

Global Analysis of the Regulon of the Transcriptional Repressor LexA, a Key Component of SOS Response in *Mycobacterium tuberculosis**^[5]

Received for publication, March 6, 2012, and in revised form, April 19, 2012. Published, JBC Papers in Press, April 23, 2012, DOI 10.1074/jbc.M112.357715

Katherine L. Smollett¹, Kimberley M. Smith², Christina Kahramanoglou, Kristine B. Arnvig, Roger S. Buxton, and Elaine O. Davis

From the Division of Mycobacterial Research, Medical Research Council National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom

Background: *Mycobacterium tuberculosis* LexA is thought to repress the expression of a small number of genes.

Results: 25 *in vivo* binding sites were identified by ChIP-seq, including nine novel sites.

Conclusion: *M. tuberculosis* LexA also shows examples of positive regulation, binding without apparent regulation, and binding to genes encoding small RNAs.

Significance: This investigation identified new aspects of LexA regulation in *M. tuberculosis*.

The DNA damage response is crucial for bacterial survival. The transcriptional repressor LexA is a key component of the SOS response, the main mechanism for the regulation of DNA repair genes in many bacteria. In contrast, in mycobacteria gene induction by DNA damage is carried out by two mechanisms; a relatively small number of genes are thought to be regulated by LexA, and a larger number by an alternate, independent mechanism. In this study we have used ChIP-seq analysis to identify 25 *in vivo* LexA-binding sites, including nine regulating genes not previously known to be part of this regulon. Some of these binding sites were found to be internal to the predicted open reading frame of the gene they are thought to regulate; experimental analysis has confirmed that these LexA-binding sites regulate the expression of the expected genes, and transcriptional start site analysis has found that their apparent relative location is due to misannotation of these genes. We have also identified novel binding sites for LexA in the promoters of genes that show no apparent DNA damage induction, show positive regulation by LexA, and those encoding small RNAs.

gen intermediates that can damage DNA (2, 3). Thus, a major critical step during infection by *M. tuberculosis* is the replication of the bacterium within macrophages, and its ability to survive these assaults depends on the DNA damage response (4). *M. tuberculosis* has two DNA damage response pathways as follows: the RecA/LexA-dependent, or SOS, response, and the RecA/LexA-independent response (5–8).

LexA is an essential component of the SOS response. Under normal growth conditions, LexA represses transcription of DNA damage-inducible genes by binding to an upstream DNA sequence termed the SOS box. Upon DNA damage, the presence of single-stranded DNA activates RecA, which in turn stimulates autocatalytic cleavage of LexA, lifting repression of the regulated genes (4, 9, 10). LexA has been well characterized in organisms such as *Escherichia coli* where the LexA/RecA SOS response seems to be the main mechanism of regulation of DNA repair genes following DNA damage.

The LexA-binding motif for *M. tuberculosis* was originally characterized by comparison with that of other bacteria and found to be similar to that of *Bacillus subtilis* (10), following which the consensus sequence was defined as TCGAAC(N₄)GTTCGA by use of a mutagenic approach (9). This information enabled the identification of genes and operators that appeared to be LexA-regulated (9); however, the number of these sites was relatively small, being 15, and some of the binding sites were found to be internal to the annotated coding sequences. Global comparison of genes induced by DNA damage found that the majority of genes remained inducible in a *recA* mutant strain, in particular the genes involved in DNA damage repair, confirming the lack of LexA/RecA regulation (7). The genes whose regulation was dependent upon RecA included 21 of those previously predicted based on an appropriately located SOS box or being co-transcribed with genes that do and three genes with no association with an SOS box. These were predicted to be artifacts due to their location downstream of highly inducible genes. A number of genes were found to show dual regulation, being regulated by both the RecA-dependent and -independent responses, showing partial

Tuberculosis, the result of infection by the bacterium *Mycobacterium tuberculosis*, causes more deaths worldwide than any other infectious disease (1). The preferred ecological niche of *M. tuberculosis* is the macrophage, a cell type that has evolved to kill most invading bacteria. A hallmark of *M. tuberculosis* infection is its ability to grow in this type of cell. When macrophages are activated, they produce reactive oxygen and nitro-

* This work was supported by United Kingdom Medical Research Council Program U1175 32056 and in part by European Union FP7 Program SystemMtb HEALTH-241587 and United Kingdom Medical Research Council Programs U117585867 and U117581288.

⌘ Author's Choice—Final version full access.

[5] This article contains supplemental Figs. S1 and S2 and Tables S1–S3.

¹ To whom correspondence should be addressed: RNAP Laboratory, Darwin Bldg., University College London, Gower Street, London, WC1E 6BT, United Kingdom. Tel.: 44-20-76790477; E-mail: k.smollett@ucl.ac.uk.

² Present address: Royal Veterinary College, Royal College St., London, NW1 0TU, United Kingdom.

induction in a *recA* deletion strain. However, the majority of these had no identifiable SOS box indicating that this regulation might not be due to direct binding by LexA (7). Comparison of the promoters of genes whose regulation is wholly or partially independent of LexA and RecA identified a consensus sequence termed the RecA nondependent promoter (RecA-NDp), which has subsequently been shown to be regulated by the *clp* gene regulator ClpR (11, 12).

In this study, we have used chromatin immunoprecipitation combined with high throughput sequencing (ChIP-seq) to identify the LexA-binding sites across the entire *M. tuberculosis* genome *in vivo*. In conjunction with promoter assays and transcriptional start site analyses, we confirmed many of the previously identified start sites for LexA-regulated genes and rectified gene misannotations. Significantly, we also identified nine novel LexA-binding sites, for most of which we were able to identify sequences similar to the consensus of the mycobacterial SOS box. Interestingly, among the newly identified LexA-binding sites, some were associated with genes not previously known to be DNA damage-inducible, and one was found to be positively regulated by LexA, and two were linked to potential small RNAs.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Culture Conditions—*E. coli* strain DH5 α (Invitrogen) was used for all plasmid constructions, strain XL1-Blue (Stratagene) was used for site-directed mutagenesis, and strain BL21 (DE3) pLysS (Stratagene) was used for protein expression (13). The mycobacterial strains used were wild-type *M. tuberculosis* strains H37Rv and 1424 and *M. tuberculosis* Δ *recA* mutant of 1424 (6, 14). Mycobacterial strains were grown in modified Dubos medium (Difco) supplemented with 4% albumin and 0.2% (w/v) glycerol in a rolling incubator at 2 rpm at 37 °C or on Difco Middlebrook 7H11-agar (BD Biosciences) plates supplemented with 4% albumin and 0.5% (w/v) glycerol. All procedures with live *M. tuberculosis* were carried out under Advisory Committee on Dangerous Pathogens Containment Level 3 conditions. Where appropriate, 50 μ g ml⁻¹ (for *E. coli*) or 25 μ g ml⁻¹ (for mycobacteria) kanamycin was included, and 250 μ g ml⁻¹ (for *E. coli*) or 50 μ g ml⁻¹ (for mycobacteria) hygromycin was included. To induce DNA damage, cultures were divided into 2 aliquots at mid-log phase ($A_{600} \sim 0.4$), and one sample was induced with mitomycin C (0.02 μ g ml⁻¹) for 24 h at 37 °C. The other sample was incubated in parallel, without treatment, to provide an uninduced control.

Plasmid Construction—The plasmids used and their construction are described in supplemental Table S1, and the primers used in this study are listed in supplemental Table S2. In general, standard DNA protocols were followed (13). Site-directed mutagenesis was performed using the QuikChange site-directed mutagenesis kit (Stratagene). All plasmids were verified by DNA sequencing.

Preparation of Cell-free Extracts, Western Blot, and β -Galactosidase Assays—Cultures were grown as described above; the bacteria were harvested, washed three times in PBS, and cell-free extract prepared as described previously (9). In the case of *M. tuberculosis* extracts, the supernatant was filtered

through a low-binding Durapore 0.22 μ m membrane filter (Ultrafree-MC; Millipore) to ensure complete removal of bacteria before removal from containment facilities.

For β -galactosidase activity, protein levels of cell-free extracts were quantified using a BCA kit (Pierce), and β -galactosidase activity was determined as described (9) and expressed in Miller units el⁻¹ mg of protein (15).

