNA fingerprints (see Figure 3.6) or spoligotyping (personal communication T. Victor). In the family tree presented in Appendix 4, Figure 4.1, these two families are placed far apart from each other, and strains from family 3 lack spacers 13 to 16, whereas, strains from family 28 lack spacers 9 to 10 in the Direct Repeat (DR) regions,

according to spoligotyping (personal communication T. Victor). This is surprising, as the G619A SNP only occurs in these two strain families (family 3 and 28), one of which has an additional T529C SNP, and it could be assumed that strain family 3 diverged from strain family 28 when the second *nat* T529C SNP occurred. However, the unique isolate 1936 and isolates from both strain family 3 and strain family 28 are classified as group 2 *M. tuberculosis* isolates, according to the *katG* and *gyrA* polymorphisms (Sreevatsan et al. 1997). These isolates also share the insertion loci ISLO480.11 and they all have the G619A SNP. On these grounds, it is likely that these isolates did evolve from a common progenitor, and that strain family 3 evolved from an isolate in which a second SNP occurred in the *nat* gene. The unique isolate 1936 only has the G619A SNP and may be a divergent isolate from strain family 28.

The identification of the new SNP, and the restriction of the SNPs to certain strain families suggests that using SNPs, instead of the IS*6110* DNA fingerprints, may be a more precise method of strain classification by family. These SNPs are only restricted to two strain families and would not be good candidates for evolutionary studies. However, an outbreak of TB cases, infected by isolates from either of these *M. tuberculosis* strain families, could potentially be identified using these SNPs, as they can be a rapid method to identify the infecting isolate. This method of PCR-RFLP is rapid, as many samples can be screened after two simple reactions, PCR and restriction using specific restriction enzymes.

Drug resistant *M. tuberculosis* strains are becoming more and more prevalent especially in cases where patients are HIV positive (Grange and Zumla 2002;Mukherjee et al. 2004). Isoniazid (INH) is one of the main drugs used in TB treatment (Torres *et al.* 2000). Once INH resistance occurs within a *M. tuberculosis* isolate, resistance to other anti-TB drugs are more likely to occur, rendering the isolate multidrug resistant (Piatek et al. 2000;Heep et al. 2001;Grange and Zumla 2002). It has been shown that a variety of SNPs in multiple genes involved in mycolic acid biosynthesis, are overexpressed as a response to buildup or cellular toxicity of INH, are found exclusively in INH resistant clinical isolates (Ramaswamy *et al.* 2003). Resistance to isoniazid is usually the result of SNPs in key genes i.e. *katG, inhA, kasA, ahpC* and *ndh* (Ramaswamy *et al.* 2003). Catalase-peroxidase (KatG) is the enzyme that activates INH (Slayden and Barry 2000). The *katG* gene encodes an 80kDa hemoprotein of 744 amino acids that has both catalase and peroxidase activity (Slayden and Barry 2000). It has been proposed that a multiple activation pathways occur, with each product having different cellular targets (Slayden and Barry 2000). It has been shown using whole cells of *M. tuberculosis* that INH is

metabolised to several products, including 4-pyridylmethanol and isonicotinic acid (Slayden and Barry 2000). One metabolite produced *in vitro* is the covalent adduct of the acyl pyridine formed from INH and the nicotinamide ring system. This adduct, proposed to be formed by the addition of an isonicotinyl radical to NAD^+, is thought to inhibit one of the cellular targets associated with mycolic acid synthesis, InhA (Slayden and Barry 2000). Enoyl-acyl carrier protein (ACP) reductase (InhA) results in INH resistance in *M. smegmatis* when overexpressed (Slayden and Barry 2000). It is thought that InhA is a component enzyme of the fatty acid synthase system (FAS II) by which mycolic acids are produced (Slayden and Barry 2000). A naturally occurring amino acid mutation (S94A), identified in both *M. smegmatis* and *M. tuberculosis,* has a reduced Km for NADH. This implies that the mutation results in a decreased affinity for binding of this adduct (Slayden and Barry 2000). Mutations in the promoter region of *inhA* resulting in overexpression of InhA have also been shown to be associated with low levels of INH resistance (Slayden and Barry 2000). The protein complex consisting of the acyl carrier protein (AcpM) and a ketoacyl synthase (KasA) has also been shown to interact with activated INH. Using mass spectrometry and incorporation of radioactive INH, this protein complex was shown to contain INH (Slayden and Barry 2000). When growing cultures of *M. tuberculosis* are exposed to INH saturated hexacosanoic acid accumulates with lesser amounts of smaller fatty acids such as palmitate, where untreated cells showed a broader range of fatty acids with abundant species having 18, 28 and 50+ carbons (Slayden and Barry 2000). This accumulation of hexacosanyl-AcpM may be the metabolic consequence of KasA inhibition by activated INH (Slayden and Barry 2000). Exposure of cells to INH also results in both AcpM and KasA to be strongly upregulated (much more so than InhA), suggesting that expression of these genes is affected by the level of mycolic acid synthesis (Slayden and Barry 2000). Mutations in *kasA* (e.g. D66N, R121K, G269S, G312S, G397D and F413L) have been associated to INH resistance in *M. tuberculosis* isolates, however, not all mutant *kasA* alleles have been strictly associated with enhanced INH resistance (Slayden and Barry 2000).

Akyl hydroperoxide reductase (AhpC) is a protein capable of detoxifying damaging organic peroxides (Slayden and Barry 2000). This protein and KatG are coordinstely regulated in response to oxidative stress. Five different mutations (G(-48)A, G(-51)A, C(-54)T, G(-74)A and C(-81)T) have been identified in the promoter region of *aphC,* which lead to overexpression of AphC, in INH resistant *M. tuberculosis* isolates (Slayden and Barry 2000). These mutations do not frequently occur, but are almost always accompanied by *katG* mutations that result in complete loss of catalase-peroxidase activity. They therefore seem to occur to counteract the

effects of loss of catalase-peroxidase activity (Slayden and Barry 2000). The *ndh* gene encodes an NADH dehydrogenase which controls the oxidative state of mycobacterial cells by regulating the NADH/NAD^+ ratio (Lee *et al.* 2001). Mutations in *ndh* have been reported to be associated with INH resistance (Lee *et al.* 2001), however, the converse has been found (Ramaswamy *et al.* 2003).

In a recent study, 76% of all INH resistant isolates tested had a mutation (most often a SNP) in the *KatG* gene (Ramaswamy *et al.* 2003). The most common alteration identified in INH resistant *M. tuberculosis* isolates is the naturally occurring substitution of serine with threonine at amino acid 315 (Slayden and Barry 2000). In this study it was important to identify how significant the polymorphisms within the *nat* gene were in relation to INH resistance. Thirty-two *M. tuberculosis* strain families, from the Western Cape Province of South Africa, were analysed to determine the prevalence of INH resistant isolates within each strain family (Table 3.3). Included in these results were strain family 3 (T529C and G619A SNPs) and strain family 28 (G619A SNP). Strain family 3 had no INH resistant isolates and 19% of the isolates found in strain family 28, were INH resistant. This percentage was fairly low when compared to other strain families e.g. 70% of isolates in strain family 29 are INH resistant. It has also been shown that identified polymorphisms within the *nat* gene occur in both INH susceptible and INH resistant isolates (Ramaswamy *et al.* 2003). These results suggest that the identified SNPs are not solely linked to INH resistance.

If NAT is important in the growth and development of *M. tuberculosis* as stated in the previous chapter, then mutations within the *nat* gene may interfere with the development of the bacterium. The SNPs are known to cause an amino acid change and the G619A SNP results in a conformational change of the NAT protein (Kawamura *et al.* 2003). These changes may slightly affect cellular processes in which NAT is involved. For example, if NAT is involved in the development of the highly complex cell envelope, then mutations within the NAT protein may result in a change in the complexity of the cell envelope. Changes in the cell envelope may result in the cell envelope being slightly more permeable, allowing more substances to enter the bacterium. An altered cell envelope could result in new drugs being used to treat TB, and may alter the host-bacterium interaction. Strain family 3 has two SNPs both resulting in amino acid changes within the NAT protein. The mutant NAT may result in the growth and development of these strains being hindered, and their cell envelope's complexity been altered. This family only consists of 6 isolates within the database available for this study. This may be an indication that

the majority of people infected with these strains are able to maintain the infection (due change in cell envelope's complexity), or that the spread of infection of these strains is rather low (due to slower growth and development).

M. bovis BCG *nat* knock-out isolates showed the bacterium to be more sensitive to the antibiotics hygromycin and gentamycin, when compared to the wild type strain (Bhakta *et al.* 2004). It is thought that these antibodies do not affect the wild type strain as they cannot penetrate the highly complex cell envelope (Barry *et al.* 1998). This implies that NAT may be involved in the synthesis of the bacterium's cell envelope, and that *M. bovis* BCG isolates without a functional NAT may have a more permeable cell envelope. Most drug targets to treat TB are those that are involved in the synthesis of the cell envelope e.g. drugs such as INH, ethionamide and thiocarlide all interfere with the synthesis of mycolic acids (Slayden and Barry 2000). Mycolic acids are one of the main constituents of the bacterium's cell envelope (Slayden and Barry 2000). If it can be verified that NAT is involved in the development of the cell envelope, then NAT would become an important candidate for drug therapy against TB.

This study has identified a new T529C SNP, which has been shown to be restricted to strain family 3. The previously identified G619A SNP was found to be restricted to two strain families, strain family 3 and strain family 28. These nsSNPs are good molecular markers for the rapid classification of isolates into strain family 3 or strain family 28. The SNPs do not appear to lend to the development of INH drug resistance in clinical isolates. However, the SNPs may result in changes within the *M. tuberculosis* cell wall making it more permeable and allowing new drugs to enter the bacterium and function as anti-TB drugs.

CHAPTER 4

EFFECTS OF INH ON NAT EXPRESSION

The aim of this chapter was to monitor the levels of expression of NAT in growing *M. bovis* BCG and *M. tuberculosis* cultures exposed to isoniazid (INH).

4.1. Introduction

Despite an increasing incidence in isoniazid resistance (cross reference to previous chapter and Table 2.1 in Appendix 2), INH is still a major component of first line antibiotic treatment against tuberculosis (Bass et al. 1994;Basso et al. 1998) as the drug remains effective for non-resistant strains of *M. tuberculosis* (Young 1994;Basso et al. 1998). It is thought that NAT competes with KatG for INH and that acetylation of INH renders it inactive in mycobacteria (Payton et al. 1999;Upton et al. 2001).

This chapter describes an investigation of the effect INH exposure has on the expression of the *nat* gene in *M. tuberculosis* and *M. bovis* BCG cultures. Q-PCR (see Chapter 2) was used to monitor *M. tuberculosis nat* in cultures exposed to INH compared to control cultures not exposed to INH. A similar experiment was done for *M. bovis* BCG; however, the Western blot technique was used to detect changes in expression instead of Q-PCR.

Knowing whether INH affects the expression of the *nat* gene *in vivo*, may lead to further investigation to identify whether there is an interaction between the NAT protein and INH, and how this interaction is affected in INH resistant isolates where no causative mutations in characterised genes, have been found.

4.2. Materials and Methods

4.2.1. SDS-PAGE Analysis and Western Blotting: Reagents and Buffers

Unless stated otherwise all reagents used in this study were from Sigma-Aldrich.

Table 4.1: The solutions used for preparing protein samples and gels for Western Blot analysis

Buffers and Gels	Constituents
Separating Buffer	1.5M Tris-HCl, 0.5% SDS (pH8.8)
Stacking buffer	0.5M Tris-HCl, 0.5% SDS (pH6.8)
Lysate Buffer (LB)	25mM Tris-HCl (pH6.8), 0.3% SDS, 1X concentration of complete, EDTA-free, protease inhibitor cocktail (Roche) and 2mM DTT
Fractionation Buffer (FB)	25mM Tris-HCl (pH6.8), 10mM EDTA, 1X concentration of complete, EDTA-free, protease inhibitor cocktail (Roche) and 2mM DTT
12% Separating gel (1 gel)	2780µl 40% acrylamide (BIORAD), 2310µl separating buffer, 36670µl ddH_2O, 500µl 1% ammonium persulfate and 20µl TEMED
Stacking gel (1 or 2 gels)	Consists of 750µl 40% acrylamide (BIORAD), 2310µl stacking buffer, 2850µl ddH_2O, 150µl 1% ammonium persulfate and 20µl TEMED
Protein loading Buffer (Urea based)	200mM Tris-HCL (pH6.8), 8M urea and 2% SDS. Just before use, DTT was added to a final concentration of 6mg/ml
Protein dye	2.5mg Coomassie blue, 2.5ml glycerol and 5ml Stacking gel buffer
Transfer buffer	43.2g Glycine, 9g Tris, 600ml Methanol and made up to a final volume of 3 litres with ddH_2O
10X TBS	90g NaCl, 60g Tris and 850ml ddH_2O, pH 7.9 and a final volume of 1 litre was made up using ddH_2O

The constituents making up the buffers and gels used in Western Blot analysis are listed in this Table.

The solutions and gels used in Western blot analysis are in Table 4.1. Proteins were separated on a 12% (w/v) polyacrylamide gel containing 0.5% SDS at pH 8.8. Proteins were either stained with Coomassie blue or transferred to nitrocellulose by electrophoresis prior to Western blotting.

4.2.1.1. Antibodies

The antibodies used in the Western blots were kindly donated by Prof. Sim, Department of Pharmacology, Oxford, U.K. M. Payton produced the antibodies in Prof. Sim's laboratory. Antisera were raised in rabbits and generated against purified recombinant NAT proteins isolated in SDS-PAGE gel slices representing *M. smegmatis* NAT (antibody #155) and *M. tuberculosis* NAT (antibody #84) (Payton et al. 2001b). Bound rabbit immunoglobulin was detected by

chemiluminescence using mouse anti-rabbit IgG mAbs (used at a dilution of 1:10 000) conjugated to horsesradish peroxidase (referred to here on as secondary antidody) (Sigma).

4.2.2. Preparation of Lysate

The titre (see section 2.2.5.) of the cells was taken (A_{600} of 1 is equivalent to 1 X 10^9 cells/ml). Approximately 1.1 X 10^{10} bacterial cells were centrifuged (benchtop centrifuge – Jouan CR412) at 3,100rpm at 4°C for 15 min and the supernatant was discarded. The pellet was washed three times by resuspending the pellet in 10ml PBS pH 7.4 (containing 1% Tween 80) and centrifuging (benchtop – Jouan CR412) at 3,100rpm at 4°C for 15 min. Prior to ribolysing, approximately 400µl of glass beads (size 1mm) were added to Apex tubes and wet thoroughly with 1ml LB buffer (Table 4.1). Apex tubes were centrifuged (benchtop centrifuge – Biofuge *pico*) at 13,000 rpm for 2min and LB buffer was removed and discarded. After the third wash, the pellet was resuspended in 500µl LB and ribolysed twice at speed 6 for 45sec. The cells were centrifuged (benchtop centrifuge – Biofuge *pico*) for 2 min at 13,000rpm and the supernatant was removed and placed in a 1.5ml Eppendorf tube that was kept on ice. A fresh aliquot of 500µl LB was added to the pellet and the ribolysation step was repeated. The cells were again centrifuged (benchtop centrifuge – Biofuge *pico*) at 13,000rpm for 2 min and the supernatant was added to the previous supernatant on ice. The lysate sample was stored at -70°C.

4.2.3. Cell Fractionation into cytosol, cell membrane and cell wall proteins

The absorbance (see section 2.2.4) of the culture at 600nm (A_{600}) was taken. The number of cells used for fractionation remained relatively constant. Mycobacteria were grown in 7H9 medium supplemented with 10% ADC (see section 2.2.2.1). Prior to ribolysing, approximately 400µl of glass beads (size 1mm) were added to Apex tubes and wet thoroughly with 1ml FB buffer (Table 4.1). Apex tubes were centrifuged (benchtop centrifuge – Biofuge *pico*) at 13,000 rpm for 2min and FB buffer was removed and discarded. The absorbance of the growing culture was determined at 600nm and approximately 7 X 10^{10} cells were centrifuged at 3,100rpm (benchtop centrifuge – Jouan CR412) for 15min at 4°C. The pellet was washed in 10ml PBS with Tween 80 to a final concentration of 1%, and then centrifuged (benchtop centrifuge – Jouan CR412) at 3,100 rpm for 15 min (centrifuge kept at 4°C). The wash was repeated three times. From this stage, cells were always kept on ice. After the final wash the pellet was resuspended in 1.5ml FB. The cells were ribolysed twice at speed 6 for 45sec. After the second ribolysation

the cells were centrifuged at 15,000rpm (benchtop centrifuge – Biofuge *pico*) for approximately 2 minutes. The supernatant was removed and placed into 50ml tubes and kept on ice. Fresh FB buffer (1.5ml) was added to the pellet and the cells were ribolysed as before. The cells were centrifuged as before, the supernatant was added to the previous supernatant in the 50ml tube and the pellet was discarded. The 50ml tube containing the supernatant was centrifuged for one hour at 15,000g at 4°C. The supernatant was removed and placed into two 1.5ml ultracentrifuge tubes and centrifuged for 1 hour at 100,000g at 4°C. The pellet was resuspended in 300µl LB (Table 4.1) and labelled 'cell wall fraction'. After ultra-centrifugation, the supernatant was removed and placed into two 1.5ml Eppendorf tubes and labelled 'cytosol' fraction. The final pellet was resuspended in 300µl LB and labelled 'cell membrane fraction'. All protein fractions were stored at -70°C.

4.2.4. SDS-PAGE gels

The gel apparatus was assembled, the underside was sealed with 1% agar, then the 12% separating gel (Table 4.1) was poured and finally the stacking gel (Table 4.1) was poured on top of the solidified separating gel. The loading wells were made as large as possible, as the final prepared loading sample was approximately 50µl. To 25µl of purified protein, 25µl of loading buffer (DTT must be added fresh to a concentration of 6mg/ml) (Table 4.1) was added; the sample was mixed well and then denatured at 95°C for 10min. For each protein fraction, the same amount of protein per cell was loaded. To the denatured protein sample, loading dye (Table 4.1) and iodoacetamide (83mg/ml) was added to a final volume of $1/10^{th}$ of the total volume of the protein sample and the mixture was heated for a further 5 min at 95°C. Along with the protein samples, a rainbow marker (10µl) (Amersham) and an aliquot of purified *M. tuberculosis* NAT protein (recombinant NAT in *E. coli* lysate) was run. The gel was run at approximately 30mA until the blue dye reached the bottom of the gel.

4.2.5. Western Blot

The protein from the gel was transferred to cellulose membrane (Amersham) overnight in transfer buffer (Table 4.1) at 250mA (transfer occurred at 4°C). After the transfer the membrane was washed three times for 5 min in 1X TBS buffer (for 10X TBS see Table 4.1) with 0.05% Tween 20. The combination of 1X TBS and Tween 20 will be referred to as TBST. After the initial washes the membrane was blocked for 2 hours in 10% skimmed milk (made in TBST).

The membrane was then washed three times for 10 min in 3% skimmed milk (made in TBST). Following the washes, the membrane was incubated in primary antibody (for *M. bovis* BCG *nat* antibody #84 and for *M. smegmatis nat* antibody # 155 was used (Table 4.1)) for $1^1/_2$ hours. The dilutions of the antibodies varied; #84 was used at a 1: 10,000 dilution and # 155 was used at 1: 4,000 dilution. Both antibodies were diluted in 3% skimmed milk (TBST). Following this incubation the membrane was washed three times for 10 min in 3% skimmed milk (TBST) and then incubated in the secondary antibody (dilution 1: 10,000) for 1 hour. The membrane was then washed three times for 10 min in 3% skimmed milk (TBST), twice for 10 min in TBST and twice for 5 min in 1X TBS. After the final wash, the membrane was blotted on tissue paper to remove excess buffer and 2ml ECL reagent (Bioscience) was added onto the protein side of membrane and left for 1min. The membrane was again blotted onto tissue paper, wrapped in Saran wrap, placed in a cassette and in the dark a film was placed over the membrane (facing upwards). Numerous films were exposed for different periods of time. Film was developed in the X-ograph. (compact x2, X-ograph Ltd).

4.2.6. SDS-PAGE and Western Blots – *M. bovis* BCG

M. bovis BCG was cultured from a glycerol stock to an absorbance at 600nm of 1. An aliquot of 2ml was transferred to 200ml of fresh 7H9 medium (see section 2.2.2.2) and the cultures were placed in 500ml bottles at 37°C in a rolling incubator. At an absorbance at 600nm of 1, the culture was split into two aliquots of 100ml, and to the one 100ml culture INH was added to a final concentration of 0.28μg/ml. Both cultures were further cultured in 500ml bottles for 2 hours, after which, 20ml was used to produce lysate proteins (see section 4.2.2) and the remaining 80ml was fractionated into cytosol, cell membrane and cell wall proteins (see section 4.2.3). Each protein sample was diluted 16-fold using two fold serial dilutions and was run on a SDS-PAGE gel (see section 4.2.4) and visualised using the Western blot technique (see sections 4.2.5).

Once these results were available, 300ml *M. bovis* BCG was grown as before in 11 bottles. This culture was divided into three aliquots of 100ml each into 500ml bottles, and to two of these cultures INH was added to a final concentration of 0.28μg/ml. After two and four hours of INH exposure, cytosol proteins were extracted (see section 4.2.3) according to correspondingly labelled bottles. Cytosol proteins were extracted from the control culture, which was not

exposed to INH, after 4 hours. The protein fractions were run on a SDS-PAGE gel (see section 4.2.4) and visualised by the Western blot technique (see sections 4.2.5).

4.2.7. Q-PCR – *M. tuberculosis*

Two *M. tuberculosis* isolates (isolate 1430 from family 3 and isolate 816 from family 133, see Appendix 2, Table 2.1) and the H37Rv reference strain were selected. These strains were selected because they represented a reference strain, an isolate with the identified SNPs (isolate 1430) and an isolate resistant to INH (isolate 816) with a INH MIC >10μg/ml. The resistant isolate had no mutations characterised, which were known to be associated to INH resistance, in the following genes: *katG, inhA, kasA* and *ndh*. However, a SNP was identified in the promoter region of the *aphC* gene. The isolates were cultured in 20 ml 7H9 medium (see section 2.2.2.2), with stirring, to an absorbance at 600nm of 0.2 to 0.4. The cultures were separated into two aliquots of 10ml and INH to a concentration of 0.01μg/ml (10X lower than the critical concentration used for susceptibility testing in BACTEC) was added to one aliquot of each of the cultures, the remaining cultures were used as corresponding controls. The cultures were placed into 50ml falcon tubes and cultured for a further 20 hours. RNA was extracted (see sections 2.2.7) after 5 minutes, 30 minutes, 2 hours, 6 hours and 20 hours. The relative amounts of cDNA for *nat* and 16S rRNA were quantified using the standard method of Q-PCR. The *nat* expression levels could be calculated by the direct ratio of *nat* to 16S rRNA. It has previously been shown that INH does not appear to alter the expression of 16S rRNA expression (Alland et al. 1998;Hellyer et al. 1999).

