Arylamine *N*-Acetyltransferases from Mycobacteria: Investigations of a Potential Target for Anti-Tubercular Therapy

by

Areej Abuhammad

Supervised by

Professor E. Sim and Professor E. Garman

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

Department of Pharmacology and University College University of Oxford

" وَأَنْزَلَ اللَّهُ عَلَيْكَ الْكِتَابَ وَالْحِكْمَةَ وَعَلَّمَكَ مَا لَمْ تَكُنْ تَعْلَمُ وَكَانَ فَضْلُ اللَّهِ عَلَيْكَ عَظِيماً"

(النساء 113)

Dedicated to
My mother

Yusra Hasoneh

Arylamine *N*-Acetyltransferases from Mycobacteria: Investigations of a Potential Target for Anti-Tubercular Therapy

Areej Abuhammad

Submitted in partial fulfilment for the degree of Doctor of Philosophy
Department of Pharmacology and University College

University of Oxford

Trinity 2012

Abstract

Reactivation of latent infection is the major cause of tuberculosis (TB). Cholesterol is a critical carbon source during latent infection. Catabolism of cholesterol contributes to the pool of propionyl-CoA, a precursor that is incorporated into cell-wall lipids.

Arylamine *N*-acetyltransferase (NAT) is encoded within a gene cluster that is involved in the sterol-ring degradation and is essential for intracellular survival. NAT from *M. tuberculosis* (TBNAT) can utilise propionyl-CoA and therefore was proposed as a target for TB-drug development. Deleting the *nat* gene or inhibiting the NAT enzyme prevents intracellular survival and results in depletion of cell-wall lipids. NAT inhibitors, including the piperidinol class, were identified by high-throughput screening. The insolubility of recombinant TBNAT has been a major limitation in pursuing it as a drug target.

Subcloning *tbnat* into a pVLT31 vector resulted in a yield of 6-16 mg/litre-bacterial-culture of pure-soluble recombinant TBNAT. The increased yield allowed for extensive screening for crystallisation conditions. However, since a structure was not obtained, the model NAT from *M. marinum* (MMNAT) was employed to further understand NAT as a target. Screening against a panel of Acyl-CoA cofactors showed that MMNAT can also utilise propionyl-CoA.

The MMNAT structure in complex with the high affinity substrate hydralazine was determined (2.1 Å) and the architecture of the arylamine pocket was delineated. A novel mechanism for the acetylation reaction of hydralazine has emerged. It is proposed that the acetyl group is transferred from acetyl-CoA to the heterocyclic aromatic nitrogen of hydralazine, which explains the immediate cyclisation of the acetylated metabolite into an *N*-methyltriazolophthalazine.

By employing mass spectroscopy, enzyme assays, computational docking and structural studies, a covalent mechanism of inhibition by the piperidinol class was established, and the inhibitor-binding pocket was identified. Inhibitors with new scaffolds were identified using the *in silico* 3D-shape screening and thermal shift assay.

Acknowledgements

With a deep sense of gratitude, I wish to express my sincere thanks to both of my supervisors, Professor Edith Sim and Professor Elspeth Garman for their guidance and support and for being always available with the right solution and the appropriate approach for every situation. I am grateful for all of their valuable suggestions, encouragement, and constructive criticism.

This work could never have been accomplished without the help, support and dedication of many people to all of whom I should like to express my sincere gratitude.

I should like to gratefully acknowledge the support of all the past and present members of the Sim laboratories (Department of Pharmacology, Oxford). In particular, I should like to thank Bob Sim and Ali Ryan for all their fruitful and valuable discussions and for introducing me to a variety of methods. I should like also to thank Nathan Lack for his help in the cloning of tbnat and for his valuable discussions and help, Judith Schweichler for her help with optimising the expression of TBNAT, Hilary Long, who carried out testing on mycobacteria, James Egleton and Sebastian Keany for testing the inhibitors on the human NATs and HsaD, Chan Ju Wang, who made time (while preparing for her viva) to introduce me to protein expression and teach me enough for a confident start, Nicola Laurieri for his useful discussions, Malgorzata Rybak-Smith for translating Polish articles. I should also like to thank Munirah Abdul Aziz, Maria Carroll, Kirsten Pondman, Anne Jaekel, Ornella Scuito, Lindsay Johnson, Susan Bayliss, Romain Ballet, Amy Varney, Nolween Juffin and Vincent Cheng for their help and support.

I sincerely thank Elizabeth Fullam (University of Birmingham, Birmingham), Isaac Westwood (Institute of Cancer Research, London), and Dimitrios Evangelopoulos (Department of Biological Science, Birkbeck, London) for their useful discussions.

I should like to gratefully acknowledge the support of all the past and present members of Garman's Research Group (Department of Biochemistry, Oxford). In particular, I should like to thank Edward Lowe for his outstanding help in all aspects of crystallography, Karthik Paithankar for introducing me to crystallography and helping in the refinement of the MMNAT-HLZ structure, Oliver Zeldin for translating French articles, Azucena Eunice Jimenez Corona for the useful information about gel-matix crystallisation, and Eugenio Delamora, Markus Gerstel, Ellizabeth Allan, Andrew Bissette and Isaac Turner for their help and support.

I sincerely thank Paul Emsely for helping me with Coot and script writing, James Parker and Sonja Baumli for their help in data collection, David Staunton for his help in protein MS, AUC and ITC analysis, Maurici Burnet for the useful discussions about PCR and cloning, Khairul Abd Halim and Phil Biggin for their help in homology modelling of TBNAT, David Harris for the useful discussions about covalent protein inhibitors and Rita De Zorzi for the useful discussions about PYMOL (Department of Biochemistry, Oxford).

I wish to thank Angela Russell for providing lab space in Chemistry and for her help and support, Peter Seden for his help in inhibitor synthesis and purification, Helen Storr for her help in HPLC purification and NMR analysis and Cyrille Thinnes and Camilo Quevedo for their help and assistance (Department of Chemistry, Oxford).

I should like to gratefully acknowledge Michael McDonough for his invaluable support and encouragement on screening for TBNAT crystallisation conditions, Nikita Loik for his help with LC/MS analysis of Cys-A-1 mixture, James McCullagh and Lingzhi Gong for their help in protein LC/MS and data interpretation, Luc Henry for providing the nonhydrolysable CoA derivatives, Akane Kawamura for translating Japanese articles and Refaat Hamed for the useful discussions about N-acetyl transferases (Department of Chemistry, Oxford).

I should like to thank Benjamin Thomas and Svenja Hester LC/MS for their help in protein LC/MS and Steven Johnson for his help with MALS analysis (Sir William Dunn School of Pathology, Oxford).

I should like to thank Giannoulis Fakis and Sotiria Boukouvala for their useful discussions about human NATs and cloning, Theodora Tsirka for her useful discussions a protein expression and Eleni Karagianni and Euanthia Kontomina for their support and assistance (Democritus University of Thrace, Greece).

I wish to thank Alex Fish (Netherlands Cancer Institute Proteomics Center, Netherlands) for useful discussions about thermostable proteins, Peter Sharratt (PNAC Facility, Cambridge) for his help in amino acid analysis, Pascal Parois (Université de Lorraine, France) for his help with small-molecules structures, Federico Forneris (Utrecht University, Netherlands) for outstanding discussions about tricks for protein crystallisation, Garrett Morris (InhibOx, Oxford) for his help in ElectroShape in-silico screening and Oleg Fedorov (SGC, Oxford) for his help and assistance in thermal shift assays.

The cooperation I received from other faculty members, colleagues, laboratory staff and administrative staff in Pharmacology, Biochemistry, Chemistry and University College is gratefully acknowledged. To all of them, I wish to express my gratitude.

I am very grateful for Edith Sim, Elspeth Garman, David Harris, Anne Jaekel, Nathan Lack, Nicola Laureiri, Akane Kawamwra, Edward Lowe, Shathat Abuhammad, Mary Hossain and Mary Montegomery, who devoted much time and effort to proof reading my thesis as it progressed, and provided invaluable feedback and encouragement.

I should like to express my deep gratitude to the Univeristy of Jordan and the Deparment of Pharmacology for their financial support and to University College for the Old members' Trust Graduate Travel bursaries, and RSC, BCA for the travel grants.

I sincerely thank all my friends in Oxford and elsewhere for their help, support and encouragement.

No words can express my gratitude and love to my family, my sincere thanks to my mother who makes it all happen and who taught me the value of helping others and of pursuing my dreams at all stages of my life, to my father who taught me how to be strong and independent, to my beloved Shatha, Khalid, Feras and Manal for always being there for me and for their endless love, support and encouragement.

Areej Abuhammad

June 2012

Table of contents

Table of figures

Table of tables

Publications

Abuhammad, A, E Fullam, ED Lowe, D Staunton, A Kawamura, IM Westwood, S Bhakta, AC Garner, DL Wilson, PT Seden, SG Davies, AJ Russell, EF Garman and E Sim (2012). Piperidinols that show anti-tubercular activity as inhibitors of arylamine N-acetyltransferase: an essential enzyme for mycobacterial survival inside macrophages. *PLoS ONE* 7(12): e52790.

Abuhammad, A, Lack, N, Schweichler, J, Staunton, D, Sim, RB, Sim, E (2011). Improvement of the expression and purification of *Mycobacterium tuberculosis* arylamine N-acetyltransferase (TBNAT) a potential target for novel anti-tubercular agents. *Protein Expression and Purification* 80(2), 246-252.

Abuhammad, A, Lowe, ED, Fullam, E, Noble, M, Garman, EF, Sim, E (2010). Probing the architecture of the *Mycobacterium marinum* Arylamine N-Acetyltransferase active site. *Protein & Cell* (1), 384-392.

Fullam, E, Kawamura, A, Wilkinson, H, **Abuhammad, A**, Westwood, I, Sim, E (2009a). Comparison of the Arylamine N-Acetyltransferase from *Mycobacterium marinum* and *Mycobacterium tuberculosis*. *The Protein Journal* (28), 281-293.

Conference proceedings

Abuhammad, A, Fullam, E, Garman, EF, Sim, E (2012). Structural Studies on Novel Antitubercular Targets. Poster in: *The Young Crystallographer Satellite meeting-British Crystallographic Association Spring Meeting*, Warwick.

Abuhammad, A, Fullam, E, Westwood, I, Russell, A, Davies, S, Sim, E (2011). Structural Studies on Novel Antitubercular Targets. Poster in: *The XXII International Congress and General Assembly of the International Union of Crystallographiny*, Madrid.

Abuhammad, A, Lowe, ED, Fullam, E, Noble, M, Schweichler, J, Garman, EF, Sim, E (2010). Structural studies on the mycobacterial arylamine N-acetyltransferases. Poster in *The ninth International Meeting of the International Society for the Study of Xenobiotics (ISSX)*, Istanbul.

Abuhammad, A, Lowe, ED, Fullam, E, Noble, M, Schweichler, J, Garman, EF, Sim, E (2010). Structural studies on the mycobacterial arylamine N-acetyltransferases. Presentation in *The Fifth International Arylamine N- Acetyltransferase Workshop*, Paris.

Abuhammad, A, Lowe, ED, Fullam, E, Noble, M, Garman, EF, Sim, E (2010). Probing the architecture of the *Mycobacterium marinum* Arylamine N-Acetyltransferase active site. Presentation in: *The Young Crystallographer Satellite meeting-British Crystallographic Association Spring Meeting*, Warwick.

Abuhammad, A, Fullam, E, Talbot, J, Kawamura, A, Westwood, I, Davies, SG, Russell, A, Sim, E (2009). Inhibitors of arylamine N-acetyltransferase as potential anti-tuberculars. Poster in: *The first RSC/SGC Symposium on Chemical Biology for Drug Discovery*: Oxford.

Abbreviations

3D	Three dimensional
4,9-DSHA	4,5-9,10-Diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid
5ASA	5-aminosalicylic acid
a.a	Amino acid
AAc-CoA	Acetoacetyl-CoA
Ac-CoA	Acetyl coenzyme A
Ar-CoA	Arachidonoyl-CoA
AV	Atomic volume
BCG	Bacillus Calmette-Guerin
CD	Circular dichroism
CF	Cord factor
CHES	2-(N-Cyclohexylamino)ethane Sulfonic Acid
De-CoA	Decanoy-CoA
DMSO	Dimethylsulfoxide
DOTS	Directly observed treatment, short-course
DS	Discovery studio
DS-TB	Drug susceptible tuberculosis
DTT	Dithiothreitol
EMB	Ethambutol
ESI-MS	Electrospray ionization mass spectrometry
EtOH	Ethanol
HEPES	N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)
His-tag	Hexahistidine-tag
HIV	Human immunodeficiency virus
HLZ	Hydralazine
HMG-CoA	Hydroxymethylgluteryl-CoA
HOPDA	2-Hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid
HTS	High throughput screening
Hx-CoA	Hexanoyl-CoA
IMAC	Immobilized metal affinity chromatography
Imz	Imidazole
INH	Isoniazid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
J	Coupling constant
Km	Michaelis constant
LAM	Lipoarabinomannan
LB	Luira-Bertani
Ma-CoA	Malonyl-CoA
MDR	Multi-drug resistant
MES	2-(N-Morpholino)ethanesulfonic acid
MIC	Minimum inhibitory concentration

Mm-CoA	Methymlonyl-CoA
MMNAT	*M. marinum* NAT
MP	Mycobacterial phagosomes
MPD	2-methyl-2,4-pentanediol
MR	Molecular replacement
MSNAT	*M. smegmatis* NAT
MTC	*Mycobacterium tuberculosis* complex
MTP	3-methyltriazolo[3,4-a]phthalazine
MW	Molecular weight
NAT	Arylamine *N*-acetyltransferase
OADC	Oleic acid-albumin-dextrose-catalase
Oc-CoA	Octanoyl-CoA
OD	Optical density
PABA	*p*-aminobenzoic acid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PDIM	Phthiocerol dimycocerate
PEG	Polyethyleneglycol
PIM	Phosphatidylinositol mannosides
POP	3-phenyl-3-oxo-propyl
Pr-CoA	Propionyl-CoA
PVK	Phenylvinyl ketone
PZA	Pyrazinamide
RIF	Rifampin
RMSD	Root-mean-square deviation
rpm	Revolutions per minute
S.D.	standard deviation
SAR	structure-activity relationship
sat. aq.	Saturated aqueous
SEM	scanning electron microscopy
SGC	Structure genomic Consortium
TACO	Tryptophan-aspartate containing coat
TB	Tuberculosis
TBNAT	*M. tuberculosis* NAT
TBSGC	TB Structure Genomic Consortium
TCA	Tricarboxylic acid cycle
TDM	Trehalose dimycolate
TDR	Totally-drug resistant
TEM	Transmission electron microscopy
TEMED	*N`*,*N`*,*N`*-tetramethylethylenediamine
TSA	Thermal shift assay
WHO	World Health Organisation
WT	Wild type
XDR	Extensively-drug resistant

1 Introduction

1.1 Tuberculosis

Tuberculosis (TB) remains one of the leading causes of death by bacterial infection. *Mycobacterium tuberculosis* (*M. tuberculosis)* was identified by Robert Koch in 1882 as the aetiological agent of TB (Daniel, 2006). *M. tuberculosis* belongs to the slowly growing members of the *Mycobacterium* genus, which in turn belongs to the Actinomycetes class. Despite belonging to the Gram-positive actinobacteria, mycobacteria are impermeable to Gram stain, due to their unusual waxy cell envelope (Section 1.2.2). This includes an outer mycolic acid layer which is considered to be the key to the success of the mycobacteria as pathogens and gives them their acid-fast property. When stained by the Ziehl-Neelsen method, mycobacteria retain the red dye after washing with acid (hence acid-fast) while other cells are stained blue (Allen, 1992; Godreuil *et al.*, 2006).

Although the disease is treatable with well-established drugs, and even preventable through vaccination, it has not been in the research programmes of most of the major pharmaceutical companies until very recently. This neglect is partly because the greatest burden of disease is among poor developing countries (Figure 1.1) (WHO, 2011). The disease affects more people today than at any other time in history (reviewed in Dye and Williams (2010)). Tuberculosis infects one-third of the human population and is responsible for approximately two million deaths annually. It is estimated that 9.8 million new infections occurred in 2010. A significant proportion of TB cases and deaths occur in human immunodeficiency virus (HIV) positive people (WHO, 2011). *M. tuberculosis* is characterised by its inherent resistance to antibiotics due to its extremely slow growth rate and the complex lipid composition of its cell wall. The current drug treatment is more than 40 years old, while the vaccine is almost 100 years old (Section 1.5.1). The ever-increasing resistance, and the side effects of currently used medications, together with the continuous

increase in HIV co-infection, necessitate the need for new, safer and more effective antituberculosis drugs.

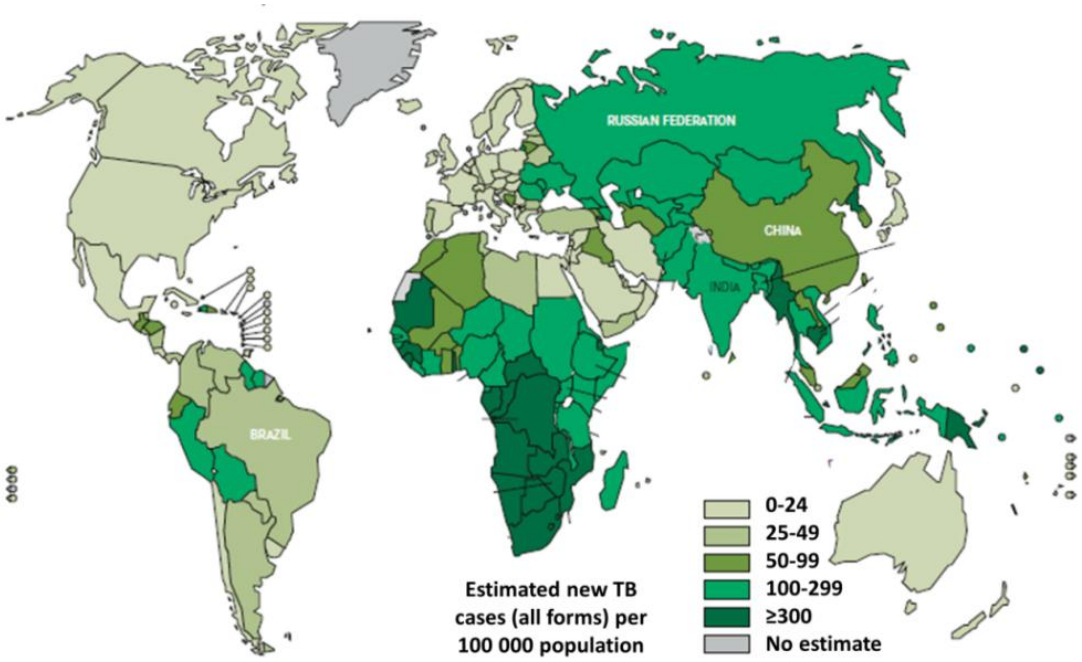


Figure 1.1: Estimated TB incidence rates in 2010 (WHO, 2011).

1.1.1 Clinical manifestations and the course of infection

TB is an airborne infectious disease that can be caused by inhaling as few as three bacilli (Riley *et al.*, 1995). In its first encounter with the lung tissues, *M. tuberculosis* utilises various routes of entry through interaction with different cells in the surrounding alveolar environment (Figure 1.2) (reviewed in (Gupta *et al.*, 2012)). Among these routes, the invasion of the alveolar macrophages has been shown to be the most favoured by *M. tuberculosis* and is considered especially advantageous for maintaining its prolonged survival (Gupta *et al.*, 2012). The outcome of this encounter is the initiation of the body's first-line defence against intruders, which is the innate immunity (Ernst, 1998). In most cases, the bacillus is immediately destroyed by the host's innate responses. However, usually this immune response is insufficient to control *M. tuberculosis* infection. Triggered by the increased level of the pathogen, the adaptive immune response is activated as a

second-line defence response which controls the primary infection (reviewed in Gupta *et al.* (2012)).

A summary of the effector mechanisms of innate and adaptive immunity and their evasion by *M. tuberculosis* is shown in Table 1.1.

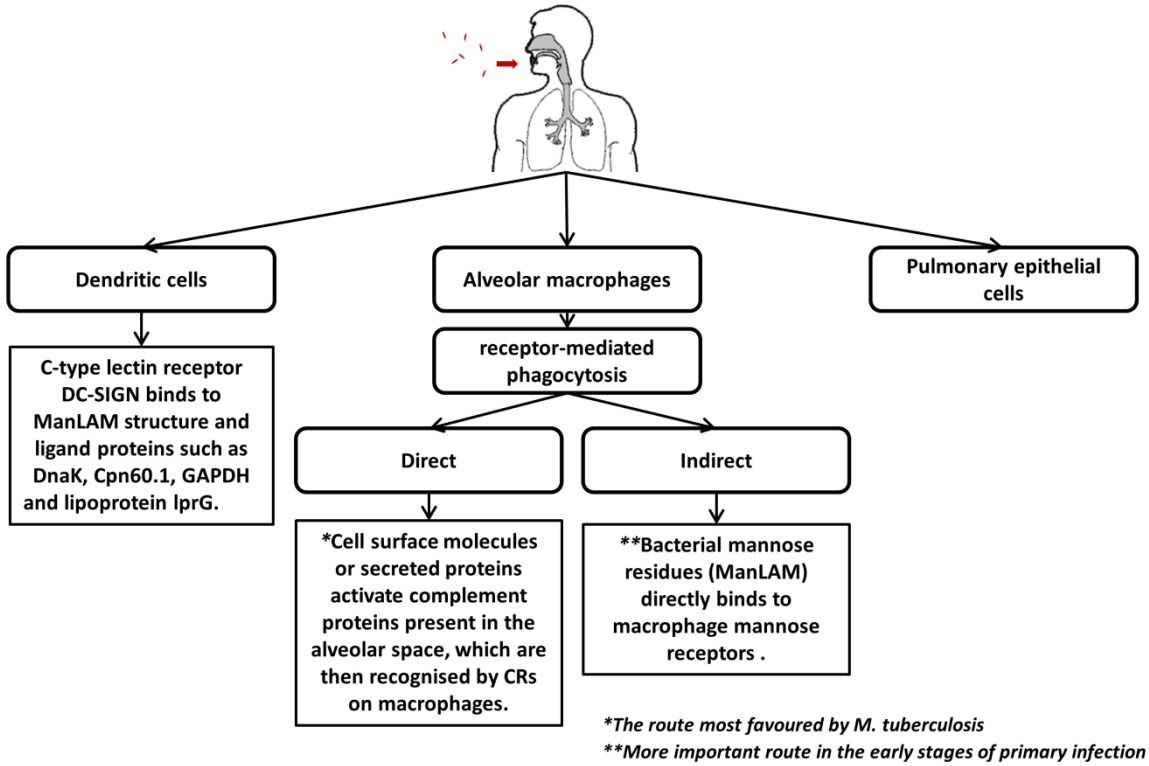


Figure 1.2: Summary of the routes of *M. tuberculosis*.

Based on information from (Gupta *et al.*, 2012).

Table 1.1: Effector mechanisms of innate and adaptive immunity, and their evasion by *M. tuberculosis*.

	Cell type	Effector mechanism	Evasion mechanism
Innate immunity	neutrophils	Degranulation	Unknown
		Production of ROIs	Genetic resistance (mechanism unknown)
		NETs	Unknown
	Macrophages	Phagocytosis	Arrest of phagosome-lysosome fusion
		Production of RNIs and NO	Genetic resistance (mechanism unknown)
	NK cells	Cytolysis of infected cells/intracellular bacteria	Unknown
		Activation of macrophage through IFN-γ	Unknown
	Complement system	Membrane attack complex, opsonisation and phagocytosis by macrophages	Unknown–possibly through binding Factor H
Adaptive immunity	CD4+ T cells	Cytolysis of infected cells/intracellular bacteria	Interference with MHC Class II antigen presentation
		Activation of macrophages through IFN-γ and TNF-α	Interference with MHC Class II antigen presentation
	CD8+ T cells	Cytolysis of infected cells/intracellular bacteria	Unknown
		Activation of macrophages through IFN-γ	Unknown
	B cells	Opsonisation and phagocytosis by macrophages	No evasion mechanism known. *M. tuberculosis* appears susceptible to Ab-opsonised phagocytosis

Abbreviations: Ab, antibody; IFN, interferon; MHC, major histocompatibility complex; NET, neutrophil extracellular trap; NO, nitric oxide; RNI, reactive nitrogen intermediate; ROI, reactive oxygen intermediate; TNF, tumor necrosis factor. Table is from (Gupta *et al.*, 2012).

Three outcomes are expected when an individual is infected (Figure 1.3): (a) in the case of primary active TB, the disease develops rapidly in a minority of the population (children and immunosuppressed hosts), and patients usually show progressive symptoms characterised mainly by a cough, (b) whilst latent TB is usually asymptomatic infection that occurs among healthy persons, and (c) post-primary TB which can arise many years after infection due to an impairment of the immune system and the reactivation of bacilli (Godreuil *et al.*, 2006; Hunter, 2011). Patients with active primary or post-primary disease are the main transmitters of infection (Figure 1.3). The severity of the active phase of the disease can range from a localised pulmonary infection to an infection of other organs associated with a wide range of symptoms (Hunter, 2011).

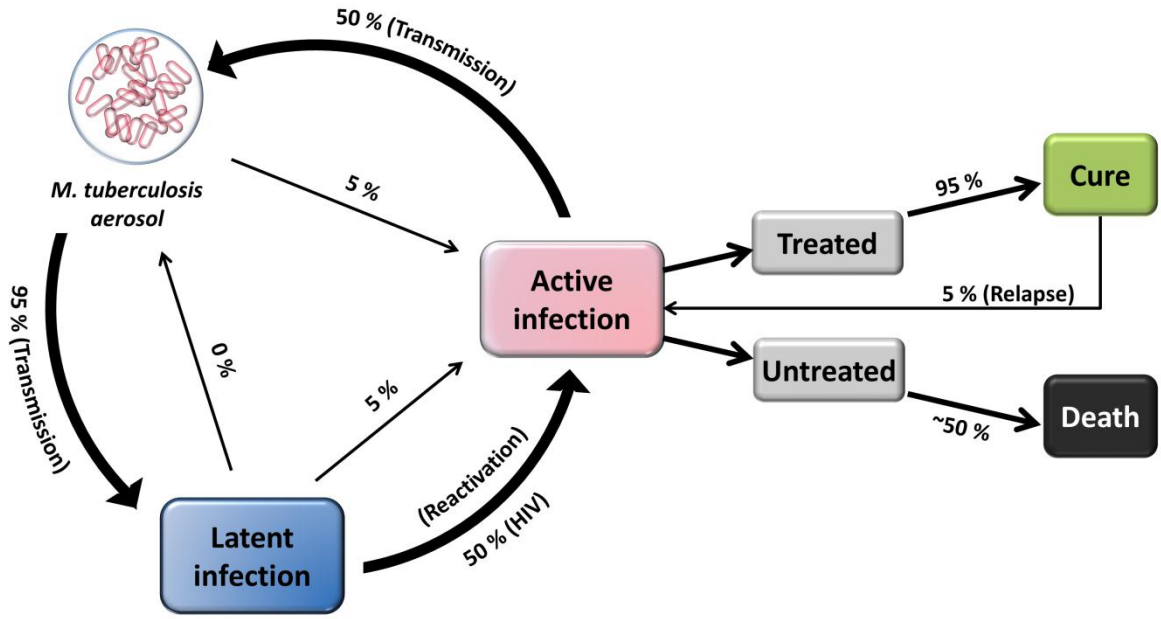


Figure 1.3: *M. tuberculosis* infection.

***M. tuberculosis* aerosol transmission can progress to infectious TB or non-infectious (latent) disease. A large pool of latently infected people may relapse into active TB years after their first exposure to the bacterium. Latent TB is commonly activated by immune suppression, as in the case of HIV. In cases of drug-susceptible (DS) TB, 95 % of patients recover upon treatment, whereas 5 % relapse. If untreated, coexisting diseases or malnutrition can result in a high mortality. The diagram and legend were modified from Koul *et al.* (2011).**

A range of scenarios that can develop upon infection with *M. tuberculosis* are summarised in Figure 1.4. The manifestations of the infection are governed by a host- and microbe-related factors. The host immune status is the major factor influencing the clinical features of TB (reviewed in Philips and Ernst (2012)). HIV infection, which dramatically increases the risk of active disease, is a major challenge in populations with a high prevalence of *M. tuberculosis* infection (reviewed in McShane (2005)).

Mycobacteria-related factors are mainly associated with the unusual lipid contents of their cell wall (reviewed in Barry (2001); Meena and Rajni (2010); and Mishra *et al.* (2011)). Among these lipids, mycolic acid is considered a key virulence factor mediating macrophage invasion by *M. tuberculosis* (Villeneuve *et al.*, 2005). Further details on the importance of the cell-wall lipids for pathogenicity are discussed in Section 1.2.2. The unique cell wall of the organism, as well as being a permeability barrier provides opportunities for specific treatment (Daffé and Draper, 1997; Russell *et al.*, 2002).

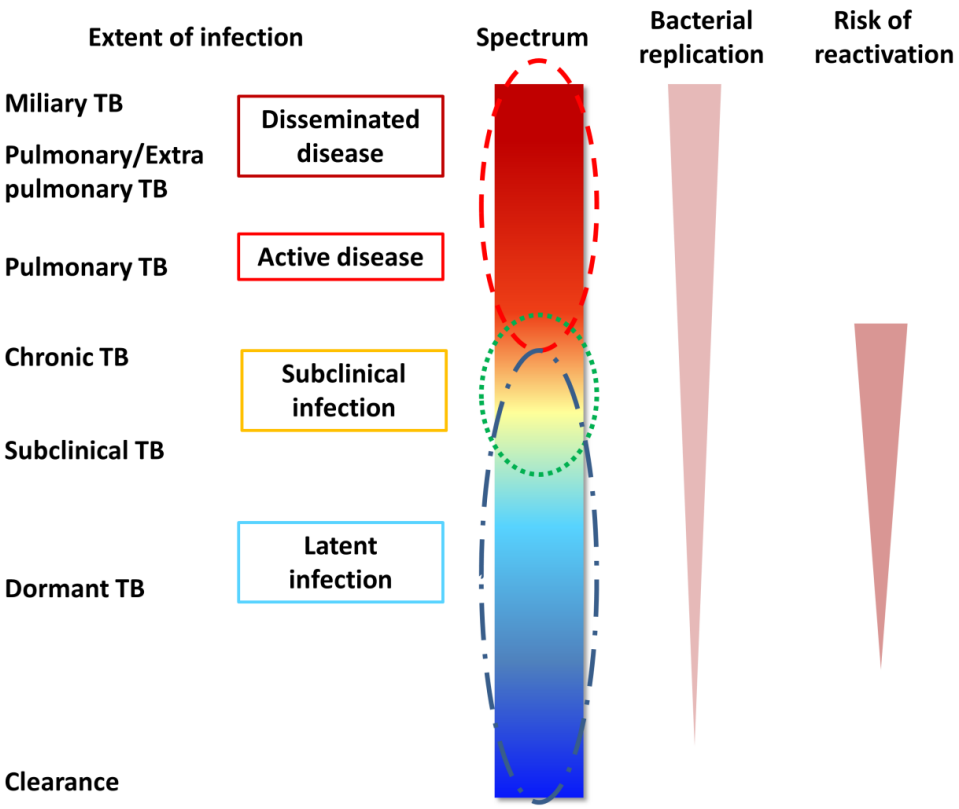


Figure 1.4: The manifestations of *M. tuberculosis* infection.

A spectrum heat-map representing the manifestations of *M. tuberculosis* infection according to the severity of the infection from complete clearance (dark-blue cold- lower end) to disseminated disease (dark-red hot-upper end). The standard classification based on the presence or absence of clinical symptoms: active (marked in red dashes) and latent TB (marked in blue dashes and dots) are marked, to indicate variability within those categories. In addition, the possible subclinical infection region is marked in green dots, which overlaps with regions of both active and latent TB. Stages associated with bacterial replication and risk of reactivation are indicated by the pink triangles; the base of each triangle corresponds to high level of replication or high risk of reactivation. The figure and legend were modified from Gideon and Flynn (2011).

1.1.2 Latent tuberculosis

One of the characteristic features of TB is the ability of *M. tuberculosis* to persist within the macrophage thus causing a latent form of the disease. Latent infection with *M. tuberculosis* is associated with the absence of clinical symptoms or the detection of any isolated bacteria from the patient. Only a positive tuberculin test or a positive result from another validated test, such as the whole-blood gamma interferon release, indicates the presence of the organism in persons with latent infection. However, the real nature of latent TB has been controversial (Cardona and Ruiz-Manzano, 2004). Evidence for the existence of *M. tuberculosis* in a latent form has been either obtained from clinical observations or concluded from *in vitro* and animal experiments (Table xx) (Cardona and Ruiz-Manzano, 2004). These findings have led to the conclusion that latent infection is a complex mixture of both slow metabolism and dormant bacilli. The fate of latent bacilli depends on many physiological factors related to the host immunity and to *M. tuberculosis* itself.

In latent *M. tuberculosis* infection, the bacteria are usually contained in a well-organized granuloma by the cell-mediated immune response (reviewed in Gideon and Flynn (2011)). Despite the role of granuloma as a host-defence mechanism, it provides a shelter that houses and protects the mycobacteria from antibiotics and further destruction by the immune responses of the host. Furthermore, it has been suggested that the development of a non-replicating state of mycobacteria, which is insensitive to antibiotics (Tudó *et al.*, 2010), is due to the accumulation of cholesterol and other lipids within granulomas (Caceres *et al.*, 2009). However, the physiology of the non-replicating *M. tuberculosis* at this stage of infection is not fully understood yet (reviewed in Pieters (2008)).

According to the WHO reports, latent infection represents the major pool of worldwide TB cases, thus making the treatment of latent TB an important strategy towards eradicating the disease (Young *et al.*, 2009).

Table 1.2: Evidence on the presence of latent bacilli.

Clinical	
	Late reactivation after a large period of chemotherapy
	Chemoprophylaxis trials: the longer the period, the lower the chance of reactivation
	Natural history: formation and healing of primary complex in basal zones of the lung. Development of post-primary TB a long time after the haematogenous dissemination to the apex
Experiments *in vitro*	
	Higher bactericidal capacity of rifampicin in "intermittent" incubation
	Bacillary survival in sealed cultures at 37uC after 12 yrs
	Adaptation of bacilli to a low oxygen pressure
	Higher resistance to stressful conditions in bacilli from stationary-phase cultures
	Transcription of genes (e.g. SigF) related to the sporulation cascade in other bacteria, under stressful conditions
	Dormant "nonculture able" bacilli can be "resuscitated" with phospholipids and a specific factor (Rpf) synthesised by growing bacilli
Experiments in animal models	
	The Cornell model in mice
	Bacilli from chronical lesions are better adapted to stressful conditions than the acute ones
	Specific *M. tuberculosis* gene knock-out strains have revealed the indispensable role of some genes expressed in stressful conditions for persistence in a chronic infection

TB: tuberculosis; Rpf: resuscitation promoting factor. Table is from (Cardona and Ruiz-Manzano, 2004).

1.2 *M. tuberculosis* and the disease process

Mycobacteria are Gram-positive, acid-fast, aerobic bacilli that are characterised by the mycolic acid content in their cell wall. Other bacteria, including Nocardia, Rhodococcus, and Corynebacterium have related properties (http://genolist.pasteur.fr/TubercuList/). The genus *Mycobacterium* is now known to comprise more than 100 species that are all closely related. *M. tuberculosis* is not the only etiological agent of TB, since other members of the *Mycobacterium tuberculosis* complex (MTC) can also cause tuberculosis. Members of the MTC include *Mycobacterium africanum, Mycobacterium bovis, Mycobacterium microti*, *Mycobacterium canetti, Mycobacterium pinnipedii* and *Mycobacterium caprae*. These slow-growing mycobacteria are very closely related and their 16s rRNA (i.e. ribosomal RNA) are 99.9% identical. While most members of the MTC rarely cause human infections, *M. africanum* and *M. bovis* are common causes of TB in Africa (Smith, 2003). *M. bovis* infection between cattle and human in the UK is thought to be via badgers as an additional host (reviewed in Corner *et al.* (2011)).

1.2.1 The *M. tuberculosis* genome

Knowledge of the complete *M. tuberculosis* genome sequence has enabled the identification of genes essential for its survival *in vitro* and *in vivo* (Cole *et al.*, 1998). Genome-wide DNA microarrays (Butcher, 2004; Li *et al.*, 2010) and mutational analysis, such as the Transposon site hybridization (TraSH) selection (Sassetti *et al.*, 2003; Beste *et al.*, 2009), were employed to study the effect of different growth conditions on gene expression. The *M. tuberculosis* H37Rv genome consists of 4.4×10^6 bp and contains approximately 4,000 genes (Cole *et al.*, 1998). Annotation of the *M. tuberculosis* genome shows that this bacterium has two unique features distinguishing it from other microorganisms (Figure 1.5). *M. tuberculosis* has around 250 lipid metabolizing enzymes, 200 genes of which encode enzymes for the metabolism of fatty acids (Figure 1.5). Among these are approximately 100 that are predicted to function in the β-oxidation of fatty acids,

compared to only 50 enzymes for such a purpose in *Escherichia coli* (*E. coli*) (Cole *et al.*, 1998). The large number of *M. tuberculosis* enzymes that putatively use fatty acids might play a role in the persistence of this organism in granuloma, where fatty acids may be the major carbon source (Section 1.3.1) (reviewed in Ehrt and Schnappinger (2007)). *M. tuberculosis* genome is also distinguished by the presence of unrelated families, of acidic, glycine- rich proteins (Cole *et al.*, 1998). These families are characterised by the presence of 110 and 180 amino acid conserved *N*-terminal domains with PE (Pro-Glu) and PPE (Pro-Pro-Glu) sequence motifs, respectively. The PE and PPE proteins (4 % of the total genome, Figure 1.5), which are cell-wall associated and surface exposed, are expected to play a role in *M. tuberculosis* virulence (Parra *et al.*, 2006; Khan *et al.*, 2008; Mukhopadhyay and Balaji, 2011). Otherwise, the genome of *M. tuberculosis* is comparable to that from other microorganisms (Cole *et al.*, 1998). Interestingly, the genome sequence of *M. bovis* is > 99.95 % identical to that of *M. tuberculosis* (Garnier *et al.*, 2003). However, the reduced genome size in *M. bovis* bacille Calmette-Guerin (BCG) is mainly attributed to gene deletion (Garnier *et al.*, 2003). Comparison with *M. leprae*, a close relative of *M. tuberculosis*, reveals a drastic gene reduction in *M. leprae* with about 70 % loss of the coding sequence, including the *nat* gene (Section 1.4.1). The lack of ability of *M. leprae* to survive outside hosts could be attributed to this massive genomic decay affecting nearly every aspect of metabolism (reviewed in Vissa and Brennan (2001)).

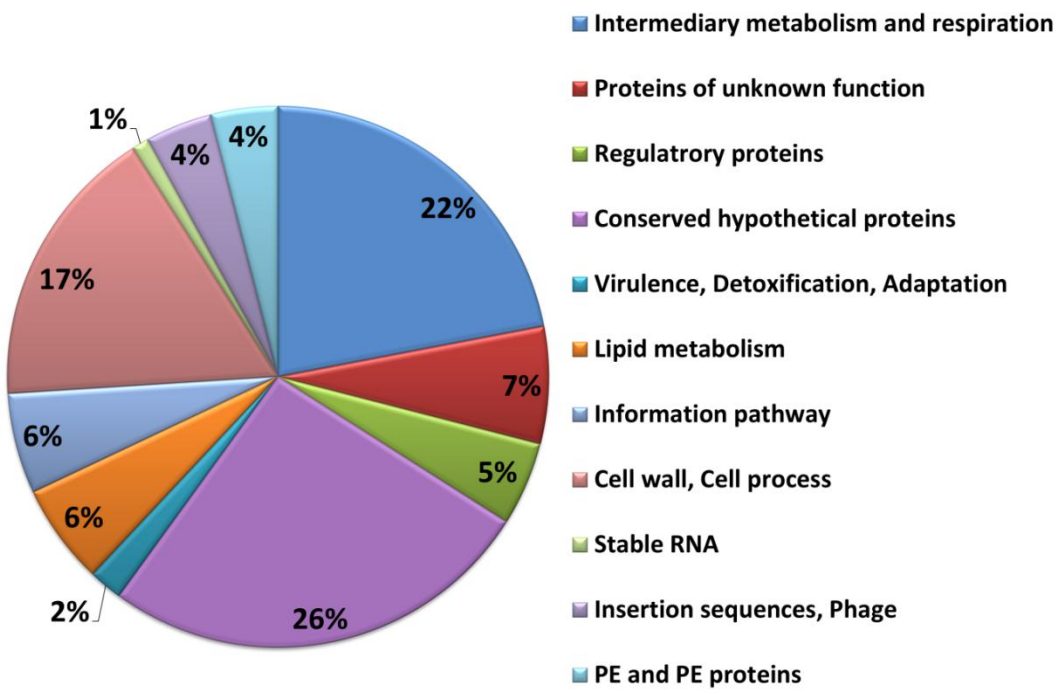


Figure 1.5: A pie chart representing the general classification of *M. tuberculosis* genes (Smith, 2003).

1.2.2 The cell envelope

The cell wall of mycobacteria is characterized by its unique architecture and unusual lipid constituents (Figure 1.6). Mycobacteria are classified as acid-fast Gram-positive bacteria due to the presence of a thick peptidoglycan layer and the lack of an outer cell membrane in their cell envelope. However, the covalently bound mycolic acid contents and the large variety of extractable lipids (usually referred to as virulence lipids) give mycobacteria their characteristic acid-fast properties (reviewed in Brennan (2003)). Mycolic acids, which constitute a major part of the waxy-cell wall, act as a permeability barrier, rendering the pathogen intrinsically resistant to many drugs.

The unique lipids of the mycobacteria cell wall, such as lipoarabinomannan (LAM) (Besra *et al.*, 1997), the cord factor trehalose dimycolate (CF) (Noll *et al.*, 1956) and phthiocerol dimycocerate (PDIM) (Minnikin *et al.*, 1983) have been shown to play an important role in the virulence and pathogenicity of *M. tuberculosis* (Chan *et al.*, 1991; Spargo *et al.*, 1991; Astarie-Dequeker *et al.*, 2009; Mishra *et al.*, 2011). LAM has been shown to contribute to the persistence of mycobacteria within the macrophage via its role in halting the phagosome maturation (Section 1.2.3) (Chan *et al.*, 1991; Mishra *et al.*, 2011). It has also been reported to play a protective role against oxygen free-radicals (Chan *et al.*, 1991). Disruption of the genes involved in the synthesis of these lipids resulted in attenuated mutants (Sirakova *et al.*, 2003). The most effective anti-tubercular agents, isoniazid and ethambutol inhibit mycolic acid and arabinan biosynthesis, respectively (Winder and Collins, 1970; Mikusová *et al.*, 1995; Karakousis, 2009).

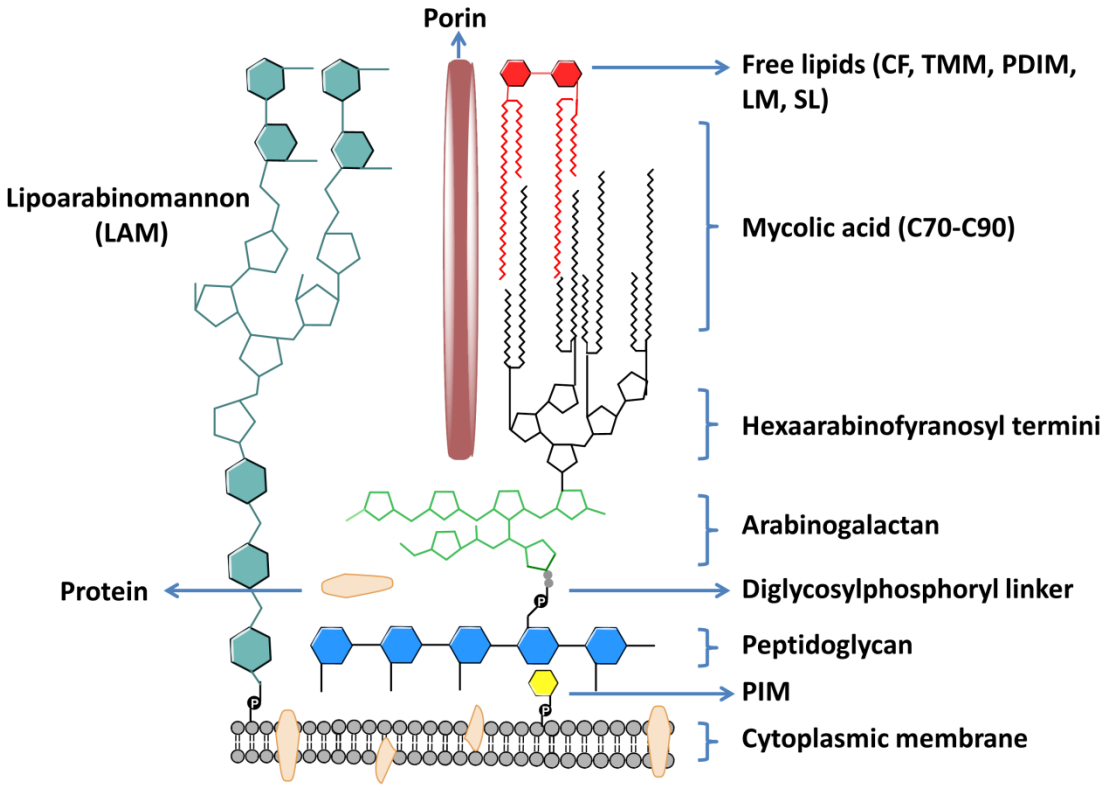


Figure 1.6: A schematic representation of the mycobacterial cell envelope.

The cytoplasmic membrane of mycobacteria is encapsulated by a layer of peptidoglycans. The peptidoglycan backbone is attached to arabinogalactan through an unusual disaccharide-phosphate-linker region. The arabinogalactan is a branched-chain polysaccharide consisting of a proximal galactose chain linked to a distal arabinose chain. Mycolic acids are covalently linked to the arabinogalactan-peptidoglycan co-polymer and are an essential component of the cell wall. Extractable lipids (free lipids) are represented in red. Another major component non-covalently associated to the mycobacterial cell wall is the immunogenic lipoarabinomannan (LAM), which is attached to the cytoplasmic membrane by a phosphatidylinositol anchor. The porin MspA mediates the uptake of small and hydrophilic nutrients such as sugars and phosphates, whereas hydrophobic compounds diffuse directly across the cell wall. PIM: phosphatidylinositol mannoside, CF: cord factor; TMM: trehalose monomycolate; PDIM: Phthiocerol dimycocerosates; LM: lipomannan; SL: sulfolipids. The figure and legend were modified from Fenton and Vermeulen (1996); and Mishra *et al.* (2011).

1.2.3 *M. tuberculosis*-host interaction and intracellular survival

Tuberculosis is a complex disease, the outcome of which depends on the dynamic interaction between pathogen and host. Genome-wide analysis using microarrays has allowed the identification of differentially expressed genes important in host-pathogen interactions (Butcher, 2004). Alveolar macrophages are likely to be the first cell type to encounter *M. tuberculosis* in a pulmonary infection. Despite the efficient mechanism within the macrophage to eradicate invading bacteria, *M. tuberculosis* has exquisitely evolved a number of very effective survival strategies within the otherwise microbicidal conditions. *M. tuberculosis* entry into the macrophages is mediated by cell-surface receptors, including CR1 and CR3 complement receptors as well as mannose receptors (Kang *et al.*, 2005; Carroll *et al.*, 2009). Cholesterol has also been shown to play an important role in the entry of mycobacteria into macrophages (Kaul *et al.*, 2004). However, the exact mechanisms by which *M. tuberculosis* enters the host cell and avoids host defences are not completely understood. It has been suggested that *M. tuberculosis* uses a variety of mechanisms to survive the hostile environment of the macrophage (reviewed in Meena and Rajni (2010)). Once formed, the mycobacterial phagosomes (MP; phagosome harbouring *M. tuberculosis*) gain the properties of early endosome (pH 6.1-6.5), fail to acidify, and become refractory to fusion with lysosomes (Sturgill-Koszycki *et al.*, 1994). The pathogenicity of *M. tuberculosis* is largely attributed to its ability to arrest phagosome maturation. A summary of the mechanisms involved in this process are shown in Figure 1.7. The ability of the *M. tuberculosis* lipids (e.g. LAM) to modulate intracellular signalling is believed to be a major key factor in the failure of MPs fusion with the lysosomes (reviewed in Warner and Mizrahi (2007); and Meena and Rajni (2010). While LAM prevents lysosomal fusion and acidification (Hmama *et al.*, 2004), PIMs (phosphatidylinositol mannosides) induce fusion with early endosomes to obtain the nutrients required for the phagosomal residence of mycobacteria (Vergne *et al.*, 2004). The virulent *M. tuberculosis* over-expresses several proteins that might also contribute to the

survival of the microorganism within the macrophage. Some of these proteins contribute to the ability of mycobacteria to escape the toxic effect of peroxides and nitrogen radicals within the macrophage (Senaratne *et al.*, 2006). Furthermore, the MP is usually characterised by a high content of tryptophan-aspartate containing coat protein (TACO). TACO is an actin binding protein known to associate with cholesterol within the plasma membrane of the host cell. The recruitment and retention of TACO on the phagosome wall are associated with a block in phagosome-lysosome fusion (Ferrari *et al.*, 1999). Moreover, TACO-mediated uptake of mycobacteria depends on cholesterol (Kaul *et al.*, 2004). The pivotal role of cholesterol in cell entry and intracellular survival is discussed in the following section.

The intracellular persistence of mycobacteria also benefits from their inherent slow growth rate and their ability to convert into a dormant state once inside the macrophage. These mechanisms will prevent mycobacteria from outgrowing their host macrophages, which would destroy their habitat (reviewed in Pieters and Gatfield (2002)).

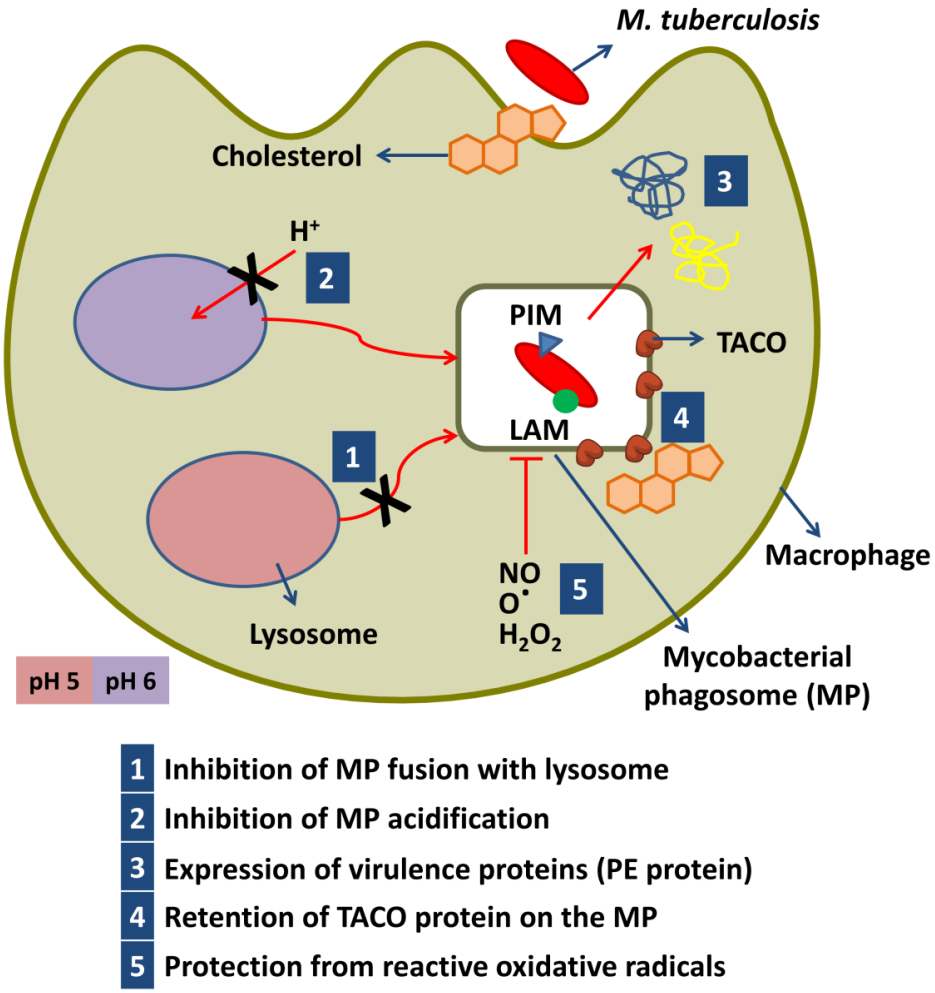


Figure 1.7: Strategies used by *M. tuberculosis* to modulate phagosome maturation.

After internalization, the bacterium uses an array of effector molecules, including the lipids phosphatidylinositol mannoside (PIM) and lipoarabinomannan (LAM) to arrest phagosome maturation. Retention of the TACO protein prevents bacterial delivery to the lysosome. TACO is a host protein that is known to be associated with cholesterol. Cholesterol has also been shown to play a role in the entry of mycobacteria into the macrophage. PE-proteins are (Pro-Glu) repetitive glycine-rich proteins. MP: mycobacterial phagosome. The figure and legend were modified from Meena and Rajni (2010).

1.3 The role of cholesterol

As discussed earlier, cholesterol is essential for the entry of mycobacteria into macrophages. This feature is specific for mycobacteria and is very likely to involve the presence of a cholesterol-binding site at the mycobacteria cell wall. Furthermore, *M. tuberculosis* is capable of using cholesterol as a carbon source. The cholesterol degradation pathway in mycobacteria has been proposed based on the determination of the genes involved in cholesterol catabolism in the *Rhodococcus* species (Figure 1.8) (Van der Geize *et al.*, 2007). Most of these genes in *M. tuberculosis* have been found to be under the control or the *kstR* regulon, which encodes a TetR-like transcriptional repressor. Several genes in the *kstR* cholesterol regulon are induced in macrophages or are essential for infection, underscoring the role of cholesterol catabolism in intracellular survival (Kendall *et al.*, 2007; Kendall *et al.*, 2010). Examples of the operons that affect cholesterol entry and catabolism in mycobacteria are shown in Figure 1.9. Among these genes, the products of the *nat* operon are the focus of this study (Figure 1.10).

It is known that *M. tuberculosis* uses fatty acids as a carbon source during chronic infection. The catabolism of fatty acid affects the propionate pool in mycobacteria and augments the production of virulence lipids (Savvi *et al.*, 2008). Cholesterol makes a potentially important contribution to the metabolic pool of propionyl-CoA (Pr-CoA) in mycobacteria (Pandey and Sassetti, 2008; Yang *et al.*, 2009). Moreover, cholesterol metabolism in mycobacteria has recently been shown to increase the average mass of the lipid virulence factor PDIM. Pr-CoA is converted to methylmalonyl-CoA (Mm-CoA), which is considered to be the building block of multimethyl-branched mycolic acids such as PDIM (Yang *et al.*, 2009) (Figure 1.11).

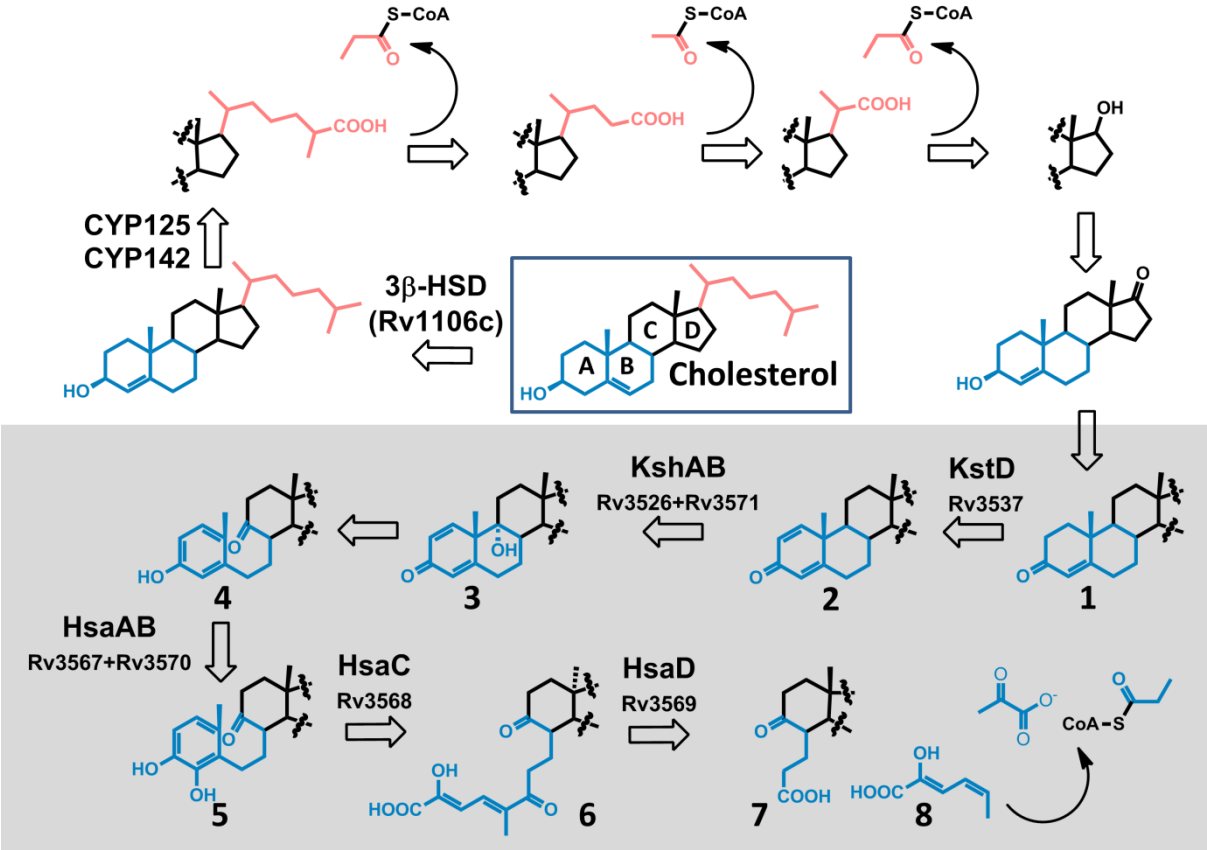


Figure 1.8: Proposed degradation pathways of the cholesterol aliphatic side chain and ring nucleus.

The carbon atoms derived from the ring nucleus are converted to CO_2 via the tricarboxylic acid cycle (TCA), whereas the Pr-CoA produced from the degradation of the side chain is assimilated into mycobacterial lipids (e.g. PDIM). The enzymes and the cholesterol metabolites are described in the text. The aliphatic side chain (pink) degradation is shown with no background while the sterol ring (blue) degradation is shown with a grey background. The fate of rings C and D (black) is still unknown. Numbers correspond to the intermediates mentioned in the text. The figure and legend were modified from Ouellet *et al.* (2011).

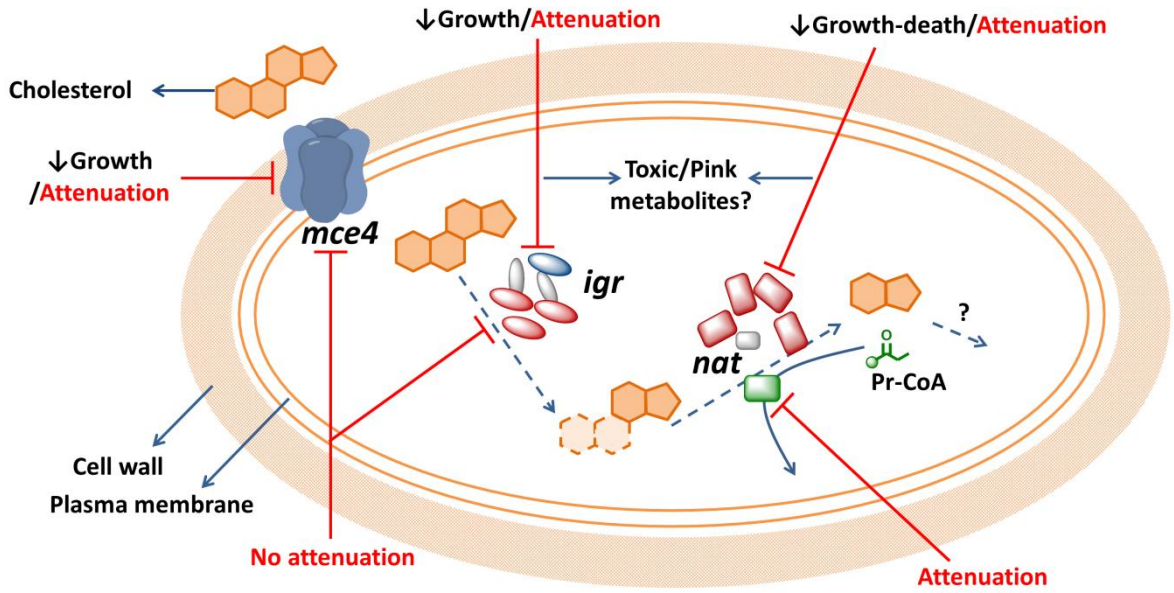


Figure 1.9: The role of essential putative operons in the intracellular survival of *M. tuberculosis* in macrophages (*mce4*, *igr* and the *nat*) in cholesterol metabolism.

The diagram is showing the established roles of the *mce4*, *igr* and the *nat* operons in cholesterol uptake and metabolism. The effect of deleting each of these operons or genes within the operon on the growth of *M. tuberculosis* in cholesterol is shown in black text, while effect on attenuation of infection in macrophage infection or in a mouse model is shown in red text. Cholesterol is transported into *M. tuberculosis* via the Mce4 transport system. The *igr* operon consists of six genes, the most important of which is cytochrome P450 (*cyp125*). Inactivation of the *igr* operon resulted in growth defects in cholesterol attenuation in a mouse model, an effect that is prevented by mutating the sterol uptake Mce4 system. The *nat* operon consists of six genes, including an arylamine-*N*-acetyl transferase (*nat*), four *hsaACDB* genes and a hypothetical protein. The *nat* gene product NAT utilises Pr-CoA, a cholesterol degradation product. Products of uncertain identity are shown as question marks. The dashed arrows indicate a multistep process. Gene products with enzymic activity are shown in red, the NAT enzyme is shown in green, hypothetical proteins in grey and transport proteins in blue. The figure was prepared using the ChemBioOffice software. The figure and legend are based on Flannagan *et al.* (2009) (Miner *et al.*, 2009).

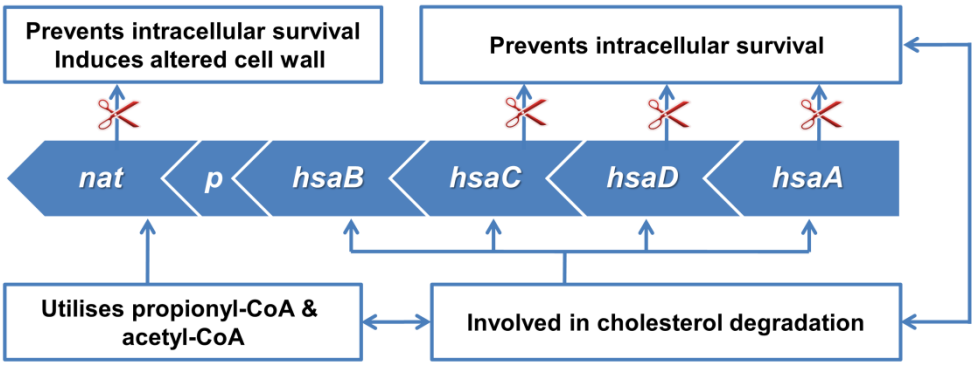


Figure 1.10: *nat* operon in *M. tuberculosis* and *M. bovis* BCG.

The accession numbers, described in http://genolist.pasteur.fr/TubercuList/, for these genes in *M. tuberculosis* H37Rv are as follows: Rv3570c (*hsaA*), Rv3569c (*hsaD*), Rv3568c (*hsaC*), Rv3567c (*hsaB*), Rv3566A (possible pseudogene) and Rv3566c (*nat*). The operon is virtually identical in *M. bovis* and *M. bovis BCG*. The scissors indicate gene knockouts. The figure is based on Bhakta *et al.* (2004); and Lack *et al.* (2009). *p*: pseudo gene.

1.3.1 Cholesterol catabolism and the *nat* operon

The cholesterol catabolic pathway in *M. tuberculosis* involves two major phases, the initial degradation of the aliphatic side chain and the subsequent degradation of the sterol A–D rings (Figure 1.8) as discussed in Ouellet *et al.* (2011). The putative degradation of the aliphatic side chain by β-oxidation reactions results in the formation of one acetyl-CoA (Ac-CoA) and two Pr-CoA molecules (Figure 1.8). The Pr-CoA derived from the side chain has been shown to be incorporated into PDIM. Several enzymes involved in this pathway have been elucidated recently including the CYP125, a product of the *igr* operon (Figure 1.9). However, the sterol ring degradation is more complicated, since it involves a large number of enzymes and is not yet completely understood (Figure 1.8). The oxidation of rings A and B by two putative dehydrogenases, 3β-HSD and KstD (intermediates 1 and 2; Figure 1.8), followed by hydroxylation by an α-hydroxylase KshAB (compound 3; Figure 1.8) leads to the aromatisation of ring A into a phenolic intermediate (compound 4; Figure 1.8), preparing the molecule for the action of the *hsaACDB* products of the *nat* operon (Figure 1.8) (reviewed in Ouellet *et al.* (2011)). The *hsaA* and *hsaB* genes encode an oxygenase and a reductase, respectively, that act together as a flavin-dependent monooxyenase that hydroxylates the phenolic metabolite into a catecholic intermediate (compound 5, Figure 1.8). The next enzyme in the pathway is HsaC, an iron-dependent extradiol dioxygenase that oxygenates and cleaves the catecholic intermediate to a β-hydroxydienoic acid derivative (compound 6; Figure 1.8) referred to as 4,9-DSHA (4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid). This intermediate is the substrate of the next enzyme in the *nat* operon, HsaD, which was identified as one of the *M. tuberculosis* genes required for its survival in macrophages (Rengarajan *et al.*, 2005). This enzyme catalyzes hydrolytic carbon–carbon bond cleavage of 4,9-DSHA to yield 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid (compound 7; Figure 1.8) and 2-hydroxy-hexa-2,4-dienoic acid (HHD, compound 8; Figure 1.8). HHD is finally metabolised to tricarboxylic acid cycle intermediates and Pr-CoA. However, the fate of the

C ring and D ring fragment of cholesterol has not yet been established (Ouellet *et al.*, 2011). Interestingly the *nat* operon encodes an arylamine *N*-acetyltransferase (NAT), hence the name, that was found to utilise Pr-CoA in addition to Ac-CoA as an acyl donor, both of which are products of the cholesterol degradation (Lack *et al.*, 2009). Whole gene studies (Sassetti *et al.*, 2003) and targeted gene approaches showed (Yam *et al.*, 2009; Lack *et al.*, 2010) each of the *hsaACDB* or the *nat* gene (Bhakta *et al.*, 2004) prevent intracellular survival and affects growth in cholesterol. The 3D-structures of HsaA, HsaC and HsaD are now available and are expected to facilitate inhibitor design (Lack *et al.*, 2008; Yam *et al.*, 2009; Dresen *et al.*, 2010). Determination of the 3D-structure of the NAT enzyme is a further approach to gain a full picture of the role of the *nat* operon and also for inhibitor design.

