EXPRESSION AND FUNCTION OF THE MUTATOR DNA POLYMERASE-ENCODING *umuC*-LIKE GENES IN MYCOBACTERIA

Robyn Brackin

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

Johannesburg, 2005

DECLARATION

I, Robyn Brackin declare that this dissertation is my own work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

……

…….day of...……..., 2005

ABSTRACT

Mycobacterium tuberculosis is an important human pathogen, claiming more lives per annum than any other single infectious organism. The host environment of *M. tuberculosis* contains DNA-damaging agents that pose a constant threat to the *M. tuberculosis* genome, and as a result, the ability to repair damaged DNA is likely to play an important role in bacterial survival. Y-family polymerases perform translesional synthesis and replicate DNA in an error-prone manner. By characterising the Y-family polymerases in mycobacteria, a better understanding the organism's adaptive mutagenesis may be established.

Through gene expression studies, it was found that UV irradiation of *Mycobacterium smegmatis* resulted in the up-regulation of *dinP3*, which was determined to be a Y-family polymerase by sequence analysis. *DinP3* expression was found to be under control of the SOS response and is the first example of a Y-family polymerase in mycobacteria forming part of the SOS regulon. However, loss of DinP3 did not change the ability of *M. smegmatis* to tolerate UV irradiation. Mutagenesis studies revealed a complex interaction between the different Y-family polymerases in *M. smegmatis*. It was shown that spontaneous mutagenesis was increased in the absence of DinP3, whereas UV-targeted mutagenesis was increased in the absence of DinP, another Y-family polymerase.

In conclusion, these results reflect the differences in control and in the mutational profiles of the Y-family polymerases in *M. smegmatis*. Moreover, these polymerases exhibit distinctive features from other bacterial Y-family polymerases, highlighting the different way in which bacteria have adapted to deal with lesions in their genetic material.

ACKNOWLEDGEMENTS

The discovery of new knowledge in science can never be undertaken alone, and as such, I am thankful to many people who have helped me along my route.

1. Foremost, I am grateful to the excellent guidance my supervisor, Prof. Valerie Mizrahi, has extended towards me during this study. Thank you for your enthusiasm and great understanding of the world of mycobacteria and DNA polymerases, allowing me in turn to pursue and better understand this world.

2. I would like to thank my co-supervisor, Dr. Stephanie Dawes, for the continual encouragement and supervision during this study. Thank you for all the practical support and assistance provided, and for always being available for discussions.

3. I would like to thank Dr. Edith Machowski for generating the Δ*dinP3* strain and the *hisD*^{5T} reporter strains, and Mr. Digby Warner for generating the *ΔdinP* strain, which were all vital for the studies presented in this dissertation. Thank for you for all the assistance provided in the laboratory and all the informative discussions.

4. I extended many thanks to the rest of my colleagues, Samantha Barichievy, Bavesh Kana, Bhavna Gordhan and Limenako Matsoso for all the insightful discussions and sharing this journey with me. I appreciate all the assistance and guidance provided by all members of the laboratory.

5. I gratefully acknowledge financial support in the form of scholarships/ bursaries from: the University, the National Research Foundation and the Medical Research Council.

6. To my parents, Debbie and Neville McGregor, thank you for always being the grounding force in my life and for all your love and support. I wish to especially thank my boyfriend, Marc Weinberg, a fellow researcher in the field of molecular biology, for continually exciting the passion in me to find the answers to the major problems of this world. I am in gratitude for your all your love and support and for undertaking the task of proofreading this dissertation.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AMV	Avian Myeloblastoma Virus
ATP	Adenosine triphosphate
BCG	Bacille Calmette-Guérin
BER	Base excision repair
BSA	Bovine serum albumin
cDNA	Complementary DNA
CFU	Colony forming unit
din	Damage-inducible
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
GASP	Growth advantage in stationary phase
GMS	Gel mobility shift
HIV	Human Immunodeficiency Virus
LA	Luria-Bertani agar
MDR	Multi-drug resistant
MMR	Mismatch repair
MSS	Ma-Sandri-Sarkar
NER	Nucleotide excision repair
OD	Optical density
PCR	Polymerase chain reaction
PNK	Polynucleotide kinase
Pol	Polymerase
RNA	Ribonucleic acid
RT	Reverse transcription
SOS	Save our souls
TB	Tuberculosis
TLS	Translesion synthesis
UV	Ultraviolet

1.0 INTRODUCTION

Mycobacterium tuberculosis claims more lives per annum than any other bacterial pathogen (Hingley-Wilson *et al.*, 2003). Although one-third of the world's population is estimated to be infected with *M. tuberculosis*, not all infected individuals show clinical signs of the disease (Stewart *et al.*, 2003). This is a result of latent tuberculosis (TB), an asymptomatic infection in which *M. tuberculosis* persists within the human host for years. In 1993, the World Health Organisation declared TB a global emergency. The emergence of drug-resistant strains of *M. tuberculosis* and its synergy with the human immunodeficiency virus (HIV) have jeopardised attempts to control the disease over the past 11 years (Campos *et al.*, 2003).

M. tuberculosis has been shown to develop drug resistance through the introduction of multiple chromosomal mutations (Ramaswamy *et al.*, 1998). Adaptive mutagenesis has long been regarded as a process in which random mutations arise in an organism, thus producing genetically different variants that can compete for available resources. It has now been demonstrated that the rate of mutation is largely under the control of genetic factors (Metzgar *et al.*, 2000). Understanding the mechanisms through which these mutations are introduced into the bacterial chromosome should provide the knowledge needed for new therapies to treat the TB epidemic.

Through exposure to the host immune response, *M. tuberculosis* encounters a range of DNA-damaging agents in the form of reactive oxygen and reactive

nitrogen intermediates (Stewart *et al.*, 2003). These agents pose a constant threat to the *M. tuberculosis* genome, and as a result, the ability to repair the damaged DNA plays an important role in bacterial survival (Mizrahi *et al.*, 1998). The specific induction of specialised DNA polymerases that are not stalled by DNA lesions allows the bacterium to tolerate limited amounts of DNA damage. The presence of a DNA-damage-inducible mutagenesis system in *M. tuberculosis*, as well as *Mycobacterium smegmatis*, was recently demonstrated by Boshoff *et al.* (2003) and was found to be mediated by DnaE2, a member of the C-family of DNA polymerases. This result is interesting for two reasons. Firstly, induced mutagenesis in other bacterial species, such as *Escherichia coli*, is mediated by polymerases belonging to the Y-family, and secondly, members of the Y-family of polymerases are present in *M. tuberculosis* and *M. smegmatis*. The primary role of the Y-family polymerases in mycobacteria therefore is an important avenue of investigation.

1.1 The Biology of Mycobacteria

1.1.1 Pathophysiology of tuberculosis disease

Mankind has suffered from TB for more than 5,000 years (Collins *et al.*, 1998). The identification of the tubercle bacillus, *M. tuberculosis*, as the causative agent in 1882 by Robert Koch firmly established the infectious nature of the disease (Koch, 1882). Still *M. tuberculosis* continues to kill approximately three million people each year, and approximately two billion people host the mycobacterium (Dye *et al.*, 1999). Despite a century of study, the ways in which *M. tuberculosis*

neutralises the host's efficient cellular defence mechanism remains to be elucidated.

Infection starts with the inhalation of small droplets containing *M. tuberculosis.* The mycobacterial droplets make their way to the lung where they are internalised by alveolar macrophages (Dannenberg and Rook, 1994). The mycobacterium spends almost its entire life inside the macrophage. *M. tuberculosis* avoids destruction within the macrophage phagosomes by actively blocking their fusion with lysosomes (Armstrong *et al.*, 1971, Malik *et al.*, 2001). The mycobacterium multiplies within the inactivated macrophage until the macrophage bursts. Lymphocytes begin to infiltrate and recognise *M. tuberculosis* antigens, which results in the activation of T-cells followed by liberation of cytokines, and tubercle formation (Dannenberg and Rook, 1994). As a result, the mycobacterium is subjected to a wide variety of growth inhibiting conditions such as low pH, oxygen limitation and reactive nitrogen and oxygen intermediates (Jackett *et al.*, 1978). An individual infected with *M. tuberculosis* has about a 10 % lifetime risk of developing active TB (Corbett *et al.*, 2003). Active TB can be brought on by various factors in which the immune response is no longer able to keep the mycobacterial growth under control. As a result, rapid multiplication occurs and the mycobacteria spread throughout the lung causing necrosis and rupture of the bronchi.

1.1.2 Tuberculosis treatment

Before the introduction of effective chemotherapeutic agents, treatment of TB involved rest, balanced diet and various surgical procedures. The development of a vaccine was realised by Calmette and Guérin after serially passaging *Mycobacterium bovis* and discovering an attenuated strain, namely bacille Calmette-Guérin (BCG), in guinea pigs, rabbits, cattle and horses (Collins, 1998). BCG was first administered to humans in 1921, and from 1948, it has been applied worldwide (Wang *et al.*, 2002). Treatment of active TB became possible during the 1940's with the discovery of *M. tuberculosis* sensitivity to streptomycin, which proved to be an effective anti-microbial agent (Ross, 1950). Isoniazid was discovered in the 1950's and was found to be a very effective anti-TB agent with few side effects (Knox *et al.*, 1952). This led to the identification of a host of other antibiotics that were effective against *M. tuberculosis*, such as pyrazinamide, ethambutol and rifampicin (Yeager *et al.*, 1952, Ferebee *et al.*, 1966 and Furesz *et al.*, 1963). Many of these antibiotics are still used today as front line anti-TB agents.

Dismantling of TB clinics and follow-up programs in developed countries occurred in the second half of last century as a decline in active TB was observed. However, during the 1980's this decline was reversed with increasing numbers of TB cases associated with HIV infection (Houston *et al.*, 1994). Active TB is greatly enhanced in HIV-positive individuals, increasing the mortality rates for both diseases (Frieden *et al.*, 2003). In persons co-infected with HIV and *M. tuberculosis* the annual risk of developing active TB is in excess of 10 % (Corbett

et al., 2003). Along with this, non-compliance of patients with the lengthy antibiotic regimes has resulted in the emergence of *M. tuberculosis* strains that are resistant to one or more of the front-line antibiotics used to treat the disease. Problems arise when individuals are infected with multi-drug-resistant (MDR) strains, where two or more of the front-line antibiotics are no longer effective. Co-infection of individuals with MDR *M. tuberculosis* and HIV results in an accelerated progression of active TB (Corbett *et al.*, 2003). Understanding the mechanisms of pathogenicity and host-bacteria interactions is proving to be vital in light of the increase in MDR TB and the rise in HIV-associated active TB.

1.1.3 *M. smegmatis* as a model organism for *M. tuberculosis*

Even though the entire genome of *M. tuberculosis* has been sequenced, attempts to identify determinants of virulence have been hampered by the bacterium's slow growth rate (twenty-four hours under optimum conditions) and the biosafety conditions under which experiments must be done. These restrictions have necessitated the development and characterisation of a surrogate bacterial model for *M. tuberculosis*. Initial studies were performed on *E. coli* (Lathigra *et al.*, 1985). However, there are significant differences between *E. coli* and *M. tuberculosis* in basic cellular and cell wall structure. Furthermore, fundamental metabolism and lifestyle differences between *E. coli* and *M. tuberculosis* led to the search for a more relevant surrogate model organism. A non-pathogenic and relatively fast-growing member of the mycobacteria genus was needed. The organism of choice was the mycobacterial saprophyte, *M. smegmatis*. *M. smegmatis* has a dividing time of three hours and *in vitro* work can be done under

standard laboratory conditions. The discovery of an efficient plasmid transformation (*ept*) mutant of *M. smegmatis* by Snapper *et al.* (1990) allowed it to be readily transformed with foreign DNA thus paving the way for its widespread laboratory use.

Doubts have been raised concerning the use of *M. smegmatis* as a comparable model of *M. tuberculosis* virulence owing largely to differences in pathogenicity and a larger gene complement. However, when studying conserved biochemical pathways, *M. smegmatis* can prove to be very useful model organism. Considerable sequence conservation exists between classes of genes that participate in sensing the environment and the appropriate response to that stress (Tyagi *et al.*, 2002). Six of the eleven two-component systems and five of the seven orphan response-regulator/histidine kinases of *M. tuberculosis* have homologues in *M. smegmatis*, as well as nine out of the thirteen *M. tuberculosis* sigma factors (Tyagi *et al.*, 2002). The study of *M. smegmatis* has proven to be useful in gaining valuable insights into specific aspects of physiological adaptation of *M. tuberculosis* in shorter periods of time and under more amenable study conditions. *M. smegmatis* is not only valuable as a model of *M. tuberculosis* biology but also for use in understanding key differences between the two organisms. By investigating these differences, the virulence determinants of *M. tuberculosis* may be identified.

1.2 Mutagenesis

The ability to maintain functional homeostasis in a continually changing environment is essential for survival of all organisms. Individual bacteria with high mutation rates have a selective advantage in a changing environment (Foster, 2004). Mutations can arise spontaneously during growth or mutagenic agents can induce them. Spontaneous mutations are a result of errors in replication, spontaneous lesions or transposition events (Kunkel *et al.*, 1984 and Hazelbauer *et al.*, 1979). Replication errors result in the formation of mutations through multiple processes. Transitions or transversions produce base pair substitutions, whereas the deletion or addition of a base pair results in frameshift mutations (Echols and Goodman, 1991). Large deletions or duplications arise by slippage or recombination based mechanisms (Lovett, 2004). Induced mutations are those caused by the action of a specific mutagen. Induced mutagenesis occurs when a base is damaged so it can no longer pair with other base (Cornell *et al.*, 2000), the base is replaced by the incorporation of base analogues (Negishi *et al.*, 1994), or mispairing occurs after DNA is exposed to alkylating agents (Drake, 1988). Loss of specific base pairing occurs after ultraviolet (UV) light damages the base (Franklin *et al.*, 1985).

Bacterial survival requires that the lethal DNA lesions are repaired; however, the extent to which the original genetic information is restored depends on the essentiality of the gene and the type of mutation (Smith *et al.*, 1998). Direct reversal of the damaged base by photoreactivation repairs thymine dimers (Tang *et al.*, 1978). Other forms of DNA lesions are repaired by specific mechanisms

aimed at excising damaged bases. These include: nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR) (Friedberg, 2003). All function by removing the damaged region and inserting of new bases to fill the gap, followed by ligation of the pieces (Friedberg, 2003). NER repairs damage located in regions that are undergoing active gene expression whereas BER is involved in global genome repair (Friedberg *et al.*, 1995). Translesion synthesis is a mechanism in which the cell tolerates damage by bypassing the DNA lesion and allowing replication to occur downstream of the damaged DNA (Goodman, 2002). This process involves the action of error-prone polymerases (discussed further in sections 1.3 and 1.4).

1.2.1 Stress-induced mutagenesis

'Adaptive', 'stationary-phase' and 'stress-induced' mutation are names for the mutation responses observed in organisms exposed to growth-limiting environments, in which mutations appear to be formed in response to the environment (reviewed in Foster, 1999 and Rosenberg, 2001). In bacteria, prolonged nutritional stress results in an increase in mutations (Wright *et al.*, 1999). Controversy exists surrounding the models that describe the origin of the observed adaptive mutagenesis (Rosenberg *et al.*, 2004 and Roth *et al.*, 2004). Two models, the directed mutation and hypermutation models, are in accordance with the theory that mutations arise under stress, producing mutants that have a selective advantage and are the basis for the second-order selection hypothesis. The directed mutation model states that mutations might be targeted specifically to genes that relieve stress, whereas the hypermutation model states that mutation

rates increase genome wide in response to stress (Rosenberg *et al.*, 2004). In contrast, the amplification mutagenesis model states that stress has no direct effect on the mutation rate and that mutations arise in cells growing under selection (Roth *et al.*, 2004). Mutations that arise in this manner form the basis of the pleiotropic hypothesis. Evidence exists for both hypotheses and no strong arguments rejecting either model are presently available, making it probable that both scenarios act simultaneously but to a different extent in stress-induced mutagenesis (Tenaillon *et al.*, 2004).

The knowledge of the genetic control of stress-induced mutagenesis arose from studies where bacterial cells were stressed using chemical and physical treatments, one of the first being UV irradiation that resulted in an increase of the mutation rate (Friedberg *et al.*, 1995). Starvation-induced mutagenesis has provided valuable insight into the understanding of the evolutionary role of genetic control of stress-induced mutagenesis (Tenaillon *et al.*, 2004). A stationary-phase-specific process for long periods of nutritional stress, “growth advantage in stationary phase” (GASP), depends on the appearance of new mutations that confer a competitive advantage to certain cells, allowing them to take over the population (Zambrano and Kolter, 1996). The appearance of a GASP phenotype indicates the dynamic state of the bacterial cells during stationary phase growth.

1.2.2 Heritable and Environment-dependent Mutators

Metzgar and Wills (2000) postulated that adaptive mutagenesis does not represent the specific generation of adaptive mutations but rather a localisation of mutations under specific environmental conditions or in particular regions of the genome. Modulation of mutation rates is achieved through either environment-dependent or heritable mechanisms. Heritable mechanisms are independent of the environment and alter either the global (genome-wide) mutation rate or target hotspots (hypervariable loci). Environment-dependent mechanisms induce or suppress mutation-inducing processes that have a global effect on the mutation rate.

Global heritable mutators appear to be adaptive under certain conditions and are likely to arise by chance. Bacterial mutator phenotypes have been traced to mutations in DNA repair genes, particularly in the MMR system (Metzgar *et al.*, 2000). Global mutators allow mutations to arise across the entire genome and have been shown to out-compete non-mutators under varying selection experiments (Miller *et al.*, 1999). Local mutators on the other hand are specific regions of the genome that are predisposed to mutation, due to their unique sequence characteristics. Tandem repeats interfere with DNA polymerase activity and result in replication slippage errors within tandem repeats (Moxon *et al.*, 1994). Mutations have also been shown to occur at higher frequencies in certain genes, indicating that these genetic sites are “hotspots” for mutations (Torkelson *et al.*, 1997). Well-known highly mutable sequences are contingency loci, generating the antigen diversity observed in pathogenic prokaryotes (Moxon *et*

al., 1994).

Environment-dependent mutators are induced in response to environmental stresses (such as nutrient deprivation). The mutator phenotype is only transient under unfavourable conditions. Mutations arise as a result of the induction of these mutators, which are part of stress response systems. The prototypical environment-dependent mutator system is the bacterial SOS response (see section 1.3) (Metzgar *et al.,* 2000). Yeiser *et al.* (2002) observed that in *E. coli,* SOS polymerase deficient cells showed a competitive disadvantage during long-term survival when cultured with wild type. The expression of the SOS polymerases during stationary phase in the absence of exogenous DNA damage presents a role for the polymerases in conferring a GASP phenotype during long-term survival (Yeiser *et al.,* 2002). The ability to survive under conditions of stress is vital for bacterial propagation. Insults on the genome due to environmental stresses pose a constant burden on the survival of the bacterium. A better understanding of the stress response systems that occur under conditions of stress may help explain the mechanisms of adaptive mutagenesis.

1.3 The SOS Response

Mutagenesis by the SOS response does not occur from the high-fidelity replicative DNA machinery but rather from specialised error-prone DNA synthesis (Goodman, 2002). The complete mechanism of the prototypical SOS response system was elucidated from several studies on the effect of UV irradiation on *E. coli*. UV-irradiated bacteriophage lambda was initially shown to

have an increased number of mutations when infected in *E. coli* that had been previously irradiated (Weigle, 1953). Later, Radman (1974) proposed that when bacteria are exposed to stress (such as UV irradiation) they produce proteins that allow the repair of damaged DNA and reactivation of DNA synthesis. It was further speculated that the 'danger' signal might be a temporary block in DNA replication or the presence of certain DNA lesions (Radman, 1974). This phenomenon was termed the SOS response to emphasise the cellular response to a distress signal, such as DNA damage.

Friedberg (1995) proposed three reasons for the presence of the SOS response. Firstly, it allows an organism to mutate in response to DNA damage in order to confer a selective advantage (still under debate). Secondly, it provides an additional mechanism for damage tolerance that may help the organism to survive lethal effects of DNA damage. Thirdly, it allows repair of particular DNA lesions that are not subject to other accurate repair systems.

1.3.1 Mechanism of the SOS Response

The *E. coli* SOS inducible repair response is the best-characterised system and thus its general mechanism of action is discussed below. Induction of the SOS response in *E. coli* after DNA damage results in increased expression of approximately forty genes (Courcelle *et al.*, 2001 and Abella *et al.*, 2004). The SOS controlled genes are scattered at different sites in the genome and their expression is dependent on the interaction between two proteins, LexA and RecA (Figure 1.1). LexA binds to a similar regulatory sequence in all SOS controlled

genes, including itself and *recA,* and represses transcription of these genes. The regulatory sequence to which the LexA repressor binds varies amongst gram-positive and gram-negative bacteria (Lewis *et al.,* 1994 and Cheo *et al.,* 1991). The prototypical regulatory sequence for gram-positive bacteria was identified in *B. subtilis* and referred to as the Cheo box (Cheo *et al.,* 1991). The Cheo box has been identified in mycobacteria to which LexA repressor interacts (Durbach *et al.,* 1997).

Increased expression of the SOS controlled genes is triggered when DNA is damaged. At points of DNA damage, single-stranded DNA accumulates. RecA, in the presence of ATP, forms filaments on the single-stranded DNA and acquires protease activity. The activated RecA/single-stranded nucleoprotein filament stimulates the otherwise latent capacity of LexA to autodigest (Little, 1984). LexA autodigests at its $Ala^{84} – Gly^{85}$ bond which inactivates its ability to act as a repressor (Horii *et al.,* 1981). The resulting decrease in LexA allows an increase in transcription of the genes under its control, resulting in the SOS response. The RecA/single-stranded DNA nucleofilament also facilitates the autocleavage the bacteriophage CI repressor as well as processing UmuD protein to its activated form UmuD' (Eguchi *et al.,* 1988, Shinagawa *et al.,* 1988). The *recA* locus was the first to be recognised as being under SOS control followed by *lexA.* Additional SOS-controlled genes were then identified and were initially termed *din* (damage-inducible). Since the discovery of the initial damage-inducible genes by Kenyon and Walker (1980), a wide range of members belonging to the SOS regulon have been identified.

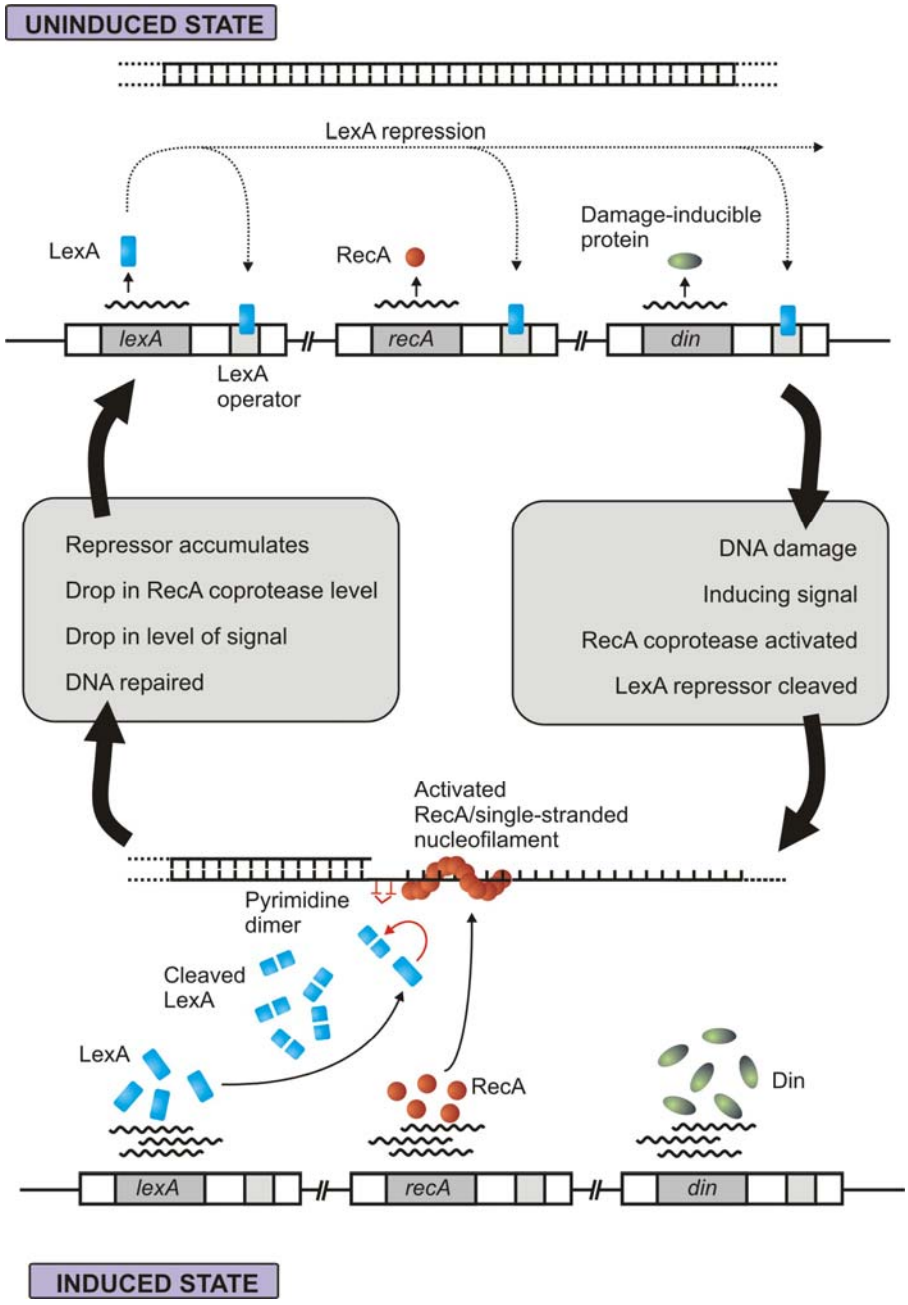


Figure 1.1: Diagrammatic representation of the SOS response system and the mechanism by which the *lexA-recA* regulon is regulated. In the uninduced state (top), LexA repressor is constitutively expressed in small amounts and binds to its regulatory sequence in the operators of all the genes under its control. Following DNA damage (pyrimidine dimer formed after exposure to UV radiation) the SOS response is induced (below). RecA binds single-stranded DNA and becomes activated, resulting in the proteolytic cleavage of LexA. Expression of the genes under the control of LexA then occurs. (Adapted from Friedberg *et al.*, 1995).

The SOS response is finely tuned to aid the cell in surviving insults on the genome, and exists in different induced states from fully induced to fully repressed (Janion, 2001). The repression of the *lexA* gene by itself leads to various induction levels of the SOS response. Firstly, it extends the range over which an intermediate state of induction can be established (Walker *et al.*, 2000). This is also observed with *recA* as its expression depends on the amount of LexA that is able to bind to its regulatory control sequence. Secondly, LexA's affinity for its own regulatory region is weak when compared to the other SOS controlled regulatory regions, allowing buffering against significant induction levels (Walker *et al.*, 2000). Thirdly, since LexA is auto-regulated, it returns quickly to its repressed state once RecA returns to an inactivated state (Walker *et al.*, 2000). LexA represses genes under control of the SOS response with varying efficiency. This allows high expression levels of genes that are under weak LexA repression. In *E. coli* the induction can range from about 100-fold for the *sulA* gene to about 4-fold for the *uvrA* gene (Friedberg *et al.*, 1995).