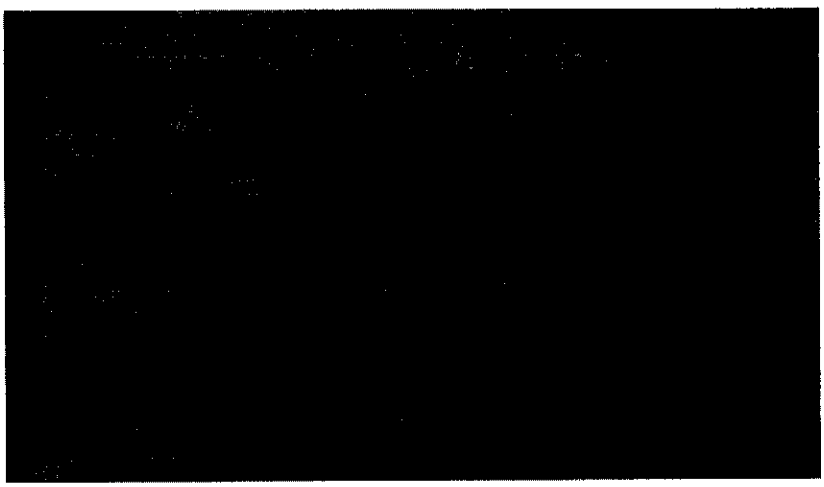
4.3. Results

4.3.1. *M. bovis* BCG

Cultures of *M. bovis* BCG with an absorbance of 1 were divided into two equal volumes. To one of the volumes INH to a final concentration of 0.28μg/m, was added and the remaining culture was maintained as the control culture (not exposed to INH). Serial dilutions of proteins from *M. bovis* cultures, with or without INH exposure, extracted from the cytosol, cell membrane and cell wall fractions were run on a SDS-PAGE gel and analysed on a western blot.

a) Lysate proteins

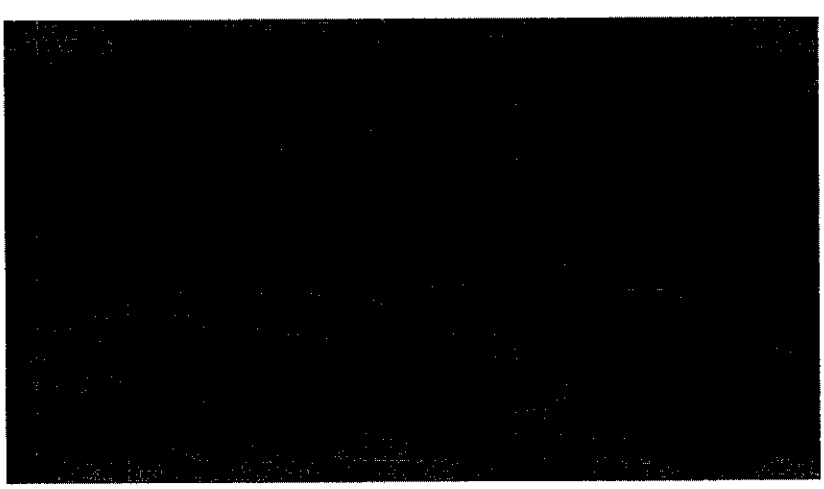
C 1 2 3 4 5 6 7 8



30kDa

b) Cytosol proteins

C 1 2 3 4 5 6 7 8



30kDa

c) Cell Wall proteins

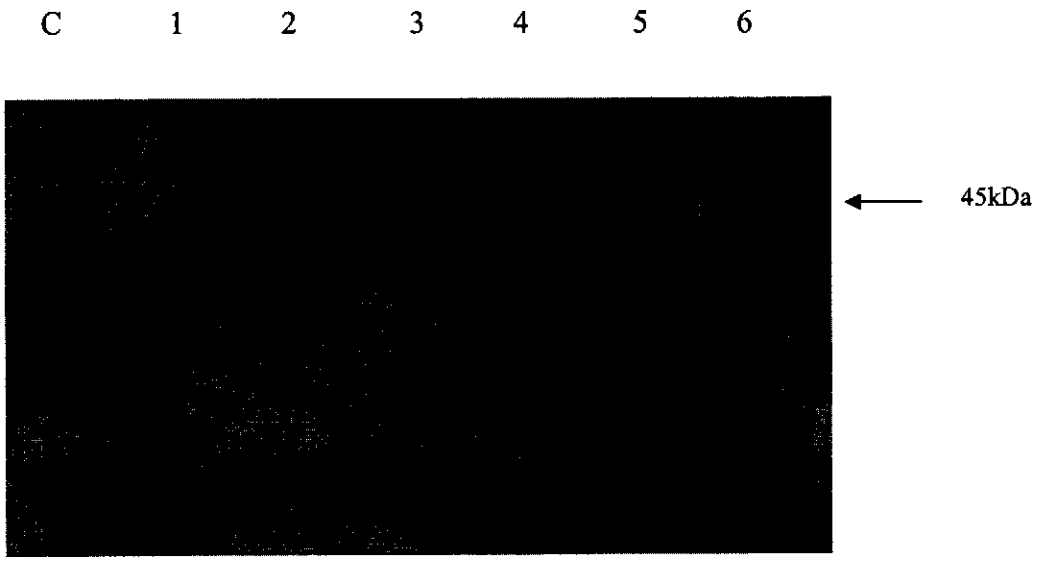


Figure 4.1: Western blot of dilutions of extracted protein fractions of cultured organisms exposed and not exposed to INH at a final concentration of 0.28μg/ml.

1:1 dilutions of extracted protein fractions: lysate protein (a), cytosol proteins (b) and cell wall proteins (c). Lane C – recombinant *M. tuberculosis* (NAT protein in *E. coli* lysate); Lane 1 – undiluted protein from cultures not exposed to INH; Lanes 3, 5, 7 - 1:1 dilutions of sample shown in lane 1; Lane 2 – undiluted protein from cultures exposed to INH and Lanes 4,6,8 – 1:1 dilutions of sample shown in lane 2. Primary antibody #84 and secondary antibody (mouse anti-rabbit IgG mAbs conjugated to horsesradish peroxidase) used at a 1:10,000 dilution. Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by the arrow.

From Figure 4.1 it is evident that after one or two 1:1 dilutions the difference in NAT product between samples exposed to INH and samples not exposed to INH becomes clearer. It is evident that organisms exposed to INH experience a slight increase in NAT production. This was more clearly visible in the cytosol fraction. The concentration of NAT protein in the cell membrane fraction was too low to make any deductions and therefore has not been shown.

The above experiment was repeated and proteins were extracted at two time points (after 2 and 4 hours of INH exposure). Only the cytosol fraction was analysed for an increase in NAT expression.

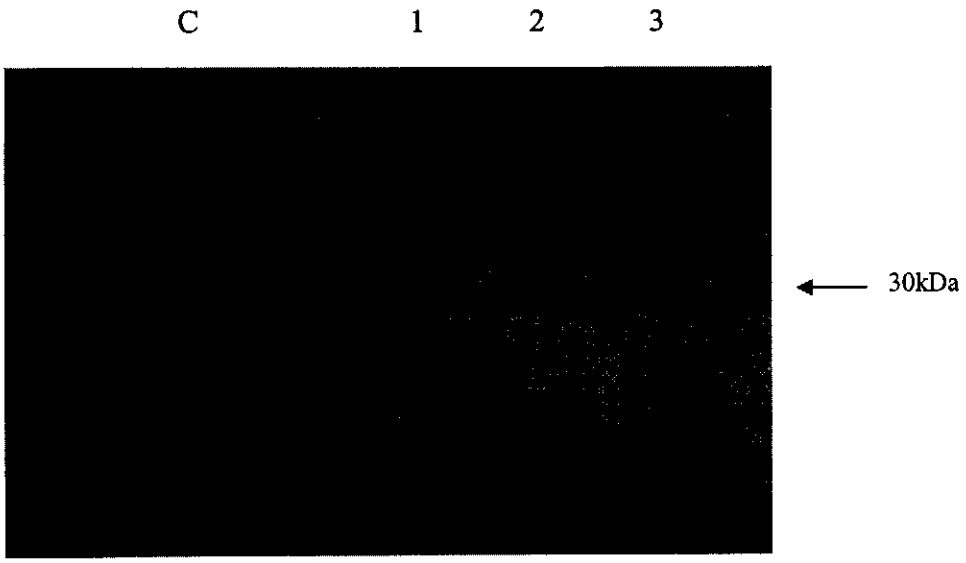


Figure 4.2: Western blot of cytosol proteins of *M. bovis* BCG exposed to INH after 2 and 4 hours.

Lane C – *M. tuberculosis* recombinant NAT protein (over-expressed in *E. coli*), Lane 1 – NAT isolated from a culture after 4 hours (no exposure to INH), Lane 2 – NAT isolated from a culture exposed to INH for 2 hours and Lane 3 – NAT isolated from a culture exposed to INH for 4 hours. Primary antibody #84 and secondary antibody (mouse anti-rabbit IgG mAbs conjugated to horsesradish peroxidase) used at a 1:10,000 dilution. It is evident that there is an increase in NAT in cultures exposed to INH. Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by arrow.

Figure 4.2 shows that NAT production increases in *M. bovis* BCG cells exposed to INH. NAT production was highest in cultures exposed to INH for four hours. The presence of INH in cultures result in an increase of NAT protein expression.

4.3.2. *M. tuberculosis*

M. tuberculosis isolates were cultured to an absorbance at 600nm of 0.3, at this absorbance INH was added to a final concentration of 0.01μg/ml. RNA was extracted from these cells after 5min, 30min, 2hr, 6hr and 20 hr of INH exposure, and Q-PCR was used to calculate the relative *nat* mRNA levels. The presence of INH in growing *M. tuberculosis* cultures results in an increase in *nat* expression; however, the extent of this increase differs between isolates (Figure 4.3). Results were discarded after 6 hours of INH exposure as the results become ambiguous due to cell death, particularly for strain H37Rv. Strain H37Rv, the isolate most sensitive to INH, shows a rapid and extensive increase in *nat* expression up to 2 hours, where isolate 1430, a strain which

contains the *nat* G619A SNP, shows a more gradual increase in *nat* expression up to 2 hours. Isolate 816, an isolate that is highly resistant to INH (MIC > 10 μg/ml), does not show much change in *nat* expression after exposure to INH. After 2 hours of INH exposure strain H37Rv showed a 30-fold increase in *nat* expression, whereas, isolate 1430 showed a 6-fold increase and isolate 816 showed an initial increase and then a slight decrease (10%) in *nat* expression (see Appendix 3, Table 3.1). It can therefore be concluded from these results, that INH exposure causes an increase in *nat* expression, in isolates susceptible to INH, whereas little change is measured in resistant isolates.

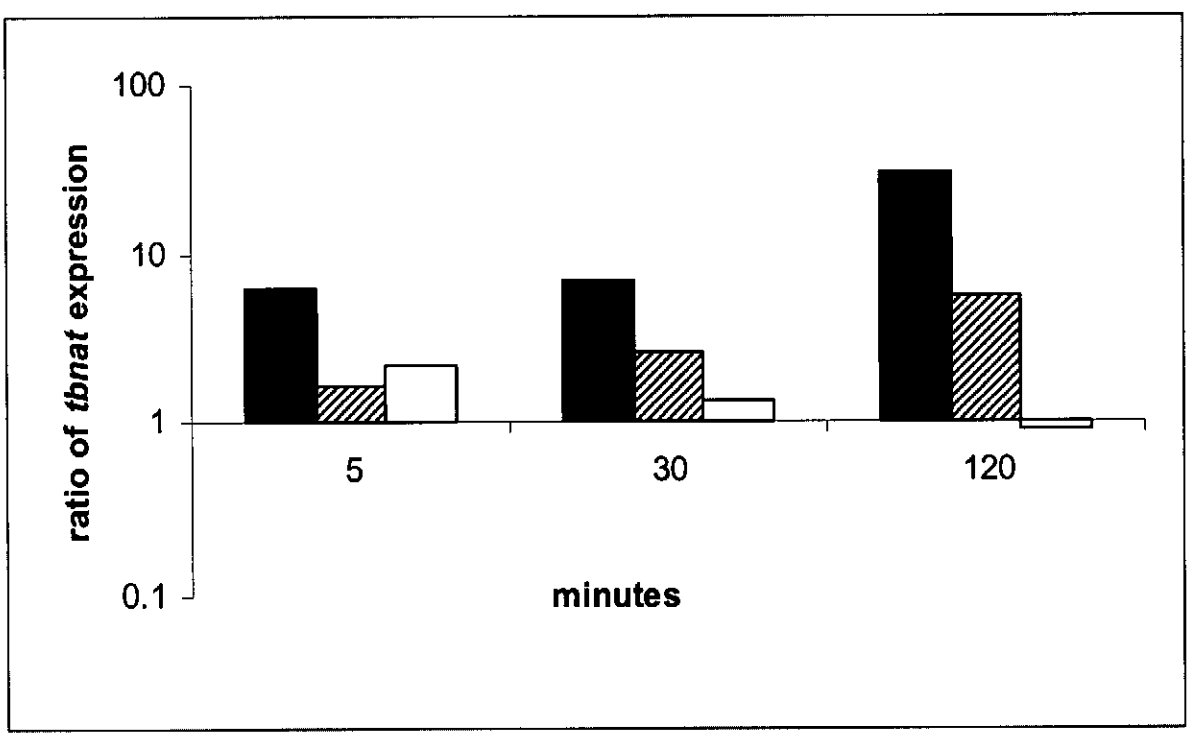


Figure 4.3: Effect of INH on *nat* expression in *M. tuberculosis* strains H37Rv, 1430 and 816. ■ represents the ratios of *nat* expression for strain H37Rv, ▨ represents isolate 1430 and □ represents isolate 816. The different *M. tuberculosis* isolates were cultured to an approximate absorbance of 0.3 at 600nm, at this point the cultures were divided into equal volumes and culturing was continued with or without INH (final concentration 0.01μg/ml) for the times indicated. RNA was extracted from the cultures at each time point and mRNA levels of *nat* were quantified using Real Time PCR with respect to 16S rRNA. The effect of INH on *nat* expression is expressed as the ratio of the amount mRNA in the presence of INH to the amount of mRNA in the absence of INH, i.e. a ratio of 10 indicates 10 times more *nat* mRNA in the presence of INH. See Appendix 3, Table 3.1 for results used to produce this figure.

4.3. Discussion

Activation of INH in *M. tuberculosis* by catalase-peroxidase (encoded for by *kat*G) involves oxidation of the hydrazine moiety, which cannot occur when INH is N-acetylated (Upton et al. 2001). Recombinant NAT from *M. tuberculosis* has been shown to N-acetylate INH *in vitro*, and when the *M. tuberculosis nat* gene is overexpressed in *M. smegmatis,* the resistance to INH of the transformed organism increases three fold (Payton *et al.* 1999). INH is thought to function as a TB drug by its activated form inferring with the fatty acid synthase system (FAS II), essential for mycolic acid synthesis (Slayden and Barry 2000). The cell wall of mycobacteria consists of a superpolymer of three covalently attached subunits. The inner most layer is a typical bacterial peptidoglycan, followed by a complex polysaccharide composed of arabinose and galactose, and finally the covalent exterior of the wall terminates in a lipophilic layer of extremely long-chain fatty acids, the mycolic acids (Slayden and Barry 2000). The inhibition of mycolic acid synthesis by activated INH causes a disruption of the hydrophobic barrier, resulting in a loss of cellular integrity (Slayden and Barry 2000). Mycolic acids are very long chain α-alkyl, β-hydroxy fatty acids containing between 70 – 90 carbons giving them a high molecular weight (Barry et al. 1998;Minnikin 2001), and they are unique to mycobacteria (Barry *et al.* 1998).

Cultures of *M. bovis* BCG and *M. tuberculosis* exposed to INH showed an increase in NAT production. This was shown using both the Western blot and Q-PCR method. *M. bovis* BCG growing cultures were exposed to INH at a concentration of 0.28µg/ml for 2 and 4 hours. The INH MIC of *M. bovis* BCG appears to be 10-fold higher than *M. tuberculosis* (Hesseling *et al.*, in Press). The concentration of INH used in the *M. bovis* BCG experiments was approximately ten times that used in the *M. tuberculosis* experiments. Preliminary experiments were done with all the protein fractions to test which fraction showed the clearer results, and it was found that the cytosol fraction produced the best results (see Figure 4.1).

When *M. tuberculosis* cultures of three different isolates were exposed to INH at a concentration of 0.01µg/ml (10X lower than the critical concentration used for susceptibility testing in BACTEC), each showed an increase in *nat* expression, although the effect was quantitatively different amongst the three isolates. A lower INH concentration was chosen for these experiments, as Q-PCR is highly sensitive and is able to detect slight changes in expression levels. Cell lysis was avoided as the half-life of mRNA is that of a few minutes (Hellyer *et al.*

1999), and mRNA degradation would cause inconsistent results. The reference strain, which is sensitive to INH (MIC of INH < 0.01μg/ml), showed a rapid increase in *nat* expression up to two hours (30-fold increase after 2 hours) of INH exposure, where the isolate containing the T529C and G619A SNPs (MIC of INH 0.02μg/ml) showed a more gradual increase in *nat* expression (6-fold increase after 2 hours) and the isolate with high-level INH resistance (MIC of INH >10μg/ml) showed little change in *nat* expression (see Figure 4.3). These results suggest that INH affects the expression of *nat*.

The isolates containing the SNPs showed a more gradual increase in *nat* expression than the reference strain H37Rv, when growing cultures were exposed to INH. These isolates also have a slightly higher INH MIC than the reference strain H37Rv, and the effects INH has on the isolate's cell environment may be slower than that of the reference strain H37Rv, resulting in a slower response in *nat* expression. It has been shown that the apparent affinity for INH of mutated 207R NAT is 10X less than the wild-type 207G NAT (Upton *et al.* 2001). This implies that the 207R NAT may not be acetylating INH as it enters the cells as quickly as the wild-type 207G NAT. This would result in less NAT being pulled from its endogenous role in isolates with the SNPs, than the reference strain H37Rv, and therefore less NAT would be needed to compensate for INH entering the growing culture of isolates containing the SNPs.

From these results it cannot be determined whether INH interacts directly with NAT or the promoter of *nat,* or whether changes in the cell environment cause the increase in *nat* expression. In the resistant isolate studied, INH had little affect on *nat* expression. Recently it has been reported, from bioinformatics analysis of the *M. tuberculosis* genome, that the *nat* gene is part of an operon involving three other genes (Payton *et al.* 2001a). A sixth putative gene has now been added to this operon (TIGR, 2003). The promoter region would be expected to lie upstream of this operon. It may be the case that more than one factor is involved in the control of *nat* expression. For example more than one promoter could be involved in *nat* expression as discovered for the *katG* gene and the rRNA operon (Gonzalez-y-Merchand et al. 1998;Master et al. 2001). The promoters may function independently and respond to varied environmental inputs and physiological demands (Gonzalez-y-Merchand et al. 1998;Master et al. 2001), such as cell wall synthesis requirements. It could also be the case that one of the other genes in the operon may need to be expressed before the expression of *nat* can occur. The entire operon may be translated and then undergo protein splicing to release the different proteins when needed (Klabunde *et al.* 1998). The expression of the other genes in the proposed operon should

therefore be investigated, and these genes making up the operon should be analysed for inteins and exteins (Klabunde *et al.* 1998).

There is evidence that INH alters the expression of a range of genes in the *M. tuberculosis* strain H37Rv (Wilson *et al.* 1999), and the majority of these genes are involved in cell envelope synthesis; such as *kasA, acpM* and those of the Antigen 85 Complex. Proteins KasA and AcpM join together to produce a protein complex, which is dramatically upregulated when *M. tuberculosis* cells are treated with INH (Slayden and Barry 2000). Hyperexpression of AcpM is toxic to mycobacteria, however, hyperexpression of KasA leads to INH resistance (Slayden and Barry 2000). AcpM and KasA, like NAT, are part of a putative operon consisting of six genes. The first gene of the operon appears to be a homolog of a regulatory protein that controls production of polyketide in *Streptomyces ambofaciens*. The second gene encodes a *fabD* homolog, which is malonyl CoA:ACP transacylase; the third and fourth genes are *acpM* and *kasA*. The fifth gene is another keto-acyl synthase (*kasB*), which shows significant homology to *kasA*. The last gene of the six gene operon is an acetyl-CoA carboxylase β-subunit homolog that is involved in production malonyl-CoA. These genes are therefore like *nat* as their expression is upregulated when cells are exposed to INH, and they are all part of an operon. Mutations (such as SNPs) in *acpM* and *kasA* have been found to be associated with INH resistance in *M. tuberculosis* isolates (Ramaswamy *et al.* 2003). Other genes important for cell envelope development in mycobacteria showing similar characteristics to *nat,* are those that encode the Antigen 85 Complex (especially *fbpC*) as they too are induced in *M. tuberculosis* strain H37Rv but not in INH resistant *M. tuberculosis* isolates (Garbe et al. 1996;Wilson et al. 1999), when cells are exposed to INH. The characteristics of both these protein complexes are similar to that of NAT (described in this study), suggesting that NAT may be involved in cell envelope synthesis, making NAT a potential candidate for drug development.

The *nat* gene did not appear to be induced in microarray studies, in which genes upregulated by INH were investigated (Wilson *et al.* 1999). The isolates and methods presented here give additional information looking at a particular candidate gene, which may not be detectable in the microarray study due to low levels of messenger RNA (Vainrub and Montgomery Petitt 2003). The combined approach of microarray analysis together with a candidate gene approach can therefore provide complimentary information.

In conclusion, it has been found that the exposure of *M. bovis* and *M. tuberculosis* cultures to INH resulted in the expression of NAT increasing. However, the relative increase varied between isolates, indicating that factors other than a direct interaction between NAT and INH may also be involved in this increase.

CHAPTER 5

LOCALISATION OF NAT WITHIN MYCOBACTERIUM CELLS

The aim of this chapter is to describe the localisation of the NAT protein in *M. smegmatis*, *M. bovis* BCG and *M. tuberculosis* cells. These mycobacteria were fractionated into cytosol, cell membrane and cell wall proteins and the NAT protein was visualised using the Western blot technique.

5.1. Introduction

The cytosol (as opposed to the cytoplasm, which also includes the organelles) is the internal fluid of the cell, and a large part of the cell metabolism occurs there. Proteins within the cytosol play an important role in signal transduction pathways, glycolysis, and act as intracellular receptors and ribosomes. In prokaryotes, most chemical reactions take place in the cytosol (Wikipedia Encylopedia 2002).

Mycobacteria have specialised cell envelopes that are rich in unusual polysaccharides and lipid components. The unique long-chain lipids provide a very hydrophobic barrier to antibiotic access. It has been postulated that the distinctive pathogenicity of *M. tuberculosis* can be correlated with this unusual cell envelope (Minnikin *et al.* 2002). The innermost layer of the capsule is the plasma membrane. This layer is similar to those found in other bacteria, consisting of a permeable lipid bilayer with interacting proteins enclosing the cell cytoplasm (Minnikin 1982;Rastogi 1991;Minnikin 1991;McNiel and Brennan 1991;Barrow 1997). The majority of these proteins moving within or upon this layer are primarily responsible for transport of ions, nutrients and waste across the membrane (Sullivan 2003).