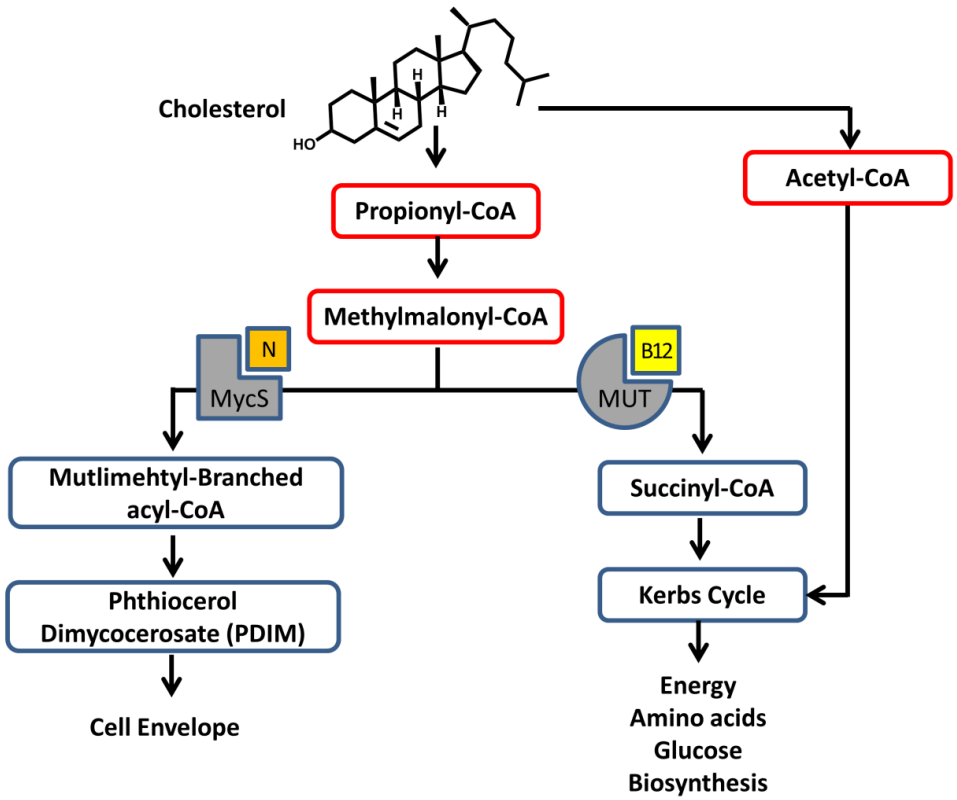


Figure 1.11: Flux of metabolites from cholesterol catabolism.

Methylmalonyl-CoA (Mm-CoA) is primarily derived from Pr-CoA, a substrate formed from the catabolism of cholesterol. MUT-B12 is a vitamin B12-dependent Mm-CoA mutase that catalyses the isomerization of Mm-CoA to succinyl-CoA, a key molecule in the Krebs (tricarboxylic acid, TCA) cycle. MycS: mycocerosic acid synthetase; N: NADPH. The figure was modified from Yang *et al.* (2009).

1.4 The arylamine *N*-acetyltransferases

Arylamine *N*-acetyltransferase (NAT) is a cytosolic enzyme that is found in *M. tuberculosis* and humans (Section 1.4.4) as well as in many other organisms (Sim *et al.*, 2008a). This enzyme catalyses the transfer of an acetyl group from Ac-CoA to an arylamine substrate through a conserved cysteine residue by a Ping-Pong bi-bi mechanism (Sinclair *et al.*, 2000) (Figure 1.12). The NAT enzyme has been identified in many mycobacterial species. However, neither the biological substrate, nor an endogenous role for the *N*-acetylation activity of NAT has been reported. It has been suggested that the enzyme plays a role in maintaining the homeostasis of Ac-CoA, a central metabolite in lipid synthesis, (Bhakta *et al.*, 2004). The *nat* genes from *M. tuberculosis* and *M. bovis* BCG are identical and are encoded in highly similar operons in both organisms (the *nat* operon). Interestingly, it has been established that the *nat* operon is essential for the survival of *M. bovis* BCG within the macrophage, although the gene-knockout strain retains the ability to grow on rich media (Section 1.4.1) (Bhakta *et al.*, 2004; Sim *et al.*, 2008a).

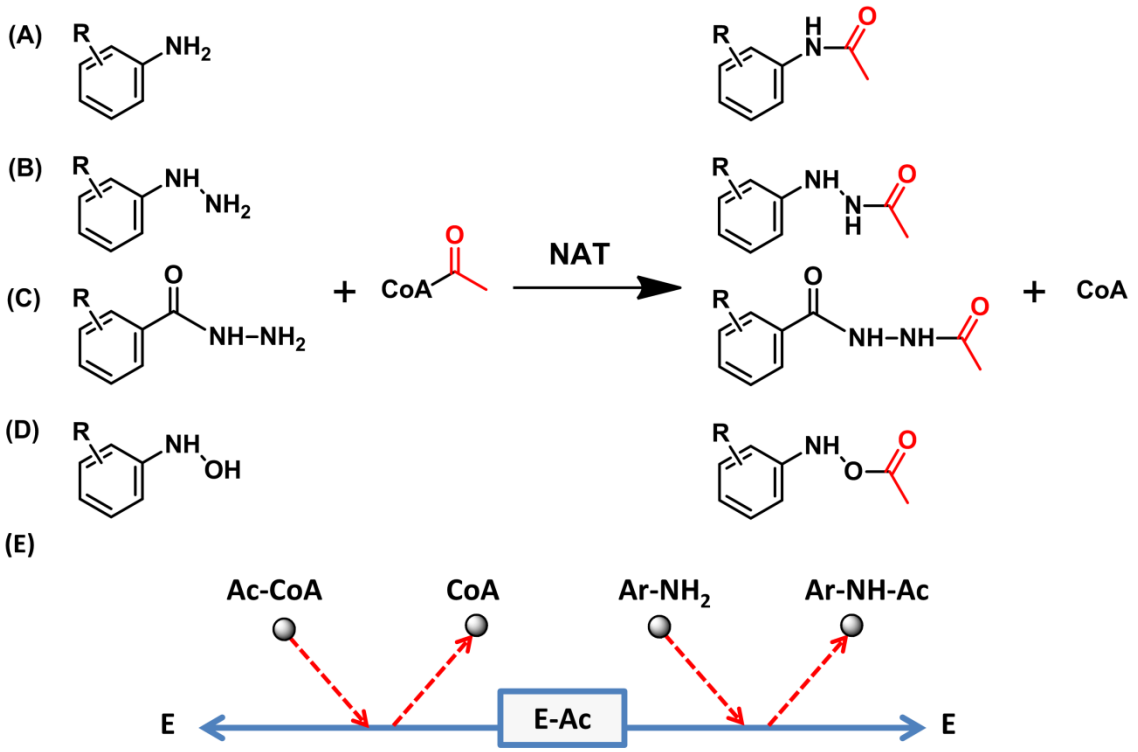


Figure 1.12: Reactions catalysed by arylamine *N*-acetyltransferases and the mechanism of the catalysis.

The range of acetyl acceptors that can be catalysed by NAT includes: (A) arylamines (e.g. *p*-aminobenzoic acid), (B) hydrazines (e.g. hydralazine), (C) hydrazides (e.g. isoniazid) and (D) hydroxylamines. (E) Schematic representation of the Ping-Pong mechanism of the NAT catalysis. Ar-NH_2: arylamine, Ar-NH-Ac: acetylated arylamine, E: enzyme, and E-Ac: acetyl-enzyme intermediate.

1.4.1 NAT as a drug target

Despite the ability of NAT to acetylate, and thus inactivate the first-line TB drug therapy INH, the role of the enzyme in isoniazid resistance is only marginal. Polymorphisms of the *nat* gene from *M. tuberculosis* clinical isolates have been observed. However, the gain-of-function mutation, that would be expected to induce isoniazid resistance, has not been found (Upton *et al.*, 2001; Coelho *et al.*, 2011).

Deleting the NAT enzyme in *M. bovis* BCG, which is identical to that from *M. tuberculosis,* resulted in delayed growth and caused morphological changes of the BCG bacilli (Figure 1.13). Moreover, the *nat* gene knockout resulted in depleted mycolic acids and virulence-lipid content (PDIM and CF; Section1.2.2) (Figure 1.14 and Figure 1.15, respectively). The ability of the enzyme to utilise Pr-CoA, the intermediate cofactor in virulence lipid synthesis, gives a good explanation of these findings. These effects render the mycobacteria sensitive to gentamicin and hygromycin that have weak activity against mycobacteria (Bhakta *et al.*, 2004). The part this enzyme plays in the virulence of mycobacteria as well as intracellular survival has made it an attractive target for TB drug therapy. NAT inhibitors have been identified by an elaborated high-throughput screening (HTS) of a 5000-compound chemical library using the homologous NAT enzymes that were available at the time (Westwood *et al.*, 2011). These enzymes include NATs from *M. smegmatis* (Sandy *et al.*, 2002), *S. typhimurium* (Sinclair *et al.*, 2000), *P. aeruginosa* (Westwood *et al.*, 2005) and *M. marinum* (Fullam *et al.*, 2009). Hits with inhibitory activity against the enzyme which showed antimycobacterial activity have been identified (compounds A-D Figure 1.16) (Westwood *et al.*, 2010). The identified inhibitors have been addressed by structure-based medicinal chemistry efforts to improve their potency and antimycobacterial activity as well as to study their structure-activity relationship (SAR) (Fullam, 2007; Westwood *et al.*, 2010; Fullam *et al.*, 2011). Among these inhibitors, compound, A-1 showed the best antimycobacterial activity (Fullam, 2007).

Efforts targeting compound B-1 (a β-amino alcohol) with medicinal chemistry, although they led to an improved understanding of the SAR, were fruitless in identifying compounds with improved activity (Fullam *et al.*, 2011). Detailed study of the NAT inhibition by compound C-1 and its analogues has also been performed and resulted in a more thorough understanding of the SAR of these compounds. The general observation from these SAR studies was that introducing a large lipophilic substituent in the structure would result in improved inhibition activity. A similar result was obtained from the SAR study of Thiazolidinedione-sultams (E-1; Figure 1.16), that were identified as NAT inhibitors from a separate screening of a proprietary small-molecule library but were not further pursued due to their lack of antimycobacterial activity (Brooke *et al.*, 2003a).

Interestingly, the NAT inhibition by compound C-1 resulted in similar changes in cell-wall lipids, morphology and intracellular survival to those observed upon deleting the gene (compound C-1; Figure 1.14 and Figure 1.15) (Westwood *et al.*, 2010). Treatment with compound C-1 resulted in a massive depletion in the amount of the mycolic acid and complex lipids in *M. bovis* BCG (Figure 1.15). Furthermore, C-1 acted synergistically with gentamycin, mimicking the effect of the *nat* knockout on the same mycobacteria (Bhakta *et al.*, 2004). Nevertheless, many of the identified inhibitors were ineffective in inhibiting the NAT from *M. tuberculosis*, despite the high similarity of TBNAT to the enzymes used in the process of screening and evaluation.

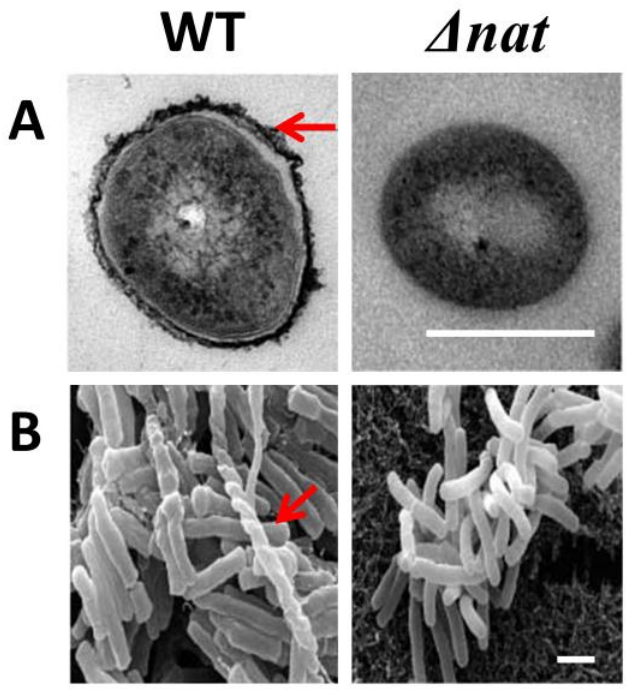


Figure 1.13: The effect of deleting the *nat* gene on the morphology and ultrastructure of *M. bovis* BCG.

Morphological changes resulted from deleting the *nat* gene (Δ*nat*). (A) Cross sectional transmission electron microscopy (TEM) images show that the size of *M. bovis* BCG (WT) is altered on deleting the *nat* gene. The outer cell wall (arrow), present in *M. bovis* BCG (WT), is absent in the Δ*nat* strain. (B) The scanning electron microscopy (SEM) images also show the difference in size of the bacilli. Cord formation (arrow) in *M. bovis* BCG is missing when the *nat* gene is deleted. Bar is 1 µm for all frames. WT: wild type. The figure and legend are from Bhakta *et al.* (2004).

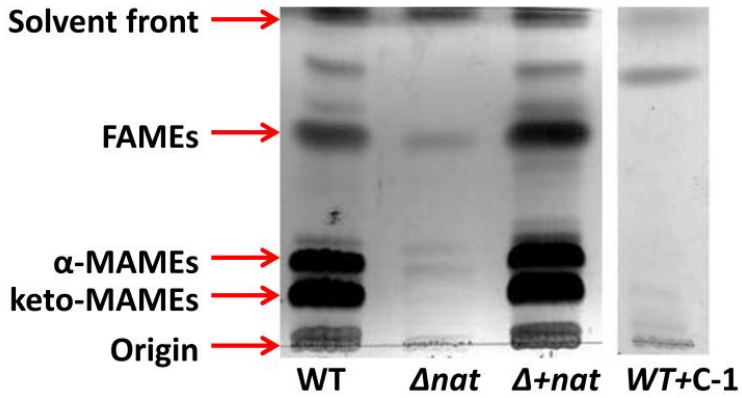


Figure 1.14: The effect of deleting the *nat* gene on the mycolate profile in *M. bovis* BCG.

Separation of mycolates from the same dry weight of de-lipidated cells by one-dimensional TLC from *M. bovis* BCG shows that the synthesis of mycolate in WT is perturbed by deletion of the *nat* gene and is fully restored when the *M. tuberculosis nat* gene is introduced (Δ+ *nat*). BCG cells grown in the presence of the NAT inhibitor C-1, that was identified by HTS, showed a similar effect as the gene knockout. MAMEs: mycolic acid methyl esters; FAMEs: fatty acid methyl esters. The figures and legend were compiled from Bhakta *et al.* (2004); and Westwood *et al.* (2010).

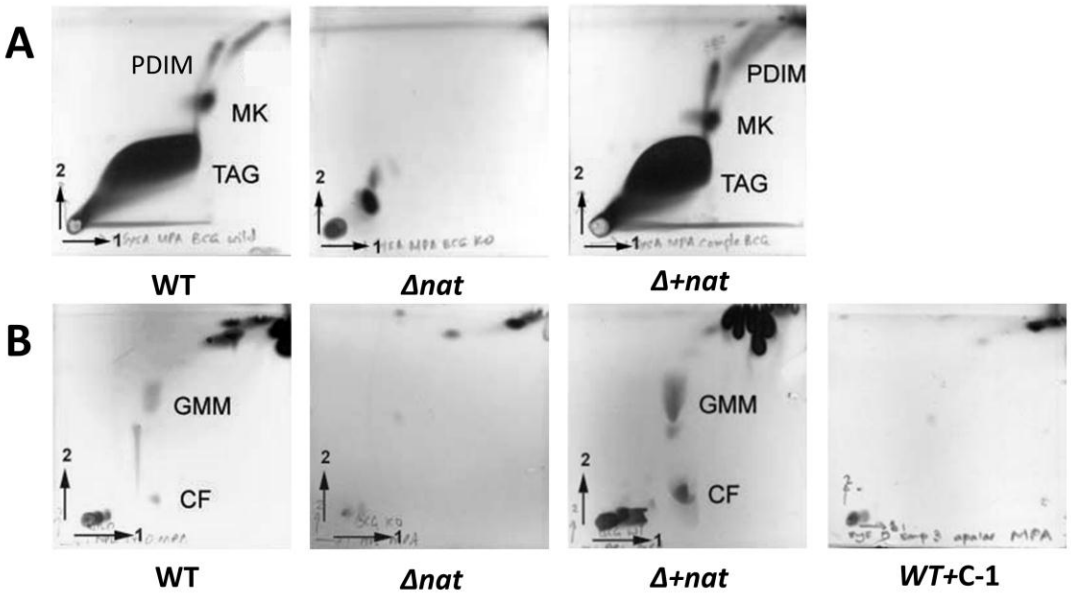
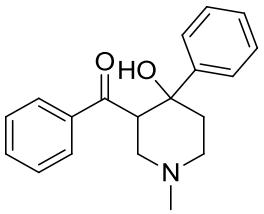
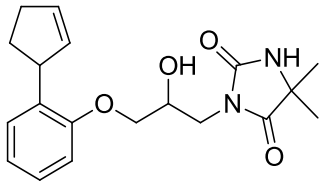


Figure 1.15: The effect of deleting the *nat* gene on the virulence lipids on *M. bovis* BCG

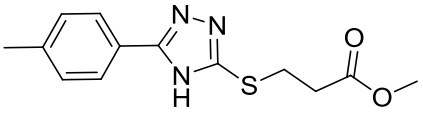
Analysis of total lipids by two-dimensional TLC shows that (A) triacyl glycerol (TAG), menaquinone (MK), and PDIM, as well as (B) CF and GMM, are present in *M.bovis* BCG (WT), but missing from the corresponding Δ*nat* BCG strain. All of these complex lipids are restored by complementation with *nat* (Δ+ *nat*). BCG cells grown in the presence of the NAT inhibitor C-1 showed similar effects as the gene knockout. The figures and legend were compiled from Bhakta *et al.* (2004); and Westwood *et al.* (2010).



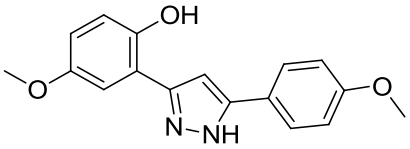
A-1
MIC < 5 µg/mL
Piperidinol



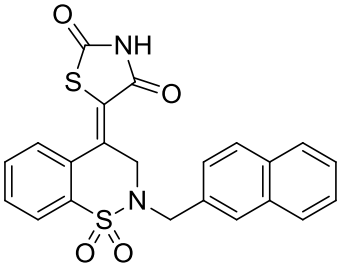
B-1
MIC 50 µg/mL
β-amino alcohol



C-1
MIC < 10 µg/mL
1,2,4 triazole-3-thione



D-1
MIC < 10 µg/mL
pyrazole



E-1
IC_{50} 78 µM
N-substituted TZD-sultam

Figure 1.16: The chemical scaffolds of NAT inhibitors.

Compounds A-D were identified by HTS against several mycobacterial NATs and their minimum inhibitory concentration (MIC) values against *M. tuberculosis* are shown (Westwood *et al.*, 2010). The MIC values against *M. tuberculosis* H37Rv of NAT inhibitors were determined as described in Anderton *et al.* (2006). Compound E-1 was identified from a separate screen using a proprietary small-molecule library (Brooke *et al.*, 2003a) but showed no inhibition of *M. tuberculosis*, which is likely to be due to lack of permeability. The IC_{50} was determined by measuring the ability of the compound to inhibit MSNAT-catalysed acetylation of INH (Brooke *et al.*, 2003a).

1.4.2 Structural models of NAT.

NAT enzymes, which are found in a wide range of organisms, were proposed through enzymic studies to have an active-site cysteine residue (Weber and Hein, 1985). In view of the small quantity of the enzyme from *M. tuberculosis* (TBNAT) that was available, NATs from other bacterial and mycobacterial sources have been used for structural studies (Table 1.3). The crystallographic structure of NAT from *Salmonella typhimurium* was the first mapping of the Cys-His-Asp catalytic triad that is essential for activity in all NATs (Sandy *et al.*, 2005b). This structure has also supported the proposed role of these residues in the acetylation reaction (Figure xx) (Sim *et al.*, 2012).

NATs from *M. smegmatis* (MSNAT) and from *M. marinum* (MMNAT) have been used as models for the TBNAT protein. In addition to the high sequence similarity (Figure 1.18), the enzymes show similar but not identical substrate profile to TBNAT. However, MMNAT is considered the best available model for TBNAT (Fullam *et al.*, 2008). The *M. smegmatis* is a fast-growing, non-pathogenic species that has been used for many years as a laboratory model of *M. tuberculosis* (Altaf *et al.*, 2010), whilst *M. marinum* is a waterborne mycobacterial pathogen that can cause granulomatous infection in fish and occasionally in humans (Bouley *et al.*, 2001; Stamm and Brown, 2004).

Studies on NATs from these mycobacteria aided in identifying the substrate (isoniazid) binding site (Sandy *et al.*, 2005a), as well as the Ac-CoA binding pocket (Fullam *et al.*, 2008). The overall NAT structure consists of three domains. Domain I, II and III form an α-helical bundle, a β-barrel and an α/β lid, respectively (Figure 1.19). Assembled together, the molecular surface of these domains presents a deep and wide active-site cleft, at the base of which founds the catalytic cysteine residue (Figure 1.19).

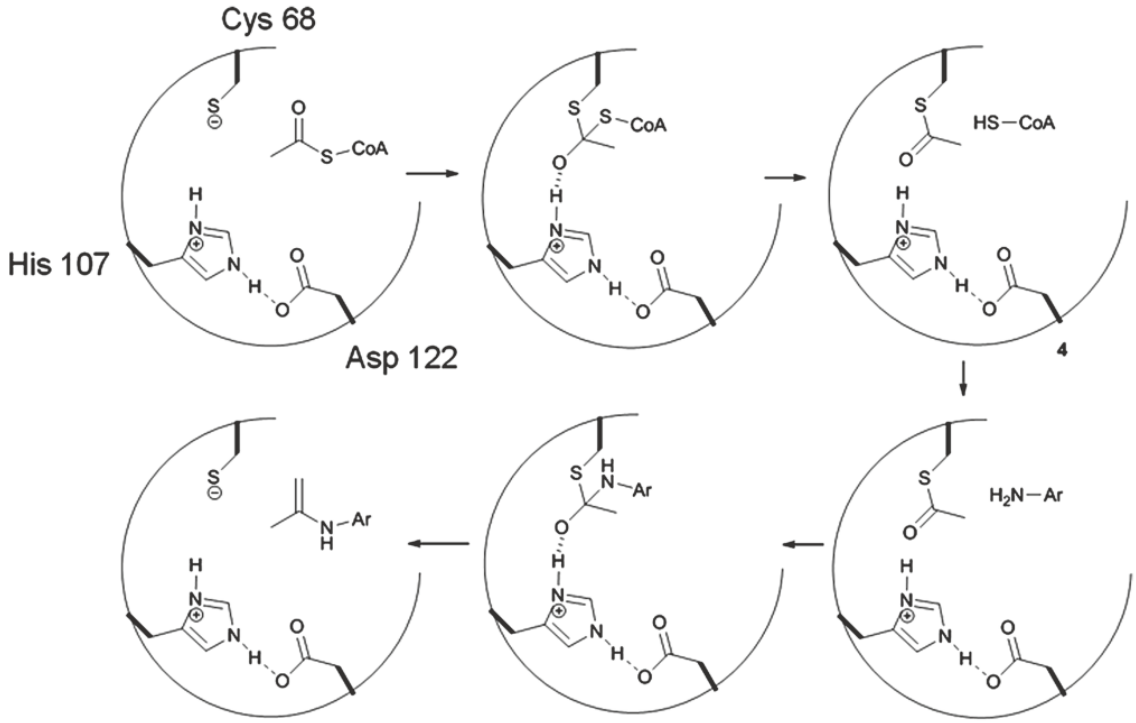


Figure 1.17: Reaction mechanism of arylamine *N*-acetyltransferases.

The proposed mechanism is shown with acetyl-CoA as the acyl donor. The residue numbering refers to the human NAT1 amino acid sequence. The formation of the acetyl-enzyme intermediate requires the stabilization of charge across a Cys-His-Asp triad. The acetyl transfer occurs following nucleophilic attack of the acetylated intermediate by the arylamine substrate. The diagram is from (Sim *et al.*, 2012).

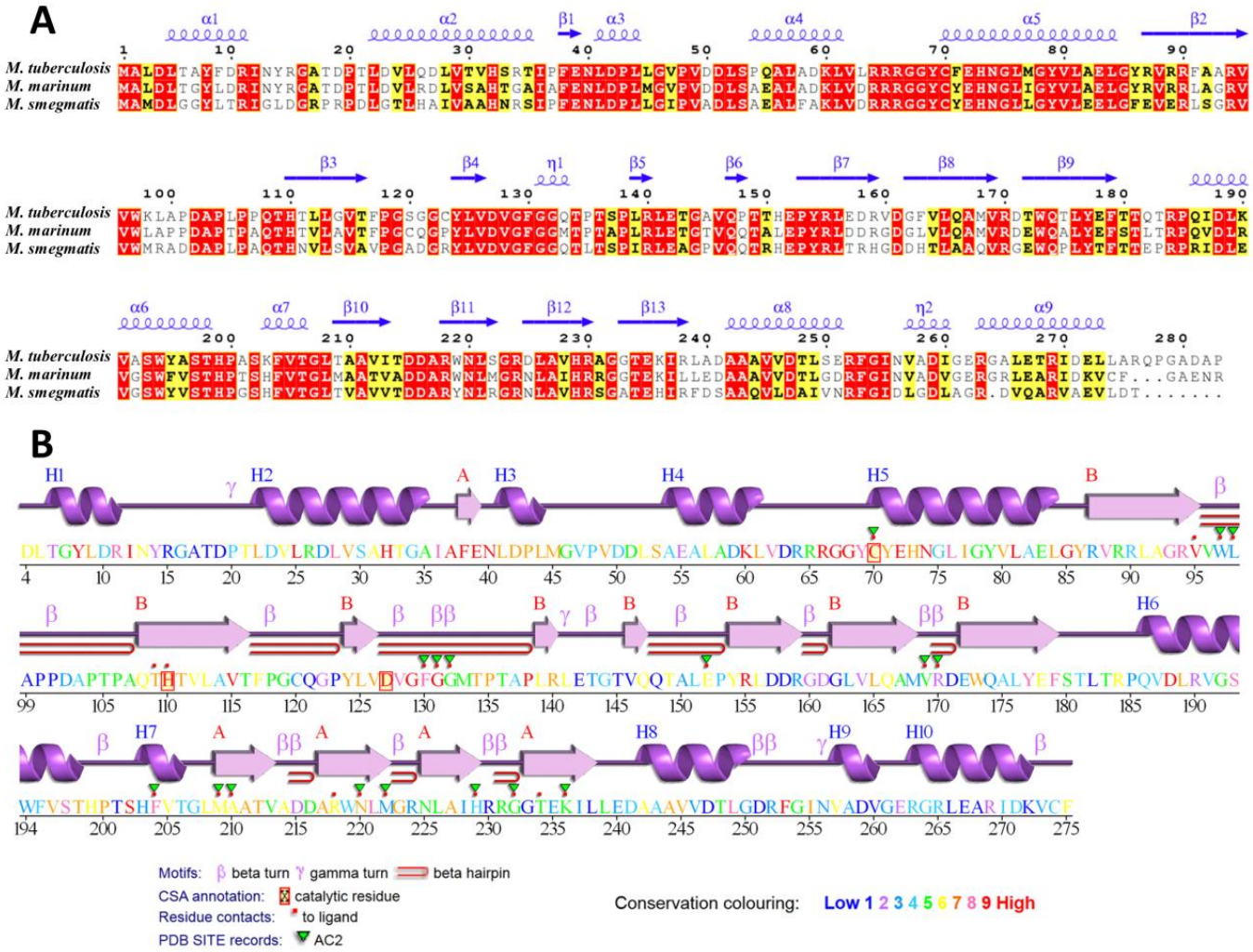


Figure 1.18: Amino acid sequence alignment of mycobacterial NATs used as models for the TBNAT.

(A)The sequence alignment of NATs from *M. tuberculosis* (TBNAT), *M. marinum* (MMNAT) and *M. smegmatis* (MSNAT) was generated by using ClustalW2 (Thompson *et al.*, 1994). Residues strictly conserved have a red background, and similar residues are indicated by black bold letters with a yellow background according to a Risler matrix implemented in ESPript (Gouet et al., 1999). The symbols above the sequence correspond to the secondary structure of MMNAT (PDB code 2VFB). The dashes indicate deletions, since MMNAT (280 a.a) and MSNAT (276 a.a) are shorter than TBNAT (283 a.a). The figure was prepared using ESPript 2.2 (Gouet et al., 1999). (B) Diagram showing the secondary-structure elements of MMNAT (2VFC) superimposed on its sequence. The labeling of secondary-structure elements is in accord with PDBsum (http://www.ebi.ac.uk/pdbsum). The sequence is coloured according to the degree of sequence conservation. The catalytic residues and binding pocket residues are highlighted.

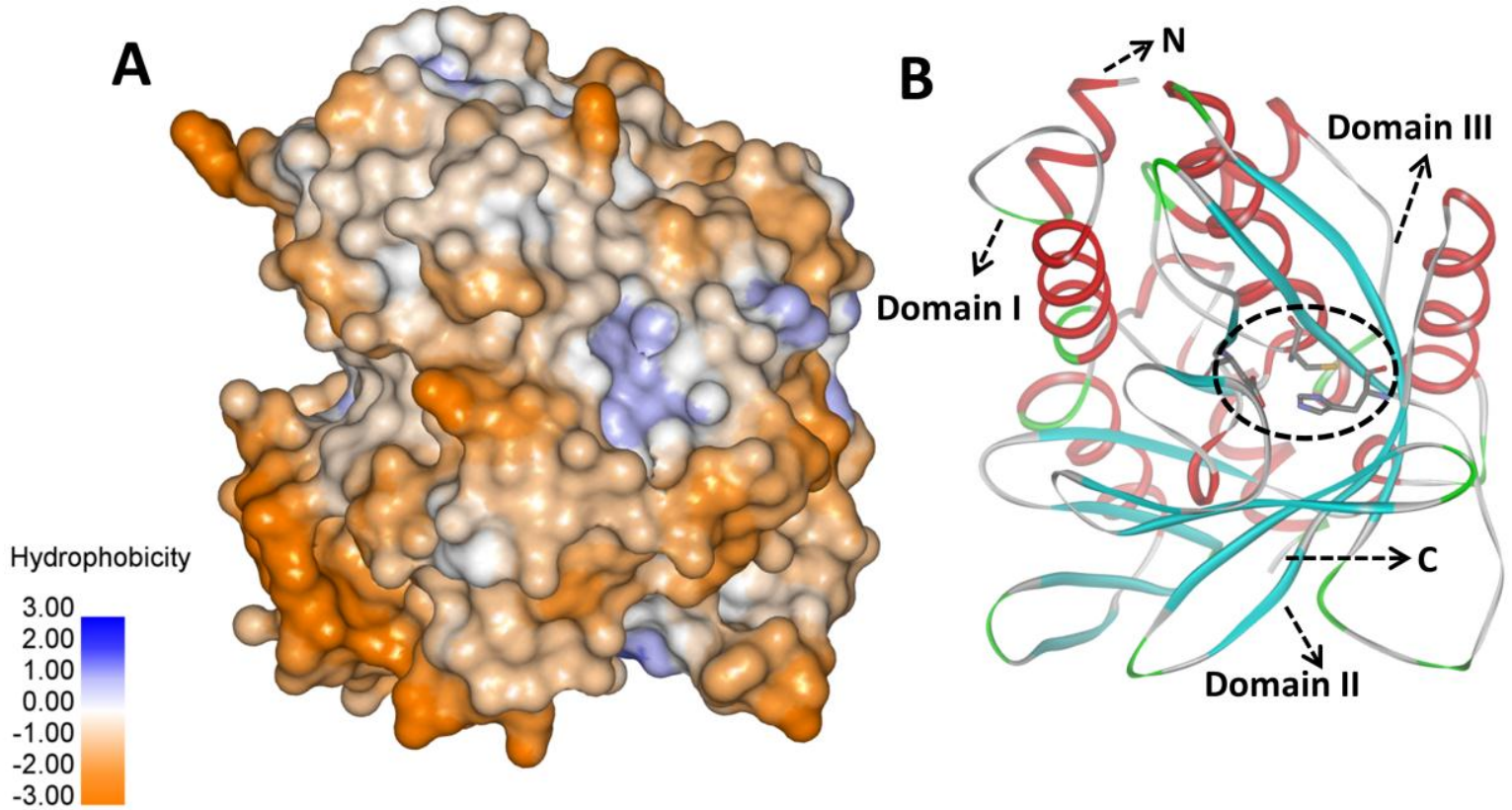


Figure 1.19: The 3D-structure of MMNAT.

(A) A hydrophobicity surface plot of the MMNAT apoprotein (PDB code 2VFB; 2.0 Å). (B) Ribbon representation showing the 3D-structure of MMNAT active site and the three subdomains. The structure is coloured according to the secondary structure. The black circle indicates the position of the catalytic triad (Cys70, His110 and Asp127) and a molecular presentation of their side chains is indicated by the scheme on the shaded area.

Table 1.3: The NAT structures available in the Protein Data Bank.[a]

Source of NAT	PDB code	Resolution (Å)	Modifications	Co-crystallised ligand	Sequence identity (%)	Structural similarity (%)	RMSD (Å)	Reference
M. marinum	2VFB	2.00	---	---	74			Fullam *et al.* (2008)
	2VFC	2.70	---	CoA				
M. smegmatis	1GX3	1.70	---	---	60	99	1.12	Sandy *et al.* (2002)
	1W6F	2.10	---	INH				Sandy *et al.* (2005a)
	1W5R	1.45	C70Q	---				Sandy *et al.* (2005b)
P. aeruginosa	1W4T	1.95	---	---	35	96	2.09	Westwood *et al.* (2005)
Mesorhizobium loti	2BSZ	2.00	---	---	41	98	1.32	Holton *et al.* (2005)
S. typhimurium	1E2T	2.80	---	---	32	97	1.63	Sinclair *et al.* (2000)
Nocardia farcinica	3D9W	2.70	---	---	32	93	2.02	Martins *et al.* (2008)
Bacillus anthracis C	3LNB	2.01	Y38F	CoA	21	97	2.03	Pluvinage *et al.* (2011)
Homo sapiens I	2PQT	1.78	2-bromo-acetanilide C70	---	26	88	1.98	Wu *et al.* (2007)
	2IJA	1.70	F125S	---				Wu *et al.* (2007)
Homo sapiens II	2PFR	1.92		CoA	23	90	2.24	Wu *et al.* (2007)

[a] Protein-sequence similarity of each protein to TBNAT is based on ClustalW2 alignment of amino-acid sequence (Thompson *et al.*, 1994). Percentage 3D-structure similarity and RMSD relative to MMNAT (PDB code 2VFB) were estimated by the jFATCAT- rigid algorithm of the RCSB PDB Protein Comparison Tool (Prlić *et al.*, 2010).

1.4.3 Thermal stability of NAT from *M. tuberculosis*

TBNAT has been reported to be a stable protein that retains full activity at high temperatures (up to 60 ºC) using circular dichroism and enzyme activity methods to monitor unfolding (Lack *et al.*, 2009). *M. tuberculosis* itself is a mesophilic bacillus with an optimal growth temperature of 37 ºC. However, this microorganism produces several thermostable proteins. HsaD, which is a product of the same operon as TBNAT, is another example of these proteins (Lack *et al.*, 2009). Interestingly, both enzymes have been shown to be essential for intracellular survival of mycobacteria within the macrophage, a stage at which mycobacteria exhibited a stationary-phase and showed resistance to high temperature (53 ºC) (Wallace, 1961). The high enzymatic heat stability of TBNAT has not been found in the NAT enzymes from any other species, including the closely related organisms, *M. smegmatis* and *M. marinum.* The melting temperature of MSNAT (i.e. temperature resulted in 50 % unfolded protein, T_m) is 50-55 ºC (Kawamura *et al.*, 2003) whilst that of MMNAT is 40-45 ºC (Fullam, 2007; Lack *et al.*, 2009). Both enzymes have been used as models of TBNAT. The first and second domains of NAT from *M. tuberculosis* appear to contribute strongly to the thermal stability of the protein, with a chimera in which domains I and II from MMNAT were genetically stitched to domain III of TBNAT causing a dramatic reduction in the heat stability (Lack *et al.*, 2009). In the first two domains of TBNAT, a number of amino acids are different to those in both MMNAT and MSNAT. In TBNAT, the residues are of a more hydrophobic character, for instance, replacement of an alanine (Ala54 and Ala106) or glutamine (Gln148) residue by proline and replacement of a glutamate by an uncharged residue (Glu55 by glutamine, Glu63 by leucine and Glu172 by threonine) (Lack *et al.*, 2009). Increased occurrence of proline, the most rigid of the amino acids, usually has a stabilizing effect on proteins (Kumar *et al.*, 2000); therefore, the increase in the number of proline residues is likely to contribute to the high stability of TBNAT.

1.4.4 Human arylamine *N*-acetyltransferase

Unlike mycobacteria, eukaryotes and some prokaryotes (e.g. *B anthracis*) have multiple NATs. In humans, there are two functional *nat* genes: *NAT1* and *NAT2*. The products of these genes HNAT1 and HNAT2, respectively, catalyse the acetylation of xenobiotics. HNAT2 is known to acetylate (and hence inactivate) of isoniazid (Figure 1.20) (Donald *et al.*, 2004). Polymorphism of *NAT2* is responsible for the slow and fast acetylator phenotypes observed in the metabolism of isoniazid and hydralazine. (reviewed in Sim *et al.* (2003); and Sim *et al.* (2012)). HNAT1's and HNAT2's substrate specificity profile (Figure 1.21) and pattern of expression, reflects their distinct functional roles (reviewed in Sim *et al.* (2008b)). HNAT2's substrate specificity resembles that of TBNAT but is distinct (Figure 1.21).

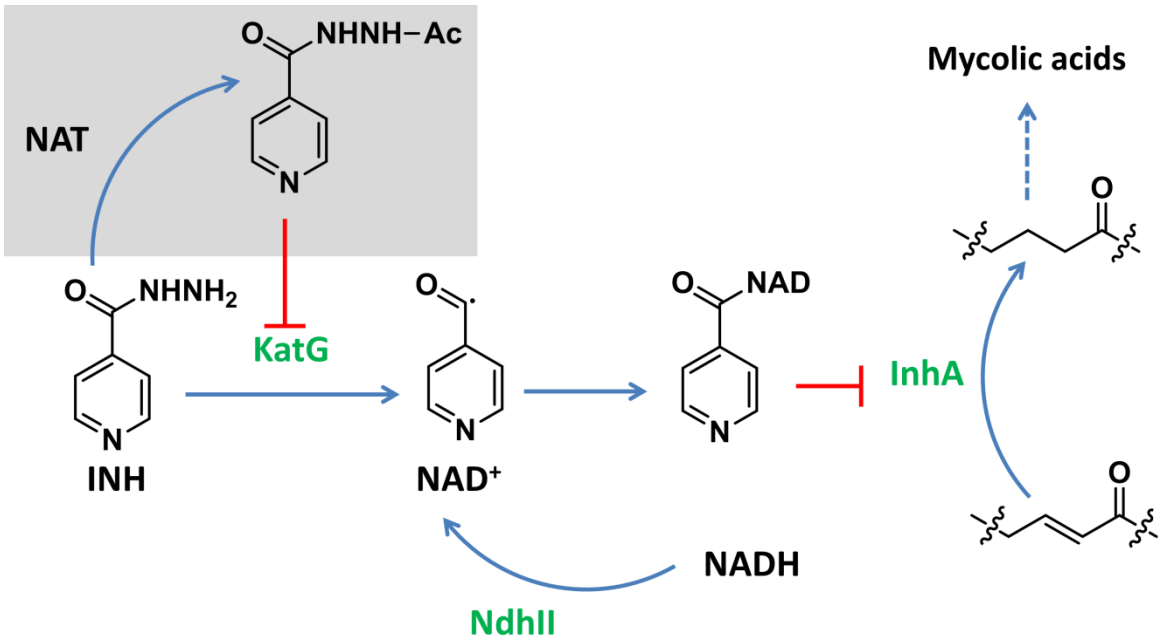


Figure 1.20: Inactivation of the first-line TB therapy isoniazid (INH) by the NAT enzyme.

The mechanism of INH activation by the mycobacterial enzymes (enzymes are shown in green): Isoniazid is a prodrug which requires activation by the catalase-peroxidase hemoprotein, KatG. The activated form of INH forms a covalent adduct with NAD^+ to generate a potent inhibitor of the InhA protein of *M. smegmatis*, an enoyl-acyl carrier protein reductase important in the mycolic acid biosynthesis. The acetylation by human NAT (HNAT2) (shaded area) prevents the activation of INH and hence reduces its mycobacterial activity. Similar inactivation of INH by TBNAT has only a marginal effect on INH resistance. The figure and legend were modified from Vilcheze *et al.* (2005).

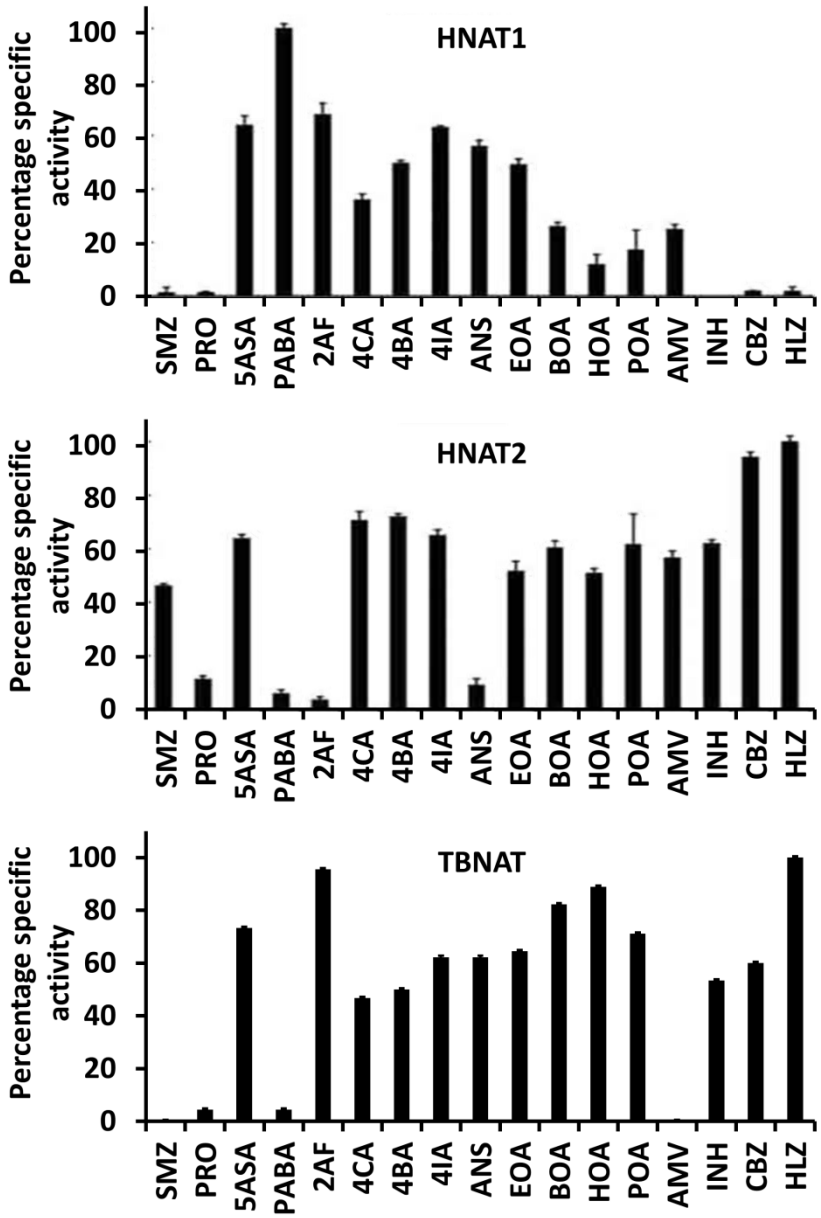


Figure 1.21: Comparison of the relative specific-activity profiles of HNAT1, HNAT2 and TBNAT.

Specific activities are presented as percent compared to the most active substrate for each enzyme. SMZ, sulfamethazine; PRO, procainamide; 5ASA, 5-aminosalicylic acid; PABA, *p*-aminobenzoic acid; 2AF, 2-aminofluorene; 4CA, 4-chloroaniline; 4BA, 4-bromoaniline; 4-IA, 4-iodoaniline; ANS, 4-anisidine; EOA, 4-ethoxyaniline; BOA, 4-butoxyaniline; HOA, 4-hexyloxyaniline; POA, 4-phenoxyaniline; AMV, 4-aminoveratrole; INH, isoniazid; CBZ, 4-chlorobenzoic hydrazide; HLZ, hydralazine. The figures of HNAT1 and HNAT2 and legend are from Westwood *et al.* (2006) and figure of TBNAT is modified from Fullam (2007).

HNAT1 has been investigated as a biomarker and a drug target for therapy for breast cancer treatment (Russell *et al.*, 2009), and has been shown to play a role in folate metabolism (Wakefield *et al.*, 2007). Interestingly, potent micromolar inhibitors of HNAT1 did not exhibit any inhibition of MMNAT or MSNAT (Laurieri *et al.*, 2010). The opposite was also true; since the HTS hit inhibitors of the mycobacterial NATs (Figure 1.16) did not show significant inhibition of the human enzymes (Fullam, 2007; Westwood *et al.*, 2010). The strategy of using the eukaryotic proteins in the HTS aimed at minimising the likelihood of cross enzyme inhibition and subsequent toxicity (Westwood *et al.*, 2011). Furthermore, structural studies on both HNAT2-CoA (2PFR) (Wu *et al.*, 2007) and MMNAT-CoA (2VFC) (Fullam *et al.*, 2008) complexes showed distinct binding sites for CoA in these two enzymes (Figure 1.22). Comparing the overlay of the two structures gives the insight that Ac-CoA approaches the active site cysteine from a different direction, and supports the distinction made between the two binding pockets (Figure 1.22) (reviewed in Sim *et al.* (2008b)). A loop region in HNAT2, which is absent in the prokaryotic enzymes, and its *C*-terminus overlap with the CoA-binding pocket in MMNAT (Figure 1.22) (Fullam *et al.*, 2008). By considering all these differences, it should be possible to design NAT inhibitors that are selectively toxic to mycobacteria.

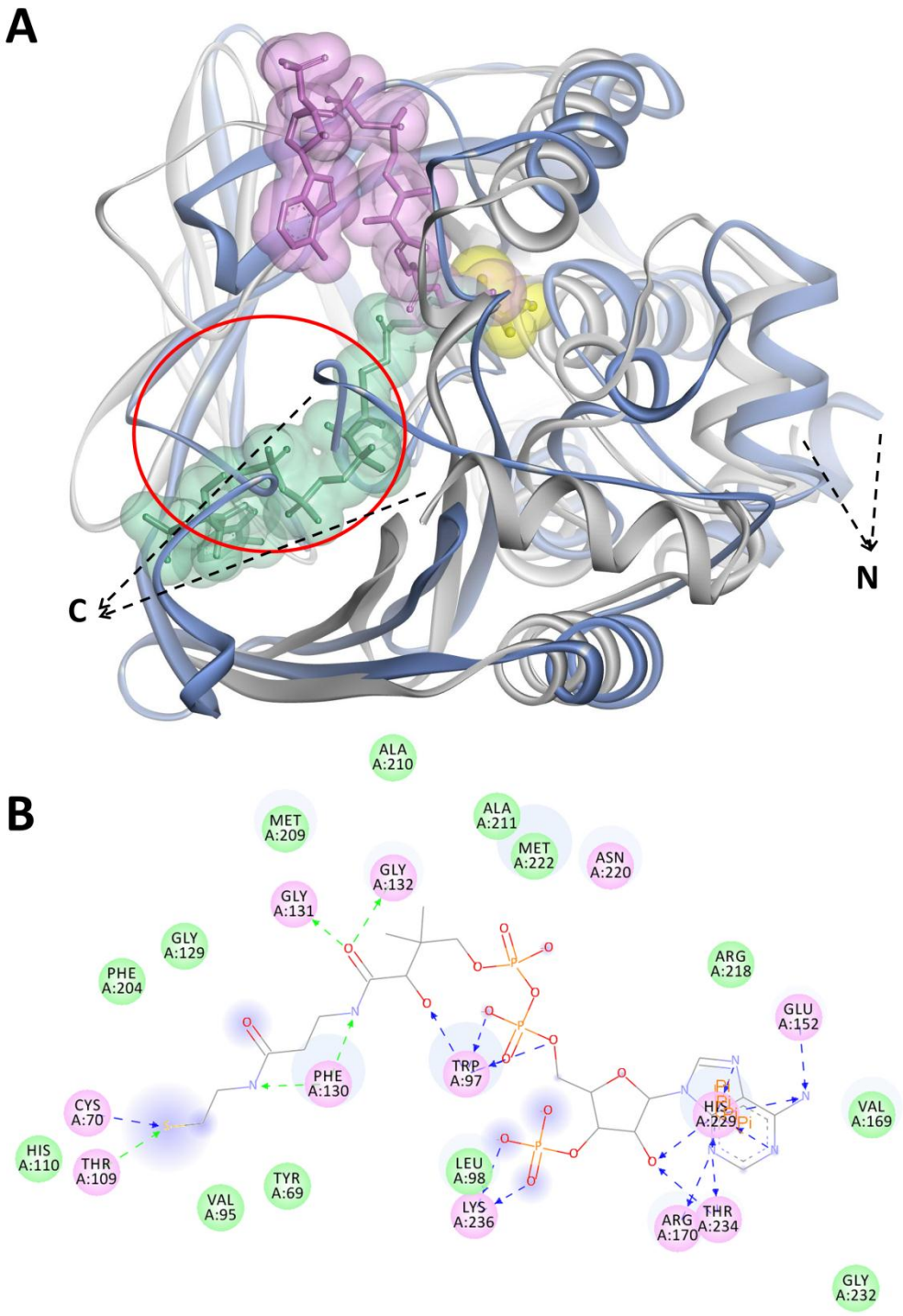


Figure 1.22: Comparison of the CoA binding site of prokaryotic and eukaryotic NAT enzymes.

(A) A ribbon presentation of the superimposition of the HNAT2-CoA complex (blue ribbons and pink CoA, 1.92 Å, PDB code 2PFR) and the MMNAT-CoA complex (grey ribbons and green CoA, 2.7 Å, PDB code 2VFC), showing the difference in the position of the CoA binding pocket. The active site cysteine is shown in yellow. The loop of HNAT2 and its *C*-terminal which overlap with the MMNAT binding pocket are indicated by the red circle. The figure was prepared using the DS visualizer. (B) Ligand interaction plot of the co-crystallised CoA molecule in the MMNAT-CoA complex. The plot depicts the 2D spatial arrangement of the ligand with respect to its key interactions. The proximity contour (dashed lines) and solvent exposed areas (solid purple spheres) of the ligand atoms are indicated as the polar (pink), hydrophobic (green), and solvent-exposed (light blue shadow) binding pocket amino acids.

1.5 TB drug development

1.5.1 Current therapy and drug resistance

Drugs available for the treatment of tuberculosis can be classified into two categories (Figure 1.23): first-line drugs such as, isoniazid (INH), rifampin (RIF), pyrazinamide (PZA), ethambutol (EMB), and second-line drugs (e,g. kanamycin, cycloserine, ethionamide, etc (Figure 1.23)) (reviewed in Zhang (2005)). Current TB therapy, also known as DOTS (directly observed treatment, short-course) consists of an initial phase of treatment with four drugs, INH, RIF, PZA and EMB, for two months daily, followed by treatment with INH and RIF for another four months, three times a week (WHO, 2011). The targets of these drugs are summarised in Figure 1.24. Although the current TB treatment is effective against drug-susceptible bacilli, the continuation of treatment for at least six months is essential in order to kill the persistent or slow growing strains (reviewed in Sarkar and Suresh (2011)). However, therapy failure is usually attributed to a lack of compliance with this long-term therapy. The treatment can also be complicated because of the side effects of these drugs and their drug-drug interactions, especially when administered with antiretroviral drugs. The emergence of multi-drug resistant (MDR) and extensively-drug resistant (XDR) TB further complicates the situation (Figure 1.23) (reviewed in Shenoi and Friedland (2009)). The patients with MDR and XDR have to be treated with second-line drugs (Figure 1.23), which are less effective, more toxic, and expensive (50-200 times the cost of first-line drugs). Recently totally-drug-resistant TB (TDR) has been reported (Udwadia *et al.*, 2011; Rowland, 2012).

The search for novel drugs that can help shorten the treatment course has become pressing in the light of the shortcomings of the current therapy. Most of the current efforts are guided by the WHO's Stop TB Strategy and involves research funded by governments, charitable foundations and non-governmental organisations (Squire *et al.*, 2006; Lienhardt *et al.*, 2012). The TB Alliance is one of these organisations that "has created and advanced

the largest TB drug pipeline in history" according to Ginsberg (2011). The mechanisms of action of new compounds that have entered clinical trials are shown in Figure 1.24. Moxifloxacin, a fluoroquinolone compound, is now in phase III trials for its use as a first-line drug and is expected to shorten the duration of treatment of pulmonary TB. In addition to the fluoroquinolones, there are six completely novel drugs now in clinical development (Figure 1.24) (reviewed in Cole and Riccardi (2011)). The most promising of these, TMC-207 is a diarylquinoline which inhibits mycobacterial ATP synthase (Andries *et al.*, 2005; Diacon *et al.*, 2009). Several other compounds have qualified for preclinical development, for example, the benzothiazinones, which inhibit cell-wall synthesis (Makarov *et al.*, 2009). Interestingly, the cholesterol catabolism inhibitor econazole has been shown to exhibit antimycobactrial activity against persistent and MDR strains of *M. tuberculosis* in a murine model (Ahmad *et al.*, 2006a; Ahmad *et al.*, 2006b). However, none of the current investigational compounds specifically targets cholesterol catabolism in mycobacteria or products of the *nat* operon. Therefore, the development of novel inhibitors targeting these enzymes would provide new therapeutic options for latent and MDR TB, and might limit the development of multidrug-resistant strains.

Despite the extensive use of the 100-year old vaccine *M. bovis* BCG, it has failed to limit the spread of the disease. BCG immunization shows better effectiveness in children compared to adults. In children, it provides 80 % protection against severe and disseminated tuberculosis (Trunz *et al.*, 2006). However, the vaccine has variable efficacy in adults (0-80 %) in different populations (Colditz *et al.*, 1994). New vaccines are in clinical development, but none has been approved for clinical use (reviewed in McShane (2011)).

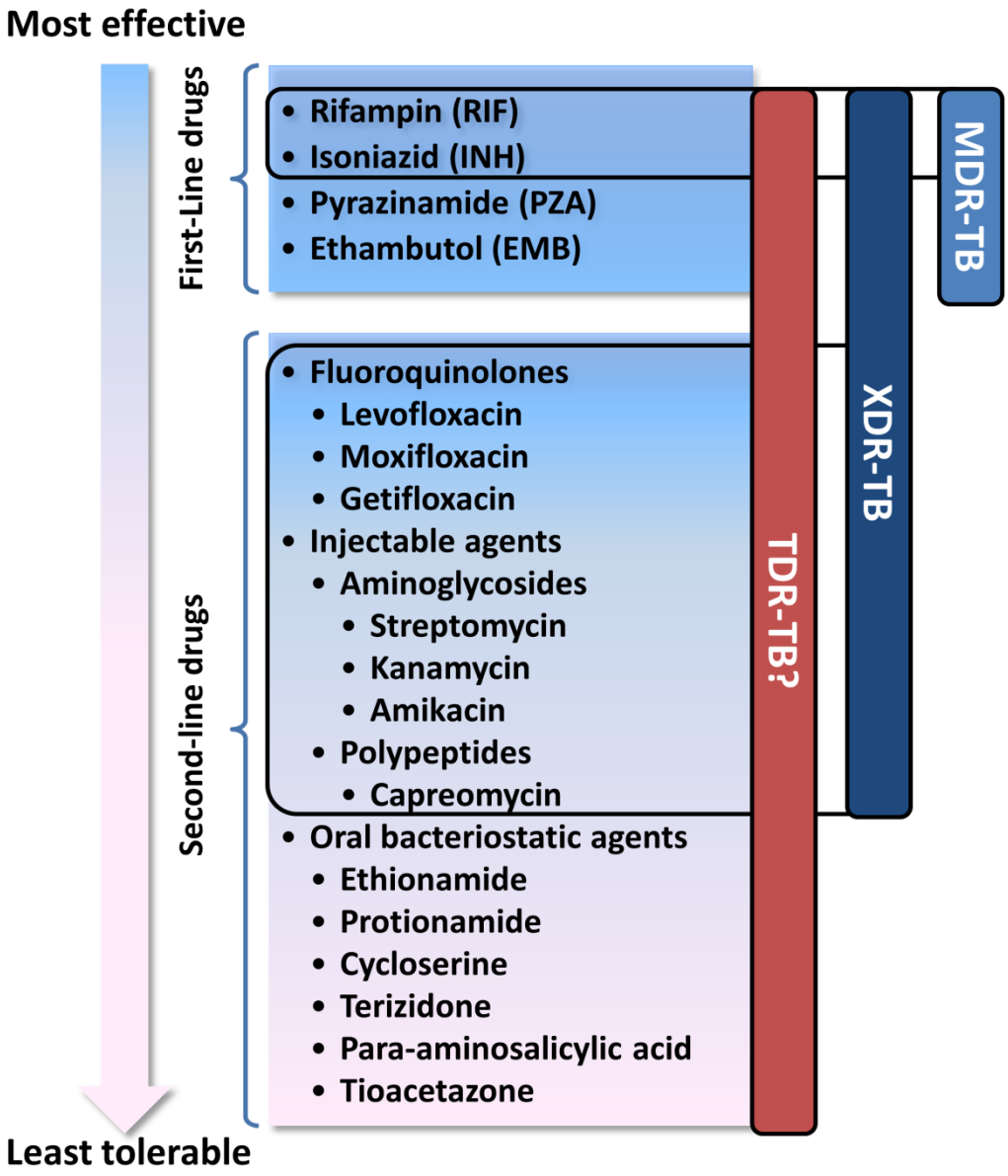


Figure 1.23: Drugs used in the current TB treatment.

The four drugs in the standard regime of first-line drugs (DOTS) are shown. Also shown are the classes of second-line drugs, which are reserved for the treatment of tuberculosis in special situations in which there is drug intolerance or resistance to first-line therapy. Multi-drug-resistant TB (MDR-TB) is caused by strains that are insensitive to isoniazid (INH) and rifampin (RIF), while extensively-drug-resistant TB (XDR-TB) is a MDR-TB with further resistance to both the fluoroquinolones and the injectable agents. Untreatable totally-drug-resistant TB (TDR-TB) has been reported recently. The figure and legend were modified from Dorman and Chaisson (2007).

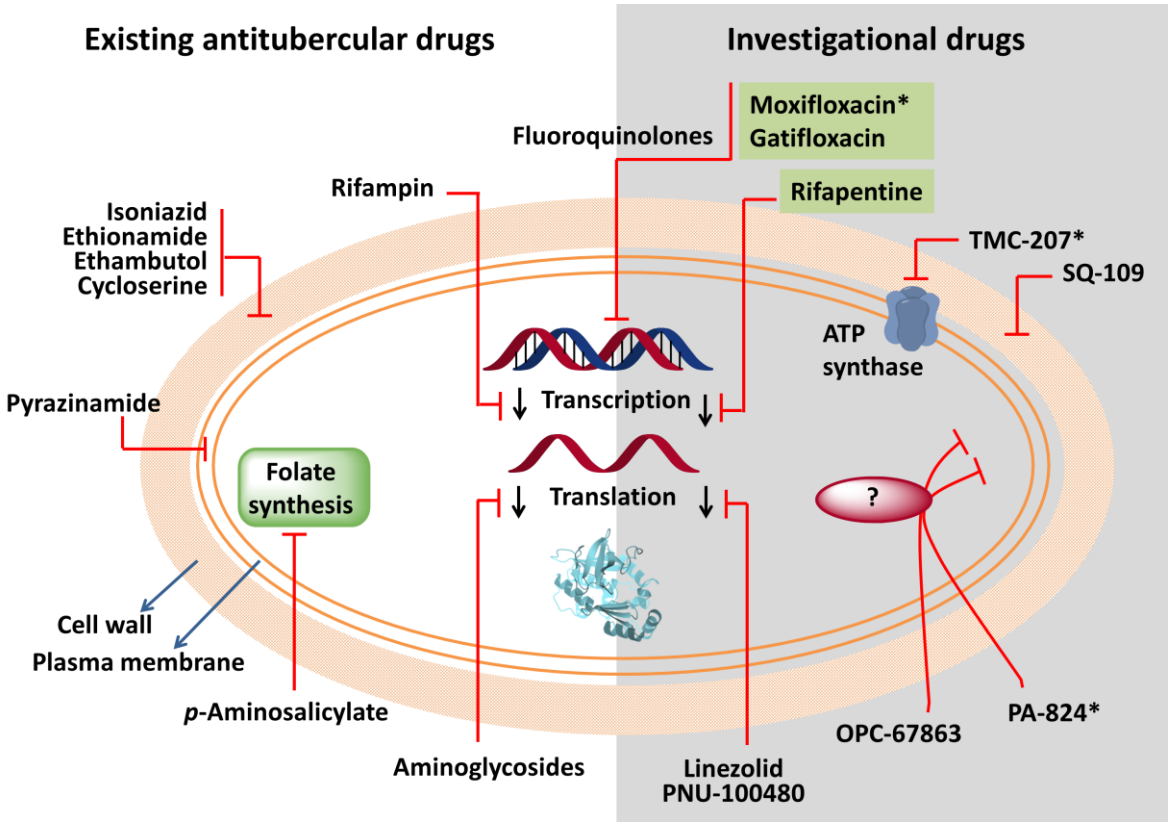


Figure 1.24: Mechanisms of action for current and investigational tuberculosis drugs.

Targets of current drugs include cell-wall synthesis (isoniazid, ethionamide, ethambutol and cycloserine), folate synthesis (p-aminosalicylate), transcription (rifampin), translation (aminoglycosides), DNA metabolism (fluoroquinolones) and the cell membrane (pyrazinamide). New compounds in clinical trials target other bacterial functions. TMC207 seems to inhibit the ATP synthase complex. SQ-109 inhibits cell-wall synthesis, and linezolid and PNU-100480 affect protein synthesis. OPC-67863 and PA-824 are prodrugs, the activation of which depends on the same cellular enzyme. The ultimate targets of these compounds remain unknown. Moxifloxacin, getifloxacin and rifapentine (green background) are in phase III clinical trials, whereas the rest are in phase II. The asterisks denote projects that are being developed or co-developed by the TB Alliance. The figure and legend were modified from Sassetti and Rubin (2007).

1.5.2 Challenges and prospects for new therapy

The combination of the special nature of *M. tuberculosis* and the complicated pathophysiology of the disease represent the inherent challenges facing TB drug development. The lack of suitable animal models that can mimic this complex disease profile, especially granuloma, latent disease and reactivation, is another challenge for understanding the disease process and screening for drugs (reviewed in Kaneko *et al.* (2011); and Sarkar and Suresh (2011)). Furthermore, relatively little funding is spent on TB research worldwide when compared with research in other therapeutic fields. The unequal distribution of global funding has led to the domination of not-for-profit foundations such as the TB Alliance (http://www.tballiance.org/) (Figure 1.24), TB Drug UK Consortium (http://www.tbd-uk.org.uk/) and the International Union Against Tuberculosis and Lung Disease (IUATLD) (http://www.theunion.org/). Projects such as the Lilly Not-For-Profit Partnership for TB Early Phase Drug Discovery have also been launched to speed TB drug discovery (http://www.tbdrugdiscovery.org/).

The urgent need to find new treatments for TB has evoked the call for new strategies in early hit-selection as compared to other therapeutic uses. Ideas such as using less strict hit-selection criteria than the standard drug-likeness rules (reviewed in Leeson (2012)) have been proposed recently (Kaneko *et al.*, 2011). Interestingly, many antibacterial agents, erythromycin for example, have unique physicochemical properties that violate the drug-likeness rules (O'Shea and Moser, 2008). Further strategies that have been proposed include the utilisation of a prodrug approach in which the *in vivo* activation will lead to the formation of bioactive moieties, such as, covalent modifiers of certain macromolecules within the *M. tuberculosis* cell (Kaneko *et al.*, 2011). Isoniazid itself, the first-line TB therapy, requires activation by the KatG (Figure 1.20).

1.5.3 The impact of crystallographic studies on TB drug discovery

The process of developing a new drug is usually long and expensive. Although the cost of a structure determination is not insignificant, the value of that structure determination to drug discovery is much higher (Figure 1.25) (Hubbard, 2005). Structural data are complementary to other experimental data, which in combination will help to define novel drugs (Thornton *et al.*, 2000). The main challenges in protein structure determination are the production of sufficient quantities of pure soluble recombinant protein for crystallisation trials, and the ability of the protein to form diffraction quality crystals (Hubbard, 2005). This is particularly an obstacle for the crystallisation of mycobacterial proteins (Arcus *et al.*, 2006). Figure 1.26 shows a comparison of the number of mycobacterial and human protein structures deposited in the Protein Data Bank (PDB) due to the efforts of two independent organisations: the TB Structure Genomic Consortium (TBSGC; USA) and the Structure Genomics Consortia (SGC; Oxford, Toronto, Stockholm), respectively. Realization of the need for new drug targets led to the increased number of 3D-structures of *M. tuberculosis* proteins in the PDB. Targets that are likely to be involved in the persistence of mycobacteria have been investigated extensively. Among the biological processes that have been investigated as targets for latent TB treatment, cell-wall synthesis and lipid metabolism are particularly attractive (Zhang, 2005).

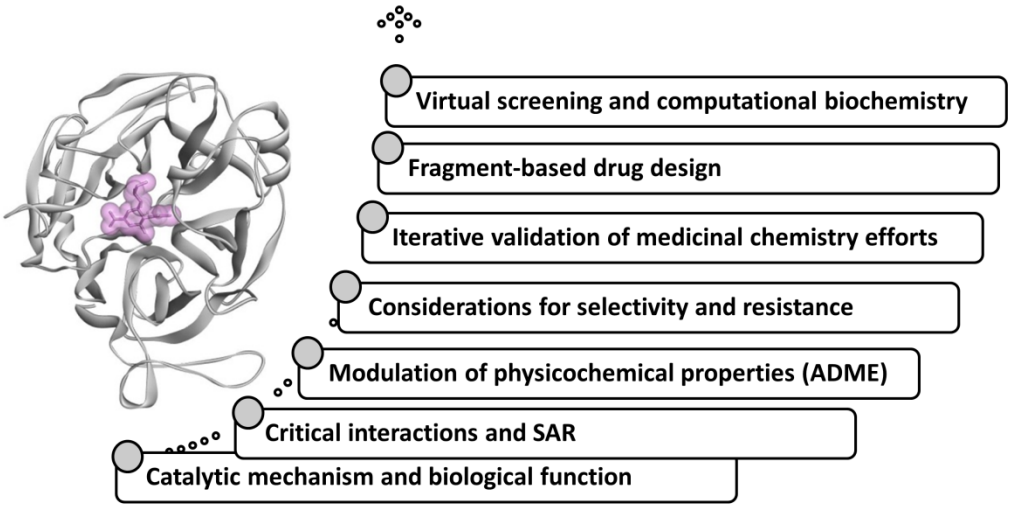


Figure 1.25: Summary of information that can be derived from a 3D-structure of a ligand-protein complex. relating to drug discovery.

The figure is based on data in review of Hubbard (2005).

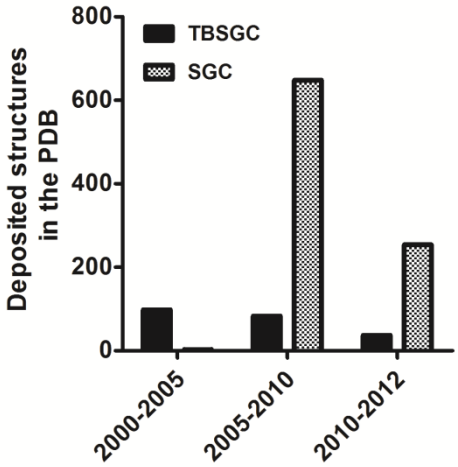


Figure 1.26: Comparison of the number of mycobacterial and human protein structures deposited in the Protein Data Bank (PDB) by the TBSGC and the SGC, respectively.

Data include all the deposited structures by the two genomic centres as accessed on 5/3/2012.

1.6 Aims of this work

Recent studies showed that NAT in mycobacteria is involved in the biosynthesis of essential cell wall components and in intracellular survival within macrophages (reviewed in Sim *et al.* (2008a)). The data supporting these findings were obtained *in vitro* with *M. bovis* BCG as a model strain, the NAT enzyme of which is identical to that from *M. tuberculosis*.

