

A Green Nonsulfur Bacterium, *Dehalococcoides ethenogenes*, with the LexA Binding Sequence Found in Gram-Positive Organisms

Antonio R. Fernández de Henestrosa,^{1*} Jordi Cuñé,¹ Ivan Erill,² Jon K. Magnuson,³
and Jordi Barbé^{1,4*}

*Departament Genètica i Microbiologia*¹ and *Centre de Recerca en Sanitat Animal*,⁴ *Universitat Autònoma de Barcelona*, and *Biomedical Applications Group, Centro Nacional de Microelectrónica*,² *Bellaterra, 08193 Barcelona, Spain*, and *Pacific Northwest National Laboratory, Richland, Washington 99352*³

Received 6 May 2002/Accepted 5 August 2002

***Dehalococcoides ethenogenes* is a member of the physiologically diverse division of green nonsulfur bacteria. Using a TBLASTN search, the *D. ethenogenes* *lexA* gene has been identified, cloned, and expressed and its protein has been purified. Mobility shift assays revealed that the *D. ethenogenes* LexA protein specifically binds to both its own promoter and that of the *uvrA* gene, but not to the *recA* promoter. Our results demonstrate that the *D. ethenogenes* LexA binding site is GAACN₄GTTC, which is identical to that found in gram-positive bacteria. In agreement with this fact, the *Bacillus subtilis* DinR protein binds specifically to the *D. ethenogenes* LexA operator. This constitutes the first non-gram-positive bacterium exhibiting a LexA binding site identical to that of *B. subtilis*.**

In *Escherichia coli*, the SOS regulon consists of a pool of at least 40 DNA damage-inducible genes whose functions are involved in DNA replication, DNA repair, mutagenesis, and control of the cell cycle (3, 5, 8, 11). This regulon is controlled by two key proteins, RecA and LexA, the positive and negative regulators, respectively (1, 12, 15, 22). Under normal conditions LexA binds specifically to an operator region located within the promoters of the SOS genes, repressing their expression (8, 22). When DNA is damaged or replication is stalled, RecA protein forms a nucleofilament with the resulting single-stranded DNA, achieving its activated conformation (RecA*). RecA* mediates the autocatalytic cleavage of LexA between residues Ala⁸⁴ and Gly⁸⁵, resulting in two parts, the N-terminal DNA binding domain and the C-terminal domain, which is involved in dimerization of the repressor (13, 14). In this process, the catalytic site formed by Ser¹¹⁹ and Lys¹⁵⁶ is responsible for the hydrolysis of the Ala-Gly bond through a biochemical reaction characteristic of serine proteases (16). As a consequence of the cleavage, LexA no longer binds DNA, the SOS genes are derepressed, and their products participate in repairing the DNA lesions to guarantee cell survival. This process is known as the SOS response and was first suggested in the 1970s by Miroslav Radman (19). Once DNA is repaired, RecA* loses its active conformation, LexA autohydrolysis ceases, and the repressor accumulates again, inhibiting the expression of the SOS genes.

The *lexA* gene seems to be widespread in the *Bacteria* domain. However, *lexA* is not found in the fully sequenced genomes of several bacterial species, such as *Aquifex aeolicus*,

Borrelia burgdorferi, *Chlamydia pneumoniae*, *Mycoplasma pneumoniae*, *Campylobacter jejuni*, *Helicobacter pylori*, and *Porphyromonas gingivalis*. Moreover, it appears that a *lexA*-like gene does not exist in the *Archaea* domain, as *lexA* homologues are absent in the 19 archaeal genomes currently available (4).

Recently, the three-dimensional structure of the *E. coli* LexA protein has been solved (16). In *E. coli*, LexA binds specifically to a DNA motif known as the SOS-box or LexA-box (22). Comparative analysis of the 31 *E. coli* LexA binding sites shows a 16-bp consensus sequence that responds to the motif CTGN₁₀CAG (5). Besides the *E. coli* SOS-box, two more LexA-boxes have been identified so far in bacteria. The directed repeat GTTCN₇GTTC is recognized by LexA protein of members of the alpha-class *Proteobacteria*, such as *Rhodobacter sphaeroides* and *Rhizobium etli* (6, 21). In gram-positive bacteria such as *Bacillus subtilis* and *Mycobacterium tuberculosis*, the consensus sequence CGAACRNRYGTTYG is the target for this repressor (2, 24).

Dehalococcoides ethenogenes is an anaerobic bacterium capable of dechlorinating tetrachloroethene, one of the most common groundwater contaminants, to ethene (17). *D. ethenogenes* is rather enigmatic taxonomically. Its cell wall composition is unlike that of gram-positive or gram-negative bacteria, instead more closely resembling the cell wall of archaea (17). However, on the basis of 16S rRNA gene sequence comparisons, *D. ethenogenes* has been assigned to the green nonsulfur bacteria, a division of physiologically diverse species with relatively few cultured representatives (10). The unusual phylogenetic position, structural properties, and metabolic capabilities of *D. ethenogenes* make it an interesting organism for further biological characterization. Here, we report the cloning of the key regulatory gene *lexA* from *D. ethenogenes*, purification of the protein, and characterization of its recognition site. Prior to this report, there were no data available pertaining to the motif recognized by LexA in this bacterial division. Finally,

* Corresponding author. Mailing address: Universitat Autònoma de Barcelona, Departament de Genètica i Microbiologia, Unitat de Microbiologia, Edifici C, Bellaterra, 08193 Barcelona, Spain. Phone: 34 93 581 1837. Fax: 34 93 581 23 87. E-mail for Antonio R. Fernández de Henestrosa: antonio.rodriguez@uab.es. E-mail for Jordi Barbé: jordi.barbe@uab.es.

