The characterization of a putative DNA repair protein in *Mycobacterium tuberculosis*, encoded *rv2414c*

Dieketseng Palesa Maleke

Supervisor: Dr. S. Senzani

A dissertation submitted to the Discipline of Medical Microbiology in the School of Laboratory Medicine and Medical Sciences, College of Health Sciences at the University of KwaZulu-Natal in fulfilment of the degree, Master of Science

Student: [redacted] Date: 30/03/2023

Supervisor: [redacted] Date: 30/03/2023

DECLARATION – PLAGIARISM

I, **Dieketseng Palesa Maleke**, declare that:

1. The research reported in this dissertation, excluding where indicated otherwise, is my original research.

2. This dissertation has not been submitted for the fulfilment of any degree or examination at any other university.

3. This dissertation does not contain other persons' data, pictures, graphs or other information, without being acknowledged as sourced from other persons.

4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) Their words were re-written but the general information attributed to them was referenced

b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signature: [redacted] Date: 30/03/2023

DEDICATION

I dedicate this work to my loving parents Mr Serumolo John Maleke and Mrs Elizabeth Diedile Maleke. Thank you very much for your continuous support and teachings. May GOD continuously bless and keep you safe my younger brother (Kopano Maleke) and I.

These are the fruits of your labour!

ACKNOWLEDGMENTS

I (Dieketseng Palesa Maleke) would like to acknowledge the following:

- GOD the almighty for blessing me with the gift of life, courage, strength, wisdom and knowledge to bring this study to successful completion
- The National Research Fund (NRF) for the highly appreciated financial assistance
- My supervisor Dr. S. Senzani for his supervision, support and continuous encouragement
- My loving husband Dr. O.A Shemfe, for his support and words of encouragement. For holding my hand and always reminding me of my worth when I was starting to lose hope
- My adorable daughter (Kgosigadi Adeola Shemfe), looking at you gave me motivation to keep going
- My parents (biological and in-laws) thank you for your endless support and prayers
- My fellow colleagues and the entire medical microbiology department, thank you for making this journey easier. I will cherish every moment we shared (the good and bad) and take away as much as I can from this wonderful experience

ABSTRACT

Mycobacterium tuberculosis (MTB) is a causative agent of the communicable disease tuberculosis (TB), which is regarded as one of the top ten causes of death worldwide. Globally, TB accounts for over 10 million infections and over 1.8 million deaths annually. These statistics are subject to a constant increase due to the emergence of drug resistant strains. Although, the recent use of next generation sequencing technology has generated complete genome sequences and functional genomic data for various organisms (MTB included), to this point, the biological functions of several proteins encoded for in the MTB genome are not known or characterized hence they are called hypothetical proteins. Characterization of these hypothetical proteins is essential, as they could be involved in key regulatory processes of MTB, which is essential for the pathogen to retain a successful life cycle and disease progression. For successful invasion of the host and disease progression, it is important for MTB to retain genomic stability. Therefore, the degree of survival of MTB in the host environment is largely dependent on the bacterium's ability to retain genomic stability. DNA repair mechanisms protect bacterial DNA from damage that can be induced by numerous stress factors. The hypothetical protein *rv2414c* encodes a gene amongst the MTB immunogenic protein identified in a study by Chiliza *et al.*, (2019) which is closely associated with genes involved in MTB DNA repair, suggesting a possible role in DNA repair pathways. Therefore, the present study is aimed at characterizing a conserved hypothetical protein encoded *rv2414c* in *Mycobacterium tuberculosis* to elucidate the proteins biological function. For in-silico characterization, three bioinformatics tools were used, namely; Mycobroswer, I-TASSER and STRING online tools. Thereafter, a CRISPR-cas9 gene silencing mechanism was developed to elucidate the biological role of *rv2414c*. CRISPR system entails the co-expression of the silenced form of RNA-guided DNA endonuclease from the type II CRISPR system (dCas9) and a small guide RNA specific to a target sequence,

leading to the DNA recognition complex resultant in transcription interference of corresponding DNA sequence. This CRIPSR mechanism was achieved by generating knockdown mutants. Phenotypic characterization of the mutants was accomplished by monitoring the mutants' growth kinetics and biological assays were done to assign a possible biological function to *rv2414c*. Bioinformatics analysis suggests *rv2414c* is involved in DNA repair based on the structural networks it forms with 3 genes (dprA, recA and cinA) involved in MTB DNA repair and 3 proteins (*rv3242c*, *rv3737* and *rv2897c*) that are involved in mainly aiding in the mechanism of DNA repair.

However, the growth kinetics showed that *rv2414c* has no impact on the MTB growth rate, as all strains grew in a similar growth pattern with no statistical significance ($p > 0.05$) observed at the different time points. Additionally, UV biological assay showed that *rv2414c* is not a major role player in DNA repair, as UV exposure did not have an effect on bacterial survival rate even in the knockdown strain. A slight decrease in cell survival rate was noticed after addition of Mitomycin C (MMC) between Δ*rv2414c* (100 ng/ml) + MMC and Δ*rv2414c* + MMC, however, the difference was not significant. This implies that *rv2414c* is not involved in MTB DNA repair.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS:

µ	:Micro
Amp	:Ampicillin
ATB	: Active TB
ATC	: Anhydrotetracycline
bp	: Base pairs
CFU	: Colony forming unit
CTAB	: Cetyltrimethylammonium bromide
DNA	: Deoxyribonucleic acid
dNTP	: Deoxynucleotide triphosphate
EDTA	: Ethylenediaminetetraacetic acid
g	: gram
HIV	:Human I immunodeficiency Virus
KanR	: Kanamycin Resistance
LTBI	: Latent tuberculosis infection
MDR	:Multidrug-Resistant Tuberculosis
MGMT	:Methylguanine Methyltransferase
ml	: Millilitre
MMC	: Mitomycin C
MTB	: *Mycobacterium tuberculosis*

MTBC	:*Mycobacterium tuberculosis* Complex
NER	:Nucleotide Excision Repair
OD_{600}	:Optical density at 600 nanometre wavelength
PCR	:Polymerase Chain Reaction
ROI	:Reactive Oxygen Intermediate
rpm	: Revolutions per minute
sdH_2O	: Sterile distilled water
SDS	: Sodium Dodecyl Sulfate
TB	:Tuberculosis
TE	: Tris-EDTA
w/v	: Weight per volume
WHO	: World Health Organisation
XDR	:Extensively-Drug Resistant
XDR-TB	:Extensively-Drug Resistant Tuberculosis
X-gal	:5-bromo-4-chloro-3-indolyl-D thiogalactopyranoside
Δ	: Delta
η	: Nano

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

1.1 **INTRODUCTION**

It has been reported that one-third of the world's population is infected with *Mycobacterium tuberculosis* (MTB) with documented cases in nearly every country. According to Lyashchenko *et al*. (2020), Tuberculosis (TB) is an infectious disease caused by microorganisms belonging to *Mycobacterium tuberculosis* complex, with a diverse range of host organisms which include humans, domesticated animals such as cattle, and wildlife species (Lyashchenko *et al*., 2020). In 2019, the World Health Organization (WHO) reported that, approximately 10.4 million new TB cases and 1.8 million TB deaths occur worldwide annually, thus making this infectious disease the leading cause of death as a result of an infectious pathogen, even when compared to the human immunodeficiency virus (HIV) as shown in Figure 1.1 (Bhargavi *et al*., 2020, Organization, 2018, Poeta *et al*., 2018).

Similarly, several studies identify TB as a global threat to mankind as the incidence and mortality rates keep increasing, particularly in the current era of many individuals being immunocompromised due to HIV/AIDS and the recently emerged COVID19 pandemic which was declared a global state of emergency (Delogu *et al*., 2013, Stein, 2011, Visca *et al*., 2021).

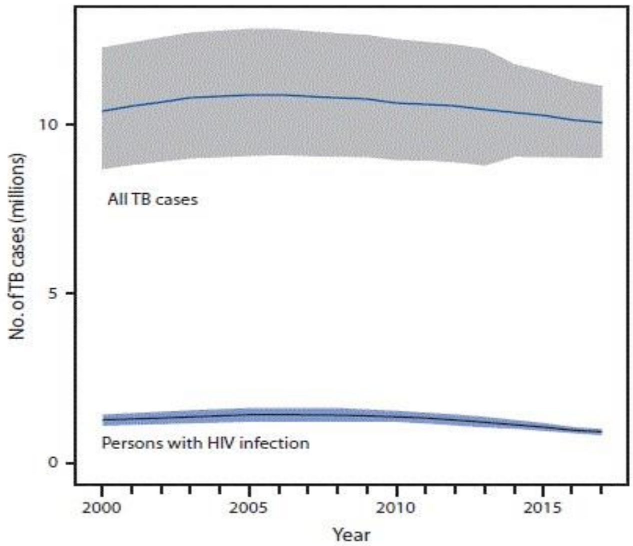


Figure 1.1: TB incidence graph showing the total number of TB cases from 2000-2017 globally and the number of persons with HIV co-infection. Image adopted from MacNeil *et al*. (2019).

In addition to TB being a health threat, it further poses social and economic burden at a global level (Delogu *et al*., 2013). However, the epidemiological distribution of this disease is strongly observed in developing countries (low and middle income countries) (Haile *et al*., 2020). Similarly, a study by Castañeda-Hernández and Rodriguez-Morales (2013) demonstrated that besides clinical factors, TB epidemiology can also be influenced by multiple factors such as: geographical, demographical and social aspects (local or national levels) which is in accordance to the reported 95% of deaths occurring in Third World countries by WHO, 2016 (Castañeda-Hernández and Rodriguez-Morales, 2013, Organization, 2018).

Later in 2018, the WHO further reported that a total of 30 high TB burden countries accounted for 87% new TB cases while 8 of the 30 countries accounted for two thirds of the total, with India in the lead, followed by China, Indonesia, Philippines, Pakistan, Nigeria, Bangladesh and lastly, South Africa (Organization, 2018). South Africa in particular has the highest TB disease incidence rate in the world (Hippner *et al*., 2019), with 438 000 estimated total number of TB patients and an incidence rate of 789 per 100 000 people in the population. Additionally, Hippner *et al*. (2019) further observed that South Africa as a developing country is negatively affected by TB with the estimated disease burden differing significantly between provinces with the rate of microbiologically confirmed TB cases as of 2012 ranging between 400/100,000/year in Limpopo Province to almost 1,200/100,000/year in KwaZulu-Natal (Hippner *et al*., 2019).

The emergence of drug-resistant TB strains often attributes to the failure of implementing proper TB control programs and correctly managing TB cases, thus worsening the threat posed by TB to mankind (Calver *et al*., 2010, Yassin *et al*., 2020). On a global scale, over 450 000 multidrug-resistant (MDR) TB cases are estimated annually, with South Africa accounting for 1%-2% (Calver *et al*., 2010). This is supported by the latest global tuberculosis report which enumerated 19 073 MDR-TB cases and 967 (extensively drug-resistant) XDR-TB cases in South Africa (Organization, 2019). While battling the HIV and COVID-19 pandemics, the incidence of TB cases continues to rise, hence the need to develop newer diagnostics and vaccines. However, the key step towards the development and search for novel therapeutic strategies is to fully characterize and understand the functionality of the proteins encoded within the MTB genome (Villela *et al*., 2017). While novel research on conserved hypothetical proteins is still under studied, the current characterization methods have limitations of being labor-intensive, time-consuming or require high technicality, and this is due to the structural complexity of MTB. Additionally, the use of immunology and proteomics (immunoproteomics) is currently a concept of interest to investigate, and evaluate pathogenicity as well as elucidate novel diagnostic biomarkers." This study will focus on determining the biological function of *Rv2414c* (conserved hypothetical protein) as it is hypothesized to play a role in MTB DNA repair which is an essential process for both the survival and progression on MTB.

1.2 LITERATURE REVIEW

1.2.1 *MYCOBACTERIUM TUBERCULOSIS*

Mycobacterium tuberculosis (MTB) the causative agent of TB, is a pathogenic bacteria belonging to family Mycobacteriaceae discovered by Robert Koch in 1882 (Cambau and Drancourt, 2014, Gordon and Parish, 2018). The causative agents of TB are a cluster of closely linked bacteria known as the *Mycobacterium tuberculosis* complex (MTBC) (Gagneux, 2018). MTBC group is made up of six closely related organisms namely: *M. africanum* (TB-like symptoms, lower pathogenicity), *M. bovis* (primarily TB in cattle), *M. canetti*, *M. caprae*, *M. microti*, and lastly MTB (Zhu *et al*., 2016). The family of these pathogenic mycobacteria originated from soil-dwelling ancestors and probably became pathogens to human and animal hosts during domestication of animals over 10,000 years ago (Green, 2017). MTB is widely known for its impermeable and thick cell wall made of peptidoglycans, polysaccharides, rare glycolipids and lipids with long-chain fatty acids, for instance mycolic acid (Daffé, 2015). At optimum conditions, which includes growth at a temperature of temperature at 37°C, and optimal accessibility to nutrition and oxygen, the bacilli has a generation period of 18-24 hours forming a white to light yellow colony on agar in 3-4 weeks (Forbes *et al*., 2018). Additionally, the bacilli, is non-spore forming, non-motile, acid-fast and an obligate anaerobe which is gram variable (Dunn *et al*., 2016). The 4.4 Mbp MTB genome is GC-rich (65.6%) and regarded as one of the largest genomes among bacteria as it codes for about 4,000 predicted proteins (Dietrich, 2013).

1.2.2 MTB PATHOGENESIS

In cases of infection, 98% of MTB is transmitted through the air in particles known as droplet nuclei that are generated when a person infected with pulmonary or laryngeal TB disease cough, sneeze, shout, or sing (Dunn *et al.*, 2016). These particles are approximately 1–5 μm in size and the normal air currents have the ability to keep them airborne for prolonged periods allowing them to spread throughout a room or building (Jensen *et al.*, 2005). MTB pathogenesis is initiated when the nuclei or droplets is inhaled and passes through the mouth or nasal passage, upper respiratory tract and bronchi to reach the alveoli (Smith, 2003). Once at the alveoli, MTB is then phagocytized by alveolar macrophages, which play a role in killing invading bacterial species by triggering the innate immune response (Bussi and Gutierrez, 2019).

The pathogenicity of MTB is greatly dependent on its virulence, ability to evade the hosts' immune system and persistence in macrophages for prolonged periods (Marimani *et al.*, 2018, Stapleton *et al.*, 2010). Subsequently, allowing MTB to replicate in the macrophage, resulting in primary infection. MTB virulence is largely dependent on its' ability to prevent macrophage maturation (Hmama *et al.*, 2015). This is facilitated by blockage of fusion among the phagosome and lysosomes, thus evading phagosome acidification (Rijal *et al.*, 2020). By so doing, MTB escapes proteolytic degradation and antigen presentation essential for activation of host adaptive immune response (Korb *et al.*, 2016). Blockage of apoptosis (programmed cell death) and utilization of macrophage necrotic route to spread to neighbouring cells are mechanism used by MTB to increase its survival chance (Awuh and Flo, 2017, Obeng, 2020). A pro-inflammatory response will be triggered by infected cells, forming granulomas by recruiting the lymphocytes, monocytes as well as fibroblasts (Etna *et al.*, 2014). However, in an immunocompetent host (healthy host with a normal immune response), infection generally has no symptoms (asymptomatic) and can be retained with no progression to active disease except optimum conditions for reactivation are established causing what is known as Latent tuberculosis infection (LTBI) (Achkar and Jenny-Avital, 2011).

Interestingly, the re-activation of dormant MTB inherent in latently MTB-infected hosts is the common cause of TB infection cases as opposed to new infections (Gengenbacher and Kaufmann, 2012). Primary infection, can advance to an active state of disease; remain as latent infection or be eliminated by the host immune system (Flynn and Chan, 2001). On average, under 10% of infected persons develop active TB in their lifetime (Brooks-Pollock *et al*., 2010). Since it is impossible to predict who has latent TB infection while appearing healthy and who will develop active TB at some point, the chances of active TB are increasing in immunecompromising cases (Barry *et al*., 2009). For instance; in anti-tumor necrosis factor (TNF) therapy for individuals with chronic inflammatory diseases, obesity/diabetes or via the coinfection with human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) (Barry *et al*., 2009).

1.2.3 DNA DAMAGE

DNA damage can be defined as any modification in the physical and/or chemical structure of DNA resulting in an altered version of a DNA molecule different from the original molecule concerning its physical, chemical and/or structural properties (Fojta, 2002). Subsequently, this results in DNA lesions, referring to a DNA molecule section containing a primarily damaged site either through base deletion, base alteration, a sugar alteration or a broken strand (Alhmoud *et al*., 2020). DNA damaging agents can arise from endogenous (stress from replication or the generation of free radicals as derives of oxidative metabolism) or exogenous (ultraviolet (UV) or ionization) sources leading to various DNA lesions (Leyns and Gonzalez, 2012). It is of high importance that these changes or lesions are corrected on time as DNA synthesis requires the pre-existing single strand of DNA to serve as a template (Chakarov *et al*., 2014). Additionally, if not corrected these changes may result in the expression of dysfunctional proteins affecting normal cell physiology (Alhmoud *et al*., 2021). DNA damage can occur due to various factors, which include oxidative damage, base alkylation, base hydrolysis, formation of bulky adduct and DNA strand breakage. Oxidative damage is caused by highly reactive oxygen radicals or reactive oxygen species (ROS) produced during cellular respiration as well as by other biochemical pathways (Chhabra and Chhabra, 2012). In ROS, hydroxyl radicals (•OH) are among the most active and electrophilic compounds that are produced via ionizing and ultraviolet radiation or other radicals rising from enzymatic reactions (Ehrt and Schnappinger, 2009). At lowered levels, ROS species implements key cellular functions such as, acting as cellular messengers in redox signalling reactions (Dupuy *et al*., 2020). On the other hand, in surplus, ROS species can result in approximately 100 various oxidative base lesions and 2-deoxyribose modifications (Chatterjee and Walker, 2017). Antibiotics are regarded as one of the sources of oxidative stress, in bacteria (Kalghatgi *et al*., 2013). Additionally, *Mycobacterium* is exposed to ROS during infection, which is released by the host innate immune response (Ehrt and Schnappinger, 2009).

Damage through the alkylation of bases

DNA alkylation is a huge threat to genetic integrity and cell life and it is defined as a chemical interaction that results in the addition of carbon atoms to a nitrogenous base (Fu *et al*., 2012). Alkylating agents are commonly found in the environment or generated in an endogenous manner as cellular metabolism by-products (Rudolph and Freeman, 2009). These alkylating agents cause lesions into DNA or RNA bases that are either neutral, mutagenic or cytotoxic to the cell (Mielecki and Grzesiuk, 2014). Cytotoxic lesions tend to result in transcription interference, hinder replication or promote the activation of apoptosis, while lesions that are mutagenic involve miscoding and cause mutations in newly synthesized DNA (Tudek *et al*., 2010). Methylation is a well-known type of alkylation. Key products involved in methylation include: N7-methylguanine (7meG), which is a significant RNA modification (occurs in epigenetic regulation), N3-methyladenine (3meA) which hinders synthesis by replicative DNA polymerases and lastly, O6-methylguanine (O6meG), a DNA repair enzyme with a crucial role in chemo-resistance of alkylating agents (Kalghatgi *et al*., 2013, Yoon *et al*., 2017, Yu *et al*., 2020).

Damage through base loss or base deletion caused by hydrolysis of bases

The covalent structure of DNA is not stable in aqueous solutions and it tends to hydrolyze in its monomeric components, which are subject to numerous hydrolytic reactions (Shapiro, 1981). These processes are slower when compared to normal chemical reactions.

An abasic site is the common term for a apurinic/apyrimidinic (AP) site, which is a region in DNA (rarely present in RNA) lacking a purine or a pyrimidine base, either spontaneously or as a result of DNA damage (Wilson III and Barsky, 2001). If this is not repaired, AP sites can result in mutation during semi-conservative replication (Hanawalt *et al*., 1979).

Damage through formation of bulky adducts

This is caused by the covalent bonds that form between biologically reactive chemicals and biological molecules (DNA and proteins), resulting in large bulky adducts or appendages, branching away from the main molecule (Turner, 2004).

DNA damage through DNA crosslinking

This type of DNA damage results in a link that occurs when numerous foreign or endogenous products react with two DNA nucleotides causing interfere with cellular metabolic processes, such as DNA replication and transcription, triggering cell death (Tretyakova *et al*., 2015, De Zio *et al*., 2013). Cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts are the two common UV products. These pre-mutagenic lesions modify the base paring and possibly molecular structure (Aralov *et al*., 2022). When not corrected these type of lesions lead to the interference of polymerases, subsequently resulting in the misread during transcription or replication, or the ultimate arrest of replication (McGlynn and Lloyd, 2002).

DNA Strand Breaks (DSBs)

Damage through DSBs, occurs in DNA due to the exposure to external agents like radiation, certain chemicals, and endogenous processes, which include DNA replication and DNA repair (Cannan and Pederson, 2016). Ionizing radiation generated by radioactive decay or in cosmic rays results in the breakage of DNA strands (Hall and Kereiakes, 2001). When ionizing radiation is at lowered levels, it prompt's irreversible DNA damage, thus causing replication and transcriptional errors easily triggering mutations (Costes *et al*., 2010). DSBs can arise when damaged chemical agents forming crosslinks within the DNA undergo DNA replication (Roos and Kaina, 2006). Additionally, DNA cross-links can result in topoisomerase enzymes being in transition state upon cleavage of the DNA backbone thus, instead of relieving supercoils and patching the backbone, the topoisomerase undergo abortive catalysis (remain covalently linked to the DNA) resulting in a single stranded break (SSB) in the case of Top1 enzymes or DSB if it is of Top2 enzymes (Allinson, 2010). During transcriptional events, the formation of DNA strand breaks and abortive catalysis can act as a damage sensory molecules within the cell and aid trigger DNA damage response signalling pathways that kick start DNA repair processes (Jackson and Bartek, 2009).

1.2.4 DNA REPAIR MECHANISMS

Genome integrity and stability is essential in all living organisms for their optimum functioning and propagation. Pathogenic organisms are constantly subjected to stressful and hostile conditions posed by the host's defense mechanisms and treatment with antibiotics (Singh, 2017). DNA stability or the ability to prevent changes to the DNA sequence resulting in mutations is essential for survival and prevention of disease states. Mechanisms responsible for preserving DNA stability include DNA polymerase proofreading and DNA repair mechanisms (Henninger and Pursell, 2014). DNA damage can be prompted by environmental stressors (arising from physical agents and chemical reactive species) and endogenous stressors (occurring spontaneously even under normal physiological conditions due to intrinsic instability of chemical bonds within the DNA molecule) (Morita *et al*., 2010). The biological extent of the DNA damage largely depends on the chemical nature of the lesion (Altieri *et al*., 2008). Hence, these organisms rely on DNA repair mechanisms for maintaining their genomic integrity. MTB as an intracellular pathogen has the capacity to repair DNA lesions caused by exposure to reactive oxygen intermediates (ROIs) generated by the phagocytic host cells thus ensuring the survival of the microbe in this hostile environment (Dupuy *et al*., 2020, Mizrahi and Andersen, 1998). Additionally, DNA repair mechanisms are crucial for the repair of lesions sustained during the dormant phase of MTB (Singh, 2017). According to Hardison and Chu (2021), there are five major classes of DNA repair mechanisms employed by bacteria and they are as follows: direct reversal of damage, nucleotide excision repair, base excision repair, mismatch repair, and lastly the repair of double-stranded breaks (Hardison and Chu, 2021).