1.3.2 Cellular SOS response for managing damaged DNA

The co-ordinated actions of the SOS controlled gene products manage many cellular events after DNA damage (Figure 1.2) (Walker *et al.*, 2000). Loss-of-function mutations in *umuC* and *umuD* showed that the products of these genes were mutagenic and processed damaged DNA (Elledge *et al.*, 1983). The *umuDC* gene product (*E. coli* Pol V) was shown to introduce mutations through a process known as translesion synthesis, allowing the replication fork to move over DNA lesions (Fujii *et al.*, 2004).

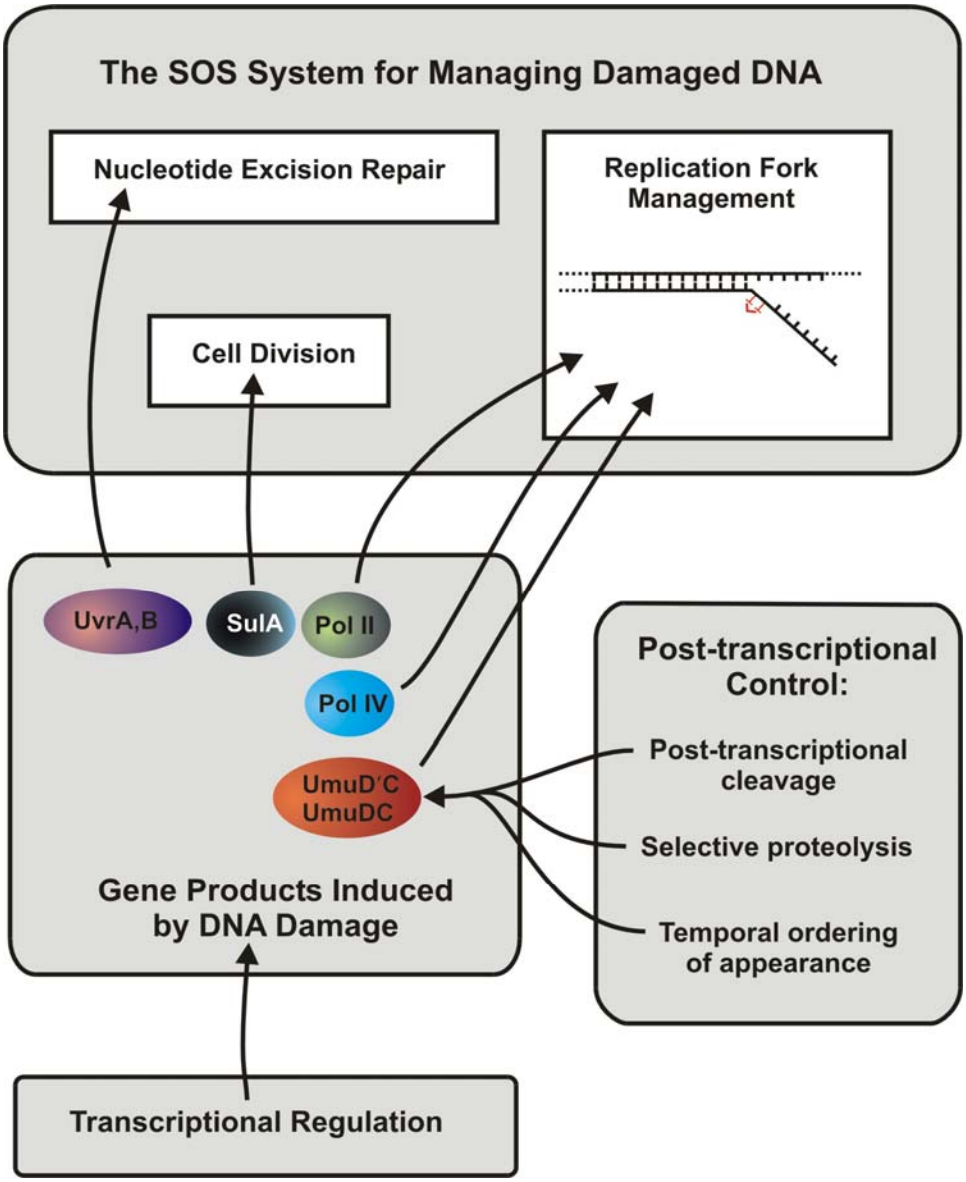


Figure 1.2: Cellular SOS response to DNA damage. After induction of the SOS response the SOS controlled genes are expressed, among them being *uvrA*, *uvrB*, *sulA*, *polB*, *dinB* (encoding Pol IV) and *umuDC* (encoding Pol V). SulA inhibits cell division. Initially after induction of the SOS response UmuD together with UmuC function to regulate DNA synthesis. First nucleotide excision repair is used and after a defined amount of time UmuD is then activated to UmuD' for translesion synthesis. (Adapted from Walker *et al.*, 2000).

Managing the stalled replication fork represents just one element of the entire cellular SOS response. Many of the gene products that are involved in NER (e.g. UvrA, UvrB and UvrD) are under the control of the SOS response. Whilst most of the proteins induced by the SOS response are involved in DNA metabolism, SulA acts to inhibit cell division by preventing polymerisation of FtsZ (Huisman *et al.*, 1981, Mukherjee *et al.*, 1998). Other polymerases (*E. coli* Pol II and IV) and recombination proteins (RecN) play additional roles in managing the replication fork. The SOS response comprises multiple levels of action through the use of gene products under SOS control. The redundancy of mechanisms used to overcome damaged DNA reveals the complex nature of the SOS response.

1.3.3 The SOS response in other bacterial species

The *E. coli* SOS response has become a paradigm for the study of inducible repair and recombination processes in many different organisms. As the number of bacterial organisms that are studied increases, it is clear that the components of the SOS response appear to be highly conserved among bacterial species. Many similarities have been found between gram-positive and gram-negative bacteria. Sequence homology between the damage inducible genes involved in translesion synthesis suggests that they shared a common ancestor (Walker, 1995). Each organism has its own set of regulated genes that are induced in response to DNA damage. The regulatory mechanisms, however, can vary among bacterial species. Four types of SOS regulon phenomena can be distinguished in *Bacillus subtilis* (Yasbin *et al.*, 1992). The regulatory sequence for SOS controlled genes, where the SOS repressor binds, varies between the gram-positive and gram-negative

bacteria (Wertman *et al.*, 1985, Cheo *et al.*, 1991). It appears that the general mechanisms of inducible repair processes are comparable between bacterial species but the key players involved need to be elucidated for each organism.

1.4 Mutator Polymerases

Organisms have high-fidelity polymerases that function to accurately replicate the genome, and low-fidelity polymerases with more specialised functions. The main DNA replicative polymerases possess high fidelity in base selection and only make an error every ~10^5 base pairs (Friedberg *et al.*, 1995). With the intrinsic exonucleolytic proofreading in DNA replication, coupled with post-replicative MMR, the overall error-rate is reduced to 1 in 10^{10} (Schaaper, 1993). The replicative polymerases however are sensitive to lesions in the DNA template (Echols and Goodman, 1991). Many organisms have equipped themselves with functionally redundant DNA repair pathways for dealing with damage to the genome, such as BER, NER, MMR and translesion synthesis (as part of the SOS response) that utilise various polymerases that allow the organism to replicate past DNA lesions (Friedberg *et al.*, 1995).

1.4.1 Various *E. coli* polymerase families

Five different DNA polymerases have been identified in *E. coli*, so far, and are numbered Pol I to Pol V. The non-inducible polymerases are Pol I (the main DNA repair enzyme) and Pol III (the main replicative polymerase) (Karkas *et al.*, 1972, McHenry *et al.*, 1977). The remaining Pol II, Pol IV and Pol V are induced during

the SOS response (Napolitano *et al.*, 2000). These five *E. coli* polymerases fall into four different polymerase families (Table 1.1). The A-family comprises of polymerases showing similar sequence homology to *E. coli* Pol I. These polymerases are regarded as high-fidelity polymerases and are necessary for the removal of RNA primers from Okazaki fragments during DNA replication (Ollis *et al.*, 1985). The B-family are homologous to Pol II, a SOS induced polymerase, and play a role in replication restart that bypasses DNA damage in an error-free manner (Pham *et al.*, 2001). The C-family polymerases are the main replicative polymerases in bacterial species and are used for DNA synthesis (Kornberg *et al.*, 1972). The final family in *E. coli* is the Y-family of polymerases. These polymerases in *E. coli* are also induced as part of the SOS response and have been shown to be error-prone (Ling *et al.*, 2001). The Y-family polymerases share common structural features that are useful for classification. However, functionally, the Y-family polymerases differ between the bacterial species and appear to be specific to the type of lesion that they can bypass (Table 1.1).

1.4.2 Y -family of DNA polymerases

Many of the polymerases that replicate damaged DNA in an error-prone manner are phylogenetically related to each other and belong to the UmuC/DinB/Rev1/ Rad30 superfamily, and are collectively known as the Y-family of polymerases (Ohmori *et al.*, 2001). Some of the polymerases belonging to this family make errors 100 times more frequently than normal replicative polymerases (McKenzie *et al.*, 2001).

Table 1.1: Bacterial DNA polymerases[a]

Family	Polymerase/ gene name	Organism	Fidelity on undamaged DNA	Function
A	Pol I/*polA*	*E.coli*	10^{-5}-10^{-6}	Okazaki maturation
B	Pol II/*polB*	*E.coli*	10^{-5}-10^{-6}	TLS, SOS mutagenesis, replication restart
C	Pol III/*polC*	*E.coli*	10^{-5}-10^{-6}	Chromosome replication
	dnaE	*S. pyogenes*	10^{-2}-10-3	Chromosome replication, TLS, SOS mutagenesis
	dnaE2	*M. tuberculosis*		Chromosome replication, TLS, SOS mutagenesis
Y	Pol IV/*dinB*	*E.coli*	10^{-3}-10^{-4}	TLS, SOS mutagenesis
	yqjH	*B. subtilis*		Untargeted mutagenesis
	dinP/dinX	*M. tuberculosis*		?
Y	Pol V/ *umuDC*	*E.coli*	10^{-3}-10^{-4}	TLS, SOS mutagenesis
	yqjW	*B. subtilis*		UV-induced mutagenesis
	?	*M. tuberculosis*		?

[a] Adapted from Rattray *et al.*, 2003
TLS – translesion synthesis
S. pyogenes – Streptococcus pyogenes, B. subtilis – Bacillus subtilis

Y-family polymerases possess five highly conserved regions, at the N-terminus, presumed to be involved in binding and catalysis (Goodman *et al.*, 2000). The C-terminal half contains sequence motifs for interactions with the replication processivity factor (Wagner *et al.*, 2000b). Y-family polymerases lack a 3´-5´ exonuclease activity (Yang, 2005). All crystal structures of Y-family polymerases consist of the catalytic core of 'finger', 'thumb' and 'palm' domains arranged in the classic polymerase 'right hand-like' configuration (Yang, 2003). However, Y-family polymerases contain an additional domain known as the 'little finger' which plays a role in holding the DNA duplex opposite the thumb domain (Ling *et al.*, 2001). Watson-Crick base pairs have a flat and smooth minor groove, but the minor groove of mismatched base pairs is uneven (Yang, 2005). The lack of a complementary interface between the polymerase and replicating base results in the high error-rate observed for low fidelity DNA synthesis.

Of particular interest in bacterial studies are Y-family polymerases belonging to DinB and UmuC subfamilies. The DinB subfamily consists of the *E. coli* DNA Pol IV, which is encoded by the *dinB* gene[1] (Wagner *et al.*, 1999). Pol IV has been demonstrated to have mutagenic properties since cells carrying a *dinB* null allele are defective in phage lambda untargeted mutagenesis (Brotcorne-Lannoye *et al.*, 1986). Pol IV has been shown to introduce -1 frameshift deletions by skipping over a lesion during replication (Kobayashi *et al.*, 2002). Pol IV adds nucleotides in a distributive manner (Wagner *et al.*, 1999). Pol IV overexpression leads to a significant increase in the mutation rate, even in the absence of DNA

[1] Ohmori (1995) sequenced a damage inducible protein and called it *dinP*. It was later discovered that *dinP* was allelic with *dinB*, originally identified by Kenyon and Walker in 1980. *dinB* is still called *dinP* in certain bacterial species, i.e. mycobacteria.

damaging agents, with frameshifts targeted to mononucleotide repeats in particular (Wagner *et al.*, 2000a). The *E. coli* Pol V belongs to the UmuC subfamily and is encoded by *umuDC* (Tang *et al.*, 1999). Pol V generates base substitutions in response to UV irradiation or abasic sites (Smith and Walker, 1998) via translesion synthesis, in which a nucleotide is misincorporated directly opposite a damaged DNA nucleotide and followed by accurate DNA synthesis downstream of the lesion. As mentioned in section 1.2.2 the Y-family polymerases also play a role in long-term survival of the bacterium.

1.4.3 Translesion synthesis

Recent findings have suggested that all organisms accomplish translesion synthesis by a common mechanism (Woodgate, 1999). In *E. coli*, translesion synthesis is intimately associated with the SOS response and involves the main DNA replicative polymerase, Pol III, in co-operation with the SOS controlled Y-family polymerases (Witkin *et al.*, 1984, Woodgate *et al.*, 1989, Napolitano *et al.*, 2000). Pol III is unable to bypass the DNA lesion and dissociates from the DNA template, leaving a 3′ primer terminus prior to the lesion. Translesion synthesis replication is then carried out by one of the error-prone polymerases. *E. coli* Pol V consists of a dimer of the proteolytically processed UmuD protein, UmuD′, and a monomer of UmuC (Tang *et al.*, 1998). It is thought that the RecA/single-stranded nucleofilament complex delivers the polymerase complex to DNA lesion to perform translesion synthesis (Baynton *et al.*, 2000). A more relaxed base pairing is accommodated by a wider active site in Pol V allowing low-fidelity replication and lesion bypass (Ling *et al.*, 2001). All of the error-prone polymerases appear to

lack significant processivity and probably dissociate from the template shortly after translesion synthesis occurs (Maor-Shoshani *et al.*, 2002). Pol V dependent translesion synthesis occurs in response to UV irradiation and the chemical carcinogen *N*-2-acetylaminofluorene (Napolitano *et al.*, 2000). However, Pol IV and Pol V dependent translesion synthesis occurs in response to benzo(*a*)pyrene (Napolitano *et al.*, 2000). The translesional DNA polymerase required for translesion synthesis appears to depend upon the structure of the lesion and the sequence context.

1.4.4 Error-prone polymerases in mycobacteria

In gram-positive bacteria two distinct subclasses of C-family polymerases have been identified, namely class I (DnaE-type, lacking proofreading abilities) and class II (PolC-like, with intrinsic 3'-5' proofreading exonuclease). *M. tuberculosis* has been shown to contain two polymerases belonging to class I, DnaE1 and DnaE2, but none belonging to class II (Cole *et al.*, 1998). Since deletion of DnaE1 is lethal it is thought to be the main replicative polymerase, whereas DnaE2 does not affect survival (Boshoff *et al.*, 2003). Boshoff *et al.* (2003) observed elevated mutational levels after UV irradiation in *M. tuberculosis* and *M. smegmatis* that appeared to be exclusively mediated by DnaE2. This is intriguing, as polymerases showing homology to Y-family polymerases have been identified in mycobacteria.

Sequence analysis of the complete genome of *M. tuberculosis* revealed two Y-family polymerases with homology to *E. coli dinB* (Pol IV), namely *dinP*

(Rv3056) and *dinX* (Rv1537)(Cole *et al.*, 1998). Three polymerases with close homology to the two *M. tuberculosis* Y-family polymerases exist in *M. smegmatis*. The relative homologues of *M. tuberculosis dinP* (MSMEG1002)[2], which was found to be duplicated in the genome, and *dinX* (MSMEG3178) were identified as well as another homologue, *dinP3* (MSMEG6405), showing close homology to *dinP*. The *M. tuberculosis* Y-family genes, *dinP* and *dinX,* are not up-regulated after treatment with DNA damaging agents such as mitomycin C (Brooks *et al.,* 2001), UV or hydrogen peroxide (Boshoff *et al.,* 2003). This suggests that members of the Y-family of error-prone polymerases in *M. tuberculosis* may not be induced as part of the SOS response. DnaE2, however, is under control of the SOS response and is involved in translesion synthesis (Boshoff *et al.*, 2003). Furthermore, no homologues of *E. coli umuD* have been found in Gram-positive bacteria, including mycobacteria, although it remains to be determined if functional substitute exist. It appears that unique SOS responses have evolved in Gram-positive and Gram-negative bacteria (Tippin *et al.*, 2004 and Duigou *et al.*, 2004). Mycobacteria also lack homologues of genes involved in MMR, which may provide an advantage under conditions of stress (Cole *et al.*, 1998; Mizrahi *et al.*, 1998). The lack of a recognisable MMR system would allow arising mutations to be preserved (thus amplifying the advent of genetic variation).

Drug resistant TB is a serious health problem and thus a thorough understanding of how drug resistance emerges in *M. tuberculosis* remains an important objective. The ability of *M. tuberculosis* to persist and develop antibiotic resistance during

[2] The *M. smegmatis* damage inducible genes will be referred to by their gene name and not their annotated gene number throughout the dissertation.

infection appears to be achieved through the action of the DnaE2-dependent mutagenesis pathway (Boshoff *et al.*, 2003). The functional role of Y-family polymerases in mycobacteria needs to be determined especially since DnaE2 mediates translesion synthesis. The presence of the three Y-family polymerases in *M. smegmatis* may help to define the role of Y-family polymerases in mycobacteria and may provide a better understanding of adaptive mutagenesis. By determining when and how the Y-family polymerases function in mycobacteria, we may begin to explain how *M. tuberculosis* persists and develops multiple-drug resistance.

1.5 Aims of Study

The role of the third Y-family polymerase, *dinP3*, in *M. smegmatis* was characterised with the following objectives:

1. To determine whether *dinP3* is damage inducible by UV irradiation.
2. To determine whether *dinP3* is regulated by LexA.
3. To generate and phenotypically characterise a *dinP3* deletion mutant of *M. smegmatis*.
4. To evaluate contribution of both inducible and non-inducible Y-family polymerases to damage tolerance in *M. smegmatis*.
5. To assess the contribution of DinP and DinP3 to mutagenesis by undertaking mutation rate studies *in vitro* that detect –1 frameshift mutations.

2.0 MATERIALS AND METHODS

2.1 Bacterial strains and growth conditions

All strains used in this study were stored at –70°C in 50 % glycerol. Lists of strains used are detailed in Table 2.1. All strains were recovered by streaking onto Luria-Bertani agar (LA) plates (see Appendix A1), and incubated for 3 days at 37°C.

Table 2.1: Bacterial Strains

Strain	***Genotype***
mc^2155[a]	High frequency transformation mutant of *M. smegmatis*
ΔdinP[b]	*dinP* mutant of mc^2155 (both *dinP* copies deleted), Hyg^r
ΔdinP3[c]	Unmarked *dinP3* deletion mutant of mc^2155
ΔdinP (hisD^{5T})[c]	*dinP* mutant of mc^2155 in *hisD* mutant background[d]
ΔdinP3 (hisD^{5T})[c]	Unmarked *dinP3* mutant of mc^2155 in *hisD* mutant background[d]

[a] Snapper *et al.*, 1990
[b] Constructed in our laboratory by D.F. Warner
[c] Constructed in our laboratory by E.E. Machowski
[d] See section 2.6.1. for description of *hisD* mutant background

2.1.1 Growth curve of *M. smegmatis* mc^2155

A growth curve was generated to assess the growth characteristics of *M. smegmatis*. This was achieved by diluting a log phase culture 100-fold in 7H9 liquid media (see Appendix A2). The culture was incubated in a New Brunswick

series 25 Orbital Incubator (New Brunswick, Inc, USA) at 37^{o}C and aliquots were taken at three-hour time points, from 0 hrs to 24 hrs. The optical density (OD), at a wavelength of 600nm, of the aliquots was recorded at each of the three-hour time points. A dilution series (see Appendix B1) was generated, and the 10^{-4} to 10^{-6} dilutions were plated, in triplicate, on 7H10 agar plates (see Appendix A3) to determine the number of colony forming units (CFU).

2.2 Gene expression analysis

2.2.1 DNA damaging treatment

The effect of DNA damage on the expression of the *M. smegmatis* Y-family gene, *dinP3*, was analysed. Mycobacterial cultures (30 mL) were grown to an OD_{600} of 1.0 in 7H9 liquid media (see Appendix A2) at 37^{o}C in a New Brunswick series 25 Orbital Incubator (New Brunswick, Inc, USA). The cells were then harvested using an Beckman Coulter JA-20 rotor centrifuge (Beckman Coulter, USA) for 5 min at 5 000 rpm at 4^{o}C and re-suspended in 5 mL of 0.5 % Tween 80 (see Appendix A4) before being transferred to a 78cm^2 petri dish. At this time, cultures were divided into two groups, one of which was exposed to UV radiation while the other was mock irradiated and used as a control. The culture was irradiated at 40 mJ/cm^2 in a UV Stratalinker 1800 (Stratagene, USA) before being re-suspended in the same original volume of 7H9 liquid media (see Appendix A2) and returned to 37^{o}C for 90 minutes. The mock-irradiated control was also re-suspended in the same original volume of media and incubated at 37^{o}C for 90 minutes. Total RNA was extracted and the mRNA expression level was assessed

by semi-quantitative Reverse Transcription (RT) Polymerase Chain Reaction (PCR).

2.2.2 RNA isolation and purification

Extraction of RNA was carried out in accordance with established extraction techniques (Manganelli *et al.*, 2001). Cells were harvested using a Beckman Coulter JA-20 rotor centrifuge (Beckman Coulter, USA) for 5 min at 5 000 rpm at 4^{o}C and then re-suspended in 1 mL of Tri reagent (Sigma-Aldrich, USA) containing 5 μL polyacryl carrier (Molecular Research Center, USA). The suspension was then transferred into 2 mL screw-capped polypropylene centrifuge tubes containing zirconia/silica beads (QBiogene, USA) and disrupted using a BIO 101/Savant FastPrep 120 cell disrupter (QBiogene, USA). The lysate was clarified using an Eppendorf 5415D centrifuge (Eppendorf International) for 1 min at 10 000 rpm. The lysate was then transferred to a 1.5 mL micro-centrifuge tube (Molecular BioProducts, USA) and 100 μL 1-Bromo-3-chloropropane (Sigma-Aldrich, USA) was added. The suspension was mixed vigorously by hand for a few seconds. Separation of phases occurred by incubating at room temperature for 10 min, followed by centrifugation (model 5415D, Eppendorf International) for 10 min at 10 000 rpm. Isopropanol (350 μL) was then added to the upper phase and precipitation of the RNA occurred at room temperature for 10 min before centrifugation for 10 min at 10 000 rpm. The RNA pellet was then re-suspended in 300 μL 75% Ethanol (see Appendix A5) and re-centrifuged for 5 min at 10 000 rpm. The pellet was air dried for 5 min before being re-suspended in 45 μL DEPC-treated distilled H_2O (see Appendix A6).

Genomic DNA was removed by subjecting the sample, in a final volume of 50 μL, to a round of DNaseI treatment, using 0.5 U DNaseI (Ambion, USA) in 1× DNaseI reaction buffer (Ambion, USA). The Tri reagent extraction procedure was then repeated using half the volume of each solution from the first extraction round. The quality of the RNA was checked by RNA electrophoresis (see Appendix B2). RNA was considered acceptable if the amount of 23S rRNA was twice that of 16S rRNA (Manganelli *et al.*, 1999). RNA was stored at –70^{o}C.

2.2.3 Primer design for RT-PCR

All primers were synthesised using standard phosphoramadite chemistry by Inqaba Biotech, Inc (South Africa). The primers used for semi-quantitative RT-PCR are listed in Table 2.2. All RT-PCR primers were designed to anneal to their target DNA at the same temperature (55^{o}C) and to amplify DNA fragments internal to the coding sequence of the relevant genes (See Appendix C1). Primers were designed using the program Primer 3 (Rozen *et al.*, 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primer 3 determines both forward and reverse primers for the PCR reactions and assesses the oligonucleotide melting temperature, size, G+C content and primer-dimer possibilities. It calculates the optimal primer pair according to the size and constraints the user specifies (parameters used; primer length: 20 nucleotides, product size: 150-200 bp, melting temperature: 60^{o}C). It also analyses the positional constraints within the source sequence when designing primers (the entire nucleotide sequence for each gene used in the PCR).

2.2.4 Reverse transcription of extracted RNA

The RT-PCR procedure was modified from that previously reported by Manganelli *et al.* (2001), as follows. For each sample, 20 ng/μL of total RNA was added to 0.25 μM (final concentration) of each antisense primer (all antisense primers were used in the same annealing mixture). DEPC-treated H_2O was added to a final volume of 20 μL. All RT reactions were carried out on an Eppendorf MasterCycler (Eppendorf International). After denaturation at 94^{o}C for 1 min and 30 s, annealing between the RNA and antisense primers was carried out for 3 min at 65^{o}C followed by 3 min at 57^{o}C. Subsequently, 10 μL of the annealing mixture was added to 1 × Avian Myeloblastoma Virus (AMV) RT buffer (Sigma-Aldrich, USA), 200 μM each dNTP (Sigma-Aldrich, USA), 4 mM $MgCl_2$ (Sigma-Aldrich, USA), 0.6 μL dimethyl sulphoxide (DMSO)(Sigma-Aldrich, USA) and 2 U Enhanced AMV RT (Sigma-Aldrich, USA) in a final volume of 20 μL. Samples were incubated for 30 min at 60^{o}C, heated at 95^{o}C for 5 min and chilled to 4^{o}C in an Eppendorf MasterCycler (Eppendorf International). The remaining 10μL from the primer annealing mixture was treated as above but without the enhanced AMV RT, to serve as a control for contaminating DNA. The complementary DNA (cDNA) samples were then diluted with 20 μL of distilled H_2O and stored at 20^{o}C.

Table 2.2: Primer sequences used to amplify the *M. smegmatis dinP3* gene

Gene	*Sense primer (5′- 3′)*	*Antisense Primer*(5′- 3′)*	*Size (bp)*
dinP3	CGAGGCGTTCCTCGACGTAT	CACCAAGTTCCTGGCCAAGG	143
sigA	TCTACGCCACGCAGAAGCTG	GTTCGCCTCCAGCAGATGGT	128
dnaE2	CTGCGCAGCAGGTTCTACGA	AGAACACCAGCGAGGCGAAG	145

*cDNA synthesis is primed using these reverse primers.