At the beginning of the project, the **AIM** was to establish the crystal structure of NAT from *M. tuberculosis* (TBNAT). In order to approach this aim it was essential to:

- **Increase the amount of *M. tuberculosis* NAT available. Experiments are described in Chapter 4 towards this AIM.**

Whilst the production of protein was being developed, and conditions for crystallisation were being explored (Chapter 6) a further series of aims was investigated using the MMNAT enzyme as a model of TBNAT, where lower amounts of TBNAT protein could be used:

- **Probe the architecture of the catalytic pocket of MMNAT by generating a co-crystallised structure of MMNAT with a high affinity substrate. Experiments towards this aim are described in Chapter 3**
- **Compare the ligand specificity of MMNAT and TBNAT to identify high affinity ligands, which could be used for co-crystallisation studies. Experiments towards this aim are described in Chapters 3 and 5**
- **Explore the underlying mechanism of NAT inhibition by a group of inhibitors belonging to the piperidinol class. Experiments towards this aim are described in Chapter 5**

In each of the chapters, a series of sub-aims is provided and the experimental data are shown. The materials and methods used throughout this work are described in Chapter 2.

2 Materials and methods

In this work, many interdisciplinary experimental techniques were utilised. This chapter details the methods and reagents involved.

2.1 Materials

2.1.1 Chemicals and reagents

All chemicals were purchased from Sigma-Aldrich or Alfa Aesar unless otherwise stated. Molecular biology reagents were obtained from Promega and New England Biolabs. Competent *E. coli* cells were purchased from Promega and Invitrogen. Mycobacterial culture media and growth supplements were purchased from Becton-Dickinson. NAT inhibitors were prepared by Dr. Elizabeth Fullam. Non-hydrolyzable coenzyme-A (CoA) derivatives were kindly provided by Mr. Luc Henry (Department of Chemistry, Oxford).

2.1.2 Bacterial strains and cell lines

E. coli and *Pseudomonas putida* strains used for plasmid maintenance and the production of recombinant MMNAT and TBNAT are described in Table 2.1.

Table 2.1: Bacterial strains used for cloning and protein expression.

Strains	Genotype/relevant characteristics	Source
E. coli JM109	endA1, recA1, gyrA96, thi, hsdR17(r_K^-,m_k^+), relA1, supE44, Δ(lac-proAB), [F', traD36, proAB, $lacI^qZ\Delta M15$]	Promega
E. coli BL21(DE3)pLysS	F^-,*omp*T, *hsd*S_B(r_B^-, m_B^-), *dcm*, *gal*, λ(DE3), pLysS, Cm^r	Promega
E. coli ArcticExpress	B F^-, *omp*T, *hsd*S(r_B^-, m_B^-), dcm^+, Tet^r, *gal*, *endA*, Hte, [*cpn*10, *cpn*60, $Gent^r$]	Stratagene
E. coli NEB Express	*fhu*A2, [Ion], *omp*T, *gal*, *sul*A11, R(mcr^- 73::miniTn10$^{--}Tet^S$)2 [dcm], R(zgb^- 210::RN10$^{--}Tet^S$), endA1, Δ($mcrC^-$ mrr)114::IS10	New England BioLabs
P. putida KT2442	hsdR, Rif^R prototrophic	

P. putida KT2442 was kindly provided by Professor Lindsay Eltis (University of British Columbia, Canada) (de Lorenzo *et al.*, 1993). *M. bovis* BCG str. Pasteur (ATCC 35734) was used for the Alamar blue assays.

2.1.3 Bacterial growth media

The media used to grow *E. coli*, *P. putida* and *M. bovis* BCG are described in Table 2.2.

Table 2.2: Growth media used throughout this study. All media were sterilized by autoclaving for 30 min at 121 °C unless otherwise stated.

Media	Components
Luira-Bertani Broth (LB)	10 g/L Bacto-Tryptone (Difco), 5 g/L Bacto-Yeast Extract (Difco), 10 g/L NaCl
LB agar	15 g/L Bacto-Agar (Difco) in LB medium
Middlebrook 7H9	5.2 g/L Middlebrook 7H9, 0.1 % (v/v) Tween-80, 0.2 % (v/v) glycerol. The mixture was autoclaved at 121 °C for 15 min. When growing *M bovis* BCG the media was supplemented with 10 % (v/v) oleic acid-albumin-dextrose-catalase (OADC).

2.1.4 Plasmids

(A) pET28b(+) (Novagen) is a 5368 bp plasmid derivative of the pET28b(+) vector which is commonly used for protein production in *E. coli* (Figure 2.1). The vector carries an *N*-terminal His-Tag/thrombin/T7-Tag configuration in addition to an optional *C*-terminal His-Tag sequence. Expression of the insert can be induced by the addition of isopropyl *β*-D-thiogalactopyransoide (IPTG) in a DE3 lysogen expression *E. coli* host strain. The presence of a thrombin cleavage site (LVPRGS) allows an *N*-terminal hexa-histidine tag to be removed from a heterologous expressed protein. The vector also contains a kanamycin resistance cassette.

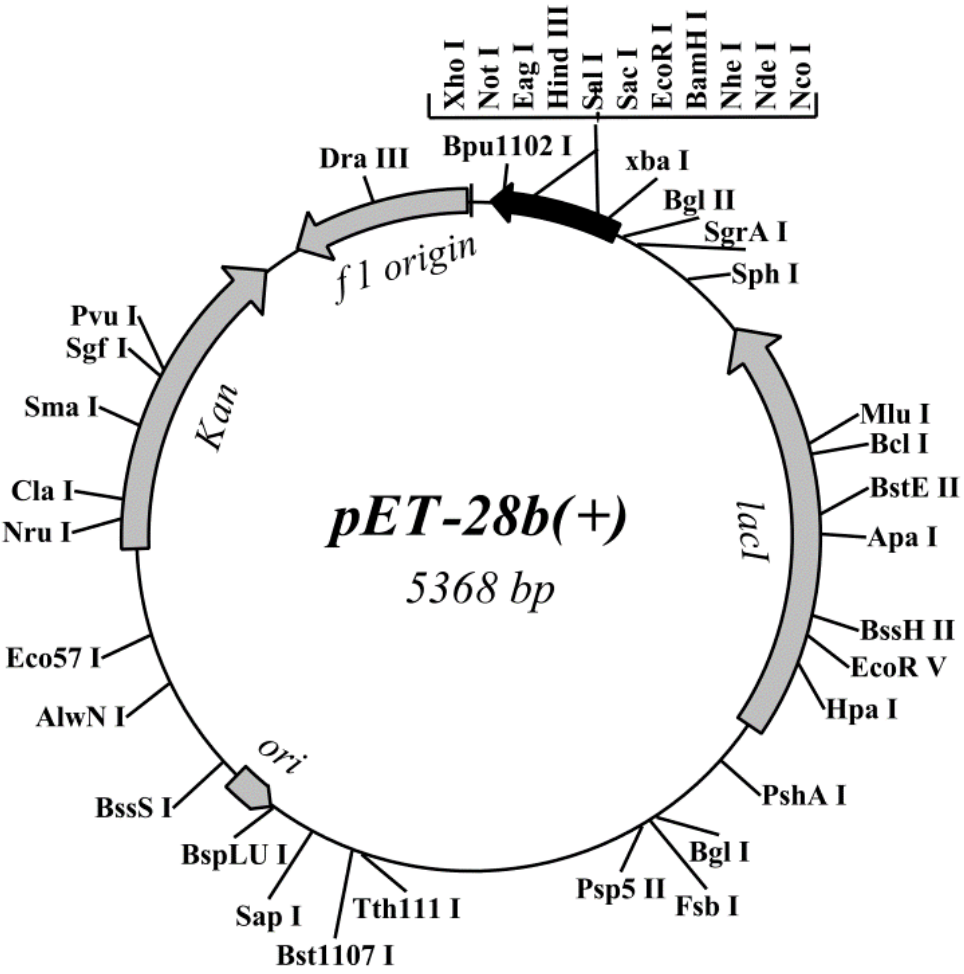


Figure 2.1: Vector map of pET-28b(+).

Figure modified from www.genomex.com/vector_maps/.

(B) pVLT31 is a 10000 bp broad-host expression vector (Figure 2.2) that induces the transcription of a cloned gene by a *ptac* promoter that is a negatively regulated by a *lacI*^{q} repressor (de Lorenzo *et al.*, 1993). The vector is a derivative of pMMB207 (ATCC 37809) that has a tetracycline resistance cassette. The plasmid was generously donated by Professor Lindsay Eltis (University of British Columbia, Canada).

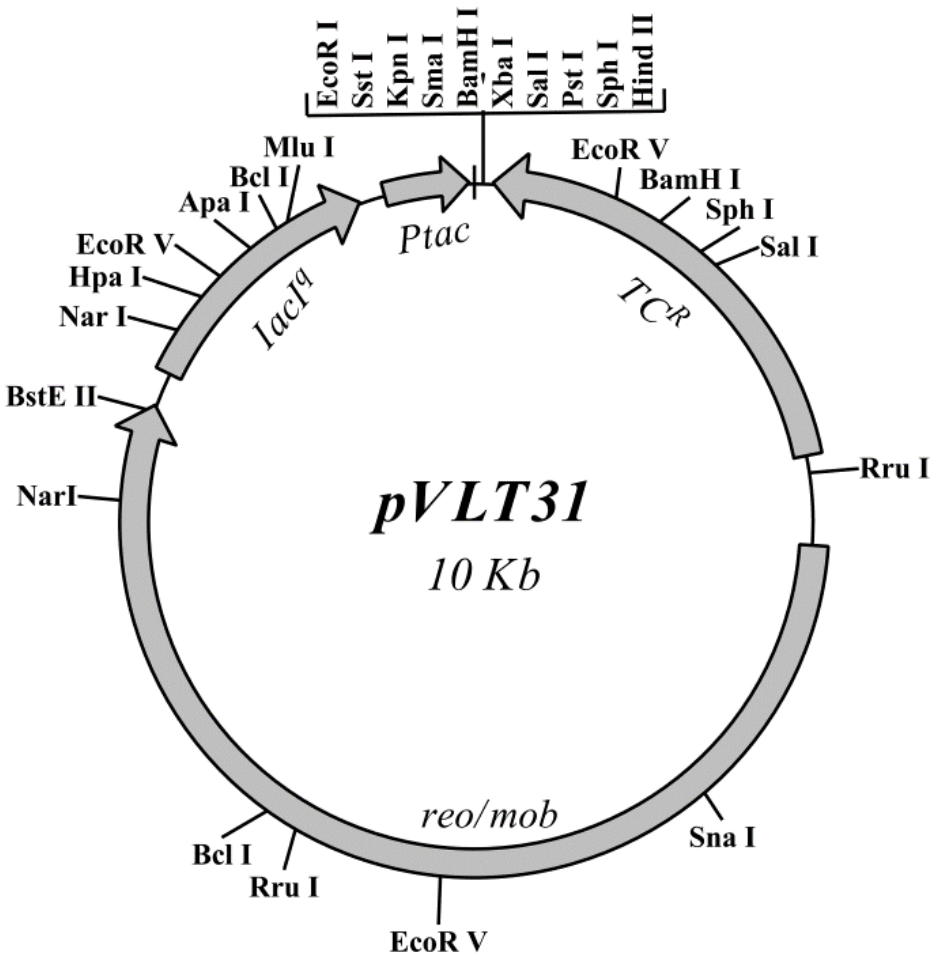


Figure 2.2: Vector map of pVLT31 (de Lorenzo *et al.*, 1993).

2.1.5 Primers

Primers were purchased from Sigma-Genosys. The oligonucleotides shown in Table 2.3 were used for sub-cloning and sequencing *nat* (Rv3566c) from *M. tuberculosis* H37Rv (*tbnat*).

Table 2.3: Oligonucleotides used in pVLT3-*tbnat* insert sequencing.

Oligonucleotide	Orientation	Sequence (5'-3')
nat-cloning	sense	ATGGCACTGGATCTGA
nat-cloning	antisense	TTACGGCGCATCGG
pVLT31-external	sense	AATAATTTTGTTTAACTTTAAG
mid-*nat*	sense	GGACACATGGCAGACA
start-*nat*	antisense	GTTCGACTGTGCACGGT

2.1.6 Commercially available crystallisation sparse matrix screens

The commercially available crystallisation screens used throughout this work are summarised in Table 2.4.

Table 2.4: List of commercially available crystallisation spare matrix screens used in this study. [a]

Screen name	Supplier	Number of conditions	Description
Crystal screen I	Hampton Research	50	Broad-based sparse matrix
Crystal screen II	Hampton Research	48	Broad-based sparse matrix
Wizard I+II	Emerald Biostructures	96	Random sparse matrices
JCSG-plus	Molecular Dimensions	96	Very broad sparse matrix
Index	Hampton Research	96	Primary, diverse reagent
PEG/Ion II	Hampton Research	48	PEG, salt and pH matrix
Crystal Screen Lite	Hampton Research	48	Primary for membrane/limited solubility proteins
PACT premier	Molecular Dimensions	96	Systematic multi-component screen in three parts, screening PEG precipitants against: pH, cations and anions.
Morpheus	Molecular Dimensions	96	Low molecular weight ligands
Structure Screen I+II	Molecular Dimensions	96	Broad-based sparse matrix
PGA	Molecular Dimensions	96	Systematic screen based on the poly-γ-glutamic acid polymer
ProPlex	Molecular Dimensions	96	Medium-high MW PEG precipitants at relatively low concentrations
MIDAS	Molecular Dimensions	96	Based around alternative polymeric precipitants
Total number of conditions		1058	

[a] MW: molecular weight, PEG: polyethylene glycol.

2.2 Methods

2.2.1 Bacterial growth

Bacterial growth in LB medium was monitored by measuring the optical density of the culture at wavelength 600 nm (OD_{600}) using sterile liquid media as a blank.

2.2.1.1 *E. coli* and *P. putida*

Liquid cultures of *E. coli* and *P.putida* were grown in LB medium (Table 2.2) with shaking (180 rpm) in either sterile 50 mL centrifuge tubes or sterile plastic conical flasks. *E. coli* was grown at 37 °C and *P. putida* was grown at 30 °C. When culturing *E. coli* and *P. putida* on solid media, the bacteria were grown on LB agar (Table 2.2) at 37 °C (*E. coli*) or 30 °C (*P. putida*). Colonies were typically observed following overnight incubation.

2.2.1.2 *M. bovis* BCG

Liquid cultures of *M. bovis* BCG were grown in roller bottles (2 rpm) (Thermo Fisher Scientific) at 37°C in 100 mL of Middlebrook 7H9 media containing 10 % (v/v) OADC (Table 2.2). Cultures were inoculated with either a bacterial glycerol stock or *M. bovis* BCG culture at mid-exponential growth phase (OD_{600} of 0.8-1.2) at a 1/100 dilution. Liquid cultures of *M. bovis* BCG typically reached an OD_{600} of 1.0 in four days. No experiments were conducted on cultures of *M. bovis* BCG that were inoculated with glycerol stocks.

2.2.2 Molecular biology

2.2.2.1 Bioinformatics and sequences

The genome sequences for *M. tuberculosis* H37Rv (Cole *et al.*, 1998) and *M. marinum* (Stinear *et al.*, 2008) were found at http://www.ncbi.nlm.nih.gov/. BLAST (Basic Local Alignment Search Tool) analyses were performed using the NCBI tools (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.*, 1990). Nucleotide and amino acid

sequence alignments were generated using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/) (Thompson *et al.*, 1994). ClustalW2 alignment results were visualized with ESPript 2.2 (http://espript.ibcp.fr/ESPript/ESPript/) (Gouet *et al.*, 1999).

2.2.2.2 Polymerase chain reaction

The polymerase chain reaction (PCR) was conducted in a total volume of 50 µL, containing 1 pmol/µL of the sense and antisense primer, 200 µM dNTPs, 6 % (v/v) dimethyl sulphoxide (DMSO), 1 µL of template DNA, 1× polymerase buffer and DNA Taq polymerase (New England Biolab). The polymerase was used according to the manufacturer's instructions.

PCR amplification was carried out using a thermal cycler (MJ Research PTC-0200 Peltier, GMI Inc) using 30 cycles with the reaction cycle outlined in (Figure 2.3).

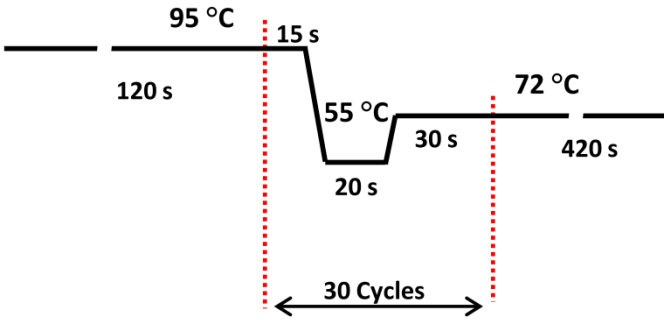


Figure 2.3: Temperature cycles used in PCR amplification.

2.2.2.3 DNA purification

Plasmid DNA samples were purified from overnight cultures of *E. coli* using QIAprep Spin Miniprep Kit, (Qiagen). DNA from agarose gels was extracted with the QIAquick Gel Extraction Kit (Qiagen). These purification systems use high concentration of chaotropic salt to precipitate proteins followed by adsorption of DNA onto silica in the presence of high salt.

2.2.2.4 Determination of DNA concentration

The relative yield of DNA was determined by spectrophotometric analysis at a wavelength of 260 nm (OD_{260}) using a BioPhotometer (Eppendorf). The concentration of the sample was determined based on the relationship: OD_{260} of 1.0 corresponded to approximately 50 μg/mL of double-stranded DNA (Sambrook *et al.*, 1989).

2.2.2.5 Agarose gel electrophoresis

Agarose gels containing 1 % agarose (w/v) were prepared in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) with 0.5 mg/mL of ethidium bromide. Before loading, DNA samples were mixed with 20 % (v/v) gel loading buffer (30 % (v/v) glycerol, 0.25 % (w/v) bromophenol blue and 0.25 % (v/v) xylene cyanol). A 1 kb Plus DNA ladder (Invtrogen) was run concurrently with the samples to determine the size of the DNA products. Electrophoresis was performed as described by Sambrook *et al.* Gels were visualized using a Gel Doc 2000 UV/Visible illuminator (BioRad).

2.2.2.6 Restriction endonucleases

DNA was digested by incubating 10 U of restriction endonucleases (New England Biolab) with 1 μg of DNA at 37 ºC for 1 h. Typical restriction endonuclease digests were carried out in 20 μL volumes using the buffers and conditions suggested by the manufacturer.

2.2.2.7 Sub-cloning of the *nat* gene

Sub-cloning of the *nat* gene from the previously available pET28b(+) vector (Fullam *et al.*, 2009) into the pVLT31 vector was performed with the assistance of Dr. Nathan Lack. To create the expression construct, the purified pET28b(+) plasmid containing the *tbnat* insert was digested with XbaI/HindIII. The fragment from the plasmid that contained *nat*, a thrombin digestion site, a ribosome binding site and a hexahistidine-tag (His-tag)-DNA sequence (H6) was gel-purified. This digested fragment was ligated with pVLT31 which had been previously digested with XbaI/HindIII. The ligation mixture was transformed into chemically competent *E. coli* JM109 using a heat-shock protocol (Hanahan, 1983) and

then plated on LB agar supplemented with tetracycline (15 μg/mL). The resulting bacterial colonies were screened for the *nat*-H6 insert by colony PCR. The identity of the resulting plasmid was confirmed by DNA sequencing. Following confirmation, the recombinant plasmid was transformed into competent *E. coli* NEB Express and *P. putida* KT2442, and plated onto LB agar supplemented with tetracycline (15 μg/mL). Individual bacterial colonies were isolated, grown in the presence of tetracycline and then sequenced to confirm nucleotide identity. Stocks of the bacterial constructs were stored at -80 ºC in 25 % (v/v) glycerol in LB medium.

2.2.2.8 Heatshock transformation

Commercially available chemically competent strains of *E.coli* (Table 2.1) were transformed using a heat-shock protocol (Hanahan, 1983). In this method, the ligation solution (2 μL) (Section 2.2.2.7) was mixed with 50 μL of thawed chemically-competent cells in a sterile 14 mL cell-culture tube and incubated for 15 min on ice. The transformation mixture was transferred to a 42 °C water bath for 45 s and then placed back on ice for 2 min. The transformed bacterial solution was diluted with 3 mL of LB media and incubated at 37 °C with 180 rpm shaking for 1 h. The transformation mixture was centrifuged; the supernatant was removed, and the bacteria were spread LB agar supplemented with the appropriate antibiotics. The plates were then incubated at 37 °C overnight.

2.2.3 Protein production

2.2.3.1 Expression of MMNAT.

The expression of MMNAT in *E. coli* (BL21(DE3)pLysS) was performed as described previously (Fullam *et al.*, 2008), but omitting the antibiotic from the expression culture. A starter culture was prepared by inoculating 20 mL LB medium containing kanamycin (30 μg/mL) and chloramphenicol (34 μg/mL) with the engineered *E coli*. The culture was

grown at 37 °C with shaking (180 rpm) for 16 h. One litre LB medium supplemented with 1 M sorbitol and 2.5 mM betaine was inoculated with 1 % starting culture and grown at 37 °C with shaking (180 rpm). At an OD_{600} of 0.5-0.8, the temperature was reduced to 25 °C and protein expression was induced with 250 μM IPTG. The cells were harvested after 16 h of incubation at 25 °C.

2.2.3.2 Expression of TBNAT.

The expression of TBNAT in *E. coli* ArcticExpress transformed with pET28b(+) vector containing the *tbnat* insert was performed as previously described (Fullam *et al.*, 2009). A starter culture was prepared by inoculating 20 mL LB medium containing kanamycin (30 μg/mL) and chloramphenicol (34 μg/mL) with the engineered *E coli*. The culture was grown at 37 °C (180 rpm) for 16 h. One litre LB culture was inoculated with 2 % of the starter culture and grown at 30 ºC for 3 h, the temperature was then reduced to 13 ºC for 30 min and the bacteria were induced with 1 mM IPTG for 16 h at 13 ºC. Expression of the TBNAT protein was performed with the assistance of Ms. Judith Schweichler.

2.2.3.3 Optimisation of the TBNAT expression

An overnight starter culture was used to inoculate a 30-mL LB medium culture, which was then incubated at either 37 °C or 30 °C for *E. coli* and *P. putida*, respectively, with shaking at 180 rpm. The growth was monitored by measuring the OD_{600}. The screening of different induction conditions was performed by varying the time and temperature of the induction, the IPTG concentration and the duration of post-induction incubation.

2.2.3.4 Cell lysis

The cells were harvested by centrifugation (6,000 g, 4 °C, 15 min) and the cell pellets were re-suspended in 25 mL lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl and Complete EDTA-free protease inhibitor (one tablet per 25 mL lysis solution; Roche)) and stored overnight at -80 °C. The cell pellets were thawed in the presence of lysozyme and DNase I

and sonicated (25 cycles of 45 s on, 45 s off) in an ice bath. The lysate was then centrifuged at 12,000 g for 25 min at 4 °C.

2.2.3.5 Protein purification

The purification of the recombinant His-tagged protein was carried out using immobilised metal ion affinity chromatography (IMAC). The supernatant soluble lysate was loaded onto either a Nickel NTA resin (Qiagen) or a Cobalt TALON resin (Clontech). The column was washed with five times the column volume of 20 mM Tris-HCl buffer pH 8.0 containing 300 mM NaCl and increased concentration of imidazole (Imz) (1-100 mM), and then eluted with three times the column volume of the elution buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM imidazole) unless otherwise stated. The bacterial lysate and various elution fractions were kept on ice for all manipulations.

2.2.3.6 Buffer exchange and protein concentration

The eluted protein was dialysed against 4 L dialysis buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA and 1 mM dithiothreitol (DTT)) for 16 h at 4 °C using a 12-14 kDa molecular weight cut-off dialysis tube. The protein was concentrated using an Amicon ultracentrifuge concentrator (10 kDa molecular weight cut-off) unless otherwise stated. The His-tag of the purified recombinant protein were cleaved by thrombin digestion (2 U/mg protein, either at 4 °C for 16 h or at 37 °C for 4 h). The protein solution was buffer exchanged (in the buffer specified in each case) and concentrated by using the Amicon ultracentrifuge concentrator. The purified protein was concentrated to a final concentration of 10-19 mg/mL and stored at -80 °C in 5 % (v/v) glycerol.

2.2.4 Protein characterisation

2.2.4.1 Determination of protein concentration

Protein concentration in the impure wash-fractions was estimated by the Bradford assay using serum albumin as protein standard (Bradford, 1976). Protein concentrations were

calculated as a mean value of triplicate measurements. The concentration of purified proteins was determined by measuring the OD_{280} using the Beer-Lambert law (Equation 2.1).

$$OD_{280} = \varepsilon_{280} \cdot c \cdot l \qquad \text{Equation 2.1}$$

where ε_{280} is the molar extinction coefficient at a wavelength of 280 nm, c is the concentration of the protein sample and l is the length of the light path in cm. The molecular weight of each protein and the extinction coefficients at 280 nm were estimated by ExPASy ProtParam http://www.expasy.ch/tools/protparam.html.

2.2.4.2 Denaturing Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Electrophoresis was performed using 12 % SDS–PAGE, according to Laemmli (1970) unless otherwise stated. The resolving gel (12 % v/v acrylamide:bis-acrylamide (29:1), 0.125 % w/v SDS, 0.05 % w/v ammonium persulfate and 0.002 % v/v N`,N`,N`,N`-tetramethylethylenediamine (TEMED) in 0.375 M Tris-HCl pH 8.8) and stacking gel (6 %v/v TEMED in 0.125 M TrisHCl pH 6.8) was cast in a vertical mini-gel apparatus (BioRad). Protein samples were prepared by mixing 10 µL of the cell lysate (30-60 µg protein) (Section 2.2.3.4) or each of the washing fraction (0.5-4.0 µg protein) (Section 2.2.3.5) with an equal volume of sample preparation buffer (0.4 % w/v SDS, 1 % v/v β-mercaptoethanol and 2 % v/v glycerol, in 10 mM Tris-HCl pH 6.8 containing 10 µg/mL bromophenol blue). The samples were then denatured at 95 °C for 5 min, and 15-20 µL loaded unless otherwise stated. Protein bands were stained with Coomassie brilliant blue and low-range molecular weight protein markers (Bio-Rad) were used as standards unless otherwise stated.

2.2.4.3 Electrospray ionization mass spectrometry

Electrospray inonization mass spectrometry (ESI-MS) was carried out by analysing samples of 2.5 mg/mL protein on a Micromass LTC mass spectrometer. ESI-MS analysis

was performed in positive ion mode after denaturation in 50 % v/v acetonitrile in water by David Staunton in the Biophysics Facility of the Department of Biochemistry in Oxford.

2.2.4.4 Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)

SEC experiments were performed with a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in 25 mM Tris-HCl, pH 7.5, 150 mM NaCl at 0.4 mL/min on an AKTA system. The column was followed in-line by a Dawn Heleos-II light scattering detector (Wyatt Technologies) and an Optilab-rex refractive index monitor (Wyatt Technologies). Molecular mass calculations were performed using ASTRA 5.3.4.14 (Wyatt Technologies) assuming a refractive index increment (dn/dc value) of 0.186 mL/g. SEC-MALS was performed by Dr. Steven Johnson (Sir William Dunn School of Pathology, Oxford).

2.2.5 Liquid chromatography-mass spectrometry (LC/MS)

Liquid chromatography-mass spectrometry of was carried out by Mr. Nikita Loik at the Department of Chemistry, Oxford. Samples of cysteine, compound A-1 and the mixture (1:1) of both (Section 5.2.1) were submitted as 100 μM solutions in MOPS buffer pH 8.0. All samples were mixed with 5 μL of the primary amine derivatization reagent AccQ Tag Ultra (6-aminoquinolyl-N-hydroxysuccinimidyl carbamate; Waters) and injected onto a reverse phase Acquity C18 column (2.1 × 100 mm,1.7 μm particles) equilibrated with 5 % AccQ Tag Ultra Eluent A on an Acquity Ultra Performance Liquid Chromatography system. Derivitized substrate and products were resolved using a gradient of increasing AccQ Tag Ultra Eluent B solution, detected by absorbance at 260 nm (Cohen and Michaud, 1993).

2.2.6 Enzymic Assays

2.2.6.1 NAT activity assay

Routinely, NAT activity of recombinant TBNAT or MMNAT was measured as a rate of CoA formation catalysed in the presence of hydralazine (HLZ) as a substrate as previously described (Brooke *et al.*, 2003b). Samples of soluble lysate/NAT (10 µL) were diluted to give an initial linear activity-time relationship in the assay buffer: 20 mM Tris-HCl pH 8.0 and preincubated with 500 µM HLZ at 24 ºC for 5 min in a 96-well plate (Corning). Acetyl-CoA (400 µM final concentration) was added to start the reaction in a final volume of 100 µL. The reaction was then quenched with 25 µL Ellman's reagent (5 mM 5,5'-dithiobis-(2-nitrobenzoate) solution in 6.4 M guanidine-HCl and 0.1 Tris-HCl, pH 7.3) and OD_{405} was measured within 2 min (Tecan Sunrise Plate Reader). The rate of the reaction was determined by reference to a standard curve obtained by plotting the OD_{405} against serial dilutions of CoA standard solutions of known concentrations. The enzyme activity was expressed as the rate of CoA formation per microgram enzyme (µM/min.µg or mM/min.µg). For the activity of the enzyme with different acyl-CoA cofactors, the assay was performed in the same way by substituting Ac-CoA with the different acyl-CoA cofactors.

2.2.6.2 NAT inhibition assay

The assay for measuring the formation of CoA was used to determine the activity of the enzyme in the presence of potential inhibitors (Brooke *et al.*, 2003b). Experimental titrations of the enzyme were performed with both HLZ and Ac-CoA in order to determine the minimum amount of enzyme required to generate a sufficient signal while ensuring a signal to noise of the absorbance measurement (OD_{405}) higher than two. All the tested compounds were dissolved in dimethylsulphoxide (DMSO) and all reactions were carried out in the presence of 5 % v/v DMSO. The enzyme was mixed with the inhibitor (at a final concentration of 0-250 µM) and incubated for 15 min at room temperature (~24 ºC) prior

to starting the reaction by adding the substrates (HLZ and Ac-CoA). The assay was performed as an end-point read out measurement (in final volume of 100 µL) by quenching the reaction after 10 min with Ellman's reagent. The assays were repeated using a 10-fold enzyme concentration to exclude promiscuous non-specific inhibitors. A specific inhibitor would not be significantly affected by such a change in the concentration of the enzyme, on the contrary, an aggregating promiscuous (non-specifinc) inhibitor would show a dramatic decrease in potency with an increased enzyme concentration (McGovern *et al.*, 2002). The activity of the enzyme in the presence of 5 % v/v DMSO was measured as a control. Inhibition percentages were determined as the ratio of the enzyme activity (expressed as the rate of CoA formation per microgram protein (µM/min.µg)) with the requisite compound to the activity of the control without inhibitor. IC_{50} values were calculated from inhibition curves obtained by non-linear fitting of the % inhibition and the inhibitor concentration (µM) using the Log(inhibitor) vs. response module of GraphPad Prism 5.0 (http://www.graphpad.com/prism/).

For simplicity, this assay will be referred to as the NAT-inhibition assay throughout the following text.

2.2.7 Growth inhibition assay of mycobacteria

2.2.7.1 Alamar blue assay

The Alamar blue assay was used to determine the antimycobacterial activity of NAT inhibitors against *M. bovis* BCG str. Pasteur (ATCC 35734) (Yajko *et al.*, 1995). The assay were preformed by Mrs. Hilary Long (Department of Pharmacology, Oxford).

M. bovis BCG (100 mL) was grown until early log-phase (OD_{600} 0.4- 0.7). The cells were harvested (2,500 g, 15 min), resuspended (50 mL phosphate-buffered saline (PBS), 0.05 % Tween 80) and harvested (2,5000 g, 15 mins). The cells were resuspended (4 mL PBS, 0.05 % Tween 80), aliquoted (100 µL per cryo-tube) and snap frozen in liquid nitrogen and

stored at -80 ^{o}C. 7H9GC-OADC (2.4 mL) was added to 100 µL of cells and diluted 100-fold in 7H9GC-OADC. The assay was carried out in 96-well plates in the presence of either compound F-4, and INH as control (0-66 µg/mL in DMSO). Plates were incubated at 37 ^{o}C and Alamar blue dye (50 µL) was added after six days. The minimum inhibitory constant (MIC) was determined visually at the concentration at which colour-change occurred (blue to pink).

2.2.8 Protein crystallography

Protein crystallisation was performed using the sitting- drop vapour-diffusion technique. Sitting crystallisation drops were set up in 96-well plates containing commercially available sparse matrix and systematic grid screens conditions (Table 2.4). Initial high-throughput screens to identify crystallization conditions were performed using a Tecan Genesis Pro Team 150 Robot (Tecan). An equal volume (100 nL) of mother-liquor and protein were set up as sitting drops using a Mosquito crystallisation robot (TTP Labtech).

2.2.8.1 Co-crystallisation of MMNAT-HLZ complex

Crystals of the MMNAT–HLZ complex were obtained by co-crystallisation of the enzyme with HLZ. A preliminary search for suitable crystallisation conditions at 293 K was carried out using the JCSG-plus screen and protein solution of 10 mg/mL in 20 mM Tris-HCl pH 8.0 containing 20 mM HLZ and 1 mM DTT. Initial crystals of MMNAT-HLZ grew within a week in condition E2 of the JCSG-plus screen (0.2 M NaCl, 0.1 M Na-cacodylate pH 6.5 and 2.0 M $(NH_4)_2SO_4$). This was followed by carrying out a series of customized fine-grid optimization screens at 277 K and 293 K around the favourable conditions found in the initial screen. In the optimisation screens, Na-cacodylate was replaced by MES. Diffraction quality bipyramidal crystals formed within three weeks at 293 K in condition E2 (0.10 M NaCl, 0.10 M MES pH 6.7 and 1.65 M $(NH_4)_2SO_4$). For cryo-protection, crystals were briefly (10-30 s) washed with a 7 M sodium formate solution containing HLZ (20 mM), and were then flash cryo-cooled into liquid nitrogen (77 K).

2.2.8.2 Co-crystallisation of MMNAT-A-1 complex

For co-crystallisation of MMNAT with compound A-1, the protein (10 mg/mL) was incubated with 5 mM compound A-1 in 20 mM Tris-HCl pH 8.0 containing 5 % DMSO, for 1 h at 24 ºC. The protein was buffer exchanged with fresh Tris-HCl buffer (20 mM Tris-HCl pH 8.0) to remove the excess inhibitor and re-concentrated to 10 mg/mL. The preliminary screen for suitable crystallisation conditions at 293 K was carried out using JCSG-plus, PACT and Morpheus sparse-matrix screens (288 conditions). Initial crystals of MMNAT grew in condition E2 of the JCSG-plus screen (0.2 M NaCl, 0.1 M Na-cacodylate pH 6.5 and 2.0 M $(NH_4)_2SO_4$). This preliminary trial was followed by carrying out a series of customized fine-grid optimization screens at 293 K around the favourable conditions known from the initial screen as shown in the scheme in Figure 2.4. Optimisation of the initial conditions resulted in visually better crystals, but with poor diffraction qualities (high resolution of only 3 Å) (Figure 2.4). For cryo-protection, crystals were briefly (10-30 s) washed with a 7 M sodium formate solution, and were then flash cryo-cooled into liquid nitrogen (77 K).

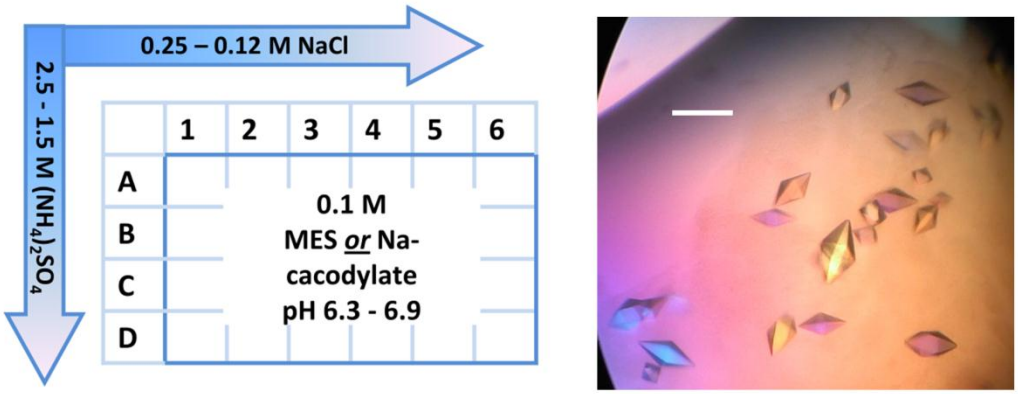


Figure 2.4: A schematic representation of the customised-screen plates designed to optimise the MMNAT-A-1 complex and the resulting crystals.

The customised screen was performed in 24-well plates and included the following conditions: 0.10 M of either Na-cacodylate or MES (pH 6.3-6.9), 1.50-2. 5 M $(NH_4)_2SO_4$ and 0.12-0.25 M NaCl. The tetragonal bipyramidal crystals shown were obtained only in conditions containing 0.23-0.15 NaCl, 2.5 M $(NH_4)_2SO_4$ and 0.1 M Na-cacodylate pH 6.5, and diffracted to 3 Å. The scale bar is equivalent to 100 µm.

2.2.8.3 Crystallisation of MMNAT, His-MMNAT and crystal soaking

Initial high-throughput screens to identify crystallization conditions were performed using the JCSG-plus, PACT, Structure Screen I+II and Morpheus sparse-matrix screens (384 conditions) (Table 2.4). Protein solutions of MMNAT and His-MMNAT (10 mg/mL in 20 mM Tris-HCl, pH 8.0, 1 mM DTT and either with or without 1 mM NaN_3) were used to set up crystallisation plates at 293 K. High quality crystals were obtained in condition F8 of the PACT primer screen (0.2 M Na_2SO_4, 0.1 M Bis-Tris propane pH 6.5 and 20 % w/v PEG-3350). These crystals tolerated the soaking step in inhibitors better than crystals from any other condition, and therefore, were used for further optimisation. The conditions were optimised by varying PEG and Na_2SO_4 concentrations in a 24-well. Crystals of His-MMNAT were obtained after at least three-days at 293 K in the optimisation condition containing 0.2 M Na_2SO_4, 0.1 M Bis-Tris propane pH 6.5 and 18-19 % w/v PEG-3350. For soaking with compound A-1, the crystals were transferred to drops of 5 mM compound A-1 in the mother liqueur containing 10 % DMSO. The crystals were harvested after 30 min, cryo-protected by brief soaking in 20 % glycerol in the mother liqueur-A-1 mixture, and flash cooled in liquid nitrogen prior to data collection. The soaking solution was prepared by substitution of water in order to retain the correct original concentrations of the mother liquor components.

2.2.8.4 Screening for crystallisation conditions of TBNAT and His-TBNAT

High-throughput screens to identify crystallization conditions for TBNAT and His-TBNAT were performed using variable combinations of the available sparse matrix screens (1060 conditions, Table 2.4), protein concentration (5-19 mg/mL), and temperature (277 K or 293 K). Further screens were performed by including different additive (details in Chapter 6) to the screen conditions or by changing the buffer in which the protein was prepared prior to setting the crystallisation trials.

Trials performed using the non-hydrolysable CoA derivatives were performed with the assistance of Dr Michael McDonough (Department of Chemistry, Oxford).

2.2.8.5 X-ray diffraction and data processing

Data collection was performed with the assistance of Dr. Edward Lowe, Dr. Sonja Baumli and Dr. James Parker (Department of Biochemistry, Oxford).

Native data were collected at 100 K at the synchrotron (details of source used for each crystal are shown later). The data were integrated and scaled using MOSFLM (Leslie, 2006) and SCALA (Evans, 2006) within the CCP4 program suite (Collaboration Computational Project, Number 4, 1994) (Collaborative, 1994). The crystal structure was solved by molecular replacement (MR) using the program MOLREP (Vagin and Teplyakov, 2000) using a previously determined native MMNAT crystal structure, stripped of heteroatoms, as a search model (PDB code: 2VFB, 1.8Å). Rigid body refinement of the MR solution and the remaining cycles of restrained refinement were carried out with REFMAC5 (Murshudov *et al.*, 1997). Molecular models of the substrate were constructed using JLigand, while model building was performed using COOT (Emsley *et al.*, 2010). The stereochemical properties and quality of the final model were assessed with the program MOLPROBITY (Davis *et al.*, 2007). Structural figures and graphical renderings were made with either PYMOL (Schrodinger, 2010) or Discovery Studio (DS) Visualizer 2.5 or 3.1 (Accelrys, 2012).

2.2.9 Thermal shift assay

The thermal shift assays (TSA) were performed with the assistance of Dr. Oleg Fedorov (Structure Genomic Consortium, Oxford) using an Mx3005p QPCR System (Agilent Technologies) as previously described (Niesen *et al.*, 2007). The reactions were performed in a total well volume of 20 µl containing 4× SYPRO Orange dye (introduced from a

5000× stock solution). The stability of the enzymes (1 μg/mL) against two types of small-molecule libraries. The first library comprised different salts, pH conditions, and commonly found small molecules in crystallisation trials (provided by the SGC, Oxford; hence referred to as SGC-stability screen).The components of this screen are listed in (Appendix A). The second library comprised compounds specific for NAT and are referred to as NAT-ligands. The latter library was generated in our laboratory and comprised substrates, physiologically relevant compounds (e.g. different acyl-CoA cofactors), and other molecules that were either identified as, or predicted to be, ligands of NAT (totalling 26 compounds). The 96-well PCR microplate was heated from 25.0 to 99.9 °C at a heating rate of 0.5 °C/min. The fluorescence intensity (The wavelengths for excitation and emission are 490 and 575 nm, respectively) was measured every 1 ºC and was plotted as a function of temperature. Data analysis was performed as previously described (Niesen *et al.*, 2007).

2.2.10 Computational approaches

2.2.10.1 3D-shape screening

In silico screening for potential NAT inhibitors was performed using the ElectroShape approach by Dr. Garrett Morris (InhibOx Ltd) (Armstrong *et al.*, 2010). The 3D-shape of compound A-1 was used as the query molecule to screen a virtual database of 7.3 million commercially available molecules (InhibOx Ltd). The identified hits were ranked based on the ElecroShape score and 12 compounds of the top 100 hits were purchased and tested for their NAT inhibition activity.

2.2.10.2 Docking simulation of NAT inhibitors

The LigandFit module of Discovery Studio 2.5 (Accelrys) was used for *in silico* docking of the NAT inhibitors. LigandFit is a grid-based method for calculating receptor-ligand interaction energies, which considers the flexibility of the ligand and treats the receptor as

rigid. The LigandFit protocol consists of three essential steps for docking ligands to the specified site: (1) definition of the active site: in the current docking experiments, the binding site was generated from the co-crystallised ligand HLZ structure, employing the default settings of LigandFit, (2) analysis of the ligand conformations and docking of ligands to the selected site: LigandFit utilizes Monte Carlo method to generate the ligand conformations, and (3) scoring of the predicted poses: estimation of binding affinity (score) by a grid-based energy calculation of the energy between the ligand and the receptor. LigandFit employs a consensus scoring of several scoring functions to predict the binding affinity.

Prior to the docking experiment, the 3D-structures of the inhibitors were prepared using the DS Visualizer. Two protonation states were assumed for each inhibitor: ionized and unionized. In the ionized forms, the molecules were assumed to exist in neutral conditions (pH 7.0), and the amino groups were protonated and given formal positive charges. The 3D-coordinates of the MMNAT-HLZ were used for the docking experiment (2.1 Å, PDB code 3LTW). Hydrogen atoms were added to the protein structure, and explicit water molecules were either kept or removed according to the required docking conditions, i.e. docking in the presence or absence of water.

In the current docking experiments, HLZ and all 12 ligands were docked into the enzyme binding site employing the default LigandFit parameters. Docking conditions, which succeeded in reproducing the co-crystallised HLZ conformer were selected for the ligand docking. The default docking scores were calculated for all docked conformers using LignadFit. The top-ranked poses were utilised in the discussion.

2.2.10.3 Homology modelling

Homology modelling was performed with assistance of Ms. Khairul Abd Halim (Department of Biochemistry, Oxford). The homology model of the TBNAT-HLZ complex was constructed using the Modeller-9.10 program (Eswar *et al.*, 2007). The

comparative protein structure modelling was performed using the 3D structure of MMNAT-HLZ complex (2.1 Å, PDB code 3LTW) as a template. ClustalW2 alignment of MMNAT with the TBNAT sequence was used as the input to MODELLER (Eswar *et al.*, 2007). Ten models were built and evaluated by the DOPE (Discrete Optimized Protein Energy) assessment scores. The model with the lowest DOPE was subjected to energy minimisation in vacuum using GROMACS 4.5 (http://www.gromacs.org/).

2.2.11 Chemical synthesis

2.2.11.1 Dehydration of compound A-1

The chemical dehydration of compound A-1 was carried out with the assistance of Dr. Peter Sendon (Department of Chemistry, Oxford). Dehydration of compound A-1 in the presence of acetic anhydride was performed as described previously (Plati and Wenner, 1951).

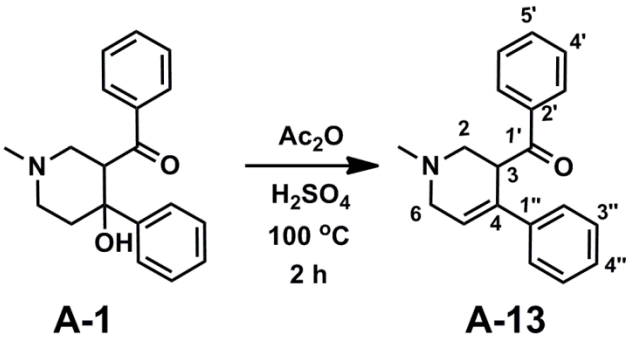


Figure 2.5: The chemical dehydration of compound A-1.

Compound A-1 (100 mg, 0.34 mmol) was suspended in acetic anhydride (1.0 mL) and treated with concentrated H_2SO_4 (1 drop). The mixture was heated to 100 ^oC for 2 h then carefully added to $NaHCO_3$ solution (70 mL sat. aq.). Solid $NaHCO_3$ was added until the mixture was made basic and the aqueous phase extracted with ethylacetate (EtOAc) (3 x 20 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo.* The residue was purified by flash chromatography (40-50 % EtOAc/Petrol + 0.1 %

triethylamine (Et_3N)) to give 1-methyl-3-benzoly-4-phenyl-1,2,5,6-tetrahydropyridine (A-13**)** as an orange oil (58 mg, 62 % yield): ν_{max} (neat)/cm^{-1} 2939 (CH), 2785 (CH), 1685 (C=O), 1447, 1027, 693; δ_H (400 MHz, $CDCl_3$) 2.34 (3H, s, NMe), 2.89 (1H, dd, *J* 11.5, 5.4, 2-*H*H), 2.96 (1H, dd, *J* 11.5, 5.4, 2-H*H*), 3.09 (1H, app. dt, *J* 17.0, 2.9, 6-*H*H), 3.36 (1H, ddd, *J* 17.0, 2.9, 2.1, 6-H*H*), 4.86 (1H, m, 3-H), 6.32 (1H app. td, *J* 2.9, 0.7, 5-H), 7.17 (1H, app. tt, *J* 7.4, 1.4, 4’’-H), 7.23 (2H, app. t, *J* 7.4, 3’’-H), 7.29 (2H, app. dd, *J* 7.4, 1.4, 2’’-H), 7.48 (2H, app. t, *J* 7.5, 4’-H), 7.58 (1H, app. tt, *J* 7.5, 1.5, 5’-H), 8.01 (2H, app. dd, *J* 7.5, 1.5, 5’-H); δ_C (100 MHz, $CDCl_3$) 45.6 (NMe), 46.9 (C-3), 54.9 (C-6), 55.8 (C-2), 125.3 (C-2’’), 125.7 (C-5), 127.0 (C-4’’), 128.4 (C-3’’), 128.5 (C-3’), 128.7 (C-4’), 133.0 (C-5’), 134.0 (C-4), 136.1 (C-2’), 140.1 (C-1’’), 199.0 (C-1’); *m/z* (ESI) 300 ($[MNa]^+$ 25), 278 ($[MH]^+$, 83).

3 MMNAT-HLZ complex structure

3.1 Introduction

Arylamine *N*-acetyltransferase from *M. tuberculosis* (TBNAT) has been proposed as a drug target for latent tuberculosis treatment (Westwood *et al.*, 2010) (Section 1.4.1).

Until the present work, TBNAT has been difficult to generate in comparable yields to NAT enzymes from *M. marinum* or *M. smegmatis* (Sikora *et al.*, 2008; Fullam *et al.*, 2009). To date, NAT from *M. marinum* (MMNAT) is considered the best available model for TBNAT. The enzyme shows 74 % sequence similarity to TBNAT (Figure 1.18), and both enzymes show comparable but not identical substrate specificity profiles despite the differences in their physical properties (Fullam *et al.*, 2008; Fullam *et al.*, 2009). Both MMNAT and TBNAT have selectivity to hydralazine (HLZ; phthalazin-1-ylhydrazine) as an acetyl group acceptor, with K_m values of 640 μM and 360 μM for TBNAT and MMNAT, respectively (Sikora *et al.*, 2008; Fullam *et al.*, 2009). The product of the biological acetylation of HLZ is a cyclic metabolite known as 3-methyltriazolo[3,4-a]phthalazine (MTP) (Israili and Dayton, 1977) (Figure 3.1).

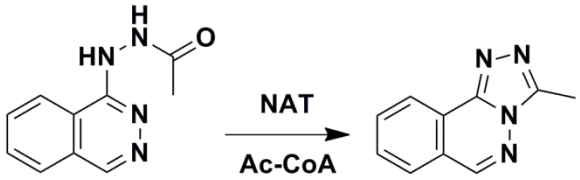


Figure 3.1: Biological acetylation of HLZ.

Recently, the 3D structures of the apo-MMNAT protein (2 Å, PDB code 2VFB) and also in complex with CoA (2.7 Å, PDB code 2VFC) have been determined (Fullam *et al.*, 2008; Fullam *et al.*, 2009). Given a high resolution structure of a protein, new ligands can be designed based on the modification of known ligands for the target protein (Babine and Bender, 1997).

The 3D structure of the MMNAT-HLZ complex can facilitate the understanding of small ligand binding, features of the binding-pocket and the mechanisms of the acetylation reaction which results in the cyclic MTP compound as a product.

Therefore, the specific **aims** of the work presented in this chapter (see Introduction, Section 1.6) were:

- **To test the affinity of MMNAT to different acyl-CoA donors**
- **To co-crystallise MMNAT with a high-affinity arylamine substrate**
- **To determine the structure of the MMNAT-HLZ complex**
- **To study the ligand protein interactions**
- **To study the mechanism of the HLZ acetylation**

Much of the work presented in this chapter has already been published in Abuhammad *et al.* (2010).

3.2 Results and discussion

3.2.1 MMNAT expression and purification

MMNAT was prepared as described in Section 2.2.3.1. The hexa-histidine tagged protein was purified using Nickel-ion affinity chromatography to apparent homogeneity, and the hexa-histidine tags were cleaved by thrombin digestion as described in Section 2.2.3.5 (Figure 3.2). Crystallographic quality protein was obtained at a yield of 28.8 mg/L culture. This yield is slightly more than in the previous work, which is likely to be due to the exclusion of antibiotic from the expression conditions in this study (Fullam *et al.*, 2008).

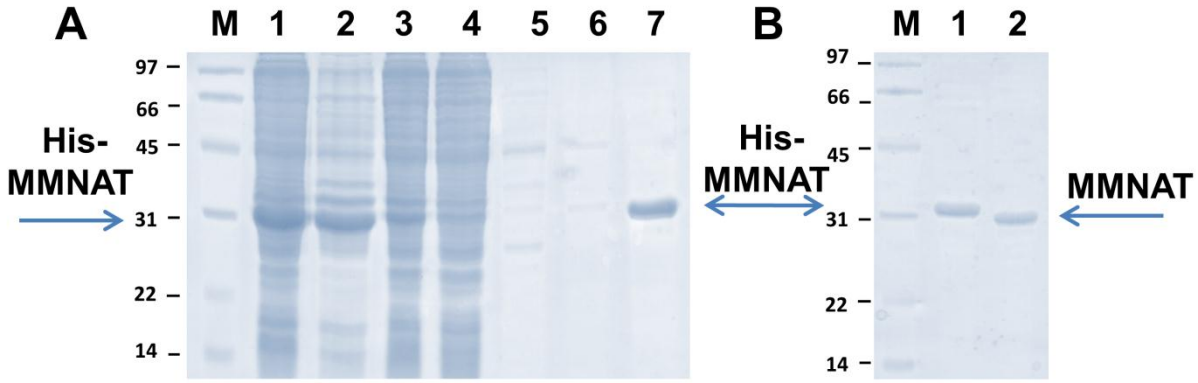


Figure 3.2: Expression and purification of recombinant MMNAT in *E. coli* transformed with pET28b(+)-*mmnat*.

Coomassie blue-stained 12 % SDS PAGE showing: (A) the over-expression of MMNAT: whole-cell lysate (Section 2.2.3.4) (lane 1), insoluble fraction (lane 2) and soluble fraction (lane 3). The purification of MMNAT was carried out by washing a Ni-NTA column loaded with the His-MMNAT with a series of steps of increasing imidazole (Imz) concentration (see Section 2.2.3.5): the unbound fraction (lane 4), 10 mM Imz (lane 5), 25 mM Imz (lane 6), 250 mM Imz (lane 7). (B) Coomassie blue-stained 12 % SDS PAGE shows the MMNAT before (lane 1) and after (lane 2) thrombin cleavage of the His-tags. Also shown is the Bio-Rad low-range molecular mass protein standard (lane M). The gels were prepared and loaded as described in Section 2.2.4.2.

3.2.2 Affinity of MMNAT to different acyl donors

Previous studies have demonstrated that both TBNAT and MMNAT have high affinity for HLZ as an acetyl acceptor (Fullam, 2007; Sikora *et al.*, 2008). Later studies showed that TBNAT catalyses Pr-CoA hydrolysis in the presence of HLZ (Lack *et al.*, 2009) and thus

Pr-CoA was proposed as a possible substrate for TBNAT. For comparison with TBNAT, various CoA derivatives were tested as cofactors of MMNAT using hydralazine as a substrate (Figure 3.3). The results demonstrated that Pr-CoA will also be hydrolysed in the presence of hydralazine (Figure 3.3). Interestingly, although acetoacetyl-CoA (AAc-CoA) was hydrolysed slightly, none of the other acyl-CoAs, including malonyl-CoA (Ma-CoA) acted as a cofactor. The same findings with AAc-CoA and Ma-CoA have been demonstrated for the TBNAT enzyme (Lack *et al.*, 2009). The observation that both enzymes catalyse the hydrolysis of propionyl CoA may be important in their physiological role.

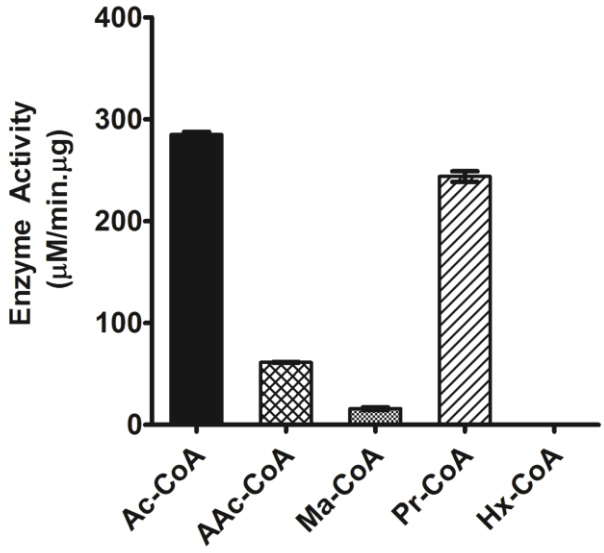


Figure 3.3: Acyl-CoA hydrolysis activity by MMNAT.

The activity of MMNAT was determined by measuring the rate of CoA formation in the presence of hydralazine. The rate of CoA formation by MMNAT was measured in the presence of 500 µM hydralazine and various acyl-CoA species: acetyl-CoA (Ac-CoA), acetoacetyl-CoA (AAc-CoA), malonyl-CoA (Ma-Coa), propionyl-CoA (Pr-CoA) and hexanoyl-CoA (Hx-CoA) at 400 µM final concentrations of each in a total volume of 100 µL of assay buffer (20 mM Tris–HCl, pH 8.0) at 24 ºC for 15 min. Results are shown as the mean ± S.D. of triplicate measurements. The error bars are within the columns.

3.2.3 Co-crystallisation of the MMNAT-HLZ complex

Crystals of the MMNAT-hydralazine complex were obtained by co-crystallisation using the sitting-drop vapour diffusion technique (Section 2.2.8.1). Crystals of the complex grew

in condition E2 of the JCSG-plus screen and a series of customised fine-grid optimisation screens, were then set up at 277 K and 293 K around the favourable conditions determined from the initial screen (Figure 3.4). Pale yellow crystals of the complex (~100 µm, Figure 3.5) formed within three weeks in condition E2 of the customised screen. The conditions in which the MMNAT-HLZ complex crystallised are shown in Table 3.1.

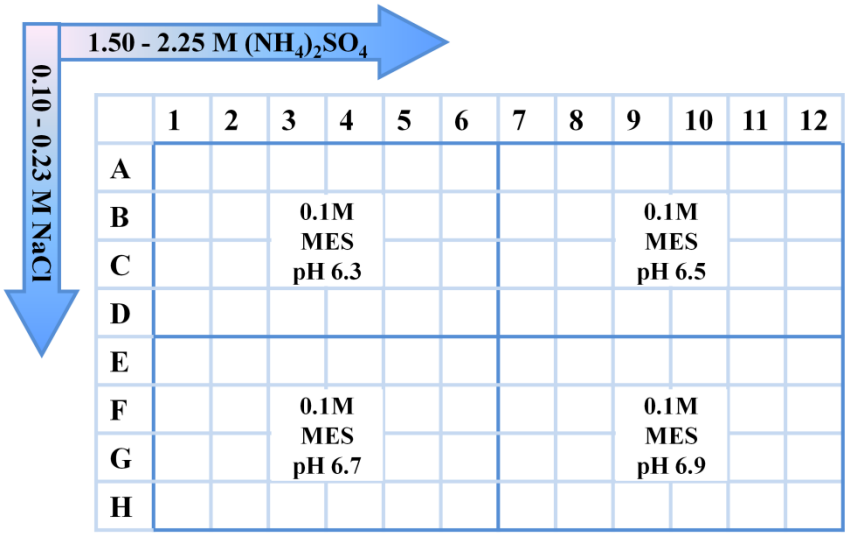


Figure 3.4: A schematic representation of the customised-screen plate designed to optimise the MMNAT-HDZ complex crystallisation.

In the optimised screens, sodium-cacodylate buffer was replaced by the less toxic MES buffer. The customised screen included the following conditions: 0.10 M MES (pH 6.3-6.9), 1.50-2.25 M $(NH_4)_2SO_4$ and 0.10-0.23 M NaCl.

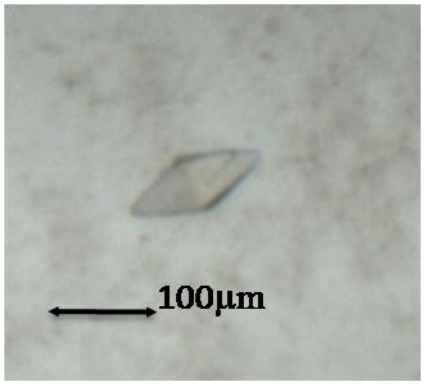


Figure 3.5: An MMNAT-HLZ complex crystal.

The crystal was obtained by the sitting-drop vapour diffusion method in the presence of HLZ. Sitting drops containing 100 nL of protein (10 mg/mL) - hydralazine (20 mM) mixture and 100 nL precipitant solutions were equilibrated against 100 µL of precipitant solution at 293 K.

Table 3.1: The customised-screen conditions in which the MMNAT-HLZ complex formed crystals of around 100 µm in size.

Code	Crystal Shape	Formulation	Temperature K	Time (days)
F9	bipyramidal	0.14 M NaCl 0.10 M MES pH 6.9 1.80 M $(NH_4)_2SO_4$	293	4-5
E10	bipyramidal	0.10 M NaCl 0.10 M MES pH 6.9 1.95 M $(NH_4)_2SO_4$	277	4-5
E2	bipyramidal	0.10 M NaCl 0.10 M MES pH 6.7 1.65 M $(NH_4)_2SO_4$	293	21

3.2.4 Data collection and processing

X-ray diffraction data were collected using synchrotron radiation sources as shown in Table 3.2. The table also shows the statistics for the data- collection and processing. The crystals, belonging to the same space group as the apo-protein (P 4_1 2_1 2), diffracted to 1.9-2.2 Å resolution. Difference electron densities corresponding to the bound HLZ molecules were observed only upon the addition of 20 mM HLZ to the cryoprotectant solution for the crystal grew in condition E2 (Table 3.2).

Table 3.2: Data collection and processing statistics for the MMNAT-hydralazine complex crystals. [a]

	Crystal		
	F9/293K	**E10/277K**	**E2/293K**
Source	DLS, ID I03	ESRF, ID 29	remote ESRF, ID23.EH2
λ (Å)	0.976	0.976	0.873
Temperature (K)		100	
Space group		P 4_1 2_1 2	
Unit cell dimensions:			
a = b, c (Å)	51.6, 176.1	51.8, 175.7	51.9, 176.7
α = β = γ (º)		90	
Resolution (Å)	38.7-2.2 (2.3-2.2)	51.8-1.9 (2.0-1.9)	38.9-2.1 (2.2-2.1)
Measured reflections	75348 (2578)	155398 (5820)	93293 (13464)
Unique reflections	12349 (476)	21163 (809)	15018 (2110)
Mosaicity (º)	1.94	0.56	0.42
Overall completeness (%)	97.0 (97.3)	100.0 (99.8)	100 (100)
Overall R_{merge}	0.071 (0.29)	0.093 (0.38)	0.13 (0.58)
Overall $R_{p.i.m.}$	0.047 (0.19)	0.06 (0.23)	0.053(0.25)
Overall R_{meas}	0.074 (0.25)	0.099 (0.41)	0.136 (0.63)
Overall I/σ(I)	16.5(6.8)	13.5(4.5)	8.9 (2.9)
Multiplicity	6.3(5.4)	7.3(7.2)	6.2 (6.4)
Solvent content (%)	36.1	36.5	36.7
Presence of HLZ	No	No	Yes

[a] **Values in parentheses are statistics of the highest resolution shell. Definitions of the statistical terms are listed in Appendix B.**

3.2.5 Molecular replacement and structure refinement

The structure was solved by molecular replacement (MR) using the software Molrep (Vagin and Teplyakov, 1997) in CCP4 (Bailey, 1994). The structure of the apo-MMNAT (PDB code 2VFB, 2 Å) was used as the starting model for MR. The final model contains only one molecule per asymmetric unit with a Matthews coefficient of 1.95 $Å^3$ Da^{-1}, corresponding to a solvent content of ~37 % (Matthews, 1968). Upon refinement of the model (Table 3.3), high difference electron density (at ~3σ) was observed close to the conserved catalytic triad (Cys70, His110 and Asp127) which is consistent with the presence of HLZ (Figure 3.6A). The quality of the electron-density map allowed HLZ to be unambiguously oriented in the model. An additional volume of relatively weaker difference electron density was observed at another position located in the space between the pairs of β-hairpins of the opposite β-sheet flaps of the structure. Another HLZ molecule was docked into this secondary binding pocket with an alternate conformation in order to fit the negative peaks in the F_o-F_c electron density maps (Figure 3.6).

The binding pocket can be seen as partially obstructed by adjoining-molecules (Figure **3.7**). This packing arrangement favours co-crystallisation over soaking, as a strategy for the complex-structure solution. Still, the solvent accessibility of the HLZ-binding site, as shown in Figure **3.7**, allows soaking to be an alternative strategy.

Table 3.3: Refinement statistics for the MMNAT-HLZ complex structure (PDB code 3LTW). [a]

	E2/20ºC
Refinement resolution (Å)	38.9-2.1
R_{cryst} (Fo>4sFo; Fo)	0.216
R_{free} (Fo>4sFo; Fo)	0.265
Refined residues[b]	271/280
Refined water molecules	96
Refined HLZ molecules	2
Overall isotropic B factor ($Å^2$)	18.83
RMSD for bond lengths (Å)	0.006
RMSD for bond angle (º)	0.885
Ramachandran plot	
Most favoured regions (%)	96.3
Allowed regions (%)	3.7
Disallowed region (%)	0

[a] Definitions of the statistical terms are listed in Appendix B. [b] The *N*- and *C*-termini residues could not be modelled due to their mobility as observed previously in the apo-protein structure (PDB code 2VFB).

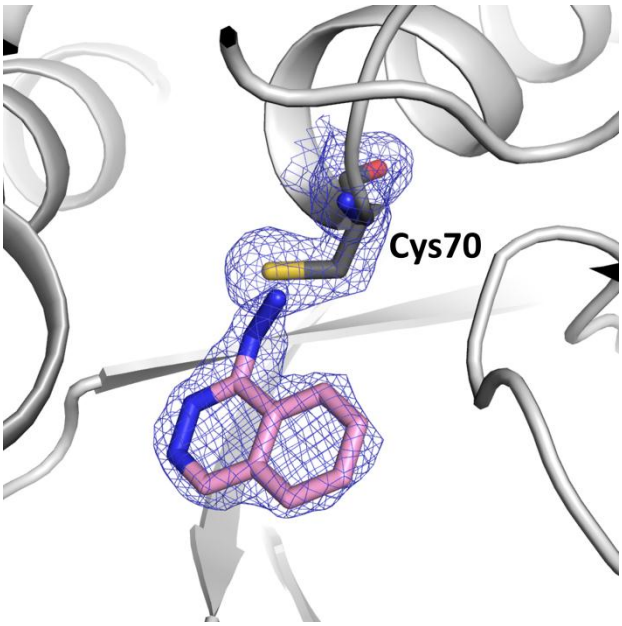


Figure 3.6: Electron density map of the MMNAT-HLZ complex.

The HLZ molecule bound in the active site is shown in pink with the 2Fo–Fc electron density shown in blue and contoured at 1 σ.

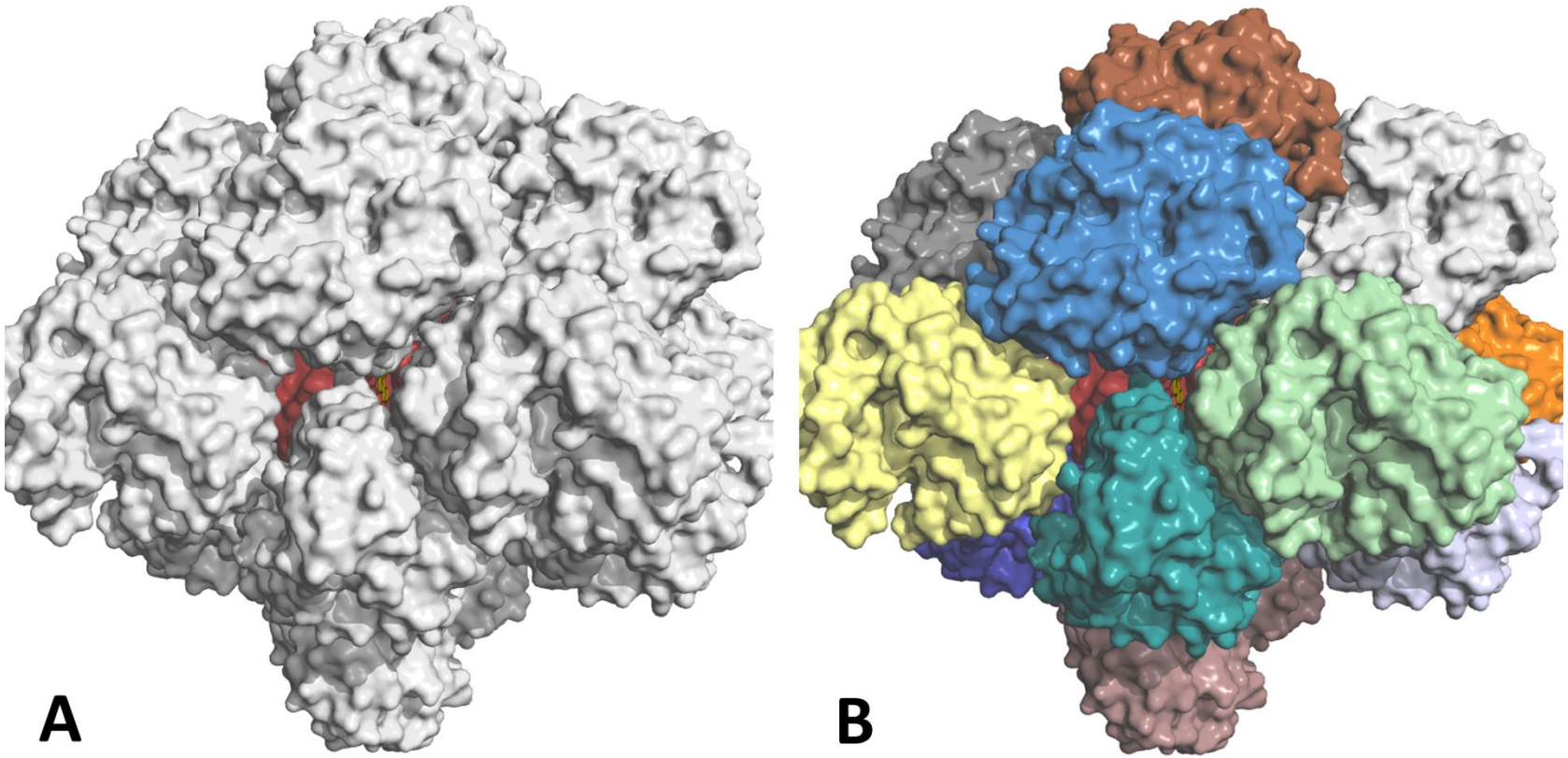


Figure 3.7:The crystal packing of the MMNAT-HLZ complex in the P $4_1 2_1 2$ space group.