1.2.4.1 Direct reversal of damage

Direct reversal repair is the removal of damage from DNA and RNA by means of chemical reversion independent of a nucleotide template, phosphodiester backbone breakage or DNA synthesis (Nay and O'Connor, 2013). According to Couvé *et al*. (2013), it is one of the easiest, free of errors and inexpensive DNA repair mechanisms to have evolved; hence its main advantage is the preservation of genetic information (Couvé *et al*., 2013). This repair mechanism is stimulated by DNA alkylation damage or RNA methylation resulting from epigenetic modifications and it involves the removal of damaged bases in a single enzymatic reaction (Alhmoud *et al*., 2020).

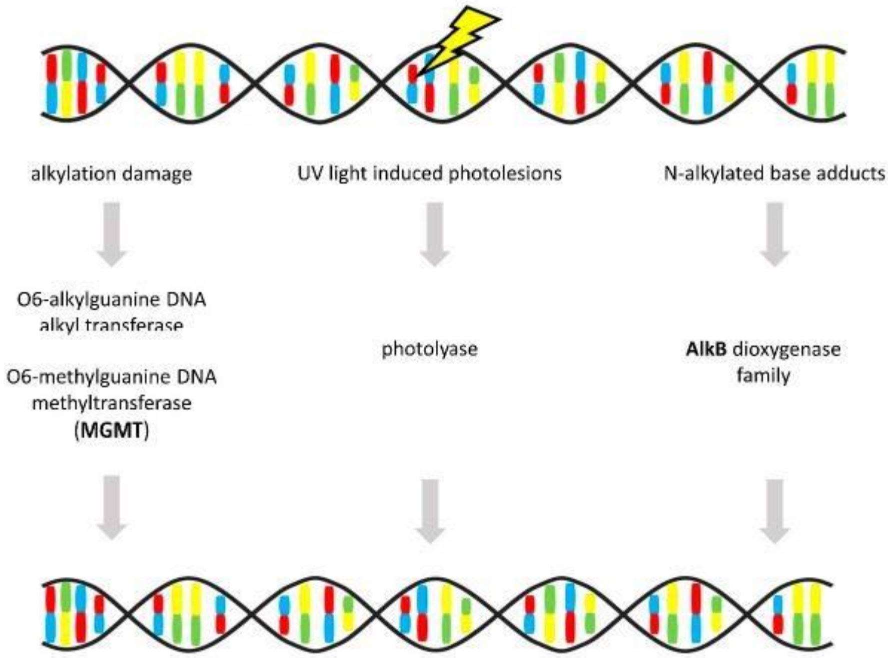


Figure 1.2: showing the simple pathway of direct repair mechanism, via direct lesion reversal eliminating alkylation and UV-prompted damage, and N-alkylated base adducts (gene alternations in OvC in **bold**). Protein names: O6-methylguanine DNA methyltransferase (MGMT), alpha-ketoglutarate- dependent dioxygenase AlkB. Image adopted from (Alhmoud *et al*., 2020).

Cells have developed direct reversal mechanisms for several types of DNA damaging events, for instance: alkylation, inter/intra-strand cross-links as shown in Figure 1.2 above (Huang and Zhou, 2021). The protein Ogt is an O^6-alkylguanine-DNA alkyltransferase, responsible for transferring an alkyl group from the O^6-position of a guanine to a cysteine and is present in both MTB and Msm (Alhmoud *et al*., 2020). In contrast, the methylation of O^6 position guanine results in DNA damage leading to incorrect complementary base pairing with thymine in place of cytosine from O^6-methylguanine product (Couvé *et al*., 2013). To counter-act this damage the O^6-methylguanine methyltransferase (MGMT) enzyme is responsible for repairing the damage in its active via the transfer of the methyl group from O6-methylguanine to a cysteine residue (Alhmoud *et al*., 2020, Christmann and Kaina, 2012, Cooper *et al*., 2007).

1.2.4.2 Nucleotide excision repair

Nucleotide excision repair (NER) was first identified in bacteria in the mid-1960s by David Pettijohn and Philip Hanawalt, with the observation of DNA synthesis that is nonsemiconservative during the excision of cyclobutane pyrimidine dimers (CPDs) (Pettijohn and Hanawalt, 1964). NER is one of the most essential DNA repair systems; as a result, it is highly conserved from prokaryotic species to higher eukaryotes. This mechanism entails a wide class of bulky, helix-distorting lesions known for obstructing transcription and normal replication, for instance, UV-induced CPDs and 6-4 photoproducts [(6-4) PP], adducts, and crosslinks prompted by chemical agents such as benzo[a]pyrene and cisplatin (Altieri *et al*., 2008, Tatum and Li, 2011). The significance of this repair mechanism is to restore the distortion of the normal DNA helix configuration caused by UV-exposure and dimers (Hardison and Chu, 2021). Since UV/ solar exposure is constantly occurring, these DNA lesions need to be corrected before the process of DNA replication is initiated to prevent the DNA polymerase from mis-reading and inserting incorrect nucleotides leading to mutations.

Therefore, NER can be summarized as follows: a DNA lesion will be identified (either as a result of solar radiation or thymine dimers), thereafter NER enzymes will recruited to the site of the DNA lesion. These enzymes will then open the natural helix structure of the DNA double stranded molecule and cleave of the region containing the lesion. After cleavage, DNA polymerase will rebuild a new strand utilizing the un-damaged DNA strand as a template. This mechanism of DNA repair is extremely important as thymine dimers are constantly occurring and although it is found in all organisms, it was initially examined in *Escherichia coli* (*E. coli*) (Žgur-Bertok, 2013). Amongst many similar studies, a study by Hardison and Chu (2021) reported that, are the UvrABC exconuclease and UvrD helicase are the best characterized enzymes for the catalysis this process in *E. coli* (Hardison and Chu, 2021). Additionally, genes encoding the UvrABC excinuclease, uvrD and polA were also easily identified in MTB.

The NER mechanism is split into two pathways namely; global genomic repair (GGR) and transcription-coupled repair (TCR) (Svejstrup, 2002). GGR can repair DNA damage anywhere within the genome, thus it is transcription independent while TCR solely repairs DNA strands due for transcription (Kimura and Sakaguchi, 2006). NER involves the following key processes; recognition, unwinding, excision and gap filling. As shown in Figure 1.3 below, in GGR-NER, the xeroderma pigmentosum group C protein complex (XPC-HR23B) and UVdamaged DNA-binding protein (UV-DDB) complex are responsible for recognising DNA damage while in TCR-NER DNA damage blockage of RNA polymerase triggers recognition (Peng *et al*., 2011). Thereafter, the transcription factor (TFIIH), DNA repair protein complementing cells (XPA) and RPA protein are recruited to the lesion site for lesion verification in both pathways. The XPF-ERCC1 complex and XPG (primary protein of NER) mediate the dual unwinding around the lesion followed by the single strand gap filling by DNA polymerase δ with the aid of the proliferating cell nuclear antigen (PCNA) which is a DNA clamp serving as a DNA polymerase δ co-factor and its recruited by the replication factor C

(RFC). Finally, ligation by the DNA ligase III-XRCC1 activity, ligating the gap after damage removal.

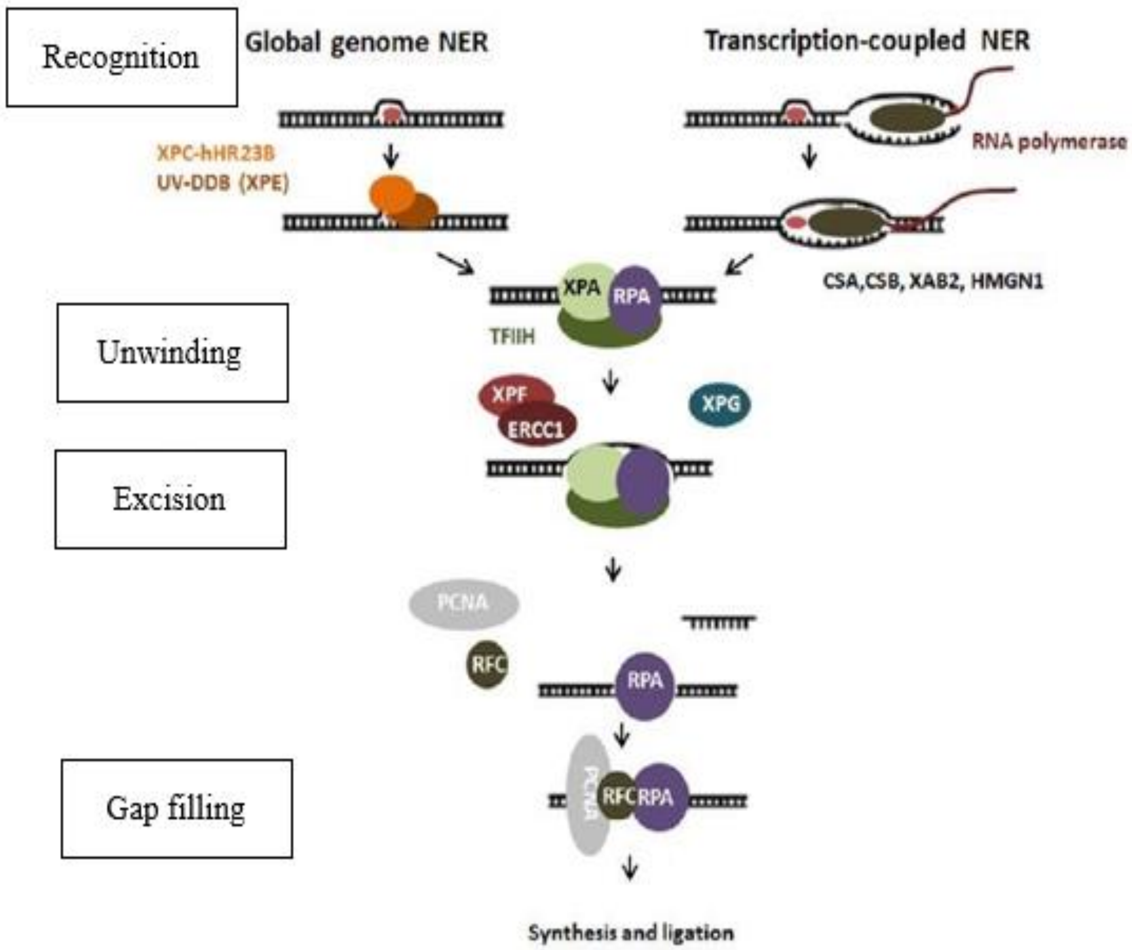


Figure 1.3: NER can be split into 2 different pathways; GG-NER and TC-NER. NER mechanism involves the elimination of a short stretch DNA containing a lesion and the repair of this lesion with the non-damaged DNA serving as a template. GG-NER is for the removal lesions in the non-transcribed domains of the genome or non-transcribed strands in the transcribed domains. TC-NER is for the removal of lesions from the transcribed strand of active genes (image adapted from (Leyns and Gonzalez, 2012))

1.2.4.3 Base excision repair

DNA also faces attacks by chemical reactions such as oxidations, deamination's and alkylation's which damage the nitrogenous bases in DNA. If left untreated this can cause DNA polymerase to misread and insert the wrong nucleotides. Base excision repair (BER) is responsible for repairing lesions caused by oxidative damage, alkylation, deamination and depurination/depyrimidination and single strand breaks (SSB) (Leyns and Gonzalez, 2012). Thus, BER is responsible for fixing non-bulky lesions such as those found when a nitrogenous base is chemically modified. BER can be divided into two pathways, namely; the short-patch and long-patch pathway. The short-patch BER pathway repairs a single nucleotide, while the longpatch BER pathway produces at least two nucleotides (Robertson *et al*., 2009). The induction of short-patch BER or long-patch BER is dependent on the condition of the 5' deoxyribose phosphate (5'dRP) terminus whereby reduced or oxidized AP sites will undergo long-patch BER, while unaltered AP sites undergoes the short-patch BER (Leyns and Gonzalez, 2012).

BER consists of the following steps: damage recognition, base removal and incision (Leyns and Gonzalez, 2012). This first step in both BER pathways is facilitated by a specialised enzyme known as DNA glycosylase that scans the DNA molecule for alternated nitrogenous bases such as deaminated, alkylated and oxidised bases (Brooks *et al*., 2013). After recognition of damaged or inaccurate bases the DNA glycosylase will remove the bases via the hydrolyzation of the N-glycosidic bond responsible for linking the base and the sugar phosphate backbone using the relevant glycosylase, leaving behind an AP site (Chatterjee and Walker, 2017). Thereafter, DNA glycosylase will recruit an enzyme known as AP

Endonuclease (in bacteria it is the Endonuclease IV commonly known as EndoIV, *Nfo*) to the AP site, which cleaves the phosphodiester backbone of DNA and removes the ribose and phosphate group that were left behind by the glycosylase, thus creating a gap in the DNA backbone (Chen *et al*., 2022).. Thereafter, DNA polymerase will start gap-filling with the accurate nucleotide and finally DNA ligase activity completes the repair process, thus integrity restoration of the helix by closing the nick (Norjmaa *et al*., 2016). Different enzymes in the two pathways facilitate this step. In short-patch repair pathway the ligation is completed by DNA ligase III (LIG3), while in the long-patch repair pathway the ligation is completed by the interaction of PCNA and DNA ligase I (LIG1) as shown in Figure 1.4 below (Caldecott, 2020).

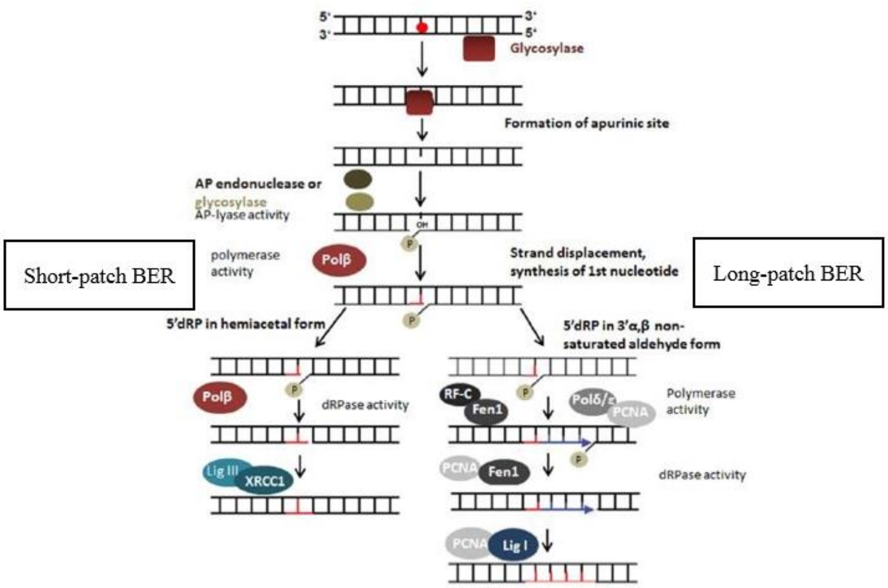


Figure 1.4: Scheme diagram of BER mechanism: the short-patch repair (left) and long-patch repair (right). Recognition is the first step, followed by base removal and incision. Selection of which pathways to use between the short-patch BER and the long-patch BER is largely dependent on the state of the 5' deoxyribose phosphate (5'dRP) terminus. Imaged adopted from (Christmann *et al*., 2003, Leyns and Gonzalez, 2012)

In some rare cases, the thymine in DNA is converted to uracil which is only required in RNA molecules hence the need to be converted back to thymine. In this case BER will identity the uracil in the DNA molecule, cleave and remove it and replace it with the thymine (Kim and Jinks-Robertson, 2010).

1.2.4.4 Mismatch repair

Mismatch repair (MMR) is a conserved post replicative pathway that largely contributes to the replication fidelity by more or less 100-folds, by restoring mismatched bases thereby avoiding the acquisition of spontaneous mutations (Merchut-Maya *et al*., 2019). This pathway guarantees genomic stability and integrity, and prevents insertions and deletions of abnormal DNA at microsatellites (Pećina-Šlaus *et al*., 2020). Errors causing base match can occur in various ways for instance: during DNA replication, formation of a hetero-duplex between homologous DNA strands during a recombinational events and the insertion-deletion loops (IDLs) in repetitive DNA sequences resultant from strand slippage events (Chatterjee and Walker, 2017). It is paramount to note that recognition is strictly on the parental strand (Castañeda-García *et al*., 2017).

The MMR pathway is initiated by binding of a MutS homo-dimer to the mismatch. Followed by the location and cleavage of a hemi-methylated dGATC site 5′ or 3′ in the mismatch by the following enzymes; MutS, MutL, MutH, and ATP (Marinus, 2012). MutH generates a strandspecific nick at hemi-methylated dGATC marking the starting point for excision of the mispaired base (Lee, 2014). Upon the availability of MutL, helicase II (UvrD) is responsible for loading at the nick point and unwinding the duplex from the nick to the mismatch (Zhang *et al*., 2007). Based on the position of the strand break relative to the mismatch, excision is cried out by exonuclease (ExoI or ExoX from 3′→5′), or ExoVII or RecJ (5′→3′ exonuclease) (Beam, 2004). The resultant single-stranded gap will undergo repair DNA re-synthesis and ligation is carried out by DNA polymerase III holoenzyme, single-stranded DNA-binding protein (SSB) and DNA ligase (Li, 2008) .

1.2.4.5 Repair of double-stranded breaks

Various chemical and physical DNA damaging agents induce highly toxic double-strand breaks (DSBs), which are implicated in various mutations, disorders and cancers when unresolved (Chatterjee and Walker, 2017). Under the DSB repair mechanism, there are two pathways namely: homologous recombination repair (HRR) and non-homologous end joining (NHEJ). In comparison to NHEJ, HRR is highly conserved and free of errors as a result of its dependence on the presence of sister chromatids, hence essential for re-establishing genomic stability (Trenner and Sartori, 2019). On the contrary, this is a limiting factor to the HRR pathways as it can only repair cells in the cell cycle S/G2 phase where sister chromatids are present, while the NHEJ pathway has no restrictions and can repair DSBs throughout the cell cycle (Huang and Zhou, 2021).

1.2.4.5.1 HRR

This pathway consists of three major processes, namely: double-strand break recognition (DSBR), synthesis-dependent strand annealing (SDSA) and break-induced replication (BIR) respectively. HRR is spilt into three steps namely: pre-synapsis, synapsis, and post-synapsis. Pre-synapsis involves the recognition and processing of the DSB ends to a 3′-OH ending singlestranded tail as shown in step(s) 1-2 of Figure 1.5 below. During synapsis, invasion of a DNA strand is facilitated by the Rad51-ssDNA filament forming a D-loop as shown in the 3rd step whereby at least three different pathways are proposed after the D-loop intermediate (Bhattacharjee and Nandi, 2017). The first step of synthesis-dependent strand annealing

(SDSA), involves detaching the invading strand after DNA synthesis followed by annealing with the second end, resulting in localized conversion without crossovers as shown in steps 4a - 5a - 6a (Saponaro *et al*., 2010).

Secondly, break-induced replication (BIR) involves the D-loop assembly into a complete replication fork, copying the whole distal part of the chromosome resulting in loss-of heterozygosity (LOH) as shown in steps 4a- 5b – 6b (Ruff *et al*., 2016). Thirdly, in the doublestrand break repair (DSBR) engages both ends of the DSB, either through independent strand invasion or second end capture, resulting in the double Holliday junction formation as shown in steps 4b - 5c - 6c-e – 7 (Nimonkar and Kowalczykowski, 2009). This resultant junction is processed by a resolvase into non-crossover or crossover products (see steps 6c and d) or dissolved by a mechanism involving BLM-mediated branch migration and TOPOIIIα-catalyzed dissolution of a hemi-catenane (step 6e), leading exclusively to non-crossover products as shown in step 7.

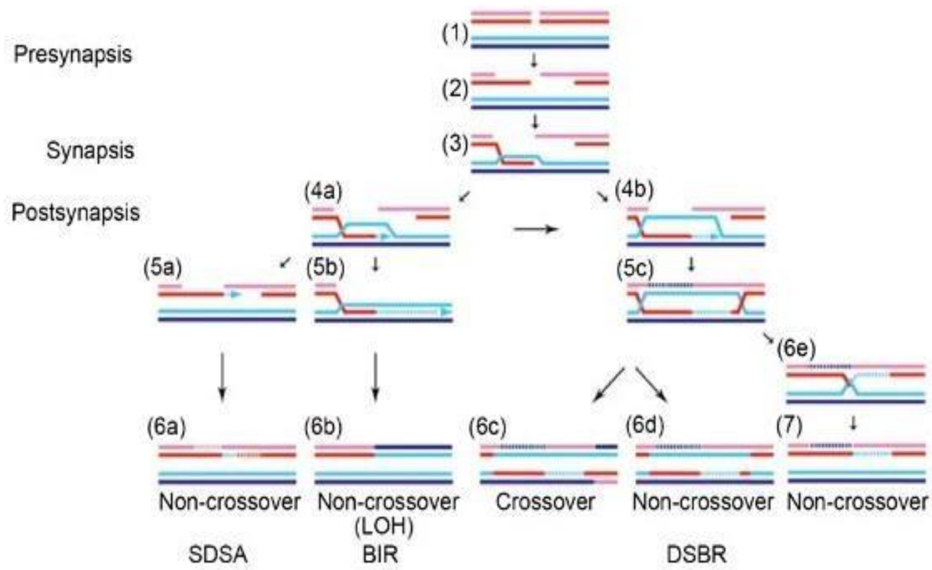


Figure 1.5: image showing the events involved in HRR pre-synapsis, synapsis and postsynapsis. Image adopted from (Li and Heyer, 2008)

1.2.4.5.2 NHEJ

The NHEJ pathway utilizes various proteins for identification, removal, polymerization and ligation of DNA ends. The NHEJ pathway is a vital repair mechanism in DSBs, activated in various cell types and cell cycle phases as it does not require the presence of sister chromatids (Tomasova *et al*., 2020). NHEJ pathway employs the following mechanism: firstly, a set of enzymes facilitates capturing of both ends of the broken DNA molecule, secondly, formation of a molecular bridge, bringing the two DNA ends back together and lastly, the re-ligation of the broken DNA molecule (Weterings and Chen, 2008). The Ku heterodimer (Ku70 and Ku80) is the first to recognize and bind to both ends of the broken DNA molecule thus the region of the DSBs within seconds and recruits additional elements required for NHEJ. Ku has the ability to bind both blunt ends and overhangs in turn stimulating the ligase activity of LigD via direct interaction (Sowa *et al*., 2022).

