Characterization of *Mycobacterium tuberculosis* LexA: recognition of a Cheo (*Bacillus*-type SOS) box

Farahnaz Movahedzadeh, M. Joseph Colston and Elaine O. Davis

The gene coding for the *Mycobacterium tuberculosis* homologue of LexA has been cloned and sequenced. Amino acids required for autocatalytic cleavage are conserved, whereas those important for specific DNA binding are not, when compared with *Escherichia coli* LexA. The transcriptional start site was mapped and a DNA sequence motif was identified which resembled the consensus Cheo box sequence involved in the regulation of DNA-damage-inducible genes in *Bacillus subtilis*. The *M. tuberculosis* LexA protein was overexpressed in *E. coli* and purified by means of a His tag. The purified LexA was shown to bind to the Cheo box sequence found upstream of its own gene.

**Keywords:** *Mycobacterium tuberculosis*, LexA, RecA, DNA-damage-inducible genes, DNA-binding protein

**INTRODUCTION**

The Gram-positive bacterium *Mycobacterium tuberculosis* is the causative agent of tuberculosis, a disease which is responsible for more deaths worldwide than any other single infectious agent (Kochi, 1991). *M. tuberculosis* is an intracellular pathogen which survives and replicates 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 *M. tuberculosis* to DNA damage could be important in pathogenesis.

Bacterial responses to DNA damage are highly conserved (Walker, 1984), in particular the RecA protein, which also has a central role in recombination (Miller & Kokjohn, 1990). In the SOS response of *Escherichia coli* a group of over 20 genes is induced by a common regulatory mechanism in which RecA and LexA are the key players. 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.

Under normal conditions LexA binds to a specific DNA sequence, the SOS box, upstream of the genes it regulates, and represses transcription (Little & Mount, 1982). Exposure to DNA damage activates RecA coprotease activity, which then triggers the autocatalytic cleavage of LexA (Little, 1984, 1991), resulting in its release from its binding sites and increased expression of the genes it regulates, including recA and lexA themselves (Little *et al.*, 1994). The sequence of the SOS box (consensus taCTGTatatananaCAGta: Walker, 1984) is conserved amongst most Gram-negative bacteria, but a different sequence is found upstream of DNA-damage-inducible genes in the Gram-positive bacterium *Bacillus subtilis*, termed the Cheo box [consensus GAACNNNGTTCC, Cheo *et al.* (1991)]. The recA genes of *M. tuberculosis* and *M. leprae* have been cloned and sequenced (Davis *et al.*, 1991, 1994) and in both cases a sequence identical to the Cheo box consensus has been identified in the upstream regions. In order to determine whether this is the site to which LexA binds in mycobacteria, a source of pure mycobacterial LexA protein was required for *in vitro* binding assays. Towards this end, we now report the cloning, sequencing, expression and purification of *M. tuberculosis* LexA, and show that it does indeed bind to a Cheo box discovered in its own promoter region.

**METHODS**

**Bacterial strains, plasmids and growth conditions.** *Escherichia coli* strain DH5α (Sambrook *et al.*, 1989) was used for all plasmid constructions and strain BL21(DE3)pLysS (Studier *et al.*, 1990) was used for expression of LexA from the T7 promoter. The plasmid vectors used in this study are listed...
Table 1. Plasmids used in this study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant characteristics or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFM8</td>
<td>1.3 kb SalI fragment of ( M.) ( \text{tuberculosis} ) ( \text{lexA} ) gene in ( \text{SalI} ) site of ( \text{pUC19} )</td>
</tr>
<tr>
<td>pFM9</td>
<td>1.5 kb ( \text{SalI} ) fragment of ( M.) ( \text{tuberculosis} ) ( \text{lexA} ) gene in ( \text{SalI} ) site of ( \text{pUC19} )</td>
</tr>
<tr>
<td>pFM14</td>
<td>0.7 kb ( \text{AflIII} )–( \text{SalI} ) fragment of ( \text{pFM9} ) in ( \text{SalI} ) site of ( \text{pTZ18R} )</td>
</tr>
<tr>
<td>pFM16</td>
<td>1.3 kb ( \text{SalI} ) fragment of ( \text{pFM18} ) in 3 kb ( \text{SalI} ) fragment of ( \text{pFM14} )</td>
</tr>
<tr>
<td>pFM18</td>
<td>0.65 kb ( \text{Ndel} )–( \text{BamHI} ) fragment of ( \text{PCR} ) product of ( \text{lexA} ) gene from ( \text{pFM16} ) in ( \text{Ndel} )–( \text{BamHI} ) site of ( \text{pET-15b} )</td>
</tr>
<tr>
<td>pTZ18R</td>
<td>Mead et al. (1987)</td>
</tr>
<tr>
<td>pUC19</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>pET-15b</td>
<td>Studier et al. (1990)</td>
</tr>
</tbody>
</table>

