In the name of GOD, the almighty

REGULATION OF MYCOBACTERIUM TUBERCULOSIS
RECA EXPRESSION

Farahnaz Movahedzadeh

A thesis submitted to the university college of London
for the degree of Doctor of Philosophy

October 1996

Division of Mycobacterial Research
National Institute for Medical Research
Mill Hill
London
if you were to ask me
if I liked my teachers
more than the stars
I would have to say,
a little more

To my parents

To Elaine and all my teachers
ACKNOWLEDGEMENTS

I would like to extend the love and prayer with which I achieved my PhD to all the people who have helped me in so many different ways through my study.

First of all, to my dear supervisors, Dr Jo Colston and Dr Elaine Davis.

To Dr Colston, thank you for helping me to obtain a grant from 'WHO', for enabling me work in your Division and, in this way, giving me the opportunity of working on the interesting world of the molecular biology of mycobacteria. I will be forever grateful for the support you have given me over the years.

To Elaine, for being such a marvellous supervisor - thank you for everything. You have been very special to me, and I will always be proud to have been your student.

Thanks a lot Sundaram for your critical comments and all your help through my Ph.D. Thank you Pat for your technical advise and for your kindly manner and also to Peter Jenner for your help in purification of LexA protein (I would never forget my wonderful gel); and a special thank you to Fran Johnson and all my colleagues in the Division of Mycobacterial Research.

To the World Health Organisation for awarding me a grant to study for a PhD (it was a miracle).

To the Photographics Department, especially Neil Papworth, for my figures.

To my neighbours in Fir Island, Veronica, Alejandro, Suammy, Davinda, Sulochana, for your company.

To all my Iranian friends, for your support and love, which has brought me through.

To my very special friend, Sharaf, for all the efficient Prayer you taught me.

And Finally, to my wonderful parents, my brothers, and my forever friends; my husband Mohammad and my son Reza, for your love, encouragement and everlasting support throughout my life.
ABSTRACT

The SOS response involves the coordinated expression of more than 20 genes in response to DNA damage and makes an important contribution to the survival of bacteria in hostile environments. The key regulatory components of the SOS response are RecA and LexA, the activator and repressor respectively. The \textit{recA} gene of \textit{M. tuberculosis} has previously been cloned and sequenced (Davis \textit{et al.}, 1991). However, nothing was known about the regulation of this system in mycobacteria. In this study the expression of the \textit{recA} gene was shown to be inducible in response to various DNA damaging agents by using a transcriptional fusion to the reporter gene chloramphenicol-acetyl transferase (CAT). Furthermore, by producing constructs containing various lengths of upstream sequence I have been able to identify the key regulatory elements involved in the induction. A segment of DNA around 300 bp upstream of the coding region was shown to be required for expression. This stretch of DNA contains a putative LexA-binding site, based on homology to the "Cheo box" binding site of \textit{Bacillus subtilis} (but unlike the "SOS box" of \textit{Escherichia coli}). I have also cloned and sequenced the \textit{lexA} gene of \textit{Mycobacterium tuberculosis} and this too has a "Cheo box" like sequence within its upstream region. Primer extension analysis has been used to identify promoter sequences involved in \textit{recA} and \textit{lexA} expression. The LexA protein was overexpressed, purified and demonstrated to bind to the Cheo boxes.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>5</td>
</tr>
<tr>
<td>List of Figures</td>
<td>9</td>
</tr>
<tr>
<td>List of Tables</td>
<td>12</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>13</td>
</tr>
<tr>
<td>1. <strong>Introduction</strong></td>
<td>14</td>
</tr>
<tr>
<td>1.1 The mycobacteria</td>
<td>14</td>
</tr>
<tr>
<td>1.1.1 Characteristics</td>
<td>14</td>
</tr>
<tr>
<td>1.1.2 Pathogenicity</td>
<td>15</td>
</tr>
<tr>
<td>1.2 Molecular biology of the mycobacteria</td>
<td>18</td>
</tr>
<tr>
<td>1.3 Cloning and expression of mycobacterial genes</td>
<td>20</td>
</tr>
<tr>
<td>1.3.1 <em>Escherichia coli</em> as a cloning host</td>
<td>20</td>
</tr>
<tr>
<td>1.3.2 <em>Streptomyces lividans</em> as a cloning host</td>
<td>21</td>
</tr>
<tr>
<td>1.3.3 Mycobacteria as cloning hosts</td>
<td>22</td>
</tr>
<tr>
<td>1.4 Mycobacterial plasmid vectors</td>
<td>23</td>
</tr>
<tr>
<td>1.5 Gene regulation</td>
<td>25</td>
</tr>
<tr>
<td>1.5.1 RNA polymerase</td>
<td>25</td>
</tr>
<tr>
<td>1.5.2 Regulatory proteins</td>
<td>27</td>
</tr>
<tr>
<td>1.5.3 Gene regulation in mycobacteria</td>
<td>27</td>
</tr>
<tr>
<td>1.6 The SOS response</td>
<td>28</td>
</tr>
<tr>
<td>1.6.1 Roles of RecA protein</td>
<td>30</td>
</tr>
<tr>
<td>1.6.2 Homologous recombination</td>
<td>30</td>
</tr>
<tr>
<td>1.6.3 Homologous recombination in mycobacteria</td>
<td>32</td>
</tr>
<tr>
<td>1.7 Regulation of the SOS response</td>
<td>32</td>
</tr>
<tr>
<td>1.8 The SOS response in mycobacteria</td>
<td>34</td>
</tr>
<tr>
<td>1.8.1 Unusual structure of the <em>recA</em> gene of pathogenic mycobacteria</td>
<td>34</td>
</tr>
<tr>
<td>1.8.2 The common features of inteins</td>
<td>36</td>
</tr>
<tr>
<td>1.9 Aims</td>
<td>37</td>
</tr>
<tr>
<td>2. <strong>Materials and Methods</strong></td>
<td>41</td>
</tr>
<tr>
<td>2.1 Plasmid preparation</td>
<td>41</td>
</tr>
<tr>
<td>2.1.1 Small scale plasmid preparation</td>
<td>41</td>
</tr>
<tr>
<td>2.1.2 Large scale plasmid preparation</td>
<td>41</td>
</tr>
</tbody>
</table>
2.1.3 Plasmid preparation from mycobacteria ............................................. 41

2.2 Cloning ..................................................................................................... 42

2.2.1 Restriction enzyme digestion and alkaline phosphatase treatment .............................................................. 42

2.2.2 Agarose gel electrophoresis ............................................................... 42

2.2.3 Purification of DNA fragments by agarose gel electrophoresis .......... 43

2.2.4 Ligations ............................................................................................. 43

2.2.5 Transformation of E. coli ................................................................. 44

2.2.6 Transformation of mycobacteria by electroporation ....................... 45

2.3 Southern hybridization ......................................................................... 46

2.3.1 Probe labelling ................................................................................ 46

2.3.2 DNA transfer and hybridization .................................................... 46

2.4 Cloning, expression and purification of M. tuberculosis LexA ............. 46

2.4.1 PCR for cloning in pET-15b ............................................................. 46

2.4.2 IPTG induction using E. coli host strain BL21 (DE3) ..................... 48

2.4.3 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE) ................................................................................. 48

2.4.3.1 Preparation of gels .................................................................. 49

2.4.3.2 Rapid staining ....................................................................... 49

2.4.4 LexA protein purification ............................................................... 50

2.5 Sequencing ............................................................................................ 51

2.5.1 Sequencing reactions ..................................................................... 51

2.5.2 Preparation and electrophoresis of sequencing gels ..................... 52

2.6 Identification of the transcription start site ......................................... 54

2.6.1 RNA preparation ......................................................................... 54

2.6.2 Primer extension ......................................................................... 55

2.7 DNA binding ........................................................................................ 55

2.7.1 Gel shift assay .......................................................................... 55

2.7.2 Footprinting .............................................................................. 56

2.8 Induction of RecA-CAT fusion in pEJ258, pFM3, pFM7 in M. smegmatis mc<sup>2</sup>155 ................................................................. 58

2.9 Bacterial strains and plasmids ............................................................. 59

3. Optimization of the conditions for induction of M. tuberculosis RecA .............................................................................. 64

3.1 Introduction ........................................................................................... 64
3.2 Confirmation of the presence of an inducible promoter for the
M. tuberculosis recA Gene .................................................................64
3.3 Optimization of the conditions required for M. tuberculosis recA
induction in M. smegmatis .................................................................65
3.4 Discussion ...................................................................................73

4. Identification of upstream DNA required for M. tuberculosis
RecA induction .................................................................................75
4.1 Introduction ..................................................................................75
4.2 Constructs to localize the inducible promoter of the M. tuberculosis
recA Gene ..........................................................................................75
4.3 CAT assays of mycobacterial transformants containing deleted
constructs .........................................................................................77
4.4 Sequence determination of the recA upstream region...............79
4.5 A single Cheo box is located upstream of the M. tuberculosis recA
gene .................................................................................................82
4.6 Discussion ...................................................................................82

5. Transcription start sites of M. tuberculosis recA gene ............85
5.1 Introduction ..................................................................................85
5.2 Transcription start point (promoter region) of M. tuberculosis recA
expressed from a plasmid ................................................................. 85
5.3 Transcription start point of M. tuberculosis recA expressed from the
chromosome ......................................................................................86
5.4 Sequences with homology to known promoters near the mapped
transcription start sites .................................................................87
5.5 Discussion ...................................................................................91

6. Cloning, sequence determination and transcription start
site of the Mycobacterium tuberculosis lexA gene .....................95
6.1 Introduction ..................................................................................95
6.2 Cloning of the M. tuberculosis lexA gene ...............................96
   6.2.1 Southern hybridization to a cosmid bearing lexA .............96
   6.2.2 Cloning of the lexA gene in pUC19 .................................96
6.3 Sequencing of the lexA gene .....................................................98
6.4 Comparison of the M. tuberculosis lexA gene with that of other bacteria
at the protein level .........................................................................98
6.5 Predicted LexA binding site (Cheo box) upstream of M. tuberculosis and
M. leprae lexA genes.................................................................................................101
6.6 Localization of the promoter region of the *M. tuberculosis* *lexA* gene .......104
6.7 Discussion .........................................................................................................107

7. **Expression and purification of *M. tuberculosis* LexA** .........................111
7.1 Introduction ........................................................................................................111
7.2 Expression studies of *M. tuberculosis* LexA in the vector pTZ18R ..........111
   7.2.1 Cloning of the *lexA* gene in the vector pTZ18R ....................................111
   7.2.2 Expression of the *M. tuberculosis* *lexA* gene from the T7
       promoter of pTZ18R......................................................................................112
7.3 Expression studies of *M. tuberculosis* LexA in the vector pET-15b ........114
   7.3.1 Cloning of the *lexA* gene in the vector pET15-b....................................114
   7.3.2 Overexpression of *M. tuberculosis* LexA protein .............................116
7.4 Purification of the LexA protein.......................................................................117
7.5 Discussion .........................................................................................................123

8. **Interaction of LexA protein with the recA operator**
   Region ................................................................................................................125
8.1 Introduction ........................................................................................................125
8.2 LexA binds to the promoter region of *recA* ................................................125
8.3 LexA binds to an oligonucleotide containing the Cheo box .........................128
8.4 Mapping the sites of interaction of LexA with the *recA* promoter region
    by DNase I footprinting....................................................................................129
8.5 Confirmation of the importance of the Cheo box in the DNA-LexA
    protein interaction in *M. tuberculosis* .......................................................133
8.6 The *lexA* gene product represses its own promoter ..................................135
8.7 Discussion .........................................................................................................137

9. **General Discussion** .....................................................................................141
9.1 Gene regulation in the mycobacteria .........................................................141
9.2 Inducible expression of *M. tuberculosis* RecA ........................................141
9.3 Promoter region of the *M. tuberculosis* *recA* .........................................144
9.4 *M. tuberculosis* LexA protein, the repressor of the SOS response ..........149
9.5 Regulatory function of *M. tuberculosis* LexA protein ..............................151
9.6 Promoter region of *M. tuberculosis* *lexA* ..............................................155
9.7 Future work ....................................................................................................157

References ............................................................................................................160
Appendix: media ....................................................................................................192
LIST OF FIGURES

1.1 Roles of RecA protein in regulation of the SOS response ........................................29
1.2 Homologous recombination as a laboratory tool ..................................................31
1.3 Production of the mature \textit{M. tuberculosis} RecA protein by post-translational
splicing ..................................................................................................35
2.1 The promoter probe vector pEJ257 ...................................................................63
3.1 The effect of growth stage on the induction of \textit{M. tuberculosis} RecA in \textit{M. smegmatis}
as measured by CAT production ...................................................67
3.2 The effect of ofloxacin concentration on the induction of \textit{M. tuberculosis}
RecA in \textit{M. smegmatis} as measured by CAT production .........................68
3.3 The effect of mitomycin C concentration on the induction of \textit{M. tuberculosis}
RecA in \textit{M. smegmatis} as measured by CAT production .........................69
3.4 The effect of nalidixic acid concentration on the induction of \textit{M. tuberculosis}
RecA in \textit{M. smegmatis} as measured by CAT production .........................70
3.5 The effect of uv irradiation on the induction of \textit{M. tuberculosis} RecA in
\textit{M. smegmatis} as measured by CAT production ..................................71
3.6 The effect of time of induction on the expression of \textit{M. tuberculosis} RecA in
\textit{M. smegmatis} as measured by CAT production ..................................72
4.1 Constructs prepared to locate the inducible promoter in the DNA upstream
of the \textit{M. tuberculosis} \textit{recA} gene .................................................................76
4.2 Induction of constructs containing DNA from upstream of the
\textit{M. tuberculosis} \textit{recA} gene by ofloxacin in \textit{M. smegmatis} mc^2155 ................78
4.3 Restriction site map and sequencing strategy for the sequence upstream of the \textit{recA} gene of \textit{M. tuberculosis} .................................................................80

4.4 Complete nucleotide sequence of the 670 bp upstream of the \textit{recA} gene of \textit{M. tuberculosis} contained in pFM7 .................................................................81

5.1 Identification of the transcription start sites of the \textit{M. tuberculosis recA} gene expressed from a plasmid .................................................................88

5.2 Identification of the transcription start sites of the \textit{M. tuberculosis recA} gene expressed from the chromosome .................................................................89

5.3 Transcription start sites and putative promoter elements upstream of the \textit{M. tuberculosis recA} gene .................................................................90

6.1 Southern hybridization of 400 bp \textit{lexA} probe to T616 cosmid DNA digested with various restriction enzymes .................................................................97

6.2 Cloning and sequencing strategy of the \textit{M. tuberculosis lexA} gene .................................................................99

6.3 Complete nucleotide sequence of the \textit{lexA} gene of \textit{M. tuberculosis} .................................................................100

6.4 Comparison of the \textit{M. tuberculosis} LexA protein sequence with those of other bacteria .................................................................102

6.5 Comparison of the \textit{M. tuberculosis} LexA protein sequence with LexA of \textit{E. coli} .................................................................103

6.6 The upstream sequence of the \textit{M. leprae lexA} .................................................................105

6.7 Identification of the transcription start site of the \textit{M. tuberculosis lexA} gene .................................................................106

7.1 The strategy for preparing the construct pFM16 .................................................................113
7.2 The strategy for preparing pFM18 construct ................................................. 115

7.3 SDS-PAGE analysis of *E. coli* BL21 (DE3) cell free extracts expressing the
LexA protein .................................................................................................. 118

7.4 SDS-PAGE analysis of LexA protein purification using Ni-NTA resin
(Batch procedure) ......................................................................................... 120

7.5 SDS-PAGE analysis of LexA protein purification using Ni-NTA resin
(column procedure) ..................................................................................... 121

7.6 Flow diagram of the LexA purification procedure ................................…… 122

8.1 Gel shift assay with 350 bp fragment from upstream of *recA* ................... 127

8.2 Gel shift assay with 32 bp oligonucleotide containing the Cheo box .......... 130

8.3 DNaseI footprinting of the *recA* promoter region binding to LexA......... 132

8.4 Gel shift assay using mutated and wild-type oligonucleotides ............... 134

8.5 Gel shift assay using an oligonucleotide containing the Cheo box .......... 136
LIST OF TABLES

2.1  Plasmids used in this work .................................................................60

9.1  Comparison of the recognition site for LexA in the Gram-negative bacteria
     with that of B. subtilis, M. tuberculosis and M. leprae ......................... 156
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>CAT</td>
<td>Chloramphenicol acetyl transferase</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Deoxynucleoside 5'-triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kbp</td>
<td>Kilobase pair</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>Km'</td>
<td>Kanamycin resistance</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>orf</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>RFLPs</td>
<td>Restriction fragment length polymorphisms</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-hydroxymethylpropane-1,3-diol</td>
</tr>
<tr>
<td>X-gal</td>
<td>4-bromo-3-chloro-2-indolyl-β-galactoside</td>
</tr>
<tr>
<td>uv</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 The mycobacteria

Historically, members of the genus *Mycobacterium* have been some of the most important players in the drama of infectious disease and immunology of man. The generic name *Mycobacterium* was originally established by Lehmann and Neumann in 1896 to include the pathogenic tubercle and leprosy bacilli, but now encompasses over 50 approved species, with new species still being discovered.

1.1.1 Characteristics

Mycobacteria belong to the high guanine plus cytosine subdivision of Gram-positive bacteria. The G+C content of their DNA ranges from 62-70% (Goodfellow & Wayne, 1982) with the exception of that of *M. leprae*, which is 56% (Clark-Curtiss et al., 1985). *M. tuberculosis* has a genome size similar to that of *Escherichia coli* (4.4x10^9 M' (daltons)), while the genome of *M. leprae* is somewhat smaller, estimates ranging from 1.3x10^9 to 2.2x10^9 M' (Clark-Curtiss et al., 1985; Imaeda et al., 1982). Mycobacteria are aerobic, acid-alcohol fast actinomycetes, usually slightly curved or straight rods. All have meso-diaminopimelic acid, arabinose and galactose in their walls, produce straight chain fatty acids and characteristic long-chain mycolic acids (Goodfellow and Wayne, 1982). Mycobacteria can be classified into two major categories, depending on their growth characteristics under optimal conditions *in vitro*: the rapid growers which yield visible colonies in two to four days, and the slow growers which take ten days or longer to yield visible colonies.
(except *M. leprae* which has not been grown *in vitro* and can only be propagated in the laboratory in nine-banded armadillos or footpads of mice). The optimum temperatures for species in both of these categories range from 28-45°C.

### 1.1.2 Pathogenicity

Mycobacteria were among the first bacterial pathogens of man to be studied. Hansen (1874) identified the leprosy bacillus, *M. leprae*, in the lesions of leprosy patients and Koch (1882) clearly established that *M. tuberculosis* was the causative agent of tuberculosis (TB). Today in spite of significant progress in diagnosis, treatment and prevention, TB and leprosy still remain major public health problems worldwide. While mortality is the major concern in tuberculosis, physical deformity is the major problem in leprosy.

Leprosy afflicts 5.5 million people around the world today (Noordeen *et al.*, 1992). The causative agent remains one of the few major pathogens not grown successfully in culture. Effective chemotherapy exists, but a preventive vaccine would be preferable, since the disease is common in areas remote from medical services. BCG appears to have a limited role in the prevention of leprosy and the need for more effective vaccines remains an important target for research (Noordeen and Godal, 1988).

Twenty years ago, tuberculosis was thought to be a disease that was no longer to be a menace in the world with the presence of a vaccine and the development of
effective chemotherapies to control this infection. However, in the last few years the world has seen an alarming resurgence of tuberculosis. More than 20% of the world’s population is infected with the tubercle bacillus. Every year there are 8 to 10 million new cases and 2 to 3 million deaths (Kochi, 1991) resulting from this dreaded illness. In addition, an increasing number of reports of infections with drug-resistant *Mycobacterium tuberculosis* makes this resurgence so much more frightening (Jacobs, 1992).

*M. tuberculosis* and *M. leprae* are closely related and show several important biological similarities. They both grow inside host cells for long periods of time, and both infections are controlled by cell-mediated mechanisms, rather than by antibody responses. In addition both species show the phenomenon of persistence, where bacilli in tissues can survive for long periods of time in the presence of bactericidal concentrations of anti-tuberculous or anti-leprosy drugs. The obvious similarities between *M. tuberculosis* and *M. leprae* mean that research on one of these organisms is often applicable to the other.

The BCG (Bacille Calmette Guerin) vaccine owes its origin to the *in vitro* attenuation, by Calmette and Guerin of *Mycobacterium bovis* and is the most widely used vaccine in the world. It is also a potentially useful vaccine vehicle for delivering protective antigens of multiple pathogens. This vaccine has powerful adjuvant activity, being able to stimulate both the humoral and cell-mediated immune systems with a low rate of serious complications. The ability of BCG to protect against tuberculosis has varied considerably in a number of human vaccine
trials. For example, BCG has been shown to have poor efficiency in southern India and the southern United States, whilst it is effective in the United Kingdom and East Africa (reviewed by Fine and Rodrigues, 1990). The high incidence of tuberculosis infections around the world and the inability of BCG to protect certain populations clearly shows that a better vaccine is required. An improved BCG vaccine might also provide greater protection against leprosy.

Non-tuberculous mycobacteria (NTM) were only recognized as pathogens in the 1950s. *M. avium* complex are opportunistic pathogens common in the environment and are of renewed interest today because they are one of the most common causes of systemic bacterial infections in HIV-positive individuals, with 3% of HIV-deaths caused by *M. avium* infection (Horsburgh, 1991). *M. paratuberculosis* is a very closely related organism which is the causative agent of Johne’s disease (or paratuberculosis) in cattle and sheep and also has been implicated in the etiology of human Crohn’s disease (Chiodini, 1989; Elsaghier *et al.*, 1992), although this remains controversial. *M. marinum* causes disease in fish and self-limiting skin granulomas in man, whereas *M. ulcerans* causes progressive malignant ulcers of skin and deeper tissues in several tropical regions. A variety of fast growing environmental mycobacteria may also manifest as post-surgical and post-traumatic infections in immuno-compromised patients. Few vertebrates are immune to mycobacterial infections and some individual species, such as *M. bovis* and *M. avium* not only infect several animal species, but may be transmitted between species.
Despite the early discovery of mycobacteria, the failure of *M. leprae* to grow *in vitro* and the slow growth rate of *M. tuberculosis* have constrained the study of the basic biology and host-pathogen relationship of these bacteria. Greater understanding of the mechanisms of mycobacterial pathogenesis, immunity and molecular biology will certainly be required in order that rapid methods for diagnosis, and more effective vaccination and treatment programmes be developed.

### 1.2 Molecular biology of the mycobacteria

The mycobacteria were among the first bacteria to be associated with a human disease, and during the past century, research on mycobacteria has led to major insights into cellular immunology and the physiology and biochemistry of the mycobacteria. In contrast relatively little was known about the genetics or molecular biology of mycobacteria until about 10 years ago. The lack of progress in mycobacterial genetics was primarily because i) mycobacteria grow very slowly (*Mycobacterium tuberculosis* has a generation time of 18-24 hours, and *M. leprae* has yet to be cultivated *in vitro*); ii) mycobacteria are hydrophobic and tend to grow in clumps, which makes it difficult to purify individual cells for genetic analysis; iii) there is no known occurring genetic exchange between mycobacteria; and iv) very few genetic markers have been identified in mycobacteria (Shinnick *et al.*, 1995).

During the last few years the more classical genetic approach to the study of mycobacteriology has been gradually superseded by molecular techniques. Mainly through the development of transformation methods and suitable vectors,
recombinant DNA techniques became applicable to mycobacterial systems. Candidates for useful genes to be cloned include those coding for antigens that might be harnessed for diagnosis or for protection through stimulation of antibody production or T-cell activity, and structural genes for targets against which novel drugs might be directed. Over 50 mycobacterial antigens have now been characterised and several identified as proteins with important functions (Young et al., 1992).