Antibody Preparation—Recombinant LexA was produced by expression of His-tagged *M. tuberculosis* LexA from plasmid pFM18 in *E. coli* strain BL21 (DE3) pLysS (10). Purified LexA was then used to immunize rabbits to produce polyclonal anti-LexA antibody by BioServ UK Ltd. (Sheffield University); specificity was determined by Western blot against *M. tuberculosis* cell-free extract.

Chromatin Immunoprecipitation (ChIP)—ChIP was performed using a method adapted from Sala *et al.* (16). Mycobacterial cultures were grown to mid-log phase ($A_{600} \sim 0.4$) before cross-linking by addition of 1% formaldehyde for 5 min at 37 °C. The reaction was then quenched with 125 mM glycine, and cells were washed in TBS and pellets stored at -80 °C. Pellets were resuspended in IP buffer (50 mM Hepes-KOH, pH 7.5; 150 mM NaCl; 1 mM EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; protease inhibitor cocktail Roche Applied Science) and sonicated to lyse cells and shear DNA using a Bioruptor (Diagenode). Samples were sonicated until the DNA fragment size was ~ 250 bp as determined by extracting the DNA from 50 μ l and running on a 2% agarose gel. Once sheared, 100 μ l was removed as the input control, and the rest of the sample was split in two. Anti-LexA antibody was added to one aliquot, and the other was a no antibody mock control; the samples were incubated overnight at 4 °C. Immunoprecipitation was performed using Dynabeads M280 sheep anti-rabbit IgG-conjugated magnetic beads (Invitrogen). Dynabeads and antibody/extract samples were mixed for 90 min at 4 °C; the beads were washed twice with IP buffer, once with IP buffer with 500 mM NaCl, once with wash buffer III (10 mM Tris-HCl, pH 8.0; 250 mM LiCl; 1 mM EDTA; 0.5% Nonidet P-40; 0.5% sodium deoxycholate), and once in Tris-EDTA buffer, pH 7.5. Elution and DNA extraction were then performed using IPURE DNA extraction kit (Diagenode). Samples were used directly in qPCR or whole genome sequencing or stored at -20 °C until required.

Library Preparation and DNA Sequencing—Concentrations of DNA samples were measured using Qubit HS DNA kit (Invitrogen). ChIP-enriched or input DNA was prepared for sequencing using ChIP-seq DNA sample preparation kit (Illumina). The quality of the prepared library was assessed using an Agilent BioAnalyzer 2100 (Agilent Technologies) before sequencing using an Illumina GA IIX platform (Illumina).

High Throughput Sequencing Analysis of Protein Binding Regions—Sequences were mapped to *M. tuberculosis* H37Rv genome using Bowtie 0.12.7 (17) allowing for one mismatch and removing any sequence that did not map uniquely. Sequences were further analyzed using SAMtools (18) and BEDtools (19) to create genome coverage plots, and statistical analysis was performed using R version 2.13.2. For each base position, the number of reads that mapped to that position was

calculated, then normalized to the total number of mapped reads $\times 10$ million, followed by subtraction of the input controls. Genome coverage plots were visualized and binding peaks determined using Artemis Genome Browser (20). The cutoff value to determine binding peaks was calculated from the number of reads per base as the mean $+ 2.5$ S.D. A binding peak was defined as any region where the normalized number of reads went above this value for over 50 bases. In addition, binding peaks were also determined using BayesPeak (21, 22). Nucleotide sequences of peaks were analyzed using Multiple Expectation Maximization for Motif Elicitation (MEME) to give position weight matrices of potential binding motifs (23). Other occurrences of the MEME-derived motif in the *M. tuberculosis* H37Rv genome were determined using Find Individual Motif Occurrences (FIMO)³ (24).

RNA Isolation and cDNA Synthesis—RNA was prepared from mycobacteria using the FastRNA Pro Blue kit (Qbiogene). Contaminating DNA was removed using TURBO DNA-free kit (Ambion). RNA quality was determined using an Agilent Bio-Analyzer 2100 (Agilent Technologies). cDNA synthesis was performed using Superscript III (Invitrogen) and random hexamer primers.

Quantitative PCR—Real time quantitative PCR was carried out using Fast SYBR Green master mix (Applied Biosystems) on an Applied Biosystems 7500 Fast instrument and analyzed with 7500 Fast SDS software version 1.4. Gene-specific primers (supplemental Table S2) were designed using Primer Express version 3.0 (Applied Biosystems). For each gene, cDNA samples (and their RT negative controls) or DNA extracts were run along with a set of genomic DNA standards to give a quantity of specific cDNA/DNA per sample. For quantitative RT-PCR, values for the RT negative controls were subtracted, followed by normalization to the corresponding value of the housekeeping genes *rrs*, encoding 16 S rRNA, or *sigA* to give relative expression level.

Transcriptional Start Site Mapping—Transcriptional start sites were mapped using the GeneRacer kit (Invitrogen) for RNA ligase-mediated rapid amplification of 5' cDNA ends or 5'-rapid amplification of cDNA ends version 2 (Invitrogen) along with gene-specific primers (supplemental Table S2). Amplified cDNA ends were cloned into pCR 4-TOPO (Invitrogen) for sequencing.

Northern Blot—Northern blot analysis was performed as in Ref. 25. Briefly, total RNA was separated on 8% denaturing acrylamide gels and electroblotted onto Brightstar membranes (Ambion). Riboprobes were made as in Ref. 25 using the mirVana probe construction kit (Ambion), [³³P]UTP (PerkinElmer Life Sciences), and the primer MTS2823 probe (supplemental Table S2). Membranes were incubated overnight with probes in Ultrahyb and exposed to phosphorimaging after washing. Transcript sizes were compared with RNA marker low (20–500 nucleotides, Abnova).

RESULTS

Chromatin Immunoprecipitation Analysis of LexA Binding Regions

A previous computational search of the *M. tuberculosis* genome for the LexA binding consensus sequence TCGAAC-(N₄)GTTCGA identified 24 potential LexA-binding sites with zero or one mismatch (9), although only 15 of these were thought to be able to bind LexA. To clarify LexA binding regions *in vivo*, ChIP experiments were performed. As LexA is a transcriptional repressor that binds to DNA during normal growth of wild-type *M. tuberculosis* strain H37Rv, cross-linking was performed at mid-log phase ($A_{600} \sim 0.4$), followed by lysis, DNA shearing, and immunoprecipitation using anti-LexA antibody and then purification of DNA fragments. Quantitative PCR was performed to confirm that known LexA binding regions were enriched in the immunoprecipitated DNA samples using primers that amplified the SOS boxes of *Rv3074*, *Rv3776*, and *dnaE2*, in contrast to the control primers located within the coding regions of *sigA* and *dnaE2*, which showed no enrichment (supplemental Fig. S1). High throughput sequencing was then performed on three independent replicates, using the input as a control. For each sample we obtained 17–43 million reads of 39 bp that could be mapped to the *M. tuberculosis* H37Rv genome. Approximately 2% of the genome contained no mapped reads due to the inability to map reads that aligned more than once. The number of reads per nucleotide position was normalized to the total number of mapped reads (see supplemental Fig. S2 for individual genome coverage plots). The coverage across the genome in the input control gave little variation with some exceptions, in particular the entire ribosomal RNA operon, consisting of *rrs*, *rrl*, and *rrf* encoding 16 S rRNA, 23 S rRNA and 5 S rRNA, respectively, was above background with an average of 213.8 ± 59.07 reads per base, and maximum of 588.7. Therefore, to correct for this, each immunoprecipitated sample was compared with its input control. To ensure correct detection of binding peaks, two different peak-calling methods were used. The first method is based on peak height. The genome coverage maps containing the normalized number of reads per nucleotide position for the input was subtracted from the immunoprecipitated sample and then the cutoff for binding was calculated as the means $+ 2.5$ S.D. for each replicate. LexA DNA binding regions were identified as any peak where the normalized number of reads went above this value for over 50 bases. The second method, BayesPeak, calculates the probability of peaks using a Hidden Markov model and also takes into account the orientation of reads (21, 22). This gives a posterior probability (PP) score for each peak, with values closer to one being highly probable specific peaks (details of peaks identified in each replicate are in supplemental Table 3). LexA binding peaks were then defined as any peak occurring in at least two replicates by height cutoff and one replicate by BayesPeak. This gave 25 individual LexA binding peaks, which were ranked according to their average peak height (Table 1 and Fig. 1B). The average peak height gives an indication of the strength of binding or binding in a higher proportion of cells.

