An article to *Microbiology*

Section: Genes and Genomes

Reconstruction of the evolutionary history of the LexA binding sequence

by

Gerard Mazón[1], Ivan Erill[2], Susana Campoy[3], Pilar Cortés[1], Evelyne Forano[4] and Jordi Barbé [1,3,5]

[1] *Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona 08193 Bellaterra, Spain*

[2] *Biomedical Applications Group, Centro Nacional de Microelectrónica, 08193 Bellaterra, Spain*

[3] *Centre de Recerca en Sanitat Animal (CReSA), 08193 Bellaterra, Spain*

[4] *Unité de Microbiologie, INRA, Centre de Recherches de Clermont-Ferrand-Theix, 63122 Saint-Genès-Champanelle, France*

[5] To whom reprint requests should be addressed.

E-mail: jordi.barbe@uab.es

Phone: 34 - 93 – 581 1837

Fax: 34 – 93 – 581 2387

Running Title: Evolution of the LexA binding sequence

Keywords: *lexA*, SOS response, evolution, lateral gene transfer

SUMMARY

In recent years, the recognition sequence of the SOS repressor LexA protein has been identified for several bacterial clades, such as the Gram-positive, Green-non Sulfur bacteria and Cyanobacteria phyla, or the Alpha, Delta and Gamma Proteobacteria classes. Nevertheless, the evolutionary relationship among these sequences and the proteins that recognize them has not been analyzed. *Fibrobacter succinogenes* is an anaerobic Gram-negative bacterium that branched from a common bacterial ancestor immediately before the Proteobacteria phylum. Taking advantage of its intermediate position in the phylogenetic tree, and in an effort to reconstruct the evolutionary history of LexA binding sequences, the *F. succinogenes lexA* gene has been isolated and its product purified to identify its DNA recognition motif through electrophoretic mobility assays and footprinting experiments. After comparing the available LexA DNA binding sequences with the here reported *F. succinogenes* one, directed mutagenesis of the *F. succinogenes* LexA binding sequence and phylogenetic analyses of LexA proteins have revealed the existence of two independent evolutionary lanes for the LexA recognition motif that emerged from the Gram-positive box: one generating the Cyanobacteria and Alpha Proteobacteria LexA binding sequences, and the other giving rise to the *F. succinogenes* and *Myxococcus xanthus* ones, in a transitional step towards the current Gamma Proteobacteria LexA box. The contrast between the results here reported and the phylogenetic data available in the literature suggests that, some time after their emergence as a distinct bacterial class, the Alpha Proteobacteria lost its vertically received *lexA* gene, but received later through lateral gene transfer a new *lexA* gene belonging to either a Cyanobacterium or a bacterial species closely related to this Phylum. This constitutes

the first report based on experimental evidence of lateral gene transfer in the evolution of a gene governing such a complex regulatory network as the bacterial SOS system.

INTRODUCTION

Preservation of genetic material is one of the most fundamental functions of any living being and it is perhaps in the Bacteria Domain where this aspect has been most thoroughly studied. As in the case of many other biological processes, *Escherichia coli* has been the principal subject of this research, and many *E. coli* genes involved in preservation of genetic material have been identified through the years. Some of them encode proteins that are able to repair different types of DNA injuries, whilst others aim at guaranteeing cell survival in the presence of such lesions. Many of these genes act in a coordinate manner, constituting specific DNA repair networks, and the broadest and most thoroughly studied of these regulons is the LexA-mediated SOS response (Walker, 1984). In *E. coli*, the LexA protein controls the expression of some 40 genes (Fernández de Henestrosa *et al.*, 2000; Courcelle *et al.*, 2001), including both the *lexA* and *recA* genes, which are, respectively, the negative and positive regulators of the SOS response (Walker, 1984). The *E. coli* LexA protein specifically recognizes and binds to an imperfect 16-bp palindrome with consensus sequence $CTGTN_8ACAG$, designated as the *E. coli* SOS or LexA box (Walker, 1984). Both *in vitro* and *in vivo* experiments have shown that binding to single-stranded DNA fragments generated by DNA damage-mediated inhibition of replication activates the RecA protein (Sassanfar & Roberts, 1990). Once in its active state, RecA promotes the autocatalytic cleavage of LexA, resulting in the expression of the genes regulated by this repressor (Little, 1991). Hydrolysis of the *E. coli* LexA protein is mediated by its Ser_{119} and Lys_{156} residues, in a mechanism similar to that of proteolysis by serine proteases (Luo *et al.*, 2001). After DNA repair, the RecA protein ceases to be

activated and, consequently, non-cleaved LexA protein returns to its usual levels, repressing again the genes that are under its direct negative control.

Even though some notable exceptions have been reported, the increasing availability of microbial genome sequences has revealed that LexA is present in many bacterial species and in most phyla. So far, all the identified and characterized LexA proteins display two conserved domains that are clearly differentiated. The N-domain, ending at the Ala-Gly bond where the protein is cleaved after DNA damage activation of RecA (Little, 1991), has three α helices that are necessary for the recognition and binding of LexA to the SOS box (Fogh *et al.,* 1994; Knegtel *et al.,* 1995). Conversely, the C-domain contains amino acids that are essential for the serine-protease mediated auto-cleavage and for the dimerization process necessary for repression (Luo et al., 2001).

The sequence of the LexA box is strongly conserved among related bacterial species. In fact, the LexA box has been shown to be monophyletic for several bacterial phyla, and this feature has been successfully exploited in phylogenetic analyses (Erill *et al.,* 2003). Thus, in the Gram-positive Phylum the LexA binding motif presents a CGAACRNRYGTTYC consensus sequence (Winterling *et al.,* 1998) that, with slight variations (Davis *et al.,* 2002), is conserved among all its members and is also found in the phylogenetically close Green Non-sulfur Bacteria that, nonetheless, are Gram-negative bacteria (Fernández de Henestrosa *et al.,* 2002). Apart from the Gamma Proteobacteria, in which the consensus sequence $CTGTN_8ACAG$ is monophyletic and seems to extend to those Beta Proteobacteria that present a *lexA* gene (Erill *et al.,* 2003), alternative LexA binding sequences with a high degree of conservation have also been described in other groups. So far, for instance, the direct repeat $GTTCN_7GTTC$ is the LexA binding sequence of the Alpha Proteobacteria harboring a

lexA gene, a group that includes the *Rhodobacter, Shinorizobium, Agrobacteium, Caulobacter* and *Brucella* genera (Fernández de Henestrosa *et al.*, 1998; Tapias & Barbe, 1999). Still, in other phyla where the LexA binding motif has been identified more data is required to gauge the conservation of the LexA box. Such is the case of the Delta Proteobacteria, for which a CTRHAMRYBYGTTCAGS consensus motif has been identified in one of its members, the fruiting body forming *Myxococcus xanthus* (Campoy *et al.*, 2003).

The existence of different LexA recognition motifs and the monophyletic or paraphyletic nature of those studied so far indicate that the appearance of new LexA binding motifs marks turning points in the evolutionary history of both this protein and its respective host species. Previous work has demonstrated that the Cyanobacteria LexA box (RGTACNNNDGTWCB) derives directly from that of Gram-positive bacteria (Mazón *et al.*, 2004). Nevertheless, a huge gap is still apparent in the further evolutionary pathway of the LexA box that leads from the Cyanobacteria up to other bacterial phyla of later appearance, such as the Proteobacteria. Protein signature analyses have established that *Fibrobacter succinogenes* branched from a common bacterial ancestor immediately before the Proteobacteria phylum (Griffiths & Gupta, 2001). *F. succinogenes* is an anaerobic gram-negative bacterium that inhabits the rumen and caecum of herbivores and, for a long time, this organism was included in the *Bacteroides* genus. Recent 16S rRNA analyses, however, have granted *Fibrobacter* a new Bacterial Phylum of its own (Maidak *et al.*, 1999; Ludwig & Schleifer, 1999).

In an effort to recreate the evolutionary history of the LexA protein through the changes in its recognition sequence, and taking advantage of the fact that the *F. succinogenes* genome is now partially sequenced, the *lexA* gene of this bacterial

species has been isolated and its encoded product has been purified to determine its DNA recognition sequence. The results here obtained are in accordance with the newly established branching point of *F. succinogenes*, and introduce a novel element that allows a finer drawing of the evolutionary path of the LexA recognition sequence from Gram-positive bacteria to Gamma Proteobacteria.

METHODS

Bacterial strains, plasmids, oligonucleotides and DNA techniques. Bacterial strains and plasmids used in this work are listed in Table 1. *E. coli* and *F. succinogenes* ATCC19169 strains were grown at either in LB (Sambrook *et al.,* 1992) or in a chemically defined medium (Gaudet *et al.,* 1992) with 3 g l^{-1} of cellobiose, respectively. Antibiotics were added to the cultures at reported concentrations (Sambrook *et al.,* 1992) . *E. coli* cells were transformed with plasmid DNA as described (Sambrook *et al.,* 1992). All restriction enzymes, PCR-oligonucleotide primers, T4 DNA ligase and polymerase, and the "DIG-DNA labelling and detection kit" were from Roche. DNA from *F. succinogenes* cells was extracted as described (Forano *et al.,* 1994).

The synthetic oligonucleotide primers used for PCR amplification are listed in Table 2. To facilitate subcloning of some PCR-DNA fragments, specific restriction sites were incorporated into the oligonucleotide primers. These restriction sites are identified in Table 2. Mutants in the *F. succinogenes lexA* promoter were obtained by PCR-mutagenesis, using oligonucleotides carrying designed substitutions (Table 2). The DNA sequence of all PCR-mutagenized fragments was determined by the dideoxy method (Sanger *et al.,* 1977) on an ALF Sequencer (Amersham-Pharmacia). In all cases the entire nucleotide sequence was determined for both DNA strands.

