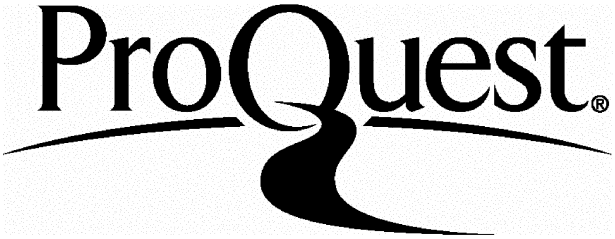
Transcription of the *Mycobacterium tuberculosis* *recA* gene

Krishna Kumar Gopaul

A thesis submitted in partial fulfilment of the requirements of University College London for the degree of Doctor of Philosophy

Division of Mycobacterial Research
National Institute for Medical Research
The Ridgeway
Mill Hill
London



ABSTRACT

The human pathogen *Mycobacterium tuberculosis* is an intracellular parasite that infects macrophages. Within these cells, the pathogen is subjected to a number of DNA damaging compounds produced as part of the host's defence mechanism. Furthermore, *M. tuberculosis* can remain in a latent but viable phase for a long period of time after primary infection and within this time DNA damage might be expected to occur. Damage to DNA affects viability and therefore it is necessary for the pathogen to possess adequate DNA repair mechanisms. The RecA protein is central to many processes of DNA repair, having a regulatory role in the expression of other DNA repair genes as well as a direct role in recombinational repair. Therefore, it is of interest to study how control of *M. tuberculosis recA* transcription is facilitated. Primer extension and transcriptional fusions to a reporter gene have previously shown that the region upstream of the *M. tuberculosis recA* gene possesses two promoters. Putative promoter motifs were identified based on known *E. coli* promoter sequences. Through direct and random base changes the promoter elements were defined. In addition, studies of both promoters in DNA damage free and induced states highlighted differences in the activity of the elements. Another approach to study the regulation of *recA* transcription was to disrupt the activity of sigma factors within the pathogen. To this end, two novel sigma factor mutants, Δ*sigC* and Δ*sigD* were created. These mutant strains were transformed with constructs carrying the two promoters and induced with a DNA damaging agent to see if there was any change in the activity of either promoter which in turn could link the sigma factor to that DNA element. In addition, the effect of the Δ*sigC* mutation on the expression of other *M. tuberculosis* genes was assessed by microarray analysis.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank the people who have helped me in my time here at NIMR. I would like to thank my supervisor, Dr. Elaine Davis for allowing me to undertake this study in her lab and also for the advice, time and support she has given me throughout this project. I would also like to thank Dr K. G. Papavinasasundaram and Dr Kristine B. Arnvig for all the help they have given me in my work. I would also like to take this opportunity to thank Dr. M. J. Colston for performing the animal work in this study and to Ms. Pat Brooks for her help in with equipment. In addition I would also like to thank Dr. Burkhard Springer for all the advice he has given..

On a personal level I'd really like to thank Sapna, Cris, Fran, Rahizan, Bosco, Bob, Jos, and Lisa for being there when things have been tough. You've all been so good to me. I'd also like to say a big thank you to Candidio who really came through for me when things were getting difficult. I just hope I can repay the favour some day! I would also like to thank Richard, Colin, Clara and Jef for being there in the beginning. Last but definitely not least, I would also like to thank my family in Grays, Essex and in Mauritius for their love and support.

I would like to dedicate this thesis to my parents and brother Ravi, who have always been my source of strength and have kept me going throughout my period of study here.

INDEX

LIST OF FIGURES FOUND IN THIS STUDY

LIST OF TABLES USED IN THIS STUDY

LIST OF ABBREVIATIONS USED IN THIS THESIS

A	Adenine
Amp^{100}	Ampicilin (100μg/ml) for *E. coli* work.
BCA	Bicinchoninic acid
BCG	Bacille Calmette-Guerin
BER	Base Excision Repair
bp	base pair
BSA	Bovine serum albumin
C	Cytosine
CAP	Catabolite Activator Protein
cfu	Colony forming unit
DCO	double crossover
dCTP	Deoxycytosine 5'-triphosphate
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoynucleoside 5'-triphosphate
DTT	Dithiothreitol
ECF	Extracytoplasmic functions
EDTA	Ethylenediaminetetraacetic acid
G	Guanine
Hyg	Hygromycin
Hyg^{50}	Hygromycin (50 μg/ml) for mycobacterial work

Hyg^R	Resistant to hygromycin
Kan	Kanamycin
Kan^{25}	Kanamycin (25 μg/ml) for mycobacterial work
Kan^R	Resistance to kanamycin
kb	kilobase pair
ko	knock-out strain
LAM	Lipoarabinomannan
mV	milli- volts
NER	Nucleotide excision repair
NO	Nitrous oxide
OD_{600}	Optical density reading at 600 nanometres
ONPG	o-Nitrophenol β-D-galactopyranoside
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
rpm	Revolutions per minute
RNA	Ribonucleic acid
RNI	Reactive Nitrogen Intermediates
ROI	Reactive Oxygen Intermediates
RT-PCR	Reverse Transcription polymerase chain reaction
SCO	single crossover
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SSC	Saline-sodium citrate buffer

TAE	Tris(hydroxymethyl)aminomethane-acetic acid- Ethylenediamine -tetraacetic acid buffer
TBE	Tris(hydroxymethyl)aminomethane-boric acid- Ethylenediamine -tetraacetic acid buffer
TBS	Tris(hydroxymethyl)aminomethane- buffered saline
TE	Tris(hydroxymethyl)aminomethane-ethylenediaminetetraacetic acid buffer
TEMED	N,N,N',N'tetramethylene-ethylene-diamine
UV	Ultraviolet radiation
wt	wild-type strain
X-Gal	4-bromo-3-chloro-2-indolyl-β-galactoside
$X\text{-}Gal^{100}$	X-Gal at 100μg/ml for selection work

CHAPTER 1: INTRODUCTION

1.1. What are mycobacteria?

The genus *Mycobacterium,* which is comprised of all 71 known mycobacterial species, belongs to the greater family of actinomycetes. It is stated that members of this group are slow-growing, rod shaped, aerobic (or micro-aerophilic) organisms that do not stain well by Gram's method (but are considered Gram-positive) (Rastogi *et al.,* 2001). With the exception of *Mycobacterium leprae*, they have a high G+C content in their DNA, which ranges between 62-70% (Rastogi *et al.,* 2001). They do not form endospores, conidia or capsules and do not have aerial hyphae (Rastogi *et al.,* 2001). However, their most distinguishable feature is that these organisms can retain stain when challenged with acidified alcohol, a characteristic that occurs due to the composition of the bacterial cell wall (Rastogi *et al.,* 2001).

1.2 The mycobacterial cell wall.

The walls of mycobacteria are lipid rich and contain the compound mycolic acid. It has been shown that mycolic acid interacts and tightly binds the stain, which becomes inaccessible to the action of acid de-stain. However, this compound is not unique to this genus and can also be found in the genera *Nocardia* and *Rhodococcus*, which are also part of the actinomycetes group (Rastogi *et al.,* 2001). Indeed, these genera are also acid fast but, unlike *Mycobacteriae,* are still decolourised by the action of alcohol. However, what differentiates mycobacterial mycolic acids from those found in the other two genera is the

length of the carbon chains involved in these acids. In the mycobacterial version these chains are generally longer having between 60-90 carbon atoms in comparison with the 34-60 carbon atoms seen in the backbone forming these acids in the other two species (Rastogi *et al.*, 2001). In addition, the mycobacterial mycolic acids are unsaturated, with there always being between 1 and 3 double bonds present, which is usually not the case in the mycolic acids found in the other two groups. Furthermore when subjected to pyrolysis, the fatty esters released from mycobacterial mycolic acids are much longer than those produced from the other genera (Rastogi *et al.*, 2001).

There are other important components that make up the cell wall in mycobacteria. The cell wall itself is made up of four layers that lie over the cell membrane (Grange, 1996). The first layer, directly above the cell membrane is made up of peptidoglycan, which is closely related to that found in the cell walls of other bacteria (Grange, 1996). A layer of arabinogalactan lies on top of the first layer, with the side chains of this polysaccharide linking with mycolic acids which are found in the third layer of the wall (Grange, 1996). The final layer is made up a variety of lipids which are collectively referred to as mycosides (Grange, 1996). In addition, another component of the cell wall is lipoarabinomannan (LAM) which has its lipid part attached to the cell membrane and its polysaccharide section extended through the other four layers of the cell wall (Grange, 1996). It has been suggested that the role of this molecule is to attach the four layers of the cell wall to the cell membrane (Grange, 1996). There are two forms of LAM present in the mycobacterial cell wall, which are differentiated by whether their side chains end with either arabinose or mannose (Grange, 1996). Of these two types, the LAM with side chains terminating in

arabinose has been shown to be able to elicit an immune response involving compounds such as tumour necrosis factor (TNF) (Roach *et al.,* 1993).

1.3. Mycobacterial grouping

All mycobacteria are slow growing in comparison to other bacterial species. It can take between 2 days and 8 weeks incubation on solid media to see colonies (in the case of *Escherichia coli* it is in the magnitude of hours rather than days). However, one means of sub-grouping mycobacteria is through growth rate. Classification on this point is given by Bergey's manual as follows: fast growers are those that take up to one week to produce visible colonies on a plate whilst slow growers take much longer (Goodfellow *et al.,* 1986).

Examples of slow growers are *M. tuberculosis* (the causative agent of pulmonary tuberculosis) and *M. bovis* (the causative agent of bovine tuberculosis), whilst *M. smegmatis* and *M. fortuitum* (both saprophytes, not known to cause disease) are typical fast growers (Grange, 1996). An interesting note is that although *M. leprae* has also been classified as a slow grower, this species is actually non-cultureable *in vitro* and can only be isolated from infected mouse footpads or from the livers of the nine-banded armadillo.

Determination of individual mycobacterial species can also be made through comparison of 16S RNA but this also has its problems as shown by Kirschner *et al.,* (1993) who showed that this method could not distinguish between the two different species, *Mycobacterium marinum* and *Mycobacterium ulcerans* (Kirschner *et al.,* 1993).

1.4. Genome sequences of mycobacteria

The study of mycobacteria has been aided greatly with the advent of genomic sequencing. The full genomic sequence of *M. tuberculosis* H37Rv was published in 1998 and showed that this pathogen had around 4000 genes encoded in 4441 kb of DNA with a G+C content of 65.6% (Cole *et al.*, 1998). Of these genes, around 84% have been assigned a function based on either homology with genes of other species with specific functions or through previous studies, leaving 16% of the complete genome with an unknown function (Cole *et al.*, 1988). Around 350 genes are involved in both the metabolism and synthesis of fatty acids (Cole *et al.*, 1998; Domenech *et al.*, 2001). This number is far greater than that found in *E. coli* and is a reflection of the fact that the cell envelope of *M. tuberculosis* is lipid rich (Domenech *et al.*, 2001).

Comparative genomics between *M. tuberculosis* and the vaccine strain *M. bovis* BCG (which is also part of the *M. tuberculosis* complex), showed that there were seven deleted regions in the avirulent species, implying the necessity of genes in these regions for pathogenicity (Domenech *et al.*, 2001).

Since the determination of the complete *M. tuberculosis* genome sequence, the complete sequence of *M. leprae* has become available (Cole *et al.*, 2001), whilst the sequencing of the saprophyte, *M. smegmatis* is nearly completed. In the case of the *M. leprae* sequence project, it was found that the genome of this pathogen was smaller than that of *M. tuberculosis* at a size of 3268 kb, and less G-C rich at a value of 57.8% (Cole *et al.*, 2001). It has only a possible 1604 genes accounting for about 49% of the entire genome (Cole *et*

al., 2001). The reason that this pathogen contained far fewer genes than *M. tuberculosis* was that *M. leprae* contained 1116 pseudo-genes (Cole *et al.,* 2001). These are sequences homologous to open reading frames (ORFs) present in an active form in *M. tuberculosis*, which are inactivated by multiple mutations in *M. leprae* (Cole *et al.,* 2001). The presence of pseudogenes accounts for around 27% of the entire genome. In addition, there is about 23% of the genome that is made up of gene deletions and regulatory sequences. Of the functional genes present in the pathogen, 89.7% are common to both *M. leprae* and *M. tuberculosis.*

1.5. Mutational analysis of *M. tuberculosis*

The knowledge gained form sequencing *M. tuberculosis* has helped determine the presence or absence of genes with a known function through homology. However, ascertaining the actual role of these genes in the viability of the pathogen is best achieved by knocking out the gene of interest and seeing what effect this has on the viability of the mutant in comparison with the wild-type. Allelic exchange, whereby a copy of the gene containing a deletion is exchanged for the wild-type copy through homologous recombination, is a common method to create knock out mutants in many types of bacteria (Baliko and Venetianer, 1993). However, in the case of *M. tuberculosis* the event of homologous recombination is not as efficient as it is in other species such as *E. coli.* In a study where linear fragments of DNA between 4 and 7 kb long were transformed into both *M. tuberculosis* and *M. bovis* BCG, it was found that these fragments were incorporated readily into the genome (Kalpana *et al.,* 1991). However, the majority of these incorporations were through illegitimate recombination events (Kalpana *et al.,* 1991).

Nevertheless, in a later study it was shown that homologous recombination to create a knock out could partially be achieved in *M. bovis* BCG with a DNA fragment of around 5kb (Aldovini *et al.,* 1993). In this study it was shown that a single crossover in the correct position could be achieved in an attempt to knock out the *uraA* gene (Aldovini *et al.,* 1993). In this paper it was argued there were a number of possible reasons for the failure of Kalpana *et al* to derive knock out mutants with their procedure. These possibilities included that the gene to be knocked out might be essential in the pathogen or that there was much non-homologous vector DNA present in the construct that could affect homologous recombination (Aldovini *et al.,* 1993). However, with only 2 colonies out of 10 tested having the single crossover and with no double crossover being obtained, it was clear that this procedure could be further refined. A later study then showed that the rate of homologous recombination leading to gene replacement could be facilitated in *M. tuberculosis* if the region of DNA introduced was much longer, at around 40kb (Balasubramanian *et al.,* 1996). Another study undertaken in that same year showed that the size of this fragment could be substantially reduced to achieve the same task, when it was shown that the *mas* gene in *M. bovis* BCG could be replaced by a DNA fragment of 7.9 kb which included a hygromycin resistance marker (Azad *et al.,* 1996). In this study the efficiency of homologous recombination mediated gene exchange was poor with only 1 knock out strain isolated from 38 potential mutant colonies screened (Azad *et al.,* 1996).

In the previously mentioned systems, selection for a possible mutant was based on a one step process. After transformation, there was a direct selection for the mutant. However in a study done by Pelicic *et al* (1997) it was found that a more efficient means of allelic exchange could be achieved if a two-step selection process was followed. The first

selection would be for the integration of the mutant construct into the chromosome, whilst the second would be for the loss of the plasmid used to introduce the mutation, with the hopeful outcome being the disruption of the gene. The plasmid used in this allelic exchange study was temperature sensitive, replicating in mycobacteria at 32°C with a non-permissive temperature of 39°C. It also contained the *B. subtilis sacB* gene, which was shown to convey sucrose sensitivity to mycobacteria, and a *xylE* gene, which makes mycobacteria turn yellow in the presence of catechol. The gene to be disrupted, which was present on a 2.5kb DNA fragment, was marked with the kanamycin resistance cassette. First selection with the permissive temperature on plates with kanamycin gave rise to colonies which had acquired the plasmid. Re-plating at the non-permissive temperature on plates with kanamycin and selection for sucrose-resistant white colonies gave rise to 200 potential mutants (Pelicic *et al.,* 1997). Several colonies were then screened by Southern blot for the allelic exchange event and all were found to be positive for this knock out, proving the efficiency of the system (Pelicic *et al.,* 1997).

Another method that has been use to undertake allelic exchange in mycobacteria is through the use of streptomycin resistant strains of various mycobacterial species through a mutation in the *rpsL* gene (Sander *et al.,* 1995; Sander *et al.,* 2001; Springer *et al.,*2001). There is a copy of the wild-type *rpsL* gene on the suicide vector used in this procedure, in addition to the mutated gene, which is marked with the *aph* gene to confer kanamycin resistance, and flanking sequence. The basis of this technique is that transformation and incorporation of the wild-type *rpsL* gene can convert the mutant strain from streptomycin resistant to streptomycin sensitive (Sander *et al.,* 1995; Sander *et al.,* 2001; Springer *et al.,*2001). Single crossovers were selected for their resistance to kanamycin whilst the

second crossovers were selected for their resistance to streptomycin (Sander *et al.*, 2001; Springer *et al.*,2001). This technique has been used successfully to delete the *recA* gene in both *M. bovis* and *M. tuberculosis* (Sander *et al.*, 2001; P. Sander, personal communication) as well as the *M. tuberculosis ahpC* gene (Springer *et al.*, 2001).

Since these studies, most allelic exchange strategies in *M. tuberculosis* have made use of a two-step selection process. The temperature sensitive selection procedure used by Pelicic *et al* has proved problematic and more recently suicide vectors have been utilised to carry the mutated gene construct (Parish and Stoker, 2000). In addition, it has been shown that the efficiency of homologous recombination can be improved if the construct DNA has been treated with either UV or alkali (Hinds *et al.*, 1999). A probable reason for this observation was that DNA repair mechanisms and particularly RecA, which is the main protein associated with recombination, were activated with the presence of this "damaged DNA" (Hinds *et al.*, 1999).

1.6. The use of micro-arrays to determine gene function

With the elucidation of the *M. tuberculosis* H37Rv genome sequence and the ability to undertake allelic exchange to knock out genes, it has become possible to see what role an individual gene has in global gene expression in the pathogen using a micro-array. The micro-array is a glass slide onto which around 4000 DNA probes are spotted, each of which is specific for one gene in the *M. tuberculosis* genome. Both mutant and parental wild-type bacteria are subjected to the same growing conditions and then RNA is extracted from both and converted to cDNA, which is tagged with one of two fluorescent dyes to differentiate

between mutant and wild-type products. The labelled products are then co-hybridised on a micro-array slide and analysed. In examples where gene activity in the mutant is the same as the wild-type, there should be equal concentrations of both dyes at those particular spots. However, where there is differential expression between the mutant and wild-type in one particular gene, then there is a bias for one dye depending on which strain has the higher expression level. An example of a study that was undertaken in this way was the analysis of a *M. tuberculosis* σ^E mutant under normal conditions and following exposure to SDS (Manganelli *et al.*, 2001). Using this technique it was possible to identify 23 genes that were regulated either directly or indirectly by the sigma factor with exposure to SDS (Manganelli *et al.*, 2001).

1.7. *M. tuberculosis* infection and host response.

Pulmonary tuberculosis is one of the oldest diseases known to man and cases have been documented as far back as the time of the Romans. In one of his texts, Celsus the medical historian gave the classic characteristics of the disease. He stated " the malady usually arises in the head, thence it drips into the lung; there ulceration supervenes, from this a slight feverishness is produced, which even after it has become quiescent nevertheless returns; there is frequent cough, pus is expectorated, sometimes blood-stained" (Celsus, 1st Century AD). However, until Koch's experiments in 1892 with infected tissue from tuberculosis, where he managed to isolate the bacillus, *M. tuberculosis*, it was thought that a virus caused this disease. Generally, the immune response mounted against *M. tuberculosis* infection is sufficient to control but not eliminate the pathogen. Evidence for this is the fact that although 1.5 million people die annually from tuberculosis, it is

estimated that about one third of the world's population is infected with *M. tuberculosis* (Flynn and Chan, 2001b).

1.7.1. Infection of the host with the pathogen

Infection with *M. tuberculosis* can occur through many pathways as demonstrated by Koch in his animal work of 1892 and more recently in human samples (Scott-Wilson, 1962). However, the primary route of infection in humans comes through the respiratory tract as an aerosol. Upon entering the respiratory tract, the pathogen is intercepted and engulfed by alveolar macrophages (Russell, 2001). These infected macrophages initiate congregation with other cells involved in the immune response to form a granuloma, which contains and isolates the infection from the rest of the body (Russell, 2001). However, it is known that the pathogen may not be eliminated in the granuloma and *M. tuberculosis* is able to persist in these structures in a latent phase (Flynn and Chan, 2001a). Indeed, it has been shown in the frog pathogen, *Mycobacterium marinum*, that there are genes present in this mycobacterial species that are specifically activated in granulomas and which play a role in the pathogen's long term survival in this structure (Chan *et al.*, 2002). Latency of *M. tuberculosis* in the host differs depending on what organism is infected. Although mice can develop a chronic infection, this is dependent on a high initial dose of the pathogen, and the infection is ultimately lethal (Manabe and Bishai, 2000). In the case of human infection, latency is coupled with low amounts of normally unculturable bacilli and this state can be maintained over the lifetime of an infected patient (Manabe and Bishai, 2000; Flynn and Chan, 2001a). It has been shown that tuberculosis lesions in patients with latent

tuberculosis infections rarely contain acid-fast bacteria but are still capable of causing infection in a guinea pig model (Manabe and Bishai, 2000).

The mechanisms for latent infections of *M. tuberculosis* have yet to be confirmed but there are a number of possible avenues. One such possibility involves the transcription factor, *sigF* (also mentioned under *M. tuberculosis* sigma factors). It was found that although initially infection of mice with wild-type *M. tuberculosis* and a Δ*sigF* mutant proceeded without difference, after a period of eight weeks post infection the deletion mutant was not able to persist in the host as well as the wild-type (Chen *et al.,* 2000). It is at around this time that it is thought that granulomas are formed and the absence of the mutant at this time indicates that σ^F may have a role in persistence (Manabe and Bishai, 2000). Another possible means of *M. tuberculosis* entering and maintaining a state of latency is through a modulation of metabolism (Wayne and Hayes, 1996). In a granuloma, the pathogen is thought to reside in an environment that is poorly aerated and depletion of oxygen has been shown to activate the glyoxylate shunt pathway *in vitro* (Wayne and Lin, 1982; McKinney *et al.,* 2001). Evidence to support the role of the glyoxylate shunt pathway in latency has been shown in the *M. tuberculosis icl* mutant, which is defective in persistence (McKinney *et al.,* 2001).

Reversion to an active infection can be initiated if the immune system becomes compromised (Flynn and Chan, 2001a; Russell, 2001). When this occurs, the centre of the granuloma degenerates (caseation), causing a cavity in the lung wall and freeing the pathogen (in the form of a pus) into an airway, where it can become transferred to others in

the form of an aerosol (Russell, 2001). Pus formed in the initial infection can also cause infection of other parts of the lung, larynx and intestine (Keers and Rigden, 1953)

In addition to secreting granoloma-forming signals, infected macrophages also secrete the cytokine interleukin (Il)-12 (Ladel *et al.,* 1997). This signal acts upon T-cells and natural killer cells to make the cytokine, interferon gamma (IFN-γ) (Ladel *et al.,* 1997; Sano *et al.,* 1999; Flynn and Chan, 2001). In turn, IFN-γ stimulates a T-helper (Th) 1- based response, a response suited for intracellular pathogens, whereby IFN-γ can activate other macrophages (Flesch and Kaufmann, 1990; Sano *et al.,* 1999; Flynn and Chan, 2001b). It has also been shown that another cytokine, tumour necrosis factor (TNF)- α, can influence and enhance the activity of IFN-γ (Flesch and Kaufmann, 1990; Bekker *et al.,* 2001; Flynn and Chan, 2001b).

1.7.2. The killing mechanism of the host against the pathogen.

Activation of macrophages leads to the production of both reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI) (Ding *et al.,* 1988). Both these compounds have been shown to have antimycobacterial activity (Jackett *et al.,* 1978; Chan *et al.,* 1992). However, there is some confusion as to which mechanism is favoured to deal with the pathogen.

Initially, it was thought that the antimycobacterial activity mediated by activated macrophages was due solely to an oxidative mechanism. Mouse bone marrow macrophages, when activated with *M. bovis* stimulated spleen cell factor, had an enhanced rate of killing intracellular *M. microti* (part of the *M. tuberculosis* complex) in comparison

with untreated cells, mediated by a peroxidative killing system (Jackett *et al.,* 1978; Walker and Lowrie, 1981). By adding catalase to the stimulated macrophages and then infecting them this enhanced killing was lost, proving that the antimycobacterial activity was due to the production of hydrogen peroxide (Walker and Lowrie, 1981). In addtion, *M. tuberculosis* contains the glycolipid lipoarabinomannan (LAM), which has been shown to "scavenge" cytotoxic oxygen free radicals within macrophages hence limiting the threat of ROI mediated damage (Chan *et al.,* 1991). This finding indicates that the pathogen is already equipped to deal with ROI. Furthermore, although there is a strong RNI response mediated by mouse macrophages upon infection *in vitro* the same is not true with human macrophages (Rockett *et al.,* 1998). However a possible reason for not seeing any response in human macrophages was that these immune cells were not activated with IFN-γ.

There is also evidence to suggest that antimycobacterial activity in humans can be mediated through RNI. It was observed in patients with a genetic defect leading to the inability to produce NADPH synthase (Chronic Granuloma Disease), and therefore unable to produce ROI, that they were still able to control a tuberculosis infection (Jones *et al.,* 1990). This observation has been questioned by some who have shown that CGD patients infected with *M. tuberculosis* have an increased susceptibility to the pathogen in comparison with people able to mount an ROI response (Lau *et al.,* 1998). However, it has been shown that peripheral blood mononuclear cells infected with both virulent and avirulent *M. tuberculosis* have increased concentrations of nitrous oxides within them (Kwon, 1997).

More evidence in support of the idea that antimycobacterial activity in humans is mediated by RNI comes from a study done where cultures of *M. tuberculosis* H37Rv were stressed with superoxide generating compounds or RNI and protein synthesis observed with 2-D gel analysis (Garbe *et al.,* 1996). It was found in this study that there were a number of

proteins produced specifically by the pathogen in response to nitrous oxide (Garbe *et al.,* 1996). In this respect there is a difference in response between what is seen in *M. tuberculosis* and *E. coli.* The *soxRS* system in the enteric bacterium, in addition to being active under superoxide stress, is also activated in the presence of nitrous oxide (Nunoshiba *et al.,* 1993). Conversely, it was found that proteins made by the pathogen in oxidative stress were actually heat shock proteins, implying a type of cross protection mechanism (Garbe *et al.,* 1996). Furthermore, in *M. tuberculosis* there is no equivalent of the *soxR* or *soxS* gene (Garbe *et al.,* 1996).

Given all the evidence backing both ROI and RNI mediated antimycobacterial activity, it is hard to determine which response is the main effector of killing. The presence of genes to deal with both RNIs and ROIs indicate that both play a part and may work synergistically to attempt clearance of the pathogen.

1.8. The DNA repair responses

The environment in which the *M. tuberculosis* cells reside in the host, namely within the macrophage, is very hostile to the pathogen. Oxidative stress produced through reactive oxygen intermediates (ROI) and from nitrous oxide (NO) can give rise to DNA damage. Nicks or lesions within DNA can halt replication that, in turn would kill the pathogen. It is, therefore, necessary to overcome this problem with adequate repair mechanisms. Study of these systems in *M. tuberculosis* is still ongoing, but an indication of the processes involved can be found in the study of DNA repair mechanisms in *E.coli.* In addition, information from sequencing the pathogen has located a number of genes that are important in DNA

damage reversal and repair (Mizrahi and Andersen, 1998). There are three main types of repair present within bacteria that deal with damage caused by nicks or lesions in DNA.

Nucleotide excision repair (NER) involves the proteins UvrA, B, C, and D and Pol I. A complex consisting of a dimer of UvrA protein and the UvrB protein can detect disruptions in the DNA double helix, which in turn, indicates a lesion. The complex binds to the DNA then separates the DNA strands at the site of the lesion. The UvrA dimer then dissociates from the complex leaving UvrB bound to the DNA. The protein UvrC then binds to the UvrB-DNA complex and promotes the excision of a small fragment of the single stranded DNA surrounding the lesion (Lin and Sancar, 1992; Moolenaar *et al.*, 2002). The action of UvrD (a helicase) releases the fragment and allows synthesis of DNA by Pol I to fill the gap excised. The final step of ligation is undertaken by DNA ligase (Volkert and Landini, 2001). A schematic of this type of repair is shown in figure 1.1. The genes *uvrA*, *uvrB* and *uvrD* are part of the SOS response (mentioned later) and are induced by DNA damage in *E. coli*, whereas the *uvrC* gene is not (Van Houten *et al.*, 2002). However, a recent study in *E. coli* has shown that there is a homologue of *uvrC*, called *cho* (standing for uvr**C** **ho**mologue) also known as *ydjQ*, whose product has been shown to have nearly an identical role as UvrC but in addition is DNA damage inducible (Moolenar *et al.*, 2002). It was found that Cho was able to excise certain DNA lesions better than UvrC but could not fully compensate for the lack of *uvrC* in a Δ*uvrC* mutant (Van Houten *et al.*, 2002). A version of this gene exists in *M. tuberculosis* along with a copy of *uvrC* (Van Houten *et al.*, 2002). However, it has been found that the mycobacterial equivalent protein is longer with an extra domain that has been suggested to have a proofreading function (Van Houten *et al.*, 2002).

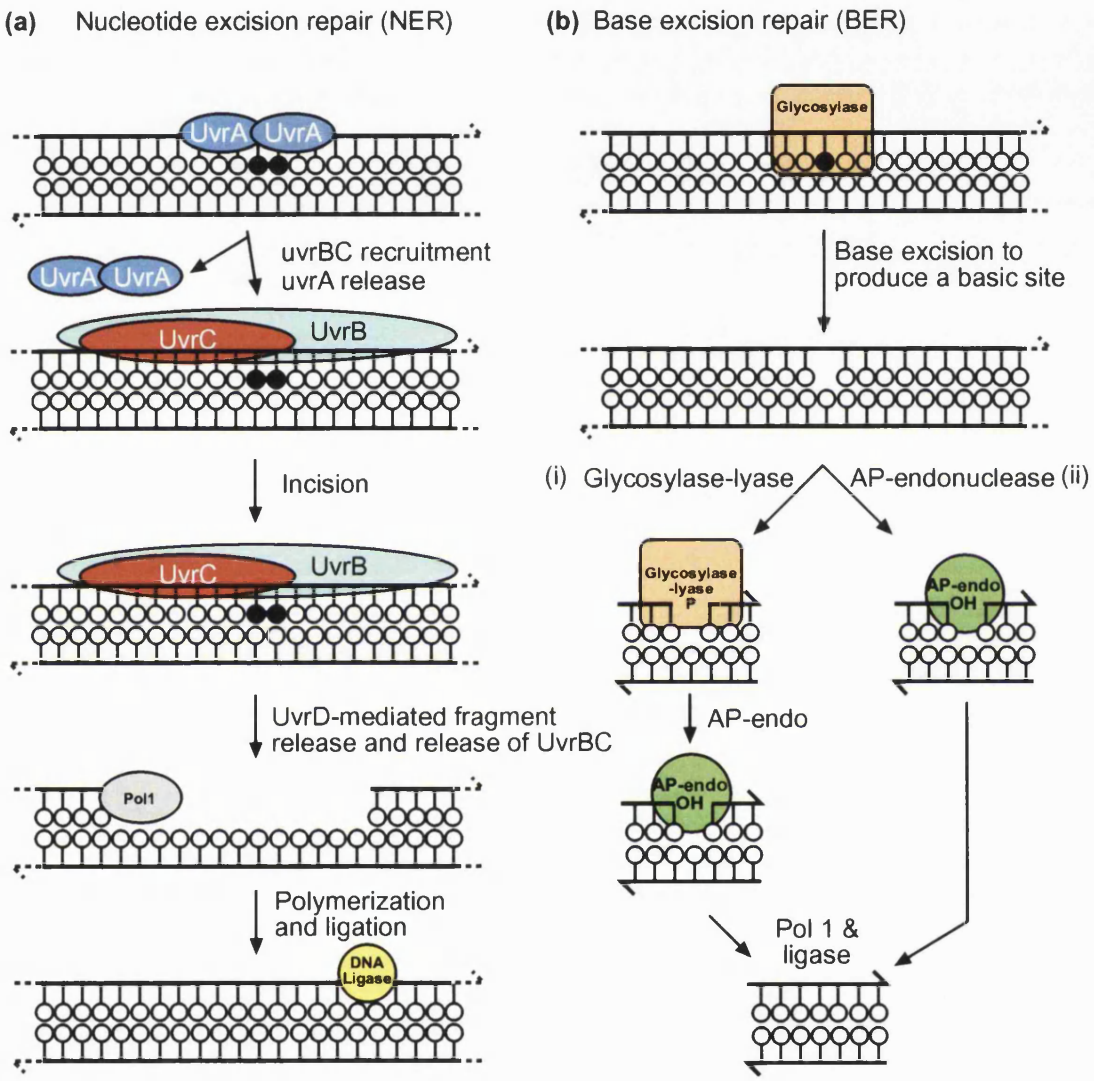


Figure 1.1. These cartoons illustrate the mechanics how nucleotide excision repair (NER- LEFT) and base excision repair (BER- RIGHT) which are involved in DNA repair, function. This diagram is taken from Volkert and Landini (2001).

Base Excision Repair (BER) involves a glycosylase protein, an example being Fpg (also known as MutM) which binds and removes the damaged base by cleaving the glycosylic bond from the sugar phosphate backbone forming an apurinic/apyrimidinic site (AP) (Lindahl *et al.,* 1988). Under the action of 5' AP endonucleases, an example being XthA, the DNA phosphodiester backbone that is at the 5' end of the AP site is cleaved to create a free 3'-OH end for DNA polymerase repair synthesis (Mol *et al.,* 2000). This type of repair is also associated with alkylation of DNA with its mediation in *E. coli* coming from the adaptive response. A schematic of this type of repair is shown in figure 1.1

Recombination repair, also known as post-replication repair, is initiated by the protein RecA (West *et al.,* 1981; Webb *et al.,* 1997). This protein is also involved in the SOS response, which will be discussed later. Upon DNA replication through a lesion, the DNA replication apparatus stalls, giving rise to single stranded DNA (Webb *et al.,* 1997). In the case of a double stranded break, the action of another protein complex, RecBCD, unwinds and degrades the DNA in a 3' →5' action simultaneously from the site of the break until it reaches a Chi site which stalls the complex (Anderson and Kowalczykowski, 1998). The ssDNA generated by either means is coated in ssDNA binding protein (SSB) (Kowalcsykowski and Krupp, 1987; Webb *et al.,* 1997). The binding of SSB to ssDNA leads to the removal of secondary structures in the DNA that could impede the action of RecA (Kowalczykowski and Krupp, 1987). However, this binding is also of hindrance to RecA as this protein stops RecA from undergoing the nucleation event that allows the RecA filaments to coat the DNA (Shan *et al.,* 1997). However, two other proteins, RecO and RecR can bind the SSB-ssDNA complex, modifying but not displacing the complex

structure and allowing the binding of RecA (Umezu and Kolodner, 1994). Once RecA has a "foothold" on the ssDNA, its extension in a 5' to 3' fashion displaces SSB on the ssDNA in an ATP dependent manner (Kowalczykowski and Krupp, 1987; Shan *et al.*, 1997). The binding and filamentation of RecA is controlled by the action of both RecF and RecO, which both act as a complex to stop extension past the single-stranded region (Webb *et al.*, 1997). The function of the RecA bound to the ssDNA is to align this DNA to homologous sequence in double stranded DNA and allow strand exchange to occur thus forming a Holliday junction (Lindsley and Cox, 1990; Webb *et al.*, 1997). It has also been shown that in DNA repair, RuvA and RuvB can act with RecA but not alone to allow strand exchange to occur over long gaps in DNA (Iype *et al.*, 1994). Branch migration is then facilitated by the action of RuvA and RuvB, which act together as a complex (Shiba *et al.*, 1991) or by RecG (West, 1994). With the completion of the extension of the new strand of DNA, the recombination event is ended through the resolving of the Holliday junction by the action of RuvC, which, as a dimer, nicks strands of like polarity (West, 1994). The action of DNA ligase then repairs this nick (West, 1994).

All the above repair systems themselves rely upon control mechanisms that respond to DNA damage. Two major control mechanisms are the SOS response (fig 1.2) and the adaptive response.

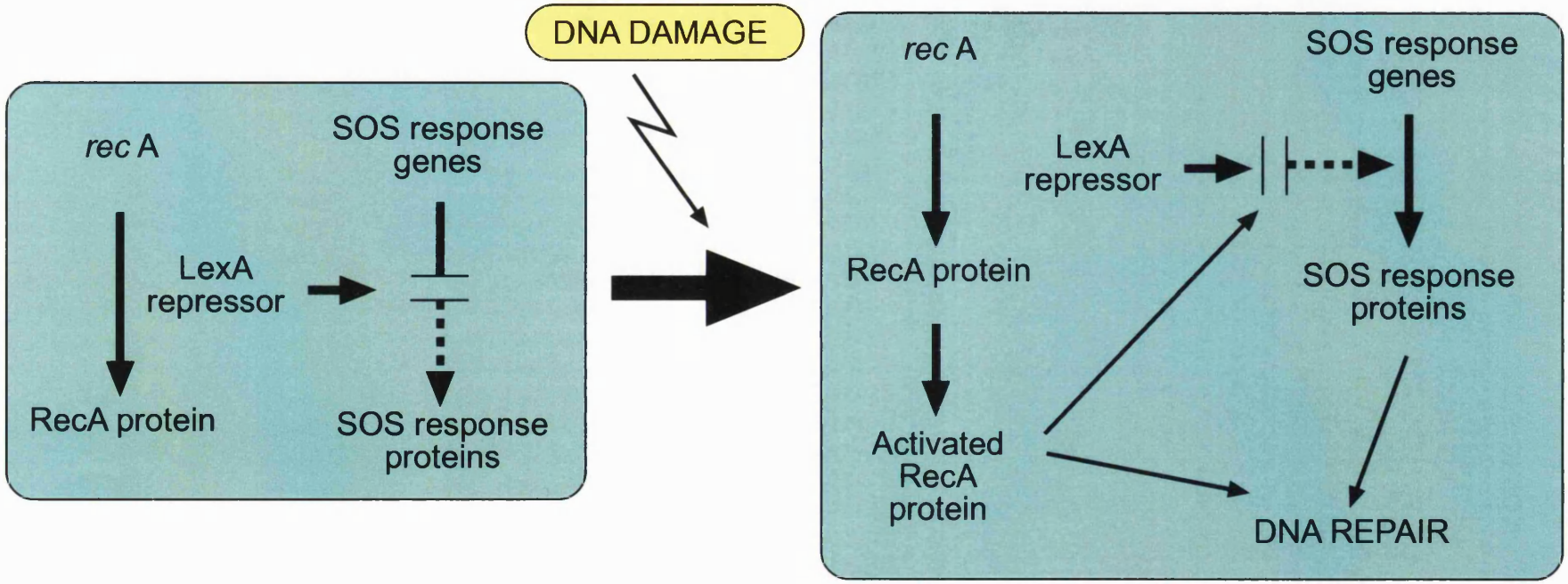


Figure 1.2. This schematic illustrates how the SOS response is thought to function. In times of DNA damage, a signal (the binding of RecA to ssDNA) converts RecA to RecA* which can then promote autocatalytic cleavage of the LexA repressor, allowing the expression of SOS response genes. This diagram is taken from Colston and Davis (1994).

1.8.1. The SOS response

The SOS response is the best studied DNA repair system present in *E. coli* and is illustrated in fig.1.2. The term "SOS" has often been misinterpreted as indicating that this mechanism is a last ditch effort from the cell to repair DNA damage and avert senescence. However, what is supposed to be suggested is that this response becomes activated upon a "distress signal", namely single-stranded DNA (ssDNA) (Radman, 1975; Gudas and Pardee *al.*, 1975).

The response itself is made up of over 30 genes that are under the regulation of two proteins; the repressor LexA and the activator RecA (Fernandez de Henestrosa *et al.*, 2000). Upstream of the SOS genes (including *recA* and *lexA*) is a palindromic sequence called the SOS box, to which LexA binds and represses transcription under normal, damage free conditions (Wertman and Mount, 1985). However, the binding affinity LexA has for the SOS box is variable depending on its precise sequence (Courcelle *et al.*, 2001). The repressor has a high affinity for the SOS boxes upstream of genes such as *sulA*, which are tightly regulated. However, in the case of genes such as *recA*, the SOS boxes upstream of their coding regions have a weaker affinity for LexA, which allows basal transcription (Walker, 1984). This form of repression allows genes that have a function other than in the SOS response to be expressed for these tasks, whilst genes solely required for DNA damage, which can be harmful to the cell under normal conditions, are expressed only when needed.

When DNA damage occurs, regions of ssDNA arise to which RecA present in the cell binds. This RecA forms filaments on the ssDNA and becomes active in the process. This active form (RecA*) can then facilitate autocatalytic cleavage of LexA (Little *et al.,* 1980). Self-cleavage of LexA can also occur at high pH but RecA* accelerates the process at intracellular pH values. The repressor is cleaved at a proteolytically sensitive hinge region, which separates the two domains of the protein. As LexA becomes cleaved, there is not enough "whole" repressor to bind SOS boxes and, therefore, genes with upstream regions that have low affinities for LexA become free to be expressed at a higher level (Friedberg *et al.,* 1995). Amongst this group of expressed genes is *lexA* itself, which in turn leads to the production of more LexA protein (Little *et al.,* 1980). However, whilst ssDNA is still exposed, RecA will still become RecA* and in turn, the newly synthesised LexA will still be cleaved. As this happens, genes with SOS boxes with higher affinities for the repressor will now become expressed and aid in repair (Friedberg *et al.,* 1995). This cascade of expression continues with the continuing presence of DNA damage.

Some genes that are associated with both NER and recombinational repair are also part of the SOS system and become active with the presence of RecA* (Friedberg *et al.,* 1995; Anderson and Kowalczykoski, 1998). When the DNA becomes repaired, the signal that activated RecA has been removed. This in turn means that the LexA synthesised can now accumulate and rebind SOS boxes, again based on affinity for individual sequence (Little *et al.,* 1980).

Although the SOS response was first characterised in *E. coli,* studies since then have showed that many bacteria, both Gram-negative and Gram-positive, have some kind of

comparable system. In the case of *B. subtilis* (the most studied Gram-positive) the LexA homologue is termed DinR (Winterling *et al.,* 1997). The consensus sequence to which this protein binds is different from that of *E. coli* and has been referred to as the Cheo box (Cheo *et al.,* 1991). However, this repressor protein also undergoes autocatalytic cleavage in the same way as LexA under the action of RecA* (Winterling *et al.,* 1997). In addition, it was found that the *E. coli* RecA protein could complement a *B. subtilis recA* mutant (Wojciechowski *et al.,* 1991). There is also evidence to suggest that a third type of LexA-like repressor and binding site may be utilised in certain bacteria which is incompatible with the *E. coli* SOS box (Riera *et al.,* 1994). It was shown that the *recA* genes present in three bacterial species belong to the same subset of the Gram-negative Proteobacteria class, although lacking the conventional SOS box, were still inducible with mitomycin C (Riera, 1994). In addition, regulation of each *recA* gene could be facilitated in all three species, showing a common pathway (Riera *et al.,* 1994). Even so, it was also shown that the *E. coli recA* gene could not be induced in the three proteobacteria or vice-versa (Riera *et al.,* 1994) showing that the system is different from the SOS response in enteric Gram-negatives.

However, there are instances where *recA* is not regulated as part of the SOS response. The oral pathogen *Porphyromonas gingivalis* was shown to possess a *recA* gene homologue that showed a high degree of similarity to *recA* genes from other Gram-negative bacteria. Through the construction of a *recA* deletion mutant, it was shown that this gene was implicated in DNA damage repair within this species (Fletcher *et al.,* 1997). However, unlike the other Gram-negative homologues, there was no recognisable LexA binding site upstream of this gene (Fletcher *et al.,* 1997). Furthermore, it was shown that the *P.*

gingivalis recA promoter did not have inducible activity upon the addition of DNA damaging agents such as mitomycin C and nalidixic acid (Liu and Fletcher, 2001). However, the explanation given for this observation was that the "basal" activity of the promoter had to be high to cope with the harsh environment which the organism colonises (Liu and Fletcher, 2001). Interestingly, it was also shown in this study that the *recA* promoter could be stimulated with high concentrations of both calcium and iron (Liu and Fletcher, 2001).

In addition, there are other Gram-negative examples such as *Neisseria gonorrhoeae* (Black *et al.,* 1998), *Thiobacillus ferooxidans* (Ramesar *et al.,* 1989) and *Bacteriodes fragilis* (Goodman *et al.,* 1990) which all possess *recA* genes lacking an apparent upstream LexA binding site and which are also not induced by DNA damage.

1.8.1.1. The SOS response in mycobacteria

The first mycobacterial version of the *recA* gene was isolated by Davis *et al.*(1991) from *M. tuberculosis* by probing the genome with the *E.coli recA* gene. However, it was found that the homology this gene shared with the *E. coli* version was located in two distinct regions at either end, which were separated by 1320bp of sequence that encoded an intein (Davis *et al.,* 1991). Inserts of this nature had previously been found only in the yeast *Saccharomyces cerevisiae* (Kane *et al.,* 1990). This type of element is excised from the precursor protein with concomitant re-ligation of the flanking regions to form the mature protein (Perler *et al.,* 1994). When the *recA* genes of 17 other mycobacteria were examined, inteins were only found in *M. bovis* (part of the *M. tuberculosis* group) and *M. leprae*, which are both

slow growing pathogenic bacteria (Davis *et al.,* 1994). However, it must be noted that the inteins found in *M. tuberculosis* and *M. leprae* do not share sequence homology (Davis *et al.,* 1994). Subsequently, inteins similar to the *M. leprae* intein have been found in the *recA* genes of 5 other mycobacterial species (Saves *et al.,* 2000).

Although it was possible for *M. tuberculosis* RecA to be processed through protein splicing in *E. coli*, the same was not true in the case of the *M. leprae* product (Davis *et al.,* 1994). Nevertheless, it was shown that splicing of the *M. leprae* intein was achievable in *M. smegmatis* (Frischkorn *et al.,* 2000). It has been suggested that the poor efficiency of homologous recombination in slow growing mycobacteria can be attributed in part to the presence of the intein (McFadden, 1996). However, studies have shown that the presence of this intein does not affect RecA function (Frischkorn *et al.,* 1998; Papavinasasundaram *et al.,* 1998).

Primer extension studies into the sequence upstream of the *M. tuberculosis recA* gene suggested that it contained two regions of promoter activity (Movahedzadeh *et al.,* 1997a). This initial finding was confirmed by fusing these regions to a promoterless *lacZ* reporter gene (E. O. Davis, personal communication). These two regions have been termed P1 (that closer to the *recA* gene) and P2 (that further upstream). In addition, there is a mycobacterial SOS box located between the –10 and –35 regions of P2 (Movahedzadeh *et al.,* 1997a).

It has been shown that in both *M. tuberculosis* and *M. smegmatis* induction of *recA* can be achieved with the addition of the DNA damaging agent, mitomycin C (Movahedzadeh *et al.,* 1997a; Papavinasasundaram *et al.,* 1997; Durbach *et al.,* 1997). The RecA proteins of

M. tuberculosis, *M. leprae*, and *M. smegmatis* are highly conserved with a 91% identity (Papavinasasundaram *et al.*, 1997). In addition, the three mycobacterial species share the presence of another gene, *recX*, which is co-transcribed with *recA* (Papavinasasundaram *et al.*, 1997). It has been shown in a *M. smegmatis ΔrecA* strain, where *recX* expression was also prevented, that over-expression of RecA from a replicating vector was toxic (Papavinasasundaram *et al.*, 1998). This toxicity was reversed when a copy of *recX* was added to the complementing plasmid, indicating its role as a regulator of RecA function (Papavinasasundaram *et al.*, 1998).

The *recA* gene has been successfully deleted in *M. smegmatis*, *M. bovis* BCG and *M. tuberculosis* indicating that it is not essential for viability (Papavinasasundaram *et al.*, 1998; Sander *et al.*, 2001; P. Sander, personal communication). Indeed, infection in a mouse model with the *M. bovis* BCG mutant proceeded as for the wild-type (Sander *et al.*, 2001). However, in both *M. smegmatis* and *M. bovis* BCG, it was found that the Δ*recA* mutant was more sensitive than the wild-type to DNA damage by UV radiation, with the *M. bovis* BCG mutant also being more sensitive to alkylating agents (Papavinasasundaram *et al.*, 1998; Sander *et al.*, 2001).

An *M. tuberculosis* homologue of the *lexA* gene was found through sequencing of the genome (Philipp *et al.*, 1996). This gene was cloned, sequenced and its product purified for study (Movahedzadeh *et al.*, 1997b). The *M. leprae lexA* gene was isolated and identified through a cosmid library. It too was cloned and its product purified for binding studies (Durbach *et al.*, 1997). Through sequence analysis it was found that not only was the LexA product highly conserved between the two pathogens, but that these LexA proteins also

shared homology with *E. coli* LexA and *B. subtilis* DinR (Movahedzadeh *et al.*, 1997b). It was found that *M. tuberculosis* LexA could bind sequences homologous to the *B. subtilis* Cheo box located upstream of its own gene and also upstream of the *M. tuberculosis recA* gene (Movahedzadeh *et al.*, 1997a; Movahedzadeh *et al.*, 1997b). In the case of *M. leprae* LexA, it was found that it could not only bind the SOS boxes upstream of its own gene, but also had a strong binding affinity to the upstream region of the *recA* gene in both *M. tuberculosis* and *M. smegmatis* through the presence of SOS boxes (Durbach *et al.*, 1997). Primer extension of *M. tuberculosis* RNA with primers complementary to the beginning of the *lexA* gene gave rise to one product, identifying one transcriptional start site (Movahedzadeh *et al.*, 1997a).

Work done on the LexA binding site, based on observations of this site in the upstream regions of *recA* and *lexA* in *M. tuberculosis*, *M. leprae* and *M. smegmatis* came up with a possible mycobacterial consensus of gAACA-N_2-TGTTc (Durbach *et al.*, 1997). Point mutations within the *M. tuberculosis* LexA binding site further defined this consensus as TCGAAC-N_4-GTTTCA (Davis *et al.*, 2002). This consensus for the binding region is dissimilar from that seen for both *E. coli* and other enterobacterial LexA binding sites taCTGTatatatatACAGta (Lewis *et al.*, 1994). However, this consensus is similar to that originally defined as the Cheo box (GAAC-N_4-GTTC) in *B. subtilis* (Cheo *et al.*, 1991) and subsequently redefined as CGAAC-RNRY-GTTCG (Winterling *et al.*, 1998). The *M. tuberculosis* LexA binding site consensus was then used to find other possible SOS genes in the pathogen's genomic sequence. In total, 16 sites having a sequence that bore a strong resemblance to the *M. tuberculosis* LexA binding site consensus were identified (Davis *et al.*, 2002). Micro-array data obtained from mitomycin C induced *M. tuberculosis* cultures

showed that 14 of these sites had genes associated which were inducible with the DNA damage (Davis *et al.*, 2002).

Sequencing the pathogen gave rise to 13 homologues of genes involved in the *E. coli* SOS response (Mizrahi and Andersen, 1998). To see how conserved the *E. coli* SOS response was in *M. tuberculosis* ten genes, nine of which were known to be inducible in *E. coli* with seven being homologues of *E. coli* SOS response genes, were selected to see if they were inducible in *M. tuberculosis* with DNA damage. Testing through RT-PCR showed that seven of these genes were inducible in *M. tuberculosis* with mitomycin C, with one of these genes being *ruvC* (Brooks *et al.*, 2001). However, this gene is not part of the SOS response in *E. coli* but nevertheless possessed an SOS box in its *M. tuberculosis* version (Brooks *et al.*, 2001). Of the other genes that were inducible in the pathogen with mitomycin C, two were found not to have an SOS box upstream of their coding sequence, indicating that another mechanism, independent of LexA, is utilised to activate these genes (Brooks *et al.*, 2001).

Studies have been done to see which genes are activated by the pathogen when it interacts with non-activated human macrophages and amongst the genes seen to be upregulated were *uvrA, uvrB* and *uvrD* which are involved in nucleotide excision repair (Graham and Clark-Curtiss, 1999). This up-regulation was seen from 48 hours post infection onwards showing activation of DNA repair genes after being phagocytosed by the macrophage. It was also suggested that the continual presence of the response after 48 hours of infection indicated that DNA damage was still ongoing "well after internalisation" (Graham and Clark-Curtiss, 1999).

1.8.2. The Adaptive response.

Damage of DNA brought about through the actions of RNI and alkylating agents induces a mechanism in *E. coli* known as the adaptive response, which is specialised to deal with damage caused solely by these agents (Samson and Cairns, 1977; Lindahl *et al.,* 1988). The adaptive response is mediated in *E. coli* by the *ada* gene (Sedgwick, 1983) which controls the expression of three other genes, *alkA*, *alkB*, and *aidB* (Landini and Volkert, 2000). The Ada protein is made up of two domains, an amino and a carboxyl terminal domain which both contain an active methyl acceptor cysteine residue (Teo *et al.,* 1984; Landini and Volkert, 2000).

In the normal state, free of alkylation, *ada* is expressed weakly at a basal level (Nakabeppu and Sekiguchi, 1986). Overproduction of the Ada protein has been found to be lethal in *E. coli* (Lindahl *et al.,* 1988). Damage of DNA caused by alkylation can happen in one of two ways and in either case one cysteine residue of Ada interacts with the site of damage and removes the methyl group (Landini and Volkert, 2000).

Alkylation of DNA can lead to the formation of O^6- methylguanine (O^6meG) and O^4–methylthymine (O^4meT). Both these altered bases introduce lesions into DNA that are mutagenic (Eadie *et al.,* 1984; Dosanjh *et al.,* 1991). Both O^6meG and O^4meT interact with the cysteine residue present in the carboxyl terminal of the Ada protein (Lindahl *et al.,* 1982). However, this interaction between the methyl group and the C-terminal cysteine is irreversible, rendering it unable to repeat the same task (Lindhal *et al.,* 1982).

Damage caused by phospomethyltriesters is dealt with by the cysteine present in amino-terminal end of the Ada protein (Landini and Volkert, 2000). Methylation of Ada on the N-terminal cysteine results in a protein which takes on the role of a transcriptional regulator (Teo *et al.,* 1986; Nakabeppu and Sekigiguchi, 1986).

Although it has been stated that the adaptive response mediates DNA repair of NO damage, this response does not act alone. In a recent study, Spek *et al.* showed that recombination repair mediated through both RecA and RuvC was required in addition to the adaptive response to tackle RNI mediated DNA damage in *E. coli* (Spek *et al.,* 2001). The action of RecA in this instance was independent of LexA as it was shown that a Δ*recA* strain was far more sensitive to NO treatment than a strain possessing a mutant *lexA* that produced an uncleavable LexA product (Spek *et al.,* 2001).