Insertion elements have been discovered in several mycobacterial species (Green et al., 1989; Moss et al., 1992; McAdam et al., 1990; Thierry et al., 1990a,b; Hermans et al., 1991; Collins and Stephans, 1991), and these may be of use in developing integrative vectors and systems for insertional mutagenesis, as well as being applied to the identification of mycobacteria and studying the epidemiology of tuberculosis (van Soolingen et al., 1992). Use of specific DNA probes (McFadden et al., 1990) and the polymerase chain reaction (PCR) for the detection and identification of individual mycobacterium species has been reported. Many target sequences have been used, but the most thoroughly evaluated assays target the M. tuberculosis-specific repeated DNA element IS6110 (Eisenach et al., 1990; 1991).

Shuttle vectors for use in mycobacterial cloning hosts have been constructed (reviewed by Hatfull, 1993) and the use of M. smegmatis as a cloning host has been greatly facilitated by the isolation of the mutant designated as M. smegmatis mc²155 (Snapper et al., 1990). Recombinant DNA technology is being applied to investigate and identify mycobacterial virulence determinants and mechanisms of pathogenicity,
elucidate the mechanisms of drug action and resistance, and the development of multivalent vaccines, employing the vaccine strain *M. bovis* BCG. Also, biotechnological applications of mycobacteria might benefit from knowledge of the systems involved in regulation of replication and gene expression (Hartmans et al., 1989). The ability to replace genes with mutated versions by homologous recombination would greatly aid the analysis of gene function in mycobacteria. Although this has been possible in the fast growing mycobacterium, *M. smegmatis* (Husson et al., 1990), slow growing mycobacteria have been recalcitrant to this methodology (see below), but techniques to improve their efficiency are just becoming available (Marklund et al., 1995; Norman et al., 1995; Reyrat et al., 1995; Balasubramanian et al., 1996).

### 1.3 Cloning and expression of mycobacterial genes

#### 1.3.1 *Escherichia coli* as a cloning host

The first cloning host used was *Escherichia coli* because of the wealth of genetic knowledge and of procedures for gene cloning available for this organism. The genomes of several mycobacterial pathogens, including *M. leprae* and *M. tuberculosis*, have been cloned in *E. coli* using both plasmid and phage expression vectors (Clark-curtiss et al., 1985; Young et al., 1985a,b; Jacobs et al., 1986). Young et al. (1985a,b) constructed libraries of *M. leprae* and *M. tuberculosis* genomic DNA sequences in the λgt11 expression vector, which transcribes the mycobacterial genes using a promoter present in the vector and translates the mycobacterial genes as fusion proteins with β-galactosidase. The lambda gt11
libraries have been used by numerous investigators to clone and characterize genes that encode immunoreactive proteins of mycobacteria. Using a panel of mouse monoclonal antibodies directed against *M. leprae* (Engers et al., 1985), recombinant clones expressing antigenic epitopes of five highly immunogenic *M. leprae* protein antigens were isolated (Young et al., 1985b).

Clark-Curtiss et al. (1985) constructed genomic libraries using cosmid expression vectors in *E. coli* and demonstrated that mycobacterial DNA could be expressed in *E. coli* under the control of an *E. coli* promoter, but mycobacterial promoters were mostly poorly recognised or not recognised at all by the *E. coli* transcription machinery. Only a small number of BCG promoters are known to be active in *E. coli* (Thole et al., 1985), although a few are strongly active (Sirakova et al., 1989). The defect in the expression of mycobacterial genes in *E. coli* led to the investigation of possible alternative cloning hosts. *Streptomyces lividans* has been explored as a host capable of expressing many mycobacterial genes, or perhaps groups of linked genes coding for a biosynthetic pathway or a complex antigen, from mycobacterial expression signals.

1.3.2 *Streptomyces lividans* as a cloning host

The cloning systems in *Streptomyces* are reasonably well developed (Hopwood et al., 1985) and *Streptomyces* were known to recognise promoters from several other bacterial genera, including *Escherichia, Serratia* and *Bacillus* (Bibb and Cohen, 1982). The anticipation of the expression of mycobacterial genes was due to the high
G+C content of *Streptomycetes* DNA, so the organisms have evolved the ability to bring about strand separation of high G+C DNA in promoter regions as a prerequisite for transcription. In addition the availability of several different forms of RNA polymerase with alternative sigma factors are at least partly responsible for the transcription of different promoter classes (Westpheling *et al.*, 1985). Moreover, translation of genes from other bacteria seems to present little problem because the ribosome-binding sites of *Streptomycetes* genes require comparatively little complementarity to the 16S ribosomal RNA sequence (Hopwood *et al.*, 1986). Results of experiments with *M. bovis* BCG, *M. tuberculosis* and *M. leprae* indicate that many genes from these species can be expressed in *S. lividans* from their own promoters, making the isolation of genes or gene-sets involved in pathogenicity feasible (Kieser *et al.*, 1986; Lamb & Colston *et al.*, 1986).

### 1.3.3 Mycobacteria as cloning hosts

By analogy to other bacterial genera, DNA transcription and translation are likely to respond to the same signals in all mycobacteria. Therefore, the non-pathogenic, readily cultivable mycobacteria such as *M. smegmatis* and *M. bovis* BCG should be the ideal hosts for manipulating the genes of pathogenic mycobacteria. Any use of a mycobacterial strain as a host for recombinant DNA requires the development of an efficient transformation system in combination with versatile cloning vectors. The development of phasmids (Jacobes *et al.*, 1987; Snapper *et al.*, 1988) and plasmids (Snapper *et al.*, 1988) for the introduction of recombinant DNA into mycobacteria has allowed the development of mycobacteria as cloning hosts. The use of
**M. smegmatis** as a cloning host has been greatly facilitated by the isolation of a mutant (designated *M. smegmatis* mc²155) that can be transformed at much higher efficiencies than the parental strain using electroporation (Snapper et al., 1990).

Another important requirement in developing cloning hosts is the absence of host-dependent modification and restriction enzyme systems, and *M. smegmatis* mc²155 has been shown to lack such a system (Snapper et al., 1990).

However, since both these species show a tendency to grow in clumps, and *M. bovis* BCG is a slowly growing species, a better candidate might be found in *M. aurum* that belongs to the fast growing mycobacteria and shows no tendency to form clumps in liquid culture, even in growth medium without Tween-80. Transformation systems for *M. aurum*, based on spheroplast fusion or electroporation have been developed and vectors for this species have been examined (Rastogi et al., 1983; Hermans et al., 1990; Hermans et al., 1991).

### 1.4 Mycobacterial plasmid vectors

The first vectors for introducing foreign DNA were shuttle phasmids, which are chimeras of mycobacteriophage and plasmid DNA that grow as plasmids in *E. coli* and as phage in mycobacteria (Jacobs et al., 1987). These vectors permitted the introduction of foreign DNA by infection into *M. smegmatis* and BCG for the first time.

Small extrachromosomal plasmids are convenient as cloning vehicles and because
they can be simply modified and manipulated in *E. coli*, most of the plasmid vehicles described to date are shuttle vectors that can replicate in both *E. coli* and mycobacteria. The most commonly used plasmid replicon active in mycobacteria is derived from pAL5000 which was isolated from *M. fortuitum* (Labidi *et al.*, 1985). Other shuttle plasmids, not based on pAL5000, include pYT72/pYT92, a derivative of the *M. scrofulaceum* plasmid pMSC262 (Goto *et al.*, 1991) and the broad host range plasmids RSF1010 (Hermans *et al.*, 1991) and pNG2 (Radford and Hodgson, 1991).

The study of pathogenic determinants in animal systems requires that the recombinant DNA is stably maintained. Genetic tools for integration of foreign DNA into the mycobacterial genome have been developed. The 65kDa *M. leprae* antigen-encoding gene was inserted into the *M. smegmatis* genome by homologous recombination using an integrating shuttle vector pY6002, harbouring an *E. coli* origin of replication, an *E. coli* selectable marker gene (the ampicillin-degrading β-lactamase gene) and the *M. smegmatis pyrF* gene, which is responsible for the synthesis of uracil and allows selection in *M. smegmatis* (Husson *et al.*, 1990). Insertion of a foreign gene into the vector’s copy of pyrF results in integration of that gene into the chromosome on recombination with the chromosomal copy of pyrF, thereby inactivating the pyrF gene. A similar approach was followed with inserting a KmR gene into the genome of *M. bovis* BCG by homologous recombination with the *uraA* gene (Aldovini *et al.*, 1993).

Another approach for the stable introduction of foreign genes into a mycobacterial
host involved the novel *E. coli*-mycobacteria shuttle vectors that utilize the site-specific integration system of mycobacteriophage L5. These vectors replicate in *E. coli* as plasmids, but integrate site specifically into the chromosomes of *M. smegmatis*, BCG, and *M. tuberculosis* (Lee et al., 1991). Yet another way was devised by construction of vector plasmids that carry a Km<sup>R</sup> gene flanked by two copies of the insertion sequence IS900 (England et al., 1991). These artificial transposons, pUS701 and pUS702, may have other foreign sequences next to the Km<sup>R</sup> gene, and can be introduced into *M. smegmatis*, where they integrate into the chromosome in the same way as a natural transposon (Dellagostin et al., 1993). This gene replacement technology should be valuable for investigating mycobacterial pathobiology and for the development of candidate mycobacterial vaccine vehicles.

1.5 Gene regulation

1.5.1 RNA polymerase

The initiation of transcription is the main point for the regulation of gene expression in prokaryotes. Initiation requires the core RNA polymerase (which is composed of two α chains, one β chain and one β’ chain (Chamberlain, 1982)), σ factor, the appropriate nucleoside triphosphate (usually a purine), and a specific promoter. The specific recognition of promoter sites by RNA polymerase, largely mediated by the associated sigma factor, plays an important role in the control of DNA transcription. The sigma factor binds directly to the promoter sequences (McClure, 1985; Helmann and Chamberlain, 1988) and is released from the elongation complex, after transcription is initiated (Hansen and McClure, 1980). The interaction of RNA
polymerase with promoter DNA has been extensively studied in *E. coli* and the promoter consensus sequence at the -10 (TATAAT) and -35 (TTGACA) regions which are usually separated by 17-19 bases is well known for *E. coli* (Harley and Reynolds, 1987). This is a typical consensus prokaryotic promoter that is recognized by RNA polymerase when the enzyme is combined with sigma 70 factor.

Other factors important in the initiation of transcription include protein factors, ions, environmental factors and DNA supercoiling (Mishra *et al.*., 1990 and 1991). An important regulatory mechanism is a shift in the σ factors in the cell: the elements that select the set of transcription start sites (promoters) recognized by each RNA polymerase.

Recent analyses have suggested that the majority of *E. coli* promoters fall into three basic categories; those recognized by Eσ\(^70\), those recognized by Eσ\(^54\), and those recognized by Eσ\(^32\) (Collado-Vibes *et al.*., 1991; Garalla, 1990). In contrast in *B. subtilis*, a Gram-positive bacterium, nine types of sigma factors have been reported (Gitt *et al.*, 1985; Fujita *et al.*, 1989) that are used for gene expression at different stages of growth and development (Wang *et al.*, 1984). In *Streptomyces* seven different RNA polymerase holoenzymes have been found (Buttner, 1989). A small group of promoter sequences in this genus were similar to those recognized by the *E. coli* σ\(^70\) subunits, and a potential group of promoter sequences containing a guanine-rich sequence (GGGGGG) as the -35 region and TTGACA as the -10 region has been described (Strohl, 1992).
1.5.2 Regulatory proteins

Transcription initiation is also regulated by regulatory proteins that bind upstream from the transcription start site. They can be either repressors, for example the LexA repressor that binds upstream of the *E. coli recA* gene and prevents transcription (Little and Mount, 1982), or activators eg. catabolite activator protein (CAP) from *E. coli* that stimulates transcription of the genes encoding proteins involved in lactose metabolism. In the presence of cAMP, CAP binds to a site upstream from RNA polymerase to stimulate transcription (de Crombrugghe and Paston, 1978), possibly by exposing a strong binding site for RNA polymerase through bending the DNA (Gartenberg and crothers, 1988; liu-Johnson *et al.*, 1986).

1.5.3 Gene regulation in mycobacteria

Studies have shown that while mycobacterial translational signals seem to function in *E. coli*, transcription initiation signals functional in this heterologous background were isolated very infrequently (Burlein *et al.*, 1994). Some mycobacterial genes have been shown to be expressed in *E. coli* from their own promoters, particularly those of heat shock proteins, eg. the 65 kDa antigen of *M. tuberculosis* (Shinnick, 1987), BCG (Thole *et al.*, 1987) and *M. leprae* (Mehra *et al.*, 1986), but also including some other proteins such as the biotin carrier protein of several species (Collins *et al.*, 1987) and the *M. tuberculosis* 38 kDa antigen (Anderson *et al.*, 1988). In addition the 16S rRNA gene of *M. leprae* (Sela and Clark-Curtiss, 1991) *M. tuberculosis* (Ji *et al.*, 1994) and *M. bovis* (Suzuki *et al.*, 1991) are expressed
There is little information on specific mycobacterial promoters and only a few mycobacterial promoters appear to be active in *E. coli*. Therefore analysis of potential promoter sequences of mycobacterial genes by computer algorithms based on prokaryotic consensus sequences is only partially helpful in predicting promoter sequences (Dale and Patki, 1990). Apparently where transcriptional start sites have been mapped, e.g. for the BCG *hsp60* gene (Stover *et al*., 1991) the promoter shows at best weak homology to *E. coli* consensus sequences. Stover (1991) showed that this promoter is stress regulated *in vivo*. Apart from heat shock proteins very little is known at the molecular level of how the level of expression of specific mycobacterial genes may be modified by external conditions. Under phosphate starvation, induction has been demonstrated for the 38 kDa protein of *M. tuberculosis* (Anderson *et al*., 1990) and sequence homology has revealed a potential *pho* box homologous to that of *E. coli* which may be involved in control (Dale and Patki, 1990). It has also been proposed that the 28 kDa antigen of *M. leprae* is an iron regulated protein based on the observation that the -35 region upstream from the start site contains a potential iron control element (Dale and Patki, 1990).

1.6 The SOS response

Prokaryotic cells have several kinds of systems to deal with damage to DNA. Bacterial responses to DNA damage are highly conserved (Walker, 1984). One system, the SOS response, involves the coordinately induced expression of over 20
Figure 1.1: Roles of RecA protein in regulation of the SOS response. (Colston and Davis, 1994).
genes through a common regulatory mechanism in which RecA and LexA are the key players (Fig. 1.1). This leads to increased synthesis of proteins involved in DNA repair, DNA synthesis, homologous and site-specific recombination, and cell division, leading to an increase in cell survival.

1.6.1 Roles of RecA protein

The RecA protein, found in most bacteria, is a key enzyme in homologous recombination (Fig. 1.2) and plays a central role in the regulation of the SOS response. This protein, in addition to its regulatory function, is also involved in DNA repair, chromosomal replication, mutagenesis and induction of phage lysogens (Kowalczykowski et al., 1994).

1.6.2 Homologous recombination

RecA protein is important for homologous recombination in E. coli. It catalyses strand transfer between homologous DNA molecules in an ATP-dependent reaction. This process is a complex, multistep series of events involving binding of RecA to single stranded DNA, pairing of the RecA:ssDNA complex with double stranded DNA (Colston and Davis, 1994) to form a nucleo-protein complex, pairing of the homologous regions of the chromosomes, and finally exchange of their strands. The strand transfer reaction requires that RecA recognizes both single and double-stranded DNA, searches for a complementary region between the two, unwinds the duplex molecule, hybridizes the single strands, and also binds and hydrolyses ATP
Figure 1.2:

Homologous recombination as a laboratory tool.
Homologous recombination can be used to delete genes by incorporating a disrupted gene, in which case the gene is usually an antibiotic resistance gene or other selectable marker (a), or to introduce a mutation by double crossover events (b). (Colston and Davis, 1994).
1.6.3 Homologous recombination in mycobacteria

The genetic analysis of mycobacteria has been hampered by a lack of efficient tools for generating defined mutants by homologous recombination (Jacobs, 1992). Homologous recombination has been reported in the fast-growing mycobacterium, *Mycobacterium smegmatis* (Husson *et al.*, 1990). This method has proven particularly inefficient in *M. tuberculosis* complex where homologous recombination is less frequent than random illegitimate recombination (Pellicic *et al.*, 1996b). The unusual structure of the *M. tuberculosis* complex RecA which contains an additional internal sequence (intein) that has to be removed by protein splicing mechanisms (Davis *et al.*, 1991; 1992; Colston and Davis 1994; Davis *et al.*, 1994), might cause this inefficient homologous recombination. There have recently been some successful reports in identifying allelic exchange with low frequency in slow-growing mycobacteria (Marklund *et al.*, 1995; Norman *et al.*, 1995; Reyrat *et al.*, 1995; Balasubramanian *et al.*, 1996). The ability to perform allelic exchange in *M. tuberculosis* would allow not only studies of gene function but also the generation of new vaccine strains having precisely defined mutations (Balasubramanian *et al.*, 1996).

1.7 Regulation of the SOS response

Exposure of *E. coli* to agents or conditions that either damage DNA or interfere
with DNA replication results in the increased expression of genes that are members of the SOS regulatory network (Walker 1984). Under normal conditions (uninduced) LexA binds to a specific DNA sequence, the SOS box, upstream of the genes it regulates and represses transcription (Little and Mount, 1982). When cells are treated with agents that induce the SOS response, such as UV irradiation or mutagens, an inducing signal is produced, possibly single stranded DNA (D’Ari, 1985), that activates RecA. Activated RecA exhibits a coprotease activity, which then triggers the autocatalytic cleavage of LexA (Little, 1984, 1991) at an alanine-glycine peptide bond near the middle of the protein to yield two proteolytic fragments (Walker, 1984). This cleavage results in the release of LexA from its binding sites and increased expression of the genes it regulates, including recA and lexA themselves (Little et al., 1994). After DNA damage, the transcription rate of the lexA gene is high, although the concentration of LexA repressor is low due to its rapid cleavage. When DNA synthesis is restored, the ability to cleave LexA falls, presumably due to the disappearance of the effector, and the high transcription rate leads to a rapid accumulation of repressor and shut-off of the SOS operon (Little, 1983).

The inducible response to DNA damage (the SOS response) has been extensively studied in E. coli and other closely related Gram-negative bacteria (Little and Mount, 1982; Walker, 1984). For E. coli, the consensus sequence of the SOS box is taCTGTatata-a-aCAGta (Walker, 1984; Wertman et al., 1985). It has been suggested that the major recognition element for the LexA repressor of E. coli comprises the CTGT motifs of the SOS box (Schnarr et al., 1991; Ottleben et al.,
1991). Many of the recA genes isolated from Gram-negative bacteria have an SOS box like that of E. coli. Nevertheless, some such as Agrobacterium tumefaciens (Wardham et al., 1992) have no E. coli like SOS box, while a different sequence (GAACNNNNGTTC) termed the Cheo box is found upstream of DNA damage inducible genes in the Gram-positive bacterium Bacillus subtilis (Cheo et al., 1991).

1.8 The SOS response in mycobacteria

Some mycobacteria are intracellular pathogens which survive and replicate in macrophages, part of the normal host defence mechanism. In this environment the bacteria would be exposed to conditions such as hydrogen peroxide and reactive nitrogen intermediates which would be expected to lead to DNA damage. Therefore, the response of pathogenic mycobacteria to DNA damage could be important in pathogenesis.

1.8.1 Unusual structure of the recA gene of pathogenic mycobacteria

The key regulatory element of the SOS response, the recA gene, of M. tuberculosis has been cloned and characterized (Davis et al., 1991). It was shown that M. tuberculosis RecA is very unusual in that it is produced from a large (85 kDa) precursor protein by excision of a central, 47 kDa protein and ligation of the flanking region by a process termed protein splicing (Fig. 1.3). Subsequently it was shown that the M. leprae RecA protein is also produced by a protein splicing reaction (Davis et al., 1994). However the protein splicing sequences (the inteins)
Figure 1.3: Production of the mature *M. tuberculosis* RecA protein by post-translational splicing. (Colston and Davis, 1994)
of *M. leprae* and *M. tuberculosis* are quite different and are inserted at different loci within the *recA* genes indicating that they have been acquired independently (Colston and Davis, 1994). All other mycobacterial *recA* genes investigated so far are normal suggesting that protein splicing in the maturation of RecA has been selected for in the human pathogens (Davis *et al.*, 1994).

Although only a few examples of protein splicing have been reported, it has been shown to occur in unrelated genes in phylogenetically diverse species. Thus protein splicing has been shown to occur in a vacuolar pump ATPase subunit (VMA1) of the eukaryotes *Saccharomyces cerevisiae* (Hirata *et al.*, 1990; Kane *et al.*, 1990) and *Candida tropicalis* (Gu *et al.*, 1993), in RecA of the eubacteria *Mycobacterium tuberculosis* (Davis *et al.*, 1992) and *Mycobacterium leprae* (Davis *et al.*, 1994) and in the DNA polymerase of the archaea *Thermococcus litoralis* (Perler *et al.*, 1992; Hodges *et al.*, 1992) and *Pyrococcus sp* (Xu *et al.*, 1993).

### 1.8.2 The common features of inteins

Although the sequences of the known inteins are, on the whole quite different, they share a number of common features. There is a varying degree of conservation at the intein-extein junctions (splice sites). At the intein-C extein junction of the *M. tuberculosis* RecA and *S. cerevisiae* VMA1 proteins, there is a six amino acid motif which is identical. This C-terminal splice site also shows a high degree of conservation in other examples of protein splicing. Conservation at the N-extein-intein junction is less obvious (Colston and Davis, 1994).
In addition to the conserved motifs at the splice sites, a second pair of motifs located within the intein sequences has been identified. These motifs are characteristic of a group of endonucleases which include the HO endonuclease of yeast, the homing endonucleases of group I mobile RNA introns, and a number of independent endonucleases (Nakagava et al., 1992). Indeed site specific endonuclease activity has been found in the inteins from VMA1 (Gimble and Thorner, 1992; Bermer et al., 1992), *T. litoralis* intein-2 (Perler et al., 1992) and *Pyrococcus* sp. (F.Perler, unpublished results).

A common characteristic of this type of endonuclease is that they mediate gene conversion by initiating recombination at their cleavage sites. The endonucleases encoded by group I introns cut DNA at or near the site at which their own intron DNA has become inserted. When the intron is present, the endonuclease cannot cut the DNA; however if the intron is absent the DNA is cut and repaired using intron-encoding DNA as the template. The "invasion" of an intronless gene by the intron, as described above, is known as homing. VMA1 and *T. litoralis* intein 2 have both been shown to be site specific endonucleases which cut intein-less DNA at the site of insertion of the intein-encoding DNA. Moreover, the VMA1 intein has been shown to mediate homing or gene conversion of an intein-less allele (Gimble and Thorner, 1992).

**1.9 Aims**

Nothing was known about the regulation of the SOS response in mycobacteria,
although the \textit{recA} genes of \textit{M. tuberculosis} and \textit{M. leprae} have been cloned and sequenced (Davis \textit{et al.}, 1991, 1994). Both \textit{M. tuberculosis} and \textit{M. leprae} possess sequences homologous to the \textit{B. subtilis} SOS box upstream of their \textit{recA} genes, but whether these sequences are involved in regulation of the mycobacterial \textit{recA} genes remains to be established. The main aim of this project was to study the regulation of \textit{M. tuberculosis recA} gene expression, in particular to identify the promoter and regulatory elements involved in the expression.

There were a number of reasons for carrying out this work. Firstly, the SOS response makes an important contribution to the survival of bacteria in hostile environments such as \textit{M. tuberculosis} might be expected to encounter during infection of its host. Secondly, nothing was known about the regulation of this system in mycobacteria. Thirdly, the key regulatory elements of this response in the best studied system, \textit{E. coli}, are RecA and LexA which are themselves part of the SOS regulon. Thus the regulation of \textit{recA} expression can be studied as a model of DNA damage inducible genes of the SOS response.