Molecular cloning of the *F. succinogenes lexA* gene and purification of its encoded protein. The *F. succinogenes lexA* gene was amplified from the total DNA of the *F. succinogenes* ATCC19169 strain using the LexAup and LexAdwn oligonucleotide primers (Table 2) corresponding to nucleotides –276 to –249 and +653 to +678, with respect to its proposed translational starting point. The 954-bp PCR fragment obtained was cloned into the pGEM-T vector (Promega) obtaining the pUA1033 plasmid. To confirm that no mutation was introduced during the amplification reaction, the sequence of the fragment was determined. The plasmid pUA1038 was constructed in order to create and express a Glutathione-S-transferase (GST)- *F. succinogenes* LexA fusion protein. The first step in the construction of this plasmid was to amplify the *F. succinogenes lexA* gene from plasmid pUA1033, using the primers LexAEcoRI and LexADw. The resulting DNA fragment was cloned into pGEM-T, to give rise to pUA1037. Following excision with *Eco*RI and *Sal*I, the *lexA* gene was inserted into the pGEX4T1 expression vector (Amersham-Pharmacia), immediately downstream of the GST-encoding gene that is under the T7 promoter control. The initiation codon of the LexA protein was placed immediately downstream of the *Eco*RI sites in LexAEcoRI primer, such that the *lexA* gene could be fused to GST in frame. The insert of pUA1037 was sequenced in order to ensure that no mutations were introduced during amplification.

To overproduce the LexA-GST fusion protein, the pUA1037 plasmid was transformed into *E. coli* BL21(λDE3) codon plus strain (Stratagene). Cells of the resulting BL21 codon plus strain were diluted in 0.5 L of LB medium and incubated at 37^{o}C until they reached an O.D.$_{600}$ of 0.8. Fusion protein expression was induced at this time by the addition of IPTG to a final concentration of 1 mM. Following incubation for an additional 3 h at 37^{o}C, cells were collected by centrifugation for 15

min at 3 000 ***g***. The bacterial pellet was resuspended in PBS buffer (10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, 140 mM NaCl, 2.7 mM KCl -pH 7.4-), containing «Complete Mini» protease inhibitors cocktail (Roche). The resulting cell suspensions were lysed by sonication. Unbroken cells and debris were removed by centrifugation for 20 min at 14 000 ***g.*** The supernatant containing the GST-LexA fusion protein was incubated with PBS-Glutathione Sepharose 4B® beads (Amersham Pharmacia), for 2 h at 4°C, in order to affinity purify the fusion protein. The beads were then washed twice with PBS containing 0.1% Triton and three times with PBS without detergent.

The sequence Leu-Val-Pro-Arg-Gly-Ser is located immediately downstream of the GST coding sequence in the pGEX4T vector series, and serves as a linker between the LexA and GST moieties of the fusion proteins. This hexapeptide is recognized by the protease thrombin, which cleaves at the Arg-Gly bond. It was therefore possible to release the *F. succinogens* LexA protein from the sepharose beads by incubating a 700 µl bed volume of beads with 25 Units of thrombin (Amersham - Pharmacia) in 1 ml of PBS. The supernatants containing the *F. succinogenes* LexA protein with an additional five amino acid tail at their N-terminal (Gly-Ser-Pro-Glu-Phe), was visualized in a Coomassie blue stained 13% SDS-PAGE gel (Laemmli, 1970). Their purity was greater than 98% (data not shown).

LexA proteins from *B. subtilis, E. coli, Anabaena* PCC7120, *M. xanthus* and *R. sphaeroides* also used in this work had been previously purified (Winterling *et al.*, 1998; Tapias *et al.*, 2002; Campoy *et al.*, 2003; Mazón *et al.*, 2004).

Mobility shift assays and DNase I footprinting. LexA-DNA complexes were detected by electrophoresis mobility shift assays (EMSAs) using purified LexA proteins. DNA probes were prepared by PCR amplification using one of the primers labelled at its 5' end with digoxigenin (DIG) (Table 2), purifying each product in a

2% -3% low-melting-point agarose gel depending on DNA size. DNA-protein reactions (20 µl) typically containing 10 ng of DIG-DNA-labelled probe and 40 nM of the desired purified LexA protein were incubated in binding buffer: 10 mM *N*-2-Hydroxyethyl-piperazine-*N'* 2-ethanesulphonic acid (HEPES) NaOH (pH 8), 10 mM Tris-HCl (pH 8), 5% glycerol, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 2 µg poly(dG-dC) and 50 µg/ml of BSA. After 30 minutes at 30°C, the mixture was loaded onto a 5% non-denaturing Tris-glycine polyacrylamide gel (pre-run for 30 minutes at 10 V/cm in 25mM Tris-HCl (pH 8.5), 250mM glycine, 1mM EDTA). DNA-protein complexes were separated at 150 V for 1 hr, followed by transfer to a Biodine B nylon membrane (Pall Gelman Laboratory). DIG-labelled DNA-protein complexes were detected by following the manufacturer's protocol (Roche). For the binding-competition experiments, a 300-fold molar excess of either specific or unspecific-unlabelled competitor DNA was also included in the mixture. Protein concentrations were determined as described (Bradford, 1976). All EMSAs were repeated a minimum of three times to ensure reproducibility of the results.

DNase I footprinting assays were performed using the ALF Sequencer (Amersham Biosciences) as described previously (Patzer and Hantke 2001; Campoy *et al.*, 2003).

***In silico* phylogenetic analysis.** Preliminary sequence data of *F. succinogenes* unfinished genome was obtained from The Institute for Genomic Research (TIGR) through their website at http://www.tigr.org, and protein sequences for all other organisms were obtained from the Microbial Genome Database for Comparative Analysis website (http://mbgd.genome.ad.jp/) and the TIGR Comprehensive Microbial Resource (CMR). Identification of additional LexA-binding genes was carried out using the RCGScanner software (Erill *et al.*, 2003), using known *E. coli* LexA-governed genes (Fernández de Henestrosa *et al.*, 2000; Erill *et al.*, 2003) and

the here reported LexA box of *F. succinogenes* to scan and then filter through the consensus method putative LexA binding sites across the *F. succinogenes* genome.

For phylogenetic analyses, protein sequences for each gene under study were aligned using the CLUSTALW program (Higgins *et al.*, 1994). Multiple alignments were then used to infer phylogenetic trees with the SEQBOOT, PROML and CONSENSE programs of the Phylip 3.6 software package (Felsenstein, 1989), applying the maximum-likelihood method on 100 bootstrap replicates. The resulting phylogeny trees were plotted using TreeView (Page, 1996).

RESULTS

Determination of the *F. succinogenes* LexA recognition DNA sequence

Electrophoretic mobility shift assays (EMSA) with the purified *F. succinogenes* LexA protein were carried out to determine the binding ability of this protein to its own promoter. As it can be seen in Fig. 1a, the addition of increasing concentrations of LexA to a fragment extending from –154 to +169 of the *F. succinogenes lexA* gene promoter (with respect to its proposed translational starting point) produces one retardation band whose intensity is directly related to the amount of protein used. The formation of this DNA-LexA complex is specific, since it is sensitive to competition by an excess of unlabelled *lexA* promoter, but not to competition by non-specific DNA (Fig. 1b). Moreover, EMSAs performed using different sized-fragments containing the *lexA* promoter as a probe demonstrated that the LexA recognition sequence must lie in a region included between positions –72 and –57 of this promoter (data not shown).

To precisely identify the *F. succinogenes* LexA box, additional footprinting experiments with a 160-bp fragment extending from positions –154 to +6 were

performed. The results obtained show that a 37 bp core region was protected by the LexA protein when both *lexA*-coding and non-coding strands were analyzed (Fig. 2). A visual inspection of this DNA sequence revealed the presence of the imperfect palindrome TGCCCAGTTGTGCA in its central region. To determine whether this motif was really involved in LexA binding, the effect of single substitutions in each nucleotide of this palindrome on the formation of the LexA protein-*lexA* promoter complex was analyzed. Results (Fig. 3) indicate that a single substitution in any position of the TGC tri-nucleotide, as well as in the last C of the TGCCC motif, abolishes LexA binding. Likewise, mutagenesis of any position of the GTGCA motif does also inhibit DNA-LexA complex formation. On the contrary, the single substitution of nucleotides immediately surrounding either the TGCCC or the GTGCAT motifs does not affect LexA binding. Taken together, these results show that the TGCNCNNNNGTGCA imperfect palindrome is the *F. succinogenes* LexA-binding sequence of the lexA gene, since it is required for the binding of the LexA protein to the *lexA* promoter. Additional single substitutions, generating TGCAC-N4-GTGCA and TGCCC-N4-GGGCA perfect palindromes were carried out (data not shown), demonstrating that a TGCAC-N4-GTGCA perfect palindrome is the most likely consensus for the LexA box of *F. succinogenes*.

