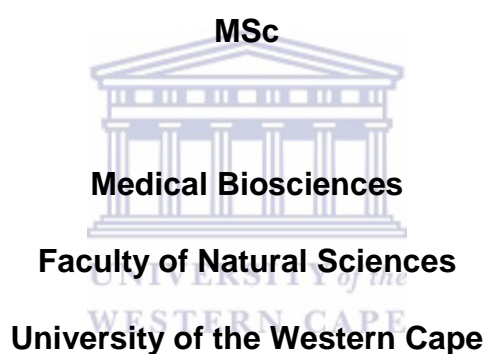


The role of arylamine *N*-acetyltransferase 1 (NAT1) in the clinical therapy of tuberculosis

By

Gratia-Lizé Willemse

Thesis submitted in Fulfillment of the Requirements for the Degree Magister Scientiae



Supervisor : Prof B.C. Fielding

Department of Medical Bioscience
University of the Western Cape

Co-supervisor: Prof A. Christoffels

MRC Bioinformatics Unit
South African National Bioinformatics
Institute
University of the Western Cape

2017

KEYWORDS

Acetylation

Mycobacterium tuberculosis

NAT1

PAS

PABA

Protein expression

Single nucleotide polymorphism



ABSTRACT

Despite attempts to develop new drugs to reduce the worldwide mortality rate attributable to tuberculosis (TB), the illness remains a threat. Isoniazid (INH) has been used as a frontline drug for decades. However, several resistant strains of the organism – *Mycobacterium tuberculosis* (*M. tuberculosis*) – still emerge. The drug is mainly metabolised by a family of enzymes, arylamine N-acetyltransferases (NAT). The three human NAT genes – NAT1, NAT2 and the pseudogene, NATP – are found on chromosome 18p22. NAT1 and NAT2 are isoenzymes which differ at certain amino acid positions. Subsequently, the differences affect substrate specificity. NAT1 shows specificity to p-aminobenzoic acid (PABA) and p-aminosalicylate (PAS). Previously, computer algorithms were used to predict the efficacy of the enzyme with regard to the acetylation phenotype it confers. The two which were focused on, Sorting Intolerant From Tolerant (SIFT) program and Polymorphism phenotyping version 2 program (PolyPhen-2), showed conflicting results for the effect of SNPs on the acetylation rate and subsequent enzyme function. Further structural prediction methods were used to test the effect of V231G on the structure and consequent function of the native protein, NAT1.

Six novel non-synonymous, independently cloned SNPs, as received from GenScript, were validated and the resultant NAT1 protein samples (~35 kDa) were successfully expressed in a BL21 *Escherichia coli* bacterial system. The rate of acetylation was tested using PNPA as acetyl-donor and NAT1 specific substrates, PAS and PABA.

In this study, all SNPs tested affect the acetylation rate when compared to the wild type NAT1. The SNPs increase acetylation rate in the presence of PAS, except SNP V231G. Contrastingly, the acetylation rate decreased for all SNPs in the presence of PABA. The results also show that the general absorbance readings for PABA (post-incubation) were higher than that of PAS. Thus, it can be concluded that NAT1 has a higher overall affinity for

PABA. This makes PAS more resistant to inactivation by NAT1 and more suitable for TB treatment in comparison to PABA. The experimental data for SNPs T193S and T240S disprove the prediction and, in fact, does affect the rate of acetylation in the presence of both substrates. The predictions for the remaining SNPs (V231G, V235A, F202V and D229H) were confirmed for the acetylating function of NAT1.



DECLARATION

I, Gratia-Lizé Willemse, declare that this thesis on “***The role of arylamine N-acetyltransferase 1 (NAT1) in the clinical therapy of Tuberculosis***” is my own work, that it has not been submitted for any degree or examination at any other university, that it is free of plagiarism and that all the sources used have been indicated and acknowledged by complete references.

Full name : _____

Date : _____

Signed : _____



ACKNOWLEDGEMENTS

With God all things are possible – Matthew 19:26

- To my parents, Dr. Willy Willemse and Mrs Maria Willemse, my love for you is infinite. So much of me is made of what I learned from you, everything I stand for. Thank you for teaching me to smile and believe in myself. You are my blessing!
- To my family, my circle of strength, thank you for your love and support.
- My supervisor, Prof Fielding, you have the innate ability to spark curiosity within your students and open their minds, and that makes you a teacher. Your influence in shaping me is appreciated.
- Virology Lab. Thank you for the laughs and making the dreary times worthwhile, for sharing reagents and knowledge.
- Aasiyah Chafekar. It is an extraordinary privilege to meet someone that touches your life so effortlessly and so deeply. Thank you for being such an important part in this, my story.

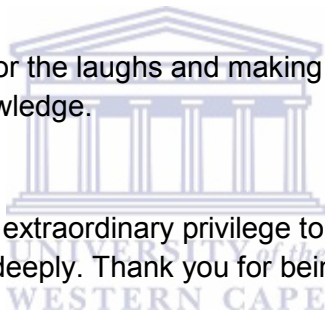


TABLE OF CONTENTS

KEYWORDS	I
ABSTRACT	II
DECLARATION	IV
ACKNOWLEDGEMENTS	V
LIST OF ABBREVIATIONS	1
LIST OF FIGURES.....	6
LIST OF TABLES	7
LIST OF APPENDICES	8
CHAPTER 1 - INTRODUCTION	9
1.1 BACKGROUND	10
1.2 TUBERCULOSIS	10
1.3 ISONIAZID (INH).....	12
1.4 ARYLAMINE N-ACETYLTRANSFERASE (NAT)	13
1.4.1 Acetylation phenotype	13
1.4.2 Mycobacterial NAT	14
1.4.3 Human Arylamine N-acetyltransferases (NAT) genes.....	15
1.4.4 Arylamine N-acetyltransferase 1 (NAT1)	16
1.5 TOXICOLOGICAL EFFECTS OF NAT.....	21
1.5.1 Carcinogenesis	21
1.6 TB TREATMENT	22
1.6.1 Anti-tubercular drugs and hepatotoxicity.....	22
1.6.2 NAT acetylation and TB treatment efficacy.....	23
1.7 HYPOTHESIS AND AIMS.....	24
1.7.1 Hypothesis.....	24
1.7.2 Aims of the study	25
2.1 PLASMID AND REAGENTS	27
2.2 BACTERIAL STRAINS	28
2.3 TISSUE CULTURE	28
2.4 REVERSE TRANSCRIPTION	28
2.5 CLONING.....	29
2.5.1 Transformation of NAT1-pET-30a(+) into BL21(DE3)pLysS.....	29
2.5.2 Colony picking and starter cultures	30
2.5.3 Extraction of plasmid DNA and DNA quantification	30
2.5.4 NAT1 verification	30
2.5.4.1 Gene-specific PCR optimization.....	30
2.5.4.1.1 cDNA gene-specific PCR amplification	31
2.5.4.1.2 NAT1-pUC57 PCR amplification	32
2.5.4.2 Gene – specific PCR amplification of NAT1-pET-30a(+) SNPs	35
2.5.4.3 Restriction endonuclease digest.....	36
2.5.4.3.1 RE digest of pGex-4T-2	36
2.5.4.3.2 RE digest of NAT1-pET-30a(+)......	37
2.5.4.4 Agarose gel electrophoresis	38

2.5.4.5	Preservation	38
2.6	PROTEIN EXPRESSION STUDIES	39
2.6.1	<i>Standard protein expression protocol</i>	39
2.6.2	<i>Protein expression optimisation</i>	40
2.6.2.1	Induction with 1mM IPTG at specific OD readings with fermentation at 30 °C	40
2.6.2.2	Induction with 0.1M IPTG at specific OD readings with fermentation at 27 °C	40
2.6.2.3	Addition of 1% glucose to the 500ml culture	41
2.6.2.4	Increased CAM+ concentration and decreased induction time	41
2.6.2.5	Expression of empty pET-28a(+) vector.....	42
2.6.3	<i>Protein analysis</i>	42
2.6.3.1	Protein extraction and cell lysate preparation	42
2.6.3.2	Protein purification.....	43
2.6.3.3	Protein quantification.....	44
2.6.3.4	Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	45
2.6.3.5	Western Blot Transfer	46
2.7	NAT1 ACETYLTRANSFERASE ACTIVITY ASSAY	47
CHAPTER 3 – RESULTS.....		49
3.1	NAT1-pUC57 AND PGEX-4T-2	50
3.2	TRANSFORMATION OF NAT1-PET-30A(+) INTO BL21(DE3)PLYSS	50
3.3	PROTEIN EXPRESSION STUDIES	52
3.3.1	<i>Bradford assay for purified protein quantification</i>	53
3.3.2	<i>SDS-PAGE and Comassie staining</i>	55
3.3.3	<i>Western blot transfer</i>	57
3.4	NAT1 ACETYLTRANSFERASE ACTIVITY ASSAYS	58
3.4.1	<i>Optimization of NAT1 acetyltransferase activity assays</i>	59
3.2.2	<i>NAT1 acetyltransferase activity assay for PAS</i>	61
3.2.3	<i>NAT1 acetyltransferase activity assay for PABA</i>	63
CHAPTER 4 – DISCUSSION AND CONCLUSION		65
4.1	DISCUSSION	66
4.2	CONCLUSION	71
REFERENCES.....		74
APPENDIX I – NAT1 SEQUENCES		82
	NAT1 OPEN READING FRAME	82
	NAT1 AMINO ACID SEQUENCE	82
APPENDIX II – VECTOR MAPS.....		83
	pET-30A(+).....	83
	pET-28A(+).....	84
	pUC57	84
	PGEX-4T-2	85
APPENDIX III - PROTOCOLS		86
	PUREYIELD™ PLASMID MINIPREP SYSTEM	86
	MAGNEHIS™ PROTEIN PURIFICATION SYSTEM.....	87
APPENDIX IV – PCR OPTIMIZATION.....		88
	PCR AT 50°C	88
	PCR AT 52°C.....	89

PCR AT 54°C	89
PCR AT 56°C, 58°C, 60°C AND 62°C	90
DNA TEMPLATE PROFILE PCR.....	90
PCR NUCLEOTIDE MIX (dNTP) PROFILE PCR	91
MgCL2 PROFILE PCR.....	91
APPENDIX V – RE DIGEST OF PGEX-4T-2.....	92
APPENDIX VI – PROTEIN EXPRESSION OPTIMIZATION	93
INDUCTION WITH 1MM IPTG AT SPECIFIC OD READINGS WITH FERMENTATION AT 30 °C	93
INCREASED CAM ⁺ CONCENTRATION TO 50µG/ML AND DECREASED INDUCTION TIME FROM 4 HOURS TO 2 HOURS	94



LIST OF ABBREVIATIONS

A	adenine nucleotide base or Alanine amino acid
aa	amino acid
Abs	absorbance
Ac-CoA	acetyl co-enzyme A
bp	base pair
BSA	Bovine serum albumin
C	cytosine nucleotide base
C ₁	starting concentration of a substance
C ₂	final concentration of a substance
CAM ⁺	Chloramphenicol
cDNA	complementary DNA
CO ₂	carbon dioxide
D	Aspartic acid protein
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
DIH	drug induced hepatotoxicity
DMABP	3,2_-dimethyl-4-aminobiphenyl
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
<i>E.coli</i>	<i>Escheria coli</i>
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide

F	Phenylalanine protein
FBS	Foetal bovine serum
g	gram
G	guanine nucleotide base or Glycine protein
GST	glutathione S-transferase
His or H	Histidine amino acid
HIV	Human
INH	isoniazid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
<i>katG</i>	catalase peroxidase gene
KAN ⁺	Kanamycin
kb	kilobase pair
kDa	kiloDalton
<i>lac</i>	lactose gene
<i>lacI</i>	lactose operon repressor gene
<i>LacZ</i>	LacZ gene encoding β -galactosidase
LB broth	Luria-Bertani broth
M	Molar
MBP	Maltose-binding protein
MDR	multidrug resistant
mg	milligram
mM	milliMolar
ml	millilitre
<i>M.tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
MTBC	Mycobaterium Tuberculosis Complex
NaCl	sodium chloride
NATs	arylamine N-acetyltransferase enzymes



NAT1	arylamine N-acetyltransferase enzyme type 1
NAT2	arylamine N-acetyltransferase enzyme type 2
<i>NAT/NAT1</i>	<i>italicised letters represent gene names</i>
NCBI	National Institute for Biotechnology Information
NdeI	endonuclease commonly used as a restriction enzyme isolated from <i>Neisseria denitrificans</i>
ng	nanogram
Ni ²⁺	Nickel
nM	nanomolar
nm	nanometer
NotI	Restriction endonuclease
OD ₆₀₀	optical density or absorbance read at 600nm
OmpT	outer membrane protease
ORF	open reading frame
PABA	para-aminobenzoic acid
pABGlu	para-aminobenzoylglutamate
PAS	para-aminosalicylic acid
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Phe	Phenylalanine protein
PNPA	para-nitrophenol acetate
PolyPhen-2	Polymorphism phenotyping version 2 program
R	Arginine protein
RE	restriction endonuclease
RNA	ribonucleic acid
rpm	revolutions per minute
S	Serine amino acid

SDS	Sodium dodecyl sulphate
SIFT	Sorting Intolerant From Tolerant program
SNP	single nucleotide polymorphism
T	thymine nucleotide base or Threonine amino acid
TB	Tuberculosis
TBE	Tris-borate-EDTA
TE	Tris-EDTA
T _m	melting point temperature
Tris	trisaminomethane
U	unit measurement
µg	microgram
UI	uninduced sample
µl	microliter
UTR	untranslated region
UV	ultra violet
V	Valine amino acid or volt
V ₁	starting volume of a substance
V ₂	final volume
WHO	World Health Organization
WT	wild type
XDR	extensively drug resistant
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XhoI	type II restriction enzyme isolated from <i>Xanthomonas vasicola</i>
Y	Tyrosine protein
α	alpha
β	beta
°C	degrees celcius



%	percentage
5'	5 prime end
3'	3 prime end

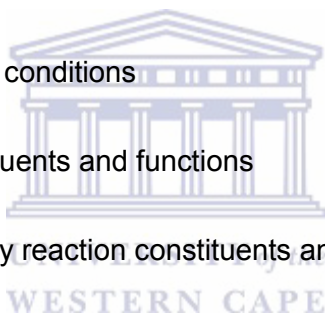


LIST OF FIGURES

- Figure 1:** Immunological cycle of *M.tuberculosis*.
- Figure 2:** *NAT1* wild type showing the 9 exons.
- Figure 3:** Overall structure of human NAT1 and NAT2
- Figure 4:** *NAT1* polymorphisms
- Figure 5:** Schematic representation of the Bradford assay setup on a 96 well plate
- Figure 6:** PCR amplification of *NAT1* constructs
- Figure 7:** RE digests of *NAT1* constructs
- Figure 8:** Bradford assay standard curve for protein quantification
- Figure 9:** SDS-PAGE gel of Bradford assay BSA standards
- Figure 10:** SDS-PAGE Comassie stained gel of *NAT1* samples
- Figure 11:** Control elements of the pET protein expression system
- Figure 12:** Western Blot Transfer of *NAT1* samples
- Figure 13:** Bar graph of optimization assay for PABA
- Figure 14:** Bar graph of optimization assay for PAS
- Figure 15:** Bar graph for *NAT1* acetyltransferase activity assay for PAS
- Figure 16:** Bar graph for *NAT1* acetyltransferase activity assay for PABA

LIST OF TABLES

- Table 1:** Initial *NAT1* forward and reverse primers
- Table 2:** DNA template PCR profile constituents
- Table 3:** PCR nucleotide mix PCR profile constituents
- Table 4:** MgCl₂ PCR profile constituents
- Table 5:** *NAT1* forward and reverse primers
- Table 6:** PCR conditions for *NAT1* amplification
- Table 7:** RE digest reaction constituents for pGEX-4T-2
- Table 8:** RE digest reaction constituents for *NAT1*-pET-30a(+)
- Table 9:** RE digest reaction conditions
- Table 10:** Lysis buffer constituents and functions
- Table 11:** *NAT1* activity assay reaction constituents and concentrations
- Table 12:** DNA concentrations from Nanodrop2000
- Table 13:** *NAT1* purified protein concentrations in µg/ml



LIST OF APPENDICES

Appendix I	NAT1 sequences
Appendix II	Vector maps
Appendix III	Manufacturers' protocols
Appendix IV	PCR optimization
Appendix V	RE digest of pGEX-4T-2
Appendix VI	Protein Expression optimization





Chapter 1 - Introduction

1.1 Background

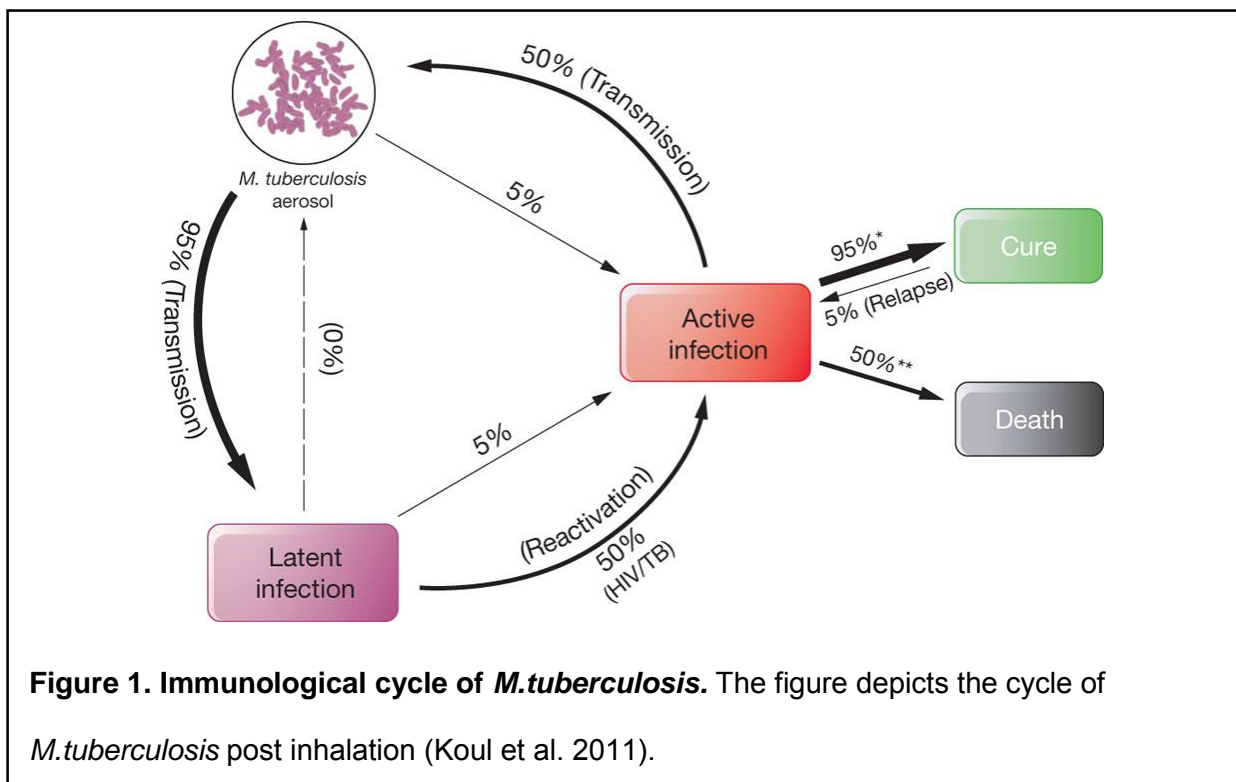
Tuberculosis (TB) is major cause of deaths worldwide and the general treatment regimen consists of a combination of drugs, with Isoniazid (INH) being one of the frontline drugs. However, INH is inactivated by NAT enzymes which aids the emergence of resistant *M. tuberculosis* strains (Upton et al, 2001; Cattamanchi et al, 2009). This study focuses on the single nucleotide polymorphisms (SNPs) within NAT1 and their effect on the rate of drug acetylation.

1.2 Tuberculosis

TB remains one of the leading causes of death worldwide, second to HIV. The disease has plagued the human population for thousands of years. Approximately 9.2 million deaths occurred in Asia and Africa with Africa accounting for 31% of TB incident cases in the world. TB statistics reported that 9 million new cases surfaced and 1.5 million deaths occurred in 2013. In 2015, an estimated 10.4 million new TB cases surfaced, of which 1.2 million were attributed to HIV-coinfected individuals worldwide. China, India, Indonesia, Nigeria, Pakistan and South Africa accounted for a collective 60% new TB incidences. In South Africa, TB accounted for 25 000 deaths in 2015. The worldwide statistics for TB fatalities for 2014-2015 were 1.4 million, 0.4 million of which were HIV-coinfected. (Global Tuberculosis Report, 2016).

Despite the availability of effective drugs to reduce the mortality rate, TB continues plague populations around the world. This is due to the fact that the causative agent, *Mycobacterium tuberculosis* (*M. tuberculosis*), has a complex aetiology. The organism is a member of a group of closely related bacterial sub-species known as the Mycobacterium tuberculosis complex (MTBC). Research suggests that the disease dates back 40 000 years (Wirth et al, 2008; Bouakaze et al, 2010).

M. tuberculosis is an airborne pathogen which uses the human macrophage as its host. Upon infection, the human immune system launches an immune response in order to contain the bacterium in a macrophage granuloma. There are three possible outcomes of TB infection. The bacteria could 1) be killed instantly by the pulmonary immune system, 2) cause active disease after infection or 3) be phagocytosed and contained in a macrophage granuloma and become latent. This containment may last for numerous years, or even the lifespan of an individual. *M. tuberculosis* could however disseminate from granulomas several years after containment, causing active tuberculosis (reactivation disease) as seen in the cycle depicted in figure 1. This is true for approximately 10% of such cases (Stockton, 2004; Pieters, 2008).



Apart from environmental influences, host genetics contribute to the possible development of TB disease phenotype of an individual (Moller, 2007). Stockton (2004) showed that only 10% of infected, immuno-competent people develop the active disease. This could be attributed to the host genetic factors (Pantelidis, 2005). It is believed that the interaction between the host, pathogen and environment would then lead to the ultimate development and portrayal of the disease phenotype. This is worrying since TB is a curable disease and many anti-tuberculosis drugs have been developed over the past few decades. The treatment success

in South Africa only reaches 68% as recorded in 2002 (Global Tuberculosis Report, 2016). TB treatment regimens are a lot more complex than any other mycobacterial infection. Thus, the rate of non-compliance among patients is high which leads to treatment failures and the emergence of resistant strains of *M. tuberculosis* - multidrug resistant (MDR), extensively drug resistant (XDR) and totally drug resistant (TDR) (Hearn and Cynamon, 2004). According to the WHO, global statistics show that 480 000 people developed MDR-TB by 2015 and 9% went on to develop XDR-TB (Global Tuberculosis Report, 2016). This supports the hypothesis that host-pathogen interactions play a major role in the prevalence of certain strains within certain population groups. It has been documented that researchers have investigated this age-old disease and its propagation throughout the world (Möller, 2007). These studies aimed at elucidating the pathogenesis and structure of *M. tuberculosis* and gave aid to the development of new and improved TB vaccines as well as the improvement of existing anti-tuberculosis drugs.

1.3 Isoniazid (INH)

Anti-tuberculosis treatment regimens are a combination of antibiotics with isoniazid (INH) being the frontline antibiotic (Cattamanchi, 2009). It was only 40 years after its first discovery that the drug was proven to be bacteriocidal (Evans, 1960). The exact mechanism of action of INH is not well known. However, the accepted explanation is that INH is metabolized and activated by the catalase and peroxidase activity of the *katG* gene products (Sandy, 2002). The activated drug then disrupts the mycolic acid pathway and, consequently, the maintenance of the bacterial cell wall. The oxidation of hydralazine is critical in the activation of INH as a pro-drug. This, however, cannot take place if it has been previously N-acetylated (Jacobson, 2011). Metabolism of INH varies between populations. This is especially important since it plays a critical part in understanding the outcome of TB treatment with respect to the response of *M. tuberculosis* to the drug and the emergence of new, resistant, strains (Evans, 1960). Resistance to INH is increasing among TB patients worldwide (Chakraborty, 2013). The differences in drug metabolism between different ethnic groups was pioneered by studies on the family of drug metabolising enzymes known as arylamine N-acetyltransferases (NAT) (Sim, 2014).

1.4 Arylamine N-acetyltransferase (NAT)

1.4.1 Acetylation phenotype

Acetylation is the major route of biotransformation of therapeutic drugs and is catalysed by the NAT enzymes (Butcher, 2002). The 2-step reaction is said to have a “ping pong bi bi” mechanism in which acetylation occurs (Riddle and Jencks, 1971; Hein, 2009). The mechanism was solved using a pigeon model of NAT in which they aimed at elucidating the intermediate formed in the acetylation reaction (Riddle and Jencks, 1971; Pompeo, 2002). The enzyme was studied using a range of different *in vitro* and *in vivo* assays to test the activity and content of human NAT in the liver (Grant, 1990). Research suggested that although animals contain similar quantities of NAT in the liver, they differ with respect to their kinetic response to different arylamine substrates. Scientists concluded that the kinetic and thermal characteristics of human liver NAT define the differences between acetylator phenotypes (Hearse and Weber, 1973). The acetylation phenotype is said to be bimodal (Evans, 1960). Studies in Asian populations show a trimodal acetylation phenotype. Furthermore, the human population can be divided in groups of rapid and slow acetylators of arylamines. This functional aspect was discovered using rodent models with extensive homology to human NATs (Pompeo, 2002). It is also believed among scientists that the proportions between these categories vary among different ethnicities with 10% of slow acetylators being Oriental and 40-70% being Caucasian (Lin, 1993; Nebert, 1997; Windmill, 2000). Research on the rate of INH inactivation shows that the slow acetylator phenotype is better for TB treatment and drug resistance research (Hearn and Cynamon, 2004). However, evidence exists which states that certain arylamines (*p*-aminobenzoic acid and *p*-aminosalicylic acid) cannot be classified into the different acetylator categories. They are said to be monomorphic substrates as compared to the polymorphic isoniazid substrate.

In 1973, Hearse and Weber carried out studies on rabbit subjects and found that the distribution of NAT in the body (extra-hepatic) affects the total drug-acetylation capacity of an individual. This is mostly seen to affect the slow acetylators. Results of their study also showed that extra-hepatic NAT was found to contribute as much as two thirds of the

acetylating capacity, that multiple forms of NAT exist in different tissues and that certain tissues contained more than one form (Hearse, 1973). Previous research results observed two different peaks of enzyme activity when testing human livers for the kinetic activity of NAT. They concluded that two isozymes exist, NAT1 and NAT2, and that these isozymes are responsible for the ultimate acetylation of monomorphic and polymorphic substrates. In light of this notion, scientists believed that the *NAT2* gene locus is responsible for the acetylation polymorphism and that NAT1 is expressed independently (Grant et al, 1990; Goodfellow, 2000). Grant (1990) tested 50 human livers and results showed that slow acetylation is directly proportional to the decrease in quantity of both NAT1 and NAT2. Arslan and colleagues (2003) continued research done on NAT1 mutations and their effect on the enzyme activity. Meyer (1994) used mouse and human livers as models and proved that NAT2 has a higher affinity for polymorphic substrates, such as INH and sulphamethazine (SMZ), compared to NAT1 which shows higher affinity for monomorphic substrates such as p-aminobenzoic acid (PABA) and p-aminosalicylate (PAS). Furthermore, results identified seven possible substitution sites – position 191, 282, 341, 481, 590, 803 and 852 in NAT1. These substitutions are between one and three nucleotides. Allelic variation affects substrate specificity and catalytic activity of the isozymes, as well as the inability to distinguish the acetylator phenotype. This increases the susceptibility of humans to many infectious and malignant diseases (Meyer, 1994).