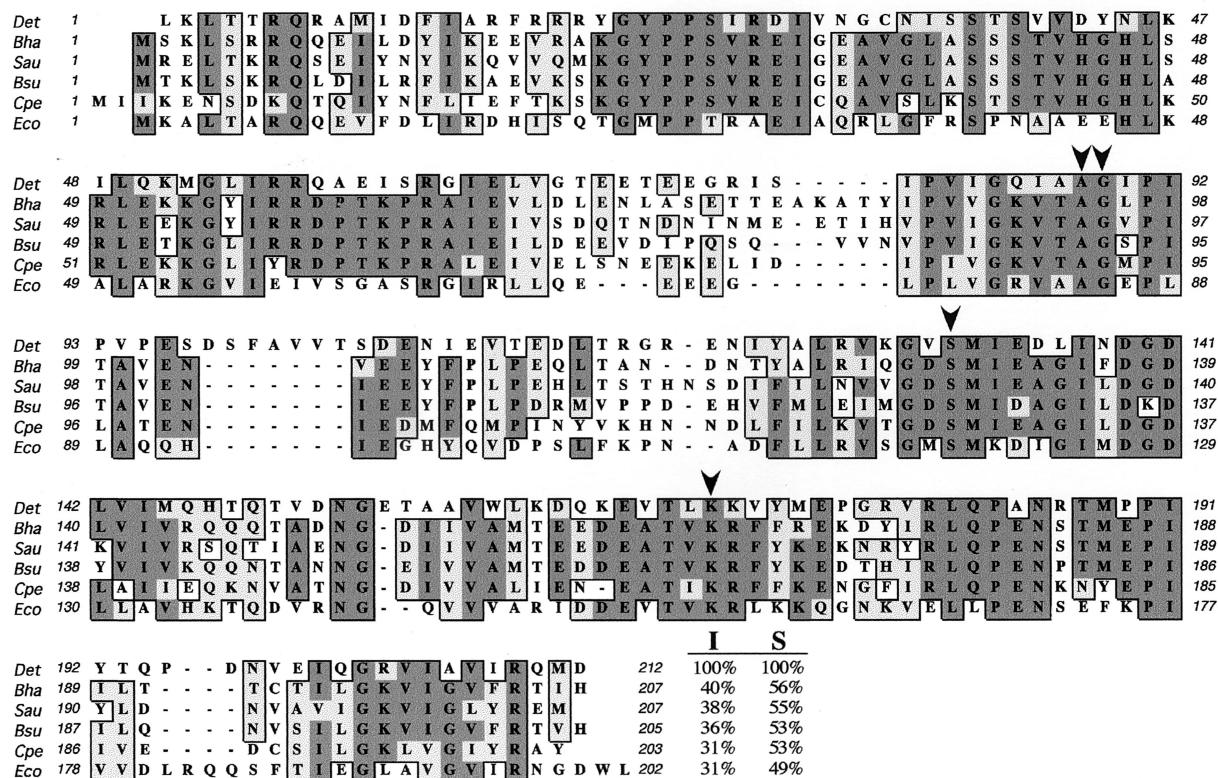


FIG. 1. CLUSTAL W alignment performed with Mac Vector 6.5 (Oxford Molecular) with LexA proteins from *Dehalococcoides ethenogenes* (Det), *Bacillus halodurans* (Bha), *Staphylococcus aureus* (Sau), *Bacillus subtilis* (Bsu), *Clostridium perfringens* (Cpe), and *Escherichia coli* (Eco). Identities have a dark gray background, and similarities are light gray. The perfectly conserved residues Ala⁸⁸ and Gly⁸⁹ forming the LexA cleavage site and the amino acids Ser¹³¹ and Lys¹⁷⁰ required for protein hydrolysis are indicated with arrowheads on the *D. ethenogenes* LexA sequence. At the end of each LexA sequence appears the percent identity (I) and percent similarity (S) that each protein shows with the *D. ethenogenes* LexA repressor. Accession numbers in the Entrez protein database at NCBI are as follows: *B. halodurans*, Q9KAD3; *S. aureus*, BAB57501; *B. subtilis*, P31080; *C. perfringens*, BAB80867; and *E. coli*, P03033.

we report a genomic analysis of the composition of the *D. ethenogenes* LexA network.

Identification and cloning of *D. ethenogenes* LexA gene. The *D. ethenogenes* strain 195 genome sequence is currently unassembled, but contigs are available for performing BLAST analysis at the Institute for Genomic Research and at the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html).

The *Escherichia coli* LexA amino acid sequence was used to query the *D. ethenogenes* genome with the TBLASTN program, revealing a protein homologue of 212 amino acids. Although *D. ethenogenes* is not a gram-positive organism, its LexA protein is highly homologous to the LexA proteins belonging to several members of this bacterial phylum, such as *Bacillus halodurans*, *Staphylococcus aureus*, and *Clostridium perfringens*. Specifically, *D. ethenogenes* LexA protein shows the highest level of identity with *Bacillus halodurans* LexA (40%), while the *D. ethenogenes* repressor is only 31% identical to *E. coli* LexA (Fig. 1). Furthermore, *D. ethenogenes* LexA contains all of the conserved residues involved in repressor autocleavage (Ala⁸⁸, Gly⁸⁹, Ser³³¹, and Lys¹⁷⁰) as deduced from CLUSTAL W alignment of different LexA proteins (Fig. 1). It is important to note that nine codons downstream of the proposed initiation site for the 212-amino-acid open reading

frame there is a 203-amino-acid protein starting with methionine. However, this truncated version of *D. ethenogenes* LexA would lack several N-terminal residues likely to be essential to the LexA structure, since it would lack important residues belonging to the first α -helix characterized by nuclear magnetic resonance spectroscopy in *E. coli* LexA (7).

The BLAST analysis returned 1 kb of DNA flanking the *lexA* coding sequence, which was used to design the primers AR31 and AR34 (Table 1). These primers were employed to PCR amplify the entire *D. ethenogenes* *lexA* and its promoter region with genomic DNA obtained from a 10-liter tetrachloroethene-methanol-fed anaerobic enrichment culture. The resulting 866-bp DNA fragment was then cloned into pGEM-T (Promega). To confirm that no mutation was introduced during the PCR, the sequence of the insert was determined by labeling DNA samples with the fmol DNA cycle sequencing system (Promega) and analyzed on an ALF sequencer (Amersham Pharmacia Biotech). This plasmid, called pUA969, was used as a template for further studies.