2.2.5 Polymerase chain reaction of cDNA samples

All PCR reactions were performed in an Eppendorf MasterCycler (Eppendorf International) and conditions were identical for all reactions. The 50 μL reaction consisted of 1× FastStart PCR buffer (without $MgCl_2$) (Roche Applied Science, Germany), 4 mM $MgCl_2$ (Roche Applied Science, Germany), 250 μM each dNTP (Roche Applied Science, Germany), 5 μL DMSO (Roche Applied), 0.5 mg/mL bovine serum albumin (BSA)(Roche Applied Science, Germany), 0.5 μM of each primer, 2.5 U Faststart Taq DNA polymerase (Roche Applied Science, Germany) and 2 μL of either cDNA from the RT reactions or *M. smegmatis* genomic DNA (see Appendix B3). The PCR consisted of an initial heat activation step at 95^{o}C for 10 min followed by 15 cycles of stringent amplification (95^{o}C denaturation for 30 s, 65^{o}C annealing for 30 s and 72^{o}C polymerisation for 30 s). A second round of amplification of 25 cycles was then performed (95^{o}C denaturation for 30 s, 57^{o}C annealing for 30 s and 72^{o}C polymerisation for 30 s). The amount of contaminating chromosomal DNA was tested in parallel using the negative AMV RT control samples. Samples were fractioned on a 2 % agarose gel (see Appendix B4) using a Gel Doc Imaging System (BioRad, USA) and product sizes

were assessed in comparison to molecular weight marker VI (Roche Applied Science, Germany). To quantify the PCR products, a known concentration control is needed. Genomic DNA was extracted from exponentially growing mycobacteria and its concentration was determined using PicoGreen (Molecular Probes, Inc, USA). Every PCR reaction contained a known concentration genomic DNA control and was electrophorised with the RT-PCR samples. Semi-quantification was carried out using the Discovery Series Quantity One 1D Analysis software (BioRad, USA) on the Gel Doc Imaging System (BioRad, USA). The Discovery Series software quantifies differences in band intensity from agarose gel electrophoresis. The software quantifies values as a percentage of a known concentration band. The degree of induction was determined from four independent experiments.

2.3 Gel Mobility Shift Assay

2.3.1 Preparation of *dinP3* probe

Binding of cell lysate to the mycobacterial Cheo box was assessed as indicative of LexA binding. A 149 base pair (bp) probe corresponding overlapping the *dinP3* Cheo box was prepared by PCR using the primers described in Table 2.3 (see Appendix C2). The antisense primer was 5´-[γ^{32}P] end-labelled using polynucleotide kinase (PNK), the following were added together to a final volume of 10 μL on ice, 25 pmol antisense primer, 1× phosphorylation buffer (Roche Applied Science, Germany), 15 μCi of [γ^{32}P] ATP (~3000 μCi/mmol; AEC Amersham International) and 10 U PNK (Roche Applied Science, Germany). The

labelled primer was used directly in the PCR. The PCR reaction (50 μL) consisted of 1× FastStart PCR Buffer (Roche Applied Science, Germany), 250 μM of each dNTP (Roche Applied Science, Germany), 0.5 μM sense primer, 1 μM [γ^{32}P] labelled antisense primer and 2.5 U FastStart Taq DNA polymerase (Roche Applied Science, Germany). The PCR consisted of an initial heat activation step at 95^{o}C for 10 min followed by 30 cycles of amplification (95^{o}C denaturation for 30 s, 58^{o}C annealing for 30 s and 72^{o}C polymerisation for 30 s). A final amplification step at 72^{o}C for 5 min was then performed. The labelled *dinP3* probe was separated on a 1 % low melting agarose gel (see Appendix B5) along with molecular weight marker VI (Roche Applied Science, Germany). The amplified probe was purified from the agarose gel using β-agarase (see Appendix B6) and ethanol precipitation (see Appendix B7) to a final volume of 10 μL. Incorporation of the [γ^{32}P] label was measured using a scintillation counter, 1 μL of [^{32}P]-labelled *dinP3* probe was added to 5 mL of scintillation fluid.

Table 2.3: Sense and Antisense primers used to generate the *dinP3* probe

Name	***Sequence (5′- 3′)***
MS*dinP3* P1	CTCAACAGCATCTACGGCAC
MS*dinP3* P2*	TCGACCGAGGCGTAGAACG

*Primer used for [γ^{32}P] end labelling by PNK.

2.3.2 Preparation of cell lysate extracts

Cell free extracts (total cell protein) were prepared according to Durbach *et al.*, 1997. *M. smegmatis* cultures were grown in 7H9 liquid media (see Appendix A2) at 37^{o}C in a New Brunswick series 25 Orbital Incubator (New Brunswick, Inc, USA) until well grown (OD_{600} greater than 1.5) before being harvested in a Beckman Coulter JA-20 rotor centrifuge. The cell pellets were washed twice with 1 mL Gel Mobility Shift (GMS) assay buffer (see Appendix A7) and re-centrifuged. Pellets were then re-suspended in 700 μL GMS assay buffer (see Appendix A7) and transferred into 2 mL orange screw-caped BIO 101 lysis tubes containing lysing matrices (QBiogene, USA). The cells were subjected to mechanical lysis in the BIO 101/Savant FastPrep 120 cell disrupter (QBiogene, USA). The supernatant was clarified using an Eppendorf 5415D centrifuge (Eppendorf International) for 1 min at 5 000 rpm and stored on ice. The cell free extracts were immediately used in the binding reactions. Protein concentrations relative to BSA were determined using a Bradford assay (see Appendix B8).

2.3.3 DNA binding assay

DNA binding reactions (20 μL) containing 5 000-10 000 cpm [^{32}P]-labelled *dinP3* probe, 1 mg/mL total cell protein lysate, 0.1 mg/mL salmon sperm DNA (Sigma-Aldrich, USA), 1 μM DTT (Roche Applied Science, Germany) and 1× GMS binding buffer (see Appendix A8) in distilled H_2O, were incubated at room temperature for 30 min. The samples were immediately separated by electrophoresis on a 4 % non-denaturing polyacrylamide electrophoresis gel

(see Appendix B9). Samples were electrophorised at 20 mA for 45 min at 15^{o}C and subjected to autoradiography at –70^{o}C.

2.4 Southern analysis of Δ*dinP3* strain

Southern analysis was used to confirm the complete knockout of the *dinP3* gene generated by homologous recombination.

2.4.1 Electro-blotting of genomic DNA

Genomic DNA was purified from potential double crossover mutants using the CTAB mycobacterial genomic DNA extraction method (see Appendix B3) and purity was analysed on a 1 % agarose gel (see Appendix B4) in relation to molecular weight marker III (Roche Applied Science, Germany). The extracted genomic DNA (1 μg) was digested, in separate reactions, with 1 U *Mlu*I (Roche Applied Science, Germany) and 1 U *Pst*I (AEC Amersham International) at 37^{o}C for 3 hrs in a 20 μL reaction with the digestion buffer consisting of 50 mM Tris-HCl, pH 7.5, 25 mM KCl, 100 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT. The restricted products were fractionated on a 1 % agarose gel (see Appendix B4) along with molecular weight marker III (Roche Applied Science, Germany). The fractioned molecular weight marker bands were recorded in relation to a ruler from a set point on the agarose gel. The gel was soaked in 0.25 M HCl for 15 min, washed with distilled H_2O, then soaked in 0.5 M NaOH/1.5 M NaCl for 15 min. The gel was washed briefly in 1× TBE buffer (see Appendix A9) and a nylon membrane (Hybond^{TM}-N) was placed on top. This was then placed between 3MM

Whatman filter paper and sponges in a TE 22 Mini transfor plastic cassette (Hoefer Scientific, USA). The fractioned DNA was transferred onto the nylon membrane using the TE 22 Mini Transfor (Hoefer Scientific, USA) at 0.5 A for 2 h in 1 × TBE buffer. The transferred DNA was cross-linked to the membrane using a UV Stratalinker 1800 (Stratagene, USA).

2.4.2 DIG (dioxigenin-11-dUTP) detection procedure

Hybridisation for Southern blot analysis was performed with the DIG- High Prime DNA Labelling and Detection System (Kit II; Roche Applied Science, Germany).

Labelling of double-stranded dinP3 DNA probe

Double stranded *dinP3* probe DNA was generated by digesting from plasmid DNA (see Appendix C3). Approximately 1 μg of probe DNA was re-suspended in 16 μL of water, heat denatured at 95^{o}C for 15 min and incubated overnight at 37^{o}C with 4 μL of supplied DIG-High Prime labelling mixture. The concentration of labelled probe was assessed by comparison to control DNA supplied with the kit. The concentration of the probe used for hybridisation was 25 ng of probe per mL of hybridisation solution.

Hybridisation of dinP3 DNA probe with target sequence

The nylon membrane with the cross-linked DNA, from section 2.4.1, was pre-incubated with DIG Easy Hybridisation solution (10 ml/100 cm^2 membrane) for 30 min. The labelled probe was denatured by boiling for 5 min and rapidly cooling in ice/water. Denatured labelled DNA probe (25 ng/mL) was added to

fresh DIG Easy Hybridisation solution (3.5 mL/100 cm^2 membrane) and pre-warmed to 42^oC. The pre-hybridisation solution was discarded and replaced with the probe-containing hybridisation solution. Hybridisation of DIG labelled probe to target DNA was allowed to occur overnight at 42^oC.

Detection of probe DNA bound to target membrane

Initially the membrane was washed with two washes of 2 × SSC; 0.1 % SDS (see Appendix A10) followed by two washes of 0.5 × SSC; 0.1 % SDS (see Appendix A11). A series of washes was then performed to develop the chemiluminescent signal. The membrane was washed in washing buffer (0.1 M Maleic acid, 0.15 M NaCl, pH 7.5, and 0.3 % Tween 20) for 5 min. The membrane was then transferred into Blocking buffer (Roche Applied science, Germany) and incubated at room temperature for 30 min, followed by 30 min in Antibody solution (Roche Applied Science, Germany). The membrane was again washed in washing buffer for 15 min, followed by equilibration in 0.1 M Tris-HCl, pH 9.5 with 0.1 M NaCl. The chemiluminescent substrate was applied and the membrane was placed in a developing cassette with 3MM medical X-ray film. Exposure was performed at room temperature for 45 min. The position of the observed bands on the X-ray film was measured by a ruler and the appropriate size determination from the molecular weight marker was then calculated.

2.5 UV Survival Assay

UV tolerance of the Y-family *dinP3* and *dinP* knockout mutants was determined using a UV sensitivity assay. Mycobacterial cultures were grown to an OD_{600} of

1.0 in 7H9 liquid media (see Appendix A2) at 37^{o}C in a New Brunswick series 25 Orbital Incubator (New Brunswick, Inc, USA) before being serially diluted (see Appendix B1). Dilutions (10^{0} to 10^{-6}) were plated in triplicate on 7H10 agar plates (see Appendix A3) and irradiated in a UV Stratalinker 1800 (Stratagene, USA) from 0-45 mJ/cm^{2} (Figure 2.1). Viable counts of the serially diluted cultures were determined by plating on 7H10 agar plates and incubating at 37^{o}C for four days.

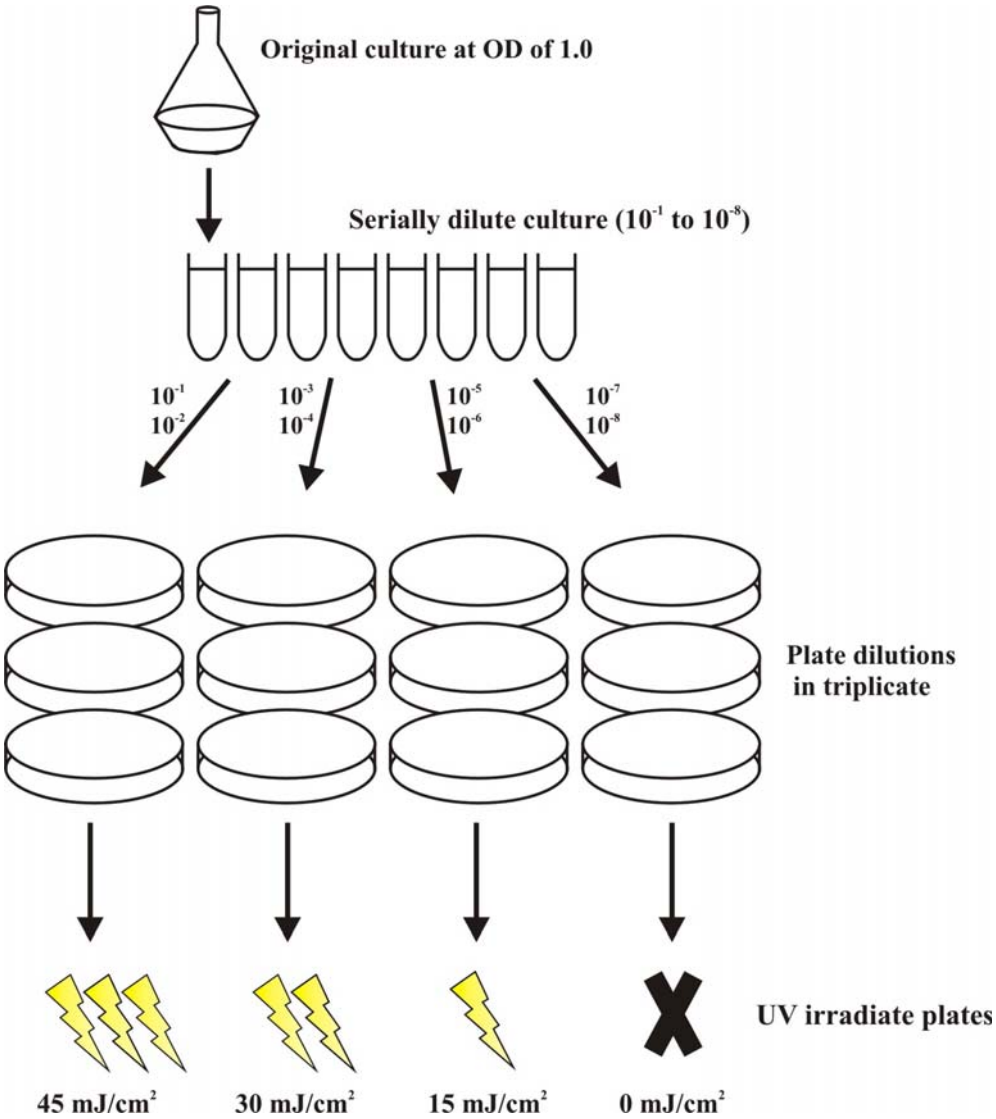


Figure 2.1: Schematic representation of UV survival assay.

2.6 Mutation Rate Analysis

The mutation rates of reversion, from histidine auxotrophy to prototrophy, of the Δ*dinP3 (hisD*^{5T}*)* and Δ*dinP (hisD*^{5T}*)* strains were determined. A modified version of the fluctuation analysis assay described by Roche and Foster (2000) was used. A fluctuation assay consists of a number of parallel cultures, inoculated with a small number of identical cells (containing no mutants) that are allowed to grow. The number of mutants in each culture that have arisen during growth is determined by plating on selective media. The total number of cells is determined by plating an aliquot of the cultures on non-selective media. The mutation rate is then determined from the distribution of the number of mutants present in the cultures.

2.6.1 Experimental method of fluctuation assay

Revertant colonies recorded in the fluctuation assay represented a –1 frameshift mutation in the *hisD* gene of *M. smegmatis*. The *hisD* gene encodes histidinol dehydrogenase and is the first gene in the operon for *de novo* histidine biosynthesis. Analysis of this gene revealed two nucleotide homopolymer tracts, a 4T tract and a 5C tract. Auxotrophs of the *hisD* gene were generated by a two-step suicide plasmid strategy to introduce site-directed +1 frameshift mutations into these homopolymer tracts (E.E. Machowski, unpublished data). Auxotrophs carrying a mutant his allele containing the 5T homopolymer tract (*hisD*^{5T}) were used in a fluctuation analysis. The mutation rates measured in the fluctuation assays represented –1 frameshift mutations in the homopolymer T tract, from 5T

(histidine auxotrophy) to 4T (histidine prototrophy)(see Appendix C4).

The overall procedure for the fluctuation assay experimental method is outlined in Figure 2.2. Freezer stocks of the *M. smegmatis* histidine auxotroph strains were grown on 7H10 plates with histidine (see Appendix A12) and 7H10 plates without histidine (see Appendix A3) to confirm histidine auxotrophy before initiation of the fluctuation assay. Genomic DNA was extracted (see Appendix B3) from a representative colony and the region of the hisD gene containing the homopolymer T tract was sequenced to confirm histidine auxotrophy (see Appendix C4). Single colonies from plates confirmed to be auxotrophs were inoculated into 5 mL 7H9 liquid medium containing histidine (see Appendix A13) and grown overnight in a New Brunswick Series 25 Orbital Incubator (New Brunswick, Inc, USA) at 37^{o}C to an OD_{600} of 1.5, to generate a pre-culture. To determine the actual CFU/mL, a serial dilution series (see Appendix B1) of the pre-culture was generated. Aliquots (100 μL) from the dilution series (10^{-4} to 10^{-8}) were plated in duplicate on 7H10 plates supplemented with histidine (see Appendix A12). The number of revertants that may have arisen during the overnight growth was also assessed. This was achieved by plating 1 mL aliquots, in duplicate, from the pre-culture on 7H 10 plates without histidine (see Appendix A3). Before the cells were plated, the 1 mL aliquots were washed with 0.5 % Tween 80 (see Appendix B10). For each single colony inoculated, two pre-cultures were grown. The first pre-culture was used directly in the fluctuation assay as described below. The second pre-culture was transferred into a petri dish and UV irradiated at 40 mJ/cm^2 in a Stratalinker (Stratagene, USA). The UV irradiated culture was then used in the fluctuation assay as further described.

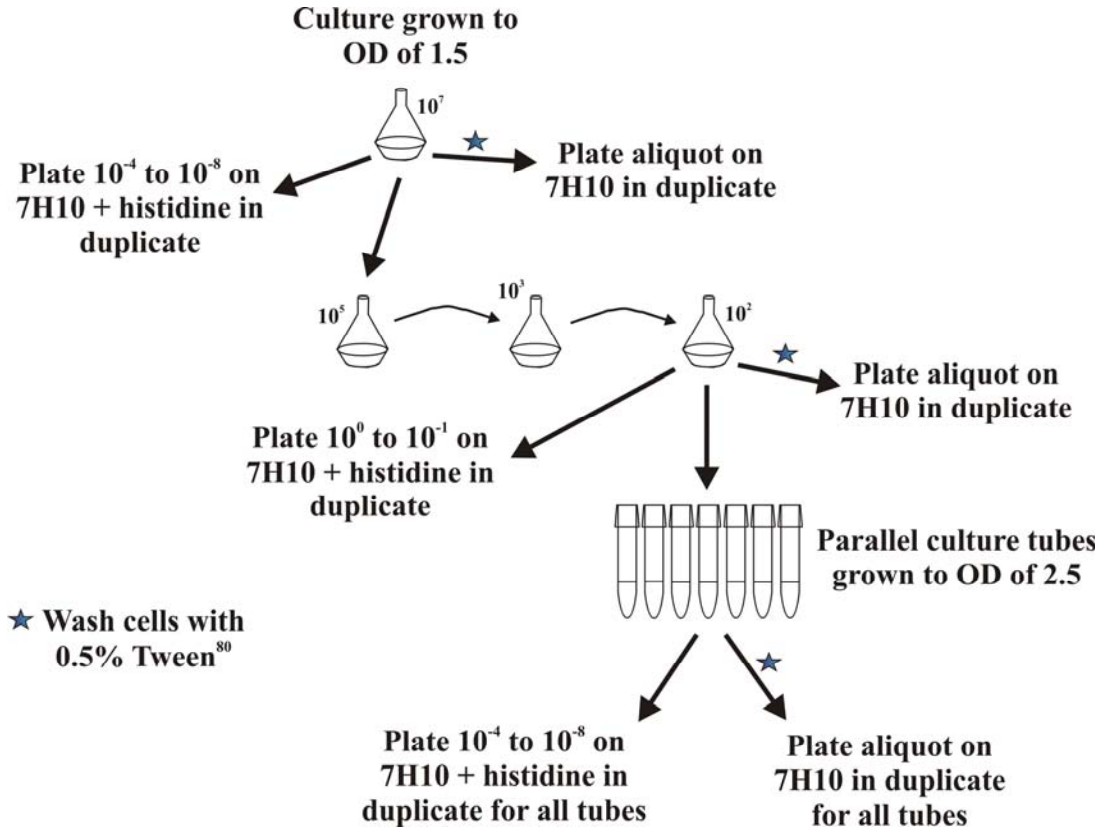


Figure 2.2: Breakdown of fluctuation assay experimental procedure used to generate histidine revertants. The mutation rate can be determined from the number of revertants generated on the 7H10 agar plates without histidine.

The fluctuation assay begins with a small number of cells being inoculated into parallel tubes. To achieve this the pre-culture, containing approximately 10^7 CFU/mL (consistent with a culture at an OD of 1.5), was serially diluted (see Appendix B11) to ~10^2 CFU/mL to generate the initial inoculum. To confirm the number of cells in the initial inoculum a dilution series (see Appendix B1) was established and a 100 μL aliquot from the initial inoculum and a 100 μL aliquot from the 10^{-1} dilution was plated, in duplicate, on 7H10 plates with histidine (see Appendix A12). The results from these plates gave an indication of the initial number of cells, N_0 (always found to be around 10^2 CFU/mL). Along with this, 1 mL aliquots from the initial inoculum were washed in 0.5 % Tween 80 (see Appendix B10) and plated, in duplicate, on 7H10 plates without histidine (see Appendix A3). This was to confirm that no revertant colonies were present in the N_0 culture. The initial inoculum was then distributed between seven parallel tubes, by dispensing 1.5 mL aliquots into each culture tube. The tubes were incubated in a New Brunswick series 25 Orbital incubator (New Brunswick, Inc, USA) at 350 rpm at 37^{o}C for three days until stationary phase (OD of 2.5) was reached. The total number of cells (N_t) in each parallel culture tube was determined by serially diluting (see Appendix B1) the culture in each tube, and then plating 100 μL aliquots (10^{-4} to 10^{-8}) onto 7H10 plates with histidine (see Appendix A12), in duplicate. The number of revertants in each parallel culture was determined by plating the entire remaining culture onto selective medium. The remaining culture was first washed with 0.5 % Tween 80 (see Appendix B10) to remove the histidine supplement and then plated onto 7H10 plates without histidine (see Appendix A3). All plates were incubated at 37^{o}C for four days before scoring for CFU's.

2.6.2 Calculation of mutation rate

Two terms are commonly used when measuring mutations, frequency and rate. The mutation rate, rather than the mutation frequency, is the most reliable measure as it records the risk of mutation per cell division. Mutation frequency on the other hand is a measure of all the mutants present in a given population and is significantly affected by a "jackpot" mutation occurring early in the culture. The definition of a mutation rate is context dependant. In molecular evolutionary biology, the mutation rate is an estimate of the generational rate of mutation per nucleotide, locus or genome in total. The mutation rate associated with the histidine reversion assay is a function of phenotypic selection. In other words, the mutation rate is defined as the probability of a cell sustaining a favourable –1 frameshift mutation at the 5T tract in *hisD* that gives rise to histidine prototrophy during its lifetime.

A fluctuation assay consists of determining the distribution of mutant numbers in parallel cultures and by analysing that distribution, the mutation rate is obtained. An important parameter used to determine the mutation rate is *m*, and it represents the mutational events occurring. Luria and Delbrück (1943) were the first to describe that an estimation of *m* and it depended on the theoretical distribution of the sizes of mutant clones. Lea and Coulson (1949) then devised a method to calculate the distribution. Their model was based on certain assumptions about the mutation process:

(i) the probability of mutation is constant per cell lifetime

(ii) the probability of mutation per cell lifetime does not vary during the growth of the culture

(iii) the proportion of mutants is always very small

(iv) the initial number of cells is negligible relative to the final number of cells

(v) the growth rates of mutants and non-mutants are the same

(vi) reverse mutations are negligible

(vii) death is negligible

(viii) all mutants are detected

(ix) no mutants arise after selection is imposed.

There are a number of methods for estimating the mutation rate for a fluctuation assay. Before the mutation rate can be determined, the probable number of mutations per culture that gave rise to the observed revertants must be calculated. This parameter is known as *m*, and represents the number of mutational events per culture. The method used to determine *m* was the MSS Maximum-Likelihood method. The MSS (Ma-Sandri-Sarkar) algorithm (see Appendix C5) is a recursive equation that efficiently computes the Luria-Delbrück distribution based on the Lea-Coulson generating function (Roche and Foster, 2000). The MSS algorithm was used to calculate the probability, P_r, of observing all the experimental values of the number of mutants, *r*, for a given *m*. The likelihood function (see Appendix C5) determines the product of the P_r's. Initially *m* is estimated by the method of the median (see Appendix C5), adjacent *m*'s are then used to recalculate the P_r's until an *m* is identified that maximizes the likelihood function. A spreadsheet was created in our laboratory that calculated all the P_r's for r = 0 to150, for a given *m*. The mutation rate was calculated by dividing *m* (the mean number of mutations

per culture) by the total number of cells in the culture, N_t. A mutation rate was determined from three independent fluctuation assay experiments under each growth condition.

The MSS algorithm accommodates departures from the Lea-Coulson assumptions that often occur under biological situations. Deviations from these assumptions that occurred in this study were that the revertants appearing represent only –1 frameshift events in the 5T homopolymeric tract, and that we are comparing UV irradiated cultures to non-UV irradiated cultures. Curve fitting can assess the deviations that occur when estimating *m*. One can plot log (*Pr*) versus log (*r*), where *Pr* is the proportion of cultures that contain *r* or more mutants. With a perfect Luria-Delbrück distribution, the plot of log (*Pr*) versus the log (*r*) approaches a straight line with a slope of –1. If the slope is close to this value, then an estimation of *m* based on the Lea-Coulson generating function is valid. To assess the data for deviations from the Luria-Delbrück distribution plots of log (*Pr*) versus log (*r*) were determined for all fluctuation assays.

To compare mutation rates among different bacterial conditions one needs to determine what confidence should be placed in the observed values. As estimates of *m* and μ -are not distributed normally, it is invalid to use normal statistics to determine confidence limits for the mutation rate (Stewart, 1994). Unfortunately, statistical methods to compare results of fluctuation assays that did not use the same *Nt*'s, or to compare mutations rates themselves, have not been developed. Relevant differences can be observed from standard deviations of multiple experiments compared to one another. The mutation rates were calculated from

three independent fluctuation assay experiments, each containing seven independent cultures tubes, for each condition studied.

3.0 RESULTS

3.1 Growth characteristics of *M. smegmatis* mc^{2}155

In order to investigate gene expression profiles and to undertake mutagenesis studies, the results that were generated need to be normalised to the growth stages of *M. smegmatis*. To characterise the different growth stages of *M. smegmatis* and relate OD units to the number of CFU, a growth curve was generated. The increase in OD as well as the number of CFU was determined at different time points (Figure 3.1). A generation time of 3 hours was observed during the logarithmic phase of growth. For all experiments conducted with *M. smegmatis*, the logarithmic phase was determined to be between an OD_{600} of 1.0 to 1.5 and stationary phase was determined to be at an OD_{600} of greater than 2.0.