1.4.2 Mycobacterial NAT

Over the years, scientists attempted to identify homologous NAT sequences in prokaryotes. One of the major reasons was to determine whether these enzymes have an endogenous role (Sim, 2003). A homolog of this *NAT2* gene in *M. tuberculosis* was over expressed in *M. smegmatis* (*M. smeg*) and INH resistance increased three fold. Upton and colleagues (2001) also found that the survival of the host improved. This was due to the N-acetylation of the drug within the bacterium. In addition, they found that 13 out of 73 random people residing in the Western Cape have a G/A mutation at position 619 of *NAT2* open reading

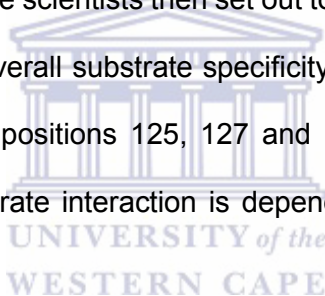
frame. The G619A SNP affects the overall structure of the enzyme and the subsequent stability and function. They considered the fact the human and *M.tuberculosis* NAT have similar substrate specificity. Thus, the bacterium would metabolise similar compounds and plays an endogenous detoxification role.

1.4.3 Human Arylamine N-acetyltransferases (NAT) genes

Three human NAT genes have been characterised - NAT1, NAT2 and, a pseudogene, NATP. These genes are found on chromosome 8p22 and occur relatively close to each other. NAT1 and NAT2 are isoenzymes and products of single exons of 870 base pairs (bp) open reading frames (ORFs) (Blum et al. 1990, Vatsis et al. 1991, Payton, 1998). The two genes share 87% homology and yield 55 amino acid differences (Sinclair, 1997; Akaguru, 2012). This family of polymorphic enzymes is believed to be the first enzymes to confer inter-individual variation in drug responses (Pompeo, 2002; Patin et al, 2006). The NAT family is known for their role in acetylating arylamines, arylhydroxylamines and arylhydrazines with acetyl coenzyme A as the acetyl donor (Liu, 2007). They are capable of N-, O-, and limited N,O-acetyl transfer reactions (Goodfellow, 2000; Akaguru, 2012). The model structure used for NAT was first discovered in *Salmonella typhirium* (*S.typhirium*). Consequently, the functional impact of the polymorphisms are more easily elucidated with a direct structure in place (Pompeo, 2002). The polymorphic marker, D8S21, is found within the NAT2 locus. After comparing the sequences of the two isoenzymes, scientists noted that most of the amino acid differences are borne in the C-terminal of the molecules. The distribution of NAT2 is limited to the liver and intestinal epithelial cells. In contrast, the distribution of NAT1 is ubiquitous which gives plausibility to the suggested endogenous role of the resultant enzyme which is believed to be specific for the substrate para-aminobenzoylglutamate (pABGlu). Knowledge about the accurate localization of the NAT enzymes will give insight into the specific function. In addition to this, NAT1 is notably present in fetal and neonatal tissues. NAT1 is present early on and its involvement in metabolic pathways is understated, albeit important (Spielberg, 1996; Windmill, 2000; Patin et al, 2006). Contrary to this, NAT2

is only detected 12 months after birth (Butcher, 2000). *NatP* has been shown to have many terminator codons and is not expressed in humans (Windmill, 2000).

The structure of NAT defines the mechanism of action. The amino acid sequence of proteins plays a major role in substrate specificity and subsequent enzyme function (Goodfellow, 2000). The difference in substrate specificity is explained by the observation that the entire transcript of *NAT1* is derived from a single exon whereas the transcript of *NAT2* is derived from a protein-coding as well as a second non-coding exon. The latter spans a length of 100 bp and is situated 8 kilobases (kb) upstream from the translation start site (Hein, 2000). Goodfellow (2000) sought to determine the substrate specificity of the respective enzymes by observing their kinetic characteristics. They found a highly conserved central region in the respective wild type alleles for each enzyme which differs at three positions and affects the kinetics of the enzyme activity. The scientists then set out to investigate the contribution these three amino acids make to the overall substrate specificity and stability of NAT1 and NAT2. The results showed amino acid positions 125, 127 and 129 has a major effect substrate specificity and the enzyme-substrate interaction is dependent on weak charge interactions (Goodfellow, 2000).



Murine NAT2, the homologue to human NAT1 is shown to be important in development. This is due to its expression in embryonic neuronal tissue as well as the heart and gut (Kawamura, 2005). Mouse models were used often because of the close relation between the homologues.

1.4.4 Arylamine N-acetyltransferase 1 (NAT1)

NAT1 was first identified and sequenced in 1990. Subsequent expression studies were performed and many polymorphisms were identified which alter catalytic activity of the protein it encodes (Blum, 1990). The NAT1 protein is approximately 33 kD in size (Minchin, 2007). In contrast to its isoenzyme, many variations within NAT1 have not been fully characterized (Grant, 1997). Human NAT1 is widely distributed in the body (Minchin, 2007). However, heterogeneity exists in the levels of NAT1 expression. Expression studies showed

that NAT1 and NAT2 were independently expressed (Butcher, 2003). The coding region of the gene spans a length of 870bp and the protein is 289 amino acids in length. The open reading frame of the gene is found on a single exon, exon 9, while the remaining 8 exons are found in the non-coding 5' UTR region, as seen in figure 2. Two, distinct promoters are present in the gene. These are named NATa and NATb and are found 51.5 kb upstream of the coding region and upstream of exon 4, respectively. NATa seems to produce transcripts in higher levels in the liver, lungs, kidneys and trachea. Contrastingly, NATb produces a bigger selection of transcripts which present in most human tissues (Hein, 2009). The official transcription start site for the NAT1 protein is found 11.8 kb upstream of the official ATG at the start of the coding region (Husain, 2004). Many studies report multiple transcripts containing various combinations of the untranslated 5'UTR exons as a result of these promoters (Minchin, 2007).

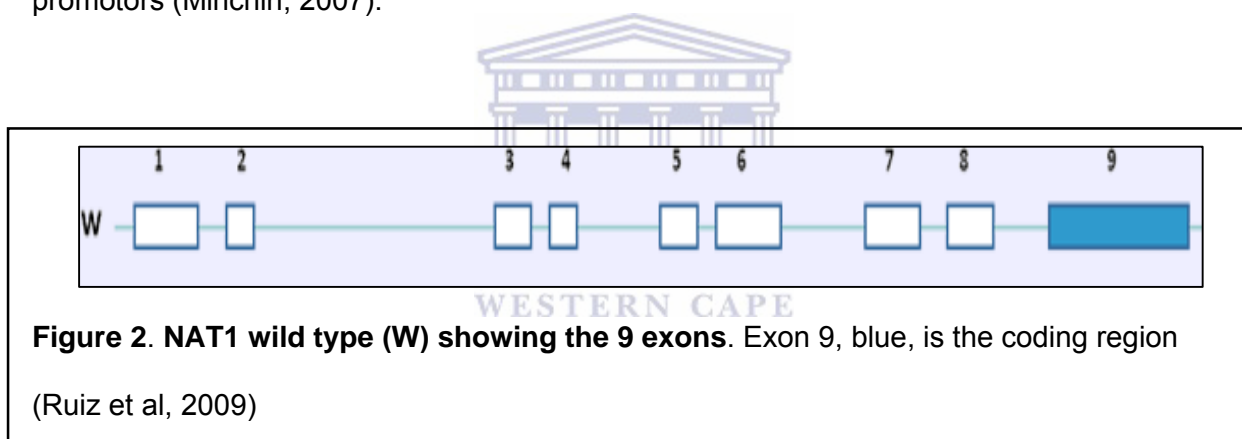


Figure 2. NAT1 wild type (W) showing the 9 exons. Exon 9, blue, is the coding region (Ruiz et al, 2009)

Wu and colleagues (2007) managed to produce crystal structures of both human NATs, including NAT1 in complex with the irreversible inhibitor 2-bromoacetanilide, a site-directed NAT1 mutant, NAT1_F125S, and a NAT2-CoA complex.

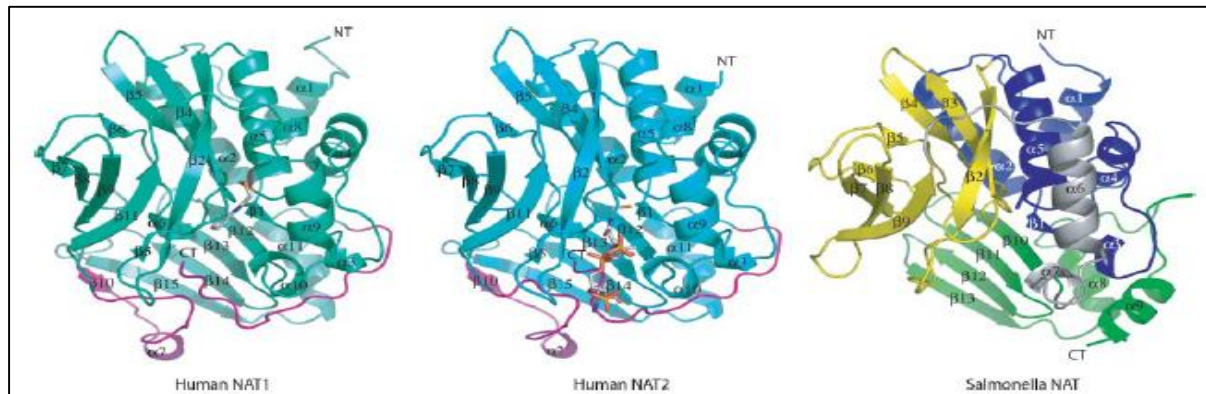
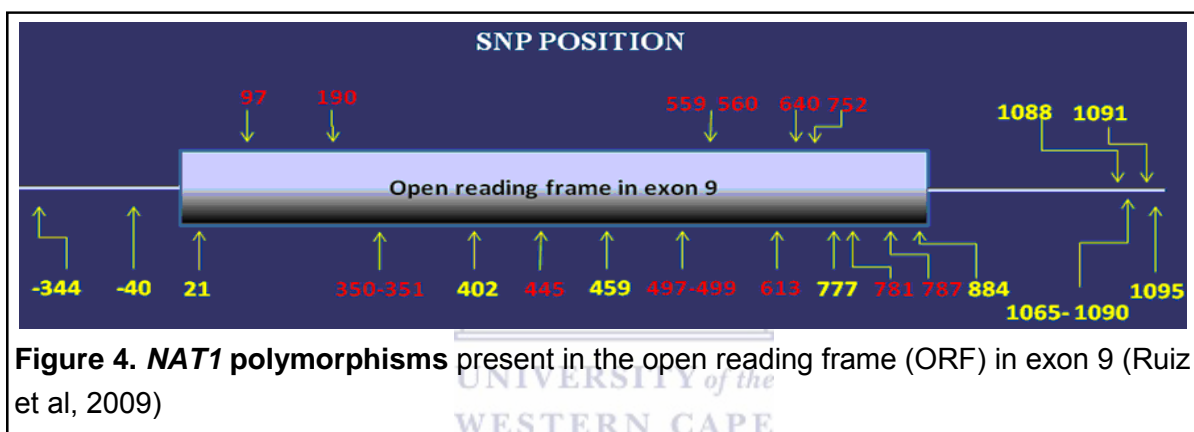


Figure 3. Overall protein structure of human NAT1 and NAT2. Ribbon diagram showing the structure of the human NAT1 (left, with acetanilide covalently bound to catalytic cysteine), NAT2 (middle, in complex with CoA), and *Salmonella typhimurium* NAT (right). The three structures are presented in the same orientation. All strands and helices are labelled. The amino and carboxyl termini are labelled as NT and CT, respectively (Wu et al, 2007).

In *M.smeg*, the model organism for *M.tuberculosis*, the active site of the enzyme contains a catalytic triad of Cysteine, Aspartate and Histidine residues (Sim, 2003). This was validated in the human NAT protein modification techniques which were used to investigate the catalytic activity of the respective NAT enzymes. The Cys68 residue was directly associated with acetyl transfer (Goodfellow, 2000). Rodrigues-Lima (2001) discovered a highly conserved residue, Arg64 and shown to be important in the conformational structure of NATs in prokaryotes. Using NAT isolated from *S.typhimurium* (*st*NAT) as a template, Rodrigues-Lima produced a computational structure of human NAT1. A catalytic triad (Cys68, His107 and Asp122) to be present in the alpha-beta motif of human NAT1 and believed that it was conserved from *st*NAT. This proved that the 3-dimensional structure of the N-terminal of *st*NAT and human NAT1 are similar. A loop in the catalytic pocket was discovered which originated at the Arg122 residue. It is believed that this loop is important in substrate selectivity of NAT1 as well as the binding of AC-CoA (Rodrigues-Lima et al, 2001). The C-terminal of NAT1 has a coiled structure which enables it to interact with the catalytic pocket, thereby regulating its size and shape. This in turn adds to the substrate specificity of the enzyme (Walraven, 2008).

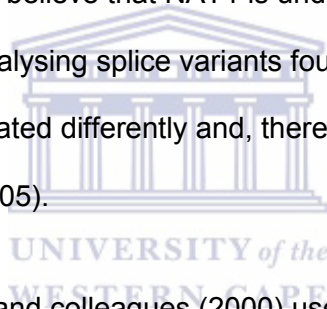
Allelic variation plays a major role in determining rate of drug metabolism (Grant, 1997). The most common genetic variation found in humans are called SNPs. SNPs are characterized by a difference in a single nucleotide in a stretch of DNA. It is believed that if a SNP occurs in the regulatory or coding region of a gene, that it plays a more direct role in disease predisposition. As research progressed, SNPs were discovered in the NAT gene. The “A” of ATG is designated as number 1, positions upstream are annotated by negative numbers and with positive numbers for those located downstream of number 1. The *NAT1*4* allele is considered to be the wild-type allele. According to Hein (2009), the evolution of molecular modelling aided development of crystallization structures of *NAT*. Subsequently, the function of SNPs in NAT1 can now be better understood. Figure 4 shows some of the SNPs already discovered.



The *NAT1*4* allele is considered to be the wild-type allele and found to occur at higher frequencies among the Caucasian population. Comparatively, *NAT1*10* occurs at higher frequencies in the African, Asian and Chinese populations. Furthermore the wild type allele was seen to appear at a lower frequency in Chinese populations (Zhangwei et al, 2006). This allele has been of great interest to scientists as it was believed to be a biomarker for disease and results in the rapid acetylator phenotype of an individual (Sim et al, 2008). Understanding the functional effect of the SNP gives insight to how the individual is predisposed to the disease.

Due to the affinity of NAT1 for PAS, the substrate was used as a probe drug in functional studies conducted in Toronto. Grant (1997) were able to detect novel variants – *NAT1*14* and *NAT1*15*. The former represents ARG¹⁸⁷→Gln amino acid change which decreases the affinity for PAS. The remaining variant represents a premature stop codon (ARG¹⁸⁷→Stop)

which truncates and inactivates the protein. They also discovered a previously identified variant (*NAT1*11*). Uncertainty concerning the effect of this variant on NAT1 function still remains since the amino acid change (Ser214Ala) does not affect the activity of the enzyme when expressed. Another study reported by Zhangwei and colleagues (2006) also used PAS as a probe drug and found a slow acetylator phenotype to be prominent in the Caucasian Australian population. The study proved that acetylation by NAT1 was co-dominant. In contrast to the study conducted by Grant (1997) and Sim and Payton (1998) designed a study in a non-related, Caucasian population using PABA as substrate in order to identify existing and possible novel variants. Previous functional studies observed the activity of NAT1 in peripheral blood mononuclear cells and noted a time-dependent loss in activity of the enzyme. This led scientists to believe that NAT1 is under substrate-dependent regulation. Studies involved in analysing splice variants found that the variants influence cell-specific expression and are regulated differently and, therefore, ultimately influence NAT1 enzyme activity (Butcher et al, 2005).



The study conducted by Butcher and colleagues (2000) used cells cultured in PABA-supplemented media and tested NAT1 activity. They proved the hypothesis that specific NAT1 substrates have the ability to down regulate their enzymatic activity. An interesting finding in this study was that NAT1 activity was also down regulated when they tested it in THP-1 monocytic cells. They found that loss of activity occurred when growth arrest was evident. Thus, the down regulation was observed to be growth dependent. Butcher's study also proved that this down regulation was not specific to PABA but was observed in other arylamine substrates. The only exception was sulfamethoxazole SMX. This could be attributed to the structural differences between PABA and SMX. Badawi and colleagues (1995) studied NAT1 activity in bladder mucosa cells and also used PABA as a probe drug. They discovered that the *NAT1*10* allele was associated with the rapid phenotype of an individual. This is supported by a similar study done using colorectal tissue as test cells (Badawi et al, 1995).

1.5 Toxicological effects of NAT

1.5.1 Carcinogenesis

Research shows that NAT enzymes also play a major role in the bioactivation of carcinogens. Studies have reported a relationship between *NAT1* and *NAT2* variants and cancers related to aromatic amines (Wu, 2007). Minchin and colleagues (2007) found that *NAT1*10* is associated with the rapid acetylation phenotype and increased susceptibility to certain cancers. Polymorphism in *NAT1* and *NAT2* affect the relative risk of developing cancer after being exposed to certain carcinogens (Hein, 2000). Hein established that the different acetylation genotypes of individuals play a role in the degree of susceptibility they will have toward cancer. Evidence exists from studies over the decades which show an association between acetylation phenotypes and spontaneous, drug-induced cancer and other toxicological effects (Blum, 1990). Blum tested the theory using Syrian hamsters as model organisms. The aromatic amine carcinogen 3,2'-dimethyl-4-aminobiphenyl (DMABP) was administered to rapid and slow acetylator hamsters and the adducts formed in the different tissues was measured. Hein (2000) found that *NAT2* slow acetylators are more susceptible to bladder and prostate cancer. Since *NAT1* and *NAT2* are prominent in the colon, it is to be expected that individuals would have an increased risk to colorectal cancer. Hein (2002) investigated the effect of rapid *NAT1* and *NAT2* acetylator genotypes on the risk of developing colorectal cancer. He found that *NAT1*10* was associated with increased risk within the Chinese population. However, more research will need to be done to fully understand the relationship between environmental influences and acetylation phenotypes (Hein, 2002). As previously mentioned, knowledge about the precise localization of NAT enzymes is important. Windmill and colleagues (2000) stated that it'll give insight into site-specific tumorigenesis and designed an experiment in which he used hybridization histochemistry to characterize the expression of the two NAT enzymes at the mRNA level. They found *NAT1* and *NAT2* to be present in the epithelial cells of the small intestine and colon. Results showed that the transcripts of these enzymes appear to be prominent in the gastrointestinal tract and in the urothelium of the human. In addition, *NAT1* and *NAT2* are

expressed in the mammary glands and NAT1 is overexpressed in breast cancer cells. Many studies analyse the expression of NAT1 in breast cancer cells. One such study was done by Adam and colleagues (2003) that NAT1 was unregulated in invasive ductal carcinoma and invasive lobular carcinoma.

1.6 TB treatment

1.6.1 Anti-tubercular drugs and hepatotoxicity

TB continues to threaten populations worldwide because of the drug resistant *M.tuberculosis* strains. It was observed that drug resistance is highest in industrialised countries (Upton, 2001). It is clear that early diagnosis of drug resistant strains of *M.tuberculosis* would allow scientists and clinicians to adjust drug therapy. Scientists agreed that the early detection of INH serum levels would aid the determination of any resistance to the drug and the possible propagation of the disease. Serum levels are major influences of acetylation by the NAT enzymes (Huang et al, 2002).

To optimize combination therapy, the efficacy of monotherapy will have to be optimized (Gumbo et al, 2007). The treatment regimen for TB, as stipulated by the WHO, consists of a multidrug approach including isoniazid, rifampicin and pyrazinamide. More recently, it has become a five-drug regimen, including streptomycin and ethambutol (Babalik et al, 2012). Due to their toxic potential, these anti-tubercular drugs commonly cause drug-induced hepatotoxicity (Sharma et al, 2005). Scientists found that treatment regimens induce acute or chronic hepatitis. Anti-tuberculosis drug-induced hepatotoxicity is primarily attributed to INH metabolism and the acetylator status of the patient. This is mainly because of the intermediates formed during INH metabolism (Huang et al., 2002). Drug-induced liver injury (DILI) caused by TB is rated anything between 5% and 33%. Furthermore, it was reported in a study done on TB patients in a 3-year cohort, that different factors predispose patients to DILI and that it can be managed (Babalik, 2012). Some of the factors include age, sex, alcohol consumption, malnutrition, predisposition to liver injury, hepatitis infections, activity of

glutathione S-transferase (GST), N-acetyltransferase and acetylator phenotype of an individual (Sharma, 2005).

1.6.2 NAT acetylation and TB treatment efficacy

The acetylation phenotype of NAT was first discovered in the treatment of TB with and the metabolism of INH, the anti-tubercular drug (Hein, 2009). It was reported that scientists should be aware of the specific mutations which cause INH resistance, in terms of NAT expression. Studies were undertaken that highlighted the difference in the drug-metabolising capacity of TB patients. Many of these variations are monogenic and polymorphic. A study also highlighted the ethnic differences in this metabolising mechanism and acetylation phenotype (Hearn, 2004).

Previous knockout studies using *M.smeg* showed an increased lag phase for bacterial growth. This proves that NAT from *M.tuberculosis* is a strong drug target for anti-tubercular therapy. A study was conducted among INH-mono-resistant TB patients in rural Western Cape in order to determine the treatment outcome. Results showed that 16% had poor outcomes and 61% developed MDR-TB. In addition, it was found that resistance conferring mutations made the outcomes worse (Jacobson et al., 2011).

In Southern Africa, the success rate is much lower and more worrying when compared to the rest of the world. This could be attributed to the synergy between HIV immunocompromised patients infected with TB (Blöndal, 2007). Between 2000 and 2015, the TB death rate decreased by 22% and 37 million lives were saved through diagnosis and effective treatment. The End TB Strategy, approved by the World Health Assembly in 2014, aims at reducing TB fatalities by 90% and TB incidence by 80% by 2030 (Global Tuberculosis Report, 2016). Management and control systems were put into place and research continues. To date, scientists conclude that early diagnosis, an adequate number of TB specialist nurses as well as the control of transmission is essential to the success rate of TB treatment (Caylà and Orcau, 2011). Furthermore, for new drug discovery and optimum efficacy, the pharmacodynamic-pharmacokinetic

relationship needs to be well understood, as well as the acetylation phenotypes and functions affected by NAT polymorphisms (Gumbo et al, 2014).

1.7 Hypothesis and aims

1.7.1 Hypothesis

N-arylamine acetyltransferases, NAT1 and NAT2, are important drug metabolising enzymes and play a major role in TB resistance to treatment efficacy. More importantly, the *NAT* gene confers acetylator phenotypes to individuals. Subsequently, we are able to understand their function via expressing the resultant proteins and determining their level of activity with respect to the different alleles they present. Previously, PolyPhen-2 and SIFT, predicted that NAT1 SNPs V231G, V235A and D229H affects the stability of the protein and thus, it's function. Cloete and colleagues (2017) used these computational predictions and continued the investigation using structural prediction methods. The findings for SNP V231G were reaffirmed in the structural analyses conducted by Cloete and colleagues (2017). SNP F202V was predicted to affect the protein function but not the structure. Two SNPs, T193S and T240S, were predicted to be benign. We then hypothesised that these results will persist after experimental validation.

1.7.2 Aims of the study

- To construct the wild type NAT1, and the 6 novel NAT1 SNPs using site-directed mutagenesis
- To clone and express the wild type NAT1 as well as the different 6 novel non-synonymous SNPs (T193S, T240S, V231G, V235A, F202V, D229H) in a bacterial system
- To determine acetyltransferase activity by studying the rate of acetylation and consequent drug degradation in the presence of the SNPs using PABA and PAS as probes and PNPA as the cofactor.



Chapter 2 - Materials and Methods



UNIVERSITY *of the*
WESTERN CAPE

2.1 Plasmid and reagents

For the initial purpose of cloning, previously purchased pGEX-4T-2, with restriction endonuclease (RE) cut sites for *NotI* and *XhoI*, was obtained for the Medical Virology laboratory at the University of the Western Cape. It was streak plated on Luria Broth (LB) agar plates and incubated overnight at 37°C. A single colony was then inoculated into LB broth for a further 16 hour incubation at 37°C. Cells were then harvested by centrifugation at 4°C for 10 minutes at 4000 revolutions per minute (rpm). A midi-prep protocol was followed using a plasmid purification kit and a RE digest with *XhoI* was done to confirm the vector is empty and could be used for subsequent cloning. In addition to this, *NAT1*-pUC57 was purchased along with JM109 competent cells, from Promega™, for transformation purposes and final ligation into pGEX-4T-2 for protein expression. However, the confirmation of *NAT1*-pUC57, by polymerase chain reaction (PCR), was unsuccessful and the cloning process was not ensued.

MCF-7 cells, passage 37, was obtained from the Medical Biosciences Department at the University of the Western Cape. The cells were maintained and once confluent, were passaged and used for further molecular extractions.

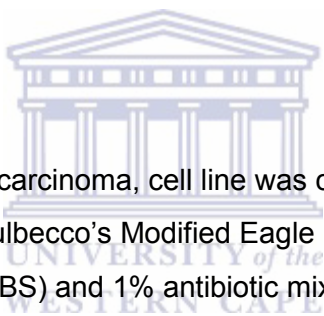
After the failed attempts at making the constructs and due to time constraints, the independently cloned plasmids were purchased from GenScript USA Inc. The *NAT1* wild type (WT) open reading frame was obtained from the National Centre for Biotechnology Information (NCBI) with the accession number shown here (Genbank AJ307007). The WT as well as the six SNPs (T240S, V231G, F202V, V235A, D229H, T193S) were independently cloned into the multiple cloning site (MCS) of the pET-30a(+) vector, with the addition of a polyhistidine tag (poly-His tag), in the site between the RE sites, *NdeI* and *XhoI*. The letters indicates the protein change and the number indicates the position at which the change occurs. The *NAT1* WT nucleotide sequence as well as the protein sequence is seen in appendix I. The plasmids were received as individual 4 µg dry pellets and were resuspended in 20 µl distilled water and stored at -20°C.

The kits used for, DNA isolation via mini-preps, protein purification and SDS-PAGE gels as well as Western Blot transfers were obtained from Promega™ and Sigma-Aldrich®.

2.2 Bacterial strains

The bacterial strain used for this study, *E.coli* BL21(DE3)pLysS, was supplied by Promega™. The reason for this choice is attributed to its specificity for the pET expression system. The pET-30a(+) vector contains a Kanamycin (KAN⁺) resistant gene and is inducible by lactose or its analogue, Isopropyl β-D-1-thiogalactopyranoside (IPTG). The strain is deficient of both the lon and ompT outer membrane proteases. Consequently, the possibility of protein degradation is diminished during the purification step. The pET-30a(+) vector is under the control of a T7 promoter which is highly specified for the use of T7 RNA polymerase. This specific strain of *E.coli* cells contain the gene for this polymerase, thereby aiding transcription effectively. The strain also contains pLysS plasmid which encodes for t7 lysozyme in order to control the level of basal expression of t7 RNA polymerase. Furthermore, the pLysS plasmid also contains a Chloramphenicol (CAM⁺) resistant gene (BL21 competent cells instruction manual).