Additionally, the ability of LigD to possess a nuclease and polymerase domain allows it to make use of both ribonucleotides (rNTPs) and deoxyribonucleotides (dNTPs), hence it is vital to DNA repair during dormancy and during dNTP limiting conditions (Delmastro, 2015). Upon initial steps of NHEJ-mediated DSB repair, the DNA-Ku scaffold leads to attraction of the DNA-dependent protein kinase catalytic sub-unit (DNA-PK_{CS}) broken site (Valikhani *et al*., 2021). This kinase has several roles, which include the formation of a synaptic complex to bring both DNA ends together (Zhao *et al*., 2019). After capturing and tethering of the two DNA ends in a protein complex consisting of Ku and DNA-PK_{CS}, the non-ligatable DNA termini requires processing before final repair of the DSB can take place (Chen *et al*., 2021). Nucleases and polymerases are part of the numerous enzymes that have been identified, and are able to either remove or fill-in single-stranded which have non-compatible overhangs

(Pryor *et al*., 2015). Finally, the ligase IV/XRCC4 complex catalyzes ligation of the processed DNA ends, which may be enhanced by the presence of the recently discovered XLF/Cernunnos protein (Recuero-Checa *et al*., 2009).

1.2.5 Characterization of hypothetical proteins

Until recent years, studies of MTB bacterium have been neglected due to the difficulties encountered in understanding and working with MTB genome (Pillai and Chouhan, 2018). However, the use of MTB H37Rv chromosome circular maps has improved and made the study of MTB genetics tangible as it shows the various portions of genes and proteins essential in the survival and replication of MTB. According to Andreu and Gibert (2008), MTB H37Rv was the first strain of MTB to undergo whole-genome sequencing (Andreu and Gibert, 2008). MTB H37Rv has been used as a reference strain since the sequencing of the *Mycobacterium* by Cole *et al*., 1998 (Carey *et al*., 2018). Genome sequence availability of pathogens such as MTB has provided a vast amount of information that can aid and improve drug target as well as vaccine target identification. A study conducted by Mazandu and Mulder (2012) concluded that regardless of the use of high-throughput sequencing technology to yield complete genome sequences, there is still a gap in our current knowledge on gene functions within the DNA sequencing of organisms (Mazandu and Mulder, 2012).

Despite the annotation of the MTB H37Rv strain being improved through numerous studies, there is still a portion of the genes that are un-annotated and subsequently the proteins are labelled as “hypotheticals” or “unknown” or “uncharacterized”. According to Ijaq *et al*. (2019), hypothetical proteins (HPs) are defined as proteins that are predicted to be expressed from an open reading frame (ORF), but for which there is no experimental evidence of translation (Ijaq *et al*., 2019). While “Conserved Hypothetical” proteins are proteins with phylogenetic lineages with no known definitive function (Chandrasekaran *et al*., 2019, Stephens *et al*., 1998). In a quest to improve the functional annotation of the MTB proteins, Chiliza et al. (2019) explored possible proteins recognized by patients' antibodies through immuoscreening of MTB F15/LAM4/KZN against TB patients sera for unique ATB and LTBI specific biomarkers. In this study, the selection of common ATB and LTBI biomarkers showed that most secreted proteins were responsible for cell wall and cell processes followed by conserved hypotheticals, then intermediary metabolism and respiration, then PE/PPE, then lipid metabolism, information pathways, then virulence, detoxification and adaptation, then unknowns, followed by regulatory proteins and lastly, insertion seqs and phages. Conserved hypotheticals were the second highest group with 1042 proteins annotated as conserved hypotheticals with 23 of these proteins being common between LTBI and ATB, which includes the protein encoded *Rv2414c* as shown in figure 1.6 below.

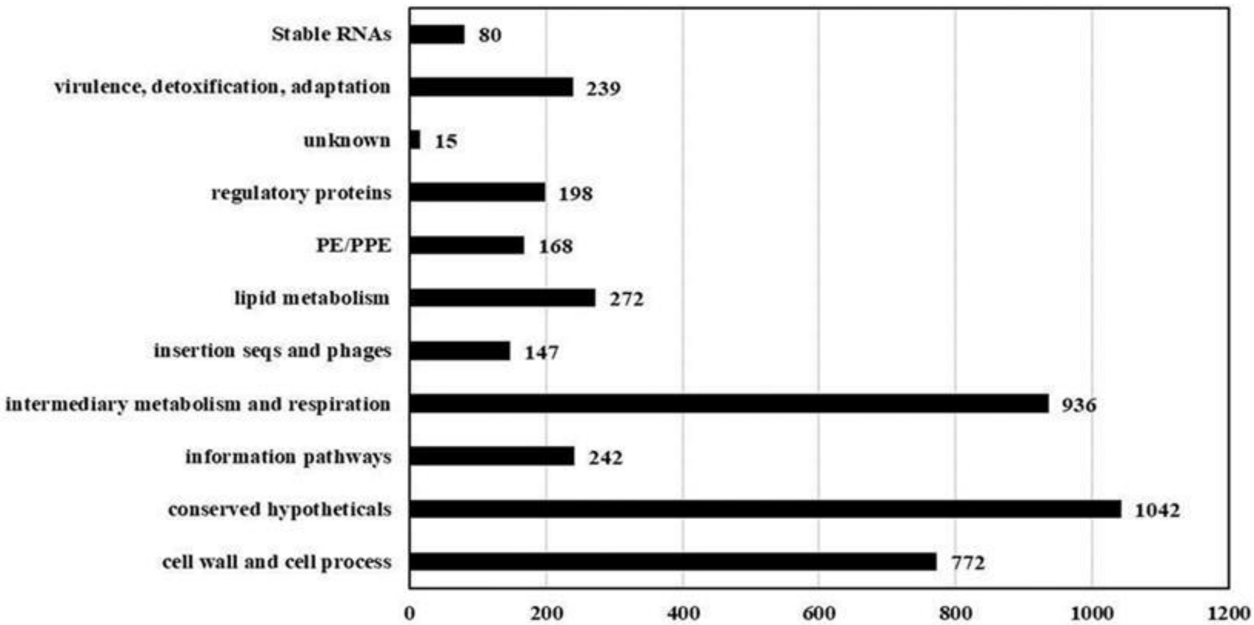


Figure 1.6: Functional categories of MTB proteins unique in ATB and LTBI (Chiliza *et al*., 2019)

The *rv2414c* protein was part of those identified in both ATB and LTBI biomarkers. *Rv2414c* is a conserved hypothetical protein, with the following genes associated to it; *recA*, *cinA* and *dprA*. As reported by Sowin and Cox (2017), *recA* is a recombinase protein which plays an important role in recombinational DNA-repair systems since it acts as a DNA pairing enzyme which is responsible for promoting an exchange of homologous DNA strands (Lusetti and Cox, 2002). On the other hand, *cinA* exhibits functions in the MTB virulence, detoxification as well as adaptation and its further reported to have strong similarity to competence damage proteins *CinA* of *Bacillus subtilis* and *S. pneumoniae* (Irshad and Munir, 2017). Lastly, *dprA* is reported to be a putative DNA processing protein and it is predicated to facilitate the loading of *recA* onto ssDNA (Brimacombe *et al*., 2015). The functionality of these genes associated with *Rv2414c* suggest that it may have a crucial role in MTB DNA repair.

In a quest of finding novel TB diagnostic tools, for improved drug design and vaccine development the characterization of HPs is of paramount importance. Furthermore, genome stability is crucial in the survival and multiplication of bacterial cells. Living organisms (both prokaryotes and eukaryotes) have various DNA repair pathways, which protect the MTB DNA from numerous errors that can occur because of internal and external factors, which are summed up as DNA damaging mechanisms. The continuous changes experienced in host defence systems and antibiotic treatments result in pathogenic bacteria being constantly exposed to a number of harsh and hostile conditions (Dos Vultos *et al*., 2009). Hence its significant for bacteria to poses DNA damage as well as repair mechanisms to efficiently impede the detrimental effects of such challenges (Dos Vultos *et al*., 2009, Warner and Mizrahi, 2006). In addition, for successive proliferation to occur, the cells genome integrity is essential for both survival and stability of microorganisms (Müller *et al*., 2018).

1.3 PROBLEM STATEMENT

TB is a major threat to humanity hence improved control measures and development of more effective therapeutics is a major priority (Schrager *et al*., 2020). An estimated 10.6 million individuals were diagnosed with TB globally (Organization, 2021). South Africa (SA) as a developing country contributes significantly to this Figure as it falls under the top 30 countries with the highest TB incidence rate and individually SA it is solely responsible for 3% of the infections (Van der Walt and Moyo, 2021). In a similar manner, a study by also showed that TB is more widespread in middle- low income populations, classified with poverty and SA is one of them (Das *et al*., 2020). Despite the numerous intervention measures put in place for TB control, which include drugs to cure the disease, diagnostic tools for patient identification and a vaccine to prevent severe forms of childhood TB (Swaminathan and Rekha, 2010). As reported by WHO, these measures seem to be inadequate and face limitations in the fights against TB due to the following reasons:

1. TB transmission is not easily interrupted as this pathogen is airborne and can easily be transmitted by simple activities like coughing, sneezing and talking from an infected individual putting their close contacts at an infection risk (Ngamvithayapong-Yanai *et al*., 2019).
2. Diagnosis of TB is usually late and not entirely correct due to limited resources in low income countries (Popescu *et al*., 2020).
3. Lack of an effective vaccine against TB in adults. Despite the early establishment of BCG it only protects against TB in early childhood (Fatima *et al*., 2020). Numerous vaccine candidates are in the development pipeline but they are still not effective in treating TB in human models (Afkhami *et al*., 2020).

4. TB treatment of requires 3–4 drugs given for a 6 months period or longer and usually compliance is poor leading to early termination of drug therapy, resulting in the emergence of resistant strains (Organization, 2018)
5. Increase of drug resistance within patients. Patients diagnosed with MDR-TB are facing a huge challenge since this form of the disease is resistant to the first-line of defence drugs being isoniazid and rifampicin (Khawbung *et al*., 2021). Treatment to the type of TB requires highly toxic and costly drug combinations with a pro-longed period of 24 months (Dwivedi and Giri, 2021). On the other hand, XDR-TB is quickly arising, resulting in increased mortality rates due to limited treatment (Tiberi *et al*., 2022).
6. HIV co-infection. Being immunocompromised due to HIV infection leads to the speedy progression of TB latent infection to TB disease by 10 to 100-folds as per the level of impairment to the immune system (Achour and Chebbi, 2022).
7. Prevention is not easily achieved. Upon TB infection, the chances of active TB development on an average lifetime is 10% (Gao *et al*., 2018). However, this can be increased by numerous factors for instance; immune compromising diseases (such as diabetes and HIV), as well as poor living conditions and unhealthy lifestyles. A combination of a class of drugs known as prophylaxis are used to prevent latent TB but they are highly toxic (Cerrone *et al*., 2020). There is still a huge gap in latent TB diagnostic biomarkers (Sudbury *et al*., 2020).
8. Inadequate health systems. Poverty is a major challenge in developing countries this directly leads to the poor response of public health systems to TB control measures as health facilities in these arears are understaffed, suffer from shortage of equipment and they are under-funded (Matatiele *et al*., 2021).

A study conducted by Chin and Hanson (2017), supported the above-mentioned shortcomings as it observed and reported in SA just above half of the population (53%) of TB patients were successfully diagnosed and treated (Chin and Hanson, 2017). These findings and reports show that there is still a huge gap in the TB control measures. Additionally, this calls for the urgent need of improved diagnostic tools because early diagnosis can reduce the spread of the airborne pathogen and more effective therapeutics as the pathogen in constantly evolving in resistant strains. This can only be fully achieved, if the MTB genome is extensively studied and well understood, to elucidate the importance of each biological process in the survival of MTB. For instance, the ability of the pathogen to combat host-immunity by employing DNA repair mechanisms that will maintain the bacterial integrity by ensuring that any form of pathogen DNA damage is corrected for survival and progression of the bacilli once within the host environment. Dispute the MTB genome having been extensively studied for over two decades, the role and functions of 27% (1051/3906) of encoded proteins in the MTB genome has not yet been determined, hence these proteins are annotated as hypothetical proteins. Subsequently narrowing our understanding on the pathogenicity and virulence of these strains and this is a major concern to the global health status as MTB continues to claim lives. In a quest of having a better understanding of MTB and developing more advanced treatment methods, it is of paramount importance to understand the role of these proteins in MTB infection. This study is aimed at the characterization of a conserved hypothetical protein encoded *Rv2414c* based on the predicted biological function of being involved in DNA repair.

1.4 HYPOTHESIS

It is hypothesised that the *rv2414c* protein has a function in MTB DNA repair. It is further hypothesised that the downregulation of the gene encoding the protein *rv2414c* is associated to MTB growth conditions.

1.5. AIM AND OBJECTIVES

Aim

To characterize a conserved hypothetical protein encoded *rv2414c* in *Mycobacterium tuberculosis* to elucidate the proteins biological function through the following **objectives:**

1.5.1 Bioinformatics analysis for in silico characterization of *rv2414c.*

1.5.1.1 To characterize the protein structure of *rv2414c*.

1.5.1.2 To determine the structural connectivity of *rv2414c* or networks formed with neighbouring proteins and genes using STRING bioinformatics tool.

1.5.1.1 To predict the protein structure and possible protein function using I-TASSER.

1.5.2 Construction of the *rv2414c* gene knock down strain.

1.5.2.1 To generate a knockdown plasmid.

1.5.3 Phenotypic characterization of the *rv2414c* gene deletion strains to elucidate enzyme function

1.5.3.1 To perform MTB growth kinetics analysis.

1.5.3.2 To perform DNA damaging (Ultraviolet and Mitomycin C) assays to assess the protein functionality.

1.6 KEY QUESTIONS

1.6.1 Can the *rv2414c* gene be effectively knocked down in MTB H37Rv wildtype strain?

1.6.2 Does the r*v2414c* gene knockdown affect growth and virulence of MTB?

1.6.3 Is *rv2414c* involved in MTB DNA repair processes?

CHAPTER TWO: MATERIALS AND METHODS

2.1. Bioinformatics analysis

Bioinformatics analysis was carried out using three database and online bioinformatics tools. MTB H37Rv DNA and protein sequences were downloaded from Mycobrowser (https://mycobrowser.epfl.ch/) using the following parameters: the species tab were specified to show MTB H37Rv hits only, with *rv2414c* as the specified gene name. Thereafter, the gene (*rv2414c*) connectivity was determined using STRING (https://string-db.org/) bioinformatics tool and lastly I-TASSER (https://zhanggroup.org/I-TASSER/) was used for structural prediction of the *rv2414c* hypothetical protein. For STRING analysis, these were the parameters used: *rv2414c* was used as the protein name and the organism tab was left on "auto-detect". For the I-TASSER job, these were the parameters: sequence obtained from mycobrowser (FASTA format) was used as an input of the I-TASSER page, and the protein name used as an ID to run the job.

General media preparations

A brief summary of the study experimental layout is provided in Appendix E.

All Media and chemicals used in this study, i.e. Middlebrook 7H9 broth, Middlebrook 7H10 agar, albumin/dextrose/catalase (ADC), oleic acid/albumin/ dextrose/catalase (OADC) supplements were purchased from Fisher Scientific. Antibiotics kanamycin, anhydrotetracycline (ATC), xgal and Mitomycin C were purchased from Sigma-Aldrich (Merck). For MTB selective media, final antibiotic concentrations were 50 µg/ml kanamycin for *Mycobacteria* unless stated otherwise.

2.2 Bacterial strains, plasmids and growth conditions

All bacterial strains and plasmids used in this study are listed in Table 2.1 and 2.2 respectively. Glycerol stocks were prepared for each strain with media containing 6.6% (v/v) glycerol and stored at -80 °C.

Table 2.1.1: Bacterial strains used in this study.

Strain	**Description or Genotype**	**Source or reference**
***Mycobacterium tuberculosis* H37Rv**	Wildtype Virulent laboratory reference strain of MTB ATCC27294	Obtained from UKZN (Medical School) storage.
Δ*rv2414c*	Derivative of MTB H37Rv carrying PLJR965Δ*rv2414c*, Kan^R	This study
XL10-Gold Ultracompetent cells	Chemically competent cells for large DNA molecules transformation with high efficiency. Hte phenotype increases transformation efficiency of large ligated DNA molecules. Ideal for the construction of plasmid DNA libraries as they decrease size bias and produce larger, more complex plasmid libraries. Strain is endonuclease deficient (endA), this significantly improves the mini-prep DNA quality, and recombination deficient (recA), ensuring insert stability.	Agilent

Kan^R: Kanamycin resistant

Table 2.1.1: plasmids used in this study.

Plasmids	Genotype or Description	Source or reference
PLJR965	MTB CRISPR plasmid (L5 integrating) Contains of the Sth1 dcas9 allele and an Sth1 sgRNA (with a modified Shine-Dalgarno motif predicted to reduce Sth1 dCas9 translational efficiency).under the control of an optimized, synthetic Tet repressorregulated promoters, a mycobacterial codon optimized Tet repressor, a single-copy L5-integrating backbone, a pBR322-derived *E. coli* replication origin, and a kanamycin-selectable marker KanR	Addgene
PLJR965Δ*rv2414c*	Derivative of PLJR965 carrying the annealed *rv2414c* guide RNA template, Kan^R	This study

Kan^R: Kanamycin Resistance

2.1.1 Growth conditions of MTB H37Rv

All MTB strains, unless otherwise stated, were grown in Difco™ Middlebrook 7H9 supplemented with 0.2% (v/v) glycerol, 10% (v/v) Middlebrook Oleic Acid Dextrose Catalase (OADC) enrichment (Difco™) and 0.05% (v/v) Tween80 or grown on Middlebrook 7H10 supplemented with 0.5% (v/v) glycerol and 10% (v/v) Middlebrook OADC. All strains were grown at 37 ^{0}C with liquid cultures grown in polystyrene tissue culture flask with filtered cap (250 ml) at 37 ^{0}C, while shaking (IncoShake Incubator, Labotec). Liquid cultures were grown in tissue culture flasks placed flat and incubated at temperature 37 ^{0}C. A class II laminar flow was used to perform all culturing and manipulations of MTB in a Biosafety Level 3 laboratory at negative pressure (160 – 170 kPa).

2.1.2 Growth conditions of PLJR965 plasmid

The cloning vector PLJR965 and its derivatives, was grown in a sterile 250 ml conical flask containing 50 ml of nutrient broth (NB) supplemented with 500 µl of kanamycin (50 µg/ml) overnight in a 37 ℃ incubator shaking at 250 rpm. Thereafter, 0.2% glycerol stocks thus 800 µl of PLJR965 overnight culture and 200 µl of sterile glycerol were prepared and stored at -80 ℃. Mycobacterial strains used in this study was MTB H37Rv, with all genes cloned from MTB H37Rv genomic DNA.

2.2 DNA manipulations

2.2.1 Mycobacterial chromosomal DNA extraction

The Cetyltrimethylammonium bromide (CTAB) DNA extraction method was followed whereby bacterial strains were grown on Middlebrook 7H10 agar for two days and the resulting cells were harvested by scraping the surface of the agar plate, and re-suspending the cells in 500 µl 1X TE buffer(pH 8.0) followed by incubation at 65 °C for 35 min. Thereafter, the cells were, cooled down by incubation at room temperature. Thereafter, a 50 µl aliquot of lysozyme (10 mg/ml) was added to the sample and incubated at 37 °C for 1 hour, subsequently 70 µl of 10% SDS and 6 µl of proteinase K (10 mg/ml) was added and the sample was incubated at 65 °C for 2 hours. After incubation, 100 µl 5 M NaCl and 80 µl pre-warmed CTAB/NaCl was added, mixed and incubated at 65 °C for 10 min. An equal volume of 24:1 Chloroform Isoamyl alcohol was added, followed by mixing and centrifugation at 11,000 x g for 5 min at room temperature. The aqueous layer was subsequently transferred into a clean Eppendorf tube followed by addition of 0.6 volume equivalent isopropanol and centrifugation for 20 min at room temperature. The resultant supernatant was discarded, the pellet washed in 70% ethanol and dried using an Eppendorf Concentrator 5301. The DNA pellet was re-suspended in 50 µl sterile distilled water (sdH2O), quantified using NanoDrop and stored at 4 °C.

2.2.2 Nucleic acid quantification

DNA used in this study was quantified using the NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies) used in conjunction with software provided by the manufacturer (Coleman Technologies).

2.2.3 Plasmid extraction of PLJR965

After successful growth of the PLJR965 plasmid, plasmid extraction was done using the GeneJet Bulk Extraction following the manufacturer's instructions. A 50 ml bacterial culture was grown to an OD_{600} (overnight) and cells were harvested by centrifugation for 10 min at

5,000 × g. Thereafter, the resultant supernatant was discarded. The remaining pellet was re-suspended in 2 ml resuspension solution (ensuring that no clumps are left behind), followed by the addition of solution 2 which is a lysis buffer (4 ml) and the tube was inverted 6 times and incubated at room temperature for 3 minutes until the mixture was viscous and slightly clear. Incubation was strictly 3 minutes to avoid the denaturation of supercoiled plasmid DNA. Thereafter, 3 ml of solution 3 (neutralization solution) was added, mixed and tube was allowed to sit on ice for 10 minutes. It is important to note the solution(s) 1, 2, 3 were added in the ratio (1: 2: 1.5) respectively. To pellet cell debris and chromosomal DNA, the mixture was centrifuged for 40 minutes at 4,000 – 5,000 × g. Thereafter, the supernatant was carefully transferred into a sterile 15 ml tube without disturbing the white precipitation followed by the addition of 3 ml 96% ethanol and mixing by inverting the tube 5-6 times. Part of the sample (~ 5.5 ml) was transferred to the pre-assembled column with a collection tube (15 ml) without overfilling the column. The column contents centrifuged for 3 min at 2,000 × g in a swinging bucket rotor and the resultant flow-through was discarded and the column was placed back into the same collection tube. Centrifugation was repeated to remove any remaining lysate. Thereafter, 4 ml of wash solution I was added to the purification column followed by centrifugation for 2 min. at 3,000 × g in a swinging bucket rotor and the resultant flow-through was discarded and the column was placed back into the same collection tube. Another 4 ml of wash solution II was added to the purification column followed by centrifugation for 2 min. at 3,000 × g in a swinging bucket rotor and the resultant flow-through was discarded. The column was transferred to into a fresh 15 ml collection tube followed by the addition of 0.35 ml of elution buffer to the centre of the purification column membrane and incubated for 2 minutes at room temperature. Followed by centrifugation for 5 minutes at 3,000 × g in a swinging bucket rotor for elution of plasmid DNA. After centrifugation the purification column was discarded. The plasmid DNA was electrophoresed on a 2% agarose gel (to ensure the accurate plasmid size), and stored at -20 °C for downstream applications.