The next layer consists of the peptidoglycan/arabinogalactan moieties that make up the basic structural component of the cell wall. Adjacent to this layer is a layer that appears electron transparent and is made up predominantly of mycolic acids and other complex lipids (Barrow 1997;Brennan 2003). The outermost layer is comprised of a variety of components that are mycobacterial species specific (Barrow 1997;Minnikin et al. 2002). This layer generally appears fibrillar in nature when observed by freeze fracture or negative staining (Draper 1974;Barksdale and Kim 1977;Barrow 1997). The outer layers are important with regard to initial host

interaction, as they contain mycobacterial components that can induce various host responses (Barrow 1997;Brennan 2003). Proteins found in the cell wall and outermost layer of the cell envelope could be involved in host pathogenicity or form part of the complex biochemical reactions required for the production of such a complex cell envelope.

The localisation of NAT was determined by fractionating the cells of *M. smegmatis*, *M. bovis* BCG and *M. tuberculosis* into cytosol, cell membrane and cell wall proteins. The NAT protein in the different protein fractions was visualised using the Western blot technique. The localisation of the NAT protein could assist with the prediction of an endogenous function.

5.2. Materials and Methods

Unless stated otherwise all reagents used in this study were from Sigma-Aldrich.

5.2.1. Optimisation of *M. smegmatis* NAT antibody # 155

Pure NAT protein samples were, as recombinant proteins in *E. coli* following heterologous expression of *nat* genes, collected from the following bacteria: *Salmonella typhimurium*, *Amycolatopsis mediteranei*, *Pseudomonas aeruginosa*, *M. smegmatis* and *M. tuberculosis*. Protein samples were supplied by Prof. E. Sim's lab, Department of Pharmacology, Oxford University. Each protein (600ng) was run on an SDS-PAGE gel (see section 4.2.4) prior to Western blotting (see section 4.2.5). Four different antibody concentrations (1:500; 1:1,000; 1: 5,000 and 1:10,000) were tested. The protocol followed for stripping and re-probing was as described in the ECL Reagent kit manual (Bioscience).

5.2.2. Location of NAT

Once growing cultures reached the desired absorbance at 600nm (0.7 – 0.8), the lysate was prepared batchwise (see section 4.2.2) from 2 X 10^{10} cells, and the lysate from 8 X 10^{10} cells was fractionated into cytosol, cell membrane and cell wall (see section 4.2.3), and protein concentration was determined for each fraction (see section 2.2.6). A sample (25µl) of each fraction was run on a SDS-PAGE gel (see section 4.2.4), which was followed by Western blotting (see section 4.2.5).

5.3. Results

5.3.1. *M. smegmatis*

5.3.1.1. Optimisation of antibody # 155

A rabbit polyclonal antibody raised against recombinant *M. smegmatis* NAT protein, designated 155 was characterised to determine its cross reactivity with other proteins. This *M. smegmatis* antibody # 155 had been developed by others (E. Sim, University of Oxford, Oxford), however, it had not been previously used. Before the antibody could be used in this study, the concentration at which the antibody worked best, needed to be established. The specificity of the antibody also needed to be known. Recombinat NAT proteins (in lysates of appropriately genetically modified *E. coli*) were run on an SDS-PAGE gel. Proteins were transferred to a cellulose membrane and differing concentrations of antibody was applied to establish which concentration gave the best results.

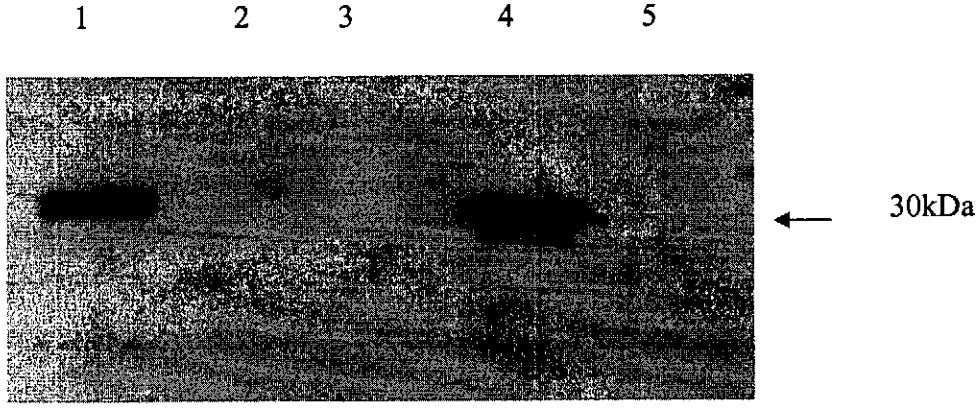


Figure 5.1: Western blot of recombinant NAT proteins.

The primary antibody # 155 was used at a dilution of 1:4,000 (secondary antibody was kept at 1:10,000 dilution). 1 – *S. typhimurium*, 2 – *Amycolatopsis mediteranei*, 3 – *Pseudomonas aeruginosa*, 4 – *M. smegmatis* and 5 – *M. tuberculosis* (600ng of each protein was loaded). Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by arrow.

Figure 5.1 shows the results of the *M. smegmatis* antibody used at a dilution of 1:4,000. This dilution produced the best results and was used in the rest of the *M. smegmatis* Western blots. The only other NAT protein detected weakly with this antibody was that of *S. typhimurium*, showing that this antibody is fairly specific to *M. smegmatis nat*.

5.3.1.2. Cell Fractionation

To determine where the protein was located within the bacterium, the cells were fractionated into cytosol, cell membrane and cell wall. The different fractions were analysed by the Western blot technique. The fractions were isolated from different growth stages of the bacterium i.e. A_{600} of 0.5 was representative of cells in early to mid-log phase and A_{600} greater than 2 was representative of cells in stationary phase.

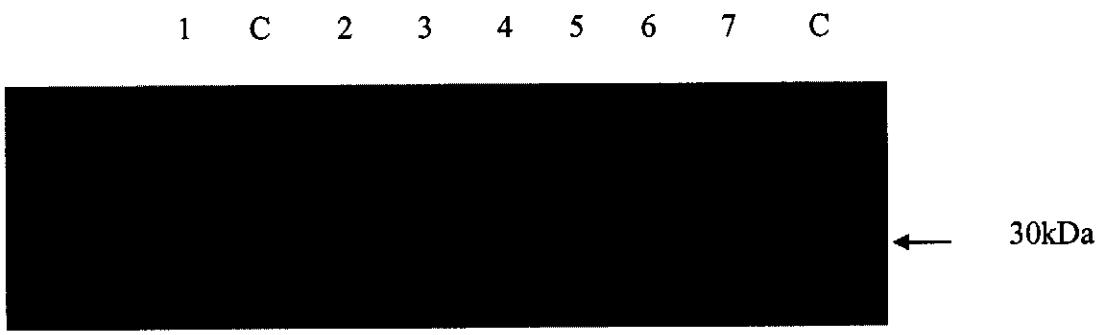


Figure 5.2: Western blot following fractionation of *M. smegmatis* cells harvested at early to mid-log phase.

C – purified recombinant *M. smegmatis* NAT protein, 1 – *M. smegmatis* lysate, 2 and 3 – cell membrane, 4 and 5 – cell wall and 6 and 7 – cytosol. A volume of 25µl of each cell fraction was loaded onto the gels. Primary antibody # 155 used at a dilution of 1:4,000 and secondary antibody used at a dilution of 1:10,000. Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by arrow.

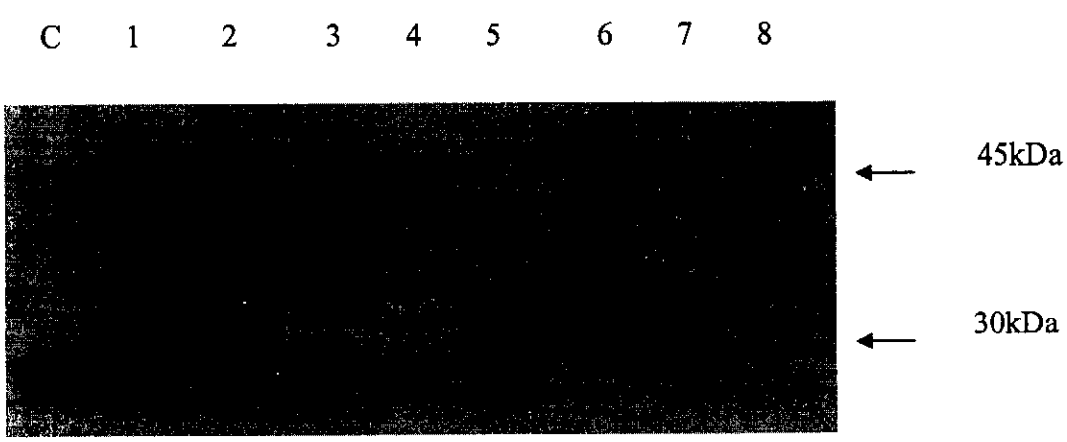


Figure 5.3: Western blot following fractionation of *M. smegmatis* cells harvested at stationary phase.

C – purified *M. smegmatis* NAT protein, 1 and 2 – *M. smegmatis* lysate, 3 and 4 – cytosol, 5 and 6 – cell wall and 7 and 8 – cell membrane. A volume of 25µl of each cell fraction was loaded onto the gels. A higher band (also seen *M. bovis* BCG), in the cell membrane protein fraction, becomes evident in cells moving into stationary phase. Primary antibody # 155 used at a dilution of 1:4,000 and secondary

antibody used at a dilution of 1:10,000. Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by arrow.

Figures 5.2 and 5.3 show the results of a Western blot of fractionated cells of *M. smegmatis* at early exponential and stationary growth phase. All protein fractions extracted from the early exponential phase show that NAT is present (see Figure 5.2), however, the protein fractions from cells nearing stationary phase do not show NAT in the cytosol fraction (see Figure 5.3). Both the lysate and cell wall fractions show the NAT protein but a fainter band is also present that runs at a slightly higher molecular weight (MW) to that of the NAT protein. The cell membrane fraction does not show the slightly higher MW band, but a band is present with a MW of approximately 45kDa.

5.3.2. *M. Bovis* BCG

5.3.2.1. Cell Fractionation

M. Bovis BCG cells were fractionated into cytosol, cell membrane and cell wall. The different fractions were analysed by Western blotting using antibody # 84. The fractions were isolated from different growth stages of the bacterium i.e. A_{600} of 0.5 was representative of cells in early to mid-log phase, A_{600} of 1 was representative of cells in late mid-log to early stationary phase and A_{600} greater than 2 was representative of cells in stationary phase. Antibody # 84 has previously been optimised to work best at a 1:50,000 dilution (Upton et al. 2001). However a 1:10,000 dilution was used in the Western blots of this study, as this dilution produced the best results for this study.

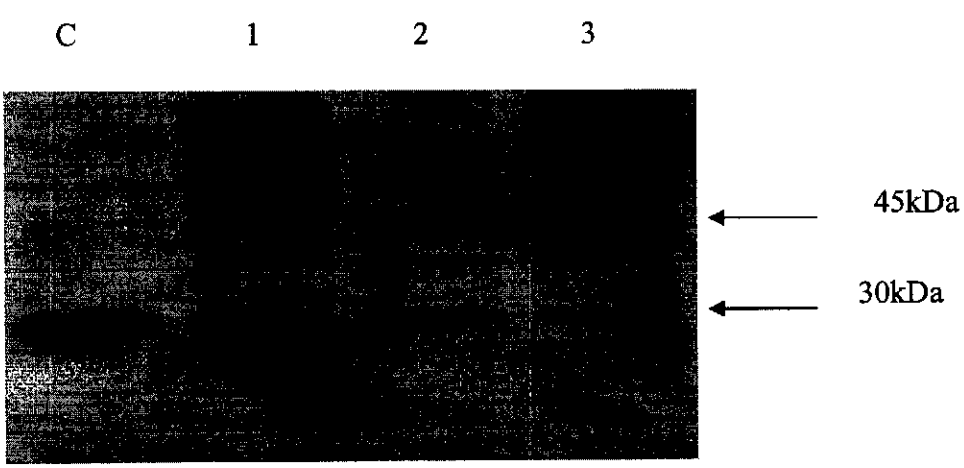


Figure 5.4: Western blot following fractionation of *M. bovis* BCG cells harvested at late mid-log phase to early stationary phase.

C – recombinant NAT (in *E. coli* lysate). Tracks 1 – 3 are fractions of *M. bovis* BCG as follows: 1 – cytosol fraction, 2 – cell wall fraction and 3 – cell membrane fraction. A volume of 25µl of each cell fraction was loaded onto the gels. Primary antibody #84 and secondary antibody (mouse anti-rabbit IgG mAbs conjugated to horsesradish peroxidase) used at a 1:10,000 dilution. Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by arrow.

Figure 5.4 shows NAT is present in the cytosol and cell wall fraction. The cell membrane fraction does show a faint band of the NAT protein, but it cannot be certain whether this is carry-over protein from the cytosol fraction. The cell wall fraction shows the NAT protein at a higher MW of approximately 45kDa. A faint band is present at 30kDa, but this too could be carry-over from the cytosol fraction.

5.3.3. *M. tuberculosis*

5.3.3.1. Cell Fractionation

Cell fractions extracted from *M. tuberculosis* were run on a SDS-PAGE gel and analysed by Western blotting. The fractions were isolated from different growth stages of the bacterium i.e. A_{600} of 0.5 was representative of cells in early to mid-log phase, A_{600} of 1 was representative of cells in late mid-log to early stationary phase and A_{600} greater than 2 was representative of cells in stationary phase. To visualise the NAT protein, antibody # 84 was used at a 1:4,000 dilution.

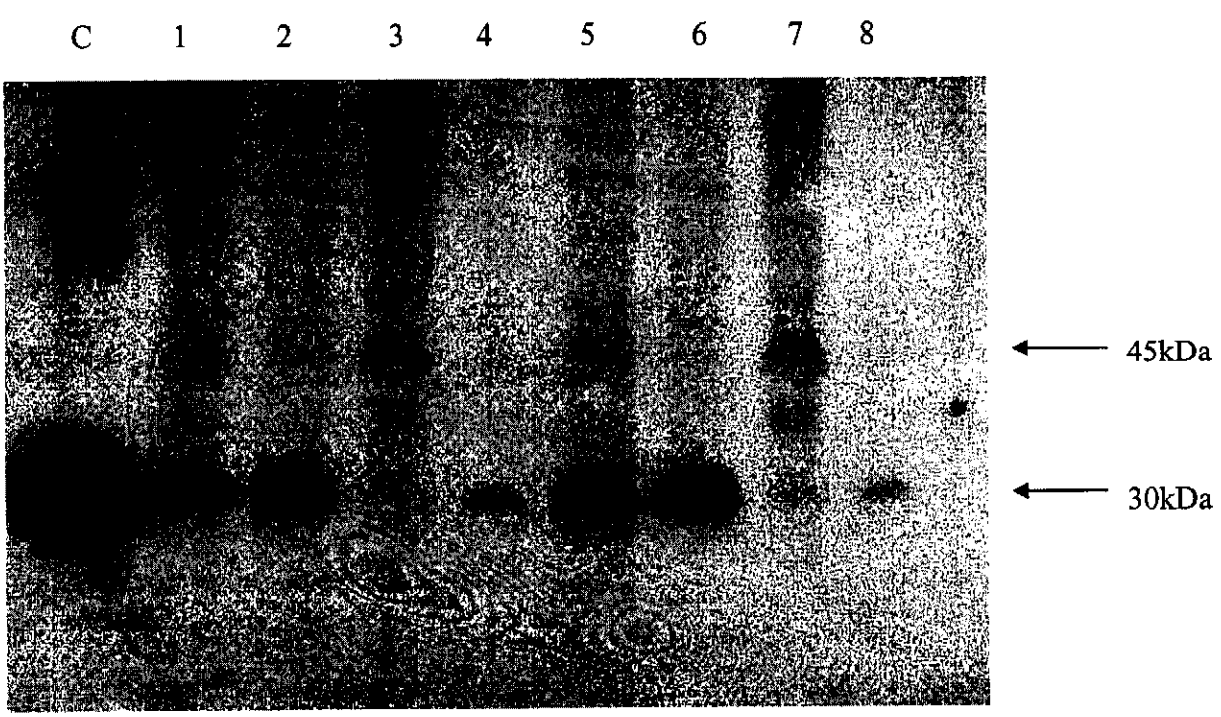


Figure 5.5: Western blot following fractionation of *M. tuberculosis* cells harvested at late mid-log phase to early stationary phase.

C – recombinant NAT (in *E. coli* lysate), 1,5 – lysate; 2,6 – cytosol fraction, 3,7 – cell wall fraction and 4,8 – cell membrane fraction. A volume of 25µl of each cell fraction was loaded onto the gels. Primary antibody #84 and secondary antibody (mouse anti-rabbit IgG mAbs conjugated to horsesradish peroxidase) used at a 1:10,000 dilution. Apparent molecular weight from Rainbow markers (Amersham biosciences) is shown by arrow.

Figure 5.5 shows that the NAT protein appears in all the protein fractions. The cell membrane proteins show a very faint band for the NAT protein, the protein in this fraction could be carry-over from the cytosol fraction. The cell wall fraction shows a band running at a higher molecular weight of 45kDa. This figure is in agreement with results from *M. bovis* BCG.

5.4. Discussion

The NAT protein in the three cellular fractions of *M. smegmatis*, *M. Bovis* BCG and *M. tuberculosis* shows a different pattern on the Western blots. *M. smegmatis* was fractionated following harvest during exponential phase, the NAT protein appeared in all fractions as a single clear band of an apparent molecular weight of 30kDa (see Figure 5.2). However, when cells were harvested in or very close to stationary phase, the NAT protein was no longer evident in the cytosol fraction (see Figure 5.3). In the other fractions, the NAT protein appeared clearly but with a fainter band of a slightly higher molecular weight (MW). The cell membrane fraction showed a band of a much higher molecular weight (approximately 45kDa). This increase in apparent molecular weight is likely to represent a very strong bond between the NAT protein and another molecule which is not broken by the preparation of the gel samples in denaturing agents (see section 5.2.2) The disappearance of the NAT protein in the cytosol fraction could be accounted for by translocation of the NAT to the cell envelope. It may be deduced that the NAT protein is produced in the cytosol and then translocated to where it is needed, which is most probably the cell envelope region. It may be speculated that by stationary phase the NAT is no longer required and production is probably stopped in the cytosol and, therefore none is being made to replace that being translocated out.

However, in both the *M. bovis* BCG and *M. tuberculosis* cytosol fractions the NAT protein was always detected at 30kDa. The *M. bovis* BCG and *M. tuberculosis* NAT protein appears mainly in the cytosol and cell wall fraction. A faint band correlating with the NAT band is evident in

the cell membrane fraction, but it cannot be certain whether this is a carry-over from the cytosol fraction. The NAT protein in the cell wall fraction appears much larger than in the other fractions. The NAT protein has 275 amino acids and runs at a molecular weight of 32kDa (Payton *et al.* 1999). In the cell wall fraction the NAT protein runs at a molecular weight of approximately 45kDa (see Figure 5.4 and 5.5). For *M. bovis* BCG, a faint band is also evident at approximately 30kDa, but this may be carry-over from the cytosol fraction (see Figure 5.4). This differs from *M. smegmatis* as the higher MW band appeared in the cell wall fraction and not the cell membrane fraction (as seen in *M. smegmatis*). Non-specific banding cannot be completely ruled out for the higher molecular weight band, but seeing that the Western blot is clean with minimal non-specific binding and that it appears when different antibodies are being used, it can be deduced that the band is that of a modified NAT protein. To verify that this band is not non-specific binding, fractionation of a *M. bovis* BCG NAT knock-out strain needs to be done. A pure NAT knock-out strain should be investigated (Bhakta et al. 2004).

The higher molecular weight protein in the cell envelope fractions is very faint, which could indicate a low concentration of NAT in this protein fraction, or it could be due to the protein now having less affinity towards the specific antibody. The antibody was made specifically for the 30kDa NAT protein.

The NAT protein is hydrophilic (Payton *et al.* 1999) and therefore possibly needs some kind of post-translational modification (hence the presence of the 45kDa band) making the protein hydrophobic to enable it to be transported to the cell membrane. The conversion of the NAT protein from a hydrophilic protein to a hydrophobic protein may result in the protein becoming less soluble. This too would explain the low concentrations of NAT evident on the Western blots of the cell wall protein fraction. It is known that several mycobacterial proteins do undergo post-translational modifications, such as glycosylation (Garbe et al. 1993;Dobos et al. 1996). Glycoproteins are present on the mycobacterial cell surface and are available for interaction with the host cells during infection. Glycosylation on its own will not make the protein hydrophobic, but will interact with cell membrane proteins enabling the uptake of the attached protein into the cell membrane. Apa (also known as the 45/47kDa complex) is an immunodominant antigen secreted *in vitro* by the *M. tuberculosis* complex (i.e. *M. tuberculosis* and *M. bovis* BCG) (Dubos *et al.* 1953). The core protein of Apa is 28.8kDa and due to glycosylation this protein's native form is shown to have a molecular weight of 45 and 47kDa. It has been shown that the

deglycosylated antigen is 10-fold less active than native molecules in eliciting delayed-type hypersensitivity reactions in guinea pigs immunized with BCG (Romain *et al.* 1999).

The presence of NAT in the cell membrane and cell wall fractions implies that NAT may be involved in the chemical synthesis of the cell envelope. If this can be identified then NAT would become a potential candidate for drug therapy. Once the 45kDa NAT protein in the cell wall fraction has been verified, deglycosylation techniques could be applied to the protein to determine whether glycosylation is the cause of the higher MW. Proteins associated with cell surface or secreted into culture medium have been shown to be modified by covalent attachment of glycan moieties. These proteins in *M. tuberculosis* are regarded as good candidates for new vaccines against TB.

CHAPTER 6

SUICIDE DELIVERY VECTOR CONSTRUCT

The aim of this chapter was to construct a suicide delivery vector containing the T529C and G619A SNPs. This vector would be used in on-going studies at Prof. E. Sim's lab, Department of Pharmacology, Oxford University.

6.1. Introduction

Plasmids have been used frequently to manipulate the genomes of various organisms (Brown 1990). Until recently, the tools available for the manipulation of the *M. tuberculosis* genome have been poor, and therefore progress in vaccine and drug development has been hampered (Parish and Stoker 2000). A new efficient system for introducing specific mutations in *M. tuberculosis* has been developed. This method makes use of a suicide delivery vector (Parish and Stoker 2000).

Suicide plasmids are not able to replicate in *M. tuberculosis* as they lack a mycobacterial origin of replication (*oriM*) (Parish and Stoker 2000). Construction of the suicide delivery vector has often been problematic as several cloning steps are required to include the relevant markers, and finding appropriate restriction sites for inserting these genes has been limiting. To overcome this problem two series of vectors have been developed. By separating the construction of the mutated version of the target gene from that of the inclusion of the required marker genes, vector construction is rapid and flexible (Parish and Stoker 2000).

The first series of plasmids (pNIL) is for manipulating the gene of interest. The plasmids consist of a simple cloning vector with an origin of replication for *E. coli* (*OriE*), a kanamycin resistance gene (*kan*) and multiple cloning sites (MCSs) with different restriction enzyme sites, but including a single *Pac*I restriction site. The second series (pGoal) is used to generate and store a number of cassettes of marker genes. These consist of vectors containing *oriE*, an ampicillin resistance gene (*amp*) and different combinations of marker genes (*hyg*, *lacZ* and *sacB*) flanked by two *PacI* sites. Thus the markers genes can be excised as a *PacI* cassette and inserted into any of the pNIL series in a one-step cloning process. The benefit of using *PacI* sites is that none

are present in the *M. tuberculosis* genome (Parish and Stoker 2000). This strategy and series of plasmids could be used in the following study.