Purification of *D. ethenogenes* LexA protein. A glutathione S-transferase (GST)-LexA fusion protein was generated. The primers AR61 and AR34 (Table 1) were used to amplify the *D. ethenogenes* *lexA* gene from pUA969. The oligonucleotide AR61 incorporates an *Eco*RI restriction site upstream of the

TABLE 1. Oligonucleotide primers used in this work^a

Primer	Sequence	Position	Application
AR31	5'-CTCAACCAAGCCGCTAAATAACTCC-3'	-121	Upper primer for cloning <i>D. ethenogenes</i> <i>lexA</i> gene and LexA1 probe
AR33	5'-AGAGGTAGAGCTGATATTACAGCCG-3'	+117	Lower primer to amplify unlabeled specific DNA competitors
AR33DIG	5'-DIG-AGAGGTAGAGCTGATATTACAGCCG-3'	+117	Lower primer to amplify digoxigenin-labeled probe fragments of <i>lexA</i> promoter
AR34	5'- <i>GTCGACCTAATCCATCTGGCGATTACAGC</i> -3'	+636	Lower primer for cloning <i>D. ethenogenes</i> <i>lexA</i> gene
AR35	5'-TACCTGCCTGAATGACTTGACACC-3'	-295	Upper primer to amplify <i>D. ethenogenes</i> <i>recA</i> promoter
AR36	5'-CTTAGTTTCATTATTGAACCTTGC-3'	+85	Lower primer to amplify <i>D. ethenogenes</i> <i>recA</i> promoter
AR38	5'-TATGCCGGGACAGAGTTGAC-3'	-15	Upper primer to amplify LexA4 probe
AR39	5'-TTATGTAGAACAAGTGTCTTATGC-3'	-35	Upper primer to amplify LexA3 probe
AR40	5'-TTGTGATTGACAGGGAAGG-3'	-68	Upper primer to amplify LexA2 probe
AR61	5'- <i>GAATTCATGAACTGACCACCAGACAGCG</i> -3'	+1	Upper primer for directed cloning of <i>D. ethenogenes</i> <i>lexA</i> in pGEX4T1
AR62	5'-GCCGAGGCATGACCCAGCTGGGGC-3'	-515	Upper primer to amplify <i>uvrA</i> promoter region
AR63	5'-TTTTGAGGTTGTGTTCCGCGGCACC-3'	+51	Lower primer to amplify <i>uvrA</i> promoter region
AR64	5'-TAGTGGAGAAATACCACTCTCAGGC-3'	-531	Upper primer to amplify <i>nuvA</i> promoter region
AR65	5'-TTGCCGGGACAGAGTTGAC-3'	+40	Lower primer to amplify <i>nuvA</i> promoter region
AR66	5'-GGGTGAACTGGTAGACAAGTTCCGC-3'	-525	Upper primer to amplify <i>nuvB</i> promoter region
AR67	5'-GGCCAGTTTCCCGGATATAAGACGC-3'	+42	Lower primer to amplify <i>nuvB</i> promoter region
AR68	5'-AGCCGGTACCGCCGGATGCACCTGC-3'	-532	Upper primer to amplify <i>recN</i> promoter region
AR69	5'-CCAGGTGATATCCTCAATAATGCCG-3'	+53	Lower primer to amplify <i>recN</i> promoter region
AR81	5'-ACTCATGATCATAACCTCCAACAGG-3'	-161	Upper primer to amplify <i>B. subtilis</i> <i>dinR</i> promoter region
AR82	5'-GGACGCAAGCCGACAGCTCTCCG-3'	+116	Lower primer to amplify <i>B. subtilis</i> <i>dinR</i> promoter region

^a Added restriction sites are shown in italics. The position of either the 5' or 3' end of the oligonucleotide with respect to the proposed translational starting point of each *D. ethenogenes* gene is shown except for AR81 and AR82, where the position of either the 5' or 3' end of the oligonucleotide with respect to the translational starting point of the *B. subtilis* *dinR* gene is shown. Primers AR41 to AR60 and AR76 are derivatives of AR39 but with the desired nucleotide changes.

lexA translational start codon, permitting fusion of the protein in frame with the GST protein. The 650-bp PCR fragment was cloned into pGEM-T, generating plasmid pUA970. Afterwards, pUA970 was digested with *EcoRI* and *SalI*, and the resulting 0.65-kb fragment was cloned into pGEX4T1 (Amersham Pharmacia Biotech), creating the GST-LexA fusion.

This plasmid, designated pUA971, was transformed into *Escherichia coli* BL21 Codon Plus cells (Stratagene) to overproduce the protein. An overnight culture of the BL21 Codon Plus strain containing plasmid pUA971 was diluted in one-half liter of Luria-Bertani (LB) medium and incubated at 37°C until it reached an optical density at 600 nm of 0.8. At this time isopropylthiogalactopyranoside (IPTG) was added (1 mM final concentration), and the culture was incubated for 4 more hours. Afterwards, cells were collected by centrifugation for 15 min at 5,000 rpm.

The bacterial pellet was resuspended in phosphate-buffered saline (PBS) (10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl [pH 7.4]) containing complete Mini protease inhibitor cocktail (Roche) and sonicated to break the cells. The cell lysate was centrifuged at 15,000 rpm for 30 min, and the supernatant containing the soluble GST-LexA protein fusion was incubated for 2 h at 4°C with glutathione-Sepharose 4B beads (Amersham Pharmacia Biotech), previously equilibrated in PBS. The beads were washed twice with PBS-0.1% Triton X-100 and three more times with PBS to eliminate contaminant proteins. The washed glutathione-Sepharose beads containing bound GST-LexA protein were equilibrated in 0.1 M Tris-HCl (pH 8)-0.1 M NaCl buffer, and GST-LexA was eluted in the same buffer containing 20 mM glutathione (Sigma), resulting in about 85% pure GST-LexA (Fig. 2).

The GST-LexA fusion protein has the amino acid sequence

Leu-Val-Pro-Arg-Gly-Ser between GST and LexA. This motif is recognized by the thrombin protease, which cleaves the Arg-Gly peptide bond. In this case, proteolysis results in *D. ethenogenes* LexA protein with the peptide Gly-Ser-Pro-Glu-Phe attached to the N terminus. GST-LexA bound to glutathione-Sepharose was incubated at room temperature for 16 h with 50 U of thrombin protease (Amersham Pharmacia Biotech) in 1 ml of PBS. The liquid phase containing LexA was collected, and a portion was separated by electrophoresis on a denaturing 13% polyacrylamide gel. The purity of the *D. ethenogenes* LexA protein was greater than 95% (Fig. 2).

Determination of *D. ethenogenes* LexA binding motif. Analysis of the DNA contig containing *D. ethenogenes* *lexA* revealed that it was located between the *rpsT* and *fucA* genes (Fig. 3A). From the TAA translational stop codon of *rpsT* to the putative TTG start codon of *lexA*, there are 104 bp that presumably contain the LexA binding site. Hence, a PCR fragment containing this sequence was amplified with primers AR31 and AR33DIG (Table 1). The resulting digoxigenin-labeled probe (LexA1) was used as a target to observe whether purified *D. ethenogenes* LexA was able to bind its own promoter in electrophoretic mobility shift assay experiments carried out as previously described (6).

