ANNEXOS

ARTICLE I

Microbiology (2002), **148**, 3583–3597 Printed in Great Britain

A new regulatory DNA motif of the gamma subclass *Proteobacteria*: identification of the LexA protein binding site of the plant pathogen *Xylella fastidiosa*

Susana Campoy,[1] Gerard Mazón,[1] Antonio R. Fernández de Henestrosa,[1] Montserrat Llagostera,[1,3] Patricia Brant Monteiro[2] and Jordi Barbé[1,3]

Author for correspondence: Jordi Barbé. Tel: +34 93 581 1837. Fax: +34 93 581 2387. e-mail: jordi.barbe@uab.es

[1] Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

[2] Fundo de Defesa da Citricultura (Fundecitrus), 14807-040, VI. Melhado-C. P. 391, Araraquara, Sao Paulo, Brazil

[3] Centre de Recerca en Sanitat Animal (CReSA), Universitat Autònoma de Barcelona-Institut de Recerca i Tecnologia Agroalimentària (UAB-IRTA), Bellaterra, 08193 Barcelona, Spain

***Escherichia coli* LexA protein is the repressor of a gene network whose members are directly involved in the repair of damaged DNA and in the survival of bacterial cells until DNA lesions have been eliminated. The *lexA* gene is widely present in bacteria, although the sequences of only three LexA-binding sites are known: Gram-positive, alpha *Proteobacteria* and some members of gamma *Proteobacteria* represented by *E. coli*. Taking advantage of the fact that the genome sequence of the plant-pathogenic bacterium *Xylella fastidiosa* has been determined, its *lexA* gene has been cloned and overexpressed in *E. coli* to purify its product. After demonstration that *X. fastidiosa lexA* and *recA* genes are co-transcribed, gel mobility shift assays and directed mutagenesis experiments using the promoter of the *lexA–recA* transcriptional unit demonstrated that the *X. fastidiosa* LexA protein specifically binds the imperfect palindrome TTAGN$_6$TACTA. This is the first LexA binding sequence identified in the gamma *Proteobacteria* differing from the *E. coli*-like LexA box. Although a computational search has revealed the presence of TTAGN$_6$TACTA-like motifs upstream of *X. fastidiosa* genes other than *lexA*, *X. fastidiosa* LexA only binds the promoter of one of them, XF2313, encoding a putative DNA-modification methylase. Moreover, *X. fastidiosa* LexA protein does not bind any of the other genes whose homologues are regulated by the LexA repressor in *E. coli* (*uvrA*, *uvrB*, *ssb*, *ruvAB*, *ftsK*, *dinG*, *recN* and *ybfE*). RT-PCR quantitative analysis has also demonstrated that *lexA–recA* and XF2313 genes, as well as the *X. fastidiosa* genes which are homologues to those of *E. coli* belonging to the LexA regulon, with the exception of *ssb*, are DNA damage-inducible in *X. fastidiosa*.**

Keywords: DNA damage, gene expression, SOS system

INTRODUCTION

Bacteria have different pathways to repair DNA damage, the SOS system being one of the most important (Walker, 1984). In *Escherichia coli*, the SOS response has been widely studied, with at least 40 genes constituting this regulon (Fernandez de Henestrosa *et al.*, 2000; Courcelle *et al.*, 2001). The *recA* and *lexA* gene products control the expression of the *E. coli* SOS network to which both belong, LexA protein being the negative regulator (Walker, 1984). It has been demonstrated that the RecA protein is converted into an active conformation after binding to single-stranded DNA regions generated by DNA damage-mediated inhibition of replication or by enzymic processing of broken DNA ends (Sassanfar & Roberts, 1990). After activation, RecA promotes autocatalytic cleavage of LexA, inducing the SOS genes (Walker, 1984). *E. coli* LexA cleavage occurs in the Ala$_{84}$-Gly$_{85}$ bond and is mediated by its Ser$_{119}$ and Lys$_{156}$ residues by a similar mechanism to that of serine proteases (Little, 1991; Luo *et al.*, 2001). Once DNA damage has been repaired, RecA is no longer activated and the LexA protein level increases again,

Abbreviation: DIG, digoxigenin.

inhibiting expression of SOS genes. Recently, it has been suggested that the product of *dinI* is involved in the arrest of RecA protein activation (Voloshin *et al*., 2001). Likewise, it has also been shown that some *E. coli* genes are DNA damage-inducible in a *lexA*-independent pathway, suggesting that other regulatory pathways could be involved in the *E. coli* DNA damage mediated response (Courcelle *et al*., 2001; Khil & Camerini-Otero, 2002).