2.3 Tissue culture



The MCF-7, human breast adenocarcinoma, cell line was cultured in 75 cm² tissue culture flasks. The cells were grown in Dulbecco's Modified Eagle Media (DMEM) supplemented with 10% Foetal Bovine Serum (FBS) and 1% antibiotic mix of Penicillin/Streptomycin (PenStrep). The culture flasks were incubated in a 37°C incubator with a constant 5% carbon dioxide (CO₂) supply. Once the cells were microscopically confirmed to be 80-90% confluent, they were subcultured. The cells were first washed with 3 ml sterile 1X phosphate buffered saline (PBS) and then treated with 2 ml 0.25% Trypsin. The flasks were then incubated at 37°C in 5% CO₂ until the cells detached from the culture flask surface. The solution was then neutralized with 2 ml DMEM, added to a 50 ml centrifuge tube and centrifuged at 2500 rpm for 3 minutes to pellet the cells. The supernatant was discarded and the cells were resuspended in 200 µl 1X PBS. The RNA was then extracted from the cells using the High Pure RNA Isolation Kit from Roche Molecular Systems, Inc.

2.4 Reverse transcription

Complementary DNA (cDNA) was transcribed from the RNA isolated from the MCF-7 cells according to the manufacturers' instructions in the GoScript Reverse Transcription System. The reaction mixture was made up of 25 mM MgCl₂, 5X reverse transcription buffer, 10 mM

dNTP mix, recombinant RNasin ribonuclease inhibitor, 25 U/μl reverse transcriptase, 0.5 μg/μl Oligo dT primer, experimental RNA and nuclease-free water to make up a final volume of 20 μl in a sterile microcentrifuge tube. The control contained all the reagents except the experimental RNA which was replaced by nuclease-free water. All RNA work was done on ice. The transcription conditions were set to 5 minutes at 25°C, 60 minutes at 42°C and inactivation was set to 15 minutes at 75°C. The product was then stored at -4°C until needed.

2.5 Cloning

2.5.1 Transformation of NAT1-pET-30a(+) into BL21(DE3)pLysS

Transformation of the plasmid was done according to the protocol presented in the BL21 competent cell Instruction Manual from Agilent Technologies. Competent cells were thawed on ice for 5 minutes. Parallel to this, 2 μl of each plasmid (wild type and six individual SNPs) was dispensed into individual 1.5 ml eppendorf tubes and kept on ice. After the cells have thawed, 50 μl was added to each of the 2 μl of plasmid DNA. The reaction mixtures were then incubated on ice for 30 minutes. The cells in their respective tubes were then heat shocked at 42°C for 45 seconds in a dry heating block and immediately returned to ice for 2 minutes. Hereafter, 950 μl of sterile LB broth which was preheated to 37°C was added to the tubes and incubated at 37°C in a shaking incubator at approximately 180 rpm for 3 hours.

During the shaking incubation period, LB agar plates were prepared with 1% Bacto-Tryptone, 0.5% Bacto-Yeast extract, 0.5% NaCl and 1.5% agar to a final volume of 500ml. The mixture was then heated to a boil in order to dissolve all the ingredients, after which it is then autoclaved. Once the mixture has cooled, KAN⁺ and CAM⁺ antibiotics were added to final working concentrations of 30 μg/ml and 34 μg/ml respectively. Hereafter, the plates were poured in a fume hood to ensure a sterile environment while they are left to cool and set on a flat surface. Once the plates have set, the following solutions were spread plated and allowed to set: 100 μl of 0.1 M IPTG, 20 μl X-gal. These solutions all aid the process of blue-white colony screening. Two extra plates were included, one for competent cells only and one for the control DNA provided by Promega™. After all the plates have set well with their solutions, 100μl of the transformation product was spread plated onto their respective plates and specifically labelled. All spread plating was done with a hockey stick under a fume hood with the aid of a Bunsen burner to ensure sterility. Extra care should be taken to work aseptically (e.g. Working with gloves). Plates were then incubated for 12-16 hours at 37°C.

2.5.2 Colony picking and starter cultures

Blue/white screening is a molecular technique verifying the successful insertion of a gene of interest into a vector for further experimentation. The principle relies on the histochemical dye, X-gal, and the *lacZ* operon of the *E.coli* strain. While *E.coli* contains a mutant *lacZ* gene, most vectors contain a functional *lacZ* gene. Thus, the vector *lacZ* sequence complements the mutant deficiency of the host and can activate the gene to ultimately produce β -galactosidase. This process is known as α -complementation. However, once foreign DNA is inserted into the hosts MCS, α -complementation cannot occur (Keese and Graf, 1996). In the case of α -complementation, the β -galactosidase hydrolyses the X-gal spread onto the agar plates. The by-product, 5,5'-dibromo-4,4'-dichloro-indigo, allows the blue pigment of the non-recombinant colonies. Thus, the colonies containing the DNA insert appear white. IPTG is an analogue of galactose, therefore inducing expression of *lacZ* (Padmanabhan, 2011).

After the incubation period, white colonies (transformed) were picked from each of the plates. The control plates only have blue colonies. Starter cultures were prepared using a 15ml polypropylene tube with 10ml sterile LB broth containing the necessary antibiotics, one tube for each colony of the respective plates. Special care was taken to label each experiment correctly and work aseptically to avoid any form of confusion or contamination. These tubes were then inoculated with these individual white colonies. The labelled tubes with the respective colonies were then placed in a shaking incubator for a further 12 hours at 37°C.

2.5.3 Extraction of plasmid DNA and DNA quantification

Overnight cultures were purified by means of the PureYield™ Plasmid Miniprep System from Promega™. Briefly, cultures were centrifuged at 4000 rpm for 10 minutes. The supernatant was discarded and the pellet was, gently, resuspended in 100 μ l of nuclease free water. Purification then followed using the manufacturers' specifications (Appendix III). The plasmid DNA was quantified spectrophotometrically using a nanodrop2000. The principle behind this method is that the absorbance (Abs) of 1 μ l DNA is measured through a 1 cm path length at 260 nm. The complete protocol can be seen in Appendix III.

2.5.4 NAT1 verification

2.5.4.1 Gene-specific PCR optimization

PCR is a simple enzymatic assay which allows for the amplification of specific DNA fragments (Garibyan and Avashia, 2014). The introduction of thermal stable DNA

polymerase from *Thermus aquaticus*, *Taq* DNA polymerase, catapulted this innovative technique in many biological disciplines. The technique, although rapid, is still erroneous. As with any other biological assay, the protocol requires optimization and troubleshooting where necessary (Lorenz, 2012).

Many different PCR reactions were performed to optimize the temperature conditions and volumes of the PCR components to ultimately verify the presence of *NAT1*. Initial primers were designed for the entire *NAT1* gene with RE sites for *NotI* (5'-GCGGCCGC-3') and *XhoI* (5'-CTCGAG-3') incorporated into the sequence. Melting temperatures for the forward and reverse primers were 72.83°C and 66.04°C, respectively. The primers were diluted and 10µM working aliquots were made and stored at -20°C (Table 1). The expected product size was ~1615 bp.

Table 1. Initial *NAT1* forward and reverse primers. The primers are written 5' to 3' and include the cut sites for restriction endonucleases, *NotI* and *XhoI* (bold) and the start and stop codons (underlined and italicized).

PRIMER NAME	PRIMER SEQUENCE	LENGTH
NAT1 forward	GCGGCCGC <u><i>GATGCTTTGTATAAGGCTCAGC</i></u>	30
NAT1 reverse	CCCCTCGAGT <u><i>GAGAATTCAACAATAAACC</i></u>	29

2.5.4.1.1 *cDNA gene-specific PCR amplification*

For the cDNA validation, PCR reactions were performed at annealing temperatures of 50°C and 55°C. The volume of plasmid DNA used varied according to the concentration needed for the reaction. The final PCR mixture was made up of 0.2 mM dNTP mix, 1.5 mM MgCl₂, 5X GoTaq Flexi buffer, 0.5 µl *Taq* polymerase enzyme, 1µl forward and reverse primers and nuclease free water to a final volume of 25 µl. PCR products were electrophoresed on 1% agarose gel for visualisation under UV light. The results for these reactions can be seen in appendix IV.

The PCR results showed no band of interest. Thus, the initial premise of constructing the SNPs failed. Time constraints forced the purchase of *NAT1*-pUC57 for cloning and subsequent expression studies.

2.5.4.1.2 *NAT1-pUC57 PCR amplification*

In another attempt to amplify *NAT1* for cloning purpose, *NAT1-pUC57* was purchased and received as a 4 µg dry pellet and resuspended in 20 µl nuclease-free water. Different optimization reactions were performed. A single set of PCR conditions will not work for all PCR reactions. Thus, the concentrations of individual components will need to be optimized for the specific purpose of the individual PCR. The standard PCR mixture for this project contained 0.2 mM dNTP mix, 1.5 mM MgCl₂, 5X GoTaq Flexi buffer, 0.5 µl Taq polymerase enzyme, 1 µl forward and reverse primers and nuclease free water to a final volume of 25 µl. For this project sample, to overcome smears and primer dimers and non-specific binding, dNTP (PCR nucleotide mix), MgCl₂ and template profile PCR reactions were performed at the most suitable annealing temperature of 54°C. Furthermore, the 4 µg *NAT1-pUC57* template was diluted to a ratio of 2:8 with nuclease free water. The already diluted *NAT1-pUC57* was then further diluted to a ratio of 1:10 with nuclease free water. Both dilutions were PCR amplified and it was found that the further diluted sample gave a clearer band. Thus, all the subsequent PCR amplification reactions were performed with said sample. A negative control (no DNA) as well as a positive control (previously amplified *NAT1*) was included in the PCR reactions. All the PCR products were electrophoresed on 1% agarose gels and viewed under ultra violet (UV) light. The results for these PCR reactions can be seen in appendix IV. The specifications for these optimization reactions can be seen in the tables below.

The first step was to optimize the temperature and so 1 µl of the 4 µg *NAT1-pUC57* solution was PCR amplified at annealing temperatures of 50°C, 52°C, 54°C, 56°C, 58°C and 60°C with the standard mixture for the other constituents of the reaction. DNA denaturation is often seen in PCR reactions. The changes in annealing temperature will enhance specificity of the primers for the template and improve stability of the product. Once the annealing temperature and the proper dilution was set, a template profile was done for the *NAT1-pUC57* sample. For the template profile PCR, the volume of the DNA template varied from 0-5 µl. The constituents for this PCR profile can be seen in table 2.

For the PCR nucleotide mix profile PCR, the volumes for the PCR nucleotide mix varied from 0-2 µl with all other volumes staying constant except the nuclease-free water. The change in the volume affects the final concentration of PCR nucleotide mix added to the PCR mixture. For this reaction, the PCR constituents with the necessary changes in volume and concentration can be seen in table 3.

The volumes for MgCl₂ in the MgCl₂ profile PCR varied from 0-3.5 µl. Too little magnesium, in the presence of dNTP will give a lower yield of PCR product. However, too much free magnesium will promote misincorporation and more artificial products forming.

The

manipulation of these components typically affects the fidelity of the PCR product. The PCR mixture can be seen in table 4, with the necessary changes in volume and concentration for MgCl₂.

All concentrations and volumes were calculated using the $C_1V_1=C_2V_2$ formula where C_1 was the working concentration, V_1 was the starting volume, C_2 was the final concentration and V_2 was the final volume. None of the amplification reactions results were clear enough to use for further studies. It was concluded that the primers were erroneous since they were ordered with the specifications to amplify the complete *NAT1* gene and not the ORF, which was the aim for this project.

Table 2. DNA template PCR profile constituents. The table shows the volume changes (red) for DNA template as well as the subsequent concentration changes made to the PCR mixture. The concentration which yielded the best result is highlighted in blue.

CONCENTRATION (µg)								
DNA	0	0.16	0	0.16	0.32	0.48	0.64	0.8
VOLUMES (µl)								
DNA	0	1	0	1	2	3	4	5
Buffer	2	2	2	2	2	2	2	2
forward primer	1	1	1	1	1	1	1	1
Reverse primer	1	1	1	1	1	1	1	1
MgCl ₂	3	3	3	3	3	3	3	3
Taq	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
dNTP	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Water	17	16	17	16	15	14	13	12

The best PCR results contained 1µl of the 4 µg DNA template resulting in a final concentration of 0.16 µg in the PCR mixture.

Table 3. PCR nucleotide mix PCR profile constituents. The table shows the volume changes (red) for the PCR nucleotide mix as well as the subsequent concentration changes made to the PCR mixture. The concentration which yielded the best result is highlight in blue.

	CONCENTRATION (mM)					
PCR nucleotide mix	0.2	0	0.2	0.4	0.6	0.8
	VOLUMES (µl)					
cDNA	1	1	1	1	1	1
Buffer	2	2	2	2	2	2
forward primer	1	1	1	1	1	1
Reverse primer	1	1	1	1	1	1
MgCl ₂	3	3	3	3	3	3
Taq	0.5	0.5	0.5	0.5	0.5	0.5
dNTP	0.5	0	0.5	1	1.5	2
Water	16	16.5	16	15.5	15	14.5

Along with the previously described DNA template volume and concentration, adding 0.5 µl of 10 mM PCR nucleotide mix stock solution to a final concentration of 0.2 mM yielded a more conclusive PCR result.

Table 4. MgCl₂ PCR profile constituents. The table shows the volume changes (red) for MgCl₂ as well as the subsequent concentration changes made to the PCR mixture. The concentration which yielded the best result is highlighted in blue.

CONCENTRATION (mM)									
MgCl ₂	1.5	0	0.5	1	1.5	2	2.5	3	3.5
VOLUMES (µl)									
cDNA	1	1	1	1	1	1	1	1	1
Buffer	2	2	2	2	2	2	2	2	2
forward primer	1	1	1	1	1	1	1	1	1
Reverse primer	1	1	1	1	1	1	1	1	1
MgCl ₂	1.5	0	0.5	1	1.5	2	2.5	3	3.5
Taq	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
dNTP	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Water	17.5	19	18.5	18	17.5	17	16.5	16	15.5

In addition to the DNA template and PCR nucleotide mix concentrations, 3 µl of 25 mM MgCl₂ stock solution was added to the PCR master mix for an even better result.

2.5.4.2 Gene – specific PCR amplification of NAT1-pET-30a(+) SNPs

In order to verify the presence of the purchased NAT1-pET-30a(+) SNPs plasmid DNA, a gene-specific PCR was done of the mini-prepped samples. Forward and reverse primers for NAT1 were specially designed using the human NAT1 ORF sequence from NCBI as a template. The specific sequences for the restriction endonucleases, *XhoI* (5'-CTCGAC-3') and *NdeI* (5'-CATATG-3'), were added to the primer sequence as well as the specific start and stop codons necessary for optimal amplification (Table 5). The PCR mixture was made up to a final volume of 25 µl. The volume of plasmid DNA used varied according to the concentration needed for the reaction. The final PCR mixture contained ~0.16 µg template DNA, 0.2 mM PCR nucleotide mix, 3 mM MgCl₂, 2 µl of 5X GoTaq Flexi buffer, 0.5 µl Taq polymerase enzyme, 1 µl forward and reverse primers and nuclease free water to a final volume of 25 µl. The annealing temperature, 60°C, for a PCR is typically set 5°C below the highest melting temperature (T_m) of the primers. In this case the forward primer T_m was

66.40°C and that of the reverse primer was 64.07°C. The primers' sequences can be seen in table 5 and the PCR conditions can be seen in Table 6.

Table 5. NAT1 forward and reverse primers. The primers are written 5' to 3' and include the cut sites for restriction endonucleases, *NdeI* and *XhoI* (bold) and the start and stop codons (underlined and italicized).

PRIMER NAME	PRIMER SEQUENCE	LENGTH
NAT1 forward	CGGCATATGATGGACATTGAAGCATATGCC	30
NAT1 reverse	CCGCTCGAGAAATAGTAAAAAATCTATCGCC	30

Table 6. PCR conditions for NAT1 amplification. The table below describes the different conditions at which the PCR reaction was performed.

CONDITION	TEMPERATURE (°C)	TIME (minutes)	CYCLES
Predenaturation	95	3	1
Denaturation	95	0.45	30
Annealing	60	0.45	30
Extension	72	0,45	30
Extension	72	15	1
Soak	4	∞	∞

2.5.4.3 Restriction endonuclease digest

2.5.4.3.1 RE digest of pGex-4T-2

A 1 µl of this specific midi-prepped DNA was used in this reaction. The RE used, 10 U/µl *XhoI*, was compatible with 10X reaction buffer H and performed optimally at 37°C. The reaction mixture also contained 10 mg/ml BSA. The reaction was done to linearize the sample and, therefore, the expected product size was ~6000 bp. The reaction product was run on a 1% agarose gel with Ethidium Bromide (EtBr) for visualisation under UV light. The reaction constituents and conditions can be seen in tables 7 and 9 respectively. The result can be seen in Appendix V, figure 8.

Table 7. RE digest reaction constituents for pGEX-4T-2

REAGENT	VOLUME (μ l)
template DNA	1
BSA	2
Buffer H	2
Restriction enzyme (<i>Xho</i> I)	1
Nuclease free water	14
FINAL VOLUME	20

2.5.4.3.2 RE digest of NAT1-pET-30a(+)

Of the mini-prepped DNA, 2 μ l were cut using the *Nde*I and *Xho*I restriction endonucleases. Both enzymes have affinity for buffer D and both perform optimally at 37°C. The reaction mixture was made up of plasmid DNA, 10 mg/ml Bovine serum albumin (BSA), 10X buffer, 10 U/ μ l restriction endonucleases (*Nde*I and *Xho*I) and nuclease free water to a final volume of 30 μ l (table 8). The reaction took place in a thermocycler under the conditions specified in table 9. The reaction mixture can also be seen in table 8.

Table 8. RE digest reaction constituents for NAT1-pET-30a(+)

REAGENT	VOLUME (μ l)
Plasmid DNA	2
BSA	2
Buffer D	2
Restriction enzyme (<i>Nde</i> I and <i>Xho</i> I)	0.2 each
Nuclease free water	23.6
FINAL VOLUME	30

Table 9. RE digest reaction conditions

TEMPERATURE	TIME
37°C	2 hours
4 °C	10 minutes
65 °C	20 minutes
4 °C	∞

2.5.4.4 Agarose gel electrophoresis

Agarose gel electrophoresis separates DNA fragments according to their size. The polymers that make up the gel form pores through which DNA molecules can sieve in the presence of an electrical current and suitable buffers. Lee stated in the article that “The rate of migration of a DNA molecule through a gel is determined by the following: 1) size of DNA molecule; 2) agarose concentration; 3) DNA conformation; 4) voltage applied, 5) presence of ethidium bromide (EtBr) , 6) type of agarose and 7) electrophoresis buffer”. Visualization is aided by the addition of EtBr which can be seen under UV light (Lee et al, 2012).

PCR and restriction digest products were run on 1% agarose gels made up of 1 g Seakem® agarose dissolved in 100 ml 1X TBE buffer and prepared in gel tanks. The mixture is boiled to dissolve all the constituents and allowed to cool. EtBr was added at a concentration of 0,001% (v/v) before pouring the mixture into gel plates to solidify. The DNA products were diluted in a 1:5 ratio with 6X Blue/Orange loading dye (0.4% (v/v) orange G, 0.03% (v/v) bromophenol blue, 0.03% (v/v) xylene cyanol FF, 15% (v/v) Ficoll® 400, 10mM Tris-HCl (pH7.5) and 50 mM EDTA (pH8). The migration pattern of the DNA fragments were observed and compared to a commercially, predetermined DNA molecular weight marker from Promega™ (1 kb or 100 bp). The gels was electrophoresed at 80V for 90 minutes 1X TBE as conductance medium. DNA is negatively charged and will, thus, migrate to the positive anode. The gel was viewed under UV light. Thus, the sizes of the bands of interest could be confirmed and verify the presence of *NAT1* based on its size within the vector.

2.5.4.5 Preservation

Glycerol stocks were made in 2 ml cryovials. From the individual overnight cultures, 250 µl were added to individually labelled cryovials. Glycerol was autoclaved and when used, it was preheated to 37°C and 750 µl was added to the vials. The vials were then gently inverted to mix the culture with the glycerol. This was then stored at -80°C until used again.

2.6 Protein expression studies

2.6.1 Standard protein expression protocol

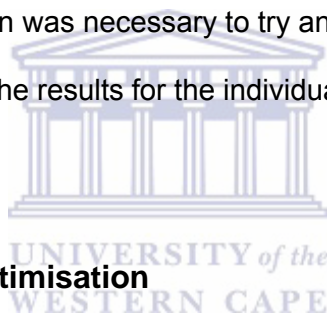
The expression vector used for this study, pET-30a(+), contains a T7 promotor and has specific affinity for BL21 *E.coli* strains. In this case, we opted for BL21(DE3)pLysS. The specific *E.coli* BL21 (DE3) strain has become more popular due to its growth efficiency and its ability to remain non-pathogenic (Khow and Suntrarachun, 2012). The BL21(DE3)pLysS contains T7 polymerase for easy transcription. The pLysS plasmid encodes for the T7 lysozyme gene to reduce basal expression. The T7 polymerase gene is controlled by *lacUV5*. This makes this plasmid inducible by IPTG to produce our target protein for further research. All the factors attribute to the stability and efficiency of protein expression through stringent regulation on all levels of the process. For successful expression, we need to troubleshoot and optimize the protocols with regards to temperature, concentration, time, growth media and induction parameters.

For the protein expression in this study, we used the protocols pertaining to our host strain and vector as independent templates. The pET Expression System, was developed by Studier and Moffatt (1986) and has revolutionized protein expression capability over the years due to the ease at which the vector system can be manipulated, the promotor-specificity and its ability for mass-production of proteins. This study utilises the protocols stipulated in the BL21 competent cells instruction manual by ©Agilent Technologies as well as the pET system instruction manual by Novagen Inc.

Briefly, we inoculated 10ml LB medium containing the KAN⁺ (30 µg/ml) and CAM⁺ (34 µg/ml) with the glycerol stocks that were previously made. Thus, 6 SNPs, the WT as well as an uninduced sample (UI) were chosen for subsequent protein expression. The extra sample was added to serve as the uninduced experiment and negative expression control. The culture was placed in a shaking incubator to shake at approximately 180 rpm for 12 hours at 37°C. Hereafter, starter cultures were diluted to a ratio of 1:100 into 500ml LB medium

containing the antibiotics. The OD₆₀₀ reading was taken every hour until OD₆₀₀ was between 0.4 and 0.8. This was done spectrophotometrically at 600 nm in an untreated 96 well flat bottom plate and readings were taken using an Elisa plate reader. Once OD₆₀₀ readings were optimal, the cultures of the 6 SNPs and the WT were induced with IPTG to a final concentration of 1 mM and allowed to shake for a further 4 hours. The UI sample was harvested immediately after optimal OD₆₀₀ readings were reached without the addition of IPTG. Once the induction fermentation period of 4 hours was complete, cells were harvested by centrifugation in 50 ml polypropylene tubes at 2500 rpm for 10 minutes. Cell pastes were then stored at -80°C until used for further experimentation.

Protein analysis revealed the band of interest in the uninduced state, before the addition of IPTG. Thus, further optimisation was necessary to try and eliminate the background expression seen in our system. The results for the individual optimization protocols can be seen in Appendix VI.



2.6.2 Protein expression optimisation

2.6.2.1 Induction with 1 mM IPTG at specific OD readings with fermentation at 30 °C

The protocol was followed precisely as before. However, once OD₆₀₀ readings were optimal and the culture was induced, the cultures were now shaken at 30 °C for the 4 hours. Cells were then harvested precisely the same and store at -80 °C. Background expression in the uninduced state was still detected (Appendix VI, Figure 9 and 10). Further modifications to the protocol was then made.

2.6.2.2 Induction with 0.1 M IPTG at specific OD readings with fermentation at 27 °C

Again, the only difference made to this protocol was the temperature adjustment from 37 °C to 27 °C. This did not change the level of background expression and as such, we continued

with optimisation. The temperature profiles offered no help and so we returned expression temperatures to 37°C.

2.6.2.3 Addition of 1% glucose to the 500 ml culture

For this modification, we started the expression protocol in exactly the same manner. The glycerol stocks we previously prepared was inoculated into 10 ml LB medium supplemented with the appropriate antibiotics at the correct concentrations. The cultures were placed in a shaking incubator and allowed to shake at 37°C for 12 hours at approximately 180 rpm.

The starter cultures were then diluted in a 1:100 ratio in 500 ml LB medium. The change was observed when the medium was now supplemented with 1% glucose in addition to the antibiotics. This allows for the reduction of basal expression since the glucose is the preferred carbon source during growth. Thus, induction will not take place until the glucose is completely utilized. This mechanism of control is known as catabolite repression (Rosano and Ceccarelli, 2014). Once optimal OD₆₀₀ readings were reached, the cultures were induced with IPTG to 1mM and continued to shake at 37°C for a further 4 hours. Cells were harvested the same as before and stored at -80°C until further use. Upon further analysis, no change was detected in the levels of background expression in the uninduced state.

2.6.2.4 Increased CAM⁺ concentration and decreased induction time

In this instance all media was supplemented with KAN⁺ at 30 µg/ml while the concentration of CAM⁺ was increased to 50 µg/ml. Starter cultures were made by inoculating 10 ml LB medium with the glycerol stocks as described previously. After the 12 hour shaking incubation period at 37°C, the cultures were diluted into a 500 ml culture to a ratio of 1:100. OD₆₀₀ readings were taken and once a reading between 0.4 and 0.6 was reached, the cultures were induced with IPTG to a final concentration of 1 mM. However, instead of the standard 4 hours, the induced cultures were only allowed to shake for 2 hours at 37°C before the cells were harvested and stored at -80°C for further use. This, like every other attempt at optimisation, did not reduce the background expression seen in the uninduced state. At the higher CAM⁺

concentration, in the purified sample, no bands were seen (Appendix VI, Figure 11). Thus, we continued expression at the initial CAM⁺ concentration of 34µg/ml.

2.6.2.5 Expression of empty pET-28a(+) vector

In a final attempt to understand the background expression as well as add a negative control for this experiment, we had obtained an empty vector courtesy of the Department of Human Genetics and Molecular Biology at Stellenbosch University. The standard protocol was followed for this expression process and cells were harvested as described previously and stored at -80°C.

In response to all the optimisation and modifications made, we decided to use the standard protocol for protein expression and subsequent protein analysis for the purpose of this project.



2.6.3 Protein analysis

2.6.3.1 Protein extraction and cell lysate preparation

Cell pastes were thawed and resuspended in 3 ml/g suitable lysis buffer. After troubleshooting for the subsequent protein quantification and activity assays, a suitable lysis buffer was obtained. The constituents of the lysis buffer as well as their functions and concentrations can be seen in table 10.

Table 10. Lysis buffer constituents and functions

CONSTITUENT	FINAL CONCENTRATION	FUNCTION
Tris (pH8)	50mM	Increases outer membrane permeability by interacting with the lipopolysaccharides
NaCl	150mM	Provides osmotic shock to cells
Lysozyme	0.4mg/ml	Breaks down bacterial cell wall to improve protein extraction

It must be noted that all work pertaining to protein extraction and analysis was done on ice or at 4°C. Once cells are resuspended with the aid of a vortex, the cells were lysed by alternate cycles of sonication and the freeze-thaw method using liquid nitrogen. The lysed suspension was then allowed to rest on ice for 30 minutes before further steps were followed. Hereafter, the suspension was centrifuged at 14000 rpm for 10 minutes to separate the soluble and insoluble fractions. The soluble fraction was extracted and stored at -20°C.