in Table 1. \( E.\) \( \text{coli} \) cultures were grown at 37 °C in LB medium (Sambrook et al., 1989) with the addition of ampicillin and methicillin, where appropriate, at 50 μg ml⁻¹ each. \( M. \) \( \text{tuberculosis} \) was grown in modified Dubos medium (Difco) at 37 °C.

Recombinant DNA techniques. Standard molecular and recombinant DNA techniques were used (Sambrook et al., 1989). Plasmid mini-preps were prepared by the boiling method (Holmes & Quigley, 1981) or using a Qiagen miniprep kit. Maxipreps were prepared using the Qiagen plasmid/cosmid purification kit. DNA was digested with restriction enzymes (New England Biolabs and Life Technologies) according to the manufacturers’ instructions. DNA fragments were purified from agarose gels using Geneclean (Bio 101).

Southern hybridizations. DNA fragments for use as probes were labelled by random primer extension with the ECL DNA labelling kit (Amersham) following the manufacturer’s instructions. Hybridization was performed using Hybond-N membranes (Amersham) at 60 °C overnight prior to washing twice for 30 min at 60 °C with 2X SSC/0.1% SDS. Then blocking, antibody incubation and washing steps were performed following the manufacturer’s instructions.

DNA sequencing and analysis. Sequencing was performed using a Sequenase kit (USB/Amersham) with the inclusion of \( E.\) \( \text{coli} \) single-strand-binding protein in the reactions using both universal and specific primers. The reactions were resolved on 6% (w/v) polyacrylamide gels, and the sequences compiled and analysed using the Lasergene (DNAStar) and GCG (Devereux et al., 1984) programs.

Preparation of \( M. \) \( \text{tuberculosis} \) RNA. RNA was isolated as described by Gonzalez-y-Merchand et al. (1996).

Primer extension reactions. The oligonucleotide primers were end-labelled with [γ-\(^{32}\)P]ATP at their 5' termini by means of T4 polynucleotide kinase as described in the primer extension kit (Promega). Labelled primer (1 pmol) was annealed to 40 μg total RNA at 52 °C (based on the melting temperature of the primer being used) for 1 h. Extension was carried out using AMV reverse transcriptase (Promega), according to the manufacturer’s instructions, at 42 °C for 1 h. The products were ethanol precipitated, resuspended in 5 μl Tris/EDTA pH 7.5 plus 5 μl loading dye [98% (v/v) formamide, 10 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue] and denatured at 90 °C for 10 min before being separated on an 8% (w/v) polyacrylamide-urea gel.