Several questions of interest arise. (i) Is \textit{M. tuberculosis recA} expression inducible? (ii) If this is the case, what upstream sequences are required? (iii) Is a LexA homolog involved in regulation of \textit{M. tuberculosis recA} expression? (iv) If so, what is the recognition site for LexA in \textit{M. tuberculosis}? Upstream of the \textit{M. tuberculosis recA} gene there is a sequence motif with homology to the Cheo box LexA regulatory site of \textit{B. subtilis} which may be involved, whilst there is no similarity to the SOS box of \textit{E. coli}.
As a first step toward investigating the mechanisms that regulate DNA damage inducible gene expression in *M. tuberculosis*, a construct in which the DNA sequence upstream of *recA* has been cloned into a reporter plasmid containing a promoterless chloramphenicol acetyl transferase (*CAT*) gene was used in *M. smegmatis*. To confirm the presence of an inducible promoter in the DNA upstream of the *M. tuberculosis recA* gene, the effect of DNA damage inducing agents on *CAT* expression was investigated. The optimal concentrations for RecA induction were determined for a variety of DNA-damaging agents, including ofloxacin, mitomycin C, nalidixic acid and UV irradiation. Using constructs of the CAT reporter system with varying lengths of *recA* upstream sequence I have been able to localize sequences required for the expression in the DNA upstream of the *M. tuberculosis recA* gene.

Studies of gene expression in mycobacteria are important because many mycobacterial genes including *recA* are not expressed in *E. coli* and so presumably have promoters which are not recognized by *E. coli* RNA polymerase but what these promoters are like is not known. Therefore primer extension analysis was used to indicate the transcription start site(s) of the *M. tuberculosis recA* and *lexA* genes and their putative -10 and -35 like promoter elements.

In addition the role of LexA in the regulation of the *M. tuberculosis* SOS response was investigated. The *M. tuberculosis lexA* was identified by probing a cosmid from the genome analysis project, cloned and sequenced. Its protein product has been overexpressed in *E. coli* and purified by means of a His-tag. The purified LexA
protein was used *in vitro* assays to investigate its function in the regulation of RecA and LexA by determining its binding sites.
2. MATERIALS AND METHODS

2.1 Plasmid preparation

2.1.1 Small scale plasmid preparation

Minipreps were prepared by the boiling method (Holmes and Quigley, 1981) or using a QIAGEN miniprep kit.

2.1.2 Large scale plasmid preparation

Maxipreps were prepared using the QIAGEN plasmid/cosmid purification kit.

2.1.3 Plasmid preparation from mycobacteria

5 to 10 ml of mycobacterial cell culture was grown in Dubos medium (overnight for fast growers, several days for slow growers). The cells were harvested and resuspended in 150 μl of GTE (25 mM Tris-HCl, pH 8, 10 mM EDTA, 50 mM glucose) containing 10 mg/ml of lysozyme, and incubated at 37°C, overnight. Two volumes of fresh 0.2 M NaOH, 1% sodium dodecyl sulphate was added, mixed by inversion, and incubated at 4°C for 10 minutes. 1.5 volumes of 5 M potassium acetate was added, mixed by inversion and incubated at 4°C for 10 minutes. The lysates were centrifuged in a microcentrifuge for between 30 minutes and 1 hour. The supernatant was removed and extracted with an equal volume of CPI (chloroform-phenol-isoamyl alcohol, 24:25:1). The aqueous phase was transferred to a clean tube, two volumes of room temperature ethanol was added and the
mixture was incubated for 5 minutes at room temperature. The DNA pellet was spun down, washed with 70% ethanol and dried briefly under reduced pressure, resuspended in 30 µl of TE buffer, and incubated at 65 °C for 15 minutes (Jacobs et al., 1991). Because of the low yield, 5-10 µl of dissolved DNA was transformed to E. coli strains, DH5 or DH5-α. Subsequently the DNA was extracted from the transformed E. coli.

2.2 Cloning

2.2.1 Restriction enzyme digestion and alkaline phosphatase treatment

Restriction enzyme digestions of DNA were carried out using a ten-fold excess of the required enzyme in the buffer specified by the manufacturer (New England Biolabs) for 2-3 hr at 37°C. Following digestions of vector DNA for subcloning, alkaline phosphatase (CIP; Boehringer-mannheim) was added in order to remove the terminal phosphate group, as described by Sambrook et al., 1989. This treatment prevents the religation of the vector unless an insert has been ligated into it.

2.2.2 Agarose gel electrophoresis

Electrophoresis in agarose gels was used to separate, identify and purify DNA fragments, according to the methods described in Sambrook et al. (1989). Several different sized gel systems were used with a variety of different comb sizes. Various concentrations of agarose were used to pour gels, depending on the size of the DNA fragments to be resolved; the gel thickness was kept to 3-4 mm. Ethidium bromide
was included in the gel itself at a final concentration of 0.25 \( \mu g/ml \) in order to visualize the DNA. Prior to electrophoresis 1/5 volumes of loading buffer (0.25% Bromophenol blue, 0.25% xylene cyanol FF, 30% v/v glycerol) was added to each sample. Electrophoresis was carried out in a Tris-acetate buffer system (TAE 1 x solution: 40 mM Tris acetate pH7.7, 1 mM EDTA). After electrophoresis gels were placed on an ultraviolet transilluminator (300 nm, ultra-violet products, INC). The ethidium bromide labelled DNA bands were visualized and photographed using a Foto/Analyst camera (Fotodyne) and printed from a videocopy processor (Mitsubishi).

2.2.3 Purification of DNA fragments by agarose gel electrophoresis

After electrophoresis in a TAE (40 mM Tris-acetate pH 7.7, 1 mM EDTA) buffer system, DNA was extracted from the agarose gel slice using the glass milk binding method following the manufacturer’s instructions (Bio 101 Geneclean kit).

2.2.4 Ligations

Ligations were set up in a total volume of 10 \( \mu l \) containing 25-100 ng of vector DNA (depending of the vector used), 50-200 ng of insert DNA, 1 \( \mu l \) of 10 x commercial ligation buffer (Boehringer-Mannheim), 1 unit of T4 DNA ligase (Boehringer-Mannheim) for cohesive-end ligation or 5 units for blunt-end ligation. The reactions were left at 15\( ^\circ \)C overnight for blunt end ligation or a few hours for cohesive-end ligations.
2.2.5 Transformation of *E. coli*

Competent *E. coli* cells were prepared using an adaptation of the method described by Cohen *et al.* (1972). A single colony of *E. coli* (maintained on minimal medium for strains carrying an F') was inoculated into 5 ml of LB broth (Luria-Bertani, see appendix A) and grown overnight at 37°C. A 1/100 dilution of culture was used to inoculate 50 ml LB broth (in a 250 ml conical flask) and grown at 37°C with rapid rotary shaking (250 rpm) until the optical density of the culture at 650 nm was between 0.3-0.4. The cells were chilled on ice for 10 minutes prior to harvesting by centrifugation (4°C, 2000 x g for 10 minutes). The supernatant was decanted and the cells were resuspended in 20 ml of ice cold 0.1 M CaCl₂. The cells were then sedimented by centrifugation and resuspended in 2 ml of 0.1 M CaCl₂. The suspension was stored on ice for at least 2 hr before 100 μl aliquots were used for each transformation. Between 5 ng-100 ng of DNA was added to each aliquot of cells and the mixture left on ice for a further 30 minutes. The cells were heat-shocked at 42°C for 90 seconds and then 1 ml of LB broth and 10 μl of 1 M glucose was added to the mixture, which was incubated at 37°C for 1 hr. The cells were spun briefly, all but approximately 100 μl supernatant was removed and the cells resuspended. Transformed cells were plated on to selective media in dilutions which gave 100-300 colonies per plate.

When pUC19 or pTZ18R was the cloning vector and the host was TG2 or DH5α plates contained 50 μg/ml ampicillin + 50 μg/ml methicillin, 0.2 M isopropyl-β-D-thiogalactopyranoside (IPTG) and 40 μg/ml 4-bromo-3-chloro-2-indolyl-β-
galactoside (X-Gal). This combination permits blue-white screening: colonies with recombinant plasmids are white and parental vector transformants blue. For the other vectors, selection was based on resistance to 50 μg/ml ampicillin + 50 μg/ml methicillin or 50 μg/ml of kanamycin.

2.2.6 Transformation of mycobacteria by electroporation

*M. smegmatis mc²155* was grown in 5 ml Lemco broth (see appendix A) at 37°C for 24 hour (in the case of *M. tuberculosis* this step was omitted), then subcultured 1/250 into Dubos A+B media (see appendix A) containing 0.2% glycerol and incubated at 37°C for 24 hour (final OD at 600 nm approx. 1). The culture was incubated on ice for 1 hour before harvesting by centrifugation at 7000 rpm for 10 minutes. The cell pellet was washed by suspension in 10% glycerol (1/3 of volume) and incubated on ice for 25 minutes, then recentrifuged and finally resuspended in 10% glycerol (1/10 of volume). The cell suspension was incubated on ice for two hours before use. 1 μg of DNA was electroporated into 0.4 ml of cells using a 0.2 cm gap cuvette electrode (25 μF, 2.5 kV, 1000 Ω), this was immediately added to 4.6 ml of prewarmed Dubos A+B+0.2% glycerol and incubated at 37°C for 3 hours. Finally 100 μl of the samples were plated out on 7H11 agar media containing the appropriate antibiotic.
2.3 Southen hybridization

2.3.1 Probe labelling

DNA was labelled by random primer extension with the ECL DNA labelling kit (Amersham) following the manufacturer's instruction.

2.3.2 DNA transfer and hybridization

After electrophoresis, DNA was Southern blotted to Hybond N membranes overnight. The filters were prehybridized at 60 °C for 1-4 hour in hybridization mix (5 x SSC, 0.5% blocking reagent, 0.1% SDS, 5% Dextran sulphate, 100 mg/ml herring sperm DNA). The denatured labelled probe was added and left to hybridize at 60 °C overnight. The filters were washed twice (2 x SSC, 0.1% SDS) for 30 minutes at 60 °C. Then blocking, antibody incubation and washing steps were performed following the manufacturer's instructions.

2.4 Cloning, expression and purification of *M. tuberculosis* LexA

2.4.1 PCR for cloning in pET-15b

The oligonucleotide primers, Nde.lex and Bam.lex complementary to the construct pFM16 containing the complete *M. tuberculosis* lexA gene (EMBL accession number, X91407), were obtained. Nde.lex and Bam.lex were designed to contain NdeI and BamHI recognition sites respectively (Chapter 7, section 3.1). PCR reactions were set up using the Expand High Fidelity PCR system (Boehringer-
Mannheim) in a final volume of 100 μl in 0.5 ml microcentrifuge tubes. Each reaction contained two master mixes. Master mix one contained dNTPs, primers (upstream and downstream), 15 ng of template DNA (e.g pFM16), made up to 50 μl with sterile distilled water. Master mix 2 contained 2 mM MgCl₂, 10 μl of 10 x Expand HiFi buffer, 1 μl of enzyme mix (containing thermostable Taq and Pwo DNA polymerases), made up to 50 μl with sterile distilled water. Then the two master mixes were pipetted together on ice (giving final concentration of 200 mM dNTPs and 300 mM primers) in thin walled PCR tubes, mixed and overlayed with 50 μl of mineral oil (Sigma), and placed into the Omnigene temperature cycler (Hybaid).

The temperature cycle used was as follows: an initial 5 minutes at 94°C to denature high G+C DNA, then 10 cycles of 30 seconds at 94°C as the denaturation step, an annealing step of 37°C for 30 seconds and an extension temperature of 72°C for 1.5 minutes. Then 20 cycles of 30 seconds at 94°C, an annealing step of 37°C for 30 seconds and an extension temperature of 72°C for 1.5 minutes + 20 seconds per cycle and finally a last extension step of 72°C for 7 minutes to complete primer extension.

Samples (10 μl) of each reaction were analysed on a 1% agarose gel. For subcloning, the PCR fragment was digested with BamHI and NdeI and purified from a 1% agarose gel by using Geneclean (see section 2.2.3), then ligated into vector pET-15b (cut with BamHI, NdeI).
2.4.2 IPTG induction using *E. coli* host strain BL21 (DE3)

BL21 (DE3) containing pFM18 was grown overnight at 37°C in 5 ml of LB broth containing 200 µg/ml ampicillin (in the presence of pLysS or pLysE, 34 µg/ml of chloramphenicol was added as well). A 1/100 dilution of culture was used to inoculate 50 ml of LB broth containing the above antibiotics (in a 250 ml conical flask) and grown at 37°C with rapid rotary shaking (250 rpm) until the optical density of the culture at 600 nm was about 0.5. The cells were induced with 0.1 mM IPTG (it is very important to add 200 µg/ml ampicillin at this stage) and grown at 37°C for 3 hr. The cells were harvested, resuspended in 1 ml of sonication buffer (50 mM phosphate pH8, 1 M NaCl) and broken in a mini bead-beater (Bio Spec.products) using 150-212 micron glass beads (Sigma) at 3400 rpm for 1 minute. For large scale preparation the cells were sonicated at 100 W 4 times for 30 seconds (4 x 30 sec). After centrifugation 10 µl of the supernatant was analysed by SDS-polyacrylamide gel electrophoresis.

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

Electrophoresis of proteins were carried out in polyacrylamide gels according to the method of Laemmli (1970). The gel system used was the Model SE250 (Hoeffner Scientific) which has gel dimensions of 10 x 8cm x 1.5mm thick. Electrophoresis was carried out at 40mA for about 40 minutes, with cooling.
2.4.3.1 Preparation of gels

Acrylamide was obtained as a 30% stock, acrylamide/ bis-acrylamide ratio 37.5:1 (PROTOGEL) and stored at 4°C. Resolving gel mixes were prepared from 3.75 ml 1.5 M Tris-Cl pH8.7, 7.5 ml 30% acrylamide stock and made up to 15 ml with water. Mixes were polymerised by the addition of 10 \( \mu l \) TEMED (Bio-RAD) and 75 \( \mu l \) of 10% (w/v) APS (Bio-RAD) (stored frozen immediately after preparation). After pouring, the resolving gel was overlaid with water-saturated butanol. Gels were left to set for 1 hr, washed free of butanol with deionized water and dried with 3MM paper (Whatman). A 3.9% stacking gel solution was then prepared containing 0.83 ml 1.5 M Tris-Cl pH6.8, 1.3 ml 30% acrylamide stock and made up to 10 ml with water. Polymerization of the stacking gel was initiated by addition of TEMED and APS as for the resolving gel, and the solution poured on top of the resolving gel. A comb was inserted and the gel left to polymerize for 1 hr. Samples were mixed with sample buffer (6 x sample buffer: 6 ml glycerol, 2 g SDS, 1.86 g DDT, 2.4 mg Bromophenol blue dissolved and made up to 20 ml, stored in 1 ml aliquots at -20 °C), heated in boiling water for 3 minutes and immediately loaded for electrophoresis. Low range protein molecular markers (Bio-RAD) ranging from 112 kDa to 20 kDa were also run. Electrophoresis was carried out in Tris glycine buffer (0.192 M glycine, 0.025 M Tris-Cl pH8.3, 0.1% SDS).

2.4.3.2 Rapid staining

After electrophoresis gels were fixed using 25% iso-propanol and 10% acetic acid.
Then proteins were visualised using Coomassie Brilliant blue R-250 (Sigma) (0.006% frozen, freshly made Coomassie Brilliant blue in 10% acetic acid). Gels were stained for 10-20 minutes and destained overnight in 5% (v/v) acetic acid, 7% (v/v) methanol. Gels were photographed on a light box (Kodak coldlight illuminator Model 3).

2.4.4 LexA protein purification

The LexA protein was purified using Ni-NTA resin (QIAGEN). The cell free extract (CFX) was adjusted to pH8 with sonication buffer (50 mM phosphate pH8, 1 M NaCl). The Ni-NTA resin was equilibrated with sonication buffer three times. Then the resin was incubated with the cell free extract being mixed by inversion using a turn-table mixer (Baird & Tatlock) at room temperature for 1 hr. After gentle spinning at 2000 rpm for 2 minutes, the supernatant was removed and the resin washed with sonication buffer 3 times. At each step some samples were kept for running on a polyacrylamide gel. The resin was further washed with wash buffer (50 mM phosphate pH5, 1 mM imidazole, 10% glycerol) 4-8 times, and the protein was eluted by adding 1 ml of elution buffer (50 mM phosphate pH8, 1 M NaCl, 0.25 M imidazole). This step was repeated with more elution buffer. All the samples from each step were run in a polyacrylamide gel along with the original CFX sample.

The final purification method used a column procedure. The Ni-NTA column (5cm x 150mm) was packed and washed and the cell lysate was applied slowly to the
column at approximately 1 ml/minute.

2.5 Sequencing

2.5.1 Sequencing reactions

Sequencing reactions were carried out using the Sequenase 2.0 kit (US Biochemical Corporation). The denatured DNA was annealed as follows: 7 µl DNA, 2 µl 5 x reaction mix and 1 µl appropriate primer were incubated at 65°C for 5 minutes, cooled gradually at room temperature and kept on ice. The labelling mix was diluted 5-fold with sterile distilled water. The sequenase® enzyme was diluted 8-fold in enzyme dilution buffer immediately prior to use. The single stranded binding protein from *E. coli* (3.1 mg/ml) was diluted 6-fold in enzyme dilution buffer. Once the template-primer reaction had reached a temperature below 30°C, the reaction was transferred onto ice (for each reaction in a microtiter plate four wells were labelled as A, C, G and T). 2.3 µl of the primer/template DNA was aliquoted to each well and kept on ice. The following reagents were then added to a tube on ice: 1 µl of 0.1 M DTT, 2 µl of sterile distilled water, 1 µl of diluted single stranded binding protein, 2 µl of diluted labelling mix, 0.5 µl of [α-35S] dATP, 2 µl of diluted sequenase (added last). After mixing thoroughly, 2 µl was added to each well and incubated on ice for 5 minutes.

2 µl of appropriate dideoxy NTP termination mix was transferred to each well (e.g ddATP termination mix to the well labelled A) and incubated at 45°C for 7 minutes. 4 µl of stop solution and 2 µl of proteinase K (0.05 µg/µl) were added to each well,
the contents were mixed thoroughly, incubated at 65°C for 20 minutes and stored on ice, ready to load on to a sequencing gel (they can be stored at -20°C for up to 1 week). When the gel was ready for loading, the samples were heated to 80°C for 3 minutes, and loaded immediately using 3-4 μl per lane.

2.5.2 Preparation and electrophoresis of sequencing gels

Sequencing reactions were run on denaturing polyacrylamide gels using a 58 x 21.5 cm gel apparatus (BASE RUNNER nucleic acid sequencer, International Biotechnologies, Inc.). The glass plates were thoroughly cleaned with Decon 90 (Decon laboratories Ltd), rinsed several times with distilled water followed by ethanol and left to dry. The notched plate was siliconised by spreading about 10 ml dimethyldichlorosilane solution (2% solution in 1,1,1-trichloroethane, BDH) in the fume hood using Kleenex towel paper and rinsed with ethanol. The plates were assembled using strips of Whatman 3MM paper as spacers and held in place with clips. Acrylamide was obtained as a 40% stock solution (38% acrylamide solution, 2% bis acrylamide, ACCUGEL™) and TBE buffer was prepared as a 10 x stock and diluted to 1 x as required (1 x TBE: 89 mM Tris borate, 2.5 mM EDTA, pH8). The gel solution consisted of 6% acrylamide, 7 M urea and 1 x TBE and was primed with 112 μl 25% (w/v) ammonium persulphate (APS) and 112 μl N,N,N’,N’-tetramethylethylenediamine (TEMED). The gel was poured and a comb fitted in place. Gels were left to polymerize for 1 hr before the gel apparatus was assembled.
Once the gels were in contact with the running buffer (1 x TBE) the combs were removed and the gels were prerun at constant power (60 W) using a power pack (Bio RAD model 3000/300) for 30 minutes. Sequencing reactions (4 μl) were carefully loaded using a Gilson P20 micropipette and "duck bill" tips (Sorenson, Bioscience, Inc) and then the gels were run at the same constant power as for the prerun. Samples were loaded in order A,C,G,T from left to right. For sequences from 0-200 bases the gel was run until the bottom dye reached the base of the plate and for sequences from 200-400 bases the gels were run until the second dye reached the base of the plate.

After electrophoresis the plates were separated with the gel remaining on the unsiliconised plate and fixed in 10% (v/v) methanol, 10% (v/v) glacial acetic acid for 15 minutes. The gel was then removed from the backing plate by overlaying it with a sheet of 3MM paper and carefully lifting this off. The gel was overlaid with "cling film" and dried under vacuum on a gel drier (Bio RAD laboratories) at 80°C for 1 hr. The dried gel was exposed in direct contact with X-ray film (X-OGRAPH BLUE) for 24-48 hr before the autoradiograph was developed (Fuji Automatic X-ray film developer). Autoradiographs were read using a gel reader (IBI Gel reader, International Biotechnologies Inc.) and sequences fed directly into the MacVector computer programme (Apple).
2.6 Identification of the transcription start site

2.6.1 RNA preparation

For fast growers 5 ml of mycobacterial cell culture was grown in Lemco media for 24 hr and then subcultured by 1/50 dilution into 200 ml of Lemco broth for 12-15 hr. *M. tuberculosis* was grown in Dubos medium until the OD at 600 nm reached 0.7. The cells were harvested for 10 minutes at 7000 rpm (HS-4 rotor), resuspended in 1 ml of guanidinium chloride buffer (6 M guanidinium chloride, 0.1% Tween 80, 10 mM EDTA, 1 mM 2-mercaptoethanol) and left at -20°C for 15 minutes or overnight. The cells were then broken in a mini bead-beater (Bio spec. products) using 150-212 microns glass beads (Sigma), at 3400 rpm for 3 minutes in total (after each minute, a minute rest on ice) and spun at 12000 rpm in a microcentrifuge for 5 minutes; the supernatant was transferred to a fresh tube, the pellet was resuspended in 300 μl of guanidinium chloride buffer again, bead-beaten for another 30 seconds and the supernatant added to the above supernatant. The combined supernatant was deproteinized by repeated extraction (3 times) with chloroform/isoamyl alcohol (24:1, v/v). RNA was precipitated by the dropwise addition of a 0.5 volume of ethanol, the sample was left at -70°C for 1 hr, then spun at 13000 rpm in a microcentrifuge for 30 minutes. The pellet was washed with 70% ethanol, resuspended in 1 x formaldehyde gel-running buffer (5 x formaldehyde gel running buffer: 0.1 M MOPS (pH7), 40 mM sodium acetate, 5 mM EDTA, pH8, Sambrook *et al.*, 1989), and kept at -70°C.
2.6.2 Primer extension

The oligodeoxyribonucleotide primers were end-labelled with [γ²³P]ATP at their 5’termini by means of T4 polynucleotide kinase as described in the primer extension kit (Promega). 0.1-1 pmole of the labelled primer was added to 5 μl of nuclease free water containing 40 μg of total RNA and 5 μl of avian myeloblastosis virus (AMV) primer extension 2 x buffer (Promega). The annealing step was performed at 42-65°C (based on the melting temperature of the primer being used) for 1 hr. 5 μl of AMV primer extension 2 x buffer, 1.4 μl of 40 mM sodium pyrophosphate and 1 unit of AMV reverse Transcriptase (Promega) were added to each reaction. The reaction mixture was incubated at 42°C for 1 hr, ethanol precipitated, washed with 70% ethanol and resuspended in 5 μl of Tris-EDTA pH7.5. 5 μl of loading dye (98% formamide, 10mM EDTA, 0.1% Xylene cyanol, 0.1% Bromophenol blue) was added. The extension products were heated at 90°C for 10 minutes, separated on a 6-8% polyacrylamide-urea gel and visualized by autoradiography.