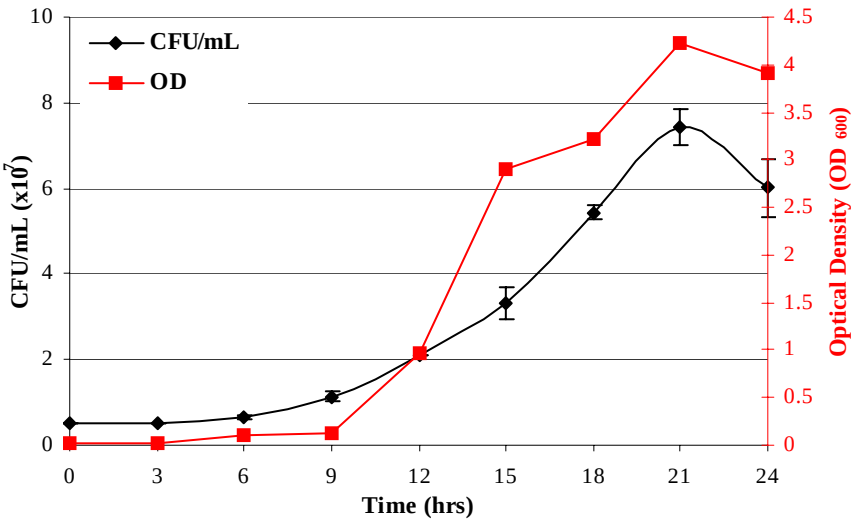


Figure 3.1: Growth characteristics of *M. smegmatis* mc^{2}155. Growth curves were taken over a period of 24 hours. Optical density reading was taken every third hour and samples were plated in triplicate to determine CFU/mL. CFU/mL, colony-forming units per millilitre.

3.2 Sequence homology of DinP3 to the Y-family polymerases

3.2.1 Classification of DinP3 by protein sequence alignment

To date, the key players in mycobacterial mutagenesis, other than DnaE2 (Boshoff *et al.*, 2003), remain to be elucidated. Mutagenesis occurs when a low-fidelity polymerase introduces a mutation into the genome, a process for which the Y-family of polymerases has been demonstrated to play an integral role (Yang, 2005). Clearly, a better understanding of the mutagenesis process can be determined by characterising the key polymerases involved. However, before any polymerase/s can be characterised, putative candidates need to be identified. The two *M. tuberculosis* genes that show homology to the Y-family error-prone polymerases, *dinP* (Rv3056) and *dinX* (Rv1537), were identified when the genome was fully sequenced and annotated (Cole *et al.*, 1998). Additionally, the genome sequence of *M. smegmatis* strain mc^2155 has recently been determined and fully annotated by The Institute for Genomic Research (TIGR) (http://www.tigr.org). Further analysis of the genome revealed conserved loci for *dinP* (MSMEG1002), found to be duplicated in the genome, and *dinX* (MSMEG3178) and revealed another homologue, *dinP3* (MSMEG6405). These genes have been classified by TIGR (http://www.tigr.org) as being part of the ImpB/MucB/SamB family of proteins, which fall under the Y-family of polymerases (Ohmori *et al.*, 2001) (see Section 1.4.2).

Classification of DinP3 as a possible DNA damage inducible protein was achieved by comparing it to all proteins represented in the TIGR Comprehensive

Microbial Resource. Alignments were generated using the program Praze (G. Sutton, unpublished data) at the TIGR comprehensive microbial resource (http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl). Praze produces an optimal gapped alignment using an implementation of the Smith-Waterman algorithm (Smith *et al.*, 1981) which compares protein sequence segments of all possible lengths, location and size and assigns a hierarchical score to the sequences, thereby maximising the optimal alignment. The first ten alignments, as ranked by Praze, are represented in Table 3.1.

The results highlight the likelihood of DinP3 being classified as a damage inducible protein by showing homology to the damage inducible protein P (known as either DinP or Pol IV). Interestingly, the alignment revealed a homologue of DinP3 in *Mycobacterium avium paratuberculosis*, showing a similarity of 88.4 % between amino acid sequences. However, the role of this protein in *M. avium paratuberculosis* remains unknown. Alignment of DinP3 to the *M. smegmatis* and *M. tuberculosis* Y-family polymerases was performed with MegAlign (DNAstar, Inc, USA) using the clustal method with the PAM250 residue weight table (Figure 3.2). This analysis shows that amino acid sequence conservation is observed amongst the *din* polymerases in *M. smegmatis* and *M. tuberculosis*, with a high degree of conservation in the N-terminus of the protein sequences (Figure 3.2). The N-terminus of polymerases contains the five highly conserved regions and is involved in nucleic acid binding and catalysis (Goodman *et al.*, 2000 and Zhou *et al.*, 2001). The similarity between the protein sequences was estimated using Praze. The *M. smegmatis* and *M. tuberculosis* DinX protein sequences show a 54 % and 53 % similarity, respectively, to DinP3. The *M. smegmatis* and

M. tuberculosis DinP protein sequences show a 48 % and 45 % similarity, respectively, in amino acid composition to DinP3.

Table 3.1: Similarity of DinP3 to TIGR comprehensive microbial resource proteins

Organism	Identification	TIGR Locus	Identity	Similarity	p-value[a]
Mycobacterium smegmatis MC2	ImpB/MucB/SamB family protein	MSMEG6405	100.0%	100.0%	1.4e-205
Mycobacterium avium paratuberculosis	Hypothetical Protein	NT03MA0182	81.8%	88.4%	2.4e-164
Mesorhizobium loti MAFF303099	DNA damage inducible protein P	NT02ML1533	43.11%	61.4%	2.4e-77
Pirellula sp strain 1	DNA damage inducible protein P	NT02PS1983	41.9%	57.0%	6.6e-66
Desulfovibrio vulgaris Hildenborough	DNA polymerase IV	DVU0071	41.4%	56.8%	4.2e-64
Rhodopseudomonas palustris CGA009	DNA damage inducible protein P	NT02RP3258	37.2%	55.9%	2.8e-58
Bacteroides thetaiotaomicron VPI-5482	DNA damage inducible	NT01BT4585	41.2%	58.5%	1.2e-57
Sinorhizobium meliloti 1021	Putative DNA damage inducible	NT01SMA1810	40.4%	58.8%	1.2e-57
Geobacter sulfurreducens PCA	ImpB/MucB/SamB family protein	GSU1616	36.5%	56.4%	2e-57
Bordetella pertussis Tohama I	DNA damage inducible protein P	NT03BP0928	41.8%	56.7%	3.2e-57
Bordetella bronchiseptica RB50	DNA damage inducible protein P	NT01B3052	41.8%	56.7%	3.2e-57

[a] p-values based on alignments between DinP3 and all the Comprehensive Microbial Resource proteins at TIGR using Praze.

```
                          ♦ ♦
MTB DinP      -----MPTAAPRWILHVDLDQFLASVELLRHPELAGLPVIVGGNGDPTEPRKVVTCASYEARAYGVRAGMPLRTAAR-RCPEATFLPSNPAAYNAASEEVVALL 98
MS DinP       ---------MTKWVLHVDLDQFLASVELRRRPDLRGQPVIVGGSGDPSEPRKVVTCASYEAREFGVHAGMPLRAAAR-RCPDATFLPSDPAAYDEASEQVMGLL 94
MTB DinX      ------------VLHLDMDAFFASVEQLTRPTLRGRPVLVGG----LGGRGVVAGASYEARAYGARSAMPMHQARRLIGVTAVVLPPRGVVYGIASRRVFDTV 87
MS DinX       MEGTVARTASRRWVLHLDMDAFFASVEQLTRPTLRGRPVLVGG----LGGRGVVAGASYEARRYGARSAMPMHQARRLVGAPAVVLPPRGAVYGVASRRALDTV 100
MS DinP3      ---MFVSAAESASILHADLDSFYASVEQRDDPALRGRPVIVGG--------GVVLAASYEAKAYGVRTAMSGGQARA-LCPQAIVVPPRMAAYTQASRDVFAVF 92
                        ♦♦
MTB DinP      RDLGYPVEVWGWDEAYLAVAP---GTPDDPIEVAEEIRKVILSQTGLSCSIGISDNKQRAKIATGLAKPAGIYQLTDANWMAIMGDRTVEALWGVGPKTTKRLA 199
MS DinP       RDLGHPLEVWGWDEAYLGADLPDESDPVE---VAERIRTVVAAETGLSCSVGISDNKQRAKVATGFAKPAGIYVLTEANWMTVMGDRPPDALWGVGPKTTKKLA 195
MTB DinX      RGLVPVVEQLSFDEAFAEPPQLAGAVAEDVETFCERLRRRVRDETGLIASVGAGSGKQIAKIASGLAKPDGIRVVRHAEEQALLSGLPVRRLWGIGPVAEEKLH 191
MS DinX       RSVVPVLEQLSFDEAFGEPSELAGAEAADVEAFCERLRAKVLEHTGLVASVGAGSGKQIAKIASGLAKPDGIRVVRREEEXVLLHGLPVRKLWGIGPVAEDRLH 204
MS DinP3      HDTTPLVEPLSVDEAFLDVSGLARVSGTPVE-IAARLRARVREQVGLPITVGIARTKFLAKVASQEGKPDGLLLVPPDRELAFLHPLPVRRLWGVGAKTAEKLR 195

MTB DinP      KLGINTVYQLAHTDSGLLMSTFG-PRTALWLLL-AKGGGDTEVSAQ-AWVPRSRSHAVTFPRDLTCRSEMESAVTELAQRTLNEVVASSRTVTRVAVTVRTATF 300
MS DinP       AMGITTVADLAVTDPSVLTTAFG-PSTGLWLLLLAKGGGDTEVSSE-PWVPRSRSHVVTFPQDLTERREMDSAVRDLALQTLAEIVEQGRIVTRVAVTVRTSTF 297
MTB DinX      RLGIETIGQLAALSDAEAANILGATIGPALHRLARGIDDRPVVER---AEAKQISAESTFAVDLTTMEQLHEAIDSIAEHAHQRLLRDGRGARTITVKLKKSDM 292
MS DinX       RIGIETIGAFAALTEAEAANVLGSTIGPALHRLARGIDDRPVAER---AEAKQISAESTFPEDLTTLAQLQDAIVPIGEHAHRRLEKDGRGARTVTVKLKKSDM 305
MS DinP3      AHGIETVADVAELSEATLGSMVG-GAMGRQLFTLSRNIDRRRVTTGVRRSSVGAQRALGRRGNSMSAAEVDAVVVNLVDRITRRMRAAGRTGRTVVLRLRFDDF 298

MTB DinP      YTRTKIRKLQAPSTDPDVITAAARHVLDLFELDRPVRLLGVRLELA*                                                          346
MS DinP       YTRTKIRKLPAPSTDAGQIVDTALAVLDQFELDRPVRLLGVRLELAMDDVPRPAVTAGT*                                            356
MTB DinX      STLTRSATMPYPTTDAGALFTVARRLLPDPLQIGPIRLLGVGFSGLSDI------------RQESLFADSDLTQETAAAHYVETPGAVVPAAHDATMWRVGDD 383
MS DinX       STLTRSATLAYATTEASTLIGTARRLLLDPVEIGPIRLVGVGFSGLSDI------------RQESLFPDLE-QPEEFPDAAPQVESVQTGPSSPTTQWRVGDD 395
MS DinP3      GRATRSHTMPWATASTDVILCAARELVAAAAPLIAERGLTLIGFAVSNI--------DRGGTQQLELPFAEQ-PDPVAIDSAIDQVRQRFGNAVLTRGVLVGRD 393

MTB DinP
MS DinP
MTB DinX      VAHP--ELGHGWVQGAGHGVVTVRFETRGSGPGSARTFPVDTGDISNASPLDSLDWPDYIGQLSVEGSAGASAPTVDDVGDR*                     463
MS DinX       VAHT--ELGHGWVQGAGHGVMTVRFETRASGPGPARTFPEDSPEISRADPVDSLDWAEYLSSLADYQSTP*                                  463
MS DinP3      PGLEMPMLPD*                                                                                              403
```

Figure 3.2: Sequence alignments of the various Y-family polymerases in *M. tuberculosis* (MTB) and *M. smegmatis* (MS) using MegAlign (DNAStar, Inc, USA). Identical amino acids amongst all sequences are shown in red. Identical amino acids between DinP3 and both *M. tuberculosis* and *M. smegmatis* DinX shown in blue. Identical amino acids between DinP3 and both *M. tuberculosis* and *M. smegmatis* DinP is shown in green. * indicates terminal amino acid in protein sequence. ♦denotes highly conserved, catalytic residues for DNA polymerase activity (Gerlach *et al.*, 1999).

3.2.2 Putative structure of DinP3

Classification of proteins can also be determined from their 3D/tertiary protein structure. The putative 3D/tertiary protein structure of DinP3 was predicted using PredictProtein (Rost, 1996) (http://cubic.bioc.columbia.edu/).The overall structure shows a similarity in the folding patterns to that observed in the DinB subfamily of polymerases, confirming the alignment observations. Amino acid alignments of the UmuC and DinB orthologues show that five conserved motifs are present in the N-terminus (Zhou *et al.*, 2001).

These five conserved motifs were identified in DinP3 and are shown in Figure 3.3A. Motifs I and III supply three invariant carboxylates in the active site (shown in red in Figure 3.3B). Motifs II and IV interact with the incoming nucleotide and motif V interacts with the primer strand (Ling *et al.*, 2001). These five motifs are distributed within the first three structural domains of the polymerase. The structural domains are termed the “palm”, “finger” and “thumb” (Yang, 2003). A unique fourth domain is found in the Y-family of polymerases and is termed the “little finger”. In the predicted structure from PredictProtein the sequence that makes up the little finger was not included but sequence similarities between DinP3 and the other Y-family polymerases indicates that the finger domain should appear the same.

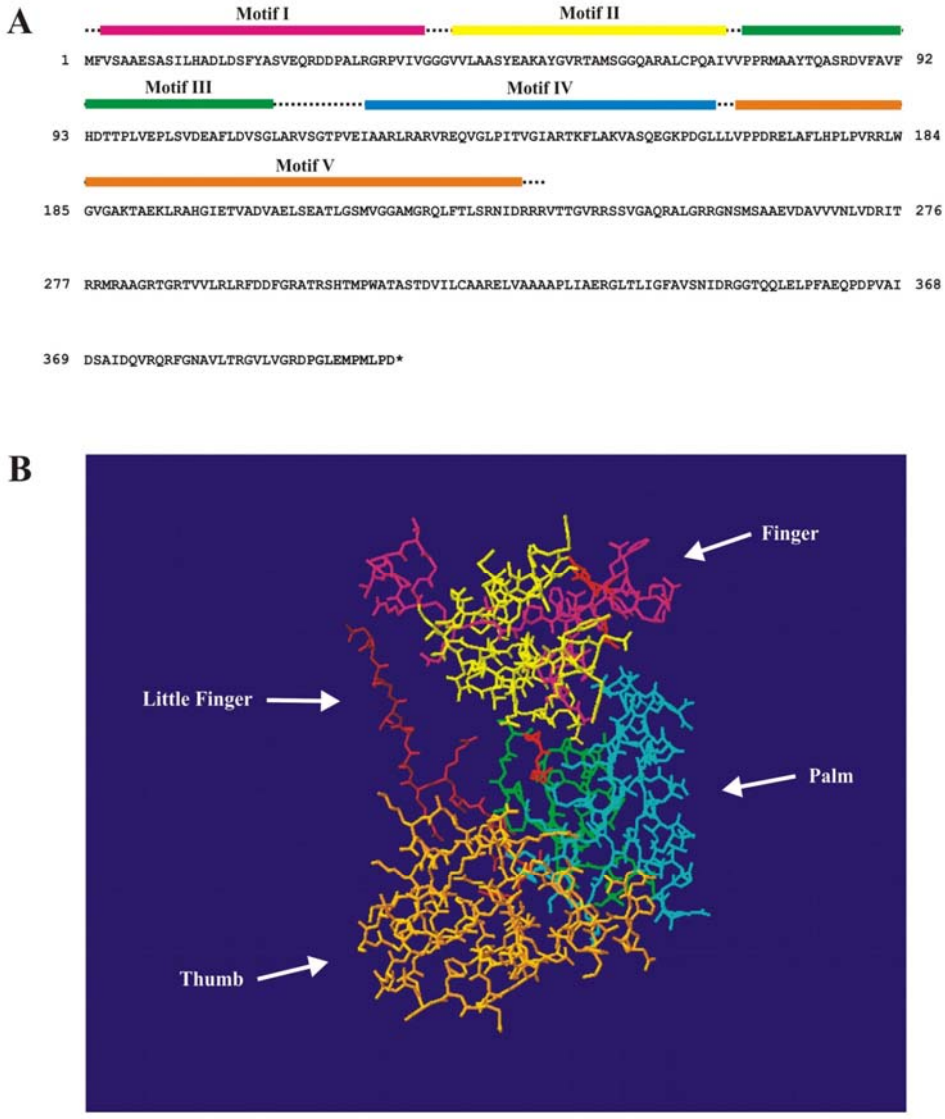


Figure 3.3: Analysis of protein sequence and predicted tertiary structure of DinP3. (A) Amino acid sequence of DinP3 showing predicted conserved motifs in the N-terminal domain. The five conserved sequence motifs are represented as follows: motif I – pink, motif II – yellow, motif III – green, motif IV – blue and motif V – orange. (B) Predicted 3D/tertiary protein structure of DinP3. The protein is coloured as described in (A). The structure was predicted using PredictProtein programme (Rost, 1996). The different structural domains have been labelled. Not all amino acids in the little finger domain are represented.

Mycobacteria appear to have Y-family polymerases that are homologous to the *E. coli* Pol IV based on the observed protein similarities. The possible function and role of these mycobacterial polymerases remains to be elucidated. Additionally, it remains intriguing that three copies of a Y-family polymerases exist with high amino acid sequence homology in *M. smegmatis* opposed to two copies in *M. tuberculosi*s. To provide clarity on this question, the response of DinP3 to stress and its role in mutagenesis were tested.

3.3 Expression analysis of *dinP3*

To assess the mutagenic ability of a polymerase, the stress conditions that induce each polymerase need to be determined. The Y-family polymerases, *dinP* and *dinX*, in *M. tuberculosis* are not up-regulated in response to DNA damage by UV, Hydrogen Peroxide and Mitomycin C (Brooks *et al.*, 2001, Boshoff *et al.*, 2003). Unpublished experiments conducted in our laboratory suggest that the *M. smegmatis* counterparts, *dinP* and *dinX*, are also not up-regulated after DNA damage by UV irradiation (D.F. Warner, unpublished data). The expression pattern of the third *M. smegmatis* Y-family polymerase, *dinP3*, was explored in response to DNA damaging agents. This was accomplished by UV irradiation of *M. smegmatis* and testing for DNA damage-mediated induction by detecting increased *dinP3* gene expression. *M. smegmatis* cultures were grown to exponential growth phase and were UV irradiated at 40 mJ/cm^2. The cultures were incubated for 90 min before total RNA was extracted. High expression levels of *dnaE2* in response to UV irradiation have been reported at 90 minutes (Boshoff *et al.*, 2003). The *dinP3* and *dnaE2* expression levels were simultaneously evaluated

by using the extracted total RNA as template to synthesise cDNA. The amount of cDNA produced is a function of the amount of transcript present in the RNA sample and was determined semi-quantitatively by PCR.

3.3.1 Evaluation of RT-PCR procedure

A PCR contains multiple variables that need to be optimised in order to obtain efficient amplification. Initially, the sensitivity of amplification of each primer set was tested on purified *M. smegmatis* genomic DNA. As the copy numbers of the genes being assessed was unknown, the primers need to detect at least 100 copies per μL. All primer sets used for RT-PCR showed efficient amplification over a range of genomic DNA standards (10^2-10^6 copies/μL)(Figure 3.4A). Evaluation of PCR conditions then followed. The *M. smegmatis* genome contains a high GC rich genome (approximately 67 %). DMSO reduces secondary structures that may form in the template DNA, especially in GC rich DNA. It was observed that optimal amplification occurred in the presence of both BSA and DMSO (Figure 3.4B). No amplification was observed when BSA and DMSO were not present.

In order to estimate mRNA levels, the RNA samples must be assessed for all contaminating DNA. The purity of the RNA extraction, in terms of contaminating DNA, was determined for each sample by including a control that lacked the AMV RT enzyme during cDNA synthesis. Samples that showed no amplification in the negative RT enzyme control but amplification in the positive RT enzyme control were used in the PCR reactions.

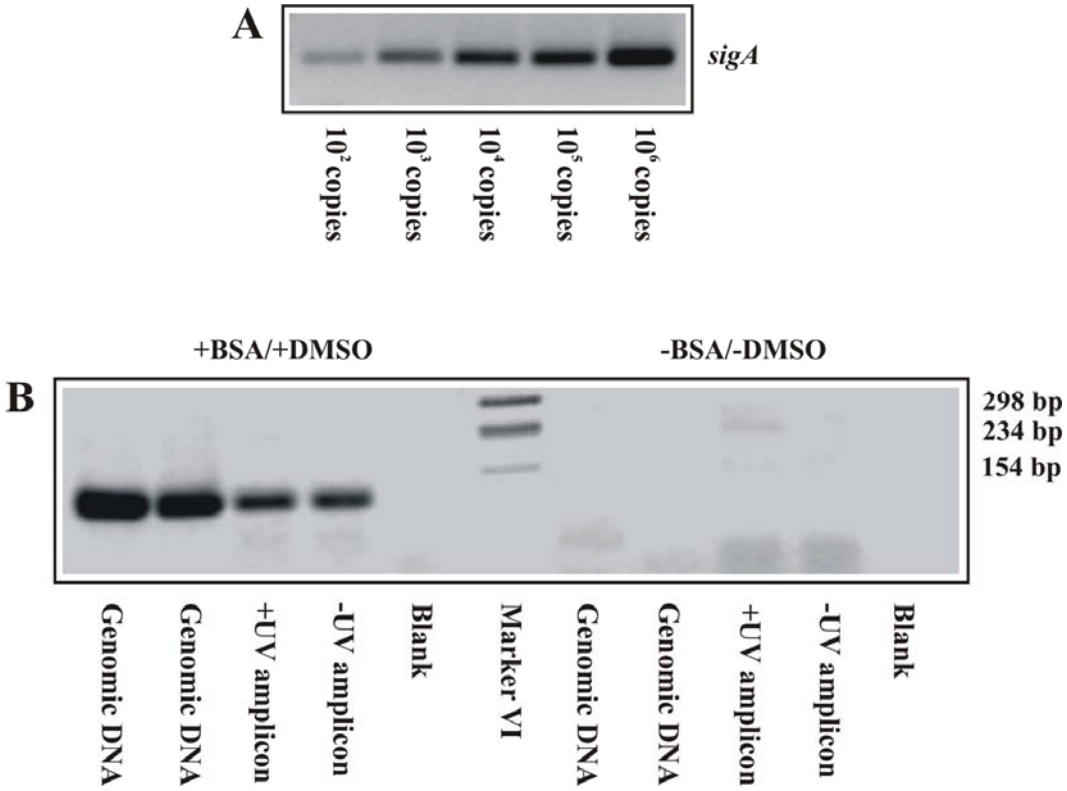


Figure 3.4: Sensitivity and specificity of primers used for all RT-PCR experiments. Agarose gels are representative for *sigA* amplification. (A) Agarose gel of *sigA* amplification on a range of genomic DNA standards (10^2-10^6). (B) Amplification of *sigA* amplicon either in the presence of both BSA and DMSO, or in the absence of both BSA and DMSO. Results were compared to 1 μg of Marker VI (Roche Applied Science, Germany), which was used as a molecular weight marker.

3.3.2 DinP3 is DNA damage inducible

Expression analysis was conducted on mRNA purified from UV irradiated and non-UV irradiated *M. smegmatis* cultures upon optimisation of the RT-PCR procedure. Up-regulation of *dinP3* was observed in samples of *M. smegmatis* that had been UV irradiated (Figure 3.5).

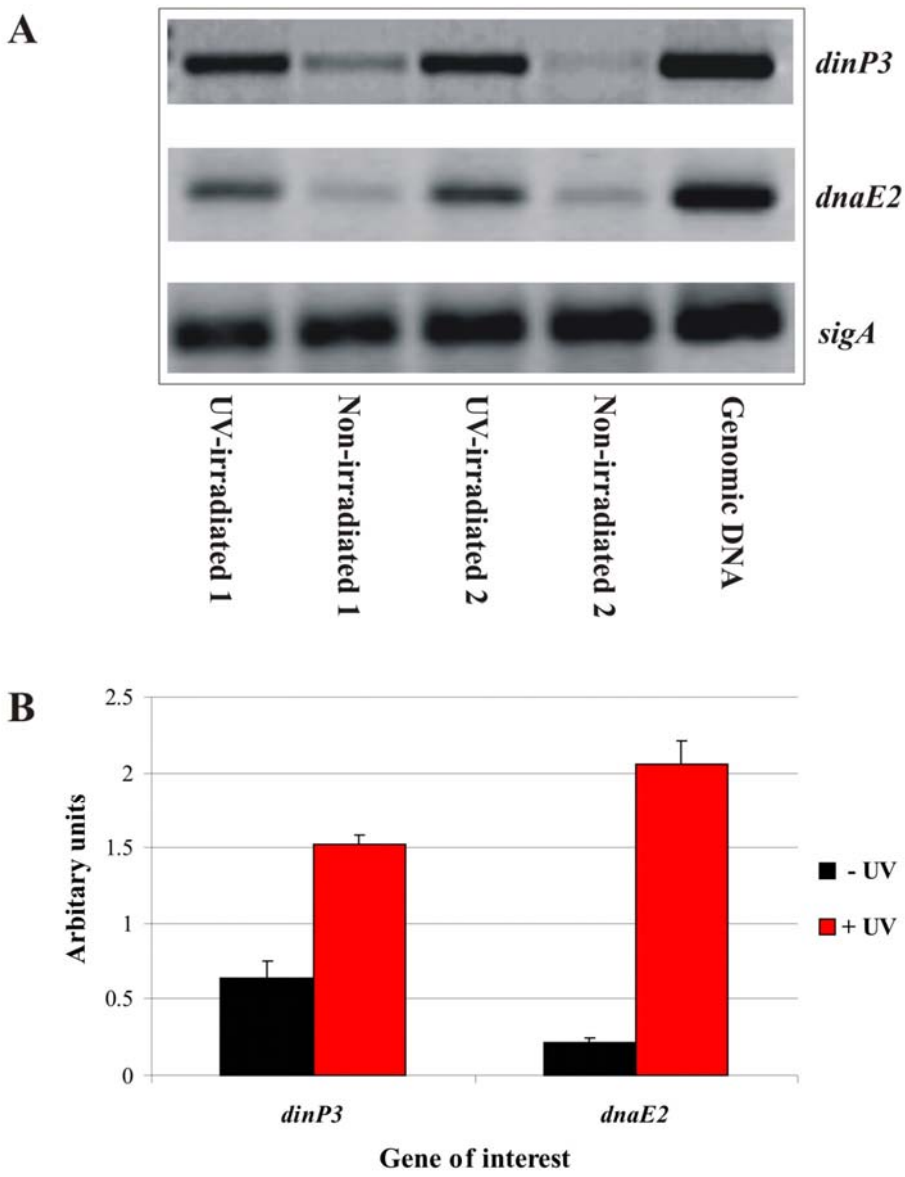


Figure 3.5: Expression of *dinP3* and *dnaE2* genes in response to UV irradiation. RT-PCR assays were performed on four independent *M. smegmatis* (mc^2155) cultures that were either subjected to UV-irradiation or grown under normal growth conditions. (A) Difference in expression of *dinP3*, *dnaE2* and *sigA* are shown in comparison to genomic DNA (10^5 copies/μL) on a 2 % agarose gel. The gel represents two of the four experiments that were used in the semi-quantification analysis. (B) The relative differences in expression due to UV irradiation are shown for *dinP3* and *dnaE2*. The values have been normalised to *sigA* expression levels. Black bars, non-irradiated and red bars, UV irradiated cultures.

