The role of the GLY loop of endonuclease VIII DNA glycosylase in maintaining genome integrity in *Mycobacterium smegmatis*

JESSICA SIAN BROTHWELL

1114578

A dissertation submitted to the Faculty of Health Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine.

2021

Declaration

I, Jessica Brothwell (student number: 1114578) declare that this dissertation is my own work. It is being submitted for the degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

J. Brothwell **30/07/2021**

(Jessica Brothwell) **Date**

This dissertation is dedicated to Victoria Rogers.

May your soul continue to live on through all of us.

Abstract

During infection, *Mycobacterium tuberculosis* bacteria are engulfed by macrophages as part of the host's innate immune response. This response generates oxidative stress on the mycobacteria, causing DNA damage and mutagenesis. Thus, DNA repair pathways are essential in the maintenance of genome integrity and cell survival. A previous study conducted in *Mycobacterium smegmatis*, a non-pathogenic close relative of *M. tuberculosis,* indicated a novel interplay between the endonuclease III (Nth) DNA glycosylase and two homologues of the endonuclease VIII (NeiI and NeiII) DNA glycosylases of the base excision repair pathway, in controlling DNA damage under oxidative stress conditions. A previous study indicated that the combinatorial loss of the *nth* and *neiI* genes has a larger impact on *M. smegmatis* survival under oxidative stress compared to when the *nth* and *neiII* genes were simultaneously lost. Bioinformatic analysis of NeiI and NeiII DNA glycosylases show that the two enzymes differ by three amino acid residues in the DNA intercalation loop that is responsible for the catalysis of binding damaged DNA to the DNA glycosylase. The NeiI enzyme has GLY residues whilst the NeiII homologue has the KME residues. In *E. coli,* mutation of the intercalation loop of the DNA- binding protein (SSB) resulted in a 100-fold decrease in DNA binding efficiency in the mutant compared to the parental strain. Hence, in this study we attempted to assess whether the GLY residues of the NeiI protein are indeed responsible for DNA glycosylase activity. Recombinant vectors containing the *neiI* gene with the GLY loop and another with the GLY loop changed to a GAF loop were genetically engineered and electroporated into the double deletion mutant lacking the *neiI* and *nth* genes (*ΔnIΔnth::nI and ΔnIΔnth::ΔGAF*). Growth kinetics of these strains indicate a similar growth pattern for the wildtype, *ΔnIΔnth::nI and ΔnIΔnth::ΔGAF* strains, and no significant difference in growth between the strains was noted. The resulting mutants were then phenotypically characterized together with the respective parental strains to assess if the GLY loop is responsible for the increased DNA glycosylase activity of NeiI. Deletion of the *nth* gene combined with the *neiI* gene led to an increase in spontaneous mutagenesis in response to rifampicin when compared to the wildtype mc^2155 strain. However, the increased spontaneous mutations were also observed in the *ΔnIΔnth::nI and ΔnIΔnth::ΔGAF* strains when compared to the parental strain. Thus, alteration of the GLY residues to GAF residues of the NeiI endonuclease did not significantly impact the genome integrity of the NeiI endonuclease.

Acknowledgements

I would like to thank my financial sponsors: The University of the Witwatersrand (PGMA) and the DST/NRF Centre of Excellence for Biomedical TB Research (CBTBR).

To my supervisor, Dr. Bhavna Gordhan, thank you for your guidance and support over the course of my research. Your scientific knowledge is exceptional, and I am grateful for all that I have learnt from you. To Professor Bavesh Kana, your advice and guidance was deeply appreciated in all aspects of this project. Thank you for the privilege of being able to work in your lab. To Dr. Christopher Ealand and Dr. Dale Liebenberg, thank you for your continued help throughout my project and especially with regards to the RT- qPCR work.

To all of my friends and colleagues at the CBTBR, thank you for the endless support, laughter and help in the lab. I will never forget the late nights and early morning shared with you by my side. To Nono, you have been by my side since day 1 and our time together will always have a place in my heart.

Lastly and most importantly, I would like to thank my mom and brother. Without your love, patience, and presence, I would not be who I am today. Thank you for all that you do for me. I love you dearly.

Table of Contents

List of Figures

List of Tables

Nomenclature

·OH	Hydroxyl radicals
8-OxoG	8-oxo-7,8-dihydroguanine
A	Adenine
AIDS	Acquired Immune Deficiency Syndrome
Amp	Ampicillin
AMIK	Amikacin
AP	Apurinic/Apyrimidinic
ARV	Anti-retroviral
BCG	Bacille Calmette - Guerin
BER	Base Excision Repair
bp/s	Base pair/s
BSA	Bovine Serum Albumin
C	Cytosine
CAP	Capreomycin
CFU	Colony Forming Unit
CTAB	Cetyltrimethylammonium bromide
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleotide triphosphates
DSBs	Double strand breaks
EDTA	Ethylenediaminetetra acetic acid
EMB	Ethambutol
EPTB	Extra-pulmonary TB
Fapy	Formamidopyrimidine
Fpg	Formamido-pyrimidine-DNA-glycosylase
G	Guanine
g	Gravitational force
G+C	Guanine plus Cytosine
H_2O_2	Hydrogen peroxide
HIV	Human Immunodeficiency Virus

HR	Homologous recombination
hr/s	hour/s
IFN	Interferon
INH	Isoniazid
KAN	Kanamycin
Kb	Kilo base pairs
LA	Luria-Bertani agar
LAMP	Loop- mediated isothermal amplification
lacZ	Gene encoding β-galactosidase
LB	Luria-Bertani broth
LTBI	Latent tuberculosis infection
M	Molar
MDR-TB	Multidrug-resistant tuberculosis
Min	Minutes
MutY	Adenine DNA glycosylase
m-value	Number of mutational events
NaCl	Sodium chloride
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NaOH	Sodium hydroxide
Nei	Endonuclease VIII
NHEJ	Non- homologous end joining
Nth	Endonuclease III
O_2	Oxygen
OD_{600nm}	Optical density at 600 nanometer wavelength
Ori	Origin of replication
P	Phosphate
PCR	Polymerase Chain Reaction
PPD- S	Purified protein derivative S
PZA	Pyrazinamide
RIF	Rifampicin
RNA	Ribonucleic acid
RNase	Ribonuclease

RNI	Reactive nitrogen intermediates
RNS	Reactive nitrogen species
ROI	Reactive oxygen intermediates
ROS	Reactive oxygen species
Rpm	Revolutions per minute
RRDR	Rifampicin- resistance determining region
s	Seconds
sdH_2O	Sterile distilled water
SSBs	Single strand breaks
T	Thymine
TB	Tuberculosis
TE	Tris EDTA
TST	Tuberculin skin test
TU	Tuberculin units
Tween	Polyoxyethylenesorbitan mono-oleate
UV	Ultraviolet
v/v	Volume/volume
WHO	World Health Organization
XDR-TB	Extensively drug- resistant tuberculosis
X-gal	5-bromo-4-chloro-3-indolyl- β-galactosidase
α	Alpha
β	Beta
m	Mutation rate

1. Introduction

1.1 Tuberculosis

Tuberculosis (TB) disease caused by *Mycobacterium tuberculosis*, leads to more deaths than any other infectious disease worldwide (Kanabus, 2017). *M. tuberculosis* forms part of the *Mycobacterium tuberculosis* complex which comprises several genetically related species including *M. bovis, M. africanum, M. caprae, M. cannetti* and *M. microti* (Malone and Gordon, 2017). These mycobacteria differ phenotypically as well as in their methods of host infection (Malone and Gordon, 2017).

1.2 History

Robert Koch discovered the tubercle bacillus in 1882 with the first forms of antibiotic treatment being developed in the 1950s (Burke, 1993). The first evidence of TB in humans was discovered in skeletal remains located in Europe and Egypt and are dated at approximately 5000 years old (Sager *et al*., 1972). Evidence of pre-Columbian TB was discovered from the remains of a Peruvian mummy in South America from circa 700 AD (Allison *et al*., 1973). Various factors such as colonisation, industrialisation and urbanisation have all affected the spread and occurrence of TB over the years.

1.3 Symptoms

Signs and symptoms of active TB disease include a cough persisting for more than 3 weeks, night sweats, fatigue, weight loss, fever and chest pains (WHO, 2019). Another sign of an active TB infection is coughing up blood or sputum. These mycobacteria can spread to other parts of the body such as the spinal cord, causing extrapulmonary tuberculosis (EPTB). Therefore, symptoms can differ depending on the area of the body that is affected (WHO, 2019). In contrast, patients with a latent TB infection (LTBI) often do not display any symptoms as they have immune systems that are able to control the infection. However, these patients are still at risk of developing active TB and are more susceptible if they have an HIV infection, an immune-compromising illness, or have previously been treated for TB incorrectly (Wilk, 2007).

1.4 Diagnosis

Individuals who contract tuberculosis can either have an active or latent *M. tuberculosis* infection. The IFN-γ release assay (IGRA) and the Mantoux tuberculin skin test (TST) are used to detect *M. tuberculosis* infection but do not differentiate between a TB infection and active

TB disease. If the patient presents with positive IGRA and TST result, with no symptoms of TB and a normal chest radiograph, the individual is said to have a latent TB infection (LTBI) (Villegas *et al*., 2000). The TST involves two tuberculin units (TU) of RT-23 or five TU of purified protein derivative S (PPD-S) injected intradermally into the patient (WHO, 2019). Complications often arise with the PPD TST due to its low specificity as well as a low sensitivity in individuals with immunosuppression (WHO, 2019). The IGRA test can be approached via an enzyme- linked immunosorbent assay (ELISA) or an enzyme- linked immunosorbent spot assay (ELISPOT). Unlike the TST assay, the IGRAs are unaffected by the bacille Calmette-Guérin (BCG) vaccination status of an individual, making it useful for the detection of latent TB infection in vaccinated individuals (WHO, 2019).

Active TB cases are detected by growth of mycobacterial cultures, which is the gold standard for TB diagnosis. The culture method can be conducted using solid (Lowenstein Jensen media) or a liquid broth system (MGIT 320) (Agrawal *et al*., 2016). More recently, a nucleic acid amplification assay (GeneXpert) has been used to detect active TB. GeneXpert is a molecular test that presents with results in approximately 2 hours, which is comparatively faster than the culture technique that takes weeks before results are available (Agrawal *et al*., 2016).

Various technologies have been endorsed by the WHO for the molecular detection of TB. The Xpert MTB/ Rifampicin (RIF) and Xpert Ultra (Cepheid, USA) have been endorsed as the initial diagnostic tests for TB and RIF resistance. In instances where acid- fast bacilli smear positive sputum or MTB cultures are available, line probe assays can be used (Hain Lifescience, Germany; Nipro, Japan). Similarly, TB loop- mediated isothermal amplification (LAMP) can be used for the detection of TB (Eiken, Japan). The line probe assay uses polymerase chain reaction (PCR) and reverse hybridisation techniques to detect mutations that may be associated with drug resistance, whilst TB LAMP is a temperature- independent technique that allows for the amplification of DNA and its visualisation under ultraviolet (UV) light (Raizada *et al*., 2014).

1.5 Statistics and epidemiology

TB is extremely contagious and individuals with compromised immune systems are at greater risk of developing the disease (WHO, 2019). Thus, people living in developing countries are at a higher risk of contracting TB due to factors such as malnutrition and the high rates of Human Immuno-deficiency Virus (HIV). In 2018, 10 million people developed TB, with 8.6 % of these individuals also living with HIV (WHO, 2019). The World Health Organization

reported that 83 % of those individuals with both TB and HIV were located in Africa. Two thirds of the TB cases reported in 2018 were found to be in eight developing countries (India, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa), whilst only 6 % were found in the European countries and North America (Figure 1.1) (WHO, 2019).

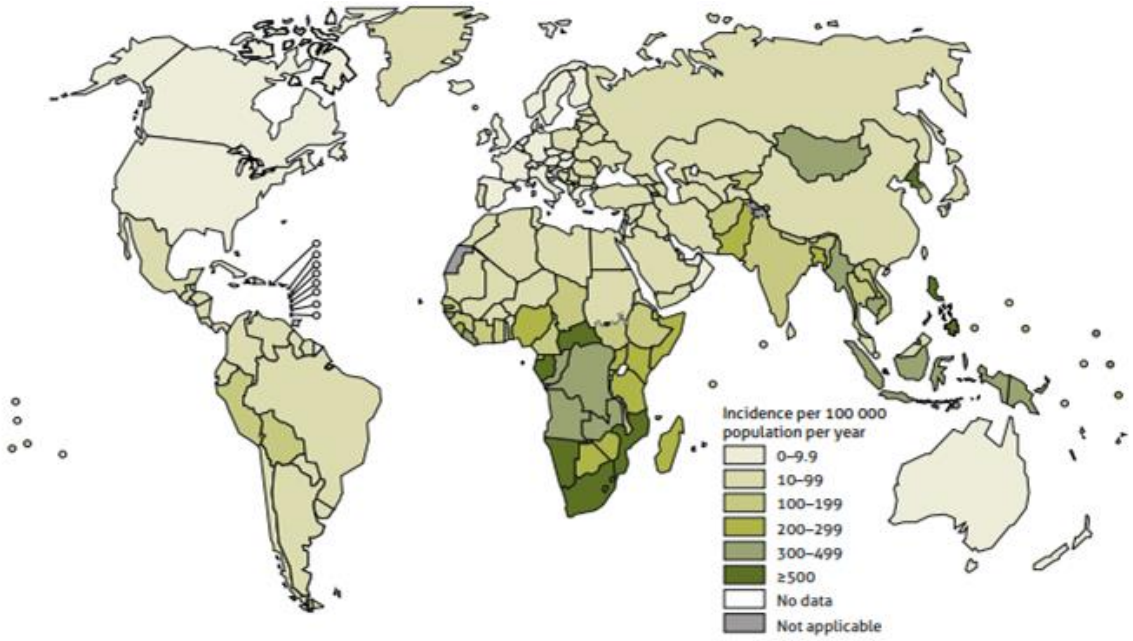


Figure 1.1: Estimated number of TB incidences in 2018; map taken from WHO, 2019. *In the most high- income countries, there were less than 10 incident cases of TB per 100 000 population, 150-400 incident cases in the 30 highest TB- burden countries and more than 500 in South Africa, the Central African Republic, the Philippines, the Democratic People's Republic of Korea, Namibia, Lesotho and Mozambique.*

Tuberculosis is the leading cause of death in people with HIV (WHO, 2019). The TB/HIV co-infection leads to difficulties in diagnostics and therapeutics, particularly in developing countries where budgets for healthcare are strained. Both HIV and *M. tuberculosis* can infect the alveolar macrophages of the host (Russell *et al*., 2010), where the mycobacteria cause increased replication of the HIV in the cells. Previous studies show that the mycobacteria cause an increase in the rates of infection and replication of HIV in monocyte-derived macrophages (MDMs), enhance the efficiency of HIV transfer from MDMs to T- cells and up-regulate CXCR4; a chemokine co-receptor that HIV uses to infect CD4+ T cells, thereby exacerbating replication of X4 HIV variants (Jemikalajah et al., 2014; Rantsi, 2017). A comprehensive understanding of the interaction between TB and HIV has not yet been achieved and further studies are necessary to establish preventive measures (Pawlowski *et al*., 2012).

1.6 Pathogenesis

Tuberculosis is spread easily and rapidly through the air as droplet nuclei that can remain airborne for many hours (Nardell, 2004). Multiplication of *M. tuberculosis* occurs rapidly in the lung due to the high concentration of molecular oxygen that the bacteria require for growth (Rastogi and David, 1988). In the lung, alveolar macrophages phagocytose the mycobacteria, which continue to invade the adjacent layer of epithelium. An inflammatory reaction by the host causes nearby blood vessels to provide mononuclear cells (Russell *et al*., 2010). These cells form the granuloma which is a defining characteristic of TB (Iliopoulos *et al*., 2006) (Figure 1.2). In the beginning stages, the granuloma is shapeless and consists of neutrophils, monocytes and macrophages with the macrophages differentiating into various specialised cells (Russell *et al*., 2010). As a complex immune response develops, lymphocytes gather, and the granuloma becomes organised with its centre consisting of a high density of macrophages (Russell *et al*., 2010).

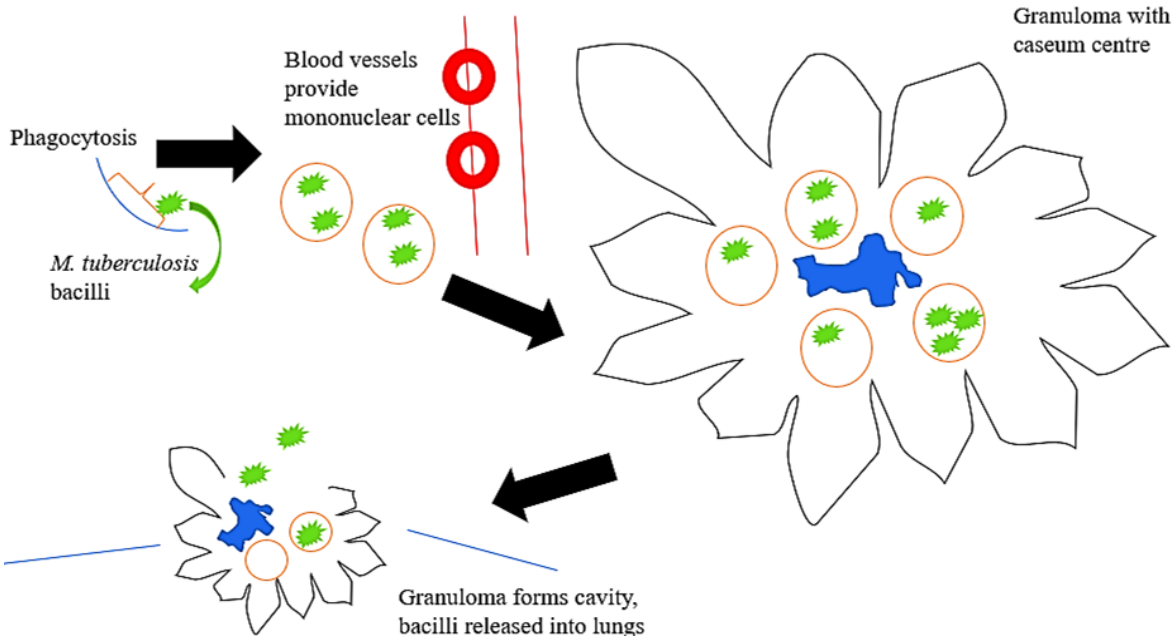


Figure 1.2: Life cycle of Mycobacterium tuberculosis derived from Russel* et al., *2010. *The* M. tuberculosis *bacilli are inhaled and undergo phagocytosis by alveolar macrophages. An inflammatory response occurs, and the cells invade adjacent epithelium. Blood vessels then provide mononuclear cells and the macrophages of the granuloma become specialised through differentiation. TB is caused when there is a decrease in vascularisation, a surge in necrosis and the build-up of caseum in the centre of the granuloma. When the granuloma forms a cavity and collapses in the lungs, the TB bacterium can be released into the airways.*

Approximately 2 to 3 weeks post-infection, lymphocytes specific to *M. tuberculosis* appear, indicating the end of the bacterial replication phase (Russell *et al*., 2010). The containment

state follows wherein the number of bacteria neither increases nor decreases. Granulomas with a thickened fibrous sheath and a decreased number of blood vessels are indicative of the progression of TB. There is a marked surge in the number of foamy macrophages, causing an increase in caseum in the centre (Russell *et al*., 2010), which can lead to hypoxia in the late-stage of infection. Finally, the granuloma ruptures and *M. tuberculosis* bacilli are released into adjacent airways (Figure 1.2). The risk of contracting an *M. tuberculosis* infection is dependent on various factors including the bacillary load inhaled, the strength of the host's immune system, as well as how infectious the source is (Zhang, 2005).

M. tuberculosis bacteria that infect the lungs of the host cause pulmonary tuberculosis (PTB). However, these bacteria can also spread to other parts of the body such as the kidneys and lymph nodes, causing extrapulmonary tuberculosis (EPTB). Between 75- 80 % of all TB cases reflect PTB rather than EPTB infection (Sharma and Mohan, 2004). In 2018, 85 % of the 7.0 million new and relapse TB cases were pulmonary TB, whilst only 15 % of the incident cases were EPTB (WHO, 2019). However, more than 50 % of individuals with an HIV/ TB co-infection present with EPTB (Sharma and Mohan, 2004). The most common sites of EPTB infection are the lymph nodes, followed by pleural effusion between the layers of the pleura of the lungs. Patients with EPTB have been shown to respond to the standard treatment regiments used for PTB disease (Sharma and Mohan, 2004).

1.7 Treatment and resistance

Currently, the bacille Calmette-Guérin (BCG) vaccine is the only licensed vaccine that can be used in the prevention of TB (WHO, 2019). The BCG vaccine is used as a preventative measure in children, but does not work to prevent the disease in adults. There are numerous treatment regimens available for TB with over 20 different drugs developed for the disease. First-line drugs for the treatment of susceptible TB include Rifampicin (RIF), Isoniazid (INH), Pyrazinamide (PZA) and Ethambutol (EMB). These drugs need to be taken as indicated for a minimum period of six months (Kanabus, 2017). This becomes challenging in cases where these drugs induce severe side effects and patients stop the therapy, thereby resulting in the emergence of drug resistance (Kanabus, 2017).

Multi-drug resistant TB (MDR-TB) is defined as resistance to both RIF and INH (Rendon *et al*., 2016). In 2018, 186 772 new cases of MDR-TB were reported, with the greatest number of cases occurring in China, India and Russia (WHO, 2019). Of these 186 772 cases, 6.2 % were identified as extensively drug resistant TB (XDR-TB). Extensively drug resistant TB confers resistance to

RIF and INH coupled with resistance to a fluoroquinolone and an injectable second-line drug such as Kanamycin (KAN), Amikacin (AMIK), or Capreomycin (CAP) (Dartois, 2014; Rendon *et al*., 2016). Recommended treatments for MDR-TB include a combination of rifamycin, ethambutol, and pyrazinamide, with or without a fluoroquinolone, for a period of between 6 and 12 months (Moadebi *et al*., 2007).

M. tuberculosis mutations, usually single nucleotide polymorphisms (SNPs), prevent injectable drugs from binding to the gene targets on the pathogen or change the drug's mechanism of action (Georghiou *et al*., 2012). These mutations have been largely associated with KAN, AMIK and CAP. The *rrs* and *tylA* genes of *M. tuberculosis,* encoding 16S rRNA and 2′-O-methyltransferase respectively, contain the most widely studied mutations that confer resistance to injectable drugs streptomycin and capreomycin respectively(Georghiou et al., 2012). The 2′-O-methyltransferase acts on nucleotide C1920 in helix 69 of 23S rRNA and nucleotide C1409 in helix 44 of 16S rRNA (Georghiou *et al.*, 2012). The most common mutations in the *rrs* gene that confer resistance to streptomycin are the A1401G, C1402T and G1484T mutations.

Similarly, mutations in the target genes to first- line drugs include the *rpoB* gene for rifampicin (RIF) as well as *katG, ahpC* and *inhA* promoter for isoniazid (INH) (Jnawali *et al*., 2013). The majority of the mutations of the *rpoB* gene reside in the 81- bp RIF-resistance determining region (RRDR) (Ramaswamy and Musser, 1998). The most commonly mutated codons in the RRDR region of the *rpoB* gene are 516, 526 and 531 (Zaw *et al*., 2018). The most common mutation in the *katG* gene is S315T, whilst that in the *ahpC* gene is C15T and in the *inhA* promoter the mutation is G46A (Dabboussi *et al*., 2015).

Recently, 'precision medicine' in the form of individualised treatment has been employed for the treatment of TB (Alffenaar *et al.*, 2020). Within this scope of treatment, therapeutic drug monitoring (TDM) is used to deliver the precise amount of anti- TB drug necessary for a single, defined patient (Alffenaar *et al*., 2020). Patients that benefit most from TDM are those that respond slowly to treatment, have a form of drug- resistant TB, pose a risk to drug- drug interactions or have another disease, such as HIV, that could significantly impact the TB disease (Peloquin, 2002). The increase in prevalence of TB, the number of HIV/TB co-infection cases as well as the surge in resistance constitutes a need for extensive research interventions for the control of this deadly epidemic.

There are approximately 23 drugs for the treatment of drug- susceptible TB, MDR- TB and latent TB in Phase I, II or III drug trials. There are 13 new compounds, of which seven belong to a new chemical class (macozinone, OPC-167832, Q203, BTZ-043, GSK-3036656, SPR720 and TBA-7371) (WHO, 2019). Another three drugs (delamanid, bedaquiline and pretomanid) have received regulatory approval. In conjunction, seven drugs (clofazimine, linezolid, levofloxacin, moxifloxacin, nitazoxanide, rifampicin [high dose] and rifapentine) that have been repurposed are undergoing continued testing (WHO, 2019).

1.8 Drug- drug interactions

Counteractive interactions between TB drugs as well as TB drugs with other drugs, used in the treatment of conditions such as HIV and diabetes, are currently a major problem (Bhutani *et al*., 2005). Thus, there are limitations in the number and types of drugs that can be used for patients with comorbidities as only certain drugs will be effective in the treatment of TB. In patients with both HIV and TB, rifabutin can replace rifampicin as first- line treatment. This is due to fewer drug- drug reactions that previously occurred when rifampicin was used, resulting in fewer numbers of adverse reactions (Chien *et al*., 2014). Rifabutin is a TB drug that induces the cytochrome P-450 enzyme pathway and P- glycoproteins (Gonzalez-Montaner *et al.*, 1994). Rifabutin can be co- administered with anti-retroviral (ARV) treatments and therefore, is a good option for those patients suffering from HIV/ TB co- infections (Khachi *et al*., 2009).

1.9 TB and mortality

The mortality rate of patients who do not receive treatment for active TB is extremely high (WHO, 2019). Previous studies, conducted prior to the availability of TB drug treatments, indicated that approximately 70 % of patients with sputum smear- positive pulmonary TB died within a decade of diagnosis (Tiemersma *et al*., 2011). Similarly, 20 % of patients with culture positive and smear- negative pulmonary TB also died within 10 years of diagnosis (Tiemersma *et al*., 2011).

In 2018, there were approximately 1.2 million deaths from TB in people who tested negative for HIV and an estimated 251 000 deaths from TB in HIV- positive people (WHO, 2019). The majority of these deaths can be prevented by correct diagnosis and following the recommended treatment regimen. There is a large variation between high- income countries such as Canada and Australia (ranging from less than one TB death per 100 000 population) and that of the WHO African Region as well as two high TB burden countries in Asia (40 or more deaths per 100 000 population) (WHO, 2019).

1.10 DNA damage

The ability of the mycobacteria to acquire mutations that confer resistance to first and second-line drugs, indicates the robustness of the bacteria for survival. Similarly, mycobacteria employ mechanisms to safeguard its DNA in the DNA damage- inducing environment of the host's macrophage (Kurthkoti and Varshney, 2011). DNA damage occurs in all cellular organisms due to various exogenous and endogenous stressors. If the altered DNA is not repaired, there may be an increase in DNA mutation rate and a decrease in cell viability (Kurthkoti and Varshney, 2011). DNA damage can be caused by multiple types of radiation, ultraviolet (UV) rays, free oxygen radicals and chemicals found in the environment (Cerutti, 1974). Ionising radiation and chemical damage of DNA can lead to single- strand breaks (SSB) or double-strand breaks (DSB), whilst mismatches of bases can arise due to a failure of DNA proofreading during replication (Ward *et al*., 1987). Covalent modifications of the four DNA bases: adenine (A), thymine (T), guanine (G) and cytosine(C), are known contributors of DNA damage (Friedberg *et al*., 2005).

1.10.1 Oxidative damage

Aerobic bacteria use molecular oxygen (O_2) to respire and oxidise nutrients for energy (Cabiscol *et al*., 2000). Reactive Oxygen Intermediates (ROI) as well as Reactive Nitrogen Intermediates (RNI) such as nitric oxide and its derivatives are toxic molecules found in the host's immune system that are responsible for the control of microbial pathogens (Bogdan *et al*., 2000). Macrophages produce ROI and RNI via NADPH oxidase (NOX2/gp91^{phox}) and inducible nitric oxide synthase (iNOS) (Ehrt and Schnappinger, 2009) in response to the *M. tuberculosis* infection. These oxidative stressors can lead to DNA damage and drive mutagenesis (Adams *et al*., 1997).

Oxidative DNA damage results in apurinic/ apyrimidic (AP) sites, strand breaks and damage to the rings of DNA bases (Cheng *et al*., 1992). Guanine, commonly targeted by oxidative damage, will convert to 8-oxo-7, 8- dihydroguanine (8-oxoG); a pro-mutagenic lesion that can incorrectly pair with adenine instead of cytosine if unrepaired, causing a C to A transversion. Similarly, oxidative damage may cause cytotoxic lesions that convert cytosine to 5-hydroxycytosine, causing cytosine to base pair with thymine instead of guanine (Purmal *et al*., 1994), causing a G to T transversion.

1.11 DNA repair pathways

DNA can be repaired through direct reversal, excision, or post- replication repair (Drapkin *et al*., 1994; Lindahl and Wood, 1999). DNA base excision is achieved via three repair pathways: base excision repair (BER), nucleotide excision repair (NER), and mismatch repair (MMR). Smaller lesions caused by oxidative stress, alkylation or deamination are resolved by the BER pathway, whilst larger lesions that often result in helix- distortions are repaired by the NER pathway (Lindahl and Wood, 1999). The mismatch repair pathway restores DNA that has mismatched or formed loops due to insertions or deletions during replication, whilst recombination repair helps restore double- stranded breaks (DSBs). In some bacteria, the SOS regulatory response directs the repair of DNA damage and mutagenesis (Dos Vultos *et al*., 2009).

The *M. tuberculosis* genome has a high GC content (67 %), thus guanine and cytosine are more susceptible to damage, resulting in pro-mutagenic and cytotoxic lesions (O'sullivan *et al*., 2005). However, *M. tuberculosis* has an innate repair response to DNA damage which is crucial for the maintenance of genome integrity. Although mycobacteria lack homologues of the DNA mismatch repair (MMR) mechanism (Mizrahi and Andersen, 1998), homologues of other repair pathways such as NER, BER, homologous recombination (HR) and non-homologous end joining (NHEJ) are present (Mizrahi and Andersen, 1998) and discussed briefly below.

1.11.1 Nucleotide excision repair

The nucleotide excision repair (NER) pathway removes bulky DNA lesions that have been caused by environmental mutagens and UV radiation (Friedberg and Gerlach, 1999). The NER pathway has been predominantly studied in *E. coli* and is mediated by helicase UvrD and the UvrABC exonuclease complex (Güthlein *et al*., 2009). The UvrABC complex is formed due to the low specificity of individual Uvr proteins. A DNA lesion is recognised by the UvrA and UvrB proteins which form the UvrAB complex (Kurthkoti and Varshney, 2011). UvrC displaces the UvrA protein and the UvrBC complex is formed. This complex cleaves and excises regions of damaged DNA, and is eventually displaced by the UvrD protein, forming a gap in the DNA (Kurthkoti and Varshney, 2011). DNA polymerase then fills the gap and DNA ligase seals the nicks.

1.11.2 Recombination repair

Homologous recombination (HR) repair occurs when a damaged DNA molecule attacks an undamaged DNA molecule of similar or identical sequence (Li and Heyer, 2008). Complex

DNA damages such as DSBs, DNA gaps ad inter- strand crosslinks are targeted by the HR pathway (Moore *et al*., 1986). This repair pathway is integral in the maintenance of the gene by supporting DNA replication and telomere maintenance (Slijepcevic, 2008).

In bacteria such as *E. coli*, the HR repair pathway requires the RecBCD enzyme; a complex helicase- nuclease that is regulated by a distinct DNA sequence known as Chi (Saikrishnan *et al*., 2012). The helicase- nuclease mechanism unwinds the double- stranded DNA which is then bound by the RecA protein (Anderson and Kowalczykowski, 1997). The RecA protein then signals the induction of the RuvABC complex in the bacterial SOS response pathway. Single-stranded DNA, coated in RecA, recognises the homologous DNA and attacks the duplex recipient DNA (Anderson and Kowalczykowski, 1997; Warner and Mizrahi, 2011). Following this, a protrusion is formed by a recipient strand and is cleaved by a RuvABC complex endonuclease, which then exchanges the strands, forming a Holliday junction. Finally, the RuvABC complex endonuclease ligates the swapped strands and seals the gap (Anderson and Kowalczykowski, 1997).

M. tuberculosis contains two HR repair pathways; RecFOR and RecBCD (Dos Vultos *et al*., 2009). Damage to *M. tuberculosis* DNA leads to increased expression of RuvA and RuvC; proteins involved in DNA repair (Brooks *et al*., 2001). The RecR and RecF proteins of the RecFOR pathway have displayed increased expression in response to CAP treatment and translational inhibition, indicating the role of RecFOR in *M. tuberculosis* (Dos Vultos *et al*., 2009).