Identification of additional LexA-binding *F. succinogenes* genes

The characteristic amino acid residues of LexA proteins (an Ala-Gly bond separated about 34 positions from a Ser residue that is 37 positions away from a Lys residue) are also present, at least, in two other prokaryotic protein families: UmuD (encoding DNA polymerase V which is involved in error-prone DNA repair) and lytic cycle prophage repressors (such as the λcI protein) (Little, 1984; Burckhardt *et al.*,1988;

Nohmi *et al.*, 1988). Nevertheless, of these two proteins only the prophage repressors are able to bind DNA specific sequences. To discard the possibility that the *F. succinogenes* LexA was, in fact, a residual prophage repressor, a phylogenetic analysis of relevant LexA proteins was performed. The results (Fig. 4), together with phylogenetic trees including the λ CI repressor as an outgroup (data not shown), indicate that the here identified *F. succinogenes* LexA protein is most probably a descendent of a Gram-positive LexA protein, and rule out the possibility of LGT from such an unspecified source as a residual prophage. To further validate this hypothesis, an *in silico* analysis of the *F. succinogenes* genome sequence was carried out using the RCGScanner program (Erill *et al.*, 2003) in search of other genes with significant TGCNCNNNNGTGCA-like palindrome motifs upstream of their coding regions. Imperfect palindrome motifs were found upstream the *recA* (TTGCACAAAAGTTCAC), *uvrA* (CTATTCAAATGTTCAC), *ssb* (CTGCCTCCTCGAGCAG) and *ruvAB* (GAGCTCAAAGGCGCAT) genes, and competitive EMSA experiments demonstrated that their promoters also bind *F. succinogenes* LexA (Fig. 5). Since these genes are under control of the LexA protein in many bacterial species, the possibility that the *F. succinogenes lexA* gene here identified was the result of the convergent evolution from a residual prophage repressor was definitively discarded.

Comparative analysis of the *F. succinogenes* LexA protein and its recognition sequence

A phylogenetic tree (Fig. 4) was constructed from a multiple alignment of available LexA protein sequences from relevant members of the Gram-positive and Cyanobacteria phyla and the Alpha, Beta and Gamma Proteobacteria classes, and

those of both *F. succinogenes* and *M. xanthus*. As expected, the resulting tree reveals that all these LexA proteins share a common ancestry, which all the available phylogenetic analyses (Gupta & Griffiths, 2002) situate in an ancestor of the Gram-positive bacteria. However, closer examination of the phylogenetic tree indicates that at least two divergent paths originated from the Gram-positive LexA protein: one leading to the Delta, Beta and Gamma Proteobacteria LexA with the *F. succinogenes* LexA as an intermediate step, and the other giving rise to the Cyanobacteria LexA and, unexpectedly, the Alpha Proteobacteria LexA.

To analyze whether the relationships between the different LexA proteins displayed in the phylogenetic tree were also reflected in their respective binding sites, a sequence comparison between the aforementioned LexA-binding sequences and that of *F. succinogenes* was carried out. This comparison reveals the presence of marked resemblances among several nucleotide positions (Fig. 6) that are consistent with a common phylogenetic origin. Moreover, and in accordance with the dual branching hypothesis prompted by LexA protein phylogeny, on close inspection these resemblances suggest again two putative evolutionary lanes emerging from the Gram-positive LexA box: one giving rise to the Cyanobacteria and Alpha Proteobacteria LexA box and the other leading to both the *F. succinogenes* and *M. xanthus* LexA boxes and, ultimately, resulting in the Beta and Gamma Proteobacteria LexA box.

Genesis of different LexA boxes through directed mutagenesis of the *F. succinogenes* LexA binding sequence

To further confirm the putative relationship between the LexA proteins described above, the vertical evolutionary path leading from Gram-positive bacteria to Gamma Proteobacteria was experimentally analyzed taking *F. succinogenes* LexA recognition

sequence as a starting point to generate, through directed mutagenesis, the LexA binding sequences of Gram-positive, *Myxococcus*, Beta and Gamma Proteobacteria. As it can be seen (Fig. 7a), the *B. subtilis* LexA protein is able to bind the *F. succinogenes* LexA box with the introduction of only five substitutions (the T as well as the two internal Cs of the TGCCC motif, plus, the internal G and A of the GTGCAT one), a number that, considering the evolutionary distance between both species, is remarkably low. Similarly, the *M. xanthus* LexA protein is able to bind a *F. succinogenes lexA*-derivative promoter in which only the flanking bases at both ends of the TGCCCAGTTGTGCA palindrome have been substituted for a C and a G, respectively, and the T of the internal GTGCAT motif has been replaced by a C. Finally, the *E. coli* LexA protein can effectively bind to the *lexA* promoter recognized by the *M. xanthus* LexA if only three additional changes to the mutant promoter are made: substitution of the CC duet for TA on the TGCC tetra-nucleotide and a change from T to A in the TTC tri-nucleotide. The fact that both these generated motifs are very close to experimentally validated LexA-binding motifs of *B. subtilis* and *E. coli* (Fig. 7XXX) indicates that a mutational transition similar to the one here proposed could certainly have taken place between the LexA-binding sequences of these species.

Derivation of the Alpha LexA binding sequence from the Cyanobacterial LexA box

To complete the above described analysis on the evolutionary relationship of LexA proteins through their binding sequences, a similar study was conducted to check the feasibility of the remaining branching line from Gram-positive bacteria (i.e. the one giving rise to Cyanobacteria and Alpha Proteobacteria LexA proteins). In

concordance with the hypothesis presented in Fig. 6, it was found that the simple addition of three nucleotides (chosen in accordance with the Alpha LexA-box consensus sequence) between the AGTAC and GTTC motifs of the Cyanobacterial LexA box was sufficient to enable the binding of the *R. sphaeroides* LexA protein to this mutant LexA box in the *Anabaena lexA* gene promoter (Fig. 7b). Furthermore, and although significant binding of the *R. sphaeroides* LexA protein to the *Anabaena lexA* promoter could be easily accomplished with the single insertion event described above, the introduction of an additional single-point mutation (substitution of T for A in the GTAC tetra-nucleotide) to the mutant *Anabaena lexA* promoter dramatically increased the recognition ability of the *R. sphaeroides* LexA repressor (Fig. 7b). Again, the fact that experimentally confirmed LexA-binding motifs closely resembling the here generated motifs are present in *R. sphaeroides***?palustris?** (Fig. 7XXX) gives further support to the evolutionary pathway here proposed.

DISCUSSION

In this work we have demonstrated that, through a programmed set of nucleotide changes, both the Gram-positive and *E. coli*-like LexA boxes can be obtained from the *F. succinogenes* LexA binding sequence. Furthermore, our results point out that the G and C corresponding to the most external positions of the GAACN4GTTC motif recognized by the Gram-positive LexA repressor are enclosed in the CTGT and ACAG sequences, respectively, found in the *E. coli*-like LexA box. In this way, the origin of the *E. coli* LexA recognition sequence (constituted by 16 nucleotides) could be explained by a 2-bp size increase of the Gram-positive LexA binding sequence (12 nucleotides long) through each one of its ends. Nevertheless, this extension of the LexA recognition motif does not seem to have carried a significant increase in the size

of the N-domain region of the LexA protein that contains the three α helices involved in DNA binding (Fig. 8). A straight comparison of the N-terminal domain of *F. succinogenes* and *M. xanthus* LexA protein sequences with the consensus sequences of this region of Gram-positive, Cyanobacteria and Alpha Proteobacteria LexA proteins reveals no amino acid insertions in those residues that, in *E. coli*, have been shown to participate directly in DNA binding activity, nor in their immediate neighbors (Fig. 8).

Moreover, this comparative analysis of LexA protein sequences shows several fully conserved residues amongst those that constitute the three predicted α helices that are involved in DNA-binding. This suggests that, since their respective LexA boxes are markedly different, these amino acids must be required for the maintenance of the overall DNA recognition complex instead of being used for specific binding. This is the case for T5, Q8, E10, P26, S39, L50, G54 and R64, following the numeric position in the *E. coli* LexA protein. Likewise, other residues present a low degree of substitutions that, besides, correspond to amino acids of the same family: L4, I15, E30, L47, K53, I56 and I66. This fact suggests that these residues must also be related to structural functions of the LexA HTH complex rather than to the specific recognition of the DNA binding sequence. It has been suggested that, in *E. coli*, the third α helix of the LexA HTH complex plays the leading role in specific DNA recognition (Knegtel *et al.*, 1995). However, other residues in the remaining α helices or in between must also play a significant part in specific DNA recognition, since a *F. succinogenes* LexA protein derivative in which the sequence of the third α helix has been replaced through directed mutagenesis with that of *E. coli* LexA can not bind the *E. coli*-like CTGTN8ACAG motif (data not shown).

Furthermore, we have also demonstrated that a functional Alpha Proteobacteria LexA binding sequence may be easily generated from the Cyanobacterial one through a single insertion event while, in turn, the Cyanobacterial LexA box derives directly from the Gram-positive one (Mazón *et al.*, 2004). The use of DNA recognition motifs in combination with other phylogenetic evidence has been proposed earlier as a measure of divergence to refine phylogenetic analyses and as a milestone to highlight branching points in evolution (Rodionov *et al.*, 2001; Rajewsky *et al.*, 2002; Erill *et al.*, 2003). Therefore, the experimental evidence of relatedness between Alpha and Cyanobacteria LexA boxes takes new relevance when combined with the fact that these two groups do also cluster together in the phylogenetic tree of LexA proteins (Fig. 4). This close relationship between Alpha Proteobacteria and Cyanobacteria is clearly at odds with the traditional positioning of the Alpha Proteobacteria class in the bacterial evolutionary tree, as prompted by RecA protein (Fig. 9; Eisen, 1995) and 16S rRNA and signature protein phylogenies (Woese *et al.*, 1984; Gupta & Griffiths, 2002), since these three phylogenetic analyses place the Alpha Proteobacteria very close to the Beta Proteobacteria and far removed from either Cyanobacteria or Gram-positive bacteria. The most feasible explanation for this combined divergence with conventional phylogenetic data is to suppose that, after their branching from other Proteobacteria classes, Alpha Proteobacteria lost their vertically-transmitted *lexA* gene, but incorporated later a novel *lexA* copy through lateral gene transfer (LGT) from either a Cyanobacterium or a bacterial species closely related to this Phylum. This LGT addition, however, must have occurred very early in the evolutionary history of the Alpha Proteobacteria, since the same protein is present in all Alpha Proteobacteria that have not suffered major reductions in chromosome size (e.g. *Rickettsia*), and GC percentage and codon usage of the extant *lexA* genes are in