2.6.3.2 Protein purification

In order to conduct any further experimental work, it is necessary to purify the samples so that only the protein of interest remains. In this case downstream activity assays will be done.

For this study, the protein purification protocol from the Promega MagneHis™ purification kit was followed. The proteins were purified based on the type of affinity tag attached to it. In this case it was a poly-His tag which is easily purified using magnetic nickel (Ni²⁺) particles. The protocol is written for 1 ml of the bacterial culture and purification is done on the principle of magnetism. Briefly, 500 mM NaCl is added to 1ml of unpurified bacterial culture. MagneHis™ Ni²⁺ particles are then added and the suspension is inverted and mixed before being put on a magnetic stand. The magnetic stand captures the magnetic particles and the supernatant can then be removed. Wash buffer is then added to the remaining particles and the samples are placed back on the magnetic stand and the supernatant is removed. The washing step was repeated thrice according to the protocol. After the final wash step, MagneHis™ elution buffer is added to capture the resultant protein from the magnetic particles. After a 2 minute incubation at room temperature, the samples were again placed in the magnetic stand and the protein supernatant is removed and transferred to a clean, sterile 1.5 ml eppendorf tube and stored at -20°C until used. The full protocol can be seen in appendix III.

2.6.3.3 Protein quantification

All proteins were quantified using the Bradford assay. Thus, as previously mentioned, a suitable lysis buffer was needed so as to not interfere with the assay constituents. It should be noted that for the purified samples, the MagneHis™ elution buffer will serve as the diluent.

The Bradford principle is based on the binding of the Comassie Blue G250 dye to a protein to create a colour change from red to blue and, thus, shift the absorption spectrum to 595 nm. The method is quick and convenient in that the colour change is rapid and stable for up to 1 hour. It was observed that certain compounds interfere with the colour change, thereby influencing the accuracy of the estimated protein concentration (Bradford, 1976). For the purpose of this study, the 1 mg/ml BSA standard was used and the Bio-Rad Bradford reagent as the dye.

In a 96 well flat bottom plate, 20 µl 1 mg/ml BSA standard is added to the well A1. Following this trend, 20 µl of each sample is added to well "A" in independent columns. Depending on the sample being quantified, the correct diluent is then added to the rest of the wells of the columns being used. In this case, the diluent was either the MagneHis™ elution buffer or the cell lysis buffer. A doubling dilution was then done, per column for each sample, using 10 µl from first well and transferring it to the next one below (row B). This was then mixed with the pipette and another 10 µl is transferred. So the dilution series continues until the last well which should contain only diluent. Once the dilution is done, 100 µl of Bradford Reagent is added to each well and the colour change was observed. A schematic representation of how the dilutions for the Bradford assay for protein quantification is done, can be seen below.

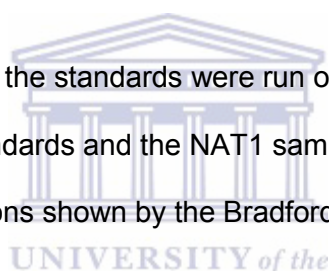
The plates were then read at 595 nm using an Elisa plate reader. Once the absorbance readings are obtained, a standard curve was produced using the known concentrations from the BSA standards. This was then used to calculate the concentrations of the rest of the samples, purified and unpurified.

	UI	WT	T193S	T240S	V231G	V235A	F202V	D229H	pET28a(+)
1	20 10µl	20	20	20	20	20	20	20	20
2	10 10µl	10	10	10	10	10	10	10	10
3	10 10µl	10	10	10	10	10	10	10	10
4	10 10µl	10	10	10	10	10	10	10	10
5	10 10µl	10	10	10	10	10	10	10	10
6	10 10µl	10	10	10	10	10	10	10	10
7	10 10µl	10	10	10	10	10	10	10	10
8	10	10	10	10	10	10	10	10	10

discard
10µl

Figure 5. Schematic representation of the Bradford assay setup on a 96 well plate

In addition to the Bradford assay, the standards were run on as SDS-PAGE gel to compare band intensities between the standards and the NAT1 samples (Figure 9). This will give some insight into the concentrations shown by the Bradford assay.



2.7.6.5 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE gels are commonly used to characterize the size, purity and integrity of macromolecules such as protein. This method is fairly quick and very sensitive. (Hames, 1998). This specific technique allows the separation of proteins with the aid of an electrical current. Electrophoretic separation is accomplished through a molecular sieve principle. The increase in temperature through electrophoresis denatures the proteins and the addition of SDS confers a negative charge to the protein polypeptide chains by coating it. The negatively charged proteins then migrate to the positive pole in the presence of a suitable buffer (Roy and Kumar, 2012).

For this project, an 8% stacking gel and 15% separation gel was prepared in accordance with the protocol stipulated in the SDS gel preparation kit (Sigma-Aldrich®). A Tris-Glycine SDS running buffer (25 mM Tris, 192 mM Glycine, 0.1% SDS(pH 8.3)) was used. The current

was set at 40 mA/gel for stacking and 20 mA/gel for separation. The samples loaded on this gel were at a concentration of 30 µg total protein and diluted in a 1:1 ratio with 1X Laemmli buffer (0.1% 2-Mercaptoethanol, 0.0005% Bromophenol Blue, 10% Glycerol, 2% SDS and 63 mM Tris-HCl (pH 6.8)). As reference, 8µl PageRuler Plus Prestained Protein Ladder (Thermo Scientific™) was loaded. The protein bands were then visualized by staining it in Coomassie Brilliant Blue stain (45% methanol, 10% glacial acetic acid, 45% water and 3 g/L Coomassie brilliant blue R250). The gel was allowed to stain overnight and was then treated with Coomassie destain solution (50% w/v methanol and 10% v/v acetic acid) until bands are clearly visible. In addition to this, another gel was run for the purpose of Western blot transfer to identify the specific His-tagged NAT1 protein.

2.7.6.6 Western Blot Transfer

This technique involves the transfer of denatured proteins to a solid nitrocellulose membrane for detection with specific antibodies to the target protein (Jensen, 2012). The nitrocellulose membrane was equilibrated in 100% methanol. A gel sandwich is then prepared using sponges and Whatman paper to protect the membrane and the gel during the transfer process and also to ensure close contact between the gel and the membrane for transfer efficiency (Mahmood and yang, 2012). Proteins were transferred in a submersion system perpendicular to the surface in a transfer buffer (27 mM Tris, 191 mM glycine, 20% methanol) at 100V for 90 minutes. Once the transfer is complete, the membrane was blocked in 5ml Western Blot Blocker Solution on a rocker for 30 minutes. Next, the primary antibody (mouseαHis) was prepared in a 1:1000 dilution with the blocking reagent. After blocking the membrane, it was treated with the antibody and left to incubate on a roller at 4°C overnight. The following day, the membrane was washed twice with PBS-Tween wash solution (0,05% Tween-20 in PBS) for 15 minutes at a time. The membrane was then treated with HRP-conjugated secondary antibody (goatmouse) prepared in a 1:1000 dilution with the blocking reagent. The membrane was incubated on a roller for 1 hour at room temperature. The membrane was again washed as described in the text already. For detection of the protein, the membrane was treated with preheated peroxidase substrate.

2.8 NAT1 acetyltransferase activity assay

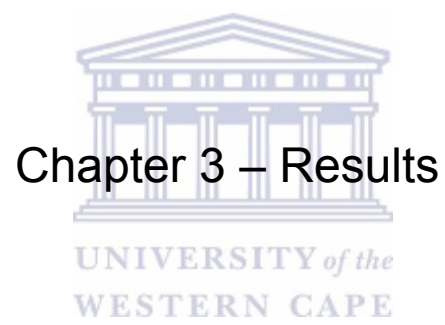
The assays was performed in a 96 well flat bottom plate. The components needed for the assay can be seen in table 11 as well as the necessary concentrations. The protocol template corresponds with that of Dairou and colleagues for the spectrophotometric estimation of NAT1 activity (Dairou et al, 2003). Briefly, all components are added to the 96 well plate, at the correct concentrations for a final volume of 300 µl per well, and a zero reading was taken before the addition of PNPA. Once PNPA is added, the plate was incubated for 10 minutes at 37°C. Thereafter, the reaction was quenched using 1% SDS. The plates were then read at 410 nm in an ELISA plate reader. Various controls were introduced to test the eligibility of the assay. These controls include a drug control (no substrate added), an enzyme control (no NAT1 added) and a co-factor control (no PNPA added). The assays were performed thrice per day over three consecutive days per substrate or drug (PABA and PAS). It has been shown that NAT1 shows specific substrate specificity for these two drugs (Wu et al, 2007). The averages of the absorbance values were calculated per assay per day. Those averages were then tabulated and a separate average calculation was done to get a standard average over three days for the assays. The results were then plotted on a bar graph.

Table 11. NAT1 activity assay reaction constituents and concentrations.

COMPONENT PER WELL	FINAL CONCENTRATION
NAT1	15nM
SUBSTRATE (PAS OR PABA)	500uM
PNPA	125µM
TE BUFFER	25mM
SDS	1%
FINAL VOLUME PER WELL	300µl

The assays were initially performed on crude WT and UI samples to test whether the system will work with all the background and basal level expression present in the crude samples. The crude assays were done as an optimization step to consolidate the concentrations described by Dairou and colleagues. Furthermore, the assay evaluated the effect purification would have on the overall result. We found that the assay did not work on crude samples since the result could not be clearly analysed thus, the assay was performed on the purified protein samples.





Chapter 3 – Results

3.1 **NAT1-pUC57 and pGEX-4T-2**

NAT1-pUC57 was purchased after failed attempts at cDNA production from a MCF-7 cell line. Thorough optimisation reactions were performed to ultimately verify the presence of *NAT1* in the vector for further ligation into pGEX-4T-2. Parallel to this, the empty pGEX-4T-2 vector was RE digested to determine whether it is empty and viable for ligation and further studies.

Appendix IV shows all the electrophoresis gel pictures for the different optimization reactions documented in section 2.5.4.1.2 and tables 2, 3, and 4 – temperature optimizations, MgCl₂ profiling, dNTP profiling, DNA template profiling. The reactions were performed according to specification previously stipulated in the text under the conditions in table 6. The primers used (Table 1), were designed to amplify the entire *NAT1* gene and not just the ORF. The expected product size, based on the primers used, was ~1600 bp.

Some of the gel pictures show specific reactions yielding a band of interest at that approximate (Appendix IV, Figure 3-6). However, for further protein expression and functional studies the entire protein was not the sample we needed. In order to successfully express the resultant *NAT1* protein, the coding region needed to be amplified for transcription to be possible further down the line. Thus, with the time constraints now pressing for this study, *NAT1*-pET-30a(+) was individually cloned and purchased. In this light, new primers were designed to amplify the ORF of the *NAT1* gene with a size of ~879 bp.

The parallel linearization of pGEX-4T-2 yielded a positive result when RE digested with XhoI. The band of interest is clearly visible confirming that the vector is empty and can be used for ligation (Appendix V, Figure 8)

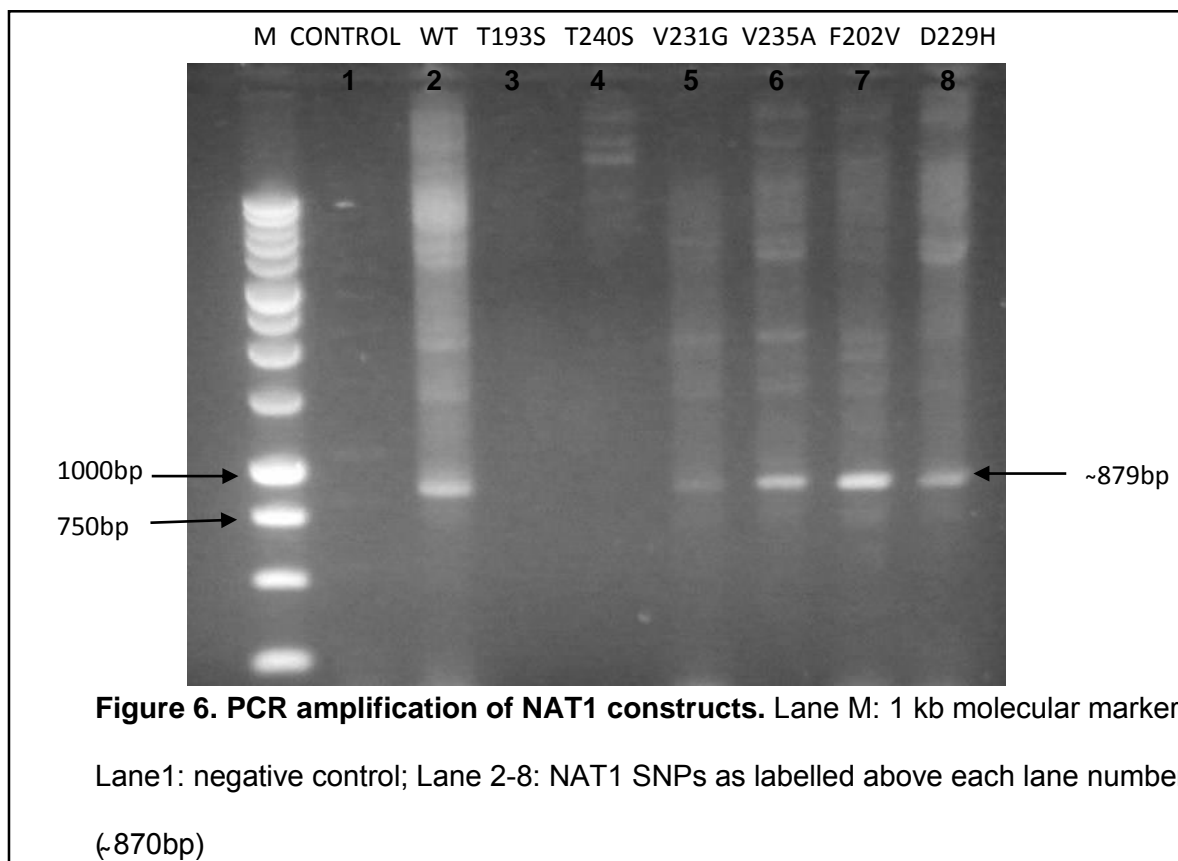
3.2 **Transformation of NAT1-pET-30a(+) into BL21(DE3)pLysS**

The *NAT1*-pET-30a(+) SNP constructs were transformed into BL21(DE3)pLysS strain of *E.coli* cells. The DNA was then extracted and quantified. PCR amplification with gene specific primers and conditions described in table 5 and 6, respectively. However, PCR amplification did not show all the bands of interest under the UV light. Thus, a RE digest was

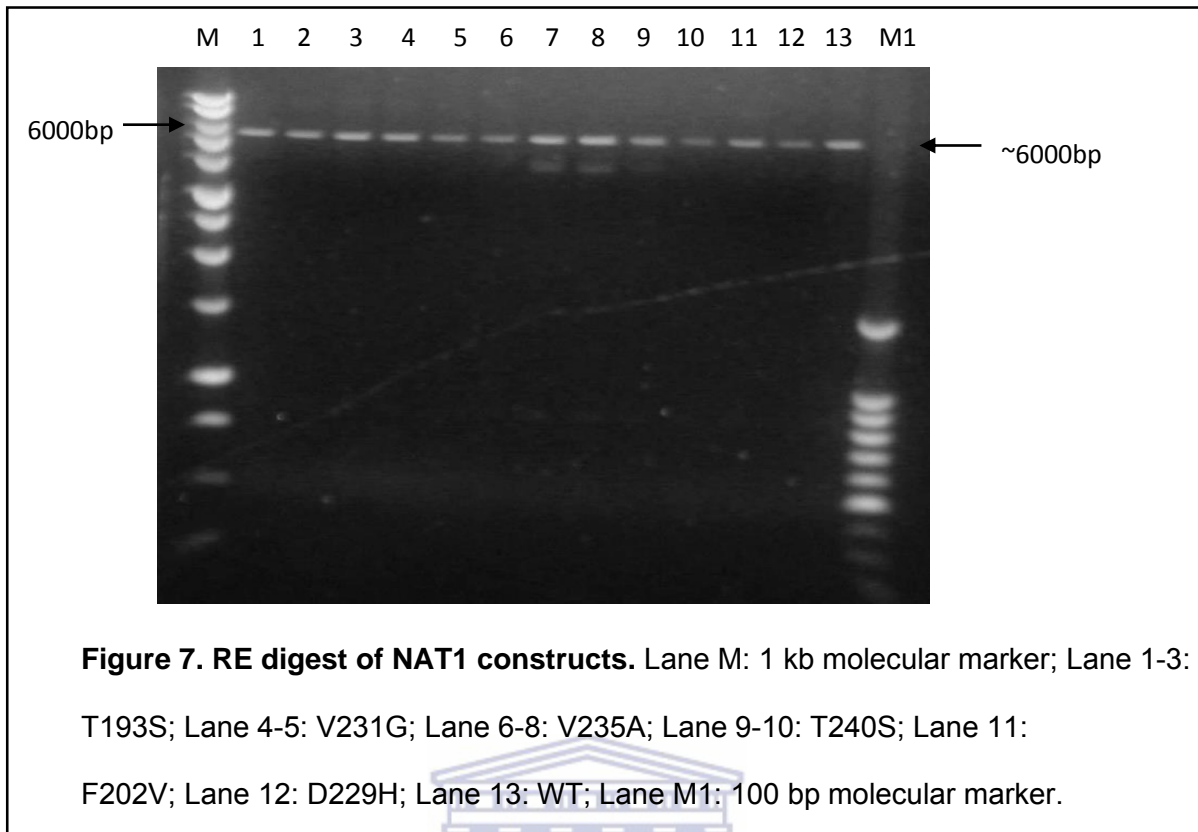
also done, to further confirm the presence of *NAT1* DNA at the correct size, according to the specifications described in table 8 and 9. The products of both reactions were run on separate 1% agarose gels at appropriate DNA concentrations for visualisation. The DNA concentrations, quantified using a Nanodrop2000, are tabulated below in table 12 and the agarose gel pictures can be seen in figures 6 and 7.

Table 12. DNA concentrations from Nanodrop2000 in ng/μl

NAT1 SNPs						
WT	T193S	T240S	V231G	V235A	F202V	D229H
185.1	19.3	86.1	55.1	150.7	145.1	93.3



Samples T193S and T240S may have been loaded at too low a concentration to be visible on this gel. However, the rest of the samples show up at the correct size of ~879 bp. In order to check for the insert in those two faulty lanes, a RE digest was performed as well.



The RE digest shows all the samples linearized at the correct size of the *NAT1*-pET-30a(+) construct of 6000 bp. Incomplete digestion is due to only one enzyme being used instead of both. Thus, the transformation success was confirmed and protein expression could ensue.

3.3 Protein expression studies

Bacterial systems offer more ease for protein expression due to their availability and the low cost, high yield principle. For unrivalled success, researchers need to carefully consider their target protein and the most suitable host strains, promoters and subsequent vectors for the best protein production (Terpe, 2006). In this study, the successfully transformed constructs were used for protein expression and characterisation. This was done using the protocol of the pET expression system described previously in the text (Studier and Mofatt, 1986). Once expressed, protein samples were quantified using the Bradford assay and visualized on an SDS-PAGE gel under treatment of a Commassie stain and further through the western blot transfer technique.

3.3.1 Bradford assay for purified protein quantification

Proteins were purified using the protocol stipulated in appendix III under the MagneHis™ purification system specifications. Thus, since the target protein (NAT1) and all the SNPs contain a polyHis-tag, the concentrations can be assumed to be that of the NAT1 present in the total protein independently produced by *E.coli* cell mechanics. The control column consists of only diluent (lysis buffer in table 10) purified with the same protocol. So it would be correct that it is the lowest of all the concentration values. The calculation used for the quantification makes use of the formula for the curve represented by figure 8 where “y” is the known absorbance of the BSA standard dilution series and “x” is the unknown calculated to represent the values in table 13. These values were then used to calculate how much of the sample would be needed to make up a final concentration of 30µg of purified protein to run on the SDS-PAGE gels for visualisation and confirmation that the NAT1 was successfully expressed in this system.

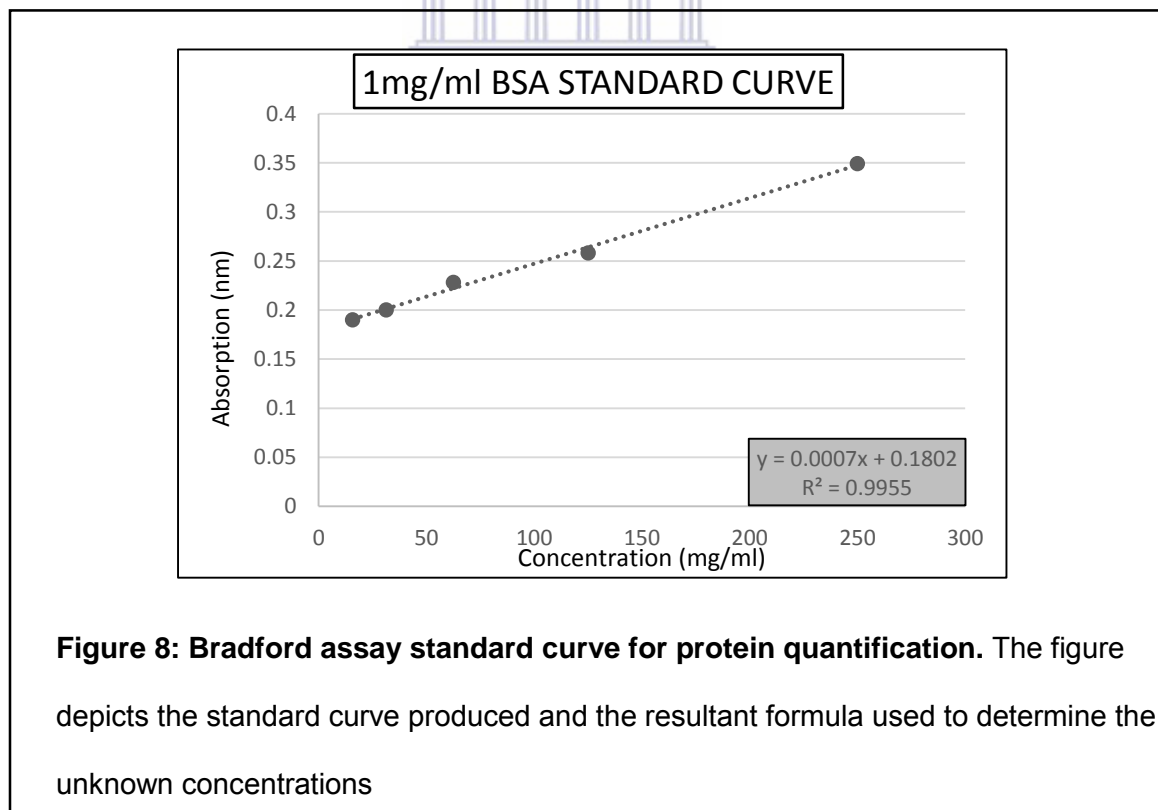
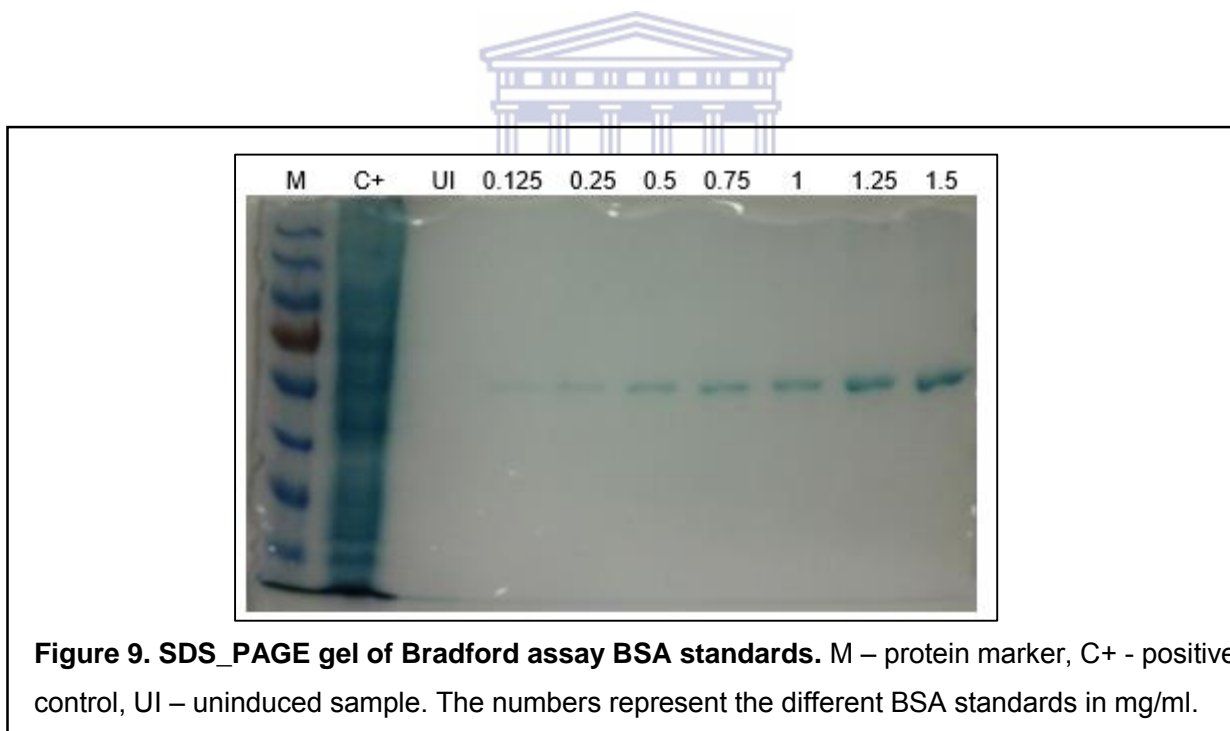


Table 13: NAT1 purified protein concentrations in $\mu\text{g/ml}$

PURIFIED NAT1 SNPs	PROTEIN CONCENTRATION ($\mu\text{g/ml}$)
UI	728.4571
WT	966.8571
T193S	751.4286
T240S	1187.2
V231G	671.1429
V235A	1604.286
F202V	1221.029
D229H	648.8571
pET-28a(+)	1420.19
CONTROL	258.9048



3.3.2 SDS-PAGE and Coomassie staining

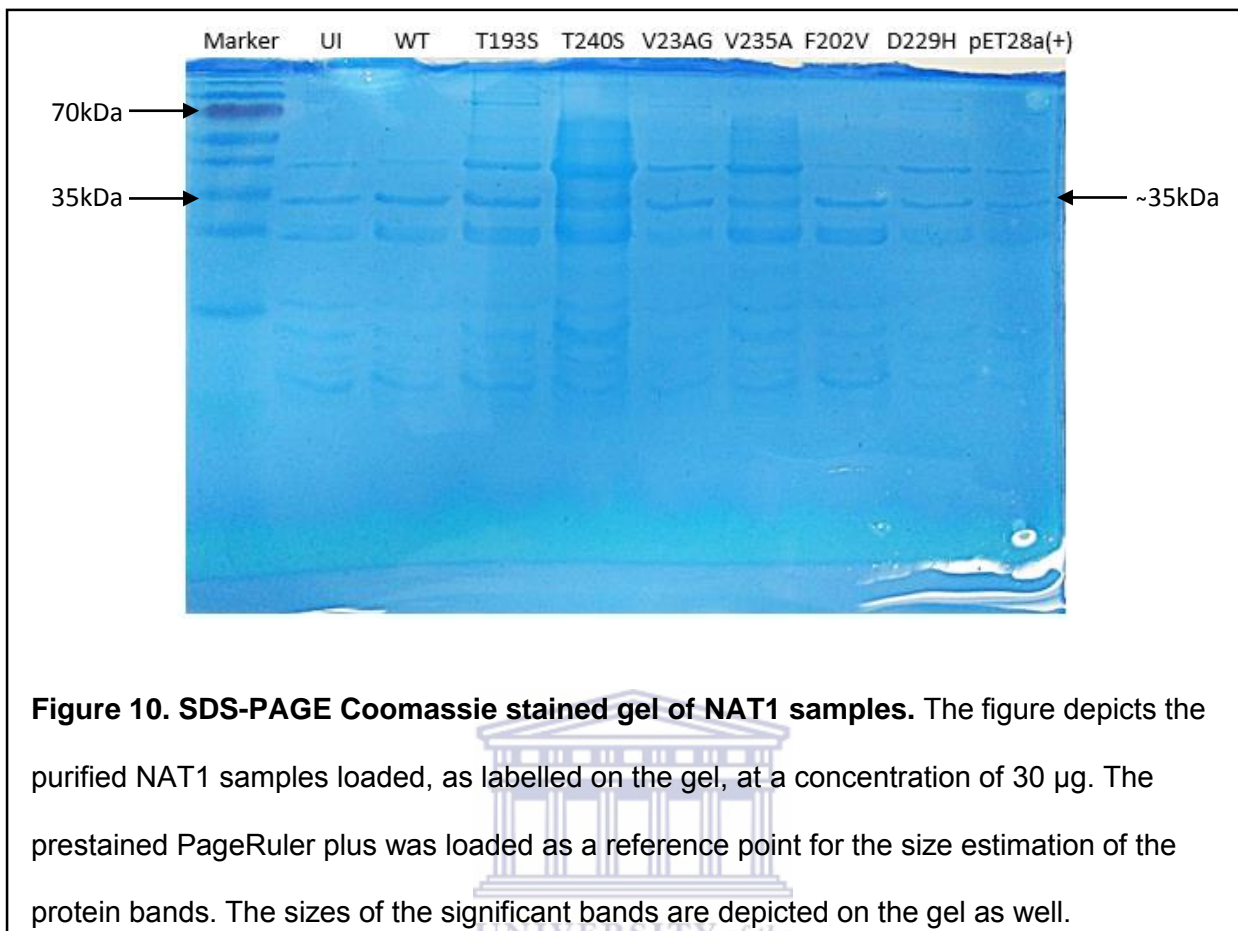


Figure 10. SDS-PAGE Coomassie stained gel of NAT1 samples. The figure depicts the purified NAT1 samples loaded, as labelled on the gel, at a concentration of 30 μ g. The prestained PageRuler plus was loaded as a reference point for the size estimation of the protein bands. The sizes of the significant bands are depicted on the gel as well.