Basically, 20- μ l reaction mixtures containing 10 ng of digoxigenin-labeled DNA probe and 25 ng (50 nM) of *D. ethenogenes* LexA were incubated in binding buffer: 10 mM HEPES NaOH (pH 8), 10 mM Tris-HCl (pH 8), 5% glycerol, 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 1 μ g of bulk carrier DNA, and 50 μ g of bovine serum albumin per ml. After 30 min of incubation at 30°C, the mixture was loaded onto a 5% non-denaturing Tris-glycine-polyacrylamide gel. Thus, *D. ethenogenes* LexA was able to shift the mobility of the LexA1 fragment (Fig.

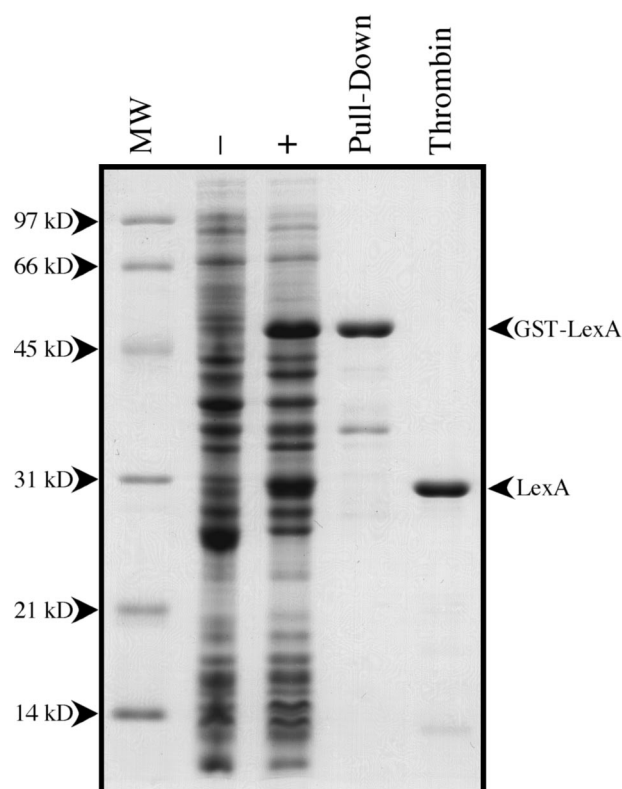


FIG. 2. Purification of the *D. ethenogenes* LexA protein. *D. ethenogenes* LexA was purified to greater than 95%. Each line shown in the denaturing 13% polyacrylamide gel represents each of the different protein purification steps employed: crude extract of BL21 Codon Plus/pUA971 (-); crude extract of BL21 Codon Plus/pUA971 induced with 1 mM IPTG (+); *D. ethenogenes* GST-LexA affinity purified with glutathione-Sepharose 4B (pull-down); purified *D. ethenogenes* LexA protein after thrombin protease digestion of GST-LexA fusion (thrombin). Lane MW, molecular size markers.

3B). Furthermore, the retarded band was abolished when unlabeled LexA1 was used as a competitor (100-fold molar excess over the digoxigenin-labeled probe). However, the shifted band could not be eliminated when nonspecific DNA competitors were used (Fig. 3B), suggesting that the binding of the *D. ethenogenes* LexA protein to its promoter is specific. It is important to note that the GST-LexA fusion protein retained its DNA binding activity in electrophoretic mobility shift assay experiments (Fig. 3B).

To establish the bounds of the LexA binding site, serial deletion fragments of the upstream promoter region were analyzed in electrophoretic mobility shift assay experiments with purified LexA. The DNA probes LexA1, LexA2 and LexA3 exhibited retarded bands in the presence of LexA protein, but the mobility of the LexA4 DNA was not shifted (Fig. 3C). These results indicate that the LexA binding motif is located within the sequence 35 bp upstream of the *lexA* translation start codon. An examination of this region revealed the presence of the inverted repeat GAACAAGTGTTTC 28 bp upstream of the putative TTG translational start codon (Fig. 3A). This sequence matches the DinR recognition motif (formerly Cheo-box), CGAACN₄GTTTCG (23). Changing the GAAC submotif to TTTT in the LexA3 fragment abolished LexA

protein binding, demonstrating that this tetranucleotide is part of the *D. ethenogenes* LexA binding site (Fig. 3C).

The importance of the rotational orientation of the inverted repeats relative to one another was examined by introducing additional nucleotides into the center of the GAACN₄GTTTC motif, lengthening the linker region between the inverted repeats (Fig. 3A). With modified LexA3 PCR primers (AR58 to AR60, Table 1), one to three adenines were inserted at position -21 relative to the translational start codon (Fig. 3A). The introduction of a single adenine resulted in a sequence that retained some residual binding of the LexA protein (Fig. 3D). However, the introduction of two or three nucleotides into the linker region completely abolished LexA binding (Fig. 3D), consistent with studies of LexA binding sites in other microorganisms (6).

To gain insight into the importance of each of the 12 bases of this imperfect palindrome and the neighboring nucleotides in the LexA binding process, single nucleotide changes were introduced into the LexA3 probe by PCR amplification with primers AR41 to AR57 and AR76 (Table 1). The changes introduced were the least permissive based on the thorough characterization of the DinR binding site performed by Winterling et al. (23). Changes introduced in the main motifs GAAC and GTTC severely decreased or eliminated LexA binding (Fig. 4A). Also critical for binding were the flanking bases of the extended direct repeat, the adenine residues at positions -29 and -24, and the thymine residues at positions -16 and -21 (relative to the TTG translational start codon). These results indicated that the tetranucleotide GAAC and the inverted repeat GTTC are critical to LexA binding, consistent with previous studies of the DinR recognition site (23). Collectively, these results demonstrate that the sequence AGAACN₄GTTCT is the specific recognition site for the *D. ethenogenes* LexA repressor.

The *Bacillus subtilis* DinR repressor recognizes the *D. ethenogenes* LexA binding site. The electrophoretic mobility shift assay experiments performed with the wild-type and mutated variants of the *D. ethenogenes* *lexA* operator revealed that A and T at positions 1 and 14, respectively, of the AGAACN₄GTTCT sequence are important bases for recognition by LexA. Nevertheless, the consensus LexA-box sequence established for *B. subtilis* and other gram-positive bacteria is CGAACN₄GTTTCG (23). Indeed, these two terminal nucleotides are part of only a few gram-positive LexA-boxes, especially the T at position 14, which is observed in only 1 of 20 boxes, the *Streptomyces lividans* *recA* operator (23).