perfect agreement with the average values for each of the Alpha Proteobacteria hosting them. In this context, it should be stressed that the loss of the *lexA* gene does not seem to be a very unusual event in bacterial evolution, as it has already been described in several genera (such as *Aquifex, Borrelia, Campylobacter, Clamydia, Helicobacter, Mycoplasma* or *Rickketsia*). Up to now, a common characteristic of those bacteria for which the lack of a *lexA* gene had been described was that they had undergone a major reduction in chromosome size, suggesting that massive genome reduction was a convergent evolutionary cause for the loss of the *lexA* gene. However, and given that Alpha Proteobacteria species here analyzed do not present significant reductions in genetic material, our data concerning their LexA protein breaks with this traditional assumption and hints at the possible existence of losses and lateral acquisitions of the *lexA* gene among bacteria. Although further work is still necessary to elucidate whether similar LGT processes have taken place in other Bacterial Phyla, the reported evidence of lateral transfer of the *lexA* gene sheds new light on the evolutionary history of a complex regulatory network like the LexA-governed SOS response and validates the previously reported use of regulatory motifs, in combination with phylogenetic and protein signature studies, as reliable indicators of phylogenetic history.

ACKNOWLEDGEMENTS

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, and by the Consejo Superior de Investigaciones Científicas (CSIC). We are deeply indebted to Dr. Roger Woodgate for his generous gifts of *E. coli* and *B.*

subtilis LexA proteins. We wish to acknowledge Joan Ruiz for his excellent technical assistance and collaboration. The free access of The Institute for Genomic Research (TIGR) to the *F. succinogenes* preliminary sequence data is also acknowledged. Partial sequencing of *F. succinogenes* was accomplished with support from U.S. Department of Agriculture.

REFERENCES

Bradford, M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72,** 248 – 254.

Burckhardt, S. E., Woodgate, R., Scheuermann, H.R., & Echols, H. (1988). UmuD mutagenesis protein of *Escherichia coli*: overproduction, purification and cleavage by RecA. *Proc Natl Acad Sci* USA **85,** 1811 - 1815

Campoy, S., Fontes, M., Padmanabhan, S., Cortes, P., Llagostera, M. & Barbe, J. (2003). LexA-independent DNA damage-mediated induction of gene expression in *Myxococcus xanthus*. *Mol Microbiol* **49,** 769 - 781.

Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. (2000). NPS@: Network Protein Sequence Analysis. *Trends Biochem Sci* **25,** 147 – 150.

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

Davis, E.O., Dullaghan, E.M. & Rand, L. (2002). Definition of the Mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis. J Bacteriol* **184,** 3287 – 3295.

Eisen, J.A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* **41,** 1105 - 1123.

Erill, I., Escribano, M., Campoy, S. & Barbé, J. (2003). *In silico* analysis reveals substantial variability in the gene contents of the Gamma Proteobacteria LexA-regulon. *Bioinformatics* **19**, 2225 - 2236.

Felsenstein, J. (1989). PHYLIP: phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.

Fernández de Henestrosa, A.R., Rivera, E., Tapias, A. & Barbé, J. (1998). Identification of the *Rhodobacter sphaeroides* SOS box. *Mol Microbiol* **28,** 991 – 1003.

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

Fernández de Henestrosa, A.R., Cuñé, J., Erill, I., Magnuson, J.K. & Barbé, J. (2002). A green nonsulfur bacterium, *Dehalococcoides ethenogenes*, with the LexA binding sequence found in gram-positive organisms. *J Bacteriol* **184**, 6073 – 6080.

Fogh, R.H., Ottleben, G., Rüterjans, H., Schnarr, M., Boelens, R. & Kaptein, R. (1994). Solution structure of the LexA repressor DNA binding domain determined by ^{1}H NMR spectroscopy. *EMBO J* **13,** 3936 – 3944.

Forano, E., Broussolle, V., Gaudet, G. & Bryant, J.A. (1994). Molecular cloning, expression and characterization of a new endoglucanase gene from *Fibrobacter succinogenes* S85. *Current Microbiol* **28,** 7 – 14.

Gaudet, G., Forano, E., Dauphin, G. & Delort, A.M. (1992). Futile cycling of glycogen in *Fibrobacter succinogenes* as shown by in situ 1H-NMR and 13C-NMR investigation. *Eur J Biochem* **207,** 155 – 162.

Griffiths, E. & Gupta, R.S. (2001). The use of signature sequences in different proteins to determine the relative branching order of bacterial divisions: evidence that *Fibrobacter* diverged at a similar time to *Chlamydia* and the *Cytophaga-Flavobacterium-Bacteroides* division. *Microbiology* **147,** 2611 – 2622.

Gupta, R.S. & Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor Pop Biol* **61,** 423-434.

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22,** 4673-4680.

Knegtel, R.M.A., Fogh, R.H., Ottleben, G., Rüterjans, H., Dumoulin, P., Schnarr, M., Boelens, R. & Kaptein, R. (1995). A model for the LexA repressor DNA complex. *Proteins* **21,** 226 – 236.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680 – 685.

Little, J. W. (1984). Autodigestion of *lexA* and phage lambda repressors. *Proc Natl Acad Sci USA* **81,** 1375-1379

Little, J.W. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* **73,** 411 – 422.

Little, J.W., Mount, D. & Yanish-Perron, C.R. (1981). Purified LexA protein is a repressor of the *recA* and *lexA* genes. *Proc Natl Acad Sci* USA **78,** 4199 – 4203.

Luo, Y., Pfuetzner, R.A., Mosimann, S., Paetzel, M., Frey, E.A., Cherney, M., Kim, B., Little, J.W. & Strynadka, C.J. (2001). Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell* **106,** 585 – 594.

Ludwig, W. & Schleifer, K.H. (1999). Phylogeny of *Bacteria* beyond the 16S rRNA Standard. *ASM News* **65,** 752 – 757.

Maidak, B.L., Cole, J.R., Parker, C.T., Jr and 11 other authors. (1999). A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res* **27,** 171 – 173.

Mazon, G., Lucena, J.M., Campoy, S., Fernández de Henestrosa, A.R., Candau, P. & Barbé J. (2004). LexA-binding sequences in Gram-positive and cyanobacteria are closely related. *Mol Gen Genomics* **271,** 40 - 49.

Nohmi, T., Battista, J.R., Dodson, L.A. & Walker, G.C. (1988). RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation. *Proc Natl Acad Sci* USA **85,** 1816-1820

Norioka, N., Hsu, M.Y., Inouye, S. & Inouye, M. (1995). Two *recA* genes in *Myxococcus xanthus. J Bacteriol* **177,** 4179 – 4182.

Page, R.D.M. (1996). TreeView: an application to display phylogenetic trees on personal computers. *CABIOS* **12**, 357-358.

Patzer, S.I. & Hantke, K. (2001). Dual repression by Fe^{2+}-Fur and Mn^{2+}-MntR of the *mntH* gene, encoding an NRAMP-like Mn^{2+} transporter in *Escherichia coli. J Bacteriol* **183,** 4806 – 4813.

Rajewsky, N., Socci, N., Zapotocky, M. & Siggia, E.D. (2002). The evolution of DNA regulatory regions for proteo-gamma bacteria by interspecies comparisons. *Genome Res* **12,** 298 – 308.

Rodionov, D.A., Mironov, A.A. & Gelfand, M.S. (2001). Transcriptional regulation of pentose utilisation systems in the *Bacillus/Clostridium* group of bacteria. *FEMS Microbiol Lett* **205,** 305 – 314.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1992). *Molecular Cloning. A Laboratory Manual,* 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, S. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74,** 5463 – 5467.

Sassanfar, M. & Roberts, J.W. (1990). Nature of SOS-inducing signal in *Escherichia coli.* The involvement of DNA replication. *J Mol Biol* **212,** 79 – 96.

Tapias, A. & Barbé, J. (1999). Regulation of divergent transcription from the *uvrA-ssb* promoters in *Sinorhizobium meliloti. Mol Gen Genet* **262,** 121 – 130.

Tapias, A., Fernández, S., Alonso, J.C. & Barbé, J. (2002). *Rhodobacter sphaeroides* LexA has dual activity: optimising and repressing *recA* gene transcription. *Nucleic Acids Res* **30,** 1539 – 1546.

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

Winterling, K.W., Chafin, D., Hayes, J.J., Sun, J., Levine, A.S., Yasbin, R.E. & Woodgate, R. (1998). The *Bacillus subtilis* DinR binding site: redefinition of the consensus sequence. *J Bacteriol* **180,** 2201 - 2211.

Woese, C.R., Stackebrandt, E., Weisburg, W.C., Paster, B.J., Madigan, M.T., Fowler, V.J., Hahn, C.M., Blanz, P., Gupta, R., Nealson, K.H. & Fox, G.E. (1984). The phylogeny of purple bacteria: the alpha subdivision. *Syst Appl Microbiol* **5,** 315 – 326.

LEGENDS OF FIGURES

Fig. 1. (a) Electrophoretic mobility of the DNA fragment containing the *F. succinogenes lexA* promoter in presence of increasing concentrations of purified *F. succinogenes* LexA protein. (b) Effect of 300-fold molar unlabelled *F. succinogenes lexA* promoter (lane 3) and pBSK(+) plasmid DNA (lane 4) on the migration of the *F. succinogenes lexA* promoter in the presence of its purified LexA protein (at 40 nM). The migration of this same fragment without any additional DNA (+) is also shown (lane 2). In both panels, the mobility of the *F. succinogenes lexA* promoter in the absence (-) of purified LexA protein is also presented as a negative control (lane 1).