To determine the effects the T529C and G619A SNPs have on the growth of *M. tuberculosis* and the *in vivo* appearance of the NAT protein, a suicide delivery vector will be constructed. This vector will be used in Prof. Sim's lab to transform the wild-type *M. bovis* BCG strain to that of the 177 His and 207 Arg NAT mutant.

6.2. Materials and Methods

6.2.1. Buffers and Reagent mixes

Unless stated otherwise all reagents are from Sigma-Aldrich.

6.2.1.1. SOC Medium

This medium contains 80ml dH_2O, 2g Bacto-tryptone (Biolab), 500mg Bacto-yeast extract (Biolab) and 50mg NaCl (Biolab). The ingredients were allowed to dissolve before 1ml 250mM KCl (186mg in 10ml dH_2O) was added and the pH was adjusted to 7 with 4M NaOH. Using dH_2O, a final volume of 100ml was made up and the medium was autoclaved. This was stored at room temperature. Immediately before use, 5ml of autoclaved 2M $MgCl_2$ (1.9g $MgCl_2$ and dH_2O to make a final volume of 10ml) and 2ml 1M glucose solution (1.8g glucose and dH_2O to make a final volume of 10ml) was added.

6.2.1.2. LB medium and agar

This medium contains 450ml dH_2O, 5g Bacto-tryptone (Biolab), 3g bacto-yeast extract (Biolab) and 2.5g NaCl (Biolab). The pH was adjusted to 7 with 4M NaOH and a final volume of 500ml was made up with dH_2O. To make LB agar the protocol for LB medium is followed, to which 7.5g of Bacto-agar (Biolab) is added. Both mediums were autoclaved before use.

6.2.2. PCR

Table 6.1: Primers used to amplify the *nat* fragments

Product (bp)	Primer name	Primer sequence	Primer name	Primer Sequence
1250	Bam F	5' TGGAACCTATCTG GCCGCGACCTT 3'	Bam R	5' CGCGGATCCATG GGGATCTGGTGA 3'
1629	n Hind F	5' CCCAAGCTTGGGTTCG GCTCGAAGGAGCCGGA 3'	n Hind R	5' GATCCCGAACC GTTCGCTCA 3'
	Natseq F	5'GGGCACTTTGTTTGC CGTC 3'		

The Bam R and n Hind F primers were designed to incorporate a BamH1 and HindIII site, respectively, into the amplified PCR product. The primers all had the same annealing temperature of 56°C. For the PCR reaction conditions see section 3.2.1.

6.2.3. Dephosphorylation Reaction

This reaction mix contains 20µl restriction reaction (see section 3.2.2), 1U of alkaline phosphatase (Promega) and dephosphorylation buffer (Promega) added to a final concentration of 1X. Reaction was incubated at 37°C for 30 min.

6.2.4. Ligation Reaction

Restricted DNA was cleaned using the Qiagen PCR purification kit and vacuum dried. The DNA to be ligated was resuspended in 8µl ddH_2O to which 1U ligase (Promega), and ligation buffer (Promega) to a final concentration of 1X, was added. The reaction mix was left at 4°C overnight.

6.2.5. Transformation

The ligation reaction mix (see section 6.2.4) was cleaned using Qiagen PCR purification kit. For transformation, 50µl JM109 competent cells (Promega) were added to 5µl of cleaned ligation reaction mix, and left on ice for 30 min. The mix was heat-shocked at 42°C for 45 sec and returned to ice for 2 min. To the mix, 950µl SOC medium (see section 6.2.1.1) was added and left shaking at 37°C for 1 hour. After one hour, 300µl of the medium was plated onto selective medium i.e. LB agar (see section 6.2.1.2) supplemented with selective antibiotic.

6.2.6. Suicide Delivery Vector Construct

6.2.6.1. PCR reactions

Two fragments were amplified: Bam fragment using the BamF/R primers and Hind fragment using the n Hind F/R primers (see Table 6.1). Primers Bam R and n Hind F have incorporated a BamH1 and HindIII restriction site, respectively. DNA used in the PCR reactions was that of *M. tuberculosis* isolate 1430 that naturally contains both the G619A and the T529C SNP (see section 3.2.1 and section 3.2.3).

6.2.6.2. PCR reaction mix

This mix differs from section 3.2.1, in that *Pfu* polymerase was used instead of *Taq* polymerase, and DMSO is added to the reactions mix:

1X buffer (Promega), 0.2μM forward primer, 0.2μM reverse primer, 0.1U/μl *Pfu* polymerase (Promega), 0.8mM dNTPs (Promega), DMSO (Sigma) to a final volume of 6%, 1ng/μl genomic DNA and make up to final volume (100μl) with ddH_2O.

The annealing temperature of these primer sets was 56°C. Other changes made to the PCR cycle as seen in section 3.2.1.2 were that the elongation step and the last 10 minute hold was kept at 74°C. The elongation step was also changed from 2 minutes to 4 minutes.

6.2.6.3. Vectors

The p2NIL (4,753bp) and pGoal19 (10,435bp) vectors were kindly donated by Prof. E. Sim's group at the Pharmacology department, University of Oxford (see Figure 6.1).

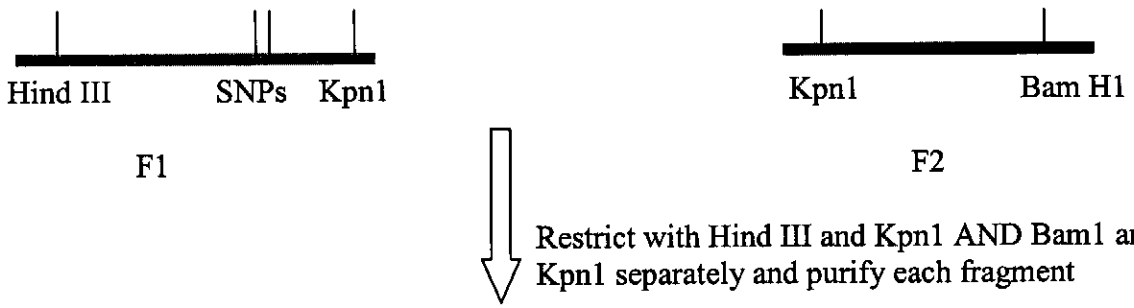
6.2.6.4. Construction of the F1F2-p2NIL plasmid

For stepwise illustration please see Figure 6.1. Fragment 1 (F1 – 1250bp) was the PCR product of the Bam primer set, and fragment 2 (F2 – 1629bp) was the PCR product of the Hind primer set (see Table 3.1). Both these amplified fragments were cleaned using the Qiagen PCR purification kit. Both cleaned fragments were restricted with Kpn1 (see section 3.2.2) at 37°C for 3 hours and then cleaned using the Qiagen PCR purification kit. The cleaned fragments were ligated (see section 6.2.4) overnight and then cleaned using the Qiagen PCR purification kit. The ligated fragment and the p2NIL vector were restricted with both BamH1 and HindIII at 37°C for 3 hours. These restrictions were cleaned using the Qiagen PCR purification kit and ligated together forming the F1F2-p2NIL plasmid. This plasmid was used to transform JM109

competent cells (see section 6.2.5), which were selected for on LB agar plates (see section 6.2.1.2) supplemented with 50μg/ml kanamycin. Single colonies were picked and grown in liquid LB medium (see section 6.2.1.2), supplemented with 50μg/ml kanamycin, with overnight shaking at 200rpm at 37°C. The following day the constructed plasmid was extracted using the SV plasmid extraction kit (Promega). The purified plasmid was verified by restricting it with both BamH1 and HindIII and incubating the restriction mix at 37°C for 3 hours. This should release the insert from the p2NIL vector, which was excised from a 1% agarose gel, purified and the *nat* insert was sequenced using the natseq primer and the *nat* 3 primer (see Table 3.1 and Table 6.1). The sequences generated were aligned to the TIGR *nat* sequence (TIGR, 2003) using the computer program DNAssist. Once the F1F2-p2NIL plasmid was verified, glycerol stocks were made of the transformed JM109 cells.

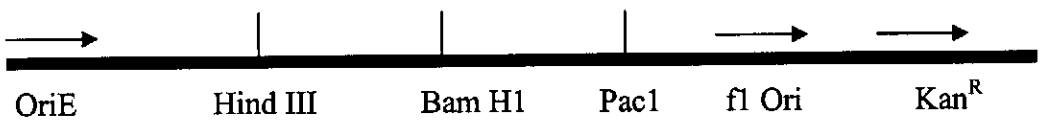
6.2.6.5. Construction of the F1F2-p2NIL-pGoal suicide vector

For stepwise illustration please see Figure 6.1. The F1F2-p2NIL plasmid and the pGoal19 vector was restricted with PacI (see section 3.2.2) and left at 37°C for 3 hours. When restricting with Pac1, BSA (usually comes as 100X) must be added to a final concentration of 1X. The restricted F1F2-p2NIL plasmid was dephosphorylated (see section 6.2.3). The dephosphorylated restriction of F1F2-p2NIL was cleaned using the Qiagen PCR purification kit and the restricted pGoal19 vector was run on a 1% agarose gel and the upper band (7945bp) was excised and cleaned using the Qiagen gel purification kit. F1F2-p2NIL was ligated to the upper pGoal19 restricted fragment (see section 6.2.4). The ligated fragment that was now >14kb was not cleaned, but used directly to transform JM109 cells. The newly constructed suicide vector was selected for on LB agar plates (see section 6.2.1.2) supplemented with kanamycin (30μg/ml), IPTG and X-Gal. Single blue colonies were selected from plates and grown in liquid LB medium supplemented with kanamycin, 30μg/ml. The suicide vector constructed was purified using the SV plasmid purification kit (Promega) and verified by restricting with BamH1 (producing fragments 9340/9052bp, 3072 and 2224/2512bp), and PstI (producing fragments 5196, 3974/3841, 2196/2288, 1778 and 1533). The fragment sizes were dependent on the orientation of the binding of the two fragments. Once verified, glycerol stocks were made of the transformed JM109 cells.



F1 F2

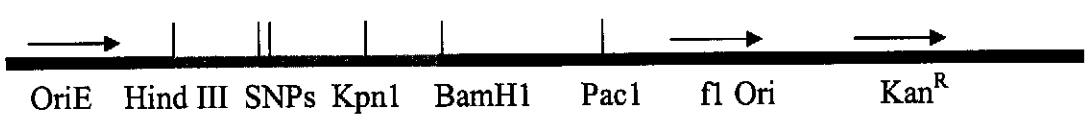
Restrict with Hind III and Kpn1 AND Bam1 and Kpn1 separately and purify each fragment



(p_2NIL (4,753bp))

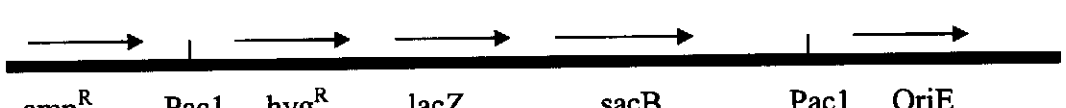
Restrict with Hind III and Bam H1, purify and dephosphorylate

Mix the restricted fragments and let them undergo ligation and transformation. Select on LB kan (50µg/ml) plates.



(F1 – F2 – p2NIL)

+



(pGoal 19 (10,430bp))

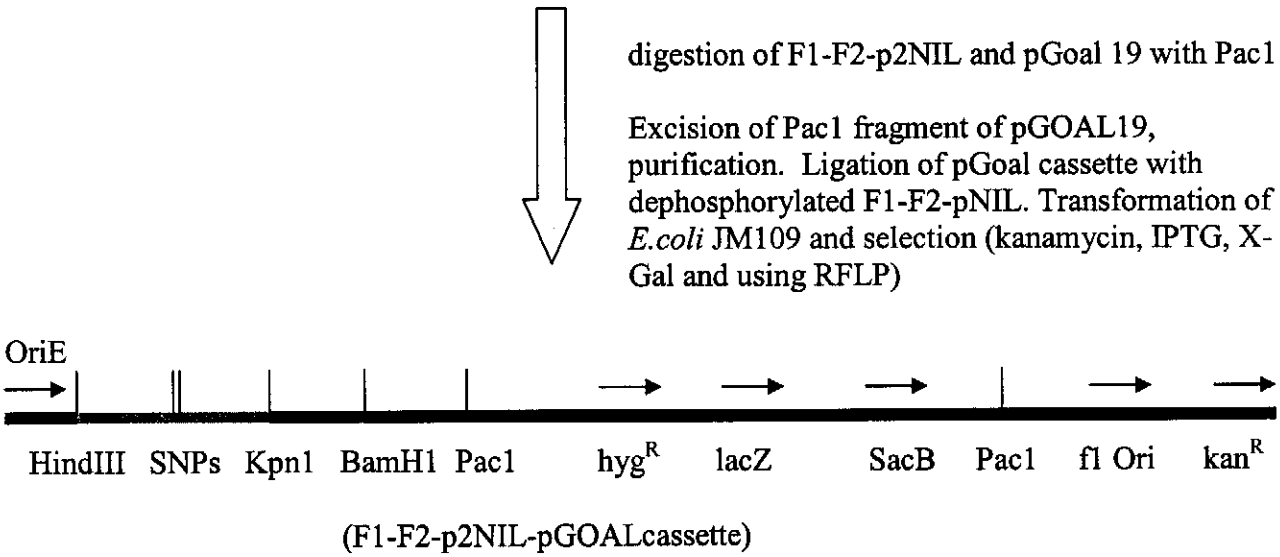


Figure 6.1: The steps involved in the construction of the suicide delivery vector.
Two vectors were used to construct the suicide delivery vector, a p2Nil vector and a pGoal19 vector. The *nat* coding region containing the SNPs was first cloned into the p2Nil vector which was subsequently cloned into the pGoal vector.

6.3. RESULTS

6.3.1. F1F2-p2NIL verification

Ligated F1 and F2, after restriction with BamH1 and HindIII, was inserted into the p2NIL vector forming the F1F2-p2NIL plasmid construct. This construct was multiplied in JM109 cells, selected for on LB kan (50μg/ml) plates and purified. Purified plasmids were restricted with BamH1 and HindIII to verify that the isolated plasmid contained the insert.

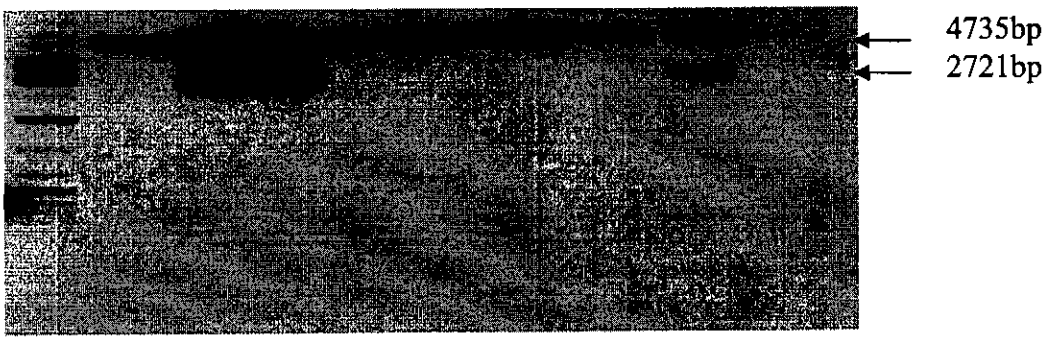


Figure 6.2: Purified Plasmid construct restricted with BamH1 and HindIII run on a 1% agarose gel.

The purified plasmids were restricted with BamH1 and HindIII, and two bands (4735bp and 2721bp) were produced in plasmids containing the putative insert (2721bp). Restricted fragments were run with a 1kb DNA ladder (GibcoBRL).

Three of the eight clones grown on selective medium had the plasmid with the putative insert, which was excised from the plasmid by the restriction digest using BamH1 and HindIII (see Figure 6.2). The smaller fragment (2721bp) was excised from the gel, cleaned and sent for sequencing.

The aligned sequences show that the original G619A SNP and the newly identified T529C SNP has been included into the inserted *nat* sequence. No other nucleotide changes were identified; therefore, the correct insert was inserted into the p2NIL vector.

6.3.2. F1F2-p2NIL-pGoal verification

Due to time constraints and technical problems (thus far unexplained) encountered, the F1F2-p2NIL vector was not ligated to the pGoal19 marker cassette. This will be attempted in Prof. E. Sim's lab, Department of Pharmacology, Oxford University.

6.4. DISCUSSION

Unfortunately, due to time and funding constraints the construct was not completed. The insert had been cloned into the p2NIL vector and this construct has been verified by restriction and sequence analysis. The remaining step of restricting the p2NIL construct with Pac1 and ligating it to the Pac1 marker cassette of the pGoal19 vector, will be done by others in Prof. Sim's lab.

A *nat* knock-out suicide delivery vector has been constructed (Personal communication E. Sim). This vector was used to knock-out the *nat* gene in *M. bovis* BCG, and phenotypic changes were monitored in the knock-out strain (Bhakta et al. 2004). Once the suicide delivery vector containing the *nat* gene with the SNPs is completely constructed, it will be used to complement the *nat* knock-out strain to a strain containing the *nat* gene with the identified SNPs. Changes in phenotypic characteristics will be monitored between the *nat* knock-out strain and the newly transformed strain with the SNPs. A similar experiment will be done with the wild-type *M. bovis* BCG strain and the *M. tuberculosis* H37Rv strain.

The results obtained in these experiments could give an indication of the function for the *nat* gene in *M. bovis BCG* and *M. tuberculosis,* and indicate whether the presence of the SNPs result in phenotypic changes in the bacterium.

CHAPTER 7
FINAL DISCUSSION

A third of the world's population is infected with *M. tuberculosis*, of which 10% will develop active TB. However, this figure is rising due to the AIDS epidemic (WHO 1999; Mukherjee *et al.* 2004) amongst many other reasons, and treatment is being hindered due to the emergence of MDR TB (Grange and Zumla 2002; Mukherjee *et al.* 2004; WHO 1999). INH is one of the front line drugs used in TB control (Torres *et al.* 2000). It is a prodrug and needs to be activated before it can act therapeutically. This occurs by an oxidation step controlled by catalase-peroxidase, encoded for by the *M. tuberculosis katG* gene (Upton *et al.* 2001b).

The enzyme N-acetyltransferase (NAT) can interfere with this oxidation step by acetylating the hydrazine moiety, and once INH is acetylated this oxidation step cannot occur (Upton *et al.* 2001b). Therefore once NAT it is acetylated, it can no longer function therapeutically. In humans, a variety of SNPs within the homologous *nat* gene results in humans having different phenotypes i.e. slow, intermediate or fast acetylators (Parkin *et al.* 1997). Fast and slow acetylators respond differently to INH treatment (Payton *et al.* 1999). *M. tuberculosis* produces NAT, which has been shown to acetylate INH *in vitro,* and when overexpressed in *M. smegmatis*, the resistance to INH of the transformed organism increases three-fold. Prior to this study a G619A SNP was reported in *M. tuberculosis*, however a further SNP was identified in this study and the known G619A SNP was analysed further. The previously identified G619A SNP, which resulted in a G207R amino acid change (Upton *et al.* 2001b), was screened for in 37 *M. tuberculosis* strain families found in the Western Cape region of South Africa. This SNP was restricted to 2 strain families, viz. strain family 3 and strain family 28 (Warren *et al.* 2000), as well as one unique isolate. In this study, an additional SNP was identified in the ORF of *nat* in *M. tuberculosis* isolates. This T529C SNP resulted in a tyrosine to histidine change at amino acid 177 (Y177H) of NAT, and when screened for in the 37 strain families was shown to be restricted to one strain family, viz. strain family 3

Molecular population genetic analysis of pathogenic bacteria has resulted in the identification of a trend, in which biomedically relevant traits, such as host range and virulence, are not randomly distributed among phylogenetic lineages (Gutacker *et al.* 2002; Musser 1996). Pathogenic bacteria have a high level of genetic variation owing to differences in gene content and

nucleotide variation in or between structural genes. Allelic variation arises from random nucleotide mutations, which can be followed by selection and amplification. When the nucleotide changes survive, they are termed Single Nucleotide Polymorphisms (SNPs). These SNPs are classed into two groups, nonsynonymous SNPs (nsNPS) and synonymous SNPS (sSNPs) (Gutacker *et al.* 2002). Nonsynonymous SNPs result in amino acid replacements, and may be followed by evolutionary selection (Gutacker *et al.* 2002). Alternatively, sSNPs do not alter the structure of the encoded protein and are in most cases evolutionarily neutral or nearly so. Both the G619A and T529C SNPs (described in this study) are nsSNPs, as they result in an amino acid change. For evolutionary studies, sSNPs are more widely used as they are functionally neutral and easy to assay (Gutacker *et al.* 2002). However, some nsSNPs have been used in evolutionary studies, for example nsSNPs in codon 463 of the *katG* gene and codon 95 of the *gyrA* gene allows all *M. tuberculosis* strains to be assigned to one of three principal genetic groups (Sreevatsan *et al.* 1997a). These three groups are referred to as Group 1, which is ancestral to group 2, and similarly group 2, which is ancestral to group 3.

The SNPs identified in this study were not widely dispersed throughout the *M. tuberculosis* strains, and therefore would not make good evolutionary markers. However, they could be used as rapid diagnostic tools to screen for these strain families in new TB cases, which would prove extremely useful if a mini outbreak of TB occurred from isolates belonging to these strain families. These SNPs can be screened for using the PCR-RFLP technique described in this study, which is quick and can screen numerous isolates at one time. Recently, a new SNP has been identified resulting in a G67A amino acid change (Ramaswamy *et al.* 2003). Isolates containing this SNP belong to Group 2 (classified according to their *katG* and *gyrA* nsSNPs), similarly the two SNPs analysed in this study are also only found in strain families classified as group 2. In the future this G67A SNP should be screened for in all the 37 strain families from the Western Cape region, to determine how widely dispersed this SNP is. It would be interesting to know whether this SNP follows a similar pattern to that of other SNPs described in this study. The three SNPS should be analysed together to see whether they could be used to identify an evolutionary pattern that could be missed when analysing each SNP as individual markers. Furthermore, a measurable phenotype may be conferred by one or a combination of these SNPs.

To date only three SNPs have been identified in the *M. tuberculosis nat* gene ((Upton *et al.* 2001b; Ramaswamy *et al.* 2003) and this study) and two of these SNPs (T529C and G619A) are only found in isolates from two distinct *M. tuberculosis* strain families, implying that this gene is

not highly polymorphic, but instead, rather conserved. Not only is the *nat* gene conserved within the *M. tuberculosis* genome, but it is also found in a wide range of other prokaryotes as well as a wide range of eukaryotes (Butcher *et al.* 2002). The conservation of the *nat* gene suggests that this gene's function is important and required in most living organisms.