Information present from the sequencing of *M. tuberculosis* H37Rv has shown that there are no homologues of the *alkB* and *aidB* genes in the pathogen. However, in the case of *alkA* and *ada*, there is a fusion of these genes in the pathogen with the N- terminus of the Ada-like protein being fused to the C-terminus of the AlkA protein to form the gene product of Rv1317c (Mizrahi and Andersen, 1998). The gene adjacent to this fusion, Rv1316c, shows homology to the *E. coli ogt* gene, which is also involved in BER and is in the same operon as Rv1317c. Gene Rv1317c has been knocked out in *M. tuberculosis* and was shown to be more sensitive to the alkylating agent, N-methyl-N-nitro-N-nitrosoguanidine (MNNG) than the wild-type (K. G. Papavinasasundaram, personal communication). This observation confirmed the role of Rv1317c in protection against alkylation. However, there was no attenuation of the mutant in mice infection studies,

where the mutant was under the stress of cell-mediated killing, (M. J. Colston, personal communication).

1.9. Gene transcription

The first step in the expression of any gene is that of transcription. Control of transcription is, therefore, an important means of controlling gene expression. In the case of the SOS and adaptive DNA damage responses mentioned in the last section, control in both cases was mediated through control of transcription. Transcription in eubacteria is dependent on a number of factors, but the two most important are the protein complex involved in the transcription process, RNA polymerase, and the point of interaction on the genomic DNA for this complex, the promoter.

1.9.1. DNA elements involved in transcription

The promoter is a region of sequence upstream of an open reading frame that has a strong interaction with the RNA polymerase holoenzyme, which in turn initiates transcription. Although the structure of these elements varies from gene to gene and indeed from bacterial species to species, the basic promoter is made up of two elements, which interact with different parts of the holoenzyme (Gaal *et al.*, 2001). These elements are called the –10 and –35 boxes, with their numbers indicating how far upstream each box is from the transcription initiation site (+1) (McClure, 1985).

Characterisation of the -10 region began with Pribnow and Schaller, who both noticed a conserved heptamer (TATAATG) present in a small number of promoter sequences that

was centred around 10 base pairs upstream of the transcriptional start point (Pribnow, 1975a, b; Schaller *et al.,* 1975; Rosenberg and Court, 1979). In the case of the –35 region the same type of observations were made in a number of studies, providing the hexamer, TTGACA (Rosenberg and Court, 1979). These findings were substantiated in a later study of over 100 promoter sequences isolated in *E. coli* in which the majority followed a variation of the theorised consensus –10 and –35 with a spacing of 17± 1 nucleotides between the two elements (Hawley and McClure, 1983). It has been shown that although the spacer element is not entirely conserved in most promoters, the distance between the –10 and –35 elements is important for promoter activity (Aoyama *et al.,* 1983; Warne and deHaseth, 1993; McKane and Gussin, 2000).

The promoter consensus from these observations was based on that utilised by the major *E. coli* sigma factor, σ^{70}, and observations of promoters associated with the major sigma factor in *B. subtilis* showed that this consensus was also conserved in this species (Moran *et al.,* 1982). Other studies involving the RNA polymerase from a number of different species showed that all the different RNA polymerases isolated recognised the same promoter element, which by inference indicated that this consensus was universal for the major sigma factor (Wiggs *et al.,* 1979). However, the presence of two sets of different consensus sequences in the *B. subtilis* bacteriophage SP01 indicated that there was a specific promoter consensus for a specific sigma factor (Lee and Pero, 1981). This theory was substantiated by the findings of Cowing *et al.* (1985), which noted that, in *E. coli,* there were a group of promoters for genes that were associated with heat shock that shared a consensus that was not the same as that for the major sigma factor. These promoters were not recognised by *E. coli* σ^{70} but were recognised by *E. coli* σ^{32} (Cowing *et al.,* 1985). From this observation, it was suggested that the –10 consensus was the key determinant of sigma factor utilisation

with the sigma factor making initial contact with the –10 region and the –35 interaction happening afterwards (Cowing *et al.,* 1985). This type of observation has also been made in *B. subtilis,* where changes in the -10 region change its specificity for two alternative sigma factors (Qiu and Helmann, 2001).

It has been shown in subsequent studies that the –35 region in some promoters is dispensable for promoter activity (Belyaeva *et al.,* 1993; Kumar *et al.,* 1993). It was found that for these promoters, a TG motif immediately upstream of the –10 region was conserved (Burr *et al.,* 2000). This type of motif has also been observed in a promoter recognised by the major sigma factor in *B. subtilis* and was suggested to be of importance in weakly expressed promoters (Voskuil *et al.,* 1995).

In addition to the –10 and –35 regions, there are regions upstream of the –35 element that can enhance promoter activity. This adenine-thymine (AT) rich DNA sequence upstream of the promoter region is called an UP element (Busby and Ebright, 1994; Gourse *et al.,* 2000). It was observed that the –40 to –60 region of rRNA promoters contributed to the high activity of the *rrnB* promoter (Newlands *et al.,* 1992; Rao *et al.,* 1994). It has since been shown that the UP element interacts with the carboxyl terminal domain of the α–subunit (Busby and Ebright, 1994).

In addition to the presence of UP elements upstream of the –35 element, in some cases there are protein-binding sites, called activator sites that interact with accessory proteins and enhance transcription (Busby and Ebright, 1994). Examples of this can be seen with *E. coli* promoters whose activities are enhanced by the cyclic AMP receptor protein (Busby *et al.,* 1994; Joung *et al.,* 1994).

Study of mycobacterial promoters has shown that the majority of promoters in both *M. smegmatis* and *M. tuberculosis* have a conserved –10 region that is similar to that found for the *E. coli* σ^{70} consensus but, at the same time also have far more variations in the –35 region (Bashyam *et al.,* 1996). The variation in the –35 region is one probable reason why mycobacterial promoters are weakly expressed in *E. coli* (Das Gupta *et al.,* 1993; Bashyam *et al.,* 1996). Interestingly, however, *E. coli* promoters can be expressed in mycobacteria, indicating that mycobacteria can tolerate variation in the –35 region better than *E. coli* (Bashyam *et al.,* 1996; Mulder *et al.,* 1997). In addition, an extended –10 region that can compensate for the lack of an adequate –35 region has also been found in some mycobacterial promoters (Bashyam and Tyagi, 1998).

A compilation of 102 promoter sequences that have been isolated in mycobacteria has been made and these sequences have been assigned to one of four groups based on sequence homology but not spacing between the elements (Gomez and Smith, 2000). Group A promoters (65 out of 102) are those which share homology in both the –10 and –35 elements with that found in the *E. coli* σ^{70} consensus and are found in all mycobacteria (Gomez and Smith, 2000). Group B (22 out of 102) promoters are more loosely homologous in the –10 region with the *E. coli* σ^{70} consensus but lack consensus in the –35 region and are again found in all mycobacteria (Gomez and Smith, 2000). Group C (7 out of 102) has no consensus in either region while group D (8 out of 102) has a novel consensus in both –10 and –35 and is found in only *M. paratuberculosis* (Gomez and Smith, 2000).

1.9.2. The RNA polymerase

The RNA polymerase used in transcription comes in two forms. The core polymerase is made up of five subunits (two α, β, β' and ω) (Minakhin *et al.,* 2001a). This core is sufficient to allow non-specific transcription to proceed and has been shown to initiate transcription from nicked DNA and 3' overhangs. However, the addition of a sixth sub-unit, the sigma factor, reduces non-specific DNA binding and promotes transcription from the specific promoter site.

1.9.2.1. The α subunit

The α subunits are made up of two domains that are linked through a long flexible linker (Busby and Ebright, 1994). The amino (N) terminal domain interacts with the rest of the polymerase and has been shown to direct the formation of the RNA polymerase core in solution (Ishihama, 1981). The carboxyl (C) terminal end of the subunit can interact with a number of DNA elements and proteins to enhance transcription (Busby and Ebright, 1994; Gourse *et al.,* 2000). It has been shown that this part of the subunit can also interact with regions far upstream of the promoter from –37 to –95 (Naryshkin *et al.,* 2000). However, the subunit-DNA or subunit-protein interactions alone cannot direct transcription.

1.9.2.2. The β and β' subunits

The *E. coli* β and β' subunits are large sub-units with the β' being 156 kDa and the β, 151 kDa. It has been shown that both these subunits have regions of proximity with the DNA binding part of the holoenzyme, the sigma factor (Burgess and Anthony, 2001). However, it was shown that a deletion of the β' subunit lead to an inability of the core enzyme to bind the sigma factor, indicating that this subunit is a point of binding for the sigma subunit (Burgess and Anthony, 2001). Through further analysis of the β' subunit with co-immobilisation assays, it was shown that the region 260-309 of the β' unit bound the σ^{70} subunit (Burgess and Anthony, 2001; Young *et al.,* 2001). The β subunit has been implicated in the formation of the open complex at the promoter, enabling transcription to proceed (Mekler *et al.,* 2002). It has also been shown that the β subunit is the point of action of the antibiotic rifampicin and mutations in the *rpoB* gene are associated with resistance to this drug. It has been postulated that the formation of the RNA polymerase is dependent on these two subunits since there is an excess of α subunit (Bedwell and Nomura, 1986). It was also found that an increase in the presence of both these subunits lead to a stimulation of σ^{70} production by the cell (Bedwell and Nomura, 1986).

1.9.2.3. The ω subunit

It is not entirely clear what the role of ω is in the transcription process. In fact, until recently, it was not known that this subunit existed. It has been shown that the gene encoding this protein, *rpoZ*, is not essential for viability and that its presence or absence in the RNA polymerase has no noticeable effect on the enzyme complex (Minakhin *et al.,* 2001b). However, it has also been shown that deletions of *rpoZ* are slow growing in comparison to the wild-type and that this gene is conserved in a number of bacterial

sequences (Minakhin *et al.,* 2001b). Recent studies have shown that this subunit may have some role in RNA polymerase core formation and has been shown to co-purify with the β' subunit with which it solely interacts (Naryshkin *et al.,* 2000; Ghosh *et al.,* 2001).

1.9.2.4. The core RNA polymerase subunits in mycobacteria

Through a number of studies it has been shown that core polymerase subunits are conserved in all known prokaryotes (Fukuda *et al.,* 1977; Wiggs *et al.,* 1979). However, in a study of *M. tuberculosis* H37Rv RNA polymerase although the three key components of the polymerase were present, it was found that both the β and β' subunits were of a smaller size in the pathogen (Harshey and Ramakrishnan, 1976). It was also shown in this study that *M. tuberculosis* RNA polymerase was more susceptible to the action of rifampicin than its *E. coli* counterpart but was not affected by the action of any other anti-tuberculosis drugs (Harshey and Ramakrishnan, 1976). Later studies involving the sequencing of cosmids identified homologues of *rpoB* and *rpoC* in *M. leprae* (Honore *et al.,* 1993). Studies of *in vitro* transcription with purified *M. smegmatis* RNA polymerase showed that its activity like that of the *E. coli* version, was inhibited by rifampicin (Levin and Hatfull, 1993). Further study with the *M. smegmatis* polymerase holoenzyme showed that it could direct *in vitro* transcription from both native genes and also those from *B. subtilis* (Predich *et al.,* 1995). In addition, it was shown that the *M. smegmatis* RNA core polymerase could associate with the major sigma factor of *B. subtilis* and undertake *in vitro* transcription (Predich *et al.,* 1995). Identification of the ω subunit was done through homology at the DNA level with the *E. coli rpoZ* gene. In this way it was found that gene *Rv1390* was a

possible candidate for the production of this subunit (Minakhin *et al,* 2001b). However, no studies have been done at this time to confirm or rule out this observation.

1.9.2.5. The sigma factor

In all studied prokaryotes there are a number of sigma factors present but most of the transcriptional events are mediated by just one major factor. In the case of *E. coli*, which has seven sigma factors, the major factor is σ^{70}. However, this protein on its own is incapable of binding DNA and requires interaction with the β' subunit to induce a conformational change in the sigma factor to facilitate binding (Burgess and Anthony, 2001). The structure of the sigma factor has been studied in great detail and is shown in figure 1.3. There are two regions associated with its interaction with the promoter. Region 4.2 is associated with the –35 region whilst region 2.4 is concerned with the –10 region (Lonetto *et al.,* 1992). Region 1.1 provides the sigma factor with a means of stopping core polymerase independent DNA binding and it has been shown that a deletion of this region enables sigma 70 binding at the promoter region (Dombroski *et al.,* 1993).

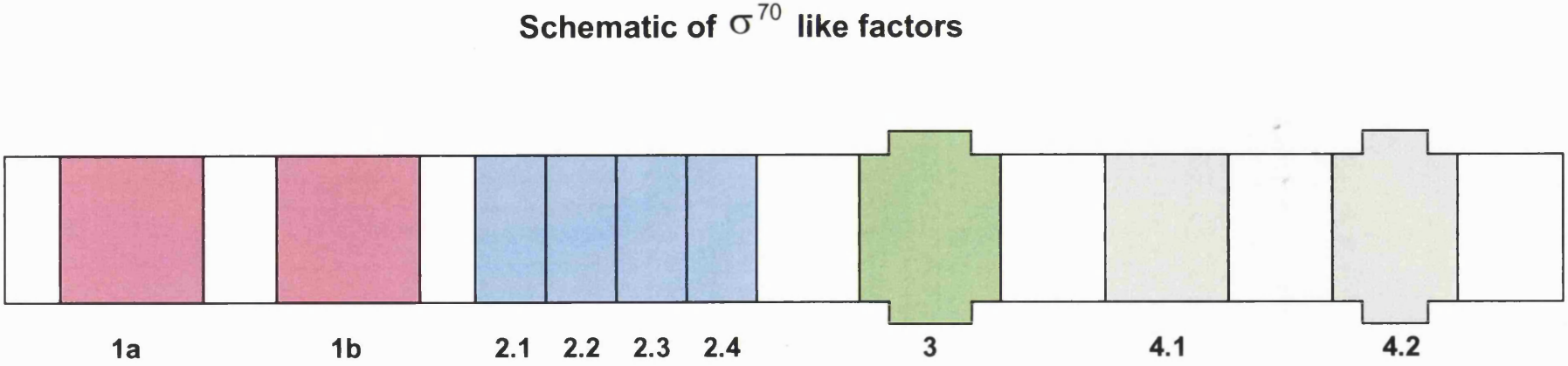


Figure 1.3. This cartoon illustrates the regions that make up σ^{70}-like sigma factor subunit of the RNA polymerase holoenzyme. In region 1, which is the amino terminus of the protein, region 1a is only found in primary σ factors. Region 2 (which is separated into four regions), region 3, and the two sections of region 4 are found in all σ^{70}-like sigma factors. Region 2.1 has been shown to be involved in binding the RNA polymerase core (Gross *et al.,* 1992), whilst region 2.4 and 4.2 are involved in interactions with the –10 and –35 regions respectively. It has also been shown that both regions 3 and 4.2 have homology with helix-turn-helix motifs, although no function as of yet has been given to region 3. However, it has been suggested the region 3 may also be involved in interactions with the RNA polymerase core (Gross *et al.,* 1992). This figure is adapted from Gross *et al.,* (1992).

Based on the findings in *E. coli* sigma factors, all other sigma factors have been broken down into two distinct families; the σ^N and the σ^{70} group (Lonetto *et al.,* 1992).

The σ^N group (also known as the σ^{54} group) encompasses certain alternative sigma factors in a wide variety of species (Studholme and Buck, 2000, Studholme *et al.,* 2000). Unlike sigma factors belonging to the σ^{70} family, the promoter structure recognised by these proteins is made up of only one element centred around the -12 region (termed the -12 consensus element) with the consensus YTGGCACCrNNNTTGCW (Barrios *et al.,* 1999; Studholme and Buck, 2000; Studholme *et al.,* 2000). Another shared characteristic that separates members of the σ^{54} family form those in the σ^{70} family is their inability to form an open complex at the promoter region in the absence of an activator protein (Guo *et al.,* 2000).

In the case of the σ^{70} family, there are three main groups. Groups 1 and 2 are made up of the primary sigma factors with group 2 being less homologous to the major *E. coli* sigma factor (Lonetto *et al.,* 1992). Group 3 is made up of sigma factors that have low homology with members of group 1, with a maximum identity of 27%, and are involved in alternative sigma factor functions (Lonetto *et al.,* 1992). Within this third group, there are clusters for the various functions; these being flagellar, heat shock, sporulation and extracytoplasmic functions (Lonetto *et al.,* 1992; Lonetto *et al.,* 1994).

1.9.2.6. Mycobacterial sigma factors.

From genome sequencing, it has been confirmed that *M. tuberculosis* H37Rv possesses 13 sigma factors given the names σ A-M (Cole *et al.*, 1998). On the basis of the σ^{70} family studies done by Lonetto *et al* (1992 and 1994), 2 (σ^{A} and σ^{B}) have been designated as group 2 primary sigma factors, with σ^{F} being classified as a class 3 sporulation factor. The remaining 10 have been classified under class 3 extracytoplasmic function (ECF) factors (Cole *et al.*, 1998; Gomez and Smith, 2000).

1.9.2.6.1. The σ^{A} Factor

The factor sigma A was first identified in *M. smegmatis* through the purification of the RNA polymerase holoenzyme (Predich *et al.*, 1995). Using an antibody raised against a region of the *E. coli* σ^{70} promoter conserved in the σ^{70} class of proteins, a western blot was done on the *M. smegmatis* RNA polymerase. From this blot it was seen that two proteins, corresponding to 65kDa and 40kDa had been isolated (Predich *et al.*, 1995). In the same study, the *hrdB* gene, which encodes the major sigma factor in *Streptomyces coelicolor* was used as probe to identify sigma factor genes in the saprophyte. This gene was used as the *Streptomyces* and *Mycobacterium* families are closely related (Doukhan *et al.*, 1995; Predich *et al.*, 1995). Using this same gene to probe genomic DNA of both *M. tuberculosis* and *M. leprae*, a *sigA* gene in each was identified (Doukhan *et al.*, 1995). Comparisons of σ^{A} in these mycobacterial species showed that there was a very high degree of identity in the three proteins in all but the1.1 sub-region (Doukhan *et al.*, 1995). However, the

sequence linking the 1.2 and 2.1 regions, which is normally variable between different sigma factors, was conserved in all three species (Doukhan *et al.,* 1995).

Since σ^A in all three species could be linked with the gene encoding the major sigma factor in *S. coelicolor*, it was assumed that σ^A was indeed the mycobacterial major sigma factor (Gomez *et al.,* 1998). This assumption was proved in the case of *M. smegmatis*, where a deletion of σ^A was shown to be lethal (Gomez *et al.,* 1998). However, an attenuated strain of *M. tuberculosis* had been isolated with a point mutation in σ^A that changes an arginine residue into a histidine (Collins *et al.,* 1995). It was noted that the point change happened in a region of the protein that interacted with the promoter. Complementing this avirulent strain with a normal copy of the gene restored its virulence, fulfilling Koch's postulate and indicating that this protein is a virulence factor (Collins *et al.,* 1995). At first sight, what the finding of such a mutant indicates is that σ^A is dispensable in *M. tuberculosis*. However, the authors of the study indicated that another possible explanation was that the mutated form of σ^A might not be able to interact specifically with promoters coupled to certain genes associated with virulence, but could still function for the majority of genes (Collins *et al.,* 1995). This idea was substantiated in a study done in *M. smegmatis*, where a point mutation equivalent to that found in the avirulent *M. tuberculosis* strain was found to still be viable (Gomez and Smith, 2000). However, this mutant strain was also more susceptible to certain stresses in comparison with the wild-type (Gomez and Smith, 2000). Further studies have also shown that *sigA* is constitutively expressed in a wide range of growth and stress conditions (Hu and Coates, 1999a; Manganelli *et al.,* 1999). However, in cases of exposure to H_2O, low aeration and entrance into stationary phase, there is a down regulation of this factor (Gomez *et al.,* 1998; Manganelli *et al.,* 1999)

1.9.2.6.2. The σ^B factor

In addition to locating the *sigA* gene in *M. smegmatis*, the study of Predich *et al* identified a second similar sigma factor downstream of *sigA* at the same locus, which was termed *mysB* and later *sigB* (Predich *et al.,* 1995). Probing the *M. tuberculosis* and *M. leprae* genomes with the *hrdB* gene also located the *sigB* gene in these two species (Doukhan *et al.,* 1995). Comparisons between the σ^B proteins of each species noted that differences seemed to be located in regions 3 and 4, but that regions 2.4 (which interacts with the –10 region of the promoter) and 4.2 (which interacts with the –35 region) were conserved in all three species (Doukhan *et al.,* 1995). In addition, region 1.1 that was variable in σ^A for all three species has been conserved in each σ^B variant (Doukhan *et al.,* 1995).

Even though *sigB* may be similar to both *sigA* and *hrdB* it has been shown through a knockout mutation in *M. smegmatis* to not be essential for viability under normal *in vitro* growth conditions (Gomez *et al.,* 1998). However, the *M. smegmatis sigB* mutant is more susceptible to oxidative stress than the wild-type (Gomez and Smith, 2000).

Studies of *M. tuberculosis sigB* involving RT-PCR have shown that this factor is upregulated under the stresses of heat shock, cold shock, low aeration and exposure to the detergent SDS (Manganelli *et al.,* 1999). In addition, *sigB* mRNA has also been found in both exponential and stationary phase (Manganelli *et al.,* 1999). However, at stationary phase, transcription is greater than in exponential phase (Hu and Coates, 1999a). This is an interesting point, as it has been suggested that *M. tuberculosis* σ^B is involved primarily with entry of the pathogen into stationary phase (Predich *et al.,* 1995; Hu and Coates, 1999). It

has been shown that σ^B may indeed be linked (at least at a functional level) with the *E. coli* KatF protein, which is the main sigma factor for this organism in stationary phase (Predich *et al.,* 1995). It was found that overexpression of the *M. tuberculosis sigB* gene lead to increased activity of the *M. tuberculosis katG* promoter mimicking the relationship of KatF and KatG in *E. coli* (Mulder *et al.,* 1999).

1.9.2.6.3. The σ^E Factor

The sigma factor σ^E is associated with the regulation of diverse functions in many bacteria (Wu *et al.,* 1997). A sequence bearing similarity to a gene associated with this factor was located in a *M. leprae* cosmid. Using primers based on this sequence a σ^E protein was found to be highly conserved in four species of mycobacteria (*M.avium*, *M. leprae*, *M. smegmatis* and *M. tuberculosis*) (Wu *et al.,* 1997). It was found that the closest non-mycobacterial matches to this protein were Ybbl of *B. subtilis* and SigE of *E. coli*, both of these being ECFs (Wu *et al.,* 1997).

In both *M. smegmatis* and *M. tuberculosis* it has been possible to knockout *sigE* without affecting viability under normal *in vitro* growth conditions (Wu *et al.,* 1997, Manganelli *et al.,* 2001). However, it has been shown that the *M. smegmatis* Δ*sigE* mutant is substantially more sensitive to acid, oxidative, detergent and heat stresses (Wu *et al.,* 1997). In related studies, it has been shown that the transcription of *sigB* in a *M. tuberculosis* Δ*sigE* knockout strain is greatly decreased in comparison with the wild-type in both exponential growth and detergent (SDS) stress conditions, although not following heat shock (Manganelli *et al.,* 1999; Manganelli *et al.,* 2001). This indicates that *sigB* has a partial

dependence on *sigE* for its expression. (Manganelli *et al.,* 1999; Manganelli *et al.,* 2001; Raman *et al.,* 2001)

In *M. tuberculosis* it has been shown that expression of *sigE* increases under heat, detergent and low aeration stresses *in vitro* (Manganelli *et al.,* 1999). In addition, work done with *M. tuberculosis* infected macrophages has shown that expression of *sigE* is induced during growth within the cell and suggests that σ^E may be required for survival and growth in this state (Jensen-Cain and Quinn, 2001).

1.9.2.6.4. The σ^F Factor

The factor σ^F was first identified in *M. tuberculosis* by PCR amplification from genomic DNA fragments through the use of degenerate primers based on the *E. coli* σ^{70} consensus (DeMaio *et al.,* 1996). Using the resulting gene to probe other mycobacterial species, it was found that *sigF* homologues were present in the slow growing *M. bovis, M. avium* and Bacille Calmette-Guerin (BCG), but not in the fast growing *M. smegmatis* or *M. abscessus* (DeMaio *et al.,* 1996). From this study, it was determined that *M. tuberculosis* σ^F had closest homology to *S. coelicolor* σ^F and *B. subtilis* σ^B and σ^F (DeMaio *et al.,* 1996; Michele *et al.,* 1999). In *S. coelicolor*, σ^F is associated with spore formation whilst in *B. subtilis*, this protein plays a role in gene regulation within the spore (Haldenwang, 1995; Gomez and Smith, 2000). In the case of *B. subtilis sigB*, transcription of this gene is activated during stationary phase or under stresses such as heat shock (Haldenwang, 1995). However, *B. subtilis sigB* has been found to be non-essential for growth and has no role in sporulation (Haldenwang, 1995).

No significant increase in *sigF* expression was detected by RT-PCR following heat shock, cold shock, low aeration or stationary phase (Manganelli *et al.,* 1999). This is in contrast to earlier studies in BCG, where it was shown that *sigF* expression was strongly increased during stationary phase and was also induced to a lesser extent in low aeration conditions (DeMaio *et al.,* 1996). Furthermore, σ^F was not detected during early exponential phase (DeMaio *et al.,* 1996). In fact, overexpression of *sigF* in BCG during exponential phase was lethal, presumably since genes required for stationary phase/ latency were being activated when exponential growth was still proceeding (Gomez *et al.,* 1997). Addition of the same *sigF* gene to *M. smegmatis* had no effect on the viability of this species (Gomez *et al.,* 1997). The reasons given for this anomalous finding in the earlier BCG study by Manganelli *et al* (1999) include a difference in the times of exposure used in either study and that a different mechanism of regulation may be at work in the two species (Manganelli *et al.,* 1999). The presence of an anti-sigma factor gene (*usfX*) directly upstream of the *M. tuberculosis sigF* could be one such regulatory factor (Gomez *et al.,* 1997).

It has been possible to knock out *M. tuberculosis sigF* by allelic exchange, which shows that this factor is not essential for viability in normal *in vitro* growth conditions (Chen *et al.,* 2000). The knock out mutant was found to be more sensitive to rifampin treatment than both the wild-type and complemented strains (Chen *et al.,* 2000). It was shown by the uptake of radio-labelled chenodeoxycholate (a substance which enters the pathogen through passive diffusion) that there was a change in the cell envelope of the mutant that made it less permeable than the wild-type (Chen *et al.,* 2000). The *M. tuberculosis ΔsigF* mutant was also less virulent (but not avirulent) in comparison with the wild-type when used to infect mice (Chen *et al.,* 2000). However, there were no obvious differences noted in the survival of the mutant compared with the wild-type over a nine day period when used to

infect human monocytes (Chen *et al.,* 2000). One possible explanation given for this finding is that there may be an overlapping stress response, much like cross protection, that can make up for the silencing of important σ^F dependent genes (Chen *et al.,* 2000). Another possibility was that the macrophages used for the infection studies with the mutant were not activated and that possible short falls in the mutant would only be made clear when macrophages were activated prior to infection (Chen *et al.,* 2000).

1.9.2.6.5. The σ^H factor

Unlike the other four previously mentioned sigma factors, σ^H was initially determined in *M. tuberculosis* through analysis of the genome sequence with the *M. tuberculosis sigE* sequence to identify other ECFs (Fernandes *et al.,* 1999). A PCR probe based on *M. tuberculosis sigE* was subsequently used to locate the *sigH* gene in both *M. tuberculosis* and *M. smegmatis* (Fernandes *et al.,* 1999). From this, one candidate gene, later named *sigH* was isolated that was similar to but distinct from *M. tuberculosis sigE* (Fernandes *et al.,* 1999). This factor was found to be up-regulated following heat shock by both primer extension analysis (Fernandes *et al.,* 1999) and RT-PCR (Maganelli *et al.,* 1999). Other studies have shown that *sigH* is also induced upon entry into macrophages and under oxidative stress (Graham and Clark-Curtiss, 1999).

It has been possible to create a viable Δ*sigH* knockout mutant in both *M. smegmatis* and *M. tuberculosis*, which indicates that this gene is not essential in regular *in vitro* growth conditions (Fernandes *et al.,* 1999; Raman *et al.,* 2001; Kaushal *et al.,* 2002). Studies undertaken with the *M. tuberculosis* mutant showed that it was more sensitive to heat and

oxidative stress that the wild-type (Raman *et al.,* 2001). In addition, it was found that transcription of *sigB* in the Δ*sigH* mutant in both normal and oxidative stress conditions was drastically reduced in comparison with the wild-type, indicating a dependence of σ^B on σ^H (Raman *et al.,* 2001). It has also been shown from micro-array analysis that the expression of both *sigB* and *sigE* decreases in late stationary phase in a *M. tuberculosis* Δ*sigH* mutant in comparison with the wild-type (Kaushal *et al.,* 2002). However, although the *M. smegmatis* Δ*sigH* mutant was also sensitive to organic peroxide stress, its survival after heat shock at 42°C was no different to that seen in the wild-type (Fernandes *et al.,* 1999)

1.9.2.6.6. Other mycobacterial sigma factors.

There is not much known about the other sigma factors present in *M. tuberculosis* at this current time. Identification of these other sigma factors is purely based on homology found in the sequence of the pathogen with known ECF σ factors (Lonetto *et al.,* 1994). However, there are results from RT-PCR for five of these factors (σ^C, σ^D, σ^G, σ^I, and σ^M) that has given some information (Manganelli *et al.,* 1999). All but σ^G were expressed at a significant level in exponential phase whilst both σ^D and σ^I were upregulated in cold shock (Manganelli *et al.,* 1999).

The expression of σ^M was also slightly increased during heat shock (Manganelli *et al.,* 1999). It has been possible to knockout *sigM* in both *M. tuberculosis* and *M. bovis* BCG (B. Springer, personal communication). However, survival experiments following heat shock

of the *M.bovis* BCG mutant has shown that there is no difference between its viability and that of the wild-type (B. Springer, personal communication).

1.10. Project aims

The primary aims of this study are to characterise the promoters of the *M. tuberculosis recA* gene and to identify the sigma factor(s) responsible for its expression.

To characterise the promoters, we wished to address the following questions. Firstly, what are the promoter elements for each promoter? Is it possible to knock out expression of each individual region by mutation? Secondly, is it possible that another promoter could exist upstream of *recA*? Thirdly, is there a difference between the two promoters in how they respond to DNA damage? What are the kinetics of induction of each promoter and how do their expression levels change during growth?

Although potential promoter sequences had been suggested from inspection of the sequences in the region of the mapped transcription start sites, it was important to determine experimentally which sequences were important for *recA* expression. This was particularly so for the *recA*P1 promoter where there were two candidate sequences for the –35 site: one based on homology and one on position (Movahedzadeh *et al.*, 1997a). Once the regions involved had been determined it would be possible to analyse the sequence requirements for promoter activity in detail and so define a consensus sequence for the promoter elements.

It is also of interest to know whether the two promoters are expressed using the same or different sigma factors. To identify which sigma factor(s) are used by each promoter, we have chosen to make knock out mutations of as many of the *M. tuberculosis* sigma factors as possible and then to test the expression from each promoter in each strain. This approach has the advantage that these strains will also be useful for the analysis of other promoters in the future. In addition, with the use of micro-array technology present within the department, further study of how alternative sigma factors affect gene expression can be undertaken.

CHAPTER 2: METHODS AND MATERIALS

2.1. General molecular biology methods

2.1.1. Bacterial strains used in this study

In this study the following bacteria and strains were used. For the purposes of general cloning and production of plasmids for promoter studies the *Escherichia coli* DH5α strain was used (Sambrook *et al.,* 1989). For site directed mutagenesis and production of plasmids involved in the creation of the sigma factor mutants, supercompetent *E. coli* XL-1 Blue (Stratagene) was used. In the case of the mycobacterial work done *Mycobacterium smegmatis* mc^2155 (Snapper *et al.,* 1990) and *Mycobacterium tuberculosis* H37Rv were used.

2.1.2. Growth media used

For work involving the *E. coli* DH5α strain, liquid growth was in L-Broth at 37°C with agitation of 300 rpm. In the case of solid media, L- agar was used for both strains. Antibiotic selection was done by the addition of one of the following compounds in the prescribed concentrations; Kanamycin (50μg per ml), Hygromycin (250μg per ml), Ampicillin (100μg per ml) or Gentamicin (20μg per ml) to the growth medium. Blue/ white plate selection on plates was achieved by adding 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (X-Gal) at a concentration of 100μg per ml.

For mycobacterial work, liquid growth for both *M. smegmatis* and *M. tuberculosis* was done in Dubos broth (constitution in appendix) supplemented with 0.2% glycerol (v/v) and 0.4% albumin (v/v). For *M. smegmatis* liquid growth was done at 37°C with agitation of 300 rpm. For *M. tuberculosis* 100ml cultures were grown in 500 ml screw top roller bottles (Nalgene) in a roller incubator set to 37°C and 2 rpm. In the case of solid media, 7H11 agar supplemented with 0.2% glycerol (v/v) and 0.4% albumin (v/v) was used for both organisms. Antibiotic selection was done by the addition of one of the following compounds in the prescribed concentrations; Kanamycin (25μg per ml), Hygromycin (50μg per ml), or Gentamycin (20μg per ml) to the growth medium. Blue/ white plate selection on plates was achieved in the same way as for *E. coli*. For selection of sucrose resistance, 2% sucrose (v/v) was added to the mixture made for solid growth.

2.1.3. Production of competent cells

The $CaCl_2$ procedure was used in the making of *E. coli* DH5α competent cells (Sambrook *et. al.*, 1989). In the case of *M. smegmatis*, once a culture of 100mls had reached the OD_{600} value of around 1.0, it was spun down in Sorvall ® GSA rotor at 4°C for 30 minutes at a speed of 10,000rpms. The cell pellet formed was re-suspended in 25mls of 10% glycerol and spun again in the same rotor under the same conditions. This procedure was repeated three times before finally suspending the pellet in a volume of 10mls 10% glycerol. The procedure to make *M. tuberculosis* competent cells was similar but with the following modification. Once a 100ml culture reached OD_{600} of around 1.0-1.2, 10mls of 2M glycine (for a final concentration of 0.2M) was added and the cells incubated for a further 24 hours prior to harvesting and washing.

2.1.4. Plasmid DNA preparations

Plasmid DNA was isolated from *E. coli* using the S.N.A.P. minprep kit (Invitrogen) or, for the screening of sigma factor colonies by restriction digestion, where the DNA quality is less important, by the STET method (Holmes and Quigley, 1981).

2.1.5. DNA digestion, precipitation and ligations.

Digests were done between a time period of two hours and overnight according to manufacturers instructions. Subsequent to digestion, enzyme removal was done by adding 1 volume of phenol: chloroform and spinning for 2 minutes at 13,000 rpms in a microfuge. The digested DNA was then precipitated out in a solution of 2.5 volumes 96% ethanol and 0.1 volumes 3M sodium acetate (pH6)

Digested vector DNA and potential insert DNA were run on a 1% agarose gel at 70mV for one hour. From an estimation of this gel, a ratio of 1:3 (50ng of vector to 150ng of insert) was used in a final volume of 21µl. Ligations were done through the use of the Rapid ligation kit (Roche) where the ligation step was done at room temperature for 5 minutes.

2.2. Methods employed in the study of the *M. tuberculosis recA* promoters

2.2.1. Use of specific oligonucleotide for directed mutagenesis of *recA*P2

For the mutations done in *recA*P2, specific oligonucleotide pairs were designed to contain the specific mutation. These were annealed by adding volumes equivalent to a final

concentration of 2.5μM of both forward and reverse oligonucleotide and making the volume up to 100μl with the addition of TEN buffer (10mM Tris, pH8, 1mM EDTA, 100mM NaCl). This mixture was heated at 95°C for 10 minute and then allowed to cool slowly to room temperature.

The annealed oligonucleotides were then phosphorylated by adding 5μl of the annealed mixture to 1μl of T4 polynucleotide kinase (Roche), 2μl 10X kinase buffer (1M Tris pH7.5, 100mM $MgCl_2$, 70mM DTT, 10mM ATP and 10mM spermidine) and made up to 20 μl with the addition of dH_2O. From the resulting kinased products (which have a concentration of 20 μM) 1μl was ligated in each case to 1μl (60ng) of pEJ444 DNA which had been *Hin*dIII / *Xba*I digested and alkaline phosphatase (Boehringer Mannheim, 1u/μl) treated. Ligation reactions were done in the same way as mentioned for general ligations.

2.2.2. Site directed mutagenesis

The designing of the specific primers required for site–directed mutagenesis and the procedure itself were done as described in the kit instructions for the QuikChange™ Site directed mutagenesis kit, with the following amendments. The sample reactions were set up as follows. In the 50μl reaction was 5μl 10X cloned Pfu buffer (Stratagene), 5μl 2.5mM dNTPs, 2.5μl DMSO, volume equating to 125ng of DNA for both forward and reverse primer, 1μl DNA from S.N.A.P. miniprep (60ng) and 1μl Pfu turbo™ enzyme (Stratagene), with the remaining volume being made up by the addition of dH_2O. After PCR, 1μl of *Dpn*I (10u/ μl) was added before incubation for 1-2 hours at 37°C to digest the template DNA.

From the *Dpn*I treated reaction, 4µl was used to transform 50µl of *E. coli* XL-1Blue competent cells (Stratagene)

2.2.3. Transformation of cells

When transforming *E. coli* DH5α or supercompetent *E. coli* XL-1 Blue cells, heat shock at 42°C for 45 seconds was used. Transformed DH5α was incubated in L-broth whilst XL-1 Blue transformants were incubated in NZY broth. In both instances incubation was in an agitated shaker (300rpm) at 37°C for one hour.

To transform both *M. smegmatis* and *M. tuberculosis*, electroporation was used. 400µl of competent cells and 2µl of SNAP minipreped (Invitrogen) derived plasmid DNA were added into a 0.2 cm Gene Pulsar® cuvette (Bio-Rad). This was firstly chilled on ice before being placed in a Gene Pulsar™ set at 2.5 kV, 1000Ω and a capacitance of 2.5 µFd. After electroporation, the cells were transferred to 4.6 ml Dubos complete (without antibiotic) and left at 37°C in a standing incubator for 2 hours before plating out 100 µl.

2.2.4. Confirmation of sequence in mycobacteria

To verify the sequence of promoter constructs that had integrated into the mycobacterial chromosome; the Instagene kit (Bio-Rad) was used to isolate genomic DNA. The modification of taking up a loopful of bacteria as opposed to the prescribed single colony was employed. The region of interest was then amplified from this DNA with primers flanking the insert in the clone. In the case of *M. smegmatis*, the PCR mixture composition

was as follows: 10µl Instagene derived DNA, 5µl 2.5mM dNTPs, 5µl cloned Pfu buffer, 2.5µl DMSO, 0.5µM of both forward and reverse primer (see appendix II), 1µl Pfu turbo polymerase (Stratagene) and made up to 50µl with the addition of dH_2O. In the case of the *M. tuberculosis* transformants, the volume of Instagene derived DNA was lowered to 5µl and the volume of dH_2O altered to give a final volume of 50 µl.

The cycle conditions were as follows: 1x 94°C at 2 minutes, 25x [94°C at 30 sec., 58°C at 30 sec., 72°C at 45 sec.], 1x 72°C for 7 minutes. After PCR, 5µl of the resulting mixture was taken and run on a 1% agarose gel to check for success of the reaction. Once checked, the remaining 45µl was made up to 100 µl with the addition of dH_2O. To remove primers and concentrate the DNA for sequencing the Microcon ® YM-50 centrifugal filter unit and protocol was used.

2.2.5 DNA sequencing

The DNA sequencing undertaken in this project was done using the ABI prism d-Rhodamine sequencing kit (Applied Biosystems). The kit protocol was followed with the following amendments. The volumes used in the sequencing reaction were as follows. A volume of 20µl was made up of 600ng (10µl) SNAP minipep DNA (or in the case of Instagene/ PCR derived DNA, 3µl), 1µl dimethyl sulphoxide (DMSO), 1µl 3.2µM primer (for promoter studies this was either LACR or PMINTF2 and these sequences can be found in appendix II) and 8µl reaction mix (Applied Biosystems - supplied in kit). In the case of sequencing the PCR products, the remaining volume was made up with dH_2O. The sequencing control provided in the kit was used with prescribed volumes. For sequencing,

the following set of cycles were used; 1x 96°C at 2 mins, 25x [96°C for 30 secs. 60°C for 4 mins15 secs.], hold at 4°C. Precipitation was done as prescribed in the protocol and the pellet was re-suspended in 3µl of loading dye (made in a 5:1 ratio deionised formamide: 25mM EDTA (pH8.0) with 50mg/ml blue dextran). These samples were loaded onto a 5% polyacrylamide gel (5ml 10X TBE, 5ml 50% stock Long ranger solution (Applied Biosystems), 18g Urea, 250µl ammonium perosulphate (BIO-RAD), 35µl TEMED (BIO-RAD) made up to 50ml with millipure dH_2O. This gel was run for nine hours on an ABI prism 377 sequencing machine. Sequences were then transferred onto ABI-Prism software and could be analysed by using the Editseq and Megalign programmes which both part of the DNAStar software package.

2.2.6. Plasmids used in the promoter studies

Table 2. 1. The table below catalogues the various plasmids utilised in the promoter work undertaken in this study. The sequence that was of interest to this study comprised of 310 bp upstream of the start of the *M. tuberculosis recA* coding region which had previously been shown to be sufficient for promoter function (Movahedzadeh *et al.*, 1997a; Papavinasasundaram *et al.*, 2001). To differentiate between the various constructs used in this study, the positions of the fragments cloned are indicated using a numbering system counting from 1 as the base 310bp upstream of the coding region, which, therefore, starts at position 311 in this system.

PLASMID NAME	PLASMID DESCRIPTION	REFERENCE/ SOURCE
pEJ414	Integrative vector containing promoterless *lacZ* reporter gene with kanamycin resistance	Papavinasasundaram *et al.*, 2001
pEJ417	As pEJ414 with entire upstream region of *recA* under study (position 1- 345)	Papavinasasundaram *et al.*, 2001

PLASMID NAME	PLASMID DESCRIPTION	REFERENCE/ SOURCE
pEJ418	As pEJ414 with fragment comprising position 211-386 containing *recA*P1 only	E. O. Davis
pEJ419	As pEJ414 with fragment comprising position 50-250 containing *recA*P2 only	E. O. Davis
pEJ425	Integrative vector containing promoterless *lacZ* reporter gene with hygromycin resistance	E. O. Davis
pEJ431	As pEJ414 with fragment comprising position 50-224	E. O. Davis
pEJ435	As pEJ414 with fragment comprising position 104-224 containing inducible *recA*P2	E. O. Davis
pEJ444	As pEJ414 with fragment from position 104-161 (no promoter element)	E. O. Davis
pEJ445	As pEJ414 with fragment from position 104-224 including constitutive *recA*P2	E. O. Davis
pEJ449	As pEJ417 but with the P2A-12C mutation	E. O. Davis
pKKG4	Hygromycin resistant equivalent of pEJ435 using pEJ425 as vector	This study
pKKG5	As pEJ445 but with P2T-10G point mutation	This study
pKKG6	As pEJ445 but with P2A- 12C point mutation	This study
pKKG7	As pEJ445 but with P2C-14A point mutation	This study
pKKG8	As pEJ445 but with P2G-16T point mutation	This study
pKKG11	As pEJ445 but with P2C-17A point mutation	This study
pKKG12	As pEJ445 but with P2A-32G point mutation	This study
pKKG13	As pEJ445 but with P2C-36A point mutation	This study
pKKG14	As pEJ445 but with P2G-15T point mutation	This study
pKKG15	As pEJ449 but with P1-10 change (TCTAGT→**GAG**AGT)	This study
pKKG16	As pEJ449 but with the P1T-13G point mutation	This study
pKKG17	As pEJ449 but with the P1A-12C point mutation	This study
pKKG18	As pEJ449 but with the P1T-10G point mutation	This study

PLASMID NAME	PLASMID DESCRIPTION	REFERENCE/ SOURCE
pKKG19	As pEJ418 but with P1-10 change (TCTAGT→**GAG**AGT)	This study
pKKG20	As pEJ418 but with the P1T-13G point mutation	This study
pKKG21	As pEJ418 but with the P1A-12C point mutation	This study
pKKG22	As pEJ418 but with the P1T-10G point mutation	This study
pKKG23	As pEJ419 but with the P2A-12C point mutation	This study
pKKG24	As pEJ449 but with the P1 –35 region by position mutated (GCAGCC→ G**ACTA**C)	This study
pKKG25	As pEJ417 but with the P1A-12C point mutation	This study
pKKG26	As pEJ449 but with the P1 –35 region by homology mutated (TTGTCA →**GAC**TCA)	This study
pKKG27	As pEJ431 but with the P2A-12C point mutation	This study
pKKG28	As pEJ435 but with the P2A-12C point mutation	This study
pKKG29	As pEJ414 with fragment comprising of position 50-103	This study
pKKG30	As pEJ414 but with fragment comprising of position 50-161	This study
pKKG31	Hygromycin resistant equivalent of pEJ449 using pEJ425 as vector	This study
pKKG2mix	Randomised P2 sequence ligated to pEJ414 through a *Hin*dIII/ *Xba*I digest.	This study

2.2.7. Induction of cell cultures

For assays of the constitutive expressing *recA*P2 constructs in *M. smegmatis* (pKKG5-14), there was no need for induction with mitomycin C. These cells were harvested between the OD_{600} values of 0.4 and 0.6, washed three times with 1ml Z buffer with no β-

mercaptoethanol (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$ pH 7.0) and assayed for β-galactosidase activity.

For assays of inducible promoter constructs (pKKG15-pKKG30) in *M. smegmatis*, cell cultures grown in a shaking incubator at 37°C, to OD_{600} reading of 0.4- 0.6. At this point two 10ml culture samples were taken. One was induced by adding a volume of 1 in 1000 of 0.2 mg/ml mitomycin C, whilst the other remained uninduced. These cultures were then placed back in the shaking incubator for another five hours before being harvested. This time was chosen as it had been shown that there was optimum induction with mitomycin C at this time in the saprophyte (Papavinasasundaram *et al.,* 2001). Harvesting involved an initial spin of 2500 rpms for 10 minutes using a MSE Mistral 3000i bench centrifuge, after which, the cell pellet was washed in 1 ml Z- buffer without β–mercaptoethanol three times by re-suspension and spinning at 13, 000 rpm for 5 minutes in a microfuge. Finally, the cell pellet was frozen at - 20°C.

For *M. tuberculosis* the same procedure with the following alterations was used. The volume of culture used for each induction was increased to 30ml, with the period of induction was 24 hours. Again, this time was chosen as it had been shown that there was optimum induction of the complete region upstream of *recA* with mitomycin C at this time in the pathogen (Papavinasasundaram *et al.,* 2001). Harvesting was done with a SS-34 rotor spun at 10,000 rpms for 15 minutes. Cells were initially washed in 10 ml Z- buffer without β- mercaptoethanol and re-spun at 10, 000rpms for 5 mins (in SS-34 rotor), before undergoing two washes in 1 ml Z –buffer without β–mercaptoethanol as done for *M. smegmatis.*

2.2.8. β- galactosidase assay

Cells were resuspended in 1 ml Z-buffer (no β- mercaptoethanol) and transferred to 2ml tubes, one third full of glass beads (size 150-200 microns) and processed in a ribolyser (Hybaid) at speed 6.5 for 30 seconds. This lysate was chilled on ice for two minutes and then spun at 13,000rpms for 5 minutes. The supernatant was then removed and re-spun to give cell free extract. To determine the total cell protein level, the Pierce BCA kit and protocol were utilised. Standards used were 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, and 1.6 mg/ml. If samples were found to give a reading too high to be registered with any accuracy, these would be diluted four fold with the addition of 1X Z- buffer without β-mercaptoethanol (the result would be multiplied by four to give the correct protein concentration).

After the BCA assay was done, 1.5µl of neat (14.63M) β-mercaptoethanol was added to each extract. For the β– galactosidase assay, samples of the lysates were made up to 500 µl with 1x Z-buffer containing β-mercaptoethanol (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$ pH 7.0, 38mM β-mercaptoethanol). These samples were then incubated at 28°C for 5 minutes to equilibrate their temperature before 100µl of 4mM o-nitrophenyl-β-D-galactopyranoside (ONPG) was added. The addition of this compound initiates a colour change from clear to yellow when in the presence of β–galactosidase. Once a visible amount of colour change had been seen or after a period of 30 minutes (whichever occurred first), 250µl of stop solution (1M Na_2CO_3) was added to arrest the colour change. The time taken for the colour change to occur in minutes (time taken from addition of ONPG to addition of stop solution) was at least 10 minutes. Duplicate assays were performed for

each sample and the A_{405} value determined using an ELISA plate reader. The units of β-galactosidase/ mg protein which would give an indication of promoter activity, were calculated using the formula (Miller, 1972):

$$\text{Units/ mg protein} = \frac{380 \times A_{405}\ \text{Reading}}{\text{Time (min)} \times \text{cell extract volume assayed (ml)} \times \text{Protein concentration (mg/ml)}}$$

2.3. Methods used in the *M. tuberculosis* sigma factor studies

2.3.1. Extraction of genomic DNA from mycobacteria

In the case of isolating *M. tuberculosis* DNA from a solid culture, bacteria covering a complete 7H11 plate was taken and re-suspended in 300 µl of TE buffer (10mM Tris pH8, 1mM EDTA). To this was added lysozyme and lipase (both prepared in TE) to get a final concentration of 2mg/ml each as well as 5 µl of DNase-free RNase (Boehringer). This mixture was incubated at 37°C for 2 hours before being snap frozen and then incubated at 75°C for 10 min. After this, the sample was cooled to room temperature before proteinase K 500 µg/ml (final concentration) and SDS 0.5% (final concentration) were added and the sample incubated at 50°C for 1 hour. After two rounds of phenol: chloroform extraction and one with just chloroform, DNA in the aqueous phase was added to 0.1 volumes of 3M sodium acetate (pH 5.2) and 1 volume of isopropanol and mixed by gentle inversion. The DNA, which precipitated out as a thread, was pelleted by being spun at 13,000 rpms for 15

min in a microfuge. The pellet was then washed with 70% ethanol, air-dried, dissolved in 400µl of TE buffer, and stored at 4°C.

For DNA extraction from liquid cultures of *M. tuberculosis*, 24 hours prior to harvesting the cells, 0.1 volumes of 2M glycine was added to the incubating culture. After pelleting of the cells (10,000 rpms for 20 min at 4°C in a GSA rotor), these cells were re-suspended in 30 ml SET solution (0.3M sucrose, 50mM Tris pH8, 10mM EDTA). This was then centrifuged for 15 min at 10, 000 rpm with GSA rotor. The pellet was resuspended in 1 ml of SET solution containing 2mg/ml of both lysozyme and lipase and left to incubate at 37°C for 30 min. After this, 4 ml of GSE solution (6M guanidinium chloride, 1% sarkosyl, 20mM EDTA) was added and the sample was left to incubate at 37°C for a further 2 hours. The sample was chloroform extracted and ethanol precipitated before being placed at -20°C overnight. After being spun for 13, 000 rpms in a microfuge for 5 minutes the pellet was re-suspended in a mixture of 400 µl TE, 2 µl RNase A (500 µg/ ml), 20 µl proteinase K (10mg/ ml) and 10 µl 20% SDS, the sample was incubated at 37°C until it was fully dissolved. After phenol:chlorofrom extraction and ethanol precipitation, the purified sample was finally re-suspended in TE buffer.

To measure the concentration and purity of DNA isolated, its optical density at 260nm and 280nm was measured. Its purity was determined by the ratio of its 260nm: 280nm readings with a value of ~1:1.8 indicating a pure sample. It concentration was calculated using the following formula:

DNA concentration (µg/ml) = A_{260} x dilution factor x 50 (constant)

2.3.2.Designing of primers and PCR reaction

Primers utilised for the amplification of the sigma factors and flanking regions were designed through the use of the DNAStar programme, PrimerSelect, and sequences were derived from the Tuberculist web server (http://genolist.pasteur.fr/TubercuList/.). For PCR the following mixture was made. In a final volume of 50 µl was added 5 µl Cloned Pfu buffer, 5 µl 2.5mM dNTPs, 2.5 µl DMSO,2X 2µl of a 2.5µM stock of forward and reverse primers, 10ng of genomic DNA, 1 µl Pfu turbo enzyme, with the remainder being made up with sterile dH_20. The annealing temperature used was 63°C and the extension time set to 4 minutes for 25 cycles with an additional time increment of 30 seconds every cycle for all the reactions.

Table 2.2. This table is composed of primers utilised in the PCR amplification of sigma factors and flanking sequence

Primer name	**Primer Sequence**
sigB1(forward primer)	5' CTAGATCTGGTGGCGTCGCTGGGCTTCTTGAT 3'
sigB2 (reverse primer)	5' CTAAGCTTCTGGCGGGCGCTGGCATTGTGT 3'
sigC1(forward primer)	5' ATAGATCTACGGCCGTCGGTGATTTTTCTTG 3'
sigC2 (reverse primer)	5' ATAAGCTTCGGGTAATACGGCGCCTTGTGGA 3'
sigD1(forward primer)	5' TTGGTACCGTTGCCGACGACCCCTGTGACC 3'
sigD2(reverse primer)	5' TTAAGCTTCCTTGCCGCGCCGTTGTTCTGGAT 3'
sigE1(forward primer)	5' ATAAGCTTGGGGCCGGCACCGTCTTTCTC 3'
sigE2 (reverse primer)	5' ATGGATCCTGTGGTGCCCTCGATGCTGATGTC 3'
sigG1(forward primer)	5' AAAGATCTAAGCCCGCCAATAGCAGCAAGAGG 3'
sigG2 (reverse primer)	5' GGATGCATAGCGCAAGCACGTCGTCCACATAA 3'

2.3.3. Production of the knock out construct plasmid.

To obtain insert fragments (PCR products or digested fragments) from agarose gels, the Qiagen gel– extraction kit and protocol were used. Samples were eluted in 30µl of dH_2O. In the case of PCR fragments for sigma factor mutants that were introduced blunt ended, 17µl of this eluate was kinased with the addition of 2µl of 10X kinase buffer (see use of designed oligonucleotides for composition) and 1µl of T4 PNK polynucleotide kinase (Roche). This mixture was incubated for one hour at 37°C and then heat-treated at 60°C to inactivate the enzyme. For ligations, 10ng of vector DNA was used with 2µl of treated mixture. In all but one case, the vector used was pBackbone (Table 2.4), the exception being the case of the Δ*sigC* construct, which was made with pBluescript KS as the vector.