In order to confirm that the LexA-box of *D. ethenogenes* is the same as that of gram-positive species, we analyzed the capacity of the DinR repressor to bind to the *D. ethenogenes* *lexA* operator. To examine binding, digoxigenin-labeled LexA1 DNA was incubated in the presence of purified *B. subtilis* DinR protein or *D. ethenogenes* LexA with or without DNA competitors containing the *B. subtilis* DinR- or *D. ethenogenes* LexA-box. The results (Fig. 4B) show that the DinR protein bound to the *D. ethenogenes* *lexA* operator and that binding was abolished by competition with unlabeled DNA containing either the *D. ethenogenes* LexA-box or the *B. subtilis* DinR-box. All these results clearly indicate that *D. ethenogenes* LexA protein specifically recognizes the motif GAACN₄GTTTC, as does DinR. Hence, *D. ethenogenes* is the first bacterium outside of

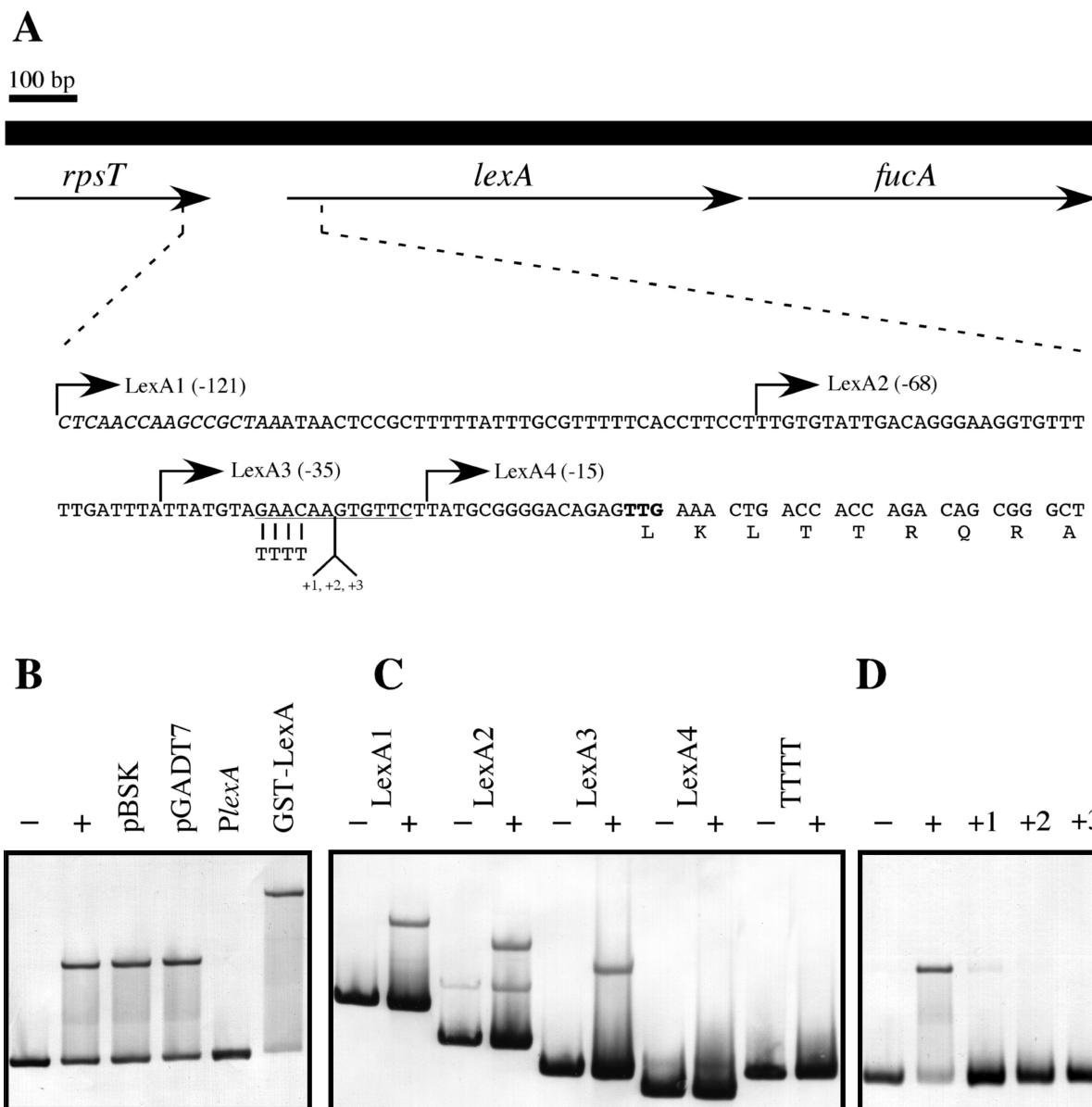


FIG. 3. (A) Schematic representation of the *D. ethenogenes* chromosomal region containing *lexA*, which is located between the *rpsT* and *fucA* genes. The sequence of the *lexA* promoter region and the first nine amino acid residues of LexA are shown. The 3' end of the *rpsT* gene is in italic type, the putative TTG translational start codon of *lexA* is in bold type, and the homologue of the DinR-box is underlined. The start points of each fragment of the *lexA* promoter (LexA1, LexA2, LexA3, and LexA4) used in electrophoretic mobility shift assay experiments are indicated by arrows, and the relative distances to the TTG are in parentheses. (B) Specific binding of *D. ethenogenes* LexA protein. Electrophoretic mobility shift assay experiments were performed with different DNAs as competitors. LexA1 fragment was incubated at 30°C for 30 min in the absence (-) and in the presence (+) of 25 ng of pure LexA protein (50 nM). To show the specificity of LexA binding, the LexA1 fragment was incubated in the presence of *D. ethenogenes* LexA protein and the following unlabeled DNA competitors: 2 µg of pBSK, 2 µg of pGADT7, or 1 µg of unlabeled LexA1 DNA fragment (*PlexA*). In the last lane, LexA1 probe was incubated in the presence of the *D. ethenogenes* GST-LexA fusion protein. (C) Establishment of the limits of the *D. ethenogenes* LexA binding site. LexA1, LexA2, LexA3, LexA4, and a derivative of the LexA3 fragment where GAAC was changed to TTTT (TTTT) were incubated in the absence (-) or presence (+) of purified LexA. (D) Electrophoretic mobility shift assays of LexA3 fragments where one (+1), two (+2), or three (+3) adenine residues were inserted at position -21 with respect to the initiation start codon. Different probes were incubated with LexA protein and loaded on a native Tris-glycine-5% polyacrylamide gel. The LexA3 probe alone (-) and the same fragment incubated in the presence of LexA (+) acted as negative and positive controls, respectively.

the gram-positive phylum whose LexA repressor binds to the DinR-box.