The asymmetric unit making up the crystal unit cell is shown as the buried red molecule in surface view. The surface view of the symmetry-related molecules are shown in (A) white, and (B) different colours. The figure was generated by PyMOL software (Schrodinger, 2010).

3.2.6 Comparison of the MMNAT-HLZ structure with the apo-MMNAT structure

The MMNAT-HLZ complex structure is similar to the previously determined apo-MMNAT structure (PDB code 2VFB, 2 Å) (Fullam *et al.*, 2008; Fullam *et al.*, 2009), and consists of three domains: domain I is an α-helical bundle, domain II forms a β-barrel and domain III forms the α/β lid (Figure 3.8A). A comparison between the Cα-backbones of the native MMNAT and MMNAT co-crystallised with HLZ shows that these backbones are almost superimposable and have an RMSD of 0.281 Å (DS Visualizer). The backbones do, however, differ in the superficial hairpin loops, which correspond to amino acids 100-102, 169-171 and 214-216, that were reported to be mobile regions (Figure 3.8B) (Fullam *et al.*, 2008).

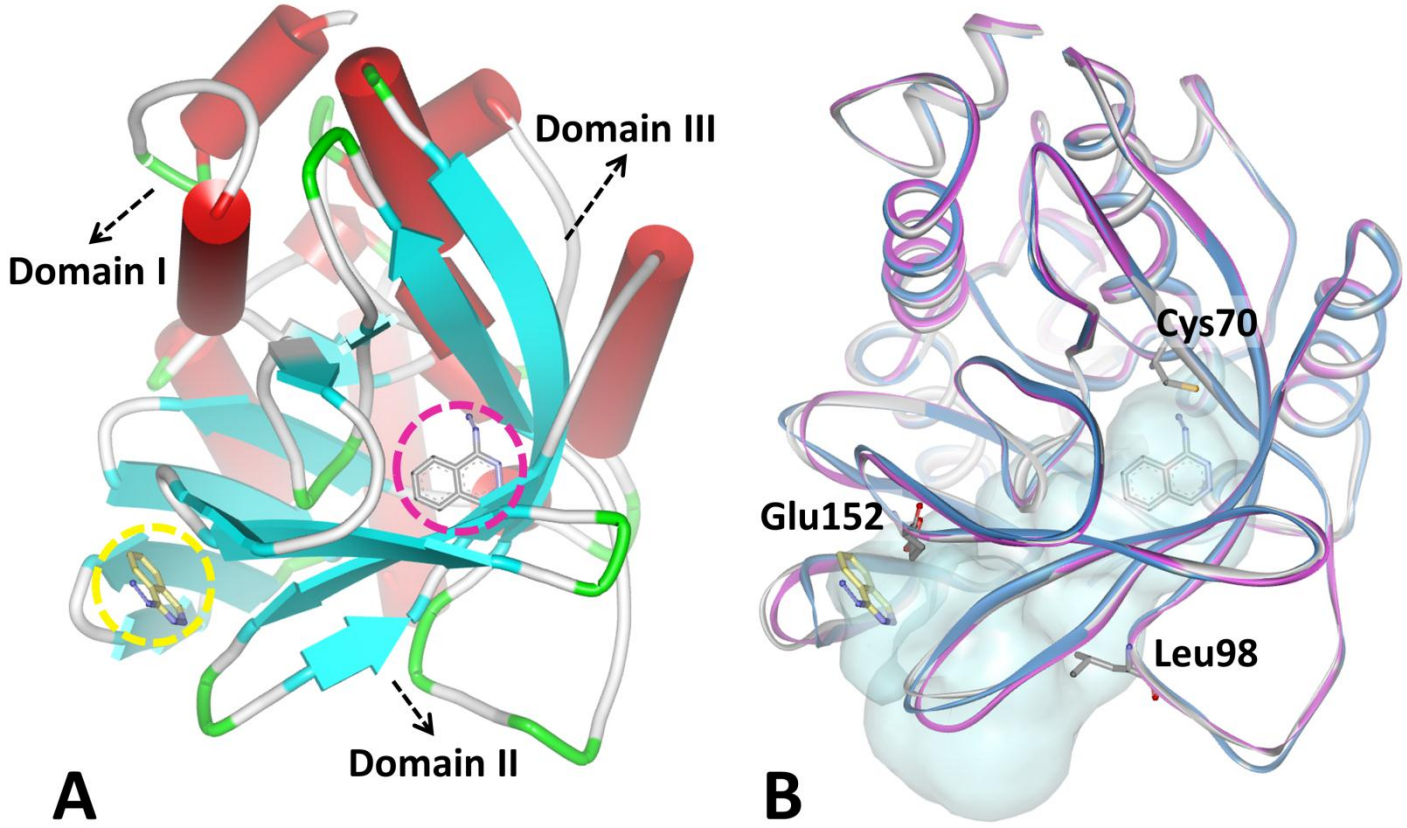


Figure 3.8: The MMNAT-HLZ complex.

(A) Schematic presentation of the MMNAT-hydralazine complex showing the secondary structure of the protein and the different subdomains. The structure is coloured according to the secondary structure. (B) A comparison of MMNAT, the MMNAT-CoA complex and the MMNAT-HLZ complex. The ribbon representation shows the superposition of the Cα-trace of the MMNAT apoprotein (PDB code 2VFB; 2.0 Å; grey), the MMNAT-CoA complex (PDB code 2VFC; 2.7 Å resolution; blue), and the MMNAT-HLZ complex (in pink). The figure also shows the space available for CoA binding in the enzyme presented by the co-crystallized CoA (PDB code 2VFC). Residues are highlighted to define the position of the CoA binding site within the protein. The hydralazine molecule bound in the active site is shown in pink (and the pink circle in (A)), and that bound to the secondary site in yellow (and the yellow circle in (A)). The figure was modified from Abuhammad *et al.* (2010).

3.2.7 Hydralazine binding pocket

NAT enzymes are characterised by having a wide active site that can accommodate the extended CoA moiety of Ac-CoA. This site occupies a volume of ~860 $Å^3$ (as determined by the DS Visualizer software) and extends between the flaps of the two β-sheets of the protein, forming an L-shaped cleft (Figure 3.9).

HLZ, which is a phthalazin-1-ylhydrazine, binds in a plane parallel to the β-sheet lid covering the active-site cleft, where it occupies an inner hydrophobic pouch of about 200 $Å^3$ within the binding pocket (Figure 3.9). This sub-pocket consists mainly of aromatic amino acids, including: Phe38, Tyr69, Val95, Trp97, Phe130, Phe204, and Met209 which form the bottom and the lid of this pocket, while the middle part is composed of the hydrogen-bond forming groups of Cys70, Thr109, Gly129 and Gly131 and a single explicit water molecule (Figure 3.9B). The water molecule is held in place by forming a network of hydrogen bonds with Val95 and Thr109 and is observed in other MMNAT structures (PDB code 2VFB, 2VFC). The residues constituting this binding space are shown in Table 3.4 with their average B-factors. All these amino acids are either conserved within mycobacterial NATs or conservatively substituted (Tyr71 and Val196 in MMNAT are Phe and Ala respectively in TBNAT) except for Met209 in MMNAT, which is replaced by Thr in TBNAT and NAT from *M. smegmatis* (**Error! Reference source not ound.**).

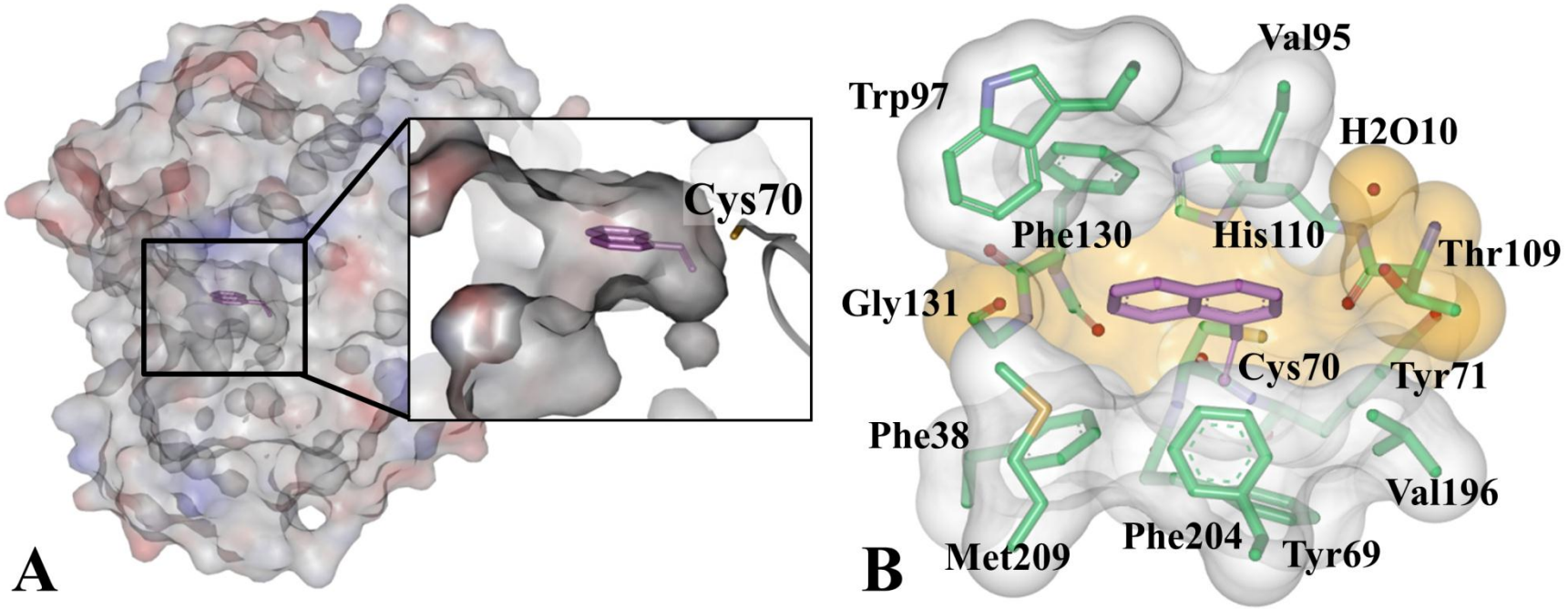


Figure 3.9: The hydralazine binding pocket within the MMNAT-HLZ complex.

(A) A solvent-accessible surface representation of the MMNAT-HLZ complex coloured by electrostatic potential (blue for positive and red for negative) shows the sub-pocket where HLZ binds. In the rectangle, an enlarged cross section shows the orientation of HLZ (in pink) in the pocket and the catalytic Cys70. (B) A surface representation of the residues in the active-site binding-pocket shows the hydrophobic areas in grey and the hydrogen-bond-forming-area including a water molecule (ID H2O10 in PDB 3LTW) in orange. The figure was prepared using DS Visualizer 2.5 (Abuhammad *et al.*, 2010).

Table 3.4: The amino acids that constitute the HLZ binding pocket within the active site with their average B-factors, and the RMS deviation of the B-factors within each amino acid. [a]

Type	ID	B Factor$_{average}$ (Å^2)	RMS (Å^2)
Phe	38	13.2	0.10
Tyr	69	16.0	0.27
Cys	70	15.6	0.28
Tyr	71	14.8	0.17
Val	95	17.1	0.19
Trp	97	20.5	0.29
Thr	109	18.4	0.22
His	110	16.8	0.22
Gly	129	16.1	0.33
Phe	130	17.3	0.25
Gly	131	18.2	0.22
Val	196	13.1	0.17
Phe	204	19.3	0.32
Met	209	19.0	0.33
HLZ	300	41.6	0.14
H_2O	10	16.6	16.65

[a] Data obtained by using BAVERAGE software implemented in CCP4 suite (Collaborative, 1994).

The phthalazine nucleus of HLZ forms strong π-π interactions with Phe130 that involve both the benzene and the pyridazine rings. This interaction is further supported by the π-σ interaction with Phe204 and Val95.

The hydrazine side chain is oriented towards the middle region of the pocket, which is made of relatively charged and hydrogen-bond-forming polar residues and can accommodate the hydrazino group of the HLZ molecule (Figure 3.9B). Although HLZ has four basic nitrogen atoms, only two of them (N2 and N4) show protolytic equilibrium as shown in Figure 3.10. In HLZ, the positive charge can be located at the imino nitrogen of the hydrazino group (N4; structure C, Figure 3.10) or at the protonated ring nitrogen (N2; structure D, Figure 3.10) (Datta *et al.*, 1976).

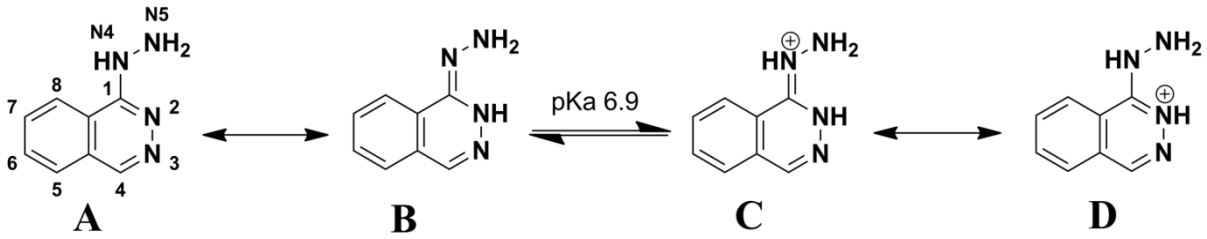


Figure 3.10: The different tautomers and protonation states of hydalazine.

(A) and (B) show the ionic states of HLZ. The pK interconnecting (A) and (B) is ≈ 7 in the free state. Protonation of (A) at N2 to yield (D) and of (B) to yield (C) can form π-cation interactions in the binding site and contribute to the higher acetylation efficiency of the hydralazine by MMNAT. The atomic numbering is according to Okabe *et al.* (1993).

HLZ can form hydrogen bonds with Cys70 and Thr109 (Figure 3.11), depending on which HLZ tautomer or ionization state is in the binding pocket. HLZ can also form a π-cation interaction with Tyr69, Tyr71, His110, Phe130 and Phe204 if it binds in an ionized protonated form (Figure 3.11C). This interaction, if present, would be expected to contribute to the observed affinity of MMNAT towards HLZ as a favourable acetyl group acceptor (Fullam *et al.*, 2009). Cation-π interactions between ligands and proteins have been reported in many 3D-structures, even in active sites that lack any anionic amino acid to compensate the positive charge of the ammonium group (Wouters, 1998; Gallivan and Dougherty, 1999). Examples include the interaction between sodium cation (Na^+) and the indole ring of tryptophan in the lysozyme structure (Wouters, 1998) (2 Å, PDB code is not available).

These interactions (i.e. cation-π) are considered much more favourable than an analogous interaction involving a neutral amine as usually observed in the interaction of ammonium ions with benzenes and olefins (Meot-Ner and Deakyne, 1985; Rodham *et al.*, 1993; Gallivan and Dougherty, 1999).

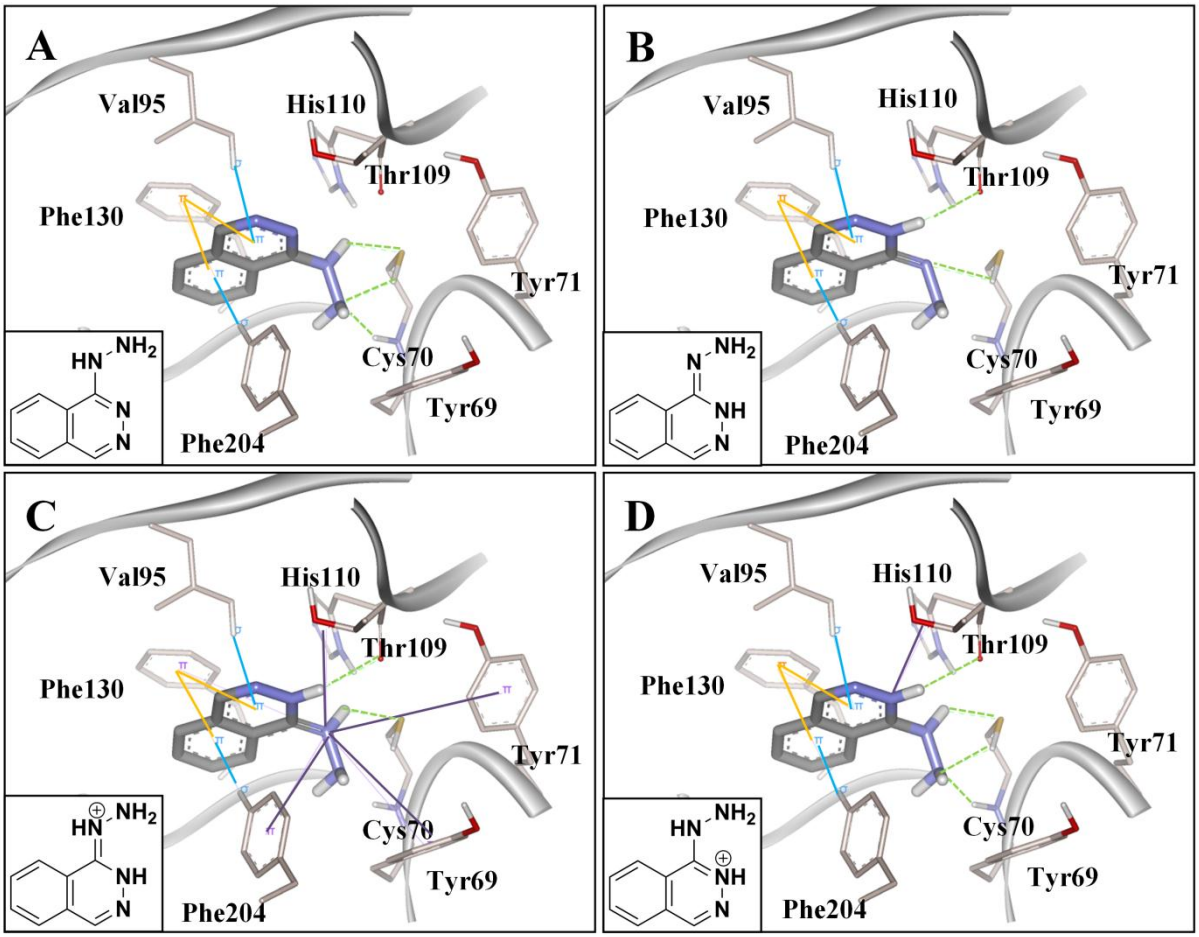


Figure 3.11: The interactions of the different tautomers and protonation states of hydalazine with the main amino acids in the MMNAT binding pocket.

Details of the main residues interacting with hydralazine in the MMNAT binding pocket. (A) and (B) The different possible HLZ tautomers (as presented in Figure 3.10). (C), and (D) The possible ionization conditions are shown. Orange: π-π interactions, blue: π-σ interactions, violet: π-cation interactions, green: hydrogen bonds. For clarity, the hydrogen atoms on some carbon atoms are shown. In each panel the chemical structure of HLZ is shown in the bottom left-hand corner. The figures were prepared using DS Visualizer 2.5 (Abuhammad *et al.*, 2010).

3.2.8 Mechanism of HLZ acetylation

The product of the biological acetylation of HLZ is a cyclic metabolite known as 3-methyltriazolo[3,4-a]phthalazine (structure D, Figure 3.12), whilst the uncyclized intermediate (compound B, Figure 3.12) has never been isolated. It is proposed that tautomer B and its protonated form C (Figure 3.12) are more likely to be the abundant species in the binding pocket of MMNAT. While the protonated form C would enhance the affinity of the enzyme MMNAT for HLZ, the unionized form B is likely to be more favourable as an acetyl group acceptor. Both nitrogen atoms of HLZ involved in the cyclic acetylation product (i.e. N2 and N5, Figure 3.12) have electronegativity in this tautomer.

The fact that the heterocyclic-nitrogen can exist in a non-aromatized form, and also given its orientation in the binding pocket relative to the catalytic Cys70, raises the possibility that acetylation could take place at the ring nitrogen N2 rather than at the hydrazine nitrogen, especially since this N is oriented close to the Cys70-sulfhydryl group (SH), where the acetyl group binds. In support of the idea that a heterocyclic N atom can be acetylated, heterocyclic-amines such as pyridine are used as catalysts in acetylation reactions. In these catalysed acetylation reactions, pyridine functions as a nucleophilic catalyst, forming an N-acylpyridinium compound (Butler and Gold, 1961), which is a more reactive species than pyridine itself, especially towards intramolecular rearrangement or cyclization. If a similar reaction occurs within the MMNAT binding pocket, this could explain the inability to identify the previously advised intermediate (B, Figure 3.12). It can also explain the higher efficiency with which HLZ is acetylated by NAT enzymes that have been tested compared with other aromatic hydrazines, e.g. isoniazid (Sikora *et al.*, 2008; Fullam *et al.*, 2009). However, at 2.1 Å resolution, it is not possible to determine which of the proposed tautomers or protonation states of HLZ is bound to the protein (Figure 3.11).

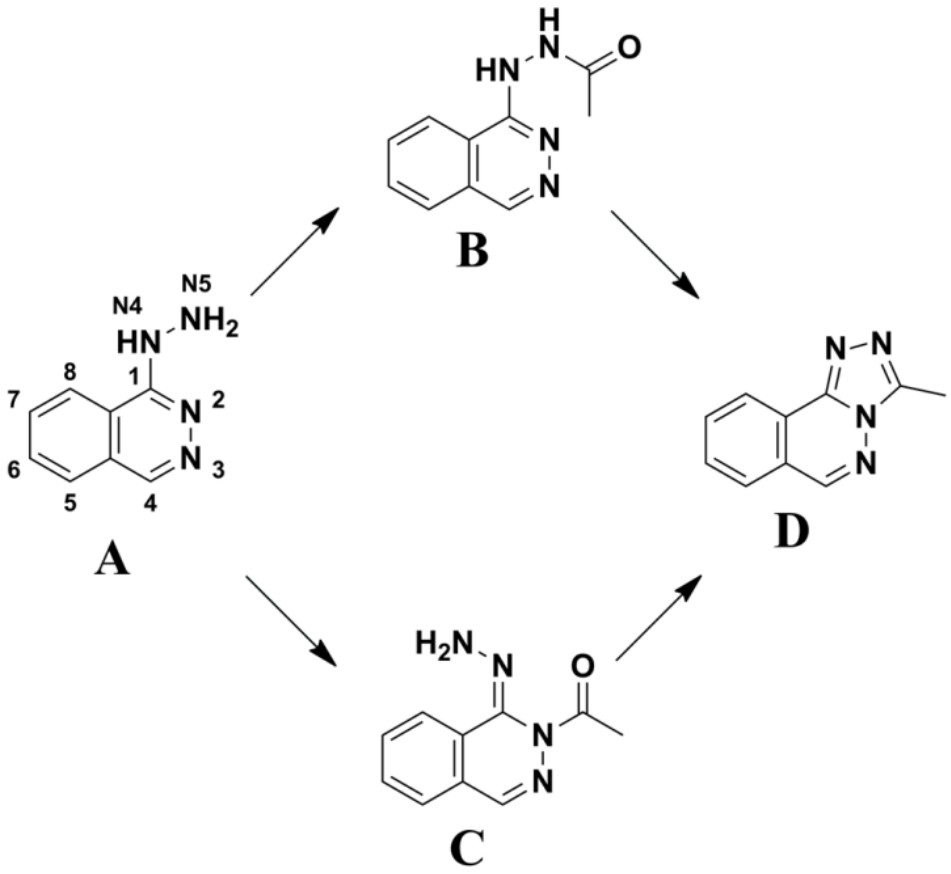


Figure 3.12: The acetylation of hydralazine.

Possible acetylation paths of hydralazine (A) by NAT enzymes leads to the formation of 3-methyltriazolo[3,4-a]phthalazine (D). The acetylated intermediate (B) has been suggested but never identified. The intermediate (C) was proposed from this work.

3.2.9 The HLZ secondary binding pocket

The second HLZ molecule binds to a superficial secondary binding-pocket located towards the end of the CoA binding site. HLZ binds to this pocket through interactions with hydrophilic residues whilst aromatic and hydrophobic interactions are not available. HLZ is hydrogen bonded to Arg170 and His229. Amino acid residues within this secondary binding pocket are conserved across NATs (**Error! Reference source not found.**). An quivalent binding pocket has been identified in an MMNAT-isoniazid complex (Fullam, 2007), where it helped in identifying the CoA binding site in the MMNAT-CoA complex (PDB code 2VFC; 2.7 Å resolution). However, the biological significance of the secondary binding site has not been established.

3.2.10 Comparison of the hydralazine structure with available NAT structures

3.2.10.1 Comparison with the MMNAT-isoniazid complex

In order to further explore the arylamine substrate-binding-pocket identified within the MMNAT active site, the interactions formed by isoniazid (INH) with MSNAT (PDB code 1W6F; resolution 2.1 Å; Figure 3.13) (Sandy *et al.*, 2005a) and with MMNAT (Fullam, 2007) were compared to that formed with HLZ. INH binds to the pocket in a comparable way to that observed for HLZ (Figure 3.13). The hydrazide group of INH is oriented towards the thiol group of Cys70, forming a hydrogen bond network with Cys70 and Thr109. The pyridine ring forms a π-π interaction with the benzene ring of Phe130. The ionization of the hydrazide group of INH is less likely because it is less basic than the hydrazino group of HLZ.

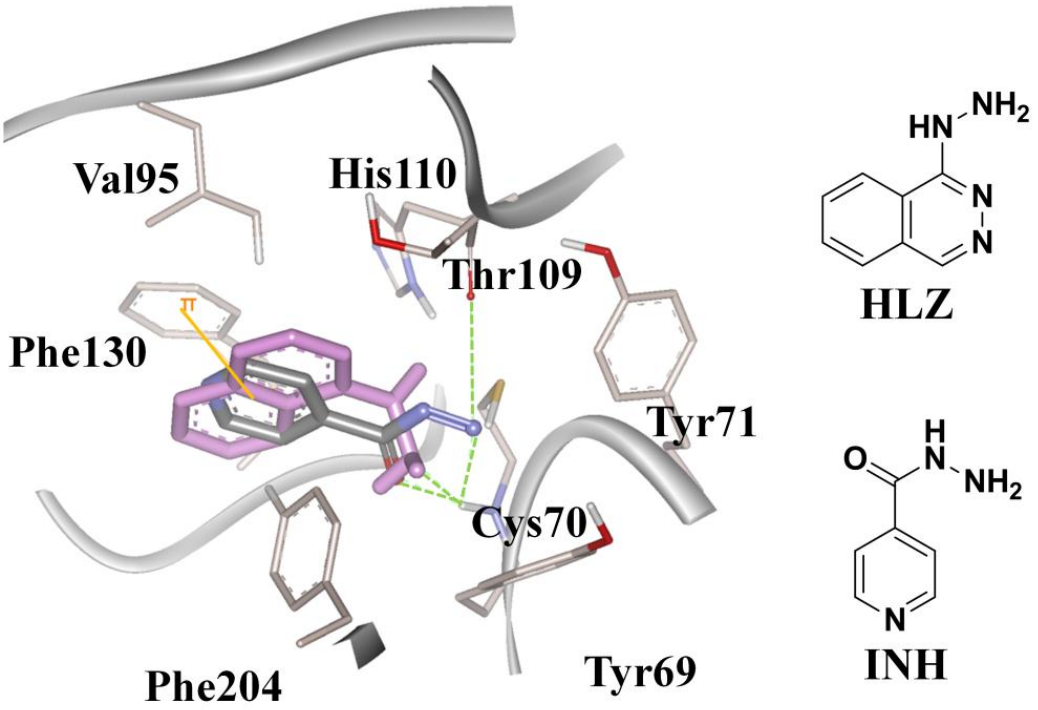


Figure 3.13: The binding of INH to the arylamine substrate-binding pocket.

A closer look at the main residues which interact with INH in the MSNAT (PDB code 1W6F; resolution 2.1 Å) binding pocket. Orange: *π*-*π* interactions, green: hydrogen bonds. The figure shows an overlay of hydralazine (HLZ) (in pink) and isoniazid (in grey) in the active site pocket. The chemical structure of both HLZ and INH is shown. The figure was prepared using Discovery Studio Visualizer 2.5 (Abuhammad *et al.*, 2010).

3.2.10.2 Comparison with the MMNAT-CoA complex and C70Q MSNAT structures

To investigate the rigidity of the active site, the HLZ-binding-pocket was compared with that of the MMNAT-CoA complex (PDB code 2VFC; 2.7 Å resolution) and also of the C70Q mutant of MSNAT (PDB code 1W5R, resolution 1.45 Å) (Figure 3.14) which approximates to the acetylated form of the protein (Sandy *et al.*, 2005b).

The -SH group of CoA (that carries the acetyl group in Ac-CoA) extends to the arylamine pocket and lies within hydrogen bonding range of the Cys70 thiol group, which indicates that both the Cys70 and CoA-SH groups are sufficiently close to exchange the acetyl group. The hydrophobic interactions between CoA and Phe204, Val95 and Tyr69 identify the role of these amino acids in substrate recognition. Figure 3.14A shows a large vacant volume in the binding pocket. The vacant volume available between Tyr69, Val196 and Phe204 is big enough to accommodate the acetyl group. Additional space is likely to be

provided by the flexibility of the side chains of these three residues, such that Pr-CoA could also be accommodated, since Pr-CoA can also act as an acyl donor to aryalmines and HLZ by MMNAT (Fullam *et al.*, 2009). We have also compared the binding pockets of the MMNAT-HLZ complex with that observed in the C70Q MSNAT mutant, which simulates the acetylated form of the enzyme (Figure 3.14C and Figure 3.15). There is a slight shift of ~ 0.8 Å of the glutamine moiety compared to Cys70 such that the amido group of the glutamine is directed towards the centre of the pocket, creating additional space (Figure 3.14C). This can be explained by the hydrophilic nature of the amido group that allows it to form hydrogen bonds with the bulk water molecules filling the pocket. The acetylated Cys is more hydrophobic and is likely to orient preferentially to the hydrophobic area created by the residues Tyr69, Val196 and Phe204. This shows that the active site pocket has only slight mobility.

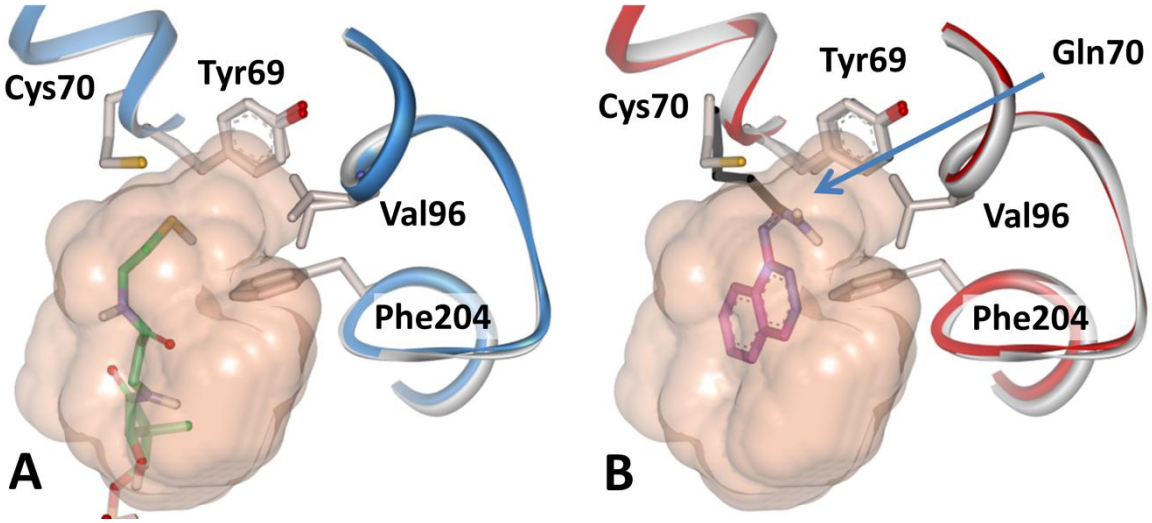


Figure 3.14: The binding pocket of hydralazine in different NAT structures.

Close-up of the active site with the space filling of the pocket cavity represented as a transparent solid surface (pink), showing the shape and volume of the pocket available for binding. The key residues around the binding pocket are labelled. (A) A superimposition of MMNAT-HLZ (white ribbons) and the MMNAT-CoA complex ((PDB code 2VFC; 2.7 Å resolution; blue ribbons) shows the space occupied by the CoA mercaptoethylamine group, and in (B) The superimposition of MMNAT-HLZ (white ribbon) and the C70Q mutant MSNAT (PDB code 1W5R; resolution 1.45 Å; red ribbon) shows the relative orientation of the thiol and hydralazine (in pink) to the glutamine side chain (shown in black and indicated with an arrow). It also shows the difference between the pocket sizes of the cavities and a vacant volume that can accommodate the acetyl group during the acetyl-transfer. The figures were prepared using DS Visualizer 2.5.

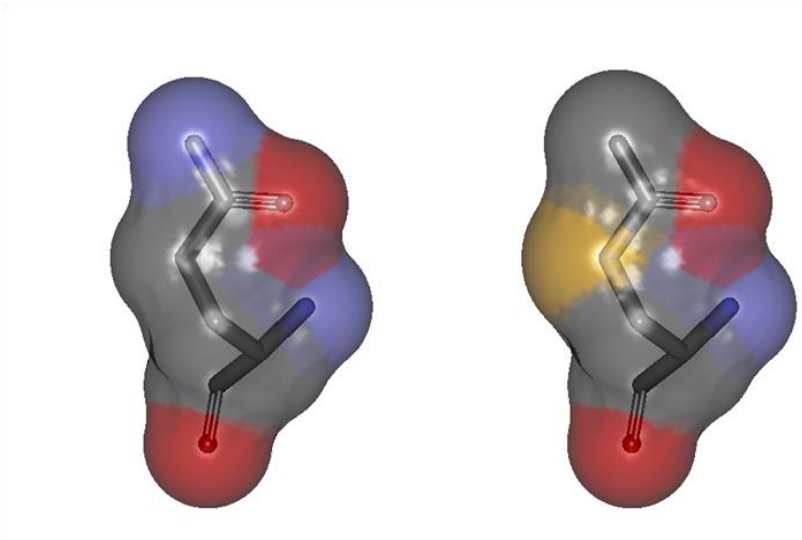


Figure 3.15: Shape comparison between a glutamine residue and the acetylated cysteine.

The close similarity in shape and volume of the two moieties is shown; the shapes are coloured according to the default colours of the atoms: carbon in grey, nitrogen in blue, oxygen in red and sulfur in yellow. The figure was prepared using DS Visualizer 2.5.

3.3 Conclusions

The mycobacterial NAT enzyme has been proposed as a possible target for antitubercular therapy, since deletion of the *nat* gene disrupts cell-wall lipid synthesis. Although the precise biochemical pathway has not yet been elucidated, small-molecule NAT inhibitors are being developed (Westwood *et al.*, 2010). MMNAT is the best available model of TBNAT and so has been used in the inhibitor-screening programme (Westwood *et al.*, 2010). Therefore, better understanding of the binding pocket of MMNAT is required as a starting point for structure-based drug design.

Comparison of the acyl-CoA profile of MMNAT to that of TBNAT revealed that MMNAT can also induce the hydrolysis of Pr-CoA in the presence of HLZ as does TBNAT. This finding supported the validity of MMNAT as a model of TBNAT. Diffraction quality crystals of the MMNAT-HLZ complex were obtained by co-crystallising the protein in the presence of excess HLZ. Optimisation of the initial crystallisation hits allowed diffraction data collection at 1.9 – 2.2 Å resolution.

Molecular replacement was used to solve the 3D-structure of the MMNAT-HLZ complex. Electron densities corresponding to the bound HLZ molecules were identified in the binding pocket as well as in a secondary pocket of the protein. The structure was refined at 2.1 Å resolution and deposited in the Protein Data Bank (3LTW).

This structure allowed the study of the binding pocket of the enzyme, and the interactions of HLZ were identified. Binding of HLZ arises through the cooperative forces of hydrophobic, aromatic, and hydrogen bonds. The key residues involved in binding have been elucidated. Aromatic residues stack over the phthalazine ring, providing a potentially strong binding interaction which is further enhanced by hydrogen bond formation and the possible π-cation interaction. The fact that the HLZ-binding-pocket within the active site is now clearly defined and is relatively immobile are good indicators for iterative structure

based medicinal chemistry investigation of inhibitors to improve their potency. An alternate mechanism for cyclization of HLZ in the acetylated form has been proposed by this work (Abuhammad *et al.*, 2010).

These studies have been carried out with the highly soluble MMNAT. It would be advantageous to carry out a similar study with TBNAT. Therefore, optimisation of TBNAT expression and purification to a quality suitable for structural studies are described in the next chapter.

4 Optimization of recombinant TBNAT production

4.1 Introduction

Despite the very high degree the of homology and the well conserved amino acid residues within the binding pocket of MMNAT and TBNAT enzymes, studies suggest that TBNAT, whilst resembling MMNAT, has its own characteristic substrate specificity and inhibition profile, and that it also shows inherent physical differences, e.g. in melting temperature (Fullam *et al.*, 2009; Abuhammad *et al.*, 2010). Solving the structure of an enzyme with an inhibitor bound is a highly desirable goal in the path to development of potent specific inhibitors as drugs. Hence, it is essential that a sufficient amount of recombinant TBNAT is available to pursue the enzyme further as a target for structure-based medicinal chemistry.

Previous attempts to generate recombinant TBNAT using various expression systems failed to yield sufficient soluble protein in the quantities and levels of purity required for structural studies or HTS (Table 4.1) (Payton *et al.*, 1999; Upton *et al.*, 2001; Sikora *et al.*, 2008; Fullam *et al.*, 2009). Therefore, the specific **aims** of the work presented in this chapter (see Introduction, Section 1.6) were:

- **To improve the TBNAT yield from the pET28b(+) system**
- **To sub-clone TBNAT in pVLT31 and express the recombinant protein in a suitable expression system**
- **To purify TBNAT to a level suitable for X-ray crystallographic studies with and without a His-tag**
- **To characterise purified recombinant TBNATprotein by mass spectroscopy**

Most of the wrok presented in this chapter has been published in Abuhammad *et al.* (2011).

Table 4.1: Expression systems used to produce TBNAT.

Vector	Expression tag	Host	Strain	TBNAT yield (mg per litre of culture)	Purity	References
Previous expression systems[a]						
pACE-1	No tag	*M. smegmatis*	MC2155	< 1	< 90 %	Payton *et al.* (1999); Upton *et al.* (2001)
pET28b(+)	His-tag	*E. coli*	BL21(DE3)pLysS	Insoluble	---	Upton (2001)
pET28b(+)	His-tag	*E. coli*	ArcticExpress RP	~2 (His TBNAT)	< 90 %	Sikora *et al.* (2008); Fullam *et al.* (2009)
Expression systems used in the present study						
pET28b(+)	His-tag	*E. coli*	ArcticExpress RP	~8 (His-TBNAT)	< 90 %	
pVLT31	His-tag	*E. coli*	NEB Express	6 -16 (TBNAT)	>95 %	
pVLT31	His-tag	*E. coli*	NEB Express	3 (His-TBNAT)	>95 %	

[a] **The enzyme quality and purity were sufficient for enzymology and functional studies.**

4.2 Results and discussion

4.2.1 Expression and purification of TBNAT in the pET28b(+) system

Production of recombinant TBNAT at low cultivation temperature (16 ºC) in the presence of protein folding chaperonins in the pET28b(+) vector (Sikora *et al.*, 2008; Fullam *et al.*, 2009) resulted in a soluble enzyme yield that outperformed other expression systems, including expression in *M. smegmatis* as host (Table 4.1) (Payton *et al.*, 1999; Upton *et al.*, 2001). Although the yield was sufficient for characterisation of the protein (1.2 mg per litre culture), it was impossible to purify the TBNAT to a quality suitable for structural studies (Table 4.1). Clearly, substantially more optimisation was necessary.

4.2.1.1 Optimization of TBNAT expression in the pET28b(+) system

Screening was conducted to optimize the induction conditions for recovery of soluble TBNAT enzyme from the pET28b(+) expression system. TBNAT generation in the ArcticExpress strain of *E. coli* was performed in small-scale cultures (30 mL) by varying the growth medium constituents and induction conditions as described in Table 4.2. The effect of the different induction conditions on the soluble enzyme yield was estimated by measuring the activity of the enzyme in the lysate without further purification, and the results are shown in Figure 4.1. Induction was performed at mid-exponential phase (Table 4.2). The conditions that resulted in higher specific enzyme activity in the soluble lysate in addition to low growth temperature (10 °C), were: (1) high level of osmolytes (1 M sorbitol and 2.5 mM betaine) in the media, (2) induction at low IPTG concentration (100 μM) and (3) extending the growth time following induction with IPTG (48 h). Osmolytes have been reported to enhance the stability of the proteins in the cytoplasm and hence may increase the folding efficiency and protect proteins from degradation (Ignatova *et al.*, 2007). In the presence of sorbitol and betaine, a slower growth rate was observed, which is likely to add to the effect of the low cultivation temperature (i.e. 10 °C) in

allowing better protein folding. Lowering the IPTG concentration from 250 μM to 100 μM resulted in at least a 1.5-fold increase in the soluble protein yield after 24 h of induction. Higher protein yield at lower IPTG concentration is not uncommon and has been observed previously (Yildir *et al.*, 1998; Olaofe *et al.*, 2010).

Table 4.2: The LB medium additives and induction conditions used in the TBNAT expression.

	pET28b(+)	**pVLT31**	
	***E. coli* ArcticExpress**	***E. coli* NEB Express**	***P. putida***
LB medium additives	1 M Sorbitol or 1 M Sorbitol + 2.5 mM Betaine	1 M Sorbitol + 2.5 mM Betaine	1 M Sorbitol
Growth level at which induction performed (OD_{600})	~0.6	0.5 or 0.9	0.9
Concentration of IPTG (μM)	100 or 250	100	100
Induction and post-induction incubation temperature (°C)	10	15	20
Duration of post-induction incubation (h)	18, 24, or 48	60	18

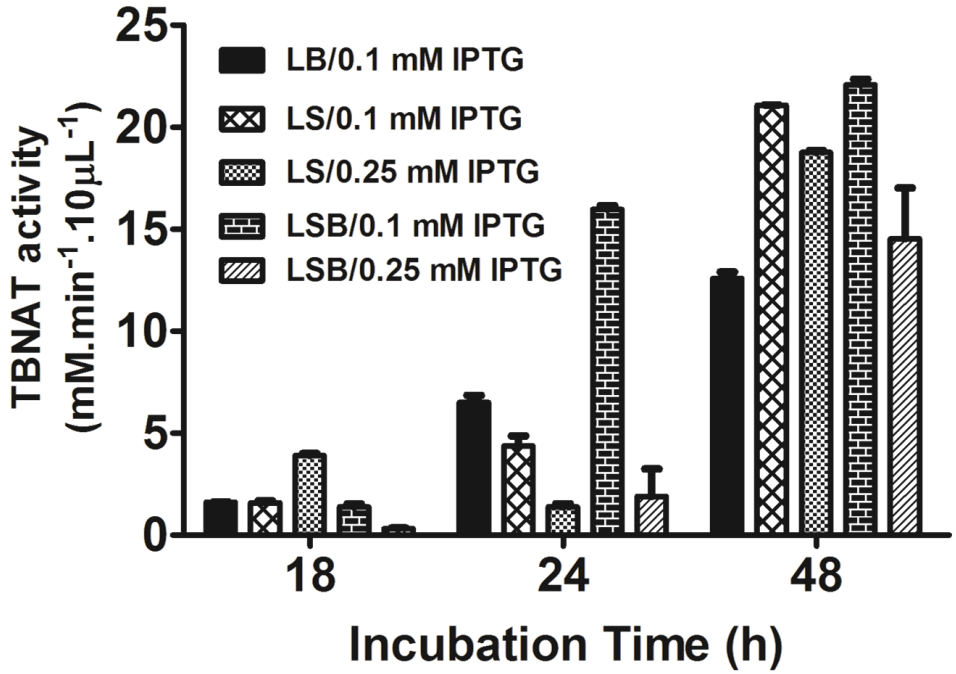


Figure 4.1: Optimisation of TBNAT production in *E. coli* ArcticExpress transformed with pET28b*(+)-tbnat.*

The performance of different induction conditions at 10 ºC that were listed in Table 4.2 was measured as described in Section 2.2.6.1. The activity of TBNAT in 10 µl of the soluble lysate from each preparation was assayed by measuring the rate of acetyl-CoA hydrolysis (400 µM) in the presence of HLZ (500 µM) and the results are shown as the mean ± S.D. of duplicate determinations of the specific activity of the enzyme. LB: Luria Bertani medium, LS: LB medium with 1 M sorbitol, LSB: LB medium with 1 M sorbitol and 2.5 mM betaine, IPTG: isopropyl β-D-1-thiogalactopyranoside.

4.2.1.2 Large-scale expression in the pET28b(+) system

Large-scale expression (one-litre culture) of *tbnat* under the conditions identified for optimal expression (LB medium supplemented with 1 M sorbitol and 2.5 mM betaine, and induced with 100 μM IPTG for 48 h at 10 °C) resulted in high TBNAT yield as the whole-cell extract SDS-PAGE results show (Figure 4.2A). However, the bulk of the protein was present in the insoluble fraction (Figure 4.2A) and less than 8 mg per litre of culture was retrieved in the soluble extract in association with His-rich co-purifying proteins (Figure 4.2B).

During subsequent manipulations, it was observed that the enzyme adhered irreversibly to different matrices and membranes. This behaviour complicated the purification process and resulted in extensive loss. For this reason, a better expression system, which required fewer purification steps was needed.

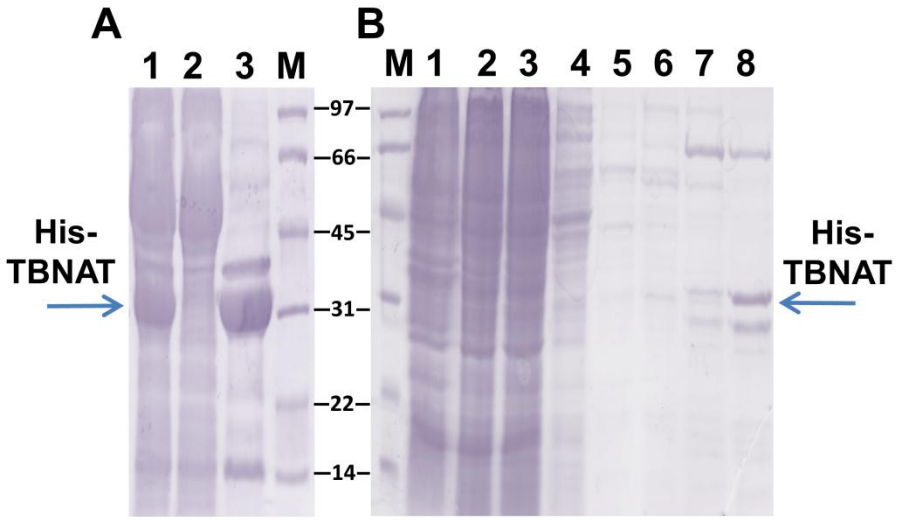


Figure 4.2: Expression and purification of recombinant TBNAT in ArcticExpress *E. coli* transformed with pET28b(+)-*tbnat*.

Coomassie blue-stained 12 % SDS PAGE showing: (A) The over-expression of TBNAT. Whole-cell lysate (lane 1) soluble fraction (lane 2) and insoluble fraction (lane 3). (B) The purification of TBNAT by washing a Ni-NTA column loaded with the His-TBNAT with a series of steps of increasing imidazole (Imz) concentration: soluble lysate (lane 1), unbound fraction (lane 2), 5 mM Imz (lane 3), 10 mM Imz (lane 4), 25 mM Imz (lane 5), 50 mM Imz (lane 6), 100 mM Imz (lane 7), 250 mM Imz (lane 8) and Bio-Rad low-range molecular mass protein standard (lane M in both A and B). Purification was performed as described in Section 2.2.3.5 and gel loading was performed as described in Section 2.2.4.2.

4.2.2 Sub-cloning of the *tbnat* gene

To improve the yield of the protein to a level that would allow further manipulation and purification, different expression systems were explored. There was previous successful experience in the laboratory of using the pVLT31 vector for another mycobacterial gene in *P. putida* as host (Lack *et al.*, 2010), where the soluble protein yield significantly outperforms that obtained in pET28b(+) vectors. pVLT31 is a broad-host expression vector that induces the transcription of a cloned gene by a *ptac* promoter, which is a negatively-controlled strong promoter regulated by the *lacI*^{q} repressor (de Lorenzo *et al.*, 1993). Despite the successful use of this vector to generate recombinant proteins from the *Pseudomonas* species (de Lorenzo *et al.*, 1993; Ueda and Wood, 2008), its performance has been less often reported in *E. coli* as the host. Therefore, it was chosen for TBNAT production using both *P. putida* and *E. coli.*

Sub-cloning of the gene encoding *tbnat* from a pET28b(+) construct into the pVLT31 vector was performed as described in Section 2.2.2.7 and is summarised in Figure 4.3. As the pVLT31 vector does not encode for a fusion-tag, the construct from the pET28b(+) plasmid that contained *nat* was further subcloned to include a thrombin digestion site, a ribosome binding site and a His-tag. After subcloning, the construct was transformed into *E. coli* JM109 and selected with tetracycline LB agar. The recombinants were confirmed by DNA sequencing analysis. Preliminary expression trials indicated that TBNAT could be produced at a high level under the control of the *lacI*^{q} promoter system in pVLT31 in JM109 (Figure 4.4). The recombinant plasmids were then used to transform *P. putida* KT2442 and *E. coli* NEB Express cells and used for protein production. Strain KT2442 of *P. putida* is a rifampicin-resistant derivative of *P. putida* KT2440, a restriction-deficient strain that has been used extensively as host for gene cloning and (Jiménez *et al.*, 2002; Chung *et al.*, 2009). *E. coli* NEB Express, is a versatile non-T7 expression strain that is

deficient in both the *lon* protease and the *ompT* protease which makes it more amenable to the production of proteins from cloned genes.

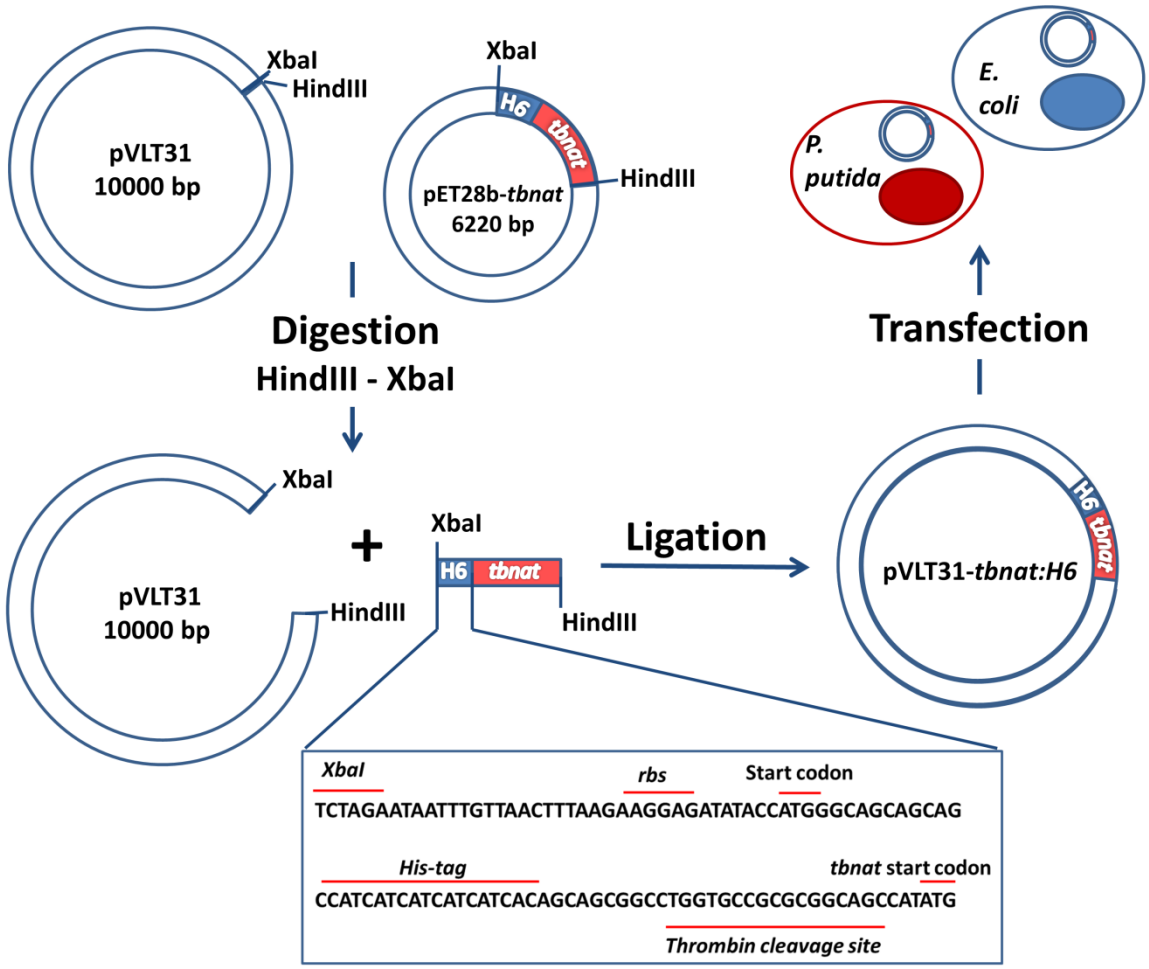


Figure 4.3: A schematic representation of the general cloning strategy used to sub-clone the *tbnat* gene from the pET28b(+) vector into the pVLT31 vector.

All techniques were carried out as described in Section 2.2.2.7. H6 refers to the His-tag. The thrombin cleavage site is at the junction of the H6 and *tbnat*. The ribosome binding site (rbs), the his-tag and the thrombin cleavage site fragment from pET28b(+)-*tbnat* is shown within the box. The cleavage is such that site leaves three extra amino acid residues (GSH) at the *N*-terminus of TBNAT following cleavage.

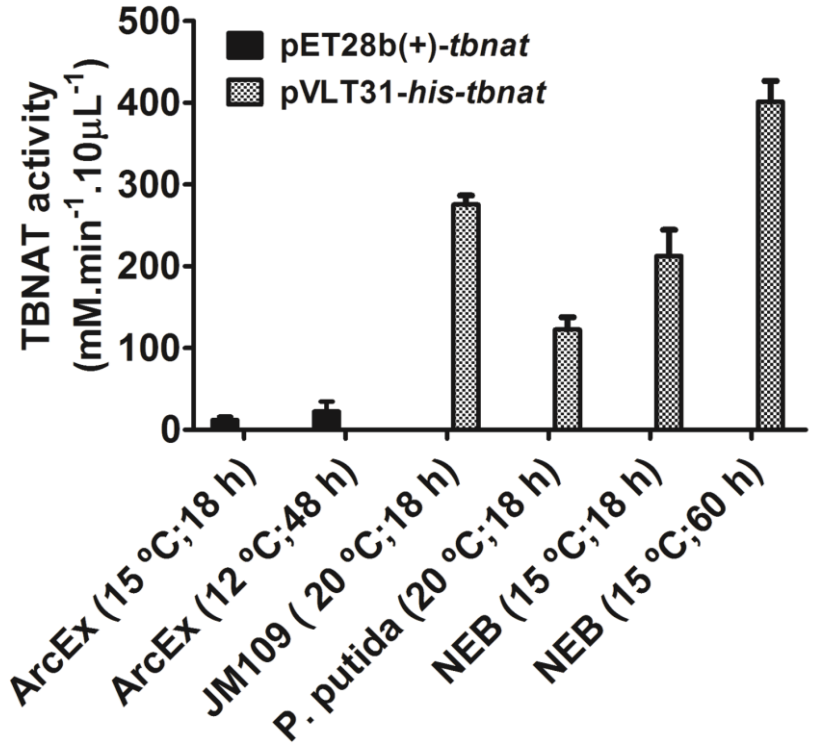


Figure 4.4: Comparison of the recombinant TBNAT yield in pET28b(+)- and pVLT31-based expression systems.

The pellets from 1 L culture were resuspended in 35 mL lysis buffer (Section 2.2.3.4) and the NAT activity in 10 µl of soluble lysate measured as described in Section 2.2.2.7, was used to estimate the yield of soluble recombinant TBNAT in the various expression systems and under the different induction conditions. The results are shown as the mean ± S.D. of triplicate measurements. ArcEx: *E. coli* ArcticExpress; JM109: *E. coli* JM109; NEB: *E. coli* NEB Express.

4.2.3 TBNAT expression in the pVLT31 system

In initial tests, the *E. coli* NEB Express strain showed better soluble TBNAT yield than *P. putida,* and thus *E. coli* NEB Express strain was used for further optimisation. Expression in *E. coli* was performed with LB medium supplemented with 1 M sorbitol and 2.5 mM betaine, which was then induced with 100 μM IPTG at 15 ºC. As observed previously with the ArcticExpress strain and pET28b(+), the low induction and cultivation temperature coupled with extended induction showed better enzyme yield. Figure 4.4 shows the activity of TBNAT in the soluble crude-extract for the selected investigated conditions. The best protein yield was obtained upon extending the induction duration to 60 h (Figure 4.4).

Changing the expression vector has been reported to increase the yields of soluble active proteins (Ueki *et al.*, 2006). The pET expression system is widely used because of its ability to mass-produce proteins. The pET vector (Figure 2.1) utilises a tightly controlled T7 promoter, which only binds T7 RNA polymerase (Novagen, 2002). Because T7 is a viral promoter, it transcribes rapidly and profusely for as long as the T7 RNA polymerase is present. The high yield of TBNAT protein in the T7 system resulted mainly in unfolded protein that formed inclusion bodies (Figure 4.2) which were difficult to avoid despite the extensive optimisation of the expression conditions.

In contrast, the *ptac* promoter in pVLT31 vector is a leaky hybrid promoter that allows expression to a lower extent compared with the T7 system. Although a strong promoter, it resulted in a lower overall yield of TBNAT, but favoured well-folded soluble recombinant protein.

4.2.4 Purification protocols

TBNAT purification was performed as summarised in the flowchart in Figure 4.5.

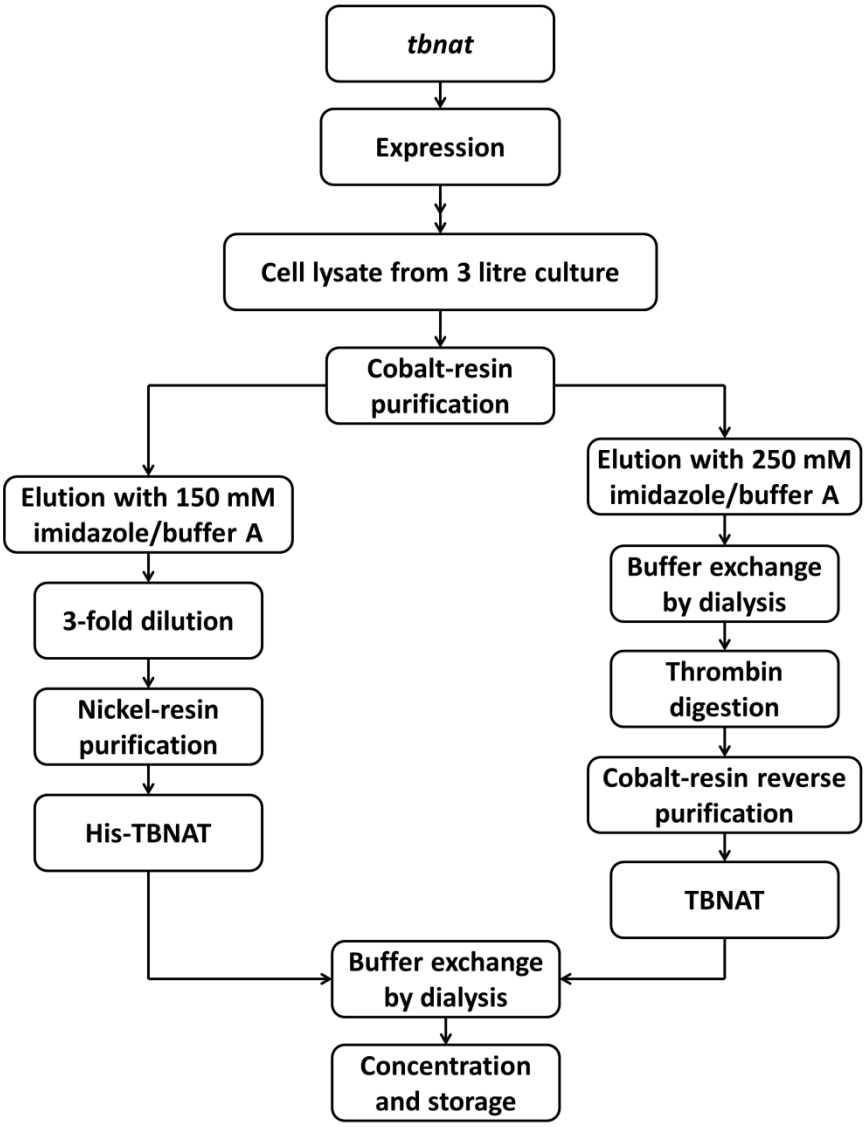


Figure 4.5: A flowchart summarising the purification protocol of TBNAT with and without a His-tag.

4.2.4.1 Purification of TBNAT

The standard IMAC purification protocol for the enzyme as reported previously (Sikora *et al.*, 2008; Fullam *et al.*, 2009) failed to produce the level of purity required for structural studies (i.e. purity more than 90-95 %) (Figure 4.2B). To achieve the desired level of purity, a comparison between the performance of Nickel- and Cobalt-metal affinity resin (Ni-NTA and TALON, respectively) was performed. Aliquots of the different washes were collected for SDS-PAGE analysis, and the results are shown in Figure 4.6. The Cobalt-column showed higher selectivity towards His-TBNAT, but with lower affinity, as reflected in the high level of His-TBNAT activity in the flow-through. In order to try to increase the binding of the protein to the column, phosphate buffer was used instead of Tris-HCl (20 mM, pH 8) and a gradient of increasing ionic strength was applied (75-300 mM NaCl). A capacity of ~0.75 mg protein per mL resin bed volume was obtained by using 20 mM phosphate buffer at pH 8 and ionic strength equivalent to 75 mM NaCl (this will be referred to as buffer A).

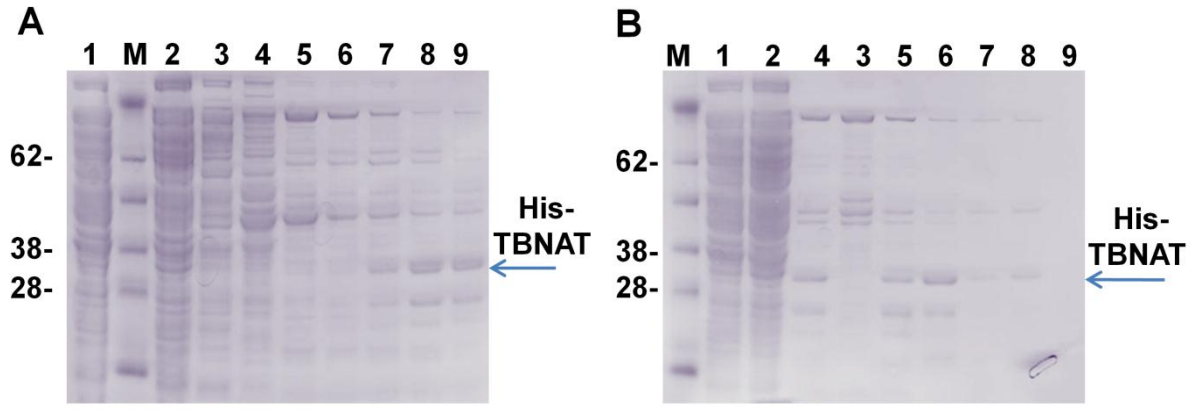


Figure 4.6: Comparison of the His-TBNAT purification by the Nickel- and Cobalt- resins.

Coomassie blue-stained 4-12 % SDS PAGE (NuPAGE Novex 4–12 % Bis-Tris Gel, Invitrogen) showing the purification of recombinant His-TBNAT. (A) TBNAT purification using the nickel-column, (B) TBNAT purification using the cobalt-column. Soluble lysate (lane 1), unbound fraction (lane 2), 5 mM Imz/buffer A (lane 3), 10 mM Imz/buffer A (lane 4), 25 mM Imz/buffer A (lane 5), 50 mM Imz/buffer A (lane 6), 100 mM Imz/buffer A (lane 7), 250 mM Imz/buffer A (lane 8), 500 mM Imz/buffer A (lane 9) and SeeBlue Plus2 marker (Invitrogen) (lane M). Buffer A: 20 mM Na-phosphate buffer at pH 8, 75 mM NaCl. The molecular weights of the marker proteins in kDa are indicated. Gel loading was performed as described in Section 2.2.4.2.

His-TBNAT was eluted with the buffer A containing 250 mM imidazole (Table 4.3). The eluted protein was exchanged by dialysis against 20 mM Tris-HCl buffer, pH 8.0 at 4 ºC using 10 kDa molecular weight cut-off Slide-A-Lyzer Dialysis Cassettes (Pierce). The His-tag was cleaved by incubating His-TBNAT with thrombin bound CleanCleave resin (Sigma) at 37 ºC for 4 h and the His-tag with the copurified His-rich proteins were removed by passing the solution through a fresh cobalt-column (reverse purification). Figure 4.7 shows the success of this step in producing high-purity protein that could be used for further structure determination as indicated by the intensity of the Coomassie stained single target band in an SDS-Polyacrylamide Gel (PAGE). The pure TBNAT (with the His-tag removed) was concentrated using 10 kDa molecular weight cut-off VivaSpin Protein Concentrators (Sartorius Stedim Biotech) and aliquots of 10-20 mg/mL in 20 mM Tris-HCl buffer pH 8.0 were stored at -80 ºC with 5 % v/v glycerol and 1 mM DTT. Final yields of 6.5-16 mg of soluble crystallographic grade purity TBNAT per litre of bacterial culture have been achieved, compared to yields of around 1 – 2 mg per litre of culture previously reported (Table 4.1). The pellets from each of three 1-litre cultures were combined for each purification run. Lower yield was observed when the cell lysate from the three cultures was combined before binding to the resin in the first IMAC run. An additional gel filtration step resulted in extensive loss (more than 90 %) of the protein due to a nonspecific adsorption on the resin and an inability to retrieve the adsorbed protein, and so was subsequently omitted.

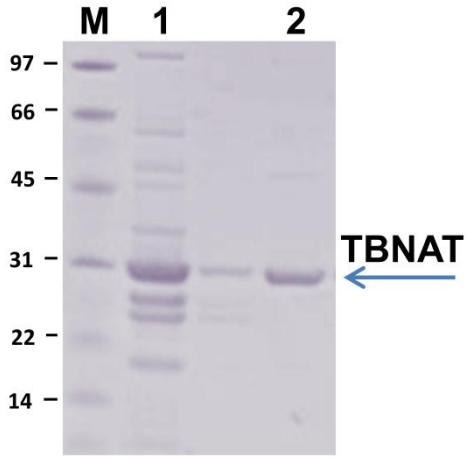


Figure 4.7: The final step in TBNAT purification.