The *E. coli* LexA repressor specifically binds regions placed upstream of SOS genes. These LexA binding sites are imperfect 16-bp palindromes whose consensus sequence is $CTGTN_8ACAG$ (Walker, 1984), which has been designated as the *E. coli* SOS box. The presence of a similar LexA regulon has been described in other bacterial groups, although the LexA-binding sequence is not conserved. Thus, the imperfect palindrome CGAA-CRNRYGTTYC and the $GTTCN_7GTTC$ direct repeat are the LexA boxes of Gram-positive bacteria and the alpha subclass *Proteobacteria*, respectively (Winterling *et al*., 1998; Fernandez de Henestrosa *et al*., 1998; Tapias & Barbé, 1999). While the *recA* gene has been found in all sequenced eubacterial genomes, *lexA* appears to be absent in some of them. In this way, a *lexA*-like gene does not seem to be present in *Aquifex aeolicus*, *Borrelia burgdorferi*, *Chlamydia pneumoniae*, *Mycoplasma pneumoniae*, *Campylobacter jejuni*, *Helicobacter pylori* and *Porphyromonas gingivalis*. Moreover, and with the exception of the plant-pathogenic bacterium *Xylella fastidiosa*, all members of the gamma subclass *Proteobacteria* whose genome sequences have been published (*E. coli*, *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Pasteurella multocida* and *Vibrio cholerae*) present an *E. coli*-like LexA box upstream of the promoter genes belonging to the SOS system. Nevertheless, and despite the absence of this *E. coli*-like LexA box, a *lexA* homologue gene is present in the genome of *X. fastidiosa* (Simpson *et al*., 2000). This fact indicates that, unlike what happens in both Gram-positive and alpha subclass *Proteobacteria* (Winterling *et al*., 1998; Tapias & Barbé, 1999), a significant heterogeneity exists in the very compact phylogenetic gamma subclass, with respect to the sequence of the control region of the common LexA regulon.

Differences among the LexA networks of several bacteria seems to be not only limited to the LexA box, but also to the regulation mechanism. So, whereas the *E. coli* LexA represses gene transcription by precluding RNA polymerase binding to the promoter region (Brent & Ptashne, 1981; Little *et al*., 1981; Bertrand-Burggraf *et al*., 1987), *Rhodobacter sphaeroides* LexA interferes with the clearance process besides also being able to act as a transcriptional activator (Tapias *et al*., 2002). All together indicate that it is not suitable to postulate a general model to explain how the LexA regulon functions in the Domain *Bacteria*, but this aspect must be studied in each one of their several phylogenetic groups. Likewise, the number of evolutionary analyses of regulatory sequences as well as of the constitution of different gene networks in the several bacterial phyla is increasing (Eisen & Hanawalt, 1999; Tan *et al*., 2001; Makarova *et al*., 2001; Panina *et al*., 2001; Rodionov *et al*., 2001; Roy *et al*., 2002). Nevertheless, and for this kind of studies to be carried out, a previous identification of new regulatory sequences, as well as of the genes which are under their control, is required.

In this context, and also to further characterize the differences existing among the LexA regulon of several groups of the gamma subclass *Proteobacteria*, the *X. fastidiosa lexA* gene has been cloned, overexpressed in *E. coli* and its product purified to determine the sequence to which it binds. Furthermore, the effect of DNA damage on the expression of *X. fastidiosa* genes whose homologues are under LexA regulation in *E. coli* has also been analysed.

METHODS

Bacterial strains, plasmids, oligonucleotides and DNA techniques. Bacterial strains and plasmids used in this work are listed in Table 1. *E. coli* and *X. fastidiosa* strains were grown at either 37 °C or 29 °C in LB or PW medium (Davis *et al*., 1981; Sambrook *et al*., 1992), respectively. Antibiotics were added to the cultures at reported concentrations (Sambrook *et al*., 1992). *E. coli* cells were transformed with plasmid DNA as described by 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 *X. fastidiosa* cells was extracted as described by Monteiro *et al*. (2001).

The synthetic oligonucleotide primers used for PCR amplification are listed in Table 2. To facilitate subcloning of some PCR-amplified DNA fragments, specific restriction sites were incorporated into the oligonucleotide primers. These restriction sites are shown in Table 2. Mutants in the *X. fastidiosa 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 (Pharmacia Biotech). In all cases the entire nucleotide sequence was determined for both DNA strands.