1.11.3 Non- homologous end joining

The non- homologous end joining (NHEJ) repair pathway is similar to homologous recombination in that it repairs DSBs. However, unlike the HR pathway, NHEJ does not require a homologous template to repair the DSBs (Boulton and Jackson, 1998). Expression of a Ku-like protein and ATP- dependent Ligase D (Lig D) has been shown to occur in *M. tuberculosis* but not in *E. coli* (Malyarchuk *et al.*, 2007). The interaction between Lig D and Ku causes the activation of Lig D in the repair of DSBs (Weller *et al*., 2002). The efficiency of repair of DSBs by the NHEJ pathway is significantly lowered when genes encoding Lig D or Ku are deleted (Shuman and Glickman, 2007). The NHEJ repair pathway is of particular importance in individuals with LTBI, and may serve as the only DSB repair mechanism in the absence of a daughter chromatid for HR (Shuman and Glickman, 2007).

1.11.4 Base excision repair

The BER pathway repairs DNA that has been damaged by various endogenous factors including alkylations, deaminations and oxidations (Seeberg et al., 1995) resulting in damaged nucleotides such as 7-methyl-G, 8-oxo-G, 5-hydroxy-C and FapyA (Singh, 2017). The mechanism of action of the BER pathway involves DNA glycosylases; the first enzymes to recognise and excise damaged DNA bases, resulting in the formation of an apurinic/ apyrimidic (AP) site. The phosphodiester bond located at the 5' end of the AP site is cleaved by an AP lyase, forming 5´-sugar phosphate and 3´-OH ends. A deoxyribose-phosphodiesterase (dRPase) removes the 5'- sugar phosphate group, and the single nucleotide gap is filled by DNA polymerase and sealed by DNA ligase (Figure 1.3).

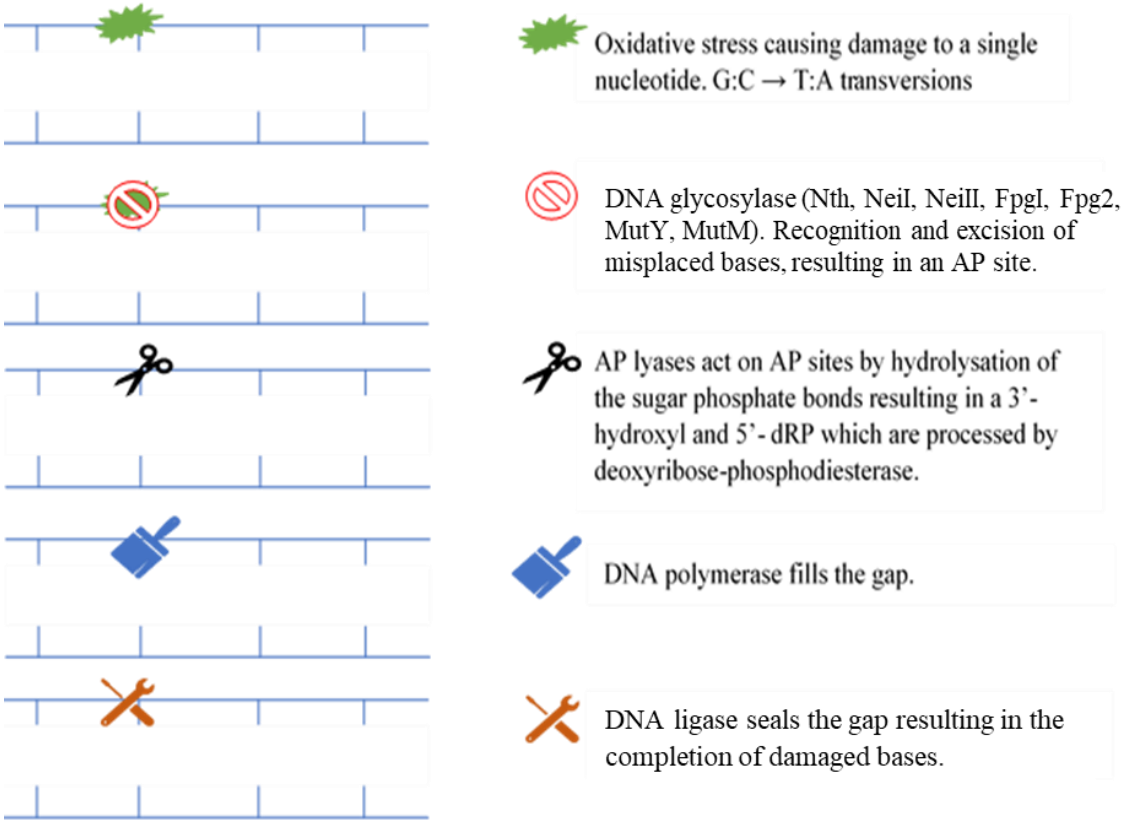


Figure 1.3: The mechanism of action of the BER pathway. *DNA glycosylases recognize and excise incorrect bases that arise due to damaged DNA. The AP lyases act on the AP sites and dRPase processes the dRP groups. DNA polymerase fills the gap and DNA ligase seals the nicks.*

Absence of the BER pathway in mycobacteria could lead to a decrease in the integrity and stability of the genome. A study conducted by Posnick and Samson in 1999 indicated that when the BER pathway of *E. coli* is imbalanced, there is an increase in spontaneous mutations and alkylation sensitivity (Posnick and Samson, 1999). In the presence of oxidative stress, pro-mutagenic and cytotoxic lesions can cause the DNA polymerase to stall resulting in the

incorporation of mutations into the bacterial genome by a non- proofreading polymerase, if not repaired (Friedberg and Gerlach, 1999). If the lesions are not corrected or incorporated into the genome, DNA replication and transcription will be stopped, resulting in DNA damage and ultimately cell death.

1.11.4.1 DNA glycosylases

The DNA glycosylases of the BER pathway are the first enzymes to recognise and excise oxidatively damaged DNA bases (Fortini *et al*., 1999). The DNA glycosylases of mycobacteria belong to two families based on structural and sequence homology: the Nth superfamily and the Fpg/ Nei family (David and Williams, 1998). Endonuclease III (Nth) belongs to the Nth superfamily which recognises and excises oxidised pyrimidines including FapyA and FapyG. Other glycosylases of the Nth superfamily include MutY which excises adenine that has mispaired with 8- oxo- G and FapyG, as well as the AlkA glycosylase that repairs alkylated bases (Leipold *et al*., 2000).

Guanine has a low redox potential, which makes it particularly sensitive to oxidative damage (Singh, 2017). Therefore, there is an increased risk of oxidative damage in mycobacterial genomes due to their high GC content. The GO repair system of the BER pathway specifically prevents the oxidation of guanine and facilitates the repair of oxidised guanine residues (Dos Vultos *et al*., 2009). The GO pathway involves a Fapy DNA glycosylase (Fpg or MutM) and an adenine glycosylase (MutY). The Fpg or MutM are involved in the removal of 8-oxo-G from the DNA, whilst the MutY is responsible for the removal of adenine that has incorrectly paired with 8-oxo-G. In addition, an 8-oxo-G triphosphatase (MutT) degrades 8-oxo-dGTP in the nucleotide pool thereby decreasing the likelihood of incorrect bases being incorporated into the DNA during replication (Singh, 2017).

Both *M. tuberculosis* and *M. smegmatis* contain two orthologues of endonuclease VIII/ Nei proteins. Endonuclease VIII belongs to the Fpg/ Nei family of glycosylases (David and Williams, 1998). The Fpg endonuclease has substrate specificity for purine lesions whilst the Nei endonuclease is similar to the Nth as it has substrate specificity for oxidised pyrimidines (Leipold *et al*., 2000). The redundancy of substrate specificity displayed by the Nth and Nei endonucleases ensures that the DNA repair function in the mycobacteria is not completely lost when one of the glycosylases is absent/non-functional/damaged for the repair of lesions.

1.11.4.2 Structural characteristics of the DNA glycosylases

The mycobacterial DNA glycosylase protein has three structural domains; a central helix-two-turn- helix (H2TH) domain, the catalytic core domain at the N- terminus and a zinc finger domain at the C- terminus (Fromme *et al.*, 2004) (Figure 1.4). The zinc finger domain of the DNA glycosylase is important in the excision of damaged DNA bases as it contains nucleic acid- binding subunits responsible for creating double- stranded breaks in damaged DNA (Sancar and Sancar, 1988) through the process of nucleophilic substitution (Duwat *et al.*, 1995).

Figure 1.4: Schematic depicting the structural domains of the mycobacterial DNA glycosylase protein adapted from Fromme* et al., *2004. *The blue region indicates the N- terminal catalytic core, followed by the H2TH domain in orange, whilst the zinc finger domain at the C- terminus is depicted in black.*

1.11.4.3 Structural characteristics of the NeiI glycosylase

The Nei DNA glycosylases consist of three domains; the N- terminal domain which is rich in β- sheets, as well as the helix- 2- turn- helix (H2TH) domain and the zinc- finger domain which are responsible for binding to damaged DNA. A three- dimensional protein structure of the *M. smegmatis* NeiI glycosylase was designed to include the various domains and motifs responsible for its mechanism of action (Figure 1.5). The NeiI glycosylase has an important "base flipping" mechanism that involves the flipping of a damaged nucleotide from the DNA (allowing the nucleophilic attack of the C1' residue of the damaged base) by the proline moiety located at the N- terminus of the glycosylase (Rantsi, 2017, unpublished). During this action, the intercalation loop of the NeiI glycosylase, consisting of the Gly- 69, Leu- 70 and Tyr-71 amino acid residues (displayed in magenta in Figure 1.5), inserts itself into the DNA helix at the site of the damaged base. This GLY- loop also creates a sharp bend in the DNA by unstacking the base opposite the lesion and forming hydrogen bonds with this base (Rantsi, 2017, unpublished).

The N- terminal domain contains β- sheets (turquoise and green sheets) which form a sandwich and are flanked by α- helices (dark- blue helices) (Figure 1.5). The C- terminal domain consists of α- helices (indicated in orange, yellow and lime- green) with two of these helices forming

the helix-two turn-helix (H2TH) motif (shown by the yellow and orange helices) as well as two anti- parallel β- strands that fold into the zinc finger motif necessary for protein function (shown as red sheets) (Figure 1.5). A Cys4- type zinc finger (displayed as yellow sticks on the yellow helix) is involved in a hydrogen- bonding network surrounding the lesion, holding the DNA in a distorted conformation. The amino acids Gln- 15 (displayed as dark- blue sticks) and Arg- 16 (displayed as grey sticks) are intricately involved in the positioning of the zinc finger (Figure 1.5) (Fromme and Verdine, 2002).

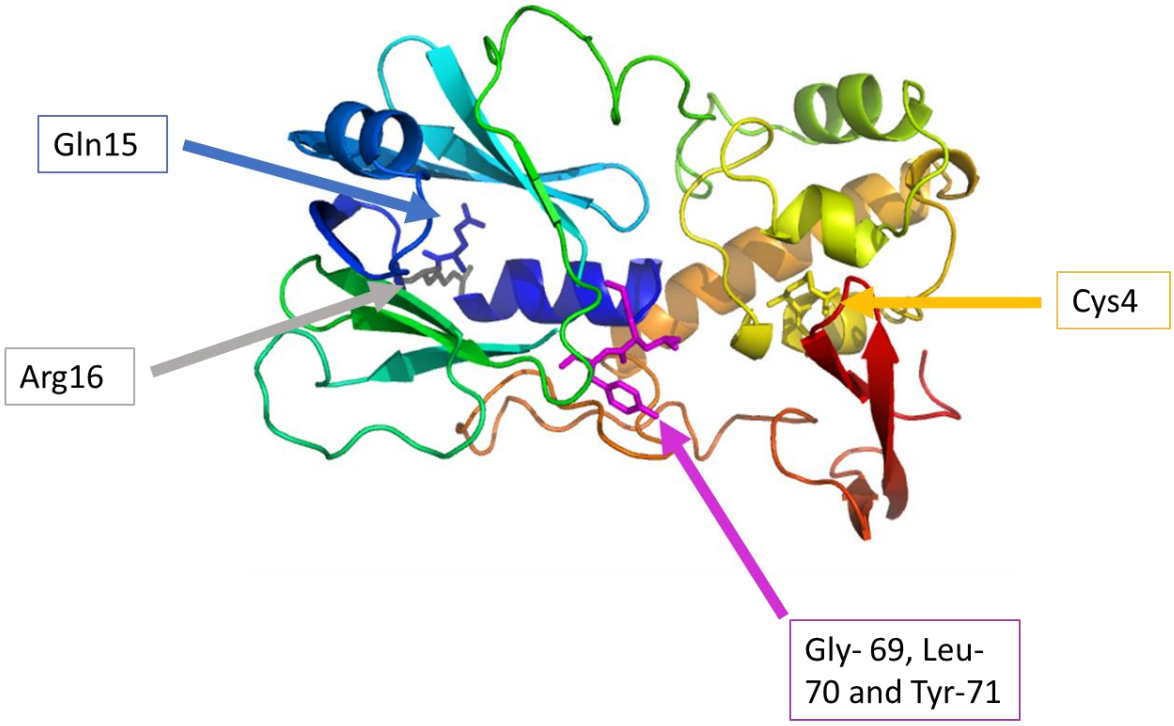


Figure 1.5: Three- dimensional protein model of the Nei1 glycosylase adapted from Rantsi, 2017, unpublished. *The structural characteristics of NeiI include an H2TH domain (yellow and orange helices), an N- terminal domain (turquoise and green sheets) as well as a zinc finger domain (red sheets). The Gln- 15 and Arg- 16 amino acids are displayed in blue and grey sticks respectively, whilst the GLY loop is represented by magenta sticks.*

1.11.4.4 AP endonucleases

Cleavage by the DNA glycosylases results in the formation of AP sites. This cleavage can result in the stalling of the replication fork during DNA replication. The abrasions created are mended by exonuclease III (XthA) and endonuclease IV (End or Nfo) proteins (Loeb, 1985). The End protein is largely responsible for the processing of AP sites in *M. tuberculosis*, and has a higher efficiency than XthA. Strikingly, *M. tuberculosis* mutants deficient in the *end* and

xthA genes have a heightened sensitivity to alkylation and oxidative stress *in vitro* (Puri *et al*., 2013). The AP endonucleases found in *E. coli* do not have a substrate preference, while those found in *M. tuberculosis* tend to favour the processing of AP sites that are formed against cytosine residues (Takeuchi *et al*., 1994). Hence, the affinity of two endonucleases in *M. tuberculosis* towards sites against cytosine, indicates an additional bias of the BER pathway towards the repair of 8-oxo-G, to attenuate the risks associated with a high GC content (Singh, 2017).

1.11.4.5 DNA polymerases

The gaps created by the AP endonucleases in the BER pathway are filled by the DNA polymerases. Although prokaryotes are known to have three major polymerases; PolI, PolII and PolIII, mycobacteria lack PolII (Cole *et al*., 1998). Instead, mycobacteria fill these gaps using PolA. The PolA polymerase is a DNA- dependent A- family polymerase I (Singh, 2017). In conjunction, mycobacteria also use the C- family PolII, which is encoded for by the *dnaE* gene as a *M. tuberculosis* PolA deficient mutant was restored by the *dnaE* gene (Rock *et al*., 2015). It has been found that *dnaE2* is upregulated *in vitro* by various DNA damaging reagents, suggesting that *dnaE2* is a primary mediator of survival through inducible mutagenesis (Boshoff *et al*., 2003). Further, specific mutations in the *rpoB* gene associated with *dnaE2* expression in rif^R *M. tuberculosis strains,* indicate a role of *dnaE2* in the adaptability of these strains. More specifically, the probability of adaption is increased when the expression of *dnaE2* is increased (Bergval *et al*., 2007).

1.11.4.6 DNA ligases

The nicks created by DNA glycosylases and filled by DNA polymerases, are sealed by DNA ligases. The ligase used by most bacteria is LigA; an NAD+ -dependent DNA ligase. In *M. tuberculosis*, LigA has been investigated as a prospective drug target (Korycka-Machala *et al*., 2007; Srivastava *et al.*, 2005). In contrast to *E. coli*, which only codes for a single ligase, *M. tuberculosis* has three other ATP- dependent ligases: LigB, LigC and LigD (Cole *et al*., 1998). Previous studies have shown that deleting the genes that encode LigB, LigC and LigD does not have a significant effect on the ligase action of *M. tuberculosis* when conducted under laboratory conditions (Gong *et al*., 2004).

1.12 Previous work and rationale

A bioinformatic analysis of the *M. tuberculosis* and *M. smegmatis* genome revealed that both organisms contain two copies of the *fpg* and *nei* genes that encode their respective

endonucleases (Moolla *et al*., 2014). Single mutant strains lacking the genes encoding the Fpg/ Nei endonucleases were generated followed by the generation of double and triple Fpg and Nei glycosylase knockout mutants (Moolla *et al*., 2014). These mutants were compared to the *M. smegmatis* mc^2155 parental strain and revealed no difference in hydrogen peroxide sensitivity, growth kinetics or increase in the rates of spontaneous mutations (Moolla *et al*., 2014) However, deletion of the Nth DNA glycosylase suggested a role for the Nth endonuclease in mycobacterial genome maintenance. Interestingly, a novel interplay between the Nth and Nei homologues but not with Fpg was observed as survival defects were exacerbated due to the combinatorial loss of the *nth* and *nei* genes under conditions of oxidative stress (Moolla *et al*., 2014).

The hierarchical role of the Nei homologues (NeiI and NeiII) was established in 2017 when it was observed that the combinatorial loss of the *nth* and *neiI* genes had a larger impact on *M. smegmatis* survival under oxidative stress compared to when the *nth* and *neiII* genes were simultaneously lost (Rantsi, 2017). The region of variability between the NeiI and NeiII homologues was thereafter identified to be located in their DNA intercalation loops (Padarath, 2017, unpublished). It was found that the NeiI glycosylase contains the glycine, leucine, and tyrosine (GLY) amino acids in its intercalation loop whilst the NeiII glycosylase has an intercalation loop consisting of lysine, methionine and glutamic acid (KME) (Padarath, 2017, unpublished). Thus, it was important to investigate whether the three amino acid residues of the NeiI glycosylase intercalation loop are responsible for increased glycosylase activity compared to NeiII.

1.13 Aims and objectives

The main aim of this study was to show that the GLY loop of the NeiI endonuclease is responsible for the increased DNA glycosylase activity in NeiI and to confirm that the NeiI DNA glycosylase has a role in mycobacterial genome maintenance. These aims were achieved by completing the following objectives:

1. The *neiI* promoter region (NP) and *neiI* gene (NI), encoding NeiI with GLY residues were amplified by PCR using DNA extracted from the parental *M. smegmatis* strain (mc^2155).
2. An altered *neiI* gene (ΔGAF) encoding NeiI with GAF residues was PCR amplified from the pUC57 replicating vector containing the *neiI* gene with the altered residues which was synthesized at Genscript (gene@genscript.com).
3. The amplified NP, NI and ΔGAF fragments were cloned into the pBlueScript replicating vector to generate the pBlueScript + NP, pBlueScript + NI and pBlueScript + ΔGAF plasmids.
4. These plasmids were verified by restriction digest analysis and sequencing to ensure absence of mutations that may be introduced during the PCR amplification process.
5. The NP, NI, ΔGAF fragments once cloned in pBlueScript were excised and cloned into the pMV306H *M. smegmatis* integrating vector.
6. The resulting two integrating vectors (pMV + NP + NI and pMV + NP + ΔGAF) containing the *neiI* promotor together with the *neiI* gene encoding either NeiI with GLY residues or the GAF residues was electroporated into the double deletion mutant lacking both the *neiI* and *nth* genes (*ΔnIΔnth*).
7. The expression of the *neiI* gene in both the *ΔnIΔnth::nI and ΔnIΔnth::ΔGAF* strains was tested by RT- qPCR using RNA extracted from these strains.
8. The newly generated mutant strains together with the respective parental strains were assessed for DNA damage by assessing spontaneous mutation rates against RIF using the fluctuation assay.

2. Materials and methods

2.1 Strains and culturing conditions

All bacterial strains and cloning vectors used and created in this study are listed in Table 2.1 and Table 2.2, respectively.

2.1.1 Bacterial strains used in this study

All bacterial strains used in this study are listed in Table 2.1 and stocks of all bacterial strains were prepared in 66 % glycerol and stored at -80 °C. Components of the culture media are listed in Appendix A.

Table 2.1: Bacterial strains used in this study and their culturing conditions

Strain	Characteristics	Source	Culture conditions
DH5α	Mutant of wildtype *E. coli*, non-pathogenic, high transformation efficiency	(Taylor *et al*., 1993)	LA/LB at 30ºC
npα	pBlueScript vector with the *neiI* promoter (MSMEG_4683) in *E. coli* DH5α cells	This study	LA/LB at 30ºC
nIα	pBlueScript vector with a *neiI* gene (MSMEG_4683) in *E. coli* DH5α cells	This study	LA/LB at 30ºC
ΔGAFα	pBlueScript vector with a mutated *neiI* gene (MSMEG_4683) in *E. coli* DH5α cells	This study	LA/LB at 30ºC
mc^2155	Mutant of mc^2 6. Displays highly efficient plasmid transformation phenotype	(Snapper *et al*., 1990)	7H10/7H9 at 37ºC
ΔnIΔnth	Mutant of mc^2155 with deleted *neiI* (MSMEG_4683) and *nth* genes (MSMEG_6184)	(Rantsi, 2015, unpublished)	7H10/7H9 at 37ºC
ΔnIΔnth::*nI*	pMV306H vector with the *neiI* promoter and *neiI* gene (MSMEG_4683) integrated into *ΔnIΔnth*	This study	7H10/7H9 at 37ºC
ΔnIΔnth.::*ΔGAF*	pMV306H vector with the *neiI* promoter and an altered *neiI* gene (MSMEG_4683) where GLY has been changed to GAF residues, integrated into *ΔnIΔnth*	This study	7H10/7H9 at 37ºC

2.1.2 Cloning vectors

All plasmids used and generated in this study are listed in Table 2.2. All corresponding maps are shown in Appendix E.

Table 2.2: A list of vectors used and generated in this study

Vectors	Size (bp)	Characteristics	Source
pBlueScript	2964	*E. coli* (DH5α) replicating plasmid, with an ampicillin resistance *bla* gene and a *lacZ* gene	Invitrogen
pUC57 + GAF	3682	pUC57 vector with altered *neiI* gene (MSMEG_4683) where GLY residues have been changed to GAF residues	Genscript (gene@genscript.com)
pMV306H	4271	*M. smegmatis* integrating plasmid with an *attP* phage attachment site, and a hygromycin resistance gene	Invitrogen
pBlueScript + NP	3492	pBlueScript vector with *neiI* promoter region inserted (MSMEG_4683) at the EcoRI site in *lacZ* gene of the plasmid	This study
pBlueScript + NI	3791	pBlueScript vector with *neiI* gene (MSMEG_4683) inserted at the EcoRI site in the *lacZ* region of the plasmid	This study
pBlueScript + ΔGAF	3791	pBlueScript vector with the altered *neiI* gene (MSMEG_4683) inserted at the EcoRI site in the *lacZ* region of the plasmid	This study
pMV + NP + NI	5574	pMV306H vector containing the *neiI* promoter and *neiI* gene (MSMEG_4683) inserted between the NcoI and HindIII sites of the plasmid	This study
pMV + NP + ΔGAF	5574	pMV306H vector with the *neiI* promoter and a mutated *neiI* gene (MSMEG_4683) where GLY residues have been changed to GAF residues inserted between the BstBI and HindIII sites of the plasmid	This study

2.1.3 Bioinformatic analysis

The genome sequences of the *neiI* promoter and the *neiI* gene sequences of *M. smegmatis* (MSMEG_4683) and *M. tuberculosis* (Rv2464c) were obtained from Mycobrowser (https://mycobrowser.epfl.ch/). Various bioinformatic tools were used to analyse and compare these sequences. Protein sequence alignments of the Nei homologues were conducted using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and three- dimensional protein structures were generated using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) and compared using PyMOL software (Schrödinger, 2020).

2.2 DNA extractions

2.2.1 Chromosomal DNA extractions

2.2.1.1 Small scale chromosomal DNA extractions

Half a colony (± 10 mm diameter) of mc^2155 was re-suspended in 50 µL sterile distilled water (sdH_2O) and boiled for 5 minutes at 95 ºC (Mak and Ho, 1992). Thereafter, 50 µL of chloroform was added and the mixture was vortexed for 10 seconds at room temperature. The mixture was then centrifuged for 5 minutes at 12 470 x g for phase separation and the aqueous phase was transferred to a fresh 1.5 mL Eppendorf tube. This DNA was used as template for PCR reactions (See section 2.3.1).

2.2.1.2 Large scale chromosomal DNA extraction

Wildtype mc^2155 DNA was extracted using the cetyltrimethylammonium bromide extraction (CTAB) method (Devi *et al*., 2013). All solutions used for the CTAB method are listed in Appendix B. *mc^2155* cells grown on a 7H9 plate were harvested and re-suspended in 500 µL TE buffer. The cells were then incubated in a dry bath at 65 °C for 10 minutes followed by 20 seconds on ice. Thereafter, 50 µL of lysozyme (10 mg/mL) was added and the suspension was incubated for 1 hour at 37 °C. A 70 µL solution of 10 % SDS and 6 µL proteinase K (10 mg/mL) was added to the mixture, followed by a 1-hour incubation at 65 °C. Following this, 100 µL of a 5 M solution of sodium chloride and 100 µL pre-warmed CTAB/NaCl mix was added to the mixture and incubated for 10 minutes at 65 °C. An equal volume of chloroform: isoamyl alcohol (24:1 v/v) was added for the removal of proteins and the mixture was centrifuged at 16168 x g for 30 minutes. The aqueous phase was transferred to a fresh 1.5 mL Eppendorf tube and precipitated with 420 µL isopropanol. The solution was centrifuged at 16168 x g for 20

minutes to pellet the DNA, and was subsequently washed with 70 % ethanol, dried in a vacuum centrifuge (5301 Eppendorf) and re-suspended in 100 µL sdH_2O. The DNA was stored at – 20 ºC until use.

2.2.2 Plasmid DNA extractions from E. coli *(Sambrook* et al.*, 1989)*

2.2.2.1 Small scale plasmid DNA extractions

A single colony of *E. coli* containing the plasmid was re- suspended in 5 mL Luria-Bertani broth (LB) with the respective antibiotic (100 mg/ mL ampicillin or 50 mg/ mL hygromycin) and grown at 37 ºC overnight with shaking at 250 revolutions per minute (rpm). A 1 mL aliquot of the pre- culture was transferred to a 1.5 mL Eppendorf tube and the cells were harvested at 12 470 x g for 5 minutes. The supernatant was discarded, and the cell pellet re- suspended in 100 µL Solution I. Following this, 200 µL Solution II was added to the suspension, mixed by gentle inversion 5- 8 times and incubated for 5 minutes on ice. Subsequently, 150 µL Solution III was added and the Eppendorf tube was incubated on ice for 5 minutes. The suspension was then centrifuged at 12 470 × g for 5 minutes, followed by the transfer of the supernatant to a 1.5 mL Eppendorf tube and the addition of 1 µL RNase A (10 mg/mL). The mixture was incubated at 42 ºC for 15 minutes. For plasmid DNA precipitation, 600 µL isopropanol was added to the mixture followed by centrifugation at 12 470 × *g* for 30 minutes. The supernatant was discarded gently, and the pellet washed with 1 mL 70 % ethanol (EtOH) and dried in a vacuum centrifuge at 60 ºC. The plasmid DNA was re-suspended in 20 – 50 µL sdH20 and stored at -20 ºC until use. All plasmid DNA extraction solutions are listed in Appendix B.

2.2.2.2 Large scale plasmid DNA extractions

The plasmid DNA was extracted using the NucleoBond® Xtra Midi kit (Macherey- Nagel, Germany). A single *E. coli* colony containing the plasmid was re- suspended in 50 mL LB with the respective antibiotic and grown at 37 ºC overnight with shaking at 100 rpm. The cells were harvested by centrifugation at 4500 × g for 15 minutes at 4 °C in a Beckman J2-21 centrifuge (Beckman, USA). The supernatant was discarded, and the pellet was re- suspended in 8 mL re-suspension buffer. Following this, 8 mL of lysis buffer was added and the suspension was incubated for 5 minutes at room temperature. Next the suspension was neutralised with the addition of 8 mL of neutralisation buffer and the NucleoBond® Xtra filter column was equilibrated with 5 mL equilibration buffer. The suspension was centrifuged at 4500 × g for 5 minutes at 4 °C in a Beckman J2-21 centrifuge, and the transparent supernatant was loaded onto the equilibrated column. Once the suspension flowed through the column, 8 mL of wash

buffer was added to the NucleoBond® filter column after which the plasmid DNA was eluted with 5 mL elution buffer, into a 15 mL Flacon tube. The eluted DNA was mixed with 3.5 mL isopropanol prior to centrifugation at 15 000 x g for 30 minutes. The supernatant was gently discarded and the pellet washed with 3 mL 70 % EtOH. The DNA/ isopropanol mix was aliquoted into 1.5 mL Eppendorf tubes and dried in a vacuum centrifuge at 60 ºC. The plasmid DNA was re-suspended in 100 – 200 µL sdH_20 and stored at -20 ºC until use.

2.3 DNA manipulations

2.3.1. Polymerase chain reaction (PCR)

2.3.1.1 Primer designs

Primers were designed using Clone Manager 9 software (https://www.scied.com) (See Appendix C) and validated using the NCBI Primer-BLAST tool (https://www.ncbi.nlm.nih.gov). Primers were synthesized at and purchased from Inqaba Biotec, South Africa.

2.3.1.2 Amplification reactions

For PCR that used FastStart Taq DNA polymerase (Sigma- Aldrich, USA), reactions were set-up to a final volume of 50 µL. The following components were used in the reaction: 1 x GC rich solution, 1 x PCR buffer, 0.2 mM dNTPs, forward and reverse primers that were diluted to a final concentration of 10 µM each, DNA template of 10- 100 ng/ µL and 1 unit of enzyme. The final volume was made up with sterile distilled nuclease free water. The reactions were carried out under the following cycling conditions: a single denaturation cycle at 94 ºC for 4 minutes, 30- 35 cycles of denaturation lasting for 30 seconds each at 94 ºC, an annealing step at 55-65 ºC for 30 seconds, elongation at 72 ºC for between 30- 90 seconds and finally an elongation step at 72 ºC for 5- 7 minutes.

2.3.1.3 Synthesis of the mutated NeiI construct

The *neiI* gene, encoding the NeiI protein with a mutated intercalation loop where the GLY (Glycine, Leucine, Tyrosine) amino acids are replaced by GAF (Glycine, Alanine, Phenylalanine) (ΔGAF), was synthesized, sequenced and cloned into a pUC57 replicating vector at Genscript (gene@genscript.com). Leucine and Alanine are structurally similar whilst Tyrosine and Phenylalanine belong to the same amino acid group due to their polarities (See section 3.1). NcoI and HindIII sites were incorporated into the sequence to assist with cloning into the *M. smegmatis* integrating vector; pMV306H (see Appendix E).

2.3.1.4 Amplification of cloning fragments

The *neiI* promoter sequence (NP), the *neiI* gene encoding the NeiI glycosylase with GLY residues (NI) and the *neiI* gene encoding the NeiI glycosylase with GAF residues (ΔGAF) were amplified by PCR using chromosomal DNA isolated from mc^2155 (See section 2.2.1.1) and primers listed in Appendix C. Appropriate restriction sites, NcoI at the 5' end and BstBI at the 3' end for the *neiI* promoter region and BstBI at the 5' end and HindIII at the 3' end for the *neiI* gene and altered *neiI* gene were introduced into the primer sequence to facilitate cloning of these fragments into the *M. smegmatis* integrating vector; pMV306H (See Appendix C).

Q5 high-fidelity DNA polymerase (New England Biolabs [#M0491S 004120714071]) was used to amplify the NP, NI and ΔGAF fragments. The use of the Q5 high-fidelity polymerase ensured that no errors were introduced into the sequences during amplification. A 50 µL reaction was made using 5 µL of the forward and reverse primers respectively, 10 µL of 5 X high G-C rich Q5 enhancer, 10 µL of 5 X Q5 Reaction buffer, 8 µL of dNTP's (1.25 mM), 1-2 µL template DNA (20-100 ng) and 0.5 µL Q5 DNA polymerase in sterile nuclease-free water.

The sample was amplified using the following parameters: denaturation for a single cycle at 98 °C for 4 minutes, followed by 30-40 cycles of denaturation at 98°C for 30 seconds, annealing of primers to the template DNA at specified Tm (See Appendix C) for 30 seconds and elongation for 45 seconds at 72 °C, followed by a final elongation step for 5-7 minutes at 72 °C. The expected amplicons of 528 bp for the *neiI* promoter region and 827 bp for the *neiI* and altered *neiI* genes were confirmed on a 1 % agarose gel by electrophoresis prior to the commencement of cloning into the *E. coli* replicating vector; pBlueScript.