2.2.4 Agarose gel electrophoresis

In this study, DNA (plasmid) was run on agarose gels containing 0.5 µg/ml ethidium bromide. Whereby, separation of high molecular weight DNA was achieved using 0.8 – 1% agarose gels made from 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA) and while 2% agarose gels were used for low molecular weight DNA separation. Electrophoresis was conducted in 1 × TAE buffer at 80 – 100 V in electrophoresis tanks (Bio-Rab laboratories). In each gel, 2 µl of 1kb gene ruler was ran and used as a molecular weight marker for size determination. Thereafter, gels were visualized using a G:Box SYNGENE system in conjunction with the GeneSnap image acquisition software (Syngene).

2.3 Construction of knockdown *rv2414c* mutant strain

There is still a huge molecular biology gap in the determination of gene functions and identification of genetic elements involved in certain biological processes. The CRISPR system is a popular genome editing method due to its simple technical nature. Besides gene disruption, this system has transformed genetic manipulation, and, with ease, has facilitated gene editing for further downstream applications in both eukaryotes and prokaryotes. In this study, this innovative approach was used to create aMTB strain where *rv2414c* has been knocked down.

2.3.1 CRISPR primer design

Primers; *rv2414c* forward BamHI and *rv2414c* reverse BamHI used in this study were designed using gene sequence obtained from mycobrowser (Table 2.3). Thereafter, a PAM site near the site of *rv2414c* was identified. PAM sequences license the R-loop formation by the Cas9-sgRNA complex which is essential for targeted gene knockdown. Thereafter the target sequence was defined by identifying the 24 nucleotides 5’ to your PAM site. These nucleotides define the guide RNA. BamHI restriction sites were added to allow screening downstream and clamps to ensure ligation. Clone manager primer design function was used to label these sequences leading to KD map generation in the accurate orientation.

Table 2.3: primers used for amplification of the *rv2414c* gene region

[redacted]	Sequence	Description
[redacted]	**GGGAGGATCC**ATCCCGCAGCCGTTGAACCGTCGC	Both 34 bps of the *rv2414c* sequence yielding an sgRNA that binds to the dCas9 of PLJR965 and anneal at 60°C
[redacted]	**AAAC**GCGACGGTTCAACGGCTGCGGGAT**GGATCC**	

Bold: Restriction sites included to facilitate fusion of the gene region and plasmid during cloning **Bold:**

clamp

2.3.2 Oligo annealing

To anneal the oligonucleotides, 4 µl of each synthesized oligo (stock concentrations 100 µM in water) was added to 42 µl of annealing buffer to a total of 50 µl in a PCR tube. Thereafter, the incubation conditions to anneal the primers were as follows: 95 °C for 2:00 and 0.1 °C/sec to 25 °C.

2.3.3 Restriction digestion

All restriction enzymes used in this study were purchased from Thermo-Fisher Scientific and all restriction enzyme digestions were performed as per the manufactures instructions using specified buffers.

2.3.3.1 Plasmid PLJR965 digestion

Approximately, 1 µg of extracted PLJR965 plasmid DNA was digested in a total reaction volume of 20 µl with the following reaction components nuclease-free water (15 µl), 10X buffer tango (2 µl), plasmid DNA (1 µl), 20 mM DTT (1 µl), and BsmBI (1 µl). Thereafter, the tube was gently inverted to mix and incubated at 37 °C for 1 hour, followed by heat inactivation at 65 °C for 20 minutes. Finally, agarose gel was ran for digested plasmid visualization and quantified by nanodrop.

2.3.4 Ligation reaction

The ligation reactions (total volume of 20 µl) were set up and incubated at room temperature for 2 hours. A total of 4 reactions was set up with the following ratios of digested plasmid to annealed oligos (1:0, 1:1, 1:2 and 1:3). Each reaction ratio consisted of annealed oligos (variable as per ratio), BsmBI-digested PLJR965 plasmid DNA (~9 ng), T4 DNA ligase buffer (2 µl), T4 DNA ligase (1 µl) and sdH2O (to volume of 20 µl).

2.3.5 Transformation

For transformation, XL 10- Gold Ultracompetent cells were used and transformation was performed as per manufactures instructions. Whereby, four 14 ml BD Falcon polypropylene round-bottom tubes were chilled on ice. Thereafter, NZY^+ broth (specified broth for XL 10-Gold Ultracompetent cells) was heated to 42 °C and the cells were thawed on ice followed by gentle mixing. A volume of 100 µl of cells was aliquoted into each of the pre-chilled tubes and 4 µl of the β-Mercaptoethanol (β-ME) was added to each aliquot of cells. The tubes were gently swirled and incubated on ice for 10 minutes (swirling gently every 2 minutes). Thereafter, 2 µl of the ligation mixture was added to each cell aliquot. The tubes were swirled gently and then incubated on ice for 30 minutes followed by heatpulsing in a 42 °C water bath for exactly 30 seconds. Thereafter, the tubes were incubated on ice for 2 minutes followed by the addition of 0.9 ml preheated (42 °C) NZY+ broth and incubated at 37 °C for 1 hour with shaking at 225-250 rpm. The transformation mixtures (100 µl) were plated on LB agar plates containing kanamycin and incubated at 37 °C overnight.

2.4 Screening the potential knockdown (KD) construct

After observing growth in the plates, colonies were selected randomly and plasmids were extracted as described in section 2.1.4. and screened via restriction digestion as described in section 2.1.5. The restriction enzyme in this case was BamHI and the reaction was set up to a total volume of 20 µl, consisting of nuclease-free water (14 µl), 10X buffer red (2 µl), extracted plasmid DNA (2 µl) and BamHI (2 µl). Thereafter the restriction reactions were incubated at 37 °C for 2 hours followed by agarose gel electrophoresis and gel purification.

2.5 Electroporation into MTB

An MTB culture (5 ml) was grown to stationary phase and inoculated into 100 ml 7H9 media then grown to an OD600nm of 1 – 1.5. Thereafter, the culture was supplemented with glycine to a final concentration of 1.5%, and incubated at 37 ˚C for 16 h. The cells were harvested by centrifugation at 2 630 × g for 10 minutes at room temperature and the pellet washed 4 times by gentle resuspension. The final cell pellet was re-suspended in 2 ml 10% glycerol followed by immediate inoculation into 400 µl aliquot of MTB competent cells. This was then transferred into a 0.2 cm electroporation cuvette (Bio-Rad laboratories), and pulsed using the following conditions: 2.5 kV, 25 µF and 1000 Ω. The cells were rescued immediately with 800 µl 7H9 for 16 h at 37 °C and thereafter plated on Middlebrook 7H11 media containing the appropriate supplements and antibiotics, followed by incubation for 3 – 6 weeks at 37 °C before scoring CFUs.

2.5.1 Screening by PCR

For electroporation confirmation, plasmid was extracted from the resultant colonies as described in section 2.2.1 and PCR was conducted using a set of universal primers (forward and reverse). Taq polymerase was used to set up a 25 µl reaction consisting of 2.5 µl Taq buffer, 2.5 µl dNTPs, 1.25 µl universal forward primer, 1.25 µl universal reverse primer, 2 µl DNA, 0.25 µl Taq polymerase and 15.25 µl sdH_2O. Thermocycling conditions used were as the manufacturers instructions. Whereby, initial denaturation was at 95 °C for 3 minutes, denaturation at 95 °C for 30 seconds, annealing at 60 °C for 30 s, extension at 72°C for 1 minute and final extension at 72 °C for 10 minutes (25 cycles). Below is a table showing the universal forward and reverse primer sequences.

Table 2.5.1.1: screening PCR primers

Primer name	Sequence	Description
Universal Forward	TCATGCCGAAATCAACACCC	Both bind to the PLJR965 plasmid backbone. The universal forward binds is 20 bps long binding at position 71 to 90 of the knockdown plasmid. While, universal reverse is 20 bps binding at position 492 to 511 of the knockdown plasmid and they are 440 bps part. Annealing at 60 °C
Universal Reverse	GGAGAAAGGCGGACAGGTAT	

2.6 Bacterial growth kinetics

2.6.1 Growth curves

Growth curve experiments were used to determine the bacterial growth rate of wild type and mutant strains. Pre-cultures for each strain were set up in 10 ml culture volumes using either 1 ml frozen stocks or a single colony. These were grown at 37 ˚C until they reached stationary phase (OD600nm ≈ 2). The pre-cultures were then diluted to a final OD600nm = 0.05 in 50 ml 7H9 containing appropriate antibiotic, when necessary, and incubated at 37 ºC, with shaking at 100 rpm. Growth was determined by recording OD600nm measurements daily and data was displayed as a scatter plot. Data was analysed using the GraphPad Prism 5.0 statistical software package. Data was analysed using the one-way analysis of variance (one-way ANOVA) test and Tukey Kramer Test post hock test. A $p < 0.05$ was considered statistically significant.

2.7 Biological function elucidation assays

2.7.1 Survival of MTB Strains Exposed to UV Light

Wild type MTB and its mutants were grown to the logarithmic (OD600 of 0.8) and stationary phase (14 days, reaching an optical density OD600 of 2.0) in 7H9 liquid media supplemented with OADC and 100 ng/ml anhydrotetracycline. Next, the cells were serially diluted, and 100 µL of each cell suspension (10-fold diluted) was spread onto 7H10 solid agar supplemented with OADC and glycerol. Then, the plates were treated with UV at doses of 0 mJ(control), 5 mJ, 10 mJ, and 15 mJ and incubated at 37 °C for 3–4 weeks and represented as a scatter plot. Data was analysed using the GraphPad Prism 5.0 statistical software package. Data was analysed using the one-way analysis of variance (one-way ANOVA) test and Tukey Kramer post hock test. A $p < 0.05$ was considered statistically significant.

2.7.2 Survival of MTB Strains in the presence of Mitomycin C (MMC)

The survival of MTB strains was assessed in 7H9/OADC/Tween 80 medium supplemented with 5 ng/ml MMC (wild-type and mutant) and 100 ng/ml ATc at an OD600 of 0.05 using standard CFU methodology at the following time points: 0 h, 48 h, 96 h, and 168 h. Serial 10fold dilutions of cells at each time point were plated on 7H10/OADC/glycerol and incubated for 3–4 weeks at 37 °C. The growth kinetics were monitored by measuring the optical density OD600 at 0 h, 48 h, 96 h, and 168 h after the addition of MMC, data was represented as a scatter plot. Data was analysed using the GraphPad Prism 5.0 statistical software package. Data was analysed using the one-way analysis of variance (one-way ANOVA) test and Tukey Kramer Test post hock test. A $p < 0.05$ was considered statistically significant.

CHAPTER THREE: RESULTS

3.1 Bio-informatics tools

Despite the use of improved molecular technologies, the mycobacterial genome is not extensively studied. This is evident as until date numerous proteins expressed in the MTB genome are labelled "hypotheticals" with no definite biological function. Additionally, these hypothetical proteins may play a huge role in the cellular processes and host-survival of the bacterium. The protein *rv2414c* is a hypothetical protein identified in a study by Chiliza, 2019 which was aimed at identifying protein biomarkers common in LBTI and ATB states. It is hypothesized to play a role in DNA repair mechanisms of MTB, which is an essential process in the survival of the bacterium in host hostile conditions. To test this hypothesize and possibly assign a biological function to this hypothetical protein (*rv2414c*) numerous bioinformatics tools were utilized in this study. Mycobroswer bioinformatics software was used obtain the *rv2414c* gene sequence. Mycobrowser provided the following information; protein type is a coding sequence (CDS) which implies that *rv2414c* has a sequence of nucleotides that correspond with the sequence of amino acids in a protein. CDS proteins are typically characterized by starting with ATG and ending with a stop codon. Mycobroswer also provided the exact position of *rv2414c* which is on the negative strand (NC_000962.3:2711332..2712876) and length of 1 545 bps. Additionally, mycobrowser shows a 27.3% identity (in 538 amino acid overlap) between *rv2414c* and come operon proteins 3 (COMEC or COME3) which is known and required for binding and uptake of transforming DNA. The sequence obtained in mycobrowser database was then used to design CRISPR primers to enable the construction of a KD mutant as well as I-TASSER for further analysis. I-TASSER server is an on-line platform that utilizes algorithms for protein structure and function predictions. In this study, it was used to generate high-quality model predictions of

3D structure and biological function of protein molecules from the protein's amino acid sequences. All images generated from the I-TASSER job are provided in Appendix D.

The *rv2414c* I-TASSER output shows that the predicted secondary structure is largely composed of alpha helixes compared to beta sheets and that these helixes are connected by coils (see output Figure D1 in Appendix D. Additionally, the confidence score under the predicted secondary function ranges at nine (9) for most of the structure, therefore the results are reliable since a higher confidence score implies that the predicated secondary structure is of high confidence.

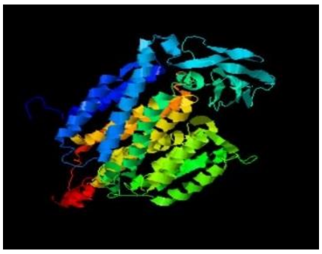
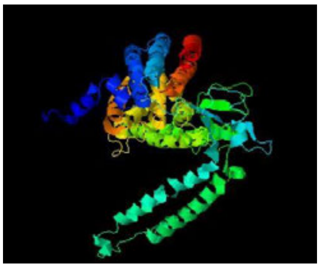
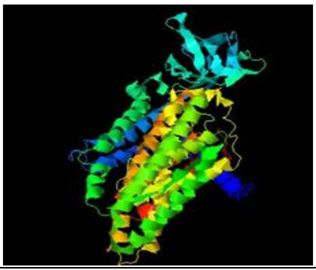
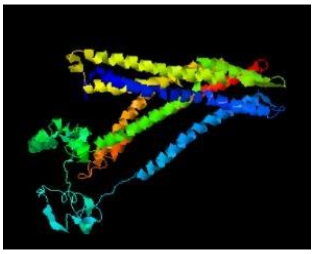
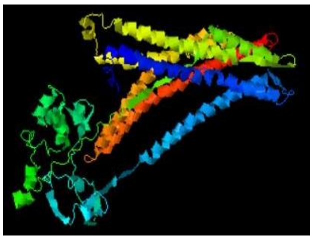
Furthermore, the predicted solvent accessibility shows which sections of the structure are hydrophilic and which are hydrophobic. As shown in the I-TASSER output provided in appendix D (Figure D2), from the beginning of the sequence to the end of the sequence the predicted solvent accessibility ranges mostly between 0-5, suggesting that *rv2414c* is made up of buried residues thus *rv2414c* is hydrophobic.

I-TASSER also provides information on the predicted normalized B-factor that shows the molecules thermal mobility, thus the ability of the amino acids to absorb heat and convert it to kinetic energy. As shown on the I-TASSER output in Figure D3 (Appendix D) *rv2414c* residue 1-100 seem to have a constant B factor, which starts to peak from residue 100-500. The highest thermal mobility is seen at the end of residue 500 within the last coils. Additionally, from residue number 0-100 the normalized B-factor is below the base line (thus below zero) which is a negative value which implies that the residue is relatively more stable in the structure. In contrast, from residue 100-200 the normalized B-factor peaks above the base line at 6 points suggesting that the residues in the range are not stable within the structure. From residue 200300 as well as from residue 300-400, the normalized B-factor remains under the base line and only peaks once in each range, suggesting that those residues can be regarded stable.

Lastly, from residue 400-500 the normalized B-factor peaks mildly (low peak) once and peaks at the terminal ends this suggests the residues in the range are likely not stable based on the high peak at residue 500.

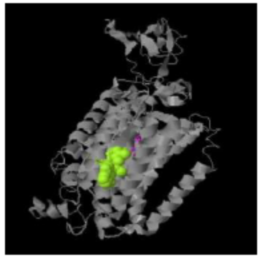
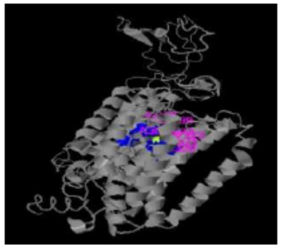
Under the top 10 threading templates as predicated by I-TASSER: *rv2414c* is closely related to the following queries in descending order: 5uj9A which is classified as a transport protein, expressed by homo sapiens. Secondly, 7s9Ya which is classified as a membrane protein, expressed by *Escherichia coli* BL21 (DE3). Thirdly, 4me3A which is responsible for replication, expressed by *Escherichia coli*. Fourthly, 7dp3A which is classified as a DNA binding protein, expressed by *Escherichia coli*. In fifth place is 5yd1A which is classified as a metal transport, expressed by Homo sapiens. In sixth position is 6rawA which is responsible for replication, expressed by Trichoplusiani. In seventh position is 6afwA which is classified as a membrane protein, expressed by *Saccharomyces cerevisiae*. In eighth place is 6zmqA which is classified as a membrane protein, expressed by *Escherichia coli*. In ninth place is 7e9Sa which is classified as a transferase, expressed by *Escherichia coli* BL21(DE3)). Lastly, in the tenth position is 6mgyA which is classified as a membrane protein, expressed by homo sapiens. Interestingly, these top 10 queries suggest that *rv2414c* is mostly related to eukaryotic proteins with most of them involved in DNA replication, which is a process closely, associated to DNA repair that is the hypothesized biological function of *rv2414c*.

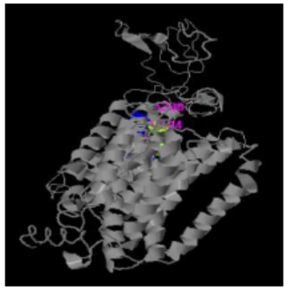
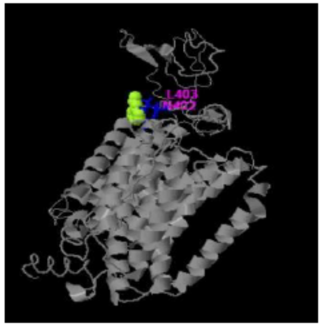
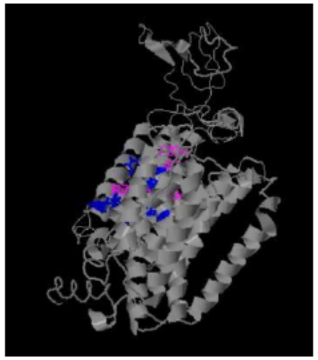
Table 3.1.1: Table showing the top 5 final models predicted by I-TASSER.

Model	**C score**	**Predicted structure**
1	-2.40 (Estimated TM-score = 0.14)	
2	-1	
3	-2.46	
4	-3.26	
5	-4.09	

The c-score is an indication of the global accuracy of the model. The c-score ranges from -5 to 2 and any score greater than -1.5 indicates a model of correct global topology. In the case of *rv2414c* four of the models generated have a score lesser than -1.5 while only one of them (model 2) is around -1.55 which is only 0.05 above the desired value of -1.5. Thus, model 2 can be safely regarded as the closet structural model to the protein of interest *rv2414c*. ITASSER output showing the top 10 identified structural analogs in the Protein Data Bank (PDB) is provided in Appendix D (Table D1), shows that the higher the TM value, the closest the structural analogue is to the protein of interest (*rv2414c*). From the obtained result, protein 7s9yA is the closet to *rv2414c* with a TM score of 0.858 hence the highest identity value of 0.079.

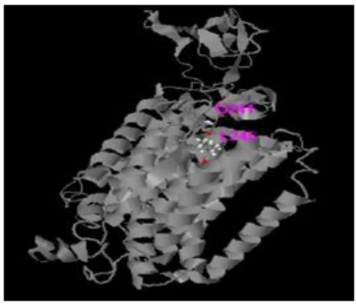
Table 3.1.2: showing the Predicted biological function using COFACTOR and COACH ligand binding site

Rank	C-score	Cluster size	PDB Hit	Lig Name	Complex	Ligand Binding Site Residues and Image (structural alignment)
1	0.08	3	3wmnA	MQ8	Rep, Mult	382, 386 
2	0.06	2	2ckjC	FES	Rep, Mult	299, 300, 303, 320, 321, 322, 323 

3	0.06	2	4brbC	78N	Rep, Mult	390, 394, 398	
4	0.06	2	3sjqC	PHU	Rep, Mult	402, 403	
5	0.03	1	N/A	N/A	N/A	207, 339, 342, 346, 385, 389	

The c-score in this section ranges from 0-1 and a higher score indicates a more reliable predication. *rv2414c* shares ligand binding sites with the following enzymes in increasing order; 3sj9C, 4brbC, 2ckjC and 3wmnA. Enzyme 3wmnA having the highest score implies that the binding site of *rv2414c* is most similar to it.

Enzyme Commission (EC) numbers and active sites

Rank	**C-score^{EC}**	**PDB Hit**	**TM-Score**	**RMSD^{a}**	**IDEN^{a}**	**Cov**	**EC number**	[redacted]
1	0.095	2x38A	0.363	6.92	0.045	0.564	2.7.1.153	N/A
2	0.093	2rd0A	0.309	7.59	0.033	0.512	2.7.1.153	241, 245 
3	0.093	1occA	0.375	6.30	0.041	0.549	1.9.3.1	N/A
4	0.092	1m56A	0.379	6.43	0.049	0.568	1.9.3.1	N/A
5	0.091	2vkzG	0.351	7.31	0.029	0.562	2.3.1.38 3.1.2.14	N/A

Cscore EC is the confidence score for the EC number prediction and it ranges from 0-1, where the highest score indicates a more reliable EC number prediction. Thus in the case, the 1st hit (enzyme 2x38A) is the most reliable with a c-score EC of 0.095.