Correct RT-PCR amplicons sizes were observed in all RT-PCR experiments (sizes were determined as described in section 2.2.5). Semi-quantitative RT-PCR was used to observe differences between the UV irradiated and non-UV irradiated cultures. The constant expression levels of *sigA* with and without UV irradiation were confirmed (Figure 3.5A). The *M. smegmatis dnaE2* gene, which is up-regulated after UV irradiation, served as a positive control (Figure 3.5A) (Boshoff *et al.*, 2003). The differences in expression levels were determined using the Discovery Series Quantity One 1D Analysis software (BioRad, USA). All PCR reactions included a 10^5-copies/μL genomic DNA control that was used for quantification (Figure 3.5A). The observed differences in expression were confirmed from four independent *M. smegmatis* cultures that were split and either subjected to UV irradiation or not. The actual copy number cannot be determined using semi-quantitative RT-PCR. However, relative differences between the UV irradiated and the non-UV irradiated cultures were determined.

The values determined for *dinP3* and *dnaE2* expression were normalized to the *sigA* gene, which has been shown to be expressed at a constant level under a range of growth and stress conditions in mycobacteria (Manganelli *et al.*, 1999; Boshoff *et al.*, 2004). The relative differences between the *dinP3* and *dnaE2* UV irradiated and non-irradiated cultures are shown in Figure 3.5B. The observable difference in *dinP3* expression after UV irradiation as compared to the non-irradiated cultures indicates that it may be involved in a damage inducible system in *M. smegmatis*.

The *M. smegmatis din* genes show a high degree of sequence similarity. To determine if the results observed for *dinP3* amplification in the RT-PCR are represent the induction of the *dinP3* gene, the amplicon was cut with the restriction endonuclease, *Sma*I that was uniquely located in the *dinP3* amplicon. Since the amplicon was digested with *Sma*I, this implies that the results observed from the RT-PCR represent the expression of *dinP3* gene and not any of the other *din* gene in *M. smegmatis* (Figure 3.6).

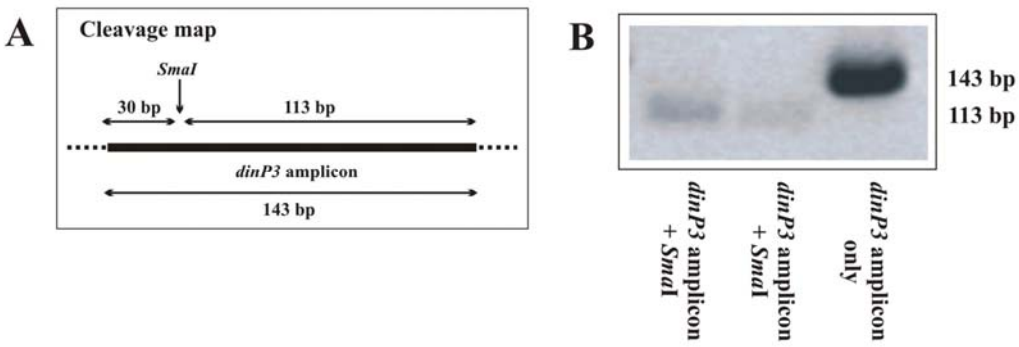


Figure 3.6: Digestion of the *dinP3* amplicon with *Sma*I. (A) Cleavage map of *dinP3* amplicon showing expected sizes after digestion with *Sma*I. (B) Digestion of the *dinP3* amplicon by *Sma*I was observed.

The expression analysis results show that the mRNA levels of *dinP3* were up-regulated after exposure to UV irradiation. In *M. tuberculosis*, damage induced mutagenesis appears to be primarily mediated by DnaE2 which is also induced in response to UV irradiation. The results of Boshoff *et al.* (2003) demonstrated that DnaE2 is the major mediator of DNA damage-induced mutagenesis in mycobacteria. However, the induction of DinP3 in response to UV irradiation

suggests that in *M. smegmatis*, this polymerase may also be involved in DNA damage induced mutagenesis. Moreover, the induction of DinP3 in response to DNA damage suggests it may be under the control of the SOS damage response system.

3.4 Identification and analysis of *dinP3* Cheo box

As DinP3 is induced in response to DNA damage by UV irradiation, it was explored whether *dinP3* expression is under control of the SOS response. LexA has been shown to be a repressor of many SOS response genes that possess the LexA binding sequence (Friedberg *et al.*, 1995). In gram-positive bacteria the LexA homologue was shown to bind a different sequence to that of the archetypal *E. coli*. This SOS response regulatory sequence is known as the Cheo box and is comprised of a palindrome 5'-GAAC$(N)_4$GTTC-3' to which the SOS repressor, LexA can bind (Cheo *et al.*, 1991).

3.4.1 Identification of Cheo box in *dinP3* promoter

The *M. smegmatis lexA* promoter contains a Cheo box that was shown to bind LexA (Durbach *et al.*, 1997; Movahedzadeh *et al.*, 1997). Analysis of the promoter region of *dinP3* led to the identification of a Cheo box, 5′-GAACatatGTTC-3′ (Figure 3.7). The sequence identified shows high conservation to Cheo boxes in gram-positive bacteria, *B. subtilis lexA* gene 5′-GAACtcatGTTC-3′, and to other Cheo boxes identified in mycobacteria, *M. smegmatis recA* gene 5′-GAACaggtGTTC-3′. The similarity between the Cheo

box sequences suggests that the binding specificities of LexA to the Cheo box may exist for DinP3.

Figure 3.7: Identification of a putative LexA binding site (Cheo box) in the promoter region of *dinP3*. The Cheo box is highlighted in red and shown in relation to the predicted translation start site of the *dinP3* gene.

3.4.2 Binding to *dinP3* Cheo box by cell lysate

Regulatory sequences in promoters bind protein regulators that control the expression of the gene. Upon identification of a Cheo box in the *dinP3* promoter a GMS assay was used to explore protein binding to this sequence (Figure 3.8). A gel electrophoretic mobility shift indicative of the formation of a protein-DNA complex was observed when the soluble fraction of a whole-cell extract of *M. smegmatis* mc^2155 was added to labelled *dinP3* probe containing the Cheo box (first panel Figure 3.8B). Competition experiments were performed to determine whether the observed protein-DNA complex was due to the sequence of the *dinP3* probe. The protein-DNA complex formation disappeared upon addition of unlabelled *dinP3* probe indicating that the observed complex is specific for the *dinP3* probe (second panel Figure 3.8B).

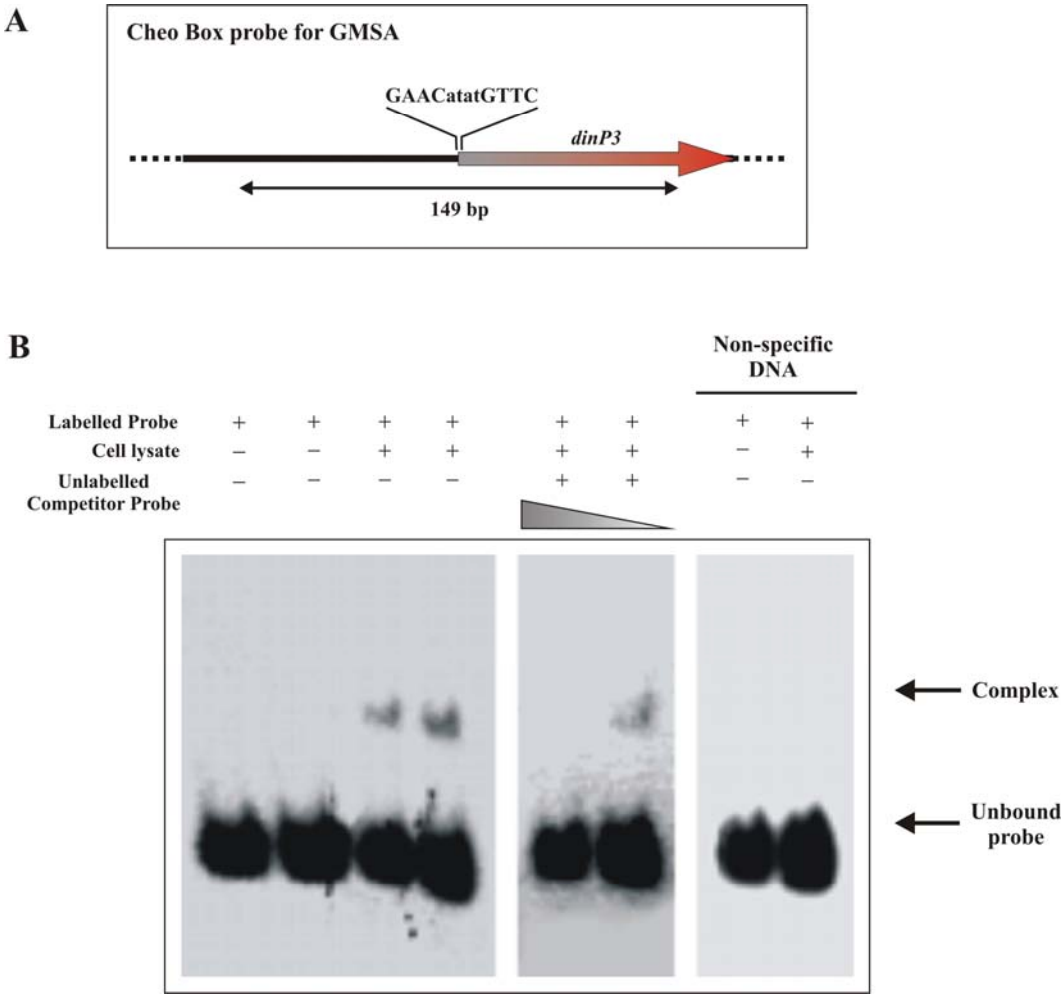


Figure 3.8: Gel mobility shift assay suggesting LexA binding to the *dinP3* Cheo box. (A) Schematic representation of probe used in GMSA. Position of the Cheo box is indicated in relation to start of *dinP3* gene. (B) Confirmation and specificity of complex formation. protein-DNA complex formation was observed in the presence of cell lysate and disappeared upon addition of unlabelled *dinP3* probe.

Partial protein-DNA complex formation was observed when labelled probe was mixed with unlabelled probe in a ratio of 2:1, respectively. Complete disappearance of the protein-DNA complex was observed when labelled probe was mixed with unlabelled probe in a ratio of 1:1. The specificity of the complex formation was confirmed when non-specific DNA was tested with the cell lysate.

The *dinP* RT-PCR amplicon was used as random DNA since this sequence does not contain any distinguishable Cheo boxes. No protein-DNA complex formed in the presence of cell lysate (third panel Figure 3.8B). Protein-DNA complex formation was only observed when a region containing the SOS regulatory sequence was incubated with cell lysate. This is highly suggestive of LexA binding to the Cheo box since binding has been demonstrated in mycobacteria (Movahedzadeh *et al.*, 1997 and Durbach *et al.*, 1997).

3.4.3 RecA-dependent induction of *dinP3*

Binding of LexA to the Cheo box is controlled by RecA. Upon activation of the SOS response, RecA causes the autodigestion of LexA and genes under LexA control are transcribed. Expression of SOS controlled genes occurs in a RecA-dependent manner. Expression of *dinP3*, in response to DNA damage by UV irradiation, was assessed for RecA-dependent induction. A mutant strain carrying an inactivated copy of RecA was used in a DNA damage gene expression experiment as described in section 2.2. The *ΔrecA* strain was previously generated by homologous recombination (E.E. Machowski, unpublished data). The absence of a functional copy of *recA* in the Δ*recA* strain was confirmed by a lack of amplification of an internal region of the RecA gene. The up-regulation of *dinP3* was found to be dependent on the presence of RecA (Figure 3.9). The dependence of *dnaE2* expression on RecA has been confirmed and was used as a positive control (Boshoff *et al.*, 2003). No difference in expression of *sigA* was observed between the UV irradiated and non-UV irradiated samples.

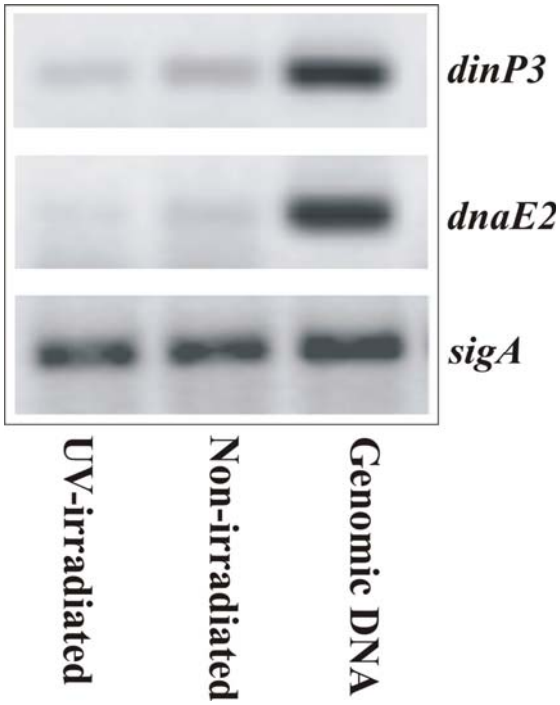


Figure 3.9: Dependence of *dinP3* expression on the presence of RecA. Cultures of mc^{2}155 Δ*recA* were either UV irradiated or grown under normal growth conditions and extracted RNA was subjected to RT-PCR as described under *Experimental Procedures*. No induction of *dinP3* or *dnaE2* was observed after DNA damaging UV irradiation.

Expression of *dinP3* in a RecA-dependent manner confirms the involvement of DinP3 in the SOS response to DNA damage in *M. smegmatis*. The fact that DinP3 is under control of RecA corroborates earlier findings that showed possible involvement of LexA. These findings strongly suggest a role for DinP3 in the damage inducible SOS response system in *M. smegmatis*. Moreover, this is the first example of a Y-family polymerase from mycobacteria that forms part of the SOS regulon. Therefore, although neither *dinP* nor *dinX* is damage-inducible in *M. smegmatis* or *M. tuberculosis*, the latter organism is distinguished by the fact that it possesses a third Y-family polymerase-encoding gene that is damage inducible in a *lexA/recA*-regulated manner. The identification of SOS response

control of DinP3 suggests that it might play a role in DNA damaged mutagenesis.

3.5 Genotype confirmation of mutant strains

To perform characterisation studies into the DNA damage induced mutagenesis of DinP3, a mutant strain containing a non-functional copy of the *dinP3* gene was needed. Assessment of both the DinP3 and DinP roles in mutagenesis were chosen in light of their differing DNA damage-induction responses. Confirmation of the absence of *dinP3* and *dinP* in their relevant mutant strains was determined prior to examining their role in mutagenesis.

3.5.1 Southern blot analysis of Δ*dinP3* strain

Single crossover mutants of *dinP3* were created by homologous recombination (E.E. Machowski, unpublished data) using an allelic exchange technique (Parish and Stoker *et al.*, 2000). The restriction sites used to confirm double crossover complete knockout mutants of *dinP3* are shown along with the Southern blot analysis in Figure 3.10. Double crossover mutants of *dinP3* were distinguished from wild-type mc^2155 by differing fragment lengths on the Southern blot, fragment lengths were determined as described in section 2.4.2. Complete knockout of the *dinP3* gene resulted in the loss of a *Mlu*I site, which is present in the *dinP3* gene, resulting in a fragment of approximately 5500 bp in length compared to 2832 bp for the wild-type (Figure 3.10B). Assessing restriction of the *dinP3* gene by *Pst*I again confirmed complete *dinP3* knockout. An approximately 1108 bp difference was observed on the Southern blot, confirming deletion of the

dinP3 gene (Figure 3.10C). A mutant strain containing an inactivated copy of DinP3 was confirmed, allowing characterisation of DinP3 by comparing mutational responses between mc^{2}155 and Δ*dinP3*.

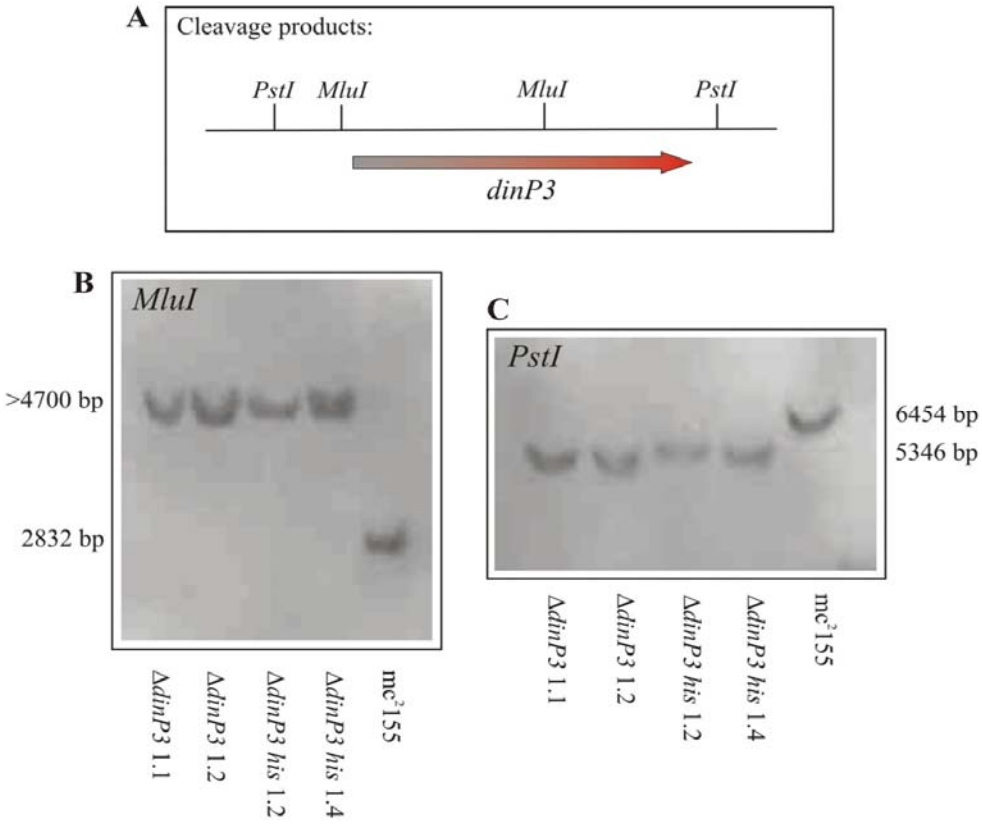


Figure 3.10: Confirmation of Δ*dinP3* and Δ*dinP3* (*his*^{5T}) strains in *M. smegmatis*. (A) Cleavage map showing the relevant positions of the restriction endonuclease sites in relation to the *dinP3* gene. (B) Southern blot analysis of the Δ*dinP3* mutant strains in comparison to mc^{2}155 after restriction with *Mlu*I. (C) Southern blot analysis of the same Δ*dinP3* mutant strains in comparison to mc^{2}155 after restriction with *Pst*I.

3.5.2 Confirmation of the Δ*dinP* strain

DinP was included in further studies to provide a comparison between a UV induced Y-family polymerase, DinP3, and non-UV induced Y-family polymerase. The Δ*dinP* strain (knockout of both *dinP* genes in the *M. smegmatis* genome) was previously generated by homologous recombination (D.F. Warner, unpublished data) and was confirmed before use to be a *dinP* complete knockout. The Δ*dinP* strain is hygromycin resistant (Figure 3.11A). An alternative procedure was used to determine complete knockout of the *dinP* by amplification of an internal region of the gene. Absence of *dinP* amplification was observed in the Δ*dinP* strain after PCR (Figure 3.11B).

3.6 UV survival analysis

The SOS response is induced to help the bacterium tolerate DNA damage. As a result, the involvement of DinP3 in tolerating UV irradiation was investigated. Strains carrying inactivated copies of *dinP3* and *dinP* were examined for sensitivity to UV irradiation. To test the effect that absence of Y-family polymerases have on the UV tolerance of *M. smegmatis*, mutant strains were UV irradiated and viable colony forming units were counted. Direct exposure of wild-type *M. smegmatis* mc^2155 to UV light was determined to be lethal (Figure 3.12; mc^2155); at UV fluences of 40 mJ/cm^2, approximately 0.01% of the bacteria survived. It has recently been shown that by removing DnaE2 from *M. smegmatis*, tolerance of DNA damage by UV irradiation was reduced to approximately 0.0001% of wild-type at 40 mJ/cm^2 (Boshoff *et al.*, 2003).

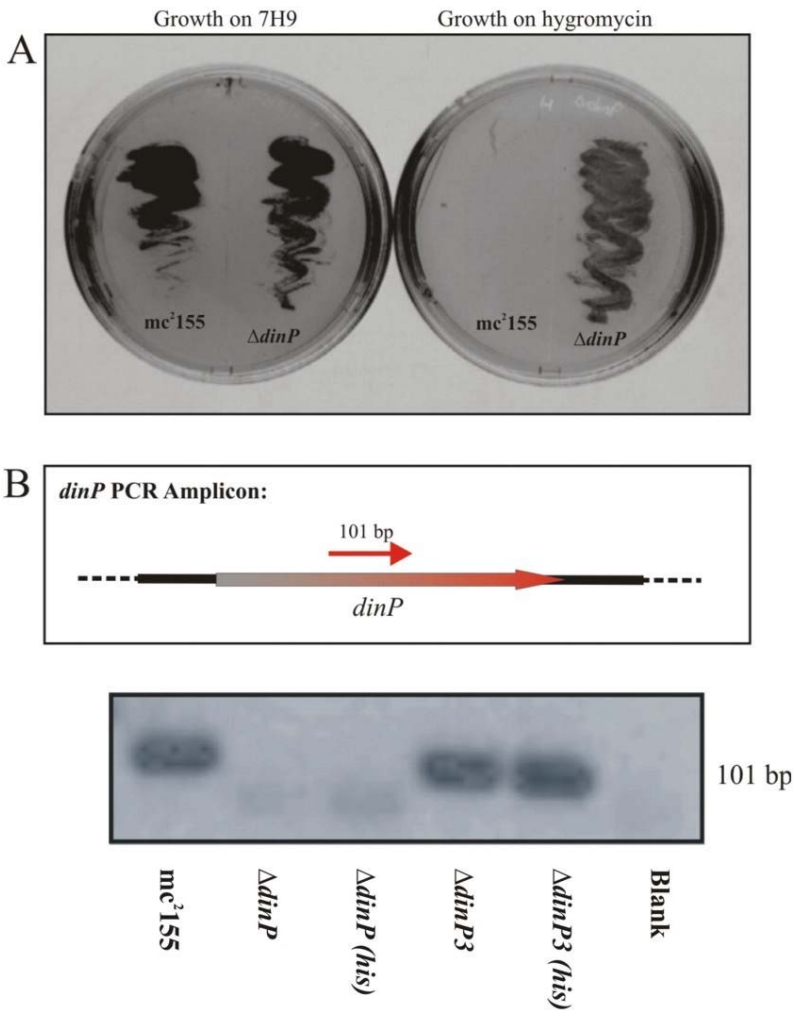


Figure 3.11: Genotypic confirmation of *dinP* mutant of *M. smegmatis* mc^{2}155. (A) Growth of wild-type mc^{2}155 and Δ*dinP* on 7H9 plates and 7H9 with hygromycin plates. (B) Amplification products after PCR of an internal segment of the *dinP* gene in *ΔdinP* and *ΔdinP (his^{5T})* strains. Upper panel shows expected *dinP* amplicon size. Lower panel shows observed PCR products on 2 % agarose gel.

The UV tolerance of the Δ*dinP3* strain was shown not to vary significantly from that of wild-type mc^{2}155, suggesting that Δ*dinP3* does not have a greater sensitivity to the cytotoxic effects of UV induced DNA damage (Figure 3.12; Δ*dinP3*). No overall UV tolerance for Δ*dinP* was observed (Figure 3.12; Δ*dinP*).

It appears that Δ*dinP* was no more sensitive to UV irradiation than mc^{2}155 and Δ*dinP3*. The absence of RecA greatly diminishes the ability of *M. smegmatis* to tolerate UV irradiation. The Δ*recA* mutant barely survived at UV fluences above 10 mJ/cm^2 (Figure 3.12; Δ*recA*). The difference in the UV tolerance between wild-type mc^{2}155 and Δ*recA* strains clearly indicates that *M. smegmatis* tolerates UV irradiation by induction of the SOS response.

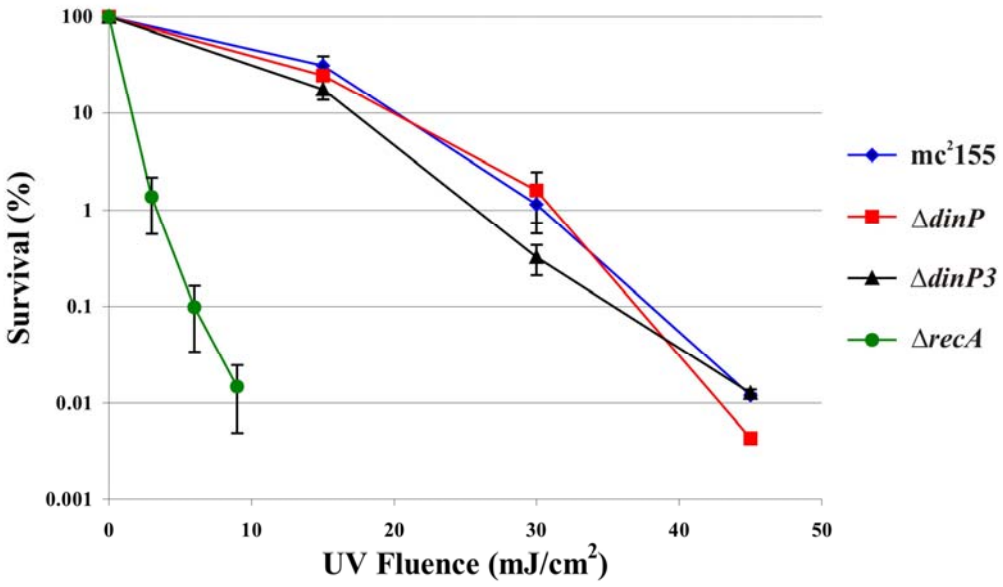


Figure 3.12: Effect of UV irradiation on survival of *M. smegmatis.* Dilutions of *M. smegmatis* were plated in triplicate and UV irradiated at different UV fluences. No overall difference in UV tolerance was observed for mc^{2}155, *ΔdinP* and Δ*dinP3.* Diamonds: mc^{2}155 (wild type); squares: Δ*dinP*; triangles: Δ*dinP3*; and circles: Δ*recA*. Results represent two independent experiments.

The lack of effect of DinP3 loss on UV sensitivity of *M. smegmatis* contrasts with the damage hypersensitivity associated with loss of DnaE2 (Boshoff *et al.*, 2003). Like DnaE2, DinP3 is induced by UV irradiation in a *recA/lexA*-dependent manner, but unlike DnaE2, DinP3 does not contribute tolerance of damage to DNA, such as TT cyclobutane dimers or 6-4 photoproducts, induced by UV irradiation. Therefore, although both polymerases are induced as part of the SOS response, DinP3 may be involved in tolerance of a different type of DNA damage.

3.7 Mutation rate analysis

To understand how DinP3 functions in response to UV irradiation, its ability to introduce mutations needed to be determined. The involvement of DinP3 in a damage-inducible mutagenesis system in response to UV irradiation was therefore explored. Mutation rates in strains lacking either DinP3 or DinP were assessed in parallel in order to compare the contributions of the damage-inducible *dinP3* and the *dinP* genes to UV-induced mutagenesis. The UV-induced mutagenesis was examined by measuring –1 frameshift mutations in the *hisD* gene that resulted in reversion to histidine prototrophy in *M. smegmatis*.