The SDS-PAGE gel (figure 10) shows the results of 30 μ g purified samples. As seen above, the NAT1 protein has been successfully expressed across the board for all the NAT1 samples. However, there is a band of interest in the UI which should not be as well as in the empty vector pET-28(+). We can speculate that this is attributed to the leaky expression and increased basal expression levels seen in the pET vectors. This is largely due to the control, or lack thereof, of the *lac* promoter. The mechanism of control within in this vector can be seen in figure 11.

The protocol followed for expression involved inoculation of a 10 ml starter culture and sub-culturing into a bigger culture in a 1:100 dilution. Cultures were grown till OD₆₀₀ readings reached a log phase of between 0.4 and 0.6. This diminishes the stress levels and the possibility of overgrowth and toxicity. The cultures were then induced with 1 mM IPTG except the UI which was harvested when OD₆₀₀ readings were optimal. The rest of the cultures, induced, were incubated at 37°C, 180 rpm for a further 4 hours before harvesting.

The optimization expression reactions performed all yielded multiple bands, whether crude or purified. Thus, we were unable to eliminate the basal and background expression seen in the gel pictures in Appendix VI for the optimisation reactions. The study was then continued using the initial protocol parameters for protein expression in our system.

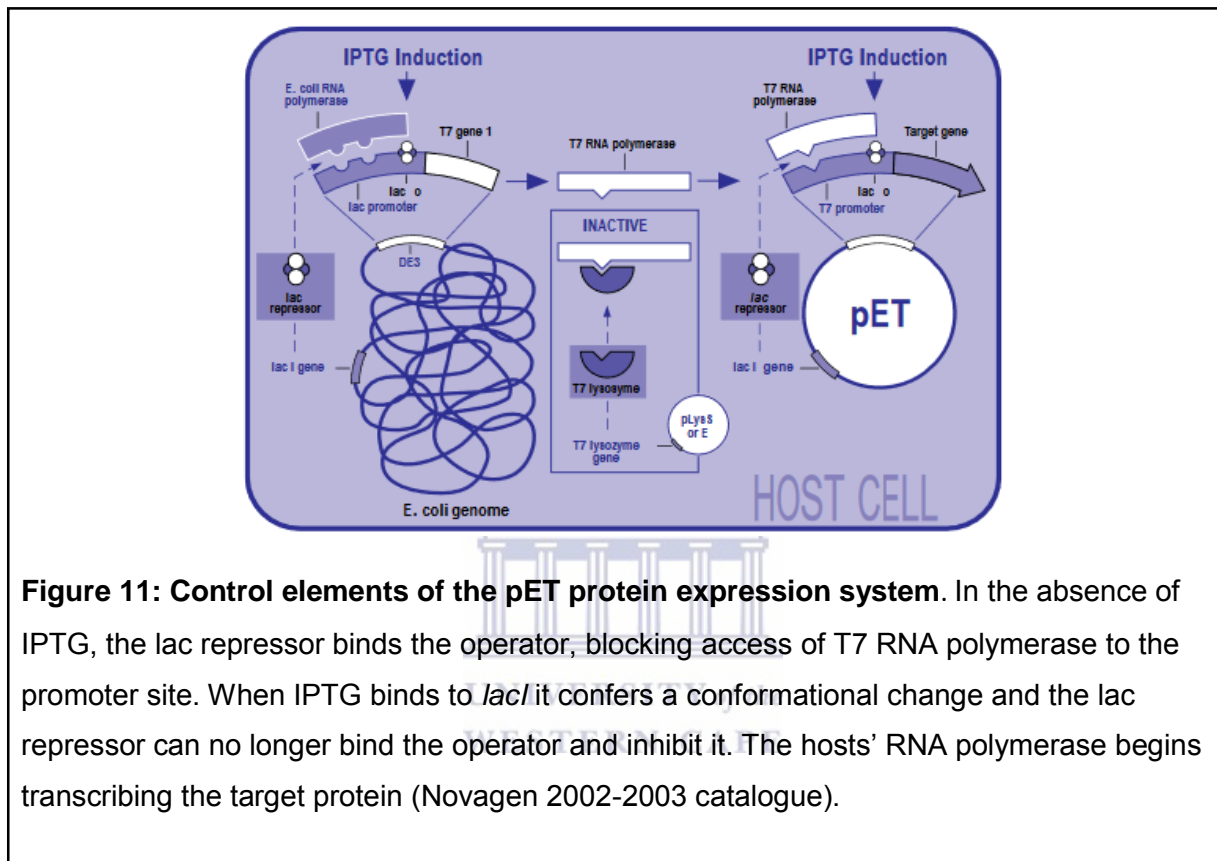


Figure 11: Control elements of the pET protein expression system. In the absence of IPTG, the lac repressor binds the operator, blocking access of T7 RNA polymerase to the promoter site. When IPTG binds to *lacI* it confers a conformational change and the lac repressor can no longer bind the operator and inhibit it. The hosts' RNA polymerase begins transcribing the target protein (Novagen 2002-2003 catalogue).

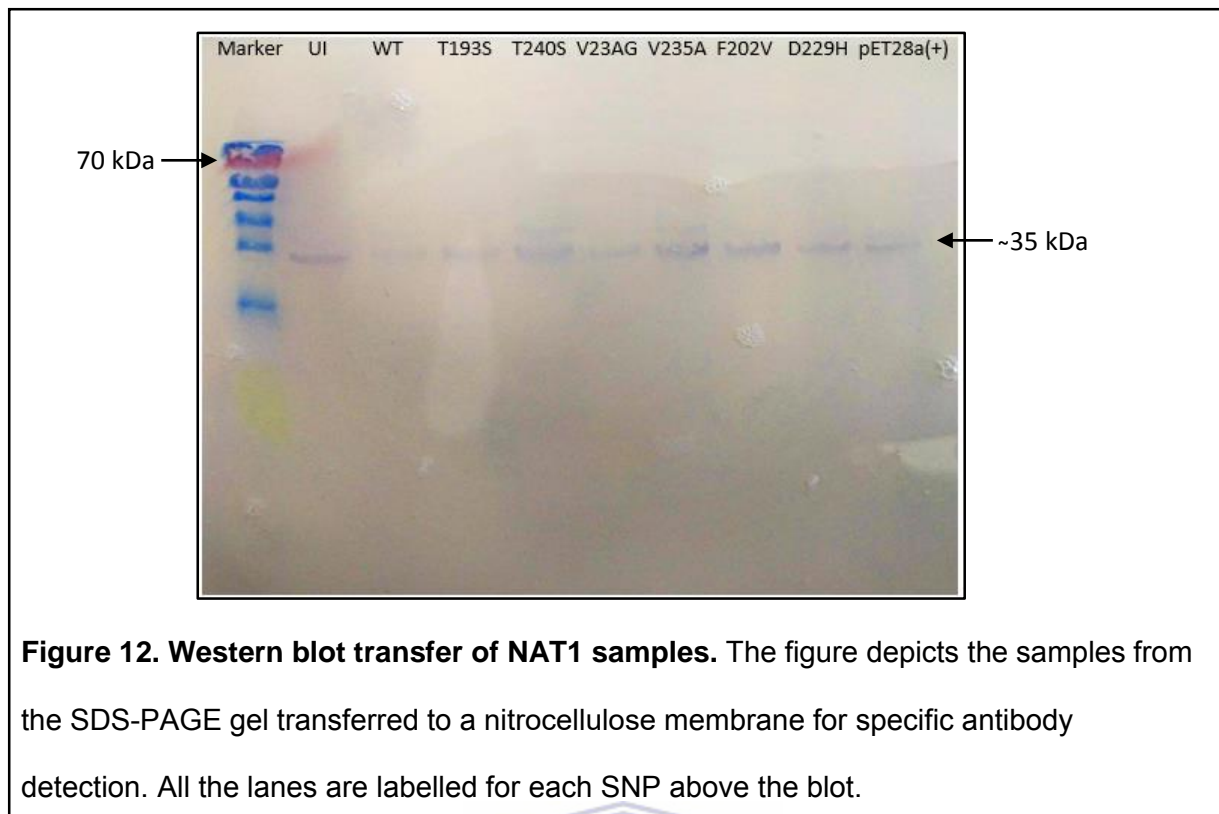
The vector for this project contains a *t7/lac* promoter which uses *t7* RNA polymerase to drive transcription. Furthermore, the pET-30a(+) and pET-28a(+) also carry the natural *lacI* repressor gene. The strain of *E.coli* chosen as host for these vectors, BL21(DE3)pLysS, encode for the *t7* RNA polymerase needed for transcription in its genome and also possessed the *lacI* gene. In addition, it is controlled by a derivative of the typical *lac* promoter, *lacUV5*, which is inducible by IPTG (Bashiri et al, 2015). IPTG is a mimic of lactose and in contrast, is not involved in any metabolic pathways of the host. Thus, the concentration added remains constant. However, in high copy number plasmids such as this, this could lead to an increase in basal level expression. Also, in high copy number plasmids, folding mechanisms can be flawed and this in turn could affect the repression

mechanism by *lacI* (Bell, 2000). The leaky expression refers to the unintended expression of t7 RNA polymerase as well as the target protein, NAT1, without induction. This can be controlled by the addition t7 lysozyme to the reaction. In our case the lysozyme is introduced through the pLysS plasmid of our bacterial strain, *E.coli* BL21(DE3)pLysS. The rationale is that t7 lysozyme binds the promotor and inhibits transcription. However, t7 RNA polymerase expression still outweighs the inhibitory effect conferred by the lysozyme (Rosano and Cecceralli, 2014).

3.1.3 Western blot transfer

SDS-PAGE depicts the total protein produced by the host. In contrast, the western blot technique detects a specific protein in the mixture produced by the host. The technique makes use of three main standards; the transfer of proteins to a solid membrane, the concept of separation by size and the use of proper primary and secondary antibodies for specific detection (Mahmood and Yang, 2012). This was the premise for further detection methods since the SDS-PAGE results were not specific for the identity of the protein.

The proteins run on the SDS-PAGE gel were transferred to a nitrocellulose membrane and were separated by electrophoresis in a Bio-Rad system. The membrane was then washed and treated with primary Mouse α His antibody. The membrane was then washed again and treated with an appropriate secondary antibody, goat α mouse antibody. The antibodies were HRP-conjugated and for detection, preheated peroxidase substrate was used. The western blot transfer result is seen in figure 12.



The result seen in figure 12 shows that the target protein, NAT1, is detected in all the lanes. The band in the UI lane can again be attributed to leaky expression of NAT1 since the WT glycerol stock was used to inoculate the starter culture. The band present in the empty pET-28a(+) vector lane, is difficult to define. As previously mentioned, we speculate that this could be due to the leaky vector but also non-specific binding. With the confirmation of NAT1 SNPs being expressed, they were used for activity assays.

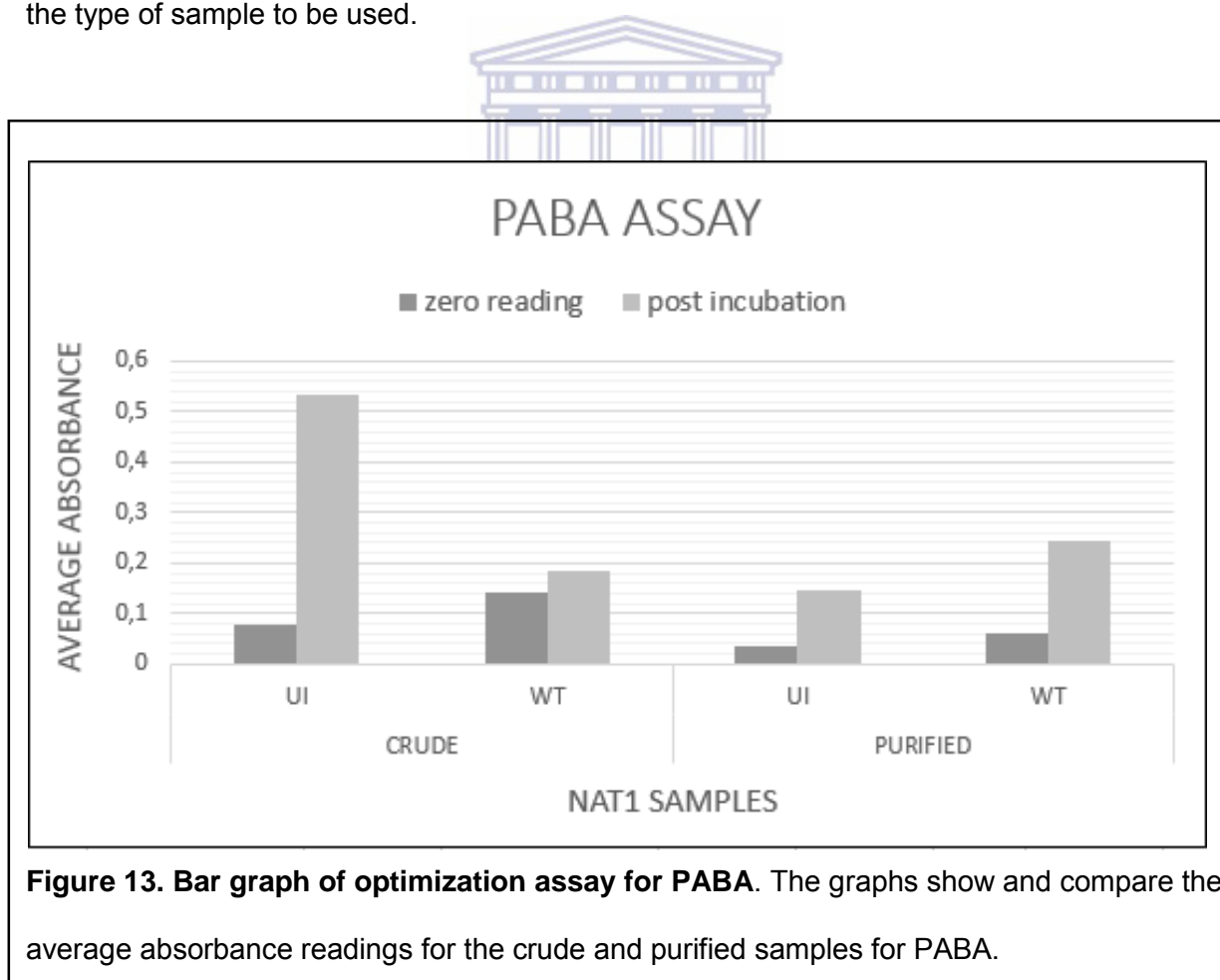
3.2 NAT1 acetyltransferase activity assays

The rationale behind the experiment is to determine the rate of acetylation that NAT1 confers to the different substrates by observing the level of hydrolysis of *p*-nitrophenol acetate (PNPA) and subsequent acetylation over a fixed time period of 10 minutes. The assay was done in 96 well flat bottom greiner plates with the components described in table 11. The process involves the hydrolysis of PNPA to release acetyl which in turn acetylates the substrate and inactivates it. The remaining by-product, *p*-Nitrophenol, released from the hydrolysis reaction causes the colour change from colourless to yellow. The absorbance is then read at 410 nm.

The acetylation of the target drugs for NAT1 occurs in a two-step reaction. The first step involves the transfer of the acetyl group from acetyl co-enzyme A (Ac-CoA) to the cysteine residue of NAT1. In the second step the acetyl group is then transferred to the arylamine drug (Levy and Weber, 2002). Research by Cloete and colleagues (2017) documents the importance of the stability of this acetylated cysteine complex for ultimate drug acetylation. In our study, Ac-CoA is substituted by PNPA and the arylamine targets refer to our drug probes, PAS and PABA.

3.2.1 Optimization of NAT1 acetyltransferase activity assays

The initial assay was performed on the crude as well as the purified WT sample. It was done as an optimization reaction for the concentrations of the components in the assay as well as the type of sample to be used.



The results for this assay was based on differences in absorbance readings. Since is was a optimization assay, statistical analysis was not done. The crude samples for the PABA assay shows the zero reading for the WT to be higher than that of the UI. The post incubation for the WT being lower than that of the UI samples. This is attributed to all the background expression seen in the UI samples. The exact protein constituents could not be identified for the UI samples. Therefore, the activity seen in the bar graph could be due to many different protein interactions with the PNPA. In contrast, the purified samples showed results in line with our hypothesis. The WT zero and post incubation readings are higher than that of the UI. This shows that the purified protein activity is more likely to be due to the increased presence of NAT1 in the WT than that of the UI since the purification process is stringent and specific for our His-tagged NAT1 protein sample.

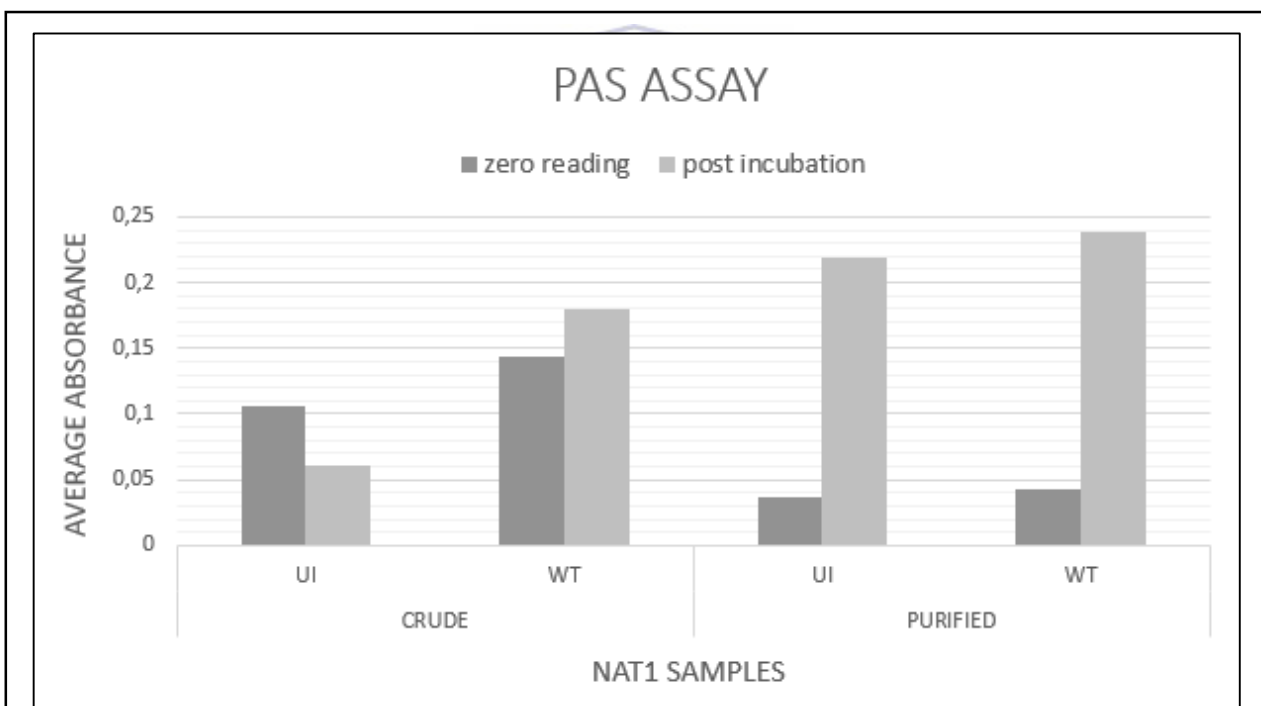


Figure 14. Bar graph of optimization assay for PAS. The graphs show and compare the average absorbance readings for the crude and purified samples for PAS before and after PNPA addition.

The crude sample for the PAS assay shows inconclusive results. The results were based only on the differences in absorbance readings between the SNPs. The post incubation reading for the UI sample is lower than its zero reading (Figure 14). This is unclear since the presence of different proteins in the UI should interact and cause activity. This, in turn,

should increase the reading after PNPA is added. The WT readings correspond with the hypothesis for these assays. However, the zero reading is high which could also be due to the protein constituents in the sample used. In comparison, the purified samples showed an increased post incubation reading when compared to the zero readings of the UI and WT samples. The zero reading of the WT is also higher than that of the UI which we hypothesised since the purified WT should have considerably higher NAT1 quantities than the UI which should not have any NAT1 presence. However, the leaky vector causes basal and background expression and consequently, NAT1 is present in the UI samples.

In light of the optimization assays, the subsequent assays for the NAT1 SNPs were done with only purified samples with different controls added to validate the credibility of the assays.

3.2.2 NAT1 acetyltransferase activity assay for PAS

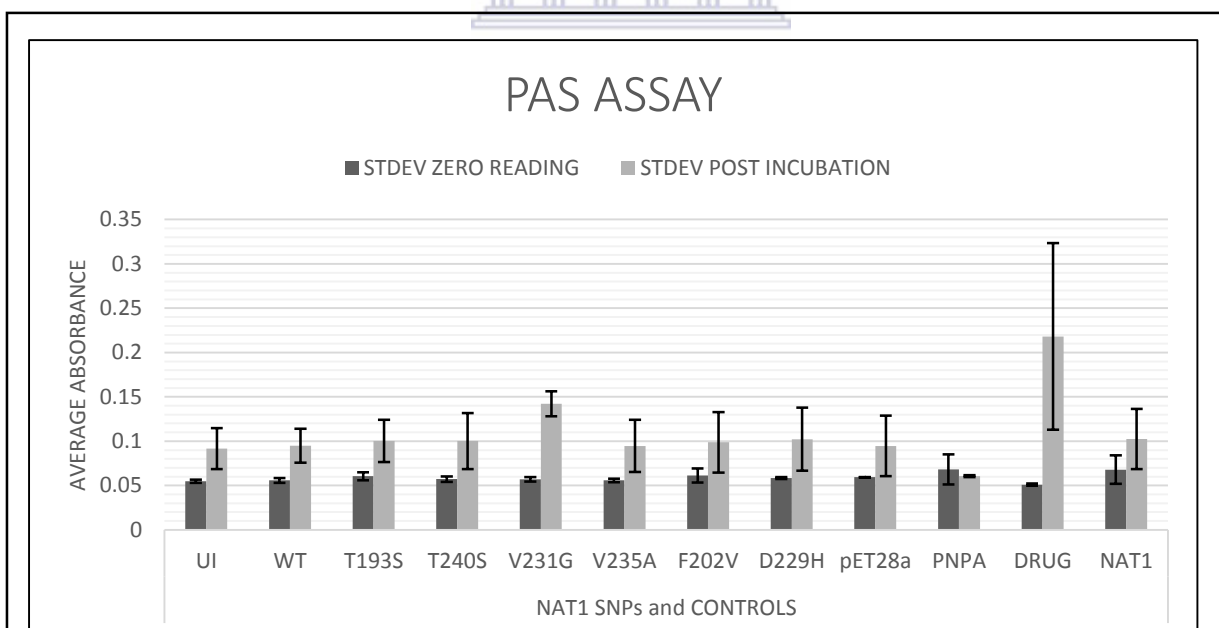


Figure 15: Bar graph for NAT1 acetyltransferase activity assay for PAS. The graph depicts the absorbance readings, showing the average absorbance for each sample before and after the addition of PNPA using PAS as substrate.

The uninduced sample (UI) should not have any difference in absorption since protein expression is stunted in this sample due to the absence of IPTG (inducer) (Figure 15). However, the glycerol stock of the (WT) was used to inoculate this culture. Furthermore, the vector used (pET-30a(+)) is proven to have leaky expression (Rosano and Ceccarelli, 2014). This is seen in the SDS-PAGE gels where the specific band of interest shows up in the UI state as well as the empty vector lanes and carries over to the western blot. Thus, there could very well be substrate-enzyme interaction which causes the increase in absorbance after PNPA. This is also evident in that the readings of UI and WT are very similar. For the empty pET-28a(+) vector, a band was detected but not confirmed to be NAT1 (Figure 10 and 12). The possibility that this antibody is not specific, does exist. Thus, if treated with a specific anti-NAT1 antibody, the results could more conclusive. However, there is an increase in absorbance which proves that there is definitely some sort of activity between the substrate and whichever protein is present. Since NAT1 has specific affinity for PAS, it is hard to explain this result.