Coomassie blue-stained 12 % SDS PAGE shows the TBNAT after thrombin cleavage of the His-tags before (lane 1) and after (lane 2) passing the cleaved TBNAT though a fresh cobalt-column (reverse purification). Bio-Rad low-range molecular mass protein standard (lane M). Gel loading was performed as described in Section 2.2.4.2.

Table 4.3: Data for purification of TBNAT expressed in NEB Express *E. coli* transformed with the pVLT31-*his-tbnat* vector.

Fraction	Total protein[a] (mg)	Total activity[b] (U)	Specific activity[c] (U/mg)	Purification factor[d]	Overall yield[e] (%)
Soluble lysate	2567.4	1812.1	0.71	1.0	100.0
Flow through	1628.6	340.06	0.21	0.3	18.8
5 mM Imz	1583.2	53.3	0.03	0.05	2.9
buffer A wash	983.7	30.4	0.03	0.04	1.7
250 mM Imz	21.7	1177.3	54.2	76.8	65.0

[a] Measured by the method of Bradford (Noble *et al.*, 2009). [b] The activity of the enzyme (10 µL of each fraction was diluted in buffer A: 20 mM Tris-HCl pH 8.0 to produce a linear initial rate) was determined by the measuring the rate of CoA formation in the presence of 500 µM HLZ. Total activity corresponds to the activity in the total volume of each fraction, and the unit of enzyme activity U is defined as the amount which produces 1mM of CoA per minute under standard assay conditions described in Section 2.2.6.1. [c] Specific enzyme activity = total activity/total protein. [d] Purification fraction = specific activity of each fraction/specific activity of soluble lysate. [e] Overall yield = total activity of each fraction/total activity of the soluble lysate.

4.2.4.2 Purification of His-TBNAT

The cleavage of the His-tag was critical for obtaining high-purity protein suitable for subsequent structural studies (Section 4.2.4.1). However, crystallisation of the purified protein proved to be challenging. Extensive screening for crystallisation conditions of the TBNAT failed to produce any crystals (Chapter 6). However, the purification of the tagged TBNAT can affect the ability of the protein to crystallise. Therefore, an additional purification step aimed at removing the co-purifying proteins but having a minimal effect on the yield of the TBNAT with the His-tag intact was required.

His-TBNAT showed different affinity towards the Nickel- and Cobalt-metal affinity resin used in the purification (Figure 4.6). The His-TBNAT was retained on the Nickel-resin at an imidazole concentration of 50 mM. However, this imidazole concentration was strong enough to elute the His-TBNAT from the Cobalt-resin (Figure 4.6). Therefore, for the purification of the enzyme with the His-tag, the elution of the enzyme from the first Cobalt-resin purification step (Section 4.2.4.1) was performed using 150 mM imidazole instead of 250 mM, and the thrombin-cleavage and the subsequent Cobalt-column steps were omitted (Figure 4.5). The reduced level of imidazole (i.e. 150 mM) was sufficient enough for complete protein elution from the Cobalt-column (Figure 4.6B). The His-TBNAT was 3-fold diluted with buffer A to lower the imidazole concentration to 50 mM. The enzyme solution was then incubated with a suitable volume of Nickel-resin that had been equilibrated with buffer A containing 50 mM imidazole: as mentioned earlier, the His-TBNAT binds to the Nickel-column in the presence of 50 mM imidazole. The column was then washed with buffer A containing increasing levels of imidazole (50 mM, 60 mM, and 75 mM) and the bound His-TBNAT was eluted with 150 mM imidazole in buffer A (Figure 4.8). In subsequent preparations, washing was performed using buffer A containing 60 mM imidazole followed by elution with 150 mM imidazole solution. A summary description of the purification protocol is shown in Figure 4.5. The additional

Nickel-resin step successfully resulted in high-purity His-TBNAT that could be used for protein crystallisation trials (Figure 4.8). Interestingly, the SDS-PAGE analysis of samples from the purification steps showed diffuse low molecular weight bands (Figure 4.8). These bands were likely to be due to lipids, since the association of TBNAT with membrane lipids has been observed previously in a whole-cell analysis (Malen *et al.*, 2010).

The enzyme showed better stability in PIPES buffer (Section 6.2.3.1) and therefore, it was used as a storage buffer. The protein was buffer exchanged with 20 mM PIPES pH 7, 150 mM NaCl. The protein was initially concentrated by dialysis against 50 mL Thermo Scientific Slide-A-Lyzer Concentrating Solution for 4-5 h. This step was followed by further concentration using Centrisart I Concentrator (10 kDa molecular weight cut-off, Sartorius Stedim Biotech) to 10 mg/mL. The protein was stored in 20 mM PIPES pH 7, 150 mM NaCl by snap freezing in liquid nitrogen followed by storage at -80 ºC. The protein was stable when stored without glycerol, and thereofore, glycerol was excluded from the storage buffer in order to avoid any further manipulations when used for setting up crystallisation trials. For His-TBNAT production, the cell lysate from three 1-litre cultures was combined and purified, and a final protein yield of 3 mg pure His-TBNAT per litre bacterial culture was obtained. The low yield obtained was expected due to the additional Nickel-resin purification step.

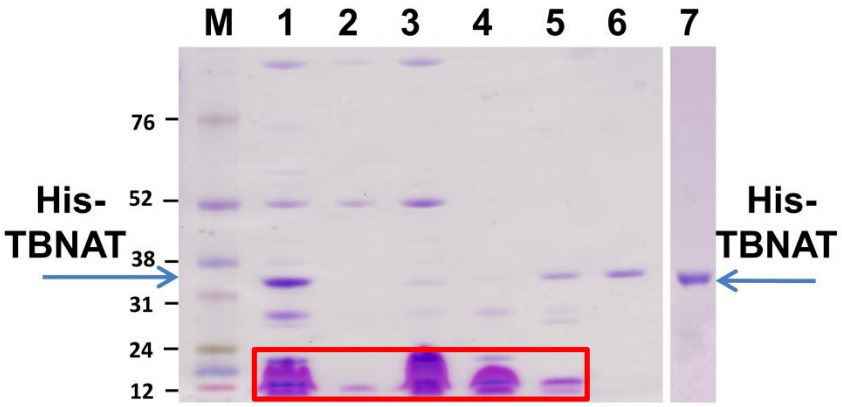


Figure 4.8: Purification of His-TBNAT.

Coomassie blue-stained 4-12 % SDS PAGE (NuPAGE Novex 4–12 % Bis-Tris Gel, Invitrogen) showing the purification of recombinant His-TBNAT using the Ni-NTA resin. His-TBNAT eluted from the Cobalt-resin in 150 mM Imz/buffer A (lane 1), unbound fraction (lane 2), 50 mM Imz/buffer A (lane 3), 60 mM Imz/buffer A (lane 4), 75 mM Imz/buffer A (lane 5), 150 mM Imz/buffer A (lane 6), His-TBNAT after the initial dialysis and concentration as described in Section 4.2.4.2 (lane 7), and High-Range Rainbow marker (GE Healthcare) (lane M). Buffer A: 20 mM phosphate buffer at pH 8, 75 mM NaCl. Diffuse bands that are likely to be due to co-purifying lipids are indicated within the red box.

4.2.5 Identification by mass spectrometry

The identity of the TBNAT (without His-tag) protein was confirmed by Electrospray Ionization Mass Spectrometry (ESI-MS) of the purified TBNAT, and the mass of the protein showed a single, intact species of molecular mass of 31314 ± 16 Da, (Figure 4.9) which is in agreement with the expected molecular mass from the sequence of 31310 Da.

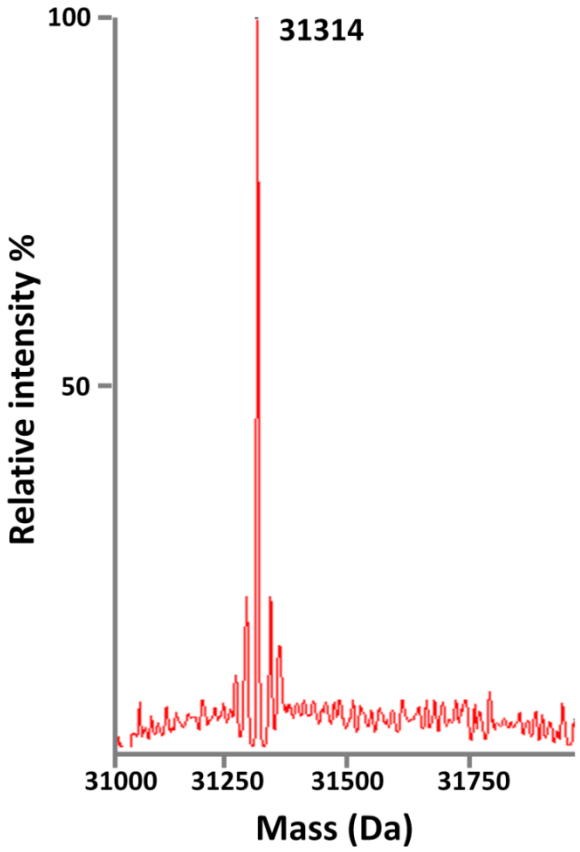


Figure 4.9: Electrospray-ionization mass spectra of TBNAT purified as described in Section 4.2.4.1 obtained after denaturation in 50 % v/v acetonitrile in water (Section 2.2.4.3).

4.3 Conclusions

Recombinant protein expression and purification are major obstacles towards protein crystallography and HTS. To allow for the exploration of TBNAT as a drug target a reliable source of sufficient amounts of the enzyme must be available. High protein yield allows for further inhibitor screening, the 3D structure determination and would facilitate drug design.

Expression of TBNAT in a pET28b(+) expression system was optimised and 4-fold improvement on the yield was achieved (8 mg compared to ~2 mg protein per litre bacterial culture). Conditions contributing to this improvement include: high osmolyte level, induction at low IPTG, and extended growth time following induction. However, this yield was not sufficient to allow for purification of the protein to a quality suitable for structural studies.

Sub-cloning of the tbnat gene into the pVLT31 vector was performed, and the engineered vector was transformed into *E. coli* and *P. putida* expression hosts. Expression in *E. coli* NEB Express resulted in a high yield of TBNAT, and the expression conditions were further optimised. A significant improvement on the yield was achieved, and the level of soluble protein yield was sufficient for purification to be undertaken.

A purification protocol for the TBNAT with and without a His-tag was developed. The use of the selectivity Cobalt-resin as a first step in the extraction of the soluble enzyme resulted in improved purification levels compared to previous purification trials, whilst the second reverse Cobalt-resin purification step resulted in the recovery of pure protein with and without the His-tag (~ 95 % purity as visually evaluated on the SDS-PAGE; size-exclusion chromatography evaluation will be described in Chapter 6). A final yield of the TBNAT (without His-tag) of 6.5-16 mg of soluble crystallographic grade purity TBNAT per litre of bacterial culture was achieved (Abuhammad *et al.*, 2011). The identity of the

TBNAT was confirmed by ESI-MS and was in agreement with the expected mass of the protein. Whilst the yield of the His-TBNAT of 3 mg per litre of bacterial culture was lower than that of TBNAT, the quantity and the quality of both forms of the enzyme are reliable enough and sufficient for pursuing protein crystallisation trials with or without the His-tag. These values are an improvement in the previously reported yield (Table 4.1). This high yield of pure protein will also allow for detailed studies on the available inhibitors to be undertaken in parallel with the structural studies.

5 Investigation of NAT as an antitubercular drug target

5.1 Introduction

Arylamine N-acetyltransferase from *M. tuberculosis* (TBNAT) has been proposed as an antitubercular drug target (Bhakta *et al.*, 2004; Westwood *et al.*, 2010). The enzyme utilises propionyl-CoA, a product of cholesterol catabolism and a building block for cell-wall lipids. Cholesterol is likely to be the fuel used by *M. tuberculosis* inside macrophages (Kendall *et al.*, 2007; Van der Geize *et al.*, 2007; Pandey and Sassetti, 2008) (Chapter 1, section 1.3.1).

It has been proven that the NAT enzyme is vital for intracellular survival of mycobacteria and mycolic acid synthesis (Bhakta *et al.*, 2004). NAT inhibitors were demonstrated to have effects similar to that of deleting the gene, which suggests that the enzyme is a valid target for anti-tubercular therapy (Westwood *et al.*, 2010). However, previous efforts to identify inhibitors utilised NATs from a range of organisms (*M. smegmatis*, *Salmonella typhimurium, Pseudomonas aeruginosa* and *M. marinum*) (Westwood *et al.*, 2010). These studies did not include TBNAT. Therefore, the specific **aims** of the work described in this chapter (Introduction, Section 1.6) were as follows:

- **To evaluate available NAT inhibitors against TBNAT (Figure 1.16)**
- **To screen for novel TBNAT inhibitors to open up new chemical space**
- **To explore the mechanism of inhibition by the piperidenol class of inhibitors for MMNAT and TBNAT**
- **To perform structural studies on NAT-inhibitor complexes**

A flowchart summarising the work described in this chapter is shown in Figure 5.1.

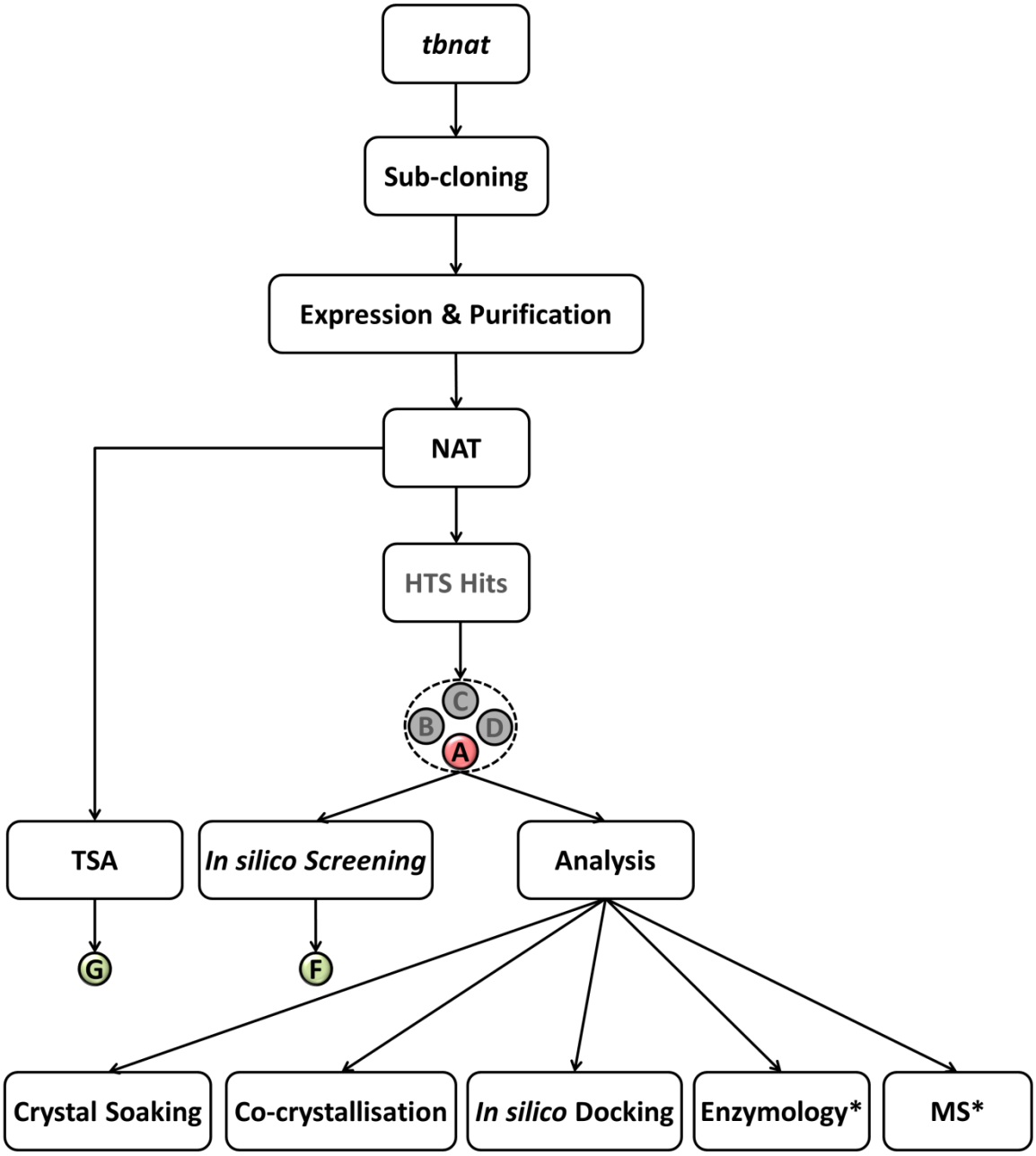


Figure 5.1: A Flowchart summarising the work described in this chapter.

Most of the work was carried out using MMNAT, but the asterisks represent experiments carried out on both MMNAT and TBNAT. The letters on coloured backgrounds represent the classes of NAT ligands/inhibitors involved in this study. The hits identified in this work are shown on a green background. High-throughput screening hits (HTS) were identified by (Westwood *et al.*, 2010). TSA: thermal shift assay. MS: mass spectroscopy. The methods used in the experiments reported have already been described in Chapter 2.

5.2 Results and Discussion

Determination of the 3D-structure of the MMNAT-HLZ complex (Chapter 3), allowed a better understanding of the observed substrate activity profile and the binding pocket specifications. Both MMNAT and TBNAT have higher affinity for HLZ than for INH or 5-aminosalicylic acid (5ASA) as the acetyl acceptors, which have been used routinely for the assessment of activity using hydrolysis of Ac-CoA to follow the reaction (Brooke *et al.*, 2003b; Westwood *et al.*, 2010; Fullam *et al.*, 2011). Since an essential part of this investigation was the measurement of the catalytic activity of NATs with various inhibitors, it was decided for the purpose of this work to modify the assay used for these measurements to have HLZ as a substrate. Details of the assay are described in Section 2.2.6.2. For simplicity, this assay will be referred to as the NAT-inhibition assay throughout the following text.

5.2.1 Evaluation of TBNAT inhibition by the available NAT inhibitors

Previous efforts to identify NAT inhibitors by high-throughput screening led to the discovery of novel classes of compounds that can inhibit the enzymes from *P. aeruginosa* and *M. marinum* (Westwood *et al.*, 2010). Evaluation of the ability of these compounds (A-1, B-1, C-1 and D-1) to inhibit TBNAT was carried out using the NAT-inhibition assay described in Section 2.2.6.2 and the results are shown in Table 5.1. The compounds were also re-evaluated for their activity against the MMNAT enzyme using the HLZ based assay. Compound A-1 among the tested hits showed the best inhibition of both enzymes with IC_{50} values of 7.7 μM and 1.3 μM against TBNAT and MMNAT, respectively (Table 5.1). Inhibitors B-1 and D-1, which showed moderate activity against MMNAT with IC_{50} values of 33 μM and 19 μM, respectively, were ineffective in inhibiting TBNAT (Table 5.1).

Since TBNAT is the focus of this work, inhibition by A-1 was studied in more detail, and these results are described in the rest of this chapter. The compound has been reported previously for its antimycobacterial activity with MIC against *M. tuberculosis* below 5 μg/mL (Jeney and Zsolnai, 1956). It has not been studied until recently, when Westwood *et al*. reproduced comparable antimycobacterial activity of A-1, in accordance with other molecules of this class, in screening for NAT inhibitors (Westwood *et al.*, 2010). This was followed by the synthesis of a group of analogues through the cyclisation of an intermediate bis-Mannich base as well as the testing of purchased analogues in an effort to improve both NAT inhibition and antimycobacterial activity (Fullam, 2007).

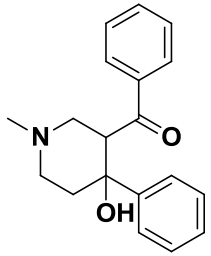
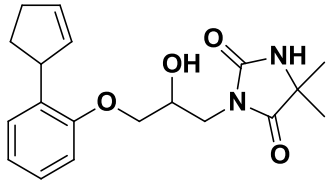
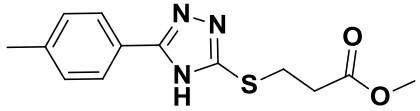
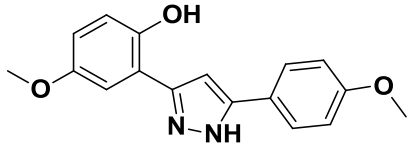
In the current study, ten analogues of A-1 (Fullam, 2007) were also tested for their inhibition of TBNAT and were re-evaluated against MMNAT with HLZ as a substrate as shown in Table 5.2. Because of the presence of a hydroxyl piperidine nucleus in these compounds, they will be referred to as the piperidinol inhibitors (Figure 5.2). The bis-Mannich base A-12 which is an intermediate in the synthesis of A-1 was also tested for

NAT inhibition activity (Table 5.2). A detailed study of the mechanism of inhibition by this class of compounds is described later in this chapter. The low affinity of compound A-1 for the eukaryotic NAT enzymes is an additional encouraging factor in pursuing it for drug design (Figure 5.2). The figure shows that compound A-1 meets the selection criteria used for hit selection in the original high-throughput screen (Westwood *et al.*, 2010): only compounds which showed over 80 % inhibition of at least two prokaryotic enzymes and showed less than 20 % inhibition of the eukaryotic enzymes at 30 μM were considered (Westwood *et al.*, 2010).

Interestingly, testing MMNAT inhibitors in the presence of HLZ as an acetyl acceptor resulted in more potent IC_{50} values than those obtained previously using 5ASA as acetyl acceptor (Table 5.1). This phenomenon has been identified frequently for other enzymes (Wetherall *et al.*, 2003). During the catalytic reaction, inhibitor and substrate(s) may compete for the binding of various enzyme species. The change of substrate perturbs the steady-state equilibrium of the system and hence affects the apparent potency of the inhibitor. This change in the apparent potency directly reflects the change in the concentration of the inhibited enzyme species in the presence of different substrates (Copeland, 2005). The choice of substrate is one of the key factors for the success of enzymic assays aimed to identify inhibitors (Copeland, 2005).

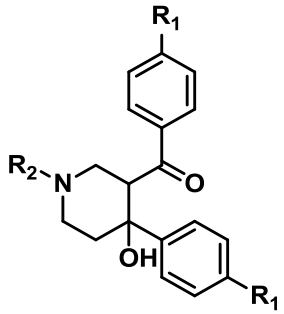
Although the data with piperidinols were encouraging, the structures of these compounds contain multiple chiral centers, so in parallel to further studies on the piperidinols, additional routes were pursued to identify other chemical scaffolds as NAT inhibitors.

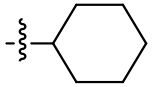
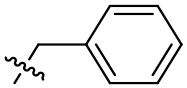
Table 5.1: Structural classes of NAT inhibitors identified through a previous HTS.[a]

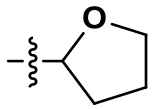
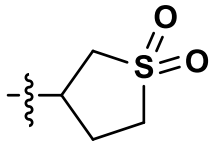
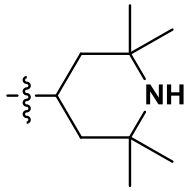
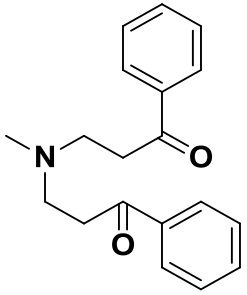
Code	Structure	TBNAT		MMNAT		
		% inhibition	IC_{50} (µM) HLZ	% inhibition	IC_{50} (µM) HLZ	IC_{50} (µM)[c] 5ASA
A-1		101 ± 1	7.7 ± 0.9	105 ± 1	1.3 ± 0.0	6 ± 1
B-1		-2 ± 1	ND	98 ± 0	1.5 ± 0.1[b]	33 ± 1
C-1		89 ± 3	21 ± 2	99 ± 1	4.9 ± 0.1	11 ± 1
D-1		27 ± 2	ND	100 ± 3	6.5 ± 0.4	19 ± 1

[a] The NAT activity was measured by the NAT-inhibition assay using 150 µM of HLZ and 120 µM acetyl-CoA as described in Section 2.2.6.2. The percentage of enzyme inhibition was measured in the presence of 50 µM inhibitor and compared to the un-inhibited control. The IC_{50}s were determined by measuring the enzyme activity in the presence of variable concentrations of each inhibitor (0-250 µM) and compared to the un-inhibited control. The results are presented as the mean ± S.D. from triplicate measurements at 24 ºC. ND is not determined. [b] From Fullam *et al.* (2011). [c] From Westwood *et al.* (2010).

Table 5.2: The inhibitory activity of compound A-1 and its analogues.[a]



Code	R_1	R_2	TBNAT		MMNAT			MIC (µg/mL)[b]	
			% Inhibition	IC_{50} (µM) HLZ	% Inhibition	IC_{50} (µM) HLZ	IC_{50} (µM)[b] 5ASA	*M. bovis* BCG	*M. tuberculosis*
A-1	H	$-CH_3$	101 ± 1	7.7 ± 0.9	105 ± 1	1.3 ± 0.0	6.0 ±1	6.3-12.5	1-10
A-2	Cl	$-CH_3$	98 ± 1	1.6 ± 0.1	103 ± 2	0.16 ± 0.01	1.4 ± 0.6	6.3-12.5	ND
A-3	Br	$-CH_3$	98 ± 3	2.9 ± 0.4	99 ± 1	0.08 ± 0.01	1.3 ± 0.4	6.3-12.5	5-10
A-4	H	$-CH_2CH_3$	72 ± 4	7.3 ± 0.3	126 ± 5	1.9 ± 0.0	8.0 ± 1	6.3-12.5	5-10
A-5	F	$-CH_2CH_3$	108 ± 1	4.4 ± 0.0	102 ± 3	0.5 ± 0.0	1.3 ± 0.5	6.3-12.5	5-10
A-6	H	$-(CH_2)_3CH_3$	100 ± 3	6.9 ± 0.4	102 ± 3	2.6 ± 1	5.0 ± 0.4	6.3-12.5	1-5
A-7	H		72 ± 60	4.4 ± 0.1	103 ± 1	ND	1.7 ± 0.2	6.3-12.5	1-5
A-8	H		58 ± 2	ND	101 ± 1	4.1 ± 0.4	ND	6.3-12.5	0-1

A-9	H		51 ± 3	ND	100.8 ± 0.5	2.5 ± 0.3	9.0 ± 0.9	ND	0-1
A-10	H		47 ± 2	ND	99 ± 0.7	13 ± 1	>30	3.1-6.3	ND
A-11	H		67 ± 4	1.1 ± 0.3	100 ± 2	2.7 ± 0.4	1.1± 0.3	6.3-12.5	1-5
A-12			101 ± 2	1.2 ± 0.1	101 ± 2	0.14 ± 0.02	3.0 ± 0.8	6.3-12.5	1-5

[a] The NAT activity was measured by the NAT-inhibition assay using 150 µM HLZ and 120 µM Ac-CoA as substrates as described in Section 2.2.6.2. The percentage of enzyme inhibition was measured in the presence of 50 µM inhibitor and compared to the un-inhibited control. The IC_{50} values were determined by measuring the enzyme activity in the presence of variable concentrations of each inhibitor (0-250 µM) and compared to the un-inhibited control. The results are presented as the mean ± S.D. of triplicate measurements. ND is not determined. Inhibition curves were obtained by non-linear fitting of the % inhibition and the inhibitor concentration (µM) using the Log(inhibitor) vs. response module of GraphPad Prism 5.0.
[b] From Fullam (2007).

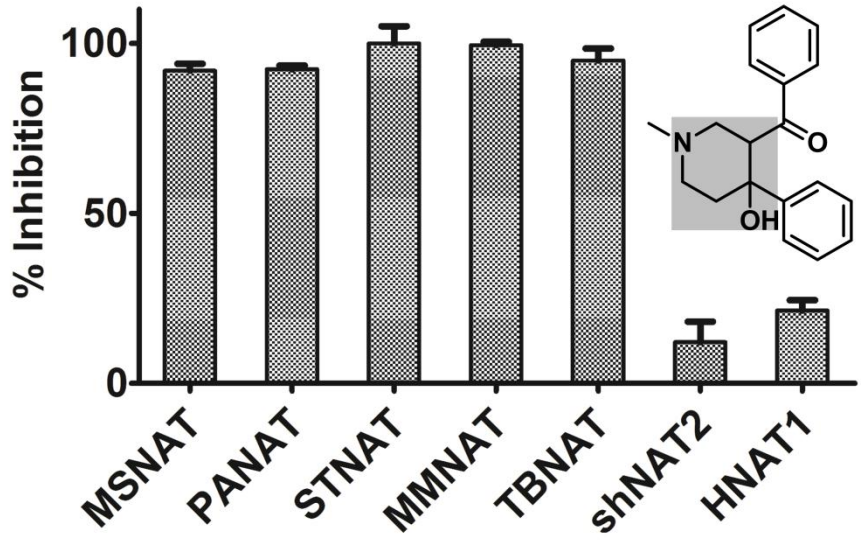


Figure 5.2: Specificity of A-1 for prokaryotic NAT enzymes.

Compound A-1 was tested at 30 µM against five pure recombinant NAT enzymes from *M. smegmatis* (MSNAT), *S. typhimurium* (STNAT), *P. aeruginosa* (PANAT) and also against two eukaryotic enzymes, hamster NAT2 (shNAT2) and human NAT1. The results are shown as the mean ± S.D. of triplicate determinations of percentage inhibition of hydrolysis of Ac-CoA in the presence of 5ASA as a substrate as described in Fullam (2007). In the current work, A-1 was tested at 30 µM against TBNAT using HLZ as a substrate. The inhibition is represented as a percentage compared to an uninhibited control. The results are presented as the mean ± S.D. from triplicate measurements. The structure of compound A-1 is shown and the piperidinol nucleus is highlighted by the shaded area. The figure was modified, from Fullam (2007).

5.2.2 Search for novel chemical scaffolds as NAT inhibitors

5.2.2.1 *In silico* screening for shape analogues of compound A-1

Screening for novel NAT inhibitors based on the 3D-shape of compound A-1 (Table 5.2) was performed using the ElectroShape approach (InhibOx Ltd, Oxford). ElectroShape is a ligand-based ultrafast virtual screening method that ranks hits based on 3D-shape, charge, and physicochemical properties (Armstrong *et al.*, 2010). The 3D-shape similarity is one of the recently developed techniques used in computer-aided drug discovery based on the knowledge of at least one active compound (i.e. ligand-based). The method relies on the observation that molecules which bind to the same biological target tend to have a similar shape (Figure 5.3) (Nicholls *et al.*, 2010).

The 3D-structure of A-1 (3R,4S-3-benzoyl-4-phenyl-1-methylpiperidinol) was subjected to several rounds of energy minimisation after which it was used as a search query against a virtual database of 7.3 million commercially-available molecules (InhibOx Ltd, Oxford). The identified hits were ranked based on the ElectroShape score and 12 compounds out of the top 100 hits were purchased and tested for their NAT inhibition activity. The chemical structures of the tested molecules and their inhibition activities are shown in Table 5.3.

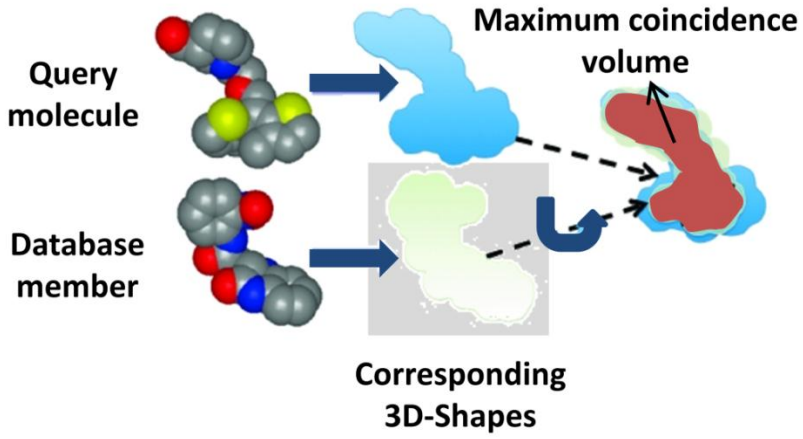
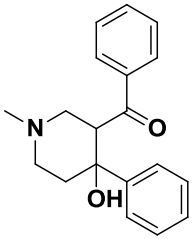
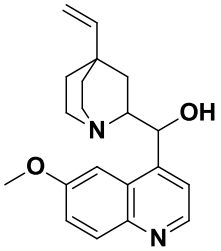
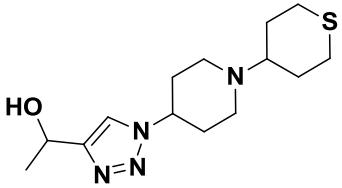
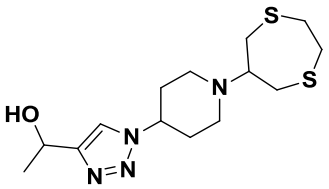
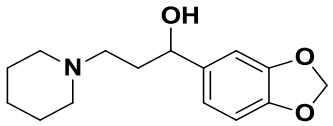
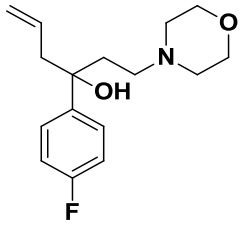
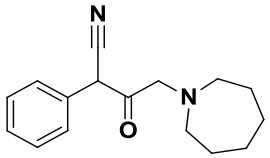
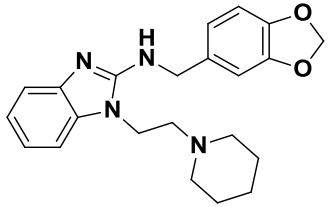
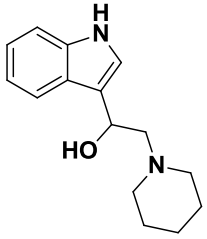


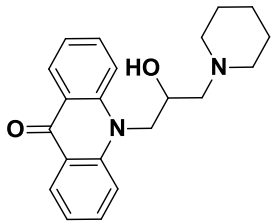
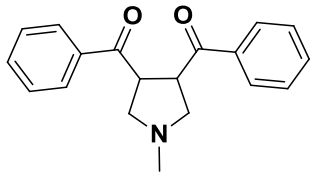
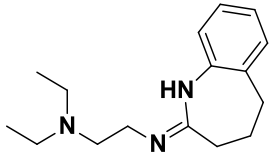
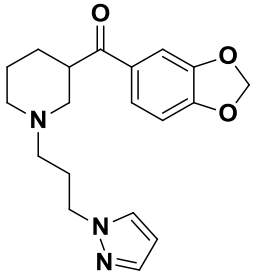
Figure 5.3: Illustration of the fundamental definition of shape similarity.

The shape similarity is derived from the alignment that achieves an optimal overlap (maximum coincidence volume) between a query molecule (represented by the blue shadow) and each molecule in the database (grey background). The figure and legend were modified from Nicholls *et al.* (2010).

Table 5.3: The chemical scaffolds of the tested 3D-shape hits with their *in silico* scores and their experimental activities against TBNAT and MMNAT.[a]

Code	Structure	ElectroShape score	*In silico* hit rank	TBNAT		MMNAT	
				% Inhibition	IC_{50} (µM)	% Inhibition	IC_{50} (µM)
A-1		0.995	1	101 ± 2	7.7 ± 0.9	105 ± 1	1.3 ± 0.0
F-1		0.949	8	16 ± 5	ND	35 ± 3	ND
F-2		0.935	46	-17 ± 1	ND	23 ± 2	ND
F-3		0.931	68	-21 ± 3	ND	31 ± 5	ND

F-4		0.964	5	86 ± 2	20 ± 3	103 ± 3	1.0 ± 0.1
F-5		0.930	70	-14 ± 3	ND	99 ± 1	6.4 ± 0.6
F-6		0.937	76	-35 ± 1	ND	12 ± 2	ND
F-7		0.923	91	14 ± 6	ND	16 ± 2	ND
F-8		0.931	66	11 ± 6	ND	10 ± 1	ND

F-9		0.923	90	9 ± 7	ND	21 ± 4	ND
F-10		0.928	76	-15 ± 4	ND	3 ± 3	ND
F-11		0.921	100	5 ± 1	ND	57 ± 1	ND
F-12		0.940	24	-13 ± 1	ND	46 ± 2	ND

[a] Compound A-1 was used as a query molecule. The rank of the hits, according to the ElectroShape scores, are shown alongside the corresponding scores. The NAT activity was measured by the NAT-inhibition assay using 150 µM HLZ and 120 µM Ac-CoA as substrates as described in Section 2.2.6.2. The percentage of enzyme inhibition was measured in the presence of 50 µM inhibitor and compared to the un-inhibited control. The IC_{50} values were determined by measuring the enzyme activity in the presence of variable concentrations of each inhibitor (0-250 µM) and compared to the un-inhibited control. The results are presented as the mean ± S.D. of triplicate measurements. ND is not determined. Inhibition curves obtained by non-linear fitting of the % inhibition and the inhibitor concentration (µM) using the Log(inhibitor) vs. response module of GraphPad Prism 5.0.

Compounds A-1, A-3 and the para-fluoro derivative of A-1, which were members of the screened virtual small-molecule library, ranked top at the hits list. Among the tested molecules, only compound F-4, which ranked fifth, showed moderate inhibition of TBNAT with an IC_{50} of 20 μM compared to 1 μM against MMNAT (Table 5.3 and Figure 5.5). The overlay of the 3D-shape of F-4 showed a high degree of overlap with the query molecule A-1, but the absence of the second aromatic ring is likely to explain its lower effectiveness as an inhibitor (Figure 5.4). Furthermore, the inhibition curves shown in Figure 5.5 suggest a different mechanism of inhibition of TBNAT and MMNAT by F-4. However, the mechanism of inhibition by F-4 was not investigated. The compound showed antimycobacterial activity against *M. bovis* BCG with a MIC of 33-66 μg/mL as assessed by the Alamar blue assay (Section 2.2.7.1, p 30). Compound F-5 is another compound of interest, although it had no activity against TBNAT, it inhibited MMNAT at an IC_{50} of 6.4 μM (Table 5.3). The compound showed high overlap with the 3D-shape of A-1 but again lacks the second phenyl group (Figure 5.4). However, this compound was not tested on *M. bovis* BCG.

The rest of the tested compounds were ineffective in inhibiting MMNAT or TBNAT despite their high Electroshape scores. This lack of inhibition despite the high shape similarity was mainly attributed to the partial overlap of the hit molecule that is either larger or smaller than the query molecules, as the case with compounds F-2 and F-8, respectively (Figure 5.4). It has been shown that the Electroshape scoring system work better as a ranking mechanism rather than as a similarity score calculator (Armstrong *et al.*, 2010).

Hits that showed enhancement of the acetyl-CoA hydrolysis in the assay (i.e. exhibited negative inhibition percentage) were evaluated for their ability to induce Ac-CoA hydrolysis in the absence of HLZ at either 100 μM or 500 μM level in the activity assay

(Section 2.2.6.1), however none showed any effect on Ac-CoA hydrolysis in the absence HLZ.

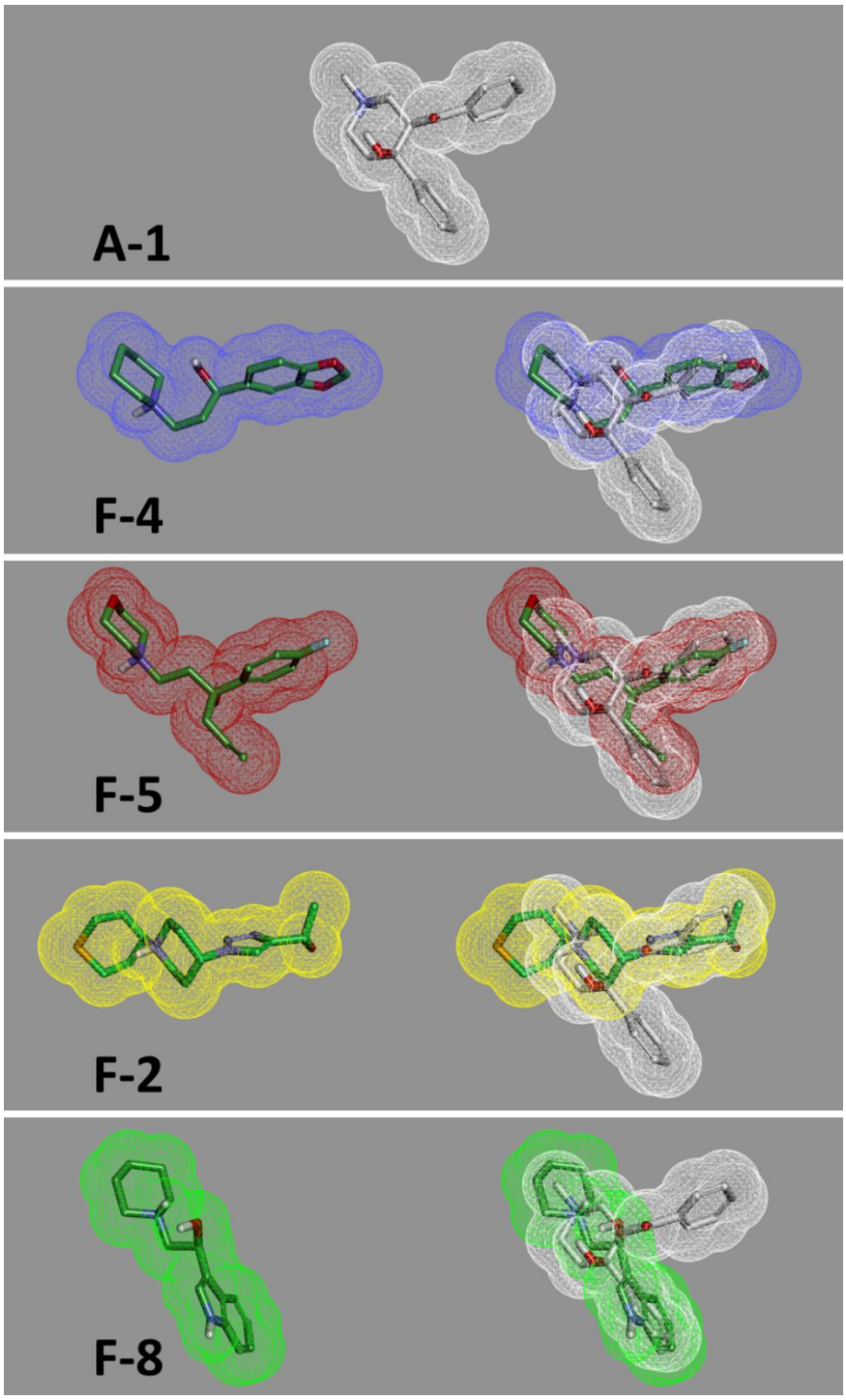


Figure 5.4: The 3D-shape of compound A-1 in comparison with the shapes of compounds F-2, F4, F5 and F8.

The mesh view of the 3D-shape in form of Van der Waals surface of the query A-1 conformer used in the *in silico* screen, the active hits F4 and F5, and examples of the inactive hits F2 and F8. An overlay of each hit with the query compound is shown. The figure was prepared by using the DS Visualizer 3.1.

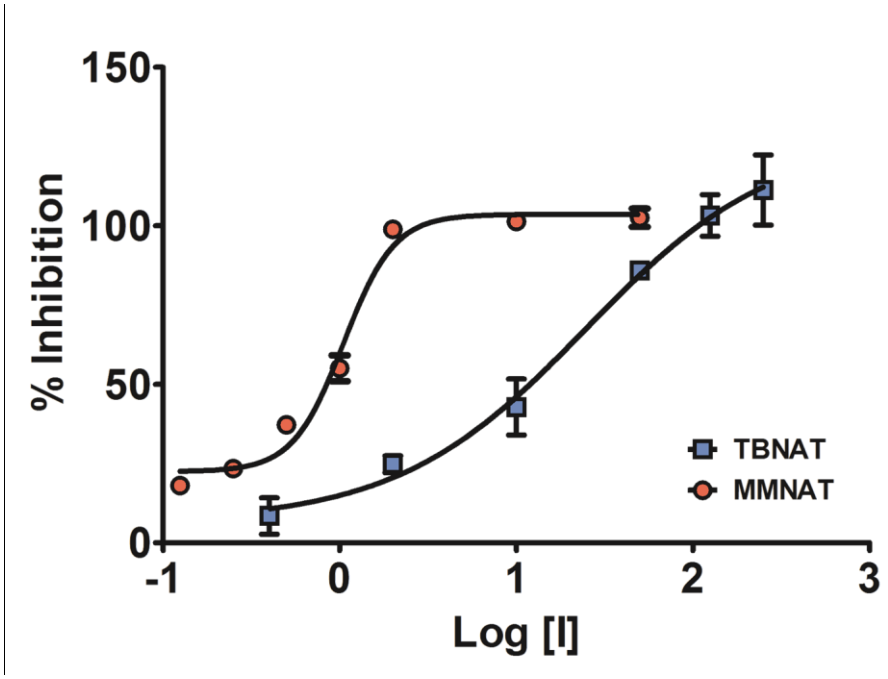


Figure 5.5: Inhibition curves of TBNAT and MMNAT by F-4.

The NAT activity was measured by the NAT-inhibition assay using 150 µM HLZ and 120 µM Ac-CoA as substrates (Section 2.2.6.2) in the presence of variable concentrations of F-4 (0-250 µM) and compared to the un-inhibited control. The results are presented as the mean ± S.D. of triplicate measurements. Inhibition curves were obtained by non-linear fitting of the % inhibition and the inhibitor concentration (µM) using the Log(inhibitor) vs. response module of GraphPad Prism 5.0. The error bars are within the symbols.

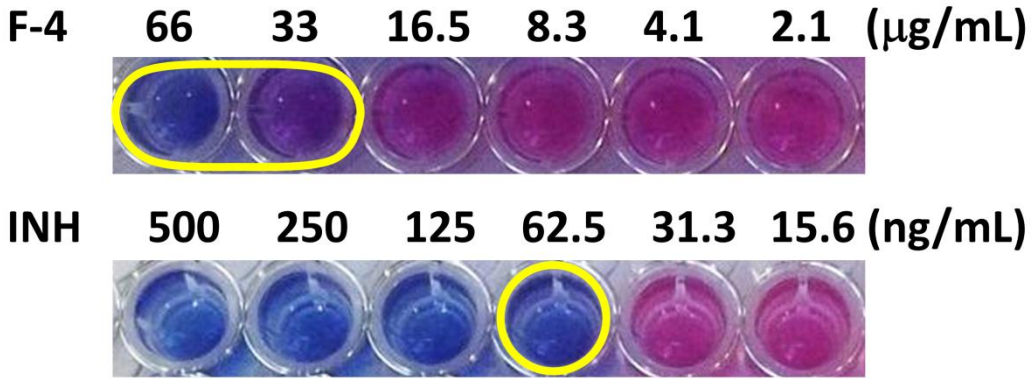


Figure 5.6: The Alamar blue assay of *M. bovis* BCG with F-4 and INH as a control.

The inhibition of *M. bovis* growth in liquid media was assessed from the colour change of the Alamar blue at 24 h of incubation as shown. A red colour indicates growth/resistance, and blue indicates no growth/sensitivity. The results are compared to those obtain with isoniazid (INH). The assay was performed as described in Section 2.2.7.1. MICs were determined visually. Note different scale.

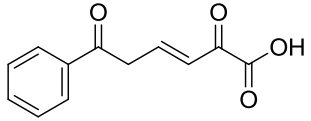
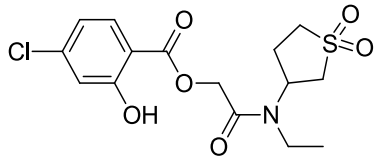
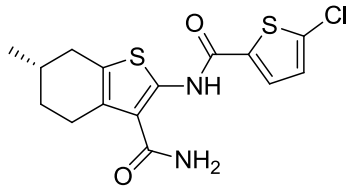
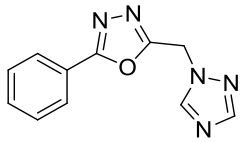
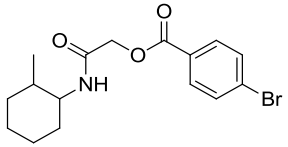
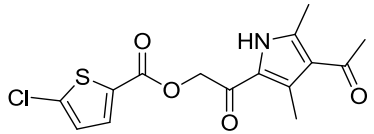
5.2.2.2 Identification of HOPDA as an inhibitor of NAT

Thermal shift assays have been used for identification of inhibitors of target proteins from compound libraries (Kemp *et al.*, 2012). Ligand binding to a target protein can stabilise the native state of the protein. This stabilisation is reflected in the increased melting temperature (T_m) and hence results in a positive shift of T_m (ΔT_m) in comparison to the protein alone (Section 6.2.3.3). The change in the protein stability is indicative of binding and proportional to the affinity of the ligand to the protein (Kemp *et al.*, 2012). The thermal shift screening strategy was used in a study aimed at screening for small-molecules that could induce TBNAT crystallisation (Section 6.2.3.3) and it highlighted possible ligands of the protein.

Compound G-1 was identified as a possible ligand of TBNAT and MMNAT with ΔT_m values of 2.4 ºC and 2.2 ºC, respectively (Table 5.4). This compound is the synthetic substrate of HsaD, an enzyme also encoded in the same cluster as TBNAT (Figure 1.10). The 3D-structure of the complex of HsaD with G-1 has been solved by crystallography (PDB code 2WUG, 1.8 Å) (Lack *et al.*, 2010) and the 3D-shape of the compound was used in an *in silico* screening experiment to identify inhibitors of HsaD (Lack and Sim, , unpublished).

Experimental evaluation of the inhibitory activity of G-1 showed that the compound weakly inhibited both MMNAT and TBNAT (Table 5.4). Further evaluation of the thirty 3D-shape analogues of G-1 was also performed. However, none of the tested compounds showed significant inhibition against TBNAT (Table 5.4). Both compounds G-3 and G-4 inhibited MMNAT to the same degree, but only compound G-3 showed moderate activity against TBNAT (Table 5.4).

Table 5.4: The experimental inhibitory activity of G-1 and related *in silico*-screening hits.

Code	Structure	TBNAT			MMNAT		
		% Inhibition	IC_{50} (µM)	ΔT_m (ºC)	% Inhibition	IC_{50} (µM)	ΔT_m (ºC)
G-1		37.7 ± 2.8	ND	2.4 ± 0.4	46.5 ± 3.2	ND	2.2 ± 0.1
G-2		42.3 ± 3.8	ND	1.7 ± 0.01	94.7 ± 2.3	> 30	1.6 ± 0.1
G-3		94.6 ± 1.0	21.7 ± 0.5	ND	99 ± 0.4	6.5 ±0.5	ND
G-4		14.3 ± 0.7	ND	ND	101.2 ± 2.3	7.7 ±0.8	ND
G-5		-8.7 ± 3.3	ND	-0.8 ± 1.0	98.4 ± 1.8	> 30	1.1 ± 0.1
G-6		-5.9 ± 5.0	ND	-0.9 ± 0.1	50.9 ± 5.9	ND	1.9 ± 0.8

[a] The NAT activity was measured by the NAT-inhibition assay using 150 µM HLZ and 120 µM Ac-CoA as substrates as described in Section 2.2.6.2. The percentage of enzyme inhibition was measured in the presence of 50 µM inhibitor and compared to the un-inhibited control. The IC_{50}s were determined by measuring the enzyme activity in the presence of variable concentrations of each inhibitor (0-250 µM) and compared to the un-inhibited control. The results are presented as the mean ± S.D. of triplicate measurements. ND is not determined. Inhibition curves were obtained by non-linear fitting of the % inhibition and the inhibitor concentration (µM) using the Log(inhibitor) vs. response module of GraphPad Prism 5.

5.2.3 In-depth analysis of NAT inhibition by piperidinol

Whilst the *in silico* routes identified potential NAT inhibitors with novel scaffolds, the inhibition profile of the piperidinol group was more promising. Therefore, a detailed understanding of NAT inhibition by the piperidinols was undertaken.

During the preliminary evaluation of the inhibitory activity of the piperidinol group, it was recognised that the compounds exhibited a time-dependent inhibition, which markedly increased with the time of incubation of NAT with the inhibitor prior to the assay. This type of inhibition is usually observed as a result of tight binding of the inhibitor to the enzyme, whether this binding involves the formation of covalent adduct with reactive residues in the protein or not. Moreover, extensive dialysis of the enzyme (against 2×10^4-fold volume excess of dialysis buffer) after incubation with A-1 was completely ineffective in restoring the activity of the enzyme, supporting the proposal of tight irreversible binding of A-1 to the NAT enzyme (Figure 5.7). These data showed the same pattern with both MMNAT and TBNAT. Therefore, mass spectroscopy (MS) on the MMNAT-inhibitor complexes was performed (Section 2.2.4.3) to confirm any covalent adduct formation. This was followed by further enzymatic and structural studies to confirm whether the inhibition involved the binding pocket in the protein. Experiments on TBNAT were conducted for comparison.

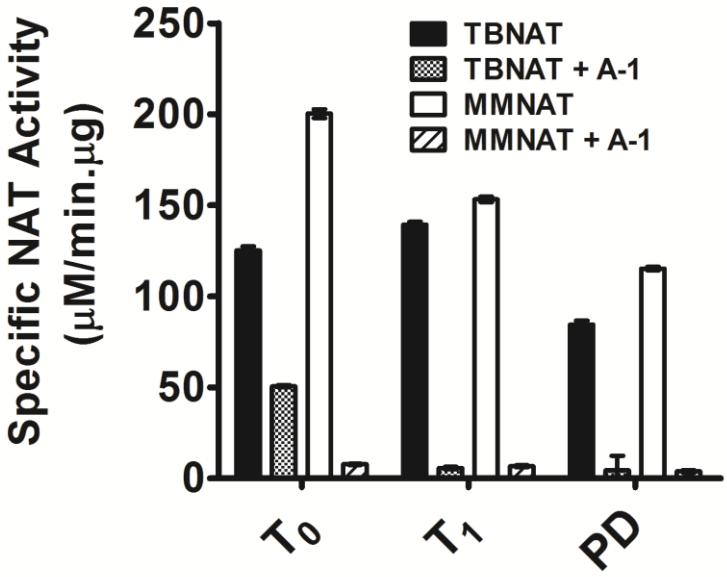


Figure 5.7: Reversibility of the inhibition of TBNAT and MMNAT by compound A-1

Each enzyme (0.07 mM, 50 µL) was preincubated either alone or with 15-fold molar excess A-1 at 24 ºC for 1 h. Each sample was then dialysed against 1 L fresh assay buffer (20 mM Tris-HCl pH 8 and 5 % DMSO) at 4 ºC for 16 h. The enzyme activities of the samples were measured at time 0 and 1 h (T_0 and T_1, respectively) before dialysis and then measured after dialysis (PD) by measuring the rate of Ac-CoA hydrolysis in the presence of HLZ as described in Section 2.2.6.1. The mean ± S.D. of three measurements of the activity is shown.

5.2.3.1 Covalent adduct formation as the mechanism of NAT inhibition

The ability of the piperidinols to form a covalent adduct with NATs was evaluated by mass spectroscopic analysis of the protein-ligand complexes, performed using compounds A-1, A-2, A-7, A-11 and A-12 to cover the aryl and *N*-substitution diversity among those inhibitors (Table 5.2). The enzyme was incubated with an equimolar amount of each inhibitor for 30-60 min in the presence of 5-10 % DMSO followed by ESI-MS analysis, and the results were compared to controls in which the inhibitors were excluded. The mass differences (Δm) between the enzyme-inhibitor complexes and the enzyme are shown in Table 5.5.

A single additional peak corresponds to a new species and an increase in the mass of each of the protein-inhibitor complexes was observed in all the spectra. These results confirm covalent adduct formation in all instances tested. Interestingly, all compounds with unsubstituted aryl groups (A-1, A-7, A-11 and A-12) resulted in species with a similar Δm of an average of 132 Da regardless of the *N*-substituents in these compounds (Table 5.5 and Figure 5.8). Although compound A-2 (the p-chloroaryl derivative of A-1) showed a peak with distinct mass difference of 164 Da, this mass is equivalent to an additional chloride atom added to the 132 Da-fragment. These findings supported the postulate that all inhibitors inactivate the enzyme by a general mechanism, which involves the formation of a reactive aryl compound that can react covalently with the enzyme. This is illustrated in Figure 5.9. The Δm expected from the addition of the 3-phenyl-3-oxopropyl moiety to the protein (C_9H_8O) is 132.07 Da, whilst the addition of 3-(4-chlorophenyl)-3-oxopropyl moiety (C_9H_7ClO) is 166.03 Da (Figure 5.9). These are in agreement with the mass difference observed upon incubating the enzyme with the piperidinols (Table 5.5).

However, incubation of MMNAT with 15-fold excess of the inhibitor for 60 min, followed by denaturation and ESI-MS resulted in the formation of multiple species (Table 5.5). Compounds, A-1 and A-2 formed three species of the modified protein, the mass of each of which matched the mono, di or tri derivatives of the 132 Da-fragment and the 166 Da-

fragment, respectively. Ratios of Δm observed for each enzyme-inhibitor complex to the mass of the corresponding 3-phenyl-3-oxopropyl fragments are shown in Table 5.5. Similarly, the treatment of TBNAT with compound A-1 resulted in two species the mass of each of which corresponds to the mass of mono and di derivatives of the 132 Da-fragment (Table 5.5). These observations indicate a reaction with a free sulfhydryl, since the number of cysteine residues in MMNAT is three, whereas in TBNAT it is only two. Whilst this does not prove that the reaction has occurred with free cysteine in MMNAT and TBNAT at the 15:1 molar ratio, it is indicative that this is the case.

Table 5.5: Mass spectrometric analysis results for MMNAT and TBNAT after incubation with the piperidinols inhibitors.

Sample	Deconvoluted mass (Da)	Δm (Da)	$\Delta m/m_{pvk}$	Molar ratio	Cys
Enzyme-to-inhibitor ratio of 1:1					
MMNAT-A-1	31047.6	132.6	0.9	1	3
MMNAT-A-2	31085.3	164.3	1	1	3
MMNAT-A-7	31046.5	131.5	1	1	3
MMNAT-A-11	31045.5	130.5	1	1	3
MMNAT-A-12	31050.6	129.7	1	1	3
Enzyme-to-inhibitor ratio of 1:15					
MMNAT-A-1	31185.9	255.9	1.9	2	3
	31062.6	132.6	1	1	3
	31324.1	394.1	3	3	3
MMNAT-A-2	31424.1	494.1	3	3	3
	31254.5	324.5	2	2	3
	31089.3	159.3	1	1	3
TBNAT-A-1	31450.9	134.1	1	1	2
	31583.4	266.6	2	2	2

For the ESI-MS studies, samples of each enzyme (0.05-0.07 mM in 20 mM Tris-HCl, pH 8.0 and 5-10 % DMSO) were incubated with the different inhibitors at an enzyme-to-inhibitor molar ratio of either 1:1 or 1:15 at 24 ºC for 30-60 min. The measurement was performed as described in Section 2.2.4.3 with confidence intervals of ± 8-12 Da. Samples of each enzyme alone (in the same buffer) were analysed in the same way as controls and used to calculate the mass difference upon incubation with the inhibitor (Δm values). m_{pvk} is the mass of the expected phenyl-oxopropyl fragments of 132 Da for A-1, A-7, A-11 and A-12, or 166 Da for A-2 and identifies the number of modifications. Molar ratio refers to the proposed number of inhibition species bound per protein molecule. Cys refers to the number of cysteine residues in the sequence.

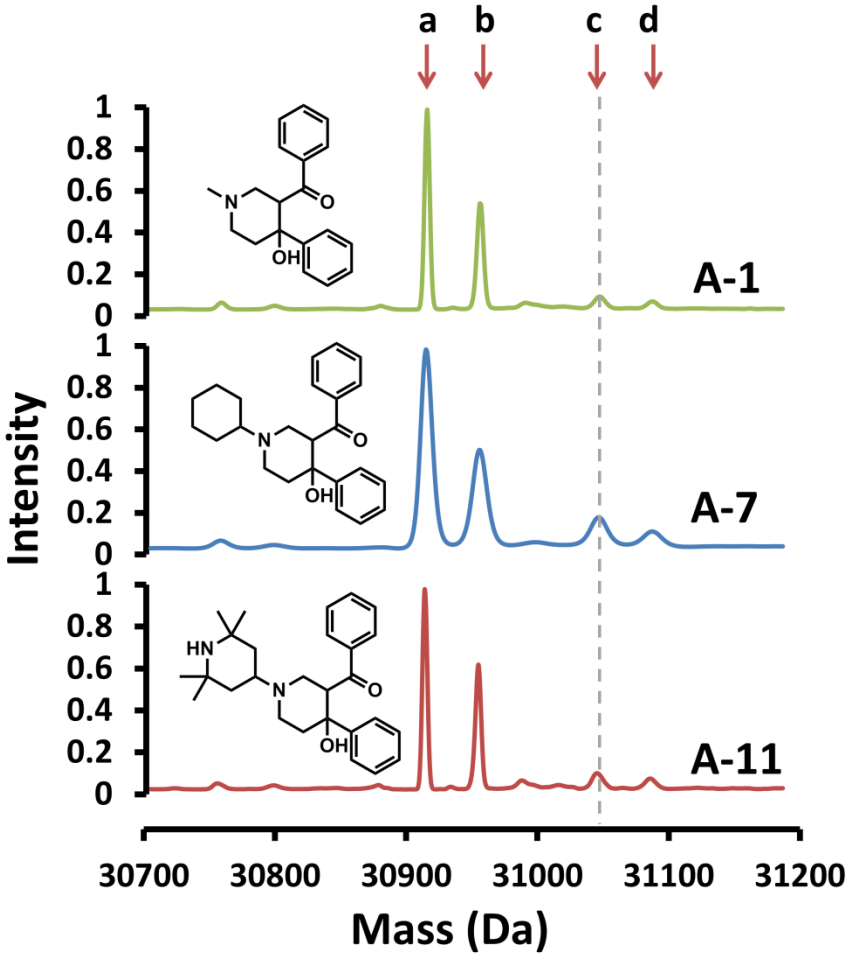


Figure 5.8: The ESI-MS of MMNAT in the presence of A-1, A-7 and A-11.

MMNAT was mixed with an equimolar sample (1:1 ratio) of each inhibitor (50 µM) in 20 mM Tris-HCl, pH 8, and the ESI-MS was performed after 30 min. The masses correspond to each peak according to compound A-7 are: a = 30915 (Δm = 0 Da), b = 30955.5 (Δm = 40 Da), c = 31046.5 (Δm = 131.5 Da), and d = 31087.5 Da (Δm = 172.2 Da). The mass corresponding to the addition of a 132 Da-fragment is marked with a dashed line. Δm of +40 Da is likely to correspond to potassium ion (38 Da).

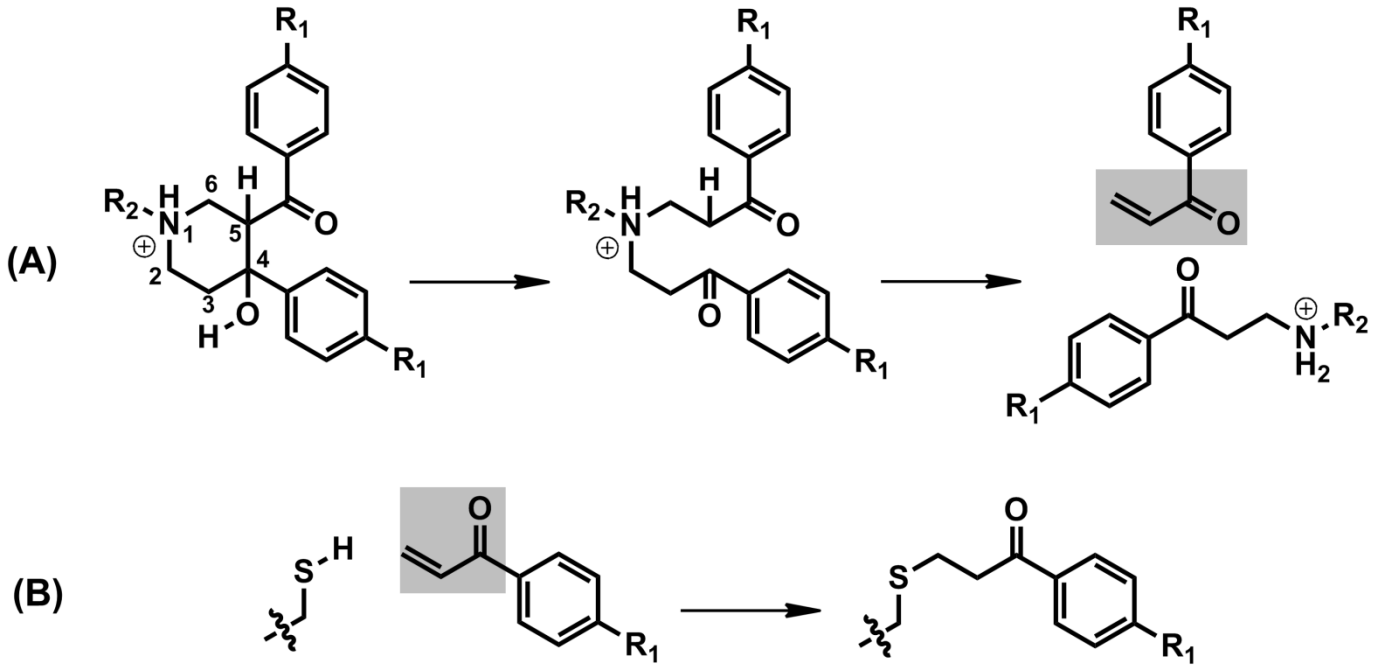


Figure 5.9: The chemical transformation of piperidinols to the corresponding phenyl vinyl ketone and the subsequent modification of a thiol containing residue by the phenyl vinyl ketone.

(A) The possible pathway of the formation of bis-Mannich bases from the rigid cyclic piperidinol. The bis-Mannich base can undergo a β-elimination of the amino group forming a reactive phenyl vinyl ketone. (B) The phenyl vinyl ketone reaction with thiols resulted in the addition of a 3-phenyl-3-oxopropyl moiety (R_1 is H) or a 3-(4-chlorophenyl)-3-oxopropyl moiety (R_1 is Cl). The expected Δm values of the added fragments is +132.07 Da and +166 Da, respectively. The shaded areas highlight the Michael acceptor moiety. R_2 corresponds to the substitution pattern shown in Table 5.2.

Whilst the data strongly suggest that all cysteine can react at a high inhibitor:protein ratio, at an equimolar inhibitor:protein ratio the question remains as to whether the active site cysteine is labelled. To confirm the reaction of A-1 with cysteine, the free amino acid was reacted with the compound under similar conditions used in the protein experiment for ESI-MS analysis in parallel with untreated samples of both cysteine and A-1. The samples were treated with 6-aminoquinolyl-n-hydroxysuccinimidyl carbamate prior to MS analysis. The amionquinolyl carbamate reagent will react with the free amino group of cysteine and ensures that only species with modified thiol groups were detected. Mass spectrometry was carried out on the mixture by analysing the different fractions collected from HPLC. The peak corresponding to cysteine (peak i, m/z = 291 Da) was absent from the reaction mixture chromatogram and the only unique peak was number viii (Figure 5.10), which had a molecular mass of 423.9 Da. The mass difference between the cysteine and the compound in the unique peak was 132 Da, which is in agreement with the Δm observed upon the reaction of A-1 with the protein.

The formation of such a species from this group of compounds has been reported previously in the alkylation of cellular glutathione in human T cells (Gul *et al.*, 2005). The proposed mechanism for the formation of reactive phenyl vinyl ketone (PVK) is through the hydrolysis to the corresponding bis-Mannich bases followed by a β-elimination reaction of the secondary amine as shown in Figure 5.9. This compound (PVK) acts as a Michael acceptor and thus is capable of forming 3-phenyl-3-oxopropyl derivatives upon reaction with nucleophiles.