To create a deletion within the cloned sigma factor gene, inverse PCR was utilised. Primers were designed by the PrimerSelect programme and incorporated an enzyme restriction site, which would be in the centre of the deleted region when the plasmid re-circularised. The PCR conditions were modified from those used to amplify up the fragments. In this case the annealing temperatures were changed to those in Table 2.3 for each deletion and the extension time was increased to 8 minutes with an additional 30 seconds being added at the end of each cycle.

Table 2.3. The table below is comprised of primers that were used in the inverse PCR reactions that were used to delete part of the target sigma factor in each instance. The sequence highlighted in bold shows the restriction site that was introduced to allow for marking of the mutation.

primer	Sequence	Anneal temp (°C)
i-SigBF	5' **GATATC**CGAGCGCGACGTGATGAGTAAGC 3'	64°C
i-SigBR	5' CTTGCCGATGCCGTTCAGATAGA 3'	
i-SigCF	5' **TACGTA**CGGACAGATAGGCGACGAA 3'	58.5°C
i-SigCR	5' ATGCATCTCACCGGCTAGGCAGACC 3'	
i-SigDF	5' **TACGTA**GAGATTGTTGCGGCAGGTGACTAT 3'	64°C
i-SigDR	5' TACTGAGACACATCCCGGGCTAACTCC 3'	
i-SigEF	5' **GTTAAC**ACCCCGAACATGGCGAGTG 3'	64°C
i-SigER	5' **GTTAAC**CAAAGTTGCGATTCCHTATTCC 3'	
i-SigGF	5' TA**CCTAGG**GCCATCGTCACCCTCATTCACCAA 3'	63°C
i-SigGR	5' TA**CCTAGG**TCGGTGTGGGCGGAGAAGT 3'	

The resulting reverse PCR samples were then treated with 1µl *DpnI* to remove the plasmid template and gel extracted as described previously. For all but the production of the Δ*sigG* mutant, the fragments were kinased as mentioned previously. In the case of the Δ*sigG* mutant, five 50µl PCR reactions were pooled together and treated with 3µl of *DpnI* for three hours at 37°C. After this incubation the digestion was purified by phenol extraction and concentrated by ethanol precipitation. The resulting pellet was dissolved in 44µl 1X DNA dilution buffer (Roche) to which 5µl of buffer 2 and 2µl *AvrII* (both New England Biolabs) were added. This digest was left to incubate at 37°C overnight.

Table 2.4. This table is composed of all the plasmids used in the sigma factor work mentioned in this study.

Plasmid	**Description**
pBluescriptKS-	Commercial vector possessing *lacZ* gene and ampicillin resistance gene (2961bp) (Stratagene)
p*Aghori*	Modified pBluescript KS with 159bp of the *lacZ* cassette removed through site directed mutagenesis, and introducing an unique *PacI* site. (2.767 kb)
pUC4K	Commercial plasmid carrying *aph* gene conferring kanamycin resistance (Pharmacia)
pBackbone	As p*Aghori* but with 1.256kbp *SalI* digested fragment from pUC4k containing Kanamycin resistance for marker (4.071Kbp)
pGoal17	Based on pBluescript but carrying the 6.367kbp double marked *lacZ- sacB* cassette which could be removed by *PacI* digestion. 8.855kbp (Parish *et al.*, 2000)
pUCHyg	Vector carrying Hygromycin resistance gene (4.312kbp) (E. Mahanthiralingham *et al.*, 1998)
pBalmedie1	As pBackbone but with a 4.1kb PCR fragment, containing *sigB* with flanking sequence inserted at *Eco*RV site.
pBalmedie2	As pBalmedie1 but with 661bp of *sigB* removed through inverse PCR and the insertion of *Eco*RV site in its place.
pBalmedie3	As pBalmedie2 but with 6.367kbp *Pac*I digested double marked *lacZ- sacB* cassette
pBluescriptsigC	As pBluescriptKS but with a 4.1KbPCR fragment containing *sigC* and flanking sequence inserted in at *Eco*RV site.
pBluescriptsigCΔsigC	As pBluescriptsgC but with 405bp of *sigC* removed through inverse PCR and the insertion of a *SnaBI* site in its place.

Plasmid	**Description**
pCulter1	As pBluescriptsigCΔsigC but with site directed mutagenesis done to remove 159bp of pBluescript KS vector *lacZ* gene.
pCulter2	As pCulter1 but with the addition of the *Xho*I digested Kan^R cassette from pUC4K.
pCulter3	As pCulter2 but with 6.367kbp *Pac*I digested double marked *lacZ- sacB* cassette
pDyce1	As pBackbone but with 3.8Kb PCR fragment containing *sigD* and flanking sequence inserted in at *Eco*Rv site.
pDyce2	As pDyce1 but with 564bp of *sigD* removed through inverse PCR and the insertion of *Sna*BI site in its place.
pDyce3	As pDyce2 but with the insertion of the hygromycin resistance gene (1.479Kb digested with *HpaI*) and inserted at *Sna*BI site.
pDyceHyg	As pDyce3 but with 6.367kbp *Pac*I digested double marked *lacZ- sacB* cassette
pElrick1	As pBackbone but with a 4.3 kb PCR fragment containing *sigE* and flanking sequence inserted in at *Eco*RV site.
pElrick2	As pElrick1 but with 681bp of *sigE* removed through inverse PCR and the insertion of a *HpaI* site in its place
pElrick3	As pElrick2 but with 6.367kbp *Pac*I digested double marked *lacZ- sacB* cassette
pGarthdee1	As pBackbone but with a 4.3 kb PCR fragment containing *sigG* and flanking sequence inserted in at *Eco*RV site.
pGarthdee2	As pGarthdee1 but with 691bp of sigG removed through inverse PCR and the insertion of a *Avr*II site in its place
pGarthdee3	As pGarthdee2 but with 6.367kbp *Pac*I digested double marked *lacZ- sacB* cassette

2.3.4. Identification of possible sigma factor mutants

2.3.4.1. Selection for possible knock out mutants

Colonies of *M. tuberculosis* transformed with the Δ*sigB*, Δ*sigC*, Δ*sigE*, and Δ*sigG* construct were initially selected for blue coloration and kanamycin resistance (single crossover phenotype). These were then streaked onto 7H11 plates without any selection pressure to allow the second crossover event to occur. The culture formed on this plate was then harvested and suspended in a 2ml screw topped tube containing 1ml of saline and glass balls (2.5-3.5 mm). This tube was then vortexed for 1 minute to homogenise the suspended culture. Serial dilutions of 10^{-1}, $10^{-2}$and 10^{-3} were made from each homogenised culture and 100μl of each dilution was plated on 7H11 plates and selected for no coloration and sucrose resistance (the phenotype of the double crossover). Colonies positive for the selection criteria were then taken and replica streaked on plates with and without kanamycin. Colonies that were found to be sensitive for this antibiotic were taken for further testing.

Unlike the lengthy selection procedure employed for the other four sigma factor mutants, colonies of *M. tuberculosis* transformed with the Δ*sigD* construct were selected for white coloration and hygromycin resistance (double crossover phenotype) directly. These were then streaked onto 7H11 plates with and without kanamycin. Colonies that were found to be kanamycin sensitive in this screen were selected for further testing.

2.3.4.2. PCR amplification of mutated region

A number of colonies for each possible sigma factor mutant had been isolated in the screening procedure mentioned above. However, before undertaking genomic analysis through Southern blot, it was decided to confirm that the colonies isolated were indeed mutants. This was done through PCR. The Instagene method of extracting DNA was annealing temperatures and extension times varied as shown in the following table .employed for a quick determination of possible sigma factor mutants. Primers for each set of reactions were designed through the DNAStar program, PrimerSelect.

Table 2.5. This table is comprised of the primers used in the PCR amplification of DNA extracted from various *M. tuberculosis* isolates through the use of the Instagene mix method (BIO-RAD)

Primer	Sequence	Anneal temp (°C)	Extension time (mins)
ins-sigBF ins-sigBR	**5'** GGGTCCCCGCGCTCATTCTGG **3'** **5'** GCCGCCGCCTGGGGTGGTTT **3'**	65°C	2 mins 30
ins-sigCF ins-sigCR	**5'** CCAAGCCGGGCATCGTATCTG **3'** **5'** CGCGGCGGCACAAACTGG **3'**	62°C	1 min
ins-sigDF**h** ins-sigDR**h** ins-sigDF ins-sigDR	**5'**GTCCAGCACCGGGGCGAGAGGTA **3'** **5'**ACTGGCCGGCGGGGTTGCTGTCAT **3'** **5'**TCAGCCCGACACGAAGCGAGAACT **3'** **5'**AAGCCGGCCCGTACCCTGTGAATA **3'**	65°C 65°C	3mins * 3mins
ins-sigGF ins-sigGR	**5'** TTGCCGGCGGAATCCTGGTGTCTG **3'** **5'** CTGGTTGCGCTCCGGCTCGTTGAA **3'**	66.4°C	2mins 30

* Since the sigD mutation was marked with the hygromycin resistance gene, two primers were based on the marker fragment whilst the other two were based on the flanking region.

Although this method provided a quick screen to determine possible mutants, a positive result could only be shown conclusively by Southern blotting.

2.3.4.3.Southern blot analysis

2.3.4.3.1. Designing of probes for Southern bolts

Probes ranging from 400bp –600bp were designed using the DNAStar program, PrimerSelect and made through PCR amplification. Primers used to produce these probes are shown below.

Table 2.6. This table is composed of the various primers used in the production of probes for the Southern blots done in this study. All probes except the one based on the deleted region of *sigC* were about 400bp long. The probe used based on the *sigC* deleted region was about 300bp long

Primer	Sequence	Probe location
SouthsigC1 SouthsigC2	**5'** CGGGCGGTGGCGGTGGTGTCC **3'** **5'** GTCGATTTCGGCGGGTTCAGGTC 3'	5' out with PCR flanking region
SouthsigC3 SouthsigC4	5'GCTCCAGCGCCCGAACTTG **3'** **5'**CCGCATAGGACAGGCCCGAGCAG **3'**	Probe based on the deleted region of *sigC*

Primer	Sequence	Probe location
SouthsigC5 SouthsigC6	**5'**AGCAACCGCAGCGCAACAAACATT **3'** **5'**CGCGGTGCTGGCCTCGGGATACT **3'**	3' out with PCR flanking region
SouthsigD1 SouthsigD2	**5'** CGCCTCGCCGTCGCAGAACAA **3'** **5'** GACCGCGTACGCCTCCGACTCC **3'**	5' within PCR flanking regions
SouthsigD3 SouthsigD4	**5'** GGAGGCCGTGGCAGGGAGAC **3'** **5'**CAGGCCGACGACAACACGCAGAAT**3'**	Probe based on deleted *sigD* region
SouthsigD5 SouthsigD6	**5'** GACCGCGACCGCTTTCAAC **3'** **5'** GGCACCACCACCGCAACCATCTT **3'**	3' out with PCR flanking region

For each isolate (double or single crossover and the wild-type H37Rv control) 3.5µg of genomic DNA was digested in a 30µl volume containing 2µl of selected restriction enzyme (selection was based on sequence analysis made using the DNAStar programme, MapDraw) overnight at 37°C. A 1% TAE based agarose midi-gel was prepared and the digested samples were run overnight at 30V to provide proper separation of the samples on the gel. In addition to the samples, λ- *Hin*dIII and λ- *Hin*dIII/ *Eco*RI ladders were run in order to measure where the different sizes of DNA would run on the gel.

2.3.4.3.2. Preparation of the gel and transfer.

After being run, measured and photographed the gel was then placed in 0.25 M hydrochloric acid and shaken moderately at room temperature for around 15 minutes (when the bromophenol blue dye in the gel has turned yellow). This step was done to make the transfer of bands greater than 10kb better. After this step, the gel was rinsed with dH_2O and placed in denaturation buffer (1.5M NaCl , 0.5M NaOH) and gently shaken at room temperature for 30 minutes. The gel was then rinsed again in dH_2O and placed in

neutralisation buffer (1.5M NaCl, 0.5M Tris-HCL pH7.2, 1mM EDTA) and gently shaken at room temperature for another 30 minutes before being rinsed for the final time in dH_2O. The gel was now ready to be transferred.

The apparatus and set up for the transfer onto Hybond-N+ nylon transfer membrane (Amersham Pharmacia Biotech) was done as prescribed by Sambrook *et al.* (1989) with the following amendment. Once dried, the transfer was fixed by placing the membrane in a UV stratolinker (Spectrolinker XL-1500 crosslinker. Spectronics Corporation) and fixed with a burst of 1.2J/ cm^2 for 20 seconds

2.3.4.3.3. Hybridisation

The fixed membrane was placed inside a screw top flask (HYBAID) with 10ml SDS-Phosphate buffer (1M Na_2HPO_4 pH7.2, 20% SDS), preheated to 65°C and placed in a rolling oven at 65°C. To label the probe, 50ng of probe DNA was denatured at 95°C for 3 minutes and then immediately placed on ice for another 2 minutes. The mixture was centrifuged briefly and 10µl of Reagent mix (Oligolabelling Kit 40 reactions – Amersham), 3µl Redivue $\alpha^{32}P$ dCTP (Amersham) and 1.5µl Klenow fragment (Oligolabelling Kit-Amersham) were added. The mixture was then gently mixed and incubated at 37°C for 1 hour. The reaction was stopped with the addition of 50µl 10mM EDTA and then denatured again at 95°C for 2 minutes. The mixture was then added to the flask containing the membrane/ SDS- phosphate buffer and left rolling at 65°C overnight.

To remove the excess radioactivity, the membrane was subjected to four washes in varying concentrations of SSC (2% for two washes and one each in 1% and 0.5%) in which 0.1% SDS (v/v) was also added. Each was lasted 20 minutes as was undertaken agitated at a temperature of 65°C. The film used to expose the blot was Kodak Scientific Imaging Film Biomax MR (Kodak) and exposures were made at 4 hours and overnight

2.3.5. Micro-array analysis of *M. tuberculosis* Δ*sigC*

Liquid cultures of *M. tuberculosis* Δ*sigC* and wild-type H37Rv were harvested at an OD_{600} value of between 0.3 and 0.35 for exponential growth and between 1.8 and 1.85 for stationary growth.

The Hybaid recovery RiboLyser kit- Blue (Hybaid) and protocol were used to extract RNA from 100ml of *M. tuberculosis* liquid culture, with the following amendment. The cell pellet was re-suspended in 600µl of reagent A and the final RNA pellet was dissolved in 50µl DEPC treated water. In addition 1.5µl of RNAsin (Promega) was added to the final RNA extract.

The RNA extract was cleaned up as follows. To the 51.5µl RNA/RNAsin mixture was added 5µl DNase (RNase free) (Promega), 2µl RNAsin (Promega), 7.5µl 0.1M $MgSO_4$, 5µl 3M Sodium acetate (pH6) and made up to 150µl with the addition of DEPC treated water. This was incubated at 37°C for an hour. After this step, another 2µl DNase and 1µl of RNAsin was added and re-incubated for a further hour at 37°C. After this second incubation the RNeasy kit (Qiagen) and protocol was used for final clean up.

To measure the concentration and purity of RNA isolated, its optical density at 260nm and 280nm was measured. Its purity was determined by the ratio of its 260nm: 280nm readings with a value of ~1:2 indicating a pure sample. It concentration was calculated using the following formula:

RNA concentration (μg/ml) = A_{260} x dilution factor x 40 (constant)

To get an accurate result from the microarray, for each mutant sample tested, two slides were used in which the knockout was labelled with either Cy3 or Cy5. For labelling each sample (whether wild-type or mutant), the following mixture was made up which comprised of 5μg RNA and 6μg Random primers (2μl of 3μg/μl stock -Invitrogen) which was made up to 11μl with the addition of DEPC treated water. This mixture was heated at 95°C for 5 minutes before being snap cooled on ice and briefly centrifuged. At this point, to each mixture was added 5μl 5X first strand buffer (Invitrogen), 2.5μl 100mM DTT, 2.3 μl dNTPs (5mM dA/G/CTP, 2mM dTTP), 1.7 μl Cy3 or Cy5 dCTP (Amersham), 2.5 μl SuperScript II (200U/μl - Invitrogen). This new mixture was then incubated at 25°C for 10 minutes followed by 42°C for 90 minutes. During this incubation time the slides were pre-hybridised by placing them in a Coplin jar containing a 50ml solution preheated to 60°C which was made up of 3.5XSSC, 0.1% SDS, 10mg/ml BSA and autoclaved water. The slides were incubated in this solution for 20 minutes at 60°C before being rinsed for 1minute, firstly in distilled water and then in propan-2-ol. Placing them individually in 50ml centrifugation tubes and spinning them at 1500 rpm for 5 minutes dried the slides. Slides were then stored until required.

After incubation, the labelled samples were briefly centrifuged before being combined (Cy3 mutant- Cy5 wild-type and vice versa) in a single tube before being resuspended in a volume of 13.5µl autoclaved water using the Qiagen PCR mini-elute kit (QIAGEN). Of this eluate, 10.5µl was taken and added to the hybridisation solution (3.2µl 20X SSC, 0.23µl 20% SDS). This mixture was then at heated at 95°C for 2 minutes before being cooled slightly and briefly centrifuged. The resulting sample was pipetted onto the edge of the array area on the pre-hybridised slide before being dragged into position with the use of a cover slip, which was then lowered gentle over this arrayed area. The slide was then place in a hybridisation cassette and immersed in a water bath at 60°C for between 16-20hrs.

After this incubation time, the slides were first washed in Wash A (1XSSC, 0.05% SDS) to remove the cover slips before being placed in a slide rack and washed by agitation in Wash A for a further 2 minutes. The slides in the rack were then washed by agitation in Wash B (0.06XSSC) for 2 minutes before being transferred to fresh Wash B for a further 2 minutes of washing by agitation. Placing them individually into 50ml centrifuge tubes and centrifuging for 5 minutes at 1500rpm then dried the slides. Slides were scanned using a ScanArray 3000 dual-laser confocal scanner and the resulting data was analysed using the Gene Pix and Gene- Spring programmes (Silicon Genetics).

CHAPTER 3: THE IDENTIFICATION OF THE *M. tuberculosis recA* PROMOTERS.

3.1. Introduction

The *M. tuberculosis recA* gene was originally isolated through hybridisation with a probe containing part of the *E. coli recA* gene (Davis *et al.*, 1991). It was shown that although the *M. tuberculosis recA* product could complement an *E. coli* Δ*recA* mutant, this could only be achieved when the mycobacterial gene was expressed from a vector promoter (Davis *et al.*, 1991), indicating that the *M. tuberculosis recA* promoter must differ from standard *E. coli* promoters. However, this finding was not unusual as there are other instances where mycobacterial promoters are not recognised by the *E.coli* transcriptional machinery (Das Gupta *et al.*, 1993).

Primer extension from the *M. tuberculosis recA* gene into the upstream region gave rise to two products, one 47bp from the translation initiation codon and the other 93bp from this site (Movahedzadeh *et al.*, 1997). This suggested that there might be two regions of promoter activity in this fragment. From comparisons with other bacterial promoter sequences, two regions, termed *recA*P1 (proximal to the *recA* gene) and *recA*P2 (distal from the *recA* gene), were theorised to be the sites of this activity (see fig. 3.1). The presence of two promoters for the *recA* gene was subsequently confirmed by cloning these regions separately before a *lacZ* reporter gene (E. O. Davis, personal communication).

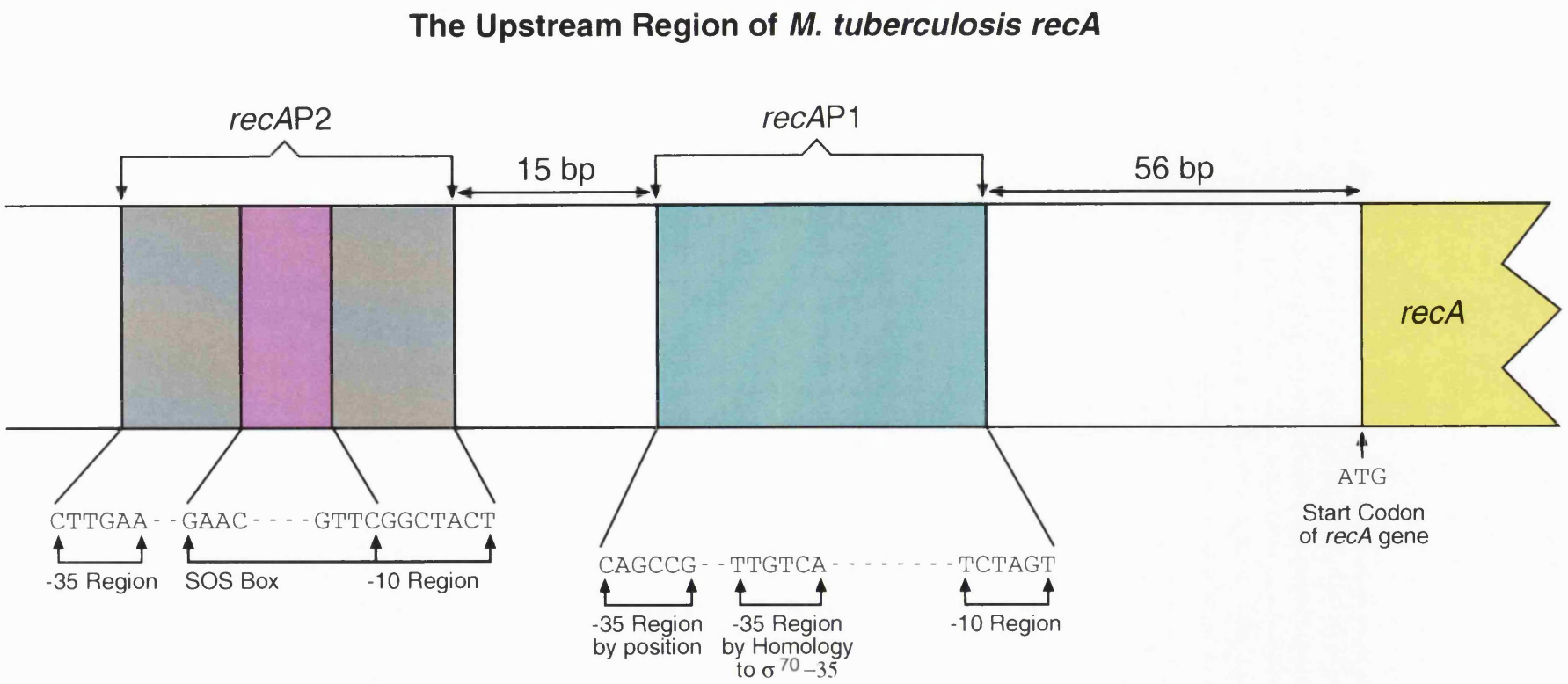


Figure 3.1. This cartoon shows the preliminary positioning of both *recA*P1 and *recA*P2 in relation to the start codon of the *M. tuberculosis recA* gene based on the findings of Movahedzadeh *et al.*, (1997a). It also shows the relative position of the SOS box, which is located within *recA*P2.

Whilst the *recA*P2 region contained a LexA binding site between the putative –10 and –35 sequences, there was no SOS box present at the *recAP1* site. However, no studies to determine which sequences are important for promoter activity had been undertaken.

3.2 Analysis of the *recA*P2 promoter site.

From the primer extension studies, promoter activity was localised to two regions in the upstream fragment. The *recA*P2 region was chosen as the –35 region it had resembled that of the σ^{32} promoter, although the same could not be said of the –10 region (Movahedzadeh *et al.,* 1997). However, the –10 region selected resembled the consensus found in a second set of promoters called the gearbox promoters although the –35 region did not match so well to this consensus (fig 3.2). To determine whether these sequences are important for expression from *recA*P2 in mycobacteria it was decided to examine a set of mutant sequences with changes at various positions within these motifs for alterations to the expression level.

The binding site for LexA was located between the putative –10 and –35 sequences for *recA*P2. Therefore, to dissociate the effects of induction and of promoter activity, promoter mutations in *recA*P2 were examined in the context of a mutation in the SOS box (GAACAGGTGTTC→GAAC**G**GG**C**GTTC) which prevents LexA binding and thus renders expression constitutive (Davis *et al.,* 2002). In addition, it was found that sequence further upstream of *recA*P2 enhanced the expression levels of this promoter.

Figure 3.2. This figure shows the comparison between the *E. coli* σ^{32} promoter consensus (top), *E. coli* gearbox consensus (bottom) and *M. tuberculosis recA*P2 sequence. From this diagram it can be seen that the mycobacterial promoter could belong to either group.

It was therefore decided to ligate these mutant sequences to a promoterless *lacZ*-integrating vector which also contained this upstream "activating" sequence (pEJ444). The complete constructs were then transformed into *M. smegmatis* and tested for promoter activity.

Individual base changes at six positions in the –10 region and two in the –35 region were assessed for their effects on promoter function in the saprophyte (fig 3.3). From the eight point mutations in the proposed *recA*P2 region, three mutants, P2A-12C (pKKG6), P2C-14A (pKKG7), and P2C-36A (pKKG13) had an effect in reducing promoter activity in comparison with the wild type. Conversely, two mutations, P2G-16T (pKKG8) and P2C-17A (pKKG11), had the effect of dramatically increasing *recA*P2 activity. In fact, the point change at P2A-12C led to an almost complete loss of activity when assayed. However, it was not clear as to whether this mutation would have the same effect on the wild-type sequence and so this change was introduced in pEJ435 to make pKKG28, which was the same as pKKG6, except for containing a functional LexA binding site. In assays in both *M. smegmatis* and *M. tuberculosis*, where this construct was induced with mitomycin C, it was found that expression was reduced to background levels (fig. 3.4). The other three changes employed, P2T-10G (pKKG5), P2G-15T (pKKG14) and P2A-32G (pKKG12) did not have a significant effect on the expression level of this promoter.

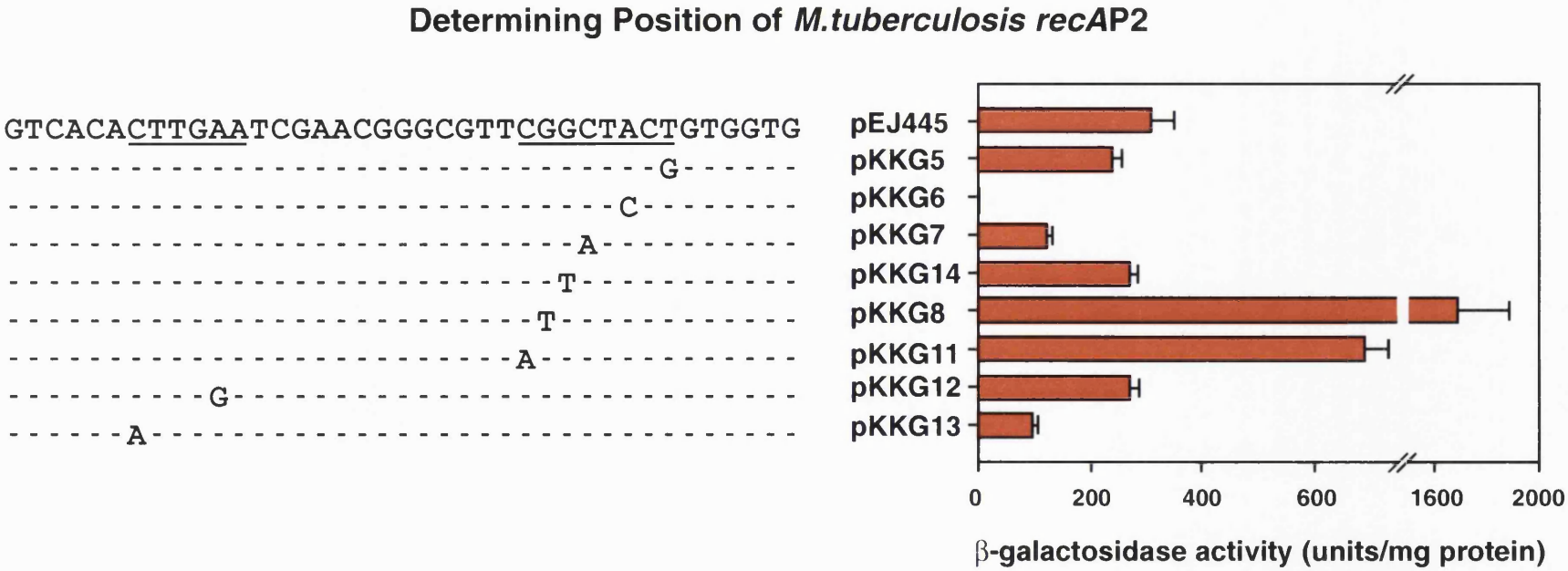


Figure 3.3. This figure shows the results of introducing point mutations into the *recA*P2 sequence and assaying for activity. Eight different oligonucleotides incorporating a single base change in either the –10 or –35 region of *recA*P2 as well as a double point mutation in the SOS box to make the resulting constructs constitutive were made and cloned into *Xba*I / *Hin* dIII digested pEJ444. These constructs were then transformed into *M. smegmatis* and assayed for activity compared to wild-type constitutive *recA*P2 (pEJ445). The error bars represent standard errors between three independent experiments per promoter construct. It was found that the *recA*P2A-12C (pKKG6) mutation virtually eliminated promoter activity whilst the *recA*P2G-15T mutation gave over a four-fold increase in activity compared with wild-type.

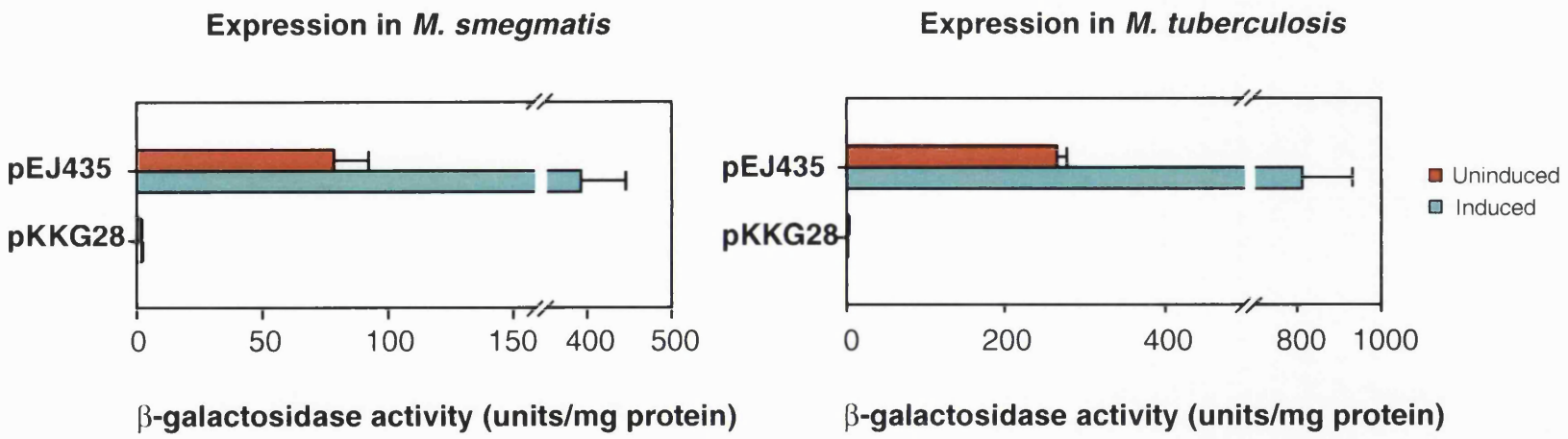


Figure 3.4. This figure shows how the expression of an inducible *recA*P2 promoter is affected by the *recA*P2A-12C mutation. It was unclear as to if the double mutation in the SOS box in the constitutively expressing pKKG6 construct had a part in the elimination of promoter activity. The *recA*P2A-12C was therefore introduced into an inducible construct (pEJ435) through site directed mutagenesis (giving pKKG28) and transformed into both *M. smegmatis* and *M. tuberculosis*. Induction was with mitomycin C and was undertaken for 5 hours in the saprophyte and 24 hours in the pathogen. The error bars represent standard errors between three independent experiments per promoter construct. When compared with the wild type in both species it is shown that the *recA*P2A-12C virtually eliminates *recA*P2 activity, confirming the importance of the –12A in the function of this promoter.

3.3. Random mutagenesis of the *recA*P2 sequence.

The individual mutations mentioned previously had helped identify and position the –10 and –35 elements of the *recA*P2 promoter. However, it was still not clear as to how much of the sequence was important in the function of this promoter. It has been shown in other studies that the region between –10 and –35, also known as the spacer region, can have a role in promoter function. This is surprising as there is no real conservation of sequence in this region between promoters that recognise the same sigma factor. It was, therefore, of interest to see how sequence sensitive was the whole *recA*P2 region. Such a study would require point mutations in each base, changing each in turn to encompass all possibilities. However, to attempt such a study of the *recA*P2 region, which comprises 36 nucleotides, would be expensive and time consuming.

An alternative method would be to use random mutagenesis within this region and to this end a method based on that used by Huang and Helmann (1998) was utilised. In their study, which was to determine a consensus sequence for the *B. subtilis* σ^X factor, they used saturation mutagenesis (Huang and Helmann, 1998). Saturation mutagenesis involves introducing small amounts of non-cognate bases in the production of a long oligonucleotide and then, through priming and extension with Klenow polymerase from the 3' end, producing a mixture of oligonucleotide pairs that would each contain on average one point mutation. In this way, Huang and Hellmann managed to produce 71 mutations in 25 positions, which were sequenced and transformed for subsequent selection of promoter activity on X-Gal plates.

In the case of the *recA*P2 sequence the 37 nucleotide sequence to be randomised comprised the –10, -35 and spacer regions plus an extra 5bp on either end. However, the oligonucleotide ordered contained an additional 18bp at either end comprising of 10bp of conserved upstream or downstream sequence plus an *Xba*I site with extra 2bp for digestion at the 5' end or a *Hin*dIII site with an extra 2bp for digestion at the 3' end. Initially a percentage mix of 97.3% cognate to 2.7% non-cognate bases was utilised for the random oligonucleotide and the resulting double stranded mutant sequences were ligated in the same plasmid used in the directed analyisis of *recA*P2 (pEJ444). The resulting constructs were then transformed into *M. smegmatis* mc^2155 cells and plated on 7H11 Kan^{25} $X\text{-}Gal^{100}$ plates. White colonies, a phenotype indicating a promoter null mutant, were picked and sequenced to identify the mutations they contained. However, it was found that the null-mutants isolated were caused by multiple as opposed to single base changes (data not shown).

In an attempt to shift the production of oligonucleotides towards more single and less multiple mutations the ratio of cognate to non-cognate bases used to make the oligonucleotide was increased to 98.5:1.5. The resulting double-stranded sequences from this new mixture were again ligated into pEJ444 and transformed directly into *M. tuberculosis* H37Rv. Plates appeared to have slightly more blue than white colonies and 120 white colonies were selected for DNA extraction, PCR amplification and sequencing. However, no amplification products were obtained from these mutants, even after a number of DNA dilutions had been used. It was, therefore, decided to determine the sequences from blue colonies to obtain an indication of the important bases through elimination of mutations that did not adversely affect promoter activity. However, it was found by

sequencing that none of 8 blue colonies contained the random oligonucleotide. A possible explanation for this was that an *Xba*I site originally used to clone the upstream enhancing region had been changed to an *Nco*I site in this construct. Although the clone containing this fragment with the *Xba*I site had no promoter activity it is possible that this change might have resulted in promoter like activity.

After this result, it was decided to replace the pEJ444 vector with the basic integrative promoter-less *lacZ*-carrying vector, pEJ414 (Papavinasasundaram *et al.,* 2001). Following transformation of *M. tuberculosis* there was roughly a 1:1 ratio of blue to white colonies on the 7H11 Kan^{25} X-Gal plates. From these 96 blue and 96 white colonies were taken for analysis. However, although it was possible to amplify DNA for sequencing from the white colonies, it was found that their lack of promoter activity was due to a deletion of part of the oligonucleotide. Therefore, it was decided to continue as previously and screen blue colonies to identify mutations that did not have a negative effect on activity. Sequence was obtained from 82 colonies, which revealed 42 different point mutations at 27 sites, which could be tolerated without eliminating promoter function (see fig. 3.5).

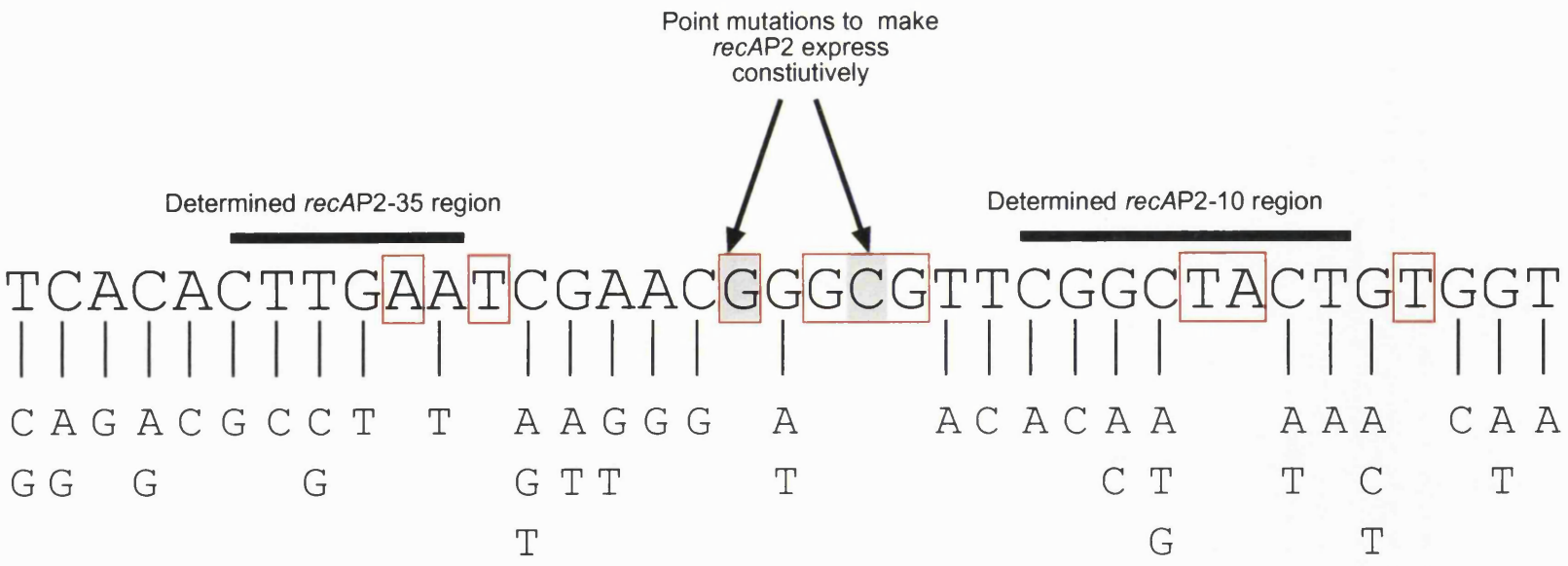


Figure 3.5. This figure illustrates the random mutagenesis of *recA*P2 study. A sequence of 37 nucleotides, which incorporated the *recA*P2 –10, -35 and spacer region as well as 5 bases flanking at either end was used as the basis for the "spiked oligonucleotide" used in this study. Through saturation mutagenesis, on average one point mutation was incorporated into each oligonucleotide. Each oligonucleotide also contained a double mutation in the SOS box to make the construct constitutive (shaded). The resulting mixture of oligonucleotide were then introduced into *XbaI*/*Hin*dIII pEJ444 and transformed into *M. tuberculosis*. Positive screening for blue colonies on X-Gal plates was used and the DNA from these colonies isolated and sequence. From this study it was observed that only the regions boxed in red remained unchanged in all the *recA*P2 "active" colonies suggesting that some of these nucleotides may be important in promoter function. The bases shown underneath the main sequence indicate changes at that position that did not effect promoter activity

A number of interesting observations were made in the analysis of the data provided by the sequencing. Firstly, it could be seen that the majority of the selected sequence could undergo point changes without complete loss of activity. These bases included most of the -35 and -10 regions as well as the –16 region, which has been implicated in promoter activity (Voskuil *et al.,* 1995). Secondly, there were a small number of bases in which no mutations were identified, which may be important for promoter function. Although it will be necessary to mutate these bases individually and assess the effect on expression to determine if this is actually the case, it is noteworthy that the A at -12 which was previously shown to be important for promoter function is amongst this group.

3.4. Analysis of the *recA*P1 promoter sequence

Although analysis of the *recA*P2 promoter was relatively straightforward, the characterisation of *recA*P1 was more complicated. From observations made in the *recA*P1 region, probable –10 and –35 sites had been identified, based on sequence homology with the –10 and -35 regions of the *E.coli* σ^{70} promoter. The spacing between these two elements in the mycobacterial sequence was much shorter than that found in the *E.coli* σ^{70} promoter (Movahedzadeh *et al.,* 1997). In contrast, the –35 sequence based on position relative to the probable –10 region, bore little resemblance to that of other promoters.

The studies done on *recA*P2, had identified a point mutation which eliminated its activity. This was important as it was thought that there was a region upstream of *recA*P1 that, as in *recA*P2, increased that activity of this promoter. Indeed, a construct based on the complete upstream region of *M. tuberculosis recA* containing the *recA*P2A-12C mutation (pEJ449)

was found to have a greater activity than a shorter construct (pEJ418) that was made up of just the presumed *recA*P1 site. However, pEJ449 had less activity than the original upstream sequence (pEJ417) which contained both promoter regions (see fig 3.9a and b. E. O. Davis, personal communication).

It was decided to use both pEJ449 and pEJ418 to analyse the *recA*P1 –10 region. Unlike the *recA*P2 study, where oligonucleotides were specifically designed to incorporate the mutation and then ligated into a promoterless vector, in this experiment point mutations were introduced through site directed mutagenesis. Briefly, an oligonucleotide and its complement, with a melting temperature of around 78°C and the point mutation at its centre were designed. This oligonucleotide pair would then be used to amplify up a plasmid, incorporating the desired mutation. The original plasmid would then be digested away with the use of the restriction enzyme *Dpn*I, which cleaves only methylated DNA, leaving a mutated construct. Following transformation of *E. coli*, sequencing of clones would identify the correct change required. These constructs were initially transformed into *M. smegmatis.*

In addition to three point mutations in the –10 region, one triple mutation was made in case the other mutations did not have any effect on activity. Unlike the *recA*P2 studies, where the constructs were constitutive, the resulting *recA*P1 strains were induced for 5 hours with mitomycin C before being assayed. Uninduced cultures were assayed as a control.

From the assay results it was found that the triple mutation almost eliminated *recA*P1 activity in both the short (pKKG19) and long construct (pKKG15) (fig. 3.6). In addition the

point mutation *recA*P1A-12C had the same effect in both constructs (pKKG21 in the short construct and pKKG17 in the long construct- see fig 3.6). This reduction in activity was also seen when pKKG17, pKKG19 and pKKG21 constructs were transformed into *M. tuberculosis* and assayed (fig 3.7). The other two constructs P1T-10G (pKKG22 in the short and pKKG18 in the long) and P1T-13G (pKKG20 in the short and pKKG16 in the long) did not reduce the activity of *recA*P1.

Using this information that a P1A-12C base change greatly reduced *recA*P1 activity, this mutation was introduced into the pEJ417 sequence to produce a construct of the original upstream sequence but containing only a functional *recA*P2 (pKKG25). This construct was assayed in both mycobacterial species and gave rise to some interesting observations. In the saprophyte it was found that this construct was more active than the complete upstream region (fig 3.8a) but in the pathogen it was seen that its activity was only about half that of the complete upstream region (fig. 3.8b).

With the identity of the –10 element clarified, the remaining question was the positioning of the –35 element. Using site directed mutagenesis a four base change mutation was introduced into the –35 region based on position (pKKG24) and a 3 base change into the –35 region based on sequence homology (pKKG26). These constructs were assayed in both *M. smegmatis* mc^2155 and *M. tuberculosis* H37Rv induced with mitomycin C and uninduced (Fig 3.9a and 3.9b).

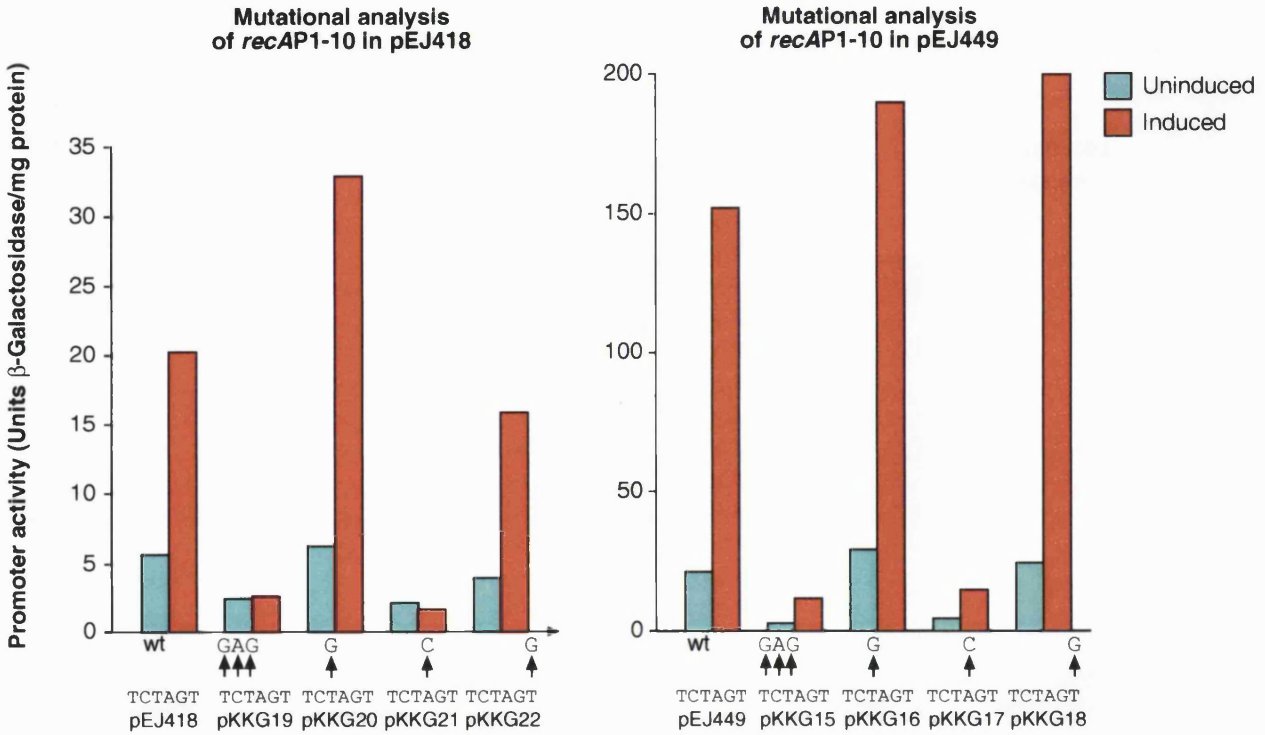


Figure 3.6. This figure illustrates the results of determining the position of the *recA*P1 –10 region. Through site directed mutagenesis, changes in the presumed –10 region were introduced into two wild-type *recA*P1 constructs; pEJ418 (just the presumed *recA*P1 region) and pEJ449 (the complete upstream region of *M. tuberculosis recA* with *recA*P2A-12C to eliminate *recA*P2 activity. These constructs were transformed into *M. smegmatis* and induced for 5 hours with mitomycin C before being assayed. The findings of this experiment were that both the triple change mutant (pKKG15 and pKKG19) as well as the *recA*P1A-12C mutant (pKKG17 and pKKG21) had the effect of virtually eliminating *recA*P1 activity. This showed this sequence and particularly the –12A base as important in *recA*P1 function.

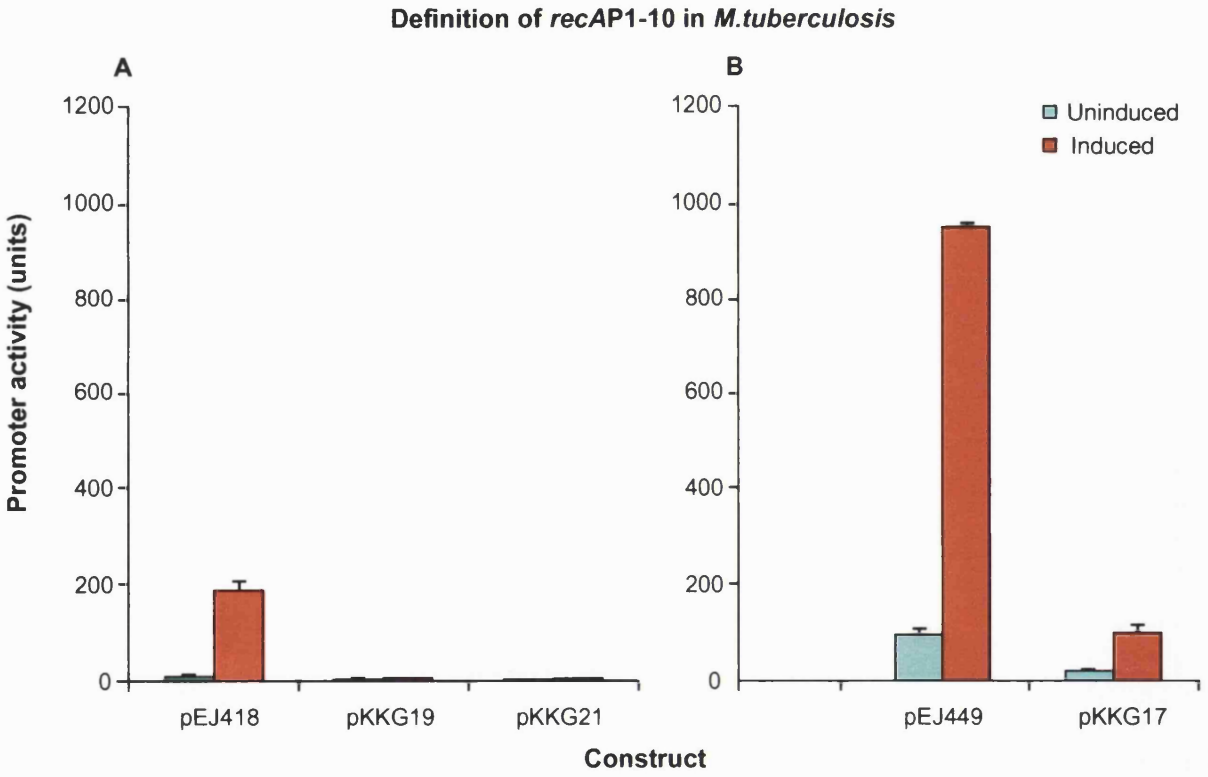


Figure 3.7. This figure illustrates confirmation of the results seen in the determination of the *recA*P-10 region using *M. smegmatis*. Constructs pKKG17, pKKG19 and pKKG21, which were shown to have no inducible activity from *recA*P1 were transformed into *M. tuberculosis*. These constructs were then induced with mitomycin C for 24 hours and assayed. Comparisons were made with the wild-type *recA*P1 constructs pEJ418 and pEJ449. The results of these assays confirmed that the sequence, and in particular the –12A base was important for the function of this promoter in *M. tuberculosis*.

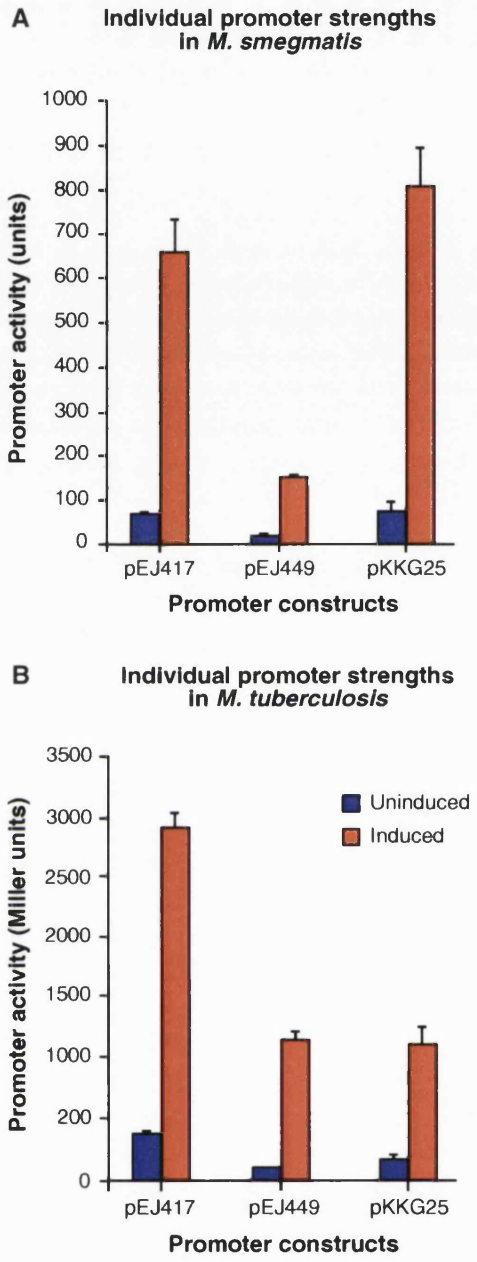


Figure 3.8. This figure illustrates the individual and combined activities of the *recA* promoters in both *M. smegmatis* (A) and *M. tuberculosis* (B). Based on information gathered in this study, -12A point mutations were introduced into the complete upstream region wild-type sequence (pEJ417) in either the *recA*P1 or *recA*P2 promoter. This left two constructs with either just a functioning *recA*P1 (pEJ449) or *recA*P2 (pKKG25). These constructs were then transformed into both *M. smegmatis* and *M. tuberculosis*, induced with mitomycin C and assayed. The error bars represent standard errors between three independent experiments per promoter construct. What was shown is that whilst induced *recA*P2 activity can account for the induced activity of the upstream sequence in the saprophyte, in the case of the pathogen complete induced activity of the upstream region is through a combination of both the *recA* promoters.