The LexA regulon is different in *B. subtilis* than in *D. ethenogenes*. In *Escherichia coli*, at least 40 genes are directly under the negative control of the LexA repressor (3, 5). The amino

acid sequences of each of these *E. coli* SOS genes were used as queries in BLAST searches of the unfinished *D. ethenogenes* genome to seek homologous genes in this bacterial species. Homologues of *recA*, *uvrA*, *ruvA*, *ruvB*, *uvrB*, *uvrD*, *ssb*, *dinB*, and *recN* were identified. However, homologues to the rest of

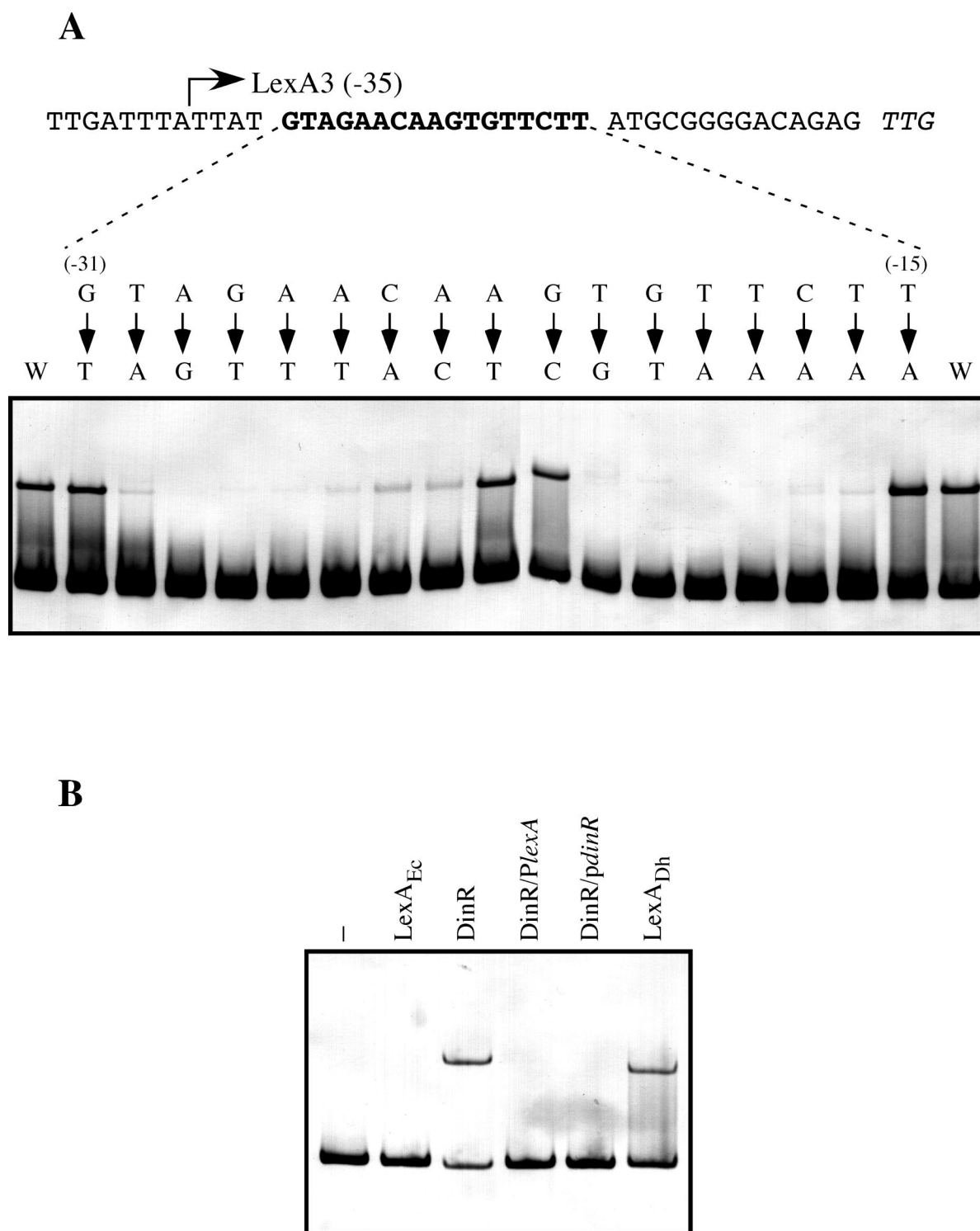


FIG. 4. (A) Determination of the specific LexA binding site in the *D. ethenogenes* *lexA* promoter. Directed mutagenesis via PCR was used to change single nucleotides in the region from -31 to -15 (bold type) relative to the putative start codon of the *D. ethenogenes* *lexA* gene (italics). Electrophoretic mobility shift assays were used to determine the affinity of the purified LexA protein for the resulting digoxigenin-labeled DNA fragments. Arrows point to the nucleotide used to replace the native nucleotide. The relative positions of nucleotides with respect to the translation initiation codon are indicated in parentheses. The wild-type LexA3 fragment was used as a positive control (W). (B) *Bacillus subtilis* DinR recognizes the *D. ethenogenes* LexA binding site. Digoxigenin-labeled LexA1 fragment was incubated in the absence (-) or presence of 25 ng of either purified *E. coli* LexA (LexA_{Ec}), pure *B. subtilis* DinR (DinR), or *D. ethenogenes* LexA (LexA_{Dh}). Furthermore, the LexA1 fragment and the DinR protein were incubated in the presence of a 100-fold molar excess of either the *B. subtilis* *dinR* (DinR/PdinR) or *D. ethenogenes* *lexA* (DinR/PlexA) promoter sequence.