Drug resistant *M. tuberculosis* strains are becoming more and more prevalent in cases where patients are HIV positive (Grange and Zumla 2002; Mukherjee *et al.* 2004). Once INH resistance occurs within a *M. tuberculosis* isolate, resistance to other anti-TB drugs are more likely to occur, rendering the isolate multidrug resistant (Heep *et al.* 2001; Mukherjee *et al.* 2004; Piatek *et al.* 2000; Grange and Zumla 2002). INH is thought to act as an anti-TB drug by targeting the cell wall synthesis of the bacterium (Slayden and Barry 2000). INH is thought to interfere with the fatty acid synthase (FAS II) system by which long-chain fatty acids (the mycolic acids) are formed (Slayden and Barry 2000). Mycolic acid are the main components of the outer layer of the cell wall, making it an extremely hydrophobic barrier (Barry *et al.* 1998).

The Minimum Inhibitory Concentration (MICs) of Isoniazid (INH) of *M. tuberculosis* strains with one or both SNPs identified in the *nat* gene were higher than that of the H37Rv reference strain, but lower than 0.1μg/ml. *M. tuberculosis* isolates are regarded as resistant to INH if their MIC of INH is greater than 0.1μg/ml. *In vitro* experiments show that the mutant 207R (G619A) NAT protein has less affinity towards INH than the wild type, 207G NAT (Upton *et al.* 2001b). It could be expected that the MIC of INH for isolates coding for 207R NAT would be lower than isolates containing the wild type (207G) NAT, but the opposite is seen. NAT is expressed in fairly low quantities as evident in western blots and Q-PCR data. It is speculated that the affinity of NAT for its endogenous substrate is greater than the affinity of NAT for INH (i.e. NAT may not be spared to acetylate INH), and mutations within the gene may affect the endogenous role, but are unlikely to affect NATs interaction with INH. The increase in the MIC of INH in 177H/207R NAT isolates could be due to a secondary effect. The mutated NAT may inadvertently affect a pathway in which INH is activated, or the expression of another gene that interacts with INH may be altered. The *nat* gene appears to be part of a five gene operon, which is conserved in *M. avium, M. bovis* BCG and *M. tuberculosis* (Payton *et al.* 2001b). Analysis of these genes may be needed to help understand the endogenous role of NAT, and help identify pathways that NAT may be a constituent of. Changes within the *nat* gene in these mycobacterium species may result in the other genes within the operon being up or down-

regulated. These changes may interfere with the INH activation process resulting in a slight rise in the MIC value. However if a mutation was identified that caused the expression of NAT to increase, then it would more likely have a direct effect on INH activity as more NAT will be present within the cell environment. To test this hypothesis, the experiment done previously, where recombinant *M. tuberculosis nat* was overexpressed in *M. smegmatis* resulting in an increase in the MIC of these isolates (Payton *et al.* 1999), could be repeated in *M. tuberculosis*. *M. tuberculosis* isolates could be transformed with an expression vector containing the *M. tuberculosis nat* gene. The MIC values of these isolates, with recombinant *M. tuberculosis* NAT overexpressed, should be compared with that of the wild type isolate.

Given the increase in MIC of INH in *M. tuberculosis* strains with *nat* SNPs, one might expect an increase in drug resistance in these strain families when compared to other strain families. Thirty-two *M. tuberculosis* strain families, from the Western Cape Province of South Africa, were analysed to determine the prevalence of INH resistant isolates within each strain family. These strain families included strain family 3 (containing both the T529C and G619A SNPs) and 28 (containing the G619A SNP). It should be noted that drug resistant isolates are found in many strain families that do not have these polymorphisms. Surprisingly perhaps, no isolates in strain family 3 show INH resistance, and only 19% of the isolates in strain family 28 are INH resistant. When these percentages are related to other strain families, they can be considered fairly low e.g. 70% of isolates from strain family 29 are INH resistant. It has also been shown that the identified polymorphisms within the *nat* gene occur in both INH susceptible (MIC of INH 0.02-0.05μg/ml) and INH resistant isolates (Ramaswamy *et al.* 2003). Therefore, it does not appear that these SNPs have any major effects on INH resistance, however, the response of these isolates may be different *in vivo*. These SNPs may also be part of a subset of genomic changes where each confers some marginal increase in resistance to INH. As of yet it is unknown how these *M. tuberculosis* isolates with the *nat* SNPs would react to INH *in vivo*. Mice could be infected with 177H/ 207R NAT *M. tuberculosis* isolates and control mice with wild-type isolates. The spread of infection could be monitored, as well as the response of infected mice to INH. In this way it could be tested whether the 177H/ 207R NAT *M. tuberculosis* isolates respond differently to INH *in vivo,* or whether the host-pathogen interaction is different with these isolates (e.g. are these isolates better contained by the hosts immune system).

It is known that *M. tuberculosis* cells exposed to INH results in an increase in expression of a range of genes, the majority being those involved in cell wall synthesis (Wilson *et al.* 1999). The proteins AcpM and KasA form a protein complex that is involved in the fatty acid biosynthetic (type II fatty acid synthase [FAS II] system) pathway. These genes are dramatically upregulated when *M. tuberculosis* cells are exposed to INH. Mutations (SNPs) in these genes have been found exclusively in INH resistant *M. tuberculosis* isolates. Cultures of *M. bovis* BCG and *M. tuberculosis* exposed to INH resulted in an increase in expression of NAT. The extent of the increase seems to be dependent on the INH sensitivity of the isolates. *M. tuberculosis* H37Rv is highly sensitive to INH and when exposed to INH, *nat* expression increased 30-fold, whereas, an isolate with both SNPs that has less sensitivity to INH resulted in an 6-fold increase in *nat* expression, when cells were exposed to INH. These isolates have a slightly higher MIC of INH than the reference strain H37Rv, and the effects INH has on the cell environment may be slower than that of the reference strain H37Rv. The apparent affinity for INH of mutated 207R NAT is 10X less than the wild-type 207G NAT (Upton *et al.* 2001b). Therefore 207R NAT may not acetylate INH as it enters the cell as quickly as 207G NAT. This would result in less NAT being pulled from its endogenous role in isolates with the SNPs, and therefore less NAT would be needed to compensate for INH entering the cells. The INH resistant isolate showed a small initial increase and then a slight decrease in *nat* expression when exposed to INH. This may occur as INH may not be activated in these resistant isolates, and therefore no major adverse effect occurs within the cells environment to which NAT expression could respond. INH only once activated may be responsible for the expression of NAT to increased and therefore if INH is not activated no changes in NAT expression would occur. It is not certain whether INH, directly or indirectly, affects the expression of *nat.* INH (or its activated form) may interact directly with the promoter of *M. tuberculosis nat,* or changes within the cell environment, induced by the exposure of cells to INH, may be the cause of the increase in *nat* expression. The difference of *nat* expression between the sensitive and resistant isolates can only be explained once the mechanism by which INH interacts with the *nat* gene is understood. It may be that INH interacts directly with one of the other genes that forms part of the *nat* operon. This reiterates the importance of studying the other genes in the operon, to determine whether they play a role in the expression of NAT.

This study shows that the expression of *nat* is evident during the early growth phase of the growth cycle of both *M. tuberculosis* and *M. bovis* BCG, and reaches maximum expression levels at approximately mid-log phase of the growth cycle of the bacterium. This detection of NAT early on in the growth of the bacterium implies that NAT may be involved in the growth

and development of the bacterium. This work supports previous work, where the *nat* gene was knocked-out of *M. smegmatis* and a noticeable delay in growth was observed, i.e. a much extended lag phase was observed (Payton *et al.* 1999). Similarly, when working with *M. tuberculosis* isolates containing the SNPs, it was noted that these isolates seemed to take much longer to grow and reach the desired cell density, than the other isolates without the SNPs. The growth cycle of these isolates with the SNPs should be compared to the reference strain, as well as other isolates without the SNPs. If NAT in *M. tuberculosis* is involved in the growth and development of the bacterium, then NAT could become a potential candidate for drug development. An inhibitor of NAT could result in controlling the TB infection in some patients by hindering the growth, and therefore spread of the bacterium within these patients. NAT may be involved in the development of the complex cell envelope, and therefore without a functional NAT alterations may occur within the cell envelope, enabling the host's own immune system to successfully fight the infection.

Preliminary studies done on the *nat* knock-out *M. bovis* BCG isolate showed that the knock-out isolate had increased drug sensitivity to gentamycin and hygromycin (Bhakta *et al.* 2004). These results imply that NAT may be involved in cell wall formation or structure, and that cells lacking NAT may have a more permeable cell wall allowing the antibiotics to enter the cell and become effective as killing agents. This experiment could be repeated on *M. tuberculosis* isolates with the SNPs, to see whether these isolates have an increased sensitivity to these drugs when compared to isolates without the SNPs. If NAT is involved in the development of the cell wall and its permeability, then a drug combination of an inhibitor of NAT and an antibody (e.g. gentamycin or hygromycin) that before could not enter through the non-permeable cell wall could be used as anti-TB therapy.

Results reported here show that the NAT protein is found throughout the cell, however, the protein in the cell envelope appears at a higher molecular weight than that of the protein in the cytosol. This higher molecular weight protein, evident in the cell envelope fractions of *M. smegmatis, M. bovis* BCG and *M. tuberculosis,* should be examined. Before this is done the higher molecular weight protein should be verified that it is NAT, this can be done once a pure *nat* knock-out isolate is available. At present *nat* knock-out *M. bovis* BCG isolates do exist, however, due to their slow growth pattern the slightest wild type contamination will out compete the knock-out strains when growing large volumes of cells. Large volumes of cell culture are needed to isolate enough NAT to be visualised on the Western blots. A pure *nat* knock-out *M.*

bovis BCG isolate should be available in the near future. Once a pure *nat* knock-out isolate is available, cells from this isolate should be fractionated, and the proteins from each fraction should be analysed by the Western blot technique. The NAT specific antibody should not detect any protein in any of the fractions. No protein detection in the cell wall fraction will prove that the higher molecular weight protein in the cell envelope fractions is indeed NAT. The modification, leading to the increase in molecular weight, will have to be determined in due course. NAT is hydrophilic and in order to move into the cell envelope, which is hydrophobic, some kind of post-translational process is probably needed to make the protein hydrophobic. It is likely that this post-translational process is that of glycosylation. Once the higher molecular weight protein, within the cell wall fraction, is confirmed, this fraction should undergo deglycosylation to confirm that the post-translational process is in fact glycosylation. Many glycoproteins are present on the mycobacterial cell surface, and are available for interaction with the host cells during infection. Protein glycosylation may therefore provide novel drug targets (Cooper *et al.* 2002). Glycosylation is the covalent attachment of glycan moieties to proteins (Dobos *et al.* 1996; Erickson and Hertzberg 1993). The immunodominant antigen Apa (also known as the 45/47kDa complex) is secreted *in vitro* by the *M. tuberculosis* complex (Dubos *et al.* 1953). The core protein of Apa is 28kDa and due to glycosylation this proteins native form is shown to have a molecular weight of 45 and 47kDa. In this study, NAT has shown to be two sizes that of 32kDa and that of 45kDa (as appears in cell envelope). It has been shown that the deglycosylated form of Apa, as an antigen, is 10-fold less active than its native form in eliciting delayed-type hypersensitivity reactions in guinea pigs immunised with BCG (Romain *et al.* 1999). If the NAT protein is required in the cell envelope, it would explain why the expression of *nat* is required early on during the growth cycle of the bacterium and then reaching maximum levels at exponential phase of growth (when the bacterium is dividing more rapidly). As the bacterium's metabolism slows down and division is not so frequent, NAT is no longer required in high quantities and so the expression of *nat* would be expected to decrease. If NAT is a surface protein of the cell envelope or excreted (like Apa), it may become a potential candidate for vaccine development against TB infection. *M. tuberculosis* proteins associated with cell surface or secreted into the culture medium have been shown to be modified by glycan moieties. These proteins in *M. tuberculosis* are regarded as good candidates for new TB vaccines.

Once the higher molecular weight protein in the cell wall fraction is verified to be that of NAT, the proteins of the *M. tuberculosis* isolates with the SNPs could be fractionated and viewed using the Western blot technique. The higher molecular weight protein, seen in the cell envelope

fraction, should be analysed for any size shifts. Any conformational changes of the protein, caused by the SNPs, may affect any post-translational process enabling the NAT protein to move into the cell envelope. These changes may be evident as size changes on the Western blot. The G619A SNP is known to result in a conformational change of the NAT protein in *M. smegmatis* (Kawamura *et al.* 2003). It is not known whether the T529C SNP also results in a conformational change of the NAT protein. As done for the G619A SNP, using site directed mutagenesis the T529C SNP can be placed into the genome of *M. smegmatis.* The NAT protein of *M. smegmatis* has a tyrosine at amino acid 177 like *M. tuberculosis* NAT. The 3D structure of this mutant protein (Y177H) can then be analysed. Thus far, recombinant NAT of both *M. bovis* BCG and *M. tuberculosis* has not generated sufficient soluble amounts for 3D structural studies, and so *M. smegmatis* recombinant NAT has been used instead (Kawamura *et al.* 2003).

NATs are found from prokaryotes through to eukaryotes (Payton *et al.* 2001b), and the full range of its functional activities is a mystery. It is known to act as a detoxifier of certain substances (Payton *et al.* 2001b), but this may not be its only importance within the organism. The evidence shown in this study suggests other functions. NAT appears to be found in the cell envelope of certain mycobacterium, suggesting that NAT may play a role in the development of the cell envelope. The significance of NAT in the cell envelope should be further investigated to determine how important NAT is to the development of these cells, and to decide whether NAT could be a possible drug target or used for vaccine development for future TB therapy.

7.1. Conclusion

In conclusion, it can be stated that NAT in *M. smegmatis, M. bovis* BCG and *M. tuberculosis* is expressed throughout the cell, however, in the cell envelope NAT appears to have an approximate 15kDa increase in size to that of the cytosolic NAT. It has been hypothesised that a post-translation process (e.g.glycosylation) needs to occur to convert the hydrophilic NAT into a hydrophobic form, so that it is able to transfer into the cell envelope. The expression of *nat* in both *M. bovis* BCG and *M. tuberculosis* was detected early in the growth stage of the bacterium, and maximum expression levels were reached when the cells were in approximately mid-log phase of their growth stage. These results imply that NAT is needed during cell division and growth of these mycobacterium cells.

M. tuberculosis nat is polymorphic, and a new T529C SNP was identified. This SNP results in a Tyr to His amino acid change at position 177, and resulted in the addition of a Bsg1 restriction

site. Isolates could be rapidly screened for this SNP by using PCR-RFLP. This SNP, as well as the previously identified G619A SNP, were screened for in 37 strain families found in the Western Cape region of South Africa. The G619A SNP was found to be restricted to two strain families, family 3 and family 28. One unique isolate 1936 also had this G619A SNP. The T529C SNP was found to be restricted to one strain family, family 3. It has been hypothesised that the G619A SNP came from a progenitor, and the second T529C SNP occurred at a later stage. These SNPs are not widely dispersed throughout the strain families, and will probably not be good evolutionary markers. However, if there is an outbreak of TB cases caused by isolates from either of these two strain families, using these SNPs and the PCR-RFLP techniques, the outbreak could be easily monitored.

The INH MICs of these isolates with the SNPs, showed a higher MIC than that of the reference strain H37Rv. However, these INH MIC values were still lower than 0.1μg/ml, the concentration at which isolates are considered as INH resistant. The percentage of INH resistant isolates in these strain families were also relatively low when compared to the other strain families analysed in this study. These SNPs may not play a significant role in INH resistance, but this does not mean that there is no interaction between the expression of *nat* and INH. *M. bovis* BCG and *M. tuberculosis* cells exposed to INH showed an increase in *nat* expression. The increase in *nat* expression in *M. tuberculosis* cells varied between isolates. It appeared that isolates more sensitive to INH showed a greater degree of increase in expression than isolates less sensitive to INH.

Finally, the two identified SNPs have been inserted into the p2NIL vector, which now needs to be ligated to the marker cassette of the pGoal 19 vector to complete the suicide delivery vector. This step will be the future work of Prof. Sim's lab. The final construct will be used to transform wild type *M. bovis* BCG and *M. tuberculosis* H37Rv isolates, so that the *nat* gene contains the newly identified SNPs. These transformed isolates would be monitored for changes in growth and other phenotypic characters. These results would aid the on going work been done in Prof. Sim's lab, Department of Pharmacology, Oxford University.

All the results produced in this study, and results published in previous studies, suggest that NAT may be involved in cell growth and development of the mycobacteria complex. It appears that NAT may be involved in the synthesis of the cell envelope. The cell envelope of *M. tuberculosis*

is highly specialised and is the reason the bacterium is such a successful pathogen. If NAT is involved in the growth and development of the bacterium, then NAT would be an excellent candidate for drug development against TB.

Reference List

1. Adams A, Petit JF, Weitzerbin-Falzpan J, Sinay P, Thomas DW, Lederer E (1969) Mass spectrometric identification of N-glycolymuramic acid, a constituent of *Mycobacterium smegmatis* walls. *Febs Letters* **4,** 87-92.

2. Agundez JA, Ladero JM, Olivera M, Abildua R, Roman JM, Benitez J (1995) Genetic analysis of the Arylamine-N-acetyltransferase polymorphic in breast cancer patients. *Oncology* **52,** 7-11.

3. Alland D, Kramnik I, Weisbrod TR, Otsubo L, Cerny R, Miller LP, Jacobs RFJr, Bloom BR (1998) Identification of differentially expressed mRNA in prokaryotic organism by customized amplification libraries (DECAL): The effect of INH on gene expression in *Mycobacterium tuberculosis*. *Proc.Natl.Acad.Sci.USA* **95,** 13227-13232.

4. Allen BW (1969) *Mycobaterium tuberculosis* strain H37Rv. *J.Med.Lab.Techn.* **26,** 389-390.

5. Ames BN, Lee FD, Durstan WE (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc.Natl.Acad.Sci.USA* **70,** 782-786.

6. Applied Biosystems. User Bulletin #2. http://www.appliedbiosytems.com . 2004.

7. Auravita. Co-enzyme A, Xynergy Health products. http://www.auravita.com/products/AURA/XYHE11830.asp . 2004.

8. Azad AK, Sirokova TD, Fernandes ND, Kolattukudy PE (1997) Gene knockout reveals a novel gene cluster for synthesis of cell wall lipids unique to pathogenic mycobacteria. *J.Biol.Chem.* **272,** 16741-16745.

9. Baggott, J. Acetyl CoA - The centre of Lipid Metabolism: Function of CoA. http://medlib.med.utah.edu/NetBiochem/FattyAcids/2_5.htm1 . 2003.

10. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Sun Um K, Wilson T, Collins D, de Lisle G, Jacobs WR (1994) *inhA*, a Gene Encoding a Target for isoniazid and Ethionamide in *Mycobacterium tuberculosis*. *Science* **263,** 227-232.

11. Bardou F, Raynaud C, Ramos C, Laneelle MA, Laneelle G (1998) Mechanism of isoniazid uptake in *Mycobacterium tuberculosis*. *Microbiology* **144,** 2539-2544.

12. Barksdale L, Kim KS (1977) Mycobacterium. *Bacteriol.Rev.* **41,** 217-372.

13. Barrow WW (1997) Processing of mycobacterial lipids and effects on host responsiveness. *Frontiers in Bioscience* **2,** 387-400.

14. Barry CEI, Lee RE, Mdluli K, Sampson AE, Schroeder BG, Slayden RA, Yuan Y (1998) Mycolic acids: structure, biosynthesis and physiological. *Prog.Lipid Res.* **37,** 143-179.

15. Bass JBJr, Farer LS, Hopewell PC, O'Brien R, Jacobs RFJr, Ruben F, Snider DEJr, Thornton G (1994) Treatment of tuberculosis and tuberculosis infection in adults and children. American thoracic society and the centres for disease control and prevention. *Am.J.Respir.Crit.Care Med.* **149,** 1359-1374.

16. Basso LA, Zheng R, Musser JM, Jacobs RFJ, Blanchard JS (1998) Mechanism of isoniazid resistance in *Mycobacterium tuberculosisi:* enzymatic characterisation of enoyl reductase mutants identified in isoniazid-resistant clinical isolates. *J.Infect.Dis.* **178.**

17. Behr M (2002) BCG-different strains, different vaccines? *The Lancet* **2,** 86-92.

18. Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ, Besra GS (1997) Role of the major antigen of *Mycobacterium Tuberculosis* in cell wall biosynthesis. *Science* **276,** 1420-1422.

19. Bernstein J, Lott W, Steinberg B, Yale H (1952) Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds. *Am.Rev.Tuberc.* **65,** 357-364.

20. Bhakta S, Besra GS, Upton A, Parish T, Sholto-Douglas-Vernon C, Gibson KJC, Knutton S, Gordan S, Da-Silva RP, Anderton MC, Sim E (2004) Arylamine N-acetyltransferase is required for synthesis of mycolic acids and complex lipids in *Mycobacterium bovis* BCG and represents a novel drug target. *The J.of Exp.Med.* **199,** 1191-1999.

21. Bhakta S, Upton A, Payton M, Pompeo F, Da-Silva RP, Besra GS, Knutton S, Busby A, Parish T, Sim E (2003a) Keystone Meeting on Tuberculosis, New Mexico, January.

22. Bhakta S, Upton A, Payton M, Pompeo F, Mo MSN, Da-Silva RP, Gordan S, Besra GS, Parish T, Sim E (2003b) Modification of *nat*, the gene encoding Arylmine N-acetyltransferase which acetylates isoniazid, affects the growth, morphology, cell wall biosynthesis and infectivity of Mycobacteria. In *Structural Biology in Drug Metabolism and Drug Discovery, Programme and Abstracts*, poster abstracts 8, p. 6, AstraZeneca R & D Charnwood, Loughborough, UK: Biochemical Society Focused Meeting.

23. Bloom BR, Murray C (1992) Tuberculosis: commentary on a reemergent killer. *Science* **257,** 1055-1067.

24. Blum M, Grant DM, McBride W, Hein M, Meyer UA (1990) Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA and Cell Biology* **9,** 193-203.

25. Bowen WP, Carey JE, Miah A, McMurray HF, Munday PW, James RS, Coleman RA, Brown AM (1999) Measurement of cytochrome P450 gene induction in human hepatocytres using quantitative real-time reverse transcriptase-polymerase chain reaction. *Drug Metab Dispos* **28,** 781-787.

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

27. Brennan PJ (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis* **83,** 91-97.

28. Brennan PJ, Draper P (1994) 'Tuberculosis: Pathogenesis,Protection, and Control.' (Am. Soc. Microbiol.: Washington DC)

29. Brockmoller J, Cascorbi I, Kerb R, Roots I (1996) Combine analysis of inherited polymorphism in arylamine N-acetyltranferase 2, glutathione S-transferase M1 & T1, microsomal expoxide hydrolase, and cytochrome P450 enzymes as modulators of bladder cancer risk. *Cancer Res* **56,** 3915-3925.

30. Brooke E, Davies SG, Mulvaney AW, Pompeo F, Sim E, Vickers RJ (2003a) The approach to identifying novel substrate of bacterial arylamine N-acetyltransferases. *Bioorg.Chem.* **11,** 1227-1234.

31. Brooke E, Sholto-Douglas-Vernon C, Sim E (2003b) The role of arylamine N-acetyltransferase in bacteria. In 'Recent Research developments in bacteriology'. (Transworld Research Network: Kerala,India)

32. Brown TA (1990) 'Gene Cloning: An introduction.' (Chapman & Hall: London)

33. Butcher NJ, Boukouvala S, Minchin RF (2002) Pharmacogenetics of the arylamine N-acetyltransferases. *The Pharmacogenomics Journal* **2,** 42.