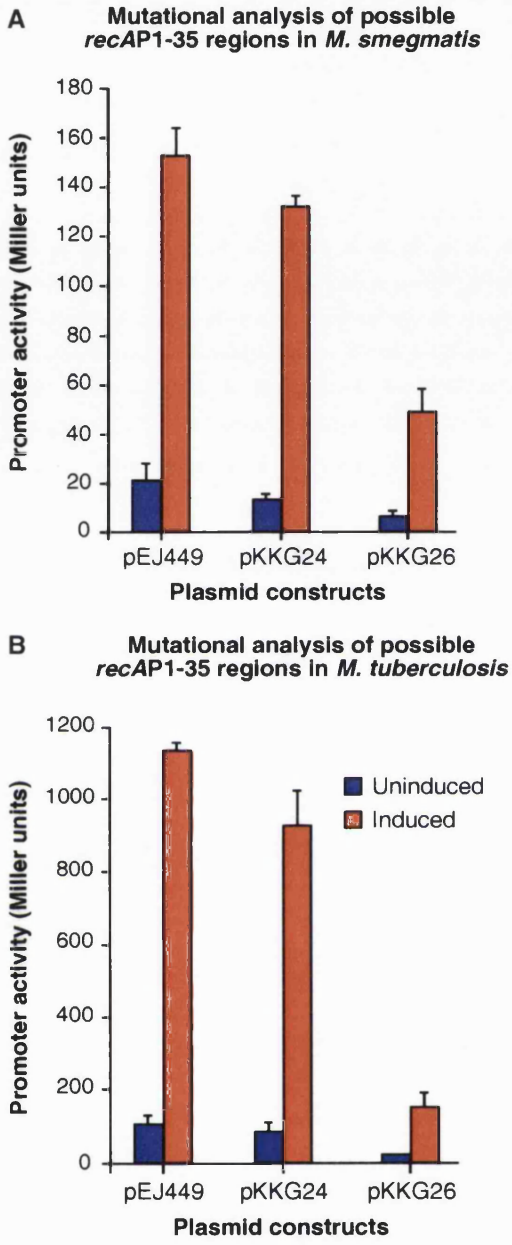


Figure 3.9. This figure illustrates the determination of the correct *recA*P1 –35 region. Movahedzadeh *et al.*, (1997a) suggested that this promoter had two possible –35 regions. One (35HOM) had sequence homology with the *E. coli* σ^{70} promoter's -35 region but was located closer to the –10 than seen in the *E. coli* σ^{70} consensus. The other (35POS) was in the correct position for this region but shared no homology with similar regions form other promoters. Two constructs were made by site directed mutagenesis of the wild-type *recA*P1 sequence (pEJ449), pKKG24 (35POS) and pKKG26 (35HOM). These constructs were transformed, induced with mitomycin C and assayed in both *M. smegmatis* (A) and *M. tuberculosis* (B). The error bars represent standard errors between three independent experiments per promoter construct. In both species it was seen that the mutation in the region of homology had a greater effect on the inducible activity of *recA*P1, indicating its importance in the functioning of this promoter.

In both species there was only a small effect on activity seen in the assays with the pKKG24 construct. However, a much greater reduction in activity was found for the pKKG26 construct, particularly in *M. tuberculosis*. These observations suggest that the motif located between the –25 and –30 region is important for the activity of *recA*P1.

3.5. The investigation of other possible promoter elements apart from *recA*P1 and *recA*P2.

As mentioned previously, the presence of two primer extension products indicated two regions of promoter activity, which were subsequently characterised in this study. An interesting finding was made with construct pKKG17, which contained both the *recA*P1 and *recA*P2 mutations, each of which had been shown to nullify promoter activity in other constructs. When this construct was induced and assayed both in *M. smegmatis* mc^2155 and *M. tuberculosis* H37Rv, it was shown to possess weak promoter activity. This suggested that there might be another promoter present in the upstream region. A construct based on the pEJ431 plasmid (a longer version of the *recA*P2 fragment than that contained within pEJ435) but containing the P2A-12C mutation (pKKG27) was made, induced and assayed and this was shown to have a weak inducible activity. However the equivalent mutation in the shorter pEJ435 (pKKG28) did not have this inducible activity (figs 3.10 and 3.11 and table 3.1).

Table 3.1. This table shows the average uninduced and induced activities of pEJ435 and pKKG28 in both *M. smegmatis* and *M. tuberculosis* based on three independent experiments. A student t-test value is included to show the significance of the difference in the activities of both constructs.

Plasmid	Species	Induced?	Activity (units)	Student t-test
pKKG28	*M. smegmatis*	No	1.5	0.002
pEJ435	*M. smegmatis*	No	91.7	-
pKKG28	*M. smegmatis*	Yes	1.8	0.003
pEJ435	*M. smegmatis*	Yes	435	-
pKKG28	*M. tuberculosis*	No	2.6	0.002
pEJ435	*M. tuberculosis*	No	261	-
pKKG28	*M. tuberculosis*	Yes	1.8	0.002
pEJ435	*M. tuberculosis*	Yes	643	-

The student t-test values in all cases show that the reduction seen in the activity of pKKG28 in comparison with pEJ435 is at least 99.7% genuine and did not happen by chance.

It was, therefore, deduced that this activity was located in the region upstream of the fragment isolated in pEJ435. A construct containing this entire region was constructed in pEJ414 (pKKG29). However this construct was found to have no activity induced or uninduced in both *M. smegmatis* and *M. tuberculosis*. Another construct, containing the fragment used in pKKG29 plus the upstream part of the fragment in pEJ435 excluding *recA*P2 cloned into pEJ414 (pKKG30) was also constructed and assayed (table 3.2). Again,

there was no detectable promoter activity in either *M. smegmatis* or *M. tuberculosis*. Thus, no further promoter was identifiable. Therefore, the most likely explanation for the weak activity seen with pKKG27 is that the promoter mutation is leaky in certain sequence contexts.

Table 3.2. This table shows the average uninduced and induced activities of pEJ414 and pKKG30 in both *M. smegmatis* and *M. tuberculosis* based on three independent experiments. A student t-test value is included to show the significance of the difference in the activities of both constructs.

Plasmid	Species	Induced?	Activity (units)	Student t-test
pKKG30	*M. smegmatis*	No	1.9	0.5
pEJ414	*M. smegmatis*	No	2.1	-
pKKG30	*M. smegmatis*	Yes	2.9	0.6
pEJ414	*M. smegmatis*	Yes	2.6	-
pKKG30	*M. tuberculosis*	No	3.7	0.6
pEJ414	*M. tuberculosis*	No	4	-
pKKG30	*M. tuberculosis*	Yes	1.8	0.5
pEJ414	*M. tuberculosis*	Yes	643	-

The student t-test values in all cases show that any reduction seen in the activity of pKKG30 in comparison with pEJ414 is at least 50% down to chance and hence there is in no significant difference in the expression between the construct pKKG30 and the vector pEJ414.

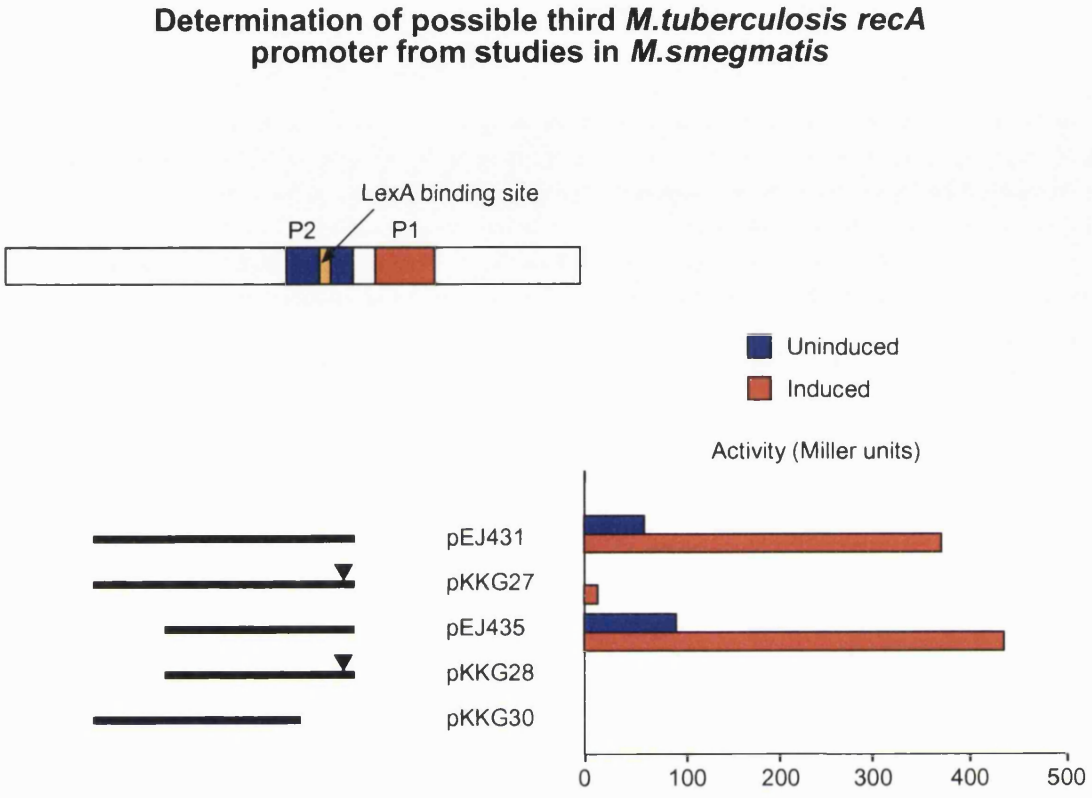


Figure 3.10. This figure illustrates the determination of a possible third promoter in the region upstream of the *M. tuberculosis recA* gene through experiments in *M. smegmatis*. It was observed that the double promoter knockouts pKKG15 and pKKG17 in *M. smegmatis* still displayed weak inducible activity although this was theoretically impossible based on the idea that there were only two promoters in this region. It was therefore of interest to determine if and where a third promoter was present. Introducing the *recA*P2A-12C point mutation (black triangle) into construct pEJ431 gave rise to pKKG27, which when transformed and induced in *M. smegmatis* had some activity. However, when a shorter *recA*P2 construct (pEJ435) was also modified to incorporate the *recA*P2A-12C point mutation (black triangle) giving pKKG28, it was seen that activity was eliminated. Through a process of elimination, where the sequence that made up the pKKG28 construct was removed, a small section of sequence remained. This sequence was ligated to a *lacZ* reporter construct (to form pKKG30), transformed into *M. smegmatis* , induced with mitomycin C and assayed but found to have no activity.

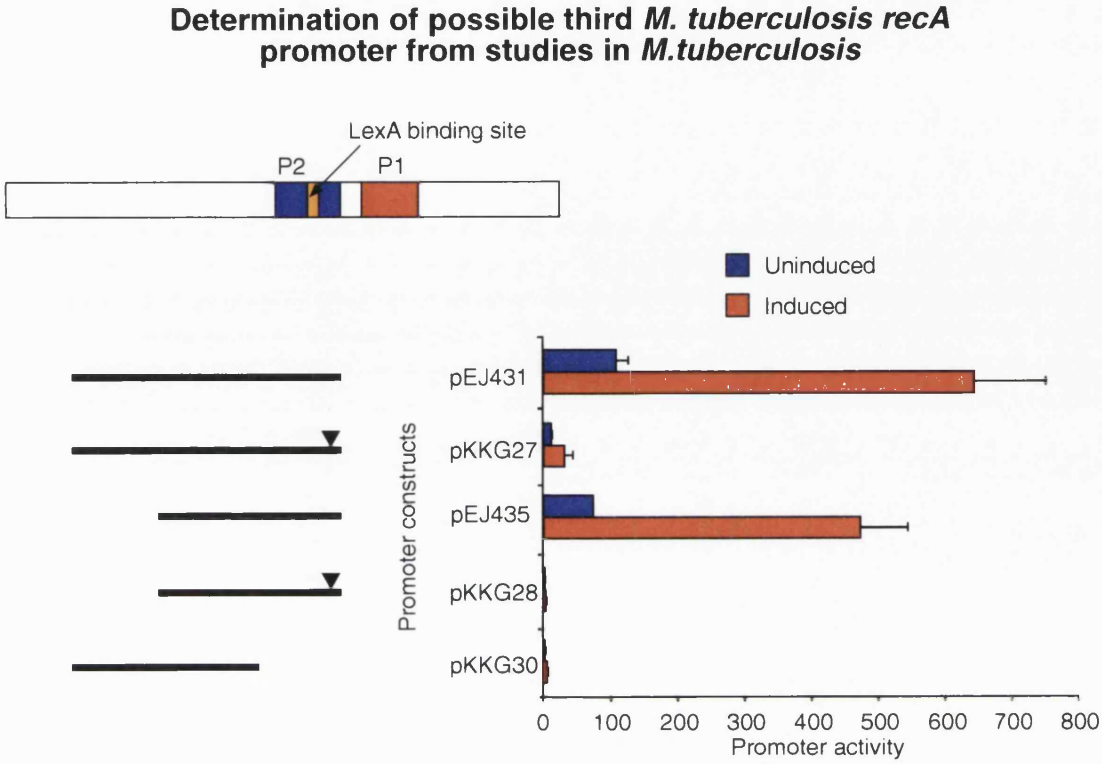


Figure 3.11. This figure illustrates the determination of a possible third promoter in the region upstream of the *M. tuberculosis recA* gene through experiments in *M. tuberculosis*. It was observed that the double promoter knock out pKKG17 in *M. tuberculosis* still displayed weak inducible activity although this was theoretically impossible based on the idea that there were only two promoters in this region. It was therefore of interest to determine if and where a third promoter was present. The error bars represent standard errors between three independent experiments per promoter construct. Introducing the *recA*P2A-12C point mutation (black triangle) into construct pEJ431 gave rise to pKKG27, which when transformed and induced in *M. smegmatis* had some activity. However, when a shorter *recA*P2 construct (pEJ435) was also modified to incorporate the *recA*P2A-12C point mutation (black triangle) giving pKKG28, it was seen that activity was eliminated. Through a process of elimination, where the sequence that made up the pKKG28 construct was removed, a small section of sequence remained. This sequence was ligated to a *lacZ* reporter construct (to form pKKG30), transformed into *M. tuberculosis*, induced with mitomycin C and assayed but found to have no activity. This result, when combined with that from the similar *M. smegmatis* study indicated that there was no third promoter present in the region upstream of *recA*. One possible explanation for the weak activity seen in pKKG17 is that the region upstream of *recA*P2 (found in the pKKG30 construct) might be able to make the mutated *recA*P2 "leaky".

3.6. Discussion

The primary aim of the experiments described in this chapter was to aid in the identification of promoter elements in the upstream region of the *M. tuberculosis recA* gene. Previous studies had localised promoter activity to two putative promoter regions, one containing an SOS box between its predicted –35 and –10 regions (*recA*P2), with the other not possessing this binding site (*recA*P1) (Movahedzadeh *et al.,* 1997a). This situation differs from what has previously been shown both in *E. coli* (Weisemann and Weinstock, 1991) and *B. subtilis* (Cheo *et al.,* 1992), which each have a single promoter upstream of their respective *recA* genes. In addition, in both these organisms the *recA* gene is controlled through the presence of an SOS box (Weisemann and Weinstock, 1991; Cheo *et al.,* 1993). Even so, the finding of multiple promoters upstream of the *recA* gene is not unique to *M. tuberculosis.* It has been shown that upstream of the *M. smegmatis* gene, there are also two regions of promoter activity (Papavinasasundram *et al.,* 1997), whilst the plant pathogen *Xanthomonas oryzae* pv. *oryzae* has three (Sukchawalit and Mongkolsuk, 2001). However, the mycobacterial *recA* promoters are unusual in being DNA damage inducible by different mechanisms (E. O. Davis, Personal communication). Therefore, it was of interest to define these two promoter regions.

3.6.1. The characterisation of *M. tuberculosis recAP2* through direct mutagenesis.

It had previously been noted that the sequence present at the potential *recA*P2 –35 was very much similar to that found in the σ^{32} *E. coli* sequence (this consensus being

TNTCNCcCTTGAA in comparison with the *recA*P2 TGTCACACTTGAA) (Davis *et al.*, 1991; Movahedzadeh *et al.*, 1997). Nevertheless the –10 region was not the same as the consensus of these promoters. However, the putative *recA*P2 also shared sequence homology with another type of promoter construct, the gearbox promoter (Aldea *et al.*, 1990) (see figure 3.2 for comparison).

Of the eight directed mutations in *recA*P2 assayed in *M. smegmatis*, five were at bases conserved in both gearbox and σ^{32} promoters, and three had the effect of reducing activity. However, the adenine residue at position –12, which, when mutated lead to the loss of promoter activity when assayed in both *M. smegmatis* and *M. tuberculosis,* is only found in the gearbox consensus.

Gearbox promoters were first identified in *E. coli* with one example being the *bol*A1*p* promoter (Aldea *et al.*, 1990). However, there are indications that these promoters may also exist in other bacterial species, including *M. tuberculosis* (Hu and Coates, 1999). These promoters have been found to regulate genes that become activated as the cell enters into stationary phase (Aldea *et al.*, 1990; Ballesteros *et al.*, 1998). In *E. coli* upon entry into stationary phase, there is a shift in the cell toward the production of the stationary sigma factor, σ^S, which has been shown to recognise these promoters (Aldea *et al.*, 1990). However, a number of studies have shown that σ^S and some genes under its control are also expressed at a high rate under certain stresses (Hengge- Aronis *et al.*, 1993, Hengge-Aronis, 1996; Muffler *et al.*, 1997). This suggests that there is a potential for gearbox promoters to be induced in stress conditions where σ^S is being expressed. In addition, it has also been shown that one of these gearbox promoters in *E. coli*, *bolA1p*, can become

activated under certain stresses in an Δ*rpoS* strain, indicating that its expression is not totally dependent on σ^S (Santos *et al.,* 1999). This finding that gearbox promoters can have stress induced activity suggests that there is a possibility that *recA*P2 could be another example of this type of promoter in mycobacteria.

However, since the two other mutations which had an effect in lowering *recA*P2 activity were also conserved in the heat shock promoter consensus, there is still a possibility that *recA*P2 could be a heat shock promoter. It has been shown in *E. coli* that heat shock can induce the SOS response through increased expression of *recA* (Fabisiewicz and Jainon, 1992). Also, in *M. tuberculosis* many genes involved in coping with oxidative damage were found to be part of the heat shock group of proteins (Garbe *et al.,* 1996).

One way of determining whether *recA*P2 is a gearbox promoter would be to see if its activity increased in stationary phase. It has been shown that expression of the sigma factor associated with heat shock in *E. coli* (σ^{32}) also increases in stationary phase as part of a starvation mechanism (Jenkins *et al.,* 1991). However, the expression of *M. tuberculosis recA*P2 in this phase of growth, coupled with the sequence information obtained in this chapter would give an indication that this element could be another mycobacterial gearbox promoter. It would therefore be of interest to see whether the activity of this promoter was affected by the change in growth phase in addition to that of DNA damage, a question which is addressed in chapter 4.

3.6.2. The characterisation of *M. tuberculosis recAP2* through random mutagenesis.

Through direct mutations in the proposed –10 and –35 region, the *recA*P2 promoter had been identified. However, to define the promoter sequence, random mutagenesis was utilised. In the study of Huang and Hellmann (1998) where they employed random mutagenesis to derive a consensus for the *Bacillus subtilis* σ^X promoter, they had initially transformed the spiked oligonucleotides into *E. coli,* which were then screened and sequenced for point mutations throughout the sequence of interest. This meant that all possible combinations of randomised sequence were available. In addition, each possible mutation was transformed and assayed individually to get a quantitative evaluation of promoter activity.

In the case of this study, a qualitative approach was employed for screening of sensitivity to sequence mutation through the fusion of the spiked oligonucleotide to a promoterless *lacZ* marker. Although subtle differences in promoter activity could not be detected using this procedure, it was hoped that through random mutagenesis of this promoter and screening for null promoter activity (white colonies with X-Gal selection), the bases crucial for promoter activity could be identified. This in turn would show what type of promoter *recA*P2 was. However, such screening in *M. tuberculosis* was not successful and instead the screening of active colonies was pursued to investigate how sensitive the sequence (the –10, -35, spacer region and five bases either side of the promoter elements) was to mutation. From this approach it was shown that this sequence could tolerate changes in the majority of positions, both in the spacer region and the promoter –10 and –35 regions. Even

so, there are nine positions in which the bases remained unchanged in all the sequences screened (fig 3.5) and these may or may not be important in the activity of *recA*P2. To confirm this suggestion it would be necessary to employ directed mutagenesis at these positions to prove the sensitivity at the particular position in question.

One finding of interest from the random mutagenesis study is that although the majority of the –10 region could be changed to a varying degree, depending on the nucleotide position, a TA motif centred around positions –13 and –12 was conserved in all the positive promoter mutants isolated. It has been shown in *E. coli* σ^{70} promoters that a motif of this nature, but located at the –12 and –11 positions is conserved in around 80% of these promoters (Lisser and Margalit, 1993). This motif has been shown to be important for the recognition of the promoter by *E. coli* RNA polymerase holoenzyme with σ^{70} (Roberts and Roberts, 1996). It has been shown that mutations of these two nucleotides reduce the binding affinity the –10 region has for the polymerase holoenzyme (Fenton and Gralla, 2001).

3.6.3. The characterisation of the *M. tuberculosis recA*P1 promoter.

The identification of the *M. tuberculosis recA*P1 promoter has raised a few interesting points. The *recA*P1 –10 region by homology fits loosely with the mycobacterial –10 consensus derived by the study of Bashyam et al. (1996). However, the identification of a region homologous to the –35 region TTGACA motif that has an effect on this mycobacterial promoter's activity is very interesting. In a previous study where 35 mycobacterial promoters were compared, it was found that none of the promoters studied

possessed the *E. coli* σ^{70} -35 TTGACA motif (Bashyam *et al.*, 1996). Another point of interest is that whilst the mutation of the *recA*P1 -35 region by homology greatly reduced promoter activity, this mutation was not as drastic as that seen for the P1A-12C mutant. This coincides with the findings of Bashyam *et al.* (1996) which showed that two mycobacterial promoters were still functional, albeit at a lower level, with the removal of their respective -35 regions. However, this in turn contradicts the findings of Kremer *et al.* (1995) which had stated that the -35 region was essential for the activity of the *M. tuberculosis* 85A promoter.

Furthermore, although the sequences of both regions of the promoter share homology with the E. coli σ^{70} consensus, the spacing between the two promoter elements is shorter. In the case of the *M. tuberculosis recA*P1, the spacer measures 9bp, which is shorter than both the 16-18bp spacer found in the E. coli σ^{70} consensus (Harley and Reynolds, 1987) or the 18-19bp spacer found in the streptomycete principal sigma factor promoter consensus (Bourn and Babb, 1995). It has been shown in studies of *E. coli* σ^{70} promoters that although the sequence between the –35 and –10 regions can be varied, the actual length of the spacer is critical to the activity of the promoter (Aoyama *et al.,* 1983). In a study done by Aoyama *et al.* (1983) where the length of an 18bp long spacer element was varied, it was found that spacer elements longer than 18 bases or shorter than 16 greatly reduced the activity of the promoter (Aoyama *et al.,* 1983). In a study of mycobacterial promoters it has been shown that spacing between the –10 and –35 region varies between 9bp and 24 bp (Mulder *et al.,* 1997). It is unclear as to whether spacing between the -10 and -35 elements in mycobacterial promoters is important in their function. In a study by Kremer *et al.* (1995), where 64 bases were inserted between the -10 and -35 region of the *M. tuberculosis* 85A

promoter, it was found that its measured activity was similar to that of a construct with a four base deletion in the spacer. From this finding, it was suggested that although the presence of a -35 region was important, the position of this element in relation to the -10 element was not (Kremer *et al.,* 1995). It was further stated by Kremer *et al.* (1995) that sequencing had showed that there was no sequence within the 64bp insert that resembled a possible -35 region. However, since the 64bp insertion was only placed in one orientation, it was not clear from this study whether there was a region within this insert that could act as a -35 element. In a study by Bashyam *et al.,* (1996) where three long inserts were placed between the -10 and -35 regions of two mycobacterial promoters, it was shown that the activity of each promoter was dependent on the orientation of the insert, regardless of insert size (Bashyam *et al.,* 1996). In addition, removal of the original –35 sequence did not significantly alter expression further. This indicated that alternative –35 sequences provided by the insert could be utilised in transcription (Bashyam *et al.,* 1996).

3.6.4. The role of the –12A in both promoters

From the experiments highlighted in this chapter, it has been shown that the majority of the structures of both *recA*P1 and *recA*P2 are dissimilar. However in the –10 region of both promoters, there is conservation of sequence from –14 to –8 with the only difference coming at position –11. What is of importance is that the TA motif mentioned earlier also exists in *recA*P1. Further mutations of *recA*P1 could affirm whether this motif is important for the functioning of this promoter. However, a point change at position –12 (A-C) like that in *recA*P2, eliminated promoter activity. It has been shown in the promoters of the *E. coli gal* operon, which are recognised by σ^{70}, that the adenine residue at position –11 plays

a vital role in the formation of the open complex at the promoter, but does not affect the formation of the polymerase-promoter complex (Lim *et al.,* 2001). In both *M. tuberculosis recA*P1 and *recA*P2 promoter there is an overall dissimilarity in promoter sequence when the –35 sites are included suggesting that a different sigma factor is probably required for recognition of either element. However, it seems likely that the conserved –12 residue in both these mycobacterial promoters plays a similar role to that seen in the *E. coli gal* operon.

3.6.5. Both promoters are inducible by mitomycin C.

It had previously been suggested that only one of the promoter sequences in the fragment of DNA upstream of the *M. tuberculosis recA* gene might be inducible with DNA damage (Movahedzadeh *et al.,* 1997a). However, this lead to a dilemma as the promoter lacking the SOS box, *recA*P1, and not *recA*P2, gave rise to the stronger primer extension product upon addition of mitomycin C (Movahedzadeh *et al.,* 1997a). It has subsequently been shown both in this study and others (E. O. Davis personal communication) that both promoters independently are inducible upon the addition of mitomycin C. This is an interesting finding as it is in contrast to what occurs in the plant pathogen *Xanthomonas oryzae* pv. *Oryzae*.

In this plant pathogen, the *recA* gene present is driven by three promoters, which are all regulated by two SOS boxes present (Sukchawalit and Mongkolsuk, 2001). However whilst the two promoters with the SOS boxes are inducible individually, the third promoter, which

lacks the protein-binding motif, when assayed independently has been shown to be constitutive in expression (Sukchawalit and Mongkolsuk, 2001).

The inducibilty of both *recA*P1 and *recA*P2 upon the addition of mitomycin C raises more questions as to whether there is a different role for each promoter upon DNA damage if both work synergistically to maximise the expression of RecA when required. For this reason, it will be of importance to study the kinetics of induction for each promoter, which is addressed in chapter 4.

3.6.6. Regions upstream of both promoter elements: A third promoter or enhancer?

The presence of a third promoter was hinted at by the weak inducible activity observed in assays with pKKG17 (a variation on the full length construct with both P1A-12C and P2A-12C mutations). Support for this came from the P2A-12C mutation in pKKG27, which had some induciblity but was less markedly active than the wild type pEJ431 in *M. tuberculosis*. However, testing fragments of this upstream sequence for promoter activity was fruitless. It has, therefore, been suggested that the weak activity observed in pKKG27 might occur from the promoter mutation being "leaky" when taken together with the remaining upstream sequence. Even so, assays done with pEJ449 and pKKG25 which are based on the whole upstream region with only an active *recA*P1 or *recA*P2 site gave promoter activities that were much higher than those obtained with the shorter pEJ418 and pEJ435 constructs.

In the previous section it had been suggested that the –12 mutation may affect the formation of an open complex, hence eliminating promoter activity. However there are a number of possible reasons as to why this mutation may be "leaky" in *recA*P2 with the addition of extra upstream sequence. It has been shown with wild type *E. coli* promoters that the presence of an UP element can aid in the formation of an open complex (Tagami and Aiba, 1999; Guo *et al.,* 2000). However, the composition of the element upstream of *recA*P2 is not AT rich and the distance between the "enhancing element" and *recA*P2 is much further than expected if this region was an UP element (Gourse *et al.,* 2000). Another possibility is that there may be a protein that interacts with this region.

Although it has been shown that an enhancer protein is required for *E. coli* σ^N to convert a closed promoter complex to an open one, these proteins are not necessary for σ^{70} to undertake this task (Guo *et al.,* 2000). However, it has been stated that one such activator protein, cyclic AMP receptor protein (CRP), which can function in tandem with the carboxyl terminal domain (CTD) of the α subunit of the RNA polymerase core to stabilise binding of the holoenzyme to a particular promoter, can also stabilise a transition state between a closed and fully open complex (Hochschild and Dove, 1998). Furthermore, the action of activating proteins to stimulate the formation of an open complex is not only reserved to an interaction with *E. coli* σ^{70} holoenzyme. One such example is that of the *E. coli csiD* promoter, which is exclusively dependent on σ^S for recognition but relies upon interaction between the σ^S holoenzyme and CRP to allow the formation of an open complex *in vitro* (Marschall *et al.,* 1998).

Assaying of the *M. tuberculosis recA*P2 promoter has shown that the addition of an adjacent 60bp region is sufficient to dramatically increase its activity (E. O. Davis, personal communication). What the results of this study indicate is that there may be another activator binding site further upstream that has an effect on the activity of *recA*P2. One such example of multiple activator binding sites upstream of a target promoter can be seen in the *E. coli narGHIJ* promoter, which has binding sites for two activators located about 60bp and 200bp away from the transcriptional start site (Dong *et al.,* 1992). It has been shown in cases of multiple activators binding upstream that these factors act synergistically (Joung *et al.,* 1994; Busby *et al.,* 1994, Hochschild and Dove, 1998). This might account for no activity from the construct possessing one activator region (pKKG28) but inducible activity from the construct with both regions present (pKKG27). One way to look for the presence of an activator-binding region within this far upstream region would be to conduct band shift assays with cell-free extracts from both DNA damage uninduced and induced cultures on this fragment of DNA. This type of experiment could also be used to determine whether the potential protein activator was inducible with DNA damage. Other experiments that could be undertaken to test whether this upstream region interacted with an inducible activator could include fusion of this region upstream of a weak mycobacterial promoter that was not DNA damage inducible and then seeing if activity was enhanced upon the addition of a DNA damaging agent. One further avenue of study could be to do a "phasing" experiment whereby through the addition of bases between this region and *recA*P2 the position on the DNA helix of the enhancer region could be modulated to be either "in phase" or "out of phase" with the promoter. This procedure was used to show that the interaction the catabolite gene activator protein (CAP) had with the *E.coli uhpT* promoter was phase dependent (Merkel *et al.,* 1995).

3.6.7. Possible problems associated with promoter-probe plasmids in the studying of promoter activity and regulation.

To observe changes in promoter activity with various DNA sequences, this study employed the technique of promoter-*lacZ* transcriptional fusions. However, there are limitations to this type of study that may affect interpretation of the results obtained. To overcome any problems associated with copy number, these fusion constructs were made using an integrative vector, pEJ414 (Papavinasasundaram *et al.*, 2002), which in turn was based on the mycobacterial vector pMV306 (Stover *et al.*, 1991; Kong and Kunimoto, 1995). This vector carries both the attachment site, *attP* as well as the integrase gene, *int*, present in the mycobacteriophage L5 (Stover *et al.*, 1991).

The promoter-*lacZ* constructs used in this study integrate into the mycobacterial chromosome specifically at the *attB* site (Stover *et al.*, 1991). This means that the *M. tuberculosis recA* promoters are being studied outside the area of genome they would naturally be present in. It is possible that there are conditions such as DNA supercoiling that may have an effect on expression from the *recA* promoters in their native positions which would not be seen when taken out of this context. The opposite is also true, whereby events affecting promoter activity in this new position in the genome may not occur in the native position.

Another problem associated with this type of study is the stability of the RNA product formed. In this case, although an increase in the transcription of *lacZ* driven by either of the *recA* promoters resulting in increased levels of β-galactosidase may suggest that the level

of RecA would also increase in the same conditions, *recA* RNA may be less stabile than that from *lacZ* and hence would be degraded more quickly.

An associated problem with stability is that of β-galactosidase accumulation in the cell. Although a promoter may only be active for a short time in the process of induction, the stability of the β-galactosidase protein means that any drop in activity would not be immediately noticeable.

In the case of this study, whilst all these problems are valid concerns, it must be noted that since both wild-type and mutated promoter construct are being tested and compared in the same way, differences seen between the assays of wild-type and mutant constructs are valid.

The conclusion drawn from this chapter is that the *M. tuberculosis recA* gene is expressed from two different promoters. The core sequences required for promoter activity appear to differ for the two promoters although they also have some common features, notably the importance of an A at position -12. It is also apparent that sequences outside the core promoter region influence the expression level. In the next chapter the patterns of expression of the two promoters will be compared.

CHAPTER 4: THE PATTERN OF EXPRESSION OF THE *recA*P1 AND *recA*P2 PROMOTERS.

4.1. Introduction

From the experiments described in the previous chapter, it was confirmed that there were two promoters in the *M. tuberculosis recA* upstream sequence, only one of which had an SOS box but both of which were inducible. It was shown that there were no other specific areas of promoter activity present in the upstream fragment. However, the presence of *recA*P1 and *recA*P2 raises the question as to the necessity of two promoters for this *recA* gene. In other bacteria, such as *E. coli* and *B. subtilis*, one promoter that is associated with a SOS box expresses the *recA* gene. It was, therefore, of interest to see how the activities of both *recA*P1 and *recA*P2 changed with induction. Were the two promoters acting synergistically or was there a lag in the induction of one promoter relative to the other? The inducibility of the *recA*P1 promoter suggested that there might be another means of transcriptional control that may be LexA independent as induction occurred in the absence of a LexA binding site. It was also known that gearbox promoters become active in stationary phase, suggesting that the same attribute might be shown by *recA*P2, which showed sequence similarity to that class of promoter. For these reason it was of interest to see what effect such properties as growth phase would have on the activity of the two promoters and to compare the kinetics of induction of the two promoters. This was initially done in *M. smegmatis* but was later undertaken in *M. tuberculosis*. In the promoter studies

done in this chapter, both *recA*P1 (on construct pEJ449) and *recA*P2 (pEJ435) were considered and tested as separate elements.

4.2 Effect of culture density/ growth phase on promoter activity

In the experiments done previously, cells were harvested and induced at an optical density that reflected a culture in exponential phase. However, it was unclear as to whether the uninduced and induced activities seen for both promoters were constant regardless of phase of growth. In addition, it was of interest to see if *recA*P2, which was shown in the previous chapter to share sequence homology with both the σ^{32} and gearbox consensus sequences, shared the behaviour of the growth phase dependent elements.

To measure the effects of growth phase on uninduced cultures, samples containing the *recA*P1 and *recA*P2 constructs were harvested from both *M. tuberculosis* and *M. smegmatis* at different optical densities ranging from 0.2 to 1.8. In addition, a modified *recA*P2 construct which was made constitutive through two point mutations it the LexA binding site was also harvested over the same range of OD_{600} values in *M. smegmatis.* These samples were then tested by β-galactosidase assay to measure their activity (figs. 4.1 and 4.2).

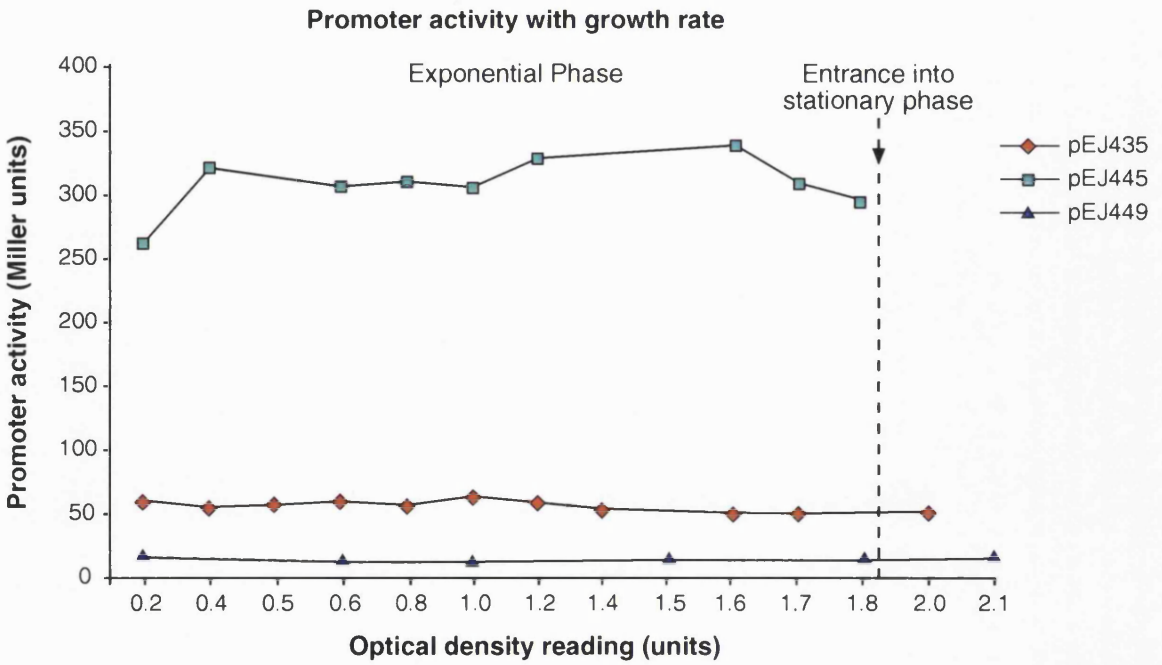


Figure 4.1. This figure illustrates the results to determine the uninduced expression of both *recA* promoters at different stages of growth in *M. smegmatis*. It has been shown that growth phase can effect the expression of *recA* in *E. coli* (Taddei *et al.*, 1995) and therefore it was of interest to see if this scenario could be repeated in *M. smegmatis*. Cells containing either the inducible *recA*P1 (pEJ449), inducible *recA*P2 (pEJ435) or constitutive *recA*P2 (pEJ445) constructs were harvested at a range of OD_{600} values from 0.2 to 2.0 and assayed. From the results of this study it can be seen that the uninduced promoter activity of both *recA*P1 and *recA*P2 remain constant regardless of growth phase. This indicates that there is no relationship between growth phase and promoter activity for either promoter. In addition, it can also be seen that the constitutive activity of *recA*P2 in the saprophyte is also constant, regardless of growth phase.

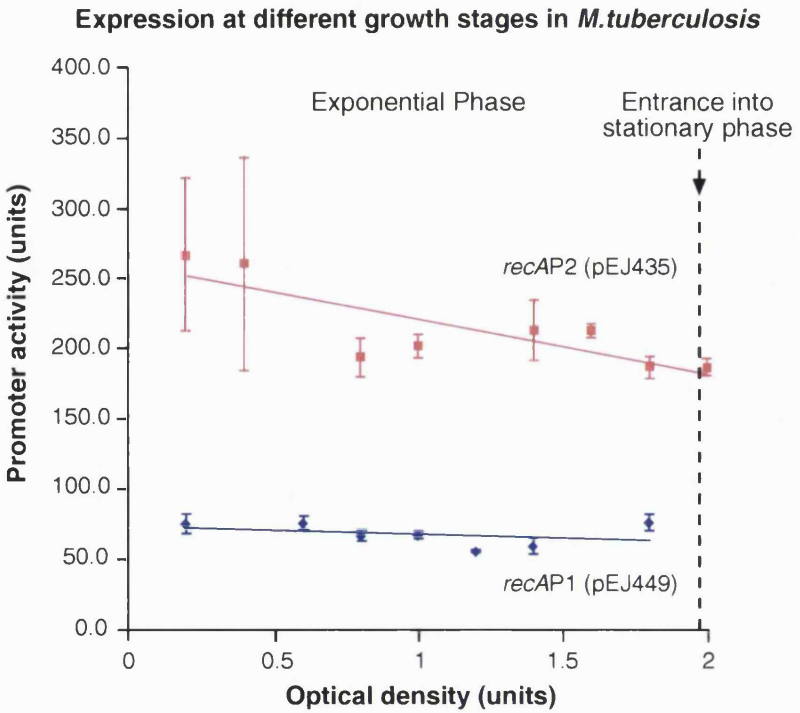


Figure 4.2. This figure illustrates the results to determine the uninduced expression of both *recA* promoters at different stages of growth in *M. tuberculosis*. It has been shown that growth phase can effect the expression of *recA* in *E. coli* (Taddei *et al.*, 1995) and therefore it was of interest to see if this scenario could be repeated in *M. tuberculosis*. Cells containing either the inducible *recA*P1 (pEJ449) or inducible *recA*P2 (pEJ435) constructs were harvested at a range of OD_{600} values from 0.2 to 2.0 and assayed. The error bars represent standard errors between three independent experiments per promoter construct The uninduced activity of *recA*P1 was constant as was observed in the *M. smegmatis* study. In the case of uninduced *recA*P2 activity, although there was an indication that there was a slight decrease in activity with changing cell density (as shown by the line plotted), since the error bars overlap, it could also be argued that there was no change associated with the change in growth phase.

In *M. smegmatis* even though there were slight variations between the different samples, there was no general trend of increasing or decreasing promoter activity for either element (fig. 4.1). Furthermore, in the constitutive *recA*P2 construct it was also demonstrated that in the absence of repression, promoter activity remained constant (fig.4.1). This in turn indicated that the uninduced activity for *recA*P1 and *recA*P2 was independent of growth phase. However, when the same study was repeated in the pathogen it was found that although *recA*P1 in *M. tuberculosis* behaved as it did in the saprophyte, there appeared to be a slight decrease in the activity of *recA*P2 that was associated with increasing cell density (fig 4.2). However, since the error bars overlap in this experiment, it could also be said that there may not be any difference in the activity of this promoter with a change in growth phase. It is therefore necessary to repeat this experiment with more samples to confirm one of these suggestions.

To test if the cell density at which DNA damage was inflicted affected the induced level of expression directed by either element, *M. tuberculosis* strains containing either the *recA*P1 or *recA*P2 construct were grown and induced at different optical densities from 0.1 to 1.0. The cells were exposed to 0.2 μg/ml (final concentration) mitomycin C for 24 hours in each case and the samples assayed. The induced promoter activity of *recA*P2 shows little variation through all cell densities tested (fig 4.3). However, the induced activity of *recA*P1 starts off very high when induction is done at low OD_{600} but constantly declines with increasing optical density of the culture (fig 4.4). This can be illustrated by the fact that at an OD_{600} of 1 the fold of induction is dramatically reduced to 2.6 fold from that seen in lower densities where it is ~11 fold.

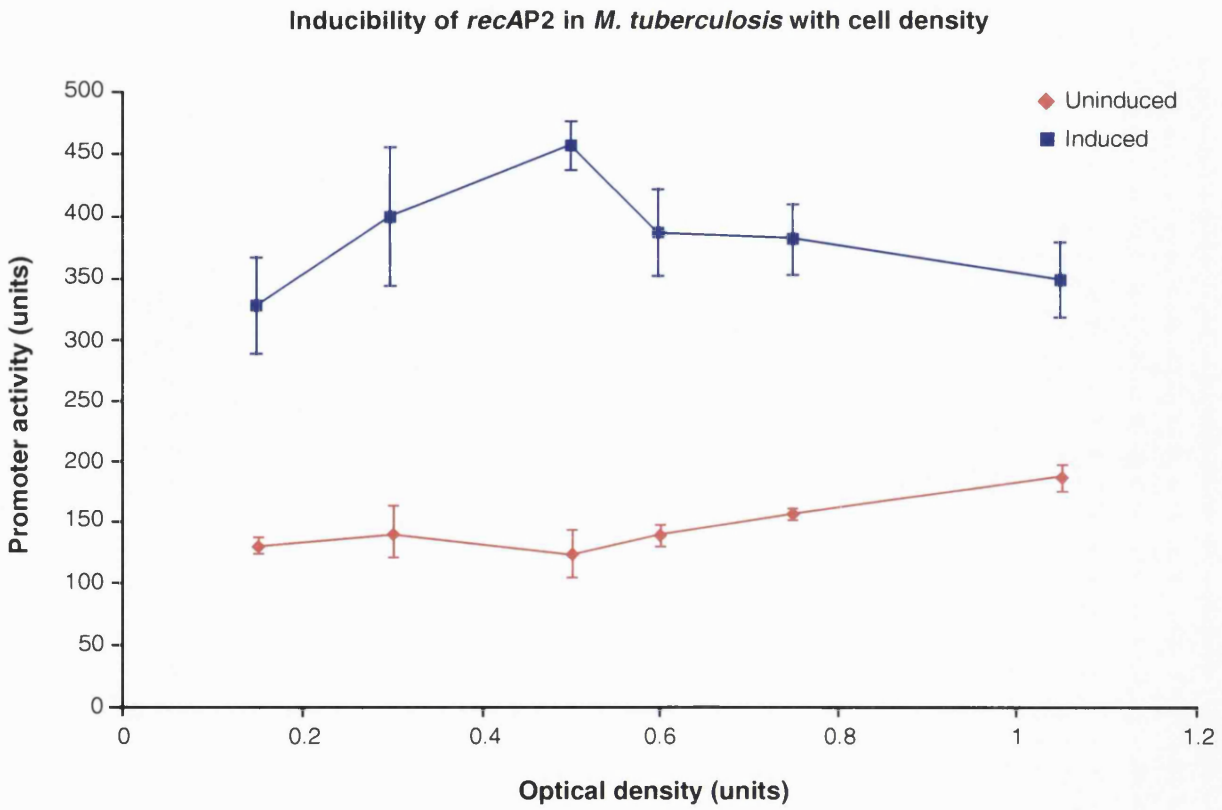


Figure 4.3. This figure illustrates the results of the study into the inducibility of *recA*P2 in *M. tuberculosis* with cell density. Cells were harvested at a range of OD_{600} values from 0.1 to 1.0 were induced with mitomycin C for 24 hours and assayed. The error bars represent standard errors between three independent experiments per promoter construct Although there is an indication of an increase in induced promoter activity between the OD_{600} values of 0.15 to 0.5, the main trend of this graph does not suggest a link between cell density and the inducibility of *recA*P2.

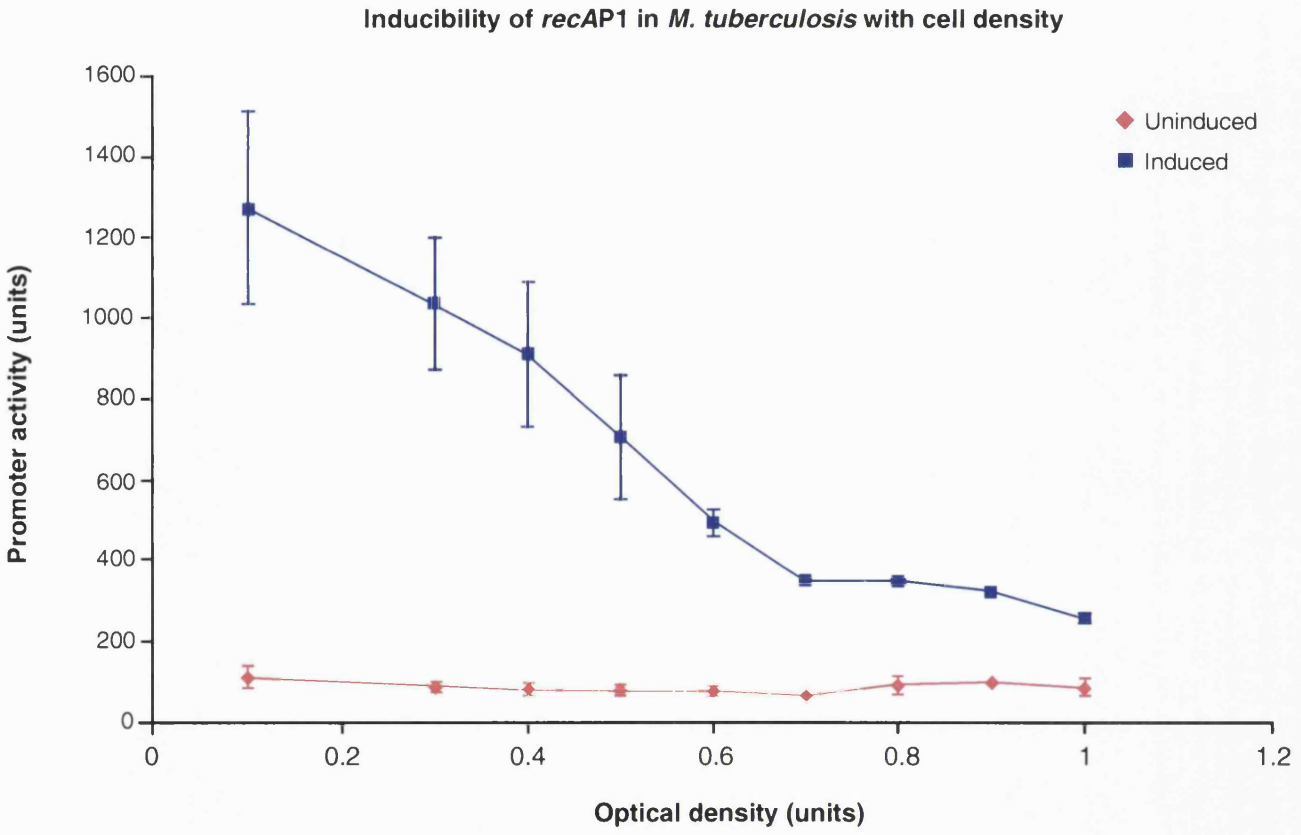


Figure 4.4. This figure illustrates the results of the study into the inducibility of *recA*P1 in *M. tuberculosis* with cell density. Cells were harvested at a range of OD_{600} values from 0.1 to 1.0 were induced with mitomycin C for 24 hours and assayed. The error bars represent errors between three independent experiments per promoter construct Unlike the inducible activity of *recA*P2 in *M. tuberculosis*, which was not affected by cell density, it was seen that the induced activity of *recA*P1 decreased with the increasing cell density of the culture. This indicates that there is a possible link between the inducibility of *recA*P1 and the cell density of the culture.

4.3. The kinetics of induction of either recA promoter with DNA damage.

This study was done to compare the kinetics of induction of the two *recA* promoters over an induction period of eight hours in *M. smegmatis* and twenty-four hours in *M. tuberculosis*. Whilst a similar analysis had previously been done with the entire upstream region of *recA* in both mycobacterial species (Papavinasasundaram *et al.*, 2001) this study would give an indication of how the individual promoters behaved.

In *M. smegmatis*, after an initial lag, both promoters reached a maximum rate of induction between one and three hours after the addition of mitomycin C (fig 4.5a), although the rate of induction is slightly higher for *recA*P2 than *recA*P1. However, after four hours post induction, there is no further induction of *recA*P1 and a plateau is reached, whilst *recA*P2 activity continues to increase through the entire time period of this study (fig. 4.5a).

The region upstream up the *recA* gene, and in particular the two promoter elements, has been shown to be highly conserved in both *M. smegmatis* and *M. tuberculosis*. It, therefore, might be assumed that the kinetics of induction of both *recA*P1 and *recA*P2 would be the same in both saprophyte and pathogen. However, when the same study was undertaken in *M. tuberculosis* it was shown that the kinetics of both *recA*P1 and *recA*P2 induction in the pathogen were quite different to that seen in *M. smegmatis*. In the pathogen, the initial post-induction activity of *recA*P2 is higher that that seen for *recA*P1 as is seen in the saprophyte (Figs. 4.5a and 4.5b) although the increase in the rate of activity seen from *recA*P2 in mitomycin C induced *M. tuberculosis* was more gradual that seen in the saprophyte. The main difference in *M. tuberculosis* compared to *M. smegmatis* is that from

6 hours post induction the rate of induction of *recA*P1 increases dramatically over a period of 12 hours resulting in the activity of *recA*P1 becoming greater than that of *recA*P2 using these constructs. However, when the more active *recA*P2 construct pKKG25 was tested in the pathogen following induction for 24 hours, the induced level of expression was comparable to that of *recA*P1 at the same time point (Fig. 3. 8).

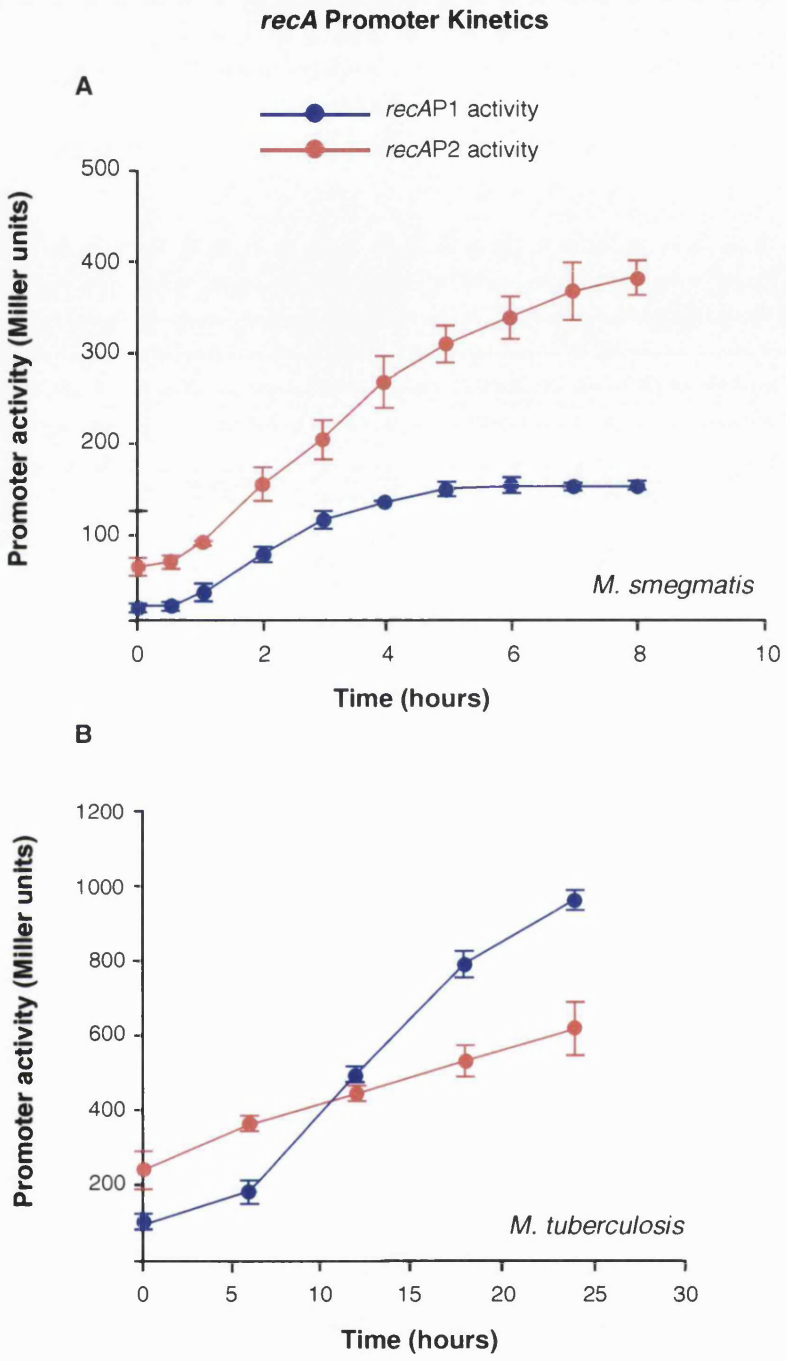


Figure 4.5. This figure illustrates the kinetics of induction study done for both *recA* promoters in *M. smegmatis* (A) and *M. tuberculosis* (B). Aliquots of identical volume were taken from one culture of either *M. smegmatis* or *M., tuberculosis* and induced with mitomycin C for a variety of incubation times ranging from 0 - 8 hours for the *M. smegmatis* study and 0-24 hours in the *M. tuberculosis* study. These samples were then assayed. The error bars represent standard errors between three independent experiments per promoter construct. From the results it can be seen that there is a difference both in the way either promoter acts in the same host and also in the way the same promoter acts in the two different mycobacteria.