2.7 DNA binding

2.7.1 Gel shift assay

Gel shift assays were carried out using a DIG Gel shift kit (Boehringer-mannheim). Complementary single-stranded oligonucleotides were annealed by denaturing at 95°C for 10 minutes, then cooling slowly to room temperature. The double stranded oligonucleotides or restriction fragments was diluted with TEN buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1 M NaCl, pH8) to 3-4 pmole/μl and labelled with
digoxigenin-11-ddUTP following the manufacturer’s instructions. The labelled oligonucleotide was diluted to a concentration of 15-30 fmole/μl in TEN buffer. The gel shift reaction was set up according to the manufacturer’s instructions with and without poly[d(I-C)] competitor DNA. Following incubation at room temperature for 15 minutes, the sample was applied immediately to a pre-electrophoresed native 6-8% polyacrylamide gel. The gel system used was a maxigel (Biometra) which has gel dimensions of 20cm x 20cm x 1mm thick. Electrophoresis was carried out in 0.25 x TBE buffer pH8, at 160 v at 4°C in a cold room. One lane was used to load sample buffer with bromophenol blue to follow the run. The dye was run 3/4 of the way to the bottom of the plate.

The gel was blotted onto a positively charged BM nylon membrane (Boehringer-mannheim) using a Semi-Dry Electroblotter A (ancos) at 400 mA for 1 hr. Then the membrane was soaked in 10 x SSC for 30 seconds and cross-linked for 1 minute at 300 nm with a transilluminator. The chemiluminescent detection was performed following the manufacturer’s instructions. The membrane was exposed to X-ray film (Hyper film-MP, Amersham) for 15-40 minutes at room temperature.

2.7.2 Footprinting

Footprinting was carried out using a method (Lin & shiu, 1995) based on the PCR technique and the dideoxy DNA sequence reaction. In this method a pair of primers were designed for producing the one-end labelled target DNA sequence by PCR. The oligodeoxyribonucleotide primers were end-labelled with [γ^32P] ATP at their
5' termini by means of T4 polynucleotide kinase as described in the primer extension kit (Promega), followed by ethanol precipitation and resuspended in water.

Two PCR reactions were performed. In one reaction the downstream primer was end-labelled and in the other reaction the upstream primer was end-labelled with \([\gamma^{32}P]\) ATP. The PCR was performed using an Expand High Fidelity PCR system kit (Boehringer-mannheim) in a final volume of 50 \(\mu\)l according to the manufacturer's instructions. 10 ng of template DNA (pEJ258) containing the upstream sequence of the recA gene and 1.5 mM of MgCl\(_2\) was used in the reaction. The temperature cycle used was as follows: an initial 2 minutes at 94°C to denature high G+C DNA, then 10 cycles of 15 seconds at 94°C as the denaturation step, an annealing step of 50°C for 30 seconds and an extension temperature of 72°C for 45 seconds. Then 20 cycles of 15 seconds at 94°C, an annealing step of 50°C for 30 seconds and an extension temperature of 72°C for 45 seconds + 20 seconds per cycle and finally a last extension step of 72°C for 7 minutes to complete primer extension.

The PCR product was purified by filtration through a Probind spin column (Millipore corporation). An optimal ratio of end-labelled DNA (10 ng of PCR product) and the LexA protein (0.3 \(\mu\)g) was prepared in a suitable buffer (refer to the gel shift assay, section 2.7.1). Due to a requirement for cations in the subsequent DNase I cleavage reaction, MgSO\(_4\) was added to a final concentration of 5 mM (1 \(\mu\)l of 100 mM MgSO\(_4\)). A concentration of 0.1 unit/\(\mu\)l DNase I (10
u/μl, Boehringer-mannheim) was prepared on ice, 1 μl added to the DNA/protein mixture, mixed carefully and left for 2 minutes at room temperature. 20 μl of the prewarmed (at 37°C) stop solution (200 mM NaCl, 20 mM EDTA, 1% SDS and 50 ng/μl tRNA) was added and vortexed. To inactivate the DNase I, the cleaved DNA fragments were heated at 90°C for 1 minute, purified by phenol/chloroform extraction and ethanol precipitation, washed with 70% ethanol (twice), and resuspended in 5 μl of sequencing gel loading buffer (95% v/v formamide, 20 mM EDTA, 0.05% Bromophenol blue 0.05% Xylene cyanol FF). The samples were heated for 10 minutes at 90°C and electrophoresed on a 6% polyacrylamide DNA sequencing gel. The same primer as that labelled in the protection assay was also used to perform the sequencing reactions of the double stranded DNA (Sequenase kit, US Biochemical corporation) which were run alongside the footprinting reaction.

2.8 Induction of RecA-CAT fusion in pEJ258, pFM3, pFM7 in M. smegmatis mc^2155

The inducing agents used were:

- ofloxacin (Sigma): 0.05, 0.12, 0.25, 0.5, 1, 1.5 μg/ml
- mitomycin C (Sigma): 0.2, 0.5, 1, 2.5, 5, 10, 30 μg/ml
- nalidixic acid (Sigma): 10, 25, 50, 100, 150, 200 μg/ml
- UV irradiation (0.1 Jm⁻²s⁻¹, Jencons): 10, 30, 50, 70, 90, 120 seconds
M. smegmatis mc²155 was grown in Lemco broth containing kanamycin (25 μg/ml), for about 36 hours, then subcultured 1/1000 in Dubos A + B + 0.2% glycerol and kanamycin (25 μg/ml), for about 24 hour. When the OD reached 0.6 to 0.7, the culture was divided into 10 ml aliquots to induce in various ways and incubated with shaking at 37°C for 5-6 hours. The bacteria were harvested, and washed in 5 ml PBS buffer, resuspended in 1 ml PBS, lysed using a mini bead-beater for 1 minute and spun for 3 minutes in a microfuge. The supernatant was taken for protein and CAT assays. The protein assay was done using a BCA protein assay kit (Pierce) and the CAT assay using a CAT-ELISA enzyme immuno-assay kit (Boehringer-Mannheim).

2.9 Bacterial strains and plasmids

The bacterial strains used in this work were M. smegmatis mc²155 (Snapper et al., 1990), E. coli strains TG2, DH5, and DH5-α (Sambrook et al., 1989) for cloning of mycobacterial DNA, and E. coli BL21 (DE3) (Studier et al., 1990) for expression of LexA protein. The media used for growing the cultures are given in the appendix.

In Table 2.1, the plasmids used in this work are explained and the promoter probe vectors pEJ257 and pEJ108 (Davis, personal communication) are illustrated in Figure 2.1.
<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFM3</td>
<td>1.3 kb</td>
<td>XhoI fragment of pEJ258 in EcoRV site of pEJ257</td>
</tr>
<tr>
<td>pFM6</td>
<td>1.3 kb</td>
<td>XhoI fragment of pEJ258 in SalI site of pUC19</td>
</tr>
<tr>
<td>pFM7</td>
<td>0.7 kb</td>
<td>HindIII-NotI fragment of pFM6 in HindIII-NotI site of pEJ257</td>
</tr>
<tr>
<td>pFM8</td>
<td>0.3 kb</td>
<td>SalI fragment of <em>M. tuberculosis</em> lexA gene in SalI site of pUC19</td>
</tr>
<tr>
<td>pFM9</td>
<td>1.5 kb</td>
<td>SalI fragment of <em>M. tuberculosis</em> lexA gene in SalI site of pUC19</td>
</tr>
<tr>
<td>pFM10</td>
<td>deletion of 0.8 kb SalI fragment from pFM8</td>
<td></td>
</tr>
<tr>
<td>pFM11</td>
<td>0.5 kb</td>
<td>SalI fragment of pFM8 in SalI site of pUC19</td>
</tr>
<tr>
<td>pFM12</td>
<td>0.7 kb</td>
<td>PvuII fragment of pFM8 in SalI site of pUC19</td>
</tr>
<tr>
<td>pFM14</td>
<td>0.7 kb</td>
<td>ApuIII-SalI fragment of pFM9 in SalI site of pTZ18R</td>
</tr>
</tbody>
</table>
pFM16 1.3 kb SalI fragment of pFM8 in 3 kb SalI fragment of pFM14

pFM17 0.35 kb PvuII-HindIII fragment of pFM6 in EcoRV-HindIII site of pEJ257

pFM18  NdeI-BamHI digested PCR product of lexA gene from pFM16 in NdeI-BamHI site of pET-15b

pEJ108  CAT promoter probe vector for use in mycobacteria based on pKK232-8 (Pharmecia) plus 2.6 kb EcoRV-HpaI fragment (myco-origin) of pYUB12 and kanamycin resistance gene from pUC4-KIXX (Pharmacia).

pEJ257  0.3 kb HindIII-BamHI superlinker fragment of pSL1180 in HindIII-BamHI of pEJ108, CAT promoter probe vector

pEJ258  4.2 kb NruI fragment of pEJ126 (Davis et al., 1991), upstream DNA and part of M. tuberculosis recA gene in EcoRV site of pEJ257

pEJ161  0.4 kb BamHI fragment of pEJ135 (Davis et al., 1991), upstream DNA and part of M. tuberculosis recA gene in BamHI site of pEJ108
pUC19: Yanisch-Perron et al., 1985

pTZ18R: Mead et al., 1986

pSL1180: Brosius et al., 1989

pET-15b: Studier and Moffat, 1986; Rosenberg et al., 1987, Studier et al., 1990
Figure 2.1. The promoter probe vector pEJ257
pEJ108 is the same as pEJ257 but with a different polylinker.
3. OPTIMIZATION OF THE CONDITIONS FOR
INDUCTION OF M. TUBERCULOSIS RECA

3.1 Introduction

Inducible DNA repair systems such as the SOS system have been extensively studied with the Gram-negative bacterium *Escherichia coli* (Little *et al.*, 1982) and to a lesser extent with the Gram-positive bacterium *Bacillus subtilis* (Cheo *et al.*, 1991). Regulation of the SOS system is controlled by the production of the RecA and LexA proteins. Nothing is known about the SOS response in mycobacteria, although the gene encoding the *M. tuberculosis* RecA (Davis *et al.*, 1991; 1992) has been identified. As a first step towards investigating the mechanisms that regulate DNA damage-inducible gene expression in *M. tuberculosis*, a reporter plasmid containing DNA from upstream of the *M. tuberculosis recA* gene was used to investigate the effects of DNA-damage inducing agents on its expression and to optimize the conditions required for induction of RecA expression.

3.2 Confirmation of the presence of an inducible promoter for the *M. tuberculosis recA* gene

In order to confirm that the complete *recA* fragment cloned from *M. tuberculosis* (Davis *et al.*, 1991) does contain an inducible promoter the 4.2 kb NruI fragment of pEJ126 (Davis *et al.*, 1991) was subcloned into a promoter probe vector (pEJ257) containing a promoterless chloramphenicol acetyl transferase (CAT) gene, giving
pEJ258 (E.O. Davis, personal communication). The effects of DNA-damage inducing agents on the CAT expression of this construct was investigated in *M. smegmatis* mc^2^155. Preliminary assays with pEJ258 which contained approximately 3 kb of upstream DNA and then with pFM3 containing 1.3 kb of upstream DNA using DNA damaging agents under arbitrarily chosen conditions of growth and duration of induction showed that expression was inducible.

### 3.3 Optimization of the conditions required for *M. tuberculosis recA* induction in *M. smegmatis*

Having established that expression of the *M. tuberculosis recA* gene is inducible by DNA damaging agents in *M. smegmatis*, the next step was to define the optimum conditions for induction of *M. tuberculosis* RecA expression in *M. smegmatis* mc^2^155, the effect of growth stage (Fig. 3.1) on levels of induction was studied. These experiments indicated that the highest levels of induction occurred if the inducing agent was added when the optical density of bacterial culture at 600 nm was 0.6-0.7.

Next the optimum concentration for various inducing agents was assessed. The gene was induced with a range of concentrations of ofloxacin, mitomycin C, or nalidixic acid or different times of uv irradiation (see Materials and Methods) followed by 6 hours incubation before being assayed. With ofloxacin, at concentrations ranging from 0.05 to 1.5 µg/ml, induction occurred at a concentration of 0.12 µg/ml and reached a maximum of about 2-fold induction at 1µg/ml but decreased at higher
Using mitomycin C, over the concentration range of 0.2-30 μg/ml, induction occurred at a concentration of 0.2 μg/ml and did not increase very much at higher concentrations up to 10 μg/ml (Fig. 3.3). During induction with concentrations of 5 μg/ml and above cell death appeared to occur and the cell death was particularly evident at levels above 10 μg/ml corresponding with the reduced levels of CAT seen in the assay. The induction ratio was 2.7 using 2.5 μg/ml mitomycin C. Therefore, concentrations of mitomycin C below 5 μg/ml were used, where the induction ratio was about 2.5 over the range 0.5-2.5 μg/ml.

With nalidixic acid, using concentrations from 10 to 200 μg/ml, induction occurred at a concentration of 100 μg/ml and reached a maximum of 2.5-fold induction at 150 μg/ml concentration before decreasing at higher concentrations (Fig. 3.4). With uv irradiation for times varying from 10-120 seconds, induction reached a maximum of 2-fold at 70-90 seconds but then decreased at higher doses (Fig. 3.5).

Finally the effect of duration of induction was investigated using a clone, pFM7, which contained 700 bp of upstream DNA. Induction was carried out for 1, 3, 5 and 7 hours using 1 μg/ml ofloxacin as the inducing agent, and the results are shown in Fig. 3.6. This experiment indicated that the optimum time of induction was 5 hours.

From this series of experiments the optimum conditions for further studies were
Extracts were prepared from *M. smegmatis* containing pFM3 (containing 1.3 kb upstream DNA) grown in Dubos medium to different optical densities at 600 nm then induced with 1 μg/ml final concentration of ofloxacin. After 6 hours the cells were harvested and the total protein concentration and the amount of CAT were measured to determine the specific amounts of CAT (ng CAT / mg total protein). The results are shown as the induction ratios corresponding to induced CAT units / uninduced CAT units. The results shown are the means of duplicate assays.
Figure 3.2

The effect of ofloxacin concentration on the induction of *M. tuberculosis* RecA in *M. smegmatis* as measured by CAT production

Extracts were prepared from *M. smegmatis* containing pFM3 grown in Dubos medium to 0.6-0.7 optical density at 600 nm then induced with 0.05, 0.12, 0.25, 0.5, 1.0, 1.5 µg/ml final concentration of ofloxacin. After 6 hours total protein concentration and the amount of CAT were determined and the results are shown as ng CAT / mg total protein. Each value is the mean of duplicate samples. The dashed line indicates the the level of expression in the absence of DNA damaging agents.
Figure 3.3

The effect of mitomycin C concentration on the induction of *M. tuberculosis* RecA in *M. smegmatis* as measured by CAT production

Extracts were prepared from *M. smegmatis* containing pFM3 grown in Dubos medium to 0.6-0.7 optical density at 600 nm then induced with 0.2, 0.5, 1, 2.5, 5, 10 and 30 µg/ml final concentration of mitomycin C. After 6 hours total protein concentration and the amount of CAT were determined and the results are shown as ng CAT/mg total protein. Each value is the mean of duplicate samples. The dashed line indicates the level of expression in the absence of DNA damaging agents.
Figure 3.4

The effect of nalidixic acid concentration on the induction of *M. tuberculosis* RecA in *M. smegmatis* as measured by CAT production.

Extracts were prepared from *M. smegmatis* containing pFM3 grown in Dubos medium to 0.6-0.7 optical density at 600 nm then induced with 10, 25, 50, 100, 150 and 200 μg/ml final concentration of nalidixic acid. After 6 hours total protein concentration and the amount of CAT were determined and the results are shown as ng CAT/mg total protein. Each value is the mean of duplicate samples. The dashed line indicates the the level of expression in the absence of DNA damaging agents.
Figure 3.5

The effect of uv irradiation on the induction of *M. tuberculosis* RecA in *M. smegmatis* as measured by CAT production

Extracts were prepared from *M. smegmatis* containing pFM3 grown in Dubos medium to 0.6-0.7 optical density at 600 nm then induced with uv irradiation (0.1 J m\(^{-2}\) s\(^{-1}\)) for 10, 30, 50, 70, 90, 120 s. After 6 hours total protein concentration and the amount of CAT were determined and the results are shown as ng CAT/mg total protein. Each value is the mean of duplicate samples. The dashed line indicates the the level of expression in the absence of DNA damaging agents.
Figure 3.6

The effect of time of induction on the expression of *M. tuberculosis* RecA in *M. smegmatis* as measured by CAT production

Extracts were prepared from *M. smegmatis* containing pFM7 (containing 700 bp upstream DNA) grown in Dubos medium to 0.6-0.7 optical density at 600 nm then induced with 1 μg/ml final concentration of ofloxacin for 1, 3, 5 and 7 hours. After each time point the total protein concentration and amount of CAT were determined and the results are shown as the ratio of induction (induced CAT units / uninduced CAT units). The results shown are the means of duplicate assays.
determined to be induction at an OD$_{600}$ of 0.6-0.7 for 5 hours with 1 $\mu$g/ml ofloxacin.

3.4 Discussion

The work in this chapter has helped to establish the following points:

1) The presence of an inducible promoter upstream of the $M. tuberculosis$ recA gene which was induced by DNA damaging agents with quite different modes of action: uv which causes the formation of pyrimidine dimers (Haseltine, 1983), mitomycin C which cross links DNA strands by alkylation (Lyer and Szybalski, 1963; 1964), and ofloxacin and nalidixic acid which both inhibit DNA gyrase (Viravau et al., 1996; Takkif et al., 1996; Sugino et al., 1977).

2) The optimum conditions required for induction of the $M. tuberculosis$ RecA expression in $A. fumigatus$ mc$_{155}$ in terms of the growth stage and the duration of induction.

3) The optimum concentrations of inducing agents required for the induction of $M. tuberculosis$ RecA in $M. smegmatis$ mc$_{2155}$.

The cloned fragment from upstream of the $M. tuberculosis$ recA gene containing the recA promoter was successfully used to drive expression of the CAT gene. Optimal induction of $M. tuberculosis$ RecA expression in $M. smegmatis$ mc$_{2155}$ occurred when the bacteria were grown in Dubos medium to 0.6-0.7 optical density at 600 nm and then incubated with the inducing agents for 5 hours. The optimum concentrations of DNA-damage inducing agents for induction were determined to
be 1 μg/ml for ofloxacin, 0.5-2.5 μg/ml for mitomycin C, 150 μg/ml for nalidixic acid and 70-90 seconds for uv irradiation.
4. IDENTIFICATION OF UPSTREAM DNA REQUIRED FOR M. TUBERCULOSIS RECA INDUCTION

4.1 Introduction

The presence of an inducible promoter upstream of the cloned M. tuberculosis recA gene has been confirmed and the optimum conditions required for induction of RecA expression investigated. This chapter describes the preparation of deletion constructs to narrow down the location of the inducible promoter of the M. tuberculosis recA gene and the sequence determination of this region.

4.2 Constructs to localize the inducible promoter of the M. tuberculosis recA gene

The work described in Chapter 3 shows that the constructs pEJ258 and pFM3 (containing 1.3 kb of DNA upstream of recA) contained an inducible promoter. To further localize the promoter within this region a series of clones was made containing progressively smaller amounts of upstream sequence (Fig. 4.1) preceding the CAT gene. The preparation of these constructs was as follows:

pFM3: 1.3 kb XhoI fragment of pEJ258 made blunt ended and ligated with pEJ257, digested with EcoRV and dephosphorylated.

pFM7: The 0.7 kb NotI-HindIII fragment of pFM6 (1.3 kb XhoI fragment of pEJ258 in pUC19) ligated with pEJ257 digested with NotI-HindIII.
Figure 4.1
Constructs prepared to locate the inducible promoter in the DNA upstream of the *M. tuberculosis recA* gene.

pEJ258 contains about 3 kb upstream of *M. tuberculosis recA*; pFM3, 1300 bp; pFM7, 667 bp; pFM17, 310 bp; and pEJ161, 269 bp.
pFM17: 0.35 kb *PvuII-HindIII* fragment of pFM6 ligated with pEJ257 digested with *EcoRV-HindIII*.

pEJ161 0.4 kb *BamHI* fragment of upstream DNA and part of *M. tuberculosis recA* gene ligated with pEJ108 digested with *BamHI*, dephosphorylated (E. O. Davis, personal communication)

Ligations for all these fragments were transformed into *E. coli* strain DH5-α. Then these constructs were electroporated into *M. smegmatis mc²155* for analysis of gene expression using the CAT assay.

4.3 CAT assays of mycobacterial transformants containing deleted constructs

Using the optimal conditions for induction defined above the constructs containing progressively smaller amounts of upstream DNA were investigated for expression level and inducibility. Uninduced and induced cultures were assayed for each construct and results are shown in Fig. 4.2. Activity for the control extracts from the parental strain *M. smegmatis mc²155* was essentially zero with no vector and with vector pEJ257 or pEJ108.

The initial constructs, pEJ258 (containing 3 kb of upstream DNA) and pFM3 (containing 1.3 kb of upstream DNA) showed 2-3 fold induction. When the amount of upstream DNA was further reduced to 667bp in pFM7, the induction ratio remained constant at 2.5. pFM17 containing even less upstream DNA (310 bp) showed a decrease in absolute expression level but an increase in induction ratio to 4.2. However removal of a further 41 bp from the 5’ end of the cloned DNA in
Figure 4.2

Induction of constructs containing DNA from upstream of the *M. tuberculosis* *recA* gene by ofloxacin in *M. smegmatis mc²155*

Extracts were prepared from *M. smegmatis* containing no vector (mc²155), parental vector (pEJ108, pEJ257) and recombinant clones. Transformants were grown as described in Materials and Methods, induced with ofloxacin (1μg/ml final concentration) for 5 hours and assayed for CAT. Total protein concentration was measured and the results are expressed as ng CAT / mg total protein. There was no detectable CAT for mc²155, pEJ108 and pEJ257. Each value is the mean of six cultures assayed in duplicate, except for the vector which is the mean of three cultures. Bars represent standard errors.
pEJ161 containing 269 bp of upstream DNA resulted in a dramatic reduction in the level of expression and the loss of induction. These results show that 310 bp of upstream DNA is necessary and sufficient for inducible expression of *M. tuberculosis* *recA* and either the promoter or other sequences essential for transcription were located in this region, either within this 41 bp or overlapping the cloning site used to make pEJ161.

### 4.4 Sequence determination of the *recA* upstream region

Although the 310 bp of upstream DNA which had been sequenced previously (Davis *et al.*, 1991) was sufficient for inducible RecA expression, the absolute levels of expression were significantly higher in a construct containing 677 bp of the upstream DNA (pFM7), so the additional 360 bp of DNA in this construct was sequenced. Based on the known sequence, the primer pro2 (5'-CGG ATC CGG CGT GCT CA-3') was designed to sequence further upstream and then based on the sequence obtained from primer pro2, primer pro3 (5'-GCC TGG GCG GCG CTA CG-3') was designed which enabled all of the fragment to be sequenced on one strand. Sequencing of the other strand was performed using primer pro4 (5'-TCA ACG AGG TGT TGT CG-3') and primer pro6 (5'-TTA TGG CTC GTC GGT GC-3') as designed from the sequence obtained on the first strand. Pro4 is located upstream of the *NotI* site in a region which is not shown in the upstream sequence in Fig. 4.4. In Fig. 4.3 the strategy for sequencing this DNA upstream of the *recA* gene is shown.
Figure 4.3

Restriction site map and sequencing strategy for the sequence upstream of the recA gene of *M. tuberculosis*

This figure shows the construct pEJ258 containing upstream DNA and part of the *M. tuberculosis recA* gene. The arrows indicate the positions of primers used for sequencing.
Figure 4.4

Complete nucleotide sequence of the 670 bp upstream of the *recA* gene of *M. tuberculosis* contained in pFM7.