The result for the PNPA control was expected since there is no PNPA in the experimental wells for the substrate to break down. As for the drug control, there is a dramatic increase in absorption. Since the substrate (PAS) is absent in these wells, it could be attributed to the increased hydrolysis of PNPA by the enzyme NAT1 since there is most likely no acetylation reaction occurring. The NAT1 control result could be caused by the gradual break down of PNPA in the presence of a substrate over time as well as other conditions affecting it such as pH.

For the 6 individual SNPs, all show definite NAT1 activity in that the absorbance reading increased after the addition of PNPA. When compared to the wild type, all the SNPs show increased absorbance readings except V235A. It can be concluded that the rest of the SNPs increases the rate of acetylation in comparison to the WT. This disproves the initial computation predictions for SNP T193S and T240S which was believed to be benign.

3.2.3 NAT1 acetyltransferase activity assay for PABA

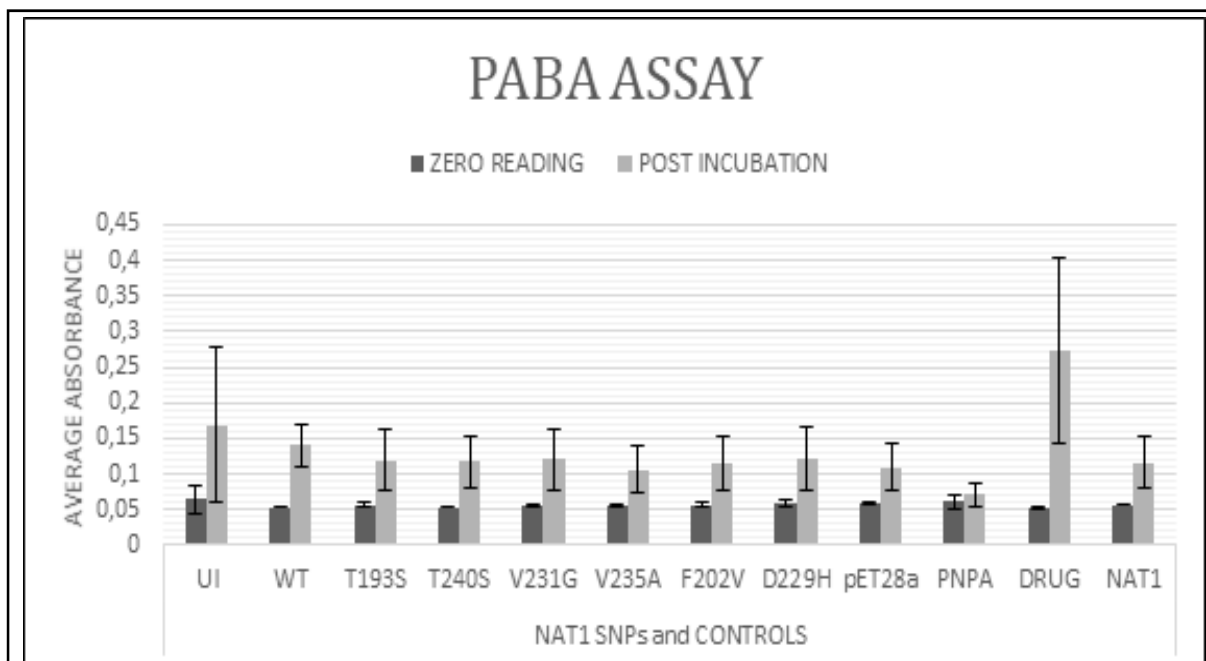


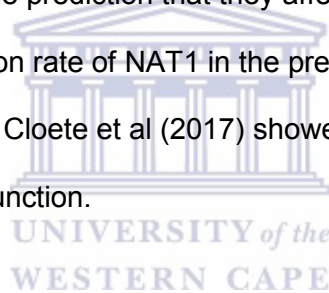
Figure 16: Bar graph for NAT1 acetyltransferase activity assay for PABA. The graph depicts the absorbance readings from table 17, showing the average absorbance for each sample before and after the addition of PNPA using PABA as substrate.

For the UI and WT, we observed a significant increase in absorbance after the addition of PNPA. However, the UI absorbance is more than that of the WT after PNPA was added. This again could just be a function of the leaky t7 vector as mentioned before. For the empty vector, the result is the same. The result for the PNPA control shows a small increase in absorption that is most likely negligible. The remaining controls, show similar results to the PAS assay.

For the 6 individual SNPs, all show definite NAT1 activity. This can be seen by the increase in absorbance after the addition of PNPA. In comparison to the PAS assay, all the post incubation readings are higher than 0.1 which is higher than those of the PAS assay. All except V231G. However, contrary to the PAS assay, all the post-incubation absorbance readings for the PABA assay are lower than that of the wild type. Thus, it can be concluded that the SNPs reduce the rate of acetylation of PABA.

Although both assays show NAT1 activity in the hydrolysis of PNPA in the presence of each substrate, it must be noted that the general absorbance readings for the PABA assay are higher than that of the PAS assay. This could mean that NAT1 has a higher affinity to PABA as compared to PAS. Also, with respect to TB treatment, NAT1 will more readily acetylate and metabolize PABA than PAS, making the latter the better option for treatment purposes. The acetylation reaction has its own parameters and condition such as pH levels which could affect readings and possibly explain the NAT1 control, no protein added, result.

With respect to the functional predictions obtained from PolyPhen-2 and SIFT for the specific SNPs, the experimental results oppose the hypothesis for T193S and T240S. The experimental data shows that these two SNPs in fact do affect the rate of acetylation in the presence of both substrates. For the remaining SNPs (V231G, V235A, F202V and D229H), the experimental data confirms the prediction that they affect the function of the protein. In this case they affect the acetylation rate of NAT1 in the presence of the specific substrates. The stability predictions made by Cloete et al (2017) showed that V231G destabilizes the native protein, thus affecting its function.



Chapter 4 – Discussion and Conclusion



4.1 DISCUSSION

The recognition of drug acetylation and the subsequent, clinical effect conferred by human variation in acetylating ability dates back more than 50 years (Spielberg, 1996). Previously, it was believed that NAT1 and its effect on PABA and PAS was genetically consistent. As research progressed, it was revealed that different mutations cause different amino acid changes, therefore possibly affecting the structure and function of NAT1 (Bell et al, 1995; de Leon et al, 2000; Arslan et al, 2003).

Allelic variations exist between individuals, whole populations and ethnicities and their geographical location, which explains the differences in the human predisposition to disease and differences in the rate of drug metabolism and human drug response. Consequently, clinical treatment is affected. This was made evident when the genetic acetylation polymorphism was investigated with the anti-tubercular INH drug (Meyer, 1994; Spielberg, 1996). The effect of allelic mutations of enzyme function is evident in our results since the comparison is drawn between the mutants and the native protein, NAT1. SNPs have been identified and the structural and functional effects were detailed in a review (Hein, 2009). Hein (2009) also documents that mutations at any given chromosomal location can also affect the stability of the enzyme and, thus, its activity. This was illustrated for SNP V231G. Structural studies confirmed that the mutation, V231G, has a destabilizing effect on the protein and so affects the acetylation capability of the anti-tuberculosis drug, PABA. The experimental results in our study reaffirms the findings by Cloete et al (2017) that SNP V231G decreases NAT1 stability and subsequent acetylation of PABA in comparison to the native NAT1 protein.

Ionic interactions are important in substrate specificity as well as the folding of a protein. The aa residues that make up the binding pocket may interact when folded but may be far apart in the linear aa chain. This inherently guides and determines enzyme-substrate interactions. The folding of a protein relies greatly on its size and the flexibility of its side chain. Thus,

SNPs can change the function without changing the shape of the 3D protein structure and vice versa (Goodfellow et al, 2000).

Previous studies have made efforts to elucidate the effect of structure on the activity of NAT1. The crystal structure of human NAT house 17 aa residues in a pocket that stabilises the protein and aids in its expression and overall binding specificity and function (Akaguru, 2012). Wu and colleagues found that the binding pocket of NAT1 is relatively small, 40% smaller than that of NAT2. This explains its affinity for the small substrates, PAS and PABA, compared to NAT2. In addition, a Phe²⁵ residue exists that forms part of the catalytic loop of NAT1 which aids the steric recognition of substrates and affects its selectivity for smaller substrates (Rodrigues-Lima et al, 2001; Wu et al, 2007). A study by Kubiak and colleagues (2013), highlighted the importance of the catalytic triad which exists in human NAT enzymes. The catalytic triad, Cys-His-Asp, is necessary for the catalysis by the enzyme family (Rodrigues-Lima, 2001; Kubiak et al, 2013). The work done by Bell and colleagues proves that even SNPs found outside the binding pocket of this enzyme can affect its functionality (Bell et al, 1995). The enzyme also contains “three key active site loop residues”, F125, Y127, and R129, which strongly influence substrate specificity of NAT1 (Hein, 2009). SNPs present in the coding region affect an individuals’ susceptibility to various disease. The novel SNPs in this study are all present in the coding region. Zhu and Hein (2008) investigated the functional effect of such SNPs and found that it affected an individuals’ phenotype and, ultimately, disease predisposition. The studies done on V231G and the data from our study show that the SNP decreases the acetylation rate of NAT1 in the presence of both PAS and PABA. With this knowledge, we can assume that the mutation would categorise an individual as a “slow acetylator”. Further structural studies are needed for the novel SNPs tested in this project.

Previous computational predictions by Polyphen-2 and SIFT showed that V231G affects NAT1 activity. These results were used as the premise in structural prediction studies by Cloet et al (2017). Different methods (eg. molecular dynamic simulations) were used to correct the contradictory and inaccurate results produced by the computational algorithms.

Their results confirm the computational findings for SNP V231G. It affects the NAT1 acetylating function. The experimental validation done in this study confirmed the prediction. It was discovered that V231G increases the rate of acetylation by NAT1 in the presence of PAS, although less than the other mutations studied, and decreases it in the presence of PABA when compared to the wild type. A study by Cloete and colleagues (2017) investigated the structural stability of V231G and its effect on the enzyme functionality. They found that the SNP destabilized the protein and increased the flexibility of the protein when compared to the wild type, meaning that the mutant protein is less thermodynamically stable than the wild type. Goodfellow and colleagues documented the importance of protein folding and flexibility with respect to different amino acid interactions as well as substrate-protein interactions which ultimately affect protein function (Goodfellow et al, 2000). Cloete and colleagues also tested the acetylation ability with distance measurements and used PABA as a probe drug. The mutant V231G showed to affect the core of the protein and showed greater distance of PABA from the catalytic triad Cysteine residue compared to the wild type which indicates weaker acetylation of PABA in comparison to the wild type NAT1. In addition, the mutant has the ability to rearrange the secondary structure of the protein and introduce loop regions close to the catalytic site. Cloete states that the protein moves more and folds less. This supports the findings of Rodrigues-Lima in 2001 which documents the effect protein folding has on its overall function (Rodrigues-Lima et al, 2001; Walraven 2008; Cloete et al, 2017). Furthermore, there could be other ionic interactions that still need to be uncovered that could be affecting the function of the enzyme. The structural findings reaffirms the experimental findings in this project for SNP V231G. Experimental data in this study for the anti-tuberculosis drug, PABA, we found a decreased acetylation rate compared to the wild type. The structural findings provide a multitude of reasons for these results (Table 17 and figure 16). The combination of findings could lead us to believe that the SNP V231G confers a slow acetylator phenotype to humans harbouring the mutant protein.

Similar to Cloete and Colleagues (2017), Akaguru tested the stability of the novel SNPs in this project using Gibbs free energy changes and based it on the interaction between the

side chains and the main chain of the protein and the change the mutation infers. It was found that SNP, T240S, V235A, D229H and V231G destabilizes the protein. This reaffirms findings by Cloete for SNP V231G. For the experimental validation done by this study, we can assume that the flexibility of the protein is affected by these mutants and that could be the reason for the effect it has on the acetylation rate changes. In addition, he found SNP T193S to be functionally neutral and SNP F202V to have a stabilizing effect. Akaguru confirmed the computational prediction made for T193S. Our data disproves their findings in that the SNP does not render the protein benign. The SNP F202V predicted to affect function and not structure. Again, Akaguru proved that the SNP does affect the structure of native NAT1 and for our experimental data, we can attribute the functional effect to the structural findings made here. The experimental results in this project show that all 6 of the novel nsSNPs tested affect the functionality of NAT1 when compared to the native wild type protein.

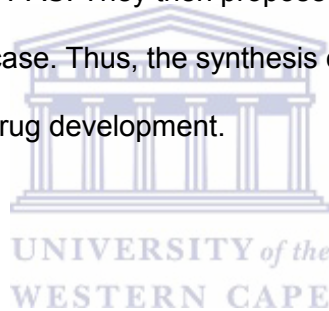
Millner and colleagues (2012), found that the acetylation phenotype conferred to any individual is substrate dependent based on the mutation the individual possesses. They used PABA as a substrate and found similar results to our study. Although we used different SNPs, we also found that NAT1 exerts a higher catalytic activity in the presence of PABA, however the SNPs do decrease the rate of acetylation and the resultant phenotype when compared to the WT. The resultant absorbance readings for the PABA assay are all higher than those for the PAS assay. Thus, PABA is more readily acetylated and rendered possibly ineffective for TB treatment. Decades later our findings, with the use of novel SNPs, supports the study done by Youmans et al (1947) where it was discovered that PAS is bacteriostatic in its effect against TB and is antagonistic to PABA (Youmans et al, 1947). Our data suggests that PAS is definitely a better treatment option against TB than PABA.

Grant et al (1990) showed that there is a correlation between the quantity of NAT1 in 50 human livers and the rate of acetylation. Thus, the reduced or absence of the enzyme reduces or eliminated the acetylation. This gives insight into the result we found for the empty vector sample and the reduced level of hydrolysis and subsequent acetylation in comparison to the WT and SNPs tested (Figures 15 and 16).

The localisation, solubility, quantity and absorption of NAT1 and the relevant drugs also play a major role in treatment efficacy. This allows for the modification of anti-tuberculosis treatments as well. Along with the discovery of PAS came the importance of combination drug treatment to overcome drug resistance in *M.tuberculosis* (O'Connor, 1948). PAS is still widely used for the treatment of XDR and MDR-TB (Minato et al, 2015). PAS is a second-line tuberculostatic drug which is effective against no other mycobacteria and is recommended for anti-tuberculosis treatment by the WHO. PAS was believed to disrupt the protein metabolism pathways of the mycobacterium and change the reaction of the host tissue in the presence of this tubercle bacillus. Studies have found that PAS inhibits Dihydropteroate synthase (DHPS) in the folate metabolism pathway of *M.tuberculosis* but does not inhibit the growth of the bacterium (Chakraborty et al, 2013; Zhao, 2014). The mode of action includes the bioactivation of PAS by the folate metabolism pathway and the subsequent inhibition of a pathway metabolite, dihydrofolate reductase (DHFR) (Rengarajan, 2004; Minato et al, 2015). Enzyme inhibition is important for new drug discovery. With the evidence that prokaryotes synthesize their own folate, it makes sense that this pathway would be a target for new drug development. However, other factors also need to be considered and so the search for new antitubercular drugs continues (Zheng et al, 2013). PAS resistance by the mycobacterium is strongly conferred by suppression of PAS bioactivation, target inhibition and limited drug accumulation in the mycobacteria (Minato et al, 2015).

PABA is a growth factor for *M. tuberculosis* and inhibits the sensitivity of the mycobacterium to PAS. This explains the findings by Chakraborty, 2013, investigation PAS inhibition but not stunted tubercle growth. General antitubercular treatments involve a combination of drugs, multidrug therapy. An interesting finding is that, when used in combination with INH and streptomycin, PAS reduces the acetylation of INH. This would be a suitable treatment for TB worldwide in addition to the introduction of intravenous treatment methods. This also gave insight into the emergence of drug resistant TB or the lack thereof (Krasnov, 2012). Decades ago, studies used PABA and PAS in conjunction to determine the effect it had on human

tubercle bacilli. Youmans et al (1947) found that when PABA is added at 0.1mg/100ml with PAS, the effect of PAS is reduced. These researchers used animal models as well and found that when PAS is used with Streptomycin, the survival of mice are prolonged. They found, however, that PAS is toxic at high concentrations. Studies to elucidate the function of PABA in TB treatment found that PABA antagonizes PAS treatment with respect to the folate metabolism pathways of *M.tuberculosis*. The two tuberculostatic drugs competitively antagonize one another. In this light, it was believed that large doses of PAS was necessary for TB treatment (Kakemi et al, 1967). Thiede and colleagues (2016) found that *M.tuberculosis* needs to maintain a basal level of PABA to maintain folate synthesis. However, a dose dependent increase in PABA was observed in the presence of PAS treatment. They concluded that this increased PABA levels plays a significant role in *M.tuberculosis* drug resistance to PAS. They then proposed that blocking PABA would eliminate drug resistance in this case. Thus, the synthesis of PABA in *M.tuberculosis* would become a viable drug target for drug development.



4.2 CONCLUSION

The expression and purification of recombinant proteins has played a pivotal role in research development as it allows for the characterization of a desired protein. The subsequent success contributes to the development of commercial and industrial processes as well as the understanding of certain complex diseases and the drugs to combat it (Rosano and Ceccarelli, 2014). Alterations in activity affect the phenotype of an individual and could very well lead to impaired treatment efficacy and disease predisposition. The investigation of NAT1 enzyme activity will give researchers an idea of the pharmacokinetics regarding the acetylation of arylamine containing drugs for clinical therapy against TB.

In this project, we aimed at and successfully expressed human, wild type NAT1 and 6 novel nsSNPs (T193S, T240S, V231G, V235A, F202V and D229H) in a bacterial system to research its acetylation function on its specific substrates, PABA and PAS.

With respect to the functional predictions obtained from PolyPhen-2 and SIFT for the specific SNPs, the experimental results oppose the hypothesis for T193S and T240S. The experimental data shows that these two SNPs affect the rate of acetylation in the presence of both substrates. For the remaining SNPs (V231G, V235A, F202V and D229H), the experimental data confirms the prediction that they affect the function of the protein. In this case they affect the acetylation rate of NAT1 in the presence of the specific substrates. The experimental data corresponds with the structural and functional predictions, for SNP V231G, made by Cloete and colleagues (2017).

We found that NAT1 has a higher overall catalytic activity in the presence of PABA when compared to the average absorbance readings of the PAS assay. The data in our study suggests that PAS is a better treatment option against TB than PABA. We also found that the SNPs increase acetylation in the presence of PAS and decrease it in the presence of PABA when compared to the wild type NAT1 protein. Furthermore, independent and combination structural studies are still necessary and more conclusive kinetic studies are necessary to fully understand the functional consequences of these novel SNPs. The assays performed on these novel SNPs provide new knowledge on its function. Thus, it will greatly influence research dedicated to TB drug discovery and allow researchers to better understand the enzyme activity and clinical function. It also gives more insight to the genotype-phenotype relationship for individuals as well as drug resistant tuberculosis. The findings in this study will allow for modifications of TB treatment strategies for human carrying these mutated proteins.

The characterization of nsSNPs present in NAT1 is important for anti-tuberculosis drug development. Tuberculosis resistance is increasing and the financial and logistical burden is high in areas where the disease is most prevalent. Thus, the urgency for better treatment strategies need to enjoy more attention. The resistance mechanism of this pathogen relies entirely on chromosomal mutations that either affect the enzyme activation of the drug or the drug itself (Sandgren, 2009).

The time constraints of this study did not allow for certain investigations to take place. The NAT1 protein should be expressed in other bacterial systems with different hosts in order to check for tighter expression control and reduce basal expression even more. This could also enhance the purification of the protein. In order to properly confirm NAT1 presence, specific anti-NAT1 antibodies should be used for western blots. Once properly expressed in a bacterial system, other organisms could be used as hosts.

As mentioned in the text, the structure of NAT1 strongly influences its substrate specificity, folding capacity and expression level. The latter directly affects the acetylation rate. Therefore, the structure-substrate interactions should be investigated for the novel nsSNPs used in this project. The effect the structure has on expression control should also be investigated. Furthermore, the novel SNPs in this study should be structurally tested against both substrates as it was done for PABA and V231G (Cloete et al, 2017). The pharmacokinetics should also be investigated for these novel SNPs as further clinical importance for TB treatment. The metabolic pathways affected by PAS and PABA should be better understood with respect to these novel NAT1 SNPs in order to further research fields focussing on drug resistant tuberculosis and TB treatment regimens and new drug development.

REFERENCES

1. (WHO/HTM/TB/2016.13) WHO. Global Tuberculosis Report. Switzerland: 2016.
2. Adam PJ, Berry J, Loader JA, Tyson KL, Craggs G, Smith P, De Belin J, Steers G, Pezzella F, Sachsenmeir KF, Stamps AC. Arylamine N-acetyltransferase-1 is highly expressed in breast cancers and conveys enhanced growth and resistance to etoposide *in vitro*. *Molecular cancer research*. 2003 Sep 1;1(11):826-35.
3. Agilent Technologies I. BL21 Competent Cells Instruction Manual. BL21(DE3) Competent Cells, BL21(DE3)pLysS Competent Cells, and BL21 Competent Cells 2010.
4. Akurugu WA. Effects of nucleotide variation on the structure and function of human arylamine n-acetyltransferase 1 (Doctoral dissertation). 2012.
5. Arslan S, Degerli N, Bardakci F. Distribution of N-acetyltransferase Type 1 (*NAT1*) genotypes and alleles in a Turkish population. *Genetics and Molecular Biology*. 2004;27(2):162-4.
6. Babalik A, Arda H, Bakirci N, Agca S, Oruc K, Kiziltas S, Cetintas G, Calisir HC. Management of and risk factors related to hepatotoxicity during tuberculosis treatment. *Tuberk Toraks*. 2012;60(2):136-44.
7. Badawi AF, Hirvonen A, Bell DA, Lang NP, Kadlubar FF. Role of aromatic amine acetyltransferases, *NAT1* and *NAT2*, in carcinogen-DNA adduct formation in the human urinary bladder. *Cancer research*. 1995 Nov 15;55(22):5230-7.
8. Bashiri S, Vikström D, Ismail N. Optimization of protein expression in *Escherichia Coli*. *BioPharm International*. 2015 May 7;28(5).
9. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. *Nature Structural & Molecular Biology*. 2000 Mar 1;7(3):209.
10. Bell DA, Badawi AF, Lang NP, Ilett KF, Kadlubar FF, Hirvonen A. Polymorphism in the N-acetyltransferase 1 (*NAT1*) polyadenylation signal: association of *NAT1* 10* allele with higher N-acetylation activity in bladder and colon tissue. *Cancer Research*. 1995 Nov 15;55(22):5226-9.
11. Blöndal K. Barriers to reaching the targets for tuberculosis control: multidrug-resistant tuberculosis. *Bulletin of the World Health Organization*. 2007 May;85(5):387-90.
12. Blum M, Heim M, Meyer UA. Nucleotide sequence of rabbit *NAT1* encoding monomorphic arylamine N-acetyltransferase. *Nucleic acids research*. 1990 Sep 11;18(17):5287-.
13. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*. 1976 May 7;72(1-2):248-54.

14. Bouakaze C, Keyser C, De Martino SJ, Sougakoff W, Veziris N, Dabernat H, Ludes B. Identification and genotyping of *Mycobacterium tuberculosis* complex species by use of a SNaPshot Minisequencing-based assay. *Journal of clinical microbiology*. 2010 May 1;48(5):1758-66.
15. Butcher NJ, Arulpragasam A, Goh HL, Davey T, Minchin RF. Genomic organization of human arylamine N-acetyltransferase Type I reveals alternative promoters that generate different 5'-UTR splice variants with altered translational activities. *Biochemical journal*. 2005 Apr 1;387(1):119-27.
16. Butcher NJ, Arulpragasam A, Catherine PO, Minchin RF. Identification of a minimal promoter sequence for the human N-acetyltransferase Type I gene that binds AP-1 (activator protein 1) and YY-1 (Yin and Yang 1). *Biochemical Journal*. 2003 Dec 1;376(2):441-8.
17. Butcher NJ, Boukouvala S, Sim E, Minchin RF. Pharmacogenetics of the arylamine N-acetyltransferases. *The pharmacogenomics journal*. 2002 Jan 1;2(1):30.
18. Butcher NJ, Ilett KF, Minchin RF. Substrate-Dependent Regulation of Human ArylamineN-Acetyltransferase-1 in Cultured Cells. *Molecular Pharmacology*. 2000 Mar 1;57(3):468-73.
19. Cattamanchi A, Dantes RB, Metcalfe JZ, Jarlsberg LG, Grinsdale J, Kawamura LM, Osmond D, Hopewell PC, Nahid P. Clinical characteristics and treatment outcomes of patients with isoniazid-monoresistant tuberculosis. *Clinical Infectious Diseases*. 2009 Jan 15;48(2):179-85.
20. Caylà JA, Orcau A. Control of tuberculosis in large cities in developed countries: an organizational problem. *BMC medicine*. 2011 Nov 28;9(1):127.
21. Chakraborty S, Gruber T, Barry CE, Boshoff HI, Rhee KY. Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in *Mycobacterium tuberculosis*. *Science*. 2012 Nov 1:1228980.
22. Dairou J, Atmane N, Rodrigues-Lima F, Dupret JM. Peroxynitrite Irreversibly Inactivates the Human Xenobioticmetabolizing Enzyme Arylamine N-Acetyltransferase 1 (NAT1) in Human Breast Cancer Cells. A cellular and mechanistic study. *Journal of Biological Chemistry*. 2004 Feb 27;279(9):7708-14.
23. de León JH, Vatsis KP, Weber WW. Characterization of Naturally Occurring and Recombinant HumanN-Acetyltransferase Variants Encoded by *NAT1*. *Molecular pharmacology*. 2000 Aug 1;58(2):288-99.
24. Evans DA, Manley KA, McKusick VA. Genetic control of isoniazid metabolism in man. *British medical journal*. 1960 Aug 13;2(5197):485.
25. Garibyan L, Avashia N. Polymerase chain reaction. *Journal of Investigative Dermatology*. 2013 Mar 31;133(3):1-4.