The LexA binding peaks identified here correspond to all 15 SOS boxes originally predicted to bind LexA, one site containing

³ The abbreviations used are: FIMO, find individual motif occurrences; MEME, maximization for motif elicitation.

TABLE 1
LexA binding regions identified by ChIP-seq

Peak no. ^a	Coordinates ^b	Length	Score ^c	PP ^d	Regulated gene ^e	SOS motif mismatches ^f	DNA damage induction ^g
1	3811492–3811767	275	1043.43	1.00000	<i>Rv3395c</i>	1	RecA dep.
2	606414–606533	145	904.10	0.99998	<i>Rv0515</i>	1	RecA dep.
3	4221005–4221184	180	750.71	1.00000	<i>Rv3776</i>	0	RecA dep.
4	400043–400176	133	712.58	1.00000	<i>Rv0336</i>	1	RecA dep.
5	3784705–3784908	203	662.63	1.00000	<i>Rv3370c (dnaE2)</i>	0	RecA dep.
6	3640223–3640410	187	582.17	1.00000	<i>Rv3260c (whiB2)</i>	1	RecA dep. (negative)^h
7	3031660–3031838	178	555.65	1.00000	<i>Rv2720 (lexA)</i>	1	RecA dep.
8	1552433–1552634	201	524.12	1.00000	<i>Rv1378c</i>	0	RecA dep.
9	3436693–3436862	168	512.89	1.00000	<i>Rv3074</i>	0	RecA dep.
10	3051452–3051646	194	497.64	1.00000	<i>Rv2737c (recA)</i>	1	RecA par.
11	1928504–1928710	206	360.23	1.00000	<i>Rv1702c</i>	1	RecA dep.
12	1117075–1117251	175	356.95	0.99999	<i>Rv1000c</i>	1	RecA dep.
13	1178795–1178934	139	346.42	1.00000	<i>Rv1057</i>	2	Positive^h
14	3031533–3031633	100	343.3	0.72267	<i>Rv2719c</i>	2, 2, 3	RecA par.
15	1547104–1547230	126	315.41	0.87757	<i>MTS1082</i>	2	Noneⁱ
16	2358446–2358578	132	310.47	0.99891	<i>Rv2100</i>	1	RecA par.
17	79429–79535	106	280.49	0.99929	<i>Rv0071</i>	1	RecA dep.
18	2925356–2925457	101	260.48	0.99911	<i>Rv2594c (ruvC)</i>	1	RecA par.
19	105113–105205	91	213.99	0.35445	<i>Rv0095c</i>	3	RecA par.
20	2903479–2903605	126	197.87	0.99754	<i>Rv2578c/Rv2579 (dhaA)</i>	1	RecA dep.
21	1789838–1789976	138	195.42	0.13059	<i>Rv1588c</i>	3	RecA par.
22	2833811–2833900	89	188.50	0.20288	<i>Rv2517c</i>	3	Not known
23	4100757–4100877	121	185.50	0.99993	<i>MTS2823</i>	1	None ⁱ
24	2925113–2925226	113	181.51	0.17685	<i>Rv2593c (ruvA)</i>	2	RecA dep.
25	1179233–1179323	96	158.83	0.73687	<i>Rv1057</i>	1	Positive ^h

^a Identified LexA binding peaks; only peaks that occurred in at least two replicates by height cutoff analysis and confirmed by BayesPeak are included and are ranked by score. LexA binding regions not previously identified or thought to be functional are shown in boldface.

^b Peak start and end genome coordinates were averaged across replicates.

^c Average height for whole peak region, averaged across replicates.

^d PP means posterior enrichment probability calculated using BayesPeak, where PP = 1 shows positive peak and PP = 0 shows no enrichment (21, 22). PP value was averaged across replicates.

^e LexA-regulated gene; if binding peak was between divergent genes, then DNA damage induction was taken into account.

^f The number of mismatches to SOS consensus identified within the LexA binding peak (see Fig. 1, C and D).

^g Unless indicated, DNA damage induction was described in Ref. 7. dep indicates induction RecA-dependent; and par. indicates induction partially RecA-dependent.

^h Induction is shown in Fig. 3.

ⁱ Induction is shown in Fig. 4.

three SOS boxes with more than one mismatches to the consensus, which has also been shown to bind LexA, and nine novel binding sites (9, 26). No LexA binding peaks were detected in the promoter regions of three genes (*Rv0516c*, *Rv1376*, and *Rv3393*) whose DNA damage induction was found to be dependent on RecA but were not associated with identifiable SOS boxes (7), confirming that these genes are not LexA-regulated.

DNA sequences contained within each peak were then analyzed using Motif-based sequence analysis tool, MEME-ChIP (23). This identified a potential DNA-binding motif for LexA with high statistical significance (E -value 6.8×10^{-58} , see Fig. 1C), which matched the previously defined SOS box (9) with some additional bases. This motif was found in 21 of the peaks with one peak (peak 14) containing three motifs (labeled *a*, *b* and *c*, Fig. 1D and Table 1). The position-specific probability matrix for the MEME-identified binding motif was used to identify occurrences in the *M. tuberculosis* H37Rv genome using FIMO (24). This identified SOS motifs overlapping two further binding peaks (peaks 19 and 21). Peak 21 (associated with *Rv1588c*) is located next to a region of no sequence alignment, which may have caused the peak to be slightly off center compared with the actual binding site, as fewer sequences were able to align on one side. Peak 19, associated with *Rv0095c*, is adjacent to a potential second peak not confirmed by BayesPeak (supplemental Table 3), and so these two peaks may represent one binding site. Both of these genes (*Rv1588c* and *Rv0095c*) are highly similar, and the SOS motifs are identical, which may also cause lower levels of alignment across these regions. No

identifiable SOS box could be found in the remaining two LexA binding peaks. FIMO analysis identified a further 15 occurrences of the LexA-binding motif throughout the *M. tuberculosis* genome (p value < 0.000001). This included two occurrences where potential binding sites were found by the initial height-based peak detection but was not confirmed by BayesPeak. These are located in the promoter regions of *rbsK* and *dnaB*, which are known to be induced by DNA damage partially dependent on RecA (7) and so may represent false negatives and be weaker LexA-binding sites. The 13 other motif occurrences had no evidence of LexA binding, and all of these were located in intragenic regions. Interestingly, of the eight SOS box consensus sequences identified containing one mismatch that were thought unlikely to bind LexA (9), only one (peak number 6 located between *whiB2* and *fbtA*, average height 582.2, posterior probability (PP) value 1) was identified using FIMO, indicating that the modified SOS motif is a much more accurate predictor of LexA binding.

Analysis of LexA-binding Sites

Sites Associated with DNA Damage-inducible Genes—Of the 25 identified LexA DNA binding peaks, 16 were previously identified as containing SOS boxes thought to be functional (9, 26) (peaks 1–5, 7–12, 14, 16–18, and 20). A further three are associated with genes known to be DNA damage-inducible wholly or partially dependent on RecA (7) (peaks 19, 21, and 24). They were associated with SOS boxes containing more mismatches to the consensus (two or three compared with zero

Mycobacterium tuberculosis LexA Regulon

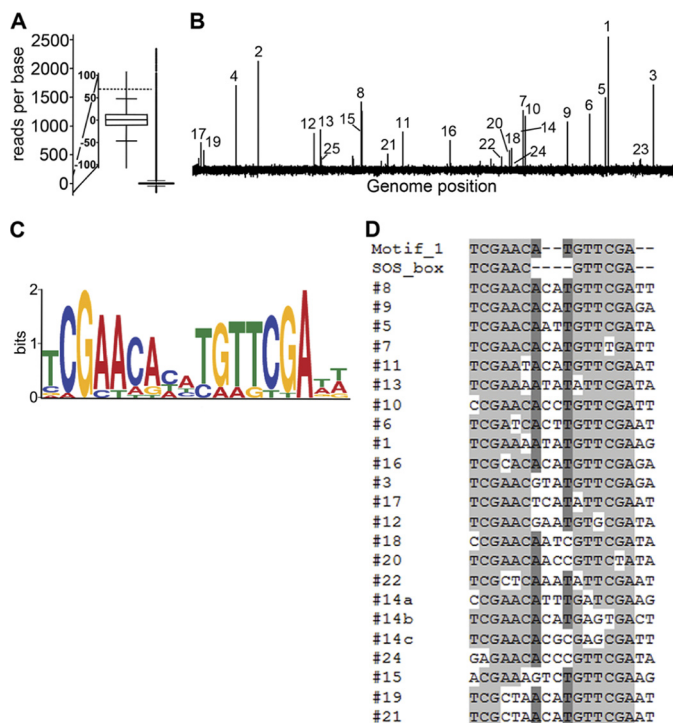


FIGURE 1. Overview of ChIP-seq results. *A*, box-plot showing average number of reads per base; dotted line indicates mean + 2.5 S.D. used as cutoff for peak calling. *B*, genome coverage plot of number of reads per base normalized to the total number of reads then normalized to the input control; peak numbers as in Table 1; average plot from three independent replicates. *C*, motif generated from MEME analysis on peak 3 sequences. *D*, alignment of motifs in different peaks with MEME-generated motif (motif_1) and previously defined SOS consensus sequence (SOS_box) (9); dark gray shows identity to motif_1; light gray shows identity to both motif_1 and SOS_box.