Mutation rates in *M. smegmatis* were measured using a fluctuation assay. The *M. smegmatis* strains were either grown under normal growth conditions or UV irradiated and then plated on selective media to determine the number of histidine revertants. Often under biologically relevant circumstances models describing the distribution of mutant numbers do not meet the Lea-Coulson assumptions (see section 2.6.2) (Rosche and Foster, 2000). To describe if any deviations affecting

the mutation rate calculations were present, log (*Pr*) was plotted as a function of log (*r*). Curves were plotted for all fluctuation assays and the slope was calculated. No distortions in the slope of the curves were observed in any of the fluctuation assays (Figure 3.13). The MSS maximum likelihood method was used to calculate mutation rates (see section 2.6.2).

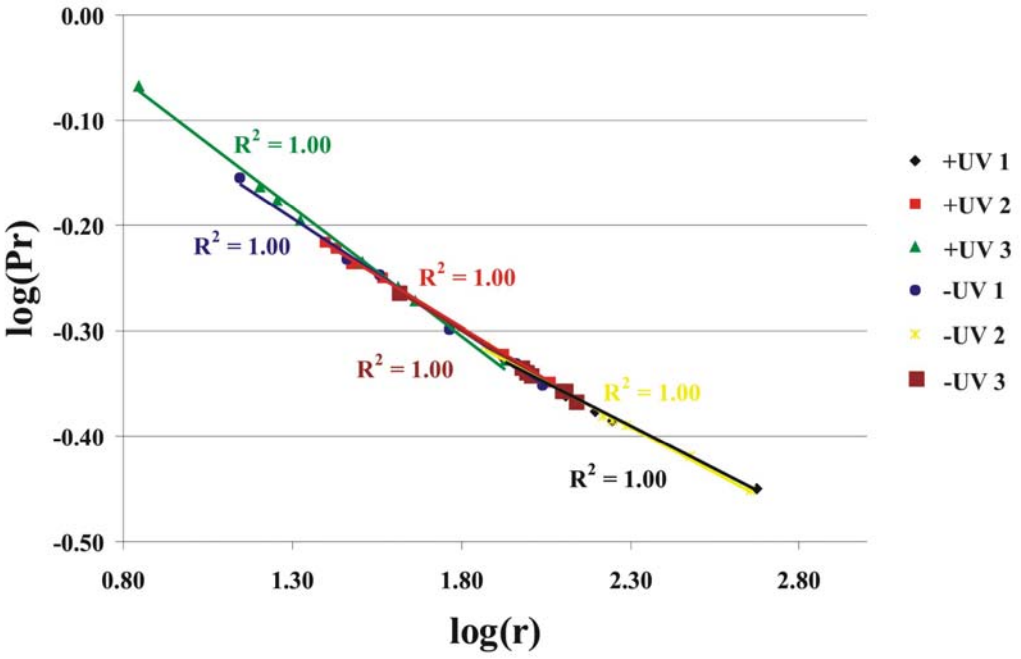


Figure 3.13: Curve showing log (*Pr*) versus log (*r*) for the mc^{2}155 mutation rate analysis. Curves are representative of both the UV irradiated and non-UV irradiated mc^{2}155 fluctuation assays. Curves were plotted for the Δ*dinP3 (hisD^{5T})* and Δ*dinP (hisD^{5T})* fluctuation assays, and show similar findings.

Unexpectedly, no difference in the mutation rate of *M. smegmatis* mc^{2}155 was observed before and after UV irradiation (Table 3.2; Figure 3.14). Previously, studies investigating mutagenesis induced by UV irradiation in mycobacteria and other organisms have shown an increase in the number of mutants following UV irradiation (Radman, 1975; Sedgwick, 1975; Wagner *et al.*, 2002; Boshoff *et al.*, 2003 and Duigou *et al.*, 2004). However, these studies were carried out using different reporter strains measuring base substitutions that conferred antibiotic resistance to rifampicin or tetracycline. The lack of effect of UV irradiation on the rate of frameshift mutagenesis using the *hisD*^{5T} reporter may be a function of the mutational event being monitored. The histidine reporter only detects –1 frameshifts in a homopolymeric (5T) tract. The results observed for mc^{2}155 suggest that *M. smegmatis* tolerates UV induced DNA lesions and replicates past the lesion without introducing –1 frameshift mutations. The ability of mycobacteria to protect against frameshift hypermutability at homopolymeric runs is noteworthy, since these organisms do not possess homologues of the highly conserved *mutS*-based MMR system (Mizrahi and Andersen, 1998; Mizrahi *et al.*, 2000). Introduction of –1 frameshift errors could be deleterious to the organism, and thus very strict control measures are needed. Boshoff's (2003) assessment of damage induced mutagenesis by DnaE2 was limited to base substitutions, one triplet deletion and a hexanucleotide insertion. The observed results appear to rule out the ability of DnaE2 in the presence of the Y-family polymerases to introduce a –1 frameshift in the *hisD*^{5T} reporter strain after UV irradiation in *M. smegmatis*.

Table 3.2: Mutation rates of *M. smegmatis* strains before and after UV irradiation

Bacterial strain	Treatment	No. independent cultures	No. cells per culture	Mutation rate per cell division (μ)[a]
mc^{2}155	Untreated	21	8.2×10^8	$(1.70 \pm 1.11) \times 10^{-8}$
	UV	21	1.0×10^8	$(1.80 \pm 1.81) \times 10^{-8}$
Δ*dinP3*	Untreated	21	5.1×10^8	$(26.3 \pm 1.46) \times 10^{-8}$
	UV	21	4.8×10^8	$(3.10 \pm 1.18) \times 10^{-8}$
Δ*dinP*	Untreated	21	9.9×10^8	$(2.60 \pm 0.83) \times 10^{-8}$
	UV	21	3.2×10^8	$(14.0 \pm 0.38) \times 10^{-8}$

[a] Breakdown of values used to calculate mutation rate is reported in Appendix C6.

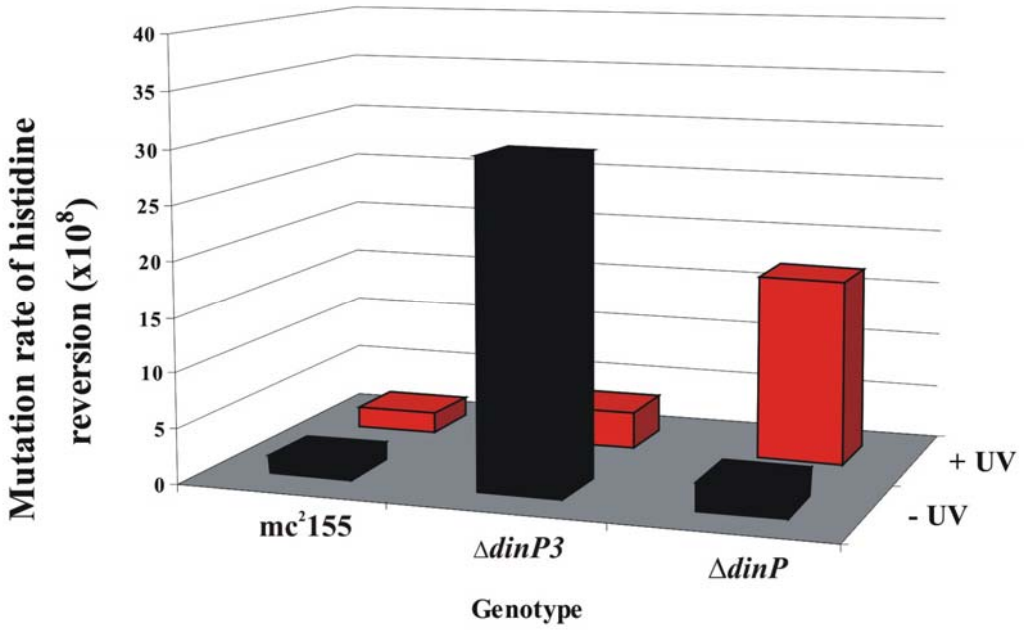


Figure 3.14: Effect of DinP3 and DinP absence on mutagenesis. The mutation rate was determined using fluctuation analysis of cultures either untreated or exposed to UV irradiation (40 mJ/cm^2). Values represent three independent fluctuation assays for each strain investigated.

Interestingly the rates of mutation to histidine prototrophy, with or without UV irradiation, changed considerably when either DinP3 or DinP was inactivated (Table3.2; Figure 3.14). When DinP3 is removed, an increased number of histidine revertants were observed under normal growth conditions (Figure 3.15). The difference in the mutation rates was approximately 8-fold, with the mutation being 26.3×10^{-8} under normal growth conditions and 3.6×10^{-8} following UV irradiation. However, as shown in Table 3.2 and Figure 3.14, loss of DinP had the opposite effect on the rate of reversion to histidine prototrophy.

The number of histidine revertants was greater following UV irradiation (Figure 3.15). The mutation rate increases from 2.6×10^{-8} under normal growth conditions to 14×10^{-8} following UV irradiation. This is an approximately 5.5-fold difference in the mutation rates. In the absence of DinP3, the high mutation rate under growth conditions suggests that the ability of *M. smegmatis* to generate spontaneous –1 frameshifts is increased. However, in the absence of DinP, the increased mutation rate following UV irradiation suggests that when one member of the Y-family of polymerases is missing, the ability of *M. smegmatis* to introduce –1 frameshifts is increased. The increase in mutation rate associated with loss of a Y-family polymerase is noteworthy, since mutational studies characterising mutator polymerases in other organisms have shown a decrease in the introduction of mutations when the investigated polymerase is removed (Napolitano *et al*., 2000; Pol II, Pol IV and Pol V in E. coli, Wagner *et al*., 2002; Pol IV in E. coli, Boshoff *et al.*, 2003; DnaE2 in *M. tuberculosis* and Duigou *et al.*, 2004; YgjW in *B. subtilis*).

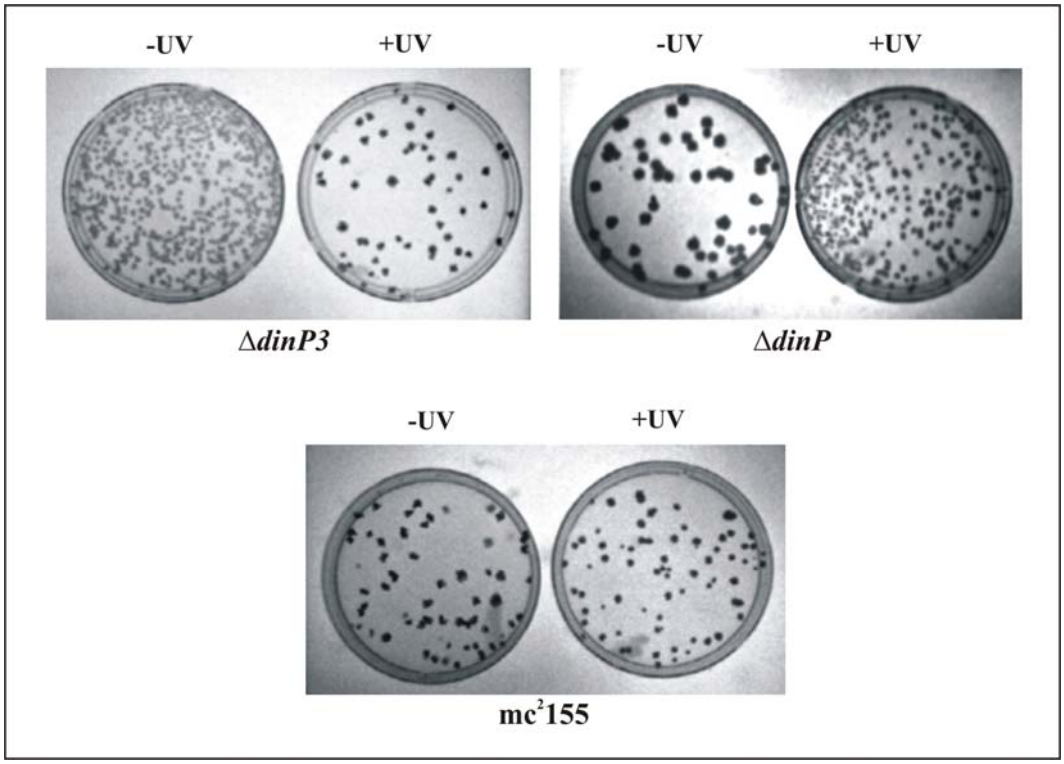


Figure 3.15: Differences in numbers of histidine revertants observed for the various strains in the fluctuation analysis. Plates represent the number of histidine revertants per 500 μL on 7H10 agar plates.

The Y-family polymerases in *M. smegmatis* appear to co-ordinate the prevention of -1 frameshift mutations. Removal of one of these members allows error-prone replication to occur. Characterisation of the mutagenic spectra of the Y-family polymerases opens an interesting avenue of investigation. The mycobacterial Y-family polymerases appear to exhibit distinctive features from the *E. coli* and *B. subtilis* Y-family polymerases, and may be due to the existence of a DNA damage inducible mutagenesis system mediated by DnaE2. Despite the overall protein sequence similarities between the Y-family polymerases, it appears that different bacteria have adapted different solutions to deal with lesions in their genetic material.

4.0 DISCUSSION AND CONCLUSIONS

The ability of the Y-family polymerases to synthesize DNA with a lower fidelity than that of the replicative polymerases requires that they be tightly regulated. Understanding how Y-family polymerases function should help in understanding the mechanisms of mutagenesis. However, mycobacteria may induce genetic variation differently from other bacteria through the action of the C-family polymerase, DnaE2. Nevertheless, mycobacteria provide an ideal system for investigating the individual and collective roles of these Y-family polymerases under circumstances in which their principal function may be performed by an alternative error-prone polymerase. The studies presented in this dissertation add to the growing body of information on Y-family polymerases in mycobacteria, and their role in the response of these organisms to DNA damage.

4.1 Biochemical characterisation of DinP3

Results presented in this study demonstrate the presence of an additional Y-family polymerase in *M. smegmatis*, which belongs to the DinB subfamily of the Y-family polymerases. To better understand the role of the *dinP3* gene in *M. smegmatis*, UV-mediated DNA damage was explored as a possible regulator of its expression. DinP3 was shown to be up-regulated in a RecA-dependent manner in response to DNA damage by UV irradiation (see section 3.4.3), confirming that its encoding gene forms part of the SOS-inducible response in mycobacteria. This finding is significant, since DinP3 is the first Y-family polymerase shown to be under the control of the SOS response in mycobacteria. Induction of DinP3 as part

of the SOS response is similar to observations found in other bacteria including *E. coli*, in which Pol II, IV and V are induced, and *B. subtilis*, in which YgjW is induced as part of the SOS response (Napolitano *et al*., 2000 and Duigou *et al.*, 2004). A difference between *M. smegmatis* and *M. tuberculosis* is observed since *M. smegmatis* contains a Y-family polymerase, DinP3, which is induced as part of the SOS response. The SOS response is induced in response to many different DNA-damaging agents, suggesting that *M. smegmatis* may utilise more than one specialised polymerase to bypass lesions caused by different DNA-damaging agents.

The possible role of DinP3 in mutagenesis was assessed by analysing the effect of *dinP3* gene loss on the rates of mutagenesis of normal vs. UV-treated cultures of *M. smegmatis*. Although DinP3 was shown to be induced by the SOS response, it does not contribute to UV tolerance in *M. smegmatis*. This could suggest that DinP3 does not play a direct role in bypassing lesions that are introduced into the DNA by UV irradiation, such as thymine dimers. DinP3 may bypass other types of lesions that also cause the induction of the SOS response. Since the SOS response is a general DNA damage response mechanism, a more specific understanding on how DinP3 functions in bypassing synthesis remains to be determined.

The role of DinP3 in mutagenesis as a result of UV induced DNA damage was explored. The mutagenesis studies lead to very interesting observations. It appears that in mycobacteria the role of the Y-family polymerases is not the same as that observed in other bacterial species. Upon removal one of the Y-family

polymerases, *M. smegmatis* acquired an increased ability to introduce –1 frameshift mutations (see section 3.7). This suggests a possible role of the Y-family polymerases in *M. smegmatis* in preventing mutations from occurring and working in an "anti-mutator" fashion. Observations from the mutagenesis studies do not clearly highlight whether DinP3 acts as an error-prone polymerase or as an "anti-mutator". As DinP3 and DinP both show protein sequence (see section 3.2) and folding similarities (section 4.3.2) to error-prone polymerases, it is not surprising that they should act as error-prone polymerases. However, the observed increase in –1 frameshift mutations in the mutagenesis studies for DinP3 and DinP do not correlate with previous studies of inactivated error-prone polymerases. When DinP3 is inactivated, increased mutagenesis is observed under normal growth conditions. However, since DinP3 is part of the SOS regulon, an increase in mutagenic events would be expected after UV irradiation. Interestingly, the opposite occurs for DinP. A putative model of how DinP3 and DinP may function in *M. smegmatis* is discussed in the next section.

4.2 Putative roles of DinP3 and DinP in mutagenesis

It is intriguing that multiple DNA polymerases involved in translesion synthesis are present in a single organism. Even though the actual process of translesion synthesis catalysed by DNA polymerases appears to be similar between different organisms, mechanisms of regulation and the type of lesions they can bypass vary greatly. It is becoming more apparent that translesion synthesis is achieved by either a specific DNA polymerase or a combination thereof, depending on the type of lesion and sequence context of the DNA. It is therefore unlikely that a

generalised model can explain how translesion polymerases function and when they are needed. However, models that are specific to a particular organism could be designed to explain how DNA polymerases interact with one another to bypass DNA lesions.

4.2.1 Proposed model for frameshift mutagenesis in *M. smegmatis*

In order to put the findings of this study in context, I outline a model of how the investigated polymerases may modulate their activity in *M. smegmatis* in Figure 4.1. This model proposes that both DinP3 and DinP act as error-prone polymerases (mutators) but that their mutagenic ability is masked in wild-type *M. smegmatis*. Exactly how this occurs remains unresolved. In mycobacteria it has been shown that DnaE2 is the major mediator of damage-induced base substitution mutagenesis (Boshoff *et al.*, 2003). However, the lack of UV-induced –1 frameshift mutations in wild-type *M. smegmatis* suggests that DnaE2 may be unable to produce targeted frameshift mutations in the *hisD*^{5T} reporter, as no difference in the mutation rate was observed in wild-type *M. smegmatis* before and after UV irradiation (see Figure 3.14 in section 3.7). The model proposed that in wild-type *M. smegmatis*, after UV irradiation, DnaE2 is up regulated as part of the SOS response and out-competes the other error-prone polymerases for access to the damaged DNA but does not introduce –1 frameshift mutations. By contrast, when either DinP3 or DinP are removed, the rates of –1 frameshift mutagenesis under normal growth conditions and following UV irradiation are altered. In the absence of DinP3, the spontaneous mutation rate was shown to increase (see Figure 3.14 in section 3.7) suggesting that DinP may gain access to the DNA.

Assuming that DinP is error-prone, -1 frameshift mutagenesis may be facilitated under normal growth conditions (see Figure 3.14 in section 3.7). However, since an increased mutation rate is not observed after UV irradiation, it is proposed that DnaE2, induced as part of the SOS response, gains access to the DNA instead of DinP. Based on the assumption that DnaE2 is unable to produce targeted frameshift mutations, preventing access of DinP to the replication fork by DnaE2 would moderate frameshift mutagenesis.

When DinP was removed in this study, an increase in –1 frameshift mutations was observed after UV irradiation (see Figure 3.14 in section 3.7). This suggests that DinP3 gains access to the replication fork after UV irradiation and is able to introduce mutations. This agrees with the observations that DinP3 is induced as part of the DNA damage mediated SOS response. DinP3 may be able to out-compete DnaE2 at the replication fork for access to the damaged DNA when DinP is absent. This model pre-supposes that polymerase switching occurs at the replication fork. It is not known yet whether this process is tightly co-ordinated or simply involves competition among the different DNA polymerases based on the ability to gain access to the replication fork. The contribution of DinX is not discussed as its involvement in frameshift mutagenesis was not established. However, the role of DinX in mutagenesis cannot be discounted and further investigations into how DinX interacts with the other Y family polymerases needs to be established.

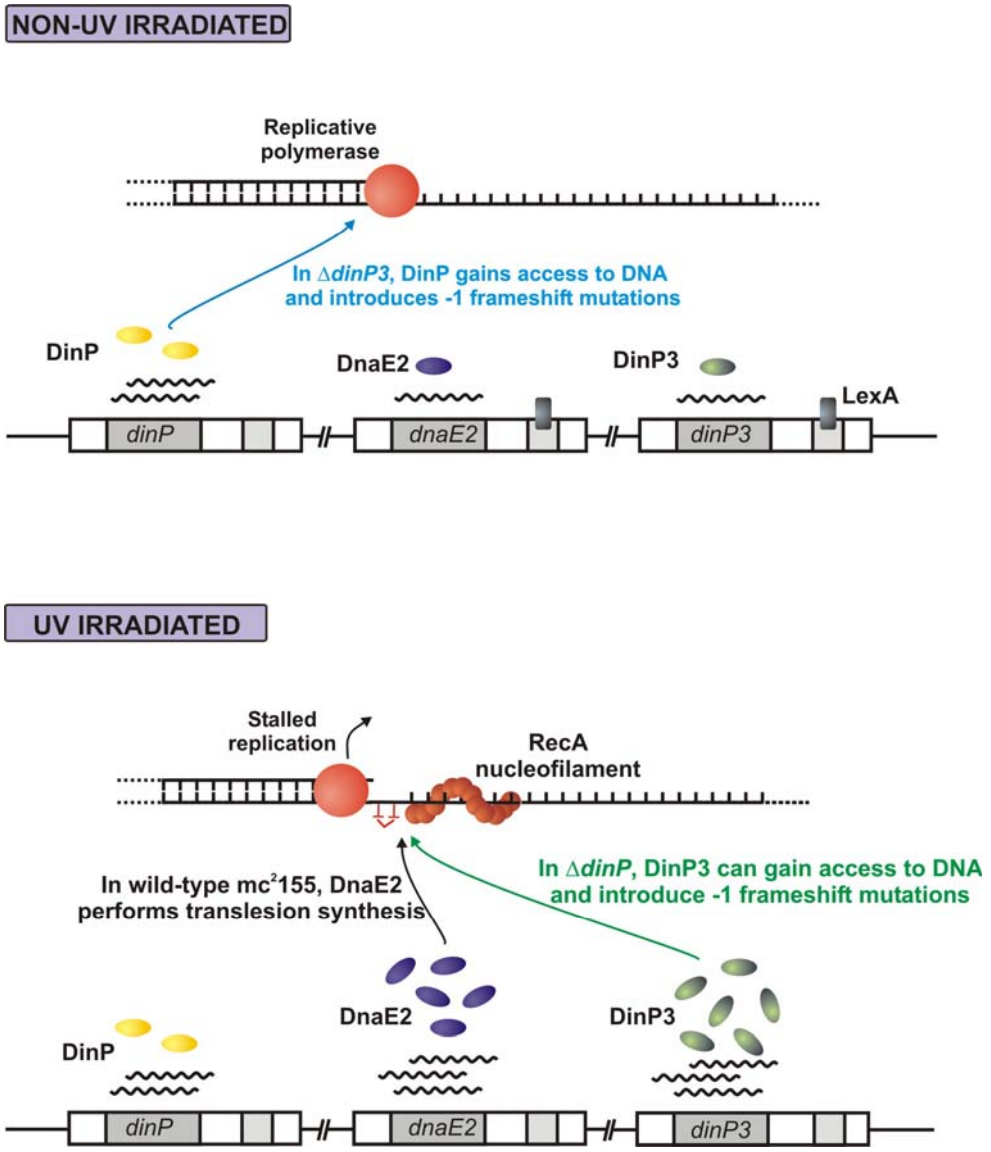


Figure 4.1: A model for –1 frameshift mutagenesis and DNA polymerase interactions in *M. smegmatis* using the *HisD*^{5T} reporter. During normal growth, the replicative polymerase traverses the DNA. In the absence of DinP3, however, DinP gains access to the DNA and introduces –1 frameshift mutations, as highlighted in blue. This suggests that the presence of DinP and DinP3 in wild-type *M. smegmatis* prevents –1 frameshift mutations from occurring. After UV irradiation the DNA becomes damaged and thus blocking the replication fork. In wild-type *M. smegmatis* DnaE2 gains access to the DNA and bypasses the lesion introducing base substitution mutations. However, when DinP is removed, DinP3 out-competes DnaE2 and gains access to the damaged DNA. DinP3 then introduces –1 frameshift mutations, as highlighted in green.

4.3 Tertiary sequence characteristics of the DNA polymerase DinP3

4.3.1 Polymerase switching and the processivity β-clamp

Protein clamps play a critical role in DNA metabolism, either as processivity factors or to manage multi-enzymatic reactions. The β-processivity clamp in bacteria communicates with multiple proteins and can regulate which polymerase gains access to the replication fork. All DNA polymerases interact with the β-clamp on the same locus, suggesting that they compete for the β-clamp at sites of DNA damage (López de Saro *et al.*, 2003). Control of polymerase selection may be exercised at the level of their affinity for the β-clamp. Dalrymple *et al*. (2001) identified a pentapeptide motif, with the consensus QL[SD]LF that allows protein-protein interactions between polymerases and the β-clamp to occur. Analysis of the C-terminal domain of DinP3 revealed a hexapeptide, 354-QLELPF-359, that showed similarity to the β-clamp consensus sequence (Dalrymple *et al.*, 2003). This may explain the ability of DinP3 to out-compete DnaE2 and gain access to the replication fork when DinP is not present. Analysis of the C-terminal domain of DinP did not reveal any β-clamp consensus sequence. However, in the proposed model, it appears that in the absence of DinP3, DinP gains access to the replication fork and introduces –1 frameshift mutations. DinP may gain access to the replication fork even in the absence of a recognisable β-clamp consensus sequence due to the amount of DinP available (levels of DinP) and its ability to localise near the replication fork when DinP3 is absent.

Different polymerases may be recruited differently at the replication fork, which may affect their ability to act as a translesion polymerase. Low processivity tends to correlate with low frameshift fidelity, possibly by allowing rearrangements in the base pairing of the primer/template (Kunkel *et al.*, 2000). The ability of DinP3 or DinP to gain access to the replication fork may suggest that due to low processivity, –1 frameshift mutations are introduced. It remains to be seen how the *M. smegmatis* polymerases interact with the β-clamp in order to gain access to the replication fork.

4.3.2 Possible substrates and site specificity

It has been proposed that the ability of Y-family polymerases to introduce mutations was a result of a relaxed geometric selection of the incoming nucleotide, or due to there being a more flexible active site (Ohashi *et al.*, 2000 and Johnson *et al.*, 2000). By assessing the Y-family polymerases, it appears that each may have an active site that is optimised to accommodate a particular subset of DNA lesions. The active site of DinP3 and one of its homologues, Dpo4, is positioned at different locations (Figure 4.2). This suggests that each polymerase has the ability to bypass a specific type of lesion. DinP3 may able to accommodate bulky lesions or an abasic bulge in the template strain.

It is interesting to observe the overall folding similarities between the two homologues. Despite this, the Y-family polymerases display remarkably different abilities to replicate pass different lesions that are present on the template DNA.