2.3.2 Restriction enzyme digestions

All restriction enzymes mentioned in this study were purchased from Roche Applied Science (Roche, Switzerland) or New England Biolabs (NEB, USA). The restriction digest reactions were all conducted as per the manufacturer's instructions. Briefly, 1 µg of DNA was digested with 1 unit of restriction enzyme in 1 x recommended buffer and made up to a final reaction volume of 20 µL sdH_2O. All reactions were incubated at 37°C for 1 hour and heat inactivated at 65°C for 10 minutes except in the case of reactions incorporating BstBI, where incubation was conducted at 65°C for one hour and no heat inactivation step was required.

2.3.3 Quantification of nucleic acids

All DNA quantifications mentioned in this study was conducted using the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA). The Nanodrop is a spectrophotometer that

conducts DNA quantification by measuring the sample's absorbance at a wavelength of 260 nm.

2.3.4 Visualisation and purification of DNA

2.3.4.1 Agarose gel electrophoresis

DNA was visualised using agarose gel electrophoresis. The molecular weight of the samples for analysis determined the percentage of agarose used in the gels. DNA of higher molecular weight was separated using 0.8- 1 % agarose gels made in 1 x TAE buffer (1 mM EDTA and 40 mM Tris-acetate). DNA of lower molecular weight was separated using 2 % agarose gels. Electrophoresis was conducted in electrophoresis tanks (Bio-Rad laboratories) in 1 x TAE buffer at 80- 90 V. For estimation of molecular size, DNA molecular weight markers were electrophoresed on the gels. The gels were visualised with the SYNGENE G-Box system.

2.3.4.2 Purification of DNA

DNA fragments from PCR amplifications or the products of restriction digests of plasmids were purified using the NucleoSpin PCR clean-up and gel extraction kit (Macherey-Nagel) as instructed by the manufacturer. Briefly, DNA was electrophoresed on agarose gel by electrophoresis and the fragment of interest was excised from the gel. The excised portion of gel was liquified by the addition of binding buffer and incubation at 50 °C for 5-10 minutes. For the clean- up of PCR fragments, 1 volume of sample was added to 2 volumes of binding buffer and vortexed. The PCR DNA fragment or liquified gel fragment was loaded onto a NucleoSpin column placed into a 2 mL collection tube. The sample was centrifuged at 12 470 x g for 30 seconds during which time the DNA was bound to the NucleoSpin column. The flow- through was discarded and 700 µL of wash buffer was added to the column and centrifuged at 12 470 x g for 30 seconds. The washing step was conducted twice and thereafter the flow- through was discarded and the column dried by centrifugation at 12 470 x g for 1 minute. The DNA was then eluted into a fresh 1.5 mL Eppendorf tube by the addition of 30 µL of elution buffer followed by centrifugation at 11 000 x g for 1 minute.

2.3.5 Modification of DNA ends

2.3.5.1 DNA dephosphorylation

The 5'- ends of linear plasmid DNA were dephosphorylated using antarctic alkaline phosphatase (NEB). Plasmid DNA (1 µg) was combined with 2 µL antarctic alkaline phosphatase buffer, 1 U antarctic alkaline phosphatase and made up to a final volume of 20 µL

with sdH_2O. The reaction mixture was incubated at 37 °C for 1 hour and heat inactivated at 65 °C for 20 minutes.

2.3.5.2 DNA phosphorylation

Products of PCR reactions were phosphorylated using polynucleotide kinase (NEB). PCR product (1 µg) was combined with 1 U polynucleotide kinase, 2 µL polynucleotide kinase buffer, 2 µL ATP (10 mM) and made to a final reaction volume of 20 µL using sdH_2O. The reaction mixture was incubated at 37 °C for 1 hour and heat- inactivated at 65 °C for 20 minutes.

2.3.6 DNA ligations

Ligation of DNA fragments was conducted using T4 DNA ligase (NEB). The reaction mixtures consisted of 50 ng plasmid DNA, a calculated volume of insert DNA, 2 µL of T4 DNA ligase reaction buffer, 1 U of T4 DNA ligase and sdH2O to a final reaction volume of 20 µL. The reaction mixture was incubated at room temperature for 20 minutes followed by heat-inactivation at 65 °C for 10 minutes. The volume of insert DNA required for a 1:1 (vector: insert) ligation reaction was calculated as follows:

$$Amount\ of\ insert\ DNA\ (ng) = \frac{50\ ng\ x\ size\ of\ insert\ (bp)}{size\ of\ vector\ (bp)}$$

2.3.7 DNA sequencing

All constructs generated by PCR for cloning in this study were sequenced for fidelity and to confirm that no mutations were introduced into the gene or region of interest. DNA sequencing was conducted by the Central Analytical Facility at the University of Stellenbosch. The sequence data was cleaned using FinchTV (www.geospiza.com/finchtv), and analysed using clone manager 9 software (Sci- Ed, 2020).

2.4 Bacterial transformations

2.4.1 Preparation and transformation of chemically competent E. coli *cells*

Chemically competent *E. coli* DH5α cells were prepared as described by Sambrook *et al*., 1989. All solutions used in the preparation and transformation of chemically competent *E. coli* cells are listed in Appendix D. A single DH5α colony was selected and inoculated into 5 mL LB and incubated at 37 °C overnight. Following this, 1 mL of the pre- culture was inoculated into 100 mL LB and incubated at 37 °C until the cells reached OD_{600nm} ~ 0.5- 06. The cells were incubated on ice for 15 minutes and then centrifuged at 3900 x g for 5 minutes at 4 °C. The

cell pellets were re- suspended in 20 mL TfbI and incubated on ice for 15 minutes. Following this, a second centrifugation step was conducted at 3900 x g for 5 minutes at 4 °C. The cell pellets were then re- suspended in 2 mL TfbII and 500 µL aliquots were dispensed into 1.5 mL Eppendorf tubes. These tubes were flash frozen in 100 % ethanol and stored at -80 °C until use.

The chemically competent cells were transformed by mixing 100 µL cells with the desired volume of plasmid DNA. The mixture was incubated on ice for 20 minutes and subsequently heat- shocked at 42 °C for 90 seconds. The cells were rescued by the addition of 750 µL 2xTY. Cells were plated on LA media supplemented with the appropriate antibiotics (see Table 2.2) and incubated at 37 °C overnight for the selection of positive transformants.

2.4.2 Preparation and transformation of electro- competent M. smegmatis *cells*

Electro- competent *M. smegmatis* cells were prepared for electroporations. Briefly, 100 mL 7H9 media was inoculated with 1 mL of a pre- culture and grown to an OD_{600nm} of 0.5- 0.8 at 37 °C. Cells were harvested by centrifugation at 4500 x g for 10 minutes at 4 °C. Following this, three washes were conducted in 30 mL ice- cold 10 % glycerol and centrifugation as above. The cells were finally re- suspended in 2 mL of ice-cold glycerol and stored on ice until use.

Electroporations were conducted using the BioRad genePulser X- Cell^{TM} system (BioRad). Briefly, 1 µg of plasmid DNA was mixed with 400 µL cells and transferred into a 0.2 cm GenePulser cuvette (Bio-Rad). The parameters of the electroporations were as follows: Voltage 2 500 V, Capacitance 25 µF, Resistance 1000 Ω and cuvette size of 0.2 cm. After pulsing, cells were rescued using 800 µL LB, transferred to a fresh Eppendorf tube and incubated at 37 °C for 16 hours. The cells were then plated onto 7H10 media supplemented with appropriate antibiotics for 3-7 days at 37 °C.

2.5 Gene expression analyses

2.5.1 RNA extraction

Mycobacterial cultures were grown to an OD_{600nm} ~ 0.6 and harvested by centrifugation at 4000 rpm for 10 minutes (Liedtke *et al*., 1994). Pellets were re- suspended in 1 mL Tri- Reagent (Invitrogen, USA). The cells were then lysed using Lysing Matrix B in a ribolyser (Thermo Fisher, USA) for three cycles at 6000 for 30 seconds, followed by incubation on ice for 5 minutes between pulses. Lysates were then centrifuged at 12 470 ×g for 10 seconds at 4 °C to

pellet the cell debris and silica beads. Subsequently, the supernatants were transferred to fresh Eppendorf tubes containing a 300 µL chloroform: isoamyl alcohol solution (24:1 v/v). These tubes were mixed vigorously for 15 seconds followed by incubation at room temperature for 2 minutes. The resulting suspension was centrifuged at 12 000 × g for 5 minutes at room temperature and the solution separated into 3 phases: red organic phase (containing protein) at the bottom, DNA at the interface and a clear upper aqueous layer of RNA. The aqueous layer containing the RNA was transferred to a fresh microcentrifuge tube and extracted with 100 µL of chloroform: isoamyl alcohol (24:1 v/v). This solution was centrifuged at 12 000 × g for 20 minutes at room temperature. The supernatant was discarded gently and the pellet washed with 1 mL of 70 % EtOH and dried in a Speed Vacuum concentrator at 60 ºC. The isolated RNA was re-suspended in 50 µL pre-heated RNAse-free water and incubated at 60°C for 10 minutes to help facilitate re-suspension.

2.5.2 RNA purification

The isolated RNA was mixed with 500 µL water-saturated 1-butanol (10 mL butanol + 10 mL water) (Sigma-Aldrich, USA) and vortexed briefly. Thereafter, the samples were centrifuged by pulsation for 10 seconds for separation of the phases. The upper, organic layer was discarded and the process of adding butanol, vortexing and centrifugation was repeated three times. The residual butanol was then removed by the addition of 500 µL water-saturated diethylether (Sigma-Aldrich, USA) (1:1 water: diethylether). Following this, the samples were vortexed briefly and centrifuged as above for the separation of phases. The upper, organic layer was discarded and the diethylether wash was repeated. The residual diethylether was removed by evaporation in a 37°C dry bath in a fume hood for 30 minutes. Approximately 50 µL RNA sample was retained and precipitated for subsequent contaminant removal.

2.5.3 Removal of contaminating gDNA

RNA samples were treated with DNase to remove any residual genomic DNA (gDNA). Briefly, 10 µL rDNase 1 buffer and 1 µL rDNase 1 (Qiagen, Germany) were added to the purified and cleaned RNA and incubated at 37°C for 30 minutes. Thereafter, an additional 1 µL rDNase 1 was added to the solution and the incubation step repeated. The DNase was removed using the RNAeasy kit (Qiagen, Germany) as per the manufacturer's instructions. The sample volume was made up to 100 µL with sdH_2O, followed by the addition of 350 µL buffer RLT and the solution mixed. Following this, 250 µL of 100 % EtOH was added and the solution was mixed once more. The entire mixture was transferred to a RNeasy Mini Spin Column with a 2.0 mL

collection tube. The tube was centrifuged at 8000 x g for 15 seconds at room temperature and the flow- through discarded. Subsequently, 500 µL buffer RPE was added to the column and centrifuged as above. The flow- through was discarded and 500 µL buffer RPE was added and centrifuged at 8000 x g for 2 minutes at room temperature. The RNeasy Mini Spin Column was then placed in a fresh 1.5 mL collection tube and 50 µL of sdH_2O was added. The tube with the column was then centrifuged at 8000 x g for 1 minute at room temperature. The column was discarded, and the RNA quantified using the Nanodrop ND-1000 spectrophotometer. The purity of RNA was gauged by the A260/ A280 ratio wherein a ratio of 1.9- 2.1 indicated pure RNA.

2.5.4 Synthesis of cDNA

During the synthesis of cDNA, the reverse primer of the gene of interest and a reference gene were used. Briefly, 23 µL of purified RNA was aliquoted into a PCR tube containing 2 µL reverse primer mix (2.5 µM each). In a separate Eppendorf tube, 10 µL of each reverse primer was added to a final concentration of 2.5 µM (genes of interest and reference gene). Thereafter, the volume of the mixture was made up to 400 µL with sdH_2O. The annealing step of the primers to the RNA was conducted in a thermal cycler using the following program: 94 °C for 1.5 minutes; 65 °C for 3 minutes; 57 °C for 3 minutes. Following this, the cDNA synthesis master mix (reverse transcription [RT]- positive and RT- negative control) was prepared using the Superscript IV kit, $MgCl_2$ (25 mM) and dNTPs (10 mM). The master mix was set up as described in table 2.3 wherein 8.5 reactions were set up.

Table 2.3 Components of the master mix used in cDNA synthesis

Component	1 Reaction	RT+ (8.5 reactions)	RT- (8.5 reactions)
$MgCl_2$ (25 mM)	4 µL	34 µL	34 µL
5x first- strand buffer	5 µL	42.5 µL	42.5 µL
Superscript IV (SSIV)	1 µL	8.5 µL	N/A
DTT (0.1 mM)	2 µL	17 µL	17 µL
dNTPs (10mM)	1 µL	8.5 µL	8.5 µL
sdH_2O	N/A	N/A	8.5 µL
Total	13 µL	110.5 µL	110.5 µL

Thereafter, 12.5 µL of the prepared primers were added to a sterile PCR tube and 12.5 µL of the RT+ master mix was added to each sample. Subsequently, 12.5 µL of the RT- master mix was added to the remaining DNAse- treated RNA. Each tube was pulsed in a microfuge to pool their contents. The PCR tubes were then incubated in a thermal cycler using the following program: 50 °C for 50 minutes followed by 80 °C for 5 minutes for the inactivation of Superscript IV (SSIV) Reverse Transcriptase.

2.5.5 Quantitative PCR using SYBR green

Primers were designed to amplify ~ 80- 150 bp DNA fragments that were located internally to the open reading frames of the *neiI* and altered *neiI* genes as well as for the *M. smegmatis sigA* housekeeping gene (Appendix D). The qPCR reactions were performed using the CFX96 Real-Time PCR detection system (Bio- Rad) in conjunction with SYBR green quantification. A linear standard curve based on ten- fold serial dilutions of the wild- type gDNA was set up using CFX Manager v3.0 for the absolute quantification of mRNA levels. Following this, data was extrapolated from the standard curves in order to determine the absolute amount of mRNA in the experimental reactions. The absolute transcript numbers were normalised to the number of *sigA* transcripts in the same sample. The normalised transcript levels of the mutant strains were compared to the normalised data of the wild- type control. All of the above analyses were performed in triplicate with the cycling conditions as follows: 98°C for 2 minutes followed by 40 cycles of 98°C for 5 seconds, 62°C for 5 seconds, 72°C for 5 seconds with SYBR Green quantification at the end of each cycle. Following this the melt curve analysis began starting from 50°C with a gradual increase of 0.5°C every 0.05 seconds to 95°C with SYBR Green quantification conducted continuously throughout this stage.

2.6 Phenotypic analysis

2.6.1 Growth kinetics under axenic conditions

To determine the growth kinetics of the wild- type, mutant and complemented strains; 1 mL freezer stocks of the various strains were inoculated in 20 mL 7H9 media to an OD_{600nm} ~ 0.01 and grown at 37 ºC with shaking at 100 rpm overnight. The culture was monitored to late log phase at 3-hour intervals over a period of 24 hours by OD_{600nm} measurements as well as colony forming units (CFU/mL). CFUs were assessed by spreading appropriate 10- fold serial dilutions (10^0-10^8) in duplicate on 7H10 media and growth scored after 3 days. The growth of the various strains was represented graphically as optical density over time, as well as colony forming units over time. The experiment was conducted in triplicate for accuracy.

2.6.2 Spontaneous mutation rates

Previously made freezer stocks of the wild- type, mutant and complemented strains were grown in 7H9 until late log phase was reached (~10^{8} cells/ mL). Thereafter, a dilution series (10^{8}-10^{0}) was carried out and spread in duplicate onto 7H10 plates to obtain the initial inoculum (N_0). Simultaneously, 1 mL of each culture was spread onto 7H10 plates containing 200 µg/mL of rifampicin (RIF) for the observation of any pre-existing mutants. The cultures were then diluted from 10^{8} cells/mL to 10^{2} cells/mL and 2 mL of each of these diluted cultures were aliquoted into 30 culture tubes and incubated at 37 ºC for 7 days with shaking at 100 rpm. Following this incubation period, 5 culture tubes were selected at random and serial dilution of 10^{8} cells/ mL to 10^{0} cells/ mL were spread in duplicate on 7H10 plates. These plates were incubated at 37 ºC for 3 days and scored for CFU/ mL to obtain the total population size (*Nt*). The remaining 25 culture tubes were individually plated on 7H10 supplemented with RIF (200 µg/mL) and incubated at 37 ºC for 7 days before scoring for RIF resistant (Rif^{R}) mutants (*m*) (Figure 2.1).

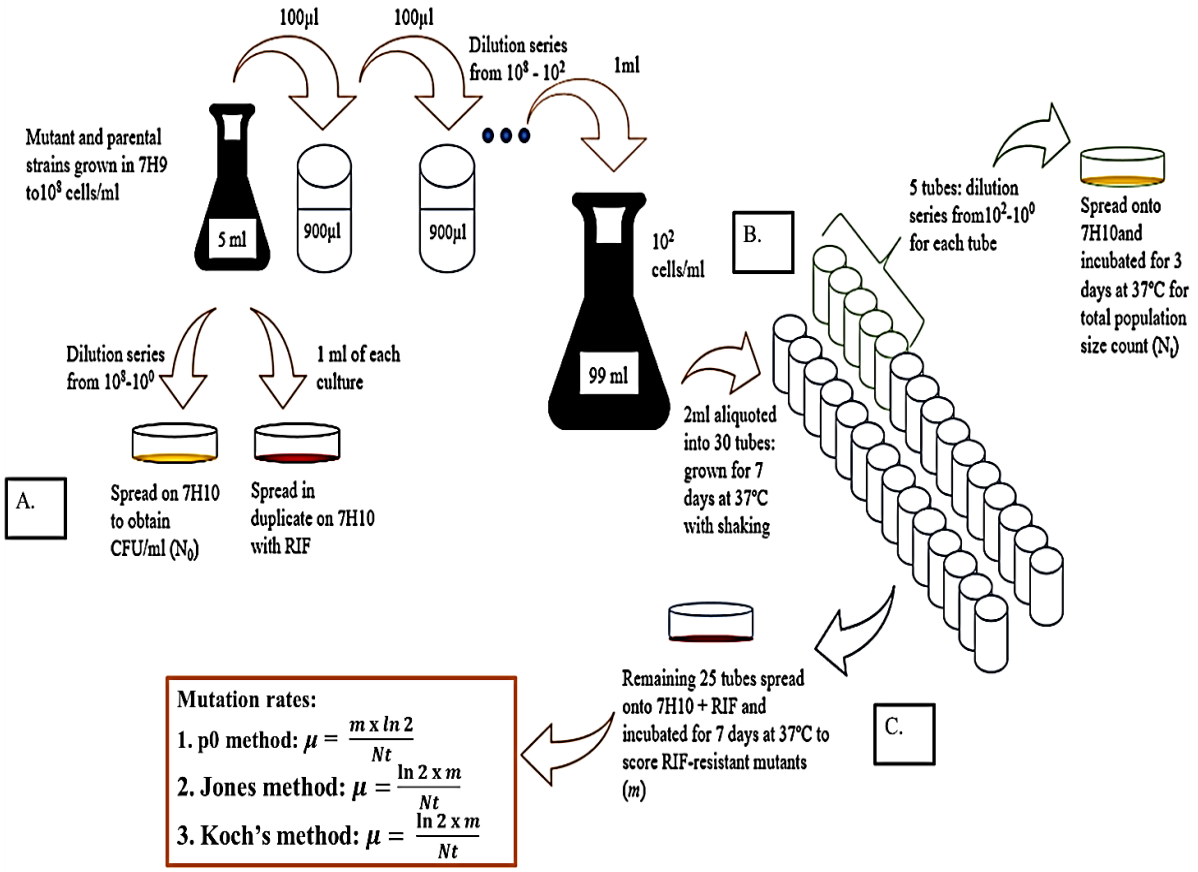


Figure 2.1: Schematic representation of the fluctuation assay. *A: Dilutions from a single starter culture (10^8-10^0) were spread on 7H10 plates to obtain the initial bacterial count (CFU/mL) and 1 mL of this starter culture was spread on 7H10 plates with RIF to assess for pre-existing mutants. B: The starter culture was diluted to 10^2 from which 30 x 2 mL parallel cultures were set-up and incubated for 7 days at 37ºC with shaking. Five of these cultures were diluted (10^8-10^0), spread on 7H10 plates which were incubated for 3 days at 37 ºC and CFUs were counted to obtain the population size (Nt). C: The remaining 25 cultures were spread on 7H10 plates with RIF, plates were incubated for 7 days at 37 ºC and scored for RIF-resistant colonies. The mutation rate was calculated using the p0 method, Jones method and Koch's method (Rosche and Foster, 2000).*

A mutation rate refers to the probability of a cell acquiring a mutation in its lifetime and accounts for the possibility of giving rise to mutations in both early and late phases of growth. The Luria-Delbrück fluctuation analysis was used to measure spontaneous mutation rates for the wild- type and mutant strains (Rosche and Foster, 2000). Parallel cultures were used to minimise variability and ensure accuracy when calculating the number of mutational events that occur.

Mutation rate is defined as the number of observed mutations per culture (*m*) in the total population (*Nt*) (Rosche and Foster, 2000). The three methods that were used to estimate the total mutation rate of a population were the Lea- Coulson method of the median (p0 method), the Jones Median Estimator and Koch's Quartiles method. The p0 method relies on the proportion of cultures which do not display any mutants, whilst the Jones Median estimator relies on the median of the observed number of mutants in a culture. Similarly, Koch's Quartiles method takes into account the median of the observed number of mutants in a culture, as well as the lower and upper quartiles in order to estimate the number of mutations per culture (Rosche and Foster, 2000).

The p0 method assumes a Poisson distribution which indicates that the mean and variance equal the product of the probability of a mutation and the number of bacteria (Rosche and Foster, 2000). This method uses an estimation of the number of mutational events that have occurred in order to determine the mutation rate with a Luria–Delbrück distribution. The number of observed mutations per culture (*m*) is influenced by how much growth has occurred as well as the mutation rate (μ) and natural logarithm (*ln*). By dividing the total number of cells (*Nt*) by the number of observed mutations per culture (*m*), the mutation rate is given. The mutation rate was calculated using the following formula:

$$\mu = \frac{m2}{Nt}$$

Using the Jones method (Jones, 1993), the mutation rate (μ) is calculated using the number of observed mutations per culture (*m*) which are dependent on the observed number of mutants in a culture (*r*). The number of observed mutations per culture (*m*) was calculated by the following formula:

$$m = \frac{r - ln\ ln\ 2}{ln\ ln\ r - ln\ ln\ (ln\ ln\ 2)}$$

Following this, the mutation rate is then calculated using the formula:

$$\mu = \frac{ln\ ln\ 2\ x\ m}{Nt}$$

Koch's quartile method considers the lower (Q1), median (Q2) and upper quartile (Q3) of the data in order to calculate the average number of observed mutations per culture (*m*). These results are then used to calculate the mutation rate (μ) using the following formulae:

$$m1 = 1.7335 + 0.4474\ Q1 - 0.002755\ (Q1)^2$$

$$m2 = 1.1580 + 0.2730\ Q2 - \ 0.000761(Q2)^2$$

$$m3 = 0.6658 + 0.1497\ Q3 - 0.0001387(Q3)^2$$

Where $Q1 = \frac{1}{4}\ (Nt + 1)$

$Q2 = \frac{1}{2}\ (Nt + 1)$

$Q3 = \frac{3}{4}\ (Nt + 1)$

The average of the *m1*, *m2* and *m3* values (*m*) are then used to calculate the mutation rate (μ):

$$\mu = \frac{ln\ ln\ 2\ x\ m}{Nt}$$

3. Results

3.1 Bioinformatic analysis

The NeiI glycosylase with GLY residues and NeiI glycosylase with GAF residues were compared using various bioinformatic tools. Bioinformatic analysis was conducted using three-dimensional protein structures that were generated using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) and compared using PyMOL software (Schrödinger, 2020). Similarly, protein sequence alignments were conducted using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). This analysis indicated the similarities and differences between the NeiI protein with a DNA intercalation region consisting of the glycine, leucine and tyrosine (GLY) amino acids and the altered NeiI protein with an intercalation loop containing the glycine, alanine and phenylalanine (GAF) amino acid residues. Although the altered NeiI protein contains an alanine in place of a leucine, and a phenylalanine in place of a tyrosine, these alterations did not cause structural changes in the NeiI protein and therefore, allowed for the investigation of the role of the GLY residues in glycosylase activity.

3.1.1 Structural similarities of the NeiI and altered NeiI glycosylase

In order to conserve the structural integrity of the NeiI protein; alanine and phenylalanine residues were chosen to replace leucine and tyrosine residues, respectively. The structural similarities of alanine and leucine allow for the correct folding of the altered NeiI protein containing the GAF residues in the intercalation loop (Figure 3.1A). Similarly, tyrosine and phenylalanine are both aromatic amino acids where tyrosine is derived from phenylalanine by hydroxylation at the para position (Figure 3.1 B). The conservative replacement of amino acids preserves the structural integrity and prevents the degradation of the NeiI protein.

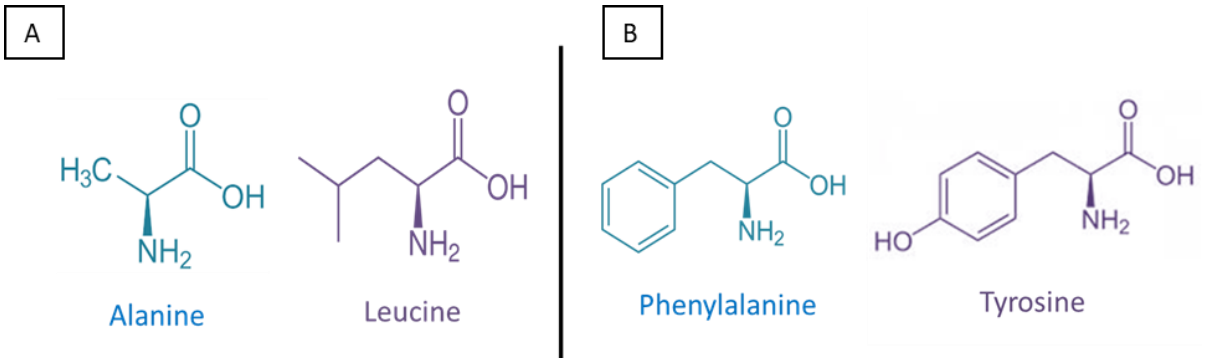


Figure 3.1: Structural similarities between amino acid residue substitutions of the NeiI proteins. *Alanine and leucine contain methyl groups that aid in their hydrophobicity (A), whilst the structurally similar aromatic rings of phenylalanine and tyrosine prevent misfolding of the NeiI protein (B).*

The NeiI protein is not structurally affected by the replacement of GLY amino acids with GAF amino acids. The template modelling score (TM-score) defines the similarity between two modelled proteins. The template modelling score (TM-score) of the NeiI protein with GAF residues = 0.88±0.07 when modelled against the NeiI protein with GLY residues. A TM= 1 would indicate a precise match between two proteins, whilst a TM< 0.2 would indicate random unrelated proteins. Thus, the NeiI protein is not hindered in its capacity to fold as the GLY and GAF amino acids are structurally and chemical similar (Figure 3.2).

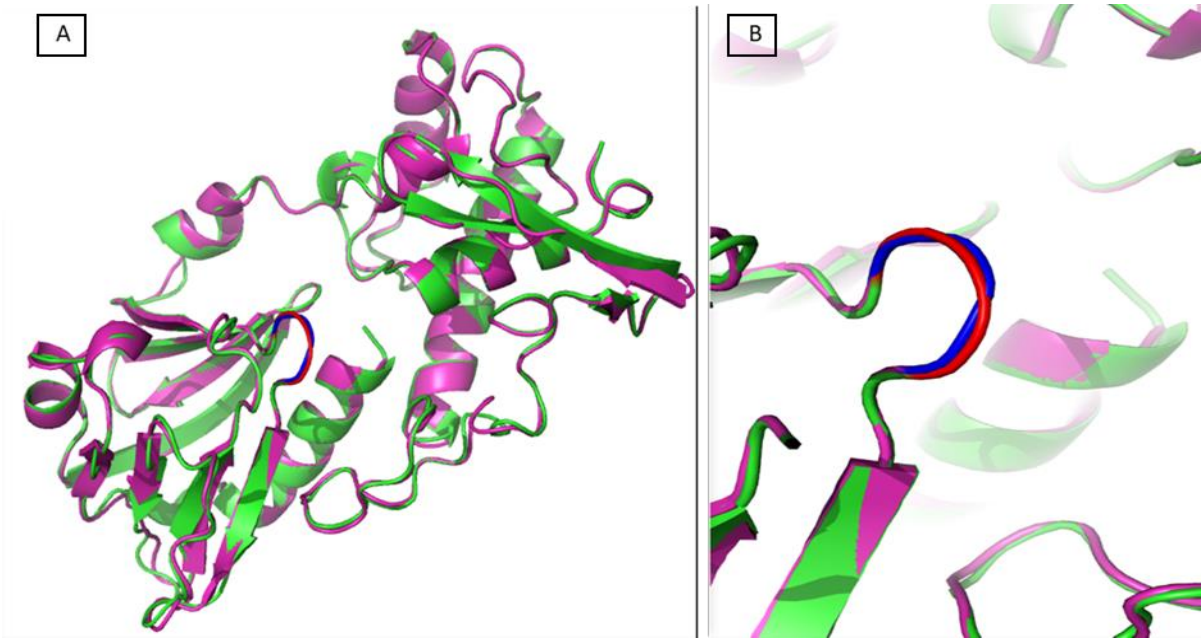


Figure 3.2: A structural comparison of the native and altered NeiI proteins*. A: The NeiI protein with the GLY amino acids (green) is structurally similar to the NeiI protein with the GAF amino acids (purple). Substituting the GLY amino acids with the GAF amino acids does not change the folding capabilities of the NeiI protein. The template modelling score (TM-score) of the NeiI protein with GAF residues = 0.88±0.07 when modelled against the NeiI protein with GLY residues, where a TM = 1 indicates a perfect match between two proteins and a TM < 0.2 indicates random, unrelated proteins. B: The intercalation loops of NeiI with GLY amino acids (red) and NeiI with GAF amino acids (blue) do not differ significantly in structure and have an estimated TM of 0.89 ±0.07.*

In conjunction, a protein sequence alignment of the NeiI endonuclease with GLY residues and that with GAF residues displayed the conservation of amino acid residues despite the alteration of the DNA intercalation loop (Figure 3.3).