A

uvrA TAGAACACATGTTTTG...**50bp**..GTG
ruvA TGGAAAGCCCGTTTGC...**287bp**..ATG
ruvB TGGAAATATTTGTGCTA...**11bp**..ATG

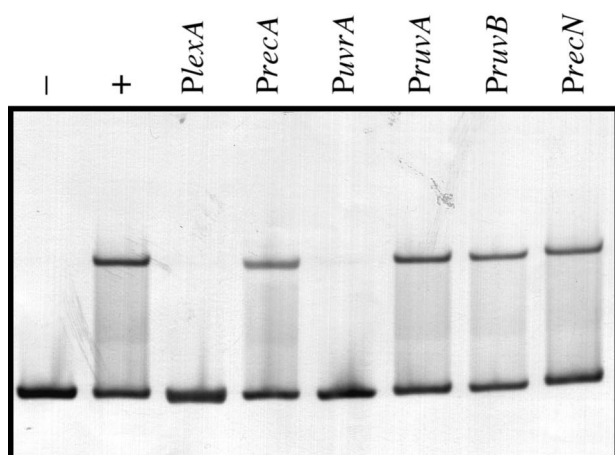
B

FIG. 5. (A) Upstream sequences of *D. ethenogenes* *uvrA*, *ruvA*, and *ruvB* genes presenting putative LexA binding sites. Potential LexA binding sites are underlined, and the putative translation initiation codons are depicted in bold letters, with the distance between them indicated. (B) Electrophoretic mobility shift assay of the LexA1 fragment incubated with purified *D. ethenogenes* LexA in the presence of different DNA competitors. LexA1 fragment was incubated at 30°C for 30 min in the absence (-) or presence (+) of pure LexA protein. Simultaneous reactions contained the LexA1 fragment, *D. ethenogenes* LexA protein, and one of six unlabeled DNA competitors (100-fold molar excess) containing the promoter regions and potential LexA binding sites of genes *lexA* (*PlexA*), *recA* (*PrecA*), *uvrA* (*PuvrA*), *ruvA* (*PruvA*), *ruvB* (*PruvB*), and *recN* (*PrecN*).

the *E. coli* LexA-regulated genes could not be identified, possibly because the *D. ethenogenes* genome has not been completely sequenced or because they simply do not exist in this microorganism. Subsequently, the promoters of the genes identified were analyzed to find potential LexA binding sites.

One clear motif, similar to the 12 bp found in the *lexA* promoter, was identified 64 bp upstream of the putative GTG translational start codon of the *uvrA* gene (Fig. 5A). Potential motifs were also found in the promoter regions of the *ruvA* and *ruvB* genes 301 and 25 bp upstream of their putative translation initiation codons, respectively (Fig. 5A). Examination of the remaining SOS gene homologues failed to reveal any potential LexA binding sites. Strikingly, at first glance, *recA* does not seem to be regulated by LexA because no LexA binding motif has been identified in its promoter region. To determine

if these potential LexA binding sites were functional, PCR-amplified fragments of the promoter regions from *uvrA*, *ruvA*, *ruvB*, *recN*, and *recA* were used as competitors in band shift assays with the LexA1 DNA fragment as the labeled probe. While a 100-fold molar excess of DNA containing the *uvrA* promoter was able to abolish the band belonging to the DNA-LexA complex, the shifted band did not disappear when the same amount of DNA containing the *recA*, *ruvA*, *ruvB*, or *recN* promoter was used as the competitor (Fig. 5B). This result clearly indicates that LexA binds the *uvrA* promoter, suggesting that the motif AGAACACATGTTTT is indeed a *D. ethenogenes* LexA recognition site.

The lack of a visible LexA binding motif in the *recA* promoter sequence is consistent with the inability of the *recA* promoter fragment to abolish the retarded band. This confirms that no LexA binding motif is present in the *recA* promoter region and indicates that the *recA* gene in *D. ethenogenes* is not under the direct control of the LexA repressor. This has also been observed for *Deinococcus radiodurans*, where the *recA* gene is induced by gamma radiation in a LexA-independent manner. However, it has been reported that *D. radiodurans* possesses a second *lexA* gene that might be involved in the direct regulation of *recA* (18). Our results demonstrate that *D. ethenogenes* LexA does not bind to the *recA* promoter, thus constituting the first report of a *recA* gene that is not regulated directly by the LexA repressor.

On the basis of the signature sequences found in various proteins, it has been recently proposed that the green nonsulfur bacterial phylum, to which *D. ethenogenes* belongs, evolved from a universal ancestor later than the gram-positive bacteria but before the cyanobacteria (9). The *Deinococcus-Thermus* group seems to have appeared before the green nonsulfur bacteria, although this order has not been definitively fixed (9). Our results demonstrating that the LexA-box sequences of *B. subtilis* and *D. ethenogenes* are nearly identical and that the DinR repressor from a gram-positive organism binds to the *D. ethenogenes* LexA box give support to the hypothesis that a close phylogenetic relationship exists between the gram-positive and green nonsulfur bacteria. However, conclusions about the order of evolution of the green nonsulfur bacteria and the *Deinococcus-Thermus* group cannot be made on this basis due to the current lack of information on the LexA binding motif for members of the latter group.

Despite having similar LexA binding sequences, gram-positive bacteria and *D. ethenogenes* display significant differences in the composition of their LexA regulons. A BLAST search of GenBank indicates that genes with an upstream LexA binding sequence are more numerous in the gram-positive bacteria *B. subtilis*, *Staphylococcus aureus*, *Mycobacterium tuberculosis*, and *Clostridium acetobutylicum* than in *D. ethenogenes* (Table 2). The reasons for these differences are not clear but could be related to the environmental conditions in which each of these organisms has evolved.

Finally, performing a TBLASTN analysis with *D. ethenogenes* LexA protein as a query at the DOE Joint Genome Institute, a LexA homologue was found in the unfinished genome of the photosynthetic bacterium *Chloroflexus aurantiacus*, which is also a member of the green nonsulfur bacteria. The putative *lexA* gene is found at the 5' end of a short contig (694), and thus the N terminus of the protein and the potential

TABLE 2. Comparison of the presence of a LexA binding sequence upstream of different genes putatively belonging to the LexA regulon from several gram-positive bacteria, *D. ethenogenes*, and *E. coli*^a

Gene	<i>B. subtilis</i>	<i>S. aureus</i>	<i>M. tuberculosis</i>	<i>C. acetobutylicum</i>	<i>D. ethenogenes</i>	<i>E. coli</i>
<i>recA</i>	+	+	+	+	–	+
<i>lexA</i>	+	+	+	+	+	+
<i>ruvAB</i>	+	–	+	–	– ^b	+
<i>uvrA</i>	–	–	–	–	+	+
<i>uvrB</i>	+	–	–	+	–	+
<i>uvrD</i>	+	–	–	–	–	+
<i>ssb</i>	–	+	–	–	–	+
<i>recN</i>	–	–	–	–	–	+

^a The presence or absence of a LexA binding site is represented as + and –, respectively. Searches of the whole genomes of several bacteria were performed with the RCGScanner program, which uses a recursive method (20) to identify a given nucleotide sequence as well as its neighboring open reading frames (Erill et al., submitted).

^b The *ruvA* and *ruvB* genes do not form an operon in *D. ethenogenes*.