Fig. 2. DNase I footprinting assays with coding and non-coding Cy5-labelled strands of the DNA fragment containing the *F. succinogenes lexA* promoter in the absence or presence of increasing amounts of purified LexA protein from this same organism. The arrows indicate the translational direction of each strand; for these, the translational starting codon is shown in bold and underlined.

Fig. 3. Single-nucleotide substitutions in the TTGCCCAGTTGTGCAT imperfect palindrome and their effect on the electrophoretic mobility of the *F. succinogenes lexA* promoter in the presence of purified *F. succinogenes* LexA protein (at 40 nM). The mobility of the wild-type *F. succinogenes lexA* promoter in the absence (-) or presence (+) of LexA from this same organism is also shown.

Fig. 5. Electrophoretic mobility of the wild-type *F. succinogenes lexA* promoter in the presence of its LexA protein (40 nM) and a 300-fold molar excess of unlabelled fragments containing about 400 bp of the upstream region of the *F. succinogenes recA, uvrA, ruvAB* and *ssb* genes. As a control, the effect on *lexA* promoter mobility upon the addition of unlabelled *lexA* and *trp* promoters is displayed in the presence of the same amount of LexA protein. The mobility of the *lexA* promoter in the absence of any additional DNA (+) or purified LexA protein (-) is also shown.

Fig. 4. Phylogenetic tree of the LexA protein-sequence. Name abbreviations are as follows: Gram-positive: *B. subtilis* (Bsu), *Clostridium perfringens* (Cpe), *Mycobacterium tuberculosis* (Mtu), *Staphylococcus aureus* (Sva), *Streptomyces coelicolor* (Sco); Cyanobacteria: *Anabaena* (Ana), *Prochlorococcus marinus* (Pmi), *Synechocystis* (Syn); *F. succinogenes* (Fbs); Delta Proteobacteria: *M. xanthus* (Myc); Alpha Proteobacteria: *A. tumefaciens* (Atc), *Bradyrhizobium japonicum* (Bja), *Brucella melitensis* (Bme), *B. suis* (Brs), *Caulobacter crescentus* (Ccr), *Mesorhizobium loti* (Mlo), *Rhodopseudomonas palustris* (Rpa), *S. meliloti* (Sme); Beta Proteobacteria: *Bordetella pertussis* (Bpe), *R. solanacearum* (Rso); Gamma Proteobacteria: *E. coli* (Eco), *H. influenzae* (Hin), *Shewanella oneidensis* (Son), *Vibrio cholerae* (Vch), *V. parahaemolyticus* (Vpa), *V. vulnificus* (Vvu), *Y. pestis* (Ype). Numbers at branch nodes indicate bootstrapping values for 100 bootstrap replicates.

Fig. 6. Schematic diagram representing the similarities between LexA recognition sites of different bacterial clades and the possible generation of several LexA boxes following the two apparent evolutionary lanes that emerge from Gram-positive bacteria (see Fig. 4). Bases belonging to the palindromic motif of the Gram-positive LexA box that are conserved through the evolutionary history of the LexA reconition sequence are marked in shadow. Changes to the LexA binding sequence are highlighted in bold at the step in which they were introduced.

Fig. 7. (a) Binding ability of *B. subtilis*, *F. succinogenes*, *M. xanthus* and *E. coli* LexA proteins to the *F. succinogenes lexA* wild type promoter (Fbs Wt) and several mutant derivatives. (b) Binding ability of *Anabena* and *R. sphaeroides* LexA proteins to the *Anabaena lexA* wild type promoter (Ana Wt) and several mutant derivatives. All changes were introduced through directed mutagenesis according to the comparative schematic diagram of LexA boxes show in Fig. 6. In all cases, (-), ($\pm$) or (+) denote, respectively: no LexA binding, LexA binding with a percentage of bound probe lower than 25% and LexA biding with a percentage of bound probe higher than that of 25%. Bases of *F. succinogenes* and *Anabaena* LexA boxes that are required for binding of their own LexA protein are overlined in each panel. For each mutagenesis step the bases either modified or added are shown in bold, and the change is indicated with an

arrow. Likewise, changes introduced in a previous step remain underlined in subsequent steps.

Fig. 8. CLUSTALW alignment of the N-terminal region of the LexA protein containing the three α helices involved in its DNA binding function. Aligned sequences correspond to the LexA protein of *M. xanthus*, *F. succinogenes* and the consensus sequence of this region for Gram-positive bacteria, Cyanobacteria and Alpha and Gamma Proteobacteria. For each LexA protein, the amino acids comprised in its predicted α helices (α1, α2 or α3) by using the NPS@ software (Combet *et al.*, 2000) are underlined. Residues of the *E. coli* LexA protein the importance of which on DNA binding has been experimentally demonstrated are overlined.

Fig. 9. Phylogenetic tree of the RecA protein sequence. Name abbreviations follow the same convention as those in Fig. 6. Myx1 and Myx2 refer to the products of the two independent copies of the *recA* gene present in *M. xanthus* (Norioka *et al.,* 1995). Numbers at branch nodes indicate bootstrapping values for 100 bootstrap replicates.

FIGURES

FIG. 1

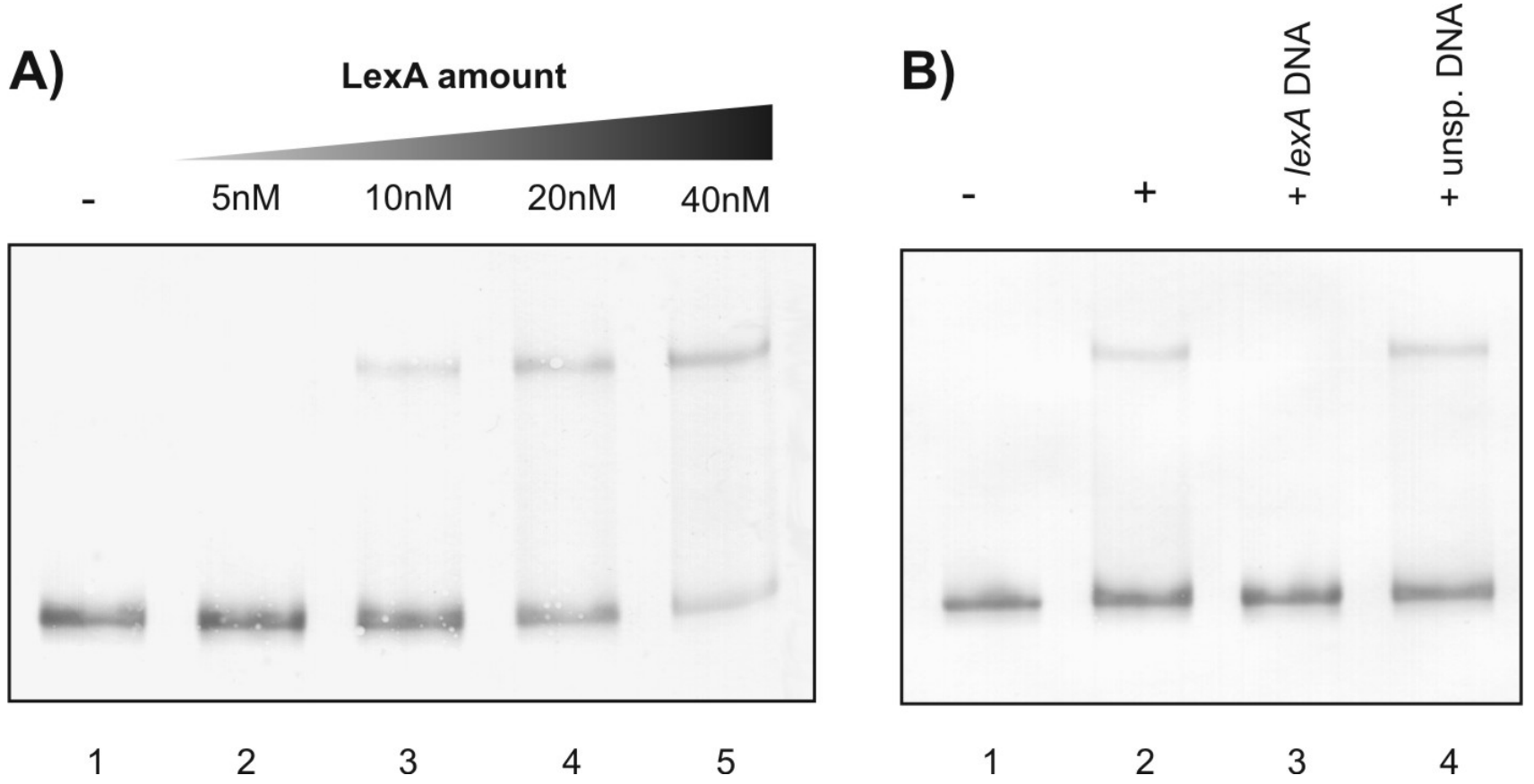

A)
LexA amount
-
5nM
10nM
20nM
40nM
1
2
3
4
5
B)
-
+
+ lexA DNA
+ unsp. DNA
1
2
3
4