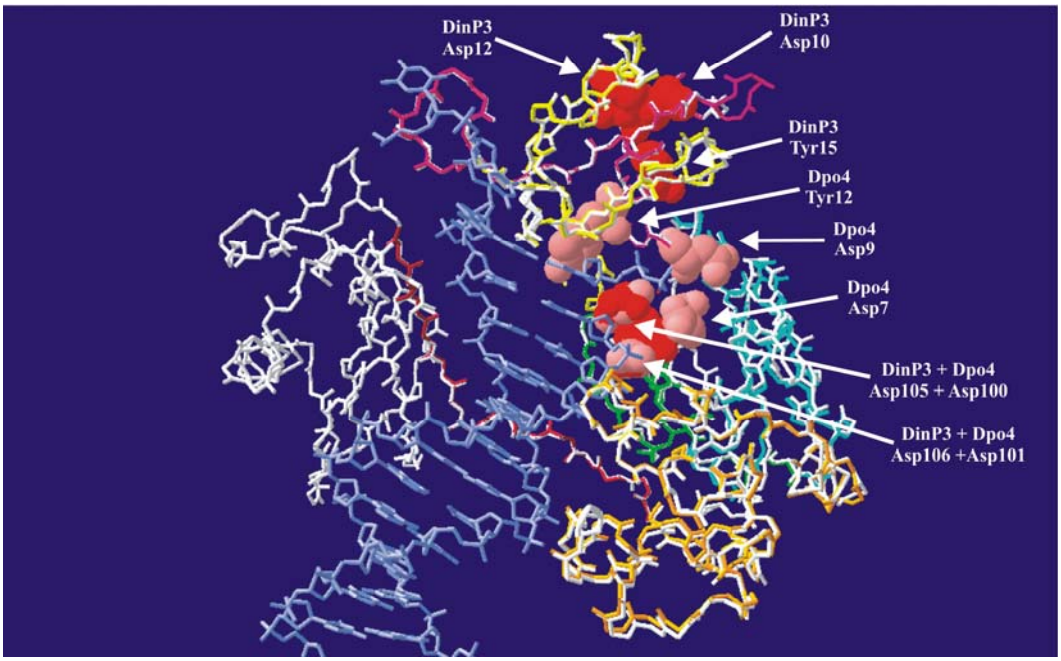


Figure 4.2: Model structure comparing the predicted 3D protein structures of DinP3 to that of Dpo4. The predicted protein structure for DinP3 has been superimposed upon the crystal structure of *Sulfolobus solfataricus* Dpo4 complexed with DNA and an incoming nucleotide (PDB code 1JX4; Yang, 2003). The DinP3 protein backbone is multi-coloured and the Dpo4 protein backbone is represented in white. The conserved amino acids that are required for catalytic function have been enlarged, red for DinP3 and pink for Dpo4.

Pol V tends to introduce base substitutions opposite thymine dimers but also introduce –1 frameshift mutations opposite an N-2-acetylaminofluorene adduct (Napolitano *et al.*, 2000). In order to better understand DinP3's ability to accommodate DNA lesions, an accurate determination of the three-dimensional structure in contact with DNA would be required, for example by using NMR or X-ray crystallography. The “little finger” domain can interact with the DNA template, and may influence the outcome of replication. The “little finger” domain is the least conserved among the Y-family polymerase but may influence the catalytic efficiency and substrate specificity (Ling *et al.*, 2004). The variation of residues in the active site that are in contact with the base pair being synthesized

may specify the type of polymerase that is required for a specific lesion (Yang, 2003).

4.4 Conclusions

All organisms have evolved multiple mechanisms for protecting the fidelity of the genome during replication of both undamaged and damaged DNA (Friedberg *et al.*, 1995). However, specialised DNA polymerases that can promote genetic heterogeneity in the presence of unrepaired base damage have been identified. One such class of DNA polymerases are the Y-family members, which are characterised by their low-fidelity of synthesis of undamaged DNA and their ability to bypass DNA lesions that would normally block replication. The transcriptional control of these polymerases, however, appears to vary greatly between different bacterial species. In *E. coli*, Pol IV is constitutively expressed, and both Pol IV and V are SOS induced (Courcelle *et al.*, 2001 and Napolitano *et al.*, 2000). In *M. tuberculosis*, DinP and DinX are not part of the SOS response (Brooks *et al.*, 2001), and whereas *dinP* is up-regulated by inhibition of the ATPase activity of DNA gyrase (Boshoff *et al.*, 2004), the conditions (if any) that lead to the induction of *dinX* have yet to be identified. In *B. subtilis*, YgjW is expressed upon SOS induction whereas YgjH is expressed constitutively during growth and is not SOS controlled (Duigou *et al.*, 2004). In *M. smegmatis*, DinP3 is induced as part of the SOS response but like *M. tuberculosis*, DinP and DinX are not. The results observed in this study reiterate the growing notion that the mechanisms by which error-prone polymerase action leads to genetic variation differ greatly between bacterial species.

TB remains a leading cause of death by an infectious agent regardless of the advances that have been made in discovering therapies to treat the disease. A hallmark feature of *M. tuberculosis* is its ability to acquire multi-drug resistance through the introduction of genomic mutations (Ramaswamy *et al.*, 1998). The rate at which *M. tuberculosis* introduces mutations depends on a variety of environmental and host responses. The ability of *M. tuberculosis* to become drug-resistant poses a major threat to the effective treatment of this disease using available drugs. An understanding of the mechanisms for genetic diversification that enable this organism to become resistant to one or more drugs by mutation of chromosomal genes associated with drug action is therefore required.

The results presented in this study have provided some insight into the role of error-prone DNA polymerase in –1 frameshift mutagenesis in *M. smegmatis* under normal growth conditions and following UV irradiation. Mycobacteria appear to have adopted a novel way of introducing base substitution mutations in response to DNA damage by primarily utilising a C-family error-prone polymerase instead of the Y-family polymerases. Since *M. tuberculosis* lacks a DinP3 homologue, the DinP3 mutant of *M. smegmatis* may model the situation of *M. tuberculosis*. The balance between error-prone versus error-free replication under a variety of environmental conditions provides an interesting avenue of investigation in mycobacterial species. The results presented in this dissertation highlight the differences in control and mutational profiles of the Y-family polymerases and add to the understanding of the key participants in lesion bypass synthesis.

5.0 APPENDICES

A. Media, Reagents and Solutions

Final volume of media (made up with distilled H_2O) is one litre. Media was sterilized by autoclaving at 121^oC for 10 min.

1. Luria-Bertani Agar (LA) plates

10 g tryptone powder (Oxoid, Ltd, UK), 5 g yeast extract (Oxoid, Ltd, UK), 10 g sodium chloride (Associated Chemical Enterprises, South Africa) and 15 g agar powder (Difco, USA).

2. Middlebrook 7H9 liquid media

4.7 g 7H9 broth base (Difco, USA) was dissolved in 900 mL distilled H_2O and 2 mL glycerol (Merck, Germany) was added. After being autoclaved, 100 mL ADC enrichment media was added along with 5 mL 20 % Tween 80.

3. Middlebrook 7H10 agar plates

19 g of 7H10 agar powder (Difco, USA) was dissolved in 900 mL distilled H_2O and 5 mL glycerol (Merck, Germany) was added. After being autoclaved, 100 mL OADC enrichment media was added.

4. 0.5 % Tween 80

Dilute 25 mL of 20 % Tween 80 (see Appendix A14) into 975 mL distilled H_2O.

5. 75 % Ethanol

Dilute 750 mL 100% Ethanol (BDH, UK) into 250 mL distilled H_2O. This solution was not autoclaved.

6. DEPC-treated H_20

Add 0.1% diethylpyrocarbonate (DEPC) to water; mix overnight, and then autoclave for 20 min.

7. GMS Assay Buffer

Combine 12 % Glycerol, 12 mM 2-[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulphonic acid (HEPES) (pH 7.9), 4 mM Tris-HCl (pH7.9), 60 mM KCl and 1 mM ethylenediaminetetraacetic acid (EDTA).

8. GMS Binding Buffer

Combine 20% glycerol, 100 mM Tris-HCl (pH 8.0), 300 mM KCl, 0.5 mg/mL BSA with 100 mM HEPES.

9. 10 × Tris-Borate (TBE) Buffer

Dissolve 108g Tris base (Sigma-Aldrich, USA) and 55 g Boric Acid (Merck, Germany) in 800 mL distilled H_2O. Add 20 mL 0.5 M EDTA and bring final volume to 1 L with distilled H_2O.

10. 2 × SSC; 0.1 % SDS

Make a 1:10 dilution of 20 × SSC (see Appendix A20) solution and add 0.1% SDS.

11. 0.5 × SSC; 0.1 % SDS

Make a 1:40 dilution of 20 × SSC solution (see Appendix A20) and add 0.1% SDS.

12. Middlebrook 7H10 agar plates with His^{100}

7H10 agar plates were made as in Appendix A3 with 1 mL 100mg/mL L-Histidine (Sigma-Aldrich, USA) being added before pouring agar plates.

13. Middlebrook 7H9 liquid media with His^{100}

7H9 liquid media was made as in Appendix A2, finally 1 mL 100mg/mL L-Histidine (Sigma-Aldrich, USA) was added.

14. 20 % Tween 80

Dilute 200 mL polyoxyethylenesorbitan monooleate (Tween 80) (Sigma-Aldrich, USA) in 800 mL distilled H_2O.

15. 50 × Tris-Acetate (TAE) buffer

Dissolve 242g of Tris base (Sigma-Aldrich, USA) in 700 mL of distilled H_2O and then add 100 mL 0.5 M EDTA, pH 8.0 and 57.1 mL glacial acetic acid (Merck, Germany). Bring the final volume to 1 L with distilled H_2O.

16. 1 × TE Buffer

Add 10 mL 1 M Tris-HCl, pH8.0, to 400 μL 0.25 M EDTA.

17. CTAB Solution

Dissolve 4.1g (0.7M) NaCl (Associated Chemical Enterprises, South Africa) in 80mL dH_2O. While stirring, add 10g (10 %) N-cetyl-N,N,N-trimethyl ammonium bromide (CTAB). If necessary, heat solution to 65^oC. Adjust volume to 100mL with dH_2O and filter sterilize.

18. 6 × DNA Loading Buffer (250 mL)

Add 0.3 g Bromophenol blue, 0.3 g Xylenol, 93.6 mL 80% glycerol, 3 mL 0.5 M EDTA, to 100 mL distilled H_2O. Make up to 250 mL with distilled H_2O.

19. 10 × TGOE Buffer

Dissolve 30.2 g 0.25M Tris (Sigma-Aldrich, USA) and 142.6 g 1.9 M Glycine (Merck, Germany) in 800 mL distilled H_2O. Adjust pH to 8.3 with Acetic acid (Merck, Germany) and bring volume to 1 L with distilled H_2O.

20. 20 × SSC

Dissolve 175.3 g NaCl (Associated Chemical Enterprises, South Africa) and 88.2 g Sodium citrate (BDH, UK) in 800 mL distilled H_2O. Adjust the pH to 7.0 with 1M HCl and bring final volume to 1L with distilled H_2O.

B. General Laboratory Methods

1. Dilution series (100 μL)

In order to calculate the number of cells present in a culture a dilution series was created. A 10 × dilution was generated by consecutively diluting 100 μL aliquots into 900 μL 0.5 % Tween 80 (see Appendix A4). The first 100 μL aliquot was taken from the culture of interest and diluted. A dilution range of 10^{-1} to 10^{-8} was created.

2. RNA gel electrophoresis

The quality of extracted RNA was analysed by resolving RNA samples on an agarose gel. For a 2 % agarose gel, 0.6 g of normal melting point agarose (FMC BioProducts, USA) was dissolved in 30 mL 1x TAE buffer (see Appendix A15) and 0.1 % SDS was added. Samples were resolved until dye front had migrated three quarters down the gel. The gel was stained for 10 min in 0.0001:1 Ethidium Bromide:H_2O whilst being agitated.

3. Mycobacterial genomic DNA extraction

All mycobacterial genomic DNA was isolated using the CTAB genomic DNA extraction method. Cells were grown in 7H9 liquid media (see Appendix A2) until well grown, before being heat killed at 75^{o}C for 35 min. Following centrifugation (5415D, Eppendorf International) for 30 s at 13 000 rpm the cell pellets were re-suspended in 500 μL TE buffer (see Appendix A16). To the suspension, 50 μL of 10 mg/mL lysoyme (Sigma-Aldrich, USA) was added before incubating at 65^{o}C for 1 and half hours. Next 100 μL 5M NaCl (Associated Chemical Enterprises,

South Africa) was added and mixed and then 80 μL CTAB solution (see Appendix A17) was added to the sample and mixed well. The sample was incubated for 10 min at 65^{o}C before an equal volume of $CHCl_3$/isoamyl alcohol was added, mixed and centrifuged (5415D, Eppendorf international) for 5 min at 13 000 rpm. The aqueous phase was removed and to this 0.6 volumes of isopropanol (Labchem, South Africa) was added and placed on ice for 30 min. After centrifugation (5415D, Eppendorf international) at 13 000 rpm for 20 min at 4^{o}C the pellet was washed with cold 70 % Ethanol and re-centrifuged for 5 min at 4^{o}C. The pellet was dried and re-suspended in distilled H_2O.

4. DNA Gel Electrophoresis

All PCR products were analysed on agarose gels. For a 2 % agarose gel, 0.6 g of normal melting point agarose (FMC BioProducts, USA) was dissolved in 30 mL 1 × TAE buffer (see Appendix A15) with 1 mL Ethidium Bromide per 30 mL buffer. Samples were loaded with DNA loading buffer (see Appendix A18) and resolved until dye front was three quarters the length of the gel. The product sizes were compared to the 1 μg DNA marker III (Roche Applied Science, Germany)

5. DNA Gel Electrophoresis (low melting point agarase)

For 1 % low melting point agarose gel dissolve 0.3 g low melting point agarose (FMC BioProducts, USA) in 30 mL 1× TAE buffer with 1 mL Ethidium Bromide per 30 mL buffer. Samples were loaded with DNA loading buffer (see Appendix A18) and resolved until dye front was three quarters the length of the gel.

6. β-Agarase procedure

The required DNA band was cut out of a 1 % low melting point agarose gel (see Appendix B4) and to this 0.04 volumes of 25 × Agarase buffer (Sigma-Aldrich, USA) was added. The sample was then incubated at 65^{o}C for 15 min until agarose had completely melted. Samples were cooled to 42^{o}C and 1 unit Agarase (Sigma-Aldrich, USA) per 100 mg agarose was added. The samples were then incubated for 1 hr at 42^{o}C before Ethanol precipitation of the DNA fragments was performed (see Appendix B7).

7. Ethanol precipitation of DNA

To a measured sample, one tenth volume of 3M NaOAc, pH 5.2 and at least two volumes of cold 100 % Ethanol (Merck, Germany) were added. The sample was then allowed to stand for more than 30 min at –20^{o}C and centrifuged (5415D, Eppendorf International) at 4^{o}C for 15 min at 12 000 rpm. The supernatant was discarded and the pellet was washed with cold 75 % Ethanol (see Appendix A5) and centrifuged (5415D, Eppendorf International) for 5 min at 4^{o}C at 12 000 rpm. The DNA pellets were air dried for 5 min and re-suspended in distilled H_2O.

8. Bradford Assay

The Bradford assay was used to determine the protein concentration in the cell free extracts. GMS assay buffer (see Appendix A7), 600 μL, was added to 200 μL of cell free extract and mixed gently. To this 200 μL of Bradford Reagent (BioRad, USA) was added and mixed gently by inverting the tubes. The samples were left at room temperature for 10 min and then the absorbance at 595 nm was recorded. BSA (Roche Applied Science, Germany) was used to generate a

standard protein concentration curve. BSA (Roche Applied Science, Germany) was diluted with GMS assay buffer (see Appendix A7) to give known protein concentrations in a volume of 800 μL. To this 200 μL of the Bradford Reagent (BioRad, USA) was also added and left at room temperature for 10 min. The absorbance at 595 nm for the known standards was recorded and a standard curve was generated. The protein concentration of the sample was then determined from the standard curve.

9. Non-denaturing Polyacrylamide Electrophoresis

Nucleic acid/protein complexes were separated using non-denaturing electrophoresis. A 4 % polyacrylamide gel was made by combining together 9.25 mL (20 %/0.33 %) acrylamide/bisacrylamide (both from BioRad, USA), 3.5 mL 10 × TGOE buffer (see Appendix A19), 1.75 mL 50 % Glycerol, 35 μL 0.5 M DTT (Roche Applied Science, Germany), 300 μL 10 % Ammonium persulphate and 30 μL N,N,N',N',-tetramethylethylenediamine (TEMED)(BioRad, USA). The volume was made up to 37 mL with distilled H_2O.

10. Histidine removal washes

To test for histidine reversion, cultures were plated on 7H10 plates (see Appendix A12) but were grown in liquid media containing histidine. Removal of all histidine from solution is necessary to prevent colonies from growing in the presence of histidine. The bacterial cultures were subjected to three rounds of washing with 0.5 % Tween 80 (see Appendix A4) to remove the media containing the histidine. Bacterial cells were pelleted by centrifuging (5415D, Eppendorf International) for 1 min at 10 000 rpm. The pellets were suspended in 0.5 %

Tween 80 and then re-centrifuged (5415D, Eppendorf International) for 1 min at 10 000 rpm. This was repeated two more times.

11. Dilution series (100 mL)

To dilute the pre-culture down to 10^2 CFU/mL a dilution series was established. The pre-culture was diluted down in 100 mL volumes. A 1 mL aliquot from the pre-culture was diluted in 99 mL 7H9 liquid media with histidine (see Appendix A13), followed by a 1 mL aliquot from this being diluted in a further 99 mL 7H9 liquid media with histidine (see Appendix A13). A 10 mL aliquot was then taken from this dilution and suspended in 90 mL 7H9 liquid media with histidine (see Appendix A13). The last dilution was used as the initial inoculum.

C. Other Appendices

1. Primer position for RT-PCR

The primers were designed to amplify a region internal to the coding sequence of each gene. Amplicons are shown relative to the start of the gene and primer positions are indicated in red.

RT-PCR for *dnaE2*:

Primer	Tm(°C)	Sequence (5'-3')
Forward	64.5	CTG CGC AGC AGG TTC TAC GA
Reverse	64.5	AGA ACA CCA GCG AGG CGA AG

RT-PCR for *dinP3*:

Primer	Tm (°C)	Sequence (3'– 5')
Forward	64.5	CGA GGC GTT CCT CGA CGT AT
Reverse	64.5	CAC CAA GTT CCT GGC CAA GG

atgttcgtgtccgctgcggaatccgcgagcatcctgcacgctgacctcgactcgttctacgc

ctcggtcgagcagcgcgacgatccggcgctgcgcggacgcccggtgatcgtcggcggcggtg

tggtgctggcggccagttacgaggccaaggcctacggcgtccgcaccgcgatgagcggcggc

caggcccgggccctgtgcccacaggccatcgtggtcccgccgcgcatggcggcctacacaca

ggccagccgcgatgtgttcgcggtcttccacgacaccaccccgctggtggaacctctgtcgg

tcga**CGAGGCGTTCCTCGACGTAT**CCGGGCTGGCCCGGGTGTCGGGCACGCCGGTGGAGATC

GCCGCGCGCCTGCGCGCACGGGTCCGCGAGCAGGTCGGGCTGCCCATCACAGTCGGCATCGC

dinP3 amplicon -143 bp

GCGCACCAAGTTCCTGGCCAAGGtcgccagccaggagggcaagcccgacgggttgctcctgg

tgccacccgac.....

RT-PCR for *sigA*:

Primer	Tm(°C)	Sequence (5'-3')
Forward	64.5	TCT ACG CCA CGC AGA AGC TG
Reverse	64.5	GTT CGC CTC CAG CAG ATG GT

gtgaagcgcaccgctaccaagacccccgccaagaaggccccagccaagcgggcagccaagagcg

ccgcggcgaaagccggcggcaaggcccccgccaagaaagccccggccaagcgggccgccaaagg

caccgctgccaagcccgaggacggcgtcaccgacgaccttgaggtgaccgacgatctcgaggct

gagcccggcgaggatctcgacgtcgaagacaccgacctggaactcgacgacctcgactccgacg

acgacaccgcagtcgaggacgaggaagaagaagctgatgccgcgacgcccgccgtcgcgaccgc

gaaggccgccgacgacgacatcgacgagccgtccgagaaggacaaggcgtccggcgacttcgtg

tgggacgaggaagagtccgaggccctgcgacaggcacgcaaagacgccgagctgaccgcatcgg

ccgactcggttcgcgcctacctcaagcagatcggcaaggtggccctgctcaacgccgaagaaga

ggtggagctggccaagcgcatcgaggcgggtc**TCTACGCCACGCAGAAGCTG**GCCGAACTCGCC

GAGAAGGGCGAGAAGCTGCCAGTGCAGCAGCGCCGCGACATGCAGTGGATCTGCCGCGACGGCG

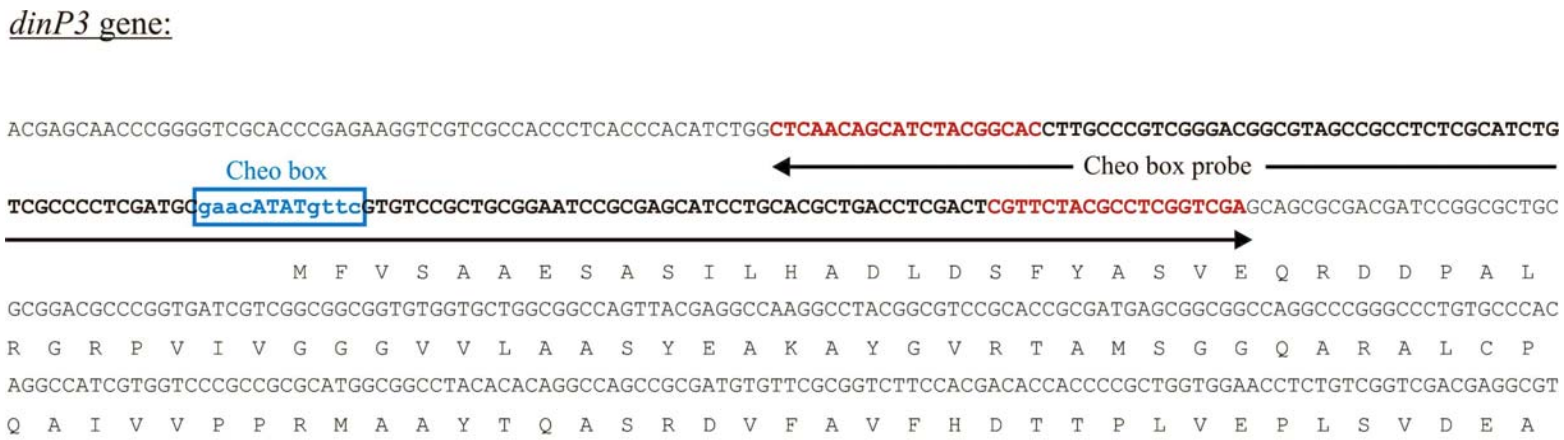
sigA amplicon - 128 bp

ACCGCGCCAAAAACCATCTGCTGGAGGCGAACctgcgcctcgtggtgtcgctggccaagcgcta

caccggccgcggcatggcgttcctcgacc.....

2. Cheo box regulator sequence in *dinP3* promoter

The *dinP3* Cheo box was investigated for putative LexA binding. A 149 bp region of the *dinP3* promoter, containing the Cheo box, was amplified. The primer positions are indicated in red and the Cheo box is highlighted in turquoise.



3. Probe for Southern analysis of *ΔdinP3* strain

Digesting a vector generated the probe for southern analysis, pGEMdinP35′probe[2], containing the 5′ region of the *dinP3* gene (Figure 5.1). The vector was digested simultaneously with *Hind*III (AEC Amersham International) and *Xba*I (Roche Applied Science, Germany) at 37°C for 1 hr in a 20 μL reaction with the digestion buffer consisting of 100 mM Tris-HCl, 1 M NaCI, 50 mM MgCl2, 10 mM 2-mercaptoethanol, pH 8.0. An 859 bp probe, the turquoise region in the map below, was resolved on a 1% Agarose gel (see Appendix B5), extracted using β-Agarase (see Appendix B6) and purified with Ethanol precipitation (see Appendix B7). This probe was used to determine the complete knockout of *dinP3* gene.

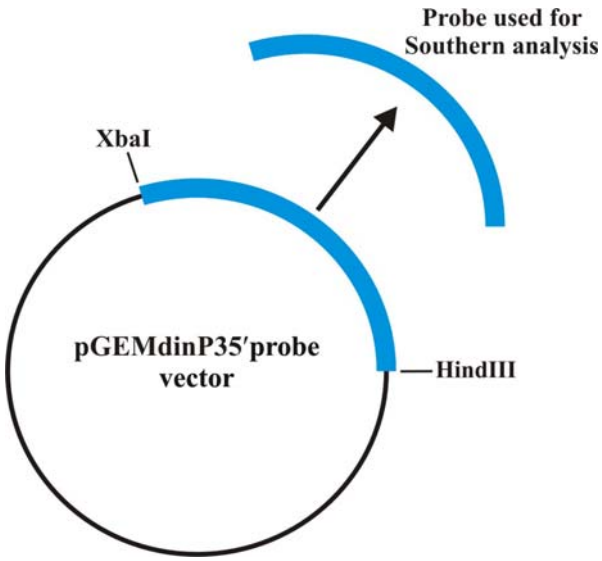


Figure 5.1: Schematic representation of probe generation for Southern analysis.

[2] S.S. Dawes in our laboratory designed the pGEMdinP35′probe vector.