Figure 3.3: Alignment of amino acid residues of the NeiI endonuclease with GLY and GAF residues. *The pink box indicates the altered amino acid residues in the intercalation loops of the NeiI protein. The amino acid residues found upstream and downstream of the GLY and GAF residues are conserved.*

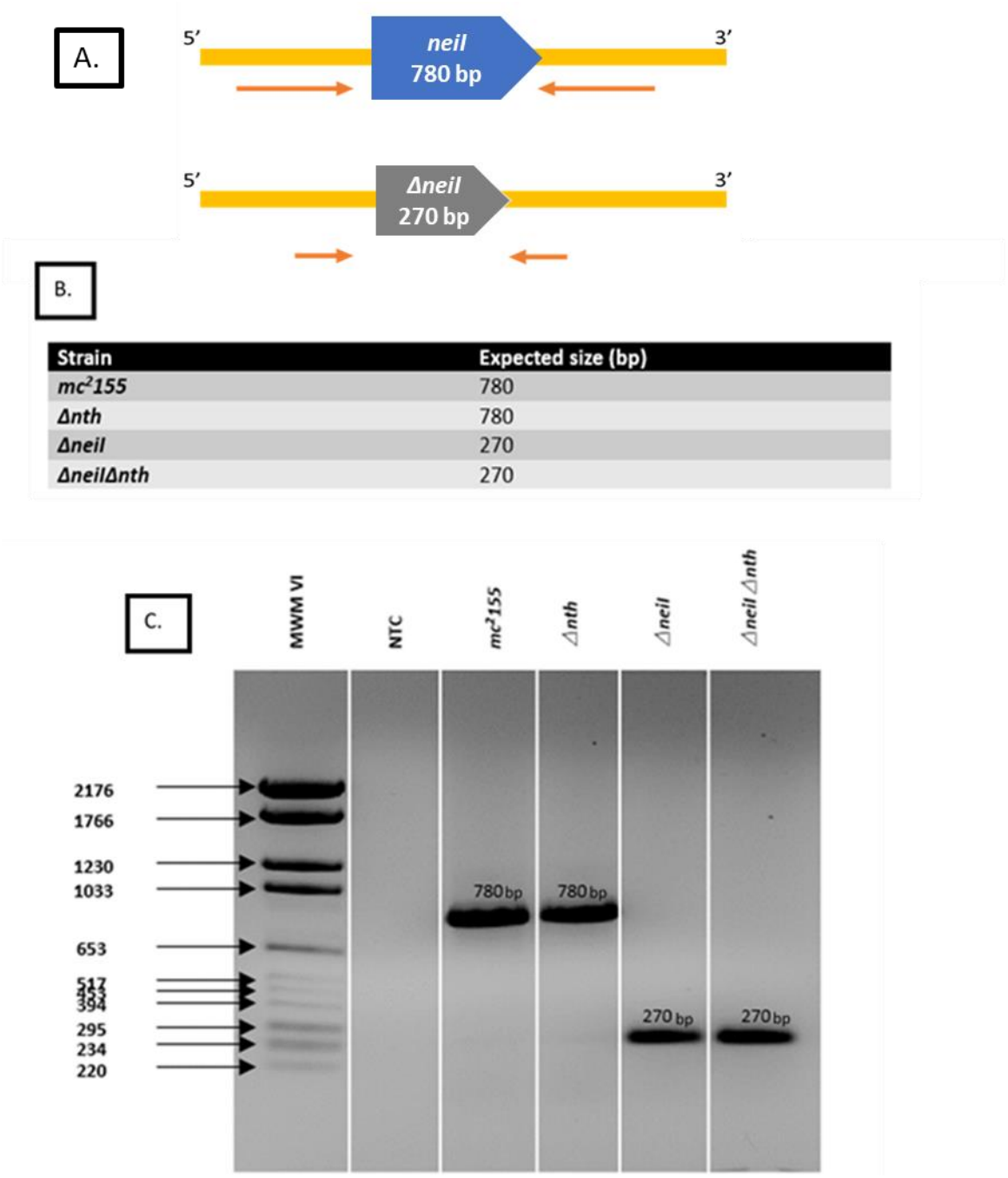
The ability of the changed NeiI protein to fold correctly based on its sustained structure allowed for further characterization of the DNA intercalation loop of the NeiI protein. The change in amino acid residues aided in determining whether the DNA intercalation loop of the NeiI protein is responsible for the increased DNA glycosylase activity in NeiI and its role in maintaining the mycobacterial genome. The bioinformatic analysis directed the investigation of the DNA intercalation loop which involved PCR amplification of the *neiI* promoter region (NP) and *neiI* gene (NI), encoding NeiI with GLY residues as well as the altered *neiI* gene (ΔGAF) encoding NeiI with GAF residues and subsequent cloning and integration into the double deletion mutant lacking both the *neiI* and *nth* genes (*ΔnIΔnth*). Phenotypic assessment of the parental strains and newly generated mutant strains for DNA damage by assessing spontaneous mutation rates would confirm if the GLY loop is indeed responsible for increased glycosylase activity in NeiI.

3.2 Confirmation of pre- existing mutants

M. smegmatis single and double mutants deficient in the *neiI* and *nth* genes (*ΔneiI, Δnth, ΔneiIΔnth*) previously constructed in our laboratory (Goosens, 2009; Moolla *et al*., 2014; Rantsi, 2017, unpublished) were confirmed using PCR (see section 2.3.1.4), together with the wildtype strain mc^2155, to ensure that the mutants were still correct in terms of their DNA repair gene complement. Specific primer sets for the *neiI* and *nth* genes were used to amplify gDNA extracted from the abovementioned strains (see section 2.2.1.2 and Appendix C) generated the expected band sizes (Figure 3.4 and Figure 3.5).

3.2.1 Confirmation of strains using neiI *primers*

The *neiI* primer set (see Appendix C) was used to amplify gDNA extracted from *ΔneiI, Δnth, ΔneiIΔnth* and mc^2155 as a positive control. A no- template negative control was included in all PCR amplifications. With the *neiI* primer set, strains that contain an intact *neiI* (wildtype and *Δnth*) gene amplified a 780 bp fragment, whilst strains that lacked a *neiI* (*ΔneiI and ΔneiIΔnth*) gene displayed the mutant band of 270 bp, confirming that the mutant strains were correct (Figure 3.4).



Strain	Expected size (bp)
mc²155	780
Δnth	780
ΔneiI	270
ΔneiIΔnth	270

Figure 3.4: Confirmation of strains using <u>neiI</u> primers. *A: Schematic representation of the* neiI *screening primers (orange arrows) and the expected fragment sizes they produce B: Table showing the expected band sizes from the* neiI *primer set with DNA from the wild type and mutant strains. C: Gel electrophoresis image displaying the various mutant and wildtype <u>M. smegmatis</u> strains. An expected band size of 780 bp was observed for <u>Δnth</u> as this strain still retains the <u>neiI</u> gene. In contrast, the <u>ΔneiI</u> and <u>ΔneiIΔnth</u> strains lack the* neiI *gene and therefore display the mutant bands of 270 bp. MWM VI indicates molecular weight marker VI and NTC depicts the no- template control.*

3.2.2 Confirmation of strains using nth *primers*

The *nth* primer set (see Appendix C) was used in the amplification of gDNA extracted from *ΔneiI, Δnth, ΔneiIΔnth* against the positive control: mc^{2}155 and a no- template negative control. When the *nth* primer set is used, those strains that contain a *nth* gene display the same band size as that of the positive control (801 bp), and those that lack an *nth* gene are represented by a mutant band (250 bp) (Figure 3.5).

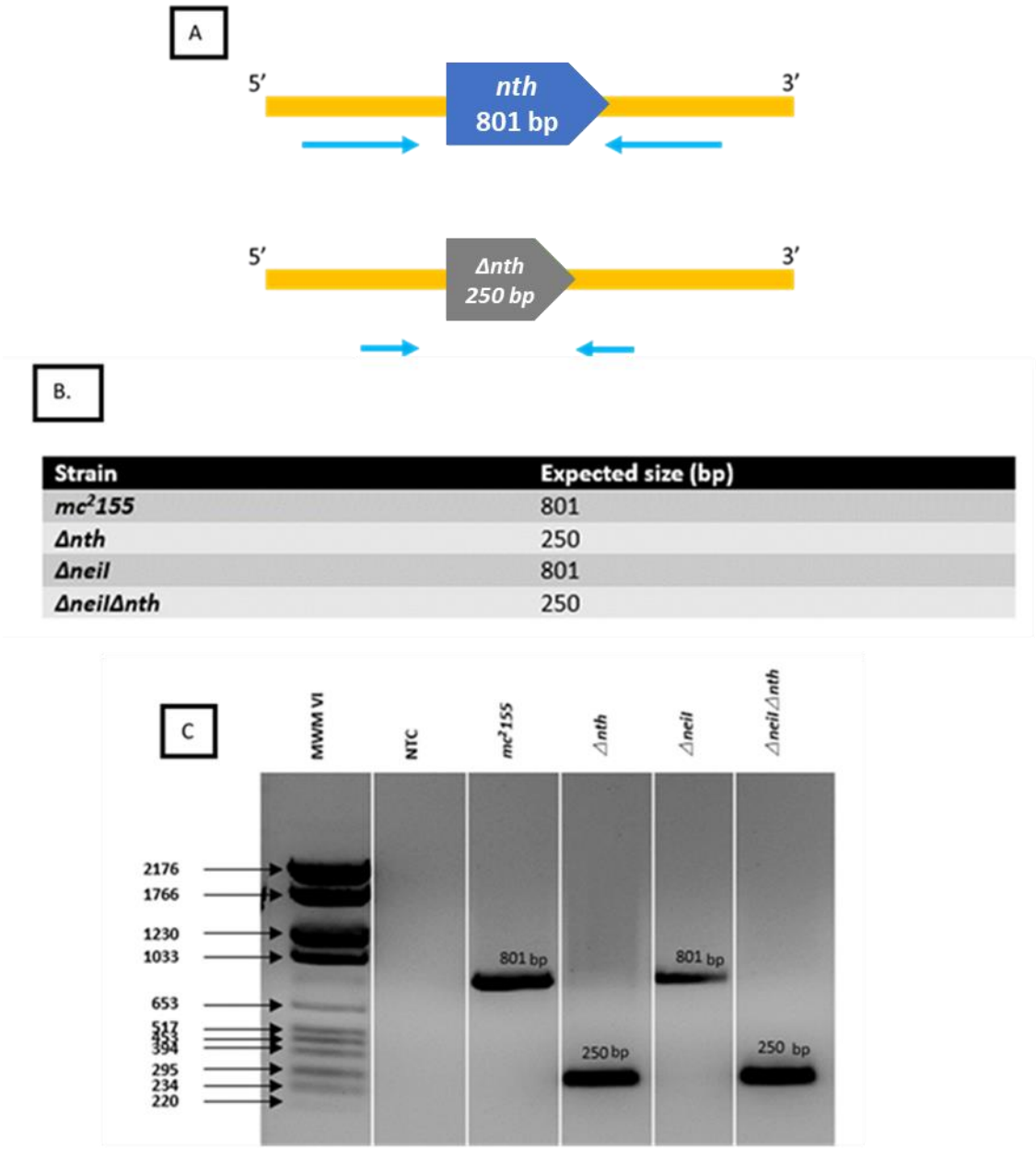


Figure 3.5: Confirmation of strains using nth primers. *A: Schematic representation of the* nth *screening primers (blue arrows) and the expected fragment sizes they produce. B: Table showing the expected band sizes with the* nth *primer set with DNA from the wild type and mutant strains. C: Gel electrophoresis image displaying the various mutant and wildtype M. smegmatis strains. An expected band size of 801 bp was observed for ΔneiI as this strain still retains the nth gene. In contrast, the Δnth and ΔneiIΔnth strains lack the neiI gene and therefore display mutant bands of 250 bp. MWM VI indicates molecular weight marker VI and NTC depicts the no- template control.*

Therefore, all previously constructed mutants were still correct in terms of their DNA repair gene complement hence, the cloning of the intact and altered *neiI* gene with the native promoter proceeded.

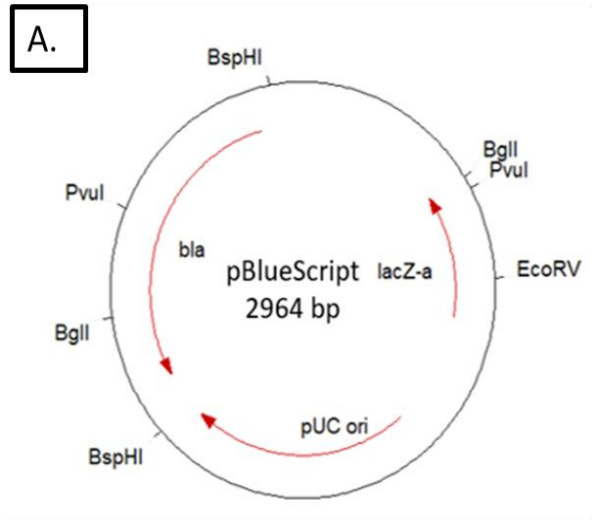
3.3 Confirmation of cloning vectors

The pBlueScript and pMV306H plasmids used for cloning the *neiI* gene were confirmed for their integrity using restriction enzyme analysis (See section 2.3.2). pBlueScript is an *E. coli* replicating vector and has an ampicillin resistant *bla* gene and a *lacZ* gene which encodes beta-galactosidase. When the *lacZ* gene is disrupted in the presence of 5-bromo-4-chloro-indolyl-galactopyranoside (X- Gal), white colonies are formed. However, if the *lacZ* gene has not been disrupted, blue colonies are visible. Thus, blue- white screening is an important selection tool when confirming possible clones. In contrast, the pMV306H vector is an *M. smegmatis* integrating plasmid with an *attP* phage attachment site, and a hygromycin resistance gene. Both these vectors were extracted from *E. coli* DH5α cells (See section 2.2.2) and analyzed by electrophoresis on an agarose gel (see section 2.3.4.1) following enzymatic digestion.

3.3.1 Confirmation of pBlueScript

The integrity of the pBlueScript vector was confirmed by restriction digest analysis with 4 different enzymes and analysed using agarose gel electrophoresis to detect the expected fragment sizes (Figure 3.6). The pBlueScript (2964 bp) vector was digested with the 4 different restriction enzymes listed in the table Figure 3.6 B. Agarose gel electrophoreses analysis confirmed that the restriction enzymes produced digested fragments of the expected sizes as can be seen in Figure 3.6 B and C.

A.



B.

Enzyme	Expected fragment size (bp)
Bgl I	1691, 1273
BspHI	1956, 1008
EcoRV	2964
PvuI	1913, 1051

C.

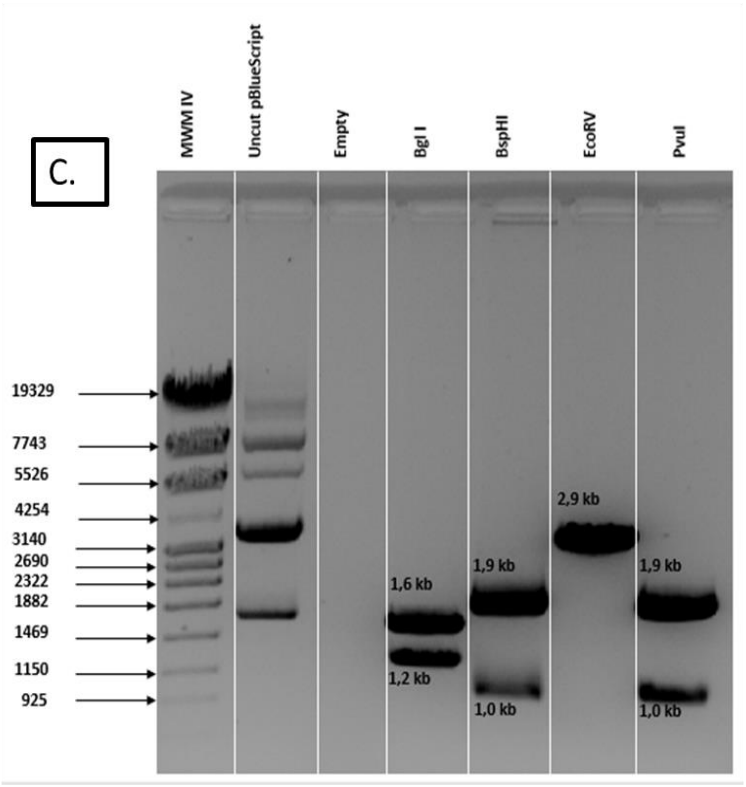
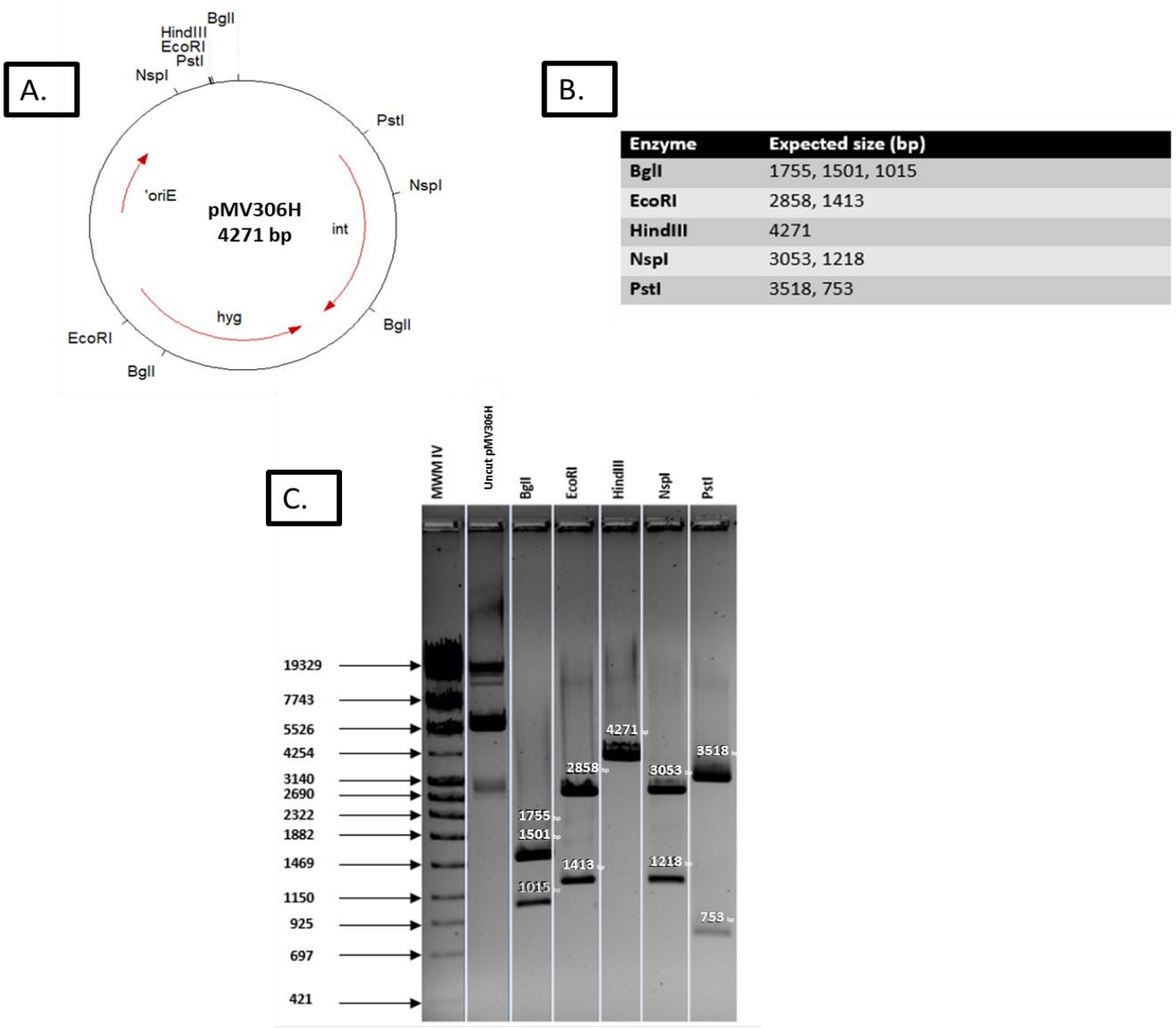


Figure 3.6: pBlueScript confirmation using restriction profiling. *A: pBlueScript vector map displaying restriction enzyme digestion sites. B: Table showing the restriction enzymes used and the expected fragment sizes for each of the enzyme. C: Agarose gel electrophoresis displaying the results of the restriction digestion. The uncut plasmid displayed several conformations which may have included: supercoiled, relaxed and nicked. All the enzymes gave the expected fragments. Uncut pBlueScript DNA (lane 2) was included to ensure that all digestion reached completion. In lane 1 MWM IV indicates the molecular weight marker IV.*

The gel electrophoresis image displays the correct fragment sizes with all the restriction enzymes confirming that the *E. coli* replicating vector, pBlueScript was correct and thus could be used for the next stage of cloning.

3.3.2 Confirmation of pMV306H

The genotype of the pMV306H vector was confirmed using 5 different restriction enzymes and the digested fragments were analysed on an agarose gel (Figure 3.7). The pMV306H (4271 bp) vector was digested with the 5 different restriction enzymes listed in the table Figure 3.7 B. Agarose gel electrophoreses analysis confirmed that the restriction enzymes produced digested fragments of the expected sizes as can be seen in Figure 3.7 B and C.



Enzyme	Expected size (bp)
BglI	1755, 1501, 1015
EcoRI	2858, 1413
HindIII	4271
NspI	3053, 1218
PstI	3518, 753

Figure 3.7: Confirmation of pMV306H by restriction profiling. *A: pMV306H vector map indicating restriction enzyme cut sites. B: Table displaying the BglI, EcoRI, HindIII, NspI and PstI restriction enzymes used and the expected fragment sizes that each of the enzymes produce. C: Agarose gel electrophoresis displaying the results of the restriction digestion. The control reaction contains uncut pMV306H DNA. When digesting with the BglI restriction enzyme, the expected fragment sizes of 1755 bp, 1501 bp and 1015bp are observed. The EcoRI enzyme yields fragment sizes of 2858bp and 1413 bp. Digestion with HindIII results in a fragment size of 4271 bp. The expected fragments of 3053 bp and 1218 bp are observed when digesting with NspI. When digesting with PstI, the expected fragment sizes of 3518 bp and 753 bp are observed. MWM IV indicates molecular weight marker IV.*

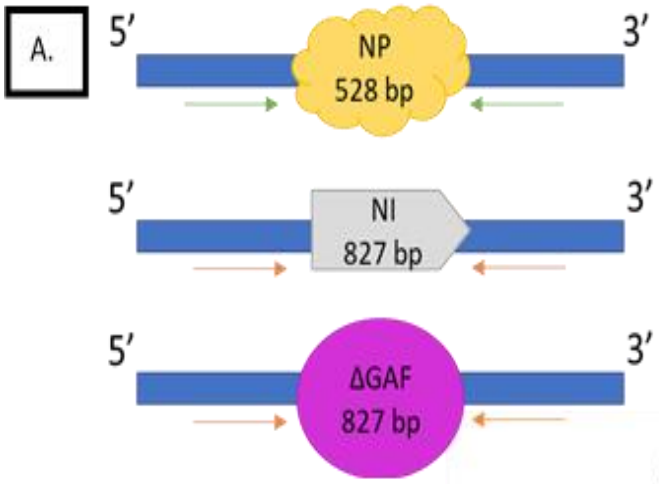
The expected fragment sizes were observed with all the restriction enzymes tested confirming that the pMV306H plasmid was correct and therefore, was suitable for cloning.

3.4 Cloning into pBlueScript

Genomic DNA isolated from mc^2155 (see section 2.2.1.1), was used to amplify the *neiI* promoter sequence (NP), the *neiI* gene encoding the NeiI glycosylase with GLY residues (NI) using primers listed in Appendix C. Using the same primers, the *neiI* gene encoding the NeiI glycosylase with GAF residues (ΔGAF) was amplified from the pUC57 vector containing the *neiI* gene with the altered amino acid residues obtained from Genscript. These primers were designed to contain the following restriction sites: NcoI at the 5' end and BstBI at the 3' end of the *neiI* promoter region, and BstBI at the 5' end and HindIII at the 3' end of the *neiI* gene and altered *neiI* gene to assist with the cloning (see Appendix F).

3.4.1 Cloning fragments

Following PCR amplification of the NP, NI and ΔGAF fragments using primers listed in Appendix C, the fragments were run on an agarose gel using gel electrophoresis (see section 2.2.1.2). The expected fragment sizes of 528 bp for NP, and 827 bp for NI and ΔGAF were observed. In preparation for cloning, the pBlueScript vector was digested at the EcoRV restriction site, yielding a blunt ended fragment of 2964 bp (Figure 3.8). The linear pBlueScript vector was dephosphorylated to prevent re- circularization on the vector. The NP, NI and ΔGAF fragments were blunt ligated into the EcoRV digested pBlueScript vectors to yield the pBlueScript + NP, pBlueScript + NI and pBlueScript ΔGAF constructs. Ligation of the amplicons into the pBlueScript vector at the EcoRV site results in the disruption of the *lacZ* gene (see Appendix F) allowing for blue- white selection of clones on Luria-Bertani agar (LA) plates supplemented with X- gal, where white colonies represent possible clones with the various inserts whilst blue colonies indicate intact or re-circularized vector.



B.

Fragment	Expected size (bp)
pBlueScript	2964
NP	528
NI	827
ΔGAF	827

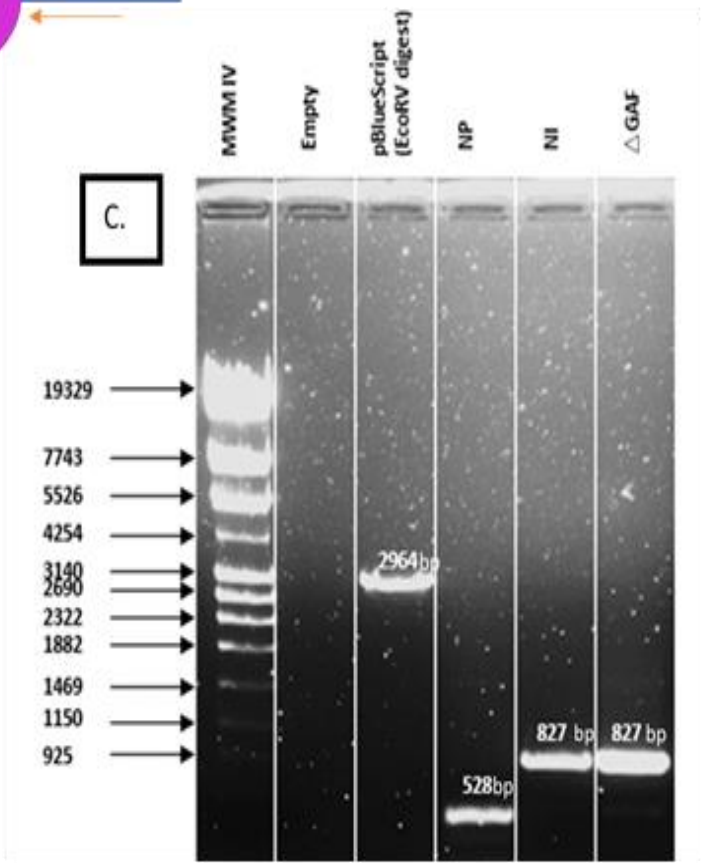


Figure 3.8: PCR fragments for ligation to pBlueScript*. A: Schematic representation of the NP and NI primers used in the amplification of NP, NI and ΔGAF, and the expected fragment sizes they produce. The primers used to amplify NP are represented with green arrows, whilst those used to amplify NI and ΔGAF are represented with orange arrows. B: Table indicating the expected sizes of the PCR products and pBlueScript vector digested with EcoRV. C: Gel electrophoresis image of pBlueScript digested with EcoRV and the NP, NI and ΔGAF PCR products. The amplification of NP yielded the expected fragment size of 528 bp when using NP primers, whilst the expected fragment size of 827 bp was observed for NI and ΔGAF when using the NI primers. A fragment size of 2964 bp was observed after the digestion of pBlueScript with the EcoRV restriction enzyme. MWM IV indicates molecular weight marker IV.*

3.4.2 pBlueScript clone confirmations

Following blunt ligation of the amplified NP, NI and ΔGAF into the pBlueScript vector at the EcoRV site (see section 2.3.6), competent *E. coli* DH5α cells were transformed with these constructs, plated on LA supplemented with X-gal (see section 2.4.1) and incubated overnight at 30ºC (see section 2.1, table 1.1). The DH5α cells only control did not display growth due to the presence of ampicillin. A "cut and ligate" control was included wherein the linear dephosphorylated pBlueScript vector was ligated and used to transform competent *E. coli* DH5α cells. This control did not display growth due to sufficient dephosphorylation of the vector. Transformations of 1:1 and 1:2 (vector: insert) ratios resulted in colonies whilst the 2:2 and 3:1 vector: insert ratios did not yield any clones (Table 3.1). Clones with a disrupted *lacZ* gene due to the insertion of the NP, NI or ΔGAF fragments appeared as white colonies on the X- gal plates, whilst colonies that did not incorporate either of the fragments were blue in colour.

Table 3.1: Results of pBlueScript DH5α transformations

vector: insert ratios	1:1		1:2		2:2		3:1	
	Blue	White	Blue	White	Blue	White	Blue	White
pBlueScript + NP	135	156	112	144	0	0	0	0
pBlueScript + NI	104	125	121	33	0	0	0	0
pBlueScript + ΔGAF	96	115	152	162	0	0	0	0
DH5α cells only	0		NA		NA		NA	
Cut and ligate	0		NA		NA		NA	

Twelve white colonies from each transformation were selected, grown to stationary phase and the plasmid extracted from these clones (see section 2.2.2.1) was subjected to gel electrophoresis (see section 2.3.4.1) (Figure 3.9).

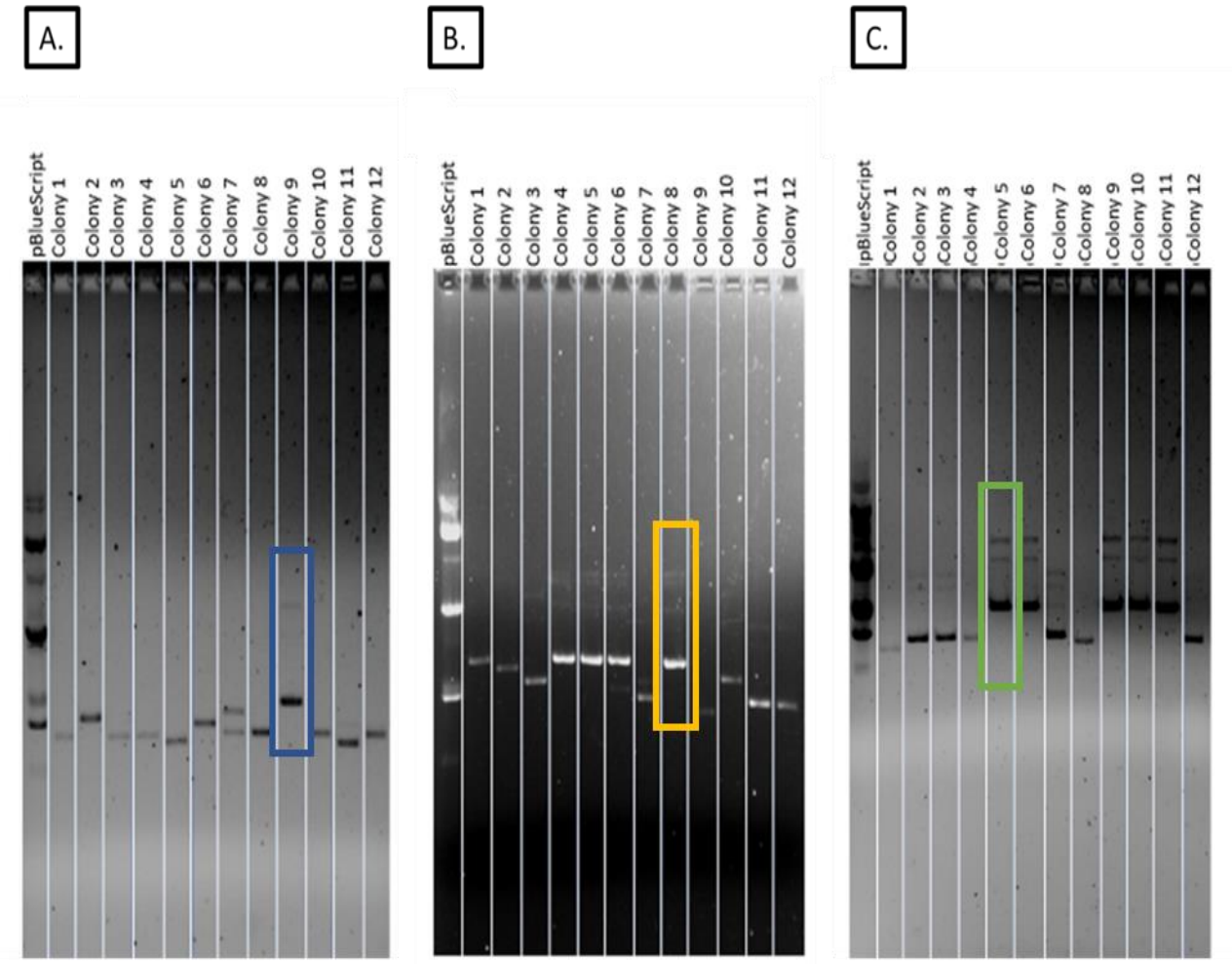


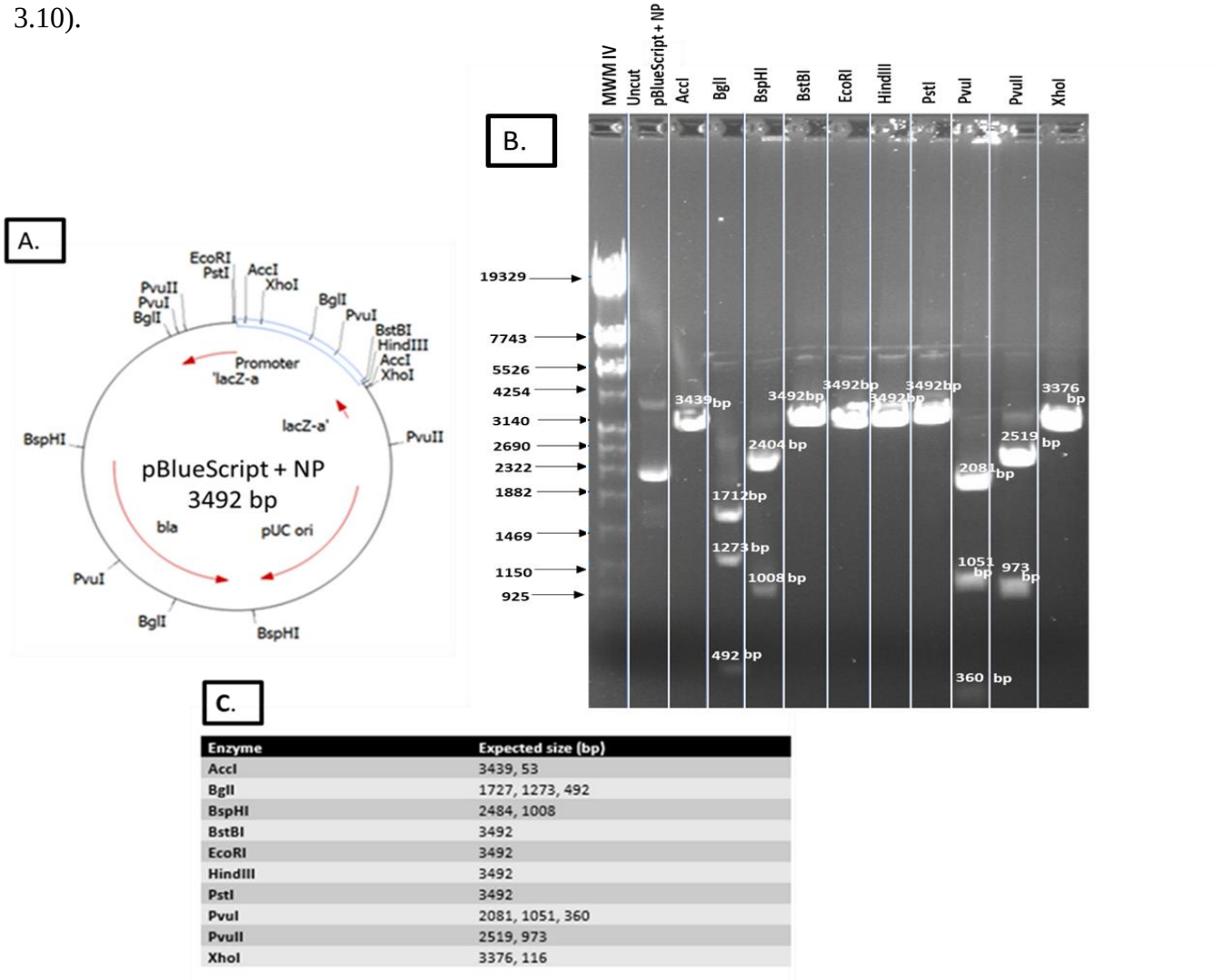
Figure 3.9: DNA extractions from potential pBlueScript clones. *A: DNA of potential pBlueScript + NP clones. B: DNA of potential pBlueScript + NI clones. C: DNA of potential pBlueScript + ΔGAF clones. All potential clones were compared to an empty pBlueScript vector. Clones that ad a higher molecular weight on the agarose gel compared to the lowest band of the empty pBlueScript vector were considered to have an insert and were considered for further screening. The blue, yellow and green blocks indicate the clones that were used in further screening. A molecular weight marker was not necessary as the agarose gel electrophoresis served to compare the various colonies to the pBlueScript plasmid, and sizing was not necessary.*

The differences in the mobility shift of the plasmid containing an insert compared to the pBlueScript vector allowed for identification of possible clones with each of the inserts. Based on this, one clone from each transformation was analysed further by restriction digest to confirm the insertion of the respective clones.