Thereafter, STRING bioinformatics tool was used to establish the networks of *rv2414c* with other genes and proteins which have an influence on the predicated role of DNA repair. Proteins and their functional interactions are crucial in forming the backbone of cellular mechanisms (Hengsawas, 2016). Additionally, cellular life is largely dependent on a complex network of functional associations. Hence, it is important to fully understand the biological phenomena behind each connectivity network. STRING is a database best suited to carry out this task as it collects, scores and integrates all public sources of protein-protein interaction information and complements them with computational predications. Physical (direct) and functional (indirect) interactions have a huge influence on the biological functions of proteins hence it is important to understand both the direct and indirect networks proteins have particularly during characterization studies. The gene *rv2414c* was searched on STRING to obtain the genes connectivity network and STRING provided the following networks as shown on Figure(s) D5 and D6 shown in Appendix D. From the STRING output as shown in Figure D5, *rv2414c* has a very close association with *rv2413c* (upstream) and *rv2415c* (downstream) which are both conserved hypothetical. From these connections, *rv2414c* has a more defined relationship with *rv2415c* protein as they share a confidence score of 0.939, which is 0.009 more than the confidence score of *rv2413c* (0.930). Additionally, *rv2414c* connects to *rv2415c* with three edges, which signify gene neighbourhood, experimentally determined and text-mining. Similarly, *rv2414c* connects with *rv2413c* with four edges, which signify text-mining, coexpression, gene fusions and gene neighbourhood. The hypothetical protein, *rv2415c* is a hypothetical protein, with certain similarities (49.2% identity in 61 amino acid overlap) within the C-terminal part to comE operon proteins. *rv2413c* is a conserved hypothetical protein, highly similar to O33133|MLCL536.07c|ML0603|Q49756|G466975|B1937_F2_36 hypothetical protein from *Mycobacterium leprae*. Additionally, *rv2414c* forms close connectivity networks with the following genes, dprA, recA and cinA (in descending order). The gene dprA is a hypothetical protein hypothesized to assist loading recA onto single stranded DNA. Additionally, dprA is suggested to be involved in the protection of incoming DNA. recA is reported as a recombinase protein which plays an essential role in the recombination DNA-repair system since it acts as a DNA pairing enzyme which is responsible for promoting an exchange of homologous DNA strands. While cinA exhibits functions in the MTB virulence, detoxification as well as adaptation and its further reported to have strong similarity to competence damage proteins CinA of *Bacillus subtilis* and *Streptococcus*. *pneumonia*. As shown in Figure D6 (Appendix D), *rv2414c* has further connections with the following genes; the enhanced intracellular survival (eis) and the DNA polymerase I (polA) as well as proteins *rv3242c*, *rv3737*, *rv2897c* and *rv2417c*. The eis gene is an effector molecule released into the host cell thereby affecting host immune response by negatively modulating inflammation.

In MTB, the eis gene is involved in the intracellular survival of *M. smegmatis*. MTB eis is reported to play an essential role in the modulation of macrophage autophagy, inflammatory responses, as well as cell death in a ROS-dependent pathway (Shin *et al*., 2010). On the other hand, polA gene is characterized as one of the most widely distributed phage genes due to its extensive role in DNA replication of ensuring accurate as well as efficient replication of the genome in order to ensure the maintenance of the genetic information for transmission through generations (Kazlauskas *et al*., 2016). Additionally, *rv2414c* has also formed network connections with the following proteins; *rv3242c*, *rv3737* and *rv2897c*. The conserved hypothetical protein, *rv3242c* has high similarity to N-terminus of Q9CCI9|ML0776 hypothetical protein from *Mycobacterium leprae* which has not been fully characterized. The protein *rv3737* has a general functional category of being involved in cell wall and cell processes, however the specific individual biological function of the gene has not been established as of yet. Furthermore, with *rv3737* being a probable conserved transmembrane protein, it shares similarities with other hypothetical proteins; AAK61331|THRE threonine export carrier from *Corynebacterium glutamicum* (*Brevibacterium flavum*) (489 aa) and Q9X8J0|SCE9.17 putative membrane protein from *Streptomyces coelicolor* (578 aa).

The conserved hypothetical proteins *Rv2897c*, has a possible Mg-chelatase function and it is highly similar to hypothetical proteins and chelatases for instance; Q9RTV0|DR1656 mg($^{2+}$) chelatase family protein from *Deinococcus radiodurans* (519 aa) and Q55372|SLR0904 hypothetical 55.1 KDA protein from *Synechocystis sp*. strain PCC 6803 (509 aa). Magnesium chelatase (E.C.6.6.1.1, Mg-chelatase) is an enzyme characterized for its ability to insert a magnesium ion (Mg^{2+}) into protoporphyrin IX to generate Mg-protoporphyrin IX which is essential for redirecting intermediates to achieve suitable conditions for bacterial photosynthetic growth. Lastly, *rv2417c* is a conserved protein which possibly binds long-chain fatty acids for instance; palmitate and it is also hypothesized to have a role in transportation of lipids or fatty acids.

3.2 Construction of knockdown *rv2414c* mutant strain

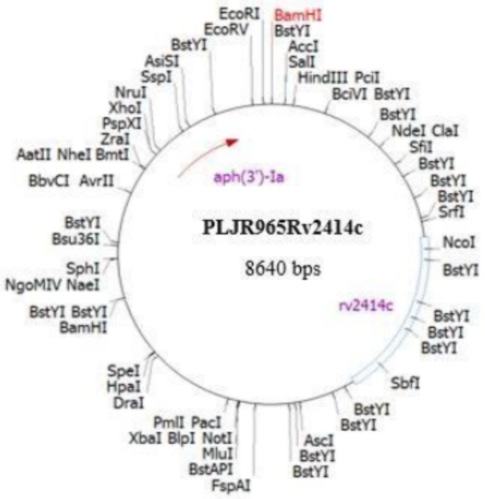
The CRISPR system was used to generate the knockdown plasmid, with in-silico cloning done on clone manager software to ensure that the process of creating a knockdown is error free. This process entailed the creation of CRISPR cloning primers followed by annealing the oligos as described in section (2.3.2). Thereafter, the cloning plasmid PLJR965 was linearized using the restriction enzyme BsmBI, which creates sticky overhangs allowing ligation to the annealed oligos as described in section 2.3.4 and shown in Figure 3.2.1 below. The size of the linearized

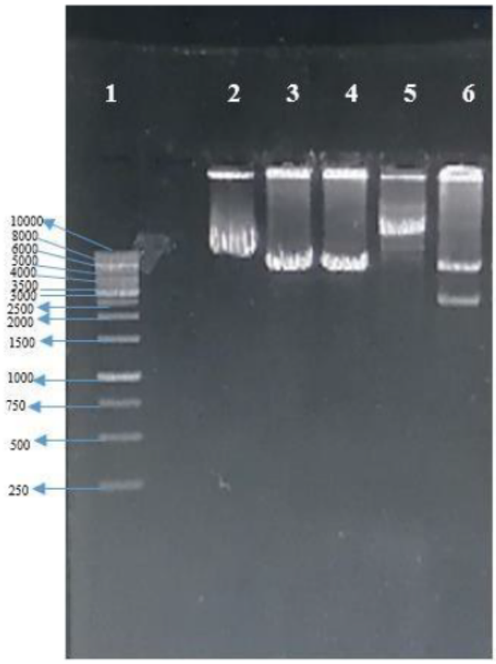
PLJR965 is 8606 bps as visualized after gel electrophoresis as shown on Figure 7 below. After ligation and transformation colonies were selected randomly and restricted with BamHI restriction enzyme as described in section 2.4. To ensure that the cloning was accurate and in the proper orientation, the restriction of the potential KD strain was done in-silico on clone manager to obtain the expected band sizes. From these colonies, a successful knockdown strain was confirmed as shown in the Figure below.

A

Restriction enzyme	Expected size (bps)
BamHI	2569 and 6071

B





C

Figure 3.2.1: Restriction profile of PLJR965Δ*rv2414c*. **(A)** Table showing the different restriction endonucleases used and the expected fragment sizes, resultant from restriction digest of the PLJR965Δ*rv2414c* construct. **(B)** Plasmid map of PLJR965Δ*rv2414c* showing the fused annealed oligos (light blue), cloned into the PLJR965 backbone. **(C)** Agarose gel showing the result of restriction enzyme profiling of PLJR965Δ*rv2414c*.**[Lane 1]** Thermo-Fisher Scientific™1kb Generuler, 250-10,000 bp, **[Lane 2]** unrestricted PLJR965Δ*rv2414c* (8640 bps), **[Lane 3 & 4]** restricted PJR965 plasmid with BsmBI (8606 bps) and **[Lane 6]** BamHI screening restriction (2569 & 6071 bps).

After colony confirmation, the KD plasmid was electroporated into MTB H37Rv to create a complete knockdown system. This electroporation was confirmed by PCR (with the expected size of 507 bps) as shown on Figure 2.8 below.

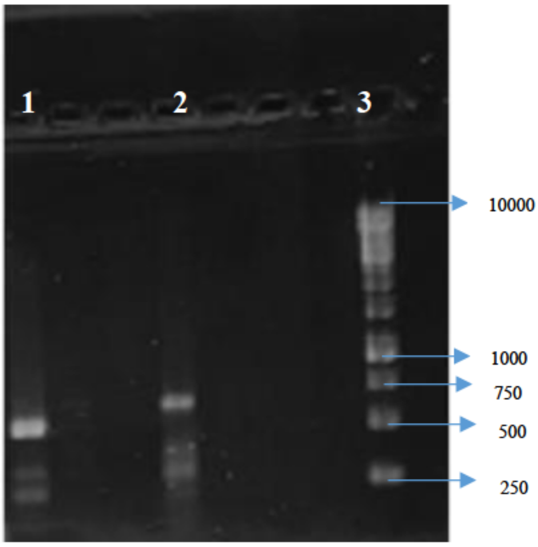


Figure 2.8: PCR screen for knockdown mutant of PLJR965Δ*rv2414c*. **[Lane 1]** 507 bps amplicon confirming the successful electroporation into MTB, **[Lane 2]** not considered in this study (incorrect fragment size) and **[Lane 3]** Thermo-Fisher Scientific™1kb Generuler, 250-10,000 bp.

3.3 Bacterial growth kinetics

3.3.1 Growth curves

To assess the effect *rv2414c* has on the bacterial growth, a growth analysis assay was carried out on the following strains; WT, Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc).

Anhydrotetracycline (ATc) is a derivative of tetracycline with no antibiotic activity, designed for use with tetracycline-controlled gene expression systems in bacteria. Additionally, it is a suitable antibiotic for use with the PROTet Vectors or other vectors or systems designed upon elements of the (Tn*10*) tet operon. In this study primers where designed to serve as a small guide RNA cloned into a PLJR965 plasmid downstream the tet operon which is known for promoting and responding to tetracycline. Upon the addition of ATc, tetracycline will induce the expression of the Cas9 enzyme, which will bind to the region specified by the sgRNA without cleaving the molecule, thus will inhibit the activity of RNA polymerase resulting in transcription interference. The ATc concentration used in this study was 100 ng/ml, which is the maximum concentration for suppression in MTB. All the above-mentioned strains were subjected to similar controlled growth conditions over a period of 12 days taken OD600nm readings daily. All strains used in this study followed a similar growth pattern as shown in Figure 3.9. The results were considered statistically significant (*) at $p < 0.05$. For most time points, the was no significant difference as the p value determined by statistical analysis was > 0.05. However, at time 120 hours the was a significant difference in the growth pattern of these strains as $p = 0.0059$ ($p < 0.05$). At this time point WT had the highest growth compared to the other strains, thus WT growth at t=120 hours was significantly higher than Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc).

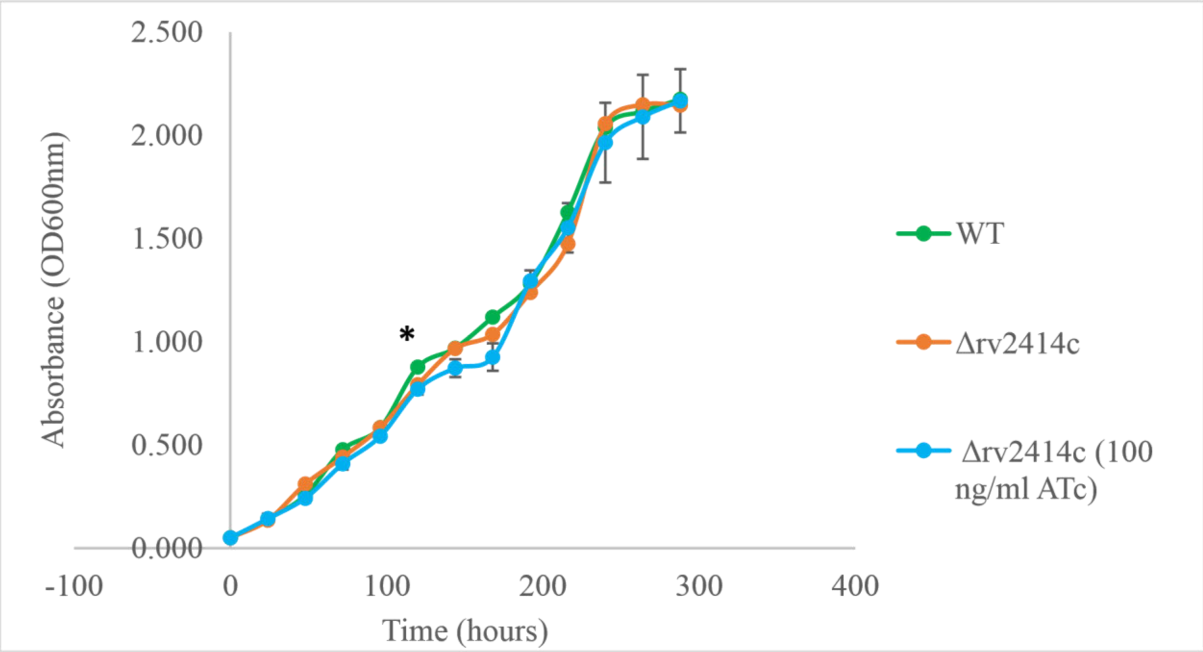


Figure 3.9: Growth analysis of the WT, Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc) in Middlebrook 7H9 containing Middlebrook OADC at 37 ˚C, over 12 days. The data represent an average of three independent experiments.

3.4 Biological function elucidation assays

As previously reported in the literature review, there are multiple ways of inducting bacterial DNA damage in mycobacterium and as a result, the bacterium will employ means to response to DNA damage. This will ensure the preservation of genetic material that is stable and has integrity allowing the bacterium to evade host immune response and successfully cause disease. In order to characterize and assign a possible function to the gene *rv2414c* (hypothesized to be involved in DNA repair), the various strains used in this study; WT, *Δrv2414c* (*rv2414c* deficient MTB H37Rv strain constructed using the CRISPR-Cas9 system) and knock down strain induced with ATc [100 ng/ml], were subjected to treatment with DNA damaging agents (exposure to UV and induction with MMC compound) and the growth was monitored by taking OD reading at OD_{600nm} daily.

3.4.1 Survival of MTB Strains Exposed to UV Light

Post incubation, the plates from both for log and stationary phases were assessed for possible growth represented by colony forming units (CFU/ml) to see the effect UV exposure has on the growth of the different strains in similar growth conditions. Additionally, this is to elucidate if the presence or absence of the gene *rv2414c* has an impact on the growth of the different strains, thereby assessing if the gene contributes significantly and is directly involved in MTB DNA repair. The results were considered statistically significant (*) at $p < 0.05$.

3.4.1.1 Logarithmic phase

The WT strain grew optimally with no exposure to UV (0 mJ), serving as a control. At 0 mJ UV; the other strains (*Δrv2414c* and *Δrv2414c* (100 ng/ml ATc) showed growth similar to the WT as shown in Figure 10 below. This observation is supported by the results from the bacterial growth curve performed in this study, which showed that *rv2414c* has no effect on the bacterial growth rate. Upon exposure to UV at different doses (5, 10 and 15 mJ) a general trend, of a slight decrease in the survival and viability rate of cells in each strain was observed. At the different UV doses of 5, 10 and 15 mJ, the p values were each > 0.05 when comparing the survival rate of different strains. At 5 mJ ($p = 0.3516$), at 10 mJ ($p = 0.0525$) and at 15 mJ ($p = 0.0885$) thus no significance was observed in the data.

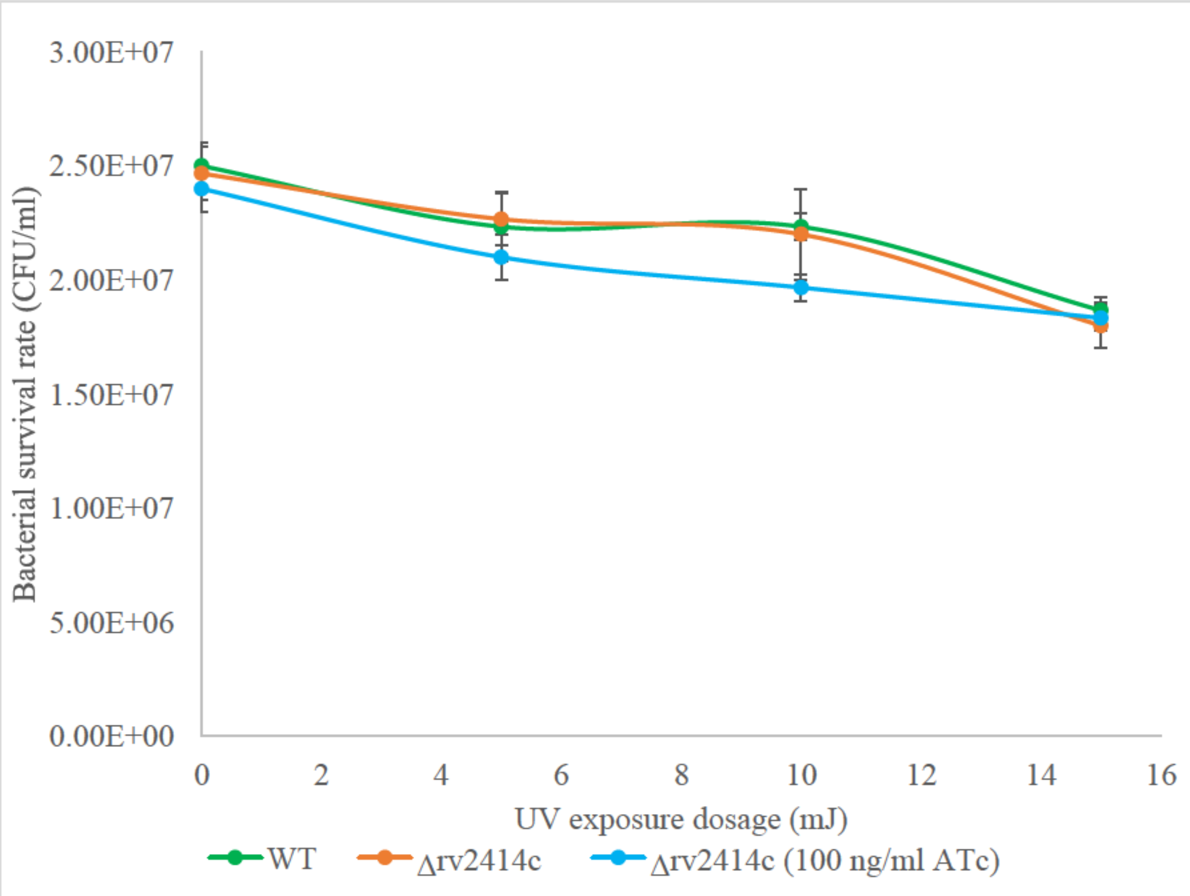


Figure 10: Viability of wildtype MTB (WT), Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc) (at logarithmic phase) exposed to UV at different dosages of 5, 10, and 15 mJ

3.4.1.2 Stationary phase

A general trend showing a decline in cell survival rate was observed for all strains (WT, *Δrv2414c* and Δ*rv2414c* (100 ng/ml ATc)) as shown in Figure 11 below. The results were considered statistically significant (*) at $p < 0.05$. At UV exposure dose of 0 mJ, the difference in the survival rate of the strains was not significant as $p > 0.05$. At UV, 5 mJ and 10 mJ a significant difference between WT and *Δrv2414c* was observed $p = 0.0473$ and $p = 0.0026$ respectively ($p < 0.05$). At UV exposure 15 mJ, survival of cells had decreased in all strains. However, the decrease was not significant as $p = 0.0770$ ($p > 0.05$).

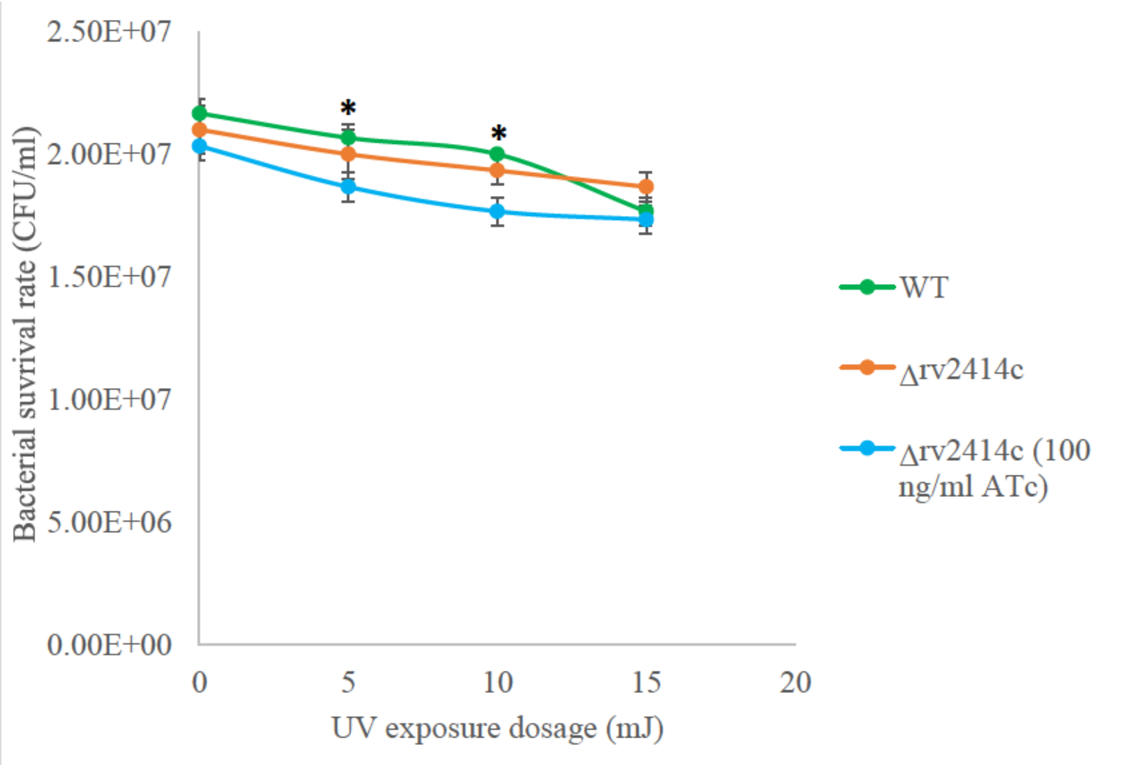


Figure 11: Viability of wildtype MTB (WT), Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc) (at stationary phase) exposed to UV at different dosages of 5, 10, and 15 mJ UV.