26. Goodfellow GH, Dupret JM, Grant DM. Identification of amino acids imparting acceptor substrate selectivity to human arylamine acetyltransferases *NAT1* and *NAT2*. *Biochemical Journal*. 2000 May 15;348(1):159-66.
27. Grant DM, Hughes NC, Janezic SA, Goodfellow GH, Chen HJ, Gaedigk A, Violeta LY, Grewal R. Human acetyltransferase polymorphisms. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 1997 May 12;376(1):61-70.
28. Grant DM, Mörike K, Eichelbaum M, Meyer UA. Acetylation pharmacogenetics. The slow acetylator phenotype is caused by decreased or absent arylamine N-acetyltransferase in human liver. *Journal of Clinical Investigation*. 1990 Mar;85(3):968.
29. Gumbo T, Louie A, Deziel MR, Liu W, Parsons LM, Salfinger M, Drusano GL. Concentration-dependent Mycobacterium tuberculosis killing and prevention of resistance by rifampin. *Antimicrobial agents and chemotherapy*. 2007 Nov 1;51(11):3781-8.
30. Gumbo T, Pasipanodya JG, Wash P, Burger A, McIlleron H. Redefining multidrug-resistant tuberculosis based on clinical response to combination therapy. *Antimicrobial agents and chemotherapy*. 2014 Oct 1;58(10):6111-5.
31. Hames BD, editor. *Gel electrophoresis of proteins: a practical approach*. OUP Oxford; 1998 Oct 1. 27.
32. Hearn MJ, Cynamon MH. Design and synthesis of antituberculars: preparation and evaluation against Mycobacterium tuberculosis of an isoniazid Schiff base. *Journal of Antimicrobial Chemotherapy*. 2004 Feb 1;53(2):185-91.
33. Hearse DJ, Weber WW. Multiple N-acetyltransferases and drug metabolism. Tissue distribution, characterization and significance of mammalian N-acetyltransferase. *Biochemical Journal*. 1973 Mar 15;132(3):519-26.
34. Hein DW. Molecular genetics and function of *NAT1* and *NAT2*: role in aromatic amine metabolism and carcinogenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2002 Sep 30;506:65-77.
35. Hein DW. N-acetyltransferase SNPs: emerging concepts serve as a paradigm for understanding complexities of personalized medicine. *Expert opinion on drug metabolism & toxicology*. 2009 Apr 1;5(4):353-66.
36. Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, Devanaboyina US, Nangju NA, Feng Y. Molecular genetics and epidemiology of the *NAT1* and *NAT2* acetylation polymorphisms. *Cancer Epidemiology and Prevention Biomarkers*. 2000 Jan 1;9(1):29-42.
37. Hein DW, McQueen CA, Grant DM, Goodfellow GH, Kadlubar FF, Weber WW. Pharmacogenetics of the arylamine N-acetyltransferases: A symposium in honor of Wendell W. Weber. *Drug metabolism and disposition*. 2000 Dec 1;28(12):1425-32.
38. Huang YS, Chern HD, Su WJ, Wu JC, Lai SL, Yang SY, Chang FY, Lee SD. Polymorphism of the N-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug-induced hepatitis. *Hepatology*. 2002 Apr 1;35(4):883-9.

39. Husain A, Barker DF, Doll MA, Hein DW. Identification of the major promoter and non-coding exons of the human arylamine N-acetyltransferase 1 gene (*NAT1*). *Pharmacogenetics and Genomics*. 2004 Jul 1;14(7):397-406.
40. Jacobson KR, Theron D, Victor TC, Streicher EM, Warren RM, Murray MB. Treatment outcomes of isoniazid-resistant tuberculosis patients, Western Cape Province, South Africa. *Clinical infectious diseases*. 2011 Aug 15;53(4):369-72..
41. Jensen EC. The basics of western blotting. *The anatomical record*. 2012 Mar 1;295(3):369-71.
42. Kawamura A, Graham J, Mushtaq A, Tsiftoglou SA, Vath GM, Hanna PE, Wagner CR, Sim E. Eukaryotic arylamine N-acetyltransferase: investigation of substrate specificity by high-throughput screening. *Biochemical pharmacology*. 2005 Jan 15;69(2):347-59.
43. Keese P, Graf L. A positive screen for cloning PCR products. *Nucleic acids research*. 1996 Sep 1;24(17):3474-5.40.
44. Khoo O, Suntrarachun S. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific journal of tropical biomedicine*. 2012 Feb 1;2(2):159-62.41.
45. Koul A, Arnoult E, Lounis N, Guillemont J, Andries K. The challenge of new drug discovery for tuberculosis. *Nature*. 2011 Jan 27;469(7331):483.
46. Krasnov VA. Intravenous Administration of PAS in TB Treatment. (2012).
47. Kubiak X, De La Sierra-Gallay IL, Chaffotte AF, Pluvinage B, Weber P, Haouz A, Dupret JM, Rodrigues-Lima F. Structural and biochemical characterization of an active arylamine N-acetyltransferase possessing a non-canonical Cys-His-Glu catalytic triad. *Journal of Biological Chemistry*. 2013 Aug 2;288(31):22493-505..
48. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. *Journal of visualized experiments: JoVE*. 2012(62).
49. Levy GN, Weber WW. 11 Arylamine Acetyltransferases. *Enzyme systems that metabolise drugs and other xenobiotics*. 2001:441.
50. Lin HJ, Han CY, Lin BK, Hardy S. Slow acetylator mutations in the human polymorphic N-acetyltransferase gene in 786 Asians, blacks, Hispanics, and whites: application to metabolic epidemiology. *American journal of human genetics*. 1993 Apr;52(4):827.
51. Liu L, Von Vett A, Zhang N, Walters KJ, Wagner CR, Hanna PE. Arylamine N-Acetyltransferases: Characterization of the Substrate Specificities and Molecular Interactions of Environmental Arylamines with Human *NAT1* and *NAT2*. *Chemical Research in Toxicology*. 2007;20(9):1300-8.
52. Lorenz TC. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of visualized experiments: JoVE*. 2012(63).
53. Mahmood T, Yang PC. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences*. 2012 Sep;4(9):429.

54. Meyer UA. Polymorphism of human acetyltransferases. Environmental health perspectives. 1994 Oct;102(Suppl 6):213.
55. Millner LM, Doll MA, Cai J, Hein DW. Phenotype of the most common “slow acetylator” arylamine N-acetyltransferase 1 genetic variant (*NAT1* 14B*) is substrate-dependent. Drug Metabolism and Disposition. 2012 Jan 1;40(1):198-204.
56. Minato Y, Thiede JM, Kordus SL, McKlveen EJ, Turman BJ, Baughn AD. Mycobacterium tuberculosis Folate Metabolism and the Mechanistic Basis for para-Aminosalicylic Acid Susceptibility and Resistance. Antimicrobial agents and chemotherapy. 2015;59(9):5097-106.
57. Minchin RF, Hanna PE, Dupret JM, Wagner CR, Rodrigues-Lima F, Butcher NJ. Arylamine N-acetyltransferase I. The international journal of biochemistry & cell biology. 2007 Dec 31;39(11):1999-2005.
58. Möller M, Nebel A, Kwiatkowski R, van Helden PD, Hoal EG, Schreiber S. Host susceptibility to tuberculosis: CARD15 polymorphisms in a South African population. Molecular and cellular probes. 2007 Apr 30;21(2):148-51.
59. Nebert DW. Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist?. American journal of human genetics. 1997 Feb;60(2):265.
60. Novagen. Protein Expression. In: Novagen, editor. pET system tutorial 2002-2003.
61. O'Connor JA. Para-aminosalicylic acid in the treatment of tuberculosis. Postgraduate medical journal. 1948 Sep;24(275):455.
62. Padmanabhan S, Banerjee S, Mandi N. Screening of bacterial recombinants: strategies and preventing false positives. In Molecular Cloning-Selected Applications in Medicine and Biology 2011. InTech.
63. Pantelidis P. Tuberculosis: an ancient disease still confusing our genes. Respiration. 2005;72(4):347-8.
64. Payton MA, Sim E. Genotyping human arylamine N-acetyltransferase type 1 (*NAT1*): the identification of two novel allelic variants. Biochemical pharmacology. 1998 Feb 1;55(3):361-6.
65. Pieters J. Mycobacterium tuberculosis and the macrophage: maintaining a balance. Cell host & microbe. 2008 Jun 12;3(6):399-407.
66. Pompeo F, Brooke E, Kawamura A, Mushtaq A, Sim E. The pharmacogenetics of NAT: structural aspects. Pharmacogenomics. 2002 Jan 1;3(1):19-30.
67. Rengarajan J, Sasseti CM, Naroditskaya V, Sloutsky A, Bloom BR, Rubin EJ. The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria. Molecular microbiology. 2004 Jul 1;53(1):275-82.
68. B. Riddle, W.P. Jencks, A. Acetyl-Coenzyme, Arylamine N-Acetyltransferase role of the acetyl-enzyme intermediate and the effects of substituents on the rate, J. Biol. Chem. 246 (1971) 3250–3258

69. Rodrigues-Lima F, Delomenie C, Goodfellow GH, Grant DM, Dupret JM. Homology modelling and structural analysis of human arylamine N-acetyltransferase NAT1: evidence for the conservation of a cysteine protease catalytic domain and an active-site loop. *Biochemical Journal*. 2001 Jun 1;356(2):327-34.
70. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in microbiology*. 2014;5.
71. Roy VK, Kumar NS, Gurusubramanian G. Proteins—structure, properties and their separation by SDS—polyacrylamide gel electrophoresis. *Science Vision*. 2012 Oct;12:170-81.
72. Ruiz JD, Agúndez JA, Martínez C, García-Martín E. NAT1 (N-acetyltransferase 1 (arylamine N-acetyltransferase)).
73. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. *PLoS medicine*. 2009 Feb 10;6(2):e1000002.
74. Sandy J, Holton S, Fullam E, Sim E, Noble M. Binding of the anti-tubercular drug isoniazid to the arylamine N-acetyltransferase protein from *Mycobacterium smegmatis*. *Protein Science : A Publication of the Protein Society*. 2005;14(3):775-82.
75. Sharma S, Mohan A. Antituberculosis treatment-induced hepatotoxicity: from bench to bedside. *medicine*. 2005:480.
76. Sim E, Abuhammad A, Ryan A. Arylamine N-acetyltransferases: from drug metabolism and pharmacogenetics to drug discovery. *Br J Pharmacol*. 2014;171(11):2705-25.
77. Sim E, Lack N, Wang C-J, Long H, Westwood I, Fullam E, et al. Arylamine N-acetyltransferases: structural and functional implications of polymorphisms. *Toxicology*. 2008;254(3):170-83.
78. Sim E, Payton M, Noble M, Minchin R. An update on genetic, structural and functional studies of arylamine N-acetyltransferases in eucaryotes and procaryotes. *Human Molecular Genetics*. 2000;9(16):2435-41.
79. Sim E, Pinter K, Mushtaq A, Upton A, Sandy J, Bhakta S, et al. Arylamine N-acetyltransferases: a pharmacogenomic approach to drug metabolism and endogenous function. *Biochemical Society transactions*. 2003;31(Pt 3):615-9.
80. Sinclair J, Sim E. A fragment consisting of the first 204 amino-terminal amino acids of human arylamine N-acetyltransferase one (NAT1) and the first transacetylation step of catalysis. *Biochemical pharmacology*. 1997;53(1):11-6.
81. Stockton JC, Howson JM, Awomoyi AA, McAdam KP, Blackwell JM, Newport MJ. Polymorphism in NOD2, Crohn's disease, and susceptibility to pulmonary tuberculosis. *FEMS immunology and medical microbiology*. 2004;41(2):157-60.
82. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*. 1986;189(1):113-30.

83. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*. 2006;72(2):211-22.
84. Thiede JM, Kordus SL, Turman BJ, Buonomo JA, Aldrich CC, Minato Y, et al. Targeting intracellular p-aminobenzoic acid production potentiates the anti-tubercular action of antifolates. *Scientific reports*. 2016;6:38083.
85. Upton AM, Mushtaq A, Victor TC, Sampson SL, Sandy J, Smith DM, et al. Arylamine N-acetyltransferase of *Mycobacterium tuberculosis* is a polymorphic enzyme and a site of isoniazid metabolism. *Molecular microbiology*. 2001;42(2):309-17.
86. Vatsis KP, Yao Y and Weber WW (1994) Allelic variation at the *NAT1* locus of N-acetyltransferase among Orientals. *Am J Hum Genet* 55(suppl):A340
87. Vatsis KP, Martell KJ, Weber WW. Diverse point mutations in the human gene for polymorphic N-acetyltransferase. *Proceedings of the National Academy of Sciences*. 1991 Jul 15;88(14):6333-7.
88. Walraven JM, Trent JO, Hein DW. Structure-function analyses of single nucleotide polymorphisms in human N-acetyltransferase 1. *Drug metabolism reviews*. 2008 Jan 1;40(1):169-84.
89. Windmill KF, Gaedigk A, de la M. Hall P, Samaratinga H, Grant DM, McManus ME. Localization of N-Acetyltransferases *NAT1* and *NAT2* in Human Tissues. *Toxicological Sciences*. 2000;54(1):19-29.
90. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, van Soolingen D, Rüsche-Gerdes S, Loch C, Brisse S, Meyer A. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS pathogens*. 2008 Sep 26;4(9):e1000160.
91. Wu H, Dombrovsky L, Tempel W, Martin F, Loppnau P, Goodfellow GH, Grant DM, Plotnikov AN. Structural basis of substrate-binding specificity of human arylamine N-acetyltransferases. *Journal of Biological Chemistry*. 2007 Oct 12;282(41):30189-97.
92. Youmans GP, Raleigh GW, Youmans AS. The Tuberculostatic Action of para-Aminosalicylic Acid. *Journal of Bacteriology*. 1947;54(4):409-16.
93. Zhangwei X, Jianming X, Qiao M, Xinhua X. N-Acetyltransferase-1 gene polymorphisms and correlation between genotype and its activity in a central Chinese Han population. *Clinica chimica acta; international journal of clinical chemistry*. 2006;371(1-2):85-94.
94. Zhao F, Wang XD, Erber LN, Luo M, Guo AZ, Yang SS, Gu J, Turman BJ, Gao YR, Li DF, Cui ZQ. Binding pocket alterations in dihydrofolate synthase confer resistance to para-aminosalicylic acid in clinical isolates of *Mycobacterium tuberculosis*. *Antimicrobial agents and chemotherapy*. 2014 Mar 1;58(3):1479-87.

95. Zheng J, Rubin EJ, Bifani P, Mathys V, Lim V, Au M, Jang J, Nam J, Dick T, Walker JR, Pethe K. para-Aminosalicylic acid is a prodrug targeting dihydrofolate reductase in Mycobacterium tuberculosis. *Journal of Biological Chemistry*. 2013 Aug 9;288(32):23447-56.
96. Zhu Y, Hein DW. Functional effects of single nucleotide polymorphisms in the coding region of human N-acetyltransferase 1. *The pharmacogenomics journal*. 2008 Oct;8(5):339.



APPENDIX I – NAT1 SEQUENCES

NAT1 open reading frame

The ORF of wild type NAT1 shown here annotates the start and stop codons (bold and underlined) as well as the nucleotide changes (underlined) for the 6 individual, novel, non-synonymous SNPs in our study.

Atggacattgaagcatatcttgaaagaattggctataag
Aagtctaggaacaaattggacttggaaacattaactgacattcttcaacaccagatccga
gctgttccctttgagaaccttaacatccattgtggggatgccatggacttaggcttagag
gccatTTTTgatcaagttgtgagaagaaatcggggtggatggtgtctccaggccaatcat
cttctgtactgggctctgaccactattggTTTTgagaccacgatggtggagggtatggt
tacagcactccagccaaaaatacagcactggcatgattcaccttctcctgcaggtgacc
attgatggcaggaactacattgtcgatgctgggTTTTggacgctcataccagatgtggcag
cctctggagttaatttctgggaaggatcagccctcagggtgcttctgtgtcttccgtttgacg
gaagagaatggattctggtatctagaccAAATcagaagggAACagtacattccaaatgaa
gaatttcttcattctgatctcctagaagacagcaaataccgAAAActactcctttact
cttaagcctcgaacaattgaagattttgagtctatgaatacatacctgcagacatctcca
tcactctgtgtttactagtaaatacttttgttccttgcagaccccagatggggttcactgt
ttgggtgggcttcaccctcaccataggagattcaattataaggacaatacagatctaata
gagttcaagactctgagtgaggaagaaatagaaaaagtgctgaaaaatatatttaatt
tccttgcagagaaagcttgtgcccacatggtgatagatttttactatttag

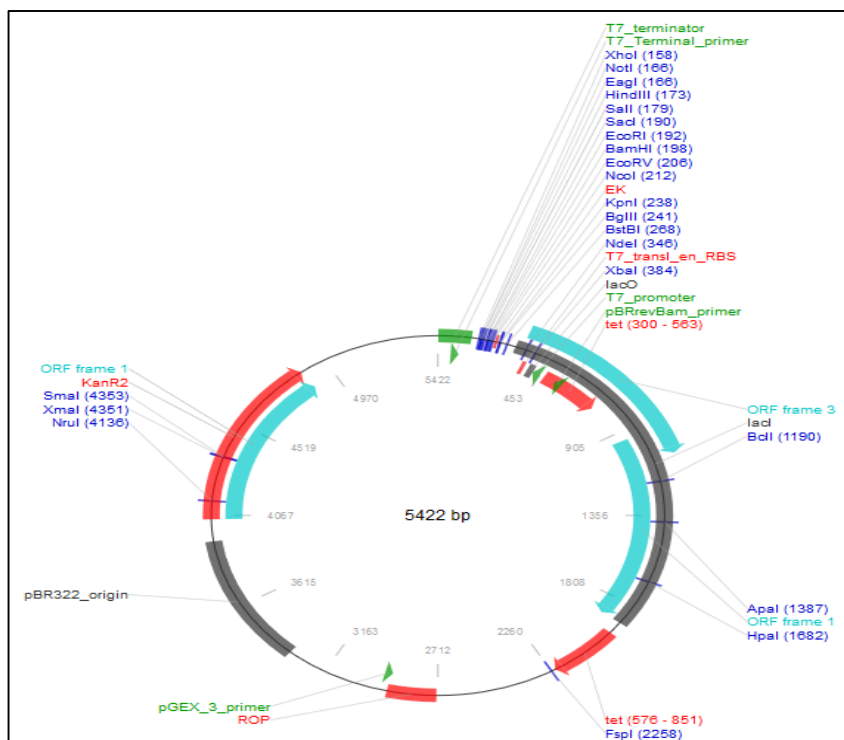
NAT1 amino acid sequence

MDIEAYLERIGYKKS RNKLDLET LTDILQH QIRAVPFENLNIHC
GDAMD LGLEAIFDQV VRRNRGGWCLQVN HLLYWALTTIGFET TMLGGYVYSTPAKKYS
TGM IHLL LQVTIDGR NYIVDAGFGRSYQ MWQPLELISGKDQPQVPCV FRLTEENGFY
LDQIRREQYIPNEEF LHSDDLLED SKYRKIYSFTLKPRTIEDFESMNTYLQTS PSSVFT
SKSFCSLQTPDGVHCLVGF TLTTHRRFNYKDNTDLIEFKTLSEEEIEKVLKNI FNISLQ
RKLVPKHGDRFFTI

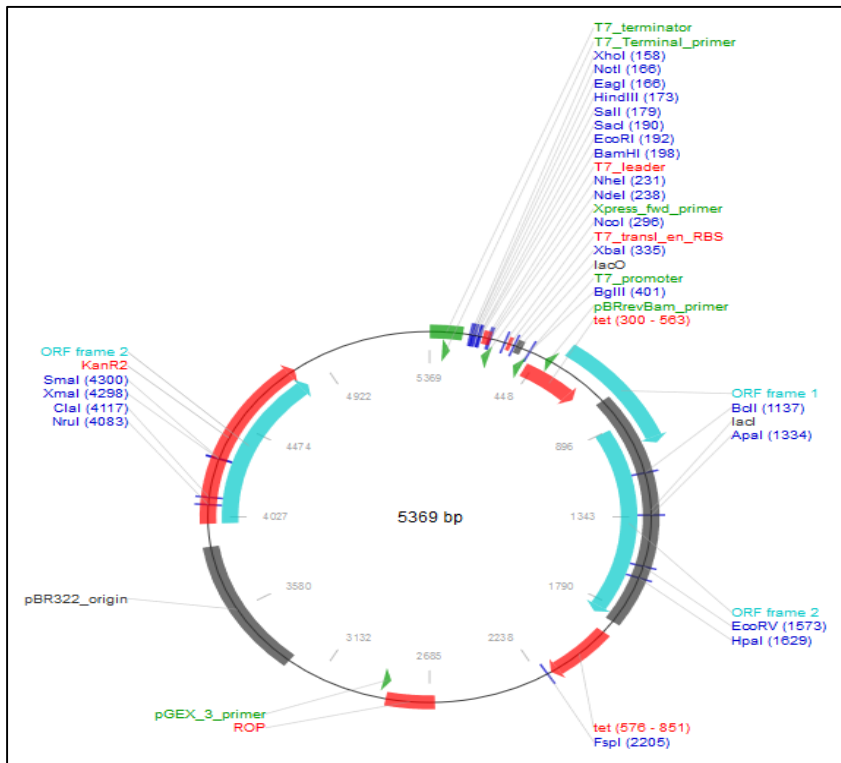
APPENDIX II – VECTOR MAPS

The vectors maps were all downloaded from Addgene vector database and show the different restriction sites, multiple cloning sites and well as the sizes of the different vectors worked with for the duration of this project.

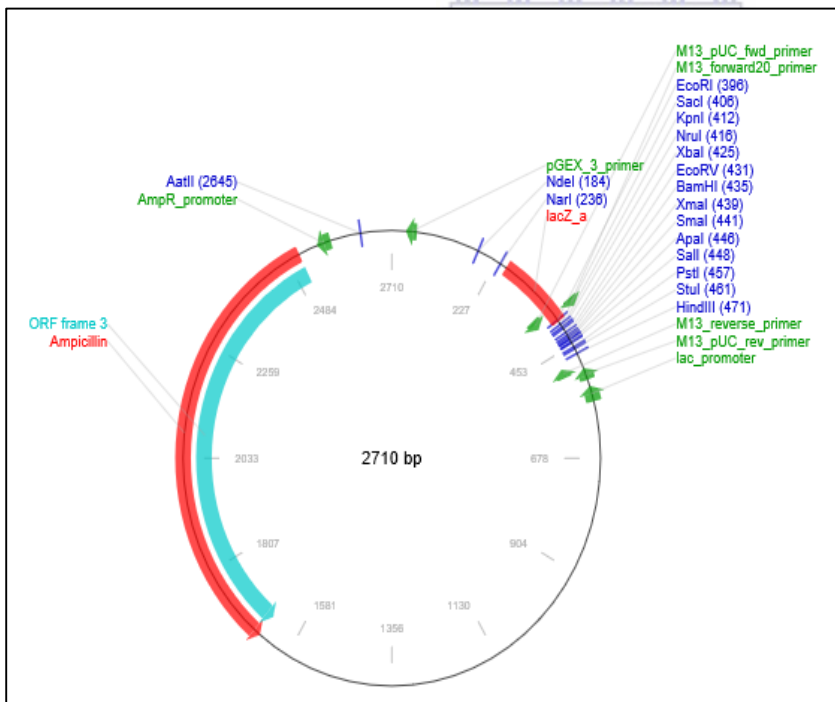
pET-30a(+)



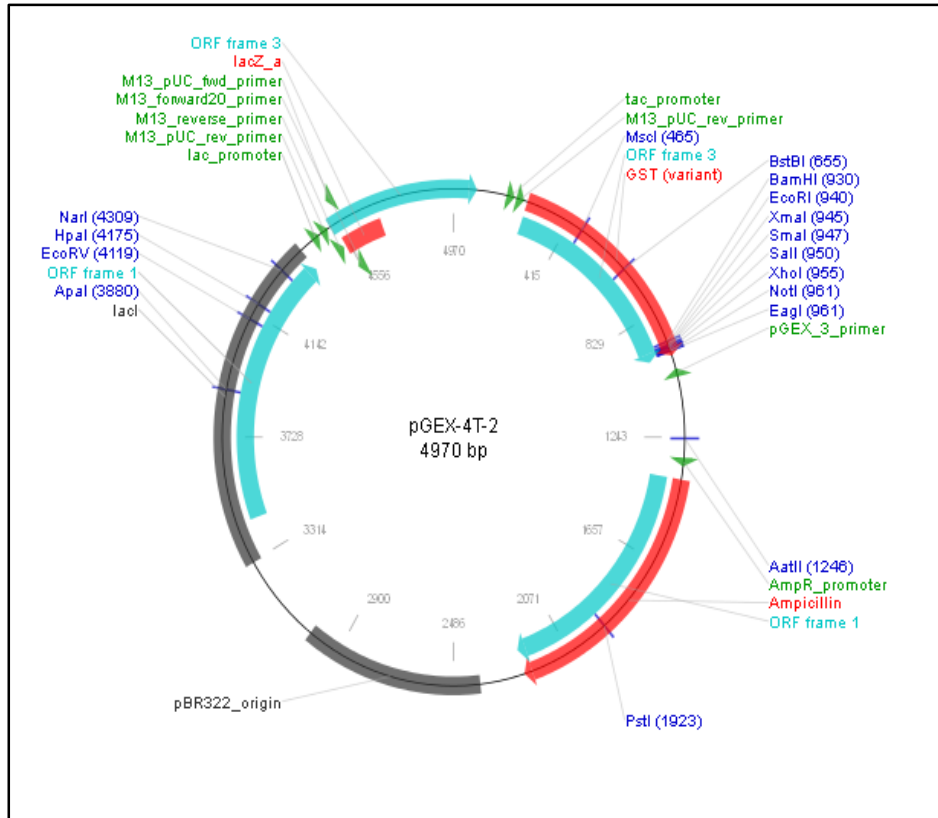
pET-28a(+)



pUC57



pGEX-4T-2



APPENDIX III - PROTOCOLS

PureYield™ Plasmid Miniprep System

DNA Purification by Centrifugation

Prepare Lysate

1. Add 600µl of bacterial culture to a 1.5ml microcentrifuge tube. harvest and process up to 3ml of bacterial culture.
2. Add 100µl of Cell Lysis Buffer (Blue), and mix by inverting the tube 6 times.
3. Add 350µl of cold (4–8°C) Neutralization Solution, and mix thoroughly by inverting.
4. Centrifuge at maximum speed in a microcentrifuge for 3 minutes.
5. Transfer the supernatant (~900µl) to a PureYield™ Minicolumn without disturbing the cell debris pellet.
6. Place the minicolumn into a Collection Tube, and centrifuge at maximum speed in a microcentrifuge for 15 seconds.
7. Discard the flowthrough, and place the minicolumn into the same Collection Tube.

Wash

8. Add 200µl of Endotoxin Removal Wash (ERB) to the minicolumn. Centrifuge at maximum speed in a microcentrifuge for 15 seconds.
9. Add 400µl of Column Wash Solution (CWC) to the minicolumn. Centrifuge at maximum speed in a microcentrifuge for 30 seconds.

Elute

10. Transfer the minicolumn to a clean 1.5ml microcentrifuge tube, then add 30µl of Elution Buffer or nuclease-free water directly to the minicolumn matrix. Let stand for 1 minute at room temperature.
11. Centrifuge for 15 seconds to elute the plasmid DNA. Cap the microcentrifuge tube, and store eluted plasmid DNA at –20°C.

MagneHis™ Protein Purification System

Purification of Polyhistidine- or HQ-Tagged Proteins from 1ml of Bacterial Culture Using MagneHis™ Ni-Particles

1. Add 500mM NaCl to HQ-tagged protein lysate (i.e., 0.03g NaCl per 1.0ml of lysate) to improve binding to MagneHis™ Ni-Particles.
2. Vortex the MagneHis™ Ni-Particles to a uniform suspension.
3. Add 30µl of MagneHis™ Ni-Particles either to cell pellet resuspended in 1X FastBreak™ Cell Lysis Reagent (from Section 3.B, Method 1, Step 7) or to 1.1ml of cell lysate (from Section 3.B, Method 2, Step 3).
4. Invert tube to mix (approximately 10 times), and incubate for 2 minutes at room temperature. Make sure the MagneHis™ Ni-Particles are well mixed.
5. Place the tube in the appropriate magnetic stand for approximately 30 seconds to capture the MagneHis™ Ni-Particles. Using a pipette, carefully remove the supernatant.
6. Remove the tube from the magnetic stand. Add 150µl of MagneHis™ Binding/Wash Buffer to the MagneHis™ Ni-Particles and pipet to mix. If NaCl was added for binding, also use NaCl during washing. Make sure that particles are resuspended well.
7. Place the tube in the appropriate magnetic stand for approximately 30 seconds to capture the MagneHis™ Ni-Particles. Using a pipette, carefully remove the supernatant.
8. Repeat the wash step 2 times for a total of 3 washes.
9. Remove the tube from the magnetic stand. Add 100µl of MagneHis™ Elution Buffer, and pipet to mix.
10. Incubate for 1–2 minutes at room temperature. Place in a magnetic stand to capture the MagneHis™ Ni-Particles. Using a pipette, remove the supernatant containing the purified protein. Analyse the samples by SDS-PAGE or by functional assay.