FIG. 2

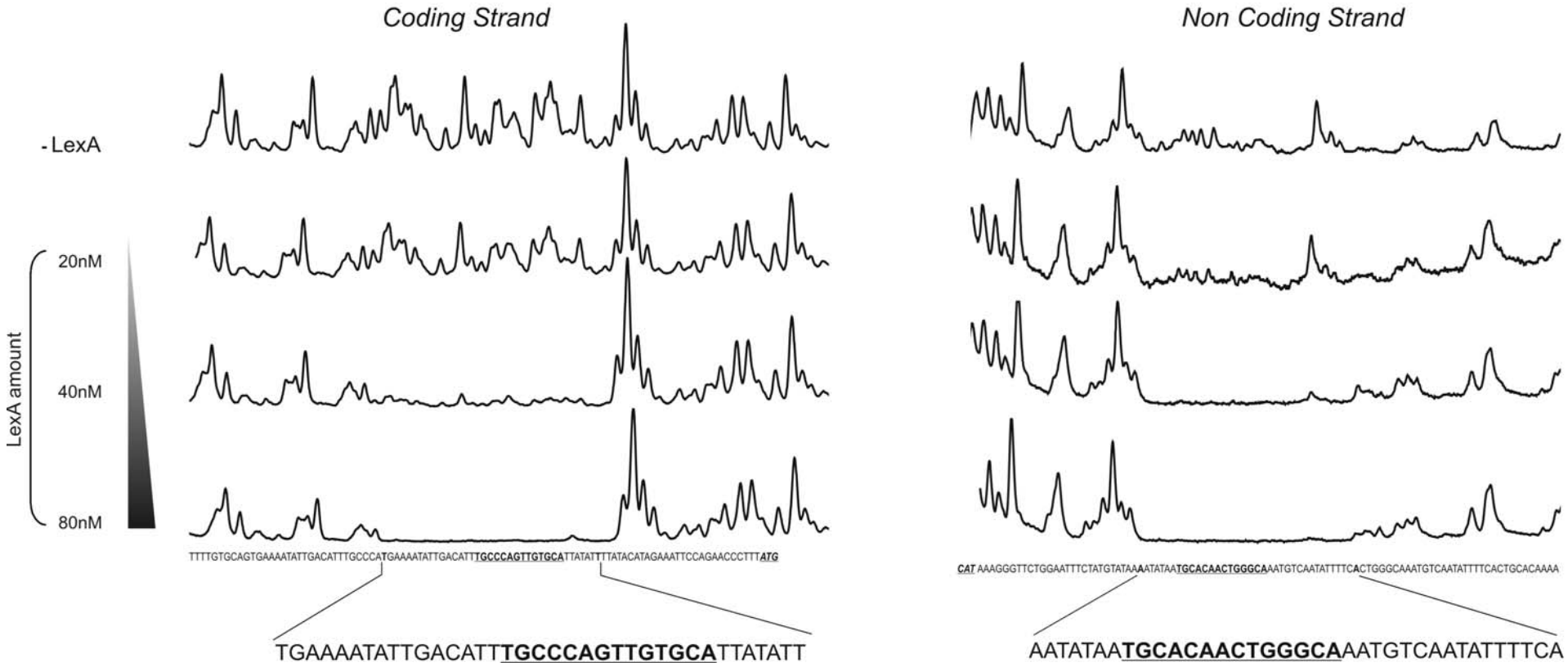

Coding Strand
Non Coding Strand
- LexA
20nM
40nM
80nM
LexA amount
TTTTGTGCAGTGAAAATATTGACATTTGCCCATGAAAATATTGACATTTGCCCAGTTGTGCATTATATTTTATACATAGAAATTCCAGAACCCTTTATG
CAT AAAGGGTTCTGGAATTTCTATGTATAAAATATAATGCACAACTGGGCAATGTCAATATTTTCACTGGGCAAATGTCAATATTTTCACTGCACAAAA
TGAAAATATTGACATTTGCCCAGTTGTGCATTATATT
AATATAATGCACAACTGGGCAAATGTCAATATTTTCA

FIG. 3

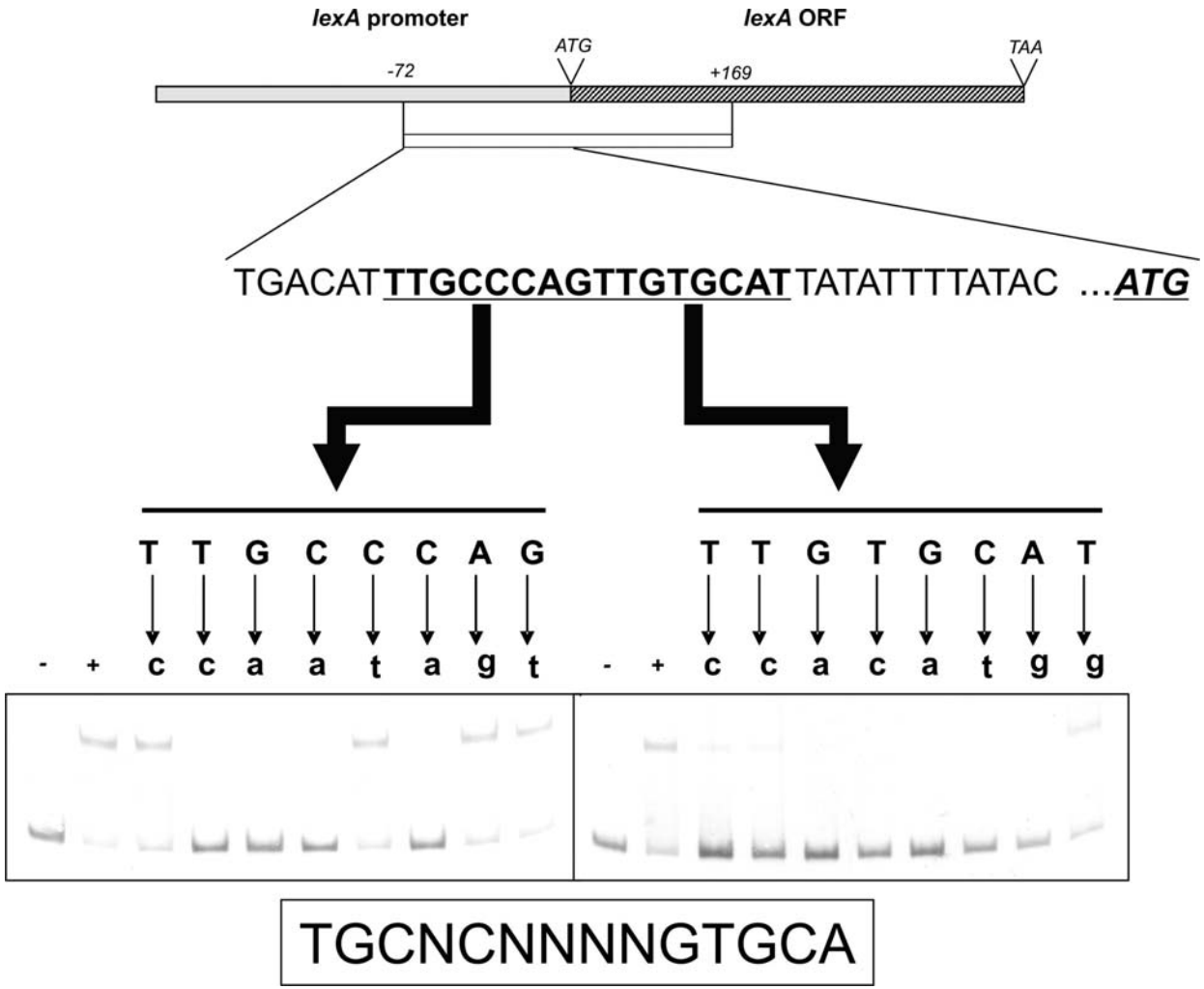

lexA promoter
lexA ORF
ATG
TAA
-72
+169
TGACATTTGCCCAGTTGTGCATTATATTTTATAC ...ATG
T T G C C C A G
- + c c a a t a g t
T T G T G C A T
- + c c a c a t g g
TGCNCNNNNGTGCA