LexA-box could not be identified. However, the truncated LexA protein of *C. aurantiacus* contains the helix-turn-helix motif, which is involved in DNA binding (7), showing 87% similarity with the *D. ethenogenes* helix-turn-helix domain. This fact suggests that *C. aurantiacus* LexA might also recognize the GAACN₄GTTC as in gram-positive bacteria. The availability of the assembled genome of this organism will be very useful for understanding the evolutionary and phylogenetic history of the green nonsulfur bacteria, for confirming the homogeneity of this phylum with respect to the LexA regulatory sequence, and for determining the composition of the LexA regulon. These data will also be useful for testing the hypothesis, cited above, that a relationship exists between environmental conditions and the genes which are under control of the LexA protein in bacteria.

This work was funded by grants BMC2001-2065 from the Ministerio de Ciencia y Tecnología (MCyT) de España and 2001SGR-206 from the Departament d'Universitats, Recerca i Societat de la Informació (DURSI) de la Generalitat de Catalunya. Antonio R. Fernández de Henestrosa is a recipient of a postdoctoral reincorporation contract from the MCyT.

We acknowledge Joan Ruiz, Susana Escribano, and Pilar Cortés for excellent technical assistance. Preliminary *D. ethenogenes* sequence data were obtained from the Institute for Genomic Research website at <http://www.tigr.org>. The *Chlorefexus aurantiacus* partial *lexA* sequence was provided free by the U.S.-DOE Joint Genome Institute (<http://www.jgi.doe.gov/>). We are deeply grateful to Roger Woodgate and Kevin Winterling for generously providing us with *B. subtilis* genomic DNA, *B. subtilis* DinR, and *E. coli* LexA protein.

REFERENCES

- Brent, R., and M. Ptashne. 1981. Mechanism of action of the *lexA* gene product. Proc. Natl. Acad. Sci. USA **78**:4204–4208.
- Brooks, P. C., F. Movahedzadeh, and E. O. Davis. 2001. Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*: apparent lack of correlation with LexA binding. J. Bacteriol. **183**:4459–4467.
- Courcelle, J., A. Khodursky, B. Peter, P. O. Brown, and P. C. Hanawalt. 2001. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. Genetics **158**:41–64.
- Eisen, J. A., and P. C. Hanawalt. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. Mutat. Res. **435**:171–213.
- Fernandez de Henestrosa, A. R., T. Ogi, S. Aoyagi, D. Chafin, J. J. Hayes, H. Ohmori, and R. Woodgate. 2000. Identification of additional genes belonging to the LexA-regulon in *Escherichia coli*. Mol. Microbiol. **35**:1560–1572.
- Fernandez de Henestrosa, A. R., E. Rivera, A. Tapias, and J. Barbe. 1998. Identification of the *Rhodobacter sphaeroides* SOS box. Mol. Microbiol. **28**:991–1003.
- Fogh, R. H., G. Oettleben, H. Ruterjans, M. Schnarr, R. Boelens, and R. Kaptein. 1994. Solution structure of the LexA repressor DNA binding domain determined by ¹H NMR spectroscopy. EMBO J. **13**:3936–3944.
- Friedberg, E. C., G. C. Walker, and W. Siede. 1995. DNA repair and mutagenesis. American Society for Microbiology, Washington, D.C.
- Gupta, R. S. 2000. The natural evolutionary relationships among prokaryotes. Crit. Rev. Microbiol. **26**:111–131.
- Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180**:4765–4774.
- Khil, P. P., and R. D. Camerini-Otero. 2002. Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. Mol. Microbiol. **44**:89–105.
- Koch, W. H., and R. Woodgate. 1998. The SOS response, p. 107–134. In J. A. Nickoloff and M. F. Hoekstra (ed.), DNA damage and repair: DNA repair in prokaryotes and lower eukaryotes, 1st ed. Humana Press, Totowa, N.J.
- Little, J. W. 1993. LexA cleavage and other self-processing reactions. J. Bacteriol. **175**:4943–4950.
- Little, J. W., S. H. Edmiston, L. Z. Pacelli, and D. W. Mount. 1980. Cleavage of the *Escherichia coli* LexA protein by the RecA protease. Proc. Natl. Acad. Sci. USA **77**:3225–3229.
- Little, J. W., D. W. Mount, and C. R. Yanisch-Perron. 1981. Purified LexA protein is a repressor of the *recA* and *lexA* genes. Proc. Natl. Acad. Sci. USA **78**:4199–4203.
- Luo, Y., R. A. Pfuetzner, S. Mosimann, M. Paetzel, E. A. Frey, M. Cherney, B. Kim, J. W. Little, and N. C. Strynadka. 2001. Crystal structure of LexA: a conformational switch for regulation of self-cleavage. Cell **106**:585–594.
- Maymó-Gatell, X., Y. Chien, J. M. Gossett, and S. H. Zinder. 1997. Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. Science **276**:1568–1571.
- Narumi, I., K. Satoh, M. Kikuchi, T. Funayama, T. Yanagisawa, Y. Kobayashi, H. Watanabe, and K. Yamamoto. 2001. The LexA protein from *Deinococcus radiodurans* is not involved in RecA induction following gamma irradiation. J. Bacteriol. **183**:6951–6956.
- Radman, M. 1974. Phenomenology of an inducible mutagenic DNA repair pathway in *Escherichia coli*: SOS repair hypothesis, p. 128–142. In L. Prakash, F. Sherman, M. Miller, C. W. Lawrence, and H. W. Tabor (ed.), Molecular and environmental aspects of mutagenesis. Charles C Thomas, Springfield, Ill.
- Rodionov, D. A., A. A. Mironov, and M. S. Gelfand. 2001. Transcriptional regulation of pentose utilisation systems in the *Bacillus/Clostridium* group of bacteria. FEMS Microbiol. Lett. **205**:305–314.
- Tapias, A., and J. Barbe. 1999. Regulation of divergent transcription from the *uvrA-ssb* promoters in *Sinorhizobium meliloti*. Mol. Gen. Genet. **262**:121–130.
- Walker, G. C. 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol. Rev. **48**:60–93.
- Winterling, K. W., D. Chafin, J. J. Hayes, A. S. Levine, R. E. Yasbin, and R. Woodgate. 1998. The *Bacillus subtilis* DinR binding site: redefinition of the consensus sequence. J. Bacteriol. **180**:2201–2211.
- Winterling, K. W., A. S. Levine, R. E. Yasbin, and R. Woodgate. 1997. Characterization of DinR, the *Bacillus subtilis* SOS repressor. J. Bacteriol. **179**:1698–1703.