4.4. Discussion

It has been shown that in the region upstream of the *recA* gene in both *M. tuberculosis* and *M. smegmatis* (Movahedzadeh *et al.,* 1997; Papavinasasundaram *et al.,* 1998), there are two promoter elements each of which share a common sequence in the two species. The presence of these two promoters could be seen as an indication that two modes of induction could be involved in activating transcription of *recA* in both species. The aim of the studies mentioned in this chapter was to obtain a better understanding of why *M. tuberculosis* has two DNA damage inducible promoter elements present upstream of its *recA* gene. It was thought that seeing how the induction of both promoters was regulated could provide insight into this.

4.4.1. Is there a link between growth phase and uninduced promoter activity?

Although genes that are associated with the SOS response become activated in conditions of DNA damage, it has been shown in *E. coli* that there are other conditions whereby SOS genes are expressed. One such condition is that of growth phase and it has been shown in *E. coli* that *recA* can be induced in stationary phase without the presence of DNA damage (Taddei *et al.,* 1995). In addition, another gene regulated by LexA in *E. coli*, *sbmC,* is induced upon entering stationary phase in a DNA damage free environment (Baquero *et al.,* 1995). It has been shown in both cases that the presence of cyclic AMP (cAMP) can stimulate the SOS response, with cleavage of the LexA repressor and activation of these genes (Baquero *et al.,* 1995; Taddei *et al.,* 1995). Conversely, the *E. coli sfiA* gene, which

also is regulated through the presence of a LexA binding site, is active in early exponential phase without the presence of DNA damage (Dri and Moreau, 1993).

In the case of both *recA*P1 and *recA*P2 in *M. smegmatis,* it was found that uninduced promoter activity remained constant in all phases indicating that there was no link between uninduced activity and growth phase for either promoter in this species (fig. 4.1). This in turn suggests that the induction of both promoters in this species is dependent on DNA damage. It was also shown that the activity of *recA*P1 in *M. tuberculosis* was independent of growth phase (fig. 4.2). In the case of uninduced *recA*P2 activity, whilst it could be argued that there may be a slight decrease in the activity of the LexA regulated promoter, the overlapping error bars seen can also be interpreted as showing no effect linked with the progression of growth.

Another finding of the growth phase study in *M. smegmatis* is that the activity of the constitutive *recA*P2 (with mutations in the SOS box) remained constant in the different growth phases. As stated earlier, gearbox promoters in a stress-free environment, are associated with increased activity in stationary phase (Aldea *et al.,* 1990; Ballesteros *et al.,* 1998). Therefore, what this finding indicates is that *recA*P2 is not a gearbox promoter. However, this contradicts the information from the point mutations done in the *recA*P2 -10 region, which showed that the -12A base (conserved in the gearbox consensus but not in the heat shock promoter consensus) is vital to the activity of the mycobacterial promoter. With these contradictions, it is obvious that it has become difficult to label this mycobacterial promoter based on the characteristics of promoters from other bacterial species. Therefore, another means of "grouping" this mycobacterial promoter would be to determine which

sigma factor is involved in it recognition. One way of achieving this aim would be to create *M. tuberculosis* sigma factor mutants and screening these strains with *recA*P1 and *recA*P2 constructs to see in which mutants promoter activity was lost.

The studies involving the constitutive *recA*P2 construct in *M. smegmatis* show that unrepressed promoter activity is constant over the whole growth process of the saprophyte. Furthermore, it was found that the activities of the constitutive and induced wild type *recA*P2 in the saprophyte were similar. This in itself does not indicate anything about the sigma factor that recognises this promoter. However, it does show that induction of this promoter element in *M. smegmatis* is dependent on the cleavage of the repressor and not on any associated factors that are produced in DNA damage conditions. The dependence on RecA* mediated cleavage of the LexA repressor has also been shown by inability of a wild type *recA*P2 construct to have enhanced expressed in a *M. tuberculosis* Δ*recA* strain upon induction with mitomycin C (E. O. Davis, personal communication).

4.4.2. Is there a link between the density of culture and inducibility of the *recA* promoters in *M. tuberculosis?*

The findings from the assays done on the inducibility of both *recA* promoter elements in *M. tuberculosis* shows that *recA*P2 activity upon DNA damage is virtually constant regardless of cell density. However, *recA*P1 activity is maximally inducible at low OD_{600} values and steadily declines from that point. This result differs from that found in *E. coli* where a *recA* promoter fused to a *luxCDABE* cassette was found to be most inducible in mid exponential phase (Gu *et al.*, 2000). What the results of assaying promoter inducibility with cell density

in the pathogen show are that there may be a difference in function between the two promoter elements in *M. tuberculosis* with *recA*P1 being important at an early phase in growth/ infection and *recA*P2 being more important at high cell densities.

4.4.3. The kinetics of induction in two mycobacterial species upon DNA damage

Given that the sequences upstream of the *recA* gene in both mycobacterial species were found to be virtually identical, with both promoter regions being highly conserved in both saprophyte and pathogen (Movahedzadeh *et al.,* 1997a; Papavinasasundaram *et al.,* 1997), it was theoretically possible that the kinetics of activity for both promoters in both mycobacterial species could be similar. From this study it has been shown that there are a number of differences in the strengths of both *recA*P1 and *recA*P2 and the way in which these promoters are regulated in each mycobacterial species.

4.4.3.1. The kinetics of induction of *recA*P2 in both mycobacterial species upon DNA damage

It has been shown that the presence of mitomycin C in the growth medium is sufficient to stimulate LexA cleavage in *M. smegmatis* and reduce its binding to the SOS box present within *recA*P2 (Papavinasasundaram *et al.,* 2001). Indeed the kinetics of *recA*P2 induction in the saprophyte, when induced with mitomycin C, is comparable to other LexA dependant *recA* promoters in both *E. coli* (Kim and Oh, 2000) and *B. subtilis* (Cheo *et al.,* 1992) when induced with the same agent. In tests with these promoters it has been shown that after

addition of mitomycin C, there is a period of 30 minutes where promoter activity increases by a relatively small amount (Cheo *et al.,* 1992; Kim and Oh, 2000). After this initial lag in activity there is a rapid increase in the rate of promoter activity, which is maintained in both bacterial species (Cheo *et al.,* 1992; Kim and Oh, 2000). This type of activity was also observed in *M. smegmatis.* This lag in activity is not seen when DNA damage is mediated through the use of ultra violet light in either *E. coli* (Kim and Oh, 2000) or *B. subtilis* (Cheo *et al.,* 1992). The use of UV light treatment provides instantaneous DNA damage and there is an immediate and massive increase in promoter activity. In the case of induction with mitomycin C, the agent has to be taken up by the cell and metabolised to become an active DNA damaging agent and the time taken to achieve this might account for initial lag.

An increase in the activity of *recA*P2 was also seen in *M. tuberculosis* upon addition of mitomycin C. However, from the kinetics of induction in the pathogen, it could be seen that the increase in the rate of *recA*P2 activity in mitomycin C induced *M. tuberculosis* was more gradual than that seen in the saprophyte (fig 4.5a and 4.5b). What this, therefore, indicates is that there is still a degree of repression of *recA*P2 activity in *M. tuberculosis* following the addition of mitomycin C. It has been shown previously that LexA was able to bind the SOS box within *recA*P2 after the addition of the DNA damaging agent to the pathogen (Papavinasasundaram *et al.,* 2001). Therefore, the repression that is seen in the kinetics of *recA*P2 induction with mitomycin C could be through the presence of intact LexA repressor still being able to bind the SOS box located within that promoter. This finding has been substantiated in a construct of *recA*P2 (pEJ443) in which a mutation was introduced into the SOS box. Assays of pEJ443 have shown that expression from this construct in *M. tuberculosis* is much stronger than that of the equivalent wild type

sequence, even after induction with mitomycin C (data not shown, E. O. Davis, personal communication)

Another observation from this study is that the strengths of *recA*P2 in both species varies dramatically (fig 4.5a and 4,5b), with promoter activity in *M. tuberculosis* being greater than that seen in *M. smegmatis.* This difference in promoter strengths of *recA*P2 was also seen in the previous chapter, but mutations placed within the hypothetical promoter had the same effect in both species. This difference in activity seen within both mycobacterial species differs from a previous study where it has been shown that a number of mycobacterial promoters are expressed at comparable levels in both species (Bashyam *et al.,* 1996).

4.4.3.2. The kinetics of *recA*P1 induction in both mycobacterial species with DNA damage.

It has already been shown that although the *recA*P1 promoter does not posses a LexA binding site, it is inducible with mitomycin C in both *M. smegmatis* and *M. tuberculosis.* However, from the kinetics of *recA*P1 induction done in both mycobacterial species, it can be seen that there is a difference in regulation of this promoter in each organism.

In the case of *M. smegmatis*, promoter activity kinetics initially resemble that seen for *recA*P2 in the same species but quickly reach a point of saturation. The activity of this promoter, which is always lower than that of *recA*P2 in the saprophyte, indicates that it plays a minor role if any in the activity seen with the induction of pEJ417. Activity of this

nature has previously been observed in the regulation of the *dnaA* gene in *E. coli*, which also is inducible by DNA damage and lacks a LexA binding site but is dependent on LexA cleavage for full expression (Quinones *et al.,* 1991). In this study it was suggested that the regulatory mechanism controlling this gene was at least indirectly linked to the SOS response (Quinones *et al.,* 1991). However, it has been possible to induce *M. tuberculosis* *recA*P1 in both a *M. smegmatis* Δ*recA* strain and also in a strain of *M. smegmatis* expressing an uncleavable LexA product (E. O. Davis, personal communication). The levels of activity of *recA*P1 in both these strains when induced was similar to that seen in the wild type strain and indicates that the activity of this promoter with DNA damage in the saprophyte is LexA independent.

In the case of the *recA*P1 kinetics of induction in the pathogen, after 6 hours post induction there is a shift upwards in the rate of activity (fig 4.5b). The relatively strong expression of this promoter in the pathogen after this time suggests that this promoter does play a significant role in regulation of *recA* in the pathogen. The period of weak inducibility is much longer than 30 minutes, which indicates that this period of induction cannot be accounted for solely by the metabolism of the DNA damaging agent. Another possibility open, is that a DNA damage inducible factor interacts with this promoter and it takes this period of time for de-repression and expression of this protein to occur. However, what can be said about this DNA inducible factor is that it must also be independent of LexA function as it has previously been shown in a *M. tuberculosis* Δ*recA* strain that *recA*P1 is inducible with the addition of mitomycin C. Furthermore, the activity of this promoter in the mutant strain is comparable with that seen in the wild type in the same conditions.

Through microarray analysis of *M. tuberculosis* cultures induced with mitomycin C for different periods of time, it has been observed that there is increased expression of σ^G after 6 hours of induction (L. Rand, personal communication). Since alternate sigma factors are known to allow the expression of genes required in certain stresses, this sigma factor may be the inducible protein that recognises *recA*P1 in the pathogen although there may be other sigma factors that recognise this promoter. To try to identify which sigma factors recognise *recA*P1 and *recA*P2, the method employed in this study, which is described in the following chapter was to create sigma factor knock out strains and study the activities of the *recA* promoters in these mutants.

CHAPTER 5: PRODUCTION OF SIGMA FACTOR MUTANTS IN *M. tuberculosis*

5.1. Introduction

From the studies undertaken the sequences that made up both *recA*P1 and *recA*P2 had been localised and partially characterised. In addition, further study of the upstream region of the *recA* gene had shown that there were no other promoter elements present. However, all this information only gave part of the story concerning the transcription of *M. tuberculosis recA*. What was needed to complete the information was to identify the sigma factors associated with both promoters. As a first step towards this, it was decided to construct knockout strains lacking individual sigma factors. These knockout strains would then be transformed with a reporter containing either *recA*P1 or *recA*P2, and then induced with mitomycin C and tested with the β-galactosidase assay. From observing the presence or lack of activity of these induced samples, the sigma factor that interacted with each promoter could be inferred. In addition, with the presence of the microarray facilities, the construction of sigma factor mutants could be used to identify genes controlled by each factor which may give an indication as to which consensus sequence is utilised in each case.

5.2. Choice of sigma factor mutants

Of the thirteen sigma factors present in *M. tuberculosis*, it has already been shown that the primary sigma factor, σ^A, was essential to the pathogen (Gomez *et al.*, 1997). In addition, the department already possessed another sigma factor mutant kindly donated by Peter Sander, in σ^M, which had been made in a streptomycin resistant strain of *M. tuberculosis* and was marked with a kanamycin resistance gene. This left eleven factors as possible targets for mutation. Although all eleven factors were included in the initial cloning step it was decided to concentrate on five factors; σ^B, σ^C, σ^D, σ^E, and σ^G for creating knockouts.

The methodology behind the production of knockouts relied on homologous recombination, and was based on the work done by Parish and Stoker (2000) in the creation of a *tlyA plcABC* mutant in *M. tuberculosis*. The first step was to amplify sigma factor sequences plus flanking regions of around 2kb on either side from *M. tuberculosis* H37Rv genomic DNA by PCR (see methods and material section). It had previously been shown that long flanking regions were preferential for introducing mutations in this fashion. In addition, restriction sites absent in the amplified region were introduced to the ends of the flanking sequence to facilitate the cloning of the fragment into the desired vector. Initially it was attempted to introduce the amplified region into the plasmid vector p2NIL, a pBluescript based vector that contained a unique *PacI* restriction site, which in turn would be used to introduce a marker fragment later in the construction process (Parish and Stoker, 2000). However, due to cloning problems with this plasmid, it was decided to use another vector. From this point on there were slight differences in the way that each mutant was made.

5.3. The construction of a *M. tuberculosis* Δ *sigC* mutant

In the case of the σ^C mutant, the fragment containing the complete gene and flanking sequence (4.1kb) was amplified and kinased as described in the methods (see fig 5.1 for region amplified and construction of the initial knockout plasmid). This blunt-ended fragment was then introduced into the pBluescriptKS vector at the *EcoR*V restriction site located in the multiple cloning site. The ligation mixture was then transformed into *E. coli* DH5α cells with blue-white screening and restriction digest testing being done to determine the presence of the insert. Primers were then designed to remove 405bp of the *sigC* coding sequence by inverse PCR and also incorporate a *SnaB*I restriction site in its place. The purpose of the site was to allow the introduction of an antibiotic resistance marker to aid in selection if the unmarked construct gave rise to many revertants to wild type. In addition, the presence of a positive marker would act as proof as to the viability of a mutant. The mutant might be out competed by the wild type and the addition of a selectable marker might promote its presence. However, in the case of the *sigC* mutant, it was decided to proceed initially with the unmarked construct.

The resulting PCR product was kinased, re-ligated and transformed into super-competent *E. coli* XL1 blue cells. The resulting construct was modified by site-directed mutagenesis to remove 159bp of the *lacZ* gene and introduce a *Pac*I site in its place. In addition to being able to introduce a screening and counter-selection cassette through the restriction site, it was hoped that the removal of the *lacZ* gene would stop any kind of homologous recombination with the *lacZ* gene contained in the cassette to be introduced later.

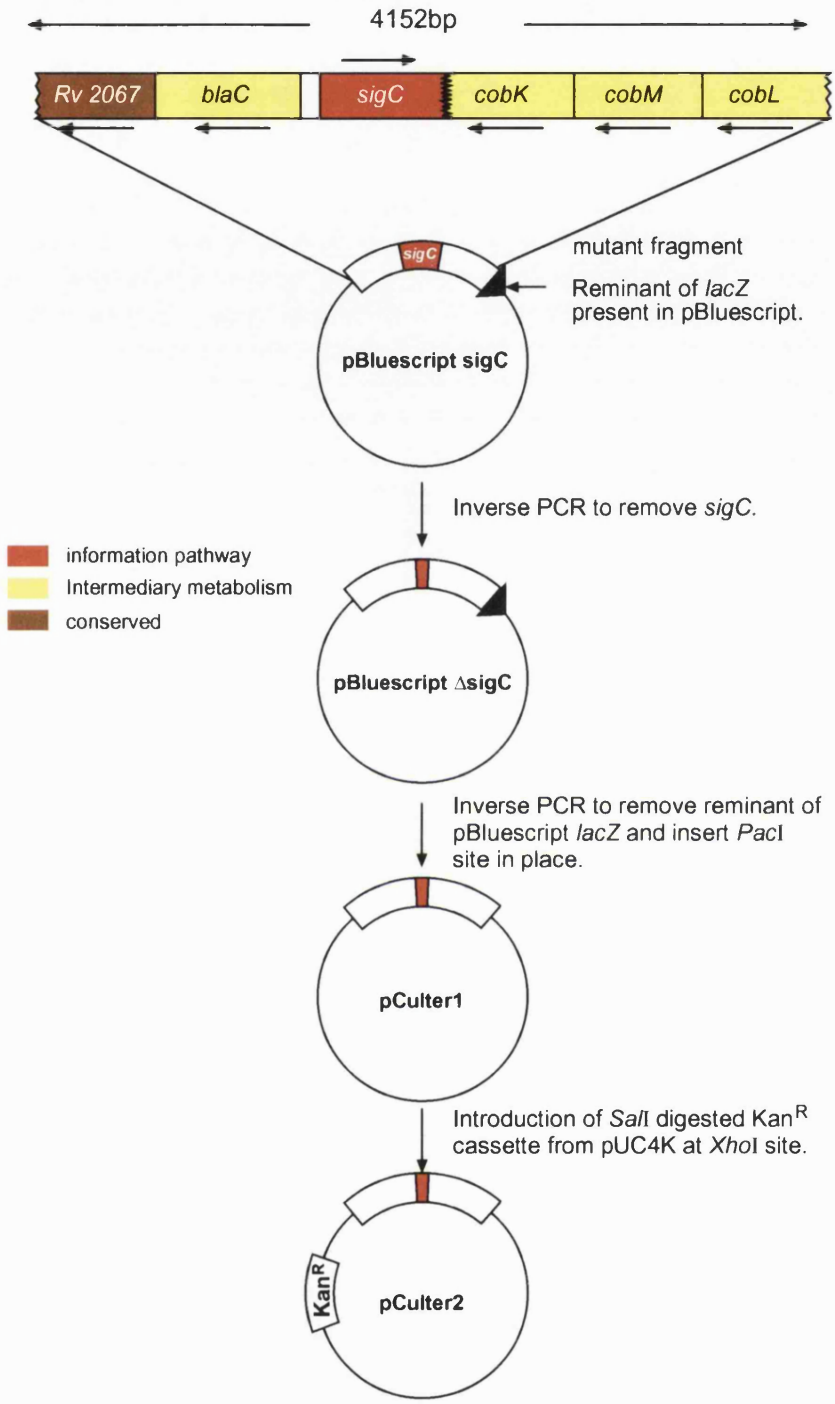


Figure 5.1. This figure illustrates the making of the *M. tuberculosis* Δ*sigC* knockout plasmid construct prior to the addition of the *lacZ* / *sacB* marker cassette from pGoal17 (Parish and Stoker, 2000). The above section shows the PCR fragment containing *sigC* and flanking region that was PCR amplified and placed in pBluescriptKS. The arrows represent which way each ORF is read whist the black region between *sigC* and *cobK* indicates that there is a small overlap in sequence between the two genes. Each region is colour-coded for function.

It had been shown previously that the presence of markers on opposite sides of the desired deletion led to an increase in frequency of the correct double cross-over identified (Hondalus *et al.,* 2000). Therefore, a kanamycin resistance gene was excised from pUC4k and digested with the restriction enzyme *Sal*I and ligated to the unique site *Xho*I site in the mutant construct.

The final step was the addition of the marker cassette (a *lacZ* / *sacB* construct) from pGOAL17 (Parish and Stoker, 2000). After this final step, the complete plasmid (pCulter3) was ready for transformation into *M. tuberculosis* H37Rv. The *lacZ* gene in the marker cassette would enable a blue-white screen to be used to identify colonies that had initially taken up the mutant construct as possible single cross-overs (signified by blue colonies). The *sacB* gene was to be used as a negative screen in the determination of double cross-overs, whereby sucrose resistant colonies grown on plates containing sucrose would signify loss of the vector. Using X-Gal plates at this step and screening for white colonies could also be undertaken as a confirmation of the loss of the vector.

It had been stated that UV or alkaline treated plasmid DNA had a greater tendency to be incorporated by homologous recombination in *M. tuberculosis* than non-treated DNA (Hinds *et al.*,1999). For this reason, it was decided to use alkaline treated as well as untreated DNA to transform *M. tuberculosis* cells.

In the transformation, 120µg of both treated and untreated S.N.A.P miniprep (Invitrogen) DNA was used to transform 400µl of *M. tuberculosis* H37Rv. From these transformations, 100µl was plated on 7H11 Kan^{25} $X\text{-}Gal^{100}$ plates to select for integration and possible single

crossover events (blue colonies), with the plates being scored after four weeks. The selection procedure is shown in Fig 5.2.

Only one colony was obtained, which happened to be blue and was from the transformation done with untreated DNA. This sole colony was streaked out on a 7H11 plate without any antibiotic pressure to allow for the double crossover event to occur. After three weeks of growth, a loopful of bacteria was taken and dilutions of $1x10^{-1}$, $1x10^{-2}$ and $1x10^{-3}$ were made. These dilutions were plated on 7H11 plates with 2% sucrose and X-Gal^{100} to select for loss of both *sacB* and *lacZ* genes (loss of the vector and, therefore, the double crossover event).

Twenty Suc^{R} white colonies were isolated from the $1x\ 10^{-3}$ plate on the basis of the right phenotype. These colonies were then patch tested on 7H11 Kan^{25} plates and screened for Kan^{S} clones, which indicated that the vector backbone had been lost. Of the twenty colonies isolated, six were found to have the Kan^{S} phenotype. Genomic DNA was isolated from these six colonies showing the right phenotype for testing by PCR amplification and Southern blot.

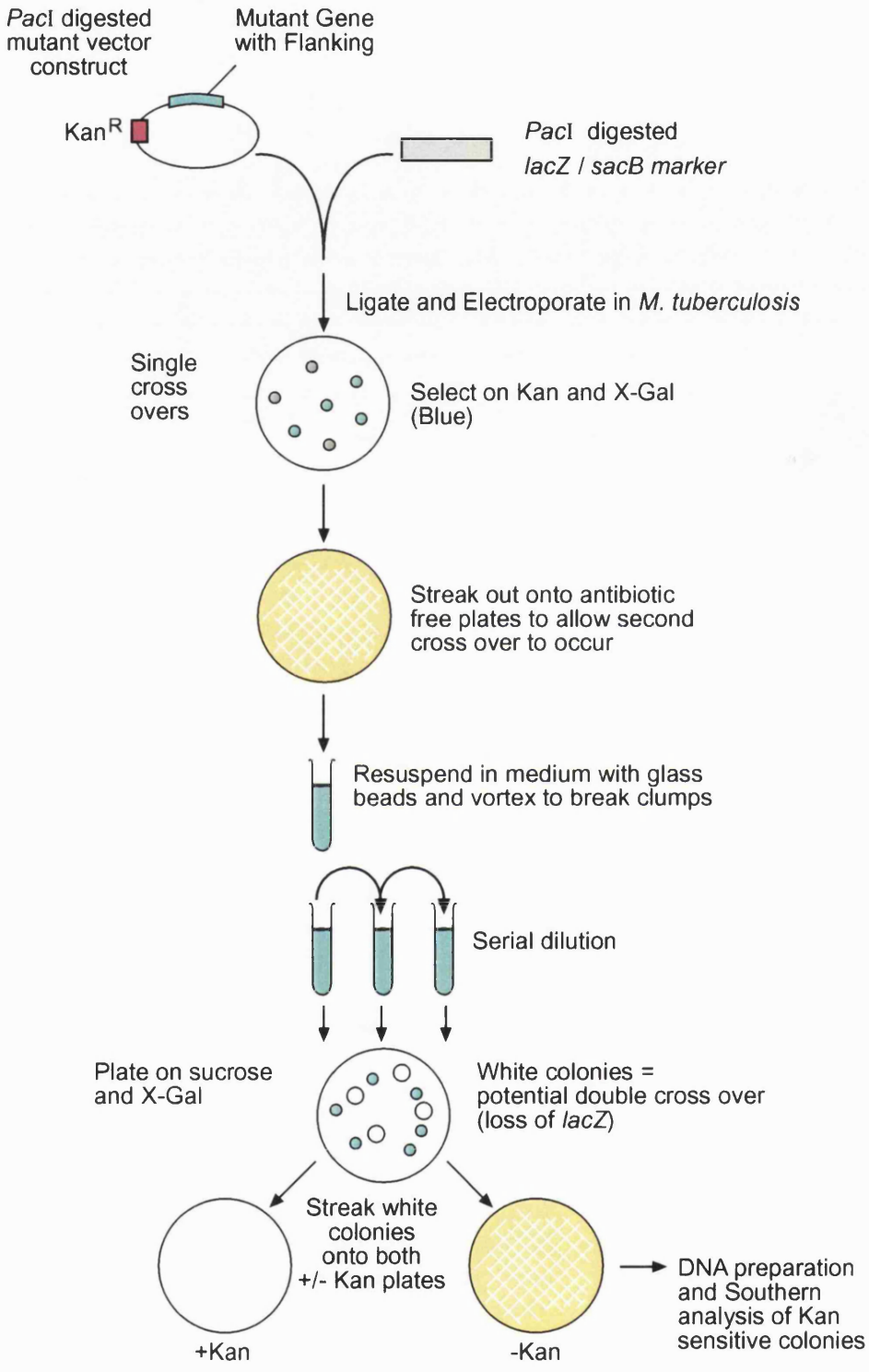


Figure 5.2. This figure illustrates the procedure by which the unmarked sigma factor mutants attempted in this study were created. This methodology was based on the work of Parish and Stoker, (2000) in the creation of a *tly plcABC* mutant in *M. tuberculosis* by homologous recombination.

The primers for PCR were based on the flanking sequence of the *sigC* gene and from this it was seen that one isolate had an amplified signal that was expected of a knockout (fig. 5.3). However, the same signal would also be given from a single crossover and, therefore, a Southern blot was done to confirm the knockout. Probes of around 400bp based on sequences outside the 5' and 3' regions initially amplified with *sigC* were made. Genomic DNA from the six strains isolated as well as the wild type were digested with *Xho*I (there was a *Xho*I site in the wild-type that was missing in the mutant allowing them to be easily distinguished (Fig 5.4). In the case of the wild type strain, band sizes of 4.1 kb or 4.9 kb would be expected depending on which probe was used while the knockout strain would give a band of 8.6 kb regardless of the probe used. From the resulting blot it was seen that one of the six isolates was indeed a knockout strain. The mutation was later confirmed by repeating the Southern blot with a new probe of about 300 bp based on the region deleted from *sigC*. In this case DNA from the knockout and mutant were digested with *Eco*RV. Whilst a band of 2.7 kb was expected when the wild-type was probed, no band was expected in the mutant. This Southern as shown in fig 5.4 confirmed that a deletion had been introduced into *sigC* in this strain.

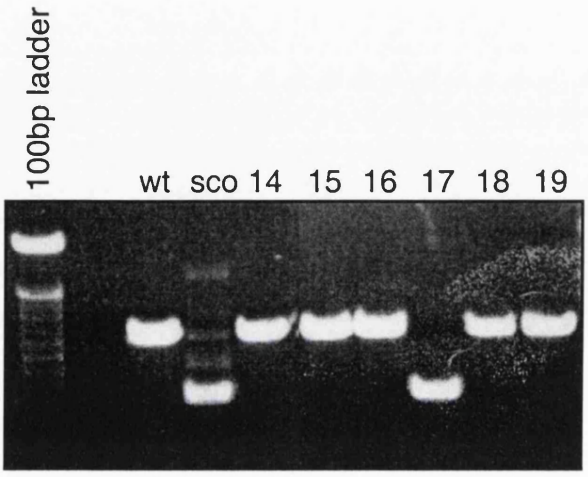


Figure 5.3. This figure illustrates the PCR amplification of DNA isolated from 6 potential *M. tuberculosis* Δ*sigC* double crossover colonies. Primers were based on sequence in the flanking genes of *sigC* which were present on the fragment used to create the knockout construct. These primers amplified up a fragment of around 1 kb in the wild-type and around 700bp in a correct *M. tuberculosis* Δ*sigC* double crossover. Wild-type DNA and DNA isolated from a single crossover were used as controls. From this figure, it can be seen that only one sample, 17, is correct by PCR. Interestingly, although it was expected for there to be two bands in the single crossover corresponding to wild-type and mutant, only one band, corresponding to that of the mutant, was seen.

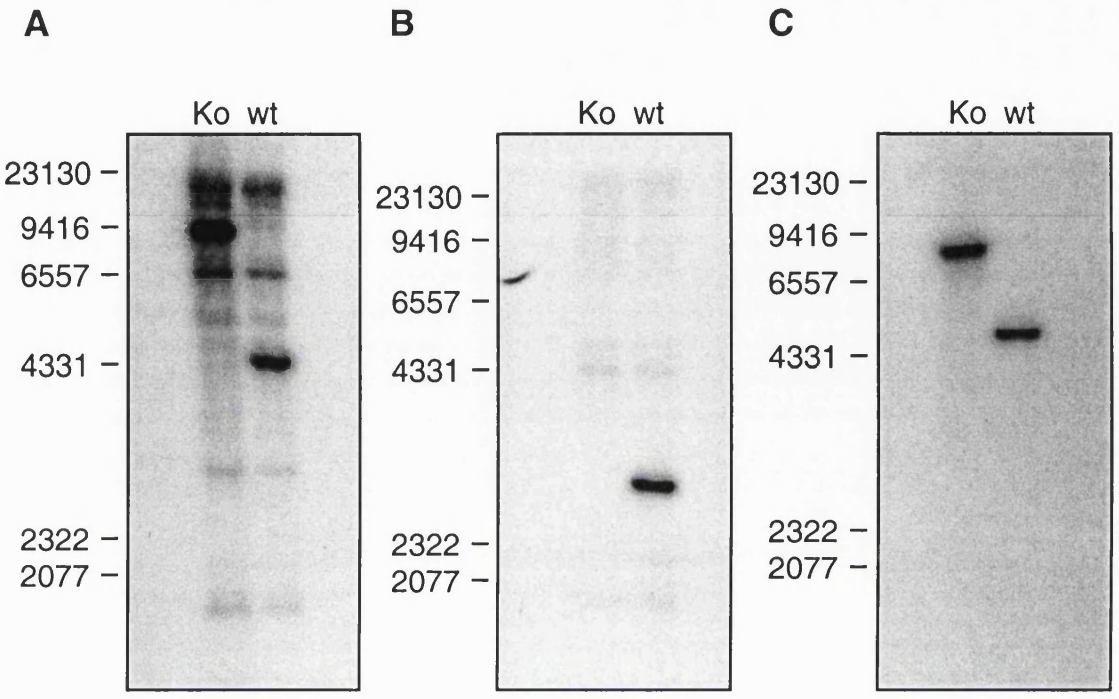


Figure 5.4. This figure illustrates the Southern blot undertaken to confirm that sample 17 was a *M. tuberculosis* Δ*sigC* strain. Three probes were used in this study. One probe (A) was based on the 5' region outside the PCR fragment that was part of the mutant construct, whilst a second probe (C) was based on the 3' region outside the PCR fragment that was part of the mutant construct. Both these probes were used in blots involving *Xho*I digested wild-type and strain 17 genomic DNA. In the case of wild-type sequence, probe A would hybridise to give a 4.1 kb band, whilst probe C would give a 4.9 kb band. In the case of a double crossover for a Δ*sigC* strain, both probes would give a band of size 8.6 kb. In addition, a third probe (B), based on the region deleted from *sigC* by inverse PCR was also used to probe both samples of genomic DNA., which in this case was digested with *Eco*RV. Whilst a band of 2.7 kb was expected from wild-type DNA, no band would be expected in a true *M. tuberculosis* Δ*sigC* strain. From the resulting blots it can be seen that strain 17 was indeed a *M. tuberculosis* Δ*sigC* strain.

5.4.The construction of a *M. tuberculosis* Δ*sigD* strain.

In the case of the Δ*sigC* mutant, the pBluescript vector had been modified after the insertion of the amplified *sigC* fragment. However, in the case of σ^D and the other mutants attempted, it was decided to start with a vector already adapted for the task, leaving only the mutation of the sigma factor and insertion of the pGOAL17 cassette to be done after the cloning of the individual σ factor. This vector was termed pBackbone.

The basis of the pBackbone vector was the commercial available pBluescript KS- vector (Stratagene). This vector was modified by the removal of part of the *lacZ* gene within that vector and replacing this with a *PacI* site. This was done through site directed mutagenesis with the primers.2205KKG3851CZ and 2205KKG3851DA (see appendix for primer sequences). This became the initial vector, pAgohri which, when transformed into *E. coli*, gave rise to white ampicillin resistant colonies on L-agar Amp^{100} $X\text{-}Gal^{100}$. The plasmid was completed with the addition of a kanamycin resistance cassette, which came from a *SalI* digested pUC4k plasmid (*Xho*I was not used because there was an *XhoI* site present in the resistance marker). This was introduced at the unique *XhoI* site of pAgohri to give rise to pBackbone.

The fragment containing the complete *sigD* gene and flanking sequence (3.8kb) was amplified and again kinased and ligated into the *EcoRV* site of pBackbone (see fig 5.5 for amplified region and construction of the initial knock out plasmid). Primers were then designed to remove 564bp of the coding sequence and introduce a *SnaB*I site.

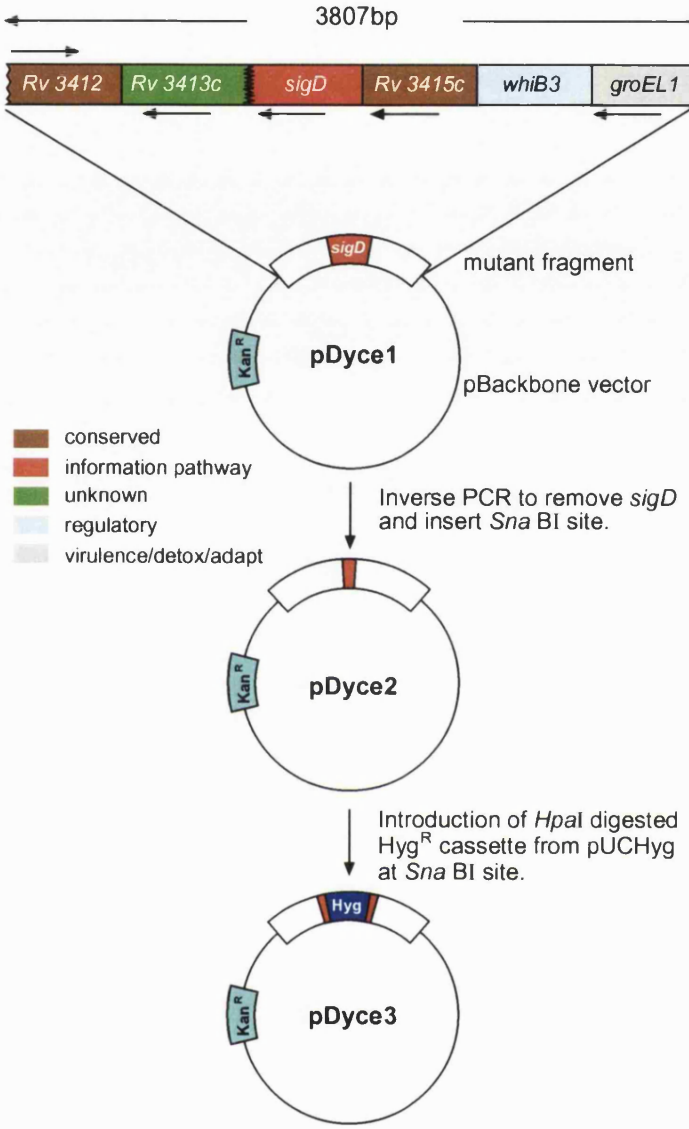


Figure 5.5. This figure illustrates the making of the *M. tuberculosis* Δ*sigD* knockout plasmid construct prior to the addition of the *lacZ* / *sacB* marker cassette from pGoal17 (Parish and Stoker, 2000). The above section shows the PCR fragment containing *sigD* and flanking region that was PCR amplified and placed in pBlackbone, a modified pBluescript KS vector that had the remaining region of *lacZ* removed by site directed mutagenesis and replaced with a *Pac*I site. In addition, a Kan^R gene had been inserted at the sole *Xho*I restriction site. The arrows represent which way each ORF is read whist the black region between *sigD* and *Rv3413c* indicates that there is a small overlap in sequence between the two genes. Each region is colour-coded for function. The Δ*sigD* construct was the only one made marked and this was done by the insertion of the Hyg^R cassette from pUCHyg (Mahenthiralingam *et al.*, 1998).

Although the Δ*sigC* mutant was created unmarked the Δ*sigD* mutant was marked with the hygromycin resistance gene from pUCHyg, which was introduced at the *SnaBI* site. In addition to the points raised whilst making the Δ*sigC* knockout, another advantage would be the ability to attempt a direct selection of the double crossover. The addition of the pGOAL17 marker cassette at the *PacI* site made the final construct, pDyceHyg3 (see fig 5.6 for the construction of this plasmid and selection of the mutant strain).

In the production of the *M. tuberculosis* Δ*sigC* mutant, some of the mutant construct was alkaline treated prior to transformation in the pathogen. However, no transformants were isolated from this treated DNA. Given that only one colony was isolated using untreated DNA, the result of the alkaline treated DNA cannot be used to pass judgement on whether this process improves transformation efficiency. However, it was decided that this step should be omitted for this and the remaining three attempted knockouts. It was also thought that the amount of DNA used in the transformation was too little and this may be the reason for the scarcity of colonies at the first selection step for Δ*sigC*. Therefore, 600μg of untreated S.N.A.P miniprep (Invitrogen) DNA of pDyceHyg3 was transformed into *M. tuberculosis.*

In addition to plating on 7H11 Kan^{25} $X\text{-}Gal^{100}$ plates to select for the single crossover event, since the mutation was marked, it was decided to attempt a direct selection of a double cross over event by plating onto 7H11 Hyg^{50} $X\text{-}Gal^{100}$. In this case selection would be for hygromycin resistant white colonies that would indicate that the *lacZ* gene, and also hopefully the vector, had been lost (fig 5.6). Hygromycin resistant blue colonies would

ideally represent single crossovers although these could also occur through random integration.

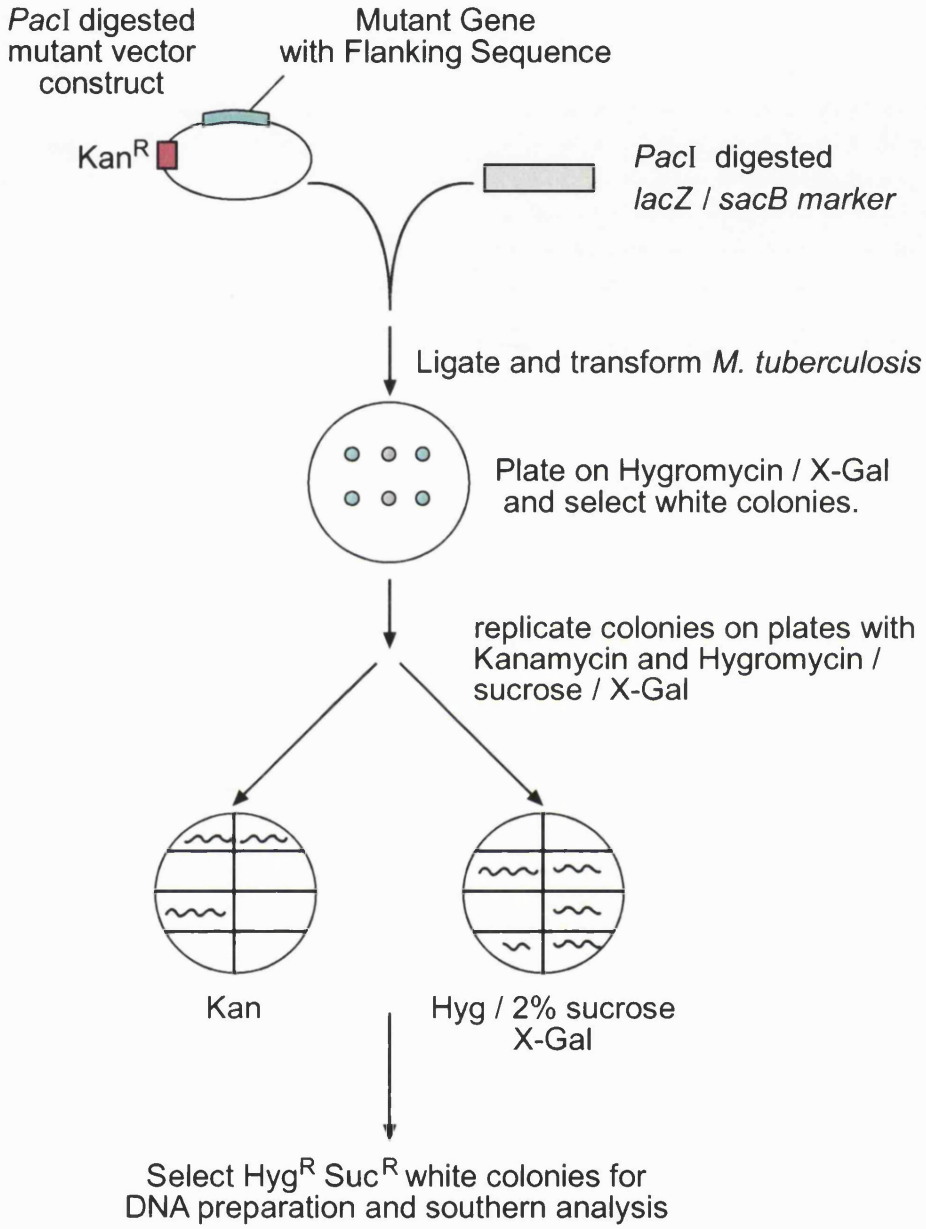


Figure 5.6. This figure illustrates the procedure by which the *M. tuberculosis* Δ*sigD* mutant was created. This methodology was based on the work of Parish and Stoker (2000) in the creation of a *tly plcABC* mutant in *M. tuberculosis* by homologous recombination. However, unlike the technique used to attempt to obtain unmarked mutants, which was a two-step approach, since Δ*sigD* was marked it was decided to attempt for direct selection of the mutant.

From this direct selection, 65 white hygromycin resistant colonies were isolated. However, no blue hygromycin resistant colonies were isolated in this procedure. This result was unexpected and, therefore, it was decided to patch-test all 65 colonies to screen for kanamycin sensitivity and sucrose resistance. From this screening, 36 were found to have the right phenotype. Genomic DNA was isolated from 10 of these strains by the Instagene method (BioRad). The DNA obtained by this method is suitable for PCR although it is not of good enough quality for Southern blots. Two sets of primer pairs were designed to amplify regions between the hygromycin resistance and sequence outside either the 5' or 3' flanking cloned regions. From the resulting amplification it could be seen that here were two possible knockouts (see figure 5.7. strains 64 and 65) of which 64 was further tested by Southern blotting.

Sample 64 was grown up in liquid culture and its DNA extracted. To test the 5' region, DNA from both the wild-type and the potential knockout were digested with *Sac*I, which was expected to give rise to a band of 12 kb for the wild-type and 4.2 kb for the mutant. For testing the 3' end, DNA was digested with *AlwN*I , with expected sizes of 3.5 kb for the wild type and 5.0 kb for the knockout strain. In addition, the deleted region was used to probe *Not*I digested DNA samples from both the wild-type and knockout strains. A band of 5.1 kb was expected in the wild-type whilst no band was expected to be seen in the knockout. The resulting Southern blots (Fig. 5.8) confirmed that sample 64 was indeed a knockout strain of *sigD.*

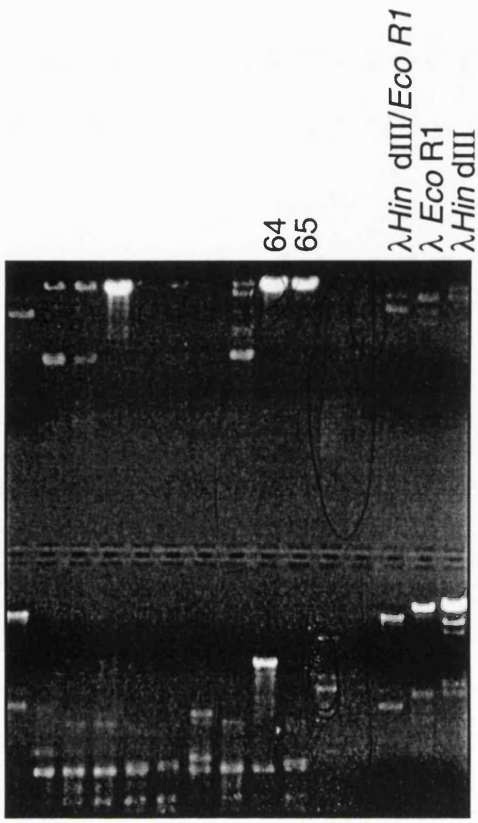


Figure 5.7. This figure illustrates the PCR amplification of 9 possible *M. tuberculosis* Δ*sigD* double crossover mutants. Two sets of primers were used to amplify from either the 5' or 3' end with one primer from each set being located within the Hyg^R cassette whislt the other was located "external DNA" that was not part of the PCR fragment used to create the Δ*sigD* mutant construct. From this PCR screening, a potential mutant would at the 5' end (top) produce a band 2.5 kb, whilst at the 3' end (bottom) produce a band of 3.2 kb. From this experiment, only two strains, 64 and 65 gave bands of the correct size with both sets of amplification.

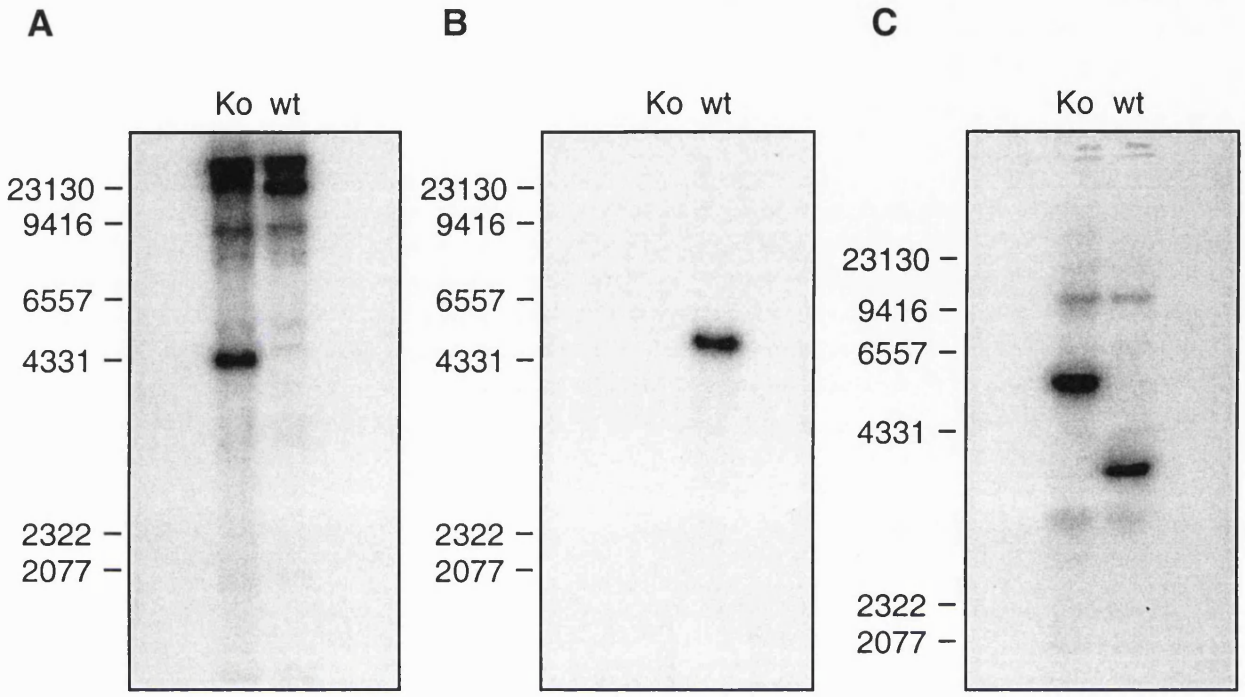


Figure 5.8. This figure illustrates the Southern blot undertaken to confirm that sample 64 was a *M. tuberculosis* Δ*sig*D strain. Three probes were used in this study. One probe (A) was based on the 5' region inside the PCR fragment that was part of the mutant construct, whilst a second probe (C) was based on the 3' region outside the PCR fragment that was part of the mutant construct. Probe A was used in blots involving *Sac*I digested wild-type and strain 64 genomic DNA and would hybridise with a band of 12 kb in the wild-type and 4.2 kb in the mutant. Probe C was used in blots involving *AlwN*I digested wild-type and strain 64 genomic DNA and would hybridise with a band of 3.5 kb in the wild-type and 5.0 kb in the mutant. In addition, a third probe (B), based on the region deleted from *sigD* by inverse PCR was also used to probe both samples of genomic DNA, which in this case was digested with *Not*I. Whilst a band of 5.1 kb was expected from wild-type DNA, no band would be expected in a true *M. tuberculosis* Δ*sigD* strain. From the resulting blots it can be seen that strain 64 was indeed a *M. tuberculosis* Δ*sigD* strain.

5.5. The attempted construction of *sigB*, *sigE*, and *sigG* knockouts.

As with the Δ*sigD* knockout, pBackbone was used as the vector for the Δ*sigB*, Δ*sigE* and Δ*sigG* knockout constructs. A fragment containing the complete gene and flanking sequence was amplified for *sigB*, *sigE* and *sigG* and was introduced at the *EcoR*V restriction site. Inverse PCR was used to eliminate 661bp of *sigB*, 681bp of *sigE* and 691bp of *sigG*. In addition, an *EcoR*V site was introduced into the Δ*sigB*, an *Hpa*I site in the Δ*sigE*, and an *Avr*II site was introduced in the Δ*sigG* mutants. It was hoped to introduce the hygromycin resistance cassette into these three constructs. However, although a number of attempts were made at this, they were unsuccessful. There was evidence to show that the Δ*sigB* and Δ*sigE* knockout constructs were viable (Wu *et al.*, 1997; Gomez *et al.*, 1998; Raman *et al.*, 2001; Manganelli *et al.*, 2001) and the Δ*sigC* mutant had been created with an unmarked construct. It was, therefore, decided to proceed with these three unmarked constructs. The *lacZ-sacB* cassette was added to all three constructs and 600μg of S.N.A.P. miniprep (Invitrogen) DNA was used in each instance to transform 400μl of *M. tuberculosis* H37Rv. All transformations were plated on 7H11 Kan^{25} $X\text{-}Gal^{100}$ solid media.

From the first selection (blue kanamycin resistant colonies) three possible single crossovers were isolated for *sigB*, one was isolated for *sigE* and four were isolated for *sigG*. Following the second selection for white sucrose resistant colonies and patch testing for kanamycin sensitivity, 27 possible double crossover mutants were isolated for *sigB* and 33 were isolated for *sigG*. However, no possible *sigE* double crossovers were isolated which indicated that the initial isolate most probably came about through random integration rather than homologous recombination.

Genomic DNA was extracted from all the possible Δ*sigB* and Δ*sigG* mutants using the Instagene method (BioRad). The isolated genomic DNA was then subjected to PCR amplification to screen for potential knockouts. However, this analysis revealed that all the colonies obtained for both target genes were actually wild type (data not shown).

5.6. Discussion

In this part of the study, it was attempted to make knockout mutants in five sigma factor genes. However, of these five, only two knockouts were successfully derived. Of the three that had not worked, there was already documented evidence to show that a disruption of *sigE* was not lethal to *M. tuberculosis* (Manganelli *et al.*, 2001). One reason why this knockout was not achievable in this study could be that the colony obtained from the Δ*sigE* construct at the first selection step was the result of a random integration, in which case the plasmid could not be excised to leave the mutation in place.

In the case of the failure to derive the other two sigma factor mutants, Δ*sigB* and Δ*sigG* it could be that these genes are essential for the viability of the pathogen in the conditions used in this study. However, at least in the case of *sigB*, this seems unlikely. As already mentioned in the introduction, the *sigB* gene from *M. smegmatis,* which is homologous to the gene seen in the pathogen, has been successfully deleted (Gomez *et al.*, 1998). In addition, a disruption of *sigE* and deletion of *sigH* in *M. tuberculosis*, which both affect the expression of *sigB* are viable (Manganelli *et al*, 2001; Raman *et al.*, 2001). Even so, it might be the case that a minimal amount of *sigB* expression may be needed for viability.

Currently work is being undertaken with hygromycin resistance marked versions of Δ*sigB* and Δ*sigG* to see if the chances of isolating a knockout are improved. However, the only means for checking if a gene is essential is to provide a second copy of that intact gene and then attempt to replace the chromosomal copy. This procedure has been shown to be effective both in *M. smegmatis* (Gomez and Bishai, 2000) as well as in *M. tuberculosis* (Zahrt and Deretic, 2000; Parish *et al.,* 2001). However, each study has achieved the same objective through different means.

In the Gomez and Bishai study (2000) it was found that the second crossover event to delete the chromosomal *M. smegmatis whmD* was not possible unless the single crossover was transformed with a shuttle vector containing a second copy of this gene. In the case of the Zahrt and Deretic (2000) study, the second copy of the intact gene was introduced into the *M. tuberculosis* single crossover strain on a temperature sensitive vector. Using this method, it was possible to delete the chromosomal copy of the *mtrA* gene (Zahrt and Deretic, 2000). However, when these strains where grown at a non-permissive temperature for the TS vector, it was found that these mutant strains were not viable (Zahrt and Deretic, 2000). An integrative vector is also the basis of the method by which Parish and Stoker, (2000) showed the necessity of *glnE* in the pathogen. After selection for the single crossover, a second copy of *glnE* was placed in an integrative vector lacking the excisionase (*xis*) gene. Although the vector can integrate into the host's chromosomal DNA, without this gene, it cannot excise itself. This vector could be transformed into a strain of *M. tuberculosis* that was already in the single crossover stage. The next stage was to allow for a double crossover whereby the chromosomal copy of the gene would be lost but the

plasmid copy retained. The final step was then to remove the integrative plasmid (and intact gene) by transforming the strain with a plasmid carrying the *xis* gene. If this final step was not possible, then this in turn would be an indication that this gene was essential for viability.