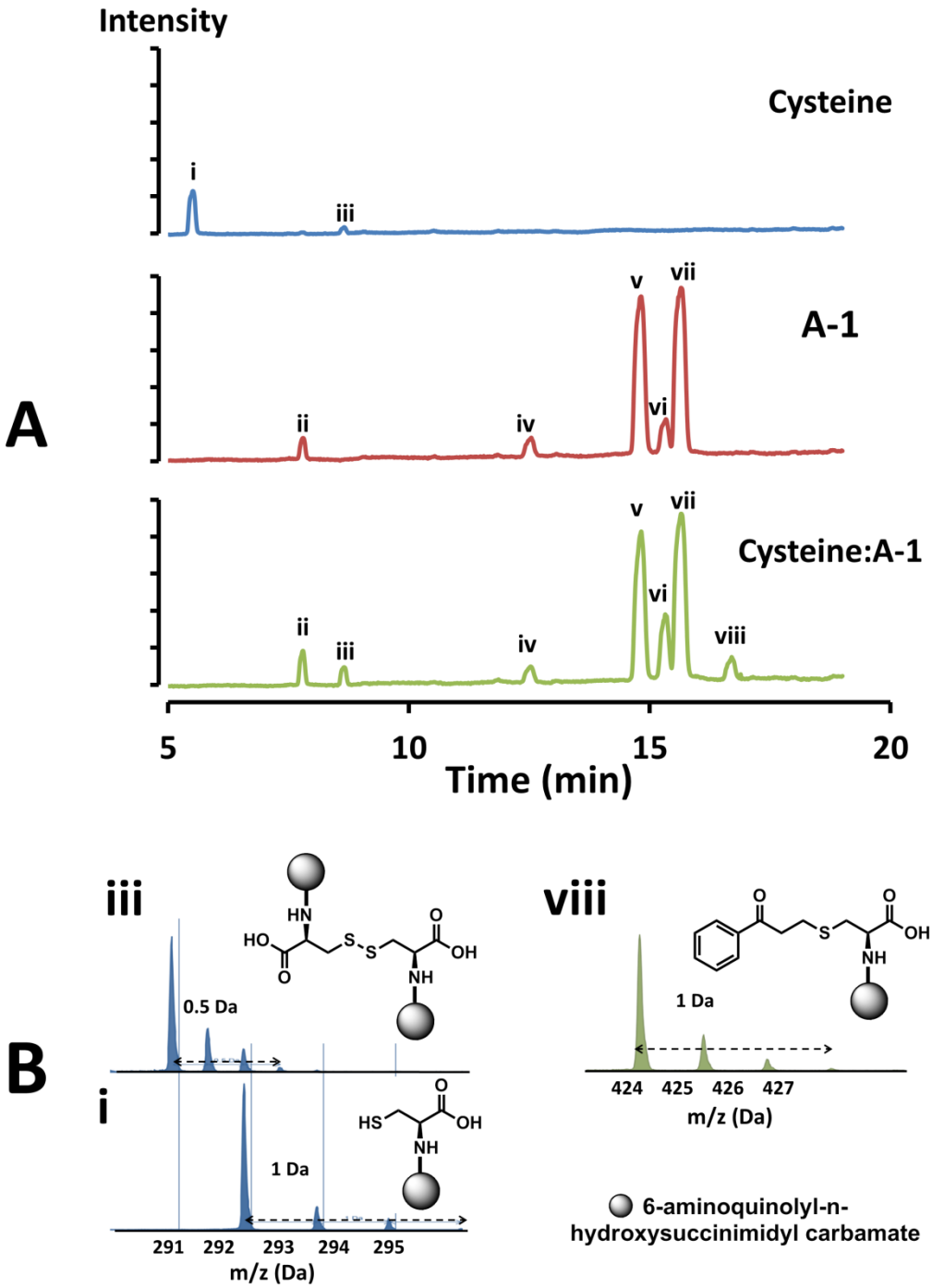


Figure 5.10: LC/MS analysis of the reaction of A-1 with free cysteine.

(A) The total ion current chromatogram (from liquid chromatography LC) of 100 µM cysteine, 100 µM compound A-1 and 100 µM cysteine:A-1 (1:1 mixture) in 20 mM MOPS buffer, pH 8 after 16 h incubation at 24 ºC. All samples were treated with 6-aminoquinolyl-n-hydtoxysuccinimidyl carbamate before analysis. (B) The ESI-MS spectra of fractions collected from the peaks in the chromatogram (in A) corresponding to i: cysteine (m/z = 291.9 Da), iii: cystine (m/z=290.9) and viii: the product of the reaction of cysteine with A-1 (m/z = 423.9 Da). The chemical structures of the compounds corresponding to each peak are shown. The round symbol represents the aminoquinolyl carbamate moiety.

5.2.3.2 Enzymatic evaluation of the piperidinols

In order to establish whether reaction was occurring with the active site cysteine, the enzyme characteristics of inhibition with piperidinols need to be described in detail. However, since the activity of piperidinols involves the formation of a reactive intermediate, the reaction rate is expected to be determined by three major factors: (a) the degree of molecular complementarity between the compound and the enzyme, (b) the ability of the compound to liberate the reactive PVKs, and (c) the ability of the PVKs to act as an alkylating agent.

The IC_{50} measurements (the concentration of a compound that is required to achieve 50 % inhibition in a biochemical assay) are of limited value for characterizing potency, selectivity and structure-activity relationships (SARs) for such inhibitors (Copeland, 2005). Therefore, the time-dependent inhibition of NAT by the piperidinols was evaluated according to the Kitz and Wilson model (Kitz and Wilson, 1962). According to this model, the covalent inactivation can be described in two steps of reversible binding followed by an irreversible pseudo-first-order reaction as follows:

$$E+I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [EI] \xrightarrow{k_{inact}} E\text{-}I$$

Based on this model, the irreversible inhibition progress with time can be estimated from Equation 5.1 (Kitz and Wilson, 1962).

$$ln\frac{[É]}{[E^o]} = -k_{obs} \times t \qquad \text{Equation 5.1}$$

[É]/[E°] can be estimated from the residual enzyme activity upon preincubation with a fixed concentration of the inhibitor at different time intervals (*t*). The detailed derivation of Equation 5.1 is described in Appendix C.

Irreversible inhibition, progressive with time in accordance with Equation 5.1, was measured for the piperidinol inhibitors by the protocol described in Figure 5.11 and the values of the apparent first order constant (k_{obs}) were determined (Table 5.6). An example of the curves obtained upon inhibition of MMNAT with variable fixed concentrations of A-1 is shown in Figure 5.12. The piperidinols exhibited similar inhibition against TBNAT when incubated at higher concentrations compared to those used with MMNAT. However, inhibition of TBNAT proceeded quickly within the first minute of incubation for most of the inhibitors, making accurate assessment of the early stage of inhibition difficult to achieve by the protocol described in Figure 5.11. An example of the time-dependent inhibition of TBNAT is shown in Figure 5.12.

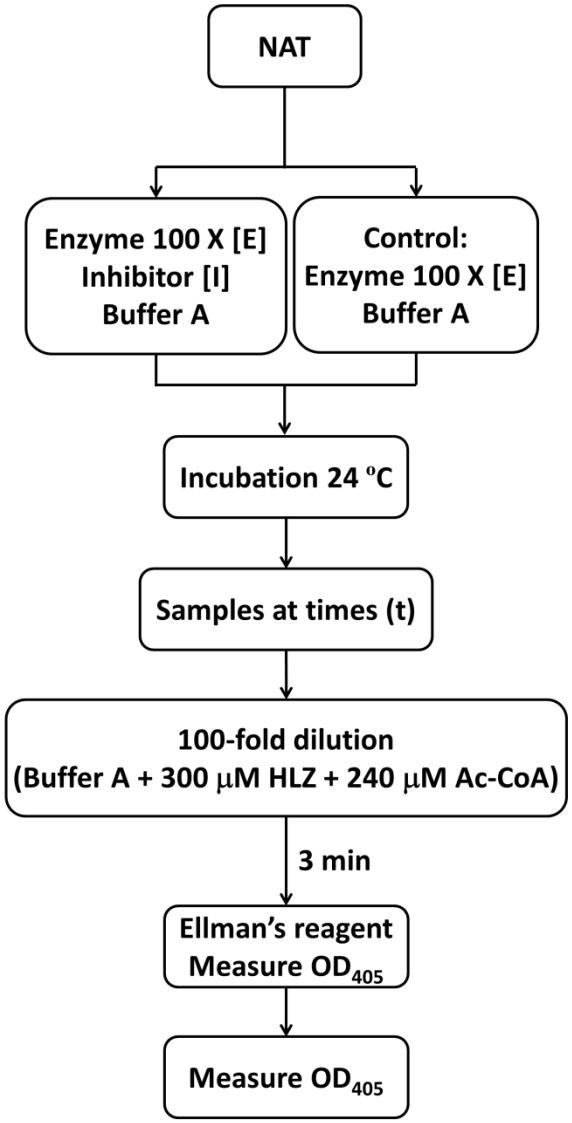


Figure 5.11: Flow chart outlining both the substrate-protection protocol and the time-dependent inhibition protocol.

Measurements were mainly performed using 0.06-0.07 mM MMNAT/TBNAT in buffer A. [E] represents the enzyme concentration required for initial linear kinetics. [I] is the reported inhibitor concentration. [S] represents 800 μM of either HLZ, Ac-CoA or the substrate derivative CoA. Buffer A consists of 20 mM Tris-HCl, pH 8 and 5 % DMSO.

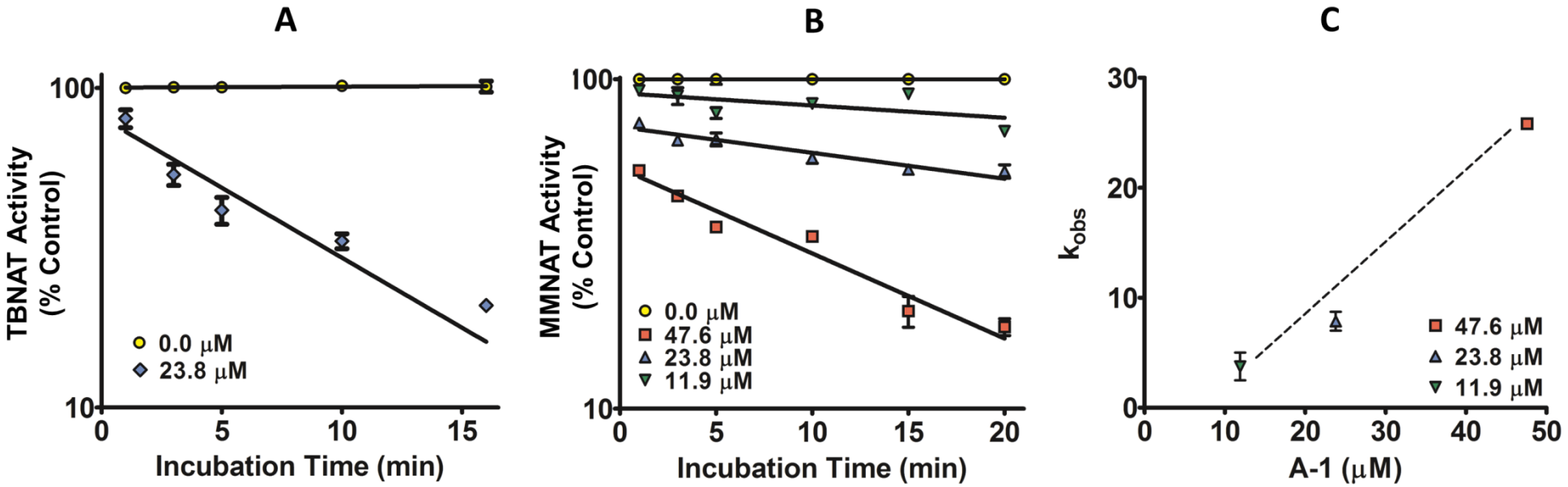
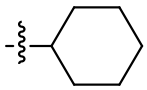
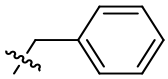
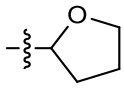
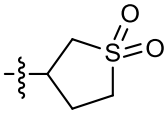
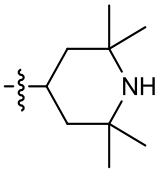


Figure 5.12: The time-dependent inhibition of the MMNAT and TBNAT by the piperidinols

Semi-logarithmic plots showing the time-dependent inactivation of (A) TBNAT by 23.8 μM A-7 and (B) MMNAT by various concentrations of A-1. The enzyme activity was measured using the protocol described in Figure 5.11. The results are presented as the mean ± S.D. of triplicate measurements. The residual activity is shown as a percentage of a control prepared as described in Figure 5.11. The data were fitted against the incubation-time using the Semilog line (X is linear, Y is Log) module of GraphPad Prism 5.0. The slope of each line is equivalent to ($-k_{obs}/2.303$) at each inhibitor concentration. The error bars are within the symbols. (B) A trend-line showing the increase of k_{obs} with the concentration of A-1.

Table 5.6: The time-dependent activity of the piperidinols against MMNAT.

Code	R_1	R_2	k_{obs} (10^{-3} min^{-1})	$t_{1/2}$ (min)	Critical volume (cm^3/mol)	LogP
A-1	H	$-CH_3$	9 ± 2	81.5	864.5	2.41
A-2	Cl	$-CH_3$	110 ± 2	6.3	962.5	3.53
A-3	Br	$-CH_3$	360 ± 26	1.9	988.5	4.07
A-4	H	$-CH_2CH_3$	15 ± 1	45.6	920.5	2.75
A-5	F	$-CH_2CH_3$	119 ± 2	5.8	956.5	3.07
A-6	H	$-(CH_2)_3CH_3$	104 ± 8	6.6	1032.5	3.66
A-7	H		573 ± 18	1.2	1077.5	3.96
A-8	H		10 ± 1	71.4	1092.5	4.15
A-9	H		163 ± 39	4.2	994.5	2.84
A-10	H		19 ± 1	37.1	1017.5	0.91
A-11	H		586 ± 115	1.2	1276.5	3.06
A-12	H	$-CH_3$	34 ± 2	20.4	913.5	2.67

The assay was performed as described in Figure 5.11 and k_{obs} values were obtained from the slope of the semilogarithmic plots of the residual activity vs incubation time at 11.9 µM except for A-3 and A-7 (5.9 µM) and A-10 (23.8 µM). The results are presented as the mean ± S.D. of triplicate measurements at 24 ºC. $t_{1/2}$ is the apparent inactivation half-life calculated from k_{obs} ($t_{1/2}$ = 0.693/k_{obs}). ND is not determined. The critical volumes (the volume of one mole of material at the critical temperature and pressure) and LogP values (the octanol-water partition constant) as predicted by ChemBioDraw Ultra 12.0 www.cambridgesoft.com are shown.

5.2.3.3 Active-site directed inhibition of MMNAT

Evidence for the formation of a covalent adduct with cysteine in the active site at the ratio of 1:1 was also obtained from MMNAT and TBNAT ESI-MS analyses. It was required to address the question of whether inhibition by the piperidinols involves modification of the cysteine residue in the binding pocket or in a different site that would affect activity by inducing change in the protein conformation. Therefore, the tests of the protective ability of the substrates (HLZ and Ac-CoA) and CoA (a reversible inhibitor) were undertaken in order to ascertain if the inhibition was directed towards the active site cysteine of the enzyme (Cys70). The rate of inactivation upon the incubation of the NAT-A-1 mixture with and without the presence of excess of either HLZ, Ac-CoA or CoA was obtained using the protocol described in Figure 5.11. If inactivation by the piperidinols was a consequence of alkylation of the active site cysteine, then the substrates Ac-CoA and HLZ or a competitive inhibitor such as CoA, would be expected to exhibit time-dependent protection from inactivation.

Interestingly, 80 % of the enzyme activity was recovered by extensive dilution when either Ac-CoA or CoA was included in the preincubation of the enzyme-inhibitor mixture (Figure 5.13). On the contrary, HLZ was completely ineffective in protecting the enzyme, despite the ability of HLZ to bind the enzyme in advance of cysteine acetylation (Abuhammad *et al.*, 2010) (Figure 5.13).

Protection of the enzyme by Ac-CoA is compatible with the Ping-Pong mechanism of the enzyme activity, in which the active-site cysteine is acetylated initially by Ac-CoA. The affinity of the enzyme for the CoA fragment could also block the access of the inhibitor to the active site residue. CoA has been reported to exhibit high affinity to the NATs: recently endogenous CoA was found co-crystallised with the NAT from *Bacillus anthracis* (PDB code 3LNB) (Pluvinage *et al.*, 2011). However, considering the protection by CoA, which was either deliberately added to the mixture or which was formed upon the

acetylation of the enzyme, the possibility that CoA acted as a trap molecule that consumed the PVK before it reacts with the active site Cys70 cannot be excluded.

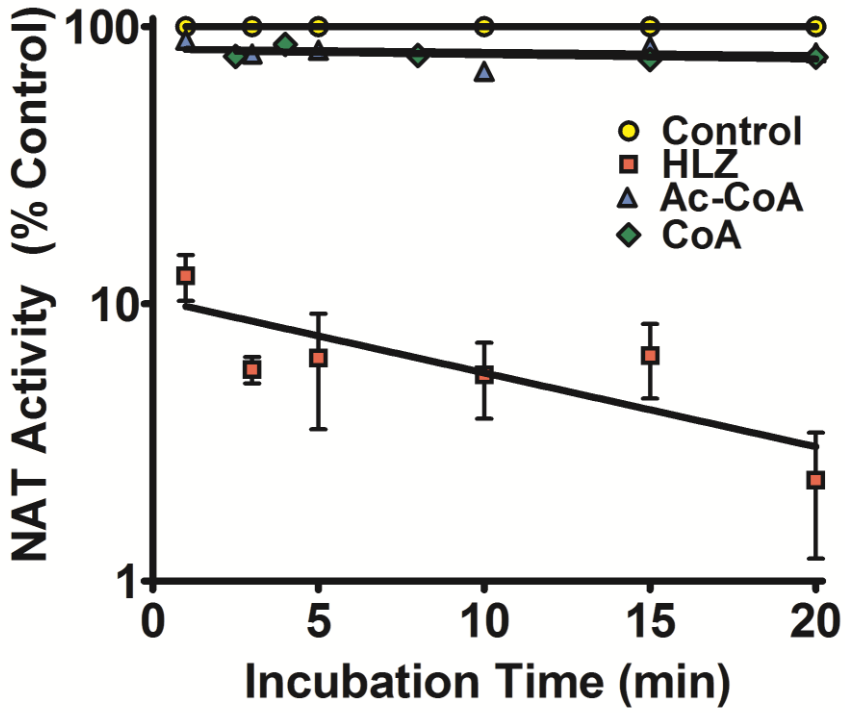


Figure 5.13: MMNAT protection by HLZ, Ac-CoA and CoA.

The substrate protection assay was performed as described in Figure 5.11 in the presence of 23.8 µM compound A-1. The mean ± S.D. of three measurements is presented. The activity is shown as a percentage of a control (Figure 5.11). The round symbols represent an overlay of all controls used (100 % activity). The error bars are within the symbols. The data were fitted against the incubation-time using the Semilog line (X is linear, Y is Log) module of GraphPad Prism 5.0.

5.2.3.4 Structure activity relationship (SAR) of piperidinols

Some degree of structural complementarity of the piperidinols with the active site can be inferred from the selectivity profile against a series of NAT enzymes, specifically to MMNAT (Figure 5.2). To test the effect of the 3D-complementarity with MMNAT on the activity of compound A-1, a dehydration derivative of A-1 was synthesised (compound A-13, Section 2.2.11.1) and tested for its inhibitory activity. As compound A-13 lacks the hydroxyl group, which was essential for the cleavage and proposed covalent binding of compound A-1 (Figure 5.9), it would be expected to stay intact in the binding pocket. Compound A-13 had a similar 3D-shape to A-1 and resulted in a 30 % inhibition of the MMNAT obtained with 50 µM inhibitor concentration (Figure 5.14). However, when the enzyme was incubated with compound A-13 for 20 min and subsequently diluted, the activity was restored, unlike the situation with A-1 (Figure 5.15).

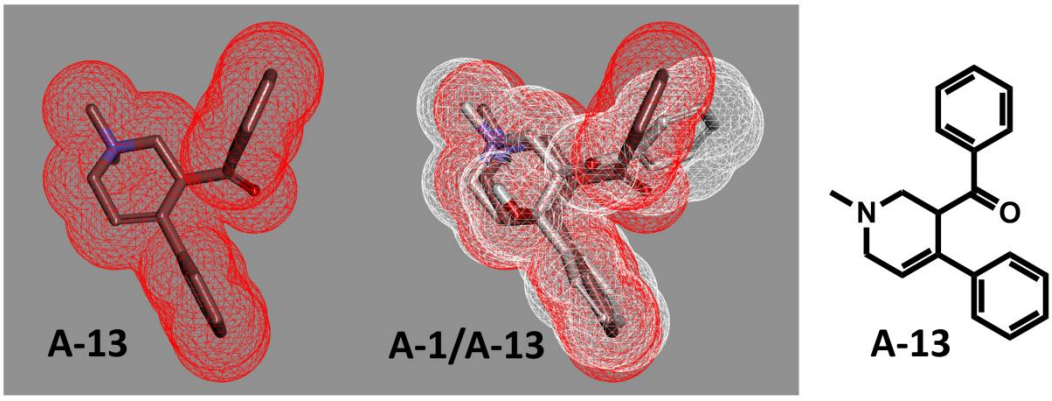


Figure 5.14: A comparison of the 3D-shape of A-1 and A-13 and their inhibition activity.

The 3D-shape of A-13 is shown in a mesh view of the Van der Waals surface. Overlapping 3D-shapes of A-1 (in white) and A-13 are also shown. Energy minimisation of A-13 was performed using the GlycoBioChem PRODRG2 online server (http://davapc1.bioch.dundee.ac.uk/prodrg/). The structure of A-13 is shown.

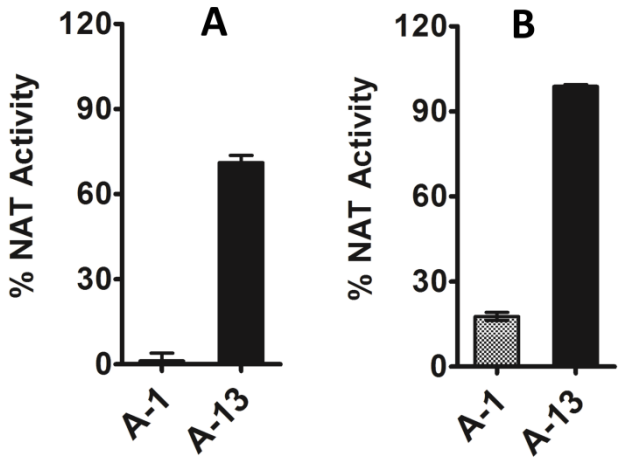


Figure 5.15: The inhibitory activity of compound A-13.

(A) The activity of MMNAT in the presence of 50 µM compound A-1 or A-13. (B) The activity of MMNAT after incubation with 50 µM of each inhibitor for 20 min followed by 200-fold dilution. The NAT activity was measured by the NAT-inhibition assay using 150 µM of HLZ and 120 µM Ac-CoA (Section 2.2.6.2). The percentage of enzyme activity was measured in the presence of 50 µM inhibitor and compared to the un-inhibited control. The results are presented as the mean ± S.D. from triplicate measurements at 24 ºC.

To the extent that can be judged from the results obtained from this limited group of compounds, it was clearly observed that the pattern of substitutions on both the benzene ring (R_1, Table 5.6) and the piperidine ring (R_2, Table 5.6) have a significant effect on the alkylation ability of the compounds. To facilitate comparison, the values of the apparent inactivation half-life of the enzyme were calculated from k_{obs} as shown in Table 5.6. Potent inhibitors would be predicted to exhibit shorter inactivation half-life. Unlike the rest of the inhibitors tested, it is to be expected that A-2, A-3 and A-4 would form the corresponding 4-halo-PVK intermediate in which the halide would be directly conjugated with the Michael acceptor. Interestingly, when the phenyl ring was *para* substituted with a chloride, the resulting piperidinol A-2 displayed a 13-fold increase in the inactivation potency as evidenced by the shorter inactivation half time (Table 5.6). This inhibitory activity was more pronounced ($t_{1/2}$ = 1.9 min at 5.9 μM concentration) when the phenyl ring was *para*-substituted with bromine. The fluoride derivative (A-5) showed an inactivation time comparable to that obtained by the chloride substituent, 5.8 min and 6.3 min, respectively. However, the ethyl substituent on A-5 is likely to have contributed to this inactivation potency since the ethyl piperidinol A-4 showed twice the activity of the methyl derivative A-1. Consequently, it appeared that the efficiency of the halide substituent at position 4 of the phenyl group in inhibiting MMNAT correlates with the atomic volume (AV) of the substituted group, i.e. bromine (AV = 23.5 cm^3mol^{-1}) > chlorine (AV = 22.7 cm^3mol^{-1}) > fluorine (AV = 17.1 cm^3mol^{-1}). Having a correlation in the reverse order of halides electron-withdrawing propterities (F > Cl > Br) indicated that the enhancement of inhibitory activity was not solely related to their effect on the reactivity of the Michael acceptor. The correlation to atomic volumes indicates that these halides also affected the complementarity of the PVK to the binding pocket.

Substituents at the *N*-position of the piperidine ring also affected the ability of these compounds to inactivate NAT (Table 5.6). Extension of the piperidine *N*-methyl in A-1 to

an ethyl group (A-4) doubled the inhibition, while further elongation to a butyl in A-6 resulted in a 12-fold increase in the inhibition of MMNAT. In contrast, aromatic substitution at the *N*-methyl, in the form of the *N*-benzyl substituted piperidinol A-8, did not confer an increase in inactivation potency. The benzyl substituent is more conformationally restrained due to the planarity of the aromatic ring, which could explain why no improvement in the inhibitory activity over A-1 was observed despite an increase in lipophilicity.

An interesting effect occurred when the *N*-substituent was replaced by a cyclohexyl ring in A-7, which exhibited the highest inhibition potency ($t_{1/2}$ = 1.2 min at 5.9 μM) comparable to that of the bromide derivative A-3. However, other cyclic substituents at this position resulted in variable activities that correlated positively with the lipophilicity of these rings. The sulpholane ring in A-10, which is the least lipophilic, resulted in significant abolition of the inhibitory activity ($t_{1/2}$ = 37.1 min at 23.8 μM). The tetrahydrofuran ring (A-9) resulted in a 19-fold increase in the inhibition whilst the hindered amine 2,2,6,6-tetramethylpiperidine, which is known for its higher lipid solubility compared to ordinary amines, resulted in further enhancement of the activity ($t_{1/2}$ = 1.2 min at 11.9 μM).

The activity of the acyclic bis-Mannich base A-12 supported the hypothesis that the action of the piperidinols was mediated by conversion to the corresponding bis-Mannich base. Compound A-12 resulted in a 4-fold increase in inactivation compared to the cyclic derivative A-1. However, the previously reported higher toxicity of these compounds compared to the cyclic piperidinol (Gul *et al.*, 2005) made them less favoured for further investigation as antimycobacterial compounds.

Given the narrow range of the piperidinols tested, there was no well-defined correlation between the electronic character of the *N*-substituent and the inhibitory potency. However, except for compounds A-3, A-7 and A-10, the inhibitor potency was found to increase linearly as the critical molar volume of the molecule increased (Figure 5.16A). Despite the improvement of inhibitor potency with the increased lipophilicity of the *N*-substituents, the overall lipophilicity of the inhibitors (as represented by LogP) did not show a correlation with the activity (Figure 5.16B).

Based on these results, the 3D-complementarity to the geometry of the binding pocket appeared to contribute to the activity of the piperidinols. These observations suggest that the enzyme might play a role in the *in situ* conversion into the reactive PVK. To test this hypothesis and to gain better insight into the binding of these ligands to the protein, a simulation of the ligand binding was required.

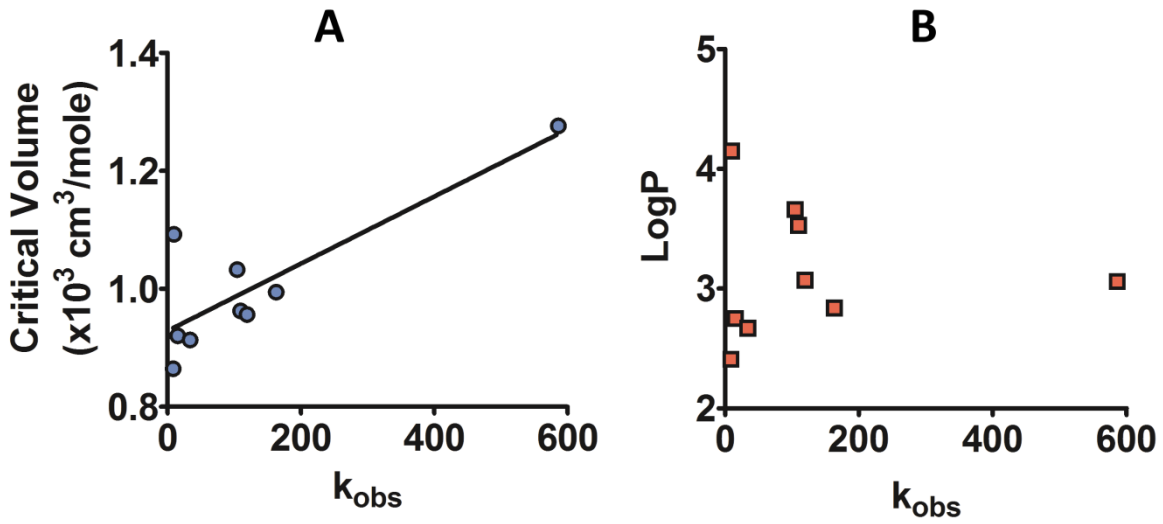


Figure 5.16: Correlation between k_{obs} with the molar critical volume and LogP of the piperidinols.

k_{obs}, the critical volume and LogP values were obtained as described in Table 5.6. (A) Significant correlation was found between k_{obs} and critical volume ($P < 0.05$) with $r^2 = 0.70$ and $p = 0.005$. (B) No correlation was found between k_{obs} and LogP, with $r^2 = 0.00$ and $p = 0.97$. Plots include all the inhibitors except A-3, A-7 and A10. The correlation coefficient (squared, r^2) and p value were calculated by the Correlation module of GraphPad Prism 5.0.

5.2.3.5 Docking simulation

Computational simulations were conducted using the automated docking program LignadFit module of Discovery Studio 2.5 (Accelrys) as described in Section 2.2.10. The MMNAT-HLZ complex structure was used as a surrogate platform for the docking simulation of the piperidinols, since it is the only structure available with a small ligand bound. Re-docking of HLZ into its co-crystallised protein structure using a regular rigid acceptor-flexible ligand approach showed that the optimal docking conditions were of ionised ligands into the hydrated binding site, and thus these were employed for the docking of the inhibitors. The docked conformations were further scored and ranked using the Consensus scoring function of LigandFit (Section 2.2.10).

Analysis of the top-ranking docking poses of the piperidinols revealed a shift towards the methionine residues of the binding pocket (Met209 and Met222, Figure 5.17). The phenyl group of the A-1 appeared to overlap with the aromatic rings of HLZ where this group can form π-π interactions with both Phe130 and Phe204 (Figure 5.17). Further aromatic interactions (π-cation) could be formed with His203 and Trp97 according to the docked conformer (Figure 5.17). The benzyl ring of A-1 was oriented towards the Met209 where it can form hyrdrophobic interactions with this amino acid. Docking results suggest that binding of the piperidinols is likely to take place at a pocket distinct but overlapping with that of HLZ. Interestingly, binding of compound A-1 at this location of the binding pocket (will be referred to as the piperidinol-binding pocket) would involve interaction with 15 residues, five of which differ between MMNAT and TBNAT (Table 5.7). This location in the binding pocket overlaps with CoA binding site, and therefore, is consistent with the protection observed upon incubation with CoA and Ac-CoA but not with HLZ. This observation could explain the suboptimal prediction of inhibitor profiles by using MMNAT as a model for TBNAT. It can also explain the lack of activity against human NATs, since the CoA binding site is different from that of MMNAT (Fullam *et al.*, 2008).

Table 5.7: Amino acids that differ between MMNAT and TBNAT and are located within 5 Å of the piperidinol docking position in the MMNAT structure.

Residue	MMNAT	TBNAT
98	Leu	Lys
196	Val	Ala
203	His	Lys
209	Met	Thr
222	Met	Ser

Comparing the binding of selective inhibitors with variable activity showed that proper alignment of the piperidine ring (specifically the bond between C4 and C5 (Figure 5.9)) of each inhibitor (A-3, A-7, A-9 and A-10) with His203 correlated well with their potency (Figure 5.17 and Figure 5.18). Since the cleavage of the bond between C4 and C5 is the first step in the conversion of piperidinols into the reactive PVK according to the proposed mechanism shown in Figure 5.9, it will hereafter be referred to as the piperidinol-cleavage-bond. The most potent compound, A-7, exhibited a very close alignment of its cleavage-bond with His203. Despite the similar potency of A-3 to A-7, the alignment of its cleavage-bond to His203 did not match that of A-7. This observation supported the hypothesis that the potency of A-3r is dependent on the high alkylating activity of the resulting 4-bromo-PVK. Interestingly, the presence of a lipophilic *N*-substituent in A-7 resulted in a flipped orientation in which the cyclohexyl group was sequestered between Phe130 and Phe204 (Figure 5.18B), whilst in A-10, which has a relatively hydrophilic *N*-substituent, the sulfolane ring was directed away from the phenylalanine residues, displacing the cleavage bond away from the His203 (Figure 5.18D).

The evaluation of the inhibitors and the parallel docking studies have provided evidence that enzyme inactivation by piperidinols depends on the 3D matching with the binding pocket and is not a consequence of nonspecific chemical reactivity. Experimental 3D-structure determination was important to confirm these ideas from docking studies.

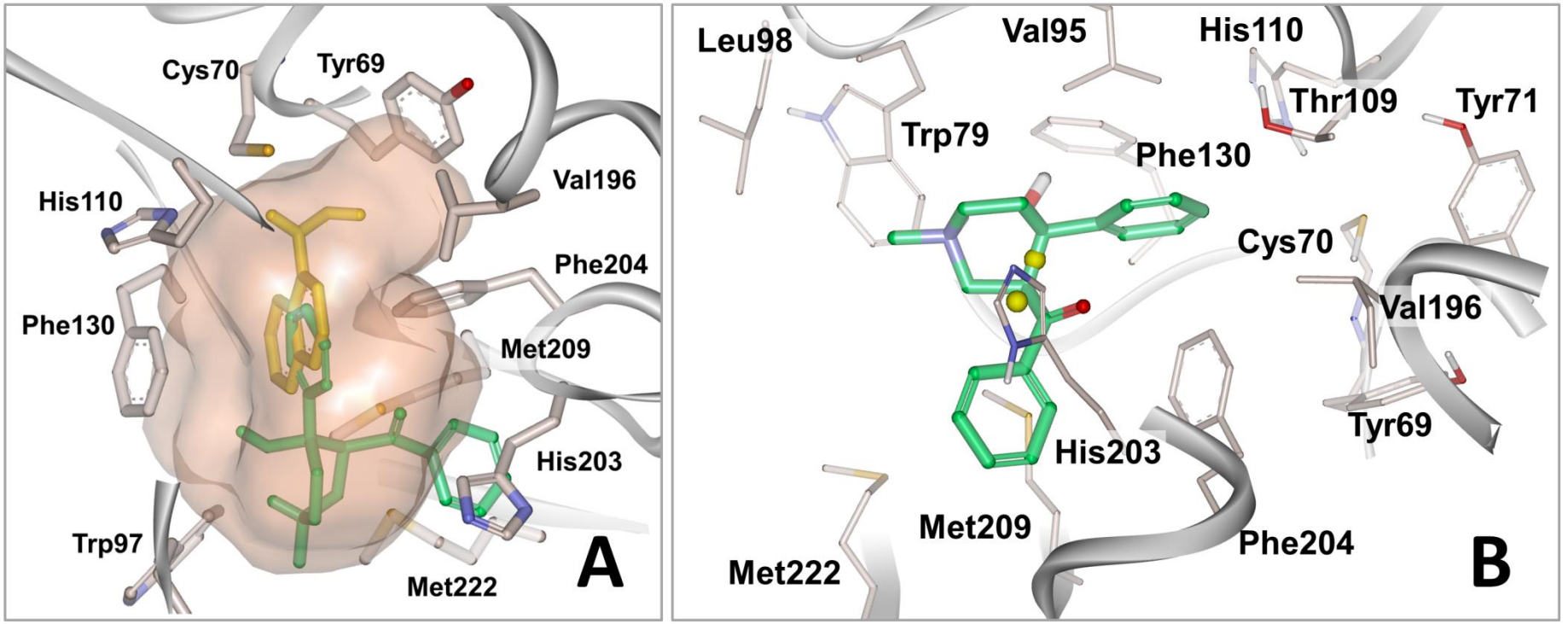


Figure 5.17: The docked pose of compound A-1 in the binding pocket of MMNAT.

Docking was performed using MMNAT-HLZ (PDB code 3LTW; 2.1 Å) as the receptor. The protein was prepared for docking by adding hydrogen and extracting the ligands. The binding site was generated from the co-crystallised HLZ structure (pink surface). Docking conditions, which resulted in accurate prediction of the HLZ binding pose (yellow) were used for docking. These conditions include docking of ionised ligand in the presence of water. Docking was performed by the LigandFit module of the Discovery Studio program (Accelrys). The resulting conformers were ranked according to the default consensus scoring function of LigandFit and the top conformer was selected. (A) The orientation of A-1 relative to the co-crystallised HLZ. (B) Interactions of the top-ranked docked pose of A-1 with the main amino acids in the MMNAT binding pocket. The yellow symbols highlight the centroids of the His203 ring and the cleavage-bond.

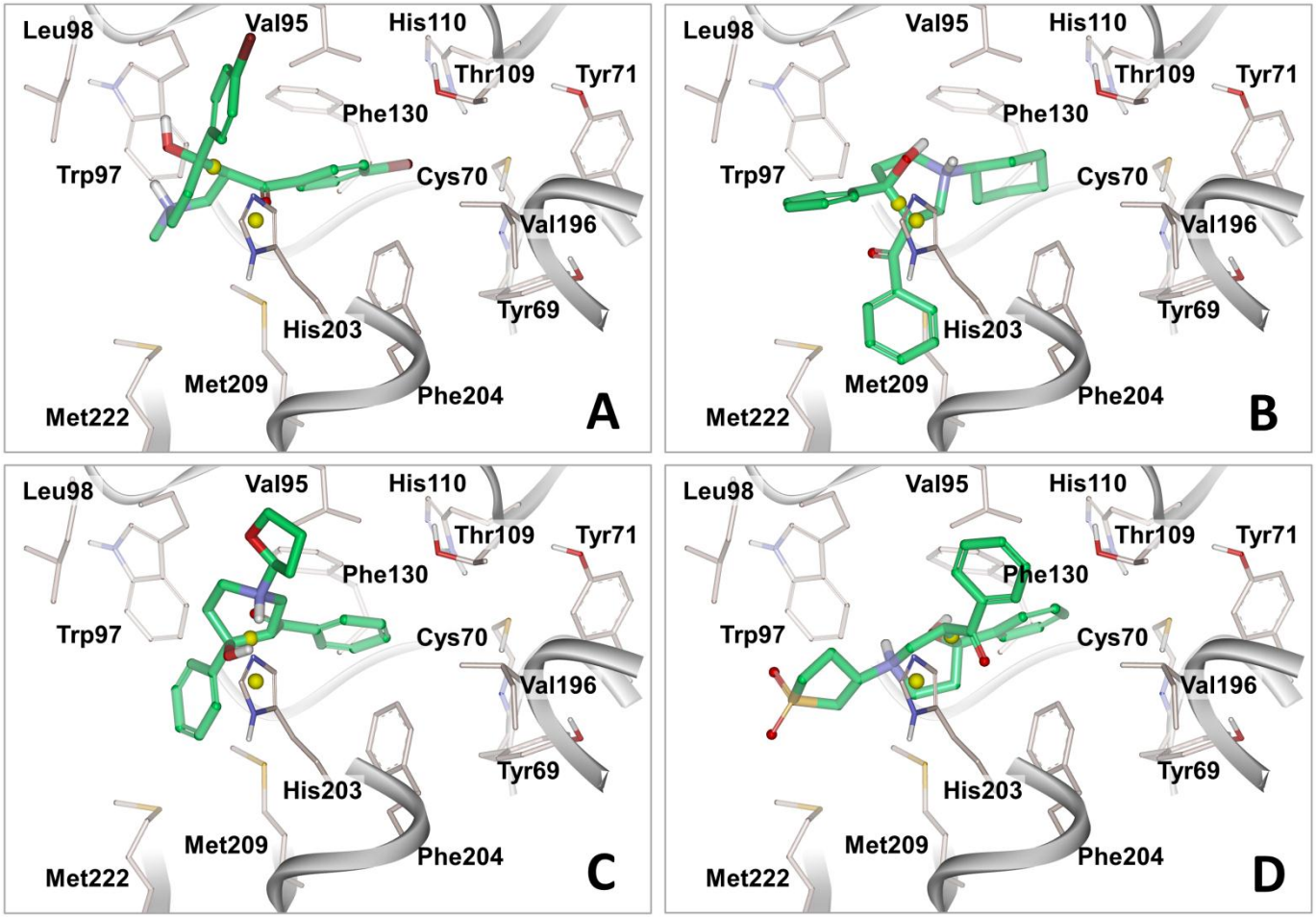


Figure 5.18: Interactions of the top-ranked docked poses of selective inhibitors with the main amino acids in the MMNAT binding pocket.

The interactions of (A) compound A-3 ($t_{1/2}$ = 1.9 min at 5.9 µM), (B) compound A-7 ($t_{1/2}$ = 1.2 min at 5.9 µM), (C) compound A-9 ($t_{1/2}$ = 4.2 min at 11.9 µM), and (D) compound A-10 ($t_{1/2}$ = 37.1 min at 23.8 µM) with the main amino acids in the binding pocket. The yellow symbols highlight the centroids of the His203 ring and the cleavage-bond.

5.2.3.6 Structural studies on MMNAT-inhibitor complexes

To validate the observations acquired from mass spectroscopy, enzymic assays and *in silico* docking studies, the 3D-structure of MMNAT was determined in complex with compound A-1, the parent piperidinol.

The crystallographic structure determination was performed by two methods: one was by soaking the ligand into existing crystals of MMNAT (crystal soaking) (Section 2.2.8.3), and the other was to form the protein-ligand complex in solution (15:1 inhibitor to protein ratio) and then to crystallise it (co-crystallisation) (Section 2.2.8.2). According to the proposed mechanism of inhibition, these methods were expected to capture different stages of the ligand binding. The crystal soaking experiment was aiming to capture the binding of the intact molecule of A-1 to the binding pocket (MMNAT-A-1), whilst the co-crystallisation method would allow time for the ligand to mature into the covalent modification of the active site cysteine with the 3-phenyl-3-oxopropyl moiety (MMNAT-POP). The crystals diffracted to 2.5 Å and 1.6 Å, respectively and data from both types of crystal were processed as described in Section 2.2.8.5. All structures were solved by molecular replacement using the apo-MMNAT structure (PDB code: 2VFB, 2.0 Å). Statistics for data collection, processing and refinement are shown in Table 5.8.

Table 5.8: Data collection, processing and refinement statistics for the MMNAT complex crystals. [a,b]

	MMNAT-A-1	MMNAT-POP	MMNAT-G-1
Data collection			
Beamline	Diamond I02	Diamond I04	Diamond I04-1
Detector	ADSC Q315 CCD	ADSC Q315r CCD	ADSC Q315 CCD
λ (Å)	0.976	0.976	0.916
Temperature (K)		100	
Space group		$P\ 4_1\ 2_1\ 2$	
Unit cell dimensions:			
a = b, c (Å)	51.8, 176.1	52.2, 177.6	52.3, 176.2
α = β = γ (°)		90	
Resolution range (Å)	38.8-1.6 (1.7-1.6)	52.2- 2.5 (2.6-2.5)	39.0-1.9 (2.0-1.9)
Measured reflections	224353 (32410)	42842 (5783)	159623 (23126)
Unique reflections	33245 (4742)	9127 (1295)	20730 (2918)
Multiplicity	6.7 (6.8)	4.7 (4.5)	7.7 (7.9)
Overall I/σ(I)	11.7 (2.0)	4.4 (1.3)	13.2 (4.5)
Overall completeness (%)	100.0 (100.0)	96.0 (96.0)	99.2 (98.2)
Overall R_{merge}	0.095 (0.83)	0.25 (1.1)	0.10 (0.43)
R_{meas}	0.10 (0.90)	0.28 (0.9)	0.11 (0.46)
R_{pim}	0.039 (0.34)	0.12 (0.3)	0.038 (0.16)
Solvent content (%)	36.4	37.9	37.7

Refinement			
Refinement resolution (Å)	38.8-1.6	52.2-2.7	39.0-1.9
R_{cryst} (Fo>4sFo; Fo)	0.19	0.24	0.19
R_{free} (Fo>4sFo; Fo)	0.23	0.30	0.23
Refined residues	271	271	271
Refined water molecules	221	57	168
Refined inhibitor molecules	1	1	0
Overall isotropic B factor ($Å^2$)	11.8	25.7	20.0
R.M.S.D for bond lengths (Å)	0.006	0.016	0.007
R.M.S.D for bond angle (º)	1.06	1.445	1.08
Ramachandran plot			
Most favoured regions (%)	97	97.8	97
Allowed regions (%)	3	2.2	3
Disallowed region (%)	0	0	0

[a] The crystals are as follows: MMNAT-A-1 was obtained by using crystal soaking, MMNAT-POP by co-crystallisation with A-1, and MMNAT-G-1 is an apo-protein structure due to the lack of electron density of support the presence of G-1. [b] Values in parentheses are statistics of the highest resolution shell. Definitions of the statistical terms are listed in Appendix B.

Upon refinement of the structure obtained from the co-crystallised complex MMNAT-POP, a continuous excess electron density connected to the active site Cys70 was observed. It was possible to fit a 3-phenyl-3-oxopropyl modification into this density, in agreement with the proposed inactivation hypothesis (Figure 5.19A). No electron density was observed in proximity to any to the other cysteine residues (Cys120 and Cys274; Figure 6.16) or any other residues, confirming the selectivity of the inhibitors to the active site. Interestingly, the POP moiety binds in exactly the same orientation as HLZ (Figure 5.20).

However, the refinement of the soaked crystal showed a slight excess density connected to the active site cysteine that was more consistent with oxidation of the thiol group. To rule out the possibility of another type of covalent modification (e.g. carbamylation of cysteine), comparison was performed with the complex of MMNAT with the reversible inhibitor G-1 (Table 5.4). Similar density was observed into which a sulfenic acid derivative of cysteine (i.e. a single oxidized sulfhydryl group -SOH) was modelled in both structures with 75 % occupancy (CSO70, Figure 5.19B). Despite that crystallisation was carried out in the presence of 1 mM DTT as a reducing agent, oxidation of the cysteine thiol was not unexpected. The crystals grew within five days at 20 ºC, and thus loss of the reducing ability of DTT had probably occurred. Patches of electron density, contiguous with the methionine residues (Met209 and Met222) and the tryptophan Trp97, were observed at a sigma level of 0.9, indicating the presence of a bound chemical entity(ies). In support of the docking studies, it was possible to partially model the aromatic rings and the methylamine moiety of A-1 with 50 % occupancy into these electron density patches (hypothetical model, Figure 5.19B). Inspection of the ligand electron density seemed to support the binding mode predicted by LigandFit in which the ligand-interaction is close to the methionine residues (Figure 5.19B).

Notably, this electron density was absent in the MMNAT-POP and MMNAT-G-1 complexes, excluding the possibility that it was due to retained chemicals from the expression system or from the crystallisation cocktail.

Trials to build the ligand using the ARP/WARP Ligands (Evrard *et al.*, 2007) module of CCP4 (Bailey, 1994) failed to find any good model. Situations in which ligands are not clearly defined in their binding site are usually attributed to low occupancy or high mobility. Efforts were made to re-collect data for the complex, but without success.

However, in the case of piperidinols, this observation was more likely to be caused by the catalytic activation of the ligands by the enzyme at this site before they mature into the reactive intermediates, which permanently inactivates the enzyme as they approach the active site cysteine. It is possible that enzyme catalysed activation will take place despite the relatively acidic pH (6.5) at which the crystals grew, as the enzyme was shown to maintain a stable catalytic (acetyl-transfer) activity within a pH range of 6-9 (Section 6.2.3.1,). However, further evaluation of the pH stability of the piperidinols would need to be pursued for drug design.

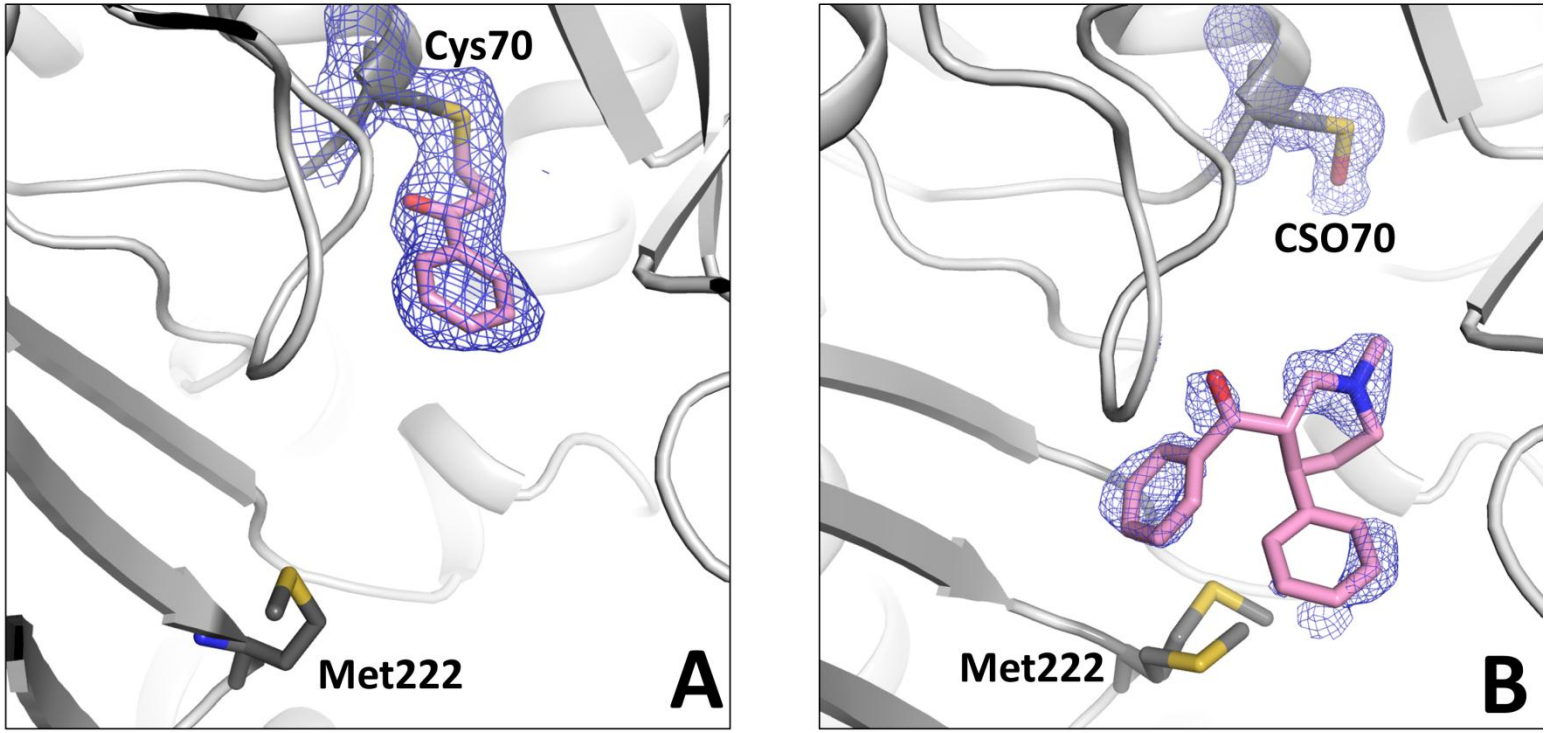


Figure 5.19: The active site electron densities observed in the MMNAT-POP and MMNAT-A-1 complexes.

(A) Upon co-crystallisation of MMNAT with A-1, excess electron density connected to Cys70 was observed, into which a 3-phenyl-3-oxopropyl (POP) modification was modelled with full occupancy. The covalent modification is shown in pink with the electron density shown using blue 2Fo–Fc electron density contoured at 1 σ. (B) The modelled molecule of A-1 in the MMNAT-A-1 complex obtained by crystal soaking (50 % occupancy) is in pink with the 2Fo–Fc electron density shown in blue contoured at 0.9 σ. The density corresponds to the oxidation of the cysteine residue Cys70 into sulfenic acid (CSO70, occupancy 75 %) and is shown using blue 2Fo–Fc electron density contoured at 1 σ. The key amino acid residues are labelled. The figures were prepared using PyMOL (Schrodinger, 2010).

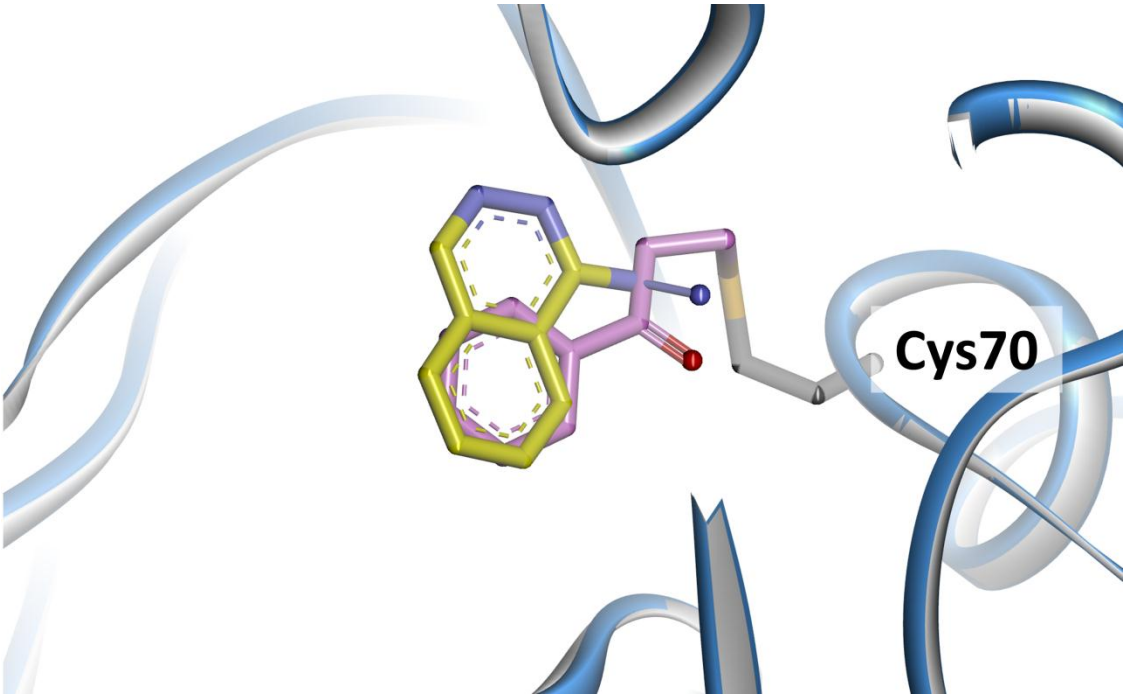


Figure 5.20: The orientation of the POP moiety in the arylamine substrate binding pocket.

Overlay of the MMNAT-POP (blue ribbons) to the MMNAT-HLZ structure (grey ribbons, PDB code 3LTW) show the orientation of the POP (pink) moiety relative to that of HLZ (yellow). The figure was prepared using the DS Visualizer 3.1.

5.3 Conclusions

In this chapter, various strategies were utilised to explore the inhibition profile of MMNAT and compare it to TBNAT as described in the flow chart in Figure 5.1.

Re-evaluations of all available NAT inhibitors using HLZ as a substrate resulted in lower IC_{50} values, emphasizing the impact of substrate choice on inhibitor-screening assays.

Efforts to find new inhibitors based on the 3D-shape of the piperidinols, or which caused a physical stabilisation of the protein, succeeded in the identification of new inhibitors with novel scaffolds. These inhibitors provide potential hits for future development, although none of the scaffolds appeared to be more potent inhibitor than the piperidinol.

In-depth analysis of the inhibition by piperidinols was performed. Covalent irreversible inactivation of the enzyme by these inhibitors was confirmed by mass spectroscopy, time-dependent inhibition kinetics and structural studies. Evidence towards adduct formation with the active site Cys70 was also obtained in this work. The inhibition by the piperidinols involves the formation of a reactive intermediate, the penylvinyl ketone that can react with the thiol group of cysteine. Inactivation of enzymes via the formation of phenyl vinly ketones is a common mechanism that was observed with acyclic Mannich bases considered for drug design (Davioud-Charvet *et al.*, 2003).

Structural studies on the MMNAT piperidinol interactions were performed using *in silico* docking and 3D-crystallographic structure determination. The piperidinols bind to a distinct pocket of the enzyme that overlaps with the HLZ binding pocket before they mature into a covalent modification at the active site cysteine. Complementarity of the inhibitors to the binding pocket affects their inhibitory activity. These studies allowed for a better understanding of the structure activity relationship of these inhibitors. A summary of the results obtained from these studies are shown in Table 5.9.

The covalent mechanism of inhibition has many advantages to offer in overcoming the shortcomings of the current TB therapies. Covalent inhibitors are usually associated with high biochemical efficiency and normally are not affected by the level of the endogenous ligand concentration (Singh *et al.*, 2011). Thus, lower doses, a sterilising effect, as well as avoiding resistance can be ensured by the design of potent and selective covalent inhibitors. Furthermore, such inhibitors usually provide a potentially longer duration of action as the gain of activity requires the synthesis of new enzyme, which would be another advantage considering the long multiplication time of mycobacteria (Singh *et al.*, 2011). However, despite the many examples of successful covalent drugs, the possible toxicity associated with this mechanism requires the careful design of highly selective agents.

The activity of these compounds against *M. tuberculosis* in which the *nat* gene is deleted (Fullam, 2007) indicates the presence of other targets of action that are likely to be cysteine dependent proteins. Further analysis is needed to elucidate the specificity against the other cell systems. Understanding the mechanism of inhibition would also allow the design of selective ligands for these targets. However, the design of highly potent and specific ligands is usually hampered by the lack of understanding of the molecular recognition, which is a prerequisite for receptor binding. Therefore, the final answers may not be reached before structural information has been obtained for TBNAT, both with and without bound ligands. The following chapter describes attempts at obtaining this information.

Table 5.9: Summary.

Mass spectroscopy (1:1 molar ratio)	The active site cysteine was modified.
Mass spectroscopy (15-fold molar excess of an inhibitor)	All cysteine residues in the protein sequence were modified.
MMNAT co-crystallisation with A-1	Electron density consistent with active-site cysteine modification was observed.
MMNAT crystals soaking in A-1	Electron density that implies binding near to Met209 in the binding pocket was observed.

6 Approaches to Crystallisation of TBNAT

6.1 Introduction

Knowledge of the 3D structure of proteins and ligands is of great assistance in the study of protein function and in drug discovery. Previous attempts to solve the 3D-structure of TBNAT had been hampered by insufficient pure protein yield (Sikora *et al.*, 2008; Fullam *et al.*, 2009). Sub-cloning of the *tbnat* gene into the pVLT31 vector provided a means of overcoming this difficulty, and resulted in a soluble recombinant TBNAT yield of 6.5-16 mg per litre culture (Chapter 4) of crystallisation quality protein.

One of the most widely used strategies to improve protein crystallisation is to screen more conditions. Most of the other strategies to improve crystallisation are based on stabilization of the protein structure. It is easier to crystallise rigid structures and they usually form better diffraction-quality crystals (Vedadi *et al.*, 2006). The success rate of crystallization can also be improved by the use of ligands or solution conditions that has a stabilising effect on the protein (Vedadi *et al.*, 2006).

In order to solve the 3D-structure of TBNAT, crystallisation of the protein was required. Therefore, the specific **aims** of the work presented in this chapter were:

- **To screen for suitable crystallisation conditions for TBNAT**
- **To study the influence of different solution conditions/additives on the thermal stability of TBNAT and compare it to those of MMNAT**
- **To study the effect of the His-tag on protein crystallisation for TBNAT and compare the results to that for MMNAT**
- **To model the 3D-structure of the TBNAT-HLZ complex by homology modelling using the MMNAT-HLZ structure as a template**

6.2 Results and discussion

6.2.1 Preliminary screening for suitable crystallisation conditions

TBNAT with the His-tag removed was obtained as described in Section 4.2.4.1, and used for crystallisation trials. A summary of the range of the crystallisation conditions used in the initial screening is shown in Table 6.1.

Table 6.1: Summary of the initial conditions used to screen for suitable crystallisation conditions for TBNAT. [a]

Screens / Protein Solution	JCSG-plus	PACT	INDEX	PEG/Ion	St. S. I+II[b]	Wizard I+II	Morpheus	Cr. S. I+II[c]	Cr. S. Lite[d]
5 mg/mL Tris-HCl pH 8	4 °C	---	---	---	---	---	---	---	---
	20 °C	---	---	---	---	---	---	---	---
8-11 mg/mL Tris-HCl pH 8	4 °C	4 °C	4 °C	---	---	---	---	---	---
	20 °C	20 °C	20 °C	20 °C	20 °C	20 °C	20 °C	20 °C	20 °C
19 mg/mL Tris-HCl pH 8	20 °C	---	20 °C	20 °C	---	---	---	---	---

[a] Crystallization trials using different protein concentration-sparse matrix screen-temperature combinations were performed by the sitting-drop vapour-diffusion technique (Section2.2.8 and Section 2.2.8.4). In each screen, an equal amount (100 nL) of freshly prepared TBNAT protein and precipitant were mixed and equilibrated against 100 µL reservoir at either 4 °C or 20 °C. Details of the commercial screens are listed in Table 2.4. [b] St.S. I+II is Structure Screen I + II. [c] Cr. S. I+II is the Crystal Screen I + II. [d] Cr. S lite is the Crystal Screen Lite.

Protein concentrations in the range of 8-11 mg/mL were found suitable to perform the screens, since approximately 70 % of the drops across this protein concentration range showed mild to moderate precipitation, which is indicative of the potential for crystallisation. The remainder of the drops were either clear or showed a heavy precipitate. Therefore, this range of protein concentration was used in subsequent screens. Dilution of the starting TBNAT solution (5 mg/mL) resulted in clear drops in approximately 80 % of the conditions in each screen, while protein concentrations above 11 mg/mL were associated with heavy precipitation. Further attempts using PGA, Proplex, MIDAS, screens were also attempted. Nevertheless, in the initial screen, crystals were obtained at a protein concentration of 19 mg/mL (Figure 6.1). Disappointingly, X-ray diffraction tests showed that these crystals were salt. Other conditions containing Zn^{2+} or Mg^{2+} as well as phosphate buffer led to salt crystal formation.

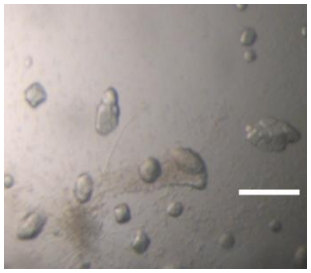


Figure 6.1: The morphology of the salt crystals formed in the initial screen.

Crystallization conditions: 19 mg/mL starting TBNAT concentration, precipitant solution consisted of: 0.2 M NaCl, 0.1 M Phosphate/citrate, pH 4.2 and 20 % w/v PEG-8000, and the incubation temperature was 20 ºC. Upon data collection, a diffraction pattern characteristic of salt was observed. The scale bar is equivalent to 100 µm.

6.2.2 Analysis of the initial screening results

The behaviour of the protein in different screens (JCSG-plus, Index, PEG/Ion II and Crystal Screen-lite.) was observed in order to understand the effect of the range of conditions within these screens. Monitoring was carried out for ten days by qualitative visual assessment of the resulting drops. None of the screened conditions resulted in crystallisation of the protein. Therefore, conditions that resulted in clear drops or drops that had mild gelatinous precipitate were of particular interest for possible optimisation. These observations usually indicated that the protein was stabilized by the content of these conditions and was observed in 193 wells (out of 288, i.e. 67 %) of the screens mentioned above. The percentages of the occurrence of different precipitants and the pH in these 193 wells are shown in Figure 6.2.

Although this study was not a systematic one, mainly due to the random nature of sparse matrix screens, it might help to highlight possible conditions in which the protein showed measured stability for subsequent optimisation. The high occurrence of PEG-3350 among the 193 wells is not unexpected due to the redundancy among the conditions in the commercially available screens (Kimber et al., 2003). The stability of the enzyme over a wide range of pH condition was also observed (Figure 6.2).

Guided by these results, optimisation was performed in the conditions containing PEG-3350, MPD (2-methyl-2,4-pentanediol) and PEG-8000 with either acetate or ammonium sulphate as additive. None of these conditions resulted in any crystals. Further trials using MMNAT crystals to microseed the optimisation plates did not result in any improvement (Sweeney and D'Arcy, 2009). Therefore, additional strategies were required to improve the crystallisability of the protein.

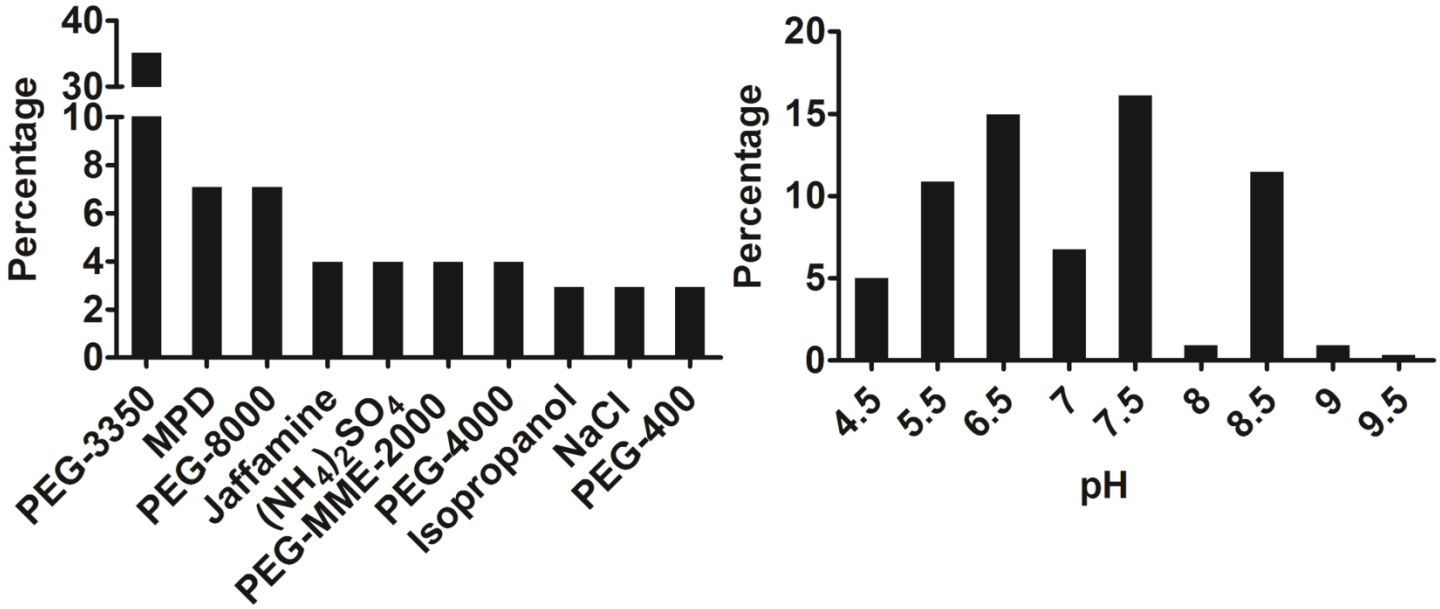


Figure 6.2: Analysis of the crystallisation trials of TBNAT in four different commercial screens.

Histograms showing the percentage of (A) precipitants and (B) pH observed in the 193 wells (67 %) which showed mild precipitation or clear drops within ten days.

6.2.3 Thermal stabilisation of TBNAT

Protein stability, homogeneity and solubility are key factors that affect protein crystallisability (Ericsson *et al.*, 2006). Therefore, optimization of these qualities is important in order to improve the success rate of crystallization.

The TSA is one of the recently developed high-throughput screening methods that has been used for buffer optimization and ligand-induced stabilization of proteins (Niesen *et al.*, 2007; Nettleship *et al.*, 2008). Improved protein crystallisability upon co-crystallisation with stabilizing additives has been reported (Ericsson *et al.*, 2006; Nettleship *et al.*, 2008). The TSA depends on the concept that folded and unfolded proteins can be distinguished via the change in fluorescence of a hydrophobic fluoroprobe (SYPRO Orange dye) that is present in the solution as summarised in Figure 6.3 (Lavinder *et al.*, 2009; Kemp *et al.*, 2012).