3.4.2.1 Confirmation of pBlueScript + NP

Colony 9, as indicated by the blue block in figure 3.9A, was chosen for further screening due to its larger band sizes compared to the empty pBlueScript vector. DNA from colony 9 was

extracted, subjected to restriction analysis and confirmed using agarose gel electrophoresis (see sections 2.2.1.2, 2.3.2 and 2.3.4.1). The pBlueScript + NP (3492 bp) vector was digested with the 10 different restriction enzymes listed in the table Figure 3.10 C. Agarose gel electrophoreses analysis confirmed that the restriction enzymes produced digested fragments of the expected sizes as can be seen in Figure 3.10 B and C. To assess whether the NP fragment had successfully ligated into the pBlueScript vector, restriction digest analysis was conducted using the following restriction enzymes: AccI (3439bp, 53 bp), BglI (1727 bp, 1273 bp, 492 bp), BspI (2484 bp, 1008 bp), BstBI (3492 bp), EcoRI (3492 bp), HindIII (3492 bp), PstI (3492 bp), PvuI (2081 bp. 1051 bp, 360bp), PvuII (2519 bp, 973 bp), XhoI (3376 bp, 116 bp). (Figure 3.10).



C.

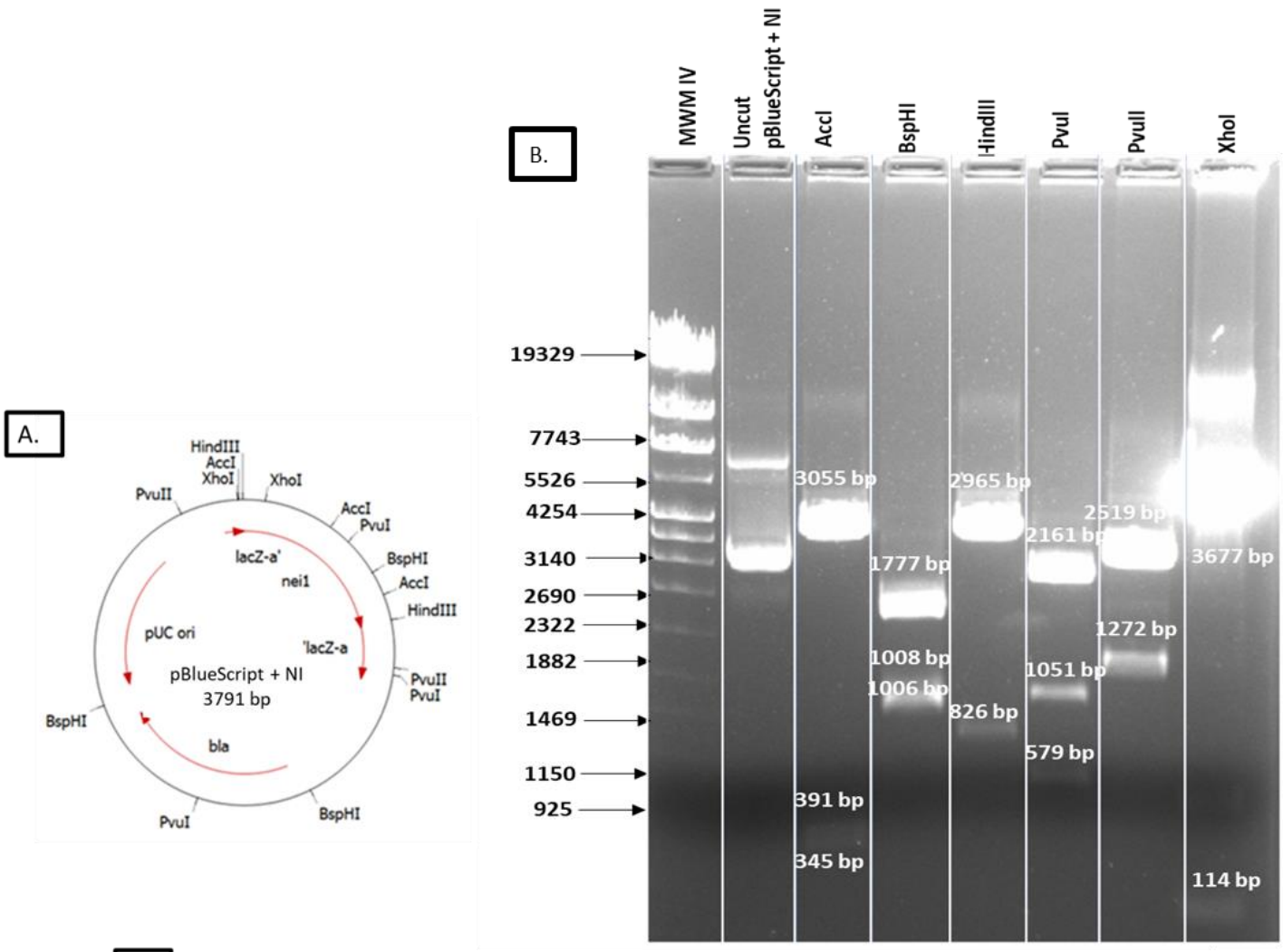
Enzyme	Expected size (bp)
AccI	3439, 53
BglI	1727, 1273, 492
BspHI	2484, 1008
BstBI	3492
EcoRI	3492
HindIII	3492
PstI	3492
PvuI	2081, 1051, 360
PvuII	2519, 973
XhoI	3376, 116

Figure 3.10: Confirmation of the* neiI *promoter in pBlueScript. *A: NP + pBlueScript vector map indicating the position of the* neiI *promoter as well as various restriction sites. B: Agarose gel indicating the correct fragment sizes of the digested NP + pBlueScript clone. AccI displayed a fragment size of 3493 bp. The BglI enzyme yielded fragments of 1712 bp, 1273 bp and 492 bp whilst BspHI, gave the expected band sizes of 2484 bp and 1008 bp. BstBI, EcoRI, HindIII and PstI all produced a single fragment of 3492 bp. Digestion with PvuI resulted in fragments of 2081 bp, 1051 bp and 360 bp. Similarly, PvuII produced 2519 bp and 973 bp fragments. For the XhoI digest the 3376 bp fragment was observed. The gel percentage did not allow for the detection of the small fragments resulting from the AccI and XhoI digests. MWM IV represented molecular weight marker IV. C: Table displaying restriction enzymes used for the confirmation of NP + pBlueScript and the expected fragment sizes for each enzyme digest.*

3.4.2.2 Confirmation of pBlueScript + NI and pBlueScript + ΔGAF

Similarly, colony 8, as indicated by the yellow block in figure 3.9B, was chosen for pBlueScript + NI and colony 5, as indicated by the green block in figure 3.9C, for pBlueScript + ΔGAF for further screening due to their larger band sizes compared to the empty pBlueScript vector. DNA from colony 8 of pBlueScript + NI and colony 5 of pBlueScript + ΔGAF was extracted, digested with various restriction digests that were analysed using agarose gel electrophoresis (see sections 2.2.1.2, 2.3.2 and 2.3.4.1) to assess whether the NI and ΔGAF fragments were ligated into the pBlueScript vector as expected (Figure 3.11 and Figure 3.12 respectively). The pBlueScript + NI (3791 bp) vector was digested with the 6 different restriction enzymes listed in the table Figure 3.11 C. Agarose gel electrophoreses analysis confirmed that the restriction enzymes produced digested fragments of the expected sizes as can be seen in Figure 3.11 B and C. The restriction digest analysis of pBlueScript + NI was conducted using the following restriction enzymes: AccI (3055 bp, 391 bp, 345 bp), BspHI (1777 bp, 1008 bp, 1006 bp), HindIII (2965 bp, 826 bp), PvuI (2161 bp, 1051 bp, 579 bp), PvuII (2519 bp, 1272 bp), XhoI (3677 bp, 114bp). (Figure 3.11).

The pBlueScript + ΔGAF vector was digested with the 10 different restriction enzymes listed in the table Figure 3.12 C. Agarose gel electrophoreses analysis confirmed that the restriction enzymes produced digested fragments of the expected sizes as can be seen in Figure 3.12 B and C. The restriction digest analysis of pBlueScript + ΔGAF was conducted using the following restriction enzymes: AccI (3313 bp, 391 bp, 345 bp), BglII (2182 bp, 1273 bp, 259 bp, 77 bp), BspHI (1413 bp, 1370 bp, 1008 bp), BstBI (3791 bp), EcoRI (3202 bp, 589 bp), HindIII (3778 bp, 13 bp), PstI (3791 bp), PvuI (2066 bp, 1051 bp, 674 bp), PvuII (2519 bp, 1272 bp), XhoI (3024 bp, 767 bp). (Figure 3.12).

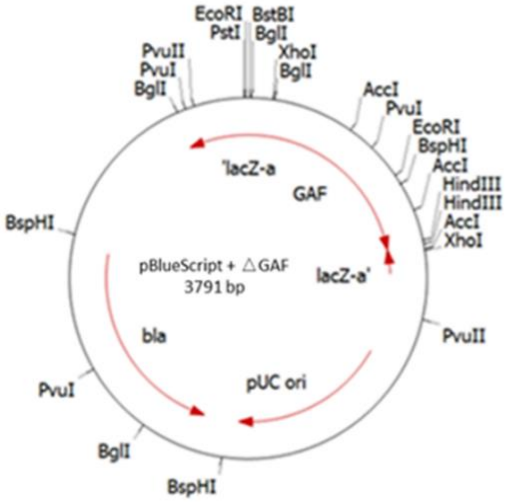


C.

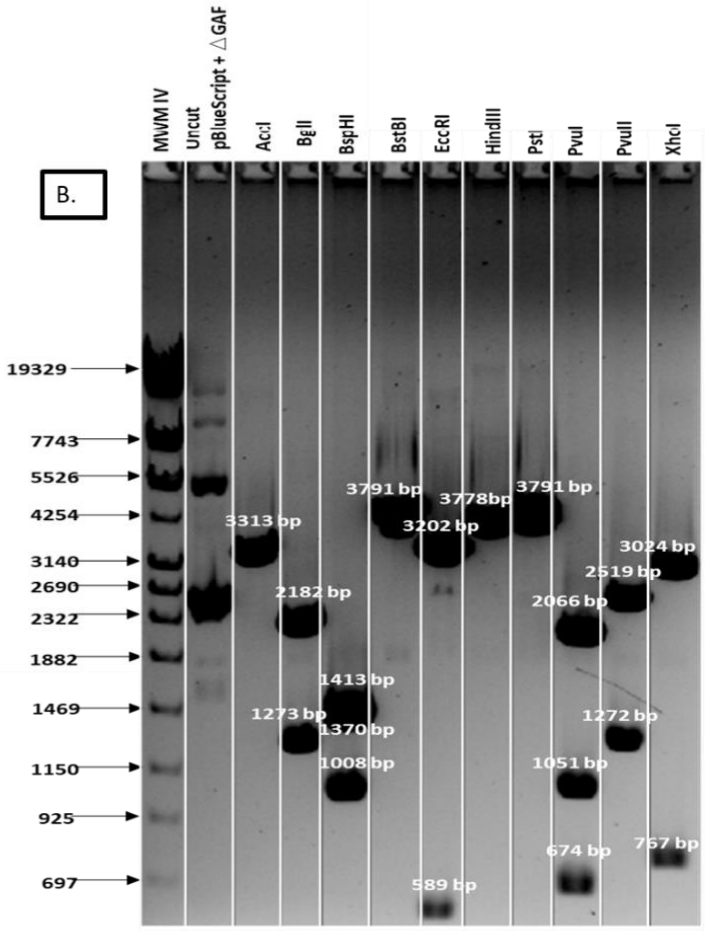
Enzyme	Expected size (bp)
AccI	3055, 391, 345
BspHI	1777, 1008, 1006
HindIII	2965, 826
PvuI	2161, 1051, 579
PvuII	2519, 1272
XhoI	3677, 114

Figure 3.11: Confirmation of the <u>neiI</u> gene in pBlueScript. *A: pBlueScript + NI vector map indicating the position of the* neiI *gene and various restriction sites. B: Agarose gel showing the correct fragment sizes of the digested pBlueScript + NI clone. AccI indicated fragment sizes of 3055 bp, 391 bp and 345 bp. The BspHI enzyme yielded fragments of 1777 bp, 1008 bp and 1006 bp. When digesting with HindIII, band sizes of 2965 bp and 826 bp can be seen. The PvuI enzyme produced fragments of 2161 bp, 1051 bp and 579 bp. Digestion with PvuII indicated fragments of 2519 bp and 1272 bp. XhoI produced 3677 bp and 114 bp fragments. The third band may be due to a partial digest of a high concentration of DNA. MWM IV represented molecular weight marker IV. C: Table displaying restriction enzymes used for the confirmation of pBlueScript + NI and the expected fragment sizes they each produce.*

A.



B.



C.

Enzyme	Expected size (bp)
AccI	3313, 345, 133
BglI	2182, 1273, 259, 77
BspHI	1413, 1370, 1008
BstBI	3791
EcoRI	3202, 589
HindIII	3778, 13
PstI	3791
PvuI	2066, 1051, 674
PvuII	2519, 1272
XhoI	3024, 767

Figure 3.12: Confirmation of the ΔGAF gene in pBlueScript. *A: pBlueScript + ΔGAF vector map indicating the position of the altered* neiI *gene and various restriction sites. B: Agarose gel showing the correct fragment sizes of the digested pBlueScript + ΔGAF clone. AccI indicated a fragment size of 3313 bp, however the expected 345 bp and 133 bp fragments were too small to be observed on the agarose gel. The BglI enzyme yielded fragments of 2182 pb and 1273 bp. Similar to AccI, the 259 bp and 77 bp fragments were not large enough to be seen on the gel. When digesting with BspHI, band sizes of 1413, 1370 and 1008 can be seen. The BstBI and PstI enzymes produced fragments of 1791 bp each. Digestion with EcoRI indicated fragments of 3202 bp and 589 bp. HindIII produced a single 3778 bp fragment as the 13 bp fragment could not be seen on the gel due to its small size. Digestion with PvuI resulted in the expected fragments of 2064 bp, 1051 bp and 674 bp. The PvuII digest displayed band sizes of 2519 bp and 1272 bp. XhoI produced 3024 bp and 767 bp fragments. MWM IV represented molecular weight marker IV. C: Table displaying restriction enzymes used for the confirmation of pBlueScript + ΔGAF and the expected fragment sizes they each produce.*

The restriction analysis indicated that the clones containing the promoter region of the *neiI* gene, the *neiI* gene and the altered *neiI* gene in the pBlueScript vector were correct. The results of the digestion of the pBlueScript + *neiI* promoter using the AccI enzyme indicated a fragment size of 3493 bp. The BglI enzyme resulted in fragments of 1712 bp, 1273 bp and 492 bp whilst BspHI, gave the expected band sizes of 2484 bp and 1008 bp. In conjunction, BstBI, EcoRI, HindIII and PstI all produced a single fragment of 3492 bp. As expected, digestion with PvuI resulted in fragments of 2081 bp, 1051 bp and 360 bp. Similarly, PvuII produced 2519 bp and 973 bp fragments. For the XhoI digest the 3376 bp fragment was observed.

The AccI enzyme digested the pblueScript + *neiI* construct which produced a fragment size of 3493 bp. Similarly, the BglI enzyme resulted in fragments of 1712 bp, 1273 bp and 492 bp whilst BspHI, indicated the expected band sizes of 2484 bp and 1008 bp. BstBI, EcoRI, HindIII and PstI all produced a single fragment of 3492 bp. When digestion was conducted with PvuI, fragment sizes of 2081 bp, 1051 bp and 360 bp were observed. Similarly, PvuII produced 2519 bp and 973 bp fragments. The XhoI digest resulted in the 3376 bp fragment. For both the pBlueScript + *neiI* promoter and pBlueScript + *neiI* constructs, the gel percentage prevented the detection of the small fragments resulting from the AccI and XhoI digests.

In relation to the pBlueScript + ΔGAF construct, AccI displayed a fragment size of 3313 bp, however the expected 345 bp and 133 bp fragments were too small to be observed on the agarose gel. Similarly, the 259 bp and 77 bp fragments of BglII were not large enough to be seen on the gel. However, the BglII enzyme did display fragments of 2182 pb and 1273 bp. Band sizes of 1413, 1370 and 1008 were observed when digested with BspHI. The BstBI and PstI enzymes produced fragments of 1791 bp each. Similarly, EcoRI indicated fragments of 3202 bp and 589 bp whilst HindIII produced a single 3778 bp. The 13 bp fragment that HindIII was expected to produce, was too small to be viewed on the gel. Digestion with PvuI resulted in the expected fragments of 2064 bp, 1051 bp and 674 bp. The PvuII digest displayed band sizes of 2519 bp and 1272 bp. XhoI produced 3024 bp and 767 bp fragments.

These clones were then grown in bulk and the plasmids were purified (see section 2.2.2.2) for subsequent sequence analysis at the Central Analytical Facility (CAF) at the University of Stellenbosch. This was done in order to ensure that no errors were introduced in the PCR fragment during amplification before proceeding with further cloning into the integrating vector, pMV306H.

3.4.2.3 Sequencing confirmations

Once the sequencing data were received from the sequencing facility, it was cleaned using FinchTV to ensure sequence accuracy and void of baseline noise before analysis against the sequence generated for pBlueScript + NP, pBlueScript + NI and pBlueScript + ΔGAF in clone manager 9 (see section 2.3.7). The alignment results for all three assembled sequences are shown below (Figure 3.13).

The sequences of pBlueScript + NP, pBlueScript + NI and pBlueScript + ΔGAF aligned well with the predicted sequence in clone manager with full coverage and no base discrepancies. There is a 100 % match for all three pBlueScript constructs with a 0 % error report. The GLY loop encoded by GGGCTCTAC bases and the GAF loop encoded by GGGGCCTTC bases were not altered during PCR amplification. These data confirmed that the PCR amplification and subsequent ligation of the *neiI* promoter region, *neiI* gene and altered *neiI* gene into pBlueScript did not cause any mutations. Thus, the pBlueScript + NP, pBlueScript + NI and pBlueScript + ΔGAF constructs were correctly generated and could be used for downstream cloning applications.

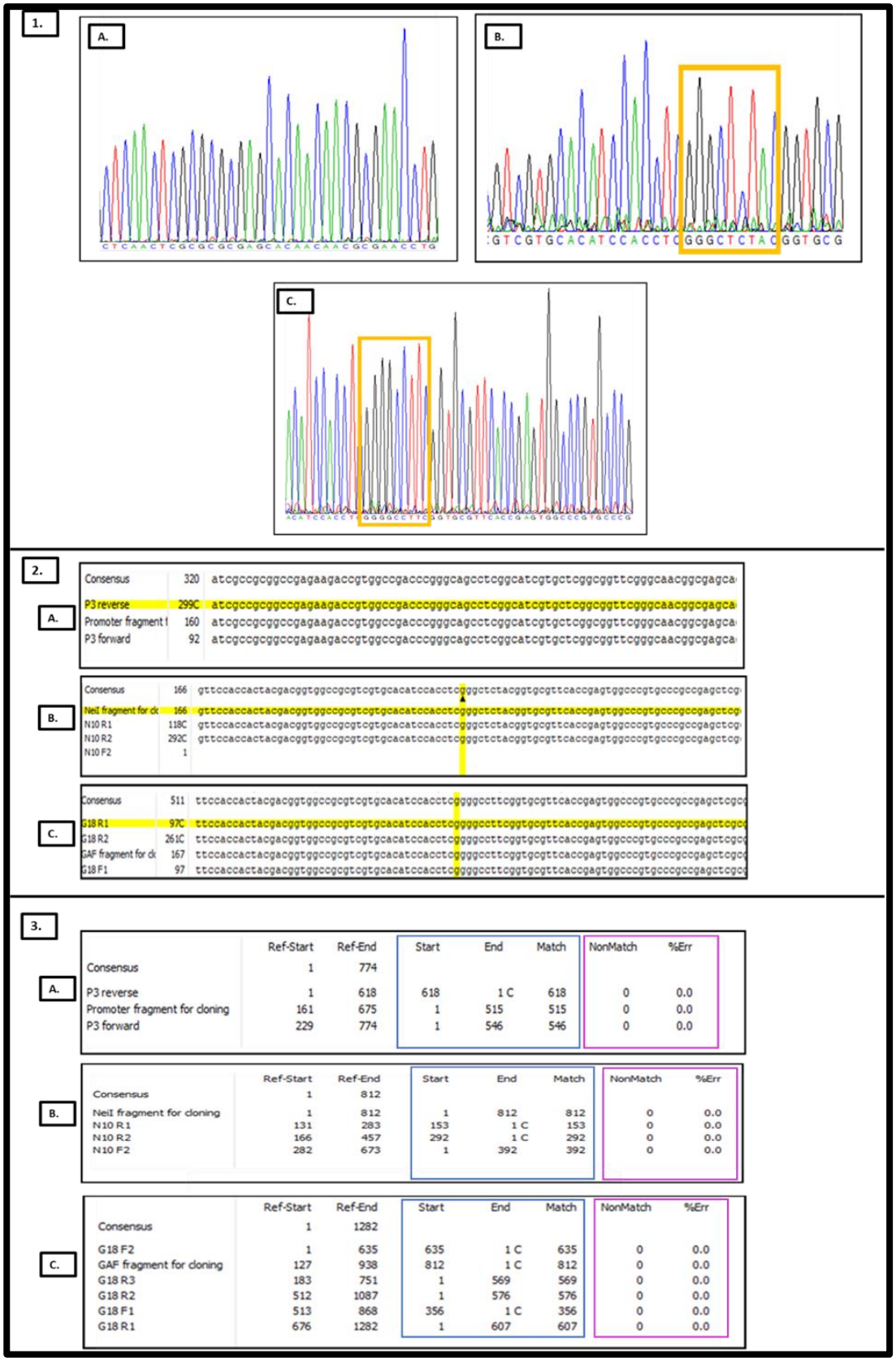


Figure 3.13: Sequencing confirmation of cloned pBlueScript plasmids. *A: pBlueScript + NP. B: pBlueScript + NI. C: pBlueScript + ΔGAF. 1: Chromatogram displaying clear peaks and accurate sequencing for each sequence alignment. The orange boxes indicate the bases that encode the GLY loop (B) and GAF loop (C). 2: Tables displaying sequence alignments of the cloned plasmids and sequences generated by CAF sequencing facility. There are no discrepancies in the alignments. The vertical yellow lines indicate the location of the bases encoding the GLY loop (B) and GAF loop (C). 3: Tables showing primer coverage and error rates. The blue boxes indicate the regions that the primers cover as well as how many bases match between the sequences. The purple boxes indicate any non- matches and the percentage of errors found.*

3.5 Cloning in integrating vector, pMV306H

Once the sequences for the NP, NI and ΔGAF fragments were confirmed, each of the fragments were excised from the replicating vector, pBlueScript with restriction enzymes (see sections 2.3.1.4 and 2.3.2). The NP fragment was excised using NcoI and BstBI restriction enzymes, whilst the NI and ΔGAF fragments were excised using BstBI and HindIII restriction enzymes. Thereafter, the *M. smegmatis* integrating vector; pMV306H was digested with NcoI and HindIII restriction enzymes. This facilitated the sticky- end ligation of NP and NI into pMV306H to form the pMV + NP + NI construct. Similarly, a sticky- end ligation was used to ligate NP and ΔGAF into pMV306H, forming the pMV + NP + ΔGAF construct (see Appendix F).

Subsequent to ligations, the new constructs: pMV + NP + NI and pMV + NP + ΔGAF were electroporated into the *ΔnIΔnth M. smegmatis* strain (see section 2.4.2) to generate the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains. The integration of the constructs was confirmed using PCR (see section 2.3.1.4) with the *neiI* promoter forward primer and the *neiI* gene reverse primer (see Appendix C). The details of the cloning are discussed below.

3.5.1 Generation of fragments for cloning into pMV306H

The pMV306H vector was digested with NcoI and HindIII enzymes to facilitate the generation of the pMV + NP + NI and pMV + NP + ΔGAF constructs (see Appendix F). The plasmid was electrophoresed on an agarose gel and purified by the NucleoSpin PCR clean-up and gel extraction kit (see section 2.3.4.2). The pBlueScript vectors containing NP, NI or ΔGAF were digested with NcoI and BstBI enzymes (to excise NP) or BstBI and HindIII enzymes (to excise NI and ΔGAF) and fragments were separated on an agarose gel (Figure 3.14).

The pMV306H vector was digested with NcoI and HindIII enzymes which produced a 4271 bp fragment. Similarly, the pBlueScript + NP plasmid was digested with NcoI and BstBI restriction enzymes which resulted in the 2964 bp pBlueScript vector as well as the 528 bp NP fragment. The pBlueScript + NI and pBlueScript + ΔGAF plasmids were both digested with BstBI and HindIII restriction enzymes, resulting in the 2964 bp vector, and the 827 bp NI fragment and 827 bp ΔGAF fragment, respectively.

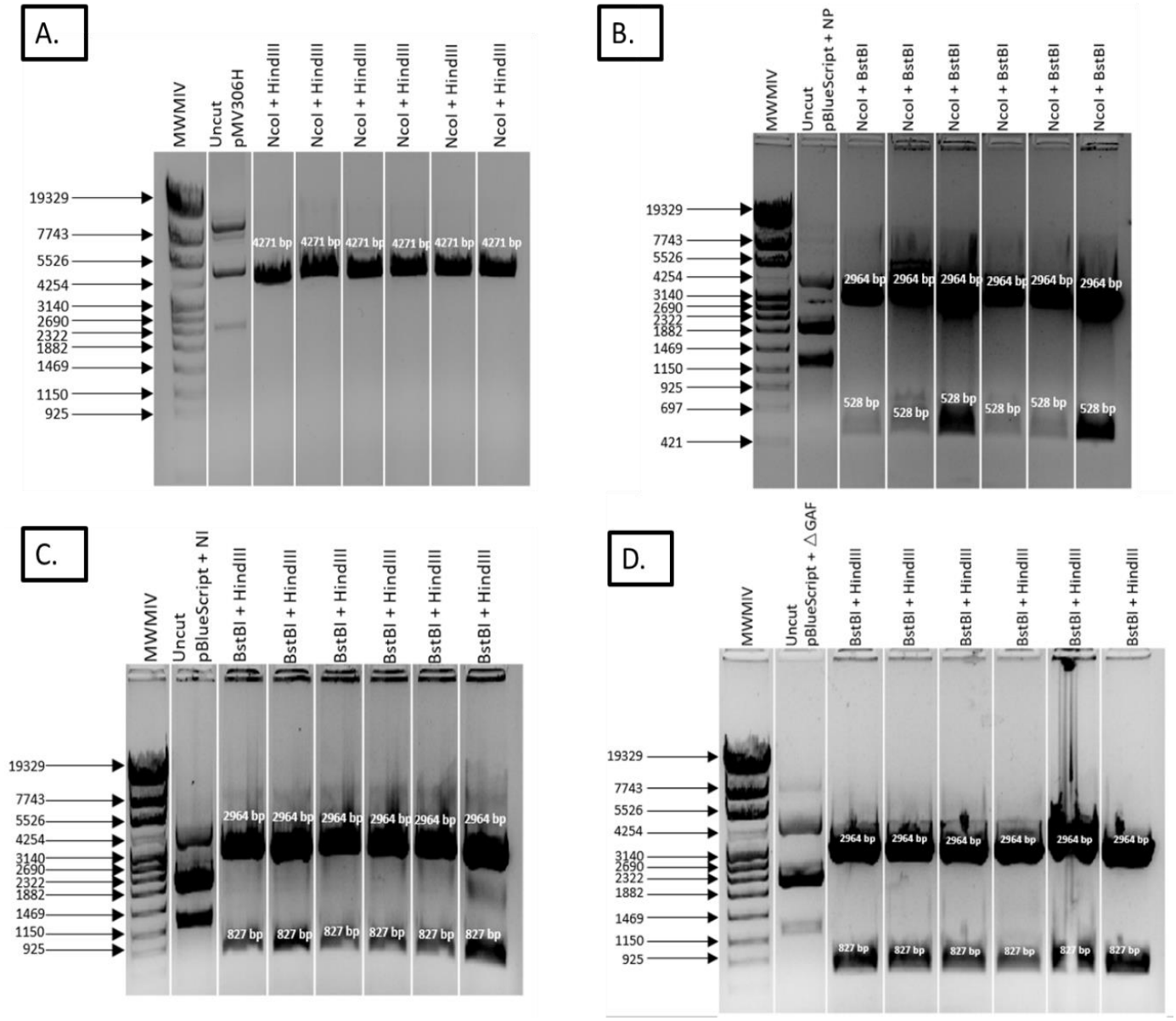


Figure 3.14: Preparation for pMV306H cloning. *A: Agarose gel showing pMV306H digestion with NcoI and HindIII enzymes producing a 4271 bp fragment. B: pBlueScript + NP plasmid digested with NcoI and BstBI enzymes to produce the 528 bp NP fragment and 2964 bp pBlueScript vector. C: pBlueScript + NI vector digested with BstBI and HindIII enzymes to produce the 2964 bp plasmid and 827 bp NI fragment. D: pBlueScript + ΔGAF plasmid digested with BstBI and HindIII restriction enzymes to produce the 827 bp ΔGAF fragment and the 2964 bp pBlueScript fragment. MWMIV indicates molecular weight marker IV.*

The respective fragments from each of the construct and the digested vector were excised out of the gel and purified by the NucleoSpin PCR clean-up and gel extraction kit (see section 2.3.4.2). The concentration of the purified fragments was measured using a nanodrop and further quantitated by gel electrophoresis in preparation for the cloning (Figure 3.15). All the cloning fragments were of the correct size as listed above, at an acceptable concentration to proceed with ligations into pMV306H.

A.