If it is shown that these two sigma factors are essential for viability in the pathogen then another means by which their affect could be studied is through the use of antisense technology. This technology allows the down regulation of necessary genes through the production of RNA that interferes with the expression of the desired gene without having to remove that gene. It has been demonstrated in *E. coli* that it is possible to down-regulate the expression of an alternative sigma factor, σ^{32}, which has been found to be essential for growth at 37°C through this procedure (Srivastava *et al.*, 2000). There are also a number of experiments that have been done in mycobacteria that have employed this technique to control gene expression. A study by Parish and Stoker (1997) showed that antisense could be used in mycobacteria when they used this technique to down-regulate the expression of the *hisD5* gene in *M. smegmatis*. It has subsequently been shown that antisense can also be used effectively to regulate gene expression of both *ahpC* in *M. bovis* (Wilson *et al.*, 1998) and *glnA1* in *M. tuberculosis* (Harth *et al.*, 2000).

CHAPTER 6: ANALYSIS OF THE SIGMA FACTOR MUTANTS

6.1. Introduction

In the previous chapter, the creation of two *M. tuberculosis* sigma factor mutants was discussed. The creation of these two mutant strains, Δ*sigC* and Δ*sigD*, enabled a number of possible avenues of study. In relation to the general theme of this project, these strains could be transformed with *recA*P1 and *recA*P2 constructs to see if the missing factors affected the activity of the two promoter regions, possibly linking the protein to the recognition of the DNA element. However, some phenotypic studies could also now be undertaken in the mutants. In addition to growth curves, to see how the growth rate was affected in the mutant in comparison with the wild type, with the advent of micro-array technology, it was now possible to see how gene functions on a global scale were affected in the knock out strains. It was, therefore, decided to proceed by examining how gene transcription was affected in the Δ*sigC* strain. In addition, preliminary mice infection studies were undertaken using the *M. tuberculosis* Δ*sigC* strain.

6.2. Growth curves of the *M. tuberculosis* mutants

It was observed during their construction that both the Δ*sigC* and Δ*sigD* knockout strains grew less well on solid 7H11 growth media than strains possessing an active copy of the relevant sigma factor. It was, therefore, decided to produce growth curves of both mutant and parental strains to see if there was any change in this phenotype in either mutant. Cells from each of the three strains grown up in liquid culture were used to inoculate Dubos broth supplemented with albumin in such a way as should give each

culture the same starting OD reading. Duplicate cultures for each strain were grown in rolling culture and samples were taken at specific time points until stationary phase had been reached. The resulting plot is shown in figure 6.1.

From these growth curves, generation time for each strain could be calculated by using the formula:

$$\textbf{Generation time (g)} = \frac{0.301}{\text{Gradient of growth curve}}$$

Using this formula it was calculated that in comparison with the wild-type strain (16.6 hours), the Δ*sigD* strain was a slightly faster growing strain (15.1 hours) whilst the Δ*sigC* strain grew slightly slower (18.1 hours). A possible reason for the differences seen between the wild-type and Δ*sigD* curve can be explained by the fact that the mutant strain was grown from a frozen stock whilst the wild-type was inoculated with culture from a rolling culture.

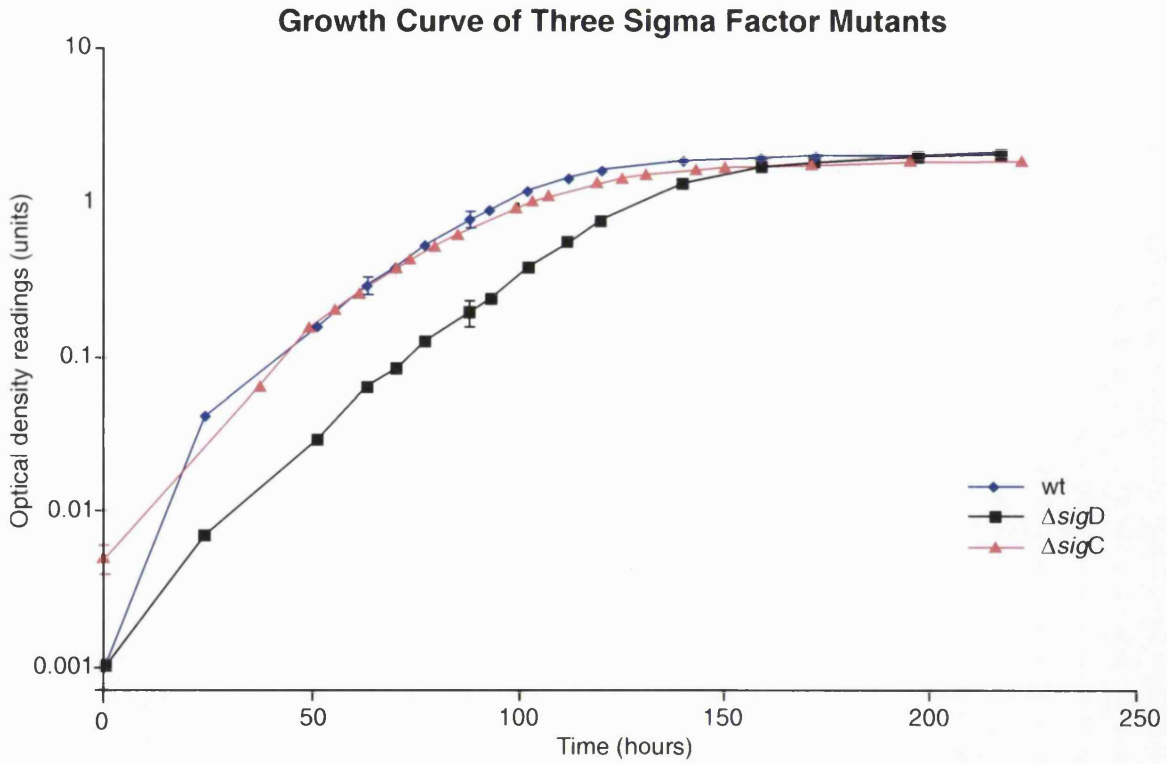


Figure 6.1. This figure shows the three plots made to determine the progression of growth of both the *M. tuberculosis* Δ*sigC* and Δ*sigD* mutants in comparison with the wild-type. Two separate cultures were used for each strain and growth was measured until the plot for each strain levelled off. Generation times for each strain were calculated as 16.6 hours for the wild-type, 15.1 hours for the Δ*sigD* mutant and 18.1 hours for the Δ*sigC* mutant. A possible explanation for the lag in growth of the Δ*sigD* mutant could be that frozen culture was used as the inoculate for this strain whilst in the case of the other two, the innoculum came from established cultures.

6.3. Assaying *M. tuberculosis recA*P1 and *recA*P2 promoter activity in both *M. tuberculosis* mutant strains.

The primary purpose of making the sigma factor mutants was to test the activity of both *recA* promoter elements in an environment lacking one particular factor. This in turn would determine the role, if any, this factor had in the recognition of that particular promoter (Fig. 6.2A & B). Both knockout strains were transformed with reporter constructs carrying each of the *recA*P1 and *recA*P2 promoters individually. In the case of those constructs transformed into the Δ*sigD* strain, these were the kanamycin resistant versions, pEJ449 (*recA*P1) and pEJ435 (*recA*P2) used previously. Although the Δ*sigC* strain was not marked, the constructs used to transform this mutant were the equivalent hygromycin resistant clones pKKG31 (*recA*P1) and pKKG4 (*recA*P2). With hindsight, it would have been better to transform both strains with the kanamycin resistant constructs. Not only could activity of both promoters in both mutants be directly compared, but also given that other studies of these promoters were done with the kanamycin versions, it would make the findings of this study comparable with what had been done previously. This would be useful to show small differences in the way promoter activity was affected in the mutant strains. However, if either sigma factor was responsible for expression from one of the promoters, it would be expected that promoter activity would be lost or dramatically reduced in the mutant strain. Since it was known that the constructs with hygromycin resistance markers were expressed at comparable levels to the kanamycin resistant equivalents in *M. smegmatis* (data not shown), their use in this study would not affect the main outcome.

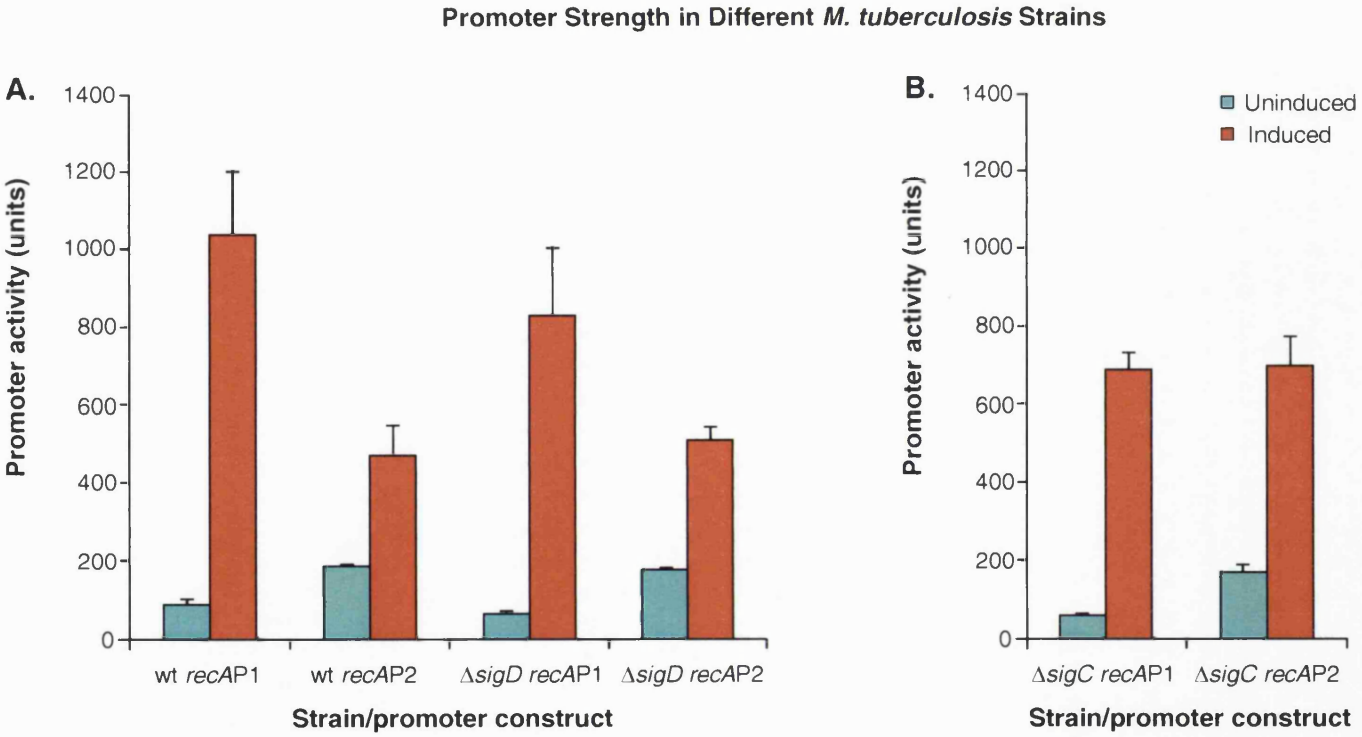


Figure 6.2. This figure illustrates the uninduced and induced strengths of the two *recA* promoters in wild-type, Δ*sigC* and Δ*sigD* strains of *M. tuberculosis*. The same kanamycin resistant *rec*AP1 (pEJ449) and *rec*AP2 (pEJ435) constructs were used to transform both the wild-type and Δ*sigD* strain (Fig 6.2A). However in the case of the Δ*sigC* mutant, hygromycin resistant versions of both *rec*AP1 (pKKG31) and *rec*AP2 (pKKG4) were used (Fig 6.2 B). In all cases, induction was with mitomycin C for a time period of 24 hours. The error bars represent standard errors in from three independent experiments per construct per strain. Although the results of the activity two promoters in the Δ*sigC* mutant cannot be directly compared with those form the other two strains because different constructs were used, it can be seen that both promoters in both mutant strains are still inducible with mitomycin C, demonstrating that neither of these sigma factors were directly involved in the recognition of either promoter.

From the assays undertaken, (Figs 6.2a and b) it could be seen that there was no elimination of activity in the case of either promoter in either mutant strain. This indicated that neither σ^C nor σ^D were directly involved in the recognition of either *recA*P1 or *recA*P2. The possibility remained that one of these sigma factors may be partially responsible for expression in the wild type, but in its absence a second sigma factor which could recognise the same promoter compensated for its absence.

6.4. Micro-array analysis of the *M. tuberculosis* Δ*sigC* strain

It was decided to undertake some micro-array analysis of one of the mutant strains. Since the Δ*sigC* strain was the first to be identified, it was used in this study. Previous work by Manganelli *et al.*(1999), where RT-PCR was used to observe variations in expression of sigma factors *in vitro*, showed that *sigC* expression was affected by growth phase, with high expression in exponential phase and reduced expression in stationary phase. This finding did not appear to correlate with the Δ*sigC* strain growth curve which suggested that *sigC* might have a role to play as growth progressed from exponential phase. For this reason it was decided to see how gene expression in the mutant varied from that in the wild type in two stages of growth. For the purposes of this study, exponential phase samples were taken between OD_{600} values of 0.3 and 0.35 and stationary phase samples were taken at an OD_{600} value of between 1.8 and 1.85. Four mutant cultures were used for the extraction of RNA at exponential phase whilst three cultures were used for the stationary phase studies. Three wild type cultures were used for the extraction of RNA in both exponential and stationary phases. In total at least 6 micro-array slides were used for each strain under each growth condition. The raw data obtained from this analysis was first "cleaned up" through the use of the Gen-

Pix™ program, which was used to remove positive controls, rRNA spots and spots that were of bad quality. The resulting filtered data was then transferred to Gene-Spring software, which was used in the analysis of the data. A minimum change of two-fold (either increase or decrease) in the expression of any particular gene was set as the "cut off point" for this study. Furthermore, values for these selected genes were then subjected to statistical analysis, using the student t-test, where a minimum confidence level of 95% was needed for any change in expression to be considered significant.

From analysis of the data, it was found that in exponential phase 111 genes had enhanced expression of two-fold or above in the mutant when compared with the wild type. By contrast, it was found that 136 genes had a two-fold or greater reduction of expression in the mutant in this phase of growth. In the case of the effects of the deletion on stationary phase gene expression, only 56 had enhanced expression of two-fold or above whilst 120 genes had reduced expression by two-fold or more in the mutant. However if the cut off value in fold change is increased from two to four-fold, it can be seen that there are more genes whose expressions are affected in stationary phase (42) in comparison with exponential growth (7). The full tables of results for this section comprise appendix III, but truncated forms with the ten most differentially expressed genes of lower or higher expression in the Δ*sigC* mutant in both exponential and stationary phase are shown in the following Tables 6.1 to 6.4. Interpretation of the results obtained in the micro-array work of the *M. tuberculosis* Δ*sigC* strain is given in the discussion section.

Table 6.1. The ten most differentially expressed genes with lower expression in the *M. tuberculosis* Δ*sigC* strain that in the wild-type in exponential growth.

Gene	Fold Change	Description
groES	4.2	10 kD chaperonin
Rv2489c	3.3	Unknown
ahpC	3.2	Alkyl hydroperoxide reductase
Rv2415c	3.2	Conserved Hypothetical
Rv0621	3.1	Cell wall process
Rv3395c	3	Unknown
Rv3184	2.9	Transposase
Rv0650	2.9	Transcription regulator (ROK family)
ephB	2.9	Probable epoxide hydrolase
Rv2811	2.9	Conserved Hypothetical

Table 6.2. The ten most differentially expressed genes with higher expression in the *M. tuberculosis* Δ*sigC* strain in exponential phase.

Gene	Fold Change	Description
desA3	12.1	Acyl-[ACP] desaturase
desA1	9.1	Acyl-[ACP] desaturase
Rv1870c	5.2	Unknown
desA2	5.1	Acyl-[ACP] desaturase
Rv0823c	4.8	Transcriptional regulator (NifR3/ Smm1 family)
Rv2949c	4.5	Unknown
kasB	3.7	B-ketoacyl-ACP synthase (meomycolate extension)
rplT	3.7	50S ribosomal protein L20
Rv2989	3.6	Transcriptional regulator (IclR family)
drrA	3.6	Similar daunorubicin resistance ABC transporter

Table 6.3. The ten most differentially expressed genes with lower expression in the *M. tuberculosis* Δ*sigC* strain in stationary phase.

Gene	Fold Change	Description
ahpC	8.6	Alkyl Hydroperoxide reductase
hsp	6.7	Possible heat shock protein
gltA1	6.6	Citrate synthase 3
ahpD	5.4	member of AhpC/ TSA family
Rv1343c	5	Possible membrane protein
groEL2	4.9	60 kD chaperonin 2
Rv1130	4.8	Conserved Hypothetical
groES	4.6	10 kD chaperonin
Rv2226	4.4	Unknown
Rv2706c	4.3	Unknown

Table 6.4. The ten most differentially expressed genes with higher expression in the *M. tuberculosis* Δ*sigC* strain in stationary phase.

Gene	Fold Change	Description
Rv3129	18.8	Conserved Hypothetical
hspX	17.3	14 kD antigen, heat shock protein Hsp20 family
Rv2030c	16.7	Conserved Hypothetical
Rv1738	12.7	Conserved Hypothetical
Rv2626c	10.4	Conserved Hypothetical
Rv3128c	9.6	Conserved Hypothetical
pfkB	9.1	Phosphofructokinase II
Rv2624c	7.8	Conserved Hypothetical
Rv2623	7.7	Conserved Hypothetical
desA1	7.7	Acyl-(ACP) desaturase

6.5. Determination of a possible consensus sequence for the *M. tuberculosis* σ^C factor.

With so many genes expressed at a reduced level in the Δ*sigC* mutant strain in comparison with the wild-type in both phases of growth, it might be possible to identify a common motif in the regions of DNA upstream of these genes representing a possible consensus sequence for the sigma factor. This was attempted using the Gene Spring programme (Silicon Genetics) with genes having reduced expression in exponential phase. However, it was soon observed by the sheer magnitude of possible regulatory elements located by the programme software that the "correct element" would be hard to locate using this method with the currently available information.

6.6. Mice studies involving the *M. tuberculosis* Δ*sigC* strain

In addition to the micro-array work done in the *M. tuberculosis* Δ*sigC* mutant, this strain was also compared with the wild-type for its ability to cause progressive infection of mice by M. J. Colston and B. Chan, who have kindly allowed the inclusion of their results in this study. A total of 11 mice for each strain were inoculated intramuscularly with the knockout strain and at time points of 1, 35, 75, and 135 days post infection, 3 mice per strain were killed (except for the last time point where only two remained). The lung and spleen were removed from these mice and viable counts of bacilli present in each of these organs were determined by colony forming units (cfu) (Fig 6.3 and 6.4). From observing this data it can be seen that numbers of both *M. tuberculosis* wild-type and Δ*sigC* cfus measured at the first three time points in both tissue types are similar. However, at the last time point in both organs it can be seen that there is a drop in the number of *M. tuberculosis* Δ*sigC* cfus measured in comparison with that of the

wild-type, with this decrease being more pronounced in lung tissue (fig 6.4). However, since counts after such a long period of infection can be variable, and the number of mice used was too low, this experiment needs to be repeated in order to draw firm conclusions.

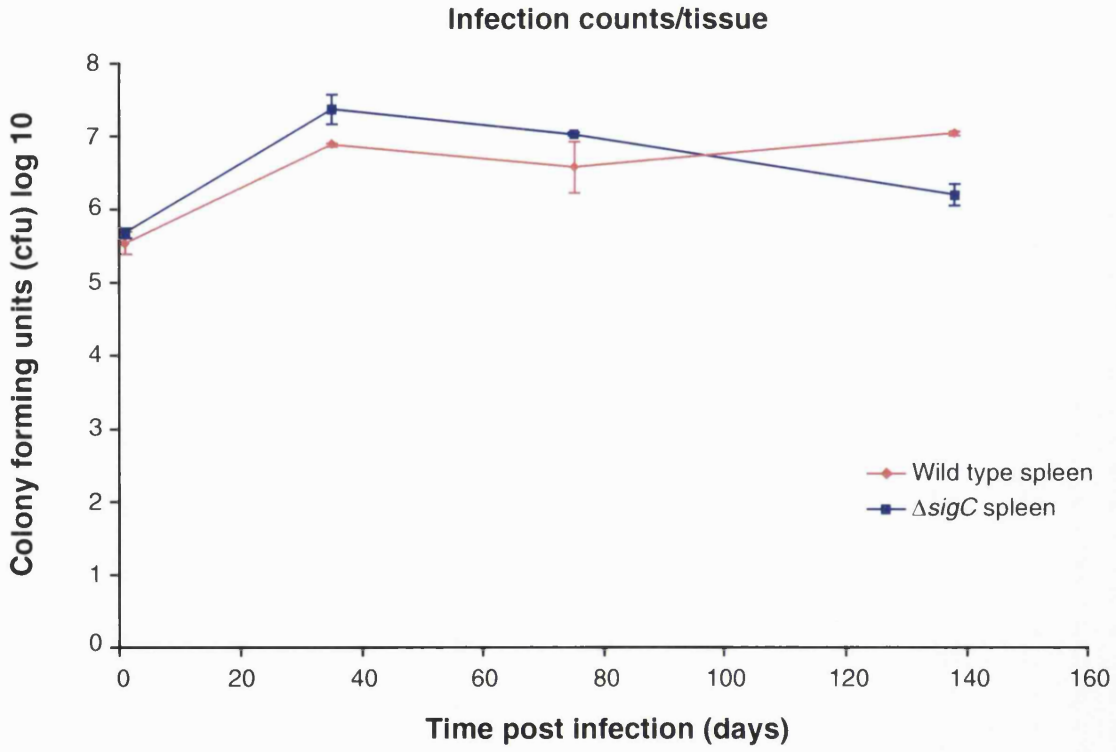


Figure 6.3. This figure illustrates the progression of infection of the *M. tuberculosis* Δ*sigC* in comparison with wild-type as observed through counts of cfus isolated from mice spleen tissue. A total of 11 mice for each strain were infected. At times of 1, 35, 75 and 135 days post-infection, 3 mice were killed (except at the last time point, where only two mice remained).and their spleens remove. From this tissue, viable counts through cfus were made at each time. From this plot it can be seen that there is little difference in the counts of either species until the last time point, where there the wild-type is more numerous than the Δ*sigC* mutant. This data was kindly made available to this study by Dr M. J. Colston and Mr. Bosco Chan from work done by these two named parties.

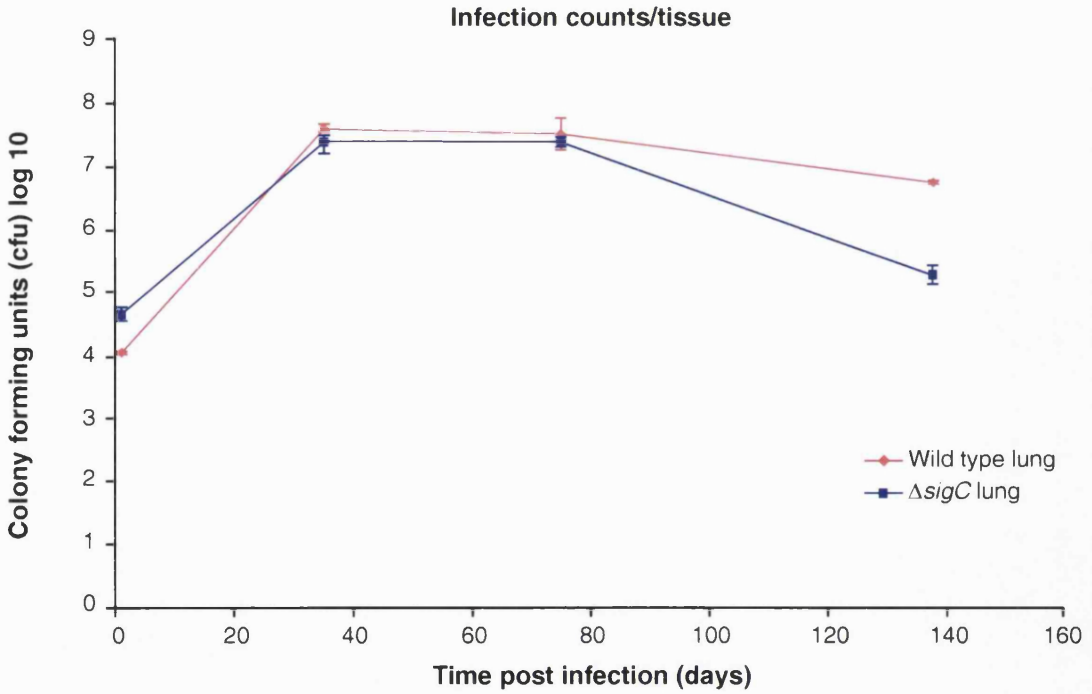


Figure 6.4. This figure illustrates the progression of infection of the *M. tuberculosis* Δ*sigC* in comparison with wild-type as observed through counts of cfus isolated from mice lung tissue. A total of 11 mice for each strain were infected. At times of 1, 35 , 75 and 135 days post-infection, 3 mice were killed (except at the last time point, where only two mice remained).and their lungs remove. From this tissue, viable counts through cfus were made at each time. From this plot it can be seen that, as with findings in the spleen, there is little difference in the counts of either species until the last time point, where there there is over a log difference between the wild-type and mutant. This data was kindly made available to this study by Dr M. J. Colston and Mr. Bosco Chan from work done by these two named parties.

6.7. Discussion

6.7.1. Activity of *M. tuberculosis recA* promoters in both sigma-factor mutants.

Since it was shown that expression of both *recA* promoters was not eliminated in either *M. tuberculosis* mutant strain, both σ^C and σ^D can be ruled out as being solely responsible for transcription from either element. However, the problem as to which sigma factors recognise these promoters still remains. It maybe the case that a second sigma factor can recognise the same promoter and compensates in the mutant strain. Although it has been suggested that there is one specific consensus sequence for each individual sigma factor (Gross *et al.*, 1992), it has been shown in *E. coli* that some promoters recognised by $E\sigma^S$ are also recognised by $E\sigma^{70}$ (Gaal *et al.*, 2001). Theoretically, this can allow the same promoter to be transcribed by the two sigma factors in certain conditions, whilst in others, only one factor would be used (Gaal *et al.*, 2001). It was shown that tolerance to deviation in the consensus sequence determined which sigma factor was utilised by the promoter (Gaal *et al.*, 2001).

The individual subunits of the *M. tuberculosis* RNA polymerase core enzyme have been expressed in a recombinant form in *E. coli* and work is currently underway to reconstitute the enzyme *in vitro* (K. Arnvig, personal communication). In addition, work is currently on going to express some *M. tuberculosis* sigma factor subunits in *E. coli*. If this work is successful, and if it was subsequently possible to form the complete RNA polymerase holoenzyme, bandshifts and *in vitro* transcription with both *M. tuberculosis* *recA*P1 and *recA*P2 could be undertaken as an alternative approach to identifying the sigma factor which recognises *recA*P1 and *recA*P2.

As mentioned in chapter 5, the creation of sigma factor deletion mutants in *M. tuberculosis* is a difficult procedure. It was suggested in this previous chapter that the use of antisense to adversely affect the expression of one sigma factor could overcome some of the difficulties deletion of the gene may pose. However, at the same time, this strategy could cause problems when it came to transform with the promoter constructs. These problems would include the types of plasmid needed to introduce the antisense and promoter fragments as well as the maintenance and induction of these constructs in the pathogen.

6.7.2. Determining a role for *sigC* expression in the pathogen.

It was stated previously in this study that both sigma factor mutant strains grew poorly on solid media in comparison with the wild type. However, it was unclear as to where the difficulty in growth occurred. It was, therefore, of interest to examine how gene expression within the Δ*sigC* strain was affected.

From micro-array analysis of the *M. tuberculosis* Δ*sigC* strain harvested at exponential and stationary phase, a number of interesting findings were made. Of particular note was that the expression of genes in stationary phase was also affected in the mutant. In the *in vitro* study of Manganelli *et al.* (1999), the RT-PCR results of *sigC* expression in the pathogen indicated that *sigC* was the most highly expressed *M. tuberculosis* sigma factor of the 10 studied in exponential growth. However, it was shown in the same study that there was a 10-fold decrease in the expression of *sigC* in stationary culture (Manganelli *et al.,* 1999). From this, it could be suggested that *sigC* might play a role in gene expression during the exponential phase of growth but was not needed beyond this point. Although, it has been stated that the majority of the *M. tuberculosis sigC* gene

had been removed by inverse PCR, around 100 bp at the 5' end of the gene remained in the mutant strain. The probe designed for *sigC* on the micro-array hybridised with the remaining *sigC* sequence, meaning that although this technique could not be used in the identification of a knock out strain, it was possible to see how expression of the knockout gene varied in the mutant. In the experiments undertaken as part of this current study, there was no significant difference in the expression of *sigC* in the wild-type when compared with the mutant strain in exponential growth suggesting that it is not involved in its own expression in exponential phase. However, in stationary phase, there was a two-fold decrease in the expression of *sigC* (appendix III. table 3) in the mutant when compared to the wild-type which suggests that in this phase of growth this factor may play a role in its own expression. Although this result could be explained by reduced stability of the truncated transcript in the mutant strain, if this were the case, one might expect the same effect to happen in exponential phase but no difference was observed there.

From all the micro-arrays done in both phases of growth, the expression of a number of conserved hypotheticals and genes of unknown function was shown to be affected in the mutant, as well as the expression of some genes of known function. Another observation in this study was that, barring a few exceptions, most of the genes affected in the mutant in one growth phase were not affected in the other phase and vice versa. This is not as surprising as it may at first seem because sigma factors can indirectly affect the expression of a vast number of un-related genes. In addition, certain genes would normally be expressed only in one phase of growth and hence could only be affected in that phase. Nevertheless, there was a small but significant number of genes whose expression was altered in both phases and one gene, *drrA*, which was expressed more in the mutant than the wild-type in exponential phase (appendix III. table 2) but

then shifted to being expressed less in the mutant in stationary phase (appendix III. table 3).

6.7.2.1. Genes with reduced expression in the *M. tuberculosis* Δ*sigC* strain

Micro-array data of both exponential and stationary phase cultures showed that in each case the expression of over 100 genes had been repressed to an extent through the deletion of *sigC*. However, only three of these genes, *ahpC*, *groEL1* and *groES* were reduced in both phases of growth.

The function of *ahpC* in other bacteria is associated with defence against hydrogen peroxide. Springer *et al.*, (2001) showed that the survival of a *M. tuberculosis* Δ*ahpC* strain was impaired in comparison with the wild-type following the addition of cumene hydroperoxide, indicating its role in detoxification. It has been shown that *ahpC* is associated with resistance of the *M. tuberculosis* complex to isoniazid (INH) treatment (Wilson and Collins, 1996). Its importance in resistance to this drug was shown when three INH resistant *M. tuberculosis* strains were found to contain mutations in the perceived *ahpC* promoter region that in turn lead to up-regulation of this gene in these strains (Wilson and Collins, 1996). However, it was also found that a *M. tuberculosis* Δ*ahpC* strain was able to infect macrophages (Springer *et al.*, 2001). This in turn indicates that this gene is not required for a primary infection of the cell. Indeed, it has been shown that whilst *ahpC* is overexpressed in the vaccine strain *M. bovis* BCG, it is not expressed in either virulent *M. bovis* or *M. tuberculosis* during growth in macrophages (Springer *et al.*, 2001). In *E. coli*, it has been shown that the expression of *ahpC* is regulated by *oxyR* (Wilson and Collins, 1996). The absence of a functional

oxyR in *M. tuberculosis* is thought to be the reason behind its poor expression in the pathogen (Deretic *et al.,* 1995; Dhandayuthapani *et al.,* 1997). In addition, it has been shown that *M. tuberculosis* AhpC forms a complex with three other proteins, AhpD, dihydrolipoamide dehydrogenase (Lpd) and dihydrolipoamide succinyltransferase (SucB) to produce a peroxidase/ peroxynitrite reductase that acts as a type of antioxidant defence (Bryk *et al.,* 2002).

As for the other two repressed genes, *groEL1* and *groES* both encode for molecular chaperonins, which are part of the heat shock response and are involved in the folding of proteins. In most bacteria, *groEL* and *groES* are found in an operon, known as the *groESL* operon (Hemmingsen *et al.,* 1988). However, in addition to this operon, it has been shown in both *M. leprae* (Rinke de Wit *et al.*, 1992) and *M. tuberculosis* (Kong *et al.,* 1993) that there is also another copy of *groEL,* known as *groEL2* (*Ml0317* in *M. leprae* and *Rv0440* in *M. tuberculosis*) whose expression is unlinked to that of *groES.*

6.7.2.1.1 Genes with reduced expression in the *M. tuberculosis* Δ*sigC* strain in exponential phase

Of the genes that only showed reduced expression in exponential phase, a number of these encoded proteins associated with mobile DNA elements. In the sequenced strain *M. tuberculosis* H37Rv, there are 56 transposons present in the chromosome (Cole *et al.,* 1998). In this study, it was found that 27 ORFs associated with IS*6110* were found to be expressed more in the wild-type strain than the Δ*sigC* mutant in exponential growth. However, there are 36 ORFs associated with IS*6110* and so it is surprising that 9 ORFs were not affected to the same extent in the mutant. One possible explanation is that there is a great deal of homology between these ORFs, which in turn could allow

for cross-hybridisation on the micro-array slide. In this case, the signal given by one spot may be made up from the expression of a number of similar genes. Therefore, if the expression of one of these ORFs is affected, the knock on affect would be that the signal given by the spots to which its product hybridised would also be reduced. Another explanation can be given in *E. coli*, where it has been demonstrated that a number of host factors can affect the transposition of the transposon Tn5 (Reznikoff, 1993). It is possible that there are factors within *M. tuberculosis* that govern the transposition of these mobile elements. These factors could include σ^C or at least have their expression affected by this sigma factor.

In addition, it was found that there was a decrease in expression of a number of genes involved in metabolism of intermediates, regulation of transcription and also a two-component response regulator system (*tcrA* and *Rv0600c*). Other genes of interest that are repressed in this phase of growth include *virS* which is a virulence factor in *M. tuberculosis* and *recC*, which encodes for an exodeoxyribonuclease involved in recombinational repair (see appendix III table 1 for full list).

6.7.2.1.2 Genes with reduced expression in the *M. tuberculosis* Δ*sigC* strain in stationary phase

In stationary phase it was noticed that the number of chaperonins that were expressed at a lower level in the *M. tuberculosis* Δ*sigC* strain had increased from 2 to 8 (with the addition of *dnaJ, dnaK, groEL2, grpE, hsp,* and *tig*). The genes *dnaJ, dnaK*, and *grpE* in *M. tuberculosis* are found together in an operon, which indicates that they would be co-expressed. In *E. coli* it has been stated that *dnaK* is the key molecule of the three, which is involved in the binding of protein requiring re-folding (Suh *et al.*, 1998). This process

requires ATP hydrolysis, which is provided when DnaK interacts with DnaJ (Suh *et al.,* 1998). It is known that GrpE is an antagonist of DnaJ, and it has been suggested that this protein might regulate the interaction between DnaK and DnaJ (Groemping and Reinstein, 2001). It has been shown in *E. coli* that the expression of *dnaJ, dnaK*, and *grpE* is affected by heat shock (Straus *et al.,* 1990). In addition, *hsp* in *M. tuberculosis* is thought to encode a heat shock protein (Cole *et al.,* 1998). Heat shock proteins in addition to being expressed during high temperatures also provide protection in a wide variety of stresses (Morimoto, 1993).

Other genes of interest that have reduced expression in the mutant in this phase of growth include a number of 50S and 30S ribosomal proteins and proteins associated with intermediary metabolism. Special note, however, must be made of certain genes involved in the citric acid pathway. It was found that the expression of *icl* (isocitrate lyase), *mdh* (malate dehydrogenase), *gltA1* (citrate synthase), *sucC* and *sucD* (succinyl-CoA synthase) were repressed in the mutant in stationary phase. Isocitrate lyase is involved in the glyoxylate shunt, malate dehydrogenase is involved in conversion of malate to oxaloacetate, citrate synthase is involved in the conversion of oxaloacetate and acetyl-CoA to citrate and succinyl-CoA synthase converts succinyl- CoA to succinate (Lehninger *et al.,* 1997). Taken out of context, the repression of these genes does not seem to make sense as both the TCA cycle and glyoxylate shunt are affected. However, when genes with enhanced expression in the stationary phase are taken into account, then a possible explanation of their repression can be made (see section 6.7.2.2.1). The reduced expression of *icl* in the *M. tuberculosis* Δ*sigC* mutant in stationary phase may have some bearing on the reduction in the cfu count in the lung seen in the last time point in the mouse infection study done. It had previously been

shown that an *icl* deletion mutant in *M. tuberculosis* had reduced persistence in mice when compared with the wild-type strain (McKinney *et al.,* 2000).

It was also observed that expression of *sodA*, which encodes a superoxide dismutase (SOD), was repressed in the mutant in stationary phase. In *E.* coli it has been stated that this protein is required in making the cell resistant to oxidative stress (Dukan and Nystrom, 1999). It has been shown in *E. coli* that expression of *sodA* is at a maximum in stationary phase (Dukan and Nystrom, 1999). Furthermore, an *E. coli* strain where both SOD genes (*sodA* and *sodB*) had been deleted exhibited greatly reduced survival in comparison with the wild type when a stationary phase culture was grown in aerobic conditions (Dukan and Nystrom, 1999). The repression of this gene in the *M. tuberculosis* Δ*sigC* mutant might, therefore, be one reason for its poor growth in this phase.

Another possible explanation as to why the *M. tuberculosis* Δ*sigC* mutant grows poorly in stationary phase is the repression of *dnaB* expression in the mutant. DnaB is a replicative helicase that interacts with DNA polymerase III at the DNA replication fork (Kim *et al.,* 1996; Yuzhakov *et al.,* 1996). It is known that DnaB is essential for viability in *E. coli* whilst its homologue in *B. subtilis*, DnaC, is important in the viability of this Gram-positive organism (Petit *et al.,* 1998).

6.7.2.2. Genes with enhanced expression in the *M. tuberculosis* Δ*sigC* strain

Although a combined total of over 150 genes were found to have enhanced expression in the mutant strain over the two phases of growth, only 4 genes, *desA1*, *desA2*, *efpA*,

and *ethA* were found to be expressed more than in the wild-type in both growth phases. Of these four genes, *desA*1 and *desA*2 encode desaturases that have been suggest by Cole *et al* (1998) to be more like those found in plants. Desaturases are enzyme associated with converting unsaturated fatty acids in membrane lipids into saturated fatty acids (Murata and Wada, 1995). It has been shown in the cyanobacterium *Synechocystis* PCC 6803 that the expression of these enzymes can be induced through a drop in temperature (Murata and Wada, 1995) as well as either exposure to light (Kis *et al.,* 1998) or salt (Allakhverdiev *et al.,* 1999). These enzymes are also present within the chloroplasts of plant cells and are used in the synthesis of essential fatty acids (Heldt, 1999). These enzymes require oxygen and reduced ferrodoxin to function and in so doing produce water as a by-product (Goodwin and Mercer, 1990; Heldt, 1999). Interestingly, a gene encoding a ferrodoxin is expressed at a higher level in the mutant in each phase of growth, although different genes are affected in the two growth phases. With this information a possible explanation for the up-regulation of the two desaturases can be formulated.

Although aerobic growth provides the most efficient means of respiration for a number of organisms, a build up of oxygen in the cell can be converted, through the Fenton reaction to hydroxyl radicals that in turn will provide oxygen free radicals. An intermediate in the Fenton pathway is hydrogen peroxide, which would normally be dealt with by the action of AhpC (Springer *et al.,* 2001) and SodA (Dukan and Nystrom, 1999). However, the expression of both *ahpC* and *sodA* in the mutant are reduced. One possible solution to the problem of the build up of hydrogen peroxide in this situation would be the over-expression of *katG*, which encodes catalase. Indeed, elevated expression of *katG* was observed in the Δ*sigC* mutant during exponential growth (appendix III, table 2). However, another means of stopping the build up of oxygen free

radicals could involve the saturation of fatty acids, converting the toxic free radical to water.

Although this model is hypothetical, the importance of desaturases in dealing with stresses brought on by aerobic growth could be confirmed by doing micro-arrays of the *M. tuberculosis* Δ*ahpC* mutant compared with the wild-type without any other stresses to see if the *desA* genes are expressed in normal growth conditions.

6.7.2.2.1 Genes with enhanced expression in the *M. tuberculosis* Δ*sigC* strain during exponential phase

During exponential phase a number of genes associated with the production of 50S and 30S ribosomal proteins were expressed at a higher level in the mutant strain. It was also found that the *alkA* gene (possibly involved in BER) in the pathogen was expressed at a higher level, as was another exonuclease, encoded by *xseB*. Other genes of interest that have elevated expression in the mutant in this phase are the efflux proteins *drrA*, *drrB* and *drrC*, which are in an operon. It was also observed that *rho*, a gene involved in transcription termination, was expressed more in the mutant. In addition, *rpoC* but not *rpoB* was expressed at an elevated level in the mutant. As mentioned in the general introduction, the action of these two genes determines the amount of RNA polymerase synthesised. It is, therefore, unclear as to why there should be an increase in the expression only of the β' subunit.

Yet another gene of interest that was observed as being up expressed in the mutant was *katG*, which encodes catalase. This gene is known as a virulence factor in *M. bovis* and

also contributes to isoniazid (INH) sensitivity in this species as well as in *M. tuberculosis* (Wilson *et al.,* 1995; Wilson and Collins, 1996). It was shown in 4 INH resistant *M. tuberculosis* strains that this gene had been repressed in some way, but in three, there was an associated up-regulation of *ahpC* (Wilson and Collins, 1996). The up-regulation of *katG* in the *M. tuberculosis* Δ*sigC* mutant suggests that this relationship might be mutual. Micro-array analysis of the *M. tuberculosis ahpC* mutant in exponential phase would have to be undertaken to prove or disprove this relationship.

6.7.2.2.2 Genes with enhanced expression in the *M. tuberculosis* Δ*sigC* strain during stationary phase.

In comparison with the number of genes seen with enhanced expression in the mutant strain in exponential growth, there were fewer genes with enhanced expression in stationary phase. As mentioned in the section dealing with repressed genes in stationary phase *aceA*, *mdh, gltA*1, *sucC* and *sucD* expression were negatively affected in mutant. If the repression of these genes is applied to the citric acid cycle (fig. 6.5), it can be seen that from citrate clockwise, the end product now becomes succinyl-CoA (Lehninger *et al.,* 1997). With the added information that fumarate reductase is expressed more in the mutant, the cycle is broken and the fumarate present in the cell will now be reconverted to succinate (Lehninger *et al.,* 1997). In effect, conditions are set to favour the production of succinate and succinyl-CoA. It is unclear as to why production of succinate is favoured, but in the case of succinyl-CoA, this molecule is known to be a precursor to porphyrins, which in turn can be used to form cytochromes required for redox reactions (Lehninger *et al.,* 1997).

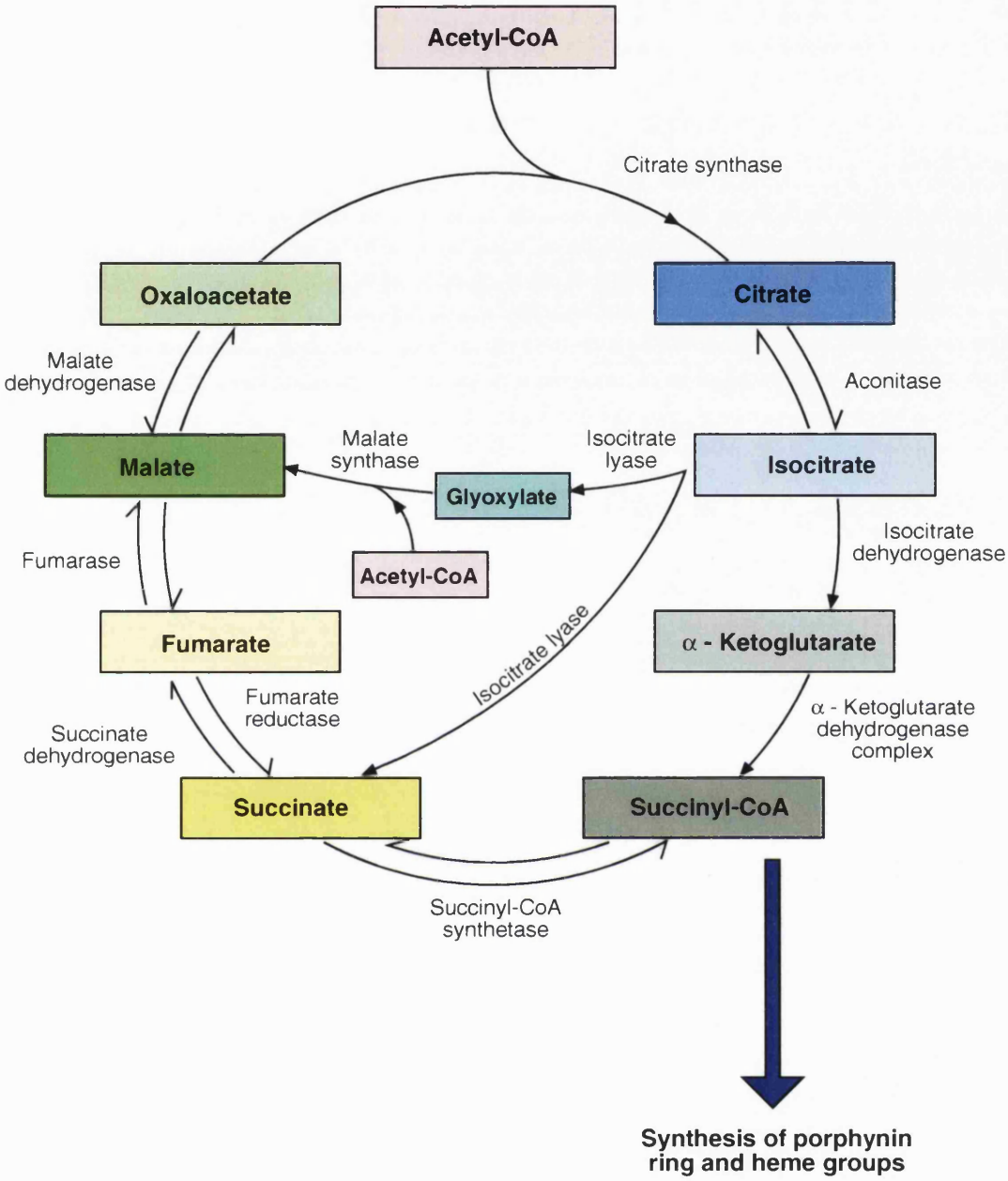


Figure 6.5. This cartoon illustrates the citric acid cycle. It was found in the micro-array data obtained from stationary phase cultures of the *M. tuberculosis* Δ*sigC* mutant that the genes *icl* (Isocitrate lyase), *mdh* (malate dehydrogenase), *glt*AI (citrate synthase 3), *sucC* and *sucD* (both involved in the succinyl CoA synthase complex) are expressed at lower levels than in the wild-type. In contrast another gene linked with the citric acid cycle, *frdA* (fumarate reductase) is expressed at a higher level than in the wild-type. The function of this protein is the conversion of fumarate to succinate. Combining all this information, it suggests that in the mutant there is pressure to produce succinyl CoA, which cane be used in the formation of cytochromes required for redox reactions. This diagram is adapted from Lehninger *et al.* (1997).

Another gene that is expressed more (17-fold higher) in the mutant in stationary phase is the chaperonin, *hspX*. The enhanced expression of this gene is of great interest for two reasons. Firstly, *hspX* is the only chaperonin present in the pathogen that was expressed at a two-fold or higher level, in contrast to the eight previously mentioned chaperonins (section 6.7.2.1.2), which were expressed at lower levels in the mutant in stationary phase. Secondly, *hspX*, is known to be highly expressed in wild-type *M. tuberculosis* in stationary phase (Hu and Coates, 1999), so its elevated expression in the mutant in comparison with the wild-type was of interest. A possible reason for this high fold change in expression of this gene could be that in the absence of σ^C, the sigma factor that recognises the *hspX* promoter may be better able to compete for binding to the core RNA polymerase.

6.7.3. Further experimentation

It is clear from the growth curve data, that the *M. tuberculosis* Δ*sigC* strain has impaired growth in late exponential - stationary phase. From the micro-array data, it has also been shown that the expression of some genes has decreased whilst for others it has been enhanced. The next step in defining the role of *M. tuberculosis* σ^C could be to undertake a proteomic approach and use 2-D gels to determine changes of proteins in the mutant in both phases of growth. Combining both a genomic and proteomic study will give a better understanding to the role of this sigma factor. In addition, micro-arrays and 2-D gel work could be done on *M. tuberculosis* Δ*sigC* cultures under a number of stress conditions to find genes that are regulated by this sigma factor in these states.

Yet another type of study that could be undertaken to determine which genes are regulated by this sigma factor could be to use a technique known as Run-off

transcription/ Macroarray Analysis (ROMA), which has been used in defining the *B. subtilis* σ^W factor consensus sequence (Cao *et al.,* 2002). Using purified RNA core polymerase and σ^W, an *in vitro* transcription reaction was set-up with genomic DNA, which was digested with either *Eco*RI or *Hin*dIII. A control reaction in which only the core polymerase was used was also done. This transcription reaction was labelled with [α-^{33}P] UTP and then used to hybridise to a nylon membrane on which all the 4107 PCR-amplified open reading frames present in the organism were spotted in duplicate. Using this technique 44 possible σ^W dependant promoters were isolated, of which 22 were later shown to be actually dependent on this sigma factor (Cao *et al.,* 2002). In the case of this type of experiment being undertaken in *M. tuberculosis*, the PCR products utilised in production of the *M. tuberculosis* micro-array mentioned in this study could be used. The data obtained from such a study combined with that from the micro-arrays and 2-D gel work would give a more thorough understanding of what genes were recognised by this sigma factor. In addition, the chance of identifying a consensus sequence for the *M. tuberculosis* σ^C factor may be increased by restricting the upstream sequences compared to those of genes which show differential expression by more than one technique. Furthermore, using a *M. tuberculosis* σ^C dependent promoter so determined, random mutagenesis of its sequence, as described for the *M. tuberculosis* *recA*P2 promoter, could be undertaken to more precisely define the consensus.

CHAPTER 7: GENERAL DISCUSSION

7.1. A summary of the findings of this study involving promoter work

From the work undertaken in this study, it has been shown that there are two separate, non-identical promoters, which were termed *recA*P1 and *recA*P2, located in the region upstream of the *M. tuberculosis recA* gene. The -10 and -35 regions of both promoters were localised through specific point mutations in the case of *recA*P1 and a combination of specific and random point mutations for *recA*P2. Specific point mutations in the two regions which could correspond to the -35 element of the *recA*P1 promoter eliminated the region based on positioning as this element. In turn, this showed that the two elements important for *recA*P1 function were located closer to each other than is usual for other promoters. Through both specific and random mutagenesis of the *recA*P2 sequence it was also seen that this region could tolerate a substantial amount of base changes throughout its length and remain active. However, through specific changes made in the -10 region, it was noted that one change, the substitution of a cytosine for an adenine at position -12, greatly reduced *recA*P2 activity. This same base change substitution in the *recA*P1 -10 region had the same effect on the activity of this promoter. The virtual elimination of promoter activity in these two otherwise dissimilar promoters could be attributed to the role of this adenine base in the formation of a promoter open complex.

In addition, it was noted that when both promoters had been effectively eliminated through the introduction of the previously mentioned mutations in the complete fragment that constituted the region upstream of the *M. tuberculosis recA* gene, a small amount of inducible promoter activity remained. A third promoter could not be detected

from further studies, suggesting that the effect was likely due to the stimulation of a weak residual activity from one or both the mutated promoters in the context of the larger DNA fragment. It may be that this region contains a binding site for an accessory protein that might in itself be inducible with DNA damage.

Whilst only one of these promoters, *recA*P2, possesses a LexA binding site, both promoters have been shown to be inducible in *M. tuberculosis*. However, whilst *recA*P1 induction appears to be dependent on the culture density, with activity decreasing with increasing density, inducible *recA*P2 activity is independent of this condition. However, it was also seen that whilst the uninduced activity of *recA*P1 remained constant regardless of growth phase, there appeared to be a small decrease in uninduced *recA*P2 activity between early exponential and late exponential phase. However, this would require further study as the overlapping error bars from this graph could more than likely indicate that there was no change in activity of this promoter associated with changes in growth phase.

Both *recA*P1 and *recA*P2 were also shown to be inducible in *M. smegmatis*, where it has been shown that the *M. tuberculosis recA*P1 and *recA*P2 sequences are conserved in the region upstream of the saprophyte's *recA* gene (Papavinasasundaram *et al.*, 1998). In studies of uninduced activity of both these *M. tuberculosis recA* promoters at different phases of *M. smegmatis* growth, it was noted that the levels of activity in both cases remained constant. However, the most interesting observation was the relationship between the two *M. tuberculosis recA* promoters in each mycobacterial species. Whilst the induced activity of each promoter in *M. tuberculosis* was roughly equal, inducible expression of *M tuberculosis recA*P1 in the saprophyte was poor. In fact, total inducible promoter activity seen from the *M. tuberculosis recA* upstream region in *M. smegmatis*

could be completely be accounted for by the activity of *recA*P2, whereas in *M. tuberculosis* the total activity included contributions from both promoters.

Studies of the kinetics of *M. tuberculosis recA*P1 and *recA*P2 induction in *M. smegmatis* showed that the patterns of their induction were almost similar. In fact, the only differences were those of activity, where *recA*P2 was far stronger that *recA*P1, and that the activity of *recA*P1 saturated after three hours of induction, whilst *recA*P2 activity only slowed after eight hours. However, the same study undertaken in *M. tuberculosis* showed a difference in the way these two promoters were induced in the pathogen. Whilst *M. tuberculosis recA*P2 induction in the saprophyte after addition of mitomycin C was strong and rapid, its response in the pathogen was more sedate. Furthermore, in the case of *M. tuberculosis recA*P1, the pattern of induction seen in the pathogen was nothing like that seen for *recA*P2 in the same organism. Instead, after a pronounced lag in induction that lasted 6 hours after the addition of mitomycin C, the rate of activity increased rapidly. This lag suggests that this promoter in the pathogen is dependent on a protein that is synthesised during mitomycin C induction. It has been shown in another study that the induction of *M. tuberculosis recA*P1 in both mycobacterial species is independent of LexA (E. O. Davis, personal communication). This, therefore, suggests that any protein that might interact with this promoter would itself be LexA independent.