FIG. 5

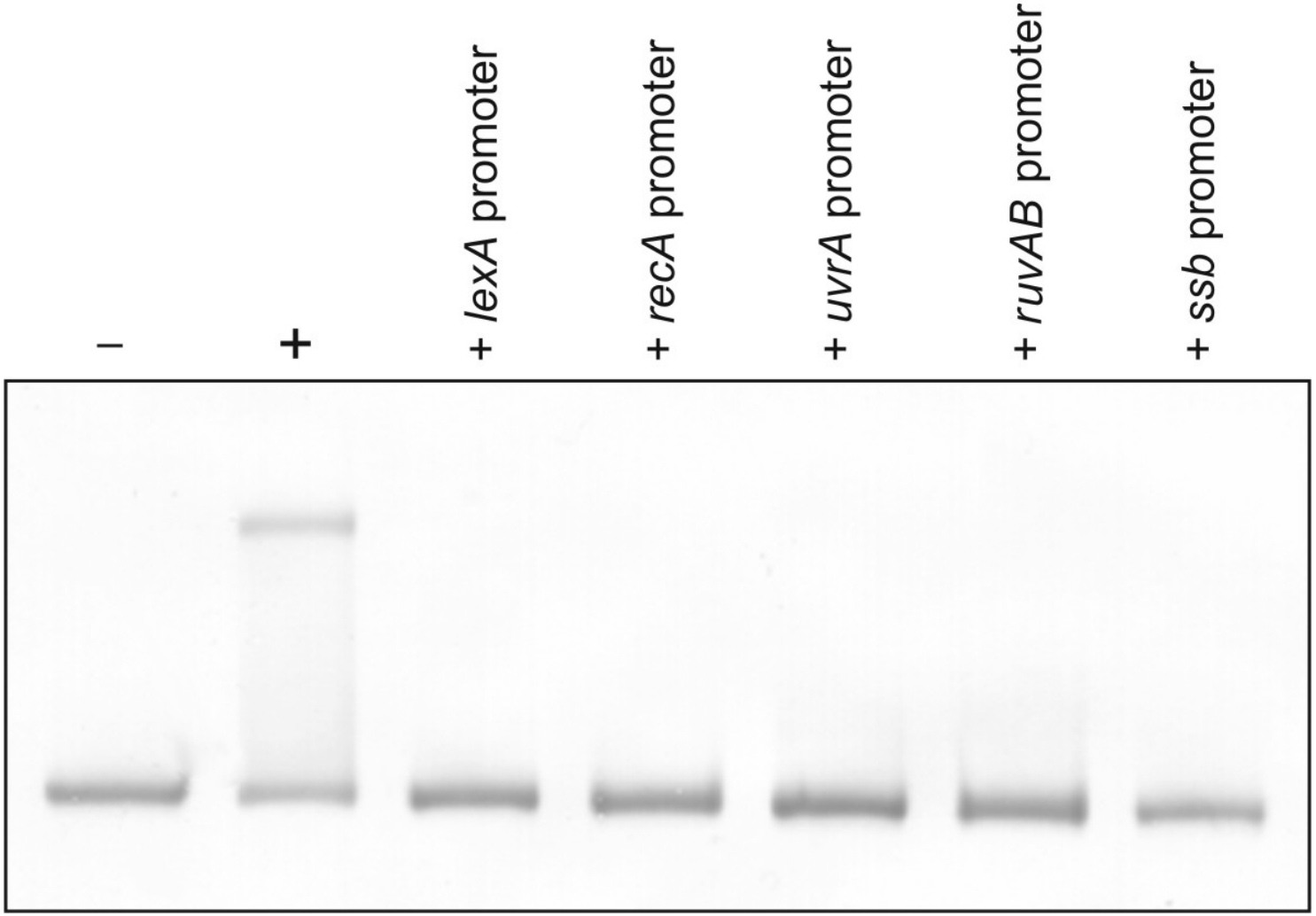

–
+
+ lexA promoter
+ recA promoter
+ uvrA promoter
+ ruvAB promoter
+ ssb promoter

FIG. 6

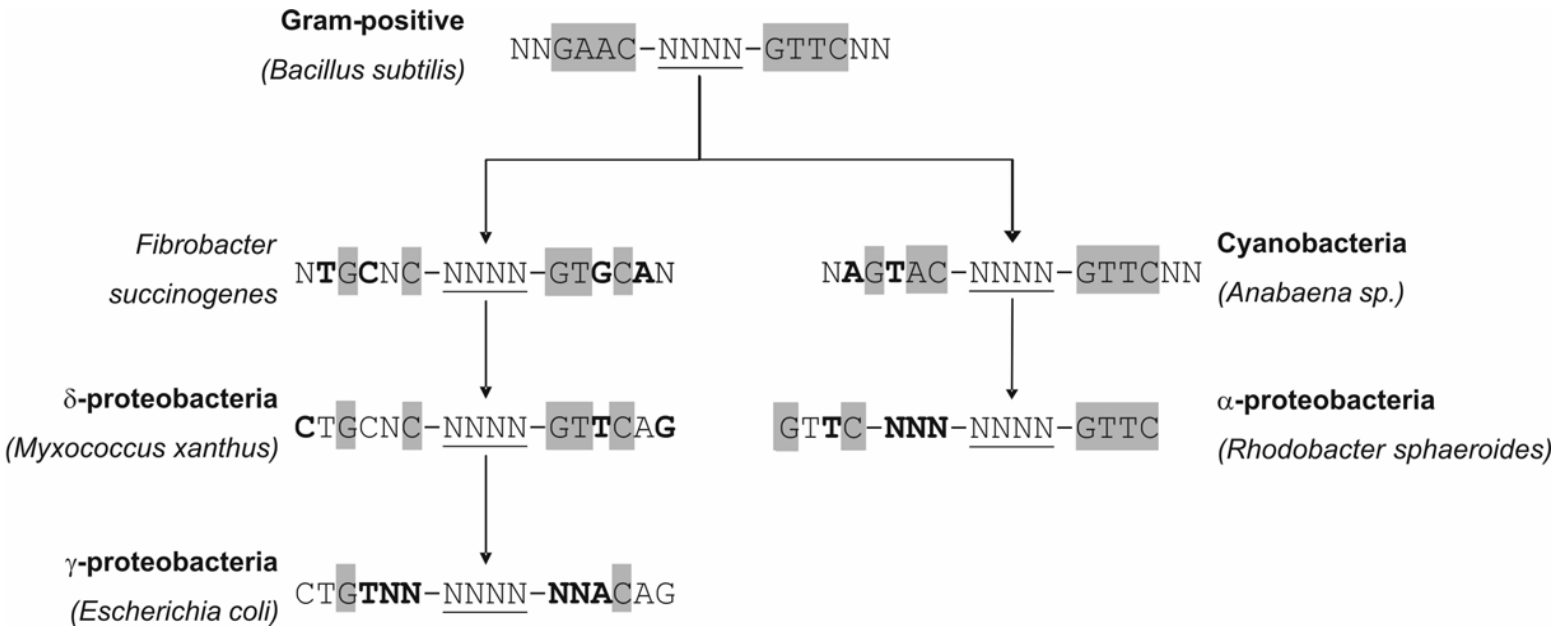

Gram-positive
(Bacillus subtilis)
NNGAAC-NNNN-GTTCNN
Fibrobacter succinogenes
NTGCNC-NNNN-GTGCAN
Cyanobacteria
(Anabaena sp.)
NAGTAC-NNNN-GTTCNN
δ-proteobacteria
(Myxococcus xanthus)
CTGCNC-NNNN-GTTCAG
α-proteobacteria
(Rhodobacter sphaeroides)
GTTC-NNN-NNNN-GTTC
γ-proteobacteria
(Escherichia coli)
CTGTNN-NNNN-NNACAG

FIG. 4

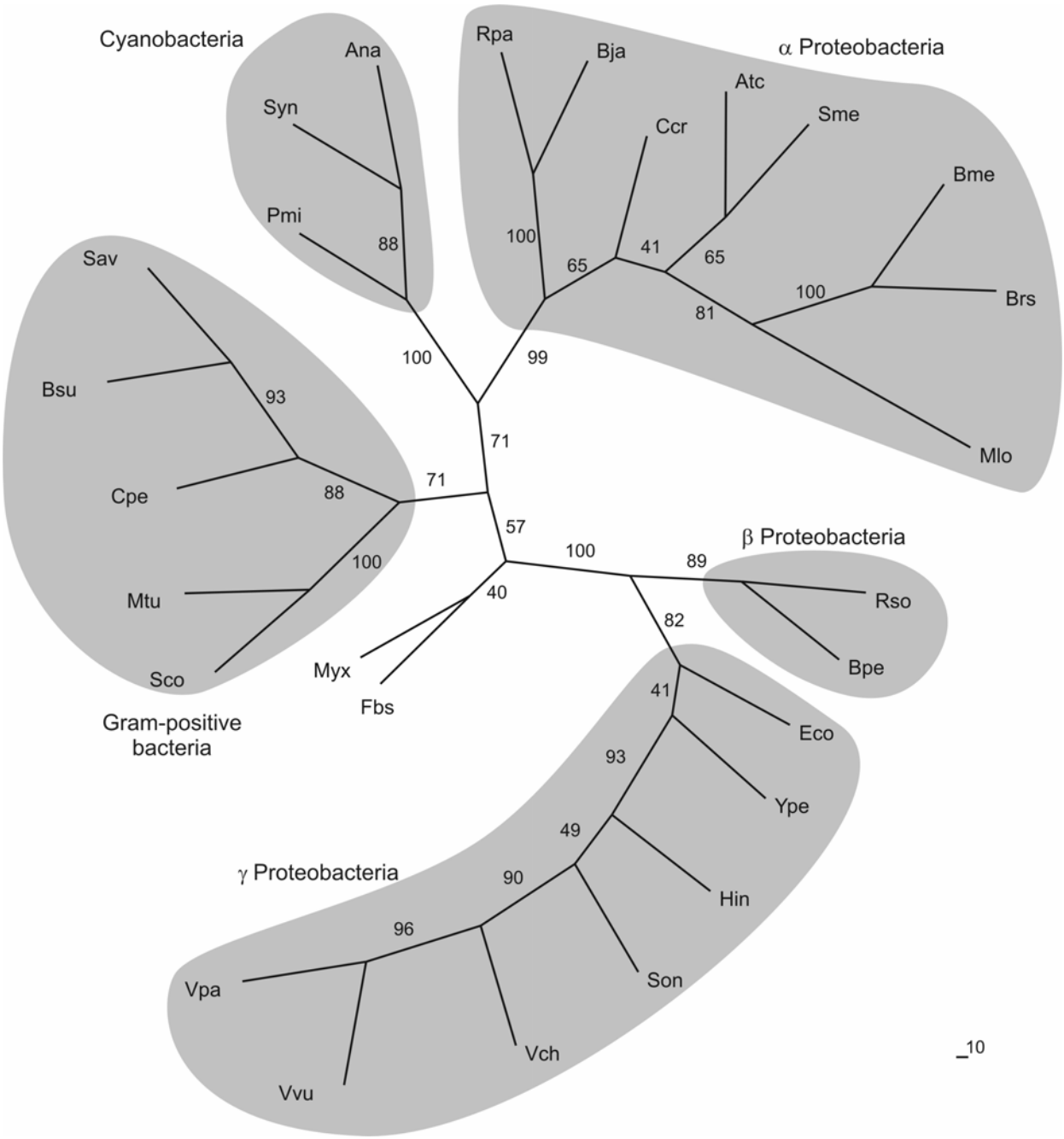
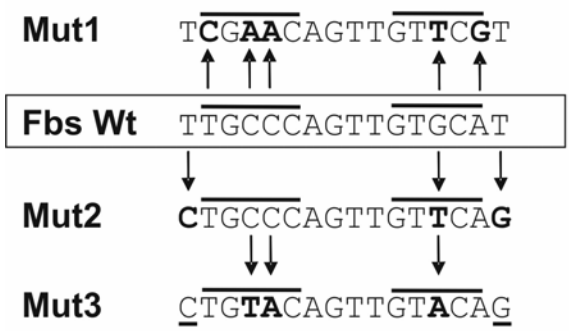