Fragment	Expected size (bp)
pMV306H	4271
NP	528
NI	827
ΔGAF	827

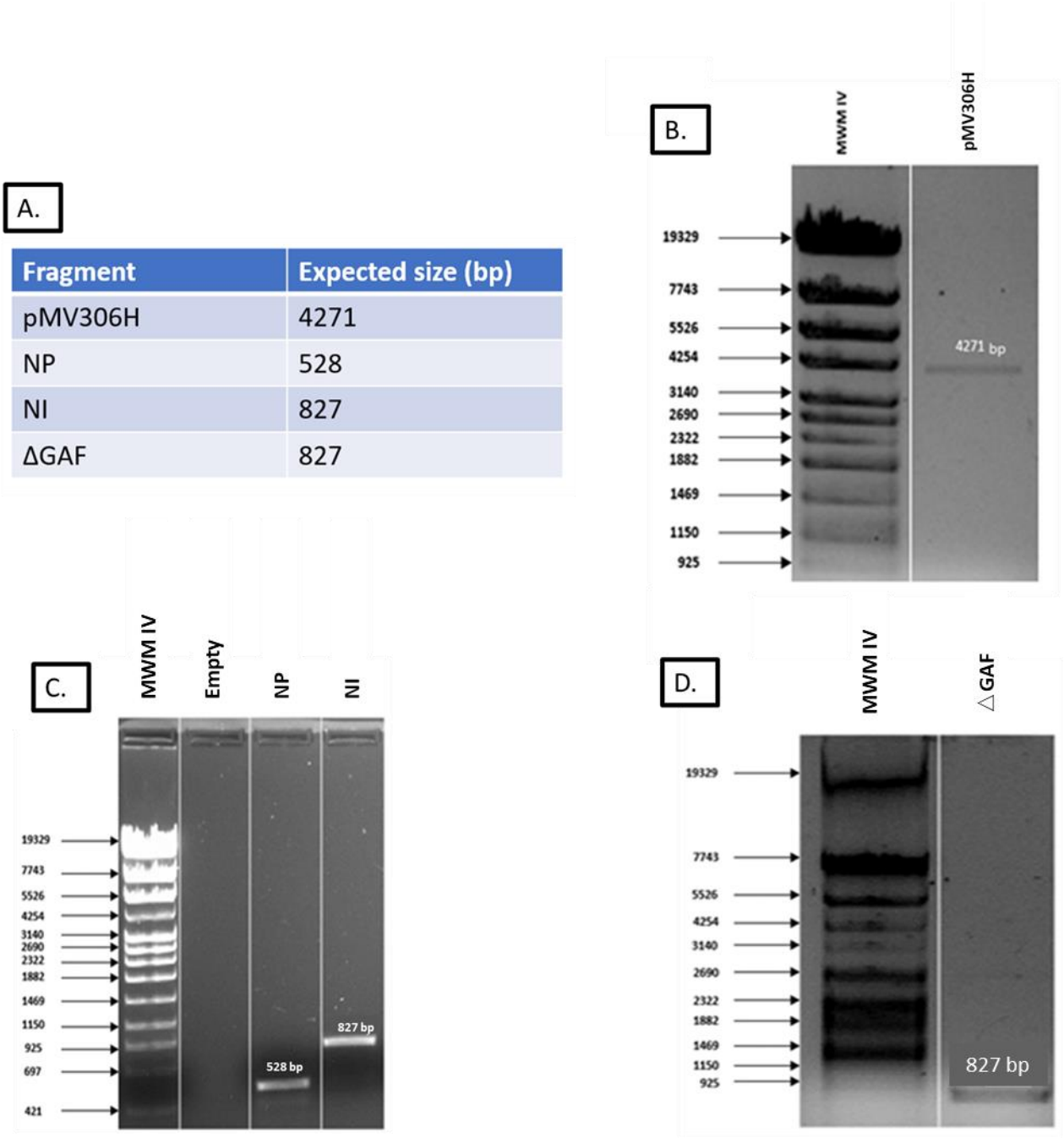


Figure 3.15: Cleaned fragments in preparation for pMV306H cloning. *A: Table indicating the expected band sizes for the digested pMV306H plasmid and cloning fragments. B: Purified pMV306H following digestion with NcoI and HindIII enzymes. C: Cleaned NP and NI fragments after digestion with NcoI and BstBI and BstBI and HindIII, respectively. D: Purified ΔGAF fragment after digestion with BstBI and HindIII restriction enzymes.*

3.5.2 Three way cloning of the NeiI promoter and the native and altered neiI *fragments into pMV306H*

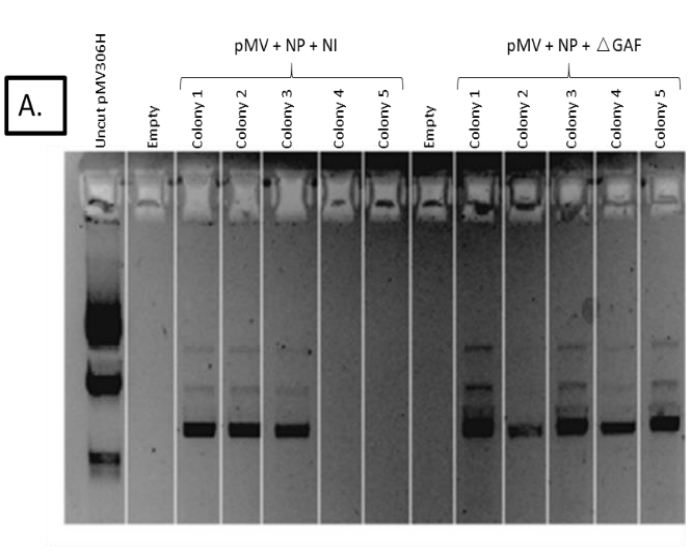
The purified NP, NI or ΔGAF fragments were ligated into the pMV306H vector digested with NcoI and HindIII restriction enzymes (see section 2.3.6). The ligated constructs were transformed into competent *E. coli* DH5α cells, plated on LA supplemented with hygromycin (LA + Hyg) (see section 2.2.2.1 and section 2.4.1) and incubated overnight at 30ºC (see section 2.1, table 1.1). The DH5α cells on the control plate did not display growth due to the presence of hygromycin, whilst the cut and ligate pMV306H control did not display growth as this vector has not transformed *E. coli* DH5α. Transformations of 1:1, 1:2, 2:2 and 3:1 (vector: insert) were conducted, all of which presented with colonies (Table 3.2).

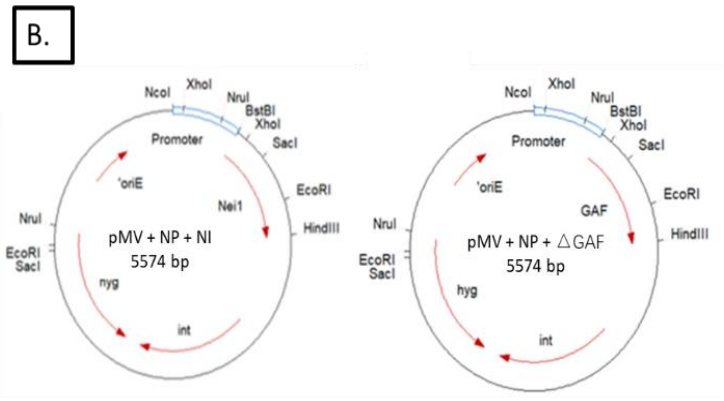
Table 3.2: Results of pMV306H DH5α transformations

	1:1	1:2	2:2	3:1
pMV + NP + NI	215	206	225	15
pMV + NP + ΔGAF	177	144	109	66
DH5α cells only	0	NA	NA	NA
Cut and ligate	0	NA	NA	NA

*NA indicates that the ratios were not applicable for these strains/ constructs

Five colonies for each of the transformed plasmids (pMV + NP + NI and pMV + NP + ΔGAF, respectively) were selected to screen for the cloned plasmids. The colonies were grown to stationary phase for plasmid extraction (see section 2.2.2.2) and the uncut plasmid was analysed on an agarose gel together with the pMV306H vector (Figure 3.16A). Three of the five pMV + NP + NI colonies produced similar results when screened against the empty pMV306H vector, while two of the five did not present with any visible DNA. In contrast, all five of the pMV + NP + ΔGAF colonies produced similar results to each other when screened against the empty pMV306H vector. For preliminary confirmation, colony 1 for both pMV + NP + NI and pMV + NP + ΔGAF was analysed with the EcoRI restriction enzyme that produced the expected band sizes of 3088 bp and 2486 bp each (Figure 3.16). These 2 clones were then grown up in bulk for large scale extraction of the respective plasmids for further analysed with additional restriction enzymes.





C.

Enzyme	Expected size (bp)
EcoRI	3008, 2486

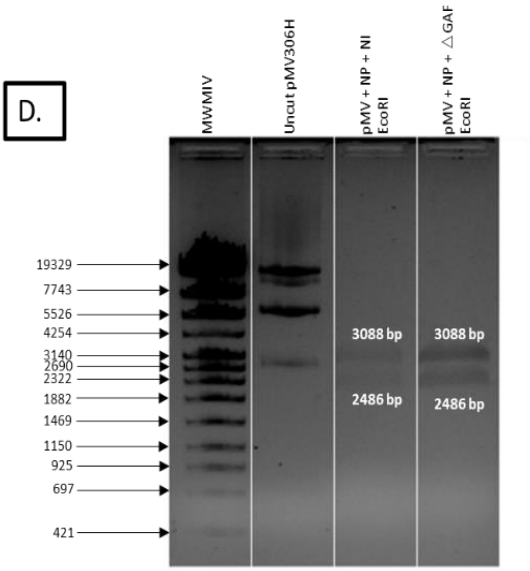


Figure 3:16: Assessment of possible pMV + NP + NI and pMV + NP + ΔGAF clones*. A: pMV + NP + NI and pMV + NP + ΔGAF colonies screened for size against empty pMV306H vector. A molecular weight marker was not necessary as the purpose of the agarose gel was to compare the various colonies against the empty pMV306H vector and thus, the exact sizes of the bands were not needed. B: Vector maps of pMV + NP + NI and pMV + NP + ΔGAF clones displaying restriction sites. C: Table displaying expected fragment sizes produced when the pMV + NP + NI and pMV + NP + ΔGAF clones are digested with EcoRI. D: Agarose gel indicating results of restriction digest. MWMIV indicates molecular weight marker IV.*

3.5.2.1 Confirmation of the pMV + NP + NI plasmid

The pMV + NP + NI clone was confirmed using restriction analysis using the enzymes BstBI (5574 bp), EcoRI (3088 bp, 2486 bp), HindII (5574 bp), NruI (3938 bp, 1636 bp), SacI (3379 bp, 2195 bp) and XhoI (5062 bp, 512 bp) (Figure 3.17). The restriction digest analysis confirmed that the pMV + NP + NI clone had the *neiI* promoter region and *neiI* gene ligated into the pMV306H vector in the correct orientation as all the expected fragment sizes were observed for all the enzymes used. The NruI enzyme appears to have partially digested the plasmid, which was attributed to old buffer rather than inconsistencies with the clone.

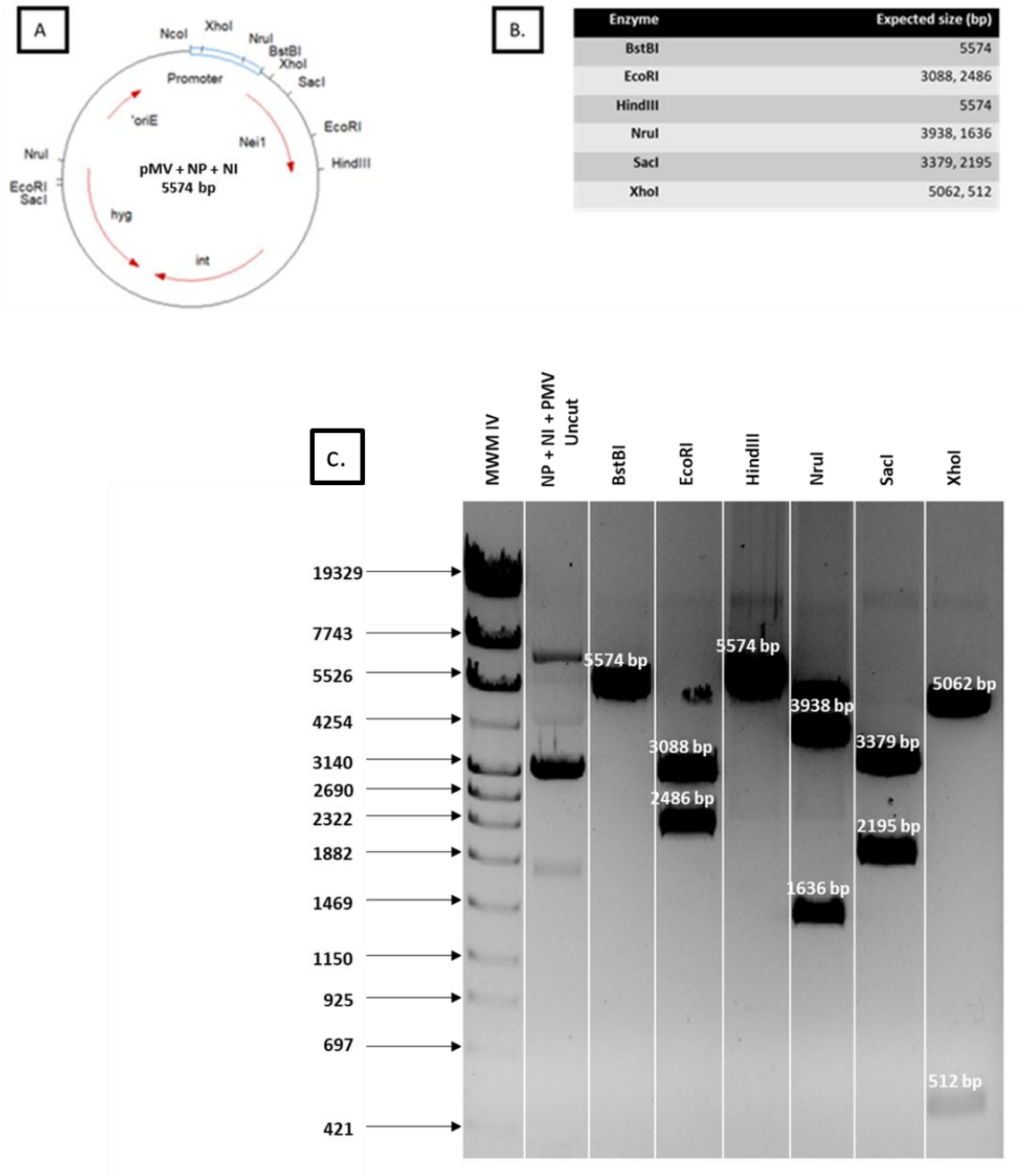


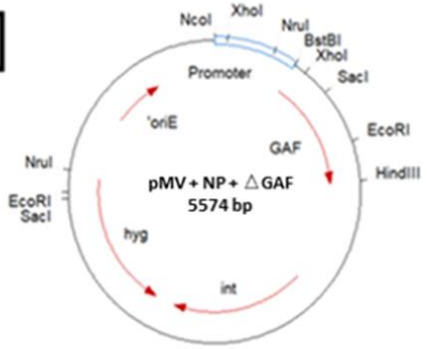
Figure 3.17: Confirmation of the pMV + NP + NI clone. *A: pMV + NP + NI vector map indicating the position of the neiI promoter,* neiI *gene and various restriction sites. B: Table displaying restriction enzymes used for the confirmation of pMV + NP + NI and the expected fragment size. C: Agarose gel showing the correct fragment sizes of the digested pMV + NP + NI clone. BstBI and HindIII produced a single fragment of 5574 bp. The EcoRI enzyme yielded fragments of 3088 bp and 2486 bp. Digestion with NruI resulted in the expected fragments of 3938 bp and 1636 bp. The SacI digest displayed band sizes of 3379 bp and 2195 bp. XhoI produced 5062 bp and 512 bp fragments. MWM IV indicates molecular weight marker IV.*

3.5.2.2 Confirmation of the pMV + NP + ΔGAF plasmid

Confirmation of the pMV + NP + ΔGAF clone was conducted using restriction analysis with the BstBI, EcoRI, HindII, NruI, SacI and XhoI enzymes (Figure 3.18). This analysis provided

confirmation that the *neiI* promoter region and altered *neiI* gene ligated into the pMV306H vector in the correct orientation to produce the pMV + NP + ΔGAF clone. The expected fragment sizes were obtained for all the enzymes used. The NruI enzyme does not appear to have partially digested the plasmid, which was attributed to the fluctuating results the old buffer produces, as described above.

A



B.

Enzyme	Expected size (bp)
BstBI	5574
EcoRI	3088, 2486
HindIII	5574
NruI	3938, 1636
SacI	3379, 2195
XhoI	5062, 512

C.

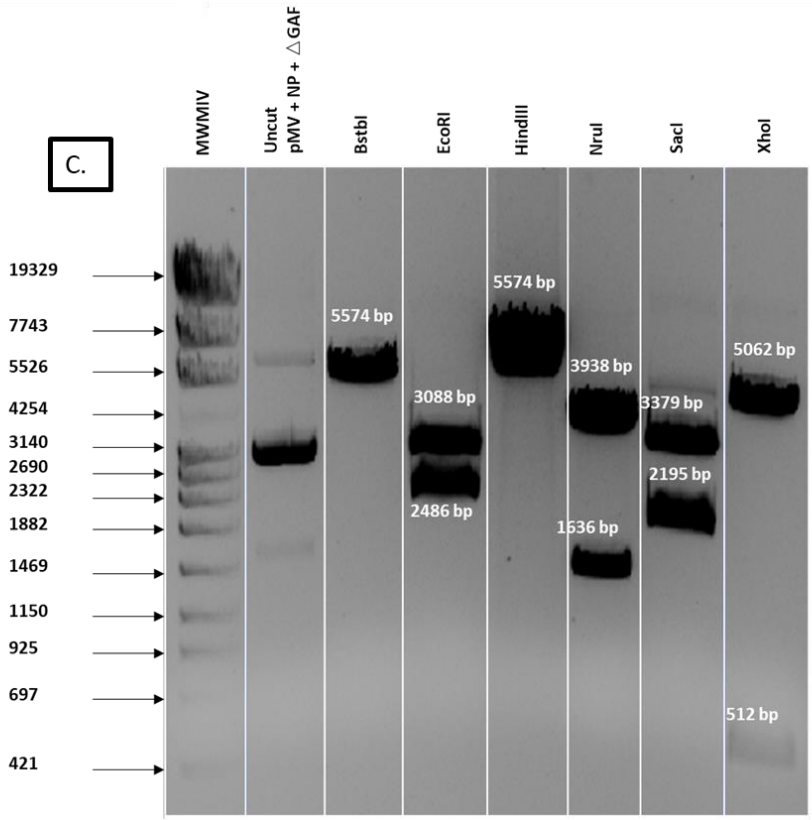


Figure 3.18: Confirmation of the pMV + NP + ΔGAF clone. *A: pMV + NP + ΔGAF plasmid map indicating the position of the neiI promoter, altered* neiI *gene and various restriction sites. B: Table displaying restriction enzymes used for the confirmation of pMV + ΔGAF + NI and the expected fragment sizes they each produce. C: Agarose gel showing the correct fragment sizes of the digested pMV + ΔGAF + NI clone. BstBI and HindIII produced a single fragment of 5574 bp. EcoRI yields fragments of 3088 bp and 2486 bp. Digestion with NruI resulted in the expected fragments of 3938 bp and 1636 bp. The SacI digest displayed band sizes of 3379 bp and 2195 bp. XhoI produced 5062 bp and 512 bp fragments. MWM IV indicates molecular weight marker IV.*

3.5.3 Electroporation of pMV306H clones into the double deletion mutant

As both the constructs pMV + NP + NI and pMV + NP + ΔGAF were correct they were electroporated into competent cells of the previously constructed *ΔnIΔnth M. smegmatis* strain to generate the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains. The electroporated cells were plated on 7H10 plates supplemented with hygromycin (see section 2.2.2.1). Approximately 100 colonies were observed for the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains, each. Genomic DNA (gDNA) from one clone for each of these strains together with the parental strain mc^2 155 was extracted by the CTAB method (see section 2.3.1.2) to ascertain whether the strains integrated the respective *neiI* gene. The gDNA was screened by PCR (see section 2.3.1.4) using primers listed in Appendix C for the *neiI* gene which should be comparable to the native gene from the wild- type strain. The expected fragment size of 1355 bp was observed for DNA from all three strains indicating the integration of the pMV + NP + NI and pMV + NP + ΔGAF constructs into the *ΔnIΔnth* strain (Figure 3.19).

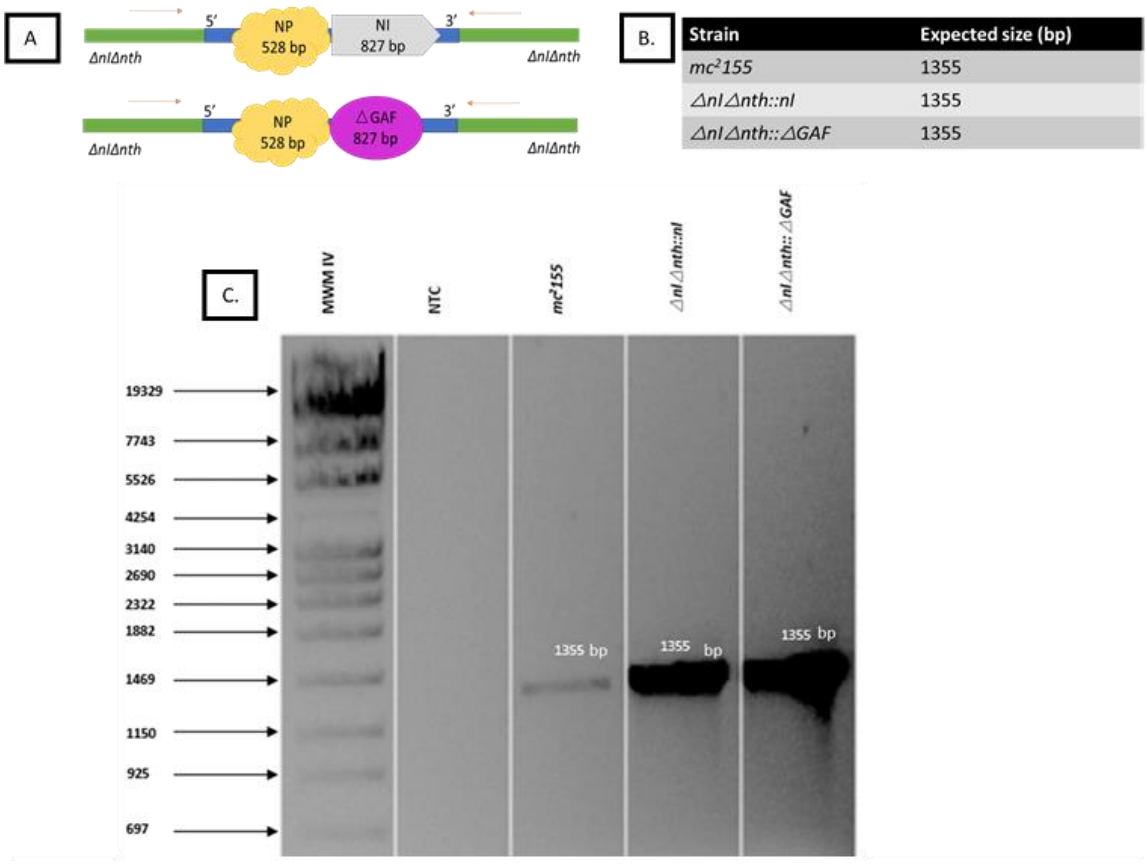


Figure 3.19: Confirmation of* ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF.** ***A: *Schematic representation of the primers used in the PCR confirmation of the NP + NI and NP + ΔGAF fragments after integration into the* ΔnIΔnth *strain. The primers used are represented in orange arrows. B: Table displaying the strains from which gDNA was extracted and used for PCR and the expected fragment sizes C: Gel electrophoresis image showing the 1355 bp fragment isolated from both the* ΔnIΔnth::nI and ΔnIΔnth::ΔGAF strains. *NTC represents the no- template negative control whilst* mc^{2}155 *acts as the positive control. MWMIV represents molecular weight marker IV.*

3.6 Gene expression of neiI in ΔnIΔnth::nI and ΔnIΔnth::ΔGAF

Subsequent to the construction of the integrating strains *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF*, it was important to assess if the native and altered *neiI* gene was expressed off the cloned promoter. RNA was extracted from mc^{2}155, *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* (see section 2.5.1) and purified with a butanol and diethylether mixture (see section 2.5.2). Thereafter, the RNA was cleared of any contaminating gDNA (see section 2.5.3) which aided in the synthesis of cDNA and successive quantification of gene expression (see sections 2.5.4 and 2.5.5).

3.6.1 Quality of the extracted and purified RNA

The extracted and purified RNA was assessed for concentration and purity using the A260/A280 and A260/A230 ratios on the nanodrop. For a sample to be devoid of contaminants, a A260/A280 ratio of approximately 2.1 is desirable whilst an acceptable A260/A230 ratio ranges between 2.0 and 2.3. In table 3.3 the results of three biological replicate extractions of RNA from the three strains is shown where orange represents experiment 1, grey represents experiment 2 and blue represents experiment 3.

Table 3.3: Concentration and purity of the extracted RNA samples

Strain	**Concentration (ng/ µL)**			**A260/A280**			**A260/A230**		
mc^{2}155	970.1	1467.1	2971.5	2.17	2.16	2.15	2.07	2.06	2.09
ΔnIΔnth::nI	477.7	510.6	482.6	2.14	2.17	2.16	1.99	2.20	2.08
ΔnIΔnth::ΔGAF	273.6	519.2	992.1	2.16	2.17	2.17	2.26	2.06	2.27

The concentration of mc^{2}155 ranged from 970.1 ng/ uL to 2971.5 ng/ uL, with an A260/A280 range from 2.15 to 2.17 and an A260/ A230 range from 2.06 to 2.09. In addition, the *ΔnIΔnth::nI* strain displayed a concentration that ranged between 477.7 ng/ uL and 510.6 ng/ uL with an A260/ A280 of between 2.14 and 2.17 and an A260/ A230 which ranged from 1.99 to 2.20. Similarly, the concentration of the *ΔnIΔnth::ΔGAF* strain ranged between 273.6 ng/ uL and 992.1 ng/ uL and displayed an A260/ A280 of between 2.16 and 2.17. The A260/ A230 of *ΔnIΔnth::ΔGAF* averaged between 2.06 and 2.27.

The RNA concentration extracted from all three strains in triplicate was of sufficient concentration and purity for downstream applications. The RNA was used in the synthesis of cDNA required for gene expression analyses.

3.6.2 Gene expression analysis

Real- time quantitative PCR (RT- qPCR) allows for assessing the change in the expression of a particular gene through the measurement of the amount of the gene- specific transcript as it is happening. The SYBR® Green dye used in RT- qPCR acts as a fluorescent reporter molecule that quantifies the amount of PCR product by binding to double- stranded DNA (dsDNA). As the quantity of the target amplicons increase, there is a directly proportional increase in the amount of fluorescence emitted from the reporter molecule.

Expression of the *neiI* gene in the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains was assessed and compared to expression of the *neiI* gene in mc^2155 parental strain. A linear standard curve based on ten- fold serial dilutions of the mc^2155 gDNA was set up in order to determine the absolute amount of RNA in the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* samples (see section 2.5.5). The absolute transcript numbers were thereafter normalised to the number of *sigA* transcripts in the same sample (see appendix G). In table 3.4 the results of three biological replicate normalizations from data of the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains was compared to the wild-type strain where yellow represents experiment 1, orange represents experiment 2 and grey represents experiment 3.

Table 3.4: Normalization of neiI *transcripts to* sigA *transcripts*

Strain/gene	**Normalization value**		
mc^2155	0.02775119617	0.000001833751254	0.03693040991
neiI	0.0003415274714	0	0
ΔGAF	0.08387729624	0.01556319862	0.02044370029

The RNA from the three strains was analysed in triplicate however, the RNA from the second biological replicate of the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains did not synthesize the cDNA correctly which affected the normalisation process. A possible explanation for this could be that the RNA samples were damaged or degraded due to multiple thawing processes. Despite, this technical problem, expression of *neiI* is evident in both the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains (Table 3.5). Although the expression in *ΔnIΔnth::nI* is comparatively lower than that of the wildtype and *ΔnIΔnth::ΔGAF* strains, the statistical difference was insignificant (p=0.4) as assessed by the two- way ANOVA test (see appendix G), suggesting that NeiI was expressed in both these strains (Figure 3.20).

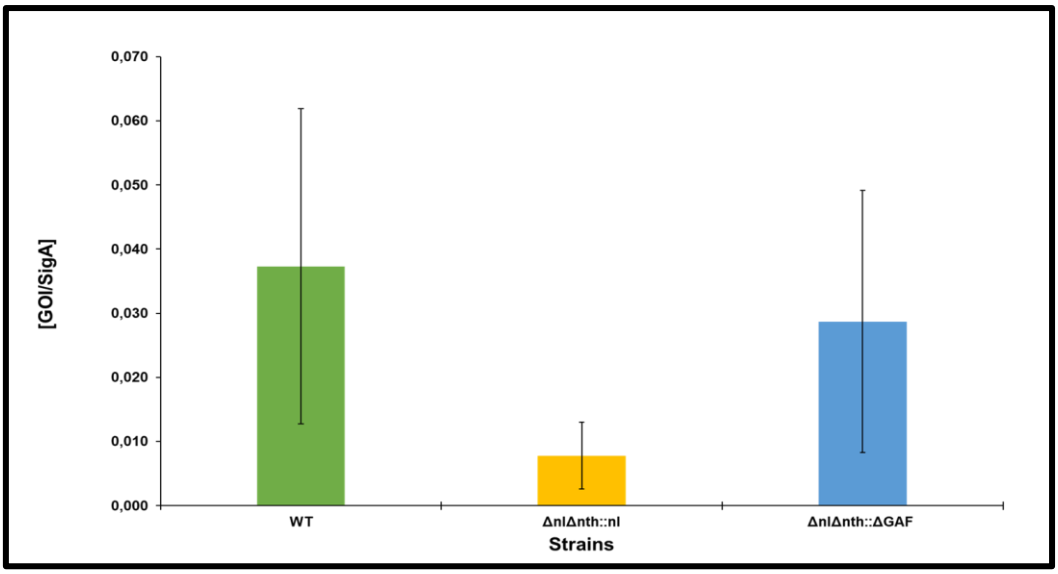


Figure 3.20: RT- PCR analysis to assess the expression of the* neiI and ΔGAF *genes in the newly constructed* ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF *strains. *The* neiI *and* ΔGAF *genes were normalised to the* sigA *gene in order to assess the levels of gene expression. The wildtype control (green) was used for comparison of the* ΔnIΔnth::nI *(yellow) and* ΔnIΔnth::ΔGAF *(blue) strains. There is no significant difference in the expression of* neiI *between any of the strains (p= 0.4) where significance was considered at p<0.05.*

Table 3.5: Average expression and standard error of mean of ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF

Strain	**Average expression**	**Standard error of mean**
mc²155	0.037	0.02458506063
neiI	0.008	0.005187427277
ΔGAF	0.029	0.02044370029

3.7 Phenotypic Characterisation

3.7.1 Growth kinetics

Growth kinetics of the mc²155, double deletion mutant (*ΔnIΔnth*) and the complemented strains (*ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF*) was carried out under axenic culture conditions in order to assess if integration of the *neiI* and altered *neiI* gene influenced the growth kinetics of the complemented strains. Log phase cultures of the mc²155, *ΔnIΔnth, ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains were inoculated in 7H9 media and the bacterial growth measured by optical density (OD_{600nm}) (Figure 3.21A) and viable cell counts (CFU/mL) (Figure 3.21B) at 3-hour intervals over a total period of 24 hours (see appendix H).

Both the optical density measurements as well as the viable cell counts showed comparable growth for the parental strain, the double deletion mutant and the complemented strains indicating that the integration of the *neiI* and altered *neiI* genes do not impact the growth kinetics of these strains (Figure 3.21).