In the final part of this study, two mutant strains of *M. tuberculosis*, Δ*sigC* and Δ*sigD*, were made and transformed with each of *recA*P1 and *recA*P2. Upon induction it was seen that there was not an elimination of promoter activity in either strain that would indicate a direct relationship between the sigma factor deleted and the respective promoter.

7.2. Why does *recA* in both *M. tuberculosis* and *M. smegmatis* require two promoters to direct its expression?

In many other bacterial species, it has been shown that only one promoter directs the expression of *recA*. In the majority of cases this promoter is associated with a LexA binding site at which bound LexA hinders transcription from that promoter. However, as already mentioned, both *M. tuberculosis* and *M. smegmatis* possess two promoters to direct expression of this gene.

In the plant pathogen, *Xanthomonas oryzae* pv. *oryzae* it has been shown that its *recA* gene is under the regulation of three promoters; two of which have associated LexA binding sites (Sukchawalit and Mongkolsuk, 2001). The Sukchawalit and Mongkolsuk study (2001) suggested that the presence of multiple elements to regulate this gene acted as a safe guard against mutation that could affect the activity of one or more of the other promoters. If this was indeed the case, then all the promoters present upstream of their respective *recA* homologue should be comparable in strength and regulation. However, the problem with this theory that not only applies to the plant pathogen, but also to the two mycobacterial species is that the expression levels of the different promoters are not the same throughout the same period of time. In the Sukchawalit and Mongkolsuk study (2001), it was stated that in independent constructs, P1 had strong, inducible activity, P2 has weak but inducible activity and P3 was of moderate, non-inducible activity. In this study it was shown that a longer *recA*P2 construct in *M. tuberculosis* had a comparable inducible activity to a *recA*P1 construct of the same length. However, in the period between addition of the inducing agent and 6 hours after this event, a shorter *recA*P2 construct was the more active of the two promoters.

Yet another example of dual promoters governing a gene belonging to the SOS response is in the case of *E. coli uvrB* (van den Berg *et al.,* 1983). Like the *M. tuberculosis recA* promoters, the more distal *E. coli uvrB* promoter, P2, has an SOS box between its -35 and -10 regions, whilst the promoter closer to the coding sequence, P1, has no associated LexA binding site (Sancar *et al.,* 1982). Although, from *in vitro* work, it had been demonstrated that only P2 was inducible with UV induced DNA damage, from the *in vivo* work, it was subsequently shown that both promoters were inducible with UV treatment (van den Berg *et al.,* 1983). Furthermore, it was shown that whilst the P1 promoter on its own has both high uninduced and induced activity, when it is in association with P2 it is less active in both states (van den Berg *et al.,* 1983). From assays done with the complete fragment in an *E. coli recA*$^-$ strain, it was shown that this induction required the presence of RecA, probably in the cleavage of the repressor. The explanation for this observation given by van den Berg *et al* (1983) is that the role of P2 is to regulate the activity of P1. In effect, although P2 possessed the SOS box, both promoters were repressed from the same binding site. However, for this very reason, this explanation does not justify the presence of two promoters in the region upstream of the *M. tuberculosis recA* gene. In a recent study where *M. tuberculosis recA*P1 activity was measured in a *M. tuberculosis* Δ*recA* mutant where LexA could not be cleaved, induced expression of this promoter with mitomycin C comparable with that seen in the wild-type was observed (E. O. Davis, personal communication).

Another possible explanation as to the necessity for multiple promoters is that they could maximise the transcriptional output of the gene. In the case of *X. oryzae* pv. *oryzae*, this plant pathogen has to overcome a strong H_2O_2 response that is generated by the plant upon infection and high RecA production would be required in this type of DNA damaging condition. In the case of the *M. tuberculosis recA* promoters, from the

results of this study concerning the individual activities of *recA*P1 and *recA*P2 (fig 3.9), it could be suggested that both promoters are working together as their inducible activities when combined are equivalent to that seen in the complete upstream region when induced. However, from the results of this current study and the work of Papavinasasundaram *et al.* (2001), there is still a degree of repression of *recA*P2 in *M. tuberculosis* induced with mitomycin C. In a recent study where a constitutive *recA*P2 construct was made its strength was almost as great as that seen in the complete *recA* upstream region when induced with mitomycin C (data not shown- E. O. Davis, personal communication). If a fully active *recA*P2 is capable of such strength, it is strange that instead of removing the repressor, the strategy employed by the pathogen is to place a second promoter downstream to "boost" transcription.

A more promising reason to explain the presence of these two promoters is that they are regulated by different stimuli and in the case of mitomycin C induction, these stimuli overlap. Dual regulation of *recA* has already been documented in *B. subtilis* where, in addition the DNA damage, enhanced *recA* expression has been linked with a physiological state called competence (Lovett *et al.,* 1989). It was shown that competence-linked induction of *recA* expression in *B. subtilis* happened without cleavage of DinR repressor (Lovett *et al.,* 1989; Hamoen *et al.,* 2001). Instead, the expression of a competence-linked transcription factor allowed the RNA polymerase holoenzyme to bypass the repressor (Hamoen *et al.,* 2001). However, it must be stated that the *B. subtilis recA* gene functions from only one promoter (Cheo *et al.,* 1992).

In the case of the *M. tuberculosis recA* promoters, the action of mitomycin C is sufficient to induce both promoters. Mitomycin C itself is not a DNA damaging agent but when taken up by the cell and reduced, a DNA reactive species is generated

(Tomasz and Palom, 1997). This modified agent causes a number of different cross-linking events in the DNA through alkylation (Tomasz and Palom, 1997). These cross-linkages are sufficient to stall DNA replication and trigger the SOS response. This triggering of the SOS response involving RecA is naturally LexA dependent. However, as stated in the introduction, it has been shown in *E. coli* that RecA can function independently of LexA as part of the recombinational repair pathway to tackle DNA alkylation from exposure to nitric oxide (NO) (Spek *et al.,* 2001). It could, therefore, be conceived that *M. tuberculosis recA*P1 is concerned with RecA production for one type of DNA damage whilst *recA*P2 is used for another type. In the case of mitomycin C induction, where two forms of DNA damage, alkylation and cross-linking, are present, both promoters would be induced.

The concept of dual regulation might also explain why the same scenario of two promoters in the region upstream of *recA* occurs in *M. smegmatis.* Although no promoter work with either *M. smegmatis recA*P1 or *recA*P2 was undertaken in this study, Papavinasasundaram *et al.,* (1998) showed that these promoter regions were very similar to those seen in the *M. tuberculosis* homologue. It is, therefore, conceivable that the functions of *M. tuberculosis recA*P1 and *recA*P2 seen in the saprophyte and the action of its own *recA* promoters upon induction with mitomycin C could be similar. If this were the case, *M. smegmatis recA*P2 would be the dominant promoter in the saprophyte, accounting for all the inducible activity seen in the region upstream of *recA.* Given this scenario, it could also be argued that the strength of the induced *M. smegmatis recA*P2 would have a negative effect on *recA*P1. This effect, known as promoter occlusion, occurs when transcription from a strong promoter "reads through" a weak promoter, stopping its transcriptional activity (Adhya and Gottesman, 1982). Furthermore, Papavinasasundaram *et al* (2001) showed that exposure to mitomycin C

was sufficient to promote LexA cleavage in *M. smegmatis*, meaning that full induction of *M. smegmatis recA*P2 is theoretically possible. However, this does not suggest that *recA*P1 is "surplus to requirement" in the saprophyte. This promoter may also be involved in driving *recA* in such tasks as recombinational repair. However, in the case of mitomycin C induction of *M. smegmatis*, the cross-linking of DNA coupled to the faster generation time of this mycobacterial species might mean that the SOS response plays a more important role in DNA repair in this given situation.

7.3. A summary of the sigma factor work done in this study

In the identification of sigma factors that recognised *recA*P1 and *recA*P2, an attempt was made to create 5 *M. tuberculosis* sigma factor deletion mutant strains (Δ*sigB*, Δ*sigC*, Δ *sigD*, Δ*sigE* and Δ*sigG*). However, only two knock out strains, Δ*sigC* and Δ*sigD* were successfully obtained by the strategy employed in this study. It was seen that the growth of the Δ*sigD* strain though initially impaired in early exponential phase, recovered in middle exponential phase to be like that of wild-type. Indeed, the final OD_{600} value of the Δ*sigD* strain was similar to that seen in the wild-type. However, in the case of the Δ*sigC* strain, the opposite was true as though initial growth was at a rate similar to that seen in the wild-type, this rate dropped as the culture proceeded from early to middle exponential phase. Furthermore, the final OD_{600} value reached by the Δ*sigC* strain was lower than that seen in the wild-type. An examination of gene expression in the *M. tuberculosis* Δ*sigC* strain in stress free conditions at exponential and stationary phase was undertaken with microarray technology. It was found that over 100 hundred genes in both exponential and stationary growth were repressed by two-fold or greater when compared to the wild-type, with *ahpC*, *groES* and *groEL1* being repressed in both phases. Interestingly, it also appears that *sigC* plays a

role in the regulation of itself in stationary phase, with its expression being lower in the mutant that the wild-type in this phase of growth. There were a number of genes that were expressed at a level of two-fold or more higher in the mutant when compared with the wild-type. Of these genes, *desA1* and *desA2,* which encode proteins similar to plant desaturases, were amongst a small number of genes that were expressed at a higher level in the mutant in both phases of growth. It has been suggested in this study that the expression of these genes as well as the repression of others involved in the citric acid cycle may be in response to increased oxidative damage brought on by reduced expression of *ahpC* or *sodA* in the Δ*sigC* mutant.

Infection studies of mice using the *M. tuberculosis* Δ*sigC* strain did not show any difference in the progress of infection when compared with the wild-type at early time points. Equivalent numbers of mutant and wild-type bacilli were isolated from both spleen and lung tissue. However, by 135 days, there was a significant drop in the cfus of the mutant isolated in lung tissue, raising the exciting possibility that σ^C may be responsible for the expression of some genes important for survival at this stage.

7.4. Future experimental work

In the case of each chapter of results, further studies to clarify points made within that particular chapter have been mentioned. However, there are other experiments that could be undertaken to continue various aspects brought up in this study.

7.4.1. Confirmation that different DNA damaging signals regulate expression of the *M. tuberculosis recA* promoters

As mentioned earlier in this chapter, a possible reason for the induction of both *M. tuberculosis recA* promoters with exposure to mitomycin C might occur through the dual DNA damaging functions this compound has when reduced. However, this theory would need to be proven experimentally. One possible set of experiments that could be done is exposing *M. tuberculosis* transformed with either of the promoter constructs to compounds that cause DNA damage through one single function. In this case, agents such as cumene hydroperoxide could be used to stimulate breaks in the DNA whilst NO could be used to produce DNA alkylation. If it could be seen that only one promoter could be induced in the presence of either agent, this would show that *M. tuberculosis recA* expression could be induced from different types of DNA damage and function in different types of DNA repair mechanisms.

7.4.2. The further characterisation of upstream region of *M. tuberculosis recA*P1

Future experimentation should also involve a proper characterisation of the *M. tuberculosis recA*P1 sequence. This type of study is not as straight forward as it was in *recA*P2 for a number of reasons. Firstly, the uninduced basal level expression of *recA*P1 in the pathogen is far less than that seen for *recA*P2. This would make the blue-white screening as utilised in the random mutagenesis protocol for screening the *recA*P2 sequence inefficient. To overcome this problem, the possibility of replacing the promoterless *lacZ* marker with a promoterless *sacB* gene and screening for sucrose resistance to determine mutations that eliminated promoter activity was investigated.

Increasing the percentage of sucrose in the solid growth media had a negative effect on the number of *M. tuberculosis* colonies expressing *sacB* from a weak mycobacterial promoter. However, there were still a few colonies present on the plates with the highest concentration of sucrose (2%). It is known that there is a high tendency for mutations in the *sacB* gene to occur that can change the phenotype from sucrose sensitive to resistant, which in turn would lead to the creation of false positives. In addition, to use a screen of this nature, the assumption that has to be made is that the same sigma factor recognises the promoter in both uninduced and induced states.

It has been stated concerning promoter sequences that "both the DNA sequences of the two conserved regions (-10 and -35) and the optimal spacing between them are unique to the particular sigma factor" (Gross *et al.,* 1992). However, it has been shown that sigma factors that are expressed in different conditions can act on the same promoter. An example of this situation is that of the *bolA1* promoter in *E.coli*, which is recognised by σ^S in uninduced stationary phase but by an unknown alternative sigma factor in stress conditions during exponential phase (Santos *et al.,* 1999). Given that *recA*P2 is repressed through LexA binding, this suggests that the sigma factor that interacts with this promoter when induced is also present in the cell in normal growth conditions. However, the kinetics of *recA*P1 induction with its extended lag period suggests that a factor that in itself is induced by DNA damage might be key in the induction of this promoter. It may be the case that in an uninduced state, its factor of choice may not be present and instead weak expression is achieved through association with another sigma factor. Determination of the proteins involved in the recognition of either promoter is also, therefore, a vital part in their characterisation and experiments to undertake this task have been mentioned in chapter 5.

7.4.3. Are the *recA* promoters, like the gene, conserved in mycobacteria?

It has been shown that the *recA* genes in a number of mycobacteria have conserved sequences (Blackwood *et al.*, 2000). However, no study to date has been to see whether the sequences upstream of this gene, and in particular the *recA*P1 and *recA*P2 regions, in these various mycobacteria are also conserved. Primer extension work could be undertaken in the species utilised in the Blackwood *et al.* (2000) study to identify regions of promoter activity in each upstream sequence. The resulting sequences could then be aligned and compared to see if the promoter regions were indeed conserved. This study also has implications for future study mentioned in the next section.

7.4. 4. Does the rate of DNA replication have an effect on the type of response mediated through the mycobacterial *recA* promoters?

Another interesting avenue of study that could be undertaken would be to test whether the differences seen in expression of both *recA* promoters in the two mycobacterial species could be explained by the differing growth rates of these organisms. A study of the *M. smegmatis* combined *recA*P1 and *recA*P2 activities using a different reporter in mitomycin C and oflaxacin induced *M. tuberculosis* revealed that together the two saprophyte promoters behaved like those of the pathogen's when inside the pathogen (F. Movahedzadeh and E. O. Davis, personal communication). Likewise, it was also shown that the pathogen's own *recA* promoters functioned like those of the saprophyte when induced in the saprophyte (F. Movahedzadeh and E. O. Davis, personal communication). Given this occurrence, there is a distinct possibility that the regulation of *recA* in mycobacteria could be like that seen in *M. tuberculosis* for slow growers and

M. smegmatis for the faster growing strains. The studies of the individual *M. smegmatis* *recA* promoters would have to be repeated using constructs based on the pEJ414 vector so results of these experiments could be comparible with those from this study. In addition, if primer extension studies of other mycobacterial *recA* genes showed conservation of *recA*P1 and *recA*P2 sequence, then fast growing mycobacteria could be transformed with the *M. tuberculosis recA* promoter constructs and assayed to see if the activities of the promoters were then like those seen in *M. smegmatis.*

REFERENCES SECTION

Adhya, S., and Gottesman, M. (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell* **29**: 939-944.

Aldea, M., Garrido, T., Pla, J., and Vicente, M. (1990) Division genes in *Escherichia coli* are expressed coordinately to cell septum requirements by gearbox promoters. *Embo J* **9**: 3787-3794.

Aldovini, A., Husson, R.N., and Young, R.A. (1993) The *uraA* locus and homologous recombination in *Mycobacterium bovis* BCG. *J Bacteriol* **175**: 7282-7289.

Allakhverdiev, S.I., Nishiyama, Y., Suzuki, I., Tasaka, Y., and Murata, N. (1999) Genetic engineering of the unsaturation of fatty acids in membrane lipids alters the tolerance of *Synechocystis* to salt stress. *Proc Natl Acad Sci U S A* **96**: 5862-5867.

Anderson, D.G., and Kowalczykowski, S.C. (1998) Reconstitution of an SOS response pathway: derepression of transcription in response to DNA breaks. *Cell* **95**: 975-979.

Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H., and Ikehara, M. (1983) Essential structure of *E. coli* promoter: effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Res* **11**: 5855-5864.

Azad, A.K., Sirakova, T.D., Rogers, L.M., and Kolattukudy, P.E. (1996) Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG produces a mutant that lacks mycosides. *Proc Natl Acad Sci U S A* **93**: 4787-4792.

Balasubramanian, V., Pavelka, M.S., Jr., Bardarov, S.S., Martin, J., Weisbrod, T.R., McAdam, R.A., Bloom, B.R., and Jacobs, W.R., Jr. (1996) Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *J Bacteriol* **178**: 273-279.

Baliko, G., and Venetianer, P. (1993) An *Escherichia coli* gene in search of a function: phenotypic effects of the gene recently identified as *murI*. *J Bacteriol* **175**: 6571-6577.

Ballesteros, M., Kusano, S., Ishihama, A., and Vicente, M. (1998) The *ftsQ1p* gearbox promoter of *Escherichia coli* is a major σ^S-dependent promoter in the *ddlB-ftsA* region. *Mol Microbiol* **30**: 419-430.

Baquero, M.R., Bouzon, M., Varea, J., and Moreno, F. (1995) *sbmC*, a stationary-phase induced SOS *Escherichia coli* gene, whose product protects cells from the DNA replication inhibitor microcin B17. *Mol Microbiol* **18**: 301-311.

Bashyam, M.D., Kaushal, D., Dasgupta, S.K., and Tyagi, A.K. (1996) A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. *J Bacteriol* **178**: 4847-4853.

Bashyam, M.D., and Tyagi, A.K. (1998) Identification and analysis of "extended -10" promoters from mycobacteria. *J Bacteriol* **180**: 2568-2573.

Bedwell, D.M., and Nomura, M. (1986) Feedback regulation of RNA polymerase subunit synthesis after the conditional overproduction of RNA polymerase in *Escherichia coli*. *Mol Gen Genet* **204**: 17-23.

Bekker, L.G., Freeman, S., Murray, P.J., Ryffel, B., and Kaplan, G. (2001) TNF-α controls intracellular mycobacterial growth by both inducible nitric oxide synthase-dependent and inducible nitric oxide synthase-independent pathways. *J Immunol* **166**: 6728-6734.

Belyaeva, T., Griffiths, L., Minchin, S., Cole, J., and Busby, S. (1993) The *Escherichia coli cysG* promoter belongs to the 'extended -10' class of bacterial promoters. *Biochem J* **296 (Pt 3)**: 851-857.

Black, C.G., Fyfe, J.A., and Davies, J.K. (1998) Absence of an SOS-like system in *Neisseria gonorrhoeae*. *Gene* **208**: 61-66.

Blackwood, K.S., He, C., Gunton, J., Turenne, C.Y., Wolfe, J., and Kabani, A.M. (2000) Evaluation of *recA* sequences for identification of Mycobacterium species. *J Clin Microbiol* **38**: 2846-2852.

Bourn, W.R., and Babb, B. (1995) Computer assisted identification and classification of streptomycete promoters. *Nucleic Acids Res* **23**: 3696-3703.

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.

Bryk, R., Lima, C.D., Erdjument-Bromage, Tempst, P., and Nathan, C. (2002) Metabolic enzymes of mycobacteria linked to antioxidant defense by a thioredoxin-like protein. *Science* **295**: 1073-1077.

Burgess, R.R., and Anthony, L. (2001) How sigma docks to RNA polymerase and what sigma does. *Curr Opin Microbiol* **4**: 126-131.

Burr, T., Mitchell, J., Kolb, A., Minchin, S., and Busby, S. (2000) DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Res* **28**: 1864-1870.

Busby, S., and Ebright, R.H. (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**: 743-746.

Busby, S., West, D., Lawes, M., Webster, C., Ishihama, A., and Kolb, A. (1994) Transcription activation by the *Escherichia coli* cyclic AMP receptor protein. Receptors bound in tandem at promoters can interact synergistically. *J Mol Biol* **241**: 341-352.

Cao, M., Kobel, P.A., Morshedi, M.M., Wu, M.F., Paddon, C., and Helmann, J.D. (2002) Defining the *Bacillus subtilis* σ^W regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches. *J Mol Biol* **316**: 443-457.

Chan, J., Fan, X.D., Hunter, S.W., Brennan, P.J., and Bloom, B.R. (1991) Lipoarabinomannan, a possible virulence factor involved in persistence of *Mycobacterium tuberculosis* within macrophages. *Infect Immun* **59**: 1755-1761.

Chan, J., Xing, Y., Magliozzo, R.S., and Bloom, B.R. (1992) Killing of virulent *Mycobacterium tuberculosis* by reactive nitrogen intermediates produced by activated murine macrophages. *J Exp Med* **175**: 1111-1122.

Chan, K., Knaak, T., Satkamp, L., Humbert, O., Falkow, S., and Ramakrishnan, L. (2002) Complex pattern of *Mycobacterium marinum* gene expression during long-term granulomatous infection. *Proc Natl Acad Sci U S A* **99**: 3920-3925.

Chen, P., Ruiz, R.E., Li, Q., Silver, R.F., and Bishai, W.R. (2000) Construction and characterization of a *Mycobacterium tuberculosis* mutant lacking the alternate sigma factor gene, *sigF*. *Infect Immun* **68**: 5575-5580.

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.

Cheo, D.L., Bayles, K.W., and Yasbin, R.E. (1992) Molecular characterizations of regulatory elements controling expression of the *Bacillus subtilis recA*$^+$ gene. *Biochimie* **74**: 755-762.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Barrell, B.G., and et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., and Barrell, B.G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007-1011.

Collins, D.M., Kawakami, R.P., de Lisle, G.W., Pascopella, L., Bloom, B.R., and Jacobs, W.R., Jr. (1995) Mutation of the principal sigma factor causes loss of virulence in a strain of the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* **92**: 8036-8040.

Colston, M.J., and Davis, E.O. (1994) The ins and outs of protein splicing elements. *Mol Microbiol* **12**: 359-363.

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.

Cowing, D.W., Bardwell, J.C., Craig, E.A., Woolford, C., Hendrix, R.W., and Gross, C.A. (1985) Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proc Natl Acad Sci U S A* **82**: 2679-2683.

Das Gupta, S.K., Bashyam, M.D., and Tyagi, A.K. (1993) Cloning and assessment of mycobacterial promoters by using a plasmid shuttle vector. *J Bacteriol* **175**: 5186-5192.

Davis, E.O., Sedgwick, S.G., and Colston, M.J. (1991) Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product. *J Bacteriol* **173**: 5653-5662.

Davis, E.O., Thangaraj, H.S., Brooks, P.C., and Colston, M.J. (1994) Evidence of selection for protein introns in the *recA*s of pathogenic mycobacteria. *Embo J* **13**: 699-703.

Davis, E.O., Dullaghan, E.M., and Rand, L. (2002) Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J Bacteriol* **184**: 3287-3295.

DeMaio, J., Zhang, Y., Ko, C., Young, D.B., and Bishai, W.R. (1996) A stationary-phase stress-response sigma factor from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **93**: 2790-2794.

Deretic, V., Philipp, W., Dhandayuthapani, S., Mudd, M.H., Curcic, R., Garbe, T., Heym, B., Via, L.E., and Cole, S.T. (1995) *Mycobacterium tuberculosis* is a natural mutant with an inactivated oxidative-stress regulatory gene: implications for sensitivity to isoniazid. *Mol Microbiol* **17**: 889-900.

Dhandayuthapani, S., Mudd, M., and Deretic, V. (1997) Interactions of OxyR with the promoter region of the *oxyR* and *ahpC* genes from *Mycobacterium leprae* and *Mycobacterium tuberculosis*. *J Bacteriol* **179**: 2401-2409.

Ding, A.H., Nathan, C.F., and Stuehr, D.J. (1988) Release of reactive nitrogen intermediates and reactive oxygen intermediates from mouse peritoneal macrophages. Comparison of activating cytokines and evidence for independent production. *J Immunol* **141**: 2407-2412.

Dombroski, A.J., Walter, W.A., and Gross, C.A. (1993) Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev* **7**: 2446-2455.

Domenech, P., Barry, C.E., 3rd, and Cole, S.T. (2001) *Mycobacterium tuberculosis* in the post-genomic age. *Curr Opin Microbiol* **4**: 28-34.

Dong, X.R., Li, S.F., and DeMoss, J.A. (1992) Upstream sequence elements required for NarL-mediated activation of transcription from the *narGHJI* promoter of Escherichia coli. *J Biol Chem* **267**: 14122-14128.

Dosanjh, M.K., Singer, B., and Essigmann, J.M. (1991) Comparative mutagenesis of o^6-methylguanine and o^4-methylthymine in *Escherichia coli*. *Biochemistry* **30**: 7027-7033.

Doukhan, L., Predich, M., Nair, G., Dussurget, O., Mandic-Mulec, I., Cole, S.T., Smith, D.R., and Smith, I. (1995) Genomic organization of the mycobacterial sigma gene cluster. *Gene* **165**: 67-70.

Dri A. M. and Moreau P. L. (1993) Phosphate starvation and low temperature as well as ultraviolet irradiation transcriptionally induce the Escherichia coli LexA-controlled gene sfiA. *Mol Microbiol* **8**, 697-706.

Dukan, S., and Nystrom, T. (1999) Oxidative stress defense and deterioration of growth-arrested *Escherichia coli* cells. *J Biol Chem* **274**: 26027-26032.

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.

Eadie, J.S., Conrad, M., Toorchen, D., and Topal, M.D. (1984) Mechanism of mutagenesis by o^6-methylguanine. *Nature* **308**: 201-203.

Fabisiewicz, A., and Janion, C. (1992) Effect of heat shock on expression of proteins not involved in the heat-shock regulon. *Eur J Biochem* **209**: 549-553.

Fenton, M.S., and Gralla, J.D. (2001) Function of the bacterial TATAAT -10 element as single-stranded DNA during RNA polymerase isomerization. *Proc Natl Acad Sci U S A* **98**: 9020-9025.

Fernandes, N.D., Wu, Q.L., Kong, D., Puyang, X., Garg, S., and Husson, R.N. (1999) A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. *J Bacteriol* **181**: 4266-4274.

Fernandez De Henestrosa, A.R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J.J., Ohmori, H., and Woodgate, R. (2000) Identification of additional genes belonging to the LexA regulon in *Escherichia coli. Mol Microbiol* **35**: 1560-1572.

Flesch, I.E., and Kaufmann, S.H. (1990) Activation of tuberculostatic macrophage functions by gamma interferon, interleukin-4, and tumor necrosis factor. *Infect Immun* **58**: 2675-2677.

Fletcher, H.M., Morgan, R.M., and Macrina, F.L. (1997) Nucleotide sequence of the *Porphyromonas gingivalis* W83 *recA* homolog and construction of a *recA*-deficient mutant. *Infect Immun* **65**: 4592-4597.

Flynn, J.L., and Chan, J. (2001a) Tuberculosis: latency and reactivation. *Infect Immun* **69**: 4195-4201.

Flynn, J.L., and Chan, J. (2001b) Immunology of tuberculosis. *Annu Rev Immunol* **19**: 93-129.

Friedberg, E.C., Walker, G.C., and Siede, W. (1995) *DNA repair and mutagenesis.* Washington D.C: American Society for Microbiology press.

Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T., and Bottger, E.C. (1998) Investigation of mycobacterial *recA* function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination. *Mol Microbiol* **29**: 1203-1214.

Frischkorn, K., Springer, B., Bottger, E.C., Davis, E.O., Colston, M.J., and Sander, P. (2000) *In vivo* splicing and functional characterization of *Mycobacterium leprae* RecA. *J Bacteriol* **182**: 3590-3592.

Fukuda, R., Ishihama, A., Saitoh, T., and Taketo, M. (1977) Comparative studies of RNA polymerase subunits from various bacteria. *Mol Gen Genet* **154**: 135-144.

Gaal, T., Ross, W., Estrem, S.T., Nguyen, L.H., Burgess, R.R., and Gourse, R.L. (2001) Promoter recognition and discrimination by $E\sigma^S$ RNA polymerase. *Mol Microbiol* **42**: 939-954.

Garbe, T.R., Hibler, N.S., and Deretic, V. (1996) Response of *Mycobacterium tuberculosis* to reactive oxygen and nitrogen intermediates. *Mol Med* **2**: 134-142.

Ghosh, P., Ishihama, A., and Chatterji, D. (2001) *Escherichia coli* RNA polymerase subunit ω and its N-terminal domain bind full-length beta' to facilitate incorporation into the $\alpha_2\beta$ subassembly. *Eur J Biochem* **268**: 4621-4627.

Gomez, J.E., Chen, J.M., and Bishai, W.R. (1997) Sigma factors of *Mycobacterium tuberculosis. Tuber Lung Dis* **78**: 175-183.

Gomez, J.E., and Bishai, W.R. (2000) *whmD* is an essential mycobacterial gene required for proper septation and cell division. *Proc Natl Acad Sci U S A* **97**: 8554-8559.

Gomez, M., Doukhan, L., Nair, G., and Smith, I. (1998) *sigA* is an essential gene in *Mycobacterium smegmatis. Mol Microbiol* **29**: 617-628.

Gomez, M., and Smith, I. (2000) Determinants of mycobacterial gene expression. In *Molecular genetics of mycobacteria.* Hatfull, G.F. and Jacobs, W.R., Jr. (eds). Washington, D. C.: ASM press, pp. 111- 129.

Goodfellow, M., Lechevalier, M.P., Prauser, H., and Wayne, L.G. (1986) Section 16- The mycobacteria. In *Bergey's manual of systematic biology.* Vol. 2. Sneath, P.H.A., Mair, N.S., Sharpe, M.E. and Holt, J.G. (eds). Baltimore: Williams and Wilkins, pp. 1435-1457.

Goodman, H.J., and Woods, D.R. (1990) Molecular analysis of the *Bacteroides fragilis recA* gene. *Gene* **94**: 77-82.

Goodwin, T.W., and Mercer, E.I. (1990) *An introduction to plant biochemistry.* Oxford: Pergamon Press.

Gourse, R.L., Ross, W., and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37**: 687-695.

Graham, J.E., and Clark-Curtiss, J.E. (1999) Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc Natl Acad Sci U S A* **96**: 11554-11559.

Grange, J.M. (1996) The biology of the genus Mycobacterium. *Soc Appl Bacteriol Symp Ser* **25**: 1S-9S.

Groemping, Y., and Reinstein, J. (2001) Folding properties of the nucleotide exchange factor GrpE from *Thermus thermophilus*: GrpE is a thermosensor that mediates heat shock response. *J Mol Biol* **314**: 167-178.

Gross, C.A., Lonetto, M., and Losick, R. (1992) Bacterial Sigma Factors. In *Transcriptional Regulation*: Cold Spring Harbor Laboratory press, pp. 129-174.

Gu, M.B., Min, J., and LaRossa, R.A. (2000) Bacterial bioluminescent emission from recombinant *Escherichia coli* harboring a *recA::luxCDABE* fusion. *J Biochem Biophys Methods* **45**: 45-56.

Gudas, L.J., and Pardee, A.B. (1975) Model for regulation of *Escherichia coli* DNA repair functions. *Proc Natl Acad Sci U S A* **72**: 2330-2334.

Guo, Y., Lew, C.M., and Gralla, J.D. (2000) Promoter opening by σ^{54} and σ^{70} RNA polymerases: sigma factor-directed alterations in the mechanism and tightness of control. *Genes Dev* **14**: 2242-2255.

Haldenwang, W.G. (1995) The sigma factors of Bacillus subtilis. *Microbiol Rev* **59**: 1-30.

Hamoen, L.W., Haijema, B., Bijlsma, J.J., Venema, G., and Lovett, C.M. (2001) The *Bacillus subtilis* competence transcription factor, ComK, overrides LexA-imposed transcriptional inhibition without physically displacing LexA. *J Biol Chem* **276**: 42901-42907.

Harley, C.B., and Reynolds, R.P. (1987) Analysis of *E. coli promoter sequences. Nucleic Acids Res* **15**: 2343-2361.

Harshey, R.M., and Ramakrishnan, T. (1976) Purification and properties of DNA-dependent RNA polymerase from *Mycobacterium tuberculosis* H37RV. *Biochim Biophys Acta* **432**: 49-59.

Harth, G., Zamecnik, P.C., Tang, J.Y., Tabatadze, D., and Horwitz, M.A. (2000) Treatment of *Mycobacterium tuberculosis* with antisense oligonucleotides to glutamine synthetase mRNA inhibits glutamine synthetase activity, formation of the poly-L-glutamate/glutamine cell wall structure, and bacterial replication. *Proc Natl Acad Sci U S A* **97**: 418-423.

Hawley, D.K., and McClure, W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* **11**: 2237-2255.

Heldt, H.-W. (1999) *Plant biochemistry and molecular biology*: Oxford University Press.

Hemmingsen, S.M., Woolford, C., van der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W., and Ellis, R.J. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333**: 330-334.

Hengge-Aronis, R., Lange, R., Henneberg, N., and Fischer, D. (1993) Osmotic regulation of *rpoS*-dependent genes in *Escherichia coli. J Bacteriol* **175**: 259-265.

Hengge-Aronis, R. (1996) Back to log phase: σ^S as a global regulator in the osmotic control of gene expression in Escherichia coli. *Mol Microbiol* **21**: 887-893.

Hinds, J., Mahenthiralingam, E., Kempsell, K.E., Duncan, K., Stokes, R.W., Parish, T., and Stoker, N.G. (1999) Enhanced gene replacement in mycobacteria. *Microbiology* **145 (Pt 3)**: 519-527.

Hochschild, A., and Dove, S.L. (1998) Protein-protein contacts that activate and repress prokaryotic transcription. *Cell* **92**: 597-600.

Holmes, D.S., and Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114**: 193-197.

Hondalus, M.K., Bardarov, S., Russell, R., Chan, J., Jacobs, W.R., Jr., and Bloom, B.R. (2000) Attenuation of and protection induced by a leucine auxotroph of *Mycobacterium tuberculosis*. *Infect Immun* **68**: 2888-2898.

Honore, N., Bergh, S., Chanteau, S., Doucet-Populaire, F., Eiglmeier, K., Garnier, T., Georges, C., Launois, P., Limpaiboon, T., Newton, S., and et al. (1993) Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. *Mol Microbiol* **7**: 207-214.

Hu, Y., and Coates, A.R. (1999a) Transcription of two σ^{70} homologue genes, *sigA* and *sigB*, in stationary-phase *Mycobacterium tuberculosis*. *J Bacteriol* **181**: 469-476.

Hu, Y., and Coates, A.R. (1999b) Transcription of the stationary-phase-associated *hspX* gene of *Mycobacterium tuberculosis* is inversely related to synthesis of the 16-kilodalton protein. *J Bacteriol* **181**: 1380-1387.

Huang, X., and Helmann, J.D. (1998) Identification of target promoters for the *Bacillus subtilis* σ^{X} factor using a consensus-directed search. *J Mol Biol* **279**: 165-173.

Ishihama, A. (1981) Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv Biophys* **14**: 1-35.

Iype, L.E., Wood, E.A., Inman, R.B., and Cox, M.M. (1994) RuvA and RuvB proteins facilitate the bypass of heterologous DNA insertions during RecA protein-mediated DNA strand exchange. *J Biol Chem* **269**: 24967-24978.

Jackett, P.S., Aber, V.R., and Lowrie, D.B. (1978) Virulence of *Mycobacterium tuberculosis* and susceptibility to peroxidative killing systems. *J Gen Microbiol* **107**: 273-278.

Jenkins, D.E., Auger, E.A., and Matin, A. (1991) Role of RpoH, a heat shock regulator protein, in *Escherichia coli* carbon starvation protein synthesis and survival. *J Bacteriol* **173**: 1992-1996.

Jensen-Cain, D.M., and Quinn, F.D. (2001) Differential expression of *sigE* by *Mycobacterium tuberculosis* during intracellular growth. *Microb Pathog* **30**: 271-278.

Jones, G.S., Amirault, H.J., and Andersen, B.R. (1990) Killing of *Mycobacterium tuberculosis* by neutrophils: a nonoxidative process. *J Infect Dis* **162**: 700-704.

Joung, J.K., Koepp, D.M., and Hochschild, A. (1994) Synergistic activation of transcription by bacteriophage λ cI protein and *E. coli* cAMP receptor protein. *Science* **265**: 1863-1866.

Kalpana, G.V., Bloom, B.R., and Jacobs, W.R., Jr. (1991) Insertional mutagenesis and illegitimate recombination in mycobacteria. *Proc Natl Acad Sci U S A* **88**: 5433-5437.

Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M., and Stevens, T.H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**: 651-657.

Kaushal, D., Schroeder, B.G., Tyagi, S., Yoshimatsu, T., Scott, C., Ko, C., Carpenter, L., Mehrotra, J., Manabe, Y.C., Fleischmann, R.D., and Bishai, W.R. (2002) Reduced immunopathology and mortality despite tissue persistence in a *Mycobacterium*

tuberculosis mutant lacking alternative sigma factor, SigH. *Proc Natl Acad Sci U S A* **99**: 8330-8335.

Keers, R.Y., Rigden, B.G., and Young, F.H. (1953) *Pulmonary Tuberculosis.* Edinburgh: E. & S. Livingstone LTD.

Kim, I.G., and Oh, T.J. (2000) SOS induction of the *recA* gene by UV-, γ-irradiation and mitomycin C is mediated by polyamines in *Escherichia coli* K-12. *Toxicol Lett* **116**: 143-149.

Kim, S., Dallmann, H.G., McHenry, C.S., and Marians, K.J. (1996) Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell* **84**: 643-650.

Kirschner, P., Springer, B., Vogel, U., Meier, A., Wrede, A., Kiekenbeck, M., Bange, F.C., and Bottger, E.C. (1993) Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory. *J Clin Microbiol* **31**: 2882-2889.

Kis, M., Zsiros, O., Farkas, T., Wada, H., Nagy, F., and Gombos, Z. (1998) Light-induced expression of fatty acid desaturase genes. *Proc Natl Acad Sci U S A* **95**: 4209-4214.

Kong, T.H., Coates, A.R., Butcher, P.D., Hickman, C.J., and Shinnick, T.M. (1993) *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc Natl Acad Sci U S A* **90**: 2608-2612.

Kong, D., and Kunimoto, D.Y. (1995) Secretion of human interleukin 2 by recombinant Mycobacterium bovis BCG. *Infect Immun* **63**: 799-803.

Kowalczykowski, S.C., and Krupp, R.A. (1987) Effects of *Escherichia coli* SSB protein on the single-stranded DNA-dependent ATPase activity of *Escherichia coli* RecA

protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J Mol Biol* **193**: 97-113.

Kremer, L., Baulard, A., Estaquier, J., Content, J., Capron, A., and Locht, C. (1995) Analysis of the *Mycobacterium tuberculosis* 85A antigen promoter region. *J Bacteriol* **177**: 642-653.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A., and Hayward, R.S. (1993) The minus 35-recognition region of *Escherichia coli* σ^{70} is inessential for initiation of transcription at an "extended minus 10" promoter. *J Mol Biol* **232**: 406-418.

Kwon, O.J. (1997) The role of nitric oxide in the immune response of tuberculosis. *J Korean Med Sci* **12**: 481-487.

Ladel, C.H., Szalay, G., Riedel, D., and Kaufmann, S.H. (1997) Interleukin-12 secretion by *Mycobacterium tuberculosis*-infected macrophages. *Infect Immun* **65**: 1936-1938.

Landini, P., and Volkert, M.R. (2000) Regulatory responses of the adaptive response to alkylation damage: a simple regulon with complex regulatory features. *J Bacteriol* **182**: 6543-6549.

Lau, Y.L., Chan, G.C., Ha, S.Y., Hui, Y.F., and Yuen, K.Y. (1998) The role of phagocytic respiratory burst in host defense against *Mycobacterium tuberculosis*. *Clin Infect Dis* **26**: 226-227.

Lee, G., and Pero, J. (1981) Conserved nucleotide sequences in temporally controlled bacteriophage promoters. *J Mol Biol* **152**: 247-265.

Lehninger, A.L., Nelson, D.L., and Cox, M.M. (1997) *Principles of Biochemistry.* New York: Worth Publishers.

Levin, M.E., and Hatfull, G.F. (1993) *Mycobacterium smegmatis* RNA polymerase: DNA supercoiling, action of rifampicin and mechanism of rifampicin resistance. *Mol Microbiol* **8**: 277-285.

Lewis, L.K., 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.

Lim, H.M., Lee, H.J., Roy, S., and Adhya, S. (2001) A "master" in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc Natl Acad Sci U S A* **98**: 14849-14852.

Lin, J.J., and Sancar, A. (1992) Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. *J Biol Chem* **267**: 17688-17692.

Lin, L.L., and Little, J.W. (1989) Autodigestion and RecA-dependent cleavage of Ind$^-$ mutant LexA proteins. *J Mol Biol* **210**: 439-452.

Lindahl, T., Demple, B., and Robins, P. (1982) Suicide inactivation of the *E. coli* o^6-methylguanine-DNA methyltransferase. *Embo J* **1**: 1359-1363.

Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988) Regulation and expression of the adaptive response to alkylating agents. *Annu Rev Biochem* **57**: 133-157.

Lindsley, J.E., and Cox, M.M. (1990) On RecA protein-mediated homologous alignment of two DNA molecules. Three strands versus four strands. *J Biol Chem* **265**: 10164-10171.

Lisser, S., and Margalit, H. (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res* **21**: 1507-1516.

Little, J.W., Edmiston, S.H., Pacelli, L.Z., and Mount, D.W. (1980) Cleavage of the *Escherichia coli lexA* protein by the *recA* protease. *Proc Natl Acad Sci U S A* **77**: 3225-3229.

Little, J.W. (1983) The SOS regulatory system: control of its state by the level of RecA protease. *J Mol Biol* **167**: 791-808.

Liu, Y., and Fletcher, H.M. (2001) Environmental regulation of *recA* gene expression in *Porphyromonas gingivalis*. *Oral Microbiol Immunol* **16**: 136-143.

Lonetto, M., Gribskov, M., and Gross, C.A. (1992) The σ^{70} family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**: 3843-3849.

Lonetto, M.A., Brown, K.L., Rudd, K.E., and Buttner, M.J. (1994) Analysis of the Streptomyces coelicolor *sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91**: 7573-7577.

Lovett, C.M., Jr., Love, P.E., and Yasbin, R.E. (1989) Competence-specific induction of the *Bacillus subtilis* RecA protein analog: evidence for dual regulation of a recombination protein. *J Bacteriol* **171**: 2318-2322.

Mahenthiralingam, E., Marklund, B.I., Brooks, L.A., Smith, D.A., Bancroft, G.J., and Stokes, R.W. (1998) Site-directed mutagenesis of the 19-kilodalton lipoprotein antigen reveals No essential role for the protein in the growth and virulence of *Mycobacterium intracellulare*. *Infect Immun* **66**: 3626-3634.

Manabe, Y.C., and Bishai, W.R. (2000) Latent *Mycobacterium tuberculosis*-persistence, patience, and winning by waiting. *Nat Med* **6**: 1327-1329.

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., Voskuil, M.I., Schoolnik, G.K., and Smith, I. (2001) The *Mycobacterium tuberculosis* ECF sigma factor σ^E: role in global gene expression and survival in macrophages. *Mol Microbiol* **41**: 423-437.

Marschall, C., Labrousse, V., Kreimer, M., Weichart, D., Kolb, A., and Hengge-Aronis, R. (1998) Molecular analysis of the regulation of *csiD*, a carbon starvation-inducible gene in *Escherichia coli* that is exclusively dependent on σ^S and requires activation by cAMP-CRP. *J Mol Biol* **276**: 339-353.

McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem* **54**: 171-204.

McFadden, J. (1996) Recombination in mycobacteria. *Mol Microbiol* **21**: 205-211.

McKane, M., and Gussin, G.N. (2000) Changes in the 17 bp spacer in the P(R) promoter of bacteriophage λ affect steps in open complex formation that precede DNA strand separation. *J Mol Biol* **299**: 337-349.

McKinney, J.D., Honer zu Bentrup, K., Munoz-Elias, E.J., Miczak, A., Chen, B., Chan, W.T., Swenson, D., Sacchettini, J.C., Jacobs, W.R., Jr., and Russell, D.G. (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**: 735-738.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**: 599-614.

Merkel, T.J., Dahl, J.L., Ebright, R.H., and Kadner, R.J. (1995) Transcription activation at the *Escherichia coli uhpT* promoter by the catabolite gene activator protein. *J Bacteriol* **177**: 1712-1718.

Michele, T.M., Ko, C., and Bishai, W.R. (1999) Exposure to antibiotics induces expression of the *Mycobacterium tuberculosis sigF* gene: implications for chemotherapy against mycobacterial persistors. *Antimicrob Agents Chemother* **43**: 218-225.

Miller, J. (1972) *Experiments in molecular genetics.* Cold Sping Harbor, New York: Cold Spring Harbor Laboratories.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001a) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98**: 892-897.

Minakhin, L., Nechaev, S., Campbell, E.A., and Severinov, K. (2001b) Recombinant *Thermus aquaticus* RNA polymerase, a new tool for structure-based analysis of transcription. *J Bacteriol* **183**: 71-76.

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.

Mol, C.D., Hosfield, D.J., and Tainer, J.A. (2000) Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means. *Mutat Res* **460**: 211-229.

Moolenaar, G.F., van Rossum-Fikkert, S., van Kesteren, M., and Goosen, N. (2002) Cho, a second endonuclease involved in *Escherichia coli* nucleotide excision repair. *Proc Natl Acad Sci U S A* **99**: 1467-1472.

Moran Jr, C.P., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshin, A.L., Pero, J., and Losick, R. (1982) Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* **186**: 339-346.

Morimoto, R.I. (1993) Cells in stress: transcriptional activation of heat shock genes. *Science* **259**: 1409-1410.

Movahedzadeh, F., Colston, M.J., and Davis, E.O. (1997a) 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.

Movahedzadeh, F., Colston, M.J., and Davis, E.O. (1997b) Characterization of *Mycobacterium tuberculosis* LexA: recognition of a Cheo (Bacillus-type SOS) box. *Microbiology* **143 (Pt 3)**: 929-936.

Muffler, A., Barth, M., Marschall, C., and Hengge-Aronis, R. (1997) Heat shock regulation of σ^S turnover: a role for DnaK and relationship between stress responses mediated by σ^S and σ^{32} in *Escherichia coli*. *J Bacteriol* **179**: 445-452.

Mulder, M.A., Zappe, H., and Steyn, L.M. (1997) Mycobacterial promoters. *Tuber Lung Dis* **78**: 211-223.

Mulder, N.J., Powles, R.E., Zappe, H., and Steyn, L.M. (1999) The *Mycobacterium tuberculosis mysB* gene product is a functional equivalent of the *Escherichia coli sigma factor*, KatF. *Gene* **240**: 361-370.

Murata, N., and Wada, H. (1995) Acyl-lipid desaturases and their importance in the tolerance and acclimatization to cold of cyanobacteria. *Biochem J* **308 (Pt 1)**: 1-8.

Nakabeppu, Y., and Sekiguchi, M. (1986) Regulatory mechanisms for induction of synthesis of repair enzymes in response to alkylating agents: ada protein acts as a transcriptional regulator. *Proc Natl Acad Sci U S A* **83**: 6297-6301.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R.H. (2000) Structural organization of the RNA polymerase-promoter open complex. *Cell* **101**: 601-611.

Newlands, J.T., Josaitis, C.A., Ross, W., and Gourse, R.L. (1992) Both *fis*-dependent and factor-independent upstream activation of the *rrnB* P1 promoter are face of the helix dependent. *Nucleic Acids Res* **20**: 719-726.

Nunoshiba, T., deRojas-Walker, T., Wishnok, J.S., Tannenbaum, S.R., and Demple, B. (1993) Activation by nitric oxide of an oxidative-stress response that defends *Escherichia coli* against activated macrophages. *Proc Natl Acad Sci U S A* **90**: 9993-9997.

Papavinasasundaram, K.G., Movahedzadeh, F., Keer, J.T., Stoker, N.G., Colston, M.J., and Davis, E.O. (1997) Mycobacterial *recA* is cotranscribed with a potential regulatory gene called *recX*. *Mol Microbiol* **24**: 141-153.

Papavinasasundaram, K.G., Colston, M.J., and Davis, E.O. (1998) Construction and complementation of a *recA* deletion mutant of *Mycobacterium smegmatis* reveals that the intein in *Mycobacterium tuberculosis recA* does not affect RecA function. *Mol Microbiol* **30**: 525-534.

Papavinasasundaram, K. G., Anderson, C., Brooks, P. C., Thomas, N. A., Movahedzadeh, F., Jenner, P. J., Colston, M. J. and Davis, E. O. (2001) Slow induction of RecA by DNA damage in *Mycobacetrium tuberculosis*. *Microbiology* **147**, 3271-3279.

Parish, T., and Stoker, N.G. (1997) Development and use of a conditional antisense mutagenesis system in mycobacteria. *FEMS Microbiol Lett* **154**: 151-157.

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 (Pt 8)**: 1969-1975.

Parish, T., Lewis, J., and Stoker, N.G. (2001) Use of the mycobacteriophage L5 excisionase in *Mycobacterium tuberculosis* to demonstrate gene essentiality. *Tuberculosis (Edinb)* **81**: 359-364.

Pelicic, V., Jackson, M., Reyrat, J.M., Jacobs, W.R., Jr., Gicquel, B., and Guilhot, C. (1997) Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis. Proc Natl Acad Sci U S A* **94**: 10955-10960.

Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J., and Belfort, M. (1994) Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucleic Acids Res* **22**: 1125-1127.

Petit, M.A., Dervyn, E., Rose, M., Entian, K.D., McGovern, S., Ehrlich, S.D., and Bruand, C. (1998) PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol Microbiol* **29**: 261-273.

Philipp, W.J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., Bergh, S., Bloom, B.R., Jacobs, W.R., Jr., and Cole, S.T. (1996) An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae. Proc Natl Acad Sci U S A* **93**: 3132-3137.

Predich, M., Doukhan, L., Nair, G., and Smith, I. (1995) Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis. Mol Microbiol* **15**: 355-366.

Pribnow, D. (1975a) Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. *J Mol Biol* **99**: 419-443.

Pribnow, D. (1975b) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A* **72**: 784-788.

Qiu, J., and Helmann, J.D. (2001) The -10 region is a key promoter specificity determinant for the *Bacillus subtilis* extracytoplasmic-function sigma factors σ^X and σ^W. *J Bacteriol* **183**: 1921-1927.

Quinones, A., Juterbock, W.R., and Messer, W. (1991) Expression of the *dnaA* gene of Escherichia coli is inducible by DNA damage. *Mol Gen Genet* **227**: 9-16.

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

Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W.R., Jr., and Husson, R.N. (2001) The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. *J Bacteriol* **183**: 6119-6125.

Ramesar, R.S., Abratt, V., Woods, D.R., and Rawlings, D.E. (1989) Nucleotide sequence and expression of a cloned *Thiobacillus ferrooxidans recA* gene in *Escherichia coli*. *Gene* **78**: 1-8.

Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., Jr., and Gourse, R.L. (1994) Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *J Mol Biol* **235**: 1421-1435.

Rastogi, N., Legrand, E., and Sola, C. (2001) The mycobacteria: an introduction to nomenclature and pathogenesis. *Rev Sci Tech* **20**: 21-54.

Reznikoff, W.S. (1993) The Tn*5* transposon. *Annu Rev Microbiol* **47**: 945-963.

Riera, J., Fernandez de Henestrosa, A.R., Garriga, X., Tapias, A., and Barbe, J. (1994) Interspecies regulation of the *recA* gene of gram-negative bacteria lacking an *E. coli*-like SOS operator. *Mol Gen Genet* **245**: 523-527.

Rinke de Wit, T.F., Bekelie, S., Osland, A., Miko, T.L., Hermans, P.W., van Soolingen, D., Drijfhout, J.W., Schoningh, R., Janson, A.A., and Thole, J.E. (1992) Mycobacteria contain two *groEL* genes: the second *Mycobacterium leprae groEL* gene is arranged in an operon with *groES*. *Mol Microbiol* **6**: 1995-2007.

Roach, T.I., Barton, C.H., Chatterjee, D., and Blackwell, J.M. (1993) Macrophage activation: lipoarabinomannan from avirulent and virulent strains of *Mycobacterium tuberculosis* differentially induces the early genes *c-fos*, KC, JE, and tumor necrosis factor-α. *J Immunol* **150**: 1886-1896.

Roberts, C.W., and Roberts, J.W. (1996) Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. *Cell* **86**: 495-501.

Rockett, K.A., Brookes, R., Udalova, I., Vidal, V., Hill, A.V., and Kwiatkowski, D. (1998) 1,25-Dihydroxyvitamin D3 induces nitric oxide synthase and suppresses growth of *Mycobacterium tuberculosis* in a human macrophage-like cell line. *Infect Immun* **66**: 5314-5321.

Rosenberg, M., and Court, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu Rev Genet* **13**: 319-353.

Russell, D.G. (2001) *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat Rev Mol Cell Biol* **2**: 569-577.

Sambrook, J., Fritsh, E.F., and Maniatis, T. (1989) *Molecular cloning : A laboratory manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Samson, L., and Cairns, J. (1977) A new pathway for DNA repair in *Escherichia coli.* *Nature* **267**: 281-283.

Sancar, G.B., Sancar, A., Little, J.W., and Rupp, W.D. (1982) The *uvrB* gene of *Escherichia coli* has both *lexA*-repressed and *lexA*-independent promoters. *Cell* **28**: 523-530.

Sander, P., Meier, A., and Bottger, E.C. (1995) *rpsL*$^+$: a dominant selectable marker for gene replacement in mycobacteria. *Mol Microbiol* **16**: 991-1000.

Sander, P., Papavinasasundaram, K.G., Dick, T., Stavropoulos, E., Ellrott, K., Springer, B., Colston, M.J., and Bottger, E.C. (2001) *Mycobacterium bovis* BCG *recA* deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model. *Infect Immun* **69**: 3562-3568.

Sano, K., Haneda, K., Tamura, G., and Shirato, K. (1999) Ovalbumin (OVA) and *Mycobacterium tuberculosis* bacilli cooperatively polarize anti-OVA T-helper (Th) cells toward a Th1-dominant phenotype and ameliorate murine tracheal eosinophilia. *Am J Respir Cell Mol Biol* **20**: 1260-1267.

Santos, J.M., Freire, P., Vicente, M., and Arraiano, C.M. (1999) The stationary-phase morphogene *bolA* from *Escherichia coli* is induced by stress during early stages of growth. *Mol Microbiol* **32**: 789-798.