34. Cangelosi CA, Brabant WH (1997) Depletion of pre-16S rRNA in starved Escherichia coli cells. *J Bacteriol* **179,** 4457-4463.

35. Chan J, Fujuwara T, Brennan P, McNiel M, Turco SJ, Sibille JC, Snapper M, Aisen P, Bloom BR (1989) Microbial glycolipids: possible virulence factors that scavenge oxygen radicals. *Proc.Natl.Acad.Sci.USA* **86,** 2453-2457.

36. Chaves F, Yang Z, El Hajj H, Alonso M, Burman WJ, Eisenach KD, Dronda F, Bates JH, Cave MD (1996) Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. *J.Clin.Microbiol.* **34,** 1118-1123.

37. Clark-Curtiss JE (1990) 'Molecular Biology in Mycobacteria.' (Academic Press Ltd: London)

38. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Barrel B (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393,** 537-544.

39. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG (2001) Massive gene decay in the leprosy bacillus. *Nature* **409,** 1007-1011.

40. Cooper HN, Gurcha SS, Nigou J, Brennan PJ, Belisle JT, Besra GS, Young D (2002) Characterisation of mycobacterial protein glycosyltransferase activity using synthetic peptide acceptors in a cell-free assay. *Glycobiology* **12,** 427-434.

41. Daffe M, Draper P (1998) The envelope layers of mycobacterium with reference to their pathogenicity. *Adv.Microb.Physiol.* **39,** 131-203.

42. Daffe M, Laneelle MA (1988) Distribution of phthiocerol diester, phenolic mycosides and related compounds in mycobacteria. *J.Gen.Microbiol.* **134,** 2049-2055.

43. Daffe M, Laneelle MA (1989) Diglycosyl phenol phthiocerol diester of *Mycobacterium leprae. Biochim.Biophys.Acta* **1002,** 333-337.

44. Deguchi T (1992) Physiology and molecular biology of arylamine N-acetyltransferases. *Biomed.Res.* **13,** 231-242.

45. Deguchi T, Mashimo M, Suzuki T (1990) Correlation between acetylator phenotypes and genotypes of polymorphic arylamine N-acetyltransferase in human liver. Journal of Biological. *Chemistry* **265,** 12757-12760.

46. Delomenie C, Fouix S, Longuemaux S, Brahimi N, Bizet C, picard B, Denamur E, Dupret J-M (2001) Identification and Functional Characterization of Arylamine N-Acetyltransferases in Eubacteria: Evidence for Highly Selective Acetylation of 5-Aminosalicylic Acid. *J.Bacteriol.* **183,** 3417-3427.

47. Dobos KM, Khoo KH, Swiderek KM, Brennan PJ, Belisle JT (1996) Definition of the full extent of glycosylation of the 45-kilodalton glyco-protein of *Mycobacterium tuberculosis. J Bacteriol* **178,** 2498-2506.

48. Draper P (1974) The mycoside capsule of *Mycobacterium avium. J.Gen.Microbiol.* **83,** 431-433.

49. Dubos RJ, Pierce CH, Schaefer WB (1953) *J.Expr.Med.* **97,** 207-220.

50. Eisen MB, Brown PO (1999) DNA arrays for analysis of gene expression. *Methods Enzymol.* **303,** 179-205.

51. Ellard GA, Gammon PT (1976) Pharmacokenetics of isoniazid metabolism in man. *Biopharm.* **4,** 83-113.

52. Ellard GA, Gammon PT, Wallace SM (1972) The determination of Isoniazid and its metabolites acetylisoniazid, monoacetylhydrazine, diacetylhydrazine, isonictinic acid and isonicotinylglycine in serum and urine. *Biochem.J.* **126,** 449-458.

53. Erickson PR, Hertzberg MC (1993) Evidence of covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. *J.Biol.Chem.* **269,** 23780-23783.

54. Evans DAP, Manley KA, McKuisick VA (1960) Genetic control of isoniazid acetylation in man. *Brit.Med.J.* **2,** 485-491.

55. Fitzgerald JR, Musser JM (2001) Evolutionary genomics of pathogenic bacteria. *Trends Microbiol* **9,** 553.

56. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidmann J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs RFJr, Venter JC, Fraser CM (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* **184,** 5479-5490.

57. Fretland AJ, Doll MA, Zhu Y, Smith L, Leff MA, Hein DW (2002) Effect of nucletide substitutions in N-acetyltransferase-1 on N-acetylation (deactivation) and O-acetylation (activation) of arylamine cacinogens: implications for cancer predisposition. *Cancer Detect.Prev.* **26,** 10-14.

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

59. Garbe TR, Hibler NS, Deretic V (1996) soniazid induces expression of the antigen 85 complex in Mycobacterium tuberculosis. *Antimicrob.Agents Chem.* **40,** 1754-1756.

60. Gelfand, D. H, Holland, P. M., Saiki, R. K., and Watson, R. U.S.Patent [5210015]. 1993.

61. Glickman MS, Cahill SM, Jacobs WR (2001) The *Mycobacterium tuberculosis cmaA2* gene encodes a mycolic acid *trans*-cyclopropane synthetase. *Journal of Biological Chemistry* **276,** 2228-2233.

62. Gonzalez-y-Merchand JA, Colston MJ, Cox RA (1998) Roles of multiple promoters in transcription of ribosomal DNA: effects of growth conditions on precursor rRNA synthesis in mycobacteria. *J.Bacteriol.* **180,** 5756-5761.

63. Goodsell D, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of Autodock. *J.Mol.Rognit.* **9,** 1-5.

64. Grange JM, Zumla A (2002) The global emergency of tuberculosis: what is the cause? *J R Soc Health* **122,** 78-81.

65. Grant DM, Lottspeich F, Meyer UA (1989) Evidence for two closely related isozymes of arylamine N-acetyltransferase in human liver. *Febs.Letters* **244,** 203-207.

66. Gutacker MM, Smoot JC, Lux Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, Graviss EA, Shashkina E, Kreiswirth BN, Musser J (2002) Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: Resolution of genetic relationships among closely related microbial strains. *Genetics* **162,** 1533-1543.

67. Hansen MC, Nielsen AK, Molin S, Hammer K, Kilstrup M (2001) Changes in rRNA levels during stress invalidates results from mRNA blotting: fluorescence in situ rRNA hybridization permits renormalization for estimation of cellular mRNA levels. *J.Bacteriol.* **183,** 4747-4751.

68. Heep M, Brandstatter B, Rieger U, Lehn N, Richter E, Rusch-Gerdes S (2001) Frequency of rpoB mutations inside and outside the cluster I region in rifampin-resistant clinical *Mycobacterium tuberculosis* isolates. *J.Clin.Microbiol.* **39,** 107-110.

69. Hein DW, Doll MA, Rustan TD, Gray K, Feng Y, Ferguson RJ, Grant DM (1993) Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis* **14,** 1633-8.

70. Hellyer TJ, DesJarden LE, Hehman GL, Cave MD, Eisenach KD (1999) Quantitative analysis of mRNA as a marker for viability of *M. tuberculosis*. *J.Clin.Microbiol.* **37,** 290-295.

71. Herrington MB (1994) Measurement of the uptake of radioactive para-amino benzoic acid monitors folate biosynthesis in *Escherichia coli* K-12. *Anal Biochem* **216,** 427-430.

72. Hickman D, Pope J, Patil SD, Fakis G, Smelt V, Stanley LA, Payton M, Unadkat JD, Sim E (1998) Expression of arylamine N-acetyltransferase in human intestine. *Gut* **42,** 402-9.

73. Higuchi R, Dollinger G, Walsh PS, Griffith R (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* **10,** 413-417.

74. Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR: Real time monitoring of DNA amplification reactions. *Biotechnology* **11,** 1026-1030.

75. Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **88,** 7276-7280.

76. Hook VY, Toneff T, Aaron W, Yasothornsrikul S, Bundy R, Reisine T (2002) Beta-amyloid peptide in regulated secretory vesicles of chromaffin cells: evidence for multiple cysteine proteolytic activities in distinct pathways for beta-secretase activity in chromaffin vesicles. *J.Neurochem.* **81,** 237-256.

77. Isenberg HD (1992) 'Clinical Microbiology Procedures Handbook, volume 1.' (American society for microbiology: Washington, D.C.)

78. Jenne JW (1965) Partial purification and properties of the isoniazid transacetylase in human liver: its relationship to the acetylation of p-aminosalicylic acid. *J.Clin.Invest.* **44,** 1992-2002.

79. Kawamura A, Sandy J, Upton A, Noble M, Sim E (2003) Structural investigation of mutant *Mycobacterium smegmatis* arylamine N-acetyltransferase: a model for a naturally occurring functional polymorphism in *Mycobacterium tuberculosis* arylamine N-acetyltransferase. *Protein Expression and Purification* **27,** 75-84.

80. Kelly CL, Rouse DA, Morris SL (1997) Analysis of *aphC* gene mutations in isoniazid resistant clinical isolates of *Mycobacterium tuberculosis*. *Antimicrob.Agents.Chem.* **41,** 2057-2058.

81. Klabunde T, Sharma S, Telenti A, Jacobs RFJr, Sacchettini JC (1998) Crystal structure of GyrA intein from *Mycobacterium Xenopi* reveals structural basis of protein splicing. *Nat.Struct.Biol.* **5,** 31-36.

82. Kubica GP, Kent PT (1985) 'Public health mycobacteriology. A guide for the level III laboratory.' (U.S. Department of health and human services, C.D.C.: Alanta, Georgia)

83. Larsen MH, Vilcheze C, Kremer L, Besra GS, Parsons L, Salfinger M, Heifets L, Hazbon MH, Alland D, Sacchettini JC, Jacobs RFJr (2002) Overexpression of *inhA*, but not *kasA*, confers resistance to isoniazid and ethionamide in *Mycobacterium smegmatis, M. bovis* and *M. tuberculosis*. *Mol.Microbiol.* **46,** 453-466.

84. Lawson SEM, Gillenwater KA, Kalargi S, Lescano AG, Du Quella G, Cummings K, Cabrera L, Torres C, Gilman RH (2003) A prospective study of bacillus Calmette-Guerin scar fromation and tuberculin skin test reactivity in infants in Lim, Peru. *Pharmac.Ther.* **112,** e298.

85. Lederer E, Adams A, Ciorbaru R, Petit JF, Wietzerbin-Falszpan J (1975) Cell-walls of mycobacteria ans related organisms - chemistry and immunostimulant properties. *Mol.Cell.Biochem* **7,** 87-104.

86. Lee AS, Teo AS, Wong SY (2001) Novel mutations in ndh in isoniazid-resistant Mycobacterium tuberculosis isolates. *Antimicrob.Agents.Chem.* **45,** 2157-2159.

87. Lee LG, Connell CR, Bloch W (1993) Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acid Research* **21,** 3761-3766.

88. Lee RE, Lim IH, Tang LL, Telenti A, Wong SY (1999) Contribution of kasA analysis to detection of isoniazid-resistant Mycobacterium tuberculosis in Singapore. *Antimicrob.Agents.Chem.* **43,** 2087-2089.

89. Lindsten, J. and Ringerte, N. The nobel prise in physiology of medicine 1901-200. www.nobel.se/medicine/articles/linsten-ringerte-rev/ . 2003.

90. Livak KJ, Flood SJA, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridisation. *PCR Methods and Applications* **4,** 357-362.

91. Marrakchi H, Laneelle G, Quemard A (2000) *InhA*, a target of the anti-tiberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system FAS-II. *Microbiology* **146,** 289-296.

92. Master S, Zahrt TC, Song J, Deretic V (2001) Mapping of Mycobacterium tuberculosis katG promoters and their differential expression in infected macrophages. *J.Bacteriol.* **183,** 4033-4039.

93. McAdam RA, Hermans PW, van Soolingen D, Zainuddin ZF, Catty D, van Embden JD (1990) Characterisation of a *M. tuberculosis* insertion sequence belonging to the IS3 family. *Mol.Microbiol.* **4,** 1607-1613.

94. McCaughan-Carucci, J. Ribonuclease protection assay. http:/micro.nwfscc.noaa.gov/protocols/RNAseProtect.htm1 . 2003.

95. McNiel MR, Brennan PJ (1991) Structure, function and biogenesis of the cell envelope of mycobacteria in relation to bacterial physiology, pathogenesis and drug resistance; some thoughts and possibilities arising from recent structural information. *Res.Microbiol.* **142,** 451-463.

96. Mdluli K, Slayden RA, Zhu Y, Ramaswamy S, Pan X, Mead D, Crane DD, Musser J, Barry CEI (1998) Inhibition of a *Mycobacterium tuberculosis* ß-ketoacyl ACP synthase by isoniazid. *Science* **280,** 1607-1610.

97. Mehra V, Brennan PJ, Rada R, Convit J, Bloom BR (1984) Lymphocyte suppression in leprosy induced by a unique *M. leprae* glycolipid. *Nature* **308,** 194-196.

98. Meyers PR, Bourn WR, Steyn LM, van Helden PD, Beyers AD, Brown GD (1998) Novel Methods for Rapid Measurement of Growth of Mycobacteria in Detergent-Free Media. *Journal of Clinical Microbiology* **36,** 2752-2754.

99. Mills JL, Conley MR (1996) Folic acid to prevent neural tube defects: scientific advances and public health issues. *Cur.Opin.Gynecol* **8,** 394-397.

100. Minchin RF (1995) Acetylation of para-aminobenzoylglutamate, a folate catabolite, by recombinant human NAT and U937 cells. *Biochem.J.* **307,** 1-3.

101. Minchin RF, Reeves PT, Teitel CH, McManus ME, Mojarrabi B, Ilett KF, Kadlubar FF (1992) N-and O-acetylation of aromatic and heterocyclic amine carcinogens by human monomorphic and polymorphic acetyltransferases expressed in COS-1 cells. *Biochem Biophys Res Commun* **185,** 839-44.

102. Minnikin DE (1982) 'The Biology of Mycobacteria.' (Academic Press: New York)

103. Minnikin DE (1991) Chemical principles in the organisation of lipid components in the mycobaterial cell envelope. *Res.Microbiol.* **142,** 423-427.

104. Minnikin, D. E. A chemical approach to the conquest of tuberculosis and leprosy. http://www.ncl.ac.uk/chemistry/research/dem.htm1 . 2001.

105. Minnikin DE, Kremer L, Dover LG, Besra GS (2002) The methyl-branched fortifications of *Mycobacterium tuberculosis*. *Chemistry & Biology* **9,** 545-553.

106. Mortlemans K, Zeiger E (2000) The Ames Salmonella/microsome mutagenicity assay. *Mut.Res.* **455,** 29-33.

107. Mukherjee JS, Rich Ml, Socci AR, Joseph JK, Viru FA, Shin SS, Furin JJ, Becerra MC, Barry DJ, Kim JY, Bayona J, Farmer P, Fawzi MCS, Seung KJ (2004) Programmes and principles in treatment of multidrug resistant tuberculosis. *The Lancet* **363,** 474-481.

108. Murray M, Alland D (2002) Methodological problems in the molecular epidemiology of tuberculosis. *Am.J.of Epidemiology* **155,** 565-572.

109. Mushtaq A, Payton M, Sim E (2002) the COOH terminus of arylamine N-acetyltransferase from Salmonella typhimurium controls enzymatic activity. *J.Biol.Chem.* **277,** 12175-12181.

110. Musser JM (1996) Molecular population genetic analysis of emerged bacterial pathogens: selected insights. *Emerg.Infect.Dis.* **2,** 1-17.

111. Musser M, Kapur V, Williams D, Kreiswirth B, van Soolingen D, Van Embden J (1996) Characterisation of the *catalase-peroxidase* gene (*katG*) and inhA locus in isoniazid resistant and susceptible strains of

Mycobacterium tuberculosis by automated DNA sequencing: restricted array of mutations associated with drug resistance. *J.Infect.Dis.* **173,** 196-202.

112. Pablos-Mendez A, Raviglione MC, Laszlo A, Binkin N, Rieder HL, Bustreo F (1998) Global surveillance for antituberculosis-drug resistance, 1994-1997. World Health Organisation-International Union against Tuberculosis and Lung Disease working Group on Anti-Tuberculosis Drug Resistance Surveillance. *N.Engl.J.Med.* **338,** 1641-1649.

113. Palandez A, Gultek D, Erdem E, Kayalp N (2003) Low level compliance with tuberculosis treatment in children monitoring urine tests. *Ann Trop.Pediatr.* **23,** 47-50.

114. Parish T, Stoker N (1998) 'Mycobacterial Protocols.' (Human Press: New Jersey)

115. Parish T, Stoker N (2000) Use of flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology* **146,** 1969-1975.

116. Parkin DP, Vandenplas S, Botha FJ, Vandenplas ML, Seifart HI, van Helden PD, van der Walt BJ, Donald PR, van Jaarsveld PP (1997) Trimodality of isoniazid elimination: phenotype and genotype in patients with tuberculosis. *American Journal of Respiratory and Critical Care Medicine* **155,** 1717-22.

117. Parthasarathy A (2003) Controversies in BCG immunisation. *Indian J.Pediatr.* **70,** 585-586.

118. Payton M, Auty R, Delgoda R, Everett M, Sim E (1999) Cloning and characterization of arylamine N-acetyltransferase genes from *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: increased expression results in isoniazid resistance. *J Bacteriol* **181,** 1343-7.

119. Payton M, Gifford C, Schartau P, Hagaemier C, Mushtaq A, Lucas S, Pinter K, Sim E (2001a) Evidence towards the role of arylamine N-acetyltransferase in *Mycobacterium smegmatis* and development of a specific antiserum against the homologous enzyme of *Mycobacterium tuberculosis*. *Microbiology* **147,** 3295-3302.

120. Payton M, Mushtaq A, Yu TW, Wu L-J, Sinclair J, Sim E (2001b) Eubacterial arylamine N-acetyltransferases - identification and comparison of 18 members of the protein family with conserved active site cysteine, histidine and aspartate residues. *Microbiology* **147,** 1137-1147.

121. Petit JF, Adam A, Weitzerbin-Falzpan J, Lederer E, Ghuysen JM (1969) Chemical structure of the cell wall of Mycobacterium smegmatis. Isolation and partial characterisation of peptidoglycan. *Biochem Biophys Res Commun* **35,** 478-483.

122. Piatek AS, Telenti A, Murray MR, El-Hajj H, Jacobs WRJ, Kramer FR (2000) Genotypic analysis of Mycobacterium tuberculosis in two distinct

populations using molecular beacons: implications for rapid susceptibility testing. *Antimicrob.Agents.Chem.* **44,** 103-110.

123. Pompeo F, Brooke E, Kawamura A, Mushtaq A, Sim E (2002a) The pharmacogenetics of NAT: structural aspects. *Pharmacogenetics* **3,** 30.

124. Pompeo F, Mushtaq A, Sim E (2002b) Expression and purification of the rifamycin amide synthase, RifF, an enzyme homologous to the prokaryotic arylamine N-acetyltransferases. *Prot.Exp.Purific.* **24,** 151.

125. Prescott LM, Harley JP, Klein DA (1999) 'Microbiology.' (McGraw-Hill: New York)

126. Quemard A, Sacchettini J, Dessen A, Vilcheze C, Bittman R, Jacobs RFJ, Blanchard JS (1995) Enzymatic characterisation of the target for isoniazid in *Mycobacterium tuberculosis*. *Biochemistry* **34,** 8235-8241.

127. Raczniak G, Ibba M, Soll D (2001) Genomic-based identification of targets in pathogenic bacteria for potential therapeutic and diagnostic use. *Toxicicology* **160,** 181-189.

128. Ramaswamy S, Musser J (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tubercle Lung Dis.* **79,** 3-29.

129. Ramaswamy S, Reich R, Dou S-H, Jasperse L, Pan X, Wanger A, Quitugua T, Gravis EA (2003) Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacteria tuberculosis*. *Antimicrob.Agents.Chem.* **47,** 1241-1250.

130. Rascher A, Hu Z, Viswanathan N, Schirmer A, Reid R, Nierman WC, Lewis M, Hutchinson CR (2003) Cloning and characterisation of a gene cluster for geldanamycin production in Stretomyces hygroscopicus NRRL 3602. *FEMS Micobiol.Letts.* **218,** 223-230.

131. Rastogi N (1991) Recent observations concerning structure and function relationships in the mycobacterial cell envelope: elaboration of a model in terms of mycobacterial pathogenicity, virulence and drug resistance. *Res.Microbiol.* **142,** 464-476.

132. Ratledge C (1982) 'The biology of the Mycobacteria.' (Academic Press: New York and London)

133. Raviglione MC, Snider DE, Kochi A (1995) Global epidemiology of tuberculosis: morbidity and mortality of a worldwide epidemic. *J.A.M.A.* **273,** 220-226.

134. Riddle B, Jencks WP (1971) Acetyl-coenzyme A: arylamine N-acetyltransferase. Role of the acetyl-enzyme intermediate and the effects of substituents on the rate. *J Biol Chem* **246,** 3250-8.

135. Riley M, Labedan B (1996) 'Escherichia coli and Salmonella.' (ASM: Wahington)

136. Rodrigues-Lima F, Delomenie C, Goodfellow GH, Grant DM, Dupret JM (2001) Homology modelling and structural analysis of human arylamine N-acetyltransferase NAT1: evidence for the conservation of a cysteine protease catalytic domain and an active-site loop. *Biochem.J.* **356,** 327-334.

137. Rodrigues-Lima F, Dupret J-M (2002) In silico sequence analysis of arylamine N-acetyltransferases: evidence for an absence of lateral gene transfer from bacteria to vertebrates and first description of paralogs in bacteria. *Biochem Biophys Res Commun* **293,** 783-792.

138. Romain F, Horn C, Pescher P, Namane A, Riviere M, Puzo G, Barzu O, Marchal G (1999) Deglycosylation of the 45/47-kilodalton antigen complex of Mycobacterium tuberculosis decreases its capacity to elicit in vivo or in vitro cellular immune responses. *Infect.Immun.* 5567-5572.

139. Rosenthal PJ, Sijwali PS, Singh A, Shenai BR (2002) Cysteine protease of malaria parasite:targets for chemotherapy. *Curr.Pharm.Des.* **8,** 1659-1672.

140. Saito K, Yamazoe Y, Kamataki T, Kato R (1983) Mechanism of activation of proximate mutagens in Ames' tester strains: the acetyl-CoA dependent enzyme in *Samonella typhimurium* TA98 deficient in TA98/1,8-DNPG catalyses DNA-binding as the cause of mutagenicity. *Biochem Biophys Res Commun* **116,** 141-147.

141. Salzberg SL, White O, Peterson J, Eisen JA (2001) Microbial genes in humans genome: lateral transfer or gene loss. *Science* **292,** 1903-1906.

142. Sampson SL, Warren R, Richardson M, van der Spuy GD, van Helden PD (1999) Disruption of coding regions by IS*6110* insertion in *Mycobacterium tuberculosis*. *Tubercle and Lung Disease* **79,** 349-359.