or one). Interestingly, one of these potential LexA binding peaks was identified within the coding region of *ruvC* (peak 24), as well as the previously identified SOS box located within the *ruvC* promoter (peak 18). This second peak (peak 24) contains an SOS box with two mismatches to the consensus sequence (Fig. 1D), and it may indicate more stringent regulation of the two downstream genes in the operon, *ruvA* and *ruvB*, the expression of both of which is dependent on RecA, whereas *ruvC* shows only limited RecA dependence (7, 27). The divergent genes *lexA* and *Rv2719c* contain two separate LexA binding peaks (peaks 7 and 14, respectively). This region contains four potential SOS boxes, only one of which was shown to regulate LexA, and this is located within peak 7. The other three, located within peak 14 (Fig. 1D), regulate *Rv2719c*, although the majority of its induction is thought to be due to the RecA non-dependent promoter, RecA_NDp (5, 26, 28).

Of the LexA binding peaks, six were found to be associated with genes belonging to the 13E12 family, including *Rv0515*, *Rv0336*, *Rv1702c*, *Rv2100*, *Rv1588c*, and *Rv0095c* (peaks 2, 4, 11, 16, 21, and 19, respectively). This family includes other genes/operons such as *Rv1128c*, *Rv1148c*, *Rv1945*, and *Rv3466*, all of which have been shown to be induced by DNA damage and contain the RecA_NDp (7, 11). Potential SOS boxes could be identified in all four of these other operons with either two or three mismatches to the consensus, but for three of these regions LexA binding could not be determined as the genome coverage plots did not align over the potential SOS boxes

because of the repeated DNA sequences within the genes. However, these SOS boxes were also not identified by FIMO using the modified SOS motif so are unlikely to represent true binding sites for LexA.

Sites Located within Annotated Coding Sequences of DNA Damage-inducible Genes—The LexA-binding sites associated with *Rv1378c*, *Rv3074*, and *Rv2100* (peaks 8, 9, and 16) were found to be within the annotated coding sequences and so were analyzed further, along with *Rv3395c*, which gave the highest ranking peak, and *Rv1000c* (peak 12) that was previously mis-annotated as being on the forward strand (29, 30). Therefore, we constructed transcriptional fusions to the reporter gene *lacZ* with DNA fragments that included the upstream regions of these genes and extended sufficiently far into the coding sequences to include the SOS boxes. We then introduced base changes into the SOS boxes by site-directed mutagenesis at the +/− four position in the consensus repeat (changing the T/A to C/G), which are known to prevent LexA binding (9). Following introduction of the constructs into *M. tuberculosis*, the expression levels directed by the various inserts with and without DNA damage were determined by assaying the resulting β-galactosidase activity. In each case, the wild-type promoter sequence exhibited DNA damage-inducible expression on the reporter gene (Fig. 2). Where DNA damage induction is known to be dependent on RecA (*Rv3395c*, *Rv1378c*, *Rv3074*, and *Rv1000c* (7)), mutation of the SOS box resulted in reporter gene expression that was no longer increased by exposure to DNA damage. In fact, the resulting constitutive level of expression was highly elevated, even relative to the induced level seen with the wild-type promoter. This indicates that, perhaps unsurprisingly, the level of DNA damage caused by the treatment used was insufficient to fully inactivate all of the LexA in the cell. Taken together, these data confirmed that the LexA-binding sites in question were required for DNA damage regulation.

DNA damage induction of *Rv2100* is partially dependent on RecA and has two potential SOS boxes each with one mismatch to the consensus, one located 16 bp downstream of the annotated start site (box 2) and the other 106 bp downstream (box 1). ChIP-seq analysis revealed the LexA binding peak only covered box 1, and only box 1 was detected by FIMO analysis using the modified SOS motif from MEME analysis of LexA binding peaks. So box 2 is unlikely to be functional. Mutation of each SOS box individually in the *lacZ* reporter promoter fusion vector for *Rv2100* resulted in reporter gene expression that could still be induced by DNA damage, as the promoter could still be regulated by the RecA-independent mechanism. However, mutation of box 1 but not box 2 resulted in higher expression levels for both uninduced and induced cultures, and the uninduced level was essentially the same as the wild-type promoter after induction (Fig. 2E). Mutation of both boxes 1 and 2 resulted in activity similar to mutation of box 1 only (data not shown). Therefore, LexA binds to box 1 of *Rv2100*, and box 2 is not functional.

Next, we mapped the transcriptional start sites of *Rv3395c*, *Rv1378c*, *Rv3074*, *Rv1000c*, and *Rv2100* by 5′-rapid amplification of cDNA ends. The transcriptional start site for *Rv3395c* was found to be 7 bp upstream of the annotated start codon, within the SOS box (Fig. 2A). Transcription of *Rv1378c* was

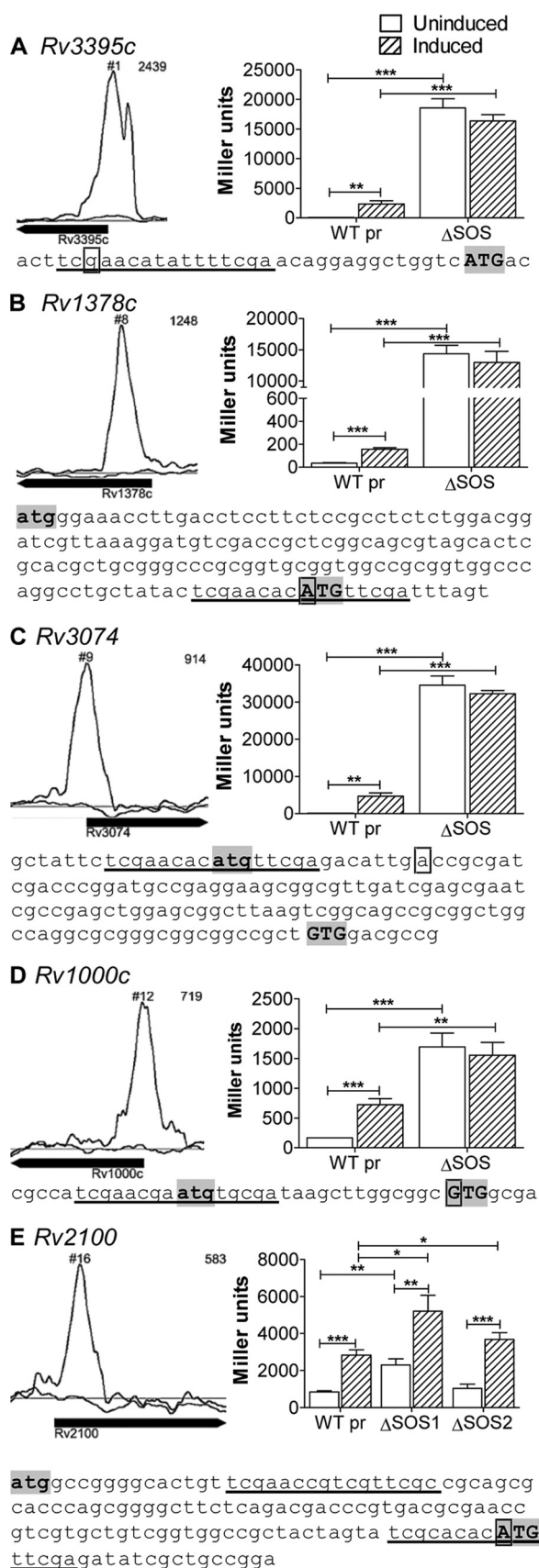


FIGURE 2. LexA-dependent DNA damage induction of *Rv3395c* (A), *Rv1378c* (B), *Rv3074* (C), *Rv1000c* (D), and *Rv2100* (E). For each gene, the localized genome coverage plot, β -galactosidase activity of promoter fusion constructs, and DNA sequence of promoter region are shown. Genome cov-