Saves, I., Laneelle, M.A., Daffe, M., and Masson, J.M. (2000) Inteins invading mycobacterial RecA proteins. *FEBS Lett* **480**: 221-225.

Schaller, H., Gray, C., and Herrmann, K. (1975) Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc Natl Acad Sci U S A* **72**: 737-741.

Scott-Wilson, H.W. (1962) *Aids to Bacteriology.* London: Bailliere, Tindall and Cox.

Sedgwick, B. (1983) Molecular cloning of a gene which regulates the adaptive response to alkylating agents in *Escherichia coli. Mol Gen Genet* **191**: 466-472.

Shan, Q., Bork, J.M., Webb, B.L., Inman, R.B., and Cox, M.M. (1997) RecA protein filaments: end-dependent dissociation from ssDNA and stabilization by RecO and RecR proteins. *J Mol Biol* **265**: 519-540.

Shiba, T., Iwasaki, H., Nakata, A., and Shinagawa, H. (1991) SOS-inducible DNA repair proteins, RuvA and RuvB, of *Escherichia coli*: functional interactions between RuvA and RuvB for ATP hydrolysis and renaturation of the cruciform structure in supercoiled DNA. *Proc Natl Acad Sci U S A* **88**: 8445-8449.

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.

Spek, E.J., Wright, T.L., Stitt, M.S., Taghizadeh, N.R., Tannenbaum, S.R., Marinus, M.G., and Engelward, B.P. (2001) Recombinational repair is critical for survival of *Escherichia coli* exposed to nitric oxide. *J Bacteriol* **183**: 131-138.

Springer, B., Master, S., Sander, P., Zahrt, T., McFalone, M., Song, J., Papavinasasundaram, K.G., Colston, M.J., Boettger, E., and Deretic, V. (2001) Silencing of oxidative stress response in *Mycobacterium tuberculosis*: expression patterns of *ahpC* in virulent and avirulent strains and effect of *ahpC* inactivation. *Infect Immun* **69**: 5967-5973.

Srivastava, R., Cha, H.J., Peterson, M.S., and Bentley, W.E. (2000) Antisense downregulation of σ^{32} as a transient metabolic controller in *Escherichia coli*: effects on yield of active organophosphorus hydrolase. *Appl Environ Microbiol* **66**: 4366-4371.

Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F., and *et al.* (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.

Straus, D., Walter, W., and Gross, C.A. (1990) DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of σ^{32}. *Genes Dev* **4**: 2202-2209.

Studholme, D.J., and Buck, M. (2000) Novel roles of σ^{N} in small genomes. *Microbiology* **146 (Pt 1)**: 4-5.

Studholme, D.J., Buck, M., and Nixon, T. (2000) Identification of potential σ^{N}-dependent promoters in bacterial genomes. *Microbiology* **146 Pt 12**: 3021-3023.

Suh, W.C., Burkholder, W.F., Lu, C.Z., Zhao, X., Gottesman, M.E., and Gross, C.A. (1998) Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proc Natl Acad Sci U S A* **95**: 15223-15228.

Sukchawalit, R., and Mongkolsuk, S. (2001) *Xanthomonas oryzae pv.* oryzae *recA* is transcribed and regulated from multiple promoters. *FEMS Microbiol Lett* **197**: 35-40.

Taddei, F., Matic, I., and Radman, M. (1995) cAMP-dependent SOS induction and mutagenesis in resting bacterial populations. *Proc Natl Acad Sci U S A* **92**: 11736-11740.

Tagami, H., and Aiba, H. (1999) An inactive open complex mediated by an UP element at *Escherichia coli* promoters. *Proc Natl Acad Sci U S A* **96**: 7202-7207.

Teo, I., Sedgwick, B., Kilpatrick, M.W., McCarthy, T.V., and Lindahl, T. (1986) The intracellular signal for induction of resistance to alkylating agents in *E. coli. Cell* **45**: 315-324.

Tomasz, M., and Palom, Y. (1997) The mitomycin bioreductive antitumor agents: cross-linking and alkylation of DNA as the molecular basis of their activity. *Pharmacol Ther* **76**: 73-87.

Umezu, K., and Kolodner, R.D. (1994) Protein interactions in genetic recombination in *Escherichia coli.* Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. *J Biol Chem* **269**: 30005-30013.

van den Berg, E.A., Geerse, R.H., Pannekoek, H., and van de Putte, P. (1983) *In vivo* transcription of the *E. coli uvrB* gene: both promoters are inducible by UV. *Nucleic Acids Res* **11**: 4355-4363.

Van Houten, B., Eisen, J.A., and Hanawalt, P.C. (2002) A cut above: discovery of an alternative excision repair pathway in bacteria. *Proc Natl Acad Sci U S A* **99**: 2581-2583.

Volkert, M.R., and Landini, P. (2001) Transcriptional responses to DNA damage. *Curr Opin Microbiol* **4**: 178-185.

Voskuil, M.I., Voepel, K., and Chambliss, G.H. (1995) The -16 region, a vital sequence for the utilization of a promoter in *Bacillus subtilis* and *Escherichia coli. Mol Microbiol* **17**: 271-279.

Walker, G.C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol Rev* **48**: 60-93.

Walker, L., and Lowrie, D.B. (1981) Killing of *Mycobacterium microti* by immunologically activated macrophages. *Nature* **293**: 69-71.

Warne, S.E., and deHaseth, P.L. (1993) Promoter recognition by *Escherichia coli* RNA polymerase. Effects of single pair deletions and insertions in the spacer separating the -10 and -35 regions are dependent on spacer DNA sequence. *Biochemistry* **32**: 6134-6140.

Wayne, L.G., and Lin, K.Y. (1982) Glyoxylate metabolism and adaptation of *Mycobacterium tuberculosis* to survival under anaerobic conditions. *Infect Immun* **37**: 1042-1049.

Wayne, L.G., and Hayes, L.G. (1996) An in vitro model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. *Infect Immun* **64**: 2062-2069.

Webb, B.L., Cox, M.M., and Inman, R.B. (1997) Recombinational DNA repair: the RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. *Cell* **91**: 347-356.

Weisemann, J.M., and Weinstock, G.M. (1991) The promoter of the *recA* gene of *Escherichia coli*. *Biochimie* **73**: 457-470.

Wertman, K.F., and Mount, D.W. (1985) Nucleotide sequence binding specificity of the LexA repressor of *Escherichia coli* K-12. *J Bacteriol* **163**: 376-384.

West, S.C., Cassuto, E., and Howard-Flanders, P. (1981) Mechanism of *E. coli* RecA protein directed strand exchanges in post-replication repair of DNA. *Nature* **294**: 659-662.

West, S.C. (1994) Processing of Holliday junctions by RuvABC--an overview. *Ann N Y Acad Sci* **726**: 156-163; discussion 163-154.

Wiggs, J.L., Bush, J.W., and Chamberlin, M.J. (1979) Utilization of promoter and terminator sites on bacteriophage T7 DNA by RNA polymerases from a variety of bacterial orders. *Cell* **16**: 97-109.

Wilson, T., de Lisle, G.W., Marcinkeviciene, J.A., Blanchard, J.S., and Collins, D.M. (1998) Antisense RNA to *ahpC*, an oxidative stress defence gene involved in isoniazid resistance, indicates that AhpC of *Mycobacterium bovis* has virulence properties. *Microbiology* **144 (Pt 10)**: 2687-2695.

Wilson, T.M., de Lisle, G.W., and Collins, D.M. (1995) Effect of *inhA* and *katG* on isoniazid resistance and virulence of *Mycobacterium bovis*. *Mol Microbiol* **15**: 1009-1015.

Wilson, T.M., and Collins, D.M. (1996) *ahpC*, a gene involved in isoniazid resistance of the *Mycobacterium tuberculosis* complex. *Mol Microbiol* **19**: 1025-1034.

Winterling, K.W., Levine, A.S., Yasbin, R.E., and Woodgate, R. (1997) Characterization of DinR, the *Bacillus subtilis* SOS repressor. *J Bacteriol* **179**: 1698-1703.

Winterling, K.W., Chafin, D., Hayes, J.J., Sun, J., Levine, A.S., Yasbin, R.E., and Woodgate, R. (1998) The *Bacillus subtilis* DinR binding site: redefinition of the consensus sequence. *J Bacteriol* **180**: 2201-2211.

Wojciechowski, M.F., Peterson, K.R., and Love, P.E. (1991) Regulation of the SOS response in *Bacillus subtilis*: evidence for a LexA repressor homolog. *J Bacteriol* **173**: 6489-6498.

Wu, Q.L., Kong, D., Lam, K., and Husson, R.N. (1997) A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol* **179**: 2922-2929.

Young, B.A., Anthony, L.C., Gruber, T.M., Arthur, T.M., Heyduk, E., Lu, C.Z., Sharp, M.M., Heyduk, T., Burgess, R.R., and Gross, C.A. (2001) A coiled-coil from the RNA polymerase β' subunit allosterically induces selective nontemplate strand binding by σ^{70}. *Cell* **105**: 935-944.

Yuzhakov, A., Turner, J., and O'Donnell, M. (1996) Replisome assembly reveals the basis for asymmetric function in leading and lagging strand replication. *Cell* **86**: 877-886.

Zahrt, T.C., and Deretic, V. (2000) An essential two-component signal transduction system in *Mycobacterium tuberculosis*. *J Bacteriol* **182**: 3832-3838.

APPENDIX I-GROWTH MEDIA

L- Broth

For 1 litre:

10g Bacto Trytone,

5g Yeast extract,

1g NaCl

Modified Dubos medium

For 1 litre:

1 g KH_2PO_4,

6.25 g $Na_2HPO_4 12H_2O$,

1.25 g Na_3-citrate,

0.6 g $MgSO_4 7H_2O$,

2 g Aspargine,

2 g Casamino acids,

5 ml 10% tween 80,

20 ml 10% glycerol.

Adjust to pH 7.2.

Autoclave and then mix with 40 ml Dubos Medium Albumin (Difco)

APPENDIX II- PRIMER LISTS

This following section is comprised of the primers not already mentioned in the materials and methods section that were used in this study.

Abbreviations used: SDM- site directed mutagenesis, PCR- polymerase chain reaction

Primer name	Primer sequence	Function
***recA*P2T-10G forward**	CTAGACGCGGCGTGTCACACTTGAATCGAACG TTCGGCTACGGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2A-12C forward**	CTAGACGCGGCGTGTCACACTTGAATCGAACG TTCGGCTCCTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2C-14A forward**	CTAGACGCGGCGTGTCACACTTGAATCGAACG TTCGGATACTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2G-16T forward**	CTAGACGCGGCGTGTCACACTTGAATCGAACG TTCTGCTACTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2T-10G reverse**	AGCTTCCGAATGATCACCACCGTAGCCGAACG CCCGTTCGATCCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2A-12C reverse**	AGCTTCCGAATGATCACCACAGGAGCCGAACG CCCGTTCGATTCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2C-14A reverse**	AGCTTCCGAATGATCACCACAGTATCCGAACG CCCGTTCGATTCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2G-16T reverse**	AGCTTCCGAATGATCACCACAGTAGCAGAACG CCCGTTCGATTCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2A-32G reverse**	AGCTTCCGAATGATCACCACAGTAGCCGAACG CCCGTTCGATCCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2C-36A reverse**	AGCTTCCGAATGATCACCACAGTAGCCGAACG CCCGTTCGATTCAATTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2G-15T reverse**	AGCTTCCGAATGATCACCACAGTAGACGAACG CCCGTTCGATTCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2C-17A reverse**	AGCTTCCGAATGATCACCACAGTAGCCTAACG CCCGTTCGATTCAAGTGTGACACGAAGAGT	*recA*P2 direct mutagenesis
***recA*P2A-32G forward**	CTAGACGCGGCGTGTCACACTTGGATCGAAC GGGCGTTCGGCTACTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2C-36A forward**	CTAGACGCGGCGTGTCACAATTGAATCGAAC GGGCGTTCGGCTACTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2G-15T forward**	CTAGACGCGGCGTGTCACACTTGAATCGAAC GGGCGTTCGTCTACTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis
***recA*P2C-17A forward**	CTAGACGCGGCGTGTCACACTTGAATCGAAC GGGAGGTAGGCTACTGTGGTGATCATTCGGA	*recA*P2 direct mutagenesis

Primer name	Primer sequence	Function
***recA*P2elimin forward**	CAGGTGTTCGGCTCCTGTGGTGATCATTC	SDM to eliminate *recA*P2 actvity
***recA*P2elimin reverse**	GAATGATCACCACAGGAGCCGAACACTG	SDM to eliminate *recA*P2 activity
LacR	TTCCCAGTCACGACGTTGTAAAA	Primer used both for sequencing and amplification of *recA*P1 and *recA*P2 sequences.
PMINTF2	ACGAGGGGCATTCACACCAGATTG	Primer used both for sequencing and amplification of *recA*P1 and *recA*P2 sequences.
***recA*P1-35HOM forward**	CATTCGGAGCAGCCGACGACTCAGTGGCTGT CTCTAG	SDM of *recA*P1-35 region; triple change in region of homology
***recA*P1-35HOM reverse**	CTAGAGACAGCCACTGAGTCGTCGGCTGCTC CGAATG	SDM of *recA*P1-35 region; triple change in region of homology
***recA*P1-35POS forward**	GTGGTGATCATTCGGAGACTCGACTTGTCAGT GGCTGTCTC	SDM of *recA*P1-35 region; four base change in region of position
***recA*P1-35POS reverse**	GAGACAGCCACTGACAAGTAGTCTCCGAATGA TCACCAC	SDM of *recA*P1-35 region; four base change in region of position.
***recA*P1-10-3X forward**	CTTGTCAGTGGCTGTCGAGAGTGTCACGGCCA ACCG	SDM of *recA*P1-10 region; triple change
***recA*P1-10-3X reverse**	CGGTTGGCCGTGACACTCTCGACAGCCACTGA CAAG	SDM of *recA*P1-10 region; triple change
***recA*P1T-15G forward**	GTCAGTGGCTGTCGCTAGTGTCACGGCC	SDM of *recA*P1-10 region; *recA*P1T-15G change
***recA*P1T-15G reverse**	GGCCGTGACACTAGCGACAGCCACTGAC	SDM of *recA*P1-10 region; *recA*P1T-15G change
***recA*P1A-12C forward**	GTGGCTGTCTCTCGTGTCACGGCCAAC	SDM of *recA*P1-10 region; *recA*P1A-12C change
***recA*P1A-12C reverse**	GTTGGCCGTGACACGAGAGACAGCCAC	SDM of *recA*P1-10 region; *recA*P1A-12C change
***recA*P1T-10G forward**	GCTGTCTCTAGGGTCACGGCCAACCG	SDM of *recA*P1-10 region; *recA*P1T-10G change
***recA*P1T-10G reverse**	CGGTTGGCCGTGACCCTAGAGACAGC	SDM of *recA*P1-10 region; *recA*P1T-1OG change

Primer name	Primer sequence	Function
Random*recA*P2(A)	GGTCTAGAACGCGGCGTGtcacacttgaatcgaacgggctg ctactggtGATCATTCGGAAGCTTGG	"Spiked" oligonucleotide used for random mutagenesis of *recA*P2 with lower case section being randomised in ratio 97.3: 2.7 (cognate : non-cognate)
Random*recA*P2(B)	GGTCTAGAACGCGGCGTGtcacacttgaatcgaacgggctg ctactggtGATCATTCGGAAGCTTGG	"Spiked" oligonucleotide used for random mutagenesis of *recA*P2 with lower case section being randomised in ratio 98.5: 1.5 (cognate : non-cognate)
RA*rec*P2rev	CCAAGCTTCCGAATGAT	Primer to bind 3' end of the Random*recA*P2 oligonucleotide and make double stranded with the action of Klenow polymerase.
2205KKG3851CZ	CGCGAATTTTAACAAAATATTAACGC TTACAATTTAATTAACGCGCGTAATA CGACTCACTATAGGGCG	Site directed muatagenesis of pBluescript KS/ insertion of *PacI* site
2205KKG3851DA	CGCCCTATGTGAAGTCGTATTACGCG CGTTAATTAAATTGTAAGTTAATATTT TGTTAAAATTCGCG	Site directed muatagenesis of pBluescript KS/ insertion of *PacI* site
pKKG29fragup	CTAGACCAGGCTAGCGGTGTTGAGCAGATCGT CGGTGATCCGGACCAGCCGCGCACGCA	Oligonucleotide based on region upstream of *recA*P2
pKKG29fragdown	AGCTTGCGTGCGCGGCTGGTCCGGATCACCGA CGATCTGCTCAACACCGCTAGCCTGGT	Oligonucleotide based on region upstream of *recA*P2

APPENDIX III- MICRO-ARRAY DATA FROM THE *M. tuberculosis* ΔsigC STRAIN.

Table 1. Genes found to be expressed at a lower level in the *M. tuberculosis* Δ*sigC* mutant when in exponential growth.

				Virulence/ Detoxification
ORF	Gene	Fold	t-test	Description
Rv1968	*mce3C*	2.3	0.03	virulence/ detoxification
Rv3417c	*groEL1*	2.6	0.05	60 kD chaperonin 1
Rv1938	*ephB*	2.9	0.05	Probable epoxide hydrolase
Rv2428	*ahpC*	3.2	0.003	Alkyl hydroperoxide reductase
Rv3418c	*groES*	4.2	0.02	10 kD chaperonin

				Lipid metabolism
ORF	Gene	Fold	t-test	Description
Rv3516	*echA19*	2	0.007	Enoyl-CoA hydratase/ isomerase
Rv3505	*fadE27*	2.1	3.00E-04	Acyl- CoA dehydrogenase
Rv1136	*Rv1136*	2.1	0.04	Enoyl- CoA hydratse
Rv1935c	*echA13*	2.1	0.04	Enoyl-CoA hydratase/ isomerase
Rv3506	*fadD17*	2.2	0.01	Acyl- CoA synthase
Rv3546	*fadA5*	2.2	0.01	Acyl- CoA synthase
Rv3562	*fadE31*	2.3	0.002	Acyl- CoA dehydrogenase
Rv3515c	*fadD19*	2.3	0.003	Acyl- CoA synthase

				Cell wall and cell processes
ORF	Gene	Fold	t-test	Description
Rv1688	*Rv1688*	2.1	0.03	3-methylpurine DNA glycosylase
Rv3649	*Rv3649*	2.2	0.04	Hypothetical ATP- dependent helicase
Rv0631c	*recC*	2.2	0.002	Exodeoxyribonuclease V
Rv3162c	*Rv3162c*	2	0.03	Cell wall process
Rv1004c	*Rv1004c*	2	0.003	Cell wall process
Rv1456c	*Rv1456c*	2	0.006	Cell wall process
Rv3635	*Rv3635*	2.1	0.02	Transmembrane protein
Rv0194	*Rv0194*	2.1	0.01	Probable ABC transporter
Rv2025c	*Rv2025c*	2.2	0.02	efflux protein
Rv2320c	*rocE*	2.3	0.007	Arginine/ ornithine transporter
Rv2265	*Rv2265*	2.3	0.01	Putative Integral membrane protein

Rv1258c	*Rv1258c*	2.3	3.00E-04	efflux protein
Rv1999c	*Rv1999c*	2.3	0.02	Transport protein similar to E. coli YjeH
Rv3163c	*Rv3163c*	2.3	0.01	Cell wall process
Rv2270	*lppN*	2.3	0.03	lipoprotein with putative signal peptide
Rv0090	*Rv0090*	2.4	0.005	Cell wall process protein
Rv0849	*Rv0849*	2.4	0.01	membrane protein similar to efflux pump
Rv2920c	*amt*	2.5	0.02	Putative ammonium transporter
Rv0622	*Rv0622*	2.5	0.006	Cell wall process
Rv3450c	*Rv3450c*	2.5	0.03	Cell wall process
Rv3448	*Rv3448*	2.6	0.003	Cell wall process
Rv2317	*uspE*	2.7	0.01	Sugar transport protein
Rv2316	*uspA*	2.7	0.005	Sugar transport protein
Rv0621	*Rv0621*	3.1	0.01	Cell wall process

Insertion sequence/ Phage protein

ORF	Gene	Fold	t-test	Description
Rv2653c	*Rv2653c*	2.1	0.04	Phage protein
Rv2815c	*Rv2815c*	2.2	0.04	Transposase
Rv3467	*Rv3467*	2.2	0.03	Insertion sequence/ Phage protein
Rv2659c	*Rv2659c*	2.2	0.02	Integrase
Rv0094c	*Rv0094c*	2.2	0.01	Member of 13 EZ repeat family
Rv2105	*Rv2105*	2.3	0.03	Transposase
Rv2648	*Rv2648*	2.4	0.02	Transposase
Rv1763	*Rv1763*	2.4	0.04	Transposase
Rv2649	*Rv2649*	2.4	0.02	Transposase
Rv2814c	*Rv2814c*	2.4	0.006	Transposase
Rv1036c	*Rv1036c*	2.5	0.03	Transposase
Rv2355	*Rv2355*	2.5	0.005	Transposase
Rv2106	*Rv2106*	2.5	0.009	Transposase
Rv3185	*Rv3185*	2.5	0.01	Transposase
Rv1369c	*Rv1369c*	2.5	0.02	Transposase
Rv3475	*Rv3475*	2.5	0.01	Insertion sequence/ Phage protein
Rv2655c	*Rv2655c*	2.5	0.05	Insertion sequence/ Phage protein
Rv2354	*Rv2354*	2.5	0.007	Transposase
Rv2479c	*Rv2479c*	2.5	0.009	Transposase
Rv0796	*Rv0796*	2.6	0.008	Transposase
Rv1756c	*Rv1756c*	2.6	0.01	Transposase
Rv1757c	*Rv1757c*	2.6	0.04	Transposase
Rv1583c	*Rv1583c*	2.6	0.02	Phage protein
Rv2167c	*Rv2167c*	2.6	0.004	Transposase
Rv2279	*Rv2279*	2.6	0.03	Transposase
Rv0795	*Rv0795*	2.6	0.03	Transposase
Rv1764	*Rv1764*	2.7	0.01	Transposase
Rv3380c	*Rv3380c*	2.7	0.03	Insertion sequence/ Phage protein
Rv3187	*Rv3187*	2.7	0.04	Transposase
Rv1370c	*Rv1370c*	2.8	0.02	Transposase

Rv3326	*Rv3326*	2.8	0.01	Transposase
Rv3325	*Rv3325*	2.9	0.03	Transposase
Rv3184	*Rv3184*	3	0.04	Transposase

				Intermediary metabolism
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv1652	*argC*	2	0.008	N-acetyl-g- glutamyl-phosphate reductase
Rv2089c	*pepE*	2	0.03	Cytoplasmic peptidase
Rv0729	*xylB*	2.1	0.02	Intermediary metabolism
Rv1622c	*cydB*	2.1	3.00E-04	Cytochrome *d* ubiquinol oxidase subunit
Rv1050	*Rv1050*	2.1	0.004	Probable oxidoreductase
Rv0848	*cysM3*	2.2	0.02	Putative cysteine synthase
Rv2072c	*cobL*	2.2	0.03	Probable methyltransferase
Rv0373c	*Rv0373c*	2.3	0.008	Possible dehydrogenase
Rv0534c	*menA*	2.3	6.00E-04	octaprenyltransferase
Rv3174	*Rv3174*	2.3	0.02	Oxidoreductase
Rv0218	*Rv0218*	2.3	0.01	Probable sulfite oxidase
Rv0553	*menC*	2.3	0.008	o- succinylbenzoate- CoA synthase
Rv1726	*Rv1726*	2.5	0.02	Probable oxidoreductase
Rv0768	*aldA*	3	0.04	L-alanine dehydrogenase

				Unknown
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv0259c	*Rv0259c*	2	0.05	Unknown
Rv3358	*Rv3358*	2.1	0.02	Unknown
Rv3541c	*Rv3541c*	2.1	0.05	Unknown
Rv3903c	*Rv3903c*	2.1	0.02	Unknown
Rv2541	*Rv2541*	2.1	7.00E-04	Unknown
Rv2687c	*Rv2687c*	2.3	0.04	Unknown
Rv2063c	*Rv2063c*	2.3	0.003	Unknown
Rv2319c	*Rv2319c*	2.4	0.006	Unknown
Rv1134	*Rv1134*	2.6	0.02	Unknown
Rv3395c	*Rv3395c*	3	0.03	Unknown
Rv2489c	*Rv2489c*	3.3	0.05	Unknown

				Regulatory mechanism
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv1267c	*embR*	2	0.02	Regulator of *embAB* genes (AfsR/DndI/RedD fan
Rv2488c	*Rv2488c*	2.1	0.003	Transcription regulatory protein
Rv3764c	*Rv3764c*	2.2	0.01	Sensor histidine kinase
Rv0022c	*whiB5*	2.2	0.05	Whi-B like regulating protein
Rv0386	*Rv0386*	2.3	0.03	(LuxR/ UhpA family)

Rv3082c	*virS*	2.4	0.01	(AraC/ XylS family)
Rv3080c	*pknK*	2.5	0.05	Serine- threonine protein kinase
Rv0494	*Rv0494*	2.6	0.01	(PbsX/ Xre family)
Rv0600c	*Rv0600c*	2.8	0.02	Sensor histadine kinase
Rv0602c	*tcrA*	2.8	0.006	two-component response regulator
Rv0650	*Rv0650*	2.9	0.006	(ROK family)

				Conserved Hypothetical
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv1051c	*Rv1051c*	2	0.03	Conserved Hypothetical
Rv0091	*Rv0091*	2	0.002	Conserved Hypothetical
Rv2307c	*Rv2307c*	2	0.01	Conserved Hypothetical
Rv1259	*Rv1259*	2.1	0.008	Conserved Hypothetical
Rv0415	*Rv0415*	2.1	0.01	Conserved Hypothetical
Rv0266c	*Rv0266c*	2.1	0.006	Conserved Hypothetical
Rv2623	*Rv2623*	2.1	0.02	Conserved Hypothetical
Rv2807	*Rv2807*	2.1	0.04	Conserved Hypothetical
Rv1930c	*Rv1930c*	2.1	0.005	Conserved Hypothetical
Rv1429	*Rv1429*	2.1	0.007	Conserved Hypothetical
Rv0260c	*Rv0260c*	2.1	0.003	Conserved Hypothetical
Rv0574c	*Rv0574c*	2.2	0.01	Conserved Hypothetical
Rv2917	*Rv2917*	2.3	0.01	Conserved Hypothetical
Rv1972	*Rv1972*	2.3	0.05	Conserved Hypothetical
Rv2624c	*Rv2624c*	2.3	0.04	Conserved Hypothetical
Rv1455	*Rv1455*	2.4	0.004	Consered Hypothetical
Rv2897c	*Rv2897c*	2.4	0.04	Conserved Hypothetical
Rv0368c	*Rv0368c*	2.5	0.02	Conserved Hypothetical
Rv0493c	*Rv0493c*	2.5	0.02	Conserved Hypothetical
Rv2625c	*Rv2625c*	2.6	0.02	Conserved Hypothetical
Rv0991c	*Rv0991c*	2.7	0.05	Conserved Hypothetical
Rv3357	*Rv3357*	2.7	0.03	Conserved Hypothetical
Rv0648	*Rv0648*	2.7	0.03	Conserved Hypothetical
Rv0614	*Rv0614*	2.7	0.005	Conserved Hypothetical
Rv2811	*Rv2811*	2.9	0.02	Conserved Hypothetical
Rv2415c	*Rv2415c*	3.2	0.02	Conserved Hypothetical

Table 2. Genes found to be expressed at a higher level in the *M. tuberculosis* Δ*sigC* mutant when in exponential growth

				Virulence/ Detoxification
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv3648c	*cspA*	3	4.00E-04	Cold shock protein , transcriptional regulator
Rv3922c	*Rv3922c*	2.7	0.02	Possible haemolysin
Rv1908c	*katG*	2.4	0.04	Catalase peroxidase

				Lipid metabolism
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv3229c	*desA3*	12.1	0.005	Acyl-[ACP] desaturase
Rv0824c	*desA1*	9.1	5.00E-04	Acyl-[ACP] desaturase
Rv1094	*desA2*	5.1	4.00E-06	Acyl-[ACP] desaturase
Rv2246	*kasB*	3.7	0.002	B-ketoacyl-ACP synthase (meomycolate extension)
Rv2247	*accD6*	3.2	3.00E-04	Acetyl/propionyl CoA carboxylase b-subunit
Rv2244	*acpM*	3.1	0.007	Acyl carrier protein (meromycolate extension)
Rv2245	*kasA*	3.1	0.008	B-ketoacyl-ACP synthase (meomycolate extension)
Rv2935	*ppsE*	3	0.002	Phenolpthocerol synthesis
Rv2950c	*fadD29*	2.8	0.002	Acyl- CoA dehydrogenase
Rv3409c	*choD*	2.6	4.00E-04	Cholesterol oxidase
Rv2243	*fabD*	2.6	0.02	Malonyl CoA-[ACP] transacylase
Rv2948c	*fadD22*	2.5	0.001	Acyl- CoA dehydrogenase
Rv2947c	*pks15*	2.2	0.003	Polyketide synthase
Rv1317c	*alkA*	2.1	0.003	DNA-3-methladenine glycosidase II

				Information Pathways
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv1643	*rplT*	3.7	1.00E-04	50S ribosomal protein L20
Rv1298	*rpmE*	3.6	6.00E-05	50S ribosomal protein L31
Rv1297	*rho*	3.3	6.00E-04	Transcription termination factor rho
Rv0705	*rpsS*	3.1	1.00E-04	30S ribosomal protein S13
Rv1642	*rpmI*	2.7	5.00E-04	50S ribosomal protein L35
Rv2783c	*gpsI*	2.7	0.02	pppGpp synthase + polyribonucleotide phosphorylase
Rv0702	*rplD*	2.6	0.001	50S ribosomal protein L4
Rv0703	*rplW*	2.5	9.00E-04	50S ribosomal protein L23
Rv3211	*rhlE*	2.4	0.001	Probable ATP-dependent RNA helicase
Rv0704	*rplB*	2.4	0.002	50S ribosomal protein L2
Rv1299	*prfA*	2.3	2.00E-05	Peptide chain release factor 1
Rv0668	*rpoC*	2.3	0.001	B'- subunit of thew RNA polymerase

Rv1641	*infC*	2.3	4.00E-04	Initiation factor IF-3
Rv0641	*rplA*	2.2	4.00E-04	50S ribosomal protein L1
Rv0722	*rpmD*	2.2	0.005	50S ribosomal protein L30
Rv0706	*rplV*	2.1	7.00E-04	50S ribosomal protein L22
Rv0709	*rpmC*	2.1	2.00E-04	50S ribosomal protein L29

				Cell wall and cell process
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv2936	*drrA*	3.6	7.00E-04	Similar daunorubicin resistance ABC transporter
Rv2846	*efpA*	3.5	8.00E-04	Putative efflux protein
Rv3921c	*Rv3921c*	3.4	4.00E-05	Membrane protein
Rv1614	*lgt*	2.8	9.00E-05	Prolipoprotein diacylglyceryl transferase
Rv2938	*drrC*	2.7	0.002	Similar daunorubicin resistance protein
Rv3918c	*parB*	2.6	0.003	Possibly involved in chromosome partitioning
Rv0655	*Rv0655*	2.6	8.00E-04	Hypothetical ABC transporter
Rv3919c	*gid*	2.5	1.00E-05	Glucose inhibited division protein B
Rv3811	*csp*	2.3	0.003	Secreted protein
Rv2937	*drrB*	2.3	0.002	Similar daunorubicin resistance protein
Rv0291	*Rv0291*	2.3	0.01	Possible protease
Rv2905	*lppW*	2.2	0.007	Lipoprotein-B -lactamase precursor
Rv2942	*mmpL7*	2.2	0.002	Conserved large membrane protein
Rv2790c	*ltp1*	2.1	0.002	Lipid transfer protein
Rv0821c	*phoY2*	2.1	1.00E-04	Phosphate transport system regulator
Rv0050	*ponA1*	2.1	1.00E-04	Penicillin binding protein

				Intermediary metabolism
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv1177	*fdxC*	3.3	4.00E-05	Ferredoxin 4Fe-4S
Rv3854c	*ethA*	3.3	0.02	Intermediary metabolism
Rv2988c	*leuC*	3	6.00E-04	3-isopropylmalate dehydratase large subunit
Rv0694	*lldD1*	2.9	8.00E-06	L-lactate dehydogenase (cytochrome)
Rv2987c	*leuD*	2.8	0.006	3-isopropylmalate dehydratase small subunit
Rv1095	*phoH2*	2.7	4.00E-04	Pho-H like protein
Rv1613	*trpA*	2.7	0.001	Trytophan synthase a- chain
Rv3487c	*lipF*	2.5	0.004	Probabale esterase
Rv2782c	*pepR*	2.5	6.00E-05	Protease/ peptidase, M16 family (insulinase)
Rv1360	*Rv1360*	2.5	4.00E-04	Putative oxidoreductase
Rv3051c	*nrdE*	2.4	0.001	Ribonucleside diphosphate alpha chain
Rv3155	*nuoK*	2.4	4.00E-05	NADH dehydrogenase chain K
Rv1644	*tsnR*	2.4	1.00E-04	Putative 23S rRNA methyltransferase
Rv1872c	*lldD2*	2.3	0.001	L-lactate dehydogenase
Rv3377c	*Rv3377c*	2.3	0.02	Probable cyclase
Rv2334	*cysK*	2.2	9.00E-05	Cysteine synthase A
Rv1300	*hemK*	2.2	1.00E-04	Protoporphyrinogen oxidase
Rv1612	*trpB*	2.2	1.81E-04	Trytophan synthase b- chain

Rv1107c	*xseB*	2.2	8.00E-05	Exonuclease VII small subunit
Rv0511	*cysG*	2.1	1.00E-03	Uroporphyrin- III o- methyltransferase
Rv2391	*nirA*	2.1	0.002	Possible nitrite reductase/ sulphite reductase
Rv3052c	*nrdI*	2.1	3.00E-04	NrdI/YgaO/YmaA family
Rv3156	*nuoL*	2.1	4.00E-04	NADH dehydrogenase chain L
Rv693	*pqqE*	2.1	1.00E-04	Heme biosynthesis proetin PQQ synth protein
Rv3727	*Rv3727*	2	4.00E-05	Phytoene dehydrogenase precursor

Unknown function

ORF	Gene	Fold	t-test	Description
Rv1870c	*Rv1870c*	5.2	9.00E-05	Unknown
Rv2949c	*Rv2949c*	4.5	7.00E-05	Unknown
Rv3093c	*Rv3093c*	3.3	0.004	Unknown
Rv3212	*Rv3212*	3	0.003	Unknown
Rv2901c	*Rv2901c*	2.6	0.001	Unknown
Rv1632c	*Rv1632c*	2.4	0.002	Unknown
Rv1265	*Rv1265*	2.3	6.00E-04	Unknown
Rv3096	*Rv3096*	2.1	9.00E-04	Unknown

Regulatory mechanism

ORF	Gene	Fold	t-test	Description
Rv0823c	*Rv0823c*	4.7	6.00E-04	Transcriptional regulator (NifR3/ Smm1 family)
Rv2989	*Rv2989*	3.6	0.003	Transcriptional regulator (IclR family)
Rv3095	*Rv3095*	3	0.003	Putative transcription regulator
Rv3583c	*Rv3583c*	2.5	5.00E-04	Putative transcription regulator
Rv0196	*Rv0196*	2.3	0.005	Transcriptional regulator (TetR/ AcrR family)
Rv3295	*Rv3295*	2.3	5.00E-05	Transcription regulator (TetR/ AcrR family)
Rv1846c	*Rv1846c*	2.1	3.00E-04	Putative transcription regulator

Conserved Hypothetical

ORF	Gene	Fold	t-test	Description
Rv3920c	*Rv3920c*	2.9	4.00E-05	Conserved Hypothetical
Rv0239	*Rv0239*	2.8	7.00E-07	Conserved Hypothetical
Rv2327	*Rv2327*	2.7	2.00E-04	Conserved Hypothetical
Rv3869	*Rv3869*	2.7	1.00E-04	Conserved Hypothetical
Rv1697	*Rv1697*	2.5	0.005	Conserved Hypothetical
Rv1871c	*Rv1871c*	2.5	0.03	Conserved Hypothetical
Rv2166c	*Rv2166c*	2.5	0.01	Conserved Hypothetical
Rv3870	*Rv3870*	2.5	4.00E-04	Conserved Hypothetical
Rv0822c	*Rv0822c*	2.4	0.005	Conserved Hypothetical
Rv1505c	*Rv1505c*	2.4	0.002	Conserved Hypothetical
Rv2009	*Rv2009*	2.4	7.00E-04	Conserved Hypothetical

Appendix III-table 2. Increased expression genes in the mutant/ EXPONENTIAL PHASE

Rv2879c	*Rv2879c*	2.4	4.00E-04	Conserved Hypothetical
Rv0177	*Rv0177*	2.3	0.009	Conserved Hypothetical
Rv2585c	*Rv2585c*	2.3	8.00E-04	Conserved Hypothetical
Rv3486	*Rv3486*	2.3	0.004	Conserved Hypothetical
Rv3871	*Rv3871*	2.3	0.001	Conserved Hypothetical
Rv1631	*Rv1631*	2.2	0.001	Conserved Hypothetical
Rv1502	*Rv1502*	2.1	3.00E-04	Conserved Hypothetical
Rv1798	*Rv1798*	2.1	2.00E-05	Conserved Hypothetical

Table 3. Genes found to be expressed at a lower level in *M. tuberculosis* Δ*sigC* mutant when in stationary growth.

				Virulence/ Detoxification
ORF	**Gene**	**Fold**	**T-test**	**Description**
Rv2428	*ahpC*	8.6	6.00E-04	Alkyl Hydroperoxide reductase
Rv0251c	*hsp*	6.7	1.00E-03	Possible heat shock protein
Rv2429	*ahpD*	5.4	0.005	member of AhpC/ TSA family
Rv0440	*groEL2*	4.9	0.03	60 kD chaperonin 2
Rv3418c	*groES*	4.6	0.004	10 kD chaperonin
Rv0350	*dnaK*	4.3	0.002	70 kD heat shock protein + chromosome replication
Rv0352	*dnaJ*	3.3	0.05	Acts with GrpE to stimulate DnaK ATPase
Rv0351	*grpE*	2.9	0.01	Stimulates DnaK ATPase activity
Rv3846	*sodA*	2.8	3.00E-04	Superoxide dismutase
Rv3417c	*groEL1*	2.5	0.006	60 kD chaperonin 1

				Lipid metabolism
ORF	**Gene**	**Fold**	**T-test**	**Description**
Rv2930	*fadD26*	2.8	0.002	Acyl- CoA synthase
Rv2940c	*mas*	2.5	0.01	Mycocerosic acid synthase
Rv1344	*Rv1344*	2.5	0.007	Possible Aycl carrier protein
Rv2289	*cdh*	2.3	0.003	CDP-diacylglycerol phosphatidyl-hydrolase
Rv0468	*fadB2*	2.3	7.00E-04	3-hydroxyacyl- CoA dehydrogenase
Rv1618	*tesB1*	2.3	0.003	Thiosesterase II
Rv3774	*echA21*	2.2	0.05	enoyl- CoA hydratase/ isomerase superfamily
Rv0905	*echA6*	2.2	0.002	*eccH* enoyl- CoA hydratase/ isomerase superfamily
Rv0673	*echA4*	2.1	0.02	enoyl- CoA hydratase/ isomerase superfamily
Rv1483	*fabG1*	2.1	5.00E-04	*mabA* 3-oxoacyl- {ACP} reductase
Rv1175c	*fadH*	2.1	0.01	2,4 Dienoyl- CoA reductase
Rv1484	*inhA*	2.1	0.002	Enoyl-[ACP] reductase
Rv0045c	*Rv0045c*	2.1	5.00E-04	Dihydrolipoamide acetyltransferase
Rv1346	*fadE14*	2	0.007	Acyl- CoA dehydrogenase
Rv2928	*tesA*	2	0.002	Thioesterase

Rv0469	*umaA1*	2	0.02	Unknown mycolic acid methyltransferase

				Information Pathway
ORF	**Gene**	**Fold**	**T-test**	**Description**
Rv0009	*ppiA*	3.6	0.003	Peptidyl-prolyl *cis-trans* isomerase
Rv0685	*tuf*	3.4	0.05	Elongation factor EF-tu
Rv1221	*sigE*	3.3	0.004	ECF sigma factor
Rv2783c	*gpsI*	2.9	0.01	pppGpp synthase+ polyribonucleotide phosphorylase
Rv0058	*dnaB*	2.8	0.02	DNA helicase (contains intein)
Rv2582	*ppiB*	2.6	1.00E-04	Peptidyl-prolyl cis-trans isomerase
Rv1388	*mIHF*	2.4	0.04	Integration host factor
Rv3456c	*rplQ*	2.4	0.02	50S ribosomal protein L17
Rv0706	*rplV*	2.3	0.04	50S ribosomal protein L22
Rv0684	*fusA*	2.2	0.03	Elongation Factor G
Rv3443c	*rplM*	2.2	0.03	50S ribosomal protein L13
Rv0707	*rpsC*	2.2	0.03	30S ribosomal protein S3
Rv0708	*rplP*	2.1	0.02	50S ribosomal protein L16
Rv2069	*sigC*	2.1	0.01	ECF sigma factor
Rv3442c	*rpsI*	2	0.001	30S ribosomal protein S9

				Cell wall and cell processes
ORF	**Gene**	**Fold**	**T-test**	**Description**
Rv1343c	*Rv1343c*	5	0.007	Possible membrane protein
Rv3524	*Rv3524*	3.2	0.001	Similar to C-terminal end of pknD
Rv2936	*drrA*	2.8	0.03	Similar daunorubin resistance ABC -transporter
Rv1980c	*mpt64*	2.6	0.003	Secreted immunogeneic protein Mpb64/ Mpt64
Rv2450c	*rpfE*	2.6	0.005	Resuscitation Promoting Factor
Rv1132	*Rv1132*	2.4	0.008	Possible membrane protein
Rv0249c	*Rv0249c*	2.3	0.02	Possible anchor subunit for Rv0247c and Rv0248c
Rv0361	*Rv0361*	2.2	0.001	Possible transmembrane stretch
Rv0206c	*mmpL3*	2	0.03	Conserved large membrane protein
Rv1072	*Rv1072*	2	0.01	Transmembrane protein

Rv2462c	*tig*	2	0.005	Chaperone protein, similar to trigger factor

				Intermediary metabolism and respiration
ORF	**Gene**	**Fold**	**T-test**	**Description**
Rv1131	*gltA1*	6.6	0.001	Citrate syntahase 3
Rv0467	*icl*	4.2	7.00E-04	Isocitrate lyase
Rv0062	*celA*	3.3	8.00E-05	Cellulase/ endoglucanase
Rv3601c	*panD*	3.3	0.02	Aspartae 1- decarboxylase
Rv3602c	*panC*	3.1	0.01	panoate- b-alanine ligase
Rv3248c	*sahH*	3	0.002	Adenosylhomocysteinase
Rv1471	*trxB*	3	0.005	Thioredoxin reductase
Rv1981c	*nrdF*	2.7	0.003	ribonuleotide reductase small subunit
Rv0952	*sucD*	2.6	7.00E-04	succinyl CoA syntahse a chain
Rv0125	*pepA*	2.4	5.00E-04	Probable serine protease
Rv2460c	*clpP2*	2.3	0.005	ATP dependent Clp protease proteolytic subunit
Rv0248c	*Rv0248c*	2.3	0.008	Probable Flavoprotein
Rv3914	*trxC*	2.3	2.00E-04	Thioredoxin
Rv2461c	*clpP*	2.2	0.01	ATP dependent Clp protease proteolytic subunit
Rv3117	*cysA3*	2.2	0.02	Thiosulphate sulfurtransferase
Rv0864	*moaC2*	2.2	0.004	Molybdenum cofactror biosynthesis, protein C
Rv1703c	*Rv1703c*	2.2	0.003	Possible o-methyltransferase
Rv3224	*Rv3224*	2.2	0.007	Oxidoreducatse
Rv0815c	*cysA2*	2.1	0.02	Thiosulphate sulfurtransferase
Rv1240	*mdh*	2.1	1.00E-04	Malate dehydrogenase
Rv2213	*pepB*	2.1	6.00E-04	Aminopeptidase A/I
Rv0247c	*Rv0247c*	2.1	0.02	Probable Iron-Sulphur protein
Rv1498c	*Rv1498c*	2.1	5.00E-05	Methyltransferase
Rv2454c	*Rv2454c*	2.1	2.00E-04	Possible oxidoreductase
Rv2959c	*Rv2959c*	2.1	0.003	Possible methylotransferase
Rv0951	*sucC*	2	3.00E-04	succinyl CoA syntahse b chain
Rv0423c	*thiC*	2	0.02	Thiamine synthesis, pyrimidine moiety

				Unknown function
ORF	**Gene**	**Fold**	**T-test**	**Description**
Rv2226	*Rv2226*	4.4	6.00E-05	Unknown
Rv2706c	*Rv2706c*	4.3	0.05	Unknown
Rv1535	*Rv1535*	3.6	0.01	Unknown
Rv3603c	*Rv3603c*	3.3	0.002	Unknown
Rv1192	*Rv1192*	2.6	0.001	Unknown
Rv2705c	*Rv2705c*	2.6	0.02	Unknown
Rv2271	*Rv2271*	2.5	0.004	Unknown
Rv3651	*Rv3651*	2.4	0.008	Unknown
Rv3654c	*Rv3654c*	2.2	0.002	Unknown
Rv2844	*Rv2844*	2	6.00E-04	Unknown

Appendix III- table 3. Decreased expression genes in mutant/ STATIONARY PHASE

				Regulatory mechanisms
ORF	Gene	Fold	T-test	Description
Rv0465c	*Rv0465c*	2.5	0.003	Putative transcription regulator (Pbs X/ Xre family)
Rv3219	*whiB1*	2.5	9.00E-04	WhiB transcriptional activator homologue
Rv2324	*Rv2324*	2.3	0.02	Transcriptional regulator (Lrp/ AsnC family)
Rv0348	*Rv0348*	2.1	0.04	Putative transcription regulator
Rv3249c	*Rv3249c*	2	0.006	Transcriptional regulator (TetR/AcrR family)

				Conserved Hypotheticals
ORF	Gene	Fold	T-test	Description
Rv1130	*Rv1130*	4.8	0.009	Conserved Hypothetical
Rv0483	*Rv0483*	3.6	7.00E-04	Conserved Hypothetical
Rv2699c	*Rv2699c*	3.6	0.002	Conserved Hypothetical
Rv1194c	*Rv1194c*	3.2	0.002	Conserved Hypothetical
Rv0047c	*Rv0047c*	3.1	0.008	Conserved Hypothetical
Rv0516c	*Rv0516c*	3	7.00E-04	Conserved Hypothetical
Rv3600c	*Rv3600c*	3	0.02	Conserved Hypothetical
Rv0637	*Rv0637*	2.8	0.01	Conserved Hypothetical
Rv0636	*Rv0636*	2.6	0.006	Conserved Hypothetical
Rv0991c	*Rv0991c*	2.5	0.009	Conserved Hypothetical
Rv1176c	*Rv1176c*	2.5	0.005	Conserved Hypothetical
Rv0010c	*Rv0010c*	2.4	0.004	Conserved Hypothetical
Rv0635	*Rv0635*	2.4	0.01	Conserved Hypothetical
Rv2698	*Rv2698*	2.4	0.005	Conserved Hypothetical
Rv0177	*Rv0177*	2.3	0.03	Conserved Hypothetical
Rv3194c	*Rv3194c*	2.3	9.00E-04	Conserved Hypothetical
Rv1073	*Rv1073*	2.2	0.005	Conserved Hypothetical
Rv0146	*Rv0146*	2.1	0.04	Conserved Hypothetical
Rv0464c	*Rv0464c*	2.1	0.03	Conserved Hypothetical
Rv2050	*Rv2050*	2.1	1.00E-03	Conserved Hypothetical
Rv2185c	*Rv2185c*	2.1	5.00E-04	Conserved Hypothetical
Rv2298	*Rv2298*	2.1	0.001	Conserved Hypothetical
Rv0308	*Rv0308*	2.1	8.00E-04	Conserved Hypothetical
Rv1078	*pra*	2	0.001	Proline rich antigen
Rv0049	*Rv0049*	2	0.008	Conserved Hypothetical
Rv0263c	*Rv0263c*	2	0.002	Conserved Hypothetical

Table 4. Genes found to be expressed at a higher level in the *M. tuberculosis* Δ*sigC* mutant when in stationary phase.

Virulence/ Detoxification				
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv2031c	*hspX*	17.3	0.001	14 kD antigen, heat shock protein Hsp20 family

Lipid metabolism				
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv0824c	*desA1*	7.7	0.005	Acyl-(ACP) desaturase
Rv1094	*desA2*	4.1	0.03	Acyl-(ACP) desaturase
Rv2590	*fadD9*	2.7	0.003	Acyl-CoA synthase

Cell wall and cell process				
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv1737c	*narK2*	6.8	0.01	Nitrite extrusion protein
Rv1733c	*Rv1733c*	5.7	0.002	Membrane protein
Rv0677c	*mmpS5*	4.8	0.02	Conserved small membrane protein
Rv2846	*efpA*	3.7	0.02	Putative efflux protein
Rv0676c	*mmpL5*	3.7	0.02	Conserved Large membrane protein
Rv0655	*Rv0655*	3.5	0.002	Hypothetical ABC transporter
Rv1739c	*Rv1739c*	3.1	0.05	Sulphate transporters
Rv1997	*ctpF*	2.8	0.01	Probable cation transport ATPase

Intermediary metabolism				
ORF	**Gene**	**Fold**	**t-test**	**Description**
Rv2029	*pfkB*	9.1	0.005	Phosphofructokinase II
Rv2007	*fdxA*	7.4	0.002	Ferredoxin
Rv3854c	*ethA*	5	0.04	Intermediary metabolism
rv0693	*pqqE*	3.9	0.009	Coenzyme PQQ syntheisis protein E
Rv0694	*lldD1*	3.4	0.03	L-lactate dehydrogenase (cytochrome)
Rv2781c	*Rv2781c*	3	0.02	2-nitropropane dioxygenase
Rv3855	*ethR*	2.8	0.02	Intermediary metabolism
Rv0570	*nrdZ*	2.7	0.01	Ribonucleotide reductase, class ii
Rv1552	*frdA*	2.5	0.03	Fumarate reductase flavoproetin subunit
Rv0458	*Rv0458*	2.4	0.03	Probable aldehyde dehydrogenase

Appendix III- table 4. Increased expression genes in knock-out/ STATIONARY PHASE

				Unknown function
ORF	Gene	Fold	t-test	Description
Rv2004c	*Rv2004c*	4.1	0.01	Unknown
Rv0572c	*Rv0572c*	3.3	0.01	Unknown
Rv3839	*Rv3839*	3.2	0.01	Unknown
Rv3613c	*Rv3613c*	3.1	0.02	Unknown
Rv2492	*Rv2492*	3	0.04	Unknown
Rv2630	*Rv2630*	3	0.01	Unknown
Rv3662c	*Rv3662c*	2.5	0.03	Unknown
Rv0193c	*Rv0193c*	2.2	0.02	Unknown

				Regulatory mechanism
ORF	Gene	Fold	t-test	Description
Rv3133c	*Rv3133c*	6.8	0.01	Two-component response regulator
Rv3132c	*Rv3132c*	5.8	0.005	Sensor histadine kinase
Rv0823c	*Rv0823c*	4.5	4.00E-04	Transcription regulator (NifR3/Smm1 family)

				Conserved Hypotheticals
ORF	Gene	Fold	t-test	Description
Rv3129	*Rv3129*	18.7	0.02	Conserved Hypothetical
Rv2030c	*Rv2030c*	16.7	0.002	Conserved Hypothetical
Rv1738	*Rv1738*	12.7	0.02	Conserved Hypothetical
Rv2626c	*Rv2626c*	10.4	0.008	Conserved Hypothetical
Rv3128c	*Rv3128c*	9.6	0.004	Conserved Hypothetical
Rv2624c	*Rv2624c*	7.8	0.005	Conserved Hypothetical
Rv2623	*Rv2623*	7.7	0.01	Conserved Hypothetical
Rv3131	*Rv3131*	7.3	0.003	Conserved Hypothetical
Rv2625c	*Rv2625c*	6.9	0.02	Conserved Hypothetical
Rv3134c	*Rv3134c*	6.7	0.03	Conserved Hypothetical
Rv3616c	*Rv3616c*	6.3	0.004	Conserved Hypothetical
Rv1734c	*Rv1734c*	5.6	0.005	Conserved Hypothetical
Rv2005c	*Rv2005c*	5.4	0.03	Conserved Hypothetical
Rv2032	*Rv2032*	5.3	0.02	Conserved Hypothetical
Rv3615c	*Rv3615c*	5.1	0.006	Conserved Hypothetical
Rv3614c	*Rv3614c*	4.6	0.006	Conserved Hypothetical
Rv1813c	*Rv1813c*	3.6	0.02	Conserved Hypothetical
Rv0574c	*Rv0574c*	2.8	0.009	Conserved Hypothetical
Rv0830	*Rv0830*	2.5	0.04	Conserved Hypothetical
Rv2956	*Rv2956*	2.4	0.006	Conserved Hypothetical
Rv1398c	*Rv1398c*	2.3	3.00E-04	Conserved Hypothetical

APPENDIX IV- POSTER ABSTRACT

Poster presentation at Taos, New Mexico for the Keystone conference on Genomics/ Post- Genomics/ Genetic diversity of *M. tuberculosis*, Friday 26th January, 2001

Krishna K. Gopaul and Elaine O. Davis, Division of Mycobacterial Research, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, England.

The RecA protein plays a central role in the recombination and DNA repair and is highly conserved among bacteria. Its expression is induced in response to DNA damage as part of the SOS response, generally being regulated by the repressor protein LexA. The *recA* gene of *M. tuberculosis* is unusual in having two promoters. The more distal of these, P2, contains a LexA binding site, whilst that proximal to the *recA* gene, P1, is DNA-damage inducible independently of LexA. Through the introduction of point mutations into the putative –10 and –35 regions of both P1 and P2, the sequences important for promoter function were analysed. Mutations eliminating the activity of each promoter were identified. Of particular interest, this investigation determined that a region with homology to the *E. coli* σ^{70} –35 sequence but located much closer to the –10 sequence was important for P1 expression in *M. tuberculosis* whilst the sequence at the standard location was not. Despite the difference in induction mechanisms of the two promoters, both promoters responded to mitomycin C in *M. smegmatis* with similar kinetics. In addition, it was found that the basal activity of each promoter was independent of growth phase.