The putative promoter elements -35 and -10 boxes are indicated. The primer rec2.ext used for primer extension experiments is underlined and the transcription start sites so identified are indicated by red circles (see Chapter 5).
4.5 A single Cheo box is located upstream of the *M. tuberculosis* recA gene

A motif homologous to the Cheo box of *Bacillus subtilis* had been identified previously 120 bp upstream of the *M. tuberculosis* recA gene (Davis *et al.*, 1991). (Fig. 4.4). In this study this sequence was shown to bind LexA repressor of *M. tuberculosis* (see Chapter 8). No further Cheo box sequences were found from analysis of the extra DNA sequenced here. Although a sequence with more limited homology to the Cheo box (6/8 matches) was found 270 bp upstream of recA, subsequent analysis by gel retardation and footprinting (Chapter 8) showed that this sequence did not interact with LexA.

4.6 Discussion

Examination of the putative promoter region of *M. tuberculosis* recA revealed no motifs like an *E. coli* SOS box, which has been highly conserved in a variety of Gram-negative bacterial species. Comparison of SOS boxes from *E. coli, Salmonella typhimurium, Erwinia carotovora, Pseudomonas aeruginosa* and *Pseudomonas putida* reveals that the consensus sequence CTG-N10-CAG is conserved in all known LexA operators of these Gram-negative bacteria (Garriga *et al.*, 1992) but some species such as *Agrobacterium tumefaciens* (Wardham *et al.*, 1992) have no *E. coli* like SOS box. It was shown that the *Anabaena variabilis* recA gene has no *E. coli*-like SOS box, but its expression is UV- and mitomycin C-inducible in *A. variabilis* itself (Owttrim and Coleman, 1989). In contrast, in *Thiobacillus ferrooxidans*, the recA gene has no SOS box and it is not inducible (Ramesar *et al.*, 1992).
Therefore, it was interesting to know whether a LexA-like protein exists in *M. tuberculosis*. A sequence identical to a consensus GAAC-N4-GTTC (Cheo box) common in various DNA-damage inducible promoters in *B. subtilis* (Cheo *et al.*, 1991) was found 120 bp upstream of the translation initiation codon in the *M. tuberculosis* *recA* DNA. Thus there may be inducible expression through the regulation of a LexA-like protein with a recognition sequence different from that of its *E. coli* counterpart. In Chapter 8, the function of this consensus sequence as an operator in *M. tuberculosis* *recA* expression will be discussed.

Using constructs of the CAT reporter plasmid with varying lengths of *recA* upstream sequence, the regulatory sequences required for inducible expression were localized to 310 bp of DNA upstream of the *recA* coding sequence (plasmid pFM17). This construct expressed *recA* under induced conditions at 4.2-fold the level of expression at uninduced conditions. This induction ratio is comparable to some of the isolated *E. coli* SOS (DNA-damage-inducible) promoters fused to the galactokinase (*galK*) gene (Lewis *et al.*, 1992) and to the expression of the *recA-lacZ* fusion of *E. coli* in several Gram-negative bacteria in the presence of DNA damage (Fernandez de Henestrosa *et al.*, 1991). The induction ratio for the expression of the *E. coli recA-lacZ* fusion in *E. coli* is 6.8 (Fernandez de Henestrosa *et al.*, 1991), and the maximum induction ratios calculated for the fused *din-galK* of *recA*, *polB*, *dinG* and *dinH* are 8.5, 6.5, 4.7, and 3.6, respectively (Lewis *et al.*, 1992).
In the Gram-positive bacterium *B. subtilis* the induction ratio of the expression of RecA protein in the presence of DNA-damaging agents is approximately 5-fold (Lovett *et al.*, 1988) as assessed from protein gels and 5.5-fold from using a recA-xylE fusion (Raymond-Denise and Guillen, 1992).

The construct pEJ161 which contains 269 bp of DNA from upstream of the *M. tuberculosis* recA coding sequence and includes the consensus repressor sequence, the Cheo box, did not show significant levels of CAT expression in either uninduced or induced conditions. This suggest that the 40 bp of DNA present in pFM17 but absent from pEJ161 contains a regulatory sequence essential for expression and induction of the *M. tuberculosis* RecA in *M. smegmatis* mc²155.

The results found with the deletion constructs analysed here for the *M. tuberculosis* RecA expression differ from those reported from an analogous deletion series for RecA in *B. subtilis*. Whereas progressive deletions led to the loss of expression with *M. tuberculosis* RecA, in *B. subtilis* there was no effect until the Cheo box was reached when constitutive expression was seen due to the loss of LexA repression (Cheo *et al.*, 1993). This could reflect either the promoter being located in a different position relative to the LexA repressor binding site or a different mode of regulation, i.e. some form of positive regulation, in *M. tuberculosis*.
5. TRANSCRIPTION START SITES OF

*M. TUBERCULOSIS* REC* A* GENE

5.1 Introduction

The DNA fragment upstream of the *recA* gene has been sequenced and constructs
with varying lengths of *recA* upstream sequence fused to a CAT reporter gene were
used to localize the region required for regulated expression to 310 bp upstream of
the RecA coding sequence. In particular, an approximately 50 bp region from 310
bp to 260 bp upstream of the *recA* gene was identified as containing sequences
essential for expression. This chapter describes the identification of the transcription
start sites of the *M. tuberculosis* *recA* gene and putative -10 and -35 like promoter
elements, to obtain an indication as to whether the promoter for the *recA* gene lay
in the region defined as essential for expression by the deletion analysis described
above.

5.2 Transcription start point (promoter region) of *M. tuberculosis* *recA*
expressed from a plasmid

The transcription start point of the *M. tuberculosis* *recA* gene was identified by the
primer extension method (see Chapter 2, section 2.6). Total RNA was extracted
from cultures of *M. smegmatis* mc²155 containing pFM7 (having 700 bp *recA*
upstream DNA) induced with ofloxacin (1 µg/ml) for 5 hours as described in
Chapter 2. This isolated RNA was used as a substrate with the primer CAT.ext (5’
TCC TGA AAA TGT CGT CGA AGG TGG GGG 3') complementary to the CAT gene which is transcriptionally fused to the recA control region in pFM7. Two products were obtained, one 47 bp and one 93 bp upstream from the translation initiation codon (Fig. 5.1), suggesting the presence of two promoters for the recA gene. Additional minor bands within a few bases of the extension product nearest the gene were also seen with the RNA from induced cells. Similar patterns have been reported previously with some other mycobacterial genes, eg. in BCG hsp60 (Stover et al., 1991). As a negative control RNA was extracted from M. smegmatis mc²155 containing the construct pEJ257 (parental vector) grown under the same conditions and the primer extension assay was performed using the above primer. No product was detected as shown in Fig. 5.1.

5.3 Transcription start point of M. tuberculosis recA expressed from the chromosome

Subsequently total RNA was extracted from M. tuberculosis itself, grown and induced as described in Chapter 2. The primer Rec2.ext (5' GGG ATG GGG GGT GTG GGT GAT 3') was designed complementary to the recA coding sequence and used in the primer extension assay with total RNA from both M. tuberculosis and from M. smegmatis mc²155 containing pFM7. The same two products for initiation of transcription were obtained with both substrates as shown in Fig. 5.2. The extension product ending nearer to the gene clearly gave the strongest signal using RNA from M. tuberculosis, suggesting that this promoter might be the most important. In order to assess whether other promoters existed further upstream
primer extension analysis was performed using two primers from different regions of the upstream sequence with both sources of RNA but no further products were identified (data not shown).

5.4 Sequences with homology to known promoters near the mapped transcription start sites

Inspection of the sequences near these mapped transcription start sites revealed some interesting similarities and differences from known E. coli consensus promoters. The main transcription start site was located 10 bp downstream of -10 (TCTAGT) and -35 (TTGTCA) elements similar to the -10 (TATAAT) and -35 (TTGACA) elements of E. coli σ70 promoters (Prinbow, 1975a, 1975b; Schaller et al., 1975; Takanami et al., 1976; Seeburg et al., 1977), but with a different spacing (Fig. 5.3). A second transcription start point was identified 8 bp downstream of -10 (CGGCTACTG) and -35 (TGTCACACTTGAA) sequences similar to those of E. coli σ32 heat shock promoters (Cowing et al., 1985) with consensus -10 (CCCCAT(Ta) and -35 (TNtCNCcTTGAA) sequences (Fig. 5.3). The features of these putative promoter elements are discussed in more detail in the following section.
Figure 5.1

Identification of the transcription start sites of the *M. tuberculosis* recA gene expressed from a plasmid

The figure shows primer extension experiments done with the labelled primer CAT.ext with: No RNA substrate (lane 1); 40 μg of RNA isolated from *M. smegmatis* mc²155 containing the parental vector pEJ257 (lane 2); 40 μg of RNA isolated from an uninduced culture of *M. smegmatis* mc²155 harbouring plasmid pFM7 (cloned upstream of recA) (lane 3); 40 μg of RNA isolated from *M. smegmatis* mc²155 harbouring plasmid pFM7 induced with ofloxacin at a final concentration of 1 μg/ml (lane 4); 40 μg of RNA isolated from *M. smegmatis* mc²155 containing pFM7 induced with mitomycin C at a final concentration of 2.5 μg/ml (lane 5). Lanes A, C, G and T represent sequencing reactions of the upstream region of the recA gene using pFM7 and primer CAT.ext. The arrows indicate the positions where transcription of the recA gene starts.
Figure 5.2

Identification of the transcription start sites of the *M. tuberculosis* recA gene expressed from the chromosome

The figure shows primer extension experiments, done with the labelled primer rec2.ext with: No RNA substrate (lane 1); 40 µg of RNA isolated from *M. tuberculosis* induced with ofloxacin at a final concentration of 1 µg/ml (lane 2); 40 µg of RNA isolated from *M. smegmatis mc²155* containing pFM7 induced with ofloxacin at a final concentration of 1 µg/ml (lane 3). Lanes A, C, G and T represent sequencing reactions of the upstream region of the recA gene of *M. tuberculosis* from pFM7 using primer rec2.ext. The arrows indicate the positions where transcription of the recA gene of *M. tuberculosis* starts.
Figure 5.3

Transcription start sites and putative promoter elements upstream of the *M. tuberculosis* recA gene

The DNA sequence upstream of the *M. tuberculosis* recA gene is shown starting from the sequence shown to be essential for expression by the CAT fusion analysis. The transcription start sites identified by primer extension are marked by red circles, and the primer rec2.ext used for some of these experiments is indicated by the arrow beneath the sequence. The putative promoter elements having homology to *E. coli* promoters identified near these transcription start sites are boxed in green, and the Cheo box is boxed in blue.
5.5 Discussion

The work described in this chapter has identified the transcription start points of the *M. tuberculosis recA* gene by means of the primer extension technique. Although it is possible that the extension products obtained result from processing of a larger transcript this seems unlikely as two products were obtained but only one of these appears to increase in abundance following exposure to DNA damaging agents. Also sequence elements resembling known promoters were identified near these proposed transcription start sites. Transcription initiated at a purine nucleotide (A) similar to the majority of the mycobacterial promoters as shown by Bashyam et al. (1996).

Previous work has shown that mycobacterial genes are not generally expressed in *E. coli* (Hopwood, 1988), possibly because their regulatory signals cannot be recognized by the RNA polymerase of the heterologous cells. The *M. smegmatis* system offers an alternative, promising approach for the study of the regulation of genes from slow growing mycobacteria (Bashyam et al., 1996). Here we have compared transcription start sites of the *M. tuberculosis recA* gene when it is expressed from a plasmid cloned into *M. smegmatis* and when it is expressed from the *M. tuberculosis* chromosome. In agreement with the observation that the efficiency, and specificity of transcriptional recognition in *M. tuberculosis* and *M. smegmatis* are conserved (Bashyam et al., 1996), two identical start sites were identified in both cases.

Although the *M. tuberculosis recA* gene is not expressed in *E. coli* from its own
promoter, a good match with the $\sigma$70 $E.\ coli$ consensus promoter was found upstream of the $recA$ gene, 10 bp away from the first and major transcription start site. This specific conserved sequence is present at nearly the same position upstream (-10) of the transcription start point for the most of the mycobacterial promoters investigated so far (Bashyam et al., 1996). These conserved sequences corresponded to the Pribnow box present in nearly all classical prokaryotic promoters. The first, second, and sixth bases (T, A, and T, respectively) of this conserved hexameric sequence which represent the functionally most important positions in $E.\ coli$ (Rosenberg and Court, 1979) and which are conserved in mycobacterial species (Bashyam et al., 1996), shows T, C, and T, respectively in the $M.\ tuberculosis\ recA$ promoter.

Although the sequences in the -35 regions of mycobacterial promoters do not bear any homology with the TTGACA motif present in promoters of $E.\ coli$ and several other bacteria (Bashyam et al., 1996), interestingly there is a similar motif (TTGTCA) centered in the -28 bp upstream of the $M.\ tuberculosis\ recA$. Significantly, however, these -35 and -10 like elements were separated by a different distance from that found in $E.\ coli$. Whereas most $E.\ coli$ promoter elements are separated by 16-19 bp, the separation of the $M.\ tuberculosis$ homologous elements was 9 bp which might explain the lack of expression of RecA in $E.\ coli$. Alternatively, the $M.\ tuberculosis$ RNA polymerase may recognise a different -35 sequence. It appears that in mycobacteria the -35 region can tolerate a larger variety of sequences than can those of $E.\ coli$ (Kremer et al., 1995, Bashyam et al., 1996). Similar flexibility in the -35 region of promoters has been reported from the
Streptomyces spp., resulting from the capacity to synthesize different sigma factors (Buttner, 1989; Westpheling et al., 1985). Therefore the heterogeneity in the -35 regions of mycobacterial promoters may result from having multiple sigma factors (Bashyam et al., 1996).

Analysis of the mycobacterial promoter sequences from positions -1 to -50 had shown that the promoters from M. tuberculosis have a higher GC content (57%) than the promoters from M. smegmatis (43%) (Bashyam et al., 1996). Interestingly the analysis of the M. tuberculosis recA promoter from the position -1 to -50 of the first transcription site point shows high GC content (58%).

In addition a sequence resembling that of σ32 E. coli heat shock promoters was found upstream of the second transcription start site (that was initiated at a purine (A) as well), about 100 bp upstream of the initiation codon. This product was fainter using RNA from M. tuberculosis compared with that from M. smegmatis carrying the clone suggesting that the promoter nearest to the gene could be the stronger one, although it has not been established whether these sequences function as promoters in mycobacteria.

LexA protein binds specifically to palindromic sites overlapping DNA-damage inducing gene promoters (Lovett et al., 1993), and the LexA binding sites in E. coli either overlap the -35 promoter region (consensus TTGACA) as in the case of the uvrA gene (Sancar et al., 1982), are situated between the -35 and -10 promoter region as in the case of the recA gene (Horii et al., 1980; Sancar et al., 1980),
overlap the -10 promoter region like the lexA gene (Brent and Ptashne, 1981; Miki et al., 1984), or occur downstream from the -10 promoter or even downstream from the +1 transcription start site as in the case of the uvrD gene (Easton and Kushner, 1983). Interestingly the consensus sequence SOS box (Cheo box) upstream of the M. tuberculosis recA structural gene is located similarly to the E. coli recA one, between the -10 and -35 promoter elements of the second recognized transcription start point that resembles heat shock promoter elements from E. coli. This overlap between the promoter and operator (LexA binding site) is typical of known SOS genes in general (Lewis et al., 1992) and suggests that LexA binding can affect binding of RNA polymerase.
6. CLONING, SEQUENCE DETERMINATION AND TRANSCRIPTION START SITE OF THE MYCOBACTERIUM TUBERCULOSIS LEXA GENE

6.1 Introduction

The SOS response in *Escherichia coli* and *Bacillus subtilis* is regulated by the molecular interaction of the LexA and RecA proteins (Walker, 1984; Wojciechowski, 1991). The LexA protein participates by repressing the many (ca.20) unlinked DNA damage-inducible (din) chromosomal operons that compose the SOS regulon including the SOS genes for LexA (*lexA*) and RecA (*recA*) themselves (Little, 1994). In mycobacteria the gene encoding RecA has been identified in *M. tuberculosis* (Davis *et al.*, 1991) and *M. leprae* (Davis *et al.* 1994) but nothing is known about the *lexA* gene.

This chapter describes:

i) Localization of the *M. tuberculosis* *lexA* gene on a cosmid

ii) Cloning of the *lexA* gene in the vector pUC19

iii) Sequence determination of the *lexA* gene and comparison with other bacteria at the protein level

iv) Identification of the transcription start site of *lexA*. 

95
6.2 Cloning of the *M. tuberculosis* *lexA* gene

6.2.1 Southern hybridization to a cosmid bearing *lexA*

Using a probe made by PCR, Southern blotting was performed to locate the *lexA* gene on a cosmid previously shown to include *M. tuberculosis* *lexA* (Philipp *et al.*, 1996). The blot consists of cosmid T616 digested with a variety of restriction endonucleases and probed with a 400 bp probe generated by PCR (both probe and T616 cosmid were kindly provided by Dr. S.T. Cole, Pasteur Institute, Paris, France). The results are shown in Fig. 6.1. A strong signal was obtained from just one fragment in most digests used. The 1.3 kb *SalI* fragment and 1.5 kb *SmaI* fragment were chosen for cloning in the hope that one of these would contain the intact *lexA* gene, since the coding region of other *lexA* genes is only about 600-700 bp.

6.2.2 Cloning of the *lexA* gene in pUC19

To clone the *lexA* gene, the 1.3 kb *SalI* and 1.5 kb *SmaI* hybridizing fragments were purified using a Geneclean kit and cloned in the *SalI* site and *SmaI* sites of pUC19 respectively (to give pFM8 and pFM9). Further constructs were made for use in sequencing as follows: plasmid pFM8 was digested with *SmaI* to give 0.8 kb and 3.1 kb fragments; the 3.1 kb *SmaI* fragment was religated to make pFM10, and the 0.8 kb fragment was Genecleaned and cloned in the *SmaI* site of vector pUC19 to make pFM11; plasmid pFM8 was digested with *PvuII* and the 0.7 kb fragment cloned in the *SmaI* site of pUC19 to make pFM12.
Figure 6.1

Southern hybridization of 400 bp lexA probe to T616 cosmid DNA digested with various restriction enzymes

This figure shows the Southern blot using 50 ng of labelled lexA probe to probe 100 ng digests of T616 cosmid with the restriction enzymes indicated in each lane. The positions of the λ EcoRI, HindIII size markers are indicated.
6.3 Sequencing of the \textit{lexA} gene

The recombinant clones in pUC19 as described in section 6.2.2 were sequenced, using universal reverse (5' TTG TGA GCG GAT AAC AAT TTC 3') and forward (5' GTT TTC CCA GTC ACG AC 3') primers. These primers hybridize to pUC19 and enabled sequencing of part of the gene. Subsequently using some specifically designed primers (Fig. 6.2) based on the sequence already obtained, the sequence of a 907 bp region of DNA containing \textit{lexA} was determined on both strands. The complete sequence is presented in Fig. 6.3. Within this region a 651 bp open reading frame coding for a protein 217 amino acids long was identified. The start codon for translation was presumed to be the GTG at position 107, and not the ATG at position 50, based on homology with other LexA proteins (see section 6.4), the presence of a potential ribosome binding sequence just upstream of the GTG (Fig. 6.3) but not the ATG codon, and the location of the transcription start site (see section 6.6).

6.4 Comparison of the \textit{M. tuberculosis lexA} gene with that of other bacteria at the protein level

The deduced protein sequence of \textit{M. tuberculosis lexA} was compared with those of \textit{M. leprae}, \textit{B. subtilis} and \textit{E. coli}. There is 96% similarity with \textit{M. leprae} LexA protein, 72% similarity with \textit{B. subtilis} LexA protein and 57% similarity with \textit{E. coli} LexA protein (Fig. 6.4). The cleavage site of \textit{E. coli} LexA is between Ala84 and Gly85 (Horii \textit{et al}., 1981a) and the nucleophile in this reaction has been
Figure 6.2

Cloning and sequencing strategy of the *M. tuberculosis* *lexA* gene

This figure shows fragments of cosmid T616 containing the *M. tuberculosis* *lexA* gene cloned in pUC19. Using universal reverse and forward primers (RP, FP), some parts of the *lexA* gene were sequenced and then primers lex1, lex2, lex3, lex4 and lex5 were made based on the known sequence and used in sequencing reactions to complete the sequence of the entire gene on both strands. The positions of these primers are shown.
Figure 6.3

Complete nucleotide sequence of the *lexA* gene of *M. tuberculosis*.

The DNA sequence of the gene and flanking region is shown, with the protein translation of LexA. The transcription initiation site (see section 6.6) is marked with a red circle and the Cheo box motif is boxed in blue. Possible promoter like sequences are boxed in green and labelled as -10 and -35 and a putative ribosome binding site is overlined. The primer *lex*.ext, used for primer extension (see section 6.6) is underlined.
identified as Ser119 with Lys156 also being required as an activator (Slilaty & Little, 1987); these residues are identical and their flanking amino acids are also conserved in all LexA proteins sequenced (Garriga et al., 1992; Horii et al., 1981b; Markham et al., 1981; Mustard et al., 1992; Raymond-Denise & Guillen, 1991; Riera & Barbe, 1993; 1995; accession numbers P44858 & U00019 position 7772) including that of *M. tuberculosis* presented here, suggesting a common mechanism of proteolytic cleavage. In contrast, the amino acids important for specific DNA binding in *E. coli* LexA are not conserved in *M. tuberculosis* LexA (Fig. 6.5); this is not surprising because *M. tuberculosis* LexA binds to a different DNA sequence to that of *E. coli* (see Chapter 8).

### 6.5 Predicted LexA binding site (Cheo box) upstream of *M. tuberculosis* and *M. leprae* lexA genes

103 bp upstream of the LexA coding sequence there is a motif (GAAC-N4-GTTT) (Fig. 6.3) similar to those found upstream of SOS-inducible genes in *B. subtilis* termed a Cheo box and having the consensus sequence GAAC-N4-GTTC (Cheo et al., 1991). An identical sequence to that in front of the *M. tuberculosis* lexA gene is found in an equivalent location (100 bp upstream) with lexA from *M. leprae* (Accession number U00019 bases 7772-8425). Two additional copies of this motif separated by only 7 bp are located further upstream of the *M. leprae* lexA gene, being 284 bp and 265 bp respectively before the coding sequence (Fig. 6.6). There are also three copies of the Cheo box upstream of the *B. subtilis* lexA gene but more evenly spaced. This distribution of the motif upstream of each of the genes
## Figure 6.4

Comparison of the *M. tuberculosis* LexA protein sequence with those of other bacteria

The *M. tuberculosis* LexA protein sequences was aligned with those of *M. leprae* (Accession number U00019, bases 7772-8425), *B. subtilis* (Accession number M64684) and *E. coli* (Accession number J01643). The alignment was generated by the program MegAlign (DNASTar) using the cluster method. Residues that are identical to the *M. tuberculosis* sequence are highlighted.
Figure 6.5

Comparison of the *M. tuberculosis* LexA protein sequence with LexA of *E. coli*

The conserved cleavage site of *E. coli* LexA (Accession number J01643) is shown by the arrow and the nucleophile Ser119 and the activator Lys156 are underlined in blue. The amino acids (Ser39, Asn41, Ala42, Glu44 and Glu45) important for specific DNA binding in *E. coli* LexA which are not conserved in *M. tuberculosis* are underlined in red.
identified to date which would be part of the SOS response in mycobacteria suggest that, as in *B. subtilis*, this sequence motif is involved in regulation of expression of DNA damage inducible genes in mycobacteria.