3.4.2 Survival of MTB Strains in the presence of Mitomycin C (MMC)

After the addition of the DNA damaging chemical MMC (5 ng/ml) to the Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc), samples were platted out immediately (t= 0 hours). The strain with MMC were labelled; Δ*rv2414c* (MMC) and Δ*rv2414c* (100 ng/ml ATc) + MMC. Thereafter, at the different time intervals of 48, 96, 168 hours, followed by incubation for 3 weeks. Post incubation, possible growth was represented by colony forming units (CFU/ml) to see the effect MMC has on the growth and survival of the different strains in similar growth conditions. Additionally, this is to elucidate if the presence or absence of the gene *rv2414c* has an impact on the growth of the different strains, thereby assessing if the gene contributes significantly and is directly involved in MTB DNA repair.

The results were considered statistically significant (*) at $p < 0.05$. The WT strain grew optimally well as, serving as a growth control as shown in Figure 12 below. The WT showed a gradual increase in cell survival at the different time intervals (0, 48, 96 and 168 hours). At 48, 96 and 168 hours, the WT p values were 0.0012, 0.0001 and 0.0001 respectively ($p < 0.05$). At all-time points the WT bacterial survival rate was significantly higher than all the other strains. Interestingly, Δ*rv2414c* was the strain with the second highest rate of surviving cells after the WT. Interestingly, Δ*rv2414c* (100 ng/ml ATc) + MMC and Δ*rv2414c* (MMC) started to show a decrease in the cell survival at 48, 96 and 168 hours. However, no significant difference observed between the strain with ATc+ MMC and those with just MMC no significance.

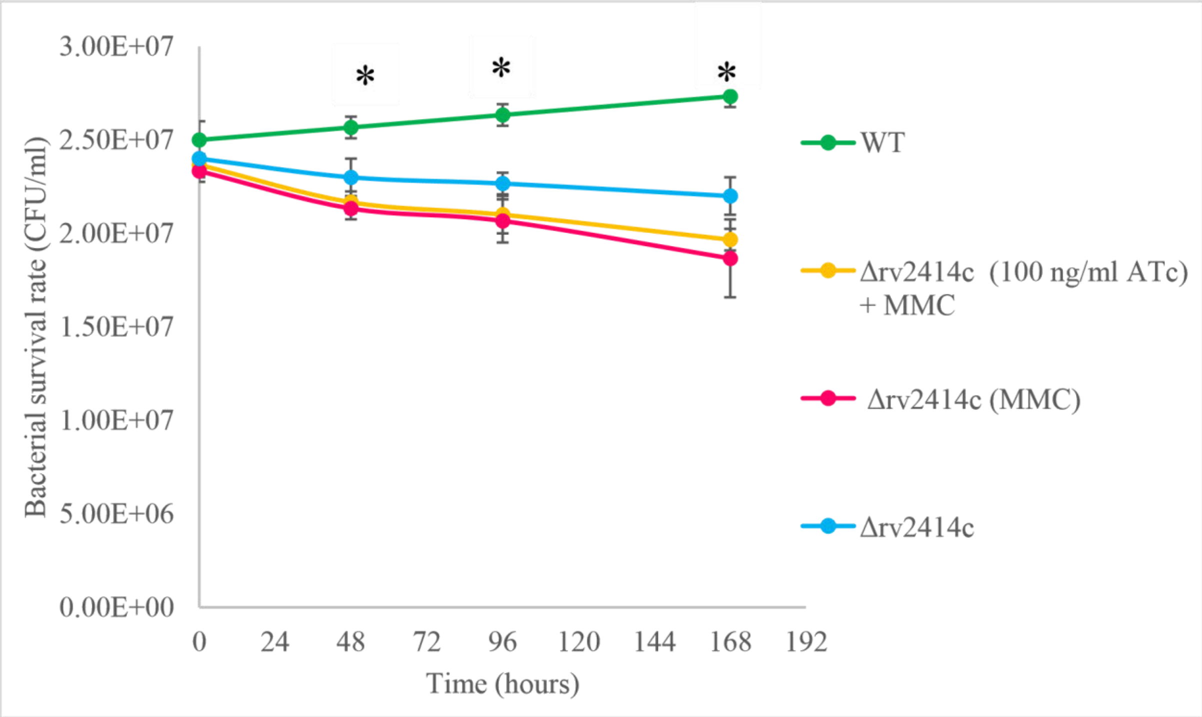


Figure 12: Viability of Δ*rv2414c*, and Δr*v2414c* (100 ng/ml ATc) after addition of Mitomycin C and the viability of WT and Δ*rv2414c*. Samples were platted at different time intervals of (0, 48, 96 and 168 hours).

4 DISCUSSION

Until date, MTB (causative agent of TB) remains a major pathogen for mortality (Elchennawi and Ollagnier de Choudens, 2022). This is an increased threat to the public health as the currently available vaccine Bacillus Calmette Guerine (BCG) is known to be poorly effective against adult TB coupled with the emergence of drug resistant MTB strains (Adesanya *et al*., 2021). Subsequently, this has led to an urgent need for identifying newer and more effective vaccine candidates. Despite the MTB genome been extensively studied over the past two decades, the biological functions of 27% (1051/3906) of proteins encoded within the MTB genome are yet to be determined and these proteins are annotated as hypothetical proteins (Yang *et al*., 2019). Characterization of hypothetical proteins is important in understanding the MTB genome functionality, as they could be involved in major pathways that lead to the overall survival of the pathogen, allowing it to successfully propagate a disease (Mazandu and Mulder, 2012).

The ability to maintain genomic integrity is vital for the survival and propagation of any bacterial species (Darmon and Leach, 2014). Once in the host cell, MTB is subject to various biological niches where the bacteria encounters stressful conditions in the form of oxidative, nutritional as well as genotoxic stress leading to a wide range of mutations within the genome and challenging the genome integrity of the bacterium (Zegeye, 2013). A typical example of this is during persistence of MTB in the phagosome of host macrophages, which is only possible because MTB has developed mechanisms to impede the host immune response (Gengenbacher and Kaufmann, 2012). DNA repair is an essential process in the maintenance of genome stability by remedying the DNA lesions that occur from DNA damaging agents (Maslowska *et al*., 2019).

In this study, we report the characterization of a putative DNA repair protein, encoded *rv2414c* by assessing the growth kinetics and biological assay response using the *in silico* approaches and generation of a knockdown mutant.

The gene *rv2414c* is a CDS protein and has similarities (amino acid overlaps) with the come operon proteins 3 (COMEC or COME3) found in the plasma membrane, responsible for metal ion binding (with a zinc ligand) in the biological process of the establishment of competence for transformation of DNA. DNA transformation is a biological process of horizontal gene transfer, whereby bacteria takes up foreign genetic material from the environment. In *Neisseria* species and other gram-negative organisms, the process of DNA uptake is regarded as a passage across the outer membrane (Makino *et al*., 1991). Whereby, the resultant DNA molecule gains access to the cytoplasm by crossing the murein layer and the cytoplasmic membrane where it might undergo recombination as well as integration into the host chromosome in a RecA-dependent manner, (Sánchez *et al*., 2019). RecA is a recombinase functioning in recombinational DNA repair in bacteria (Sheng *et al*., 2019). This similarity suggests that indeed *rv2414c* is involved (whether minor or major) in MTB DNA repair.

Protein structures can help in elucidating gene functions and I-TASSER relived that *rv2414c*'s predicted secondary structure consists over 500 alpha helixes compared to beta sheets. Secondary structures mainly consist of the alpha-helixes and beta-sheets and the percentage of each structure has a significant influence on proteins functionality (Nava and Kreinovich, 2012). Alpha helixes are stable and don't undergo mutation easily and designable compared to β-sheets in natural protein (Abrusán and Marsh, 2016).This implies that *rv2414c* is a stable protein that can be designed easily.

Another structural property reviled by I-TASSER is that *rv2414c* is a hydrophobic protein, thus it is not soluble in water nor does it interact with water molecules. Hydrophobic interactions are important for protein folding (Newberry and Raines, 2019). Additionally, hydrophobic interactions ensures that a protein is stable and biologically active, as they allow the protein to decrease in surface area, reducing the undesirable interactions with water (Ma *et al*., 2020).

I-TASSER further gives information on the normalized B–factor of a protein molecule. Determination of a B-factor is an extremely important feature as it gives information about the inherent thermal mobility of residues in proteins but often overlooked in the structure prediction procedures. At temperature of absolute zero, atoms within a protein stay at the equilibrium position of lowest energy (Sun *et al*., 2019). However, as the temperature increases, the ambient thermal energy causes the oscillation of atoms around the equilibrium position to an extent dependent on the relative location on the 3D structure and the interaction with ligand as well as solvent atoms this measured as a B-factor (Yang and Zhang, 2015). However, there are limitations to a B-factor, hence, the need to calculate a normalized B-factor with a Z-score-based transformation and I-TASSER does this. The overall analysis of the normalized B-factor for *rv2414c* suggests that the residues present in the molecule are stable.

The rest of the I-TASSER output shows that most proteins *rv2414c* is related to are involved in DNA repair. This is seen in the top ten (10) threading templates show that *rv2414c* is mostly associated to eukaryotic proteins, involved in DNA repair. This strongly supports the hypothesis that *rv2414c* could possibly be involved in DNA repair.

Additionally, the STRING output further relived that *rv2414c* forms structural networks with genes that are related to DNA repair. These genes are dprA, recA and cinA and they are key proteins in the biological process of DNA repair. Therefore, interaction will these proteins suggests that *rv2414c* could have a similar role as protein networks does affect their functionality. This is supported by a study conducted by (VanderSluis *et al.,* 2018), which was aimed at understanding the integration of large-scale genetic and protein-protein interaction networks, highlighting principles that apply to both unicellular and more complex systems, including human cells. This study validated that, genetic interactions within the same biological process or functional module is commonly connected by coherent subsets of genetic interactions, this is known as "monochromaticity". The *in vitro* part of the study was done to assess the effect of *rv2414c* in bacterial growth and assign a biological function through DNA damaging assays.

The bacterial growth curve showed that all strains had a growth curve that was within the same range, as further supported by the one-way ANOVA results as it showed no significance for 12 out of 13 time points. However, at t= 120 the WT can be considered an outlier in the data as it was notably high and $p < 0.05$. The time frame, t = 120 is a point whereby all the strains are in exponential or logarithmic phase of the bacterial growth curve. As a result, a significant increase in growth was observed for the WT. This was expected as the WT genome was intact and not edited, therefore optimum growth. The WT has a complete genome that has not been altered, thus it can employ DNA repair mechanisms to counteract the damage. The data obtained from the growth curve showed no significance, it implies that the loss of *rv2414c* resulted in no significant defects in bacterial growth rate. This answered the research question and showed that *rv2414c* is not essential for bacterial growth and survival.

The next step was then to subject the strains to two (2) DNA damaging agents (UV and MMC) and observe the response of the strains. The ideal scenario would be that UV and MMC would result in a decrease in survival of cells. Thus, strains induced with these agents would result in the least growth, while the WT would have the optimum survival rate as the entire genome is present and unaltered.

UV radiation is known to induce deleterious effects in all living organisms ranging from prokaryotic bacteria to eukaryotic lower and higher plants, animals and humans (Roy, 2017). In MTB, UV radiation induces two of the most abundant mutagenic and cytotoxic DNA lesions; cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs), this type of damage is known as the formation of bulky adducts as previously reviewed in this study.

Upon no UV exposure (0 mJ) at the logarithmic phase, all the strains (WT, Δ*rv2414c* and Δ*rv2414c* (100 ng/ml ATc), showed a similar growth pattern, of a steady decline in cell survival. Additionally, at UV (5, 10 and 15 mJ) $p > 0.05$, thus no significant difference observed in the survival of the strains upon UV exposure. Since survival was not too far from each other, this suggest that's *rv2414c* does not have a significant role to play in DNA repair. It further, shows that yes, UV does induce damage to MTB DNA but within the genome (more evident in the log phase results), the are key genes responsible for fixing DNA lesions caused by UV exposure. These key gene could be recA and LecA. This is supported by a study conducted by Brzostek *et al*. (2021), as they noticed that the recA deletion resulted in the strain being extremely sensitive to UV exposure (the percentage of survival viability significantly dropped) compared to the WT, this shows that indeed recA is involved as a key DNA repair protein in MTB (Brzostek *et al*., 2021).

In a similar manner, MMC is a compound known to induce various types of DNA damage resulting in significant cytotoxicity to cells (Joseph *et al*., 2022). In bacteria, MMC results in the following DNA lesions; DNA cross-links, DNA alkylation, and the production of reactive oxygen species (ROS) (Magerand *et al*., 2021). The WT grew well and this can be attributed to the WT having a complete genome that can easily counteract DNA damage, serving as a control. All strains with MMC, thus Δ*rv2414c* (100 ng/ml ATc) + MMC and Δ*rv2414c* (MMC) had a similar growth pattern with no significance. Whereby, those with MMC only had the least growth and this was expected as MMC at time point 48, 96 and 168 hours. Compared to the WT the was a significant difference as $p < 0.05$. However, no significance was observed between the Δ*rv2414c* (100 ng/ml ATc) + MMC and Δ*rv2414c* + MMC cell survival rate. This implies that *rv2414c* is not involved in MTB DNA repair. This is supported by a study conducted by Brzostek *et al*. (2021), as they noticed that the recA deletion resulted in the strain being extremely sensitive to MMC induction (the percentage of survival viability dropped significantly) compared to the WT, this shows that indeed recA is involved as a key DNA repair protein in MTB, which was not the case in this study (Brzostek *et al*., 2021).

5 CONCLUDING REMARKS

The study reported in this MSc represents the first extensive characterization of a putative DNA repair hypothetical protein (*rv2414c*) since its identification by(Chiliza *et al*., 2019). The purpose of identifying the possible role in MTB repair. However, the hypothesis of this study was rejected due to results observed in this study. The gene did not show any significant role in the bacterial growth rate of MTB. Additionally, this study showed that absence (Δrv*2414c* and Δ*rv2414c* (100 ng/ml ATc)) of *rv2414c* does not have an impact on repairing the DNA damage caused by UV exposure. The MMC assay showed a slight decrease in the cell survival (CFU/ml) of strains were it was added with no significance. This gives further room for this study to be developed into a PhD research project with the following recommendations.

5.1 Recommendations for future studies

Further studies can be done on the strains used in this study to determine the extent of damage caused by DNA damaging agents and the different mechanisms used for DNA repair.

Additionally, creating a knockout via homologous recombination and subjecting the strains to more DNA damaging agents will also add valuable information in the attempt to characterize *rv2414c*. Additionally, protein work (expression, purification and analysis) would also be a way of advancing the study as *rv2414c* still remains a conserved hypothetical protein. Thus its' existence still needs to be proven and assigned a biological assay.

6 Appendices

6.1 Appendix A: Culture media

Middlebrook 7H9

4.7 g Difco Middlebrook 7H9 powder, 2 ml glycerol, 900 ml dH_2O, 10 ml 100X glucose-salt, 2 ml Tween80

Middlebrook 7H10

19 g Difco Middlebrook 7H10 powder, 5 ml glycerol, 900 ml dH_2O, 10 ml 100X glucose-salt

100 × Glucose-salt

Glucose salt 10 g glucose, 4.25 g NaCl, 500 ml dH_2O

Tween80 (25%)

10 ml Tween80 dissolved in 40 ml dH_2O then sterilized by filtration

X-gal (2%)

1 g X-gal in 50 ml deionised DMF

Media was sterilized by autoclaving at 121 °C for 20 min

6.2 Appendix B: Solutions

1 M Tris-HCl	60.56 g Tris, pH 8.0 with HCl, 500 ml dH_2O final volume
0.5 M EDTA	18.6 g powder, pH 8.0 with NaOH, dH_2O
Soln I	50 nM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8.0
Soln II	0.2 M NaOH, 1.0% SDS
Soln III	5 M potassium acetate, 11.5 ml glacial acetic acid, 88.5 dH_20, pH 4.8
TE buffer	1 ml 1 M Tris-HCl, 2 ml 0.5 M EDTA
TAE buffer	242 g Tris base, 100 ml 0.5 M EDTA, 57.1 ml glacial acetic acid, pH 8.0, 1000 ml final volume dH_2O
Chloroform: Isoamyl alcohol	24 ml chloroform, 1 ml isoamyl alcohol
Phenol: chloroform	1 ml phenol, 1 ml chloroform
Sodium acetate	3M sodium acetate dissolved in dH_2O (autoclaved)
CTAB/NaCl	4.1% NaCl, 10% N-cetyl-N, N, N-trimethyl ammonium bromide dissolved in dH_2O (filter sterilized)

6.3 Appendix C: Molecular weight marker(s)

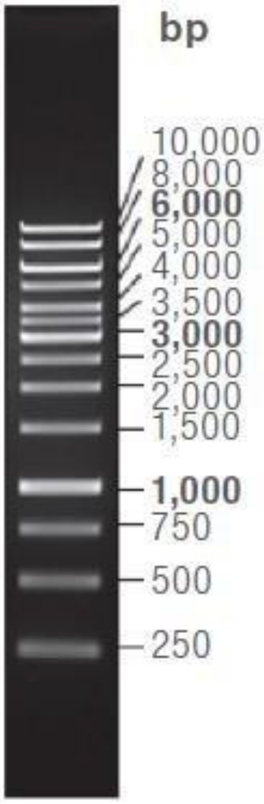


Figure C1: GeneRuler 1 kb DNA Ladder

6.4 Appendix D: Bioinformatics analysis

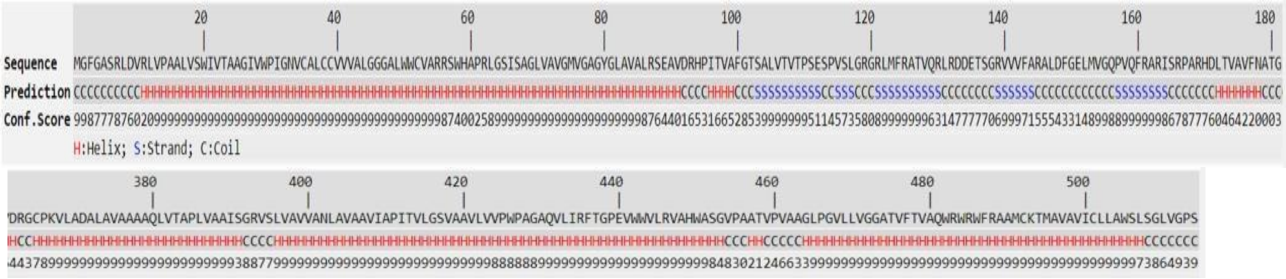


Figure D1: I-TASSER output showing the predicted secondary structure of *rv2414c*

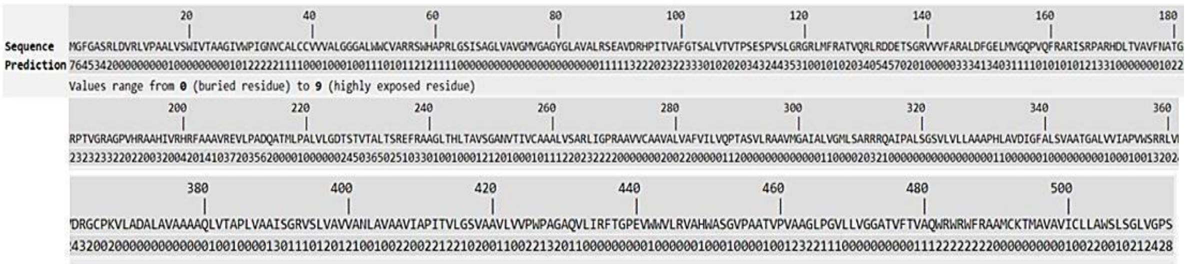


Figure D2: I-TASSER output showing predicated solvent accessibility

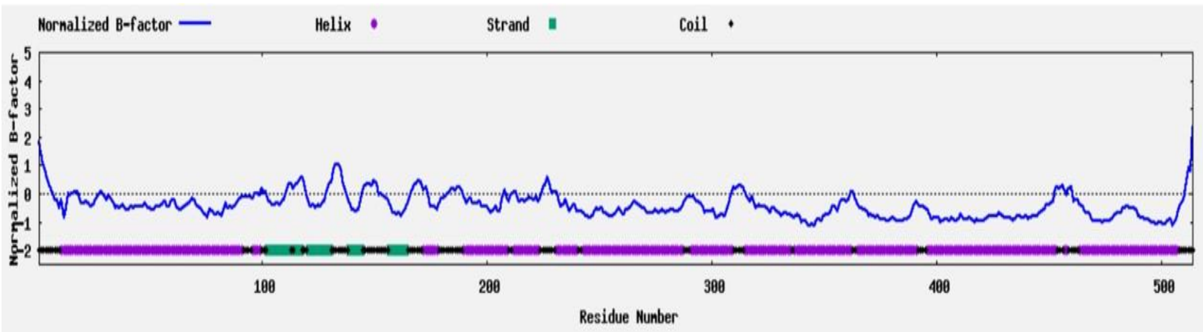


Figure D3: I-TASSER output showing the predicated normalized B-factor

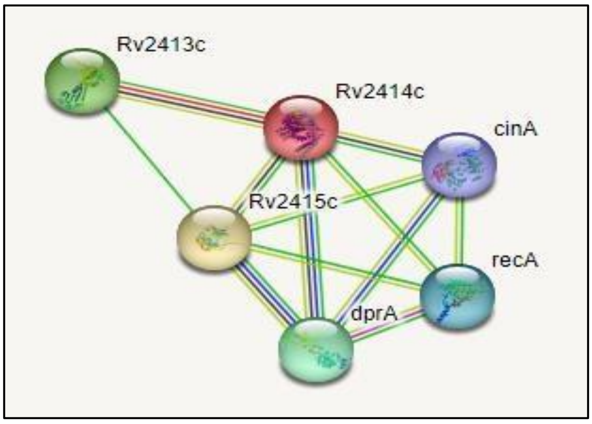


Figure D4: STRING analysis showing immediate gene and protein connections with *rv2414c*