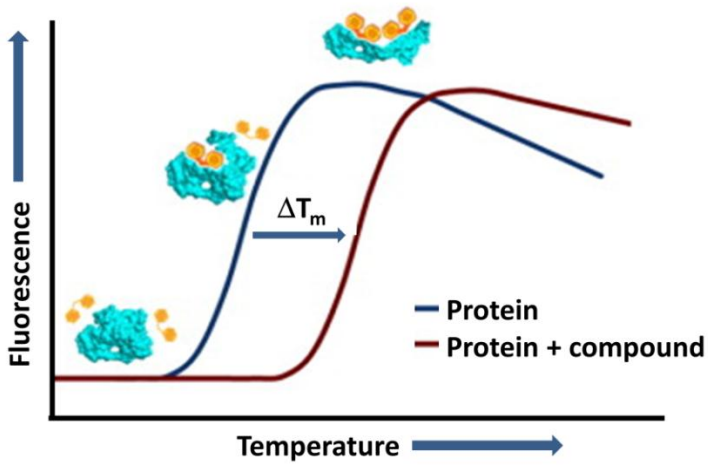


Figure 6.3: A schematic representation of the events monitored in the thermal shift assay.

The method involves incubation of protein (in blue) and ligand (in orange) with an environmentally sensitive fluorescent dye that is quenched in aqueous environments. Thermally induced denaturation will result in the exposure of the hydrophobic regions of the protein which leads to un-quenching of the dye. The increase in fluorescence is monitored at increasing temperatures (melting curve) and the midpoint of the transition between the native state and the denatured state is referred to as the T_m. Measurement of the difference in melting temperatures of the protein (i.e. ΔT_m) in different conditions or in absence/presence of ligands can be used as a measure of the stability or affinity, respectively. The figure and legend were modified from Kemp *et al.* (2012).

TBNAT itself is stable at a relatively high temperature (60 ºC) (Section 1.4.3). The TSAs were used in this study to evaluate the influence of variable solution conditions and different additives on the stability of the enzyme.

TBNAT with the His-tags removed was screened against two types of small-molecule library. The first library comprised different salts, pH conditions, and commonly found small molecules in crystallisation trials (provided by the SGC, Oxford, hence referred to as SGC-stability screen). The components of this screen are listed in Appendix A. The second library comprised compounds specific for NAT and are referred to as NAT-ligands. The latter library was generated in our laboratory and comprised substrates, physiologically relevant compounds (e.g. different acyl-CoA cofactors), and other molecules that were either identified as, or predicted to be, ligands of TBNAT (totalling 34 compounds). Parallel screens on the highly similar but crystallisable MMNAT were performed for comparison. To obtain a melting curve of each protein, the relative fluorescence of SYPRO Orange was measured while subjecting the protein to a gradually increasing temperature as described in Section 2.2.9. The temperature shift between the melting temperature ($\Delta T_{m)}$ in the presence and absence of a bound ligand was measured (Figure 6.3). Melting temperatures of ~70 ºC and ~48 ºC (in 100 mM HEPES, pH 7 and 150 mM NaCl) were observed for TBNAT and MMNAT, respectively (Figure 6.4). These results were comparable to the previously measured temperatures of ~68 ºC for TBNAT (measured in 100 mM sodium phosphate buffer, pH 7.4 and 20 mM NaCl, using circular dichroism (Lack *et al.*, 2009)) and ~48 (measured in 20 mM Tris-HCl, pH 8 and 150 NaCl using the NAT activity assay (Fullam, 2007). Evaluation of protein stability by the TSA was reported to provide similar results to that obtained by the circular dichroism (Vedadi *et al.*, 2006; Vedadi *et al.*, 2010).

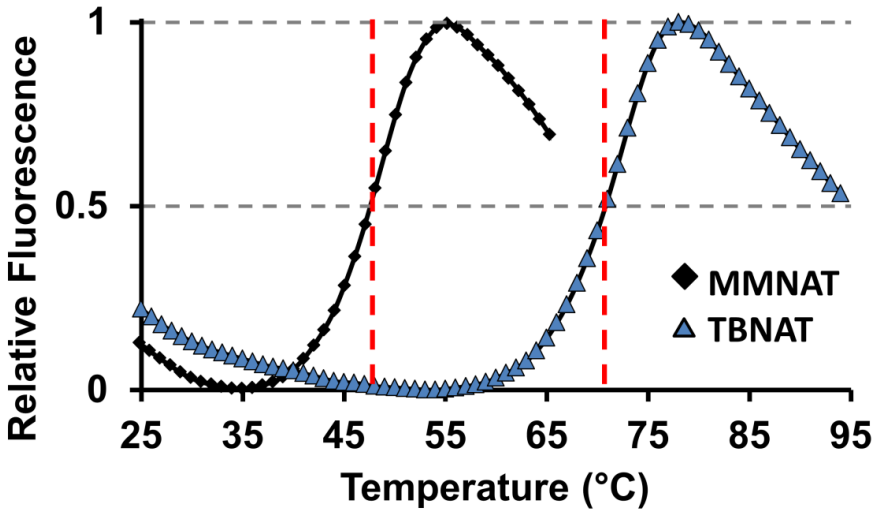


Figure 6.4: The melting curves of MMNAT and TBNAT as measured by the TSA.

The curves were obtained by plotting the relative fluorescence of SYPRO Orange against temperature for MMNAT and TBNAT in 100 mM HEPES, pH7 and 150 mM NaCl (Section 2.2.9). The melting temperature , T_m, on each curve is indicated by the red line.

6.2.3.1 The effect of pH on NAT stability

pH is one of the key parameters that affects the stability and function of proteins. The effect of the pH on the TBNAT stability was studied by the TSA in different buffer systems (Figure 6.5). Examples of the melting temperature curves are shown in Figure 6.6. This study was followed by evaluating the enzyme activity after incubation at different pH obtained by using a combined buffer (pH range 5-10) (Figure 6.7) to check for a correlation between the stability and activity of the enzyme.

TBNAT, which has an estimated pI value of 5.5 (as predicted from sequence using the PortParam online tool http://web.expasy.org/protparam/, showed thermal stability over a wide range of pH (pH 6-8.3) (Figure 6.7). The stability results from the TSA were also in agreement with the stability results from analysis of the crystallisation drops (Section 6.2.2). The particularly enhanced stability at pH 6-6.5 (Figure 6.5) is consistent with the proposed role of the enzyme in intracellular survival and persistence inside macrophages (Bhakta *et al.*, 2004). In the light of evidence that the macrophage vacuoles, in which live mycobacteria reside, fail to acidify below pH 6.1-6.5 (Section 1.2.3), this

stability would suggest that TBNAT remains active following infection. Interestingly, the enzyme also showed a gradual increase in its catalytic activity after incubation in this pH range, and only lost activity below pH 6 (Figure 6.7). The MMNAT, which is known to have higher catalytic activity (Fullam *et al.*, 2009), retained its catalytic activity over a broad pH range (6-9) and only losses 60 % of that activity when preincubated at pH 5. This is to be expected since *M. marinum*, which is an opportunistic *Mycobacterium*, can grow over a wide range of pH values with an optimal growth at acidic conditions (5-6.5) (Falkinham, 2002). The function of the NAT enzyme in *M. marinum* remains to be elucidated. High pH conditions resulted in decreased thermal stability, which agrees with the observation of reduced enzymic activity after incubation at that pH. The difference in the buffer systems used in the TSA (Figure 6.5) and the enzyme assessment of the activity (Figure 6.7) could explain the retained activity despite the lower observed stability.

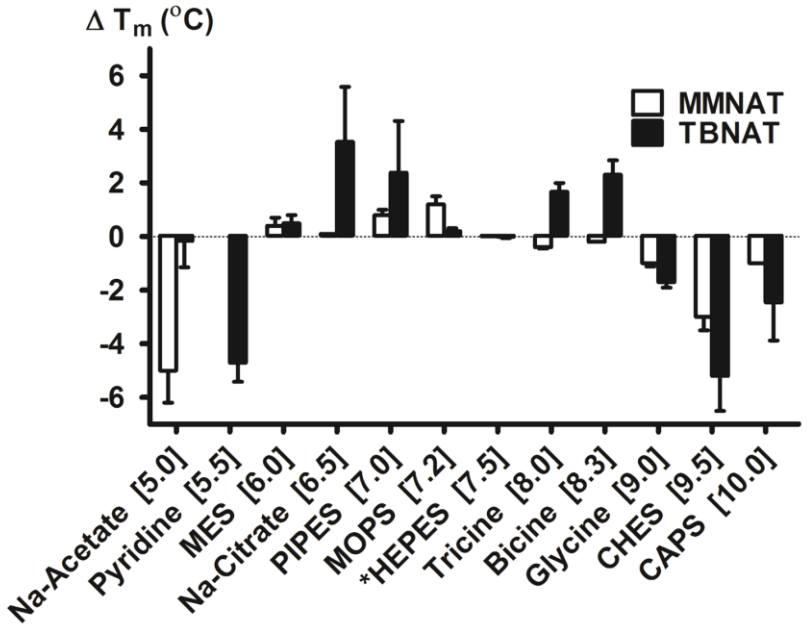


Figure 6.5: Changes in the TBNAT and MMNAT melting temperatures (ΔT_m) for different buffer conditions.

Changes in the melting temperature (ΔT_m) were calculated for measurements on TBNAT/MMNAT in different buffer systems using TBNAT/MMNAT in 100 mM HEPES, pH 7.5 and 150 mM NaCl as a reference (marked with an asterisk) as described in Section 2.2.9. All buffers were used at 100 mM concentration with 150 mM NaCl. The numbers between square brackets represent the pH of each buffer solution. Values are shown as the mean ΔT_m ± S.D of three independent measurements. A negative mean ΔT_m value signifies that the buffer destabilizes the proteins, and a positive ΔT_m value indicates that the buffer has a stabilizing effect.

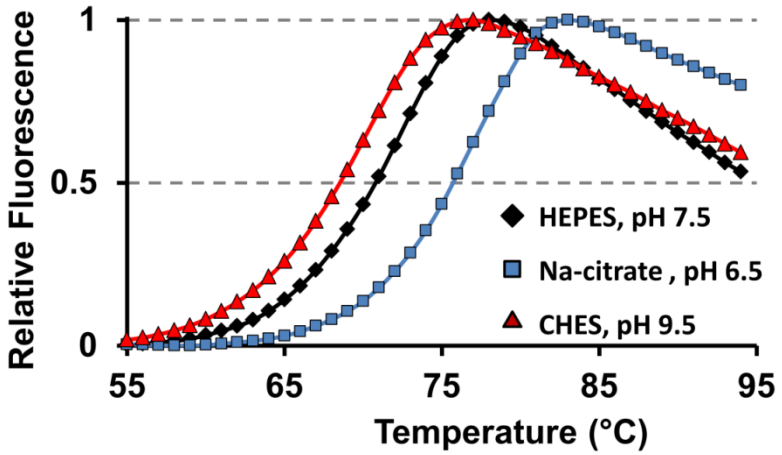


Figure 6.6: Variation in T_m of TBNAT at different pH.

Thermal stability assay melting curves demonstrating improved protein stability (the curve is shifted to the right) in a solution at pH 6.5. NAT shows better stability at pH 6.5. Data were plotted from one measurement determination of the melting temperature shift as described in Section 2.2.9.

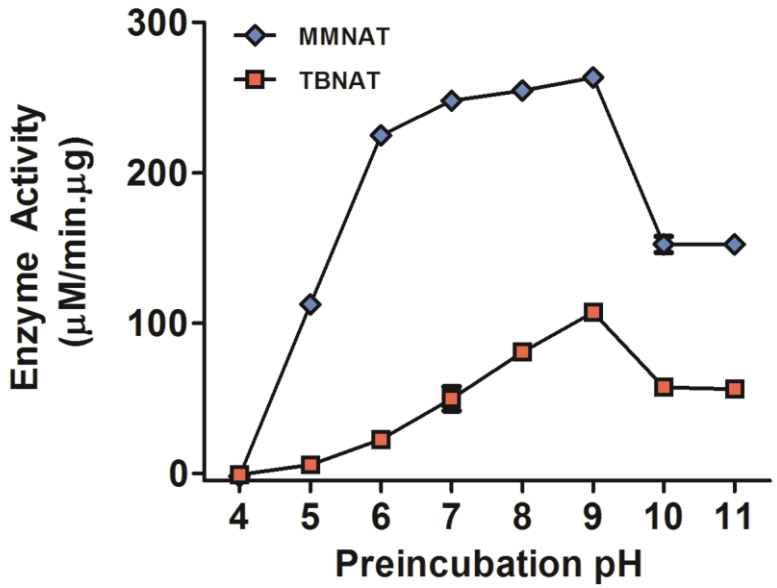


Figure 6.7: Effect of pH on the activity of both TBNAT and MMNAT.

The enzymes were incubated prior to the assay at different pHs for 30 min in an ice bath. The intended pH levels were obtained by adjusting the pH of a combined buffer (100 mM equimolar mixture of MES (pKa 6.16), HEPES (pKa 7.55), Tricine (pKa 8.16) and CHES (pKa 9.55)) by one pH unit over pH range of 4-11. pH values of 4 and 11 are not within the optimal buffer range. The activity of each enzyme was then measured at 24 ºC, pH 8 by the NAT activity assay (Section 2.2.6.1) in the presence of 500 µM HLZ. The reaction was performed in the linear range of the enzyme activity. Each point represents the mean ± S.D from triplicate determinations. The error bars are within the symbols.

6.2.3.2 The effect of different ions on NAT stability

Cations can play an essential role in crystallisation. The effect of ions can include participation in intermolecular bridges that contribute to protein stability, in addition to their possible role as ligands or cofactors (Yeh, 2009). Metals have not been reported to be involved in the activity of the different NATs. Both MMNAT and TBNAT enzymes show moderate tolerance to a variety of ions in solution with a few exceptions, as shown in Figure 6.8. Small ions with high-water activity can alter the structure of water surrounding the polar residues, which is one of the factors that determines the folded conformation of globular proteins in solution (Eisenberg and McLachlan, 1986). The enzyme instability at 500 mM NaCl is likely to be due to the disruption of the electrostatic interactions within the protein. This finding is consistent with the significant role of ionic interactions in the thermal stability of thermostable proteins (Szilágyi and Závodszky, 2000). However, a

similar effect has been observed with MMNAT even though the T_m values could not be calculated due to the instability of the enzyme at this salt concentration in the screen (Figure 6.9). The protective effect of 500 mM NaCl in the presence of glycerol (Figure 6.8) suggests that maintenance of hydration by glycerol favours the folded over the unfolded state of TBNAT and MMNAT. A similar destabilising effect was observed with iodide as the counter ion (NaI) with both MMNAT and TBNAT. Iodide is a chaotrope (large ions of low charge density) that is known to perturb the conformation of proteins by altering the structure of water and in doing so favours the transfer of apolar regions buried in the interior of proteins to the aqueous phase (Hanstein *et al.*, 1971).

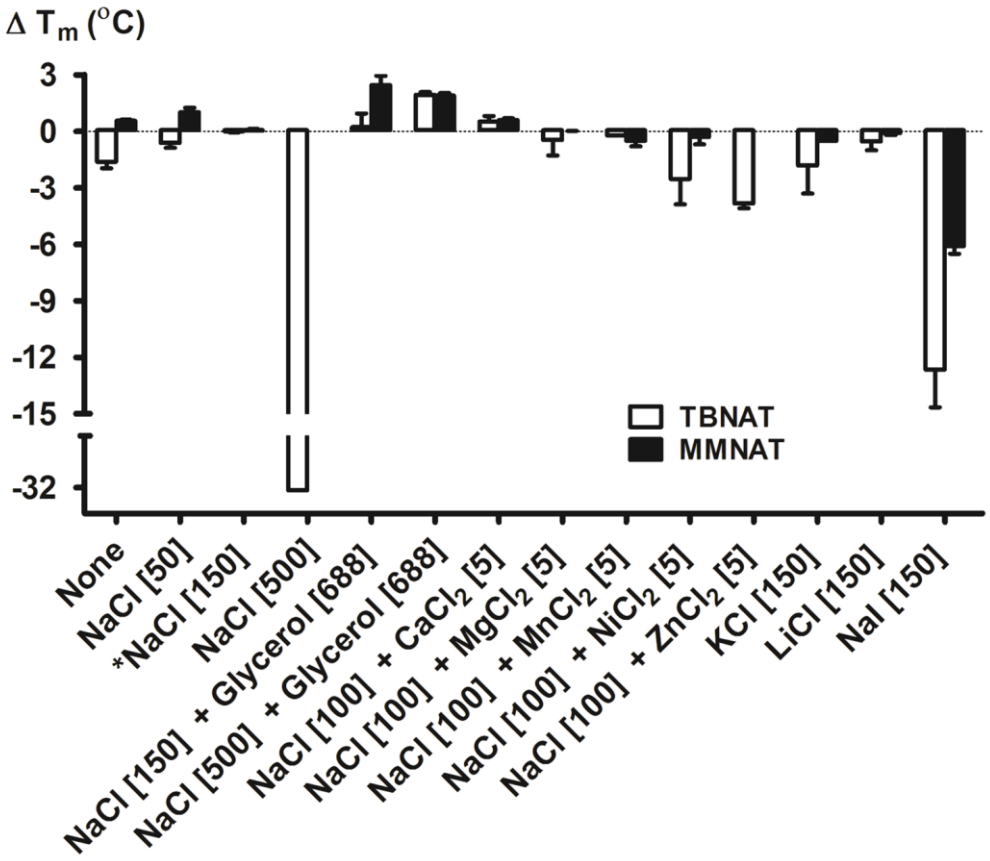


Figure 6.8: Changes in the TBNAT and MMNAT melting temperature in different salts.

Changes in the melting temperature (ΔT_m) were calculated for measurements on TBNAT/MMNAT in different salts and ionic strengths using TBNAT/MMNAT in 100 mM HEPES, pH 7.5 and 150 mM NaCl as the reference (marked with an asterisk) as described in Section 2.2.9. All solutions contained 100 mM HEPES buffer at pH 7.5. The numbers between square brackets represent the concentration of each salt/additive in millimolar. The columns represent the mean ΔT_m ± S.D. values of triplicate independent measurements. A negative ΔT_m value signifies that the buffer destabilizes the protein, and a positive ΔT_m value indicates that the buffer has a stabilizing effect.

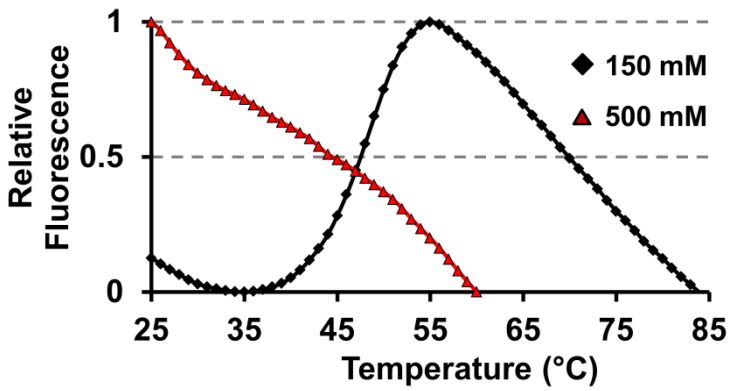


Figure 6.9: The effect of high NaCl level on the stability of MMNAT.

Thermal stability assay melting curves demonstrating the destabilising effect of the 500 mM NaCl solution on MMNAT as indicated by the high initial fluorescence background with no transition. Data were plotted from one measurement determination of the melting point shift as described (Section 2.2.9).

6.2.3.3 The effect of small molecules on the stability of NATs

Several small ligands have been shown to induce the stabilisation of proteins and hence enhance crystallisation when used in low concentration in crystallisation screens or in crystal optimisation (Yeh, 2009). These ligands can be classified into: (a) cosmotropes, such as amino acids and sugars: these molecules can improve the solubility of proteins by binding directly to the protein, which increases the net charge of the protein, making it more soluble, (b) detergents, (c) chaotropes, such as urea and guanidine: both of which improve the solubility of the proteins by interfering with ionic, hydrogen, or hydrophobic bonds, which can disturb intermolecular interactions, and (d) Cofactors, substrates and inhibitors, which bind to specific pockets on the protein and thus stabilise proteins by shifting the folding state equilibrium. This effect will subsequently affect the lattice interactions and crystallisation behaviour.

Ligands, from small molecules within the SGC-stability screen or from the NAT-ligand screen, that resulted in a positive shift in $T_m > 1$ with either TBNAT or MMNAT, are presented in Figure 6.10 and Figure 6.11, respectively. Most of the tested ligands exert better stabilisation effect on MMNAT compared to on TBNAT. Histidine showed the highest stabilising effect on TBNAT; however, the resulting curves were very broad and could indicate a false-positive measurement. Despite this, it was thought it worth trying this additive in subsequent screening trials. Other cosmotropes, such as arginine, lysine, a mixture of arginine and glutamate (1:1 ratio), glycine, choline, trimethylamine *N*-oxide (TMAO), myo-inositol, sorbitol and sucrose have also resulted in a positive shift in the melting temperature of TBNAT (Figure 6.10). The possible effect of these compounds on the ionisation state of the protein might explain this stabilisation effect, since TBNAT is a relatively thermostable protein and ionic interactions (salt bridges) are considered major contributors towards stability of thermostable proteins (Kumar *et al.*, 2000). Surprisingly, the TBNAT showed slightly lower stability in the presence of the non-ionic detergent n-

octyl glucopyranoside (OGP) compared to the ionic detergent lauryldimethylamine N-oxide (LOAD). The improved stability in the presence of LOAD might be due to formation of ionic interactions with the protein which help in stabilising it (Salameh and Wiegel, 2010). Both detergents and chaotropes induced the stabilisation of MMNAT; this effect can be attributed to the effects of these additives on the solubility of the MMNAT, which showed lower aggregation and higher enzyme activity in the presence of detergents (Fullam, 2007). Among the cofactors tested, NADPH resulted in positive $\Delta T_{m,}$ this might be due to the adenosine phosphate moiety which is a feature shared with the CoA molecule.

The effects of the more selective NAT-ligand library on the enzyme stability are shown in Figure 6.11. Despite the stabilisation of MMNAT by the NAT-inhibitors (HTS derived hits A1-3, B-1, C-1 and D-1(Chapter 5)), none of them resulted in positive thermal shifts with TBNAT. This is not surprising, since they resulted in such weak inhibition of the TBNAT enzymes (Chapter 5). Interestingly, HOPDA (G-1), the substrate of the HsaD enzyme, resulted ΔT_m of 2 ºC. G-1 and two of its analogues inhibited both MMNAT and TBNAT (Section 5.2.2.2). However, none of the arylamine substrates resulted in significant stabilisation of TBNAT: this could be attributed to the higher affinity of the arylamine to the acetylated enzyme according to the Ping-Pong catalytic mechanism, in which Ac-CoA would bind first to the enzyme (Riddle and Jencks, 1971; Delgoda *et al.,* 2003). This conclusion was supported by the high stabilising effect of Ac-CoA on both MMNAT and TBNAT. Among the other CoA factors that stabilised TBNAT were acetoacetyl-CoA and malonyl-CoA, which are both weak substrates of the enzyme (Lack *et al.*, 2009). However, arachidonoyl-CoA, which is not a substrate of either of the enzymes, when tested (in the current work) at 50 μM arachidonoyl-CoA concentration, as described in Section 2.2.6.2, was found to inhibit TBNAT (% inhibition of 63 ± 2 %) and MMNAT (% inhibition of 97.8 ± 2 %). Interestingly, it has been reported that arachidonic acid is involved in

phagosome maturation (Anes *et al.*, 2003) and exhibited antimycobacterial activity (Shimizu and Tomioka, 2009). However, inhibition of TBNAT by arachidonoyl-CoA was not further investigated in this study.

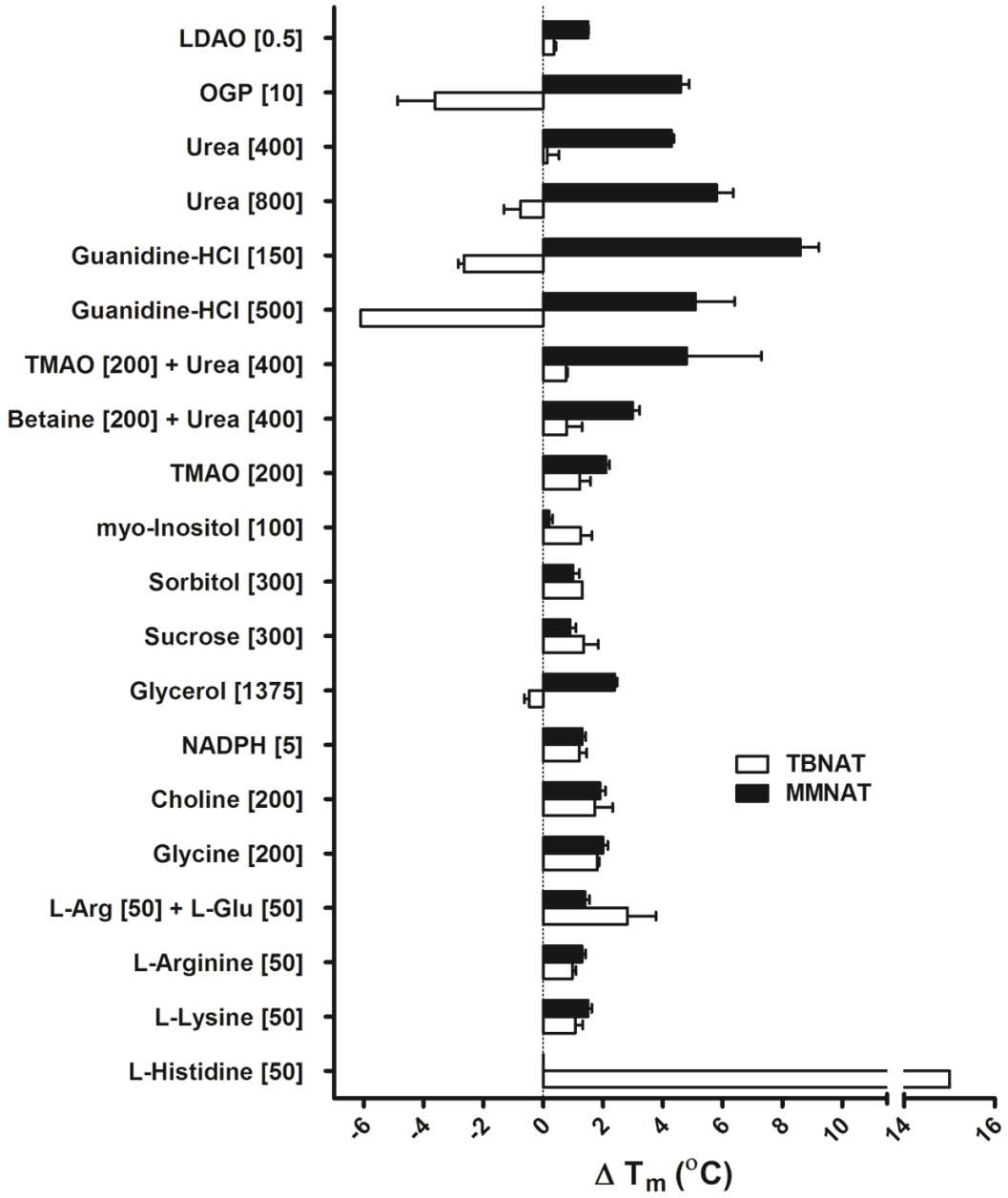


Figure 6.10: The changes in the TBNAT and MMNAT melting temperature in the presence of small-molecules from the SCG-stability screen.

Changes in the melting temperature (ΔT_m) were calculated for measurements on TBNAT/MMNAT in the presence of these additives from the SGC-stability screen which resulted in ΔT_m > 1 ºC (Appendix A) using TBNAT/MMNAT in 100 mM HEPES, pH 7.5 and 150 mM NaCl as the reference. All solutions contained 100 mM HEPES buffer at pH 7.5. The numbers between square brackets represent the millimolar concentration of each additive. The columns represent the mean ΔT_m values ± S.D of three independent measurements. A negative ΔT_m value signifies that the additive destabilizes the protein, and a positive ΔT_m value indicates that the additive has a stabilizing effect. TMAO: trimethylamine *N*-oxide, OGP: n-octyl glucopyranoside, LOAD: lauryldimethylamine N-oxide.

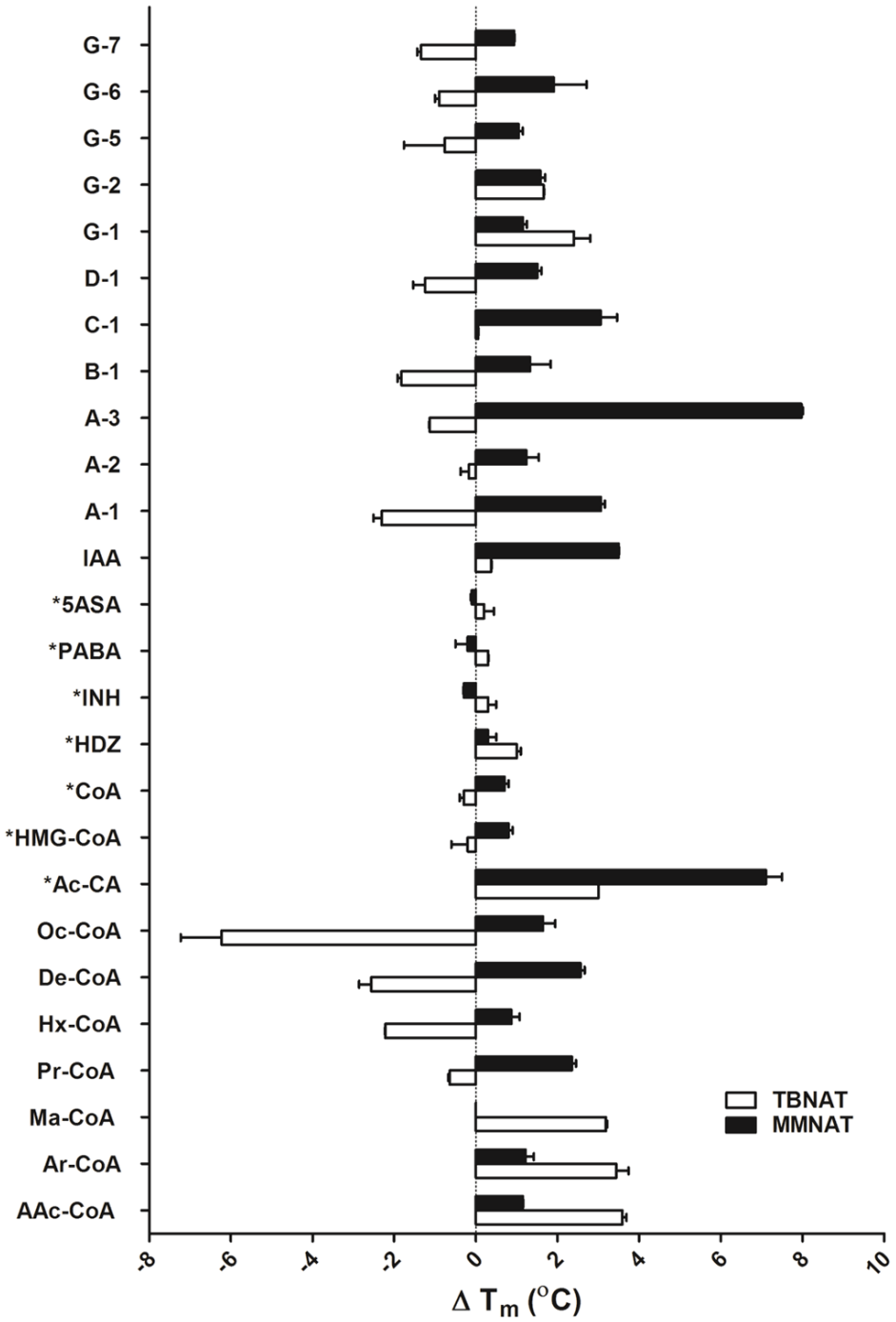


Figure 6.11: Changes in the TBNAT and MMNAT melting temperature in different additive solutions.

Changes in the unfolding transition temperature (ΔT_m) were calculated for measurements on TBNAT/MMNAT in the presence of the additives from the NAT-ligand library using TBNAT/MMNAT in 100 mM HEPES, pH 7.5, 150 mM NaCl as the reference. All solutions contained 100 mM HEPES buffer at pH 7.5, 5 % DMSO and 50 µM ligand concentration. Ligands with an asterisk were used without DMSO. The columns represent the mean ΔT_m values ± S.D. of duplicate measurements. A negative ΔT_m value signifies that the buffer destabilizes the protein, and a positive ΔT_m value indicates that the buffer has a stabilizing effect. IAA: iodoacetamide, AAc-CoA: acetoacetyl-CoA, Ar-CoA: arachidonoyl-CoA, Ma-CoA: malonyl-CoA, Hx-CoA: hexanoyl-CoA, De-CoA: decanoyl-CoA, Oc-CoA: octanoyl-CoA, HMG-CoA: hydroxymethylgluteryl-CoA, PABA: *p*-aminobenzoic acid, 5ASA: 5-aminosalicylic acid. Codes (A-G) correspond to inhibitors discussed in Chapter 5.

6.2.4 Multi-angle light scattering (SEC-MALS) and the influence of the His-tag

The hydrodynamic properties of proteins have a stronger influence on the success of crystallisation compared to protein stability (Price *et al.,* 2009). The lattice stability is governed by the protein surface properties, which control the formation of intermolecular contacts (Price *et al.,* 2009). The oligomeric state of both TBNAT and MMNAT with and without a His-tag was determined by size exclusion chromatography coupled with multi-angle light scattering detector (SEC-MALS) (Section 2.2.4.4).

The chromatogram from the gel filtration Superdex 200 SEC column in Figure 6.12 shows the abundance of a well-folded, active species with a molecular mass of 64 kDa which corresponds to a dimer of TBNAT. The results also suggested the presence of a higher order oligomeric state as observed by the presence of two minor peaks at 128 kDa and 195 kDa, which correspond to a tetramer and hexamer, respectively. The simultaneous measurement of the molecular mass by a MALS detector within each peak indicates that these higher oligomeric states undergo rapid interconversion. This is manifested as a gradient change of the measured molecular mass against the elution time within the hexamer and to a lower extent the tetramer peaks, in contrast with the flat line shown within the dimer peak (Figure 6.12A).

The presence of the enzyme in this dynamic oligomeric state is likely to represent a challenge, and to hamper attempts to crystallise this protein. In comparison, the more easily crystallisable MMNAT showed higher homogeneity, with most of the protein present as a dimer. MMNAT showed higher tendency to aggregate as shown by the large peak eluted at the void volume (Figure 6.13A). This aggregation is very likely to be a result of nonspecific hydrophobic interactions. MMNAT has a cysteine residue at the *C*-terminal of the protein that could be involved in intermolecular disulphide bridge formation. However, the presence of the His-tag in both proteins had a significant effect on

these profiles. The His-TBNAT showed a single peak corresponding to the mass of the dimer, whilst the MALS measurement showed a stable homogeneous species within the expected mass of the protein (Figure 6.12B). However, in the His-MMNAT, the presence of the His-tag resulted in a significant reduction in the aggregation of the protein (Figure 6.13B). These observations strongly suggest that the presence of the His-tag on both enzymes is very likely to improve the crystallisability of the proteins.

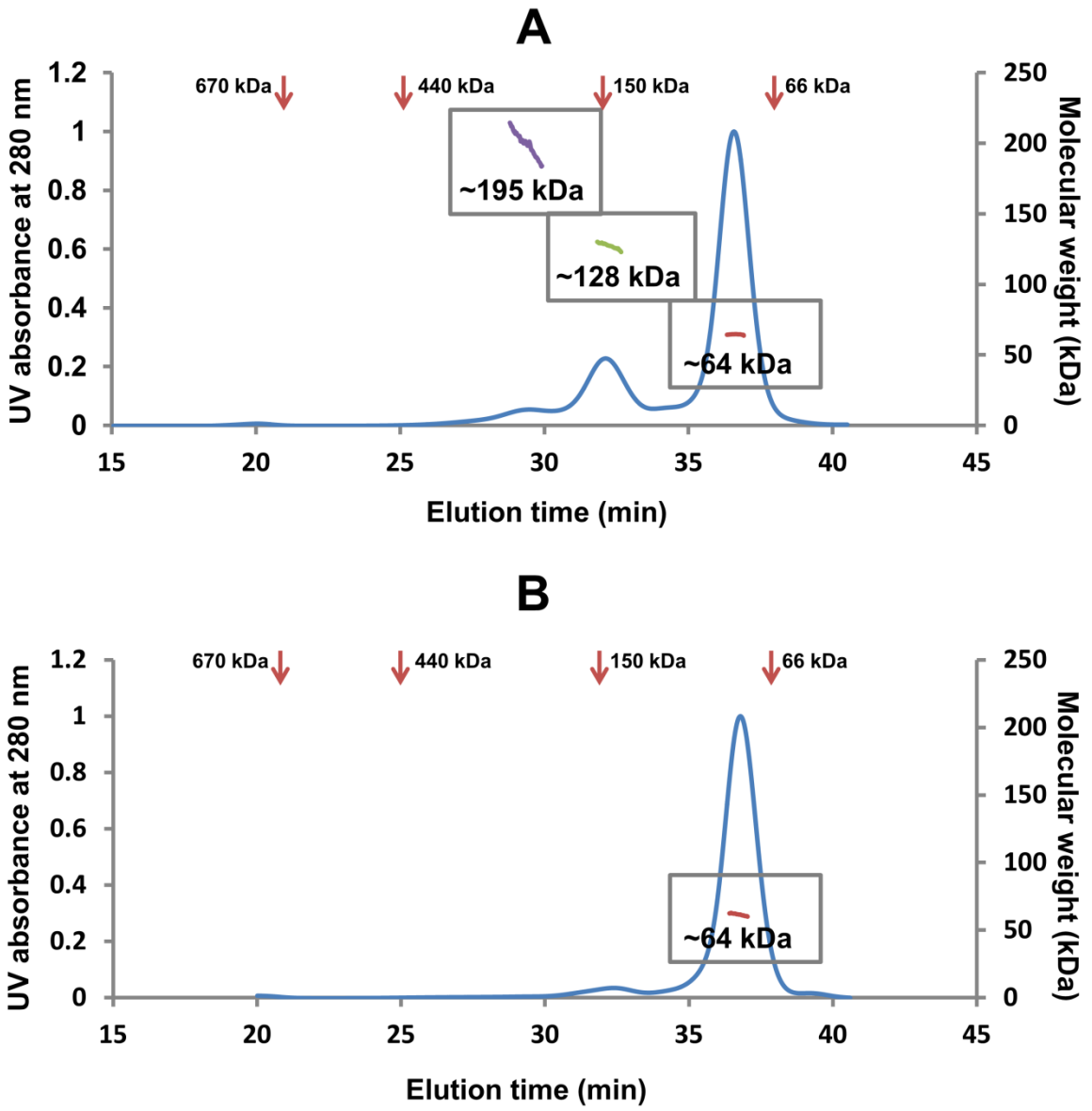


Figure 6.12: Representative multi-angle light scattering analyses of TBNAT and His-TBNAT.

Analyses of the (A) TBNAT and (B) His-TBNAT complexes by size exclusion chromatography followed by in-line multi-angle light scattering. The right axis represents the molecular mass at any given point in the elution profile. The segments of the curve within each peak are highlighted in a box and labeled with the average measured molecular mass. The elution profile of TBNAT is shown and demonstrates an abundant dimer and a lower quantity of tetramer and hexamer. The column volume is 23 mL (57 min), the void volume is 7.5 mL (18.75 min) and the peaks for molecular weight markers are indicated by the red arrows at the top of the graph.

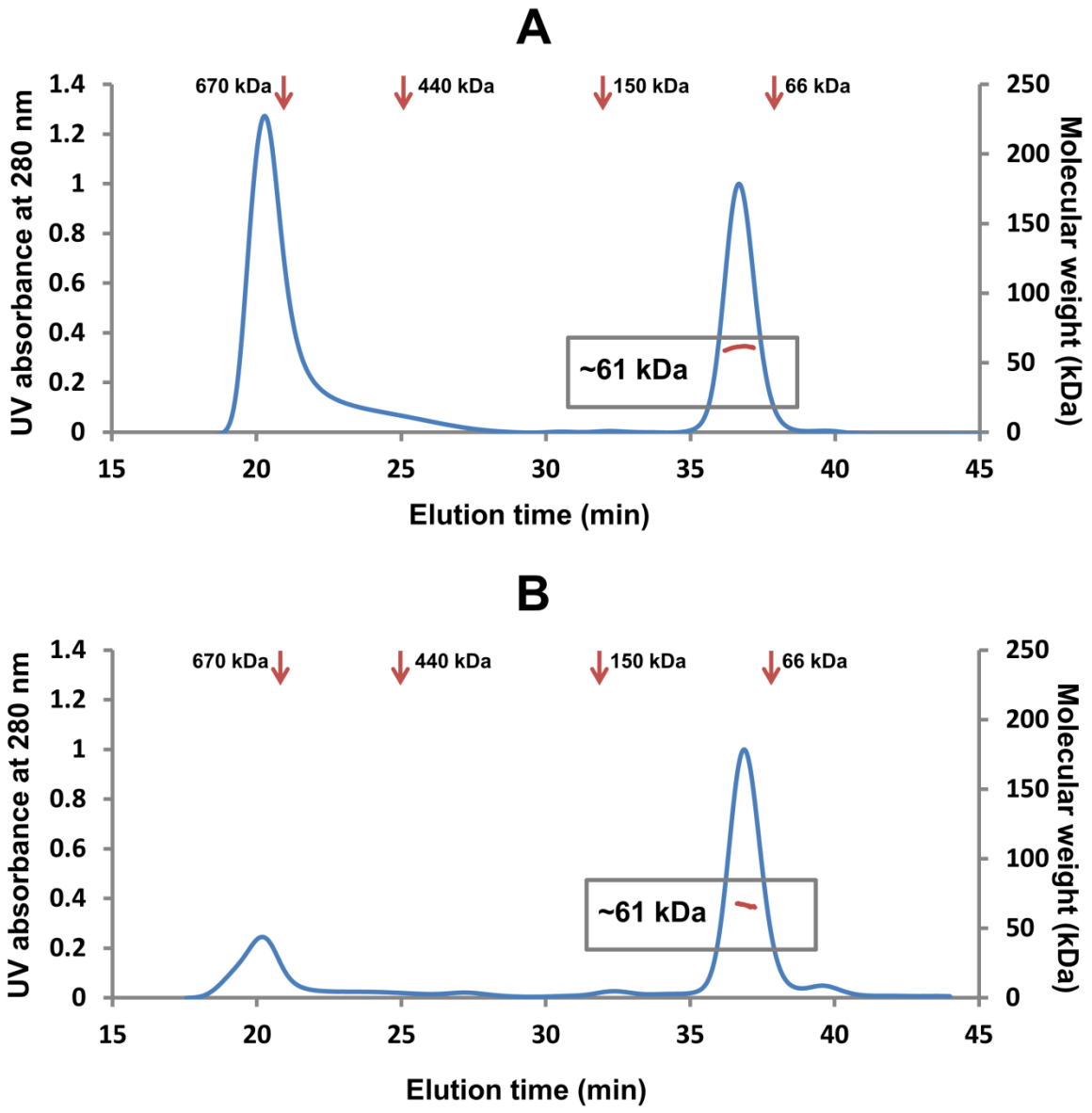


Figure 6.13: Representative multi-angle light scattering analyses of MMNAT and His-MMNAT.

Analyses of the (A) MMNAT and (B) His-MMNAT complexes by size exclusion chromatography followed by in-line multi-angle light scattering. The right axis represents the molecular mass at any given point in the elution profile. The segments of the curve within each peak are highlighted in a box and labeled with the average measured molecular mass. The column volume is 23 mL (57 min), the void volume is 7.5 mL (18.75 min) and the peaks for molecular weight markers are indicated by the red arrows at the top of the graph.

To obtain insights into the effect of the His-tag on the crystallisability of the proteins, the crystallisation of MMNAT with and without the His-tag was explored. The crystallization of MMNAT in the presence and absence of the His-tag was evaluated using the sitting-drop vapour diffusion technique. The screens were performed using the following sparse matrix screens: JCSG-plus, PACT, Structure screen I&II and morpheous (384 conditions) either with or without sodium azide in the protein stock (Section 2.2.8.3). Tetragonal bipyramidal crystals (Figure 6.14) were obtained in 15 different conditions when the His-MMNAT was used, while only one condition of the MMNAT resulted in diffraction quality crystals. The conditions in which His-MMNAT and MMNAT formed crystals are shown in Table 6.2. The His-MMNAT was at slightly lower molar concentration than the MMNAT (both proteins at approximately 10 mg/mL). This small difference is unlikely to be the cause of the improved number of conditions which gave crystals. These results suggested that the presence of the His-tag could significantly improve TBNAT crystallisation.

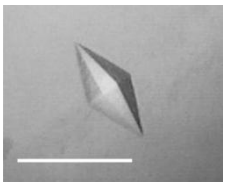


Figure 6.14: The tetragonal bipyramidal crystals of MMNAT.

Crystals grew in condition F8 of the PACT screen (0.2 M Na_2SO_4; 0.1 M Bis-Tris propane, pH 6.5; 20 % w/v PEG-3350). The scale bar represents 100 µm in length.

Table 6.2: The conditions in which MMNAT formed diffraction quality crystals.[a]

	Screen	Additive	Code	Conditions
His-MMNAT				
	St.S. I+II	NaN_3	F2	0.1 M NaCl 0.1 M HEPES, pH 7.5 1.6 M $(NH_4)_2SO_4$
		NaN_3	C9	0.2 M $MgCl_2$ 0.1 M Tris-HCl, pH 8.5 30 % w/v PEG-4000
	PACT	± NaN_3	F11	0.2 M Na-citrate 0.1 M Bis-Tris propane, pH 6.5 20 % w/v PEG-3350
		---	F8	0.2 M Na_2SO_4 0.1 M Bis-Tris propane, pH 6.5 20 % w/v PEG-3350
		---	G12	0.2 M Na-malonate 0.1 M Bis-Tris propane, pH 7.5 20 % w/v PEG-3350
		---	H12	0.2 M Na-malonate 0.1 M Bis-Tris propane, pH 8.5 20 % w/v PEG-3350
		---	F12	0.2 M Na-malonate 0.1 M Bis-Tris propane, pH 6.5 20 % w/v PEG-3350
	Morpheous	NaN_3	A12	0.06 M $MgCl_2$ + $CaCl_2$ 0.1 M Tris-Bicine, pH 8.5 37.5 % MPD + PEG-1000 + PEG-3350
		---	A8	0.06 M $MgCl_2$ + $CaCl_2$ 0.1 M HEPES + MOPS, pH 7.5 37.5 % MPD + PEG-1000 + PEG-3350
	JCSG -plus	± NaN_3	E2	0.2 M NaCl 0.1 M Na-cacodylate, pH 6.5 2.0 M $(NH_4)_2SO_4$
		± NaN_3	A2	0.1 M Citrate, pH 5.5 20 % w/v PEG-3000
		---	F10	1.1 M Na-malonate, pH 7.0 0.1 M HEPES, pH 7.0 0.5 % v/v Jeffamine ED-2001
MMNAT				
	JCSG -plus	---	E2	0.2 M NaCl 0.1 M Na-cacodylate, pH 6.5 2.0M $(NH_4)_2SO_4$

[a] The tetragonal bipyramidal shape (Figure 6.14) in the size range of 20-100 µm grew within 2-5 days of incubation at 20 ºC.

6.2.5 *In silico* prediction of disorder regions on TBNAT

The presence of disordered regions in the protein could explain the difficulties experienced in crystallising it. A sequence-based *in silico* prediction of the intrinsically disordered regions of the protein was obtained using the Oxford Protein Production Facility (OPPF) online tool RONN (Yang *et al.*, 2005) and compared with MMNAT. These comparisons allow an understanding of factors that hinder the crystallisability of the TBNAT.

The plot in Figure 6.15 shows a high probability of disorder for TBNAT in the second domain of the protein (137-158) and in the *C*-terminus (282-283) region. The figure also indicates a lower degree disorder in the same region in MMNAT (140-160) (Figure 6.15). This region is involved in the secondary binding pocket observed in the MMNAT-HLZ structure (2.1 Å, PDB code 3LTW) (Abuhammad *et al.*, 2010). Binding of a second HLZ molecule into a secondary pocket in the MMNAT-HLZ complex (Abuhammad *et al.*, 2010) may promote crystallisation. The disorder in the MMNAT structure extended to different regions of the protein (49-51, 103-109,) (Figure 6.15). Both regions are within loops of the structure that are usually associated with high B-factor and show changes in the differences for various MMNAT structures (Figure 3.8B). Figure 6.16 highlights the regions of MMNAT that were identified by using RONN to be disordered.

The successful co-crystallisation of MMNAT with CoA indicates that binding of CoA may help stabilise this region and promote crystal formation by stabilising the loop that makes up a flap of the binding pocket (103-109) (Figure 6.16). The presence of substrates has been reported to cause aggregation of NAT from pigeon liver suggesting that the formation of the active dimer of the protein was induced by substrates (Andres *et al.*, 1983). Furthermore, the recently solved structure of NAT from *Bacillus anthracis* showed endogenous CoA bound to the protein even though CoA was not included in the crystallisation solution (Pluvinage *et al.*, 2011).

The presence of these disordered regions is likely to hamper the crystallisation of the protein. Molecular engineering at those disordered regions represents a possibility towards the crystallisation of TBNAT in the future. Truncation of the *C*-terminal region of proteins has also been reported to improve crystallisation (Dale *et al.*, 2003). Since the structure is intended for the design of inhibitors, careful planning of any mutations is required to avoid changes in the catalytic profile of the enzyme. It has been shown that the *C*-terminus truncation induced the rate of acetyl-CoA hydrolysis even in the absence of arylamine substrate (Mushtaq *et al.*, 2002; Delgoda *et al.*, 2003). Even deletion of few residues from the *C*-terminus of a protein can have an effect on the activity, for example, the deletion of five residues results in reduction of the activity of the nitrite reductase from *Achromobacter cycloclastes*, while deleting 11 residues caused a complete loss of the biological activity (Li *et al.*, 2002). Moreover, the mutation of MSNAT in which Gly207 is replaced by Arg to simulate a naturally-occurring polymorphism, resulted in higher enzymic activity and changes in the structure of the binding pocket (Kawamura *et al.*, 2003).

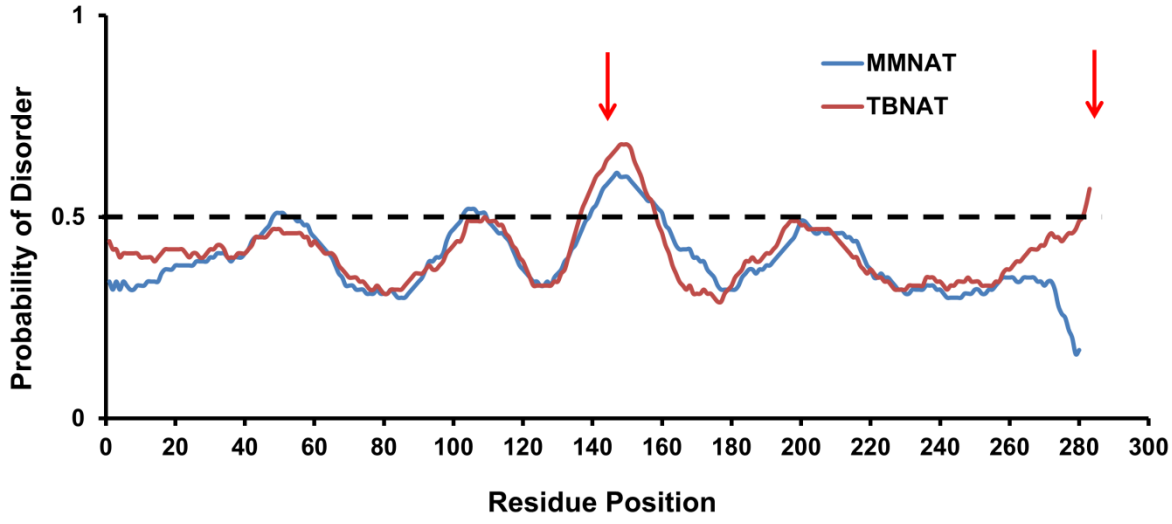


Figure 6.15: Plot of the predicted probability of disorder for each residues of TBNAT and MMNAT.

Probability values higher than 0.5 represent regions likely to be disordered. The red arrows indicate the disordered regions in TBNAT. The probabilities were predicted by RONN tool (http://www.strubi.ox.ac.uk/RONN) (Yang *et al.*).

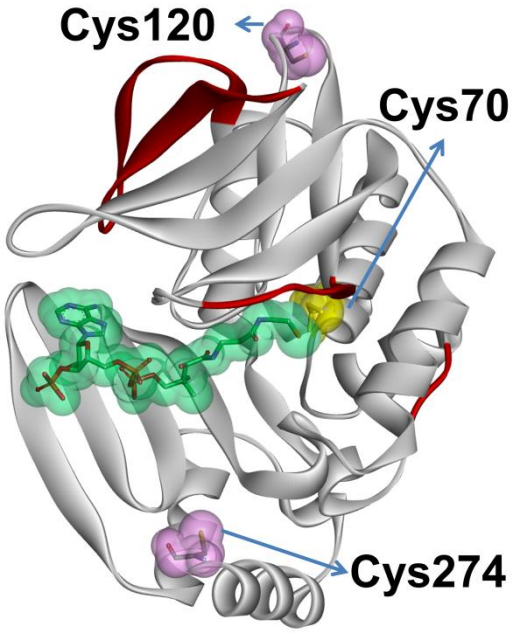


Figure 6.16: Ribbon representation of MMNAT shows the regions predicted to be disordered by RONN.

The structure of CoA-MMNAT complex shows the possible role of CoA (green) in stabilising the disordered flap of the protein (highlighted in red) (2.7 Å, PDB code 2VFC). The active site cysteine is shown in yellow, whilst other cysteine residues are shown in pink. The figure was prepared using the DS Visualizer.

6.2.6 Crystallisation of TBNAT in the presence of the ligands and/or the His-tag

Further rounds of TBNAT crystallisation trials were performed taking into consideration the knowledge gained from the stability studies and the biophysical analysis of the protein.

Crystallization conditions for TBNAT and His-TBNAT were tested using sitting-drop vapour diffusion techniques (Section 2.2.8.4). It was shown by the PCT test that a protein concentration of 10 mg/mL was suitable for the crystallisation of His-TBNAT. Screening was carried out using the commercially available sparse matrix screens: JCSG-plus and PACT Premier. The conditions which were explored are summarised in Table 6.3. Because of the qualitative nature of the TSA used for guidance, ligands that did not show positive thermal shifts but were known as high affinity substrates or inhibitors were also explored (HLZ, CoA, A-1, A-3). The 2-O-b-Mannosylglycerate (MSG) is a solute that accumulates

in some thermophilic organisms and has been reported as a thermostabiliser of proteins (Sawangwan *et al.*, 2010) (Figure 6.17), therefore, it was included in the screen. Attempts with HLZ and CoA mixture were expected to fill the binding pocket and offer stabilisation to the disordered regions in TBNAT. A similar effect was expected with the non-hydrolysable CoA derivatives malonyl-oxo(dethia)-CoA, malonyl-aza(dethia)-CoA and methylmalonyl-aza(dethia)-CoA (Figure 6.18). The presence of the different additives was associated with improved behaviour of the protein within the drops as indicated by a shift towards more increased gelatinous mild precipitate and less heavy precipitation and degradation. However, only a few crystals were obtained and submitted for data collection, and showed the diffraction corresponding to salt crystals (Figure 6.19). Other conditions in which salt crystals formed are listed in Appendix D. Conditions that may be suitable for future optimisation are summarised in Table 6.4.

Table 6.3: Additives added to the protein crystallisation solutions.

Class	Compounds	Concentration (mM)	Protein (10 mg/mL)
Substrates	HLZ	20	TBNAT
Inhibitor	A-1	5	TBNAT
	A-3	5	TBNAT
	A-12	5	TBNAT
	G-1	5	TBNAT
	G-2	5	TBNAT
	IAA	5	TBNAT
Cofactors	CoA	1	TBNAT
	CoA	10	TBNAT
	CoA:HLZ (1:1)	20	His-TBNAT
	AAc-CoA	1	TBNAT
	Ma-CoA	1	TBNAT
	MO-CoA	20	His-TBNAT
	MN-CoA	20	His-TBNAT
	MMN-CoA	20	His-TBNAT
Detergent	OGP	1	TBNAT
Reducing agents	DTT	1	TBNAT/His-TBNAT
Stabilisers	NaN_3	1	TBNAT/His-TBNAT
	MSG	50	His-TBNAT
	L-Histidine	50	His-TBNAT
	Arginien:glutamate (1:1)	50	His-TBNAT
Starting buffer	MES	20 / pH 6.5	His-TBNAT
	PIPES	20 / pH 7	His-TBNAT
	HEPES	20 / pH 7.5	His-TBNAT

MSG: 2-O-b-mannosylglycerate, MO-CoA: malonyl-oxo(dethia)-CoA, MN-CoA: malonyl-aza(dethia)-CoA, MMN-CoA: methylmalonyl-aza(dethia)-CoA.

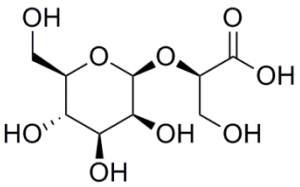
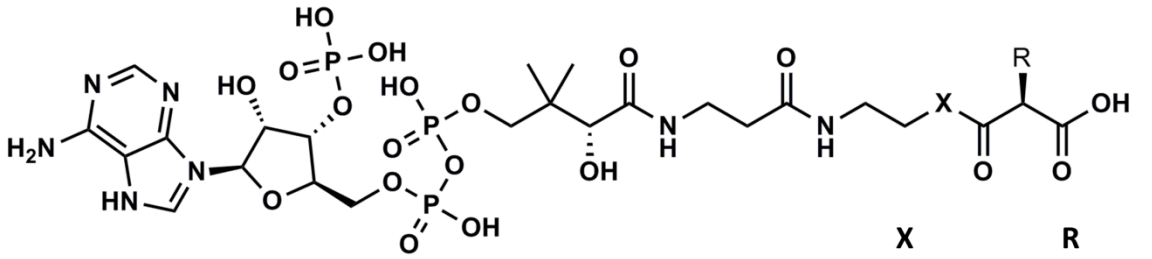


Figure 6.17: The 2-*O*-b-Mannosylglycerate.



	X	R
Ma-CoA	S	H
MO-CoA	O	H
MN-CoA	N	H
Mm-CoA	S	CH_3
MMN-CoA	N	CH_3

Figure 6.18: The non-hydrolysable CoA derivatives.

The structures of the non-hydrolysable CoA derivatives and their corresponding natural analogues. Ma: malonyl-CoA, MO-CoA: malonyl-oxo(dethia)-CoA, MN-CoA: malonyl-aza(dethia)-CoA, Mm-CoA: methylmalonyl-CoA, MMN-CoA: methylmalonyl-aza(dethia)-CoA.

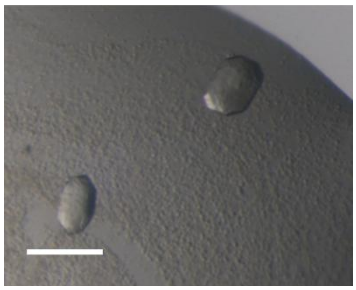
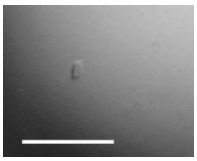
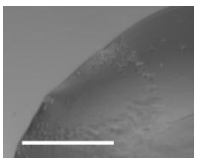
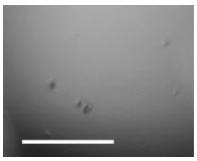
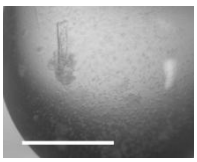
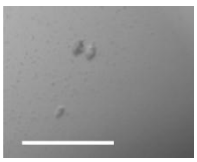
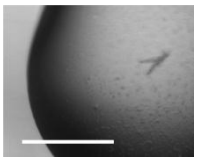


Figure 6.19: Crystals formed in the presence of 20 mM HLZ.

Crystals formed in condition B10 of the JCSG-plus screen, which contains 0.2 M $MgCl_2$, 0.1 M Na-cacodylate pH 6.5 and 50 % v/v PEG-200. The crystals gave a few extremely strong widely spaced diffraction peaks that are characteristic of salt diffraction patterns. The scale bar is equal to 100 µm.

Table 6.4: Summary of the conditions that might be suitbale for future optimisation.[a]

	Additive	Conditions	Code
JCSG-plus			
	---	0.1 M HEPES, pH 7.0 10 % w/v PEG-6000	C4
	---	0.005 M $CoCl_2$ 0.005 M $CdCl_2$ 0.005 M $MgCl_2$ 0.005 M $NiCl_2$ 0.1 M HEPES, pH 7.5 12 % w/v and PEG-3350	G5
	50 mM Firoin	0.1 M Tris, pH 8.5 20 % v/v Ethanol	C8
	50 mM Firoin	0.1 M Bicine, pH 9.0 10 % w/v PEG-20,000 2 % v/v Dioxane	C10
	50 mM Firoin	0.2 M Li_2SO_4 0.1 M Tris, pH 8.5 40 % v/v PEG-400	D7
PACT			
	50 mM Firoin	0.2 M Na-malonate 20 % w/v PEG-3350	E12

[a] the scale bar is equal to 100 µm.

6.2.7 Homology model of the TBNAT-HLZ structure

Homology modelling is a computational approach that is widely used to predict the 3D-structures of proteins from sequence when the experimental determination of the structure is impossible. The MMNAT-HLZ (PDB code 3LTW, 2.1 Å) was used as a template to simulate the HLZ binding pocket of the enzyme (Section 2.2.10.3). The resulting model has the same fold as MMNAT and the same orientation of the HLZ molecule in the binding pocket (Figure 6.20).

Examination of the orientation of HLZ in the binding pocket revealed that the location of HLZ in the binding pocket of TBNAT is very similar to that of MMNAT. The substitution of Val196 with Ala196 in TBNAT does provide more volume in the binding pocket compared to that available in MMNAT (Figure 6.21). The more extended binding pocket (indicated by the shaded areas in Figure 6.21) includes the piperidinol-binding pocket (in the vicinity of Met209 in MMNAT) and appears different in the two enzymes. The differences are as follows: the 3D-volume available for ligand interactions and in the electronic properties of the amino acid residues within this pocket (Figure 6.21). Volume was shown to be important in ligand binding (Figure 5.16). The "piperidinol-binding pocket" in TBNAT is wider than that of MMNAT and has hydrophilic character due to the substitution of the Leu98, His203, Met209 and Met222 with lysine, lysine, threonine and serine, respectively. The binding of the piperidinols to this binding pocket explains why they are less potent inhibitors of TBNAT compared to MMNAT (Table 5.2). Until the structure of the protein can be obtained experimentally, this visualisation of the binding pocket gives insight into the variability between the substrate profile and the inhibitor profile and can be used for structural drug design efforts against the TBNAT.

Figure 6.20: The homology model of the TBNAT-HLZ complex.

A ribbon presentation of the homology model of TBNAT-HLZ (green) structure aligned to the MMNAT-HLZ structure (PDB code 3LTW, white). The active site Cys70 is shown in yellow and the HLZ molecule in green. Main chain RMSD = 0.234 Å.

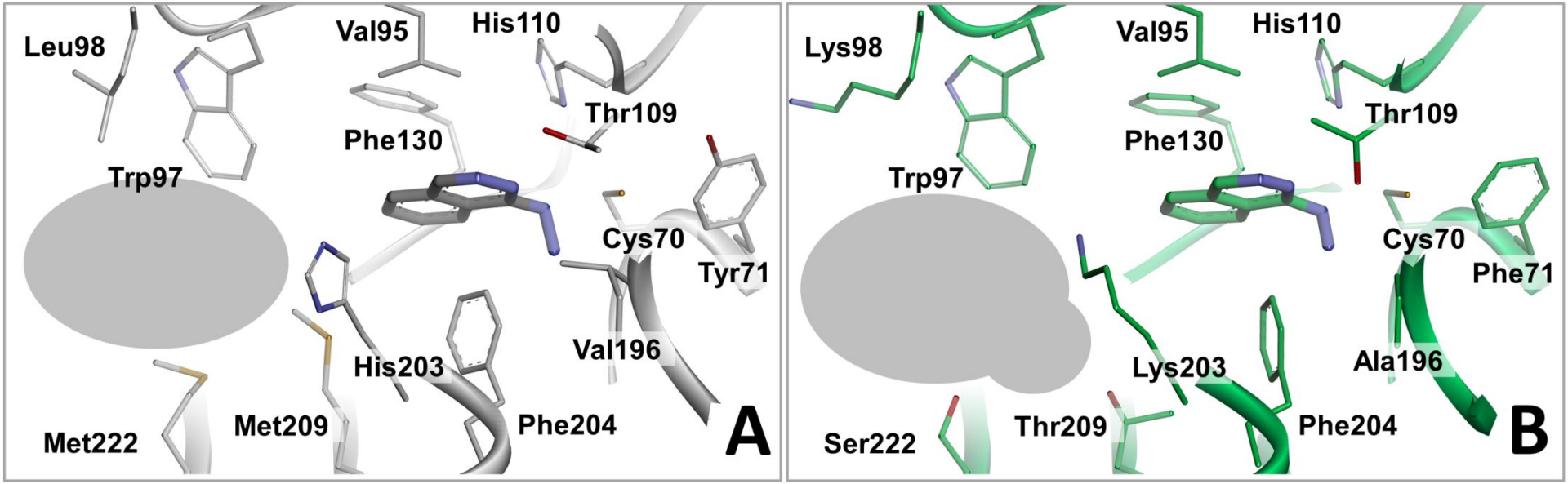


Figure 6.21: Comparison of the binding pockets of the MMNAT-HLZ crystal structure and TBNAT-HLZ homology model.

Comparison of the binding pockets of HLZ in the (A) MMNAT-HLZ complex (PDB code 3LTW) and (B) TBNAT homology model showing the HLZ binding pocket and the inhibitor binding pocket of the protein. The shaded areas indicate the proposed "piperidinol binding pocket". The key amino acids are highlighted.

6.3 Conclusions

The recombinant TBNAT yield with and without His-tag was sufficient to allow for extended screening for suitable crystallisation conditions to be undertaken. The variables screened in the initial round included the protein concentration, the incubation temperature and a wide range of the commercially available sparse matrix screens.

The TSA was applied to identify the effect of different stabilisation conditions that would be expected to enhance the crystallisability of the protein. Conditions, including the pH of the protein solution, the ionic strength and the possible ligands that can be included in the screens have been identified and compared to those that affect the stability of MMNAT.

A study of the effect of the His-tag on the behaviour of TBNAT in solution was also performed using the SEC-MALS analysis on the protein with and without the His-tag. This study showed that the presence of the His-tag resulted in an improved homogeneity of TBNAT, a criterion essential for the success of the crystallisation of proteins. It is very likely that crystallisation of the protein with a His-tag would increase the possibility of obtainging crystals of the protein. Comparison with MMNAT was performed, and showed that the His-tag improved the solubility of the MMNAT enzyme and increased the range of conditions in which crystals grew.

Since the experimental determination of the 3D-structure of the TBNAT protein was challenging due to the lack of crystals, a 3D-model of the protein was created by homology modelling using the MMNAT-HLZ complex as a template. This model allowed the visualisation of both the HLZ binding pocket and the proposed piperidinol binding pocket, and provided an explanation for the subtle differences between the substrate profile and the inhibitor profile of TBNAT with MMNAT.

Screening for suitable crystallisation conditions of TBNAT has been exhaustive but no crystals were obtained. The difficulties in crystallising proteins can be a major obstacle towards the 3D-structure determination. However, the current study gave better understanding of the properties of TBNAT and its stability in different solution conditions as well as the effect of the His-tag on its behaviour in solution. Thus, it helped narrow down the possibilities that can be explored in the future.

7 Conclusions and Future Work

7.1 Conclusions

Exploring TBNAT as a drug target was the major aim of this work. The extent to which the aims described in Section 1.6 have been met through the studies presented in this thesis are discussed below:

- **Increase the amount of *M. tuberculosis* NAT available (Chapter 4).**

  Sub-cloning of the *nat* gene in the broad-host expression vector pVLT31 resulted in the best achieved yield of the protein since the early attempts to produce it (Payton *et al.*, 1999). Efficient protocols to purify the recombinant enzyme with and without an affinity-tag were established and led to the production of the enzyme of a quality suitable for crystallographic studies and in yields between 6.5-16 mg per litre bacterial culture (Abuhammad *et al.*, 2011). This satisfactory yield of the soluble recombinant protein allowed extensive screening for suitable crystallisation conditions of TBNAT (Chapter 6). Although a structure could not be obtained, the studies performed have identified possible routes towards crystallising TBNAT. Considering techniques that would usually be used to crystallise membrane proteins is a promising possibility that has arisen from the association of TBNAT with lipid-contaminants and interaction with arachidonoyl-CoA (Chapters 4 and 6, respectively).