6.6 Localization of the promoter region of the *M. tuberculosis lexA* gene

The transcription start point of the *M. tuberculosis lexA* gene was identified by the primer extension technique (see Materials & Methods) using primer lex.ext (5'ACG TCG AGA ATA GTG CGT TG 3') complementary to, and close to the beginning of, the LexA coding sequence. This primer was annealed to total RNA (40 μg) isolated from *M. tuberculosis* which had been cultured in the absence or the presence of ofloxacin (1μg/ml) to induce SOS genes and extended with AMV reverse transcriptase as described in the Methods section. The transcriptional start site was found to be at an A 57 bp upstream of the GTG translational start codon (Figs. 6.3 & 6.7). The transcriptional start point is preceded at an appropriate distance (8 bases upstream) by the sequence elements TACATT and TTGGTG which have homology with the -10 (TATAAT) and -35 (TTGACA) regions of *E. coli* promoters respectively, but, similar to what had been found for the *M. tuberculosis recA* gene, there is a spacing of only 10 bp between these two elements in this mycobacterial promoter region instead of the 16-19 bp found for *E. coli* promoters (Rosenberg & Court, 1979). The -35 like sequence was 5 bp downstream from the Cheo-box like sequence identified upstream of the *lexA* gene (Fig. 6.3).
Figure 6.6

The upstream sequence of the *M. leprae* lexA

The three consensus sequences similar to the Cheo box are boxed (Accession number U00019).
Figure 6.7

Identification of the transcription start site of the *M. tuberculosis* *lexA* gene

Primer extensions done with labelled primer lex.ext with: No RNA substrate (lane 1); 40 μg RNA isolated from uninduced *M. tuberculosis* cells (lane 2); 40 μg RNA from *M. tuberculosis* induced with ofloxacin at 1 μg/ml final concentration (lane 3). Lanes A, C, G and T represent sequencing reactions of the cloned upstream region of the *lexA* gene of *M. tuberculosis* (pFM9) using primer lex.ext. The arrow indicates the position where the transcription of the *lexA* gene of *M. tuberculosis* starts.
6.7 Discussion

The work in this chapter describes: i) Identifying the *M. tuberculosis* *lexA* gene from a cosmid by means of Southern blotting using a probe made by PCR. ii) Cloning of the *lexA* gene in plasmid pUC19. iii) Sequencing of the *lexA* gene using universal reverse and forward primers and some specifically designed primers. iv) Comparison of the *M. tuberculosis* *lexA* gene with that of other bacteria including *M. leprae*, *B. subtilis* and *E. coli* at the protein level. v) Identifying the transcription start site of the *M. tuberculosis* *lexA* gene by means of primer extension.

The sequence of a 907 bp region of DNA containing *lexA* was determined on both strands and within this region a 651 bp open reading frame coding for a protein 217 amino acids long was identified (Fig. 6.3). This orf exhibited very high homology to LexA of the other pathogenic slow growing species of this genus, *M. leprae* (89% amino acid identity, 96% similarity), and lower but still significant homology with LexA of the Gram-positive bacterium *B. subtilis* (47% amino acid identity, 72% similarity) and the Gram-negative bacterium *E. coli* (36% identity, 57% similarity).

The start codon for translation was presumed to be the GTG at position 107, and not the ATG at position 50, based on these alignments to give a protein of similar size to other LexAs, because of the presence of a potential ribosome binding sequence just upstream of the GTG but not the ATG codon, and the location of the transcription start site (Fig. 6.3).
The cleavage site of *E. coli* LexA is between Ala84 and Gly85 (Horri *et al.*, 1981a) and the nucleophile in this reaction has been identified as Ser119 with Lys156 also being required as an activator (Slilaty and Little, 1987); these residues are identical and their flanking amino acids are also conserved in all LexA proteins sequenced (Garriga *et al.*, 1992; Horri *et al.*, 1981b; Markham *et al.*, 1981; Mustard *et al.*, 1992; Raymond-Denise and Guillen, 1991; Riera and Barbe, 1993; 1995; accession numbers P44858 and U00019 position 7772) including that of *M. tuberculosis* presented here, suggesting a common mechanism of proteolytic cleavage.

In contrast, the amino acids important for specific DNA binding (Asn41, Glu44, Glu45, Ser39 and Ala42) in *E. coli* LexA (Knegtel *et al.*, 1995) are not conserved in *M. tuberculosis* LexA; this is not surprising because *M. tuberculosis* LexA binds to a different sequence to that of *E. coli* (see Chapter 8). The importance of these five amino acids in *E. coli* is supported by mutational studies (Thliveris *et al.*, 1991; Thliveris and Mount, 1992) and they along with the flanking residues are absolutely conserved among eight Gram-negative LexA sequences (Garriga *et al.*, 1992; Horii *et al.*, 1981b; Markham *et al.*, 1981; Mustard *et al.*, 1992; Riera and Barbe, 1993; 1995; accession number P44858); however only one of them, the serine residue, is found in *M. tuberculosis* LexA or indeed LexA from *M. leprae* or *B. subtilis*, the other Gram-positive LexAs for which the sequence is known. Since these three Gram-positive LexAs appear to all bind to the same recognition sequence (see Chapter 8) it might be expected that they would share an equivalent highly conserved amino acid motif, but if it is so it is not in the same location in the sequence. There is, however, a high degree of conservation apparent amongst the
three Gram-positive sequences from residue 27 to 41 (numbering the *M. tuberculosis* sequence).

The sequence GAAC-N4-GTTT similar to a consensus GAAC-N4-GTTC (Cheobox) identified in various DNA damage-inducible promoters in *B. subtilis* was found 100 bp upstream of the translation initiation codon. In Chapter 8, the operator function of this consensus sequence in *M. tuberculosis* will be discussed.

The site of transcription initiation was identified by primer extension using a primer complementary to, and close to the beginning of, the LexA coding sequence. The RNA used was isolated from *M. tuberculosis* which had been cultured in the absence or the presence of ofloxacin to induce SOS genes (see Chapter 3). The transcription start site was found to be at an A 57 bp upstream of the GTG translational start codon. This coincided with the ATG of the possible alternative translational start site referred to before. Although there are examples of transcription and translation initiation at the same point in mycobacteria (Timm et al., 1994; Sarkis et al., 1995) these are rare and in conjunction with the considerations discussed earlier it seems likely that it is not so in this case. The transcriptional start site is preceded at an appropriate distance (8 bases upstream) by the promoter sequence elements of a -10 box (TACATT) and -35 box (TTGGTG) similar to the *E. coli* consensus sequences TATAAT and TTGACA but not with the same spacing. Whereas most *E. coli* promoter elements are separated by 16-19 bp (Rosenberg and Court, 1979) the separation of the *M. tuberculosis* *lexA* homologous elements was 10 bp, similar to the observation made for the *M. tuberculosis* *recA* gene (see Chapter 5). The -35
like sequence was 5 bp downstream from the Cheo-box like sequence identified upstream of the \textit{lexA} gene (Fig. 6.3). It remains to be determined experimentally whether these elements constitute the binding site for mycobacterial RNA polymerase. Alternatively, if a different sequence is recognised by the mycobacterial RNA polymerase in the -35 region the Cheo box would overlap the promoter as has been found for various \textit{E. coli} SOS genes, including \textit{lexA} where of the two LexA binding sites one overlaps the -10 promoter region and the other is located downstream of the +1 transcription start site (Brent and Ptashne, 1981; Miki \textit{et al.}, 1984).
7. EXPRESSION AND PURIFICATION OF

*M. TUBERCULOSIS* LEXA

7.1 Introduction

The *M. tuberculosis* *lexA* gene has been cloned and sequenced (Chapter 6). To obtain the LexA protein in order to investigate its interaction with the putative control elements identified by sequence analysis, the gene needed to be subcloned into an expression vector.

This chapter describes:

i) The cloning of the *M. tuberculosis* *lexA* gene in the expression vector pET-15b to make a His-tagged fusion protein.

ii) The overexpression of LexA protein in *E. coli* BL21(DE3) under the control of the bacteriophage T7 promoter

iii) The purification of the LexA protein by affinity chromatography using Ni-NTA resin.

7.2 Expression studies of *M. tuberculosis* LexA in the vector pTZ18R

7.2.1 Cloning of the *lexA* gene in the vector pTZ18R

DNA sequence analysis revealed that neither of the clones pFM8 (1.3 kb *Sall* fragment) and pFM9 (1.5 kb *SmaI* fragment) contained an intact *lexA* gene so the purpose of cloning the *lexA* gene in the vector pTZ18R was: 1) To put all the gene
into one clone 2) To express the gene under the T7 promoter of vector pTZ18R.

Based on the restriction map of lexA, pFM9 was digested with AflIII plus Smal and 0.7 kb restriction fragment was purified from an agarose gel using Geneclean, made blunt-ended and ligated to pTZ18R digested with Smal and dephosphorylated, to make pFM14. To complete the cloning of lexA, pFM8 was digested with SalI and the 1.3 kb restriction fragment was purified from an agarose gel. This was ligated to the 3 kb SalI fragment from pFM14, to make pFM16 (Fig. 7.1).

7.2.2 Expression of the M. tuberculosis lexA gene from the T7 promoter of pTZ18R

Once the lexA gene was cloned in pTZ18R, which has the T7 promoter, the recombinant plasmid was transferred to E. coli host strains BL21 (DE3) and B834 (DE3). The cells were grown and induced using IPTG as described in Chapter 2, but no expression of LexA was detected by Coomassie blue staining of SDS-PAGE gels. Further inductions of the construct using the above hosts containing the plasmids pLysS or pLysE, also gave no detectable expression of LexA.
Figure 7.1

The strategy for preparing the construct pFM16.

Cloning of the entire \( \text{lexA} \) gene from the clones of pUC19 carrying \( \text{lexA} \) gene fragments into pTZ18R.

113
7.3 Expression studies of *M. tuberculosis* LexA in the vector pET-15b

7.3.1 Cloning of the *lexA* gene in the vector pET15-b

DNA encoding *lexA* was cloned (Fig. 7.2) into the pET15-b plasmid under control of the strong bacteriophage T7 promoter to make a fusion protein with an N-terminal His tag. Thus the vector provided both the transcriptional and translational initiation signals in this case. In addition, the His.Tag sequence provides a convenient means of purification of the target protein. Expression was induced by providing a source of T7 RNA polymerase in the host cell.

The coding sequence for LexA was amplified by PCR using primers designed with the aid of the program PrimerSelect (DNASTar) to introduce restriction sites for *NdeI* at the N-terminus and *BamHI* at the C-terminus to permit cloning in-frame into the expression vector pET-15b (Novagen). The primers used were primer Nde.lex (5' G G A A C C T C A T A T G T T G T C C G C G A T T C 3') (the start codon in primer Nde.lex has been changed from GTG to ATG) and primer Bam.lex (5'G G A T C C T C A G A C C T T C C G T A T C A C C G T 3') where the introduced restriction site is highlighted in bold and the start and stop codons are indicated in italics. The PCR reaction was carried out as described in Chapter 2. The PCR products were visualized on a 1% agarose gel stained with ethidium bromide. After purifying DNA from the agarose gel by the Genecleaning method, the PCR products were digested with *NdeI* and *BamHI* and then cloned into pET15-b (restricted with the same enzymes *NdeI* and *BamHI* and dephosphorylated), to give pFM18.
Figure 7.2

The strategy for preparing pFM18 construct

The PCR product of lexA from pFM16 was cloned into pET-15b at NdeI, BamHI site.
7.3.2 Overexpression of *M. tuberculosis* LexA protein

The *lexA* gene had been cloned in plasmid pET-15b under the control of the strong bacteriophage T7 promoter and translation signals provide by the vector. The *lexA* gene was initially cloned using host (DH5-α) that does not contain the T7 RNA polymerase gene, thus eliminating plasmid instability due to the production of a protein potentially toxic to the host cell. For protein production, the recombinant plasmid (pFM18) was transferred to the expression *E. coli* strain BL21 (DE3) containing a chromosomal copy of the gene for T7 RNA polymerase. This host is a lysogen of bacteriophage DE3, a Lambda derivative that has the immunity region of phage 21 and carries a DNA fragment containing the *lacI* gene, the lacUV5 promoter and the gene for T7 RNA polymerase. Once a DE3 lysogen is formed, the only promoter known to direct transcription of the T7 RNA polymerase gene is the lacUV5 promoter, which is inducible by isopropyl-B-D-thiogalactoside (IPTG). Addition of IPTG at a 0.1 mM final concentration, for 3 hours, to a growing culture of lysogen (OD=0.5 at 600 nm) induces T7 RNA polymerase which in turn transcribes the cloned gene in the pET-15b vector. Strains of BL21 (DE3) were used which contained a compatible plasmid that provides a small amount of T7 lysozyme, a natural inhibitor of T7 RNA polymerase, which inhibits transcription.

After cloning of the *lexA* gene into the pET-15b vector as described in section 7.3.1, for protein production the recombinant plasmid was transformed to *E. coli* BL21 (DE3) containing plasmid pLysS or pLysE. The results of typical induction experiments are shown in Fig. 7.3 comparing the parental vector (pET-15b) and the
recombinant plasmid (pFM18) which were grown and induced under the same conditions in BL21 (DE3) pLysS or BL21 (DE3) pLysE individually. (The details of the growth and induction conditions were described in Chapter 2). There is clearly a major induced band (28 kDa) corresponding to LexA present with the clone which is absent with the vector, in the expected size range. The LexA protein was expressed to a higher level using the host BL21 (DE3) containing pLysS, (Fig. 7.3) than that containing pLysE.

7.4 Purification of the LexA protein

LexA protein was purified using a Ni-NTA resin as described in Chapter 2. Cloning of lexA in the vector pET-15b had the advantage of carrying the His.Tag sequence, a stretch of 6 consecutive histidine residues that are expressed at the N-terminus of the LexA protein. The His.Tag sequence binds at pH=8 to divalent cations (Ni^{2+}) immobilized on the His.bind metal chelation resin (Ni-NTA resin). The NTA ligand has four chelating sites which can interact with metal ions. NTA (nitrilo-tri-acetic acid) occupies four of the six ligand binding sites in the coordination sphere of the Ni^{2+} ion, leaving two sites free to interact with the 6 x His tag.

Initial experiments were performed using a batch method to determine the optimum washing and elution conditions for the His.LexA protein. The results presented in Fig. 7.4 showed that even at pHs as low as pH4 the His.LexA protein was retained by the resin and could subsequently be eluted by imidazole. The final purification scheme included washing at pH5 buffer which should remove most non-specific
Figure 7.3

SDS-PAGE analysis of *E. coli* BL21 (DE3) cell free extracts expressing the LexA protein

BL21 (DE3) pLysS and BL21 (DE3) pLysE transformed by plasmids pET-15b and pFM18 (*lexA* gene cloned in pET-15b) were induced by IPTG and a cell free extract was prepared and analyzed on a 15% SDS-Polyacrylamide gel. The results are shown as: lane 1) BL21 (DE3) pLysS cells harbouring plasmid pET-15b, lane 2) BL21 (DE3) pLysS cells harbouring plasmid pFM18 (*lexA* gene in pET-15b), lane 3) BL21 (DE3) pLysE cells harbouring plasmid pET-15b, lane 4) BL21 (DE3) pLysE cells harbouring plasmid pFM18, lane M) molecular weight marker size given in kDa. The arrow shows the expressed His Tagged-LexA protein.
binding. Then the His.Tagged-LexA protein was recovered by elution with imidazole (0.25 M, final concentration) which binds to the Ni-NTA column and displaces the Tagged protein. To prevent non specific binding to the resin, a high salt concentration (1 M NaCl) was used in all the solutions (see Chapter 2) and a very low concentration of imidazole (1 mM) was also used in the washing solution (see Chapter 2).

Figure 7.5 shows the final purification (washing steps done at pH5, using 1 mM imidazole) using a column procedure. In this procedure the Ni-NTA column was packed and washed and the cell lysate was applied slowly to the column at approximately 1 ml / minute. In the elution step, 5 ml samples were collected in individual tubes, then samples from the washing steps and the different fractions collected in the elution steps were run in an SDS-polyacrylamide gel. The purification procedure is shown schematically in Fig. 7.6. The purified protein was estimated to be >95% pure from SDS-polyacrylamide gel electrophoresis.
Figure 7.4

SDS-PAGE analysis of LexA protein purification using Ni-NTA resin (Batch procedure)

Cell free extracts prepared from an induced culture of BL21 (DE3) pLysS cells harbouring pFM18 (lexA cloned in pET-15b) were purified using a Ni-NTA resin and analyzed on a 15% SDS-polyacrylamide gel: lane M) molecular weight markers, size given in kDa, lanes 1 & 2) crude cell free extracts equilibrated in sonication buffer pH8, lanes 3 to 8) washing steps at pH6, lanes 9 & 10) elution step using 0.25 M imidazole in elution buffer at pH8, lanes 11 & 12) alternative elution at pH5 and pH4 respectively (without any imidazole), lane 13) final elution using 0.25 M imidazole in elution buffer, following low pH elutions.
Figure 7.5

SDS-PAGE analysis of LexA protein purification using Ni-NTA resin (column procedure)

Cell free extracts were prepared from an induced culture of BL21 (DE3) pLysS cells harbouring pFM18 (lexA cloned in pET-15b), purified using a Ni-NTA column procedure, and all the fractions analyzed on a 15% SDS-polyacrylamide gel: lane M) molecular weight marker, size given in kDa, lanes 1 to 3) washing steps at pH5, lanes 4 to 11) different fractions of elution steps using 0.25 M imidazole (samples of fraction 7 were used in the binding experiments described in Chapter 8).
Grow Cells to Mid-log Phase
\[ \text{OD}_{600} = 0.5 \]

Induce with 0.1 mM IPTG, 3hr

Harvest Cells and Lyse by Sonication in Sonication Buffer pH = 8

Incubation with the Ni-NTA Column for 1 hr.

Wash the Resin with Buffer pH = 5

Protein Elution Using Elution Buffer containing Imidazole (0.25M)

Separation of the Products by SDS-PAGE.

Figure 7.6
Flow diagram of the LexA purification procedure
7.5 Discussion

Purified *M. tuberculosis* LexA would facilitate studies of its role in the regulation of DNA damage inducible genes of *M. tuberculosis*, particularly *recA*. Following the lack of high level expression of the *M. tuberculosis lexA* gene cloned in vector pTZ18R, PCR amplification was applied using two primers based on the sequence of the *lexA* coding region, to clone the *lexA* gene into the T7 expression vector pET-15b to produce a fusion protein having an amino-terminal His tag. The recombinant plasmid was transferred to strain BL21 (DE3), a lysogen of bacteriophage DE3 that carries a DNA fragment containing the *lacI* gene, the gene for T7 RNA polymerase and the lacUV5 promoter. The lacUV5 promoter is inducible by isopropyl-B-D-thiogalactopyranoside (IPTG); addition of IPTG to a growing culture of the lysogen induces T7 RNA polymerase which in turn transcribes the target gene in the plasmid. The host strain used contained a compatible plasmid (pLysS) that provides a small amount of T7 lysozyme, a natural inhibitor of T7 RNA polymerase (Moffatt and Studier, 1987; Studier, 1991). T7 lysozyme is a bifunctional protein: it cuts a specific bond in the peptidoglycan layer of the *E. coli* cell wall (Inouye *et al.*, 1973), and it binds to T7 RNA polymerase, inhibiting transcription. When produced from the cloned gene, relatively high levels of T7 lysozyme can be tolerated by *E. coli*, apparently because the protein is unable to pass through the inner membrane to reach the peptidoglycan layer. The presence of pLysS (or pLysE) has the further advantage of facilitating the preparation of cell extracts, as it lyses the cells efficiently on freeze-thawing (Studier *et al.*, 1990). As strain BL21 lacks the *lon* protease and the *ompT* outer membrane protease that can degrade proteins during
purification, some target genes should be more stable in this strain than in hosts containing these proteases.

Induction of the clone pFM18 (lexA in pET-15b) in *E. coli* BL21 (DE3) pLysS gave high level expression of an approx. 28 kDa protein corresponding to the LexA fusion protein which was absent from the vector control (Fig. 7.3). Cloning of *lexA* in pET-15b had the advantage of forming a fusion protein with the His.Tag sequence, that binds to divalent cations (Ni$^{2+}$) immobilized on the His.bind metal chelating resin (Ni-NTA resin). Following the overexpression of His.Tagged-LexA protein, purification was achieved using Ni-NTA resin. Unbound protein was washed away and the His.Tagged-LexA protein was recovered by elution with imidazole which bound to the Ni-NTA column and displaced the tagged protein yielding 95% pure LexA protein.
8. INTERACTION OF LEXA PROTEIN WITH
THE RECA OPERATOR REGION

8.1 Introduction

The LexA repressor of E. coli negatively regulates the transcription of about twenty "SOS genes" including its own gene (Brent and Ptashne, 1980). The availability of purified M. tuberculosis LexA protein allows the study of its function in the regulation of recA gene expression and the investigation of any role in autoregulation.

The work in this chapter describes:

i) The demonstration that LexA binds to the promoter region of recA

ii) The precise identification of the DNA sequence upstream of recA to which LexA binds.

iii) Evidence that M. tuberculosis LexA has a role in autoregulation.

8.2 LexA binds to the promoter region of recA

The study of DNA-protein interactions has been facilitated by the "gel retardation" or "gel mobility shift" assay. This technique is based on the separation of free DNA from DNA-protein complexes due to the differences in their electrophoretic mobilities.
After purification of the LexA protein it was dialysed in buffer (Chapter 2, section 2.7.1) suitable for the gel mobility shift assay. A 350 bp region of DNA upstream of the *M. tuberculosis* *recA* (350 bp *PvuII*-HindIII restriction fragment of pFM6) was purified by GeneCleaning and labelled with digoxigenin as described (in Chapter 2). A native polyacrylamide (5% in 0.25 x TBE buffer) gel was made the day before use, to make sure that the gel is completely polymerized and the ammonium per sulphate is decomposed and pre-run for 2 hours at 150 V. The labelled DNA fragment was diluted with TEN buffer to a concentration of 30 fmol/ml. The binding reaction was set up on ice using about 60 fmole of the labelled DNA and 0.2 μg of purified LexA protein, as well as using 1 μl of poly[d(I-C)] competitor DNA which prevents non specific binding. Following incubation at room temperature for 15 minutes, the samples were immediately loaded on to the gel. One lane was used to load sample buffer with bromophenol blue to follow the run. The dye was run 3/4 of the way to the bottom of the plates. The gel was blotted as described (Chapter 2, section 2.7.1) and chemiluminescent detection was performed, following the manufacturer’s instructions. The membrane was exposed to X-ray film for 35 minutes at room temperature. The results are shown in Fig. 8.1. The DNA fragment was clearly retarded in the presence of LexA indicating that *M. tuberculosis* LexA did bind to the promoter region of *M. tuberculosis recA*. 
Figure 8.1

Gel shift assay with 350 bp fragment from upstream of recA

A 350 bp labelled DNA fragment (30 fmole) upstream of the *M. tuberculosis recA* gene was bound to the purified LexA protein (0.2 µg) in binding buffer (see Materials & Methods), loaded on a 6% native polyacrylamide gel and run at 150 V until 3/4 of the way to the bottom of the plate:

lane 1) 350 bp labelled fragment without protein
lane 2) 350 bp labelled fragment with protein
8.3 LexA binds to an oligonucleotide containing the Cheo box

Once it has been shown that the 350 bp sequence upstream of the recA gene binds to the LexA protein and that a sequence similar to the consensus LexA binding site of B. subtilis, called the Cheo box, was found in this 350 bp region, oligonucleotides were designed based on the Cheo box and including 10 bp of upstream and downstream flanking sequence as follows:

Primer gel.shift 1
(5' CAC TTG AAT CGA ACA GGT GTT CGG CTA CTG TG 3')

Primer gel.shift 2
(5' CAC AGT AGC CGA ACA CCT GTT CGA TTC AAG TG 3')

A double stranded sequence was formed as described (Chapter 2, section 2.7.1) by denaturing the oligonucleotides together at 95°C, then cooling slowly to room temperature to allow them to anneal. The double stranded oligonucleotide was then labelled with digoxigenin. The gel shift reaction was carried out using 0.1 μg protein and 30 fmole of labelled DNA. The samples were loaded in a pre-run 8% native polyacrylamide gel and run at 150 V until the bromophenol blue had run 3/4 of the way to the bottom of the plates (one lane was used to load sample buffer with bromophenol blue to follow the run).