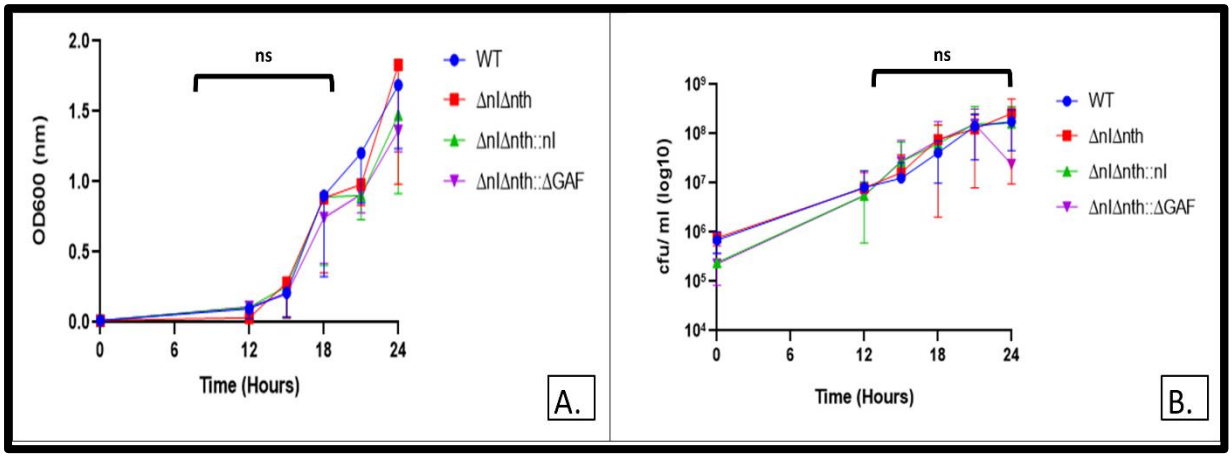


Figure 3.21: Growth kinetics of the parental, mutant and complemented strains under normal growth conditions. *Growth of mc2155, ΔnIΔnth, ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF strains was monitored over a period of 24 hours by (A) Optical density measurements showed no significant difference in the growth pattern for the ΔnIΔnth, ΔnIΔnth::nI and ΔnIΔnth::ΔGAF strains compared to the WT strain (P>0.05) (B) Viable cell count. Measurements were taken every three hours with the data represented as the mean of three separate biological assays.The CFU/ mL data indicated no significant difference in the growth rate of all of the strains (p>0.05). ns represents the p- value for non- significance.*

The optical density measurements and the cell counts indicated no significant variation between the *ΔnIΔnth, ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains compared to the WT strain (P>0.05). This indicated that the integration of the *neiI* and altered *neiI* genes do not impact the growth kinetics of these strains, as described above.

3.7.2 Evaluation of mutation rates by the fluctuation assay

The mutation rate measures the probability of a mutational event occurring during a cell's lifetime which can be measured with the use of fluctuation assays. These assays analyse the range of mutation frequencies over a series of parallel cultures. The mutation rates are calculated by assessing the total population of cells (*Nt*) and the distribution of mutations (*m*) in these parallel cultures (Rosche and Foster, 2000), as described in section 2.6.2.

In order to investigate the role of the NeiI endonuclease during mutagenesis, the spontaneous mutation rates to rifampicin (RIF) of the double deletion mutant and complemented strains were assessed by the fluctuation assay and compared to the parental strain.

Statistics can be used to make comparisons between the different strains and biological replicates, allowing for a degree of confidence to be given to these comparisons (such as

standard deviations). However, it is not possible to accurately calculate the differences between mutation rates with a degree of statistical confidence, due to the sensitivity of the mutation rate to the Nt value (Rosche and Foster, 2000). Thus, a difference in mutation rates was considered when these rates were extremely dissimilar (>10 fold). In addition, the data were analysed using three methods including the P_0 method, the Jones method and the Koch method (Table 3.6, Table 3.7 and Table 3.8 respectively) to ensure that all statistical criteria were met.

The fluctuation assay was carried out in triplicate for all four strains to assess whether the GLY loop of the *neiI* gene is responsible for DNA glycosylase activity during DNA repair. The data was firstly analysed by the P_0 method (Table 3.6) (Rosche and Foster, 2000). The P_0 method only depends on the experimental value zero, and thus is unaffected if mutants grow slower than non- mutants provided at least one of each mutant clone produces a colony following selection. The data shown in Table 3.6 show that the *ΔnIΔnth* double deletion mutant displayed increased mutagenesis compared to the parental strain in experiments 1 and 3. The *ΔnIΔnth::nI* complemented strain which reverted to the *Δnth* strain with the integration of the *neiI* gene showed increased mutagenesis in all three experiments. This deviation could be due to the integrated *neiI* not expressing at wild type levels (Figure 3.20) and therefore, is unable to compensate for the loss of the *neiI* gene. However, the low expression levels of *neiI* may have also been due to technical problems of the RNA mentioned previously, rather than an inability to be expressed at wild type levels. In the case of the *ΔnIΔnth::ΔGAF* mutant the mutation rate was expected to maintain levels observed for the *ΔnIΔnth* double deletion mutant as the GAF residues were hypothesized not to be involved in glycosylase activity and therefore, will not compensate for the loss of the *neiI* gene. However, in two of the three experiments this strain showed reduced mutation rates that were more comparable to the parental strain.

Table 3.6: Spontaneous rifampicin resistance mutation rates of the ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF *strains compared to the parental and double deletion mutant strains as determined by fluctuation analysis and assessed using the P_0 method.*

	Mutation rate			**Fold change**		
Strain	Experiment 1	Experiment 2	Experiment 3	Experiment 1	Experiment 2	Experiment 3
mc²155	1,90E-10	4,50E-12	6,20E-11	1,00	1,00	1,00
ΔnIΔnth	2,20E-09	6,80E-12	6,80E-10	11,58	1,51	10,97
ΔnIΔnth::nI	5,30E-09	7,20E-11	3,50E-10	27,89	16,00	5,65
ΔnIΔnth::ΔGAF	2,00E-10	1,30E-11	8,00E-10	1,05	2,89	12,90

Using the P_0 method, experiment 1 indicated an increase rate of spontaneous mutations for the *ΔnIΔnth* and *ΔnIΔnth::nI* strains, with only a very small increase for that of the *ΔnIΔnth::ΔGAF* strain. In contrast, experiment 2 displayed an increase in spontaneous mutations for the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains, whilst *ΔnIΔnth* displayed only a slight increase. Experiment 3 yielded increased spontaneous mutation rates for all three strains, when compared to the wildtype strain.

Similarly, the fluctuation data were analysed using the Jones and Koch methods (Tables 3.7 and 3.8). The Jones Median Estimator readily accommodates dilutions in its formulae and Koch's Quartiles method takes into account the upper and lower quartiles of the data, rather than just the mean (Rosche and Foster, 2000). However, further interrogation of these mutants by fluctuation analysis or other mutagenesis assay is required to better understand the role of the GLY loop in glycosylase activity during DNA repair.

Table 3.7: Spontaneous rifampicin resistance mutation rates of the ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF *strains compared to the* M. smegmatis *double deletion mutant and parental strains as determined by fluctuation analysis and assessed using the Jones method.*

	Mutation rate			**Fold change**		
Strain	Experiment 1	Experiment 2	Experiment 3	Experiment 1	Experiment 2	Experiment 3
mc²155	2,83E-09	5,58E-11	3,36E-10	1,00	1,00	1,00
ΔnIΔnth	2,10E-08	5,68E-11	3,45E-09	7,42	1,02	10,27
ΔnIΔnth::nI	1,73E-08	1,07E-09	2,27E-09	6,11	19,18	6,76
ΔnIΔnth::ΔGAF	8,52E-10	1,40E-10	3,10E-09	0,30	2,51	9,23

When using the Jones method, the rate of spontaneous mutations substantially increased for the *ΔnIΔnth::nI* strain in all three experiments. However, spontaneous mutations increased for the *ΔnIΔnth* strain in experiments 1 and 3, but only slightly in experiment 2. Similarly, and increase in spontaneous mutations was observed in experiments 2 and 3 for the *ΔnIΔnth::ΔGAF* strain, but not in experiment 1.

Table 3.8: Spontaneous rifampicin resistance mutation rates of the ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF *strains compared to the* M. smegmatis *double deletion mutant and parental strains as determined by fluctuation analysis and assessed using the Koch method.*

	Mutation rate			**Fold change**		
Strain	Experiment 1	Experiment 2	Experiment 3	Experiment 1	Experiment 2	Experiment 3
mc²155	2,70E-09	4,30E-11	1,20E-10	1,00	1,00	1,00
ΔnIΔnth	1,50E-08	2,70E-11	1,20E-09	5,56	0,63	10,00
ΔnIΔnth::nI	8,50E-09	1,00E-09	1,00E-09	3,15	23,26	8,33
ΔnIΔnth::ΔGAF	2,20E-10	8,90E-11	2,50E-09	0,08	2,07	20,83

An increase in spontaneous mutations can be seen in experiments 1 and 3 for the *ΔnIΔnth* strain, but a decrease was observed for this strain in experiment 2. In contrast, the spontaneous mutation rates were increased in all experiments for the *ΔnIΔnth::nI* strain, whilst only experiment 2 and 3 yielded increased spontaneous mutations for the *ΔnIΔnth::ΔGAF* strain.

4. Discussion

The cellular mechanisms used by *M. tuberculosis* to repair damaged DNA are essential components of the pathogen, as it is constantly exposed to harsh conditions that may affect the integrity of its DNA. Therefore, the replication of *M. tuberculosis* and its subsequent translation is directly dependent on its ability to successfully repair damaged DNA. Within the host macrophage, there are a variety of DNA stressors that could potentially cause damage to the *M. tuberculosis* pathogen. It is therefore important to understand the mechanisms that are required for the repair of the damaged *M. tuberculosis* DNA and ultimately its role in maintaining genome integrity under these conditions.

In a study conducted by Moolla *et al*, the importance of DNA repair in the survival of *M. smegmatis* under conditions of oxidative stress, UV exposure and during mutagenesis was reported (Moolla *et al*., 2014). In this study, although the loss of *nth, neiI* and *neiII* genes individually or in combination did not cause atypical growth under axenic conditions, deletion of the single *nth* gene in *M. smegmatis* resulted in a significant increase in the number of spontaneous mutations in the presence of RIF compared to the mutant lacking both the *nei* homologues. Furthermore, combinatorial deletion of *nth* together with the *nei* homologues resulted in an exacerbated mutation rate to RIF suggesting an anti-mutator role for Nth endonuclease and synergy between the Nth and Nei glycosylases (Moolla *et al*., 2014).

Based on these findings it was important to establish the hierarchical role of the NeiI and NeiII glycosylases under oxidative stress conditions (Rantsi, 2017). The Nei homologues were deleted individually in the *Δnth* background to generate the mutant strains *ΔnthΔneiI* and *ΔnthΔneiII* (Rantsi 2017, unpublished). Phenotypic analysis of these mutants showed that the combinatorial loss of the *nth* and *neiI* genes had a larger impact on *M. smegmatis* survival under oxidative stress compared to when the *nth* and *neiII* genes were deleted (Rantsi, 2017, unpublished). These results provided a basis for further interrogation of the two Nei homologues to better understand their differential role in glycosylase activity during DNA repair.

Bioinformatic analyses of the NeiI and NeiII proteins show a difference in three amino acids in the active site of their intercalation loop. The *M. smegmatis* NeiI enzyme contains GLY amino acid residues whilst the NeiII DNA glycosylase has KME residues. The residues in the intercalation loop are catalytically significant for DNA binding and base flipping; a mechanism in which a nucleotide is flipped outside the DNA double helix allowing access to the DNA

glycosylase for excision of the damaged base with subsequent repair of the lesion by the BER pathway (Roberts and Cheng, 1998). A study conducted in *E. coli* indicated that a mutation in the intercalation loop of the Nei glycosylase, represented by QLY residues, caused a 100-fold decrease in DNA binding efficiency of the mutant compared to the parental strain (Kropachev *et al*., 2006). A decreased binding efficiency is indicative of decreased glycosylase activity which may result in a lowered genome stability.

Thus, it was important to understand whether the GLY residues of the NeiI endonuclease play a role in genome stability and maintenance in mycobacteria. This knowledge will add to our understanding the role of the NeiI DNA glycosylase in DNA binding during repair.

4.1 DNA glycosylases structure and function

DNA glycosylases of the BER pathway act as the first enzymes to recognise and excise damaged nucleotides. The NeiI endonuclease consists of three structural domains: a catalytic core domain located at the N- terminus, an H2TH domain as well as a C- terminal zinc finger domain (Fromme *et al.,* 2004). The intercalation loop of the NeiI endonuclease is responsible for the catalysis of binding damaged DNA to the DNA glycosylases (Fortini *et al*., 1999). Bioinformatic analysis indicated that replacement of the GLY amino acid residues in the intercalation loop of the NeiI endonuclease with GAF residues does not alter the structural function or cause degradation of the NeiI protein due to structural and chemical similarities of the amino acid residues. In *E. coli*, mutating the QLY residues, with a methyl side chain of alanine, caused a 100 fold decrease in its binding efficiency (Kropachev *et al*., 2006). Thus, to determine the role of the GLY loop of the endonuclease VIII DNA glycosylase, the functional *neiI* gene encoding the NeiI glycosylase with GLY amino acids, as well as a mutated *neiI* gene encoding the NeiI glycosylase with GAF amino acids was integrated into the *ΔneiIΔnth* mutant and phenotypically assessed together with the double deletion mutant and parental strains.

4.2 The newly constructed ΔnIΔnth::nI *and* ΔnIΔnth::ΔGAF *strains are viable*

Gene expression is described as the process by which information from a particular gene is used to synthesise a functional gene product (Taft *et al*., 2010). A SYBR green dye was used to detect PCR products that accumulated during the PCR cycles. The SYBR green dye binds to double- stranded DNA which then causes it to fluoresce. As the DNA polymerase amplifies the target sequence, more amplicons are created. Thus, the fluorescence of the SYBR green increases in proportion to the amount of PCR product (Wittwer *et al*., 1997). The synthesised products are mainly proteins, but in genes that do not code proteins, such as that of rRNA genes

or tRNA genes, the product is a housekeeping or structural RNA (Taft *et al*., 2010). Evaluation of gene expression of the *neiI* and altered *neiI* genes in the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains as compared to that in the parental strain yields no significant difference in expression. However, comparatively, the *ΔnIΔnth::nI* strains presents with a lower level of *neiI* gene expression compared to the wildtype and *ΔnIΔnth::ΔGAF* strains. Although this difference is not statistically significant, it may have implications in downstream applications. Technical errors in the conduction of the experiment resulted in *SigA* deviating from its normal role as a housekeeping gene. This produced a large variation in the results which impacts their interpretation. However, when comparing the mean expression across the mc^2155, *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains, there is no significant difference. Thus, the *neiI* and altered *neiI* genes in the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains were deemed to express in the double deletion mutant. These new strains were phenotypically assessed alongside the previously constructed *ΔnIΔnth* strain (Rantsi, 2017, unpublished).

4.3 The NeiI glycosylase is not essential for growth under axenic culture conditions

Comparison of the growth kinetics of the parental, *ΔnIΔnth, ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains showed no differences. The results of the growth of the parental and *ΔnIΔnth* strains are in alignment with previously published data wherein the double *ΔnIΔnth* deletion mutants of *M. smegmatis* displayed no differences in growth under axenic culture conditions (Moolla et al., 2014; Rantsi, 2017). Furthermore, complementation of the *ΔnIΔnth* strain with the *neiI* and altered *neiI* genes to produce the *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains, respectively, did not display growth differences when compared to the parental and double deletion mutant strains. Therefore, it is clear that the NeiI glycosylases are non- essential in the survival of *M. smegmatis*. Importantly, in various other bacterial species such as *Haemophilus influenza,* no *nei* genes encoding the Nei endonuclease are present (Tomb *et al*., 1997) which strongly suggests that the retention of the two Nei DNA glycosylases in mycobacteria is a purposeful evolutionary event that aids in the maintenance of genome integrity under stressful conditions.

4.4 Substitution of the GLY residues for GAF residues of the NeiI endonuclease does not significantly impact the genome integrity

Spontaneous mutations are those mutations that arise without the pressure of exogenous agents (Foster, 2006). These mutations occur because of errors made by DNA polymerases during DNA replication, DNA repair, errors during recombination or spontaneously arising as part of cellular metabolism (Foster, 2006).

To further understand the role of the GLY loop of the NeiI endonuclease in the BER pathway, the spontaneous mutation rate against RIF for the double mutant strain as well as the newly constructed complemented strains was measured against the wildtype strain using the Luria-Delbrück fluctuation analysis. This method of measuring the mutation rate takes into account the number of mutational events that occur in a cell during its lifetime (Rosche and Foster, 2000). The Luria-Delbrück fluctuation analysis excludes pre-existing mutations that would result in "jackpot" events that lead to an over estimation of the number of mutations. The fluctuation assays were set- up by inoculating a small number of bacterial cells into a large number of parallel cultures. These cultures were then allowed to grow to saturation and each culture was plated on selective medium which allowed the mutant to form colonies. The total number of cells in each population was determined by plating dilutions of a few cultures on non- selective medium. The distribution of the mutant numbers obtained from the parallel cultures was used to calculate the mutation rate. The P_0 method (Luria and Delbrück, 1943), the Jones median estimator (Jones, 1993) and the Koch method (Koch, 1982) were all used to calculate the mutation rates of three different fluctuation assays.

Using all three of these methods, an increase in spontaneous mutations in the presence of RIF was observed for the *ΔnIΔnth, ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains relative to the wildtype strain. The 5.56- 11.58-fold increase in mutation rate of *ΔnIΔnth* correlates to previously observed mutation phenotypes wherein a 6- 8 fold increase in mutation rate of *ΔnIΔnth* has been observed (Rantsi, 2017). The *ΔnIΔnth::nI* and *ΔnIΔnth::ΔGAF* strains displayed a 3.15 – 27.89- fold and 2.07- 20.83- fold increase in spontaneous rates, respectively. The increase in spontaneous mutation rate of the *ΔnIΔnth::nI* could be due to the lower expression of the *neiI* gene in this strain, as observed in the RT- qPCR analysis. However, this low expression may have been due to the technical errors experienced during the RT-qPCR as a result of poor RNA, and therefore, may be an inaccurate representation of the actual result. If the expression level was similar to the wildtype and *ΔnIΔnth::ΔGAF* strains, the increased mutation of *ΔnIΔnth::nI* compared to *ΔnIΔnth::ΔGAF* strain may have been due to the role played by the GLY loop of the *neiI* endonuclease. Due to the increase in spontaneous mutation rates observed in *ΔnIΔnth::ΔGAF* during experiment 3 according to the P_0 method, Jones method and Koch method, the GLY loop of the *neiI* endonuclease could potentially be responsible for the glycosylase activity during DNA repair. This observation would also align with previous studies that have indicated a greater role for the NeiI endonuclease compared to the NeiII endonuclease in the maintenance of genome integrity (Rantsi, 2017). However, further

interrogation by fluctuation analysis or other mutagenesis assay is required to confirm the role of the GLY loop in glycosylase activity. In relation to the *ΔnIΔnth::ΔGAF* mutant, the observed mutation rates were similar to that of the parental strain which suggests that the alteration of the *neiI* gene does not hinder the complementation of the double deletion mutant. These indicate that the amino acids of the intercalation loop of the NeiI protein required for glycosylase activity may be redundant and lack specificity.

4.5 Conclusion

In conclusion, this study provided some insights into the role of the GLY residues of the DNA intercalation loop of the NeiI endonuclease in the BER pathway in maintaining genome integrity during repair of damaged DNA bases. Based on the data in the study, the NeiI glycosylases are non- essential in the survival of *M. smegmatis* as indicated by the similarities in growth kinetics under normal culture conditions. The preliminary mutagenesis data suggest that the GLY residues in the NeiI DNA glycosylase intercalation loop may play a role in maintaining genome stability but further in-depth analysis is required to confirm these observations.

4.6 Limitations of the study

In the gene expression analysis, the statistics indicated no significant difference, whilst the actual values indicated a much greater expression by mc^2155 and ΔGAF compared to *neiI.* The standard deviations are large which would suggest technical error. The biological assays should be repeated in future work, so that RNA can be extracted again. The assays used in this study provide a crude measurement for the role of the GLY loop of endonuclease VIII DNA glycosylase in maintaining genome integrity in *Mycobacterium smegmatis*, based on growth kinetics and mutation rates. More advanced techniques require development in order to measure specific lesions and the DNA glycosylases associated with them, in order to assess their function precisely. This would aid our understanding of the BER pathway and its glycosylases.

4.7 Future study areas

The repetition of the gene expression using newly extracted intact RNA should be conducted in future studies for accuracy. To definitively confirm that the GLY loop of the NeiI endonuclease is responsible for the glycosylase activity during DNA repair, further interrogation by fluctuation analysis or other mutagenesis assays is required. Furthermore, a purified NeiI glycosylase protein could be used in an electrophoretic mobility shift assay to

provide further insight into the protein- nucleic acid interaction. This could enhance our understanding of the mechanisms involving the acquisition of mutations which could lead to genome instability and furthermore, the development of resistance.

5. References

Adams, L., Dinauer, M., Morgenstern, D., Krahenbuhl, J., 1997. Comparison of the roles of reactive oxygen and nitrogen intermediates in the host response to *Mycobacterium tuberculosis* using transgenic mice. Tubercle and Lung Disease 78, 237-246.

Agrawal, M., Bajaj, A., Bhatia, V., Dutt, S., 2016. Comparative study of GeneXpert with ZN stain and culture in samples of suspected pulmonary tuberculosis. Journal of Clinical and Diagnostic Research: Volume : 10 | Issue : 5 | Page : DC09 - DC12.

Alffenaar, J.-W.C., Akkerman, O.W., Kim, H.Y., Tiberi, S., Migliori, G.B., 2020. Precision and Personalized Medicine and anti–TB treatment: Is TDM feasible for programmatic use? International Journal of Infectious Diseases 92, S5-S9.

Allison, M.J., Mendoza, D., Pezzia, A., 1973. Documentation of a case of tuberculosis in pre-Columbian America. American Review of Respiratory Disease 107, 985-991.

Anderson, D.G., Kowalczykowski, S.C., 1997. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. Cell 90, 77-86.

Bergval, I.L., Klatser, P.R., Schuitema, A.R., Oskam, L., Anthony, R.M., 2007. Specific mutations in the *Mycobacterium tuberculosis* rpoB gene are associated with increased dnaE2 expression. FEMS Microbiology Letters 275, 338-343.

Bhutani, H., Singh, S., Jindal, K., 2005. Drug-drug interaction studies on first-line anti-tuberculosis drugs. Pharmaceutical Development and Technology 10, 517-524.

Bogdan, C., Röllinghoff, M., Diefenbach, A., 2000. Reactive oxygen and reactive nitrogen intermediates in innate and specific immunity. Current Opinion in Immunology 12, 64-76.

Boshoff, H.I., Reed, M.B., Barry III, C.E., Mizrahi, V., 2003. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. Cell 113, 183-193.

Boulton, S.J., Jackson, S.P., 1998. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. The EMBO Journal 17, 1819-1828.

Brooks, P.C., Movahedzadeh, F., Davis, E.O., 2001. Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*: apparent lack of correlation with LexA binding. Journal of Bacteriology 183, 4459-4467.

Burke, D.S., 1993. Of postulates and peccadilloes: Robert Koch and vaccine (tuberculin) therapy for tuberculosis. Vaccine 11, 795-804.

Cabiscol, E., Tamarit Sumalla, J., Ros Salvador, J., 2000. Oxidative stress in bacteria and protein damage by reactive oxygen species. International Microbiology, 2000, vol. 3, núm. 1, p. 3-8.

Cerutti, P.A., 1974. Excision repair of DNA base damage. Life Sciences 15, 1567-1575.

Cheng, K.C., Cahill, D.S., Kasai, H., Nishimura, S., Loeb, L.A., 1992. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. Journal of Biological Chemistry 267, 166-172.

Chien, J.-Y., Chien, S.-T., Huang, S.-Y., Yu, C.-J., 2014. Safety of rifabutin replacing rifampicin in the treatment of tuberculosis: a single-centre retrospective cohort study. Journal of Antimicrobial Chemotherapy 69, 790-796.

Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S., Eiglmeier, K., Gas, S., Barry, C.r., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537-544.

Dabboussi, F., Ismail, B., Rahmo, A., Hamze, M., 2015. Contribution of katG, ahpC and inhA mutations to the detection of isoniazid-resistant Mycobacterium tuberculosis isolates from Lebanon and Syria. International Journal of Mycobacteriology 4, 45-46.

Dartois, V., 2014. The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. Nature Reviews Microbiology 12, 159-167.

David, S.S., Williams, S.D., 1998. Chemistry of glycosylases and endonucleases involved in base-excision repair. Chemical Reviews 98, 1221-1262.

Devi, K.D., Punyarani, K., Singh, N.S., Devi, H.S., 2013. An efficient protocol for total DNA extraction from the members of order Zingiberales-suitable for diverse PCR based downstream applications. SpringerPlus 2, 1-9.

Dos Vultos, T., Mestre, O., Tonjum, T., Gicquel, B., 2009. DNA repair in *Mycobacterium tuberculosis* revisited. FEMS Microbiology Reviews 33, 471-487.

Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.-C., Zawel, L., Ahn, K., Sancar, A., Reinberg, D., 1994. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. Nature 368, 769-772.

Duwat, P., de Oliveira, R., Ehrlich, S.D., Boiteux, S., 1995. Repair of oxidative DNA damage in grampositive bacteria: the *Lactococcus lactis* Fpg protein. Microbiology 141, 411-417.

Ehrt, S., Schnappinger, D., 2009. Mycobacterial survival strategies in the phagosome: defence against host stresses. Cellular Microbiology 11, 1170-1178.

Fortini, P., Parlanti, E., Sidorkina, O.M., Laval, J., Dogliotti, E., 1999. The type of DNA glycosylase determines the base excision repair pathway in mammalian cells. Journal of Biological Chemistry 274, 15230-15236.

Foster, P.L., 2006. Methods for determining spontaneous mutation rates. Methods in Enzymology 409, 195-213.

Friedberg, E.C., Gerlach, V.L., 1999. Novel DNA polymerases offer clues to the molecular basis of mutagenesis. Cell 98, 413-416.

Fromme, J.C., Banerjee, A., Huang, S.J., Verdine, G.L., 2004. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature 427, 652-656.

Fromme, J.C., Verdine, G.L., 2002. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. Nature Structural Biology 9, 544-552.

Georghiou, S.B., Magana, M., Garfein, R.S., Catanzaro, D.G., Catanzaro, A., Rodwell, T.C., 2012. Evaluation of genetic mutations associated with *Mycobacterium tuberculosis* resistance to amikacin, kanamycin and capreomycin: a systematic review. PLoS ONE 7(3): e33275. https://doi.org/10.1371/journal.pone.0033275.

Gong, C., Martins, A., Bongiorno, P., Glickman, M., Shuman, S., 2004. Biochemical and genetic analysis of the four DNA ligases of mycobacteria. Journal of Biological Chemistry 279, 20594-20606.

Gonzalez-Montaner, L., Natal, S., Yongchaiyud, P., Olliaro, P., Abbate, E., Mosca, C., Casado, G., Di Lonardo, M., Gerhart-Filho, G., Betjel, I., 1994. Rifabutin for the treatment of newly-diagnosed pulmonary tuberculosis: a multinational, randomized, comparative study versus rifampicin. Tubercle and Lung Disease 75, 341-347.

Goosens, V., 2009. The Construction and Phenotypic Characterization of Mycobacterial Mutants Deficient in DNA Glycosylases, Masters dissertation, unpublished.

Güthlein, C., Wanner, R.M., Sander, P., Davis, E.O., Bosshard, M., Jiricny, J., Böttger, E.C., Springer, B., 2009. Characterization of the mycobacterial NER system reveals novel functions of the uvrD1 helicase. Journal of Bacteriology 191, 555-562.

Iliopoulos, A., Psathakis, K., Aslanidis, S., Skagias, L., Sfikakis, P., 2006. Tuberculosis and granuloma formation in patients receiving anti-TNF therapy. The International Journal of Tuberculosis and Lung Disease 10, 588-590.

Jemikalajah, J., Okogun, G., Adu, M., Okolie, G., 2014. Evaluation of serum proteins in pulmonary tuberculosis. African Journal of Cellular Pathology 3, 20-24.

Jnawali, H.N., Hwang, S.C., Park, Y.K., Kim, H., Lee, Y.S., Chung, G.T., Choe, K.H., Ryoo, S., 2013. Characterization of mutations in multi-and extensive drug resistance among strains of *Mycobacterium tuberculosis* clinical isolates in Republic of Korea. Diagnostic Microbiology and Infectious Disease 76, 187-196.

Jones, M.E., 1993. Accounting for plating efficiency when estimating spontaneous mutation rates. Mutation Research/Environmental Mutagenesis and Related Subjects 292, 187-189.

Kanabus, A., 2017. Information about Tuberculosis: TB Statistics-Global, Regional and High Burden. Global Health Education (GHE). Volume 66, Issue 4, Pages 554–563.

Khachi, H., O'Connell, R., Ladenheim, D., Orkin, C., 2009. Pharmacokinetic interactions between rifabutin and lopinavir/ritonavir in HIV-infected patients with mycobacterial co-infection. Journal of Antimicrobial Chemotherapy 64, 871-873.

Koch, A.L., 1982. Mutation and growth rates from Luria-Delbrück fluctuation tests. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 95, 129-143.

Korycka-Machala, M., Rychta, E., Brzostek, A., Sayer, H.R., Rumijowska-Galewicz, A., Bowater, R.P., Dziadek, J., 2007. Evaluation of NAD+-dependent DNA ligase of mycobacteria as a potential target for antibiotics. Antimicrobial Agents and Chemotherapy 51, 2888-2897.

Kropachev, K.Y., Zharkov, D.O., Grollman, A.P., 2006. Catalytic mechanism of *Escherichia coli* endonuclease VIII: roles of the intercalation loop and the zinc finger. Biochemistry 45, 12039-12049.

Kurthkoti, K., Varshney, U., 2011. Base excision and nucleotide excision repair pathways in mycobacteria. Tuberculosis 91, 533-543.

Leipold, M.D., Muller, J.G., Burrows, C.J., David, S.S., 2000. Removal of hydantoin products of 8-oxoguanine oxidation by the *Escherichia coli* DNA repair enzyme, FPG. Biochemistry 39, 14984-14992.

Li, X., Heyer, W.-D., 2008. Homologous recombination in DNA repair and DNA damage tolerance. Cell Research 18, 99-113.

Liedtke, W., Battistini, L., Brosnan, C., Raine, C., 1994. A comparison of methods for RNA extraction from lymphocytes for RT-PCR. Genome Research 4, 185-187.

Lindahl, T., Wood, R.D., 1999. Quality control by DNA repair. Science 286, 1897-1905.
Loeb, L.A., 1985. Apurinic sites as mutagenic intermediates. Cell 40, 483-484.

Luria, S.E., Delbrück, M., 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28, 491.

Mak, Y., Ho, K., 1992. An improved method for the isolation of chromosomal DNA from various bacteria and cyanobacteria. Nucleic Acids Research 20, 4101.

Malone, K.M., Gordon, S.V., 2017. *Mycobacterium tuberculosis* complex members adapted to wild and domestic animals. Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control, 135-154.

Malyarchuk, S., Wright, D., Castore, R., Klepper, E., Weiss, B., Doherty, A.J., Harrison, L., 2007. Expression of Mycobacterium tuberculosis Ku and Ligase D in *Escherichia coli* results in RecA and RecB-independent DNA end-joining at regions of microhomology. DNA Repair 6, 1413-1424.

Mizrahi, V., Andersen, S.J., 1998. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? Molecular Microbiology 29, 1331-1339.

Moadebi, S., Harder, C.K., Fitzgerald, M.J., Elwood, K.R., Marra, F., 2007. Fluoroquinolones for the treatment of pulmonary tuberculosis. Drugs 67, 2077-2099.

Moolla, N., Goosens, V.J., Kana, B.D., Gordhan, B.G., 2014. The contribution of Nth and Nei DNA glycosylases to mutagenesis in *Mycobacterium smegmatis*. DNA Repair 13, 32-41.

Moore, P.D., Song, K.-Y., Chekuri, L., Wallace, L., Kucherlapati, R.S., 1986. Homologous recombination in a Chinese hamster X-ray-sensitive mutant. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 160, 149-155.

Nardell, E.A., 2004. Catching droplet nuclei: toward a better understanding of tuberculosis transmission. American Thoracic Society. Volume 169, Issue 5, Page 553.

O'sullivan, D.M., McHugh, T.D., Gillespie, S.H., 2005. Analysis of *rpoB* and *pncA* mutations in the published literature: an insight into the role of oxidative stress in *Mycobacterium tuberculosis* evolution? Journal of Antimicrobial Chemotherapy 55, 674-679.

Padarath, K., 2017. The Role of *Nth* and *Nei* homologues from base excision repair(BER) pathway in *M. smegmatis*, Honours dissertation, unpublished.

Pawlowski, A., Jansson, M., Sköld, M., Rottenberg, M.E., Källenius, G., 2012. Tuberculosis and HIV co-infection. PLoS Pathogens 8, e1002464.

Peloquin, C.A., 2002. Therapeutic drug monitoring in the treatment of tuberculosis. Drugs 62, 2169-2183.

Posnick, L.M., Samson, L.D., 1999. Imbalanced Base Excision Repair Increases Spontaneous Mutation and Alkylation Sensitivity in *Escherichia coli*. Journal of Bacteriology 181, 6763-6771.