- **Probe the architecture of the catalytic pocket of MMNAT by generating a co-crystallised structure of MMNAT with a high affinity arylamine substrate (Chapter 3).**

  MMNAT was co-crystallised with hydralazine, the high affinity substrate of both MMNAT and TBNAT. The MMNAT-HLZ complex structure was obtained at a resolution of 2.1 Å (PDB code 3LTW). Analysis of the bound HLZ molecule allowed

mapping of the arylamine interactions with the residues in the binding pocket and shed light on the acetylation mechanism of this compound by MMNAT (Abuhammad *et al.*, 2010).

- **Compare the ligand specificity of MMNAT and TBNAT to identify high affinity ligands, which could be used for co-crystallisation studies (Chapter 3 and Chapter 5).**

  Evaluation of the enzyme's activity in the presence of variable acyl-CoA co-factors demonstrates that propionyl-CoA will also be hydrolysed by MMNAT in the presence of an acceptor substrate (Chapter 3) (Fullam *et al.*, 2009). The same finding has also been demonstrated for the TBNAT enzyme. In order to extend the scope of the range of chemical entities, which have been identified as NAT inhibitors, two different approaches were taken: (a) Ligand-based *in silico* screening, and (b) biophysical screening of a small group of candidate molecules. New inhibitors of the MMNAT were identified with moderate activity against TBNAT and against *M. bovis* BCG. The results obtained emphasise the unique specificity of the TBNAT enzyme (Chapter 5).

- **Explore the underlying mechanism of NAT inhibition by a group of inhibitors belonging to the piperidinol class (Chapter 5).**

  The multidisciplinary approach for investigating the mechanism of NAT inhibition revealed that these compounds could act as covalent inhibitors. Evidence towards a selective modification of the active site cysteine was obtained using mass-spectrometric analysis and crystallographic studies of MMNAT. However, all cysteines can be modified upon denaturation in the presence of molar excess of the inhibitor (Chapter 5). The results obtained suggested that only a fragment of the inhibitor (3-phenyl-3-oxopropyl) is involved in the covalent modification. Structural

studies on the inhibitor-binding pocket interactions, both experimental and computational, indicated that structural complementarity could be important in inhibitor binding (Chapter 5).

7.2 Future work

New questions and challenges have arisen from the findings of the studies on arylamine *N*-acetyltransferases described in this thesis.

Disappointingly, the structure of the TBNAT is still unknown. This reflects the wider challenge in the protein crystallography and structure determination field. The structure of an enzyme in a complex with a ligand is also needed. As discussed in Chapter 1, the structure-based drug design process is greatly aided by the characterisation of inhibitor binding (Hubbard, 2005). Researchers have encountered countless examples of proteins recalcitrant to crystallisation in their wild type (Price *et al.*, 2009; Qiu and Janson, 2009). Although many advances have been made in protein crystallography, better understanding of the crystallisation process is still required.

The understanding of ligand-enzyme interactions, as well as the identification of substrate- and inhibitor-binding pockets would facilitate the future targeting of TBNAT with medicinal chemistry research. The wide range of potency and selectivity exhibited by the identified NAT inhibitors makes them attractive candidates for further chemical studies. Since the yield of recombinant protein is not an issue, new techniques for inhibitor screening, which usually require high quantities of the protein (e.g. isothermal calorimetry) can now be considered. Although an engineered protein, either mutated or truncated, may seem a reasonable way to solve the 3D structure, new techniques for crystallising wild TBNAT can also be explored.

TBNAT, although a soluble protein, exhibited a considerable affinity for lipids. The presence of lipid contaminants in the recombinant protein preparation that is in agreement with

findings by Malen *et al.* (2010) (Chapter 4) are consistent with the possible role of the enzyme in virulence-lipid synthesis (Section 1.4.1). However, further research is required to identify the role of TBNAT in cell wall-lipid synthesis, which has been proposed, based on findings on *M. bovis* BCG by Bhakta *et al.* (2004). In addition, the ability of TBNAT to utilise Pr-CoA in the presence of arylamine substrate reveals a potential role as a link between two pathways that are important for *M. tuberculosis* persistence in the macrophage: cholesterol catabolism and virulence-lipid synthesis. Interestingly, the products of the neighbouring genes in the *nat*-operon: HsaA, HsaC and HsaD, which are involved in cholesterol degradation (Section 1.3.1) showed similar association with lipids in a whole-cell analysis (Malen *et al.*, 2010).

Progress in small-molecule, inhibitor or ligand discovery, accompanied by the 3D-structure determination, would help to elucidate the endogenous role of TBNAT in mycobacteria. It would also enable the continuation of the iterative medicinal chemistry syntheses and biological assessments (drug design) to which the work described in this thesis has contributed.

Appendices

Appendix A: List of the additives and buffers in the SGC-stability screen.

Well	Description
A01	100 mM Na-Acetate pH 5.0, 150 mM NaCl
A02	100 mM Pyridine pH 5.5, 150 mM NaCl
A03	100 mM Na-Cacodylate pH 6.0, 150 mM NaCl
A04	100 mM MES pH 6.0, 150 mM NaCl
A05	100 mM Na-Citrate pH 6.5, 150 mM NaCl
A06	100 mM Imidazole pH 7.0, 150 mM NaCl
A07	100 mM PIPES pH 7.0, 150 mM NaCl
A08	100 mM MOPS pH 7.2, 150 mM NaCl
A09	100 mM HEPES pH 7.5, 150 mM NaCl
A10	100 mM EPPS pH 8.0, 150 mM NaCl
A11	100 mM Tricine pH 8.0, 150 mM NaCl
A12	100 mM Bicine pH 8.3, 150 mM NaCl
B01	100 mM Glycine pH 9.0, 150 mM NaCl
B02	100 mM Borax pH 9.0, 150 mM NaCl
B03	100 mM CHES pH 9.5, 150 mM NaCl
B04	100 mM Ethanolamine pH 9.5, 150 mM NaCl
B05	100 mM CAPS pH 10.0, 150 mM NaCl
B06	700 mM Glycerol, 100 mM HEPES pH 7.5, 500 mM NaCl
B07	700 mM Glycerol, 100 mM HEPES pH 7.5, 500 mM NaCl
B08	700 mM Glycerol, 100 mM HEPES pH 7.5, 500 mM NaCl
B09	100 mM HEPES pH 7.5
B10	100 mM HEPES pH 7.5
B11	100 mM HEPES pH 7.5, 50 mM NaCl
B12	100 mM HEPES pH 7.5, 500 mM NaCl
C01	100 mM HEPES pH 7.5, 150mM LiCl
C02	100 mM HEPES pH 7.5, 150 mM KCl
C03	100 mM HEPES pH 7.5, 150mM NaI
C04	100 mM HEPES pH 7.5, 150 mM NaCN
C05	100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM NiCl2
C06	100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$
C07	100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $ZnCl_2$
C08	100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $CaCl_2$
C09	100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM CoCl2, 5 mM $CdCl_2$

C10	100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MnCl_2$
D01	100 mM HEPES pH 7.5, 50 mM NO_3
D02	100 mM HEPES pH 7.5, 50 mM NH_4Ac
D03	100 mM HEPES pH 7.5, 50 mM NH_4SO_4
D04	500 mM TMAO, 100 mM HEPES pH 7.5, 150 mM NaCl
D05	200 mM TMAO, 100 mM HEPES pH 7.5, 150 mM NaCl
D06	200 mM TMAO, 0.4 M urea, 100 mM HEPES pH 7.5, 150 mM NaCl
D07	400 mM urea, 100 mM HEPES pH 7.5, 150 mM NaCl
D08	800 mM Urea, 100 mM HEPES pH 7.5, 150 mM NaCl
D09	150 mM Guanidine-HCl, 100 mM HEPES pH 7.5, 150 mM NaCl
D10	500M Guanidine-HCl, 100 mM HEPES pH 7.5, 150 mM NaCl
D11	500 mM Betaine, 100 mM HEPES pH 7.5, 150 mM NaCl
D12	200 mM Betaine, 100 mM HEPES pH 7.5, 150 mM NaCl
E01	200 mM Betaine, 0.4 M urea, 100 mM HEPES pH 7.5, 150 mM NaCl
E02	100 mM Trimethylamine-HCl, 100 mM HEPES pH 7.5, 150 mM NaCl
E03	200 mM Sarcosine, 100 mM HEPES pH 7.5, 150 mM NaCl
E04	100 mM L-Alanine, 100 mM HEPES pH 7.5, 150 mM NaCl
E05	200 mM Choline, 100 mM HEPES pH 7.5, 150 mM NaCl
E06	200 mM Glycine, 100 mM HEPES pH 7.5, 150 mM NaCl
E07	500 mM Taurine, 100 mM HEPES pH 7.5, 150 mM NaCl
E08	500 mM L-Proline, 100 mM HEPES pH 7.5, 150 mM NaCl
E09	500 mM L-Histidine, 100 mM HEPES pH 7.5, 150 mM NaCl
E10	500 mM L-Arginine, 100 mM HEPES pH 7.5, 150 mM NaCl
E11	50 mM MIX: L-Arginine & L-Glutamate (1:1), 100 mM HEPES pH 7.5, 150 mM NaCl
E12	250 mM L-Phenylalanine,100 mM HEPES pH 7.5, 150 mM NaCl
F01	50 mM L-Cysteine, 100 mM HEPES pH 7.5, 150 mM NaCl
F02	50 mM L-Glutamine, 100 mM HEPES pH 7.5, 150 mM NaCl
F03	50 mM L-Glutamic acid, 100 mM HEPES pH 7.5, 150 mM NaCl
F04	0.5 mM LDAO, 100 mM HEPES pH 7.5, 150 mM NaCl
F05	50 mM L-Lysine, 100 mM HEPES pH 7.5, 150 mM NaCl
F06	50 mM Biotin, 100 mM HEPES pH 7.5, 150 mM NaCl
F07	5 mM TMP, 100 mM HEPES pH 7.5, 150 mM NaCl
F08	300 mM Guanosine, 100 mM HEPES pH 7.5, 150 mM NaCl
F09	300 mM Adenosine, 100 mM HEPES pH 7.5, 150 mM NaCl
F10	5 mM Pyridoxal-5'-phosphate, 100 mM HEPES pH 7.5, 150 mM NaCl
F11	0.4 mM CHAPS, 100 mM HEPES pH 7.5, 150 mM NaCl
G02	5 mM ATP, 5 GDP, 100 mM HEPES pH 7.5, 150 mM NaCl

G03	5 mM GTP, 100 mM HEPES pH 7.5, 150 mM NaCl
G06	10 mM FAD, 100 mM HEPES pH 7.5, 150 mM NaCl
G07	5 mM NAD, 100 mM HEPES pH 7.5, 150 mM NaCl
G08	5 mM NADP, 100 mM HEPES pH 7.5, 150 mM NaCl
G09	5 mM NADH, 100 mM HEPES pH 7.5, 150 mM NaCl
G10	5 mM NADPH, 100 mM HEPES pH 7.5, 150 mM NaCl
G11	2 mM Folic acid, 2 mM THF, 100 mM HEPES pH 7.5, 150 mM NaCl
G12	10 mM OGP, 100 mM HEPES pH 7.5, 150 mM NaCl
H01	0.1 mM DDM, 100 mM HEPES pH 7.5, 150 mM NaCl
H02	0.5 mM DDM, 100 mM HEPES pH 7.5, 150 mM NaCl
H03	100 mM D-Maltose, 100 mM HEPES pH 7.5, 150 mM NaCl
H04	300 mM Sucrose, 100 mM HEPES pH 7.5, 150 mM NaCl
H05	100 mM L-Ascorbic acid, 100 mM HEPES pH 7.5, 150 mM NaCl
H06	100 mM Myo-inositol, 100 mM HEPES pH 7.5, 150 mM NaCl
H07	300 mM Sorbitol, 100 mM HEPES pH 7.5, 150 mM NaCl
H10	1 mM GSSG, 2 mM GSH, 100 mM HEPES pH 7.5, 150 mM NaCl

Appendix B: The mathematical formulae for the different R factors

$R_{merge} = \sum_h \sum_j \left| I_{h_j} - \langle I_h \rangle \right| / \sum_h \sum_j \langle I_h \rangle$, where I_{h_j} is the intensity of the j^{th} observation of unique reflection with indices h (Evans, 2006).

$R_{p.i.m.} = \sum_h \left(\frac{1}{n_h - 1}\right) \sum_j \left| I_{h_j} - \langle I_h \rangle \right| / \sum_h \sum_j \langle I_h \rangle$, where I_{h_j} is the intensity of the j^{th} observation of unique reflection with indices h (Evans, 2006).

$R_{meas} = \sum_h \left(\frac{n}{n_h - 1}\right) \sum_j \left| I_{h_j} - \langle I_h \rangle \right| / \sum_h \sum_j \langle I_h \rangle$ where I_{h_j} is the intensity of the j^{th} observation of unique reflection with indices h (Diederichs and Karplus, 1997)

$$R_{cryst} = \sum_{hkl} |F_o - F_c| \Big/ \sum_{hkl} |F_o|$$

R_{Free} is R_{cryst} calculated using the 5 % of data that were excluded from the refinement.

Appendix C: The mathematical derivation of Kitz-Wilson model

According to the Kitz- Wilson model (Kitz and Wilson, 1962), covalent inactivation can be described in a two-step model of a reversible binding event followed by an irreversible pseudo-first-order reaction as follows:

$$E+I \underset{k_{-I}}{\overset{k_I}{\rightleftharpoons}} [EI] \xrightarrow{k_{inact}} E\text{-}I$$

In the first step, the inhibitor *[I]* binds the active enzyme *E* forming the reversible complex *[EI]*, and then the binding matures into a covalent modification resulting in irreversible inactivation of the enzyme *[E-I]*. The k_{inact} represents the limiting rate for time-dependent inhibition, which is slow relative to catalytic turnover.

At any time point, the total amount of enzyme in the reaction mixture $[E^o]$ can be described by the following relationship:

$$[E^o] = [E] + [EI] + [E\text{-}I] \quad (1)$$

The active enzyme species can be defined by *É ([E]+[EI])*, and hence E^o becomes:

$$[E^o] = [\acute{E}] + [E\text{-}I] \quad (2)$$

The rate of the reversible reaction is defined by:

$$\frac{[I][E]}{[EI]} = K_I \quad (3)$$

$$K_I = (k_{-I}/k_I) \quad (4)$$

whilst the rate of the irreversible inactivation is

$$\frac{d[\acute{E}]}{dt} = -k_{inact}[EI] \quad (5)$$

If the enzyme-inhibitor solution was extensively diluted before the assay, the activity would correspond to the concentration of *É*. The time-dependent inactivation can be described by solving the rate equation (5):

$$ln\frac{[\acute{E}]}{[E^o]} = -\frac{k_{inact} \times t}{1 + {K_I}/{[I]}} \qquad (6)$$

where *[É]/[E^o]* can be estimated from the residual enzyme activity after incubation with the inhibitor *[I]* for time interval (*t*). In accordance with equation (6), a plot of the natural logarithm of the residual enzyme activity against incubation time will result in a straight line, the slope of which gives a measurement of the apparent rate constant of the pseudo-first rate inactivation reaction, k_{obs}.

$$-slope = k_{obs} = \frac{k_{inact}}{1 + {K_I}/{[I]}} \qquad (7)$$

Appendix D: Conditions yielding salt crystals in screening for crystallisation conditions for TBNAT .[a]

Screen	Code	Crystal shape	Condition
Index	B1	plates	0.1 M Bis-Tris pH 5.5; 0.3 M Mg-formate.
	C7	plete cluster	0.8 K-Na-tartartate; 0.1 M Tris pH 8.5; 0.5 % w/v PEG-MME-5000.
	D11	plates	0.1 M Bis-Tris pH 6.5; 28 % w/v PEG-MME-2000.
	D12	plates	0.2 M $CaCl_2$; 0.1 M Bis-Tris pH 5.5; 45 % w/v MPD.
	E1	plates	0.2 M $CaCl_2$; 0.1 M Bis-Tris pH 6.5; 45 % w/v MPD.
	E11	plates	0.02 M $MgCl_2$; 0.1 M HEPES pH 7.5; 22 % w/v poly(acrylic acid sodium salt)-5100.
	E7	plates	0.05 M $MgCl_2$; 0.1 M HEPES pH 7.5; 30 % v/v PEG-MME-550.
	E9	cubes	0.05 M $(NH_4)_2SO_4$; 0.05 M Bis-Tris pH 6.5; 30 % v/v pentaerythritol ethoxylate (15/4 EO/OH).
PEG/Ion	A4	plate	0.2 M LiCl; pH 6.8; 20 % w/v PEG-3350.
	A7	plate	0.2 M $CaCl_2$, pH 5.1; 20 % w/v PEG-3350.
	C7	plates	0.2 M Li_2SO_4; pH 6.0; 20 % w/v PEG-3350.
	D1	plates	0.2 M K-Na-tartrate; pH 7.4; 20 % w/v PEG-3350.
	E7	plate	0.1 M Na-malonate pH 7.0; 12 % w/v PEG-3350.
	E9	cubes	4 % v/v Tacsimate pH 4.0; 12 % w/v PEG-3350.
JCSG	B10	hexagonal prisms	0.2 M $MgCl_2$; 0.1 M Na-cacodylate pH 6.5; 50 % v/v PEG-200.
	C1	hexagonal prisms	0.2M NaCl; 0.1M phosphate/citrate pH 4.2; 20 % w/v PEG-8000.
	D12	plates	0.04M KH_2PO_4;16 % w/v PEG-8000; 20 % v/v glycerol.
	D2	plates	0.2M $MgCl_2$; 0.1M Na-Hepes pH 7.5; 30 % v/v PEG-400.
St. S I+II	1-38	plate cluster	0.1 M Tris pH 8.5; 8 % w/v PEG-8000.
	1-48	plates	0.2 M Mg-formate
	2-17	plates	0.1 M HEPES pH 7.5; 70 % v/v MPD.

[a] The screen in each case was performed using a 10 mg/mL solution of TBNAT without a His-tag. Tacsimate is a mixture of titrated organic acid salts containing 1.8305 M malonic acid, 0.25 M ammonium citrate tribasic, 0.12 M succinic acid, 0.3 M DL-malic acid, 0.4 M Na-acetate trihydrate, 0.5 M Na-formate, and 0.16 M ammonium tartrate dibasic1. PEG-MME: Polyethylene glycol monomethyl ether; MPD: (+/-)-2-methyl-2,4-pentanediol, St. S I+II: Structure Screen I+II.

References

Abuhammad, A, Lack, N, Schweichler, J, Staunton, D, Sim, RB, Sim, E (2011). Improvement of the expression and purification of *Mycobacterium tuberculosis* arylamine *N*-acetyltransferase (TBNAT) a potential target for novel anti-tubercular agents. *Protein Expression and Purification* 80(2)*:*246-252.

Abuhammad, AM, Lowe, ED, Fullam, E, Noble, M, Garman, EF, Sim, E (2010). Probing the architecture of the *Mycobacterium marinum* arylamine *N*-acetyltransferase active site. *Protein & Cell* 1(4)*:*384-392.

Accelrys, L (2012).DS Visualizer and ActiveX Control 3.5.

Ahmad, Z, Sharma, S, Khuller, GK (2006a). The potential of azole antifungals against latent/persistent tuberculosis. *FEMS Microbiology Letters* 258(2)*:*200-203.

Ahmad, Z, Sharma, S, Khuller, GK, Singh, P, Faujdar, J, Katoch, VM (2006b). Antimycobacterial activity of econazole against multidrug-resistant strains of *Mycobacterium tuberculosis*. *International Journal of Antimicrobial Agents* 28(6)*:*543-544.

Allen, JL (1992). A modified Ziehl-Neelsen stain for mycobacteria. *Medical Laboratory Sciences* 49(2)*:*99-102.

Altaf, M, Miller, CH, Bellows, DS, O'Toole, R (2010). Evaluation of the *Mycobacterium smegmatis* and BCG models for the discovery of *Mycobacterium tuberculosis* inhibitors. *Tuberculosis* 90(6)*:*333-337.

Altschul, SF, Gish, W, Miller, W, Myers, EW, Lipman, DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215(3)*:*403-410.

Anderton, MC, Bhakta, S, Besra, GS, Jeavons, P, Eltis, LD, Sim, E (2006). Characterization of the putative operon containing arylamine *N*-acetyltransferase (nat) in *Mycobacterium bovis* BCG. *Molecular Microbiology* 59(1)*:*181-192.

Andres, HH, Kolb, HJ, Weiss, L (1983). Purification and physical-chemical properties of acetyl-CoA: arylamine N-acetyltransferase from pigeon liver. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 746(3)*:*182-192.

Andries, K, Verhasselt, P, Guillemont, J, Gohlmann, HW, Neefs, JM, Winkler, H, Van Gestel, J, Timmerman, P, Zhu, M, Lee, E, Williams, P, de Chaffoy, D, Huitric, E, Hoffner, S, Cambau, E, Truffot-Pernot, C, Lounis, N, Jarlier, V (2005). A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 307(5707)*:*223-227.

Anes, E, Kuhnel, MP, Bos, E, Moniz-Pereira, J, Habermann, A, Griffiths, G (2003). Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria. *Nature Cell Biology* 5(9)*:*793-802.

Arcus, VL, Lott, JS, Johnston, JM, Baker, EN (2006). The potential impact of structural genomics on tuberculosis drug discovery. *Drug Discovery Today* 11(1-2)*:*28-34.

Armstrong, MS, Morris, GM, Finn, PW, Sharma, R, Moretti, L, Cooper, RI, Richards, WG (2010). ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *Journal of Computer-Aided Molecular Design* 24(9)*:*789-801.

Astarie-Dequeker, C, Le Guyader, L, Malaga, W, Seaphanh, FK, Chalut, C, Lopez, A, Guilhot, C (2009). Phthiocerol dimycocerosates of *M. tuberculosis* participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. *PLoS Pathogens* 5(2)*:*e1000289.

Babine, RE, Bender, SL (1997). Molecular recognition of protein−ligand complexes: applications to drug design. *Chemical Reviews* 97(5)*:*1359-1472.

Bailey, S (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* 50(5)*:*760-763.

Barry, C (2001). Interpreting cell wall 'virulence factors' of *Mycobacterium tuberculosis*. *Trends in Microbiology* 9(5)*:*237-241.

Besra, GS, Morehouse, CB, Rittner, CM, Waechter, CJ, Brennan, PJ (1997). Biosynthesis of Mycobacterial Lipoarabinomannan. *Journal of Biological Chemistry* 272(29)*:*18460-18466.

Beste, DJV, Espasa, M, Bonde, B, Kierzek, AM, Stewart, GR, McFadden, J (2009). The Genetic Requirements for Fast and Slow Growth in Mycobacteria. *PLoS ONE* 4(4)*:*e5349.

Bhakta, S, Besra, GS, Upton, AM, Parish, T, Sholto-Douglas-Vernon, C, Gibson, KJ, Knutton, S, Gordon, S, DaSilva, 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 Journal of Experimental Medicine* 199(9)*:*1191-1199.

Bouley, DM, Ghori, N, Mercer, KL, Falkow, S, Ramakrishnan, L (2001). Dynamic nature of host-pathogen interactions in *Mycobacterium marinum* granulomas. *Infection and Immunity* 69(12)*:*7820-7831.

Bradford, MMM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72(248-254.

Brennan, PJ (2003). Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis* 83(1-3)*:*91-97.

Brooke, EW, Davies, SG, Mulvaney, AW, Okada, M, Pompeo, F, Sim, E, Vickers, RJ, Westwood, IM (2003a). Synthesis and in vitro evaluation of novel small molecule inhibitors of bacterial arylamine *N*-acetyltransferases (NATs). *Bioorganic & Medicinal Chemistry Letters* 13(15)*:*2527-2530.

Brooke, EW, Davies, SG, Mulvaney, AW, Pompeo, F, Sim, E, Vickers, RJ (2003b). An approach to identifying novel substrates of bacterial arylamine *N*-acetyltransferases. *Bioorganic & Medicinal Chemistry* 11(7)*:*1227-1234.

Butcher, PD (2004). Microarrays for *Mycobacterium tuberculosis*. *Tuberculosis* 84(3-4)*:*131-137.

Butler, AR, Gold, V (1961). The hydrolysis of acetic anhydride. Part VII. Catalysis by pyridine and methylpyridines in acetate buffers. *Journal of the Chemical Society:*4362-4367.

Caceres, N, Tapia, G, Ojanguren, I, Altare, F, Gil, O, Pinto, S, Vilaplana, C, Cardona, PJ (2009). Evolution of foamy macrophages in the pulmonary granulomas of experimental tuberculosis models. *Tuberculosis (Edinburgh)* 89(2)*:*175-182.

Cardona, P-J, Ruiz-Manzano, J (2004). On the nature of Mycobacterium tuberculosis-latent bacilli. *European Respiratory Journal* 24(6)*:*1044-1051.

Carroll, MV, Lack, N, Sim, E, Krarup, A, Sim, RB (2009). Multiple routes of complement activation by *Mycobacterium bovis* BCG. *Molecular Immunology* 46(16)*:*3367-3378.

Chan, J, Fan, XD, Hunter, SW, Brennan, PJ, Bloom, BR (1991). Lipoarabinomannan, a possible virulence factor involved in persistence of *Mycobacterium tuberculosis* within macrophages. *Infection and Immunity* 59(5)*:*1755-1761.

Chung, A, Liu, Q, Ouyang, S-P, Wu, Q, Chen, G-Q (2009). Microbial production of 3-hydroxydodecanoic acid by <i>pha operon and <i>fadBA knockout mutant of <i>Pseudomonas putida KT2442 harboring <i>tesB gene. *Applied Microbiology and Biotechnology* 83(3)*:*513-519.

Coelho, MB, Costa, ERD, Vasconcellos, SEG, Linck, N, Ramos, RM, Amorim, HLNd, Suffys, PN, Santos, AR, Silva, PEAd, Ramos, DF, Silva, MSN, Rossetti, MLR (2011). Sequence and structural characterization of tbnat gene in isoniazid-resistant *Mycobacterium tuberculosis*: Identification of new mutations. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 712(1-2)*:*33-39.

Cohen, SA, Michaud, DP (1993). Synthesis of a fluorescent derivatizing reagent, 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate, and its application for the analysis of hydrolysate amino acids via high-performance liquid chromatography. *Analytical Biochemistry* 211(2)*:*279-287.

Colditz, GA, Brewer, TF, Berkey, CS, Wilson, ME, Burdick, E, Fineberg, HV, Mosteller, F (1994). Efficacy of BCG vaccine in the prevention of tuberculosis. *The Journal of the American Medical Association* 271(9)*:*698-702.

Cole, ST, Brosch, R, Parkhill, J, Garnier, T, Churcher, C, Harris, D, Gordon, SV, Eiglmeier, K, Gas, S, Barry, CE, Tekaia, F, Badcock, K, Basham, D, Brown, D, Chillingworth, T, Connor, R, Davies, R, Devlin, K, Feltwell, T, Gentles, S, Hamlin, N, Holroyd, S, Hornsby, T, Jagels, K, Krogh, A, McLean, J, Moule, S, Murphy, L, Oliver, K, Osborne, J, Quail, MA, Rajandream, MA, Rogers, J, Rutter, S, Seeger, K, Skelton, J, Squares, R, Squares, S, Sulston, JE, Taylor, K, Whitehead, S, Barrell, BG (1998).

Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393(6685)*:*537-544.

Cole, ST, Riccardi, G (2011). New tuberculosis drugs on the horizon. *Current Opinion in Microbiology* 14(5)*:*570-576.

Collaborative (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography* 50(5)*:*760-763.

Copeland, RA (2005). Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists. Hoboken, NJ: John Wiley & Sons, Inc.

Corner, LAL, Murphy, D, Gormley, E (2011). *Mycobacterium bovis* Infection in the Eurasian Badger (Meles meles): the Disease, Pathogenesis, Epidemiology and Control. *Journal of Comparative Pathology* 144(1)*:*1-24.

Daffé, M, Draper, P (1997). The envelope layers of *Mycobacteria* with reference to their pathogenicity. In: Advances in Microbial Physiology. Poole, RK editor: Academic Press, pp. 131-203.

Dale, GE, Oefner, C, D'Arcy, A (2003). The protein as a variable in protein crystallization. *Journal of Structural Biology* 142(1)*:*88-97.

Daniel, TM (2006). The history of tuberculosis. *Respiratory Medicine* 100(11)*:*1862-1870.

Datta, VN, Brien, RD, Kathleen, MM, Stephen, GS (1976). Fluorescence of hydralazine in concentrated sulfuric acid. *Journal of Pharmaceutical Sciences* 65(2)*:*274-276.

Davioud-Charvet, E, McLeish, MJ, Veine, DM, Giegel, D, Arscott, LD, Andricopulo, AD, Becker, K, Muller, S, Schirmer, RH, Williams, CH, Jr., Kenyon, GL (2003). Mechanism-based inactivation of thioredoxin reductase from *Plasmodium falciparum* by Mannich bases. Implication for cytotoxicity. *Biochemistry* 42(45)*:*13319-13330.

Davis, IW, Leaver-Fay, A, Chen, VB, Block, JN, Kapral, GJ, Wang, X, Murray, LW, Arendall, WB, III, Snoeyink, J, Richardson, JS, Richardson, DC (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl Acids Res* 35(2)*:*W375-383.

de Lorenzo, V, Eltis, L, Kessler, B, Timmis, KN (1993). Analysis of *Pseudomonas* gene products using lacIq/Ptrp-lac plasmids and transposons that confer conditional phenotypes. *Gene* 123(1)*:*17-24.

Delgoda, R, Lian, LY, Sandy, J, Sim, E (2003). NMR investigation of the catalytic mechanism of arylamine *N*-acetyltransferase from *Salmonella typhimurium*. *Biochimica et Biophysica Acta* 1620(1-3)*:*8-14.

Diacon, AH, Pym, A, Grobusch, M, Patientia, R, Rustomjee, R, Page-Shipp, L, Pistorius, C, Krause, R, Bogoshi, M, Churchyard, G, Venter, A, Allen, J, Palomino, JC, De Marez, T, van Heeswijk, RPG, Lounis, N, Meyvisch, P, Verbeeck, J, Parys, W, de Beule, K, Andries, K, Neeley, DFM (2009). The diarylquinoline TMC207 for multidrug-resistant tuberculosis. *New England Journal of Medicine* 360(23)*:*2397-2405.

Diederichs, K, Karplus, PA (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology* 4(4)*:*269-275.

Donald, PR, Sirgel, FA, Venter, A, Parkin, DP, Seifart, HI, Wal, BWvd, Werely, C, Helden, PDv, Maritz, JS (2004). The Influence of Human N-Acetyltransferase Genotype on the Early Bactericidal Activity of Isoniazid. *Clinical Infectious Diseases* 39(10)*:*1425-1430.

Dorman, SE, Chaisson, RE (2007). From magic bullets back to the Magic Mountain: the rise of extensively drug-resistant tuberculosis. *Natural Medicines* 13(3)*:*295-298.

Dresen, C, Lin, LY, D'Angelo, I, Tocheva, EI, Strynadka, N, Eltis, LD (2010). A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism. *Journal of Biological Chemistry* 285(29)*:*22264-22275.

Dye, C, Williams, BG (2010). The population dynamics and control of tuberculosis. *Science* 328(5980)*:*856-861.

Ehrt, S, Schnappinger, D (2007). *Mycobacterium tuberculosis* virulence: lipids inside and out. *Natural Medicines* 13(3)*:*284-285.

Eisenberg, D, McLachlan, AD (1986). Solvation energy in protein folding and binding. *Nature* 319(6050)*:*199-203.

Emsley, P, Lohkamp, B, Scott, WG, Cowtan, K (2010). Features and development of Coot. *Acta Crystallographica Section D, Biological Crystallography* 66(Pt 4)*:*486-501.

Ericsson, UB, Hallberg, BM, DeTitta, GT, Dekker, N, Nordlund, P (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry* 357(2)*:*289-298.

Ernst, JD (1998). Macrophage Receptors for Mycobacterium tuberculosis. *Infection and Immunity* 66(4)*:*1277-1281.

Eswar, N, Webb, B, Marti-Renom, MA, Madhusudhan, MS, Eramian, D, Shen, M-y, Pieper, U, Sali, A (2007). Comparative protein structure modeling using MODELLER. In: Current Protocols in Protein Science: John Wiley & Sons, Inc.

Evans, P (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography* 62(1)*:*72-82.

Evrard, GX, Langer, GG, Perrakis, A, Lamzin, VS (2007). Assessment of automatic ligand building in ARP/wARP. *Acta Crystallographica Section D* 63(1)*:*108-117.

Falkinham, JOr (2002). Nontuberculous mycobacteria in the environment. *Clinics in Chest Medicine (Philadelphia, PA)* 23(3)*:*529-551.

Fenton, MJ, Vermeulen, MW (1996). Immunopathology of tuberculosis: roles of macrophages and monocytes. *Infection and Immunity* 64(3)*:*683-690.

Ferrari, G, Langen, H, Naito, M, Pieters, J (1999). A Coat Protein on Phagosomes Involved in the Intracellular Survival of Mycobacteria. *Cell* 97(4)*:*435-447.

Flannagan, RS, Cosio, G, Grinstein, S (2009). Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nature Reviews Microbiology* 7(5)*:*355-366.

Fullam, E (2007). Arylamine *N*-acetyltransferase of *Mycobacteria* (DPil Thesis). Oxford, Oxford University.

Fullam, E, Westwood, IM, Anderton, MC, Lowe, ED, Sim, E, Noble, ME (2008). Divergence of cofactor recognition across evolution: coenzyme A binding in a prokaryotic arylamine *N*-acetyltransferase. *Journal of Molecular Biology* 375(1)*:*178-191.

Fullam, E, Kawamura, A, Wilkinson, H, Abuhammad, A, Westwood, I, Sim, E (2009). Comparison of the arylamine *N*-acetyltransferase from *Mycobacterium marinum* and *Mycobacterium tuberculosis*. *The Protein Journal* 28(6)*:*281-293.

Fullam, E, Abuhammad, A, Wilson, DL, Anderton, MC, Davies, SG, Russell, AJ, Sim, E (2011). Analysis of beta-amino alcohols as inhibitors of the potential anti-tubercular target *N*-acetyltransferase. *Bioorganic & Medicinal Chemistry Letters* 21(4)*:*1185-1190.

Gallivan, JP, Dougherty, DA (1999). Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America* 96(17)*:*9459-9464.

Garnier, T, Eiglmeier, K, Camus, J-C, Medina, N, Mansoor, H, Pryor, M, Duthoy, S, Grondin, S, Lacroix, C, Monsempe, C, Simon, S, Harris, B, Atkin, R, Doggett, J, Mayes, R, Keating, L, Wheeler, PR, Parkhill, J, Barrell, BG, Cole, ST, Gordon, SV, Hewinson, RG (2003). The complete genome sequence of *Mycobacterium bovis*. *Proceedings of the National Academy of Sciences* 100(13)*:*7877-7882.

Gideon, HP, Flynn, JL (2011). Latent tuberculosis: what the host "sees"? *Immunologic Research* 50(2-3)*:*202-212.

Ginsberg, A (2011). The TB Alliance: overcoming challenges to chart the future course of TB drug development. *Future Med Chem* 3(10)*:*1247-1252.

Godreuil, S, Tazi, L, Bañuls, A-L (2006). Pulmonary tuberculosis and *Mycobacterium tuberculosis*: modern molecular epidemiology and perspectives. In: Encyclopedia of Infectious Diseases: John Wiley & Sons, Inc., pp. 1-29.

Gouet, P, Courcelle, E, Stuart, DI, Metoz, F (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4)*:*305-308.

Gul, M, Gul, HI, Das, U, Hanninen, O (2005). Biological evaluation and structure-activity relationships of bis-(3-aryl-3-oxo-propyl)-methylamine hydrochlorides and 4-aryl-3-

arylcarbonyl-1-methyl-4-piperidinol hydrochlorides as potential cytotoxic agents and their alkylating ability towards cellular glutathione in human leukemic T cells. *Arzneimittel-Forschung* 55(6)*:*332-337.

Gupta, A, Kaul, A, Tsolaki, AG, Kishore, U, Bhakta, S (2012). *Mycobacterium tuberculosis*: immune evasion, latency and reactivation. *Immunobiology* 217(3)*:*363-374.

Hanahan, D (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166(4)*:*557-580.

Hanstein, WG, Davis, KA, Hatefi, Y (1971). Water structure and the chaotropic properties of haloacetates. *Archives of Biochemistry and Biophysics* 147(2)*:*534-544.

Hmama, Z, Sendide, K, Talal, A, Garcia, R, Dobos, K, Reiner, NE (2004). Quantitative analysis of phagolysosome fusion in intact cells: inhibition by mycobacterial lipoarabinomannan and rescue by an 1 , 25-dihydroxyvitamin D3–phosphoinositide 3-kinase pathway. *Journal of Cell Science* 117(10)*:*2131-2140.

Holton, SJ, Dairou, J, Sandy, J, Rodrigues-Lima, F, Dupret, JM, Noble, ME, Sim, E (2005). Structure of *Mesorhizobium loti* arylamine *N*-acetyltransferase 1. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 61(Pt 1)*:*14-16.

Hubbard, RE (2005). 3D structure and the drug-discovery process. In: Molecular Biosystems, pp. 391-406.

Hunter, RL (2011). Pathology of post primary tuberculosis of the lung: An illustrated critical review. *Tuberculosis* 91(6)*:*497-509.

Ignatova, Z, Gierasch, LM, Dieter Häussinger and Helmut, S (2007). Effects of osmolytes on protein folding and aggregation in cells. *Methods in Enzymology* 428(355-372.

Israili, ZH, Dayton, PG (1977). Metabolism of hydralazine. *Drug Metabolism Reviews* 6(2)*:*283-305.

Jeney, E, Zsolnai, T (1956). Studies in search of new tuberculostatic drugs. I. Hydrazine derivatives, carbolic acid, phenols, quaternary ammonium compounds and their

intermediaries. *Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene 1 Abt Medizinisch-hygienische Bakteriologie, Virusforschung und Parasitologie Originale* 167(1)*:*55.

Jiménez, JI, Miñambres, B, García, JL, Díaz, E (2002). Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environmental Microbiology* 4(12)*:*824-841.

Kaneko, T, Cooper, C, Mdluli, K (2011). Challenges and opportunities in developing novel drugs for TB. *Future Med Chem* 3(11)*:*1373-1400.

Kang, PB, Azad, AK, Torrelles, JB, Kaufman, TM, Beharka, A, Tibesar, E, DesJardin, LE, Schlesinger, LS (2005). The human macrophage mannose receptor directs *Mycobacterium tuberculosis* lipoarabinomannan-mediated phagosome biogenesis. *The Journal of Experimental Medicine* 202(7)*:*987.

Karakousis, PC (2009). Mechanisms of Action and Resistance of Antimycobacterial Agents In: Infectious Disease: Antimicrobial Drug Resistance. Mayers, DL editor: Humana Press, pp. 271-291.

Kaul, D, Anand, PK, Verma, I (2004). Cholesterol-sensor initiates *M. tuberculosis* entry into human macrophages. *Molecular and Cellular Biochemistry* 258(1)*:*219-222.

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(1)*:*75-84.

Kemp, MM, Weïwer, M, Koehler, AN (2012). Unbiased binding assays for discovering small-molecule probes and drugs. *Bioorganic & Medicinal Chemistry* 20(6)*:*1979-1989.

Kendall, SL, Withers, M, Soffair, CN, Moreland, NJ, Gurcha, S, Sidders, B, Frita, R, Ten Bokum, A, Besra, GS, Lott, JS, Stoker, NG (2007). A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Molecular Microbiology* 65(3)*:*684-699.

Kendall, SL, Burgess, P, Balhana, R, Withers, M, ten Bokum, A, Lott, JS, Gao, C, Uhia-Castro, I, Stoker, NG (2010). Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. *Microbiology* 156(5)*:*1362-1371.

Khan, N, Alam, K, Nair, S, Valluri, VL, Murthy, KJR, Mukhopadhyay, S (2008). Association of Strong Immune Responses to PPE Protein Rv1168c with Active Tuberculosis. *Clinical and Vaccine Immunology* 15(6)*:*974-980.

Kitz, R, Wilson, IB (1962). Esters of methanesulfonic acid as irreversible inhibitors of acetylcholinesterase. *Journal of Biological Chemistry* 237(10)*:*3245-3249.

Koul, A, Arnoult, E, Lounis, N, Guillemont, J, Andries, K (2011). The challenge of new drug discovery for tuberculosis. *Nature* 469(7331)*:*483-490.

Kumar, S, Tsai, C-J, Nussinov, R (2000). Factors enhancing protein thermostability. *Protein Engineering* 13(3)*:*179-191.

Lack, N, Lowe, ED, Liu, J, Eltis, LD, Noble, ME, Sim, E, Westwood, IM (2008). Structure of HsaD, a steroid-degrading hydrolase, from *Mycobacterium tuberculosis*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64(Pt 1)*:*2-7.

Lack, N, Sim, E (Unpublished work). *In silico* screening for HsaD inhibitors.

Lack, NA, Kawamura, A, Fullam, E, Laurieri, N, Beard, S, Russell, AJ, Evangelopoulos, D, Westwood, I, Sim, E (2009). Temperature stability of proteins essential for the intracellular survival of *Mycobacterium tuberculosis*. *Biochemical Journal* 418(2)*:*369-378.

Lack, NA, Yam, KC, Lowe, ED, Horsman, GP, Owen, RL, Sim, E, Eltis, LD (2010). Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism. *Journal of Biological Chemistry* 285(1)*:*434-443.

Laemmli, UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259)*:*680-685.

Laurieri, N, Crawford, MHJ, Kawamura, A, Westwood, IM, Robinson, J, Fletcher, AM, Davies, SG, Sim, E, Russell, AJ (2010). Small molecule colorimetric probes for specific

detection of human arylamine *N*-acetyltransferase 1, a potential breast cancer biomarker. *Journal of the American Chemical Society* 132(10)*:*3238-3239.

Lavinder, JJ, Hari, SB, Sullivan, BJ, Magliery, TJ (2009). High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *Journal of the American Chemical Society* 131(11)*:*3794-3795.

Leeson, P (2012). Drug discovery: Chemical beauty contest. *Nature* 481(7382)*:*455-456.

Leslie, AGW (2006). The integration of macromolecular diffraction data. *Acta Crystallographica Section D Biological Crystallography* D62(48-57.

Li, AH, Waddell, SJ, Hinds, J, Malloff, CA, Bains, M, Hancock, RE, Lam, WL, Butcher, PD, Stokes, RW (2010). Contrasting Transcriptional Responses of a Virulent and an Attenuated Strain of *Mycobacterium tuberculosis* Infecting Macrophages. *PLoS ONE* 5(6)*:*e11066.

Li, HT, Chang, T, Liu, MY, Le Gall, J, An, XM, Gui, LL, Zhang, JP, Liang, DC, Chang, WR (2002). Preliminary crystallographic studies of two *C*-terminally truncated copper-containing nitrite reductases from *Achromobacter cycloclastes*: changed crystallizing behaviors caused by residue deletion. *Biochemical and Biophysical Research Communications* 299(2)*:*173-176.

Lienhardt, C, Glaziou, P, Uplekar, M, Lönnroth, K, Getahun, H, Raviglione, M (2012). Global tuberculosis control: lessons learnt and future prospects. *Nat Rev Micro* 10(6)*:*407-416.

Makarov, V, Manina, G, Mikusova, K, Mollmann, U, Ryabova, O, Saint-Joanis, B, Dhar, N, Pasca, MR, Buroni, S, Lucarelli, AP, Milano, A, De Rossi, E, Belanova, M, Bobovska, A, Dianiskova, P, Kordulakova, J, Sala, C, Fullam, E, Schneider, P, McKinney, JD, Brodin, P, Christophe, T, Waddell, S, Butcher, P, Albrethsen, J, Rosenkrands, I, Brosch, R, Nandi, V, Bharath, S, Gaonkar, S, Shandil, RK, Balasubramanian, V, Balganesh, T, Tyagi, S, Grosset, J, Riccardi, G, Cole, ST (2009). Benzothiazinones kill *Mycobacterium tuberculosis* by blocking arabinan synthesis. *Science* 324(5928)*:*801-804.

Malen, H, Pathak, S, Softeland, T, de Souza, GA, Wiker, HG (2010). Definition of novel cell envelope associated proteins in Triton X-114 extracts of *Mycobacterium tuberculosis* H37Rv. *BMC Microbiology* 10(-)*:*132-143.

Martins, M, Pluvinage, B, Li de la Sierra-Gallay, I, Barbault, F, Dairou, J, Dupret, JM, Rodrigues-Lima, F (2008). Functional and structural characterization of the arylamine *N*-acetyltransferase from the opportunistic pathogen *Nocardia farcinica*. *Journal of Molecular Biology* 383(3)*:*549-560.

Matthews, BW (1968). Solvent content of protein crystals. *Journal of Molecular Biology* 33(2)*:*491-497.

McGovern, SL, Caselli, E, Grigorieff, N, Shoichet, BK (2002). A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of Medicinal Chemistry* 45(8)*:*1712-1722.

McShane, H (2005). Co-infection with HIV and TB: double trouble. *International Journal of STD & AIDS* 16(2)*:*95-101.

McShane, H (2011). Tuberculosis vaccines: beyond bacille Calmette-Guerin. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 366(1579)*:*2782-2789.

Meena, LS, Rajni (2010). Survival mechanisms of pathogenic *Mycobacterium tuberculosis* H37Rv. *FEBS Journal* 277(11)*:*2416-2427.

Meot-Ner, M, Deakyne, CA (1985). Unconventional ionic hydrogen bonds. 2. NH+...π complexes of onium ions with olefins and benzene derivatives. *Journal of the American Chemical Society* 107(2)*:*474-479.

Mikusová, K, Slayden, RA, Besra, GS, Brennan, PJ (1995). Biogenesis of the mycobacterial cell wall and the site of action of ethambutol. *Antimicrobial Agents and Chemotherapy* 39(11)*:*2484-2489.

Miner, MD, Chang, JC, Pandey, AK, Sassetti, CM, Sherman, DR (2009). Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian Journal of Experimental Biology* 47(6)*:*407-411.

Minnikin, DE, Dobson, G, Hutchinson, IG (1983). Characterization of phthiocerol dimycocerosates from *Mycobacterium tuberculosis*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 753(3)*:*445-449.

Mishra, AK, Driessen, NN, Appelmelk, BJ, Besra, GS (2011). Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host–pathogen interaction. *FEMS Microbiology Reviews* 35(6)*:*1126-1157.

Mukhopadhyay, S, Balaji, KN (2011). The PE and PPE proteins of *Mycobacterium tuberculosis*. *Tuberculosis* 91(5)*:*441-447.

Murshudov, GN, Vagin, AA, Dodson, EJ (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D Biological Crystallography* 53(3)*:*240-255.

Mushtaq, A, Payton, M, Sim, E (2002). The COOH terminus of arylamine *N*-acetyltransferase from *Salmonella typhimurium* controls enzymic activity. *Journal of Biological Chemistry* 277(14)*:*12175-12181.

Nettleship, JE, Brown, J, Groves, MR, Geerlof, A (2008). Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Methods in Molecular Biology* 426(3)*:*299-318.

New England Biolabs, COMPETENT CELLS Cloning & Protein Expression.

Nicholls, A, McGaughey, GB, Sheridan, RP, Good, AC, Warren, G, Mathieu, M, Muchmore, SW, Brown, SP, Grant, JA, Haigh, JA, Nevins, N, Jain, AN, Kelley, B (2010). Molecular shape and medicinal chemistry: a perspective. *Journal of Medicinal Chemistry* 53(10)*:*3862-3886.

Niesen, FH, Berglund, H, Vedadi, M (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* 2(9)*:*2212-2221.

Noble, JE, Bailey, MJA, Richard, RB, Murray, PD (2009). Quantitation of Protein. *Methods in Enzymology* 463(73-95.

Noll, H, Bloch, H, Asselineau, J, Lederer, E (1956). The chemical structure of the cord factor of *Mycobacterium tuberculosis*. *Biochimica et Biophysica Acta* 20(0)*:*299-309.

Novagen (2002). Protein expression: prokaryotic expression: pETBlue and pET system overview. Novagen Catalogue In: Novagen editor, pp. 84-91.

O'Shea, R, Moser, HE (2008). Physicochemical properties of antibacterial compounds: implications for drug discovery. *Journal of Medicinal Chemistry* 51(10)*:*2871-2878.

Okabe, N, Fukuda, H, Nakamura, T (1993). Structure of hydralazine hydrochloride. *Acta Crystallographica, Section C: Crystal Structure Communications* 49(10)*:*1844-1845.

Olaofe, OA, Burton, SG, Cowan, DA, Harrison, STL (2010). Improving the production of a thermostable amidase through optimising IPTG induction in a highly dense culture of recombinant *Escherichia coli*. *Biochemical Engineering Journal* 52(1)*:*19-24.

Ouellet, H, Johnston, JB, Montellano, PROd (2011). Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*. *Trends in Microbiology* 19(11)*:*530-539.

Pandey, AK, Sassetti, CM (2008). Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences of the United States of America* 105(11)*:*4376-4380.

Parra, M, Cadieux, N, Pickett, T, Dheenadhayalan, V, Brennan, MJ (2006). A PE Protein Expressed by *Mycobacterium avium* Is an Effective T-Cell Immunogen. *Infection and Immunity* 74(1)*:*786-789.

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. *Journal of Bacteriology* 181(4)*:*1343-1347.

Philips, JA, Ernst, JD (2012). Tuberculosis pathogenesis and immunity. *Annual Review of Pathology: Mechanisms of Disease* 7(1)*:*353-384.

Pieters, J, Gatfield, J (2002). Hijacking the host: survival of pathogenic mycobacteria inside macrophages. *Trends in Microbiology* 10(3)*:*142-146.

Pieters, J (2008). *Mycobacterium tuberculosis* and the macrophage: maintaining a balance. *Cell Host and Microbe* 3(6)*:*399-407.

Plati, JT, Wenner, W inventors (1951). Hoffmann-La Roche Inc. assignee. Tetrahydropyridines. US2537854.

Pluvinage, B, Sierra-Gallay, IL, Kubiak, X, Xu, X, Dairou, J, Dupret, JM, Rodrigues-Lima, F (2011). The *Bacillus anthracis* arylamine *N*-acetyltransferase ((BACAN)NAT1) that inactivates sulfamethoxazole, reveals unusual structural features compared with the other NAT isoenzymes. *FEBS Letters* 585(24)*:*3947-3952.

Price, WN, Chen, Y, Handelman, SK, Neely, H, Manor, P, Karlin, R, Nair, R, Liu, JF, Baran, M, Everett, J, Tong, SCN, Forouhar, F, Swaminathan, SS, Acton, T, Xiao, R, Luft, JR, Lauricella, A, DeTitta, GT, Rost, B, Montelione, GT, Hunt, JF (2009). Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nature Biotechnology* 27(1)*:*51-57.

Prlić, A, Bliven, S, Rose, PW, Bluhm, WF, Bizon, C, Godzik, A, Bourne, PE (2010). Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26(23)*:*2983-2985.

Qiu, X, Janson, CA (2009). Improving Protein Crystallizability by Modification and Engineering. In: Protein crystallization. Bergfors, TM editor. California: International University Line, pp. 249-266.

Rengarajan, J, Bloom, BR, Rubin, EJ (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 102(23)*:*8327-8332.

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. *Journal of Biological Chemistry* 246(10)*:*3250-3258.

Riley, RL, Mills, CC, Nyka, W, Weinstock, N, Storey, PB, Sultan, LU, Riley, MC, WELLS, WF (1995). Aerial dissemination op pulmonary tuberculosis: a two-year study of contagion in a tuberculosis ward. *American Journal of Epidemiology* 142(1)*:*3-14.

Rodham, DA, Suzuki, S, Suenram, RD, Lovas, FJ, Dasgupta, S, Goddard, WA, Blake, GA (1993). Hydrogen bonding in the benzene-ammonia dimer. *Nature (London)* 362(6422)*:*735-737.

Rowland, K (2012). Totally drug-resistant TB emerges in India: Discovery of a deadly form of TB highlights crisis of 'mismanagement'. *Nature News* doi:10.1038/nature.2012.9797.

Russell, AJ, Westwood, IM, Crawford, MH, Robinson, J, Kawamura, A, Redfield, C, Laurieri, N, Lowe, ED, Davies, SG, Sim, E (2009). Selective small molecule inhibitors of the potential breast cancer marker, human arylamine *N*-acetyltransferase 1, and its murine homologue, mouse arylamine *N*-acetyltransferase 2. *Bioorganic & Medicinal Chemistry* 17(2)*:*905-918.

Russell, DG, Mwandumba, HC, Rhoades, EE (2002). *Mycobacterium* and the coat of many lipids. *Journal of Cell Biology* 158(3)*:*421-426.

Salameh, MA, Wiegel, J (2010). Effects of detergents on activity, thermostability and aggregation of two alkalithermophilic lipases from *Thermosyntropha lipolytica*. *Open Biochem J* 4(-)*:*22-28.

Sambrook, J, Fritsch, EF, Maniatis, T (1989). Molecular cloning. a laboratory manual Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sandy, J, Mushtaq, A, Kawamura, A, Sinclair, J, Sim, E, Noble, M (2002). The structure of arylamine *N*-acetyltransferase from *Mycobacterium smegmatis*--an enzyme which inactivates the anti-tubercular drug, isoniazid. *Journal of Molecular Biology* 318(4)*:*1071-1083.

Sandy, J, Holton, S, Fullam, E, Sim, E, Noble, M (2005a). Binding of the anti-tubercular drug isoniazid to the arylamine *N*-acetyltransferase protein from *Mycobacterium smegmatis*. *Protein Science* 14(3)*:*775-782.

Sandy, J, Mushtaq, A, Holton, SJ, Schartau, P, Noble, ME, Sim, E (2005b). Investigation of the catalytic triad of arylamine *N*-acetyltransferases: essential residues required for acetyl transfer to arylamines. *Biochemical Journal* 390(Pt 1)*:*115-123.

Sarkar, S, Suresh, MR (2011). An overview of tuberculosis chemotherapy - a literature review. *Journal of Pharmacy and Pharmaceutical Sciences* 14(2)*:*148-161.

Sassetti, CM, Boyd, DH, Rubin, EJ (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology* 48(1)*:*77-84.

Sassetti, CM, Rubin, EJ (2007). The open book of infectious diseases. *Natural Medicines* 13(3)*:*279-280.

Savvi, S, Warner, DF, Kana, BD, McKinney, JD, Mizrahi, V, Dawes, SS (2008). Functional characterization of a vitamin B12-dependent methylmalonyl pathway in *Mycobacterium tuberculosis*: implications for propionate metabolism during growth on fatty acids. *Journal of Bacteriology* 190(11)*:*3886-3895.

Sawangwan, T, Goedl, C, Nidetzky, B (2010). Glucosylglycerol and glucosylglycerate as enzyme stabilizers. *Biotechnology Journal* 5(2)*:*187-191.

Schrodinger, LLC (2010). The PyMOL molecular graphics system, version 1.3r1.

Senaratne, RH, De Silva, AD, Williams, SJ, Mougous, JD, Reader, JR, Zhang, T, Chan, S, Sidders, B, Lee, DH, Chan, J, Bertozzi, CR, Riley, LW (2006). 5′-Adenosinephosphosulphate reductase (CysH) protects *Mycobacterium tuberculosis* against free radicals during chronic infection phase in mice. *Molecular Microbiology* 59(6)*:*1744-1753.

Shenoi, S, Friedland, G (2009). Extensively drug-resistant tuberculosis: a new face to an old pathogen. *Annual Review of Medicine* 60(1)*:*307-320.

Shimizu, T, Tomioka, H (2009). Novel type of antimicrobial mechanism in host macrophages against mycobacterial infections. *Nihon Hansenbyo Gakkai Zasshi (Japanese Journal of Leprosy)* 78(283-291.

Sikora, AL, Frankel, BA, Blanchard, JS (2008). Kinetic and chemical mechanism of arylamine *N*-acetyltransferase from *Mycobacterium tuberculosis*. *Biochemistry* 47(40)*:*10781-10789.

Sim, E, Pinter, K, Mushtaq, A, Upton, A, Sandy, J, Bhakta, S, Noble, M (2003). Arylamine N-acetyltransferases: a pharmacogenomic approach to drug metabolism and endogenous function. *Biochemical Society Transactions* 31(Pt 3)*:*615-619.

Sim, E, Sandy, J, Evangelopoulos, D, Fullam, E, Bhakta, S, Westwood, I, Krylova, A, Lack, N, Noble, M (2008a). Arylamine *N*-acetyltransferases in mycobacteria. *Current Drug Metabolism* 9(6)*:*510-519.

Sim, E, Walters, K, Boukouvala, S (2008b). Arylamine *N*-acetyltransferases: from structure to function. *Drug Metabolism Reviews* 40(3)*:*479-510.

Sim, E, Fakis, G, Laurieri, N, Boukouvala, S (2012). Arylamine N -Acetyltransferases – from Drug Metabolism and Pharmacogenetics to Identification of Novel Targets for Pharmacological Intervention. *Advances in Pharmacology and Chemotherapy:in press*.

Sinclair, JC, Sandy, J, Delgoda, R, Sim, E, Noble, ME (2000). Structure of arylamine *N*-acetyltransferase reveals a catalytic triad. *Nature Structural Biology* 7(7)*:*560-564.

Singh, J, Petter, RC, Baillie, TA, Whitty, A (2011). The resurgence of covalent drugs. *Nature Reviews Drug Discovery* 10(4)*:*307-317.

Sirakova, TD, Dubey, VS, Cynamon, MH, Kolattukudy, PE (2003). Attenuation of *Mycobacterium tuberculosis* by Disruption of a mas-Like Gene or a Chalcone Synthase-Like Gene, Which Causes Deficiency in Dimycocerosyl Phthiocerol Synthesis. *Journal of Bacteriology* 185(10)*:*2999-3008.

Smith, I (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical Microbiology Reviews* 16(3)*:*463-496.

Spargo, BJ, Crowe, LM, Ioneda, T, Beaman, BL, Crowe, JH (1991). Cord factor (alpha,alpha-trehalose 6,6'-dimycolate) inhibits fusion between phospholipid vesicles. *Proceedings of the National Academy of Sciences* 88(3)*:*737-740.

Squire, SB, Obasi, A, Nhlema-Simwaka, B (2006). The Global Plan to Stop TB: a unique opportunity to address poverty and the Millennium Development Goals. *The Lancet* 367(9514)*:*955-957.

Stamm, LM, Brown, EJ (2004). *Mycobacterium marinum*: the generalization and specialization of a pathogenic *Mycobacterium*. *Microbes and Infection* 6(15)*:*1418-1428.

Stinear, TP, Seemann, T, Harrison, PF, Jenkin, GA, Davies, JK, Johnson, PDR, Abdellah, Z, Arrowsmith, C, Chillingworth, T, Churcher, C, Clarke, K, Cronin, A, Davis, P, Goodhead, I, Holroyd, N, Jagels, K, Lord, A, Moule, S, Mungall, K, Norbertczak, H, Quail, MA, Rabbinowitsch, E, Walker, D, White, B, Whitehead, S, Small, PLC, Brosch, R, Ramakrishnan, L, Fischbach, MA, Parkhill, J, Cole, ST (2008). Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Research* 18(5)*:*729-741.

Sturgill-Koszycki, S, Schlesinger, P, Chakraborty, P, Haddix, P, Collins, H, Fok, A, Allen, R, Gluck, S, Heuser, J, Russell, D (1994). Lack of acidification in *Mycobacterium* phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* 263(5147)*:*678-681.

Sweeney, AM, D'Arcy, A (2009). Seeding. In: Protein crystallization. Bergfors, TM editor. California: International University Line, pp. 95-114.

Szilágyi, A, Závodszky, P (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* 8(5)*:*493-504.

Thompson, JD, Higgins, DG, Gibson, TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22(22)*:*4673-4680.

Thornton, JM, Todd, AE, Milburn, D, Borkakoti, N, Orengo, CA (2000). From structure to function: approaches and limitations. *Nature Structural Biology* 7(-)*:*991-994.

Trunz, BB, Fine, P, Dye, C (2006). Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. 367(9517)*:*1173-1180.

Tudó, G, Laing, K, Mitchison, DA, Butcher, PD, Waddell, SJ (2010). Examining the basis of isoniazid tolerance in nonreplicating *Mycobacterium tuberculosis* using transcriptional profiling. *Future Med Chem* 2(8)*:*1371-1383.

Udwadia, ZF, Amale, RA, Ajbani, KK, Rodrigues, C (2011). Totally Drug-Resistant Tuberculosis in India. *Clinical Infectious Diseases* 10.1093/cid/cir889.

Ueda, A, Wood, TK (2008). Potassium and sodium transporters of *Pseudomonas aeruginosa* regulate virulence to barley. *Applied Microbiology and Biotechnology* 79(5)*:*843-858.

Ueki, M, Galonic, DP, Vaillancourt, FH, Garneau-Tsodikova, S, Yeh, E, Vosburg, DA, Schroeder, FC, Osada, H, Walsh, CT (2006). Enzymatic generation of the antimetabolite [gamma],[gamma]-dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chemistry & Biology* 13(11)*:*1183-1191.

Upton, AM (2001). Arylamine *N*-acetyltransferase of *Mycobacterium tuberculosis* (DPhil Thesis). Oxford, Oxford University.

Upton, AM, Mushtaq, A, Victor, TC, Sampson, SL, Sandy, J, Smith, DM, van Helden, PV, Sim, E (2001). Arylamine *N*-acetyltransferase of *Mycobacterium tuberculosis* is a polymorphic enzyme and a site of isoniazid metabolism. *Molecular Microbiology* 42(2)*:*309-317.

Vagin, A, Teplyakov, A (1997). MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* 30(6)*:*1022-1025.

Vagin, A, Teplyakov, A (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallographica Section D Biological Crystallography* 56(12)*:*1622-1624.

Van der Geize, R, Yam, K, Heuser, T, Wilbrink, MH, Hara, H, Anderton, MC, Sim, E, Dijkhuizen, L, Davies, JE, Mohn, WW, Eltis, LD (2007). A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 104(6)*:*1947-1952.

Vedadi, M, Niesen, FH, Allali-Hassani, A, Fedorov, OY, Finerty, PJ, Jr., Wasney, GA, Yeung, R, Arrowsmith, C, Ball, LJ, Berglund, H, Hui, R, Marsden, BD, Nordlund, P, Sundstrom, M, Weigelt, J, Edwards, AM (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences of the United States of America* 103(43)*:*15835-15840.

Vedadi, M, Arrowsmith, CH, Allali-Hassani, A, Senisterra, G, Wasney, GA (2010). Biophysical characterization of recombinant proteins: a key to higher structural genomics success. *Journal of Structural Biology* 172(1)*:*107-119.

Vergne, I, Fratti, RA, Hill, PJ, Chua, J, Belisle, J, Deretic, V (2004). *Mycobacterium tuberculosis* phagosome maturation arrest: mycobacterial phosphatidylinositol analog phosphatidylinositol mannoside stimulates early endosomal fusion. *Molecular Biology of the Cell* 15(2)*:*751-760.

Vilcheze, C, Weisbrod, TR, Chen, B, Kremer, L, Hazbon, MH, Wang, F, Alland, D, Sacchettini, JC, Jacobs, WR, Jr. (2005). Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. *Antimicrobial Agents and Chemotherapy* 49(2)*:*708-720.

Villeneuve, C, Gilleron, M, Maridonneau-Parini, I, Daffé, M, Astarie-Dequeker, C, Etienne, G (2005). Mycobacteria use their surface-exposed glycolipids to infect human macrophages through a receptor-dependent process. *Journal of Lipid Research* 46(3)*:*475-483.

Vissa, V, Brennan, P (2001). The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biology* 2(8)*:*1023.1021 - 1023.1028.

Wakefield, L, Cornish, V, Long, H, Griffiths, WJ, Sim, E (2007). Deletion of a xenobiotic metabolizing gene in mice affects folate metabolism. *Biochemical and Biophysical Research Communications* 364(3)*:*556-560.

Wallace, JG (1961). The heat resistance of tubercle bacilli in the lungs of infected mice. *American Review of Respiratory Disease* 83(-)*:*866-871.

Warner, DF, Mizrahi, V (2007). The survival kit of *Mycobacterium tuberculosis*. *Natural Medicines* 13(3)*:*282-284.

Weber, WW, Hein, DW (1985). N-acetylation pharmacogenetics. *Pharmacological Reviews* 37(1)*:*25-79.

Westwood, I, Bhakta, S, Russell, A, Fullam, E, Anderton, M, Kawamura, A, Mulvaney, A, Vickers, R, Bhowruth, V, Besra, G, Lalvani, A, Davies, S, Sim, E (2010). Identification of aryalmine *N*-acetyltransferase inhibitors as an approach towards novel anti-tuberculars. *Protein & Cell* 1(1)*:*82-95.

Westwood, IM, Holton, SJ, Rodrigues-Lima, F, Dupret, JM, Bhakta, S, Noble, ME, Sim, E (2005). Expression, purification, characterization and structure of *Pseudomonas aeruginosa* arylamine *N*-acetyltransferase. *Biochemical Journal* 385(Pt 2)*:*605-612.

Westwood, IM, Kawamura, A, Fullam, E, Russell, AJ, Davies, SG, Sim, E (2006). Structure and mechanism of arylamine N-acetyltransferases. *Current Topics in Medicinal Chemistry* 6(15)*:*1641-1654.

Westwood, IM, Kawamura, A, Russell, AJ, Sandy, J, Davies, SG, Sim, E (2011). Novel small-molecule inhibitors of arylamine *N*-acetyltransferases: drug discovery by high-throughput screening. *Combinatorial Chemistry & High Throughput Screening* 14(2)*:*117-124.

Wetherall, NT, Trivedi, T, Zeller, J, Hodges-Savola, C, McKimm-Breschkin, JL, Zambon, M, Hayden, FG (2003). Evaluation of Neuraminidase Enzyme Assays Using Different Substrates To Measure Susceptibility of Influenza Virus Clinical Isolates to Neuraminidase Inhibitors: Report of the Neuraminidase Inhibitor Susceptibility Network. *Journal of Clinical Microbiology* 41(2)*:*742-750.

WHO (2011). World Health Organisation: Global tuberculosis control report 2011.

Winder, F, Collins, P (1970). Inhibition by isoniazid of synthesis of mycolic acids in *Mycobacterium tuberculosis*. *Journal of General Microbiology* 63(1)*:*41.

Wouters, J (1998). Cation-pi (Na+-Trp) interactions in the crystal structure of tetragonal lysozyme. *Protein Science* 7(11)*:*2472-2475.

Wu, H, Dombrovsky, L, Tempel, W, Martin, F, Loppnau, P, Goodfellow, GH, Grant, DM, Plotnikov, AN (2007). Structural basis of substrate-binding specificity of human arylamine *N*-acetyltransferases. *Journal of Biological Chemistry* 282(41)*:*30189-30197.

Yajko, D, Madej, J, Lancaster, M, Sanders, C, Cawthon, V, Gee, B, Babst, A, Hadley, W (1995). Colorimetric method for determining MICs of antimicrobial agents for *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* 33(9)*:*2324-2327.

Yam, KC, D'Angelo, I, Kalscheuer, R, Zhu, H, Wang, JX, Snieckus, V, Ly, LH, Converse, PJ, Jacobs, WR, Jr., Strynadka, N, Eltis, LD (2009). Studies of a ring-cleaving

dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathogens* 5(3)*:*e1000344.

Yang, X, Nesbitt, NM, Dubnau, E, Smith, I, Sampson, NS (2009). Cholesterol metabolism increases the metabolic pool of propionate in *Mycobacterium tuberculosis*. *Biochemistry* 48(18)*:*3819-3821.

Yang, ZR, Thomson, R, McNeil, P, Esnouf, RM (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21(16)*:*3369-3376.

Yeh, JI (2009). Additives and microcalorimetric approaches for optimization of crystallization. In: Protein crystallization. Bergfors, TM editor. California: International University Line, pp. 331-349.

Yildir, C, Onsan, ZI, Kirdar, B (1998). Optimization of starting time and period of induction and inducer concentration in the production of the restriction enzyme EcoRI from recombinant *Escherichia coli* 294. *Turkish Journal of Chemistry* 22(3)*:*221-226.

Young, DB, Gideon, HP, Wilkinson, RJ (2009). Eliminating latent tuberculosis. *Trends in Microbiology* 17(5)*:*183-188.

Zhang, Y (2005). The Magic bullets and tuberculosis drug targets. *Annual Review of Pharmacology and Toxicology* 45(1)*:*529-564.