The results obtained (Fig. 8.2) clearly show that the labelled DNA was retarded in the presence of the LexA protein. The proportion of retarded DNA was reduced when the reaction was performed in the presence of excess unlabelled oligonucleotide of the same sequence (Fig. 8.2, lane 3). Thus the LexA protein does
bind at or near the Cheo box. Another sequence with some similarity to the Cheo box consensus is located further upstream of the \textit{M. tuberculosis recA} gene. This motif has two mismatches from the \textit{B. subtilis} consensus but was not retarded in the presence of the LexA protein, suggesting that it does not play a role in LexA mediated regulation of \textit{recA} gene expression.

8.4 Mapping the sites of interaction of LexA with the \textit{recA} promoter region by DNase I footprinting

DNase I footprinting enables the location of a protein binding site in the DNA sequence to be identified. Once the interaction between the 32 bp oligonucleotides containing the Cheo-box sequence and the LexA protein was shown by the gel shift assay, footprinting was carried out to localize the precise protein binding site in this DNA sequence.

A method based on the PCR technique and dideoxy DNA sequence reaction was used. In this method a pair of primers were designed for producing the one-end labelled target DNA sequence by PCR. The sequence of the two primers used are as follows:

Primer Rec5 (5' AGG CGC ATC ACC GAA CCT TTG 3')
Primer foot 1 (5' ACG ATC GGT GGT GAG GTT G 3')

These two primers based on the sequence upstream of \textit{recA} were end labelled with [\(\gamma^{32}\text{P}\)] ATP at their 5' termini by means of T4 polynucleotide kinase as described in Chapter 2, section 2.7.2.
Figure 8.2

Gel shift assay with 32 bp oligonucleotide containing the Cheo box

A 32 bp labelled oligonucleotide (30 fmole) containing the Cheo-box upstream of the *M. tuberculosis recA* gene was bound to the purified LexA protein (0.1 µg) in a suitable buffer, loaded in a 8% native polyacrylamide gel and run at 150 V until 3/4 of the way to the bottom of the plate.

- lane 1) 32 bp labelled oligonucleotide without protein
- lane 2) 32 bp labelled oligonucleotide with protein
- lane 3) 32 bp labelled oligonucleotide with protein plus excess of 32 bp unlabelled oligonucleotide (7.6 pmole)
The PCR was performed using 10 ng of template DNA (pEJ258) containing the upstream sequence of the *M. tuberculosis* recA gene in two separate reactions, to give products labelled separately at either end: in one reaction labelled primer Rec5 and unlabelled primer foot 1 were used and the other reaction was carried out using labelled primer foot 1 and unlabelled primer Rec5. The 480 bp PCR products were purified through spin columns, then 10 ng of each PCR product and 0.3 \( \mu \)g of LexA protein were incubated in binding buffer (see Chapter 2, section 2.7.2), at room temperature for 15 minutes to allow binding of LexA to the DNA, then MgSO\(_4\) was added to a final concentration of 5 mM. After treating with DNase I (see Chapter 2, section 2.7.2), the cleaved DNA fragments were phenol/chloroform extracted, ethanol precipitated and run on a 5% denaturing polyacrylamide gel. A sequencing reaction on plasmid pEJ258, using the same primer which was used in PCR as a labelled primer, was run in adjacent tracks.

There is a very clear protection of a region of DNA by the binding protein (LexA protein) (Fig. 8.3). This area covers the Cheo box and five-six nucleotides flanking either side as follows: TGAATCGAACAGGTGTTTCGGCTAC. The same region was protected when the other strand was labelled (data not shown). There was no evidence of any binding of LexA to the sequence with two mismatches from the Cheo box consensus, confirming the previous result obtained by gel retardation.

Thus the *M. tuberculosis* LexA protein does bind to and around the Cheo-box but this result does not identify which bases within the protected sequence are required for specific recognition.
Figure 8.3

DNasel footprinting of the *recA* promoter region binding to LexA

One end-labelled PCR product from the sequence upstream of the *recA* gene and LexA protein were allowed to bind, treated with DNase and the cleaved DNA fragments loaded in a 5% denaturing polyacrylamide gel. The results are shown as:

- lane 1) labelled PCR product without protein cleaved with DNase I
- lane 2) labelled PCR product with protein, in the presence of poly [d(I-C)] cleaved with DNase I
- lane 3) labelled PCR product with protein, in the absence of poly [d(I-c)] cleaved with DNase I

Lanes A, C, G and T represent sequencing reactions of the cloned DNA upstream of the *recA* gene (pEJ135) using primer Rec5 (the same primer which was end labelled for the PCR reaction).

The region of DNA protected by the binding protein (LexA) is indicated.
8.5 Confirmation of the importance of the Cheo box in the DNA-LexA protein interaction in *M. tuberculosis*

DNase I footprinting showed that the *M. tuberculosis* LexA protein can bind to the Cheo box sequence and six flanking nucleotides:

TGAATCGAAGGTGGTTCGCTAC. To confirm that the Cheo box GAAC-N4-GTTC is the essential sequence for binding as in *B. subtilis*, a pair of 32 bp oligonucleotides of the same sequence as those which were used previously in the gel mobility shift assay (section 8.3) but mutated in just the Cheo box were made (altered bases shown in italic and bold):

Primer mut.gs.1

(5' CAC TTG AAT CTC CAA GGT TGG AGG CTA CTG TG 3')

primer mut.gs.2

(5' CAC AGT AGC CTC CAA CCT TGG AGA TTC AAG TG 3')

When these oligonucleotides were used in a gel shift assay exactly as before no retardation was detected (Fig. 8.4). Thus it was confirmed that the binding of *M. tuberculosis* LexA is specific for the GAAC-N4-GTTC consensus sequence.
Figure 8.4

Gel shift assay using mutated and wild-type oligonucleotides

32 bp oligonucleotides (30 fmole) of identical sequence to that used in Fig. 8.2 except for mutation of its Cheo box were designed, labelled and incubated in binding buffer with purified LexA protein (0.1 µg) then loaded on a 8% native polyacrylamide gel and run at 150 V. At the same time the original 32 bp oligonucleotide was used as a positive control in the assay. The results are shown as follows:

lane 1) 32 bp labelled wild-type oligonucleotide without protein
lane 2) 32 bp labelled wild-type oligonucleotide with protein
lane 3) 32 bp labelled wild-type oligonucleotide with protein + excess 32 bp unlabelled oligonucleotide (7.6 pmole)
lane 4) 32 bp labelled mutated oligonucleotide without protein
lane 5) 32 bp labelled mutated oligonucleotide with protein
lane 6) 32 bp labelled mutated oligonucleotide with protein + excess 32 bp unlabelled mutated oligonucleotide (7.6 pmole)

No retardation was found using the mutated Cheo box oligonucleotide.
8.6 The *lexA* gene product represses its own promoter

The *M. tuberculosis* LexA protein was shown to bind to the *recA* promoter by gel mobility shift assay, and DNase I footprinting enabled us to localize a protein binding site similar to the Cheo box found in *B. subtilis* (Cheo et al., 1991).

A sequence similar to the Cheo box, with a different flanking area and having a one nucleotide substitution (GAAC-N4-GTTC) compared with the consensus sequence (GAAC-N4-GTTC) was found in the DNA sequence upstream of the *M. tuberculosis* *lexA* gene itself (Fig. 6.3). To investigate whether *M. tuberculosis* LexA protein can bind to its own promoter, a gel mobility shift assay was performed using a 32 bp oligonucleotide containing this Cheo box and some flanking sequence as follows (the Cheo box highlighted in bold and italic):

Primer lex.gs.1 (5' CAC GCC TGT CGA ACA CAT GTT TGA TTC TTG GT 3')
Primer lex.gs.2 (5' ACC AAG AAT CAA ACA TGT GTT CGA CAG GCG TG 3')

Again a gel shift was clearly observed in the presence of LexA protein (Fig. 8.5) showing LexA binds to its own promoter. LexA protein binds to this Cheo box with one mismatch but did not bind to the sequence with two mismatches upstream of *recA*, suggesting that a Cheo box with a single mismatch is functional in regulation whereas one with two mismatches is not. A similar result has been reported in *B. subtilis* in that a Cheo box with a single mismatch from the consensus is required for repression, but deletion of one with two mismatches has no effect on *recA* gene expression (Cheo et al., 1993).
Figure 8.5

Gel shift assay using an oligonucleotide containing the Cheo box from upstream of the *M. tuberculosis* lexA gene

32 bp labelled oligonucleotide (30 fmole) containing the Cheo box upstream of the *M. tuberculosis* lexA gene was bound to the purified LexA protein (0.1 μg) in binding buffer, loaded in a 8% native polyacrylamide gel and run at 150 V. At the same time a 32 bp double-stranded oligonucleotide of unrelated random sequence (5'CAC AGT AGC CTC CAA CCT TGG AGA TTC AAG TG 3') was used as a negative control in the assay. The results are shown as follows:

lane 1) 32 bp labelled random oligonucleotide without protein
lane 2) 32 bp labelled random oligonucleotide with protein
lane 3) 32 bp labelled random oligonucleotide with protein + excess unlabelled oligonucleotide (7.6 pmole)
lane 4) 32 bp labelled oligonucleotide without protein
lane 5) 32 bp labelled oligonucleotide with protein
lane 6) 32 bp labelled oligonucleotide with protein + excess unlabelled oligonucleotide (7.6 pmole)
8.7 Discussion

The work in this chapter describes the regulatory function of the *M. tuberculosis* LexA protein. 121 bp upstream of the *M. tuberculosis recA* coding sequence there is a motif (GAAC-N4-GTTC) identical to those found upstream of SOS-inducible genes in *B. subtilis* (Cheo et al., 1991) termed a Cheo box. An identical sequence to that in front of *M. tuberculosis recA* is found 111 bp upstream of *recA* from *M. leprae* (Accession number X73822). Interestingly, there is a similar motif (GAAC-N4-GTTT) 103 bp upstream of the *LexA* coding sequence of *M. tuberculosis* and an identical sequence to that in front of *M. tuberculosis lexA* is found in an equivalent location (100 bp upstream) with *lexA* from *M. leprae* (Accession number U00019 bases 7772-8425). This distribution of the motif upstream of each of the genes identified to date which would be part of the SOS response in mycobacteria suggests that, as in *B. subtilis*, it is involved in regulation of expression of DNA damage inducible genes in mycobacteria.

To study the interaction of LexA with DNA, 350 bp of DNA upstream of *M. tuberculosis recA* was labelled and used with purified LexA protein in a gel mobility shift assay. Once the interaction had been shown by gel retardation, the DNA used was narrowed down to a 32 bp sequence containing the consensus sequence similar to that of the *B. subtilis* SOS box. As the results showed the interaction occurred using the 32 bp oligonucleotides, DNase I footprinting was used to identify precisely the DNA sequence bound by the LexA protein. The DNA protected by LexA protein covers the consensus sequence (Cheo box) and a further five-six nucleotides
flanking it on either side.

To study the importance of the consensus sequence GAAC-N4-CTTC for recognition by LexA protein, the GAAC and GTTC nucleotides were mutated to completely different nucleotides in the 32 bp oligonucleotide used in the previous gel mobility shift assay. No retardation was observed with these mutated oligonucleotides. This proved the importance of the consensus sequence for the DNA interaction with LexA protein. Thus *M. tuberculosis* LexA recognises and binds to a Cheo box type motif.

There are three motifs similar to the consensus sequence of the *B. subtilis* Cheo box centered at 121 pb and 276 bp upstream, and 21 bp downstream of the *M. tuberculosis* *recA* coding region. Only the motif 121 bp upstream containing the GAAC-N4-GTTC sequence identical to the *B. subtilis* one, showed binding to the LexA protein. By sequence determination of the promoter region of *M. tuberculosis* *lexA*, a similar sequence (GAAC-N4-GTTT) has been identified and it was shown by means of gel retardation that the *M. tuberculosis* LexA protein, like that of *E. coli* (Brent and Ptashe, 1980), can bind to its own promoter. Although LexA will bind to a motif with one mismatch from the consensus (binding of the LexA protein to its own promoter) binding did not occur to the related sequence with two mismatches from the consensus found in the upstream DNA.

In *M. leprae* there are also two additional copies of this motif separated by 7 bp, located further upstream of the *lexA* gene, being 284 bp and 265 bp respectively.
before the coding sequence. Each of these copies has only a single mismatch from
the *B. subtilis* consensus suggesting that they are equally likely to be functional as
the motif closer to the gene. There also three copies of the Cheo box upstream of
the *B. subtilis lexA* gene but more evenly spaced. Although there was only one
functional copy of this motif in the *M. tuberculosis recA* gene it will be interesting
to determine whether there are also multiple copies of the motif in the *M.
tuberculosis lexA* gene. It is pertinent to note that the Gram-negative species which
have been studied possess two binding sites for LexA separated by only 3-5 bp
upstream of their *lexA* genes, but in these cases the operators overlap the promoters
and are very close to the beginning of the genes (Brent and Ptashne, 1981; Garriga
*et al.*, 1992; Riera and Barbe, 1993; 1995). It has been shown that in *E. coli* LexA
protein binds to each of these multiple sites with lower affinity than it does to the
single binding site upstream of the *recA* gene (Brent and Ptashne, 1981). The
presence of multiple binding sites for LexA might allow a more subtle level of
regulation than would be possible with a single site, perhaps with different degrees
of expression correlating with different levels of occupancy of the sites. Similarly,
varying affinities of LexA for different promoters will result in variable timing of
their induction, presumably reflecting their individual roles in DNA repair and cell
survival.

For all known SOS genes, the LexA repressor binds close to the promoter of these
genes suggesting that LexA exerts repression mostly through an inhibition of
transcription initiation (Schnarr *et al.*, 1991). In *E. coli recA* and *lexA* genes, the
LexA binding site overlaps the promoter region and appears to work by excluding
RNA polymerase from the promoter (Little et al., 1981). These results indicate that *M. tuberculosis* LexA recognises and binds to a similar sequence to that of *B. subtilis*, another Gram-positive bacterium, but which is quite unlike the SOS box to which LexA binds in many Gram-negative bacteria. This is only the second species documented to use this "Cheo" box for LexA binding and may be indicative of a general trend amongst Gram-positive bacteria. If this is the case, it would imply that the sequence to which LexA binds together with the LexA protein itself have evolved separately after the division into Gram-positive and Gram-negative bacteria.

The difference in the number of bases between the palindromic conserved four base sequence in each motif must also imply differences in the arrangement of the two recognition helices in the LexA dimers and hence in their overall quaternary structures.

These studies will now facilitate the elucidation of the role of LexA in the regulation of DNA damage inducible genes in mycobacteria. It would be envisaged that, analogous to other LexA regulated systems, the binding of LexA to its recognition site upstream of the genes it regulates would repress transcription. Upon DNA damage RecA would be activated leading to cleavage of LexA, release of the LexA fragments from the binding site, and hence expression of the gene concerned.
9. GENERAL DISCUSSION

9.1 Gene regulation in the mycobacteria

Progress towards an understanding of gene regulation and expression in mycobacteria has been slow. The few mycobacterial promoters that have been studied so far include the 16S rRNA genes of *M. tuberculosis* (Ji et al., 1994a; Verma et al., 1994), *M. leprae* (Sela and Clark-Curtiss, 1991) and *M. smegmatis* (Ji et al., 1994b); the *bla* gene of *M. fortuitum* (Timm et al., 1994), the *askβ* gene of *M. smegmatis* (Cirillo et al., 1994); the *hsp-60* gene of *M. bovis* BCG (Levin and Hatfull, 1993); the *cpn-60* gene of *M. tuberculosis* (Kong et al., 1993); the 85 antigen gene of *M. tuberculosis* (Keremer et al., 1995); the PAN promoter from *M. paratuberculosis* transposon IS900 (Murray et al., 1992); and the three promoters responsible for transcribing the repressor-like gp71 protein of mycobacteriophage L5 (Nesbit et al., 1995). As a first step toward the understanding of gene expression and regulation, I have analyzed the promoter regions of the two key regulatory elements of the SOS response in *M. tuberculosis*: the *recA* and *lexA* genes.

9.2 Inducible expression of *M. tuberculosis* RecA

Comparison of SOS boxes from *E. coli*, *Salmonella typhimurium*, *Erwinia carotovora*, *Pseudomonas aeroginosa* and *Pseudomonas putida* reveals that the consensus sequence CTG-N10-CAG is conserved in all known LexA operators of these Gram-negative bacteria (Garriga et al., 1992) although some other Gram-
negative species such as *Agrobacterium tumefaciens* (Wardham *et al.*, 1992) have no *E. coli*-like SOS box. It was shown that the *Anabaena variabilis* recA gene has no *E. coli*-like SOS box, but its expression is UV- and mitomycin C-inducible in *A. variabilis* itself (Owttrim and Coleman, 1989). In contrast, in *Thiobacillus ferrooxidans*, the recA gene has no SOS box and it is not inducible (Ramesar *et al.*, 1989). Examination of the putative promoter region of *M. tuberculosis* recA revealed no motifs like the *E. coli* SOS box, but interestingly a sequence identical to a consensus (GAAC-N4-GTTC), termed a Cheo box, common in various DNA damage-inducible promoters in *B. subtilis* (Cheo *et al.*, 1991) was found 120 bp upstream of the translation initiation codon in the *M. tuberculosis* recA DNA.

The results described here demonstrate that the recA gene of *M. tuberculosis* is inducible by DNA damaging agents and begin to analyse the requirements for this regulated expression. The cloned upstream fragment of *M. tuberculosis* recA containing the recA promoter was successfully used to drive expression of the CAT gene. Optimal induction of *M. tuberculosis* RecA expression in *M. smegmatis mc²155* occurred when the bacteria were grown in Dubos medium to 0.6-0.7 optical density at 600 nm and then incubated with the inducing agents for 5 hours. The optimum concentrations of DNA-damage inducing agents for induction were determined to be for 1 µg/ml for ofloxacin, 0.5-2.5 µg/ml for mitomycin C, 150 µg/ml for nalidixic acid and 70-90 seconds for uv irradiation.

Using constructs of the CAT reporter plasmid with varying lengths of recA upstream sequence, the regulatory sequences required for inducible expression were localized
to 310 bp of DNA upstream of the *recA* coding sequence (plasmid pFM17). This construct expressed *recA* under induced conditions at 4.2-fold the level of expression under uninduced conditions. This induction ratio is comparable to some of the isolated *E. coli* SOS (DNA damage-inducible) promoters fused to the galactokinase (*galK*) gene (Lewis *et al.*, 1992) and to the expression of the *recA-lacZ* fusion of *E. coli* in several Gram-negative bacteria in the presence of DNA damage (Fernandez de Henestrosa *et al.*, 1991). The induction ratio of the expression of the *E. coli recA-lacZ* fusion in *E. coli* is 6.8 (Fernandez de Henestrosa *et al.*, 1991), and the maximum induction ratios calculated for the fused *din-galK* of *recA, polB, dinG* and *dinH* are 8.5, 6.5, 4.7, and 3.6, respectively (Lewis *et al.*, 1992). The induction ratio found for *M. tuberculosis recA* is also in line with that described for *B. subtilis* of 5.5-fold from a *recA-xylE* fusion (Raymond-Denise and Guillen, 1992).

Construct pEJ161 which contains 260 bp of DNA from upstream of the *M. tuberculosis recA* coding sequence including the consensus repressor sequence, the Cheo box, did not show significant levels of CAT expression under either uninduced or induced conditions. This suggests that the 40 bp of DNA present in pFM17 but absent from pEJ161 contains a regulatory sequence essential for expression and induction of *M. tuberculosis* RecA in *M. smegmatis mc²155*. This regulatory sequence could be an upstream activating sequence that binds to an activating protein, essential for expression. An alternative possibility remains that the promoter is located in this region; although mapping the transcriptional start sites suggested that this was not the case, the primer extension products observed could represent
processing sites of a longer mRNA. In *E. coli* many promoters are activated by one or more activator proteins and in most cases these proteins bind within or upstream of their promoters. For instance the catabolite activator protein (CAP) activates transcription at the *lac* promoter by binding to a DNA site centered between position -61 and -62, relative to the transcription start site (Busby and Ebright, 1994). In the case of the *M. tuberculosis recA* gene the upstream activating sequence would be about -240 bp upstream from the first transcription start site.

In comparison with the *B. subtilis recA* regulatory system, in *M. tuberculosis* there appears to be two levels of regulation because when progressive deletions were made towards the *M. tuberculosis recA* gene, loss of expression was found while in *B. subtilis* there was no effect until the Cheo box was reached when constitutive expression was seen due to the inactivation of LexA repression (Cheo *et al.*, 1993). Although it was demonstrated here that LexA bound to the Cheo box in *M. tuberculosis*, it remains to be confirmed experimentally that this does indeed cause repression of transcription as would be expected by analogy to other systems. Further studies of SOS regulation in *M. tuberculosis* will elucidate these regulatory mechanisms and advance our understanding of gene regulation in mycobacteria.

9.3 Promoter region of the *M. tuberculosis recA*

Previous work has shown that mycobacterial genes are not generally expressed in *E. coli* (Hopwood *et al.*, 1988), possibly because their regulatory signals cannot be recognized by the RNA polymerase of the heterologous cells. The *M. smegmatis*
system offers an alternative, promising approach for the study of the regulation of
genes from slow growing mycobacteria (Bashyam et al., 1996). Here we have
compared the transcription start sites of the *M. tuberculosis recA* gene when it is
expressed from a plasmid cloned in *M. smegmatis* and when it is expressed from the
*M. tuberculosis* chromosome. In agreement with the observation that the efficiency,
and specificity of transcriptional recognition in both *M. tuberculosis* and *M.
smegmatis* appear to be conserved (Bashyam et al., 1996), two identical start sites
were identified in both cases that were located 47 and 93 bp upstream of the *recA*
coding sequence. These two transcriptional initiation signals for the *M. tuberculosis
recA* suggested that two kinds of mRNA, different in size, may be synthesized. In
both cases transcription initiated at a purine nucleotide (A) similar to the majority
of the mycobacterial promoters showed by Bashyam et al. (1996).

Although the *M. tuberculosis recA* gene is not expressed in *E. coli* from its own
promoter, a good match with the *E. coli* consensus promoter was found upstream
of the *recA* gene, 10 bp away from the first and major transcription start site. This
specific conserved sequence is present at nearly the same position upstream (-10
region) of the transcription start point for most of the mycobacterial promoters
mapped so far (Bashyam et al., 1996). These conserved sequences corresponded to
the Pribnow box present in nearly all classical prokaryotic promoters. The first,
second, and sixth bases (T, A, and T, respectively) of this conserved hexameric
sequence represent the functionally most important positions in *E. coli* (Rosenberg
and Court, 1979) and are conserved in mycobacterial species (Bashyam et al.,
1996). The corresponding bases for the *M. tuberculosis recA* promoter are T, C,
and T, respectively.

Although the sequences in the -35 regions of mycobacterial promoters in general do not bear any homology with the TTGACA motif present in promoters of *E. coli* and several other bacteria (Bashyam *et al.*, 1996), interestingly there is a similar motif (TTGTCA) centered -28 bp upstream of the *M. tuberculosis* recA. Significantly, however, these elements were separated by a different distance from that found in *E. coli*. Whereas most *E. coli* promoter elements are separated by 16-19 bp, the separation of the *M. tuberculosis* homologous elements was 9 bp, which might explain the lack of expression of *M. tuberculosis* RecA in *E. coli*. So it is possible that either the mycobacterial RNA polymerase recognises the same motifs as the *E. coli* polymerase but at a different spacing or that it binds to a different sequence in the -35 region.