Molecular cloning of the *X. fastidosa lexA* gene. The *X. fastidiosa lexA* gene was amplified from the total DNA of the *X. fastidiosa* CVC 9a5c strain using the LexAup and LexAdw oligonucleotide primers (Table 2) corresponding to nucleotides −273 to −252 and +706 to +726, with respect to its proposed translational starting point (Simpson *et al*., 2000). The 999 bp PCR fragment obtained was cloned into the pGEM-T vector giving plasmid pUA945. To confirm that no mutation was introduced during the amplification reaction, the sequence of the fragment was determined. pUA945 was used as a template in PCR amplifications with LexANdeI and LexABamHI primers (Table 2), and the fragment obtained was again cloned in pGEM-T. The putative ATG initial triplet of LexA is part of the *Nde*I restriction site of the LexANdeI primer which enables IPTG-mediated overexpression of the *X. fastidiosa* LexA protein. After digestion with *Nde*I and *Bam*HI, the *lexA* gene was cloned downstream of the T7 promoter of the pET15b expression vector (Novagen) carrying a N-terminal tag containing six histidine residues according to standard procedures (Sambrook *et al*., 1992), and the ligation mix was transformed into DH5α competent cells. Once confirmed that no mutation had been introduced in the *lexA* gene contained in the pUA973 plasmid by DNA sequencing,

Table 1. Bacterial strains and plasmids used in this work

Strain/plasmid	Relevant features	Source/reference
Xylella fastidiosa		
CVC 9a5c	Wild-type strain	Fundecitrus
E. coli		
DH5α	*supE4* Δ*lacU169* (ϕ80 *lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Clontech
BL21(DE3)	F⁻ *ompT* $hsdS_B(r_B\text{-}r_B)$ *dcm gal* λ(DE3)	Clontech
Plasmids		
pGEM-T	PCR cloning vector, Ap^R	Promega
pET15b	Overexpression vector carrying a N-terminal tag containing six histidine residues; Km^R	Novagen
pUA945	pGEMT derivative carrying a 999 bp PCR fragment containing the *X. fastidiosa lex*A gene including its own promoter region	This work
pUA973	pET15b derivative carrying a 764 bp *Nde*I–*Bam*HI fragment containing the *X. fastidiosa lex*A gene	This work

this was transformed into the *E. coli* BL21(DE3) strain for overexpression of the LexA protein.

Purification of the *X. fastidiosa* LexA protein. His-tag fusion protein was purified by using the Talon Metal Affinity Resin Kit (Clontech) as described by the manufacturer. To carry out this, *E. coli* BL21(DE3) cells containing pUA973 were grown in 250 ml LB broth with shaking until OD_{550} 0·8. Protein expression was induced by adding 10 mM IPTG to the culture, which was incubated at 37 °C for an additional 3 h. Cells were harvested by centrifugation at 5000 ***g*** for 10 min at 4 °C, resuspended in 10 ml washing solution (50 mM NaH_2PO_4, 300 mM NaCl, pH 7) to which protease inhibitor cocktail (Complete Mini, EDTA free; Roche) was added to the concentration indicated by the supplier. After sonication of the cells, the lysate was centrifuged at 14000 ***g*** for 20 min, and the supernatant was collected. Metal Affinity Resin (Clontech; 2 ml), previously equilibrated with washing buffer, was added to the supernatant and the mix was agitated for 20 min at room temperature to allow binding of the tagged LexA protein to the resin. Afterwards, it was centrifuged and washed several times with washing buffer. The resin was then transferred to a gravity-flow column and washed again with 10 ml washing buffer. Tagged LexA protein was eluted by adding 10 ml elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 150 mM imidazole) and the eluate was collected in 500 µl fractions. To determine the fraction containing the majority of the tagged LexA protein, SDS-PAGE in 15 % polyacrylamide gels was performed according to standard procedures (Laemmli, 1970). The selected fraction was more than 98 % pure as determined by Coomassie blue staining as shown in Fig. 1.

Mobility shift assays. *X. fastidiosa* LexA–DNA complexes were detected by electrophoretic mobility shift assays using purified *X. fastidiosa* LexA protein (Fig. 1). DNA probes were prepared by PCR amplification using one of the primers labelled at its 5′ end with digoxigenin (DIG) (Table 2) and 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 DIG-labelled DNA probe and 40 ng purified *X. fastidiosa* LexA protein were incubated in binding buffer: 10 mM 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 BSA ml. After 30 min at 30 °C, the mixture was loaded onto a 5 % non-denaturing Tris/glycine polyacrylamide gel (pre-run for 30 min at 10 V cm^{-1} in 25 mM Tris/HCl (pH 8·5), 250 mM glycine, 1 