143. Sandy J, Mushtaq A, Kawamura A, Sinclair J, Sim E (2002) The structure of arylamine N-acetyltransferase from *Mycobacterium smegmatis* - an enzyme which inactivates the anti-tubercular drug isoniazid. *J.Mol.Biol.* **318,** 1071-1083.

144. Schoolnik G (2002) Functional and comparative genomics of pathogenic bacterias. *Curr.Opin.Microbiol.* **5,** 20-26.

145. Sequella. Sequella Global tuberculosis foundation. A resource fot the TB research community worldwide. www.sequellafoundation.org/fact.asp . 2003.

146. Sherman DRK, Mdluli K, Hickey MJ, Arain TM, Morris SL, Barry CE, III, Stover CK (1996) Compensatory *aphC* gene expression in isoniazid resistant *Mycobaterium tuberculosis*. *Science* **272,** 1641-1643.

147. Shi T, Fredrickson JK, Balkwill DL (2001) Biodegradation of polycyclic aromatic hydrocarbons by sphingomonas strains isolated from the terrestrial subsurface. *J.Ind.Microbiol.Biotechnol.* **26,** 283-289.

148. Sim E, Payton M, Noble M, Minchin R (2000) An update on genetic, structural and functional studies of arylamine *N*-acetyltransferases in eucaryotes and procaryotes. *Hum.Mol.Genet.* **9,** 2435-41.

149. Sim E, Pinter K, Mushtaq A, Upton A, Bhakta S, Noble M (2003) Arylamine N-acetyltransferases: a pharmacogenomic approach to drug metabolism and endogenous function. *Biochemical Society Transactions* **31,** 615-619.

150. Sinclair J, Sim E (1997) A fragment consisting of the first 204 amino-terminal amino acids of human arylamine *N*-acetyl transferase one (NAT1) and the first transacetylation step of catalysis. *Biochem.Pharmacol.* **53,** 11-16.

151. Sinclair JC, Sandy J, Delgoda R, Sim E, Noble ME (2000) Structure of arylamine N-acetyltransferase reveals a catalytic triad. *Nat.Struct.Biol.* **7,** 560-4.

152. Slayden RA, Barry CE (2000) The genetics and biochemistry of isoniazid resistance in *Mycobacteria tuberculosis*. *Microbes and Infection* **2,** 659-669.

153. Smelt VA, Upton A, Adjaye J, Payton MA, Boukouvala S, Johnson N, Mardon HJ, Sim E (2000) Expression of arylamine N-acetyltransferases in pre-term placentas and in human pre-implantation embryos. *Hum Mol Genet* **9,** 1101-7.

154. Smith L, Underhill P, Pritchard C, Tymowska-Lalanne Z, Abdul-Hussein S, Hilton H, Winchester L, Williams D, Freeman T, Webb S, Greenfield A (2002) Single primer amplification (SPA) of cDNA for microarray expression analysis. *Nucleic Acids Res.* **31,** 1-7.

155. Sreevatsan S, Pan X, Stockbauer KE, Conelle ND, Kreiswirth BN, Whittam TS, Musser JM (1997a) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evloutionary recent global dissemination. *Proc.Natl.Acad.Sci.USA* **94,** 9869-9874.

156. Sreevatsan S, Pan X, Zhang Y, Deretic V, Musser JM (1997b) Analysis of the oxyR-ahpC region in isoniazid-resistant and - susceptible Mycobacterium tuberculosis complex organisms recovered from diseased humans and animals in diverse localities. *Am.Soc.Microbiol.* **41,** 600-606.

157. Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR (2001) Phylogenetic analyses do not support horizontal gene transfer from bacteria to vertebrates. *Nature* **411,** 940-944.

158. Steenken W (1934) Biological studies of the tubercle bacillus. III Dissociation and pathogenicity of the human tubercle bacillus (H37). *J.Expr.Med.* **60,** 515.

159. Sullivan, J. Cells alive: Bacterial cell structure. www.cellsalive.com/cells/bactcell.htm . 2003.

160. Supply P, Magdalena J, Himpens S, Locht C (1997) Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol.Microbiol.* **26,** 991-1003.

161. Takayama K, Wang L, David HL (1972) Effect of isoniazid on the *in vivo* mycolic acid synthesis, cell growth, and viability of *Mycobacterium tuberculosis. Antimicrob.Agents Chemother.* **2,** 29.

162. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B (1990) IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res* **18,** 188-192.

163. Thurman P, Draper P (1989) Biosynthesis of phenolic glycolipids in *M. microti. Acta Leprol* **7,** 74-76.

164. Torres MJ, Criado A, Palomares JC, Aznar J (2000) Use of Real-Time PCR and fluorimetry for rapid Detection of Rifampin and Isoniazid resistance-associated mutations in M*ycobacterium tuberculosis. J.Clin.Microbiol.* **38,** 3194-3199.

165. Upton A, Johnson N, Sandy J, Sim E (2001a) Arylamine N-acetyltransferases – of mice, men and microorganisms. *Trends Pharmacol.Sci.* **22,** 140-146.

166. Upton A, Mushtaq A, Victor T, Sampson SL, Sandy J, Smith D-M, van-Helden P, Sim E (2001b) Arylamine N-acetyltransferase of *Mycobacterium tuberculosis* is a polymorphic enzyme and a site of isoniazid metabolism. *Mol.Microbiol.* **42,** 309-317.

167. Vachula M, Holzer TJ, Anderson BR (1989) Suppression of monocyte oxidation response by phenolic glycolipid I of *Mycobacterium leprae. J.Immunol.* **142,** 1696-1701.

168. Vainrub A, Montgomery Petitt B (2003) Surface electrostatic effects in oligonucleotide microarrays: Control and optimization of binding thermodynamics. *Biopolymers* **68,** 265-270.

169. Vandecasteele SJ, Peetermans WE, Merck R, van Eldere J (2001) Quantification of expression of Staphylococcus epidermidis housekeeping genes with Taqman quantitative PCR during in vitro growth and under different conditions. *J.Bacteriol.* **183,** 7094-7101.

170. Varzim G, Monteiro E, Silva R, Pinheiro C, Lopes C (2002) Polymorphisms of arylamine N-acetyltransferase (NAT1 and NAT2) and larynx cancer susceptibility. *ORL J.Otorhinolaryngol Relat.Spec.* **64,** 206-212.

171. Victor TC, Warren R, Butt JL, Jordaan AM, Felix JV, Venter A, Sirgel FA, Schaaf HS, Donald PR, Richardson M, Cynamon MH, VanHelden PD

(1997) Genome and MIC stability in Mycobacterium tuberculosis and indications for continuation of use of isoniazid in multidrug-resistant tuberculosis. *J.Med.Microb.* **46,** 847-857.

172. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT:a program to generate schematic diagram of protein-ligand interactions. *Prot.Eng.* **8,** 127-134.

173. Warren R, Richardson M, Sampson SL, Hauman JH, Beyers N, Donald PR, van-Helden P (1996) Genotyping of *Mycobacterium tuberculosis* with additional markers enhances accuracy in epidemiological studies. *J.Clin.Microbiol.* **34,** 2219-2224.

174. Warren R, Sampson SL, Richardson M, van der Spuy GD, Lombard CJ, Victor TC, van-Helden P (2000) Mapping of IS6110 flanking regions in clinical isolates of Mycobacterium tuberculosis demonstrates genome plasticity. *Mol.Microbiol.* **37,** 1405-1416.

175. Warren R, van der Spuy GD, Richardson M, Beyers N, Booysens M, Behr M, van-Helden P (2002) Evolution of the IS*6110*-based restriction fragment length polymorphism pattern during the transmission of *Mycobacteria tuberculosis*. *J.Clin.Microbiol.* **40,** 1277-1282.

176. Watanabe M, Sofuni T, Nohmi T (1992) Involvement of Cys69 residue in the catalytic mechanism of N-hydroxyarylamine O-acetyltransferase of *Salmonella typhimurium*. Sequence similarity at the amino acid level suggests a common catalytic mechanism of acetyltransferase for *S. typhimurium* and higher organisms. *J.Biol.Chem.* **267,** 8429-8436.

177. Watanabe M, Sofuni T, Nohmi T (1993) Comparison of the sensitivity of *Salmonella typhimurium* strains YG1024 and YG1012 for detecting the mutagenicity of aromatic and nitroarenes. *Mutat Res* **301,** 7-12.

178. Watterson SA, Wilson SM, Yates MD, Drobniewski FA (1998) Comparison of three molecular assays for rapid detection of rifampin resistance in *Mycobacterium tuberculosis*. *J.Clin.Microbiol.* **36,** 1969-1973.

179. Weber WW, Hein DW (1985) Arylamine N-acetyltransferases. *Pharmacol.Rev.* **37,** 25-79.

180. Weitzerbin-Falzpan J, Das BC, Azuma I, Adams A, Petit JF, Lederer E (1970) Determination of amino acid sequences in peptides by mass spectrometry. Isolation and mass spectronomic identification of peptide subunits of mycobacterial cell walls. *Biochem Biophys Res Commun* **40,** 57-63.

181. Wheeler PR, Ratledge C (1994) 'Tuberculosis: Pathogenesis, Protection, and Control.' (Am. Soc. Microbiol.: Washington DC)

182. WHO, World Health Organisation. Tuberculosis. http://www.who.int/gtb/index.htm . 1999.

183. Wikipedia Encylopedia. Cytosol. http://www.wikipedia.org . 2002.

184. Wilson M, De Risi J, Kristensen HH, Imboden P, Rane S, Brown PO (1999) Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization. *Proc.Natl.Acad.Sci.USA* **96,** 12833-12838.

185. Young D (1994) Strategies for new drug development. In 'Tuberculosis. Pathogenesis, protection and control.'. (Ed. B Bloom) pp. 559-568. (ASM Press: Washington, DC)

186. Young DB, Cole ST (1993) Leprosy, tuberculosis, and the new genetics. *J Bacteriol* **175,** 1-6.

187. Yu T-W, Shen Y, Doi-Katayama Y, Tang L, Park C, Moore B, Hutchinson C, Floss H (1999) Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc.Natl.Acad.Sci.USA* **96,** 9051-9056.

188. Zhang Y, Dhandayuthapani S, Deretic V (1996) Molecular basis for the exquisite sensitivity of *Mycobacterium tuberculosis* to isoniazid. *Proc.Natl.Acad.Sci.USA.* **23,** 13212-13216.

189. Zhang Y, Young D (1993) Molecular mechanisms of isoniazid: a drug at the front line of tuberculosis control. *Trends Microbiol* **1,** 109-113.

APPENDIX 1

1. Quantitative PCR

1.1. Amount of RNA present in each reaction mix

Table 1.1.1:

a) H37Rv trial 1

days	*16S rRNA* copy #	STDEV	DNAse step	RT (1:5)	extra cDNA dilution	Q-PCR (1:5)	RT efficiency Copy #
0	0.00E+00	0.00E+00	0.00E+00	2.00E-02		1.20E-01	2.34E+00
1	9.00E-02	6.00E-02	2.30E-01	1.15E+00		5.73E+00	1.15E+02
2	9.00E-01	4.40E-01	2.40E+00	1.20E+01		5.99E+01	1.20E+03
4	2.08E+02	1.15E+02	5.52E+02	2.76E+03		1.38E+04	2.76E+05
7	8.16E+03	1.17E+03	2.17E+04	1.08E+05	2.39E+05	1.19E+06	2.39E+07
14	1.63E+04	1.43E+03	4.34E+04	2.17E+05	4.78E+05	2.39E+06	4.78E+07
21	1.67E+04	1.98E+03	4.45E+04	2.22E+05	4.89E+05	2.45E+06	4.89E+07
28	5.72E+03	5.90E+01	1.52E+04	7.60E+04	1.67E+05	8.36E+05	1.67E+07

days	*nat* copy #	STDEV	DNAse step	RT (1:5)	extra cDNA dilution	Q-PCR (1:5)	RT efficiency Copy #
0	0.00E+00	0.00E+00	0.00E+00	0.00E+00		0.00E+00	0.00E+00
1	0.00E+00	0.00E+00	0.00E+00	0.00E+00		0.00E+00	0.00E+00
2	0.00E+00	0.00E+00	0.00E+00	0.00E+00		0.00E+00	0.00E+00
4	1.00E-02	0.00E+00	3.00E-02	1.70E-01		8.40E-01	1.68E+01
7	4.50E-01	4.30E-01	1.19E+00	5.97E+00	1.31E+01	6.57E+01	1.31E+03
14	1.70E-01	1.00E-02	4.50E-01	2.23E+00	4.90E+00	2.45E+01	4.90E+02
21	2.20E-01	3.00E-02	5.90E-01	2.97E+00	6.54E+00	3.27E+01	6.54E+02
28	3.00E-02	2.00E-02	8.00E-02	4.10E-01	9.00E-01	4.52E+00	9.04E+01

b) H37Rv Trial 2

days	*16S rRNA* copy #	STDEV	DNAse Step	RT (1:5)	extra cDNA dilution	Q-PCR (1:5)	RT efficiency Copy #
0	2.00E-02	0.00E+00	4.00E-02	2.20E-01		1.08E+00	2.16E+01
1	1.00E-02	1.00E-02	2.00E-02	9.00E-02		4.50E-01	8.98E+00
2	4.00E-02	1.00E-02	1.20E-01	6.00E-01		2.98E+00	5.95E+01
4	2.07E+01	1.77E+01	5.51E+01	2.76E+02		1.38E+03	2.76E+04
7	8.93E+03	1.15E+03	2.37E+04	1.19E+05	2.61E+05	1.31E+06	2.61E+07
14	6.66E+03	1.38E+02	1.77E+04	8.86E+04	1.95E+05	9.75E+05	1.95E+07
21	9.14E+03	7.54E+02	2.43E+04	1.22E+05	2.67E+05	1.34E+06	2.67E+07
28	2.34E+03	2.54E+02	6.24E+03	3.12E+04	6.86E+04	3.43E+05	6.86E+06

days	*nat* copy #	STDEV	DNAse Step	RT (1:5)	extra cDNA dilution	Q-PCR (1:5)	RT efficiency Copy #
0	0.00E+00	0.00E+00	0.00E+00	0.00E+00		0.00E+00	0.00E+00
1	0.00E+00	0.00E+00	0.00E+00	0.00E+00		0.00E+00	0.00E+00
2	0.00E+00	0.00E+00	0.00E+00	0.00E+00		2.00E-02	3.10E-01
4	1.00E-02	0.00E+00	3.00E-02	1.40E-01		7.10E-01	1.42E+01
7	4.60E-01	1.00E-02	1.22E+00	6.11E+00	1.35E+01	6.72E+01	1.34E+03
14	2.60E-01	2.00E-02	6.90E-01	3.46E+00	7.61E+00	3.81E+01	7.61E+02
21	9.00E-02	4.00E-02	2.30E-01	1.15E+00	2.54E+00	1.27E+01	2.54E+02
28	1.00E-02	0.00E+00	2.00E-02	1.20E-01	2.70E-01	1.33E+00	2.66E+01

RNA extracted at different days of the growth cycle. To portray more realistic numbers dilution factors and efficiency values need to be accounted for e.g. DNAse step: 1 in 2.66 dilution, multiply by 2.66. RT efficiency: assume 5%, multiply by 20. RT and Q-PCR reactions: 1in 5 dilutions, multiply by 5. Extra cDNA dilution: 1 in 2.2 dilution, multiply by 2.2. The value obtained for RT efficiency is more realistic amounts of mRNA (copy number (#)) present in the sample analysed.

1.2. Showing how the final mRNA value was obtained

Table 1.2.1: The expression value of *nat* using both the absorbance readings and protein concentration as an estimate for cell number.
a) H37Rv Trial 1

Days	OD600	[RNA] ng/ul	RT tot RNA	*rRNA* amount	total *rRNA*	*nat* amount	*rRNA*/OD	*rRNA*/[RNA]	G/H	*nat*/RNA	*nat*/cell
day 0	0.00E+00	2.10E+01	1.05E+02	2.34E+00	9.37E+01	0.00E+00	0.00E+00	2.00E-02	0.00E+00	0.00E+00	0.00E+00
day 1	0.00E+00	2.16E+01	1.08E+02	1.15E+02	4.58E+03	0.00E+00	0.00E+00	1.06E+00	0.00E+00	0.00E+00	0.00E+00
day 2	0.00E+00	1.56E+01	7.80E+01	1.20E+03	4.79E+04	0.00E+00	0.00E+00	1.54E+01	0.00E+00	0.00E+00	0.00E+00
day 4	2.50E-01	1.27E+02	6.36E+02	2.76E+05	1.10E+07	1.68E+01	4.45E+07	4.34E+02	1.03E+05	3.00E-02	2.70E+03
day 7	5.80E-01	1.95E+02	9.75E+02	2.39E+07	1.91E+09	1.31E+03	3.31E+09	2.45E+04	1.35E+05	1.35E+00	1.83E+05
day 14	1.29E+00	2.93E+02	1.47E+03	4.78E+07	3.82E+09	4.90E+02	2.95E+09	3.26E+04	9.06E+04	3.30E-01	3.03E+04
day 21	1.04E+00	6.14E+01	3.07E+02	4.89E+07	3.91E+09	6.54E+02	3.75E+09	1.59E+05	2.36E+04	2.13E+00	5.02E+04
day 28	1.38E+00	2.62E+01	1.31E+02	1.67E+07	1.34E+09	9.04E+01	9.67E+08	1.28E+05	7.57E+03	6.90E-01	5.23E+03

b)

days	[protein] (µg/ml)	total ng of protein	[RNA] ng/ul	RT tot RNA	*rRNA* amount	total *rRNA*	*nat* amount	rRNA/ [protein]	*rRNA*/[RNA]	G/H	*nat*/[RNA]	*nat*/cell
day 0	1.75E+01	1.22E+05	2.10E+01	1.05E+02	2.34E+00	9.37E+01	0.00E+00	0.00E+00	2.00E-02	3.00E-02	0.00E+00	0.00E+00
day 1	1.04E+01	7.23E+04	2.16E+01	1.08E+02	1.15E+02	4.58E+03	0.00E+00	6.00E-02	1.06E+00	6.00E-02	0.00E+00	0.00E+00
day 2	5.88E+00	4.10E+04	1.56E+01	7.80E+01	1.20E+03	4.79E+04	0.00E+00	1.17E+00	1.54E+01	8.00E-02	0.00E+00	0.00E+00
day 4	4.82E+01	3.36E+05	1.27E+02	6.36E+02	2.76E+05	1.10E+07	1.68E+01	3.28E+01	4.34E+02	8.00E-02	3.00E-02	2.00E-03
day 7	1.06E+02	7.41E+05	1.95E+02	9.75E+02	2.39E+07	1.91E+09	1.31E+03	2.58E+03	2.45E+04	1.10E-01	1.35E+00	1.42E-01
day 14	1.67E+02	1.16E+06	2.93E+02	1.47E+03	4.78E+07	3.82E+09	4.90E+02	3.28E+03	3.26E+04	1.00E-01	3.30E-01	3.40E-02
day 21	2.02E+02	1.41E+06	6.14E+01	3.07E+02	4.89E+07	3.91E+09	6.54E+02	2.78E+03	1.59E+05	2.00E-02	2.13E+00	3.70E-02
day 28	2.91E+02	2.03E+06	2.62E+01	1.31E+02	1.67E+07	1.34E+09	9.04E+01	6.58E+02	1.28E+05	1.00E-02	6.90E-01	4.00E-03

Table 1.2.2: The expression value of *nat* using both the absorbance readings and protein concentration as an estimate for cell number
a) H37Rv Trial 2

Days	OD600	[RNA] ng/ul	RT tot RNA	*rRNA* amount	total *rRNA*	*nat* amount	*rRNA*/OD	*rRNA*/[RNA]	G/H	*nat*/RNA	*nat*/cell
day 0	4.00E-02	2.10E+01	1.05E+02	2.16E+01	8.65E+02	0.00E+00	2.47E+04	2.10E-01	1.20E+05	0.00E+00	0.00E+00
day 1	4.00E-02	2.16E+01	1.08E+02	8.98E+00	3.59E+02	0.00E+00	9.21E+03	8.00E-02	1.11E+05	0.00E+00	0.00E+00
day 2	7.00E-02	1.56E+01	7.80E+01	5.95E+01	2.38E+03	3.10E-01	3.40E+04	7.60E-01	4.46E+04	0.00E+00	1.76E+02
day 4	2.00E-01	1.27E+02	6.36E+02	2.76E+04	1.10E+06	1.42E+01	5.46E+06	4.34E+01	1.26E+05	2.00E-02	2.81E+03
day 7	9.20E-01	1.95E+02	9.75E+02	2.61E+07	2.09E+09	1.34E+03	2.27E+09	2.68E+04	8.48E+04	1.38E+00	1.17E+05
day 14	1.25E+00	2.93E+02	1.47E+03	1.95E+07	1.56E+09	7.61E+02	1.25E+09	1.33E+04	9.39E+04	5.20E-01	4.88E+04
day 21	1.26E+00	6.14E+01	3.07E+02	2.67E+07	2.14E+09	2.54E+02	1.70E+09	8.71E+04	1.95E+04	8.30E-01	1.61E+04
day 28	1.32E+00	2.62E+01	1.31E+02	6.86E+06	5.49E+08	2.66E+01	4.17E+08	5.24E+04	7.96E+03	2.00E-01	1.62E+03

b)

Days	[protein] (µg/ml)	total ng protein	[RNA] ng/ul	RT tot RNA	*rRNA* amount	total *rRNA*	*nat* amount	*rRNA*/ [protein]	*rRNA*/[RNA]	G/H	*nat*/[RNA]	*nat*/cell
day 0	1.97E+01	1.37E+05	2.10E+01	1.05E+02	2.16E+01	8.65E+02	0.00E+00	1.00E-02	2.10E-01	3.00E-02	0.00E+00	0.00E+00
day 1	8.43E+00	5.88E+04	2.16E+01	1.08E+02	8.98E+00	3.59E+02	0.00E+00	1.00E-02	8.00E-02	7.00E-02	0.00E+00	0.00E+00
day 2	1.14E+01	7.92E+04	1.56E+01	7.80E+01	5.95E+01	2.38E+03	3.10E-01	3.00E-02	7.60E-01	4.00E-02	0.00E+00	0.00E+00
day 4	4.09E+01	2.85E+05	1.27E+02	6.36E+02	2.76E+04	1.10E+06	1.42E+01	3.86E+00	4.34E+01	9.00E-02	2.00E-02	2.00E-03
day 7	1.22E+02	8.53E+05	1.95E+02	9.75E+02	5.75E+07	4.60E+09	1.34E+03	5.39E+03	5.89E+04	9.00E-02	1.38E+00	1.26E-01
day 14	1.48E+02	1.03E+06	2.93E+02	1.47E+03	4.29E+07	3.43E+09	7.61E+02	3.32E+03	2.93E+04	1.10E-01	5.20E-01	5.90E-02
day 21	1.50E+02	1.05E+06	6.14E+01	3.07E+02	5.88E+07	4.70E+09	2.54E+02	4.48E+03	1.92E+05	2.00E-02	8.30E-01	1.90E-02
day 28	2.04E+02	1.42E+06	2.62E+01	1.31E+02	1.51E+07	1.21E+09	2.66E+01	8.50E+02	1.15E+05	1.00E-02	2.00E-01	1.00E-03

Tables 1.2.1 and 1.2.2 show the results for each equation used to get the final mRNA value. The total RNA used in the RT reaction (tot RNA) is the RNA concentration multiplied by 5 (number of µl used in the RT reaction). The G/H value is the rRNA/OD (or protein concentration [protein]) divided by the rRNA/[RNA] (RNA concentration). To get the amount of *nat* mRNA per cell (*nat*/cell) the G/H value is multiplied by the amount of *nat* mRNA per RNA content (*nat*/[RNA]).
The protocol used here was that of the Standard curve method explained in Applied biosystems User Bulletin #2

1.3. NAT expression during growth cycle of *M. bovis* BCG

Table 1.3.1: The relative amounts of *nat* mRNA during the growth curve of *M. bovis* BCG

Day	C_T with *nat* probe			Ave	STDEV	C_T with *16S rRNA* probe			Ave	STDEV
6	28.78	29.39	29.04	29.07	0.31	10.38	10.41	10.42	10.40	0.02
5	27.31	27.68	27.26	27.42	0.23	9.37	9.41	9.40	9.39	0.02
4	27.19	27.18	27.04	27.14	0.08	9.37	9.33	9.36	9.35	0.02
3	26.64	26.90	26.85	26.80	0.14	9.42	9.31	9.32	9.35	0.06
2	27.40	27.04	26.80	27.08	0.30	9.45	9.38	9.44	9.42	0.04
1	30.30	29.57	29.98	29.95	0.37	11.91	11.79	11.80	11.83	0.07

Day	OD600	av. *nat*	STDEV	av. *rRNA*	STDEV	av.*nat*-av.*rRNA*	STDEV
6	1.85	29.07	0.31	10.40	0.02	18.67	0.02
5	1.40	27.42	0.15	9.39	0.02	18.02	0.02
4	0.89	27.14	0.08	9.35	0.02	17.78	0.02
3	0.35	26.80	0.23	9.35	0.06	17.45	0.06
2	0.13	27.08	0.30	9.42	0.04	17.66	0.04
1	0.08	29.95	0.37	11.83	0.07	18.12	0.07

Day	difference from day 6	relative to day 6	incl STDEV1	incl STDEV2
6	0.00	1.00	0.99	1.01
5	-0.64	1.56	1.54	1.58
4	-0.88	1.84	1.82	1.87
3	-1.22	2.33	2.23	2.43
2	-1.01	2.01	1.96	2.07
1	-0.55	1.46	1.40	1.53

These are the results for each step of the equations used for the Comparative Method explained in Applied biosystems User Bulletin #2.