Puri, R.V., Singh, N., Gupta, R.K., Tyagi, A.K., 2013. Endonuclease IV Is the major apurinic/apyrimidinic endonuclease in *Mycobacterium tuberculosis* and is important for protection against oxidative damage. PloS one 8, e71535.

Purmal, A.A., Kow, Y.W., Wallace, S.S., 1994. 5-Hydroxypyrimidine deoxynucleoside triphosphates are more efficiently incorporated into DNA by exonuclease-free Klenow fragment than 8-oxopurine deoxynucleoside triphosphates. Nucleic Acids Research 22, 3930-3935.

Raizada, N., Sachdeva, K., Chauhan, D., Malhotra, B., Reddy, K., Dave, P., Mundade, Y., Patel, P., Ramachandran, R., Das, R., 2014. A multi-site validation in India of the line probe assay for the rapid diagnosis of multi-drug resistant tuberculosis directly from sputum specimens. PloS one 9, e88626.

Ramaswamy, S., Musser, J.M., 1998. Molecular genetic basis of antimicrobial agent resistance inMycobacterium tuberculosis: 1998 update. Tubercle and Lung Disease 79, 3-29.

Rantsi, T.C., 2017. Molecular basis of the interplay between the Nth and Nei DNA glycosylases of the base excision repair pathway in Mycobacterium smegmatis, Masters dissertation, unpublished.

Rastogi, N., David, H.L., 1988. Mechanisms of pathogenicity in mycobacteria. Biochimie 70, 1101-1120.

Rendon, A., Tiberi, S., Scardigli, A., D'Ambrosio, L., Centis, R., Caminero, J.A., Migliori, G.B., 2016. Classification of drugs to treat multidrug-resistant tuberculosis (MDR-TB): evidence and perspectives. Journal of Thoracic Disease 8, 2666.

Roberts, R.J., Cheng, X., 1998. Base flipping. Annual Review of Biochemistry 1998 67:1, 181-198

Rock, J.M., Lang, U.F., Chase, M.R., Ford, C.B., Gerrick, E.R., Gawande, R., Coscolla, M., Gagneux, S., Fortune, S.M., Lamers, M.H., 2015. DNA replication fidelity in *Mycobacterium tuberculosis* is mediated by an ancestral prokaryotic proofreader. Nature Genetics 47, 677-681.

Rosche, W.A., Foster, P.L., 2000. Determining mutation rates in bacterial populations. Methods 20, 4-17.

Russell, D.G., Barry, C.E., Flynn, J.L., 2010. Tuberculosis: what we don't know can, and does, hurt us. Science 328, 852-856.

Sager, P., Schalimtzek, M., Moller-Christensen, V., 1972. A case of tuberculosis spondylosa in the Danish neolithic age. Dan Med Bull 19, 176-180.

Saikrishnan, K., Yeeles, J.T., Gilhooly, N.S., Krajewski, W.W., Dillingham, M.S., Wigley, D.B., 2012. Insights into Chi recognition from the structure of an AddAB-type helicase–nuclease complex. The EMBO Journal 31, 1568-1578.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press.

Sancar, A., Sancar, G.B., 1988. DNA repair enzymes. Annual Review of Biochemistry 57, 29-67.

Seeberg, E., Eide, L., Bjørås, M., 1995. The base excision repair pathway. Trends in Biochemical Sciences 20, 391-397.

Sharma, S., Mohan, A., 2004. Extrapulmonary tuberculosis. Indian Journal of Medical Research 120, 316-353.

Shuman, S., Glickman, M.S., 2007. Bacterial DNA repair by non-homologous end joining. Nature Reviews Microbiology 5, 852-861.

Singh, A., 2017. Guardians of the mycobacterial genome: A review on DNA repair systems in *Mycobacterium tuberculosis*. Microbiology 163, 1740-1758.

Slijepcevic, P., 2008. DNA damage response, telomere maintenance and ageing in light of the integrative model. Mechanisms of Ageing and Development 129, 11-16.

Snapper, S., Melton, R., Mustafa, S., Kieser, T., Jr, W.J., 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Molecular Microbiology 4, 1911-1919.

Srivastava, S.K., Tripathi, R.P., Ramachandran, R., 2005. NAD+-dependent DNA ligase (Rv3014c) from Mycobacterium tuberculosis crystal structure of the adenylation domain and identification of novel inhibitors. Journal of Biological Chemistry 280, 30273-30281.

Taft, R.J., Pang, K.C., Mercer, T.R., Dinger, M., Mattick, J.S., 2010. Non-coding RNAs: regulators of disease. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland 220, 126-139.

Takeuchi, M., Lillis, R., Demple, B., Takeshita, M., 1994. Interactions of *Escherichia coli* endonuclease IV and exonuclease III with abasic sites in DNA. Journal of Biological Chemistry 269, 21907-21914.

Taylor, R.G., Walker, D.C., McInnes, R., 1993. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. Nucleic Acids Research 21, 1677.

Tiemersma, E.W., van der Werf, M.J., Borgdorff, M.W., Williams, B.G., Nagelkerke, N.J., 2011. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PloS one 6, e17601.

Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388, 539-547.

Villegas, M.V., Labrada, L.A., Saravia, N.G., 2000. Evaluation of polymerase chain reaction, adenosine deaminase, and interferon-γ in pleural fluid for the differential diagnosis of pleural tuberculosis. Chest 118, 1355-1364.

Ward, J., Limoli, C., Calabro-Jones, P., Evans, J., 1987. Radiation vs chemical damage to DNA, Anticarcinogenesis and radiation protection. In: Cerutti P.A., Nygaard O.F., Simic M.G. (eds) Anticarcinogenesis and Radiation Protection. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-6462-1_48.

Warner, D.F., Mizrahi, V., 2011. Making ends meet in mycobacteria. Molecular Microbiology 79, 283-287.

Weller, G.R., Kysela, B., Roy, R., Tonkin, L.M., Scanlan, E., Della, M., Devine, S.K., Day, J.P., Wilkinson, A., di Fagagna, F.d.A., 2002. Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297, 1686-1689.

WHO, 2019. Global tuberculosis report 2019. Geneva: World Health Organization.

Wilk, M.J., 2007. Tuberculosis: the frail elderly and the risk of transmission to health care workers. Perspectives (Pre-2012) 31, 5.

Wittwer, C.T., Herrmann, M.G., Moss, A.A., Rasmussen, R.P., 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22, 130-138.

Zaw, M.T., Emran, N.A., Lin, Z., 2018. Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in Mycobacterium tuberculosis. Journal of Infection and Public Health 11, 605-610.

Zhang, Y., 2005. The magic bullets and tuberculosis drug targets. Annual Review of Pharmacology and Toxicology. 45, 529-564.

6. Appendix list

Appendix A: Culture media

E. coli culture media

Luria-Bertani broth (LB)	10 g tryptone, 10 g NaCl, 5 g yeast extract in 1 *l* dH_2O
2xTY	20 g tryptone, 10 g NaCl, 10 g yeast extract in 1 *l* dH_2O
Luria-Bertani agar (LA)	10 g tryptone, 5 g NaCl, 5 g yeast extract, 15 g agar in 1 *l* dH_2O

M. smegmatis culture media

Middlebrook 7H9	4.7 g Middlebrook 7H9, 10 mL 100 x glucose salts, 2 mL glycerol, 2 mL 25 % Tween80 in 1 *l* dH_2O
Middlebrook 7H10	19 g Middlebrook 7H10, 10 mL 100x glucose salts, 5 mL glycerol in 1 *l* dH_2O
100 x glucose salts	20 g glucose, 8.5 g NaCl dissolved in 100 mL dH_2O

Appendix B: DNA extraction solutions

Cetyltrimethylammonium bromide (CTAB)	4.1 % NaCl, 10 % N-cetyl- N, N, N-trimethyl ammonium bromide dissolved in dH_2O
24: 1 Chloroform: isoamyl alcohol	24 mL chloroform, 1 mL isoamyl alcohol
TE Buffer	10 mM Tris-HCl (pH 8), 10 mM EDTA dissolved in dH_2O
Phenol: chloroform	1 mL phenol, 1 mL chloroform
Sodium acetate	3 M sodium acetate dissolved in dH_2O, pH 5.2
CTAB/ Nacl	10 % CTAB made in 0.7 M NaCl

Genomic DNA extraction solutions for mycobacteria

Plasmid DNA extraction for *E. coli* solutions

Solution I	50 mM glucose, 25 mM Tris-HCl (pH 8), 10 mM EDTA in dH_2O
Solution II	1 % SDS, 0.2 M NaOH dissolved in dH2O
Solution III	3 M potassium acetate, 11.5% acetic acid dissolved in dH2O

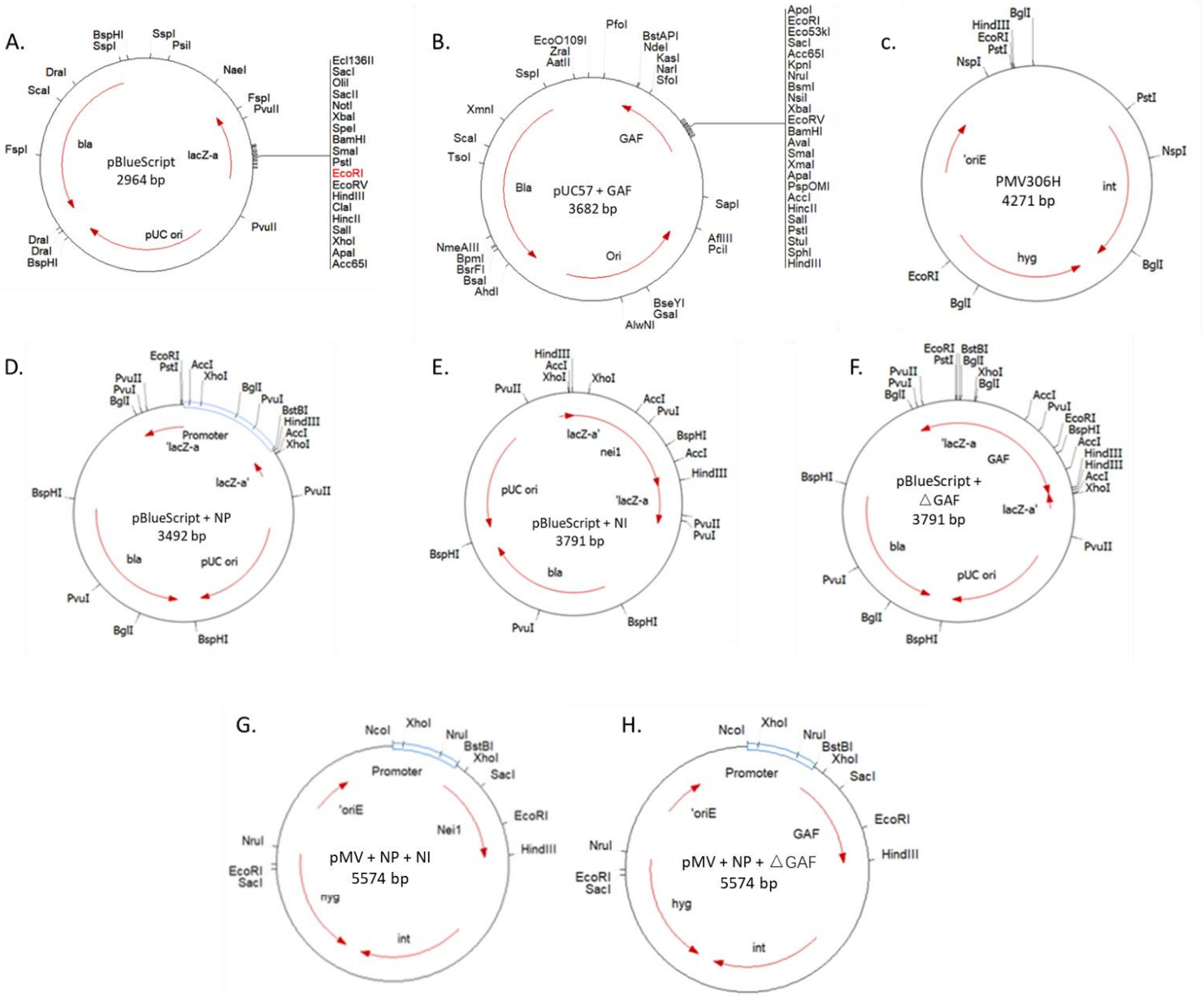
Appendix C: Table of PCR primers

Region amplified	Name of primer	Primer oligonucleotide sequence (5'- 3')	Annealing temperature (Tm) of primer (ºC)	Restriction sites
nth gene	Nth Forward Nth reverse	ggcatatgagtgcgggtgccgcccgg gcgcctgcagtcaccaaaccgccagcg c	63	N/A
neiI promoter	NP Forward NP Reverse	tagggttagcgccatgcgcgtc gcgacaccgtcctggtcaggc	67	NcoI BstBI
neiI gene	NI Forward NI Reverse	gtgcctgaggggcacaccctg tcaggactgacaggtggggca	65	BstBI HindIII
ΔGAF	ΔGAF Forward ΔGAF Reverse	gtgcctgaggggcacaccctg tcaggactgacaggtggggca	65	BstBI HindIII
*neiI*RT	RT Forward RT Reverse	acgcacctacgtctaccg agaacaggtttcgccctt	62	N/A
*ΔGAF*RT	RT Forward RT Reverse	acgcacctacgtctaccg agaacaggtttcgccctt	62	N/A
sigA gene	SigA Forward SigA Reverse	gggcgtatgtccatctcct gtatcccggtgcatggtc	62	N/A

Appendix D: Competent cells and transformation solutions

TfbI	0.588 g potassium acetate, 2.42 g rubidium chloride, 0.294 g calcium chloride, 2.0 g manganese chloride, 30 mL glycerol in dH2O. pH to 5.8 with dilute acetic acid and made up to 200 mL. Filter sterilized.
TfbII	0.21 g MOPS, 1.1 g calcium chloride, 0.121 g rubidium chloride, 15 mL glycerol in dH2O. pH to 6.5 with dilute NaOH and made up to 100 mL. Filter sterilized.
2 x TY	20g tryptone, 10g NaCl, 10 g yeast extract dissolved in dH2O to make 1*l* and autoclaved.

Appendix E: Maps of cloning vectors used and created in this study



Appendix F: Cloning strategy

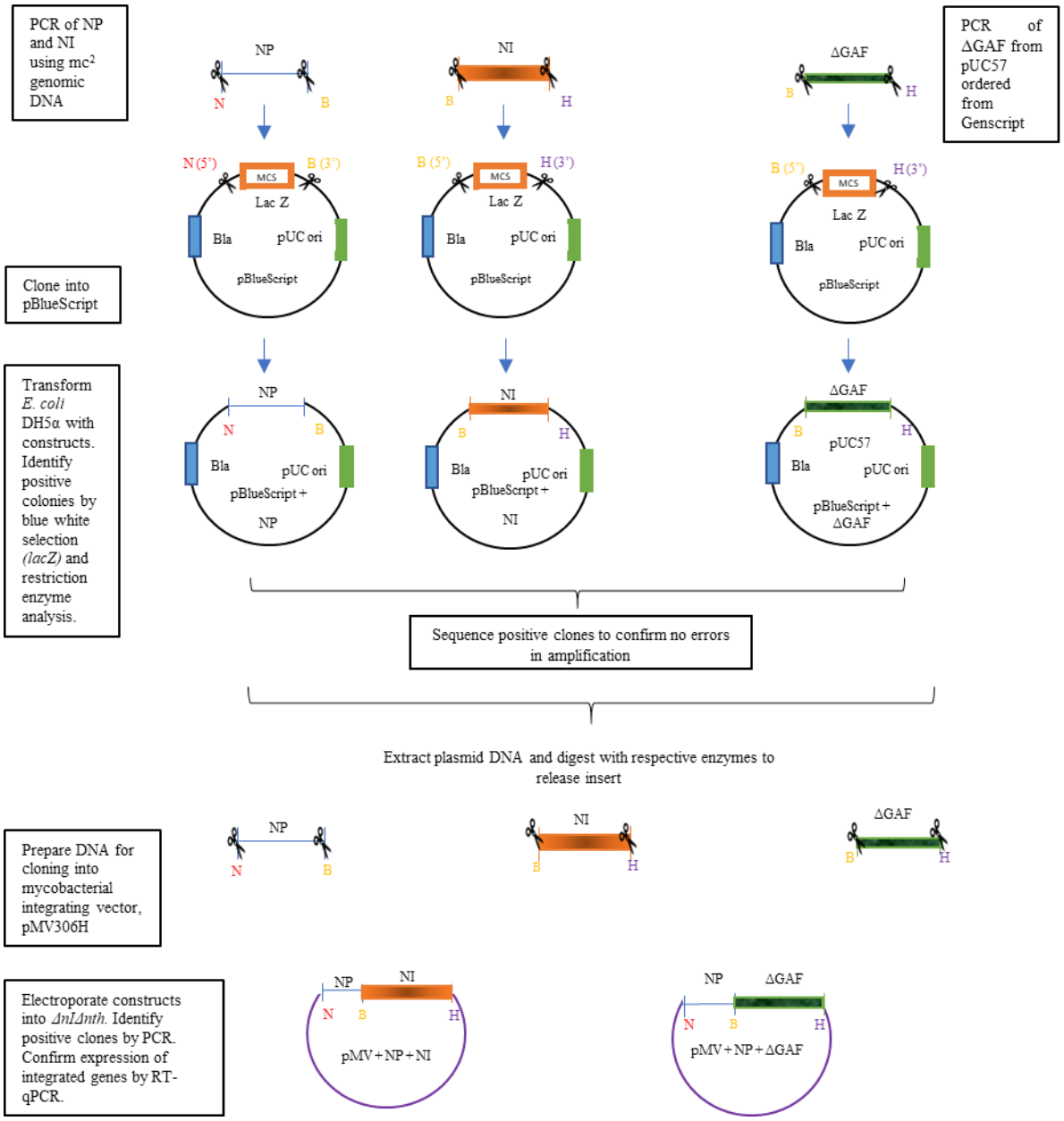


Figure F1: Cloning strategy. *The NP, NI and ΔGAF fragments were ligated into pBlueScript. Thereafter, they were excised from pBlueScript and ligated into pMV306H to generate the pMV + NP + NI and pMV + NP + ΔGAF vectors. The red "N" represents NcoI. The yellow "B" BstBI and the purple "H" HindIII restriction enzymes.*

Appendix G: Raw data from gene expression analysis

	sigA				*nei1*					
	RT+		RT-		RT+		RT-			*Normalized (nei1/sigA)*
	Ct	*Calculated Concentration*	*Ct*	*Calculated Concentration*	*Ct*	*Calculated Concentration*	*Ct*	*Calculated Concentration*	wt_1+	0,027751196
wt_1+	22,24	1,05E+06	*35,08*	*1,55E+01*	21,68	2,90E+04	*25,84*	*2,93E+02*	wt_2+	0,000341527
wt_2+	25,01	9,52E+04	*38,49*	*8,08E-01*	27,83	3,25E+01	*27,39*	*5,30E+01*	wt_3+	0,083877296
wt_3+	22,96	5,61E+05	*35,44*	*1,13E+01*	21,24	4,70E+04	*27,2*	*6,54E+01*	n1_1+	1,83375E-06
n1_1+	26,01	3,99E+04	*36,85*	*3,35E+00*	33,35	7,31E-02	*26,61*	*1,25E+02*	n1_2+	0
n1_2+	N/A	N/A	*36,49*	*4,56E+00*	N/A	N/A	*27,13*	*7,05E+01*	n1_3+	0,015563199
n1_3+	22,12	1,16E+06	*37,69*	*1,62E+00*	22,11	1,81E+04	*27,28*	*5,99E+01*	GAF_1+	0,03693041
GAF_1+	21,44	2,10E+06	*35,59*	*9,93E+00*	20,79	7,75E+04	*26,83*	*9,80E+01*	GAF_2+	0
GAF_2+	N/A	N/A	*37,56*	*1,81E+00*	N/A	N/A	*27,21*	*6,45E+01*	GAF_3+	0,0204437
GAF_3+	21,29	2,39E+06	*33,65*	*5,35E+01*	21,21	4,88E+04	*27,02*	*7,98E+01*		

Data Summary

Groups	N	Mean	Std. Dev.	Std. Error
Group 1	3	0.0373	0.0426	0.0246
Group 2	3	0.0191	0.0185	0.0107
Group 3	3	0.0052	0.009	0.0052

ANOVA Summary

Source	Degrees of Freedom DF	Sum of Squares SS	Mean Square MS	F-Stat	P-Value
Between Groups	2	0.0016	0.0008	1.0421	0.4088
Within Groups	6	0.0045	0.0007		
Total:	8	0.006			

Appendix H: Raw data from growth kinetics

H.1: Growth kinetics experiment 1

	Dilutions	mc²		Cfu/ml	ΔnlΔnth		Cfu/ml	ΔnlΔnth::nl		Cfu/ ml	ΔnlΔnth::ΔGAF		Cfu/ml
T0	10^{-2}	TNTC	TNTC	NA	TNTC	TNTC	NA	273	275	2.7×10^5	234	256	2.5×10^5
	10^{-3}	131	110	1.2×10^6	118	97	1.1×10^6	31	17	2.4×10^5	18	26	2.2×10^5
	10^{-4}	6	10	8.0×10^5	8	10	9.0×10^5	5	1	3×10^5	2	4	2.0×10^5
Cfu/ml		1.0×10^6			1.0×10^6			2.7×10^5			2.2×10^5		
OD600		0.015			0.016			0.012					
T1	10^{-2}	582	428	5.1×10^5	275	284	2.8×10^5	114	104	1.1×10^5	28	25	2.7×10^4
	10^{-3}	68	56	6.2×10^5	32	23	2.8×10^5	9	11	1.0×10^5	5	2	3.5×10^4
	10^{-4}	5	13	9.0×10^5	3	5	4.0×10^5	1	1	1.0×10^5	0	0	0
Cfu/ml		6.8×10^5			3.2×10^5			1.0×10^5			3.1×10^4		
OD600		0.025			0.021			0.024			0.024		
T2	10^{-2}	426	540	4.8×10^5	291	242	2.7×10^5	62	90	7.6×10^4	60	41	5.1×10^4
	10^{-3}	54	67	6.1×10^5	30	18	2.4×10^5	7	6	6.5×10^4	6	3	4.5×10^4
	10^{-4}	7	14	1.1×10^6	4	4	4.0×10^5	1	1	1.0×10^5	1	1	1.0×10^5
Cfu/ml		7.3×10^5			3.0×10^5			8.0×10^4			6.5×10^4		
OD600		0.039			0.033			0.030			0.031		
T3	10^{-4}	4	0	2.0×10^5	3	2	2.5×10^5	222	212	2.2×10^7	268	264	2.7×10^7
	10^{-5}	0	0	0	0	1	5.0×10^5	29	19	2.4×10^7	26	21	2.4×10^7
	10^{-6}	0	0	0	0	0	0	1	2	1.5×10^7	9	4	6.5×10^7
Cfu/ml		2.0×10^5			2.0×10^5			2.0×10^7			3.9×10^7		
OD600		0.901			0.884			0.891			0.745		
T4	10^{-5}	76	98	8.7×10^7	72	65	6.9×10^7	LAWN	LAWN	NA	LAWN	LAWN	NA
	10^{-6}	7	3	5.0×10^7	4	10	7.0×10^7	TNTC	TNTC	NA	TNTC	TNTC	NA
	10^{-7}	0	0	0	1	0	5.0×10^7	225	222	2.2×10^{10}	467	260	3.6×10^{10}
Cfu/ml		6.9×10^7			6.3×10^7			2.2×10^{10}			3.6×10^{10}		
OD600		1.204			0.981			0.902					
T5	10^{-6}	19	17	1.7×10^8	18	18	1.8×10^8	LAWN	LAWN	NA	LAWN	LAWN	NA
	10^{-7}	0	0	0	1	1	1.0×10^8	TNTC	TNTC	NA	TNTC	TNTC	NA
	10^{-8}	1	0	5.0×10^7	0	0	0	150	130	1.4×10^{11}	143	142	1.4×10^{11}
Cfu/ml		1.1×10^8			1.4×10^8			1.4×10^{11}			1.4×10^{11}		
OD600		1.235			0.982			0.915			1.214		

H.2: Growth kinetics experiment 2

	Dilutions	mc²		Cfu/ml	ΔnlΔnth		Cfu/ml	ΔnlΔnth::nl		Cfu/ ml	ΔnlΔnth::ΔGAF		Cfu/ml
T0	10^{-2}	462	304	3.83×10^5	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	73	94	8.35×10^4
	10^{-3}	76	74	7.5×10^5	68	54	6.1×10^5	32	25	2.85×10^5	4	7	5.5×10^4
	10^{-4}	12	8	1×10^6	8	11	9.5×10^5	2	1	1.5×10^5	3	0	
Cfu/ml		7.1×10^5			7.8×10^5			2.2×10^5			6.9×10^4		
OD600		0.017			0.015			0.013			0.011		
T1	10^{-3}	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC
	10^{-4}	154	122	2.76×10^7	166	158	1.62×10^7	39	44	4.15×10^6	76	66	7.1×10^6
	10^{-5}	7	12	9.5×10^6	10	27	1.85×10^7	1	27		4	4	4×10^6
	10^{-6}	1	0		0	0		2	1	1.5×10^7	1	0	
Cfu/ml		1.9×10^7			1.7×10^7			9.6×10^6			5.6×10^6		
OD600		0.553			0.488			0.159			0.110		
T2	10^{-4}	220	232	2.26×10^7	329	305	3.17×10^7	TNTC	TNTC	TNTC	TNTC	TNTC	TNTC
	10^{-5}	19	19	1.9×10^7	48	39	4.35×10^7	83	68	7.55×10^7	115	75	9.5×10^7
	10^{-6}	3	5	4×10^7	3	6	4.5×10^7	5	10	7.5×10^7	10	3	6.5×10^7
	10^{-7}	0	0		0	0		0	0		1	0	
Cfu/ml		2.7×10^7			4×10^7			7.5×10^7			8×10^7		
OD600		0.818			0.954			0.511			0.505		
T3	10^{-5}	41	72	5.65×10^7	66	65	6.55×10^7	102	107	1.05×10^8	152	105	1.29×10^8
	10^{-6}	8	6	7×10^7	8	16	1.2×10^8	28	16	2.2×10^8	26	10	1.8×10^8
	10^{-7}	1	0		1	3	2.0×10^8	0	0		2	3	2.5×10^8
Cfu/ml		6.3×10^7			1.3×10^8			1.6×10^8			1.9×10^8		
OD600		1.699			1.668			0.891			0.885		
T4	10^{-5}	152	225	1.9×10^8	209	215	2.1×10^8	277	350	3.1×10^8	327	309	3.2×10^8
	10^{-6}	17	28	2.3×10^8	36	46	4.1×10^8	34	22	2.8×10^8	17	26	2.2×10^8
	10^{-7}	5	3	4×10^8	2	1	1.5×10^8	10	6		0	0	
	10^{-8}	0	0		1	0		4	5		0	1	
Cfu/ml		2.7×10^8			2.6×10^8			3×10^8			2.7×10^8		
OD600		5.07			5.11			1.011			1.906		
T5	10^{-6}	27	55	4.1×10^8	39	31	3.5×10^8	1	14		15	1	8×10^8
	10^{-7}	2	3	2.5×10^8	11	4	7.5×10^8	1	2	3×10^8	0	0	
	10^{-8}	0	0		0	1		0	0		0	0	

H.3: Growth kinetics experiment 3

	Dilutions	mc^2		Cfu/ml	ΔnlΔnth		Cfu/ml	ΔnlΔnth::nl		Cfu/ ml	ΔnlΔnth::ΔGAF		Cfu/ml
T0	10^{-2}	209	272	2.4×10^5	268	295	2.8×10^5	187	289	2.3×10^5	27	16	2.1×10^4
	10^{-3}	39	38	3.9×10^5	28	38	3.3×10^5	36	15	2.5×10^5	8	2	5×10^4
	10^{-4}	5	4	4.5×10^5	10	9	9.5×10^5	3	1	2.0×10^5	1	0	
Cfu/ml		3.6×10^5			5.2×10^5			2.3×10^5			3.5×10^4		
OD600		0.012			0.010			0.011			0.014		
T1	10^{-3}	470	465	4.7×10^6	TNTC	TNTC		TNTC	TNTC		399	417	4.1×10^6
	10^{-4}	46	129	4.6×10^6	56	83	6.9×10^6	80	81	8.1×10^6	42	66	5.4×10^6
	10^{-5}	5	12	5×10^6	4	8	6×10^6	8	4	6×10^6	3	11	7×10^6
	10^{-6}	0	0		0	3		0	0		11	3	
Cfu/ml		4.8×10^6			6.5×10^6			7.1×10^6			5.5×10^6		
OD600		0.097			0.032			0.108			0.170		
T2	10^{-4}	97	90	9.4×10^6	94	64	7.9×10^6	112	71	9.2×10^6	36	16	2.6×10^6
	10^{-5}	15	7	11×10^6	10	5	7.5×10^6	6	2	4×10^6	0	0	
	10^{-6}	0	2		1	0	1×10^7	0	0		0	0	
Cfu/ml		1×10^7			8.4×10^6			6.6×10^6			2.6×10^6		
OD600		0.211			0.282			0.256			0.201		
T3	10^{-4}	136	141	1.4×10^7	279	208	2.4×10^7	206	250	2.3×10^7	243	221	2.3×10^7
	10^{-5}	17	20	1.9×10^7	13	32	2.3×10^7	16	23	2×10^7	18	19	1.9×10^7
	10^{-6}	2	3	2.5×10^7	2	0		0	1		0	0	
Cfu/ml		1.9×10^7			2.4×10^7			2.2×10^7			2.1×10^7		
OD600		0.325			0.352			0.406			0.420		
T4	10^{-5}	59	51	5.5×10^7	68	40	5.4×10^7	3	20	2×10^7	67	54	6.1×10^7
	10^{-6}	13	7	1×10^8	9	1	5×10^7	2	3	2.5×10^7	3	1	2×10^7
	10^{-7}	1	1	1×10^8	3	0		0	0		0	0	
Cfu/ml		8.5×10^7			5.2×10^7			2.3×10^7			4.1×10^7		
OD600		0.850			0.833			0.731			0.778		
T5	10^{-5}	106	118	1.1×10^8	137	118	1.3×10^8	33	11	2.2×10^7	25	21	2.3×10^7
	10^{-6}	9	5	7×10^7	4	7	5.5×10^7	4	3	3.5×10^7	4	1	2.5×10^7
	10^{-7}	1	0	1×10^8	0	0		0	0		0	0	
Cfu/ml		9.3×10^7			9.3×10^7			2.9×10^7			2.4×10^7		
OD600		1.688			1.830			1.476			1.363		

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Human Research Ethics Committee (Medical)

Research Office Secretariat
Faculty of Health Sciences, Phillip Tobias Health Sciences Building, 3rd Floor, Office 301/2/4, 29 Princess of Wales Terrace, Parktown, 2193
Private Bag 3, Wits 2050
Office email: HREC-Medical.ResearchOffice@wits.ac.za
Website: www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/

Ref: W-CBP-190919-03

19/09/2019

TO WHOM IT MAY CONCERN:

Waiver: This certifies that the following research does not require clearance from the Human Research Ethics Committee (Medical)

Investigator: Miss Jessica Brothwell (Student no. 1114578)

Supervisor: Dr Bhavna Gordhan

Department: Molecular Medicine and Haematology

Project title: The role of the GLY loop of endonuclease VIII DNA glycosylase in maintaining genome integrity in *Mycobacterium smegmatis*

Reason: *In vitro* laboratory study. Using non-pathogenic bacteria. No human participants will be involved in the study.

CB Penny

Dr CB Penny

Chairperson: Human Research Ethics Committee (Medical)

Copy – HREC (Medical) Secretariat: Zanele Ndlovu, Mapula Ramaila, Josh Ndlangamandla and Rhulani Mkansi

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Jessica Brothwell
Assignment title:	MA research report
Submission title:	1114578:Inline_Submission
File name:	4d3-e100-4e6e-97ea-aae5030da36...
File size:	135.51K
Page count:	19
Word count:	24,541
Character count:	110,855
Submission date:	24-Mar-2021 03:30PM (UTC+0200)
Submission ID:	1541134782

1114578:Inline_Submission

ORIGINALITY REPORT

23%	14%	8%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Witwatersrand Student Paper	10%
2	hdl.handle.net Internet Source	3%
3	theses.gla.ac.uk Internet Source	1%
4	apps.who.int Internet Source	<1%
5	etheses.bham.ac.uk Internet Source	<1%
6	wiredspace.wits.ac.za Internet Source	<1%
7	www.pubmedcentral.nih.gov Internet Source	<1%
8	e-space.mmu.ac.uk Internet Source	<1%
9	d-nb.info Internet Source	<1%