4. *HisD*^{5T} auxotrophy and sequencing

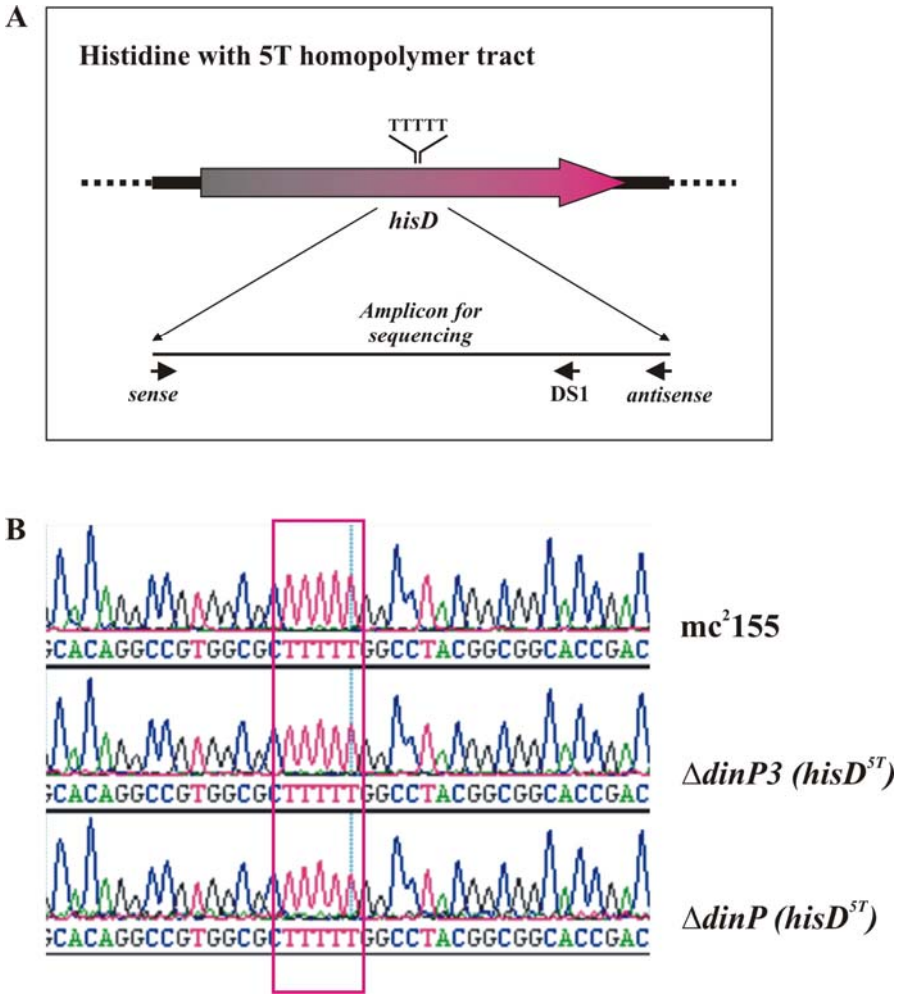


Figure 5.2: Analysis of 5T homopolymeric tract for histidine reversion assays. (A) Schematic representation of *hisD* gene. Histidine auxotrophs contain a 5T homopolymeric tract that is reverted to a 4T homopolymeric tract after a –1 frameshift mutation has been introduced. (B) Sequence analysis of *hisD*^{5T} gene from histidine auxotroph colonies.

5. Equations for mutation rate calculation

Listed below are the equations for the calculation of *m*, which is needed to calculate the mutation rate.

Table 5.1: MSS maximum likelihood method used to calculate the mutation rate

Equation name	Equation	Reference.
MSS algorithm	$p_0 = e^{-m}; p_r = \frac{m}{r}\sum_{i=0}^{r-1}\frac{p_i}{(r-i+1)}$	Ma *et al.*, 1992
Likelihood function	$f(r\|m) = \prod_{i=1}^{C} f(r_i\|m)$ where $f(r\|m) = p_r$ from Eq. [1]	Roche *et al.*, 2000
Mutation rate	$m/(1.44N_t)$	Roche *et al.*, 2000

6. Breakdown of experimental values used to determine mutation rates

Table 5.2: Breakdown of individual mutation rates observed for the various fluctuation assays

		Non-UV irradiated			UV irradiated		
		mc^{2}155	**Δ*dinP3***	**Δ*dinP***	**mc^{2}155**	**Δ*dinP3***	**Δ*dinP***
#1	**Mutation rate (μ)**	1.17E-08	3.83E-07	2.23E-08	2.51E-08	3.24E-08	1.32E-07
	Mutations per culture (*m*)	28	225	8.5	36.5	27.5	58
	N_t value	1.66E+09	4.08E+08	2.65E+08	1.01E+09	5.90E+08	3.05E+08
#2	**Mutation rate (μ)**	9.54E-09	1.01E-07	1.95E-08	1.04E-08	2.28E-08	1.07E-07
	Mutations per culture (*m*)	9.2	95	9.1	23	10.5	26
	N_t value	6.70E+08	6.51E+08	3.24E+08	1.54E+09	3.20E+08	1.69E+08
#3	**Mutation rate (μ)**	2.98E-08	3.05E-07	3.50E-08	1.88E-08	4.62E-08	1.81E-07
	Mutations per culture (*m*)	5.5	210	121	14.6	35	130
	N_t value	1.31E+08	4.79E+08	2.40E+09	5.39E+08	5.26E+08	4.99E+08

All mutation rates were determined using the MSS Maximum likelihood method. The mutation rates indicate –1 frameshift mutations occurring in a polythymine run in the *hisD* gene. The mutation rate is the probability of a cell sustaining a mutation during its lifetime, and is calculated by dividing the mean number of mutations per culture by the number of divisions occurred.

6.0 REFERENCES

Abella, M., Erill, I., Jara, M., Mazón, G., Campoy, S., and Barbé, J. (2004) Widespread distribution of a *lexA*-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum. *Mol Microbiol* 54: 212-222.

Barry, C.E. 3rd. (1997) New horizons in the treatment of tuberculosis. *Biochem Pharmacol* 54: 1162-1172.

Baynton, K., and Fuchs, R.P.P. (2000) Lesions in DNA: hurdles for polymerases. *Trends Biochem Sci* 25: 74-79.

Boshoff, H.I.M, Durbach, S.I., and Mizrahi, V. (2001) DNA metabolism in *Mycobacterium tuberculosis*: implications for drug resistance and strain variability. *Scan J Infect Dis* 33: 101-105.

Boshoff, H.I.M., Reed, M.B., Barry, C.E., 3rd, and Mizrahi, V. (2003) DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* 113: 183-193.

Boshoff, H.I.M., Meyers, T.G., Copp, B.R., McNeil, M.R., Wilson, M.A., and Barry, C.E., 3rd. (2004) The transcriptional responses of *M. tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem* 279: 40174-40184.

Bridges, B.A., and Woodgate, R. (1985) Mutagenic repair in *Escherichia coli*: products of the *recA* gene and the *umuD* and *umuC* gene act at different steps in UV-induced mutagenesis. *Proc Natl Acad Sci USA* 82: 4193-4197.

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

Brotcorne-Lannoye, A., and Maenhaut-Michel, G. (1986) Role of RecA protein in untargeted UV mutagenesis of bacteriophage lambda: evidence for the requirement for the *dinB* gene. *Proc Natl Acad Sci USA* 83: 3904-3908.

Campos, P.E., Suarez, P.G., Sanchez, J., Zavala, D., Arevalo, J., Ticona, E., *et al.* (2003) Multidrug-resistant *Mycobacterium tuberculosis* in HIV-infected persons, Peru. *Emerg Infect Dis* 9: 1571-1578.

Cheo, D.L., Bayles, K.W., and Yasbin, R.E. (1991) Cloning and characterization of DNA damage-inducible promoter regions from *Bacillus subtilis*. *J Bacteriol* 173: 1696-1703.

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

Collins, F.M. (1998) Mycobacterial pathogenesis: a historical perspective. *Frontiers Biosci* 3:123-132.

Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C., *et al.* (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 163: 1009-1021.

Courcelle, J., Khodursky, A., Peter, B., Brown, P.O., and Hanawalt, P.C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* 158: 41-64.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001) A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc Natl Acad Sci USA* 98: 11627-11632.

Dalrymple, B.P., Wijffels, G., Kongsuwan, K., and Jennings, P. (2003) Towards an understanding of protein-protein interaction network hierarchies.

Analysis of DnaN (β)-binding peptide motifs in members of protein families interacting with eubacterial processivity clamp, the β subunit of DNA polymerase III. First Asia-Pacific Bioinformatics Conference. Adelaide, Australia.

Dannenberg, A.M., Jr., and Rook, G.A.W. (1994) Pathogenesis of pulmonary Tuberculosis: an interplay of tissue-damaging and macrophage-activating immune responses-dual mechanisms that control bacillary multiplication. *In Tuberculosis: Pathogenesis, Protection and Control*. (Bloom, B.R., ed) ASM Press, Washington, D.C. pp 459-483.

Dolin, P.J., Raviglione, M.C., and Kochi, A. (1994) Global tuberculosis incidence and mortality during 1990-2000. *Bull WHO* 72: 213-220.

Drake, J.W. (1988) Clone size distributions of mutations induced by ethyl methanesulfonate in bacteriophage T4. *J Mol Biol* 202: 11-16.

Duigou, S, Ehrlich, S.D., Noirot, P., and Noirot-Gros, M.F. (2004) Distinctive genetic features exhibited by the Y-family DNA polymerases in *Bacillus subtilis*. *Mol Microbiol* 54: 439-451.

Durbach, S.I., Andersen, S.J., and Mizrahi, V. (1997) SOS induction in mycobacteria: analysis of the DNA-binding activity of a LexA-like repressor and its role in DNA damage induction of the recA gene from *Mycobacterium smegmatis*. *Mol Microbiol* 26: 643-653.

Echols, H., and Goodman, M.F. (1991) Fidelity mechanisms in DNA replication. *Annu Rev Biochem* 60: 477-511.

Eguchi, Y., Ogawa, T., and Ogawa, H. (1988) Cleavage of bacteriophage phi80 CI repressor by *recA* protein. *J Mol Biol* 202: 365-73.

Elledge, S.J., Walker, G.C. (1983) Proteins required for ultraviolet light and chemical mutagenesis: identification of the products of the *umuC* locus of

Escherichia coli. *J Mol Biol* 164: 175-192.

Ferebee, S.H., Doster, B.E., and Murray, F.J. (1966) Ethambutol: a substitute for para-aminosalicyclic acid in regimens for pulmonary tuberculosis. *Ann N Y Acad Sci* 135: 910-920.

Foster, P.L. (1993) Adaptive mutation: the uses of adversity. *Ann Rev Microbiol* 47: 467-504.

Foster, P.L. (1999) Mechanisms of stationary phase mutation: A decade of adaptive mutation. *Annu Rev Genet* 33 (1999) 57–88.

Foster, P.L. (2004) Adaptive mutation in *Escherichia coli*. *J Bacteriol* 186: 4846-4852.

Franklin, W.A., Doetsch, P.W., and Haseltine, W.A. (1985) Structural determination of the ultraviolet light-induced thymine-cytosine pyrimidine-pyrimidine (6-4) photoproduct. *Nucleic Acids Res* 13: 5317-5325.

Friedberg, E. C., Walker, G.C., and Siede, W. (1995) In *DNA Repair and Mutagenesis.* ASM Press, Washington. D.C. pp 408-428.

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

Friedberg, E.C., Feaver, W. J., Gerlach, V. (2000) The many faces of DNA polymerases: Strategies for mutagenesis and for mutational avoidance. *Proc Natl Acad Sci USA* 97: 5681-5683.

Frieden, T.R., Sterling, T.R., Munsiff, S.S., Watt, C.J., and Dye, C. (2003) Tuberculosis. *Lancet* 362: 887-899.

Fujii, S., Gasser, V., and Fuchs, R.P. (2004) The biochemical requirements of DNA polymerase V-mediated translesion synthesis revisited. *J Mol Biol* 341: 405-417.

Furesz, S., and Timball, M.T. (1963) The antibacterial activity of rifamycins. *Chemotherapia* 7: 200.

Gerlach, V.L., Aravind, L., Gotway, G., Schultz, R.A., Koonin, E.V., and Friedberg, E.C. (1999) Human and mouse homologs of *Escherichia coli* DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proc Natl Acad Sci USA* 96: 11922-11927.

Goodman, M.F., and Tippin, B. (2000) Sloppier copier DNA polymerises involved in genome repair. *Curr Opin Genet Dev* 10: 162-168.

Goodman, M.F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Ann Rev Biochem* 71: 17-50.

Harfe, B.D., and Jinks-Robertson, S. (2000) DNA mismatch repair and genetic instability. *Ann Rev Genet* 34: 359-399.

Hazelbauer, G.L., and Harayama, S. (1979) Mutants in transmission of chemotactic signals from two independent receptors of *E. coli. Cell* 16: 617-625.

Hingley-Wilson, S.M., Sambandamurthy, V.K., and Jacobs, W.R. Jr. (2003) Survival perspectives from the worlds most successful pathogen, *Mycobacterium tuberculosis. Nat Rev Immunol* 4: 949-955.

Hingorani, M.M., and O'Donnell, M. (2000) Sliding clamps: a (tail)ored fit. *Curr Biol* 10: R25-R29.

Horii, T., Ogawa, T., Nakatani, T., Hase, T., Matsubara, H., and Ogawa, H. (1981) Regulation of SOS functions: purification of *E. coli* LexA protein

and determination of its specific site cleaved by the RecA protein. *Cell* 27: 515-522.

Houston, R., Ray, S., Mahari, M., Neill P., Legg, W., Latif, A.S., *et al.* The association of tuberculosis and HIV infection in Harare, Zimbabwe. *Tuber Lung Dis* 75: 220-226.

Huisman, O., and D'Ari, R. (1981) An inducible DNA replication-cell division coupling mechanism in *E. coli*. *Nature* 290: 797-799.

Inoue, R., *et al.* (2001) Genetic identification of two distinct DNA polymerises, DnaE and PolC, that are essential for chromosomal DNA replication in *Staphylococcus aureus. Mol Genet Genomics* 266: 564-571.

Jackett, P.S., Aber, V.R., and Lowrie, D.B. (1978) Virulence and resistance to superoxide, low pH and hydrogen peroxide among strains of *Mycobacterium tuberculosis. J Gen Microbiol* 104: 37-45.

Johnson, R.E., Washington, M.T., Prakash, S., and Prakash, L. (2000) Fidelity of human DNA polymerase η. *J Biol Chem* 275: 7447-7450.

Kato, T., and Shinoura, Y. (1977) Isolation and characterization of mutants of *Escherichia coli* deficient in induction of mutations by ultraviolet light. *Mol Gen Genet* 156:121-131.

Kenyon, C.J., and Walker, G.C. (1980) DNA-damaging agents stimulate gene expression at specific loci in *Escherichia. coli*. *Proc Natl Acad Sci USA* 77: 2819- 2823.

Knox, R., King, M.B., and Woodroffe, R.C. (1952) *In vitro* action of isoniazid on *Mycobacterium tuberculosis. Lancet* 2:854-858.

Koch, R. (1882) The aetiology of tuberculosis. *Berlin Klin Wochenschr* 19: 221-230.

Kobayashi, S., Valentine, M.R., Pham, P., O'Donnell, M., and Goodman, M.F. (2002) Fidelity of *Escherichia coli* DNA polymerase IV: preferential generation of small deletion mutations by dNTP-stabilized misalignment. *J Biol Chem* 277: 34198-34207.

Kunkel, T.A., Loeb, L.A., and Goodman, M.F. (1984) On the fidelity of DNA replication. The accuracy of T4 DNA polymerase in copying phiX174 DNA in vitro. *J Biol Chem* 259: 1539-1545.

Kunkel, T.A., and Bebenek, K. (2000) DNA replication fidelity. *Annu Rev Biochem* 69: 497-529.

Lathigra, R.B., Young, D.B., Sweetser, D., and Young, R.A. (1985) A gene from *Mycobacterium tuberculosis* is homologous to the DnaJ heat shock protein of *E. coli*. *Nucleic Acid Res* 16: 1636.

Lea, D.E., and Coulson, C.A. (1949) The distribution of the numbers of mutants in bacterial populations. *J Genet* 49: 264-285.

Lewis, K.L., Harlow, G.R., Gregg-Jolly, L.A., and Mount, D.W. (1994) Identification of high-affinity binding sites for LexA which define new DNA damage-inducible genes in *Escherichia coli*. *J Mol Biol* 241:507-523.

Ling, H., Boudsocq, F., Woodgate, R., and Yand, W. (2001) Crystal structure of a Y-family polymerase in action: A mechanism for error-prone and lesion-bypass replication. *Cell* 107: 97-102.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2004) Snapshots of replication through an abasic lesion: structural basis for base substitutions and frameshifts. *Mol Cell* 13: 751-762.

Little, J. W. (1984) Autodigestion of *lexA* and phage λ repressors. *Proc Natl Acad Sci USA* 81: 1375-1379.

López de Saro, F.J., Georgescu, R.E., Goodman, M.F., and O'Donnell, M. (2003) Competitive processivity-clamp usage by DNA polymerases during DNA replication and repair. *EMBO J* 22: 6408-6418.

Lovett, S.T. (2004) Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol Microbiol* 52: 1243-1253.

Luria, S.E., and Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28: 491-511.

Ma, W.T, Sandri, Gv. H., and Sarkar, S. (1992) Analysis of the Luria-Delbrück distribution using discrete convolution powers. *J Appl Prob* 29: 255-267.

Malik, Z.A., Iyer, S.S., and Kusner, D.J. (2001) *Mycobacterium tuberculosis* phagosomes exhibit altered calmodulin-dependent signal transduction: contribution to inhibition of phagosome-lysosome fusion and intracellular survival in human macrophages. *J Immunol* 166:3392-3401.

Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F.R., and Smith, I. (1999) Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol Microbiol* 31:715-724.

Manganelli, R., Tyagi, S., and Smith, I. (2001) Real-time PCR using molecular beacons. In *Methods in molecular medicine, vol. 54: Mycobacterium tuberculosis protocols*. (Parish, T., Stoker, N.G., eds.) Humana Press Inc. Totowa, N.J. pp 295-310.

McKenzie, G.J., and Rosenberg, S.M. (2001) Adaptive mutagenesis, mutator DNA polymerases and genetic change strategies of pathogens. *Curr Opin Microbiol* 4: 586-594.

Metzgar, D., and Wills, C. (2000) Evidence for the adaptive evolution of mutation rates. *Cell* 101: 581-584.

Miller, J.H., Suthar, A., Tai, J., Yeung, A., Truong, C., and Stewart, J.L. (1999) Direct selection for mutators in *Escherichia coli*. *J Bacteriol* 181: 1576-1584.

Mizrahi, V., and Andersen, S.J. (1998) DNA repair in *Mycobacterium tuberculosis.* What have we learnt from the genome sequence? *Mol Microbiol* 29:1331-1339.

Moai-Shoshani, A., and Livneh, Z. (2002) Analysis of the stimulation of DNA polymerase V of *Escherichia coli* by processivity proteins. *Biochemistry* 41: 14438-14446.

Movahedzadeh, F., Colston, M.J., and Davis, E.O. (1997) Determination of DNA sequences required for regulated *Mycobacterium tuberculosis* RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist. *J Bacteriol* 179: 3509-3518.

Moxon, E.R., Rainey, P.B., Nowak, M.A., and Lenski, R.E. (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* 4: 24-33.

Mukherjee, A., Cao, A. C., and Lutkenhaus, J. (1998) Inhibition of FtsZ polymerization by SulA, an inhibitor of septation in *Escherichia coli. Proc Natl Acad Sci USA* 95: 2885-2890.

Napolitano, R., Janel-Blintz R., Wagner, J., and Fuchs, R.P. (2000) All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *EMBO J* 19: 6259-6265.

Negishi, K., Bessho, T., and Hayatsu, H. (1994) Nucleoside and nucleobase analog mutagens. *Mutat Res* 318: 227-238.

Ohashi, F., Bebenek, K., Matsuda, T., Feaver, W.J., Gerlach, V.L., Friedberg, E.C., *et al.* (2000) Fidelity and processivity of DNA synthesis by DNA

polymerase kappa, the product of the human DINB1 gene. *J Biol Chem* 275: 39678-39684.

Ohmori, H., Friedberg, E.C., Fuchs, R.P.P., Goodman, M.F., Hanaoka, F., Hinkle, D., *et al.* (2001) The Y-Family of DNA polymerases. *Mol Cell* 8: 7-8.

Pablos-Mendez, A., Raviglione, M.C., Laszlo, A., Binkin, N., Rieder, H.L., Bustreo, F., *et al.* (1998) Global surveillance for antituberculosis-drug resistance, 1994-1997. World Health Organisation-International Union against Tuberculosis and Lung Disease Working Group on Anti-Tuberculosis Drug Resistance Survey. *N Engl J Med* 338:1641-1649.

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

Pham, P., Rangarajan, S., Woodgate, R., and Goodman, M.F. (2001) Roles of DNA polymerases V and II in SOS-induced error-prone and error-free repair in *Escherichia coli. Proc Natl Acad Sci USA* 98: 8350-8354.

Pham, P., Seitz, E.M., Saveliev, S., Shen, X., Woodgate, R., Cox, M.M., and Goodman, M.F. (2002) Two distinct modes of RecA action are required for DNA polymerase V-catalyzed translesion synthesis. *Proc Natl Acad Sci USA* 99: 11061-11066.

Radman, M. (1974) Phenomenology of an inducible mutagenic DNA repair pathway in *Escherichia coli*: SOS repair hypothesis. In *Molecular and Environmental Aspects of Mutagenesis* (Sherman, S., Miller, M., Lawrence, C. and Tabor, W.H., eds.) Charles C. Thomas Publisher, Springfield. pp 128–142.

Radman, M. (1975) SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci* 5A:355-367.

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

Rattray, A.J., and Strathern, J.N. (2003) Error-prone DNA Polymerases: When making a mistake is the only way to get ahead. *Annu Rev Genet* 37: 31-66.

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

Rosenberg, S. M. (2001) Evolving responsively: Adaptive mutation. *Nat Rev Genet* 2: 504–515.

Rosenberg, S.M., Hastings, P.J. (2004) Adaptive point mutation and adaptive amplification pathways in the *E. coli* Lac system: Stress responses producing genetic change. *J Bacteriol* 186: 4838–4843.

Ross, H.J. (1950) Streptomycin and PAS in pulmonary tuberculosis. *Ir J Med Sci* 6: 745-752.

Rost, B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol* 266: 525:539.

Roth J.R., and Andersson D.I. (2004) Adaptive mutation: How growth under selection stimulates Lac^+ reversion by increasing target copy number. *J Bacteriol* 186: 4855–4860.

Rozen, S., and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* (Krawetz, S., Misener, S., eds.) Humana Press, Totowa, NJ. pp 365-386.

Schaaper, R.M. (1993) Base selection, Proofreading, and mismatch repair during DNA Replication in *Escherichia coli*. *J Biol Chem* 266: 23762-23765.

Sedgwick, S.G. (1975) Inducible error-prone repair in *Escherichia coli*. *Proc Natl Acad Sci USA* 72:2753-2757.

Shinagawa, H., Iwasaki, H., Kato, T., and Nakata, A. (1988) RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc Natl Acad Sci USA* 85: 1806-1810.

Smith, B.T., and Walker, G.C. (1998) Mutagenesis and more: *umuDC* and the *Escherichia coli* SOS response. *Genetics* 148: 1599-1610.

Smith, T.F., and Wateman, M.S. (1981) Identification of common molecular sequences. *J Mol Biol* 147: 195-197.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R. Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* 4: 1911-1919.

Sniegowski, P. D., Gerrish, P.J., and Lenski, R.E. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387: 703–705.

Stewart, F.M. (1994) Fluctuation tests: how reliable are the estimates of mutation rates? *Genetics* 137: 1139-1146.

Stewart, G.R., Robertson, B.D., and Young, D.B. (2003) Tuberculosis: a problem with persistance. *Nat Rev Microbiol* 1: 97-105.

Sudre, P., ten Dam, G., and Kochi, A. (1992) Tuberculosis: a global overview of the situation today. *Bull WHO* 70: 149-159.

Sung, H.M., Yeamans, G., Ross, C.A., and Yasbin, R.E. (2003) Roles of YgjH and YgjW, homologs of Escherichia coli UmuC/DinB or Y superfamily of DNA polymerases, in stationary-phase mutagenesis and UV-induced mutagenesis of *Bacillus subtilis*. *J Bacteriol* 185: 2153-2160.

Tang, M., Bruck, I., Eritja, R., Turner, J., Frank, E.G., Woodgate, R., *et al.* (1998) Biochemical basis of SOS-induced mutagnesis in *Escherichia coli*: Reconstitution of *in vitro* lesion bypass dependent on the UmuD´$_2$C mutagenic complex and RecA protein. *Proc Natl Acad Sci USA* 95: 9755-9760.

Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., and Goodman, M.F. (1999) UmuD'$_2$C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc Natl Acad Sci USA* 96: 8919-8924.

Tang, T., and Asato, Y. (1978) Ultraviolet light-induction and photoreactivation of thymine dimmers in a cyanobacterium, *Anacystis nidulans. Arch Microbiol* 118: 193-197.

Tenaillon, O., Denamur, E., and Matic, I. (2004) Evolutionary significance of stress-induced mutagenesis in bacteria. *Trends Microbiol* 12: 264-270.

Tippin, B., Pham, P., and Goodman, M.F. (2004) Error-prone replication for better or worse. *Trends Microbiol* 12: 288-295.

Torkelson, J., *et al.* (1999) Genome-wide hypermutation in a subpopulation of stationary-phase cells underlies recombination-dependent adaptive mutation. *EMBO J* 16: 3303–3311.

Tyagi, J.S., and Sharma, D. (2002) *Mycobacterium smegmatis* and *tuberculosis*. *Trends Microbiol* 10: 68-69.

Wagner, J., Gruz, P., Kim, S.R., Yamada, M., Matsui, K., Fuchs, R.P., *et al.* (1999) The dinB gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol Cell* 4: 281-286.

Wagner, J., and Nohmi, T. (2000a) *Escherichia coli* DNA polymerase IV mutator activity: genetic requirements and mutational specificity. *J Bacteriol* 182: 4587-4595.

Wagner J., Fujii S., Gruz P., Nohmi T., and Fuchs R.P. (2000b) The beta clamp targets DNA polymerase IV to DNA and strongly increases its processivity. *EMBO Rep* 1:484-488.

Wagner, J., Etienne, H., Janel-Bintz, R., and Fuchs, R.P.P. (2002) Genetics of mutagenesis in *E. coli*: various combinations of translesion polymerases (Pol II, IV and V) deal with lesion/sequence context diversity. *DNA Repair* 1:159-167.

Walker, G.C. (1995) SOS-regulated proteins in translesion DNA synthesis and mutagenesis. *Trends Biochem Sci* 20: 416-420.

Walker, G.C., Smith, B.T., and Sutton, M.D. (2000) The SOS response to DNA damage. In *Bacterial Stress Responses.* (Storz, G., Hengge-Aronis, R., eds.) ASM Press. Washington, D.C. pp 131-144.

Wang, J., and Xing, Z. (2002) Tuberculosis vaccines: the past, present and future. *Expert Rev Vaccines* 1:349-354.

Weigle, J.J. (1953) Induction of mutation in a bacterial virus. *Proc Natl Acad Sci USA* 39: 628–636.

Witkin, E.M., and Kogoma, T. (1984) Involvement of the activated form of RecA protein in SOS mutagenesis and stable DNA replication in *Escherichia coli*. *Proc Natl Acad Sci USA* 81: 7539-7543.

Woodgate, R., Rajagopalan, M., Lu C., and Echols, H. (1989) UmuC mutagenesis protein of *Escherichia coli*: purification and interaction with UmuD and UmuD′. *Proc Natl Acad Sci USA* 81: 7539-7543.

Yang, W. (2003) DNA damage repair DNA polymerases Y. *Curr Opinion Struc Biol* 13: 23-30.

Yang, W. (2005) Portraits of a Y-family polymerase. *FEBS Lett* 579: 868-872.

Yasbin, R.E., Cheo, D.L., and Bayles, K.W. (1992) Inducible DNA repair and differentiation in *Bacillus subtilis*: interactions between global regulons. *Mol Microbiol* 6: 1262-1270.

Yeager, R.L., Munroe, W.G., and Dessau, F.I. (1952) Pyrazinamide (aldinamide) in the treatment of pulmonary tuberculosis. *Am Rev Tuberc* 65: 523-546.

Yeiser, B., Pepper, E.D., Goodman, M.F., and Finkel, S.E. (2002) SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci USA* 99: 8737-8741.

Zambrano, M.M., and Kolter, R. (1996) GASPing for life in stationary phase. *Cell* 86: 181-184.

Zhou, B-L., Pata, J.D., and Steitz, T.A. (2001) Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol Cell* 8: 427-437.