Overexpression of \( M. \) \( \text{tuberculosis} \) \( \text{LexA} \). 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 \( \text{Ndel} \) at the N terminus and \( \text{BamHI} \) at the C terminus to permit cloning in-frame into the expression vector pET-15b (Novagen). The primers used (each at 300 nM final concentration) were GGAACCTCATAGTTGTCGGCAGATTCC and GGGATCTTACACCTTCCGTATCAGCT. The introduced restriction site is in bold and the start and stop codons are indicated by italics. PCR was performed using Expand High Fidelity PCR system (Boehringer Mannheim) with pFM16 (15 ng) as template. The temperature cycle used was as follows: an initial 5 min at 94 °C to denature high G + C DNA; then 10 cycles of 30 s at 94 °C, 30 s at 37 °C and 1.5 min at 72 °C; then 20 cycles of 30 s at 94 °C, 30 s at 37 °C and 1.5 min plus 20 s per cycle at 72 °C; and finally a last extension step of 72 °C for 7 min to complete primer extension. The annealing was done at a low temperature due to mismatches incorporated into the primers to prevent dimer formation.

The resulting PCR product was digested with \( \text{Ndel} \) and \( \text{BamHI} \) and cloned into pET-15b to give pFM18. This clone was transformed into \( E.\) \( \text{coli} \) BL21(DE3)pLysS. Cultures were grown in LB broth containing 200 μg ampicillin ml⁻¹ and 34 μg chloramphenicol ml⁻¹, and were induced at OD₆₀₀ 0.5 (1 cm pathlength, Unicam UV2 spectrometer) by the addition of IPTG to a final concentration of 0.1 mM (along with more ampicillin to 200 μg ml⁻¹) and incubation continued for a further 3 h before harvesting.

Purification of \( M. \) \( \text{tuberculosis} \) \( \text{LexA} \). The induced cells were harvested, resuspended in sonication buffer (50 mM phosphate pH 8, 1 M NaCl), and lysed by sonication at 100 W for 4 x 30 s. The LexA protein was purified by means of the His tag introduced from the vector on a Ni-NTA (Qiagen) column. The lysate was applied at 1 ml min⁻¹, the column washed with 50 mM phosphate pH 5, 1 mM imidazole, 10% (w/v) glycerol, and the LexA protein eluted with 50 mM phosphate pH 8, 1 M NaCl, 0.25 M imidazole. The purified protein was estimated by SDS-PAGE to be >95% pure.

Gel retardation assays. Gel shift assays were carried out using a DIG gel shift kit (Boehringer Mannheim). The double-stranded oligonucleotide (3-4 pmol μl⁻¹) was labelled following the manufacturer’s instructions, then diluted to a concentration of 15–30 fmol μl⁻¹ in TEN buffer (10 mM Tris/HCl, 1 mM EDTA, 0.1 M NaCl, pH 8). The gel shift reaction was set up according to the manufacturer’s instructions using 0.1 μg LexA protein with poly[d(I-C)] competitor DNA. Following incubation at room temperature for 15 min, the sample was applied immediately to a pre-electrophoresed native 8% (w/v) polyacrylamide gel. The gel system used was a Maxigel (Biometra) which has gel dimensions of 20 cm x 20 cm x 1 mm thick. Electrophoresis was carried out in 0.5 x TBE
buffer pH 8, at 160 V at 4 °C in a cold room. One lane was loaded with sample buffer containing bromophenol blue to follow the extent of electrophoresis; the dye was run three-quarters of the way to the bottom of the gel. The gel was blotted onto a positively charged BM nylon membrane (Boehringer Mannheim) using a Semi-Dry Electroboblotter (Auncos) at 400 mA for 1 h. Then the membrane was soaked in 10 x SSC for 30 s and cross-linked for 1 min at 300 nm with a transilluminator. The chemiluminescent detection was performed following the manufacturer's instructions. The membrane was exposed to X-ray film (Hyperfilm-MP, Amersham) for 15-40 min at room temperature.

RESULTS
Cloning the M. tuberculosis lexA gene
The lexA gene of M. tuberculosis had been located on cosmids T31, T616 and T754 (Philipp et al., 1996) as part of the M. tuberculosis genome analysis project. A 400 bp lexA probe made by PCR from total M. tuberculosis DNA was used in hybridizations to cosmid pFM16 bearing an intact lexA gene: first a 0.7 kb EcoRI-SmaI fragment was isolated from pFM9, made blunt-ended and cloned into the SmaI site of pTZ18R to give pFM14; then the 1.3 kb SalI fragment from pFM8 was cloned into SalI-digested pFM14 to give pFM16 (Fig. 1).