Cyanobacteria
Ana
Syn
Pmi
88
Rpa
Bja
α Proteobacteria
Atc
Sme
Ccr
Bme
Brs
Mlo
100
65
41
65
100
81
100
99
71
Sav
Bsu
93
Cpe
88
71
57
100
89
β Proteobacteria
Rso
Bpe
Mtu
100
40
82
Myx
Fbs
Sco
Gram-positive
bacteria
41
Eco
93
Ype
49
γ Proteobacteria
90
Hin
96
Son
Vpa
Vch
Vvu
0.10

FIG. 7

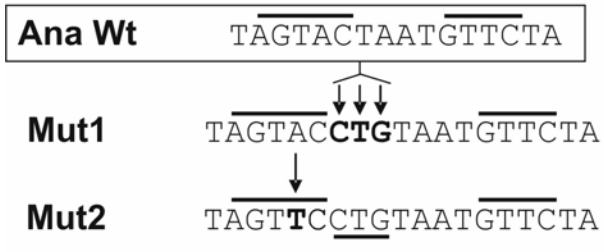
A)



LexA protein binding ability

B. subtilis	*F. succinogenes*	*M. xanthus*	*E. coli*
+	-	-	-
-	+	-	-
-	±	+	-
-	-	-	+

B)



LexA protein binding ability

Anabaena	*R. sphaeroides*
+	-
-	±
-	+

FIG. 8

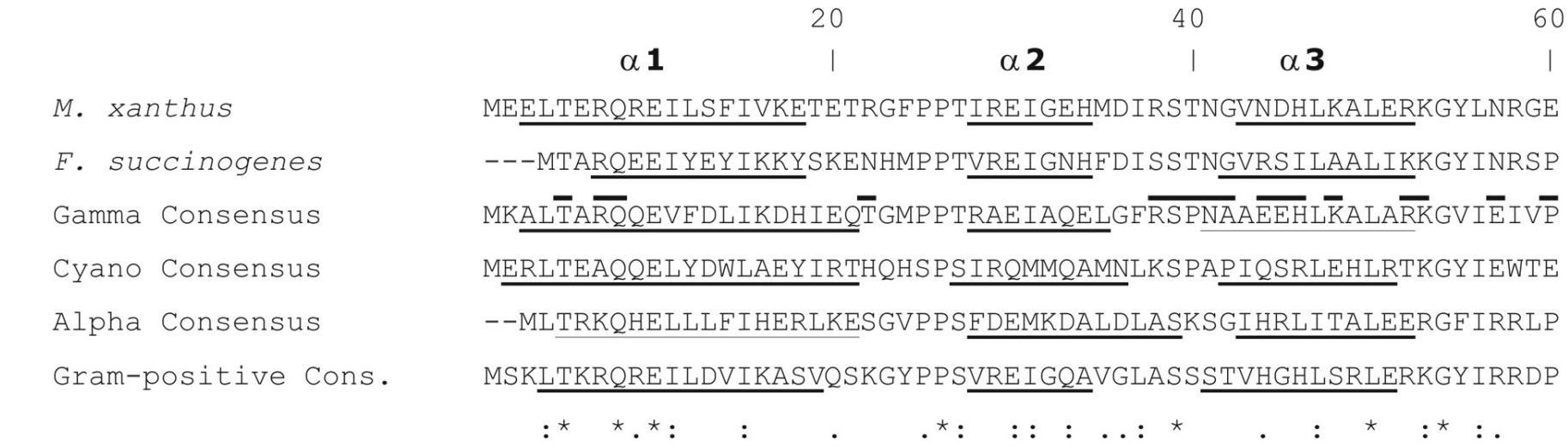
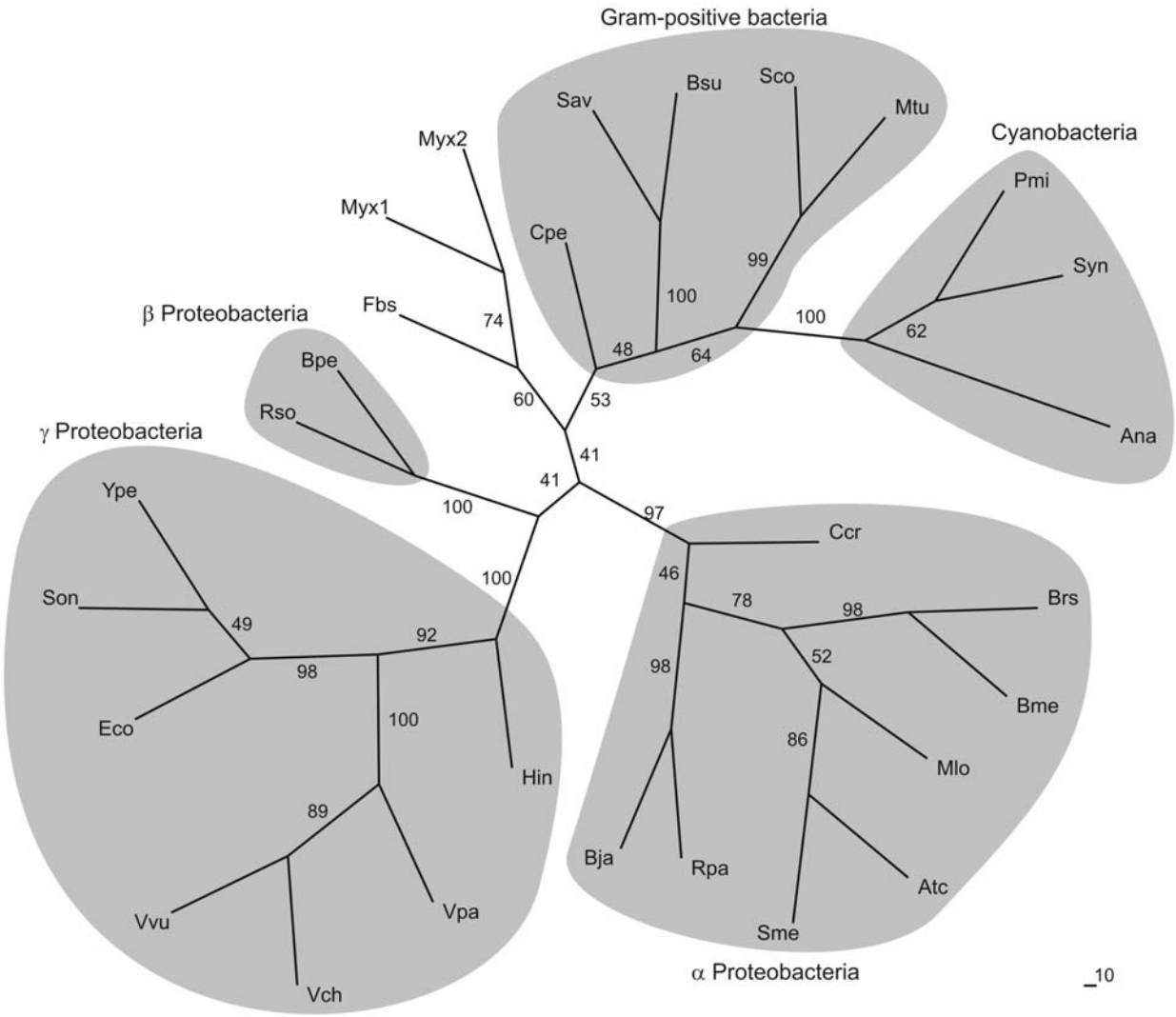

20
40
60
α1
α2
α3
M. xanthus MEELTERQREILSFIVKETETRGFPPTIREIGEHMDIRSTNGVNDHLKALERKGYLNRGE
F. succinogenes ---MTARQEEIYEYIKKYSKENHMPPTVREIGNHFDISSTNGVRSILAALIKKGYINRSP
Gamma Consensus MKALTARQQEVFDLIKDHIEQTGMPPTRAEIAQELGFRSPNAAEEHLKALARKGVIEIVP
Cyano Consensus MERLTEAQQELYDWLAEYIRTHQHSPSIRQMMQAMNLKSPAPIQSRLEHLRTKGYIEWTE
Alpha Consensus --MLTRKQHELLLFIHERLKESGVPPSFDEMKDALDLASKSGIHRLITALEERGFIRRLP
Gram-positive Cons. MSKLTKRQREILDVIKASVQSKGYPPSVREIGQAVGLASSSTVHGHLSRLERKGYIRRDP

FIG. 9


Gram-positive bacteria
Sav
Bsu
Sco
Mtu
Myx2
Myx1
Cpe
Cyanobacteria
Pmi
Syn
Ana
β Proteobacteria
Fbs
Bpe
Rso
γ Proteobacteria
Ype
Son
Eco
Hin
Vvu
Vpa
Vch
Ccr
Brs
Bme
Mlo
Bja
Rpa
Atc
Sme
α Proteobacteria
99
100
74
48
64
62
60
53
41
41
100
97
100
46
78
98
52
98
92
98
100
89
86
10