erage plots normalized to the total number of reads for individual peaks show location relative to annotated coding regions. *Black line* indicates immunoprecipitated sample; *dark gray line* indicates input control; number in *upper right corner* of each plot represents the scale; plots are the average of three independent replicates. For β -galactosidase activity of transcriptional *lacZ* fusion constructs of either the wild-type promoter sequence (*WT pr*) or promoters containing mutations of the SOS box (Δ SOS) expressed in *M. tuberculosis* H37Rv, induction is for 24 h with 0.2 μ g ml⁻¹ mitomycin C; data represent mean \pm S.D. from at least three replicates statistical significance is by two-tailed *t* test (*p* values: *, <0.1; **, <0.01; and ***, <0.001). Promoter DNA sequences show positions of the experimentally determined transcriptional start sites in a *box*; the most probable translational start site is in *shaded capital letters*; the annotated translational start site if different is shown in *shaded boldface*; and the location of the SOS box *underlined*.

shown to start 138 bp downstream of the annotated start site, indicating the annotated translational start codon to be incorrect. Translation most likely starts at the same site as the transcriptional start site, within the SOS box (Fig. 2B). Transcription of *Rv3074* starts 15 bp downstream from the annotated translational start, indicating translation most likely starts at the next GTG codon, 123 bp from the annotated start (Fig. 2C). The transcriptional start site of *Rv1000c* was shown to be 21 bp downstream of the annotated start codon and coincides with a GTG, the probable actual start codon (Fig. 2D). Transcriptional start site mapping of *Rv2100* showed that the transcription starts 248 bp downstream of the annotated translational start codon, within the functional SOS box (Fig. 2E). It is likely that translation also starts at this site. Therefore, in each case the SOS box is actually located upstream of the coding region or at the translated start site of the gene.

Sites Associated with Genes Not Linked with DNA Damage Induction—Six of the LexA binding peaks were found to be associated with genes not known to be induced by DNA-damaging agents (peaks 6, 13, 15, 22, 23, and 25). The peak located between *whiB2* and *fbiA* (peak 6), was previously shown to contain an SOS box with one mismatch to the consensus sequence but was not thought to be functional (9); however, it is the sixth highest ranking LexA-binding peak. Two LexA-binding peaks (peaks 13 and 25) were found to be associated with *Rv1057*, but only one of these (peak 13) contains an identifiable SOS box. This SOS box is located 520 bp upstream of the start codon of *Rv1057*; the intervening region contains sequences associated with DNA bending (31), suggesting that it is possible that the smaller downstream peak 25 does not represent direct binding of LexA but indirect association from the SOS box of peak 13.

To determine whether LexA regulates the expression of these three genes, induction by the DNA-damaging agent mitomycin C was assessed in both wild-type *M. tuberculosis* strain 1424 and a Δ *recA* mutant, in which autocatalytic cleavage of LexA following DNA damage does not occur. Expression of the RecA/LexA-dependent *Rv3074* and RecA/LexA-independent *uvrA* was determined as the control; *uvrA* shows induction by mitomycin C in both strains, and expression of *Rv3074* in the Δ *recA* strain is equivalent to the uninduced wild-type expression regardless of treatment (Fig. 3A). Expression of *fbiA* showed a slight induction by mitomycin C in the Δ *recA* strain but no induction in the wild-type strain; in contrast, expression of *whiB2* was reduced after exposure to mitomycin C in the wild-type strain only and was higher in the Δ *recA* strain than in the wild type (Fig. 3B). This indicates that peak 6 regulates

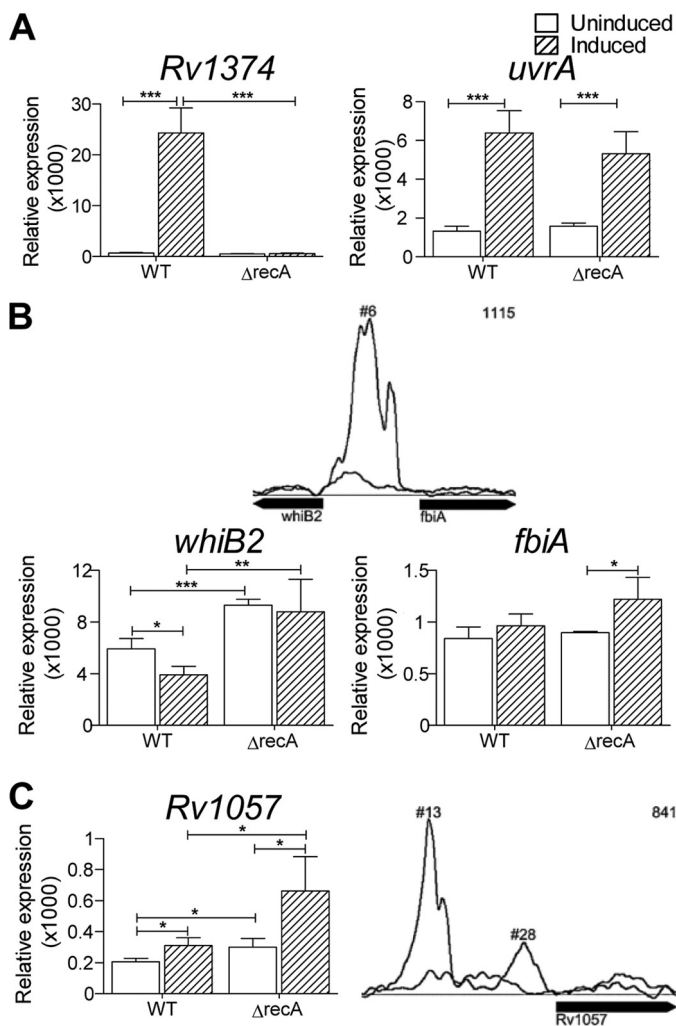


FIGURE 3. DNA damage induction in wild-type (WT) and Δ recA strains of *Rv3074* and *uvrA* (A), *whiB2* and *fbiA* (B), and *Rv1057* (C). For each relative expression, levels are shown, and the localized genome coverage plots are shown for B and C. Genome coverage plots normalized to the total number of reads for individual peaks show location relative to annotated coding regions; black line indicates immunoprecipitated sample; dark gray line indicates input control; number in upper right corner of each plot represents the scale, and plots show average of three independent replicates. Relative expression shows RNA levels normalized to 16 S rRNA of each gene in either the wild-type (WT) or Δ recA strains; induction is 24 h with $0.02 \mu\text{g ml}^{-1}$ mitomycin C; data represents mean \pm S.D. from at least three replicates, and statistical significance was by two-tailed t test (p values: *, <0.1 ; **, <0.01 ; and ***, <0.001).

whiB2 rather than *fbiA* and also that in this case LexA acts as an activator; the increased expression in the Δ recA strain indicates that even without induction some of the RecA in the wild-type strain is active. Expression of *Rv1057* showed a slight but significant induction by mitomycin C; however, expression in the Δ recA strain remained inducible and was higher than in the wild-type (Fig. 3C), indicating that regulation of *Rv1057* is more complex.

Binding Peaks Associated with Small RNAs—LexA binding peaks not found to be regulating the nearest annotated protein-encoded gene could instead be regulating a previously unannotated gene. This is possibly the case for peaks 15 and 23 that are located within recently identified small RNAs (32). Peak 15 is located within MTS1082, encoded opposite *Rv1374c*, which is

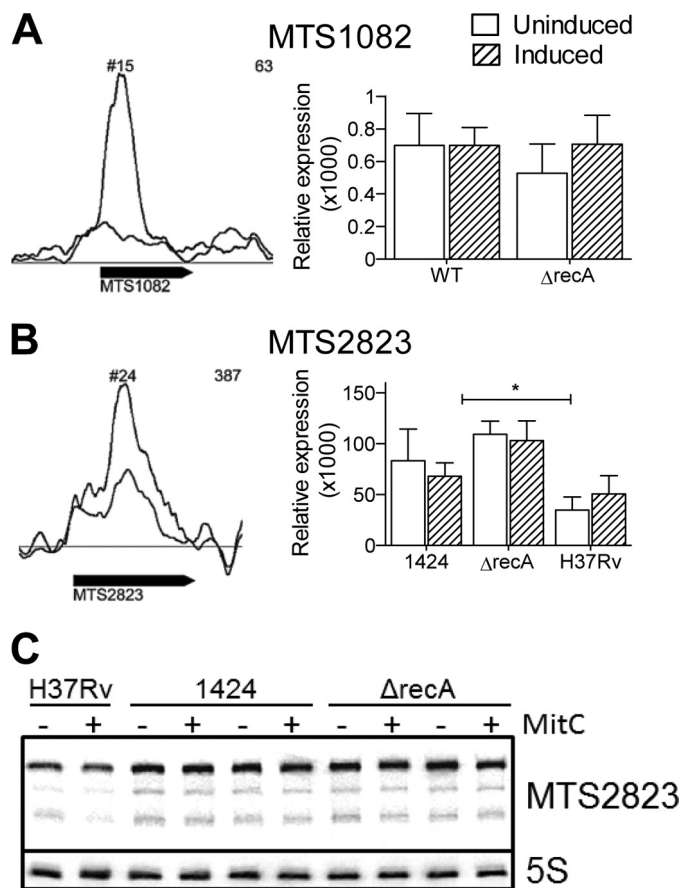


FIGURE 4. DNA damage induction in potentially LexA-regulated small RNAs MTS1082 (A) and MTS2823 (B) is shown. For each small RNA, the localized genome coverage plot and relative expression levels are shown. Genome coverage plots normalized to the total number of reads for individual peaks show location relative to the small RNA encoding gene; black line indicates immunoprecipitated sample; dark gray line indicates input control; number in upper right corner of each plot represents the scale, and plots average the three independent replicates. Relative expression shows RNA levels normalized to 16 S rRNA of each gene in either the wild-type (WT) or Δ recA strains in strain 1424 for MTS1082 (A), or wild-type 1424 and H37Rv, and Δ recA strains for MTS2823 (B). Induction is 24 h with $0.02 \mu\text{g ml}^{-1}$ mitomycin C; data represent mean \pm S.D. from at least three replicates, and statistical significance is shown by two-tailed t test (p values: *, <0.1 ; **, <0.01 , and ***, <0.001). C, Northern blot analysis showing the presence of MTS2823 in wild-type 1424, H37Rv, and Δ recA with and without addition of mitomycin C (MitC), and level of 5 S RNA in each sample is shown as a control.

thought not to be expressed (32). Expression of MTS1082 showed no significant differences following mitomycin C exposure in wild-type *M. tuberculosis* 1424 or in Δ recA according to quantitative RT-PCR (Fig. 4A). Similarly, there was no significant increase in expression of MTS2823, associated with peak 23, after mitomycin C exposure in wild-type 1424 or in Δ recA (Fig. 4B). However, both the basal and induced levels of MTS2823 were higher in Δ recA (Fig. 4B). We also compared induction by mitomycin C in 1424 and H37Rv by quantitative PCR. Expression of MTS2823 was lower in H37Rv compared with 1424. Although no induction by mitomycin C was observed in 1424, a slight, but not significant, increase in expression was seen in H37Rv. Northern blot confirmed the higher level of expression in 1424 compared with H37Rv (Fig. 4C). The binding peak located at MTS2823 does not contain an identifiable SOS box, and so LexA binding is likely occurring via another unidentified factor.

DISCUSSION

For mycobacteria, the majority of DNA damage-inducible genes, including many genes known to be involved in DNA repair, are not controlled by LexA and instead are controlled by a RecA-independent mechanism controlled by ClpR (7, 12). Previous identification of the LexA consensus binding sequence identified 15 apparently functional binding sites containing up to one mismatch. A further 351 sites in the *M. tuberculosis* genome contain two mismatches, and three functional SOS boxes have been identified that regulate *Rv2719c*, which contain two and three mismatches, and so there are many more potential LexA-binding sites (9, 26). To determine LexA-binding sites *in vivo*, we performed ChIP-seq experiments, which identified 25 LexA-binding sites throughout the *M. tuberculosis* genome, including all 16 previously known. Of these 25 binding sites, we found that all but two shared a common motif, which matched the previously defined SOS box (9). The common motif identified in this study was longer than that originally defined, showing that some of the surrounding nucleotides, although less critical in transcription reporter assays, are highly conserved. LexA is able to bind to sites with up to three mismatches to the original consensus and with mismatches not previously thought to be tolerated, which may indicate interaction with other factors.

Of the nine novel binding sites identified in this study, one was previously recognized but not thought to be functional due to the position of the single mismatch to the consensus, coupled with the lack of induction of the associated genes (peak 6, *whiB2*) (9). In this study, we now show that LexA does in fact bind to this region and positively regulates expression of *whiB2*. Although LexA generally acts as a transcriptional repressor, activation of gene expression by LexA has been reported previously. In *Corynebacterium glutamicum*, three SOS boxes were found to activate expression of four genes; in the cyanobacterium *Synechocystis* sp. PCC 6803 LexA positively regulates the *hox* operon, and in *Rhodobacter sphaeroides* positive or negative regulation of *recA* was shown to correlate with LexA abundance (34–36).

Of the remaining novel binding sites, four were associated with DNA damage-inducible genes and contained identifiable SOS boxes (peaks 13, 19, 21, and 24, regulating *Rv1057*, *Rv0095c*, *Rv1588c*, and *ruvA/B*, respectively); one was associated with DNA damage induction with no SOS box (peak 25 also regulating *Rv1057*), and three were associated with genes that are not known to be induced by DNA damage (peaks 15, 22, and 23 associated with MTS1082, *Rv2517c*, and MTS2823).

Peaks 25 and 23 do not contain any identifiable binding consensus for LexA. Binding without a consensus sequence has been shown to occur with other transcription factors. Examples include fumarate and nitrate reductase regulator (FNR) of *E. coli* in which 10 of the 63 identified binding sites had no identifiable binding consensus (37). Furthermore, with *E. coli* LexA, 19 novel sites lacking a consensus binding site were identified, and these sites were associated with genes that did not show DNA damage induction (38). Computer-generated analysis of the region upstream of *Rv1057* predicted that it contains regions of DNA curvature (31). Our evidence agrees with this as

two LexA-binding sites were identified within this region with only the more distal one (located 520 bp from the start codon, immediately upstream of the regions of DNA curvature) containing the SOS box. It is likely that the smaller second peak (which overlaps the transcriptional start site for *Rv1057*) is where LexA binding to the SOS box located at the larger peak bends around and comes into contact with the DNA. Positive regulation of the *hox* operon in the cyanobacterium *Synechocystis* was due to LexA binding 600–700 bp upstream of the translational start site, and this was also assumed to be mediated via DNA looping (36).

Binding without apparent regulation has been shown for *E. coli* transcription factors such as LexA, RutR, PurR, and FNRs, where binding sites have been identified with no associated effects on the transcriptome (37–40). Also, 60% of the identified binding sites for the alternative σ factor, SigF, of *M. tuberculosis* were not linked with effects on transcription (41). Here, five LexA binding peaks were identified with genes that are not DNA damage-inducible or the induction remained in a RecA mutant strain, in which the LexA repression should not be able to be lifted. Regulation of these genes could involve multiple factors so that removal of LexA repression alone is insufficient for induction or that LexA remains bound to these specific promoters under the conditions tested. Alternatively, such binding sites could regulate genes located distant to the binding site on the chromosome but physically located closer due to folding of the DNA or could regulate previously unannotated genes or small RNAs. One such FNR-binding site was subsequently found to regulate a small RNA (37, 42). Also, a SigF binding peak was found to be associated with the small RNA F6 (25, 41). Recent analysis of the transcriptome of *M. tuberculosis* identified a number of previously unannotated transcripts, many of which represent small RNAs, including MTS1082 and MTS2823 which are both associated with LexA-binding sites (32). In addition, LexA was shown to bind upstream of *Rv1057*, which showed some DNA damage induction in wild-type *M. tuberculosis*. However in a *recA* mutant, this induction still occurred indicating that its regulation is more complex. This gene is induced by envelope stress and regulated by two different two-component regulatory systems, MprAB and TrcRS (31, 43). Alternatively, this LexA-binding site, due to being located distant from the start codon of *Rv1057*, could regulate a different factor. There are no other annotated genes in the region between *Rv1057* and *Rv1056* of *M. tuberculosis* strain H37Rv; however, it does contain a small unannotated ORF, which is annotated in the equivalent region of *M. tuberculosis* strain CDC1551. This ORF encodes a hypothetical protein that is a member of the 13E12 repeat family; however, it is not known whether it is expressed (43, 44).

Binding of transcription factors to intragenic regions is also increasingly being recognized due to ChIP analyses. For example, 14 of 20 identified binding sites for RutR were found to be internal to coding regions (39), and 27 of 67 SigF-binding sites were also intragenic (41). In this study, we performed transcriptional reporter fusion assays to confirm LexA-dependent DNA damage induction for five of these sites where either the SOS box was found to be located internal to the coding sequence or where the gene was originally misannotated (29, 30). This con-

Mycobacterium tuberculosis LexA Regulon

firmed the re-annotation of *Rv3395c* and revealed that the annotated start sites of *Rv1378c*, *Rv3074*, *Rv1000c*, and *Rv2100* are incorrect and are most likely located 138, 123, 21, and 248 bp downstream, respectively. Because of the high GC content of the *M. tuberculosis* genome and the prevalence of GTG as a start codon, there are many more start than stop codons, and so misannotation of coding regions is common, as is also shown for LexA itself (45). The LexA-binding motif for *E. coli* was found to be restricted to promoter regions and to be sufficient for binding, as artificial insertion of the motif elsewhere on the genome led to binding (38). This indicates that the *E. coli* genome is permissive for transcription factor binding, in contrast to eukaryotic genomes where histones prevent inappropriate binding outside of promoters. Of the 40 sites identified in this study that contain the modified LexA-binding motif, 13 showed no evidence of binding LexA, all of which were intragenic, indicating that *M. tuberculosis* LexA binding may be restricted to promoter regions and that the *M. tuberculosis* genome may not be as permissive to transcription factor binding as the *E. coli* genome. Alternatively, LexA may require other factors for binding to lower affinity motifs, which are not present at these 13 sites.

There is very little overlap of LexA-regulated genes between *E. coli* and *M. tuberculosis*, as the majority of genes that are up-regulated by DNA damage in *M. tuberculosis* are regulated independently of LexA/RecA (7, 38). In both species LexA does, however, regulate itself, *recA*, *dnaE2* (*dinB* in *E. coli*), *ruvAB*, and *imuAB* (*umuDC* in *E. coli*). DnaE2 is the error-prone α subunit of DNA polymerase III essential to DNA damage-induced mutagenesis (46), and the *Rv3395/4c* operon, encoding ImuA' and ImuB, forms a complex with DnaE2 (47). The *ruvCAB* operon is involved in recombination repair. In *E. coli*, this operon lacks *ruvC*; interestingly, we show here that in *M. tuberculosis* this operon is regulated by two LexA-binding sites. One is located within the *ruvC* promoter, although the RecA_NDP is responsible for the majority of its regulation (9, 27). The second LexA-binding site is internal to *ruvC* and possibly just regulates *ruvAB*, whose DNA damage induction is dependent on RecA/LexA (7). The genes *lexA*, *dnaE2*, *imuA*, and *imuB* are thought to represent the evolutionary core regulon of LexA. Bacteria from many different phyla contain these cassettes in many different combinations, but they are all associated with LexA regulation (33, 48). For *M. tuberculosis*, these four genes are formed from three different transcriptional units, all regulated by LexA.

Of the other genes regulated by LexA in *M. tuberculosis*, a large number are members of the 13E12 repeat family, which are insertion elements containing phage attachment sites. Only one of these attachment sites, located within the biotin operon in H37Rv (starting at *Rv1588c*), has a phage inserted within it (49). DNA damage induces phage production, and in many bacteria this induction is often linked to the SOS response either through RecA-dependent cleavage of a repressor protein or LexA-mediated repression of an anti-repressor (50). In *M. tuberculosis*, the majority of the 13E12 family is controlled directly by LexA, regardless of phage insertion. The SOS response also regulates genes involved in inhibition of cell division, which functions to prevent cell separation before the DNA

is repaired and replicated (33, 51). *M. tuberculosis* LexA binds to the promoters of two genes known to function in cell division, negatively regulating *Rv2719c*, a cell division suppressor (52), and positively regulating *whiB2*, shown to be essential for cell division in *M. smegmatis* (53).

In addition, this investigation identified LexA-binding sites associated with genes encoding two recently identified small RNAs, MTS1082 and MTS2823 (25, 32). MTS1082 is located on the opposite strand of *Rv1374c*, annotated as a hypothetical protein; however, transcriptional profiling revealed that this ORF is unlikely to be expressed (32). MTS2823 is the most abundant transcript in *M. tuberculosis*, apart from the ribosomal RNA, and its expression is increased 10-fold in stationary phase. Overexpression of this small RNA results in a global down-regulation of gene expression, particularly of genes involved in energy metabolism (32). Thus, LexA regulation of this small RNA may have the consequence of diversifying the effect of LexA. Alternatively, as no SOS box was identified in this region of enrichment, LexA binding could be an artifact associated with the high level of transcription at this part of the chromosome.

In this study we identified the regulon of *M. tuberculosis* LexA on a global scale *in vivo* using ChIP-seq. Most of the identified LexA-binding sites were associated with identifiable SOS boxes and were located in the promoters of genes up-regulated by DNA damage. However, novel binding sites were identified that were not linked with SOS boxes and associated with genes that were not induced by DNA damage. Plus in one case LexA binding was found to result in down-regulation by DNA damage. This highlights that regulation by LexA in *M. tuberculosis* is much more complex than previously recognized.

Acknowledgments—We thank Dr. Alan Williams for purification of LexA for antibody production and Patricia Brooks for technical assistance. We also thank Dr. Abdul Sesay and Dr. Harsha Jani at the Medical Research Council National Institute for Medical Research high throughput sequencing service.

REFERENCES

1. World Health Organization (2011) Global tuberculosis control. *WHO Report 2010*, World Health Organization, Geneva, Switzerland
2. Burney, S., Caulfield, J. L., Niles, J. C., Wishnok, J. S., and Tannenbaum, S. R. (1999) The chemistry of DNA damage from nitric oxide and peroxynitrite. *Mutat. Res.* **424**, 37–49
3. Nathan, C., and Shiloh, M. U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8841–8848
4. Dos Vultos, T., Mestre, O., Tonjum, T., and Gicquel, B. (2009) DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS Microbiol. Rev.* **33**, 471–487
5. Brooks, P. C., Movahedzadeh, F., and Davis, E. O. (2001) Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*. Apparent lack of correlation with LexA binding. *J. Bacteriol.* **183**, 4459–4467
6. Davis, E. O., Springer, B., Gopaul, K. K., Papavinasundaram, K. G., Sander, P., and Böttger, E. C. (2002) DNA damage induction of *recA* in *Mycobacterium tuberculosis* independently of RecA and LexA. *Mol. Microbiol.* **46**, 791–800
7. Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R. S., and Davis, E. O. (2003) The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of RecA. *Mol. Microbiol.* **50**, 1031–1042