DNA sequence analysis
The sequence of a 907 bp region of DNA containing lexA was determined on both strands using both universal primers with various subclones in pUC19 and some primers designed to the sequence generated; the overall sequencing strategy is shown in Fig. 1. Within this region a 651 bp ORF coding for a protein 217 amino acids long was identified (Fig. 2). This ORF exhibited very high homology to M. leprae lexA (89% identity of amino acids) and lower but still significant homology with lexA from B. subtilis (47% amino acid identity, 72% similarity) and E. coli (36% identity, 57% similarity) (Fig. 3). 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 LexA proteins, and also...
because of the presence of a potential ribosome-binding sequence just upstream of the GTG (Fig. 2) but not the ATG codon, and the location of the transcription start site (see below).

**Mapping the transcriptional start site**

The site of transcription initiation was identified by primer extension using a primer (ACGTCGAGAATA-ATGCGTTG) complementary to, and close to the beginning of, the LexA coding sequence. This primer had been cultured in the absence or presence of ofloxacin (1 μg ml⁻¹) to induce SOS genes (F.M., unpublished observation) and extended with AMV reverse transcriptase as described in Methods. The transcriptional start site was found to be at an A 57 bp upstream of the GTG translational start codon (Figs 2, 4). This coincided with the ATG of the possible alternative translational start site referred to above; although there are examples of transcription and translation initiating at the same point in mycobacteria (Timm et al., 1994; Sarks et al., 1995), these are rare, and in conjunction with the considerations discussed in the previous section it seems most likely that this is not so in this case. The transcriptional start site 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 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). It remains to be determined experimentally whether these elements constitute the binding site for mycobacterial RNA polymerase. The −35-like sequence was 5 bp downstream from the Cheo-box-like sequence identified upstream of the *lexA* gene (see below).

**High-level expression and purification of *M. tuberculosis* LexA**

Purified *M. tuberculosis* LexA would facilitate studies of its role in the regulation of DNA-damage-inducible genes of *M. tuberculosis*, particularly *recA*. Therefore, the coding sequence for LexA was amplified by PCR using primers which introduced restriction sites at appropriate locations and, after digestion, the resulting fragment was cloned into the T7 expression vector PET-15b to produce a fusion protein having an N-terminal His tag (see Methods for details). Induction of this clone (pFM18) in *E. coli* strain BL21(DE3)pLysS gave high-level expression of an approximately 28 kDa protein corresponding to the LexA fusion protein, which was

---

**Fig. 3.** Comparison of the *M. tuberculosis* LexA protein sequence 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.
absent from the vector control (Fig. 5a). The protein was purified on the basis of the affinity of the His tag for metal ions on a Ni-NTA column as detailed in Methods, which yielded >95% pure LexA protein from this single step (Fig. 5b).

**LexA binds to the Cheo box upstream of its own gene**

Interestingly, 103 bp upstream of the LexA coding sequence there is a motif (GAAC-Nx-2-GTTC) (Fig. 2) similar to those found upstream of SOS-inducible genes in *B. subtilis* termed a Cheo box and having the consensus sequence GAAC-Nx-2-GTTC (Cheo et al., 1991). 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 777243425), and a motif identical to the *B. subtilis* consensus sequence is found 121 bp upstream of *M. tuberculosis* recA (Davis et al., 1991) and 111 bp upstream of *M. leprae* recA (accession number X73822). 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 test this hypothesis, the purified LexA was used in a gel retardation assay using a 32 bp double-stranded oligonucleotide (CACGCCTGTCGAACACATGTTTGATTCTTGGT) spanning the Cheo box (in bold) identified in the DNA sequence upstream of the *lexA* gene (see above). Retardation of this fragment due to binding of the protein was clearly observed in the presence of LexA protein (Fig. 6, lanes 4 and 5). The proportion of retarded DNA was reduced when the reaction was performed in the presence of excess unlabelled oligonucleotide of the same sequence (Fig. 6, lane 6). In contrast, no retardation was observed when LexA was incubated with a 32 bp double-stranded oligonucleotide of unrelated sequence (CACAGTAGCCTCCAACCTTTGGAGATTCAAGTG) (Fig. 6, lanes 1–3), demonstrating that the binding observed was specific. Thus *M. tuberculosis* LexA recognizes and binds to a Cheo-box-type motif.