Analysis of the mycobacterial promoter sequences from positions -1 to -50 had shown that the promoters from *M. tuberculosis* have a higher GC content (57%) than the promoters from *M. smegmatis* (43%) (Bashyam *et al.*, 1996). Interestingly the analysis of the *M. tuberculosis* recA promoter from the position -1 to -50 of the first transcription start point also shows high GC content (58%). Weaker extra bands were seen within a few bases of the transcription start site nearest to the coding region, presumably reflecting some initiation at those points as has been reported before eg. in BCG *hsp60* gene (Stover *et al.*, 1991).

In addition a sequence resembling that of heat shock promoters in *E. coli* was found
upstream of the second transcription start site about 100 bp upstream of the initiation
codon. This product was fainter using RNA from *M. tuberculosis* compared with
that from *M. smegmatis* carrying the clone suggesting that the promoter nearest to
the gene could be the stronger one, although it has not been established whether
these sequences function as promoters in mycobacteria.

These promoters appear to represent two different classes, one of which contains
features similar to the majority of *E. coli* promoters and presumably is recognized
by a form of *M. smegmatis* RNA polymerase associated with a sigma factor related
to the $\sigma^{70}$ protein of *E. coli*. However, these promoters do not appear to be
particularly active in *E. coli* (Davis *et al.*, 1991). Their poor activity in *E. coli in
vivo* is consistent with previous reports (Das Gupta *et al.*, 1993). The second
promoter appears to represent a second type of promoter that contains quite different
promoter elements similar to the $\sigma^{32}$ *E. coli* heat shock promoter.

LexA protein binds specifically to the palindromic sites overlapping DNA damage
inducible gene promoters (Lovett *et al.*, 1993). The LexA binding sites in *E. coli*
either overlap with the -35 promoter region (consensus TTGACA) as in the case of
the *uvrA* gene (Sancar *et al.*, 1982), are situated between the -35 and -10 promoter
regions as in the case of the *recA* gene (Horii *et al.*, 1980; Sancar *et al.*, 1980),
overlap with the -10 promoter region like the *lexA* gene (Brent and Ptashne, 1981;
Miki *et al.*, 1984), or are sited downstream from the -10 promoter or even
downstream from the +1 transcription start site as in the case of the *uvrD* gene
(Easton and Kushner, 1983).
Interestingly the consensus sequence SOS box (Cheo box) upstream of the \textit{M. tuberculosis} \textit{recA} structural gene, similar to the \textit{E. coli} \textit{recA} one, is positioned between the -10 and -35 promoter elements of the second recognized transcription start point that resembles heat shock promoter elements from \textit{E. coli}. This overlap between the promoter and operator (LexA binding site) is typical of known SOS genes in general (Lewis \textit{et al.}, 1992). This siting of the LexA binding site between the -10 and -35 promoter regions suggests that LexA binding affects the binding of RNA polymerase.

Mycobacterial promoters seem to have features in common with streptomycete promoters, because they both have tolerance to large variety of sequences in their -35 region (Kremer \textit{et al.}, 1995, Bashyam \textit{et al.}, 1996). This high degree of transcriptional flexibility in the \textit{Streptomyces} spp results from the capacity to synthesize different sigma factors (Buttner, 1989; Westpheling \textit{et al.}, 1985). These common features may explain the fact that \textit{Streptomyces} is a better host than \textit{E. coli} for the expression of mycobacterial genes (Kieser \textit{et al.}, 1986). Therefore the heterogeneity in the -35 regions of mycobacterial promoters may result from having multiple sigma factors (Bashyam \textit{et al.}, 1996). It remains to be determined experimentally whether these elements constitute the binding site for mycobacterial RNA polymerase.
9.4 *M. tuberculosis* LexA protein, the repressor of the SOS response

The gene coding for the *M. tuberculosis* homolog of LexA has been cloned and sequenced. The sequence of a 907 bp region of DNA containing *lexA* was determined on both strands and within this region a 651 bp open reading frame coding for a protein 217 amino acids long was identified. The start codon for the translation was presumed to be the GTG at position 107, and not the ATG at position 50, based on protein sequence alignments to give a protein of similar size to other LexAs, and also because of the presence of a potential ribosome binding sequence just upstream of the GTG but not the ATG codon, and the location of the transcription start site. Comparison of the *M. tuberculosis lexA* gene at the protein level showed the highest similarity with LexA from the other pathogenic slow growing species of this genus, *M. leprae* (96%), followed by that from the Gram-positive bacterium *B. subtilis* (72%), with lower but still significant similarity with LexA of the Gram-negative bacterium *E. coli* (57%).

In DNA damage induced cells the autodigestion of LexA protein is promoted specifically by the RecA protease. This protease is inactivated in exponentially growing cells, but is reversibly activated by signal molecules thought to be single stranded DNA, leading to derepression of target genes for as long as functional LexA protein cannot accumulate. The cleavage site of *E. coli* LexA is between Ala84 and Gly85 (Horri *et al.*, 1981a) and the nucleophile in this reaction has been identified as Ser119 with Lys156 also being required as an activator (Slilaty and Little, 1987); these residues are identical and their flanking amino acids are also
conserved in all LexA proteins sequenced (Garriga et al., 1992; Horri et al., 1981b; Markham et al., 1981; Mustard et al., 1992; Raymond-Denise and Guillen, 1991; Riera and Barbe, 1993; 1995; accession numbers P44858 and U00019 position 7772) including that of *M. tuberculosis* presented here. Both *B. subtilis* (Cheo et al., 1993) and *M. tuberculosis* LexA proteins share significant homology to the LexA protein of *E. coli*, particularly within the regions of LexA that are involved in RecA mediated autocatalytic cleavage, suggesting a common mechanism of proteolytic cleavage.

In contrast, the amino acids important for specific DNA binding (Asn41, Glu44, Glu45, Ser39 and Ala42) in *E. coli* LexA (Knegtel et al., 1995) are not conserved in *M. tuberculosis* LexA. The importance of these five amino acids is supported by mutational studies (Thliveris et al., 1991; Thliveris and Mount, 1992) and along with the flanking residues are absolutely conserved among eight Gram-negative LexA sequences (Garriga et al., 1992; Horii et al., 1981b; Markham et al., 1981; Mustard et al., 1992; Riera and Barbe, 1993; 1995; accession number P44858); however only one of them, the serine residue, is found in *M. tuberculosis* LexA or indeed LexA from *M. lepae* or *B. subtilis*, the other Gram-positive LexAs for which the sequence is known. Since these three Gram-positive LexAs appear to all bind to the same recognition sequence, which is different from that used in Gram-negative bacteria, it might be expected that they would share an equivalent highly conserved amino acid motif, but if it is so it is not in the same location in the sequence. There is, however, a high degree of conservation apparent amongst the three Gram-positive sequences from residue 27 to 41 (numbering the *M. tuberculosis*
Purified *M. tuberculosis* LexA facilitated studies of its role in the regulation of DNA damage inducible genes of *M. tuberculosis*, particularly *recA*. Following the lack of high level expression of the *M. tuberculosis* *lexA* gene cloned in vector pTZ18R, the *lexA* gene has been cloned into the T7 expression vector pET-15b to produce a fusion protein having an amino-terminal His tag. Cloning of *lexA* in pET-15b had the advantage of forming a fusion protein with the His.Tag sequence, that binds to divalent cations (Ni²⁺) immobilized on the His.bind metal chelating resin (Ni-NTA resin). Following the overexpression of His.Tagged-LexA protein, purification was achieved using Ni-NTA resin. The tagged purified protein was estimated to be 95% pure.

### 9.5 Regulatory function of *M. tuberculosis* LexA protein

121 bp upstream of the *M. tuberculosis* *recA* coding sequence there is a motif (GAAC-N4-GTTC) identical to those found upstream of SOS-inducible genes in *B. subtilis* (Cheo *et al.*, 1991) termed a Cheo box. An identical sequence to that in front of *M. tuberculosis* *recA* is found 111 bp upstream with *recA* from *M. leprae* (Accession number X73822). Interestingly, there is a similar motif (GAAC-N4-GTTT) 103 bp upstream of the LexA coding sequence of *M. tuberculosis* and an identical sequence to that in front of *M. tuberculosis* *lexA* is found in an equivalent location (100 bp upstream) with *lexA* from *M. leprae* (Accession number U00019 bases 7772-8425). This distribution of the motif upstream of each of the genes
identified to date which would be part of the SOS response in mycobacteria suggests that, as in *B. subtilis*, it is involved in regulation of expression of DNA damage inducible genes in mycobacteria.

There are three motifs similar to the *B. subtilis* Cheo-box consensus sequence centered 121 bp and 276 bp upstream, and 21 bp downstream of the *M. tuberculosis* coding region. Only the motif 121 bp upstream containing the GAAC-N4-GTTC sequence identical to the *B. subtilis* consensus showed binding to the LexA protein. Although it is known that LexA will bind to a motif with one mismatch from the consensus as in the motif upstream of the *lexA* gene, binding did not occur to the related sequence with two mismatches from the consensus found in the DNA upstream of the *recA* gene.

There are also two additional copies of this motif separated by 7 bp, located further upstream of the *M. leprae lexA* gene, being 284 bp and 265 bp respectively before the coding sequence. Each of these copies has only a single mismatch from the *B. subtilis* consensus suggesting that they are equally likely to be functional as the motif closer to the gene. There are also three copies of the Cheo box upstream of the *B. subtilis lexA* gene but more evenly spaced. Although there was only one functional copy of this motif before the *M. tuberculosis recA* gene it will be interesting to determine whether there are also multiple copies of the motif before the *M. tuberculosis lexA* gene. It is pertinent to note that the Gram-negative species which have been studied possess two binding sites for LexA separated by only 3-5 bp upstream of their *lexA* genes, but in these cases the operators overlap the
promoters and are very close to the beginning of the genes (Brent and Ptashne, 1981; Garriga et al., 1992; Riera and Barbe, 1993; 1995). It has been shown that in E. coli LexA protein binds to each of these multiple sites with lower affinity than it does to the single binding site upstream of the recA gene (Brent and Ptashne, 1981).

In E. coli the precise localization of different LexA operators varies with respect to the transcription start site (Schnarr et al., 1991) and the different operators vary in their operator strength. In addition, the number of operators used for regulation ranges from one as in the case of E. coli recA (Walker, 1987; Sancar et al., 1980) to a triple operator in the case of E. coli recN (Rostas et al., 1987). The presence of multiple binding sites for LexA might allow a more subtle level of regulation than would be possible with a single site, perhaps with different degrees of expression correlating with different levels of occupancy of the sites. Similarly, varying affinities of LexA for different promoters will result in variable timing of their induction, presumably reflecting their individual roles in DNA repair and cell survival. Additionally, where the separation of the individual motifs is more than just a few base pairs as in B. subtilis, it has been proposed that the formation of a regulatory loop involving cooperative binding of LexA might occur (Cheo et al., 1991).

It has been shown that M. tuberculosis LexA protein binds specifically to the promoter region of M. tuberculosis recA and like that of E. coli (Brent and Ptashe, 1980) to its own promoter. It binds specifically to the consensus sequence GAAC-N4-GTTC, which is also the sequence of the B. subtilis SOS box (Cheo-box). It
was shown that substitution of all eight of these conserved bases eliminated LexA binding in *M. tuberculosis*. This proved the importance of the consensus sequence similar to the *B. subtilis* one for the DNA interaction with LexA protein. Thus *M. tuberculosis* LexA recognises and binds to a Cheo box type motif.

For all known SOS genes, the LexA repressor binds close to the promoter of these genes suggesting that LexA exerts repression mostly through an inhibition of transcription initiation (Schnarr et al., 1991). In *E. coli* *recA* and *lexA* genes, the LexA binding site overlaps the promoter region and appears to work by excluding RNA polymerase from the promoter (Little et al., 1981). The mechanisms of LexA control of the SOS response appear common to several species (Sedgwick, 1985).

In exponentially growing cells, LexA protein represses *recA*, *lexA* and other genes involved in the SOS response (Little et al., 1981). Our data show that *M. tuberculosis* LexA recognises and binds to a similar sequence to that of *B. subtilis* (GAAC-N4-GTTC), another Gram-positive bacterium, but which is quite unlike the SOS box to which LexA binds in many Gram-negative bacteria (CTGT-N8-ACAG), (Table-9.1). This is only the second species documented to use this "Cheo box" for LexA binding and may be indicative of a general trend amongst Gram-positive bacteria. If this is the case, it would imply that the sequence to which LexA binds together with the LexA protein itself have evolved separately after the division into Gram-positive and Gram-negative bacteria. The difference in the number of bases between the palindromic conserved four base sequence in each motif must also imply differences in the arrangement of the two recognition helices in the LexA dimers and hence in their overall quarternary structures.
9.6 Promoter region of *M. tuberculosis lexA*

The transcription start site was found to be at an A 57 bp upstream of the GTG translational start codon. This coincided with the ATG of the possible alternative translational start site referred to before; although there are examples of transcription and translation initiation at the same point in mycobacteria (Timm *et al.*, 1994; Sarkis *et al.*, 1995) these are rare and in conjunction with the considerations discussed previously it seems that this is not so in this case. The transcriptional start site is preceded at an appropriate distance (8 bases upstream) by the promoter sequence elements of a -10 box (TACATT) and -35 box (TTGGTG) similar to the *E. coli* consensus sequences TATAAT and TTGACA but not with the same spacing. Whereas most *E. coli* promoter elements are separated by 16-19 bp (Rosenberg and Court, 1979) the separation of the *M. tuberculosis lexA* homologous elements was 10 bp, similar to the spacing observed at the *M. tuberculosis recA* gene. The -35 like sequence was 5 bp downstream from the Cheo-box like sequence identified upstream of the *lexA* gene. The sequence GAAC-N4-GTTT similar to a consensus GAAC-N4-GTTC (Cheo-box) identified in various DNA damage-inducible promoters in *B. subtilis* was located 100 bp upstream of the translation initiation codon.

In *E. coli* there are two LexA binding sites at the *lexA* gene one of which overlaps the -10 promoter region and the other of which is located downstream of the +1
Table 9.1: Comparison of the recognition site for LexA in the Gram-negative bacteria with that of *B. subtilis*, *M. tuberculosis* and *M. leprae*
transcription start site (Brent and Ptashne, 1981; Miki et al., 1984). With the *M. tuberculosis lexA* gene, the LexA binding site does not appear to have any overlap with -10 or -35 promoter elements and is situated upstream of the -35 element. However, it may be that if a different sequence is recognised by the mycobacterial RNA polymerase around the -35 position, the Cheo box would overlap the promoter. With both *E. coli* and *M. tuberculosis recA*, the LexA binding site is situated between the -10 and -35 promoter regions. It has been speculated (Cheo et al., 1991) that when a repressor molecule is bound upstream of the -35 promoter element it might not interfere with the initiation of transcription by RNA polymerase as much as when a repressor is bound between the -35 and -10 promoter elements. It remains to be determined experimentally whether these elements constitute the binding site for mycobacterial RNA polymerase.

Little is known about mycobacterial gene expression, gene regulation and the influence of environmental conditions on expression levels. Identification of sequences involved in gene regulation and characterization of transcriptional signals should be useful for a better understanding of mycobacterial pathogenesis.

9.7 Future work

These studies will now facilitate the elucidation of the role of LexA in the regulation of DNA damage inducible genes in mycobacteria. It will be interesting to discover if components of this regulatory mechanism of *M. tuberculosis* and *B. subtilis* are conserved among other Gram-positive or closely related species of bacteria. Other
functionally important features of the LexA binding site should become apparent from sequence studies of other genes and their operator mutants. It was shown that substitution of all eight conserved bases for the recognition of the LexA protein eliminated LexA binding in *M. tuberculosis*. Whether all bases within this sequence are equally important for recognition by *M. tuberculosis* LexA remains to be determined although it is known that LexA will bind to a motif with one mismatch from the consensus (Movahedzadeh *et al.*, in press). However, binding did not occur to the related sequence with two mismatches from the consensus found in the DNA upstream of *recA*.

Studies of gene expression in mycobacteria are important because many mycobacterial genes including *recA* are not expressed in *E. coli* and so presumably have promoters which are not recognized by *E. coli* RNA polymerase but what these promoters are like is not known. Although it seems likely that the promoter elements recognized for the *M. tuberculosis* *recA* and *lexA* genes are located in the vicinity of the defined transcription start sites it remains to be determined whether these elements constitute the binding site for mycobacterial RNA polymerase, and thus function as promoters.

In addition the region of DNA identified from the deletion analysis as essential for expression is considerably further upstream from the *recA* gene than the putative promoter elements. One interesting possibility is that part of this sequence, maybe extending into the adjacent DNA, forms an upstream activating sequence to which an as yet unknown activator protein must bind to permit transcription, so it would
be interesting to identify any proteins interacting with this region of DNA.

The *in vivo* studies reported here have been undertaken using *M. smegmatis* as a host and although it appears that the transcriptional machinery of *M. smegmatis* and *M. tuberculosis* are similar (Bashyam *et al.*, 1996) the reporter gene analysis is being extended to *M. tuberculosis* itself. It also remains to be directly demonstrated that binding of LexA represses gene expression eg. by mutating the binding site such that LexA no longer binds *in vitro* and analyzing the expression of the reporter gene which would be predicted to become constitutive.

Better understanding of the transcriptional signals of mycobacteria will enhance our knowledge about regulation of gene expression in mycobacteria and may help to elucidate the mechanism of the pathogenesis of these organisms at the molecular level. In addition, it can provide new tools to study the molecular biology of the mycobacteria, and for expression of foreign genes of immunological interest in BCG for providing more effective recombinant vaccines. Moreover, *M. tuberculosis* resides in host macrophages, part of the normal defence mechanism where conditions might be expected to lead to DNA damage, so further studies of the regulation of DNA damage inducible genes may be important for a better understanding of its pathogenicity.
REFERENCES

Aldovini, A., Husson, R.N. and Young, R.A. (1993). The ura4 locus and homologous recombination in Mycobacterium bovis BCG. J. Bacteriol. 175, 7282-7289


169


170


180


181


*M. leprae* putative ribosomal RNA promoter in *E. coli*. *Gene* 98, 123-127


*J. Bacteriol.* 169, 1080-1088


Snapper, S.B., Lugosi, L., Jekkel, A., Melton, R.E., Kieser, T., Bloom, B.R. 


APPENDIX: MEDIA

**Luria - Bertani Medium (LB)**

- Bacto-tryptone: 10g
- Bacto-yeast extract: 5g
- NaCl: 10g
- Distilled water: to 1000ml

sterilize by autoclaving

**Lemco Broth**

- Bacto-peptone: 10g
- Bacto Lab Lemco powder: 5g
- NaCl: 5g
- Distilled water: to 1000ml

pH to 7.2 with 2M NaOH, sterilize by autoclaving

**Modified Dubos Medium**

**Part A**

- KH$_2$PO$_4$: 1g
- Na$_2$HPO$_4$.12H$_2$O: 6.25g
- Na$_3$citrate: 1.25g
- MgSO$_4$.7H$_2$O: 0.6g
- Asparagine: 2g
- 10% Tween 80: 5ml
Casamino acids 2g
dissolve asparagine in 50ml hot water, add to rest of ingredients, make up to 950ml,
adjust to 7.2 with 2M NaOH, make up to 1000ml, sterilize by autoclaving

Part B
Bovine Serum Albumine Fraction V 5g
Distilled water to 1000ml
filter sterilize

for use add 10ml part B to 250ml part A

M9 Minimal Medium
5X M9 salts* 200ml
sterile distilled water to 1000ml

add 20ml 20% glucose and 1ml 1% thiamine (both sterilized by filtration), 2ml of
sterile 1M MgSO\textsubscript{4} and 0.1ml sterile 1M CaCl\textsubscript{2} (The MgSO\textsubscript{4} and CaCl\textsubscript{2} should
prepared separately and sterilized by autoclaving).

*5X M9 salts
Na\textsubscript{2}HPO\textsubscript{4}.7H\textsubscript{2}O 64g
KH\textsubscript{2}PO\textsubscript{4} 15g
NaCl 2.5g
NH\textsubscript{4}Cl 5g
Distilled water to 1000ml
sterilize by autoclaving
HOMOLOGOUS RECOMBINATION

The process of homologous recombination is important to all organisms in providing the genetic variation that propels the evolution of organisms and in some cases, the differentiation of the cells. It also has an important role in repairing damaged DNA. Four steps are involved: initiation, exchange of DNA strands, DNA heteroduplex extension, and resolution. Homologous recombination in *E. coli* requires the participation of a large number of proteins in three overlapping pathways.

The recBCD pathway is involved in the initiation of the recombination process between linear DNA molecules and the chromosome. This pathway is RecA dependent and, in addition, a need for RecBCD, SSB, DNA polymerase I, DNA ligase, DNA gyrase and either the RuvA, RuvB, and RuvC proteins or RecG protein in this process has been defined. The RecBCD enzyme, which is also known as exonuclease V, is encoded by *recB*, *recC*, and *recD* genes. The enzyme possesses both helicase and nuclease activities (reviewed by Kowalczykowski, 1994). The RecBCD enzyme is also needed for the recombination hotspot activity that is associated with the DNA sequence known as Chi (5' GCTGGTGG 3'); the enzyme nicks the DNA near a Chi site and produces a 3'-ended ssDNA tail which is a potential substrate for homologous pairing by RecA and ssDNA binding proteins (reviewed by Smith, 1994). The enzyme binds to the DNA and begins unwinding it (due to its DNA helicase activity) forming ssDNA loops that enlarge as they travel along the DNA. This ssDNA is the substrate for invasion of the homologous
dsDNA, this process being promoted by RecA and SSB proteins. Since RecA is a DNA-dependent ATPase, the exchange of DNA strands is normally accompanied by the hydrolysis of ATP and in a complex reaction, DNA strand exchange between ssDNA and dsDNA substrates is stimulated by the *E.coli* ssDNA binding protein (Eggleston and Kowalczykowski, 1994). After formation of the homologously paired joint molecule and alignment of homologous sequences on the ssDNA and dsDNA, a Holiday junction can form through the action of either DNA polymerase I or DNA gyrase. Strand exchange requires local denaturation of the dsDNA and exchange of the corresponding single strands of DNA through the branch migration activity of either RecA, RuvAB or RecG protein resulting in the complete exchange of DNA single strands (reviewed by Kowalczykowski, 1994). Subsequent resolution of the recombinant strands requires the nucleolytic action of either RuvA, B and C or RecG proteins.

In *E.coli* recombination can also be promoted by the recF pathway. RecF is the main pathway for plasmid recombination and is involved in intra- and intermolecular recombination of circular substrates, and also in DNA repair. This pathway is also RecA dependent (Kowalczykowski et al, 1994). The recF pathway is inducible and is blocked by mutation in *recA* or *lexA* genes that prevent induction of the SOS response (Weinstock, 1987). The products of several additional genes-*recJ, recN, recO* and *recQ* are required for the recF pathway but not for the recBCD pathway (reviewed by Smith, 1989).

Another pathway is the recE pathway which can be activated by certain mutations
in *E. coli*. The *recE* gene encodes exonuclease VIII which digests each strand of double stranded linear DNA in the 5' to 3' direction to produce 3'-OH ssDNA ends. However, the *recE* pathway unlike the *recBCD* pathway, can also promote recombination between molecules that are primarily circular: plasmid with plasmid. Depending on the nature of the substrate the *recE* pathway can either require RecA, as in conjugal recombination, or does not require RecA as in plasmid recombination. Furthermore, the *recF, recJ, recO,* and *recQ* gene products are required for the *recE* pathway as well as the *recF* pathway, but not for the *recBCD* pathway. The gene responsible for RecE is part of a cryptic λ like prophage called rac (reviewed by Smith, 1989), and therefore this system may be less general to other species of bacteria.

Organisms may have multiple pathways of recombination in order to be able to utilize various forms of DNA substrates: circular or linear, double-stranded or single-stranded. This may be important to deal with various kinds of DNA damage incurred, eg. double-stranded breaks or patches of single-stranded DNA. Studies so far suggest that recombination systems presenting dsDNA ends follow the *recBCD* pathway, whereas those without dsDNA ends follow the *recF* pathway (Smith, 1989) and RecA is required in both of these pathways.
References:


