Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites

Joseph T. Wade,[1,4] Nikos B. Reppas,[2,3,4] George M. Church,[3] and Kevin Struhl[1,5]

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard University, Boston, Massachusetts 02115, USA; [2]Graduate Biophysics Program and [3]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

Genomes of eukaryotic organisms are packaged into nucleosomes that restrict the binding of transcription factors to accessible regions. Bacteria do not contain histones, but they have nucleoid-associated proteins that have been proposed to function analogously. Here, we combine chromatin immunoprecipitation and high-density oligonucleotide microarrays to define the in vivo DNA targets of the LexA transcriptional repressor in *Escherichia coli*. We demonstrate a near-universal relationship between the presence of a LexA sequence motif, LexA binding in vitro, and LexA binding in vivo, suggesting that a suitable recognition site for LexA is sufficient for binding in vivo. Consistent with this observation, LexA binds comparably to ectopic target sites introduced at various positions in the genome. We also identify ~20 novel LexA targets that lack a canonical LexA sequence motif, are not bound by LexA in vitro, and presumably require an additional factor for binding in vivo. Our results indicate that, unlike eukaryotic genomes, the *E. coli* genome is permissive to transcription factor binding. The permissive nature of the *E. coli* genome has important consequences for the nature of transcriptional regulatory proteins, biological specificity, and evolution.

[*Keywords*: LexA; chromatin immunoprecipitation; genome organization; transcription; RNA polymerase]

Supplemental material is available at http://www.genesdev.org.

Received July 14, 2005; revised version accepted September 6, 2005.

For organisms to express genes in a developmentally and environmentally appropriate fashion, it is essential that transcriptional regulatory proteins selectively associate with biologically relevant target DNA sequences in vivo. Two general principles explain how sequence-specific DNA-binding proteins selectively associate with relevant target sites in vivo. First, intrinsic DNA-binding specificity is determined by complementary surfaces with energetically favorable contacts between amino acids and bases. For simple protein–DNA interactions, optimal binding occurs at a consensus DNA sequence motif that can be expressed as a probability matrix of independent nucleotide positions. In more complex situations, DNA-binding specificity is altered by cooperative interactions between two or more proteins bound to DNA. Second, target recognition in vivo depends on accessibility of the DNA. As a consequence of differential accessibility, the same DNA sequence in two different genomic locations can be bound to very different extents.

[4]These authors contributed equally to this work.
[5]Corresponding author.
E-MAIL kevin@hms.harvard.edu; FAX (617) 432-2529.
Article and publication are at http://www.genesdev.org/cgi/doi/10.1101/gad.1355605.

DNA accessibility plays a very significant role in eukaryotic organisms because genomic DNA is packaged by histones into nucleosomal arrays and more complex chromatin structures. In general, nucleosomes severely inhibit accessibility of DNA to proteins, with the degree of inhibition depending on the specific protein (Felsenfeld 1996; Workman and Kingston 1998; Struhl 1999). Thus, DNA sequences located in linker regions between nucleosomes are more accessible than sequences wrapped around histones. Preferentially accessible regions in many eukaryotic organisms have been identified by DNase I hypersensitivity or by other enzymatic probes of chromatin structure.

In the yeast *Saccharomyces cerevisiae*, promoter regions are relatively depleted of nucleosomes in comparison to protein-coding regions (Ng et al. 2003; Bernstein et al. 2004; Lee et al. 2004; Sekinger et al. 2005). This genomic organization ensures that transcription factors bind preferentially to cognate sites in promoters, rather than to the excess of functionally irrelevant sites in nonpromoter regions (Sekinger et al. 2005). For example, HinfI endonuclease cleavage (Mai et al. 2000) and Rap1 binding (Lieb et al. 2001) in vivo are far more efficient at promoter regions, despite the existence of numerous

consensus sequences in coding regions. Notably, this distinction between promoter and coding regions with respect to histone density is largely independent of transcriptional activity, often reflecting differences in intrinsic histone–DNA interactions (Sekinger et al. 2005). In addition, transcriptional activator proteins bound at enhancer elements can evict histones from DNA, presumably by recruiting nucleosome remodeling complexes, and histone-modifying enzymes, thereby expanding the region of low histone density beyond the direct protein-recognition site (Deckert and Struhl 2001; Boeger et al. 2003, 2004; Reinke and Horz 2003). In human cells, unbiased identification of p53, Sp1, and Myc target sites on Chromosomes 21 and 22 indicates that a very small subset (~1%) of consensus motifs are actually bound by the cognate protein in vivo (Cawley et al. 2004).

Prokaryotic cells do not contain histones, but their genomes are associated with histone-like proteins in a structure termed the nucleoid (Dame 2005), and the chromosome is organized into discrete macrodomains (Boccard et al. 2005). Histone-like proteins H-NS, HU, Fis, and IHF can repress transcription from several promoters in *Escherichia coli* (Dorman and Deighman 2003). It has been proposed that such proteins may package and compact prokaryotic genomes in a manner analogous to histones (Dame 2005) and thus may similarly restrict accessibility to DNA-binding proteins. An alternative view is that histone-like proteins do not globally restrict access of protein to DNA, and hence that prokaryotic genomes may be permissive to binding by transcription factors (Struhl 1999). However, little is known about the genome-wide association of histone-like proteins and their role in global DNA accessibility, and experimental information distinguishing between these two models is very limited.

More generally, the relationship between DNA sequence motifs, in vitro binding, and in vivo association of a DNA-binding protein has never been addressed comprehensively on a genome-wide level in a prokaryotic organism. Genome-wide identification of in vivo targets of bacterial DNA-binding transcriptional regulatory proteins has been performed in a few cases (Laub et al. 2002; Molle et al. 2003a,b; Eichenberger et al. 2004; Grainger et al. 2004), but these experiments typically involved microarrays containing PCR products representing only coding sequences, and hence do not represent an unbiased or comprehensive identification of in vivo targets of a DNA-binding protein. In experiments designed to address the topological domain structure of the *E. coli* chromosome, it was shown that expression of the restriction enzyme EcoRI in *E. coli* cells results in at least partial cleavage of the majority of recognition sites (Postow et al. 2004). However, EcoRI was overexpressed for considerable time in these experiments, and cleavage by a restriction enzyme requires only a single catalytic event that may occur over one or more generation times, such that effects on accessibility may have been masked. Indeed, similar experiments in *S. cerevisiae* have shown that the majority of EcoRI (Barnes and Rine 1985) and HinfI (Iyer and Struhl 1995; Mai et al. 2000) endonuclease recognition sites are at least partially cleaved, even though it is clear that the yeast genome is differentially permissive for binding transcription factors.

LexA directly regulates ~30 *E. coli* transcription units involved in the "SOS" response whose transcription is induced in response to DNA damage (Little and Mount 1982; Walker 1985). Under normal growth conditions, LexA binds to a specific 20-base-pair (bp) sequence within the promoter regions of these genes, repressing transcription by sterically occluding RNA polymerase (RNAP). Upon DNA damage, RecA bound to single-stranded DNA at blocked recombination forks stimulates LexA autoproteolysis, resulting in the derepression of LexA-regulated genes. Sequence analysis of LexA-regulated promoters revealed the consensus sequence $TACTG(TA)_5CAGTA$ as the binding site for LexA (Walker 1984), and computational analysis has identified additional LexA targets in the *E. coli* genome (Lewis et al. 1994; Fernandez de Henestrosa et al. 2000). In total, there are 27 identified LexA target promoters in *E. coli*. In all cases, LexA binds these targets in vitro and is required for repression of the target gene in vivo. There are also two putative LexA-binding sites, at the *minC* and *yigN* promoters, that have not been shown to be important in transcriptional regulation (Fernandez de Henestrosa et al. 2000), although microarray analysis suggests *yigN* may be regulated by LexA (Courcelle et al. 2001).

In this work, we use chromatin immunoprecipitation (ChIP) coupled with high-density microarrays (ChIP–chip) to identify targets for LexA across the whole genome. We demonstrate a remarkable correlation between the correspondence of a LexA sequence motif to the consensus, the extent of LexA binding in vitro, and LexA binding in vivo. Furthermore, we show that LexA binds comparably to an ectopic canonical LexA motif introduced at various positions in the genome. These observations suggest that a suitable recognition site for LexA is sufficient for binding in vivo, regardless of its genomic location. Thus, unlike the case in eukaryotic organisms, transcription factor association with DNA in *E. coli* is not controlled by DNA accessibility. We also identify ~20 novel LexA targets that lack a canonical LexA sequence motif and are not bound by LexA in vitro. Analysis of one such noncanonical target in the *ptrA* promoter region reveals an aberrant, but specific, version of a LexA motif, and it suggests that another factor is important for LexA binding in vivo.

Results

In vivo binding of LexA and RNAP to known LexA target promoters

Although LexA has been extensively characterized by DNA-binding experiments in vitro and mutational analysis in vivo, direct analysis of LexA binding in vivo has yet to be described. We therefore used ChIP and quantitative PCR to determine the in vivo association of LexA and the β subunit of RNAP with four known LexA target promoters—*recN*, *lexA*, *umuDC*, and *ruvA*. We

determined the association of LexA and RNAP relative to the *sgrR* coding region that should not bind LexA or RNAP and serves as a control. As expected, significant levels of LexA association are observed at all four promoters, and this association is significantly reduced following UV irradiation (Fig. 1A). Also, as expected, RNAP association with the four promoters increases significantly following UV irradiation (Fig. 1B).

To confirm that changes in LexA and RNAP binding upon DNA damage are due to LexA proteolysis, we analyzed an isogenic *lexA1* strain that harbors a mutant LexA resistant to UV-induced proteolysis. As expected, LexA association in this strain with the four promoters is high both before and after UV irradiation (Fig. 1C), and RNAP association is essentially unchanged (Fig. 1D). Taken together, the in vivo association of LexA and RNAP with these target regions is in complete accord with previous work, and provides a means to validate the whole-genome ChIP–chip analysis of LexA as described below.

Identification of LexA-bound regions on a genome-wide scale

In order to determine the in vivo DNA targets of endogenous LexA on a genomic scale, we performed a ChIP–chip analysis using high-density *E. coli* Affymetrix microarrays with 25-mer oligonucleotides that cover the entire *E. coli* genome at an average of one probe every 30 bp (Selinger et al. 2000). We compared the signal from two LexA ChIP experiments to each of two controls, identifying 49 statistically significant ($p < 0.001$) LexA

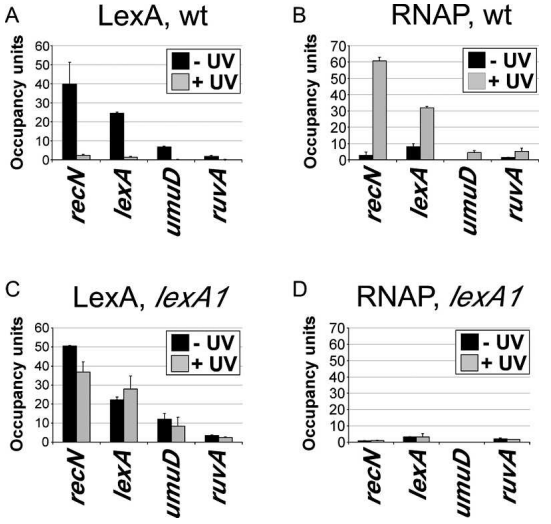

Figure 1. In vivo binding of LexA and RNAP β subunit to previously identified LexA targets. Association of LexA (*A*,*C*) and RNAP (*B*,*D*) with known LexA targets before and after UV irradiation (black and gray bars, respectively), in wild-type MG1655 (*A*,*B*) and an isogenic *lexA1* strain (*C*,*D*). Occupancy was measured as a ratio of binding of LexA or β to the tested region and to a control region located within the coding sequence of the predicted ORF *sgrR*.

targets. We define a target as a genomic region covered by a minimum of 20 consecutive oligonucleotide probes that reproducibly show binding of LexA above a defined threshold (see Materials and Methods). For each target, we computed its predicted LexA-binding site as the genomic coordinate under the maximal LexA IP/control $\log_2$ ratio, and its corresponding ChIP–chip score as essentially the value of this ratio.

The 49 LexA targets (Table 1; Supplementary Fig. 1) include all but two previously identified LexA sites. Due to the resolution afforded by the arrays, we were able to make relatively accurate predictions as to the location of the LexA-binding site at each target region identified: For previously identified LexA targets, our predictions were accurate to within an average of 167 bp of the known sites. Table 1 shows a ranked list of the LexA targets identified by ChIP–chip analysis with their corresponding ChIP–chip score, predicted LexA-binding site coordinate, and a measure of LexA-dependent transcriptional induction from Courcelle et al. (2001) for the target genes contained within. We have divided the 49 LexA targets into three classes: Class I comprises 25 targets experimentally determined in previous work; Class II comprises five novel targets with a canonical LexA motif (see below); and Class III comprises 19 novel targets that lack a canonical LexA motif (see below). For Class I and Class II LexA targets, there is a correlation between the ChIP–chip score and LexA-dependent UV-induction ($r = 0.52$).

Previous biochemical, genetic, and computational analyses identified 27 DNA targets for LexA (Fernandez de Henestrosa et al. 2000). Our genome-wide ChIP–chip analysis identified 25 of the Class I targets, the exceptions being the *dinS* and *ybfE* promoters. We determined whether LexA binds to these sites in vivo using ChIP and quantitative PCR, defining a true bound region as having more than twofold enrichment of target DNA relative to the control region in the *sgrR* coding sequence (equivalent to 1 Occupancy Unit; see Materials and Methods). This direct analysis of the *dinS* and *ybfE* promoters, in fact, shows no significant association of LexA before UV irradiation (Fig. 2A) and no significant change in RNAP association following UV irradiation (Fig. 2B) in either case. Similar results were seen with the *lexA1* strain (Fig. 2C,D). Additionally, neither the *dinS* nor the *ybfE* promoter showed any significant change in RNAP association following UV irradiation in the wild-type or *lexA1* strain (Fig. 2B,D). Thus, under the conditions tested, the *dinS* and *ybfE* promoters are not true LexA targets.

The *dinJ* promoter contains a significant match to a canonical DNA site for LexA, although previous work suggested that LexA does not bind the *dinJ* promoter in vitro and does not regulate *dinJ* expression in vivo (Fernandez de Henestrosa et al. 2000). The *dinJ* promoter was not identified as a LexA target in our ChIP–chip analysis, narrowly missing an initial cutoff. Direct analysis, however, indicates that LexA does, in fact, associate with the *dinJ* promoter in vivo, albeit at relatively low levels in comparison to other targets (Fig. 2). As expected, UV irradiation causes decreased LexA association and increased RNAP association with the *dinJ* promoter in a

Table 1. *Summary of in vivo LexA targets identified by ChIP–chip*

ChIP–chip score[a]	Predicted site[b]	Target class[c]	Target gene(s)[d]	Distance to canonical LexA site (bp)[e]	LexA-dependent UV induction[f]
2.75	1,020,250	I	*sulA*	–74	**17.3**
2.62	3,851,691	I	*ysdA*	–358	n.d.
2.48	2,079,201	I	*sbmC*	111	**5.1**
2.46	3,815,928	I	*dinD*	–200	**10.6**
2.38	2,749,644	I	*recN* (3)	113	**31.4**
2.26	1,928,879	I	*yebG*	–79	**7.5**
2.26	4,255,589	I	*lexA* (2)	–469	**4.6**
1.93	1,120,728	I	*dinI*	13	**3.1**
1.92	1,808,278	I	*ydjM* (2)	–69	**3.6**
1.85	1,225,340	II	*minC*	232	0.6
1.83	4,577,915	I	*yjiW*	32	**1.8**
1.81	1,230,172	I	*umuD*	–213	**25.7**
1.78	2,957,084	III	*ptrA*		0.9
1.75	2,821,814	I	*recA*	45	**10.0**
1.71	606,933	I	*hokE*	–62	1.0
1.50	4,272,039	I	*uvrA, ssb*	–50	**1.9**, 1.2
1.42	1,979,908	III	***otsB***, *otsA*		0.9, **1.4**
1.41	2,359,548	III	***yfaX***, *yfaW*		n.d, n.d
1.39	4,047,394	III	***polA***		0.8
1.38	4,015,135	I	*yigN*	169	**2.5**
1.37	812,662	I	*uvrB*		**3.0**
1.33	2,194,300	I	*molR*	177	**1.4**
1.30	1,852,700	III	*ydjF*		n.d.
1.29	3,646,206	I	*dinQ*	–192	n.d.
1.23	1,630,434	III	*ydfJ, ydjK*		0.6, 0.5
1.16	887,125	III	***ybjK***		**1.4**
1.12	4,356,774	III	***cadB***, *cadA*		n.d., **1.4**
1.08	691,049	III	*ybeX*		0.8
1.05	3,031,087	III	*idi*		1.2
1.03	676,333	III	***ybeR***, *ybeS*		1.0, 0.4
1.00	1,432,798	III	*ynaE*		0.5
1.00	1,903,626	III	*yebN*		0.7
0.98	1,892,090	II	*yoaC*	–78	**2.2**
0.94	3,957,919	II	***b3776***	324	**2.7**
0.87	250,770	I	*dinB*	102	**8.7**
0.85	514,280	III	***ybbK***, *ybbJ*		1.1, 0.4
0.84	564,100	III	***intD***		**1.6**, 0.6
0.82	931,917	I	*ftsK*	445	n.d.
0.82	1,821,079	I	*ydjQ*	435	**3.1**
0.82	243,381	II	*fadE*	–16	0.4
0.77	274,442	III	*trs5_1, mmuP*		n.d., 0.2
0.76	65,858	I	*polB*	–13	**2.2**
0.76	1,944,027	I	*ruvA*	31	**2.8**
0.75	1,585,011	II	***ydeQ***	354	1.1
0.69	1,696,318	III	***malI***, *hdhA*		0.4, n.d.
0.63	831,994	I	*dinG*	276	1.0
0.58	2,880,600	III	***ygcK***		n.d.
0.53	456,549	III	*clpX*		0.8
0.52	3,995,941	I	*uvrD*	–3	1.0

[a]Score indicating the strength of in vivo LexA binding (see Materials and Methods).
[b]Predicted genomic coordinate of LexA binding (see Materials and Methods).
[c]Class I represents previously known targets containing ≥1 canonical LexA motif(s) (canonical targets), Class II novel canonical targets, and Class III novel noncanonical targets (see text).
[d]For Class I and Class II targets, genes closest to the canonical LexA motif(s); almost all these canonical motifs reside in the promoter regions of the indicated genes (italics), but for two Class II targets (*b3776* and *ydeQ*), they actually reside within the ORF (bold). Numbers in parentheses indicate the number of canonical LexA motifs in the target region. For Class III targets, if the predicted site falls within an intergenic region, the gene with the closest 5′-end is designated the target gene (italics); in cases where it falls in coding sequence, we indicate the corresponding ORF in bold. For Class III predicted sites with two target genes listed, the first represents the promoter/ORF (italics/bold) at the predicted LexA site, while the second (italics) is the gene with the (next) closest promoter region. We indicate this second target because of the slight uncertainty determining the actual LexA-binding coordinate within Class III targets.
[e]Distance between predicted site and canonical LexA motif for Class I and Class II targets in base pairs. For targets containing multiple canonical motifs, we take the smallest distance.
[f]Ratio of $+UV_{20min}$/–UV expression fold-change in MG1655 to $+UV_{20min}$/–UV expression fold-change in MG1655 *lexA1* for the corresponding target gene(s) as determined previously (Courcelle et al. 2001). Ratios in bold indicate ≥1.4-fold LexA-dependent induction following UV irradiation; n.d. indicates not determined.

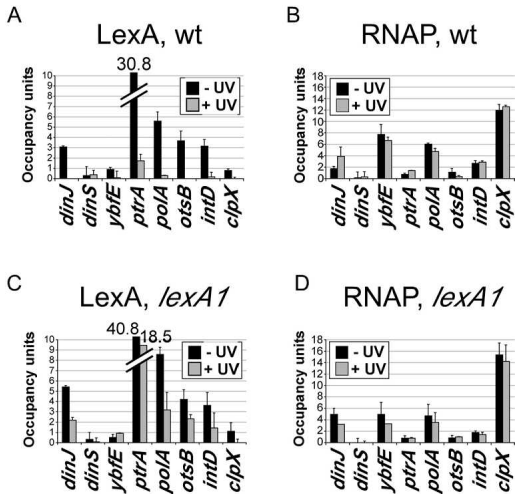

Figure 2. In vivo binding of LexA and RNAP to novel LexA targets. Association of LexA (*A*,*C*) and RNAP (*B*,*D*) with novel LexA targets identified by ChIP–chip analysis, before and after UV irradiation (black and gray bars, respectively), in wild-type MG1655 (*A*,*B*) and an isogenic *lexA1* strain (*C*,*D*). Occupancy was measured as described in Figure 1.

wild-type strain, but not in a *lexA1* strain. Thus, the *dinJ* promoter is actually a Class I LexA target. Taken together, our genome-wide analysis of LexA binding in vivo is in near perfect accord with expectations from previous biochemical and genetic studies and exhibits a low false-negative rate for Class I targets.

A canonical LexA sequence recognition motif is sufficient for LexA binding in vivo

We used MEME (Bailey and Elkan 1994) to identify common DNA sequence motifs among genomic regions representing 1000 bp centered at each of the 49 predicted LexA sites (Table 1). An appealing feature of this program is its ability to automatically compute optimal motif widths, in contrast to the majority of current motif-finding software. We did not impose a constraint that the motif be dyad-symmetric on the search. The resulting LexA position weight matrix (PWM) (Fig. 3) is based on the 30 Class I and Class II LexA targets; the remaining 19 Class III LexA targets lack a conventional LexA-binding site and will be discussed separately. The MEME-derived LexA PWM based on in vivo association is similar, but not identical, to the matrix previously described by biochemical and genetic analysis (Fernandez de Henestrosa et al. 2000). In addition to being present at all previously known LexA targets at their expected coordinates, this LexA motif is also detected by the MEME analysis at five Class II targets: in the promoter regions of *minC*, *fadE*, and *yoaC*, and the coding sequences of *ydeQ* and *b3776* (Table 1). We note that two of the genes, *yoaC* and *b3776*, are induced in a LexA-dependent manner upon UV exposure (Courcelle et al. 2001). Interestingly, our PWM fails to identify a site at the *ybfE* promoter, consistent with our ChIP analysis (Fig. 2).

We used ScanACE (Roth et al. 1998) to score the PWM derived from the ChIP–chip analysis against every 21-bp sequence in the *E. coli* genome, thereby providing a quantitative measure of how each site compares to the known consensus. As shown in Figure 4, for sites scoring >16.5, that is, canonical sites, there is a striking correlation between ScanACE and ChIP–chip scores. In fact, only four such sites, three in the *dinJ*, *dinS*, and *yciG* promoters and one in the *yehZ* open reading frame (ORF), are not identified by our ChIP–chip analysis. Importantly, the *dinS*, *yciG*, and *yehZ* motifs have ScanACE scores only just above the 16.5 threshold, with one being a confirmed true negative (*dinS*), while the *dinJ* site is a known false negative (Fig. 2). There are many sites immediately below the 16.5 threshold (10 sites >15) that are not bound in vivo.

These data strongly suggest that a suitable DNA site is sufficient for LexA association in vivo and that the strength of the in vivo LexA–DNA interaction can be well approximated by the degree of similarity to the PWM. To provide independent evidence for this conclusion, we constructed derivatives of MG1655 in which the 24-bp sequence encompassing the LexA site at the *sulA* promoter (ScanACE score of 20.6) was introduced into each of seven different coding sequences (*lacZ*, *gltA*, *araG*, *araE*, *malQ*, *melA*, and *fecD*) scattered around the genome. Of these seven regions, six are transcriptionally inactive under the conditions tested, the exception being the *gltA* gene (Wade and Struhl 2004). Strikingly, the association of LexA with each of these regions containing an ectopic DNA site differs by less than twofold relative to that of the DNA site at its natural locus (Fig. 5). In each case, LexA only binds in the presence of the ectopic DNA site. In addition, binding of LexA to the ectopic *sulA* site in the *lacZ* coding sequence is not significantly altered under conditions of high transcription (data not shown). Thus, LexA binds equivalently at each of these genomic locations, most likely independently of transcriptional activity.

Strong positive and possible negative evolutionary selection for genomic locations of canonical LexA sites

Based on the PWM derived from our genomic analysis, a canonical LexA sequence should occur by chance on average approximately nine times in the *E. coli* genome (the average number of sites with ScanACE score >16.5 in a randomized genome), with approximately two sites in intergenic regions and seven sites in protein-coding

Figure 3. Sequence logo of the LexA-binding motif determined using MEME on all 49 target regions. The corresponding position-weight matrix was used for all ScanACE analyses.

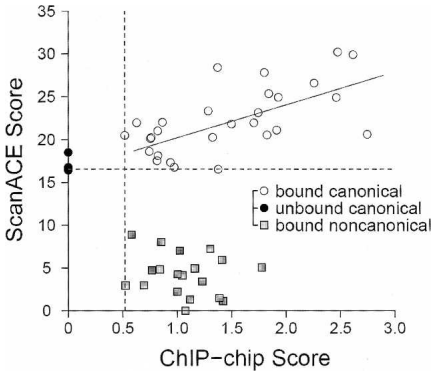

Figure 4. Relationship between LexA binding in vivo and similarity to the LexA motif. The ChIP–chip score is plotted against the ScanACE score for all 49 target regions as well as all other (i.e., ChIP–chip score = 0) genomic regions that have a ScanACE score ≥16.5. Targets containing multiple canonical LexA motifs of differing ScanACE scores were excluded. The horizontal dashed line indicates the lowest ScanACE score of a canonical LexA motif in a ChIP–chip target region (*yigN*), while the vertical one the lowest ChIP–chip score of a target region bearing a canonical LexA motif (*uvrD*). For the *upper right* sector defined by these cutoffs, a line of correlation is drawn.

regions. Thus, there is a strong evolutionary selection for the 28 canonical LexA sites found in intergenic regions, and, indeed, almost all of these 28 sites have been shown to mediate LexA-dependent repression of the adjacent gene. In this regard, the average distance between the LexA motif and the nearest translational start codon is only 58 bp, a distance very close to the promoter, and in accord with the view that LexA blocks transcription by steric interference with RNAP. Only three LexA motifs with ScanACE score >16.5 occur in coding regions, suggesting that LexA motifs in functionally inappropriate locations may be negatively selected during evolution. Indeed, across 100 coding sequence randomizations, the number of such high-scoring LexA sites found was always >6.

Unconventional in vivo targets of LexA that lack a canonical LexA motif

As mentioned above, 19 of the novel target regions did not contain a significant match to the canonical LexA motif as defined by the PWM. ScanACE scores of the closest matches to the LexA PWM for each of the 1000-bp regions surrounding the predicted site locations for the novel targets (Fig. 4) are very low (<9), and expected for scores from random 1-kb regions of the genome ($p = 0.41$). In addition, using MEME and a variety of other motif-finding programs, we were unable to identify a common DNA sequence motif among LexA targets that lack a canonical LexA motif.

Given that these putative LexA targets lacking a consensus LexA motif were unexpected, it was essential to confirm the results of the microarray analysis. Therefore, we directly analyzed LexA and RNAP association with five of these novel target regions, covering a range of ChIP–chip scores: the *ptrA*, *otsB*, and *clpX* promoters, and the coding regions of *polA* and *intD*. With the exception of the *clpX* promoter, LexA associates to varying degrees above the cutoff value with these targets, and LexA association is dramatically reduced following UV irradiation (Fig. 2A). The LexA1 mutant protein associates to varying degrees with each of the unconventional targets, both before and after UV irradiation, again with the exception of *clpX* (Fig. 2C). It is important to note that the precise location of these noncanonical LexA targets could be hundreds of base pairs away from the predicted site, and hence the location of the primers used in the quantitative PCR analysis; hence, the observed fold-enrichment may be an underestimate. Conversely, analysis of canonical sites was performed with primers flanking the LexA motif, therefore maximizing the observed fold-enrichment. In spite of this, we have confirmed seven out of eight LexA targets identified by our ChIP–chip analysis, demonstrating a very low false-negative rate.

In striking contrast to classical LexA targets (Fig. 1B), RNAP association with each of the unconventional targets tested is not substantially altered following UV irradiation in a wild-type (Fig. 2B) or *lexA1* (Fig. 2D) strain. In support of this observation, a previous study shows only minor changes (<1.5-fold) in the RNA levels of *ptrA*, *otsB*, *polA*, and *intD* following UV irradiation (Table 1; Courcelle et al. 2001). Thus, for four unconventional targets tested, and presumably most of the unconventional targets not tested, LexA binding in vivo does not appear to correlate with transcriptional repression of the target gene, at least under the conventional laboratory growth conditions used in these experiments. Interestingly, a different study did identify *ptrA* as a LexA-regulated gene (Quillardet et al. 2003), suggesting that LexA-de-

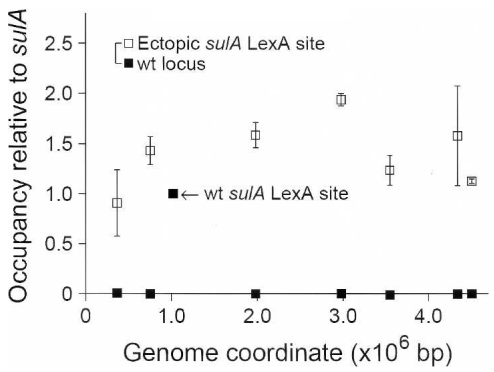

Figure 5. In vivo binding of LexA to the *sulA* LexA motif in its natural location (genome position 1,020,172) or in each of seven ectopic locations in the coding sequences of *lacZ* (365,099), *gltA* (753,123), *araG* (1,982,342), *araE* (2,979,212), *malQ* (3,547,000), *melA* (4,340,520), and *fecD* (4,509,915). White squares represent ectopic LexA sites, and black squares represent the association of LexA with the corresponding wild-type loci. Occupancy was measured as a ratio of binding of LexA to the tested region to a control region located around the natural binding site for LexA at the *sulA* promoter.

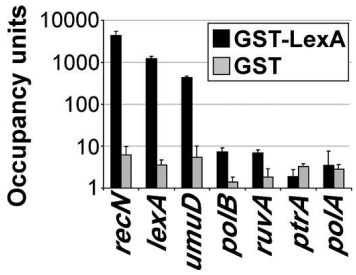

Figure 6. LexA binding in vitro. Association of GST (gray bars) and GST-LexA (black bars) with sheared genomic DNA. Occupancy was measured as a ratio of binding of GST or GST-LexA to the indicated region and to a control region located within the coding sequence of the predicted ORF *sgrR*.

pendent regulation of this gene might depend on the precise experimental conditions.

LexA binds in vitro to canonical, but not noncanonical, targets

To determine which in vivo targets are bound directly by LexA in vitro, we incubated a GST-LexA fusion protein or GST alone with sheared genomic DNA. The resulting protein:DNA complexes were purified using glutathione beads and then washed several times to remove nonspecifically bound DNA. We determined the association of GST-LexA and GST alone with the *umuD*, *recN*, *lexA*, *ruvA*, *polB*, and *ptrA* promoters and the *polA* ORF by quantitative PCR in real time (Fig. 6). In accord with their high levels of LexA association in vivo, conventional LexA targets (*umuD*, *recN*, and *lexA* promoters) associate strongly with GST-LexA but not with GST alone. The *polB* and *ruvA* promoters, conventional LexA targets that are bound relatively weakly in vivo (Fig. 1A; data not shown), show detectable, but low levels of LexA association in vitro (Fig. 6). Thus, for conventional LexA sites, the level of association in vivo is strongly related to the inherent affinity of LexA for these DNA sites in vitro. In striking contrast, GST-LexA binding to two unconventional targets (i.e., those lacking the LexA motif) is indistinguishable from that of the GST control. Thus, these unconventional targets are bound extremely poorly by LexA in vitro, presumably because they lack sites with significant correspondence to the LexA PWM.

Sequence determinants for LexA association at the ptrA *promoter*

The strongest unconventional in vivo LexA target is located near the *ptrA* promoter, and we mapped this site to a 40-bp region using ChIP and quantitative PCR with six primer pairs surrounding the predicted location (Fig. 7A; Table 1). This region does not contain the best match to the consensus LexA motif within 1000 bp of the predicted site location. It does, however, contain a sequence that is a good match to the consensus LexA motif except that the near-invariant CTG motifs in each half-site are replaced by ATG (Fig. 7). To determine whether this sequence is required for LexA association in vivo, we analyzed, in an otherwise MG1655 background, 1-, 20-, and 80-bp chromosomal deletions centered at this site. In each case, the deletion completely abolished LexA binding (Fig. 7B). We also made point mutations in either or both of the ATG triplets, converting them to AGT, and all three mutations completely abolished LexA binding (Fig. 7B). Finally, LexA association in vivo is completely abolished in derivatives with symmetric point mutations in which the nonconserved base of each of the ATG triplet was converted to TTG (Fig. 7B). Thus, LexA association at the *ptrA* promoter is mediated by an aberrant, but specific, LexA motif-like sequence in which key residues directly contacted by LexA are altered so as to preclude DNA binding in vitro.

In order to define a minimal binding site for LexA at the *ptrA* promoter, we created derivatives of MG1655 that contain an 144-, 16-, or 14-bp sequence centered around the putative LexA-binding site at the *ptrA* promoter, introduced into the coding sequence of *melA*. LexA binds at a similar level to the 144- and 16-bp sequences but does not bind the 14-bp sequence (Fig. 7C). Thus, we have defined a minimal 16-bp sequence that is sufficient for LexA binding at the *ptrA* promoter. This sequence contains each of the ATG motifs and the intervening 10 bp.

Discussion

Genome-wide identification of in vivo targets of bacterial DNA-binding transcriptional regulatory proteins have been limited to a handful of examples (Laub et al. 2002; Molle et al. 2003a,b; Eichenberger et al. 2004; Grainger et al. 2004). Moreover, these experiments typically involved microarrays containing PCR products representing only coding sequences, and hence do not represent an unbiased or comprehensive identification of in vivo targets of a DNA-binding protein. Here, we combine ChIP and high-density microarrays representing the entire *E. coli* genome to identify in vivo targets of the LexA repressor in a relatively unbiased manner. We identify 49 high-confidence LexA target regions in *E. coli* with low false-positive and false-negative rates, and map these sites to a resolution of 167 bp without any sequence information (Table 1). As is discussed separately below, LexA targets can be classified as either canonical (Class I and Class II) or noncanonical (Class III), as defined by the presence of a LexA motif and the ability to bind LexA in vitro.

The E. coli *genome is permissive for binding transcription factors*

In eukaryotic cells, association of DNA-binding proteins with target sites is strongly influenced by accessibility of DNA within nucleosome arrays. In the yeast *S. cerevisiae*, promoter regions are relatively depleted for nucleosomes in comparison to protein-coding regions (Ng et al.

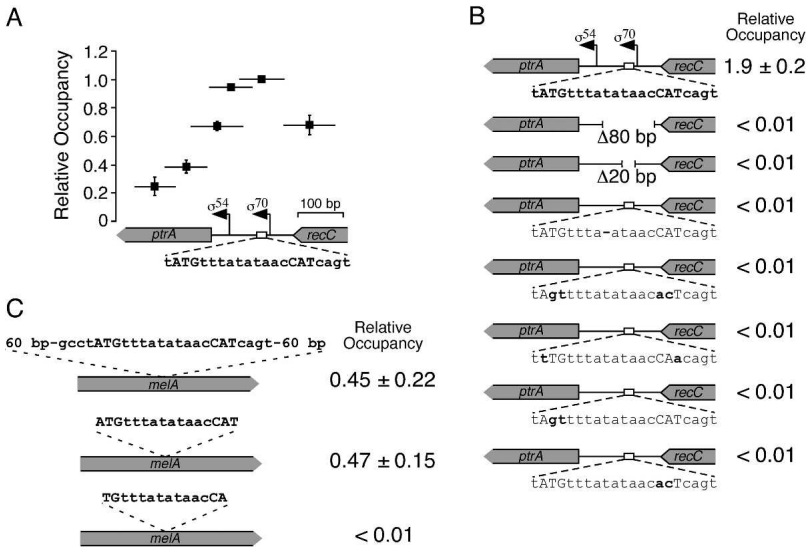

Figure 7. LexA binding to the *ptrA* promoter region. (*A*) Association of LexA with six regions covering the wild-type *ptrA* promoter was determined as in Figure 1, with values being normalized to the value for PCR product #5. The pseudocanonical LexA motif (CTG-type triplets capitalized) and transcriptional start sites for σ^{54} and σ^{70} promoters are indicated. (*B*) Association of LexA with the indicated mutated derivatives of the *ptrA* promoter region. The pseudocanonical LexA motif is shown along with 80- and 20-bp deletions (dotted gray lines) drawn to scale; 1-bp deletion and point mutations within the CTG-type triplets are shown in bold. Occupancy was measured as a ratio of binding of LexA to the tested region and to a control region located around the binding site for LexA at the *sulA* promoter. (*C*) Association of LexA with the indicated regions of the *ptrA* promoter region present at an ectopic location in the coding sequence of *melA*. Occupancy was measured as a ratio of binding of LexA to the *melA* coding sequence and to a control region located around the binding site for LexA at the *sulA* promoter.

2003; Bernstein et al. 2004; Lee et al. 2004; Sekinger et al. 2005), such that binding of transcription factors is largely restricted to appropriate sites in promoters (Lieb et al. 2001; Sekinger et al. 2005). In contrast, our genome-wide analysis of LexA binding in vivo demonstrates that, unlike eukaryotes, the *E. coli* genome is permissive for binding transcription factors. Specifically, virtually all canonical LexA sequence motifs that are bound by LexA in vitro associate with LexA in vivo, indicating that the presence of a suitable DNA sequence is sufficient for LexA association in vivo. Conversely, with a few notable exceptions (see below), LexA does not associate with any of the 4.6 million potential sites with ScanACE scores below those of canonical LexA motifs. The fact that DNA sequence is a remarkably accurate predictor of LexA binding in vivo, as indicated by the ChIP–chip score, is in stark contrast to the situation in yeast and mammalian cells, where numerous high-quality DNA sequence motifs are not bound by the relevant protein in vivo (Iyer et al. 2001; Lieb et al. 2001; Martone et al. 2003; Cawley et al. 2004; Euskirchen et al. 2004; Harbison et al. 2004).

Independent confirmation of the permissive nature of the *E. coli* genome comes from the observation that LexA binds comparably to the *sulA* site artificially introduced within seven different coding sequences representing different levels of transcriptional activity. Thus, although the vast majority of bona fide LexA target sequences are located in intergenic regions, LexA is fully capable of associating with all transcriptionally inactive or active protein-coding regions. Thus, our results demonstrate that the entire *E. coli* genome is equally accessible to binding by LexA, and hence it is very likely that it is equally accessible to other DNA-binding proteins. Thus, our results indicate that although bacterial genomes are packaged by histone-like proteins into a nucleoid structure (Dame 2005), this structure is not analogous to eukaryotic chromatin, and it does not impose a significant restriction on the association of transcriptional regulatory proteins with target DNA sequences.

Mechanistic and evolutionary implications of a permissive genome

The permissive nature of the *E. coli* genome has profound consequences for the nature of DNA-binding proteins, biological specificity, and evolution of transcriptional regulatory systems. As the entire genome is equally accessible, *E. coli* (and presumably other prokaryotic organisms) must evolve transcription factors with high DNA-binding specificity and/or tolerate binding to biologically irrelevant locations. LexA displays a high degree of DNA-binding specificity, because the PWM derived from in vivo LexA-binding sites should only occur nine times by chance in the *E. coli* genome. Furthermore, organisms with permissive genomes must either tolerate a certain level of binding to irrelevant sites or they must evolutionarily select against the presence of such sites. Our analysis of LexA provides some evidence that *E. coli* has evolutionarily selected against LexA target sites at irrelevant biological locations, but this conclusion is qualified by the small sample size involved. Lastly, canonical LexA sites occur in the genome far more frequently than expected by chance, and the vast majority of these sites are located in a very restricted region close to the transcriptional initiation site. In this regard, a computational study of 55 *E. coli* transcription factors shows that binding sites with high ScanACE scores are generally enriched in noncoding sequences (Robison et al. 1998). Thus, even though the genome is uniformly permis-

sive, LexA binding is restricted to biologically relevant targets with limited binding to functionally irrelevant regions.

The permissive nature and equal accessibility of the *E. coli* genome are in marked contrast to the restrictive nature of eukaryotic genomes, in which genomic regions are differentially accessible. A possible consequence of this distinction is that eukaryotic transcriptional regulatory proteins might intrinsically possess lower DNA-binding specificity than prokaryotic regulatory proteins. As the information content necessary to specify DNA binding has rarely been determined in a rigorous fashion, comparisons between prokaryotic and eukaryotic DNA-binding proteins are difficult. Nevertheless, it is our general impression that consensus matrices for prokaryotic DNA-binding proteins display higher information content than matrices for eukaryotic proteins. For example, the consensus matrix for high-affinity Gcn4 sites predicts ~1000–2000 sites throughout the yeast genome, most of which are in protein-coding regions (Oliphant et al. 1989; Mai et al. 2000). In any event, we propose that high DNA-binding specificity, perhaps with evolutionary selection against sites in biologically irrelevant regions, is a general feature of transcription factors in bacteria.

A new class of LexA targets

High-resolution microarrays that cover entire genomes are valuable for identifying in vivo targets of transcriptional regulatory proteins, because they permit an unbiased search that is not constrained by previous assumptions. In this vein, our analysis reveals a novel, and unexpected, class of 19 LexA targets that lack a LexA sequence motif. Almost all of the unconventional LexA targets identified by the microarray analysis are bona fide targets in vivo; five out of six such targets tested by direct ChIP analysis show clear LexA association that is reduced by UV treatment and largely unaffected by the *lexA1* mutation. On average, it appears that unconventional targets may be bound somewhat less well than conventional targets, but there is considerable overlap in LexA association levels between these two classes, and the *ptrA* site is bound at very high levels. In all cases tested, the unconventional LexA targets are not bound by LexA in vitro, indicating that LexA binding in vitro is strictly correlated with the quality of the LexA motif.

How does LexA bind these unconventional targets in vivo, given that the inherent LexA:DNA interaction is insufficient? The simplest model is that LexA binds to these sites cooperatively with another protein. Although there are very few known examples of *E. coli* proteins binding to noncanonical DNA sites, CRP binds to a noncanonical site in the *melAB* promoter in cooperation with MelR bound to adjacent DNA sites (Wade et al. 2001). Alternatively, these unconventional target sites might exist in a structural conformation in vivo (e.g., bent or supercoiled DNA) that is distinct from standard B-form DNA. At present, the molecular mechanism(s) for LexA binding to these sites is unknown, and it is certainly possible that the mechanisms (and proteins that cooperate with LexA) differ among the various unconventional sites. Interestingly, our analysis of the unconventional site at the *ptrA* promoter suggests that a particular suboptimal version of the LexA motif with symmetric mutations in each half-site, not merely a weakened binding site, is necessary and sufficient for LexA binding in vivo. These symmetric mutations might be important for binding another protein that cooperates with LexA and/or alters the precise nature of the LexA:DNA contacts.

In contrast to the conventional LexA targets, few of the novel LexA targets are associated with LexA-dependent or UV-inducible regulation of transcription (Table 1; Courcelle et al. 2001), suggesting the possibility that they are biologically irrelevant. However, several considerations suggest that these unconventional targets do not represent fortuitous binding by LexA. First, in accord with the major biological function of LexA at conventional sites, three out of the nine characterized genes adjacent to the unconventional targets have biological functions related to DNA metabolism: *polA* is a DNA polymerase, *intD* is an integrase, and *ptrA* is a protease of unknown function that is located between the *recB* and *recC* genes that encode two subunits of the RecBCD recombinase. Second, *E. coli* DNA-binding proteins do not inevitably associate with unconventional targets, because MelR binds predominantly, and possibly exclusively, with the *melAB* promoter in vivo (Grainger et al. 2004). Third, as *E. coli* has evolved to restrict canonical motifs for LexA and presumably for other DNA-binding proteins (Robison et al. 1998) almost exclusively to biologically relevant targets, it seems unlikely that this organism would be so promiscuous as to permit a similar number of unconventional target sites at meaningless genomic locations. Furthermore, binding to unconventional targets requires additional factors (e.g., cooperative interactions with other proteins and/or unusual DNA conformation), and such combinatorial requirements are typical of increased specificity, not promiscuity. Fourth, the LexA site at the *ptrA* promoter is located downstream of the known σ^{70}-dependent transcription start site and upstream of a putative σ^{54}-dependent transcription start site (Fig. 7A; Claverie-Martin et al. 1987). We speculate that LexA binding to these unconventional sites can affect transcription of adjacent genes, but only under specific conditions related to the additional factor(s) required for LexA binding in vivo.

Although previous analyses of prokaryotic DNA-binding proteins have ignored the possibility of noncanonical target sites in vivo, this phenomenon may be much more common than expected. For example, 38% of in vivo intergenic CtrA targets in *Caulobacter crescentus* do not contain a match to the derived consensus motif for CtrA (Laub et al. 2002). In addition, 15% of the regions bound by Spo0A in *Bacillus subtilis* are not bound by purified Spo0A in vitro, and 24% of the Spo0A target regions are not associated with a gene regulated by Spo0A under the

conditions tested (Molle et al. 2003a). Although direct validation experiments for in vivo binding were not performed in those cases, we think it likely that many of these nonconsensus target sites are truly bound in vivo and are not artifacts (i.e., false positives) of the microarray analysis. Finally, as exemplified by yeast TFIIIC, the promoter-recognition component of the RNA polymerase III (Pol III) transcription machinery, in vivo targets with atypical properties may nevertheless be biologically significant. Specifically, yeast cells contain nine unusual TFIIIC targets that are not bound by other components of the Pol III machinery and are transcriptionally inactive under standard conditions, yet these targets are highly conserved, both in sequence and function, among different yeast species (Moqtaderi and Struhl 2004). Our results provide additional evidence that the relationships between DNA sequence motifs, protein binding in vivo, and biological function are more complicated than previously thought.

Materials and methods

Strains

E. coli strains MG1655 and MG1655 *lexA1* (gift from J. Courcelle, Mississippi State University) were used for ChIP experiments in Figures 1 and 2. Cells were grown to mid-exponential phase (OD_{650} = 0.3–0.6) in LB. For UV irradiation, cells were exposed to UV light 20 min prior to harvesting. For the experiments in Figures 5 and 7, cells were grown to mid-exponential phase (OD_{650} = 0.3–0.6) in LB + 0.2% glucose. Ectopic and mutated LexA-binding sites were chromosomally introduced into an MG1655 background using a λRed-based recombineering scheme involving *thyA* as a marker for both positive and negative selection that will be described elsewhere (N.B. Reppas and G.M. Church, in prep.) and is similar to a recent method for in vivo BAC engineering (Wong et al. 2005).

ChIP

ChIP was based on previously described procedures (Wade and Struhl 2004). Cells were grown in appropriate media, and formaldehyde was added to a final concentration of 1%. After 20 min of incubation, glycine was added to a final concentration of 0.5 M, and cells were harvested by centrifugation and washed once with Tris-buffered saline (pH 7.5). Cells were resuspended in 500 µL of lysis buffer (10 mM Tris at pH 8.0, 20% sucrose, 50 mM NaCl, 10 mM EDTA, 4 mg/mL lysozyme) and incubated at 37°C for 30 min. Five-hundred microliters of immunoprecipitation (IP) buffer (50 mM HEPES-KOH at pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS) and PMSF (final concentration 1 mM) were added to the cell extract, and DNA was sheared by sonication to an average size of ~500 bp. Insoluble cellular material was removed by microcentrifugation for 10 min, and the supernatant was transferred to a fresh tube. Fifty microliters of this supernatant were kept for use as the "input" sample.

Proteins were immunoprecipitated by diluting a fraction of the cross-linked cell extract with IP buffer to a final volume of 800 µL. This was then incubated with 20 µL of Protein A-Sepharose beads (Amersham-Pharmacia) and either no antibody, RNAP β subunit mouse monoclonal (NeoClone), LexA rabbit polyclonal antibody (Upstate), or Gal4 DNA-binding domain rabbit polyclonal antibody (Upstate) for 90 min at room temperature with gentle mixing. Samples were then washed twice with IP buffer, once with IP buffer + 500 mM NaCl, once with wash buffer (10 mM Tris-HCl at pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet-P40, 0.5% sodium deoxycholate), and once with TE (pH 7.5). Immunoprecipitated complexes were eluted by incubation of beads with elution buffer (50 mM Tris-HCl at pH 7.5, 10 mM EDTA, 1% SDS) at 65°C for 10 min.

Immunoprecipitated samples and the corresponding "input" sample were decross-linked by incubation for 2 h at 42°C and for 6 h at 65°C in 0.5× elution buffer + 0.8 mg/mL Pronase. DNA was purified using a PCR purification kit (QIAGEN). All ChIPs were performed at least twice.

Quantitative PCR

Quantitative PCR was performed in real time using the Applied Biosystems 7000 and 7700 sequence detectors. All values were calculated by comparison of target regions with a region of the *sgrR* coding sequence as a background control. Occupancy units represent a background-subtracted value for the association of a particular protein with a target region (Aparicio et al. 2004).

Microarray analysis

Approximately 5 µg of amplified DNA (Moqtaderi and Struhl 2004) from each of the anti-Gal4, anti-MelR, and duplicate anti-LexA immunoprecipitations were terminally labeled with biotin-ddUTP and hybridized to a GeneChip *E. coli* Antisense Genome Array (Affymetrix); arrays were then washed, stained, and scanned according to the manufacturer's instructions. Arrays were background-subtracted using the MAS5 algorithm. Following quantile normalization of replicate LexA arrays, all four arrays were lowess normalized. The resulting perfect match (PM) values were used as the final probe intensities. Background subtraction, normalization, and mismatch (MM) probe removal were performed in the Bioconductor R package affy (http://www.bioconductor.org). All subsequent data manipulations were performed using perl scripts. Raw intensity data (.CEL files) can be obtained at http://www.fas.harvard.edu/~nreppas/LexA.

Because the Affymetrix array probes were designed based on an outdated build of the *E. coli* MG1655 genome, we positioned all 25-mer probes by BLASTing them against the most recent genome sequence (NCBI accession no. NC_000913.2). We removed probes that did not have an exact 25/25 nucleotide match, as well as those perfectly matching probes that had >19 nucleotides (nt) of matching sequence to two or more genomic loci (so as to minimize potential cross-hybridization effects), resulting in a list of 126,475 probes. For each probe we computed four $\log_2$ intensity ratios (LIRs): $LexA_1$/Gal4; $LexA_1$/no Ab; $LexA_2$/Gal4; $LexA_2$/no Ab (a positive LIR indicating a locus enriched in the LexA ChIP DNA; the larger the value, the greater the degree of enrichment).

We smoothed the four sets of LIRs versus genome coordinate by computing for each probe the average LIR over all probes within a window of 1250 bp. We identified candidate LexA-bound regions where ≥20 consecutive probes had all four LIRs ≥0.17; such an approach emphasizes the reproducibility of the microarray data. Each candidate region, or probe block, was summarized according to four parameters: the number of probes it contained, the start-to-end probe length, the average of the highest LIRs within each of its four data sets (the ChIP–chip score), and the average of the genome coordinates corresponding to these maxima (the predicted position of the LexA-binding site). Probe blocks within 5 kb were merged; the doublet probe

block covering the canonical LexA target sites at the *minC* and *umuD* promoters had to be manually split. The values for the width of the smoothing window, the number of consecutive probes, and the threshold LIR were chosen so as to minimize the average absolute distance between the predicted LexA target coordinates and the midpoint coordinates of known canonical LexA motifs. We determined the statistical significance of each candidate region by randomizing the four LIR data sets with respect to the probe coordinate and repeating the above analysis. A probe block's *p*-value was calculated as the number of times a probe block was identified with greater probe number, length, and score over 1000 data randomizations.

Identifying conserved motifs

One-thousand base pairs of MG1655 genomic sequence centered around the ranked 49 peak coordinates were used as input to search for motifs using MEME (Bailey and Elkan 1994). The optimal motif width is automatically computed. We did not impose a requirement that the motif be dyad-symmetric. ScanACE (Roth et al. 1998) was used to identify and score matches to the resulting motif determined using MEME. WebLogo was used to generate the LexA motif logos (Crooks et al. 2004). ScanACE analysis of randomized genomic sequence was done separately for coding and noncoding regions to account for their differing percentage of GC content.

In vitro binding assay

LexA was cloned into the XbaI and NcoI restriction sites of plasmid pGEX-KG, and GST or GST-LexA was then purified as previously described (Guan and Dixon 1991). *E. coli* genomic DNA was purified using a 20/G genomic-tip kit (QIAGEN). Genomic DNA was sheared by sonication to an average size of ~300 bp. Two micrograms of sheared genomic DNA were incubated with 10 µL of glutathione-Sepharose 4B beads (Amersham Pharmacia Biotech) and 10 mg of purified GST or GST-LexA, in 500 µL of binding buffer (10 mM Tris at pH 8, 1 mM EDTA, 50 mM NaCl, 1 mM DTT, 0.1% Tween 20) for 30 min at room temperature with rotation. Beads were then washed five times with binding buffer. Proteins were eluted by incubating with binding buffer + 5 mM glutathione for 10 min. DNA was purified using a PCR purification kit (QIAGEN). All in vitro binding assays were performed three times.

Acknowledgments

We thank Jay Shendure, Rhonda Harrison, Dan Janse, Edward Sekinger, Joseph Geisberg, Zarmik Moqtaderi, Heather Hirsch, Marc Schwabish, Jason Lieb, and Xiao Liu for helpful discussions. We thank Justin Courcelle for the MG1655 *lexA1*. This work was supported by a long-term EMBO fellowship to J.T.W. and a research grant to K.S. from the National Institutes of Health (GM30186).

References

Aparicio, O.M., Geisberg, J.V., and Struhl, K. 2004. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. In *Current protocols in molecular biology* (eds. F.A. Ausubel, et al.), pp. 21.3.1–21.3.17. John Wiley & Sons, New York.

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.

Barnes, G. and Rine, J. 1985. Regulated expression of endonuclease EcoRI in *Saccharomyces cerevisiae*: Nuclear entry and biological consequences. *Proc. Natl. Acad. Sci.* **82:** 1354–1358.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. 2004. Global nucleosome occupancy in yeast. *Genome Biol.* **5:** R62.

Boccard, F., Esnault, E., and Valens, M. 2005. Spatial arrangement and macrodomain organization of bacterial chromosomes. *Mol. Microbiol.* **57:** 9–16.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. 2003. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11:** 1587–1598.

———. 2004. Removal of promoter nucleosomes by disassembly rather than sliding in vivo. *Mol. Cell* **14:** 667–673.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Smentchenko, V., Cheng, J., Williams, A.J., et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of non-coding RNAs. *Cell* **116:** 499–509.

Claverie-Martin, F., Diaz-Torres, M.R., and Kushner, S.R. 1987. Analysis of the regulatory region of the protease III (*ptr*) gene of *Escherichia coli* K-12. *Gene* **54:** 185–195.

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

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. 2004. WebLogo: A sequence logo generator. *Genome Res.* **14:** 1188–1190.

Dame, R.T. 2005. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56:** 858–870.

Deckert, J. and Struhl, K. 2001. Histone acetylation at promoters is differentially affected by activators and repressors. *Mol. Cell. Biol.* **21:** 2726–2735.

Dorman, C.J. and Deighman, P. 2003. Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Genet. Dev.* **13:** 179–184.

Eichenberger, P., Fujita, M., Jensen, S.T., Conlon, E.M., Rudner, D.Z., Wang, S.T., Ferguson, C., Haga, K., Sato, T., Liu, J.S., et al. 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol.* **2:** e328.

Euskirchen, G., Royce, T.E., Bertone, P., Martone, R., Rinn, J.L., Nelson, F.K., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., et al. 2004. CREB binds to multiple loci on human chromosome 22. *Mol. Cell. Biol.* **24:** 3804–3814.

Felsenfeld, G. 1996. Chromatin unfolds. *Cell* **86:** 13–19.

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

Grainger, D.C., Overton, T.W., Reppas, N., Wade, J.T., Tamai, E., Hobman, J.L., Constantinidou, C., Struhl, K., Church, G.M., and Busby, S.J.W. 2004. Genomic studies with *Escherichia coli* MelR protein: Applications of chromatin immunoprecipitation and microarrays. *J. Bacteriol.* **186:** 6938–6943.

Guan, K.L. and Dixon, J.E. 1991. Eukaryotic proteins expressed in *Escherichia coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* **192:** 262–267.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds,

D.B., Yoo, J., et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104.

Iyer, V. and Struhl, K. 1995. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic structure. *EMBO J.* **14:** 2570–2579.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409:** 533–538.

Laub, M.T., Chen, S.L., Shapiro, L., and McAdams, H.H. 2002. Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl. Acad. Sci.* **99:** 4632–4637.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36:** 900–905.

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

Lieb, J.D., Liu, X.L., Botstein, D., and Brown, P.O. 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nat. Genet.* **28:** 327–334.

Little, J.W. and Mount, D.W. 1982. The SOS regulatory system of *Escherichia coli*. *Cell* **29:** 11–22.

Mai, X., Chou, S., and Struhl, K. 2000. Preferential accessibility of the yeast *his3* promoter is determined by a general property of the DNA sequence, not by specific elements. *Mol. Cell. Biol.* **20:** 6668–6676.

Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. 2003. Distribution of NF-κB-binding sites across human chromosome 22. *Proc. Natl. Acad. Sci.* **100:** 12247–12452.

Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., Gonzalez-Pastor, J.E., Liu, J.S., and Losick, R. 2003a. The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* **50:** 1683–1701.

Molle, V., Nakaura, Y., Shivers, R.P., Yamaguchi, H., Losick, R., Fujita, Y., and Sonenshein, A.L. 2003b. Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J. Bacteriol.* **185:** 1911–1922.

Moqtaderi, Z. and Struhl, K. 2004. Genome-wide occupancy of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol. Cell. Biol.* **24:** 4118–4127.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11:** 709–719.

Oliphant, A.R., Brandl, C.J., and Struhl, K. 1989. Defining sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of the yeast GCN4 protein. *Mol. Cell. Biol.* **9:** 2944–2949.

Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R. 2004. Topological domain structure of the *Escherichia coli* chromosome. *Genes & Dev.* **18:** 1766–1779.

Quillardet, P., Rouffaud, M.A., and Bouige, P. 2003. DNA array analysis of gene expression in response to UV irradiation in *Escherichia coli*. *Res. Microbiol.* **154:** 559–572.

Reinke, H. and Horz, W. 2003. Histones are first hyperacetylated and then lose contact with the activated *PHO5* promoter. *Mol. Cell* **11:** 1599–1607.

Robison, K., McGuire, A.M., and Church, G.M. 1998. A comprehensive library of DNA-binding matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.* **284:** 241–254.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16:** 939–945.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. 2005. Intrinsic histone–DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18:** 735–748.

Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J., and Church, G.M. 2000. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.* **18:** 1262–1268.

Struhl, K. 1999. Fundamentally different logic of gene expression in eukaryotes and prokaryotes. *Cell* **98:** 1–4.

Wade, J.T. and Struhl, K. 2004. Growth-regulated association of σ^{70} and NusA with elongating RNA polymerase in *E. coli*. *Proc. Natl. Acad. Sci.* **101:** 17777–17782.

Wade, J.T., Belyaeva, T.A., Hyde, E.I., and Busby, S.J. 2001. A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter. *EMBO J.* **20:** 7160–7167.

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

———. 1985. Inducible DNA repair systems. *Annu. Rev. Biochem.* **54:** 425–457.

Wong, Q.N., Ng, V.C., Lin, M.C., Kung, H.F., Chan, D., and Huang, J.D. 2005. Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli*. *Nucleic Acids Res.* **33:** e59.

Workman, J.L. and Kingston, R.E. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67:** 545–579.