**DISCUSSION**

The gene coding for the *M. tuberculosis* homologue of LexA has been cloned and sequenced. Its protein product shows significant homology to other LexA proteins, and for which the roles of particular residues have been identified, is informative with regard to its probable mode of action. The cleavage site of *E. coli* LexA is between Ala-84 and Gly-85 (Horii et al., 1981a) and the nucleophile in this reaction has been identified as Ser-119, with Lys-156 also being required as an activator.
44 and Glu-45 in they, along with the flanking residues, are absolutely conserved amongst eight Gram-negative LexA proteins spanning residues 28-52. The importance of these five bases in its recognition site, according to a model of the proteolytic cleavage presented here, suggesting a common mechanism of structure of the LexA DNA-binding domain (Fogh et al., 1994) in which a helix-turn-helix motif was identified.

In contrast, the amino acids important for specific DNA binding in E. coli LexA are not conserved in M. tuberculosis LexA; this is not surprising because M. tuberculosis LexA binds to a different DNA sequence to that of E. coli. The particular amino acids Asn-41, Glu-44 and Gln-45 in E. coli LexA are predicted to form specific H-bonds and Ser-39 and Ala-42 are predicted to form specific hydrophobic interactions with the CTGT bases in its recognition site, according to a model of the LexA DNA complex (Knegtel et al., 1995) based on the structure of the LexA DNA-binding domain (Fogh et al., 1994) in which a helix-turn-helix motif was identified spanning residues 28-52. The importance of these five amino acids is supported by mutational studies (Thliveris et al., 1991; Thliveris & Mount, 1992) and they, along with the flanking residues, are absolutely conserved amongst eight Gram-negative LexA sequences (Garriga et al., 1992; Horii et al., 1981b; Markham et al., 1981; Mustard et al., 1992; Riera & Barbe, 1993, 1995; accession numbers P44858 and U00019, position 7772), including that of M. tuberculosis presented here, suggesting a common mechanism of proteolytic cleavage.

Expression and purification of the M. tuberculosis LexA protein has been achieved using E. coli, and this has allowed us to demonstrate that it recognizes and binds to the same motif as that found upstream of SOS-inducible genes in B. subtilis known as a Cheo box. Curiously, 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. 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; it will be interesting to determine whether there are also multiple copies of the motif in M. tuberculosis and what role they play in regulation of LexA expression. There are also three copies of the Cheo box upstream of the B. subtilis lexA gene, but more evenly spaced. 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 & Ptashne, 1981; Garriga et al., 1992; Riera & 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 & 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. 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).

These results indicate that M. tuberculosis LexA recognizes and binds to a similar sequence to that of B. subtilis (GAAC-N3- GTTC), another Gram-positive bacterium, but which is quite unlike the SOS box to which LexA binds in many Gram-negative bacteria.
(CTGT-N₂-ACAG). 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 and particularly recA in mycobacteria. It may 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.

ACKNOWLEDGEMENTS

We are very grateful to Stewart Cole for providing the cosmid T616 and the M. tuberculosis lexA probe. We would also like to thank Peter Jenner for help with the protein purification, K. Papavinasasundaram for helpful discussions and Patricia Brooks for technical advice. F.M. is supported by a research training grant from the UNDP/World Bank/WHO Special Programme for Training in Tropical Diseases.

REFERENCES


Received 20 June 1996; revised 19 September 1996; accepted 26 September 1996.