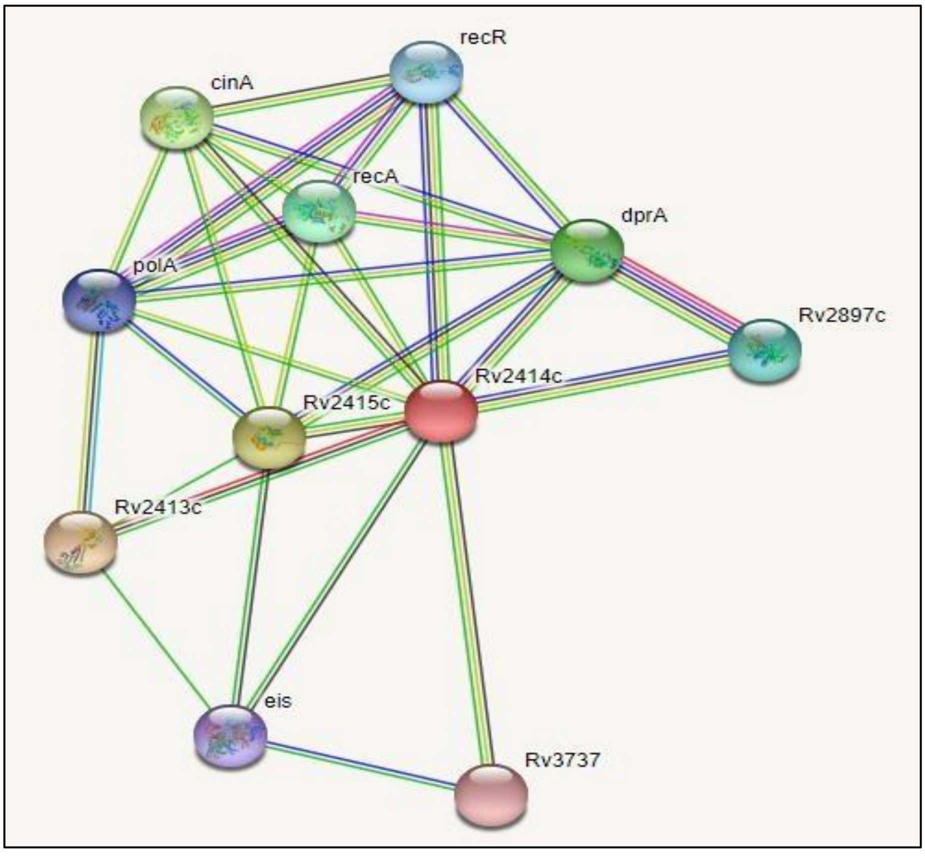
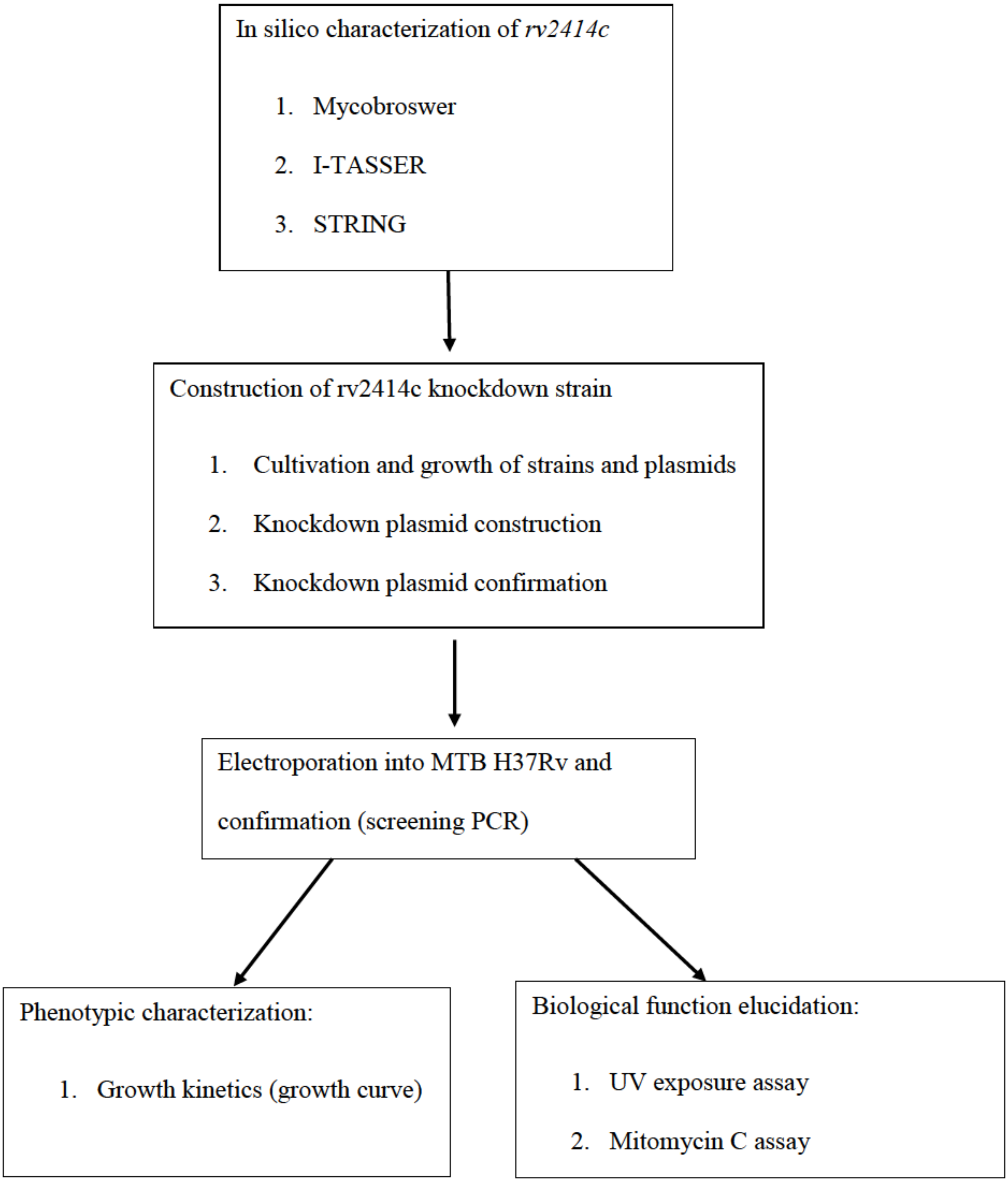


Figure D5: STRING analysis showing genes and protein connections (showing both close and further networks)

Table D1: I-TASSER output showing the top 10 identified structural analogs in PDB

Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov
1	7s9yA	0.858	2.31	0.079	0.897
2	6t15a	0.392	6.49	0.054	0.580
3	3hizA	0.389	6.94	0.048	0.597
4	4bxsV	0.386	7.64	0.037	0.642
5	7b54X	0.386	6.90	0.024	0.582
6	2y3aA	0.385	7.01	0.037	0.595
7	7qhmD	0.384	6.51	0.046	0.578
8	7rh5R	0.383	6.47	0.043	0.576
9	6zmqA	0.382	3.92	0.073	0.449
10	7d3eA	0.381	6.16	0.051	0.535

6.5 Appendix E: Flow chart



7 REFERENCES

ABRUSÁN, G. & MARSH, J. A. 2016. Alpha helices are more robust to mutations than beta strands. *PLoS computational biology,* 12**,** e1005242.

ACHKAR, J. M. & JENNY-AVITAL, E. R. 2011. Incipient and subclinical tuberculosis: defining early disease states in the context of host immune response. *Journal of Infectious Diseases,* 204**,** S1179-S1186.

ACHOUR, W. & CHEBBI, Y. 2022. Pathophysiology of Tuberculosis and Microbiological Diagnosis. *Imaging of Tuberculosis.* Springer.

ADESANYA, O. A., UCHE-ORJI, C. I., ADEDEJI, Y. A., JOSHUA, J. I., ADESOLA, A. A. & CHUKWUDIKE, C. J. 2021. Bacillus Calmette-Guerin (BCG): the adroit vaccine. *AIMS microbiology,* 7**,** 96.

AFKHAMI, S., VILLELA, A. D., D'AGOSTINO, M. R., JEYANATHAN, M., GILLGRASS, A. & XING, Z. 2020. Advancing immunotherapeutic vaccine strategies against pulmonary tuberculosis. *Frontiers in Immunology,* 11**,** 557809.

ALHMOUD, J. F., WOOLLEY, J. F., AL MOUSTAFA, A.-E. & MALKI, M. I. 2020. DNA damage/repair management in cancers. *Cancers,* 12**,** 1050.

ALHMOUD, J. F., WOOLLEY, J. F., AL MOUSTAFA, A.-E. & MALLEI, M. I. 2021. DNA Damage/Repair Management in Cancers. *Advances in Medical Biochemistry, Genomics, Physiology, and Pathology.* Jenny Stanford Publishing.

ALLINSON, S. L. 2010. DNA end-processing enzyme polynucleotide kinase as a potential target in the treatment of cancer. *Future Oncology,* 6**,** 1031-1042.

ALTIERI, F., GRILLO, C., MACERONI, M. & CHICHIARELLI, S. 2008. DNA damage and repair: from molecular mechanisms to health implications. *Antioxidants & redox signaling,* 10**,** 891-938.

ANDREU, N. & GIBERT, I. 2008. Cell population heterogeneity in *Mycobacterium tuberculosis* H37Rv. *Tuberculosis,* 88**,** 553-559.

ARALOV, A. V., GUBINA, N., CABRERO, C., TSVETKOV, V. B., TURAEV, A. V., FEDELES, B. I., CROY, R. G., ISAAKOVA, E. A., MELNIK, D. & DUKOVA, S. 2022. 7, 8-Dihydro-8-oxo-1, N 6-ethenoadenine: an exclusively Hoogsteen-paired thymine mimic in DNA that induces A→ T transversions in *Escherichia coli*. *Nucleic acids research,* 50**,** 3056-3069.

AWUH, J. A. & FLO, T. H. 2017. Molecular basis of mycobacterial survival in macrophages. *Cellular and Molecular Life Sciences,* 74**,** 1625-1648.

BARRY, C. E., BOSHOFF, H. I., DARTOIS, V., DICK, T., EHRT, S., FLYNN, J., SCHNAPPINGER, D., WILKINSON, R. J. & YOUNG, D. 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews Microbiology,* 7**,** 845-855.

BEAM, C. E. 2004. *A role for the radA/sms and radC genes in DNA repair and genetic recombination in the bacterium Escherichia coli*, Brandeis University.

BHARGAVI, G., HASSAN, S., BALAJI, S., TRIPATHY, S. P. & PALANIYANDI, K. 2020. Protein–protein interaction of Rv0148 with Htdy and its predicted role towards drug resistance in Mycobacterium tuberculosis. *BMC microbiology,* 20**,** 1-15.

BHATTACHARJEE, S. & NANDI, S. 2017. Synthetic lethality in DNA repair network: A novel avenue in targeted cancer therapy and combination therapeutics. *IUBMB life,* 69**,** 929-937.

BRIMACOMBE, C. A., DING, H., JOHNSON, J. A. & BEATTY, J. T. 2015. Homologues of genetic transformation DNA import genes are required for Rhodobacter capsulatus gene transfer agent recipient capability regulated by the response regulator CtrA. *Journal of bacteriology,* 197**,** 2653-2663.

BROOKS-POLLOCK, E., COHEN, T. & MURRAY, M. 2010. The impact of realistic age structure in simple models of tuberculosis transmission. *PLoS One,* 5**,** e8479.

BROOKS, S. C., ADHIKARY, S., RUBINSON, E. H. & EICHMAN, B. F. 2013. Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1834**,** 247-271.

BRZOSTEK, A., PŁOCIŃSKI, P., MINIAS, A., CISZEWSKA, A., GĄSIOR, F., PAWEŁCZYK, J., DZIADEK, B., SŁOMKA, M. & DZIADEK, J. 2021. Dissecting the RecA-(In) dependent response to mitomycin C in *Mycobacterium tuberculosis* using transcriptional profiling and proteomics analyses. *Cells,* 10**,** 1168.

BUSSI, C. & GUTIERREZ, M. G. 2019. *Mycobacterium tuberculosis* infection of host cells in space and time. *FEMS microbiology reviews,* 43**,** 341-361.

CALDECOTT, K. W. 2020. Mammalian DNA base excision repair: Dancing in the moonlight. *DNA repair,* 93**,** 102921.

CALVER, A. D., MURRAY, M., STRAUSS, O. J., STREICHER, E. M., HANEKOM, M., LIVERSAGE, T., MASIBI, M., VAN HELDEN, P. D., WARREN, R. M. & VICTOR, T. C. 2010. Emergence of increased resistance and extensively drug-resistant tuberculosis despite treatment adherence, South Africa. *Emerging infectious diseases,* 16**,** 264.

CAMBAU, E. & DRANCOURT, M. 2014. Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. *Clinical Microbiology and Infection,* 20**,** 196-201.

CANNAN, W. J. & PEDERSON, D. S. 2016. Mechanisms and consequences of double-strand DNA break formation in chromatin. *Journal of cellular physiology,* 231**,** 3-14.

CAREY, A. F., ROCK, J. M., KRIEGER, I. V., CHASE, M. R., FERNANDEZ-SUAREZ, M., GAGNEUX, S., SACCHETTINI, J. C., IOERGER, T. R. & FORTUNE, S. M. 2018. TnSeq of *Mycobacterium tuberculosis* clinical isolates reveals strain-specific antibiotic liabilities. *PLoS pathogens,* 14**,** e1006939.

CASTAÑEDA-GARCÍA, A., PRIETO, A., RODRIGUEZ-BELTRAN, J., ALONSO, N., CANTILLON, D., COSTAS, C., PÉREZ-LAGO, L., ZEGEYE, E., HERRANZ, M. & PLOCIŃSKI, P. 2017. A non-canonical mismatch repair pathway in prokaryotes. *Nature communications,* 8**,** 1-10.

CASTAÑEDA-HERNÁNDEZ, D. M. & RODRIGUEZ-MORALES, A. J. 2013. Epidemiological Burden of Tuberculosis in Developing Countries. *Curr. Top. Public Heal*.

CERRONE, M., BRACCHI, M., WASSERMAN, S., POZNIAK, A., MEINTJES, G., COHEN, K. & WILKINSON, R. J. 2020. Safety implications of combined antiretroviral and anti-tuberculosis drugs. *Expert opinion on drug safety,* 19**,** 23-41.

CHAKAROV, S., PETKOVA, R., RUSSEV, G. C. & ZHELEV, N. 2014. DNA damage and mutation. Types of DNA damage. *BioDiscovery,* 11**,** e8957.

CHANDRASEKARAN, M., RAMAN, C., KARTHIKEYAN, K. & PARAMASIVAN, M.

2019. Functional Annotation of Hypothetical Proteins Derived from Suppressive Subtraction Hybridization (SSH) Analysis Shows NPR1 (Non-Pathogenesis Related)Like Activity. *Agronomy,* 9**,** 57.

CHATTERJEE, N. & WALKER, G. C. 2017. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and molecular mutagenesis,* 58**,** 235-263.

CHEN, S., LEES-MILLER, J. P., HE, Y. & LEES-MILLER, S. P. 2021. Structural insights into the role of DNA-PK as a master regulator in NHEJ. *Genome Instability & Disease,* 2**,** 195-210.

CHEN, X., BRADLEY, N. P., LU, W., WAHL, K. L., ZHANG, M., YUAN, H., HOU, X.-F., EICHMAN, B. F. & TANG, G.-L. 2022. Base excision repair system targeting DNA adducts of trioxacarcin/LL-D49194 antibiotics for self-resistance. *Nucleic acids research,* 50**,** 2417-2430.

CHHABRA, N. & CHHABRA, S. 2012. *A case oriented approach towards biochemistry,* JP Medical Ltd.

CHILIZA, T. E., PILLAY, M., NAIDOO, K. & PILLAY, B. 2019. Immunoscreening of the M. tuberculosis F15/LAM4/KZN secretome library against TB patients′ sera identifies unique active-and latent-TB specific biomarkers. *Tuberculosis,* 115**,** 161-170.

CHILIZA, T. E., PILLAY, M. & PILLAY, B. 2017. Identification of unique essential proteins from a *Mycobacterium tuberculosis* F15/LAM4/KZN phage secretome library. *Pathogens and disease,* 75.

CHIN, D. P. & HANSON, C. L. 2017. Finding the missing tuberculosis patients. *The Journal of infectious diseases,* 216**,** S675-S678.

CHRISTMANN, M. & KAINA, B. 2012. O6-methylguanine-DNA methyltransferase (MGMT): impact on cancer risk in response to tobacco smoke. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 736**,** 64-74.

CHRISTMANN, M., TOMICIC, M. T., ROOS, W. P. & KAINA, B. 2003. Mechanisms of human DNA repair: an update. *Toxicology,* 193**,** 3-34.

COOPER, G. M., HAUSMAN, R. E. & HAUSMAN, R. E. 2007. *The cell: a molecular approach,* ASM press Washington, DC.

COSTES, S. V., CHIOLO, I., PLUTH, J. M., BARCELLOS-HOFF, M. H. & JAKOB, B. 2010. Spatiotemporal characterization of ionizing radiation induced DNA damage foci and their relation to chromatin organization. *Mutation Research/Reviews in Mutation Research,* 704**,** 78-87.

COUVÉ, S., ISHCHENKO, A. A., FEDOROVA, O. S., RAMANCULOV, E. M., LAVAL, J. & SAPARBAEV, M. 2013. Direct DNA lesion reversal and excision repair in Escherichia coli. *EcoSal Plus,* 5.

DAFFÉ, M. 2015. The cell envelope of tubercle bacilli. *Tuberculosis,* 95**,** S155-S158.

DARMON, E. & LEACH, D. R. 2014. Bacterial genome instability. *Microbiology and Molecular Biology Reviews,* 78**,** 1-39.

DAS, D., DAS, S., PANDEY, M. & BHATTACHARYAY, D. 2020. In silico analysis of phytochemicals from Mucuna pruriens (L.) DC against *Mycobacterium tuberculosis* causing tuberculosis. *European Journal of Medicinal Plants,* 31**,** 19-24.

DE ZIO, D., CIANFANELLI, V. & CECCONI, F. 2013. New insights into the link between DNA damage and apoptosis. *Antioxidants & redox signaling,* 19**,** 559-571.

DELMASTRO, D. 2015. Effect of the loss of Ribonuclease H activity on genome instability: from DNA damage tolerance to retrotransposition.

DELOGU, G., SALI, M. & FADDA, G. 2013. The biology of *Mycobacterium tuberculosis* infection. *Mediterranean journal of hematology and infectious diseases,* 5.

DIETRICH, M. 2013. *Investigation of changes in immunostimulatory properties of clinical Mycobacterium avium strains over time of infection.* Norges teknisknaturvitenskapelige universitet, Det medisinske fakultet ….

DOS VULTOS, T., MESTRE, O., TONJUM, T. & GICQUEL, B. 2009. DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS microbiology reviews,* 33**,** 471-487.

DUNN, J. J., STARKE, J. R. & REVELL, P. A. 2016. Laboratory diagnosis of *Mycobacterium tuberculosis* infection and disease in children. *Journal of clinical microbiology,* 54**,** 1434-1441.

DUPUY, P., HOWLADER, M. & GLICKMAN, M. S. 2020. A multilayered repair system protects the mycobacterial chromosome from endogenous and antibiotic-induced oxidative damage. *Proceedings of the National Academy of Sciences,* 117**,** 1951719527.

DWIVEDI, M. & GIRI, P. 2021. Challenges in drug discovery against tuberculosis. *Molecular Epidemiology Study of Mycobacterium tuberculosis Complex. Intech Open***,** 9-30.

EHRT, S. & SCHNAPPINGER, D. 2009. Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cellular microbiology,* 11**,** 1170-1178.

ELCHENNAWI, I. & OLLAGNIER DE CHOUDENS, S. 2022. Iron–Sulfur Clusters toward Stresses: Implication for Understanding and Fighting Tuberculosis. *Inorganics,* 10**,** 174.

ETNA, M. P., GIACOMINI, E., SEVERA, M. & COCCIA, E. M. Pro-and anti-inflammatory cytokines in tuberculosis: a two-edged sword in TB pathogenesis. Seminars in immunology, 2014. Elsevier, 543-551.

FATIMA, S., KUMARI, A., DAS, G. & DWIVEDI, V. P. 2020. Tuberculosis vaccine: A journey from BCG to present. *Life sciences,* 252**,** 117594.

FLYNN, J. L. & CHAN, J. 2001. Tuberculosis: latency and reactivation. *Infection and immunity,* 69**,** 4195-4201.

FOJTA, M. 2002. Electrochemical sensors for DNA interactions and damage. *Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis,* 14**,** 1449-1463.

FORBES, B. A., HALL, G. S., MILLER, M. B., NOVAK, S. M., ROWLINSON, M.-C., SALFINGER, M., SOMOSKÖVI, A., WARSHAUER, D. M. & WILSON, M. L. 2018. Practical guidance for clinical microbiology laboratories: mycobacteria. *Clinical microbiology reviews,* 31**,** e00038-17.

FU, D., CALVO, J. A. & SAMSON, L. D. 2012. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nature Reviews Cancer,* 12**,** 104-120.

GAGNEUX, S. 2018. Ecology and evolution of Mycobacterium tuberculosis. *Nature Reviews Microbiology,* 16**,** 202.

GAO, L., ZHANG, H., XIN, H., LIU, J., PAN, S., LI, X., GUAN, L., SHEN, F., LIU, Z. & WANG, D. 2018. Short-course regimens of rifapentine plus isoniazid to treat latent tuberculosis infection in older Chinese patients: a randomised controlled study. *European Respiratory Journal,* 52.

GENGENBACHER, M. & KAUFMANN, S. H. 2012. Mycobacterium tuberculosis: success through dormancy. *FEMS microbiology reviews,* 36**,** 514-532.

GORDON, S. V. & PARISH, T. 2018. Microbe Profile: Mycobacterium tuberculosis: Humanity's deadly microbial foe. *Microbiology,* 164**,** 437-439.

GREEN, M. H. 2017. The globalisations of disease. *Human dispersal and species movement: From prehistory to the present***,** 494-520.

HAILE, B., TAFESS, K., ZEWUDE, A., YENEW, B., SIU, G. & AMENI, G. 2020. Spoligotyping and drug sensitivity of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients in the Arsi Zone of southeastern Ethiopia. *New Microbes and New Infections,* 33**,** 100620.

HALL, E. J. & KEREIAKES, J. G. 2001. Effects of ionizing radiation on cells. *Cell Physiology Source Book.* Elsevier.

HANAWALT, P. C., COOPER, P. K., GANESAN, A. K. & SMITH, C. A. 1979. DNA repair in bacteria and mammalian cells. *Annual review of biochemistry,* 48**,** 783-836. HARDISON, R. C. & CHU, T. 2021. Working with molecular genetics. Chpt.

HENNINGER, E. E. & PURSELL, Z. F. 2014. DNA polymerase ε and its roles in genome stability. *IUBMB life,* 66**,** 339-351.

HIPPNER, P., SUMNER, T., HOUBEN, R. M., CARDENAS, V., VASSALL, A., BOZZANI, F., MUDZENGI, D., MVUSI, L., CHURCHYARD, G. & WHITE, R. G. 2019. Application of provincial data in mathematical modelling to inform sub-national tuberculosis program decision-making in South Africa. *PloS one,* 14**,** e0209320.

HMAMA, Z., PEÑA-DÍAZ, S., JOSEPH, S. & AV-GAY, Y. 2015. Immunoevasion and immunosuppression of the macrophage by M ycobacterium tuberculosis. *Immunological reviews,* 264**,** 220-232.

HUANG, R. & ZHOU, P.-K. 2021. DNA damage repair: Historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduction and Targeted Therapy,* 6**,** 1-35.

IJAQ, J., MALIK, G., KUMAR, A., DAS, P. S., MEENA, N., BETHI, N., SUNDARARAJAN, V. S. & SURAVAJHALA, P. 2019. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC bioinformatics,* 20**,** 14.

IRSHAD, M. & MUNIR, H. 2017. Structural and Functional Characterization of a Hypothetical protein of Streptococcus Pyrogenes: An In-Silico Approach. *Journal of Biochemistry, Biotechnology and Biomaterials,* 1**,** 54-63.

JACKSON, S. P. & BARTEK, J. 2009. The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071-1078.

JENSEN, P. A., LAMBERT, L. A., IADEMARCO, M. F. & RIDZON, R. 2005. Guidelines for preventing the transmission of *Mycobacterium tuberculosis* in health-care settings, 2005.

JOSEPH, A. M., NAHAR, K., DAW, S., HASAN, M. M., LO, R., LE, T. B., RAHMAN, K. M. & BADRINARAYANAN, A. 2022. Mechanistic insight into the repair of C8-linked pyrrolobenzodiazepine monomer-mediated DNA damage. *RSC Medicinal Chemistry,* 13**,** 1621-1633.

KALGHATGI, S., SPINA, C. S., COSTELLO, J. C., LIESA, M., MORONES-RAMIREZ, J. R., SLOMOVIC, S., MOLINA, A., SHIRIHAI, O. S. & COLLINS, J. J. 2013. Bactericidal antibiotics induce mitochondrial dysfunction and oxidative damage in mammalian cells. *Science translational medicine,* 5**,** 192ra85-192ra85.

KHAWBUNG, J. L., NATH, D. & CHAKRABORTY, S. 2021. Drug resistant tuberculosis: a review. *Comparative Immunology, Microbiology and Infectious Diseases,* 74**,** 101574.

KIM, N. & JINKS-ROBERTSON, S. 2010. Abasic sites in the transcribed strand of yeast DNA are removed by transcription-coupled nucleotide excision repair. *Molecular and cellular biology,* 30**,** 3206-3215.

KIMURA, S. & SAKAGUCHI, K. 2006. DNA repair in plants. *Chemical reviews,* 106**,** 753766.

KORB, V. C., CHUTURGOON, A. A. & MOODLEY, D. 2016. Mycobacterium tuberculosis:

manipulator of protective immunity. *International journal of molecular sciences,* 17**,** 131.
LEE, S. 2014. Molecular Mechanism of Human Mismatch Repair Initiation.
LEYNS, L. & GONZALEZ, L. 2012. Genomic integrity of mouse embryonic stem cells. *Embryogenesis. Intech***,** 333-358.
LI, G.-M. 2008. Mechanisms and functions of DNA mismatch repair. *Cell research,* 18**,** 8598.
LI, X. & HEYER, W.-D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell research,* 18**,** 99-113.
LUNDHOLM, I. V., RODILLA, H., WAHLGREN, W. Y., DUELLI, A., BOURENKOV, G., VUKUSIC, J., FRIEDMAN, R., STAKE, J., SCHNEIDER, T. & KATONA, G. 2015. Terahertz radiation induces non-thermal structural changes associated with Fröhlich condensation in a protein crystal. *Structural Dynamics,* 2**,** 054702.
LUSETTI, S. L. & COX, M. M. 2002. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annual review of biochemistry,* 71**,** 71-100.
LYASHCHENKO, K. P., VORDERMEIER, H. M. & WATERS, W. R. 2020. Memory B cells and tuberculosis. *Veterinary Immunology and Immunopathology***,** 110016.
MA, F. H., LI, C., LIU, Y. & SHI, L. 2020. Mimicking molecular chaperones to regulate protein folding. *Advanced Materials,* 32**,** 1805945.
MACNEIL, A., GLAZIOU, P., SISMANIDIS, C., MALONEY, S. & FLOYD, K. 2019. Global epidemiology of tuberculosis and progress toward achieving global targets—2017. *Morbidity and Mortality Weekly Report,* 68**,** 263.
MAGERAND, R., REY, P., BLANCHARD, L. & DE GROOT, A. 2021. Redox signaling through zinc activates the radiation response in Deinococcus bacteria. *Scientific Reports,* 11**,** 1-12.
MAKINO, S.-I., VAN PUTTEN, J. & MEYER, T. F. 1991. Phase variation of the opacity outer membrane protein controls invasion by Neisseria gonorrhoeae into human epithelial cells. *The EMBO journal,* 10**,** 1307-1315.
MARIMANI, M., AHMAD, A. & DUSE, A. 2018. The role of epigenetics, bacterial and host factors in progression of *Mycobacterium tuberculosis* infection. *Tuberculosis,* 113**,** 200-214.
MARINUS, M. 2012. DNA mismatch repair. *EcoSal Plus,* 5.
MASLOWSKA, K. H., MAKIELA-DZBENSKA, K. & FIJALKOWSKA, I. J. 2019. The SOS system: a complex and tightly regulated response to DNA damage. *Environmental and molecular mutagenesis,* 60**,** 368-384.
MATATIELE, M., STIEGLER, N. & BOUCHARD, J.-P. Factors associated to subjective well-being and TB infection in South Africa. Annales Médico-psychologiques, revue psychiatrique, 2021. Elsevier, 553-562.
MAZANDU, G. K. & MULDER, N. J. 2012. Function prediction and analysis of *Mycobacterium tuberculosis* hypothetical proteins. *International journal of molecular sciences,* 13**,** 7283-7302.
MCGLYNN, P. & LLOYD, R. G. 2002. Recombinational repair and restart of damaged replication forks. *Nature reviews Molecular cell biology,* 3**,** 859-870.
MERCHUT-MAYA, J. M., BARTEK, J. & MAYA-MENDOZA, A. 2019. Regulation of replication fork speed: Mechanisms and impact on genomic stability. *DNA repair,* 81**,** 102654.
MIELECKI, D. & GRZESIUK, E. 2014. Ada response–a strategy for repair of alkylated DNA in bacteria. *FEMS Microbiology Letters,* 355**,** 1-11.

MIZRAHI, V. & ANDERSEN, S. J. 1998. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Molecular microbiology,* 29**,** 1331-1339.

MORITA, R., NAKANE, S., SHIMADA, A., INOUE, M., IINO, H., WAKAMATSU, T., FUKUI, K., NAKAGAWA, N., MASUI, R. & KURAMITSU, S. 2010. Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems. *Journal of nucleic acids,* 2010.

MÜLLER, A. U., IMKAMP, F. & WEBER-BAN, E. 2018. The mycobacterial LexA/RecAindependent DNA damage response is controlled by PafBC and the Pup-proteasome system. *Cell reports,* 23**,** 3551-3564.

NAVA, J. & KREINOVICH, V. 2012. Towards Symmetry-Based Explanation of (Approximate) Shapes of Alpha-Helices and Beta-Sheets (and Beta-Barrels) in Protein Structure. *Symmetry,* 4**,** 15-25.

NAY, S. L. & O'CONNOR, T. R. 2013. Direct repair in mammalian cells. *New Research Directions in DNA Repair***,** 123-162.

NEWBERRY, R. W. & RAINES, R. T. 2019. Secondary forces in protein folding. *ACS chemical biology,* 14**,** 1677-1686.

NGAMVITHAYAPONG-YANAI, J., LUANGJINA, S., THAWTHONG, S., BUPACHAT, S. & IMSANGAUN, W. 2019. Stigma against tuberculosis may hinder non-household contact investigation: a qualitative study in Thailand. *Public Health Action,* 9**,** 15-23.

NIMONKAR, A. V. & KOWALCZYKOWSKI, S. C. 2009. Second-end DNA capture in double-strand break repair: how to catch a DNA by its tail. Taylor & Francis.

NORJMAA, B., TULGAA, K. & SAITOH, T. 2016. Base excision repair pathway and polymorphisms of XRCC1 gene. *J Mol Pathol Epidemiol,* 1**,** 1-4.

OBENG, E. 2020. Apoptosis (programmed cell death) and its signals-A review. *Brazilian Journal of Biology,* 81**,** 1133-1143.

ORGANIZATION, W. H. 2018. Global tuberculosis report 2018. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO. WHO/CDS/TB/2018.20. Available from: http://apps. who. int/iris/bitstream ….

ORGANIZATION, W. H. 2019. Global tuberculosis report 2018. 2018. *Geneva: World Health Organization*.

ORGANIZATION, W. H. 2021. WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–2025: background document.

PEĆINA-ŠLAUS, N., KAFKA, A., SALAMON, I. & BUKOVAC, A. 2020. Mismatch repair pathway, genome stability and cancer. *Frontiers in molecular biosciences***,** 122.

PENG, Y., WANG, H., SANTANA-SANTOS, L., KISKER, C. & VAN HOUTEN, B. 2011. Nucleotide excision repair from bacteria to humans: Structure–function studies. *Chemical Carcinogenesis***,** 267-296.

PETTIJOHN, D. & HANAWALT, P. 1964. Evidence for repair-replication of ultraviolet damaged DNA in bacteria. *Journal of molecular biology,* 9**,** 395-410.

PILLAI, L. & CHOUHAN, U. 2018. COMPUTATIONAL ANALYSIS OF MOLECULAR SEQUENCES OF *MYCOBACTERIUM TUBERCULOSIS* AND LEPRAE ON THE BASIS OF PATHWAY ANALYSIS FOR DRUG TARGET IDENTIFICATION. *European Journal of Biomedical,* 5**,** 578-584.

POETA, P., SILVA, V., GUEDES, A., EDUARDO PEREIRA, J., CLÁUDIA COELHO, A. & IGREJAS, G. 2018. Tuberculosis in the 21th century: Current status of diagnostic methods. *Experimental lung research,* 44**,** 352-360.

POPESCU, C. R., CAVANAGH, M. M., TEMBO, B., CHIUME, M., LUFESI, N., GOLDFARB, D. M., KISSOON, N. & LAVOIE, P. M. 2020. Neonatal sepsis in lowincome countries: epidemiology, diagnosis and prevention. *Expert review of antiinfective therapy,* 18**,** 443-452.

PRYOR, J. M., WATERS, C. A., AZA, A., ASAGOSHI, K., STROM, C., MIECZKOWSKI, P. A., BLANCO, L. & RAMSDEN, D. A. 2015. Essential role for polymerase specialization in cellular nonhomologous end joining. *Proceedings of the National Academy of Sciences,* 112**,** E4537-E4545.

RECUERO-CHECA, M. A., DORÉ, A. S., ARIAS-PALOMO, E., RIVERA-CALZADA, A., SCHERES, S. H., MAMAN, J. D., PEARL, L. H. & LLORCA, O. 2009. Electron microscopy of Xrcc4 and the DNA ligase IV–Xrcc4 DNA repair complex. *DNA repair,* 8**,** 1380-1389.

RIJAL, R., CADENA, L. A., SMITH, M. R., CARR, J. F. & GOMER, R. H. 2020. Polyphosphate is an extracellular signal that can facilitate bacterial survival in eukaryotic cells. *Proceedings of the National Academy of Sciences,* 117**,** 31923-31934.

ROBERTSON, A., KLUNGLAND, A., ROGNES, T. & LEIROS, I. 2009. DNA repair in mammalian cells. *Cellular and molecular life sciences,* 66**,** 981-993.

ROOS, W. P. & KAINA, B. 2006. DNA damage-induced cell death by apoptosis. *Trends in molecular medicine,* 12**,** 440-450.

RUDOLPH, T. K. & FREEMAN, B. A. 2009. Transduction of redox signaling by electrophileprotein reactions. *Science signaling,* 2**,** re7-re7.

RUFF, P., DONNIANNI, R. A., GLANCY, E., OH, J. & SYMINGTON, L. S. 2016. RPA stabilization of single-stranded DNA is critical for break-induced replication. *Cell reports,* 17**,** 3359-3368.

SÁNCHEZ, S., DE MIGUEL, T., VILLA, T., GORRINGE, A. & FEAVERS, I. 2019. Horizontal Gene Transfer Among Neisseria Species and Humans. *Horizontal Gene Transfer: Breaking Borders Between Living Kingdoms***,** 361-376.

SAPONARO, M., CALLAHAN, D., ZHENG, X., KREJCI, L., HABER, J. E., KLEIN, H. L. & LIBERI, G. 2010. Cdk1 targets Srs2 to complete synthesis-dependent strand annealing and to promote recombinational repair. *PLoS genetics,* 6**,** e1000858.

SCHRAGER, L. K., VEKEMENS, J., DRAGER, N., LEWINSOHN, D. M. & OLESEN, O. F. 2020. The status of tuberculosis vaccine development. *The Lancet Infectious Diseases,* 20**,** e28-e37.

SHAPIRO, R. 1981. Damage to DNA caused by hydrolysis. *Chromosome damage and repair.* Springer.

SHENG, D.-H., WANG, Y.-X., QIU, M., ZHAO, J.-Y., YUE, X.-J. & LI, Y.-Z. 2019. Functional divergence and potential mechanisms of the duplicate recA genes in Myxococcus xanthus. *bioRxiv***,** 766055.

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

SMITH, I. 2003. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews,* 16**,** 463-496.

SOWA, D. J., WARNER, M. M., TETENYCH, A., KOECHLIN, L., BALARI, P., RASCON PEREZ, J. P., CABA, C. & ANDRES, S. N. 2022. The *Mycobacterium tuberculosis* Ku C-terminus is a multi-purpose arm for binding DNA and LigD and stimulating ligation. *Nucleic Acids Research*.

SOWIN, L. & COX, M. M. 2017. Exploring the Role of ATP Hydrolysis in the RecA Protein via Stalled DNA Intermediates. *The FASEB Journal,* 31**,** 906.20-906.20.

STAPLETON, M., HAQ, I., HUNT, D. M., ARNVIG, K. B., ARTYMIUK, P. J., BUXTON, R. S. & GREEN, J. 2010. *Mycobacterium tuberculosis* cAMP receptor protein (Rv3676) differs from the *Escherichia coli* paradigm in its cAMP binding and DNA binding properties and transcription activation properties. *Journal of Biological Chemistry,* 285**,** 7016-7027.

STEIN, C. M. 2011. Genetic epidemiology of tuberculosis susceptibility: impact of study design. *PLoS Pathog,* 7**,** e1001189.

STEPHENS, R. S., KALMAN, S., LAMMEL, C., FAN, J., MARATHE, R., ARAVIND, L., MITCHELL, W., OLINGER, L., TATUSOV, R. L. & ZHAO, Q. 1998. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science,* 282**,** 754-759.

SUDBURY, E. L., CLIFFORD, V., MESSINA, N. L., SONG, R. & CURTIS, N. 2020. Mycobacterium tuberculosis-specific cytokine biomarkers to differentiate active TB and LTBI: A systematic review. *Journal of Infection,* 81**,** 873-881.

SUN, Z., LIU, Q., QU, G., FENG, Y. & REETZ, M. T. 2019. Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chemical reviews,* 119**,** 1626-1665.

SUNDE, M., PHAM, C. L. & KWAN, A. H. 2017. Molecular characteristics and biological functions of surface-active and surfactant proteins. *Annual review of biochemistry,* 86**,** 585-608.

SVEJSTRUP, J. Q. 2002. Mechanisms of transcription-coupled DNA repair. *Nature Reviews Molecular Cell Biology,* 3**,** 21-29.

SWAMINATHAN, S. & REKHA, B. 2010. Pediatric tuberculosis: global overview and challenges. *Clinical infectious diseases,* 50**,** S184-S194.

TATUM, D. & LI, S. 2011. Nucleotide excision repair in S. cerevisiae. *DNA Repair-On the Pathways to Fixing DNA Damage and Errors; Storici, F., Ed***,** 97-122.

TIBERI, S., UTJESANOVIC, N., GALVIN, J., CENTIS, R., D'AMBROSIO, L., VAN DEN BOOM, M., ZUMLA, A. & MIGLIORI, G. B. 2022. Drug resistant TB–latest developments in epidemiology, diagnostics and management. *International Journal of Infectious Diseases*.

TOMASOVA, K., CUMOVA, A., SEBOROVA, K., HORAK, J., KOUCKA, K., VODICKOVA, L., VACLAVIKOVA, R. & VODICKA, P. 2020. DNA repair and ovarian carcinogenesis: impact on risk, prognosis and therapy outcome. *Cancers,* 12**,** 1713.

TRENNER, A. & SARTORI, A. A. 2019. Harnessing DNA double-strand break repair for cancer treatment. *Frontiers in oncology,* 9**,** 1388.

TRETYAKOVA, N. Y., GROEHLER IV, A. & JI, S. 2015. DNA–protein cross-links: formation, structural identities, and biological outcomes. *Accounts of chemical research,* 48**,** 1631-1644.

TUDEK, B., WINCZURA, A., JANIK, J., SIOMEK, A., FOKSINSKI, M. & OLIŃSKI, R. 2010. Involvement of oxidatively damaged DNA and repair in cancer development and aging. *American journal of translational research,* 2**,** 254.

TURNER, J. S. 2004. *Molecular Recognition of Damaged DNA Using Synthetic Affinity Reagents.* The Open University.

VALIKHANI, M., RAHIMIAN, E., AHMADI, S. E., CHEGENI, R. & SAFA, M. 2021. Involvement of classic and alternative non-homologous end joining pathways in hematologic malignancies: targeting strategies for treatment. *Experimental Hematology & Oncology,* 10**,** 1-26.

VAN DER WALT, M. & MOYO, S. 2021. The First National TB Prevalence Survey, South Africa 2018: Short report. 2021.

VANDERSLUIS, B., COSTANZO, M., BILLMANN, M., WARD, H. N., MYERS, C. L., ANDREWS, B. J. & BOONE, C. 2018. Integrating genetic and protein–protein interaction networks maps a functional wiring diagram of a cell. *Current opinion in microbiology,* 45**,** 170-179.

VILLELA, A. D., RODRIGUES-JUNIOR, V. S., PINTO, A. F. M., FALCÃO, V. C. D. A., SÁNCHEZ-QUITIAN, Z. A., EICHLER, P., BIZARRO, C. V., BASSO, L. A. & SANTOS, D. S. 2017. Construction of *Mycobacterium tuberculosis* cdd knockout and evaluation of invasion and growth in macrophages. *Memórias do Instituto Oswaldo Cruz,* 112**,** 785-789.

VISCA, D., ONG, C., TIBERI, S., CENTIS, R., D'AMBROSIO, L., CHEN, B., MUELLER, J., MUELLER, P., DUARTE, R. & DALCOLMO, M. 2021. Tuberculosis and COVID19 interaction: a review of biological, clinical and public health effects. *Pulmonology,* 27**,** 151-165.

WARNER, D. F. & MIZRAHI, V. 2006. Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. *Clinical microbiology reviews,* 19**,** 558-570.

WETERINGS, E. & CHEN, D. J. 2008. The endless tale of non-homologous end-joining. *Cell research,* 18**,** 114-124.

WILSON III, D. M. & BARSKY, D. 2001. The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutation Research/DNA Repair,* 485**,** 283-307.

YANG, J. & ZHANG, Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research,* 43**,** W174-W181.

YANG, Z., ZENG, X. & TSUI, S. K.-W. 2019. Investigating function roles of hypothetical proteins encoded by the *Mycobacterium tuberculosis* H37Rv genome. *BMC genomics,* 20**,** 1-10.

YASSIN, M., SAMSON, K., WANDWALO, E., GRZEMSKA, M., GEGIA, M., NWANERI, N., HANSEN, P. & KUNII, O. 2020. Performance-based technical support for drugresistant TB responses: lessons from the Green Light Committee. *The International Journal of Tuberculosis and Lung Disease,* 24**,** 22-27.

YOON, J.-H., CHOUDHURY, J. R., PARK, J., PRAKASH, S. & PRAKASH, L. 2017. Translesion synthesis DNA polymerases promote error-free replication through the minor-groove DNA adduct 3-deaza-3-methyladenine. *Journal of Biological Chemistry,* 292**,** 18682-18688.

YU, W., ZHANG, L., WEI, Q. & SHAO, A. 2020. O6-methylguanine-DNA methyltransferase (MGMT): challenges and new opportunities in glioma chemotherapy. *Frontiers in oncology,* 9**,** 1547.

ZEGEYE, E. D. 2013. Biochemical Characterization of the DNA Helicases RecG, XPB and DinG in Mycobacterium tuberculosis.

ŽGUR-BERTOK, D. 2013. DNA damage repair and bacterial pathogens. *PLoS pathogens,* 9**,** e1003711.

ZHANG, W., DILLINGHAM, M. S., THOMAS, C. D., ALLEN, S., ROBERTS, C. J. & SOULTANAS, P. 2007. Directional loading and stimulation of PcrA helicase by the replication initiator protein RepD. *Journal of molecular biology,* 371**,** 336-348.

ZHAO, B., WATANABE, G., MORTEN, M. J., REID, D. A., ROTHENBERG, E. & LIEBER, M. R. 2019. The essential elements for the noncovalent association of two DNA ends during NHEJ synapsis. *Nature communications,* 10**,** 3588.

ZHU, L., ZHONG, J., JIA, X., LIU, G., KANG, Y., DONG, M., ZHANG, X., LI, Q., YUE, L. & LI, C. 2016. Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. *Nucleic Acids Research,* 44**,** 730-743.

Maleke DP MSc Dissertation

ORIGINALITY REPORT

12%	12%	6%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	4%
2	mts.intechopen.com Internet Source	1%
3	dspace.uni.lodz.pl Internet Source	1%
4	open.uct.ac.za Internet Source	1%
5	www.ncbi.nlm.nih.gov Internet Source	1%
6	pdfs.semanticscholar.org Internet Source	1%
7	www.nature.com Internet Source	1%
8	2017.igem.org Internet Source	1%
9	brcdownloads.vbi.vt.edu Internet Source	1%
10	eprints.nottingham.ac.uk Internet Source	1%

CHILIZA, T. E., PILLAY, M., NAIDOO, K. & PILLAY, B. 2019. Immunoscreening of the M. tuberculosis F15/LAM4/KZN secretome library against TB patients′ sera identifies unique active-and latent-TB specific biomarkers. *Tuberculosis,* 115**,** 161-170.
HENGSAWAS, S. 2016. *Importance of stability of pharmaceutical formulations.*
KAZLAUSKAS, D., KRUPOVIC, M. & VENCLOVAS, Č. 2016. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic acids research,* 44**,** 4551-4564.
ROY, S. 2017. Impact of UV radiation on genome stability and human health. *Ultraviolet light in human health, diseases and environment***,** 207-219.
SHIN, D.-M., JEON, B.-Y., LEE, H.-M., JIN, H. S., YUK, J.-M., SONG, C.-H., LEE, S.-H., LEE, Z.-W., CHO, S.-N. & KIM, J.-M. 2010. Mycobacterium tuberculosis eis regulates autophagy, inflammation, and cell death through redox-dependent signaling. *PLoS pathogens,* 6**,** e1001230.