APPENDIX IV – PCR OPTIMIZATION

All PCR optimization reactions were performed with primers specific to the entire NAT gene and, therefore, the band of interest was expected to be at a size of ~1615 bp. In some cases the band is observed. However, the primers were not correct for the ORF of NAT1 to amplify the expected size of ~879 bp for subsequent protein expression and analysis.

PCR at 50°C

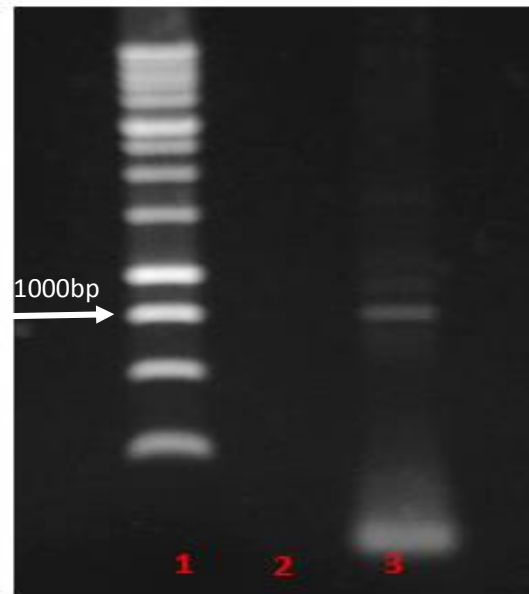


Figure 1. PCR at 50°C. Lane 1 – 1 kb molecular marker; lane 2 – negative control; lane 3 – experimental *NAT1*-pUC57

PCR AT 52°C

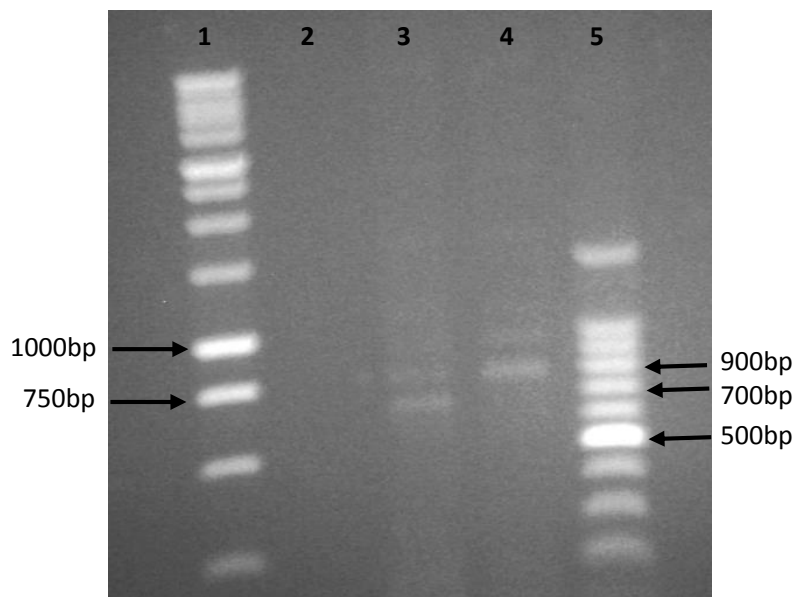


Figure 2. PCR at 52°C. Lane 1 – 1 kb molecular marker; lane 2 – negative control; lane 3 – positive control; lane 4 – experimental NAT1-pUC57; lane 5 – 100 bp molecular marker.

PCR at 54°C

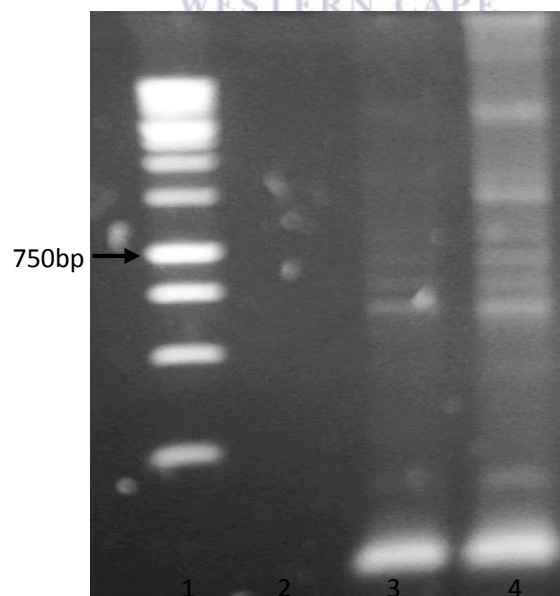
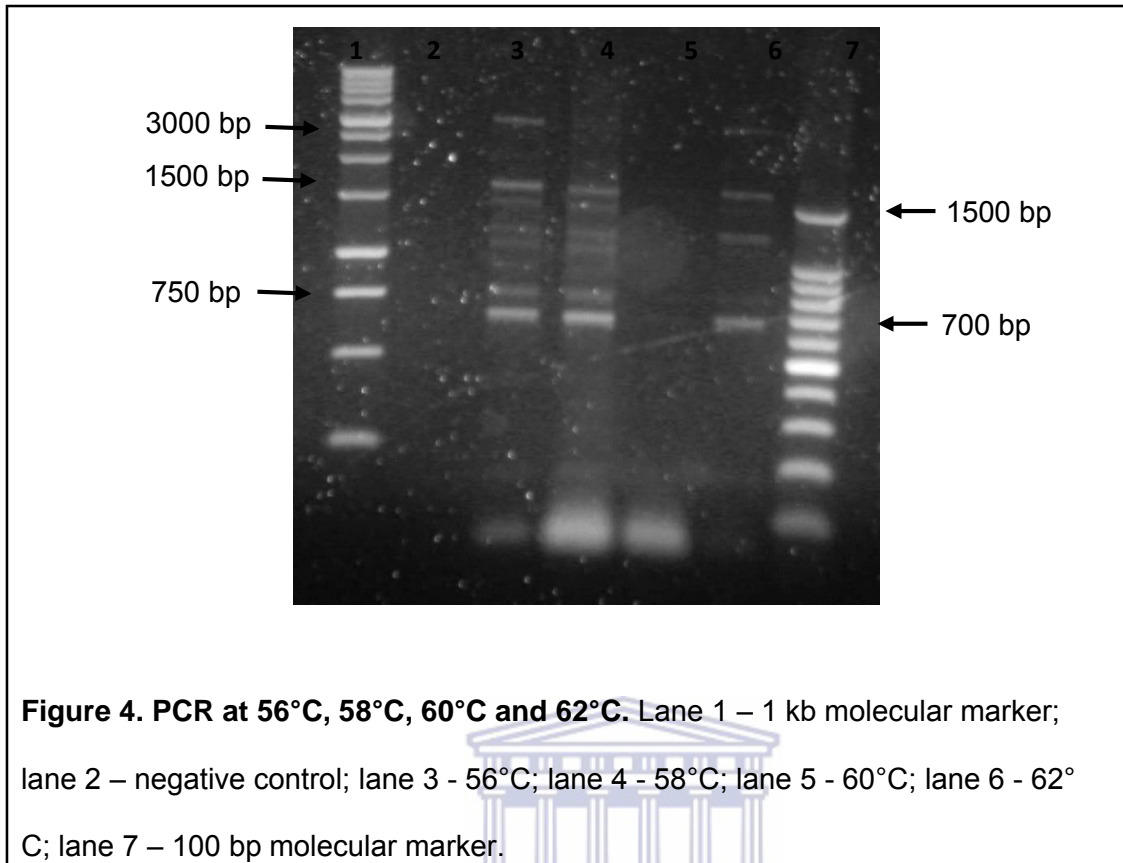
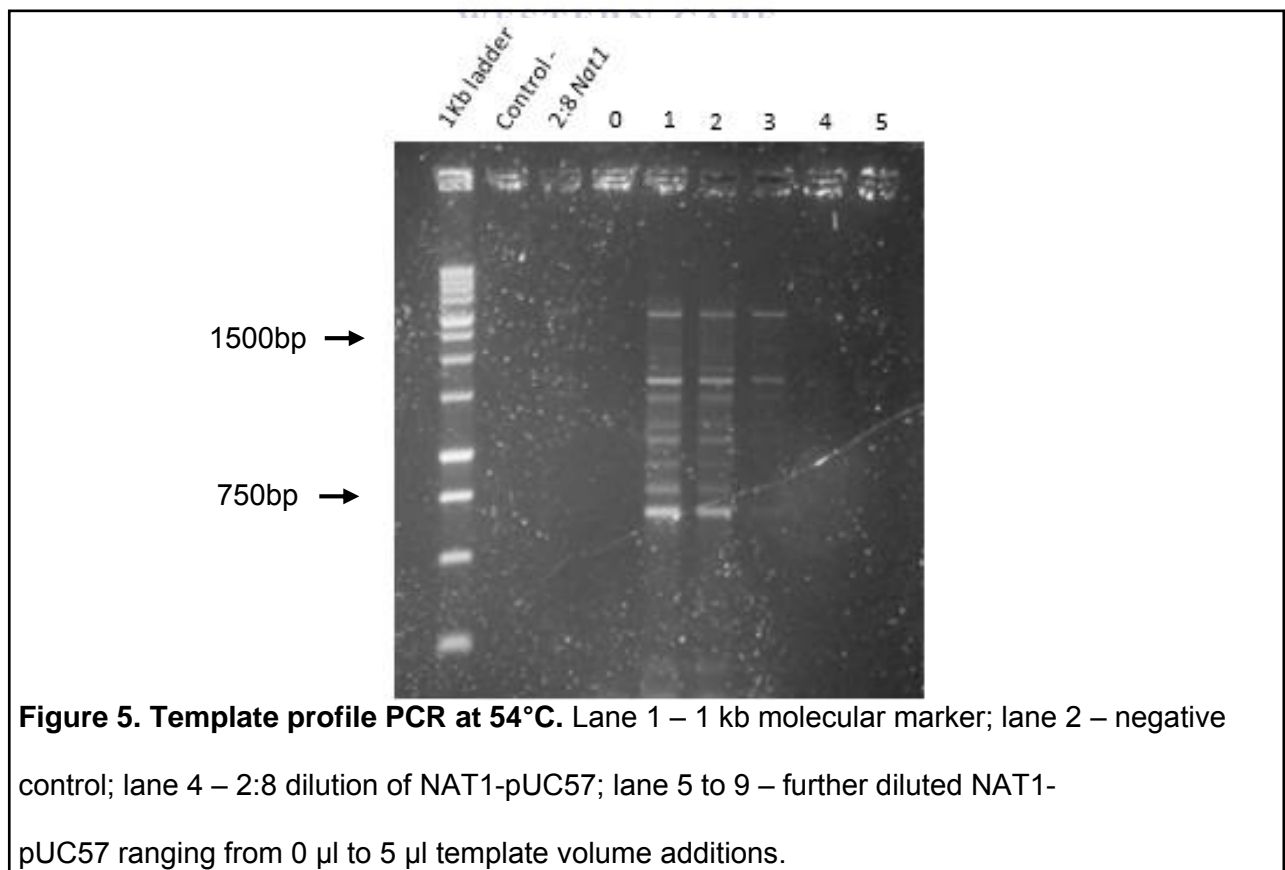


Figure 3. PCR at 54°C. Lane 1 – 1 kb molecular ladder; lane 2 – negative control; lane 3 – positive control; lane 4 – experimental NAT1-pUC57.

PCR at 56°C, 58°C, 60°C and 62°C



DNA template profile PCR



PCR nucleotide mix (dNTP) profile PCR

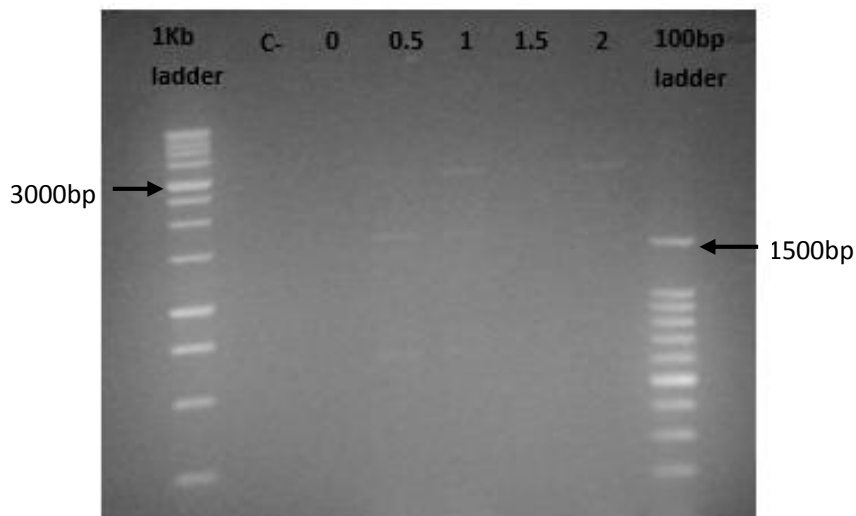


Figure 6. PCR nucleotide mix profile PCR at 54°C. Lane 1 – 1 kb molecular marker; lane 2 – negative control; lanes 3 to 7 – different volumes of PCR nucleotide mix added to PCR mix; lane 8 – 100 bp molecular marker.

MgCl₂ profile PCR

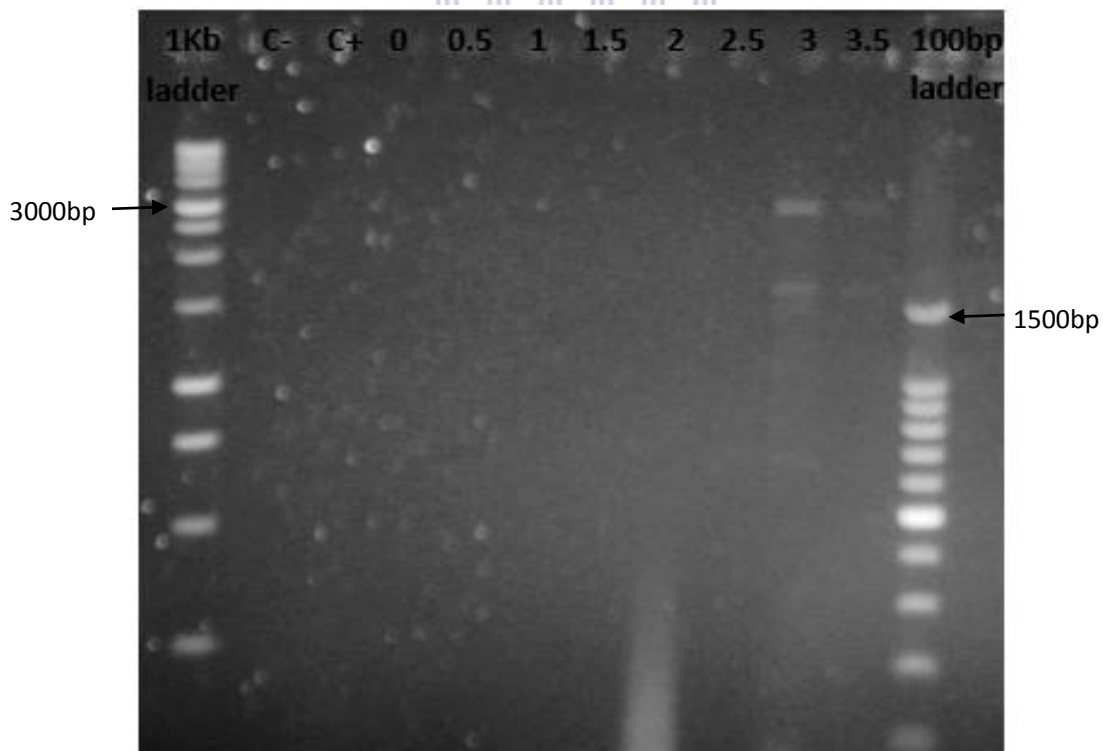


Figure 7. MgCl₂ profile PCR at 54°C. Lane 1 – 1 kb molecular marker; lane 2 – negative control; lane 3 – positive control; lane 3 to 10 – different volumes of MgCl₂ added to the PCR mix; lane 11 – 100 bp molecular marker.

APPENDIX V – RE DIGEST OF pGEX-4T-2

The RE digest result below depicts the linearization of pGEX-4T-2 with restriction endonuclease XhoI according to the conditions described in table 9, chapter 2. The expected product size was ~5000bp which proves the digest to be successful.

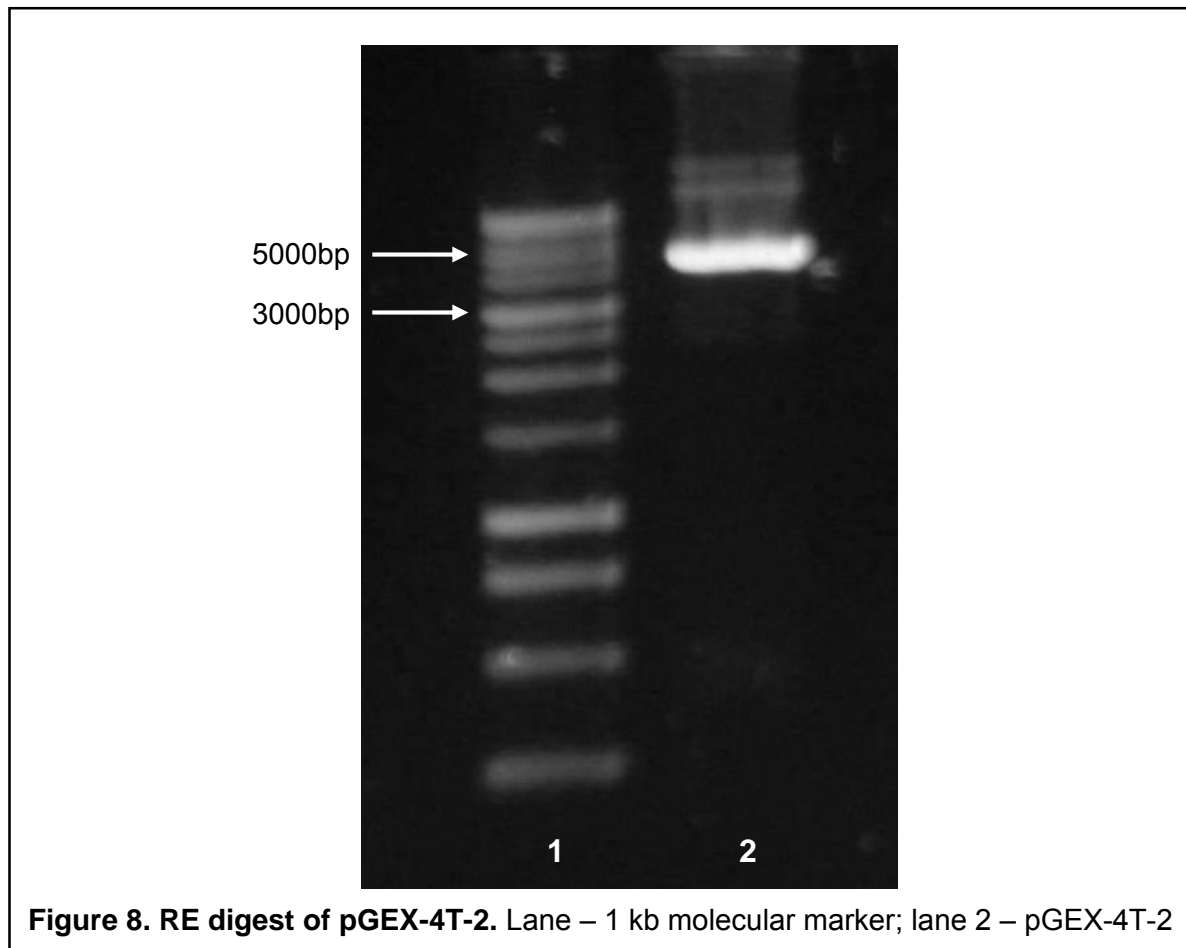


Figure 8. RE digest of pGEX-4T-2. Lane – 1 kb molecular marker; lane 2 – pGEX-4T-2

APPENDIX VI – PROTEIN EXPRESSION OPTIMIZATION

The different optimisation reactions showed the NAT1 protein successfully expressed. However, we were unable to eliminate the background and basal expression seen in the figures below.

Induction with 1mM IPTG at specific OD readings with fermentation at 30 °C

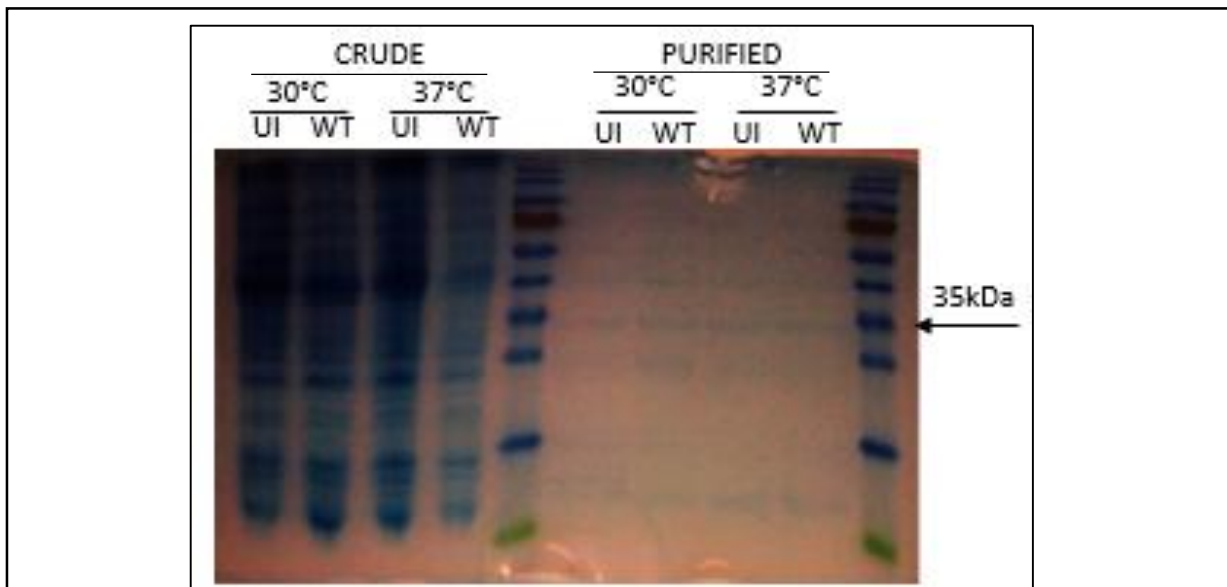


Figure 9. SDS-PAGE gel at 30°C fermentation. NAT1 uninduced (UI) and wild type (WT) samples at the normal 37°C and the new 30°C for optimization

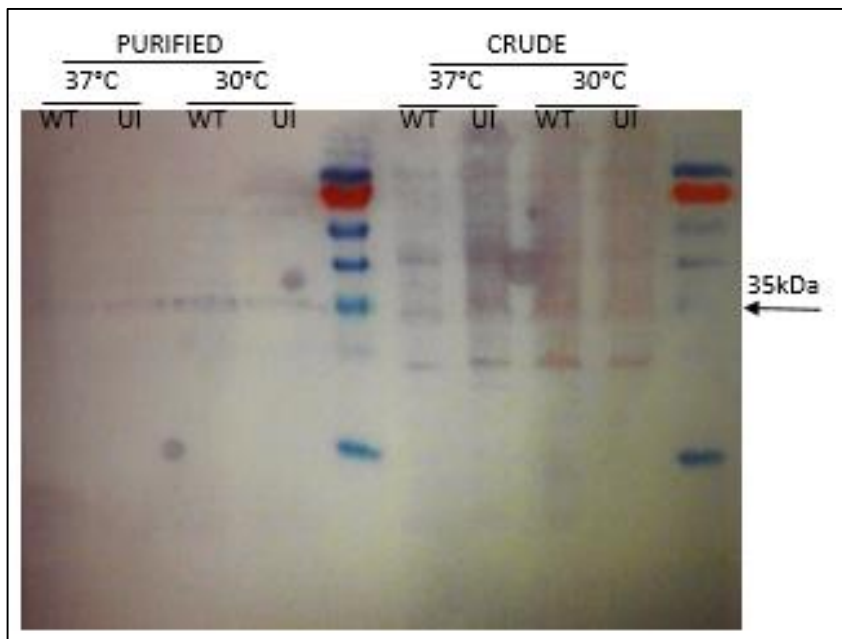


Figure 10. Western Blot transfer at 30°C fermentation. NAT1 uninduced (UI) and wild type (WT) samples at the normal 37°C and the new 30°C for optimization.

Increased CAM⁺ concentration to 50 µg/ml and decreased induction time from 4 hours to 2 hours

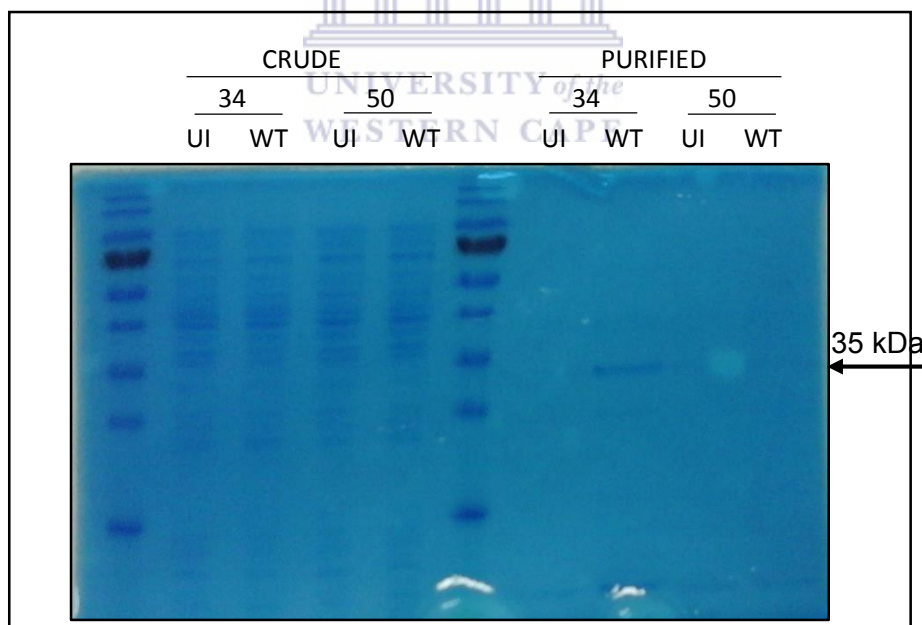


Figure 11. SDS-PAGE gel at 34 µg/ml CAM⁺ concentration and the new 50 µg/ml CAM⁺ concentration. NAT1 expression, crude and purified samples, comparing the chosen 34 µg/ml CAM⁺ concentration and the new 50 µg/ml CAM⁺ concentration (shown by the numbers above the lanes) for the uninduced (UI) and wild type (WT) samples.



UNIVERSITY *of the*
WESTERN CAPE