APPENDIX 2

2. *Mycobacterium tuberculosis* Isolates

Table 2.1 : Polymorphisms, mutations and Minimum Inhibitory Concentration (MIC) of INH in different isolates of *Mycobacterium tuberculosis*

[a]Isolate	[b]Strain family	mut (s)	*nat* G619A SNP	*nat* T529C SNP	MIC (μg/ml)
H37Rv	Not classified	-	-	-	<0.005
131	2	+ *katG* (315AGC-AAT)	-	-	5
929	3	-	+	+	0.025
1394	3	-	+	+	0.05
1398	3	-	+	+	0.05
1430	3	-	+	+	0.025
208	11	-	-	-	0.025
766	18	+ *katG* (315AGC-ACC)	-	-	1
83	28	-	+	-	0.05
404	28	+ *katG* (315AGC-ACC)	+	-	5
526	28	-	+	-	0.025
704	28	-	+	-	0.05
1077	28	-	+	-	0.05
1595	28	-	+	-	0.025
816	133	+ *aphC* (-9 G-A)	-	-	>10

A collection of DNA samples and a corresponding database with clinical and molecular information from *M. tuberculosis* clinical isolates, obtained from tuberculosis patients residing in the Western Cape Province of South Africa, was used for this study (Warren *et al.* 2000). From this collection [a]isolates (already classified into specific [b]families) were selected according to their polymorphism (pol) and mutational (mut) status (presence (+) and absence (-)) and whether they were resistant (resis) to INH or not. Isolates with a MIC of INH higher than 0.1μg/ml are regarded as resistant. No other known mutations contributing to INH resistance {Ramaswamy, 2003 57 /id} are present.

APPENDIX 3

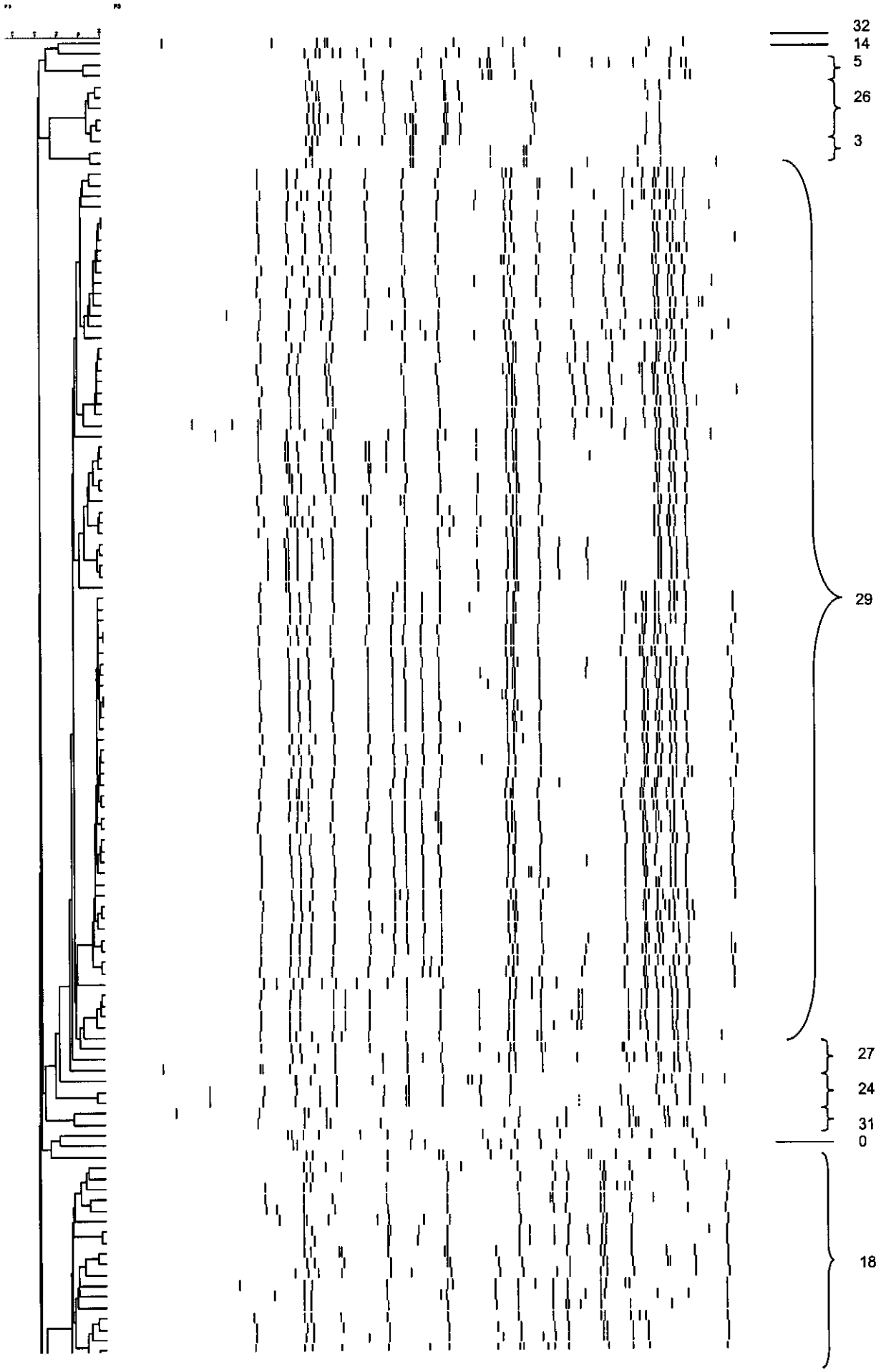
3. The effects of INH exposure on *nat* expression

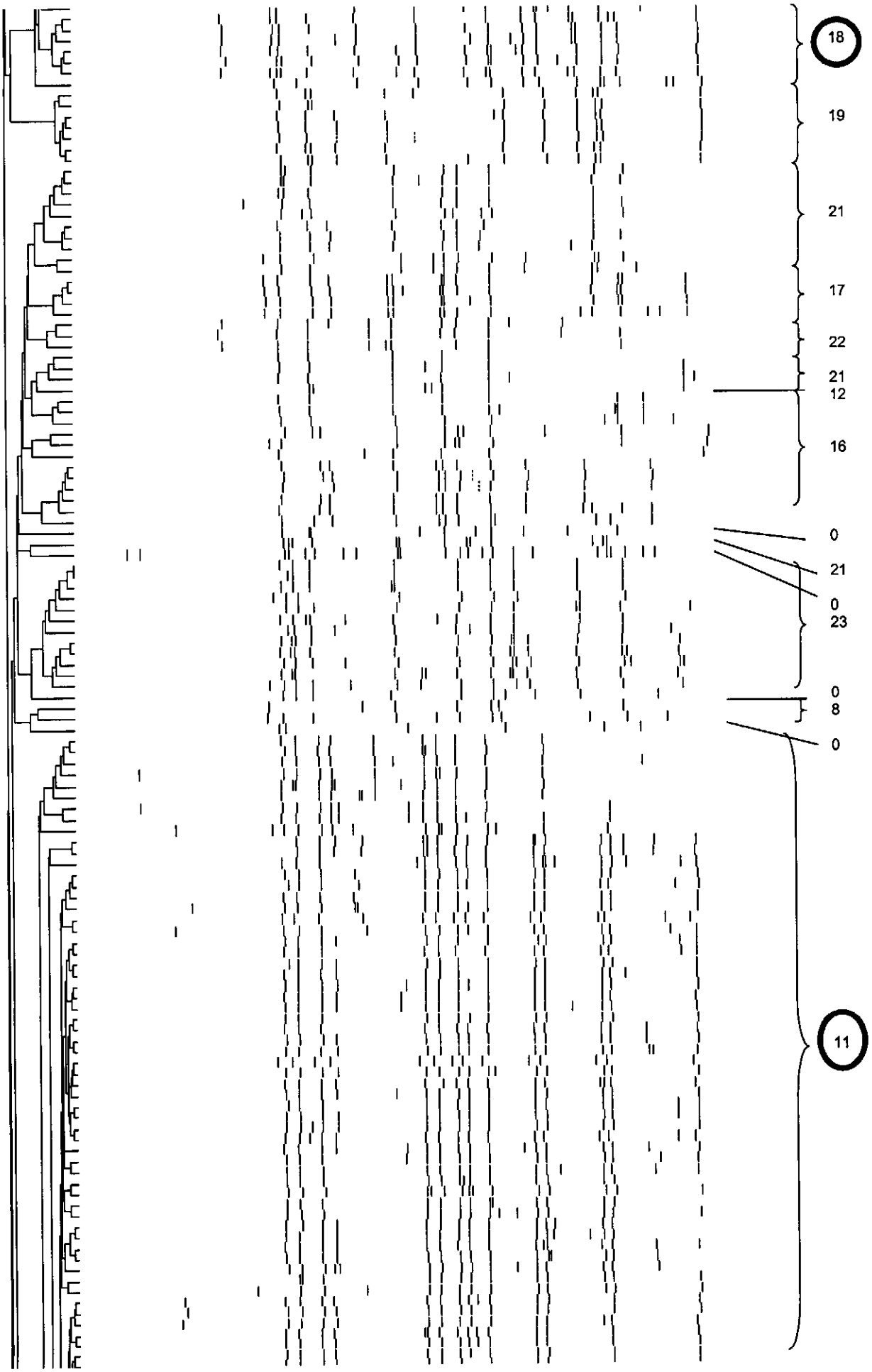
Table 3.1: Relative amounts of *nat* mRNA in *M. tuberculosis* cultures exposed to INH

		tbnat expression for each isolate of *M. tuberculosis*					
Time	Sample	H37Rv		816		1430	
		nat amount	fold incr	*nat* amount	fold incr	*nat* amount	fold incr
0 minutes	Control	2.94E-05	6.32E+00	3.38E-05	2.15E+00	4.21E-06	1.64E+00
	INH	1.86E-04		7.28E-05		6.91E-06	
30 minutes	Control	3.01E-05	7.02E+00	1.08E-04	1.32E+00	1.12E-05	2.64E+00
	INH	2.12E-04		1.42E-04		2.96E-05	
2 hours	Control	1.70E-05	3.08E+01	7.53E-05	9.27E-01	1.69E-05	5.59E+00
	INH	5.24E-04		6.98E-05		9.45E-05	
6 hours	Control	1.52E-03	1.71E-01	6.76E-05	2.10E+00	1.67E-06	5.96E+00
	INH	2.60E-04		1.42E-04		9.95E-06	
20 hours	Control	3.28E-03	1.66E-02	7.74E-05	2.36E-01	1.24E-04	2.81E-01
	INH	5.43E-05		1.82E-05		3.48E-05	

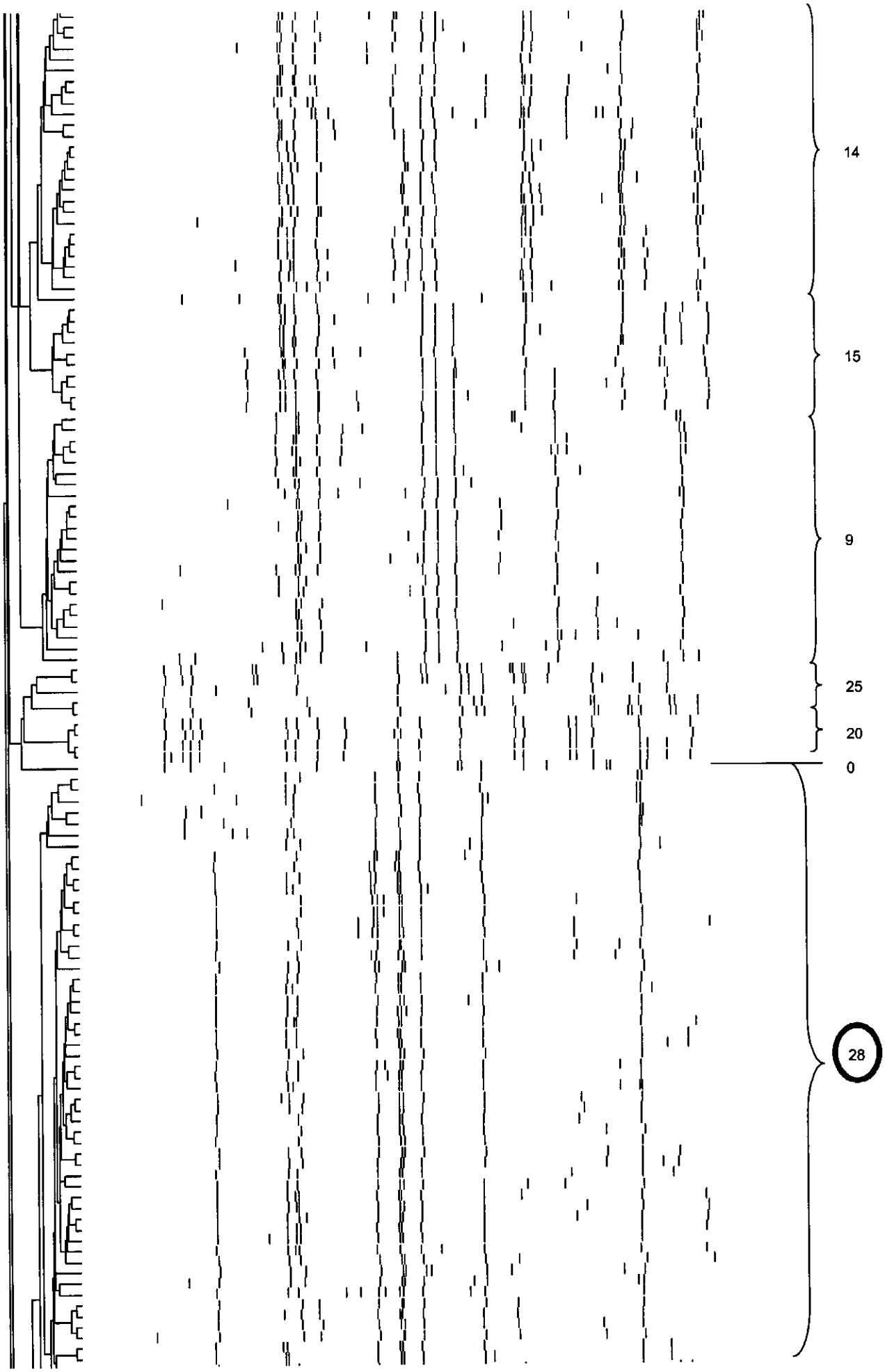
The ratio of *nat* mRNA to 16S rRNA gives the relative amount of *nat* (*nat* amount) in the Q-PCR reaction. The amount of *nat* mRNA in cultures exposed to INH (exposure time varied) were compared to the amount of *nat* mRNA in corresponding control cultures (not exposed to INH). These values were referred to as the fold increase in *nat* mRNA (fold incr).

Appendix 4




18
19
21
17
22
21
12
16
0
21
0
23
0
8
0
11

11
0
13
14


14
15
9
25
20
0
28

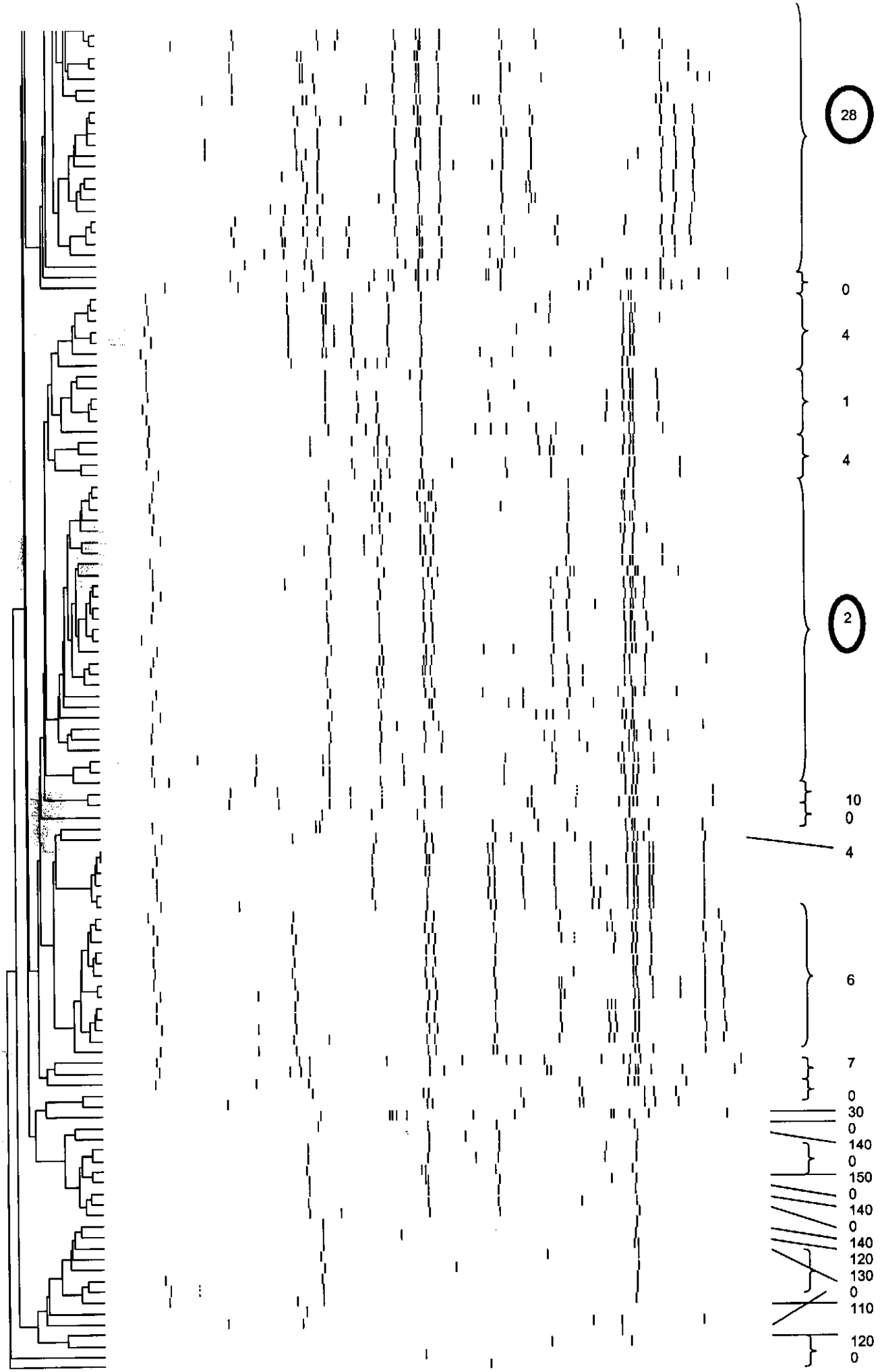

28
0
4
1
4
2
10
0
4
6
7
0
30
0
140
0
150
0
140
0
140
120
130
0
110
120
0

Figure 4.1: Family tree of the majority of strains making up the 37 different strain families, as well as the unique isolates.

A family of strains is defined as strains which are > 65% related to each other in terms of RFLP IS6110 DNA banding patterns. Unique isolates contain unique IS6110 patterns which do not fall into any of the families {Warren, 2000 639 /id}. The family numbers are presented on the right side of the family tree. The family numbers surrounded by a blue circle are those used in this study.

Publications

- Edward Brook, Carolyn Sholto-Douglas-Vernon and Edith Sim (2003) The role of arylamine N-acetyltransferase in bacteria. Published by Transworld Research Network, Kerala, India.
- Bhakta S., Besra G.S., Upton,A.,Parish T., Sholto-Douglas-Vernon C., Gibson, K.J.C., Knutton S., Gordan S., Da-Silva R.P., Anderton M.C., Sim E. (2004) Arylamine N-acetyltransferase is required for synthesis of mycolic acids and complex lipids in *Mycobacterium bovis* BCG and represents a novel drug target. The J. of Exp. Med. **199(9)**, 1191-1999.
- Sholto-Douglas-Vernon C., Victor T. and van Helden P. (2004) Arylamine N-Acetyltransferase of *Mycobacterium tuberculosis*: its expression and the identification of strain family specific SNPs. **(In progress).**

Abstracts

- Bhakta S.*, Sholto-Douglas-Vernon C.*, Mo M.S.N., Busby S., Victor T., van Helden P. and Sim E. (2003) Isoniazid increases the expression of the gene for Arylamine N-acetyltransferase (NAT) in Mycobacteria. Tuberculosis: Integrating Host and Pathogen Biology. Keystone Symposia, Taos, New Mexico. (* equal contribution) **Poster.**
- Sholto-Douglas-Vernon C., Bhakta S., Sampson S., Upton A., Sim E., van Helden P. and Victor T. Expression of N-acetyltransferase (NAT) in *M. bovis* BCG and *M. tuberculosis.* Biochemical Society Focused Meeting, Structural Biology in Drug Metabolism and Drug Discovery, AstraZeneca R & D Charnwood, Loughborough. **Poster.**