8. Davis, E. O., and Forse, L. N. (2009) in *Mycobacterium: Genomics and Molecular Biology* (Parish, T., and Brown, A., eds) pp. 79–119, Caister Academic Press, London
9. Davis, E. O., Dullaghan, E. M., and Rand, L. (2002) Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J. Bacteriol.* **184**, 3287–3295
10. Movahedzadeh, F., Colston, M. J., and Davis, E. O. (1997) Characterization of *Mycobacterium tuberculosis* LexA. Recognition of a Cheo (Bacillus-type SOS) box. *Microbiology* **143**, 929–936
11. Gamulin, V., Cetkovic, H., and Ahel, I. (2004) Identification of a promoter motif regulating the major DNA damage response mechanism of *Mycobacterium tuberculosis*. *FEMS Microbiol. Lett.* **238**, 57–63
12. Wang, Y., Huang, Y., Xue, C., He, Y., and He, Z. G. (2011) A ClpR-like regulator specifically recognizes a RecA-independent promoter motif and broadly regulates expression of DNA damage-inducible genes in mycobacteria. *J. Biol. Chem.* **286**, 31159–31167
13. Sambrook, J., Fritsch, E., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
14. Oatway, W. H. J., and Steenken, W. J. (1936) The pathogenesis and fate of tubercle produced by dissociated variants of tubercle bacilli. *J. Infect. Dis.* **59**, 306–325
15. Miller, J. (1972) *Experiments in Molecular Genetics*, pp. 352–355, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
16. Sala, C., Haouz, A., Saul, F. A., Miras, I., Rosenkrands, I., Alzari, P. M., and Cole, S. T. (2009) Genome-wide regulon and crystal structure of Blal (Rv1846c) from *Mycobacterium tuberculosis*. *Mol. Microbiol.* **71**, 1102–1116
17. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25
18. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
19. Quinlan, A. R., and Hall, I. M. (2010) BEDTools. A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842
20. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. (2000) Artemis. Sequence visualization and annotation. *Bioinformatics* **16**, 944–945
21. Spyrou, C., Stark, R., Lynch, A. G., and Tavaré, S. (2009) Bayes Peak. Bayesian analysis of ChIP-seq data. *BMC Bioinformatics* **10**, 299
22. Cairns, J., Spyrou, C., Stark, R., Smith, M. L., Lynch, A. G., and Tavaré, S. (2011) Bayes Peak. An R package for analysing ChIP-seq data. *Bioinformatics* **27**, 713–714
23. Bailey, T. L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36
24. Grant, C. E., Bailey, T. L., and Noble, W. S. (2011) FIMO. Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018
25. Arnvig, K. B., and Young, D. B. (2009) Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **73**, 397–408
26. Dullaghan, E. M., Brooks, P. C., and Davis, E. O. (2002) The role of multiple SOS boxes upstream of the *Mycobacterium tuberculosis* *lexA* gene. Identification of a novel DNA damage-inducible gene. *Microbiology* **148**, 3609–3615
27. Dawson, L. F., Dillury, J., and Davis, E. O. (2010) RecA-independent DNA damage induction of *Mycobacterium tuberculosis* *rvuC* despite an appropriately located SOS box. *J. Bacteriol.* **192**, 599–603
28. Brooks, P. C., Dawson, L. F., Rand, L., and Davis, E. O. (2006) The *Mycobacterium*-specific gene Rv2719c is DNA damage-inducible independently of RecA. *J. Bacteriol.* **188**, 6034–6038
29. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544
30. Camus, J. C., Pryor, M. J., Médigue, C., and Cole, S. T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**, 2967–2973
31. Haydel, S. E., and Clark-Curtiss, J. E. (2006) The *Mycobacterium tuberculosis* TrcR response regulator represses transcription of the intracellularly expressed Rv1057 gene, encoding a seven-bladed β -propeller. *J. Bacteriol.* **188**, 150–159
32. Arnvig, K. B., Comas, I., Thomson, N. R., Houghton, J., Boshoff, H. I., Croucher, N. J., Rose, G., Perkins, T. T., Parkhill, J., Dougan, G., and Young, D. B. (2011) Sequence-based analysis uncovers an abundance of noncoding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog.* **7**, e1002342
33. Erill, I., Campoy, S., and Barbé, J. (2007) Eons of distress. An evolutionary perspective on the bacterial SOS response. *FEMS Microbiol. Rev.* **31**, 637–656
34. Tapias, A., Fernández, S., Alonso, J. C., and Barbé, J. (2002) *Rhodobacter sphaeroides* LexA has dual activity. Optimizing and repressing *recA* gene transcription. *Nucleic Acids Res.* **30**, 1539–1546
35. Jochmann, N., Kurze, A. K., Czaja, L. F., Brinkrolf, K., Brune, I., Hüser, A. T., Hansmeier, N., Pühler, A., Borovok, I., and Tauch, A. (2009) Genetic make-up of the *Corynebacterium glutamicum* LexA regulon deduced from comparative transcriptomics and *in vitro* DNA band shift assays. *Microbiology* **155**, 1459–1477
36. Gutekunst, K., Phunpruch, S., Schwarz, C., Schuchardt, S., Schulz-Friedrich, R., and Appel, J. (2005) LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator. *Mol. Microbiol.* **58**, 810–823
37. Grainger, D. C., Aiba, H., Hurd, D., Browning, D. F., and Busby, S. J. (2007) Transcription factor distribution in *Escherichia coli*. Studies with FNR protein. *Nucleic Acids Res.* **35**, 269–278
38. Wade, J. T., Reppas, N. B., Church, G. M., and Struhl, K. (2005) Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites. *Genes Dev.* **19**, 2619–2630
39. Shimada, T., Ishihama, A., Busby, S. J., and Grainger, D. C. (2008) The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res.* **36**, 3950–3955
40. Cho, B. K., Federowicz, S. A., Embree, M., Park, Y. S., Kim, D., and Palsson, B. (2011) The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res.* **39**, 6456–6464
41. Hartkoorn, R. C., Sala, C., Uplekar, S., Busso, P., Rougemont, J., and Cole, S. T. (2012) Genome-wide definition of the SigF Regulon in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194**, 2001–2009
42. Boysen, A., Møller-Jensen, J., Kallipolitis, B., Valentin-Hansen, P., and Overgaard, M. (2010) Translational regulation of gene expression by an anaerobically induced small noncoding RNA in *Escherichia coli*. *J. Biol. Chem.* **285**, 10690–10702
43. Pang, X., Cao, G., Neuenschwander, P. F., Haydel, S. E., Hou, G., and Howard, S. T. (2011) The β -propeller gene Rv1057 of *Mycobacterium tuberculosis* has a complex promoter directly regulated by both the MprAB and TrcRS two-component systems. *Tuberculosis* **91**, S142–S149
44. Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Umayam, L. A., Ermolaeva, M., Salzberg, S. L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs, W. R., Jr., Venter, J. C., and Fraser, C. M. (2002) Whole genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* **184**, 5479–5490
45. Smollett, K. L., Fivian-Hughes, A. S., Smith, J. E., Chang, A., Rao, T., and Davis, E. O. (2009) Experimental determination of translational start sites resolves uncertainties in genomic open reading frame predictions. Application to *Mycobacterium tuberculosis*. *Microbiology* **155**, 186–197
46. Boshoff, H. I., Reed, M. B., Barry, C. E., 3rd, and Mizrahi, V. (2003) DnaE2

***Mycobacterium tuberculosis* LexA Regulon**

- polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* **113**, 183–193
47. Warner, D. F., Ndwandwe, D. E., Abrahams, G. L., Kana, B. D., Machowski, E. E., Venclovas, C., and Mizrahi, V. (2010) Essential roles for *imuA'*- and *imuB*-encoded accessory factors in DnaE2-dependent mutagenesis in *Mycobacterium tuberculosis*. *Proc. Nat. Acad. Sci. U.S.A.* **107**, 13093–13098
 48. Erill, I., Campoy, S., Mazon, G., and Barbé, J. (2006) Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic Acids Res.* **34**, 66–77
 49. Gordon, S. V., Heym, B., Parkhill, J., Barrell, B., and Cole, S. T. (1999) New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **145**, 881–892
 50. Waldor, M. K., and Friedman, D. I. (2005) Phage regulatory circuits and virulence gene expression. *Curr. Opin. Microbiol.* **8**, 459–465
 51. Huisman, O., D'Ari, R., and Gottesman, S. (1984) Cell division control in *Escherichia coli*. Specific induction of the SOS function SfiA protein is sufficient to block septation. *Proc. Nat. Acad. Sci. U.S.A.* **81**, 4490–4494
 52. Chauhan, A., Lofton, H., Maloney, E., Moore, J., Fol, M., Madiraju, M. V., and Rajagopalan, M. (2006) Interference of *Mycobacterium tuberculosis* cell division by Rv2719c, a cell wall hydrolase. *Mol. Microbiol.* **62**, 132–147
 53. Gomez, J. E., and Bishai, W. R. (2000) *whmD* is an essential mycobacterial gene required for proper septation and cell division. *Proc. Nat. Acad. Sci. U.S.A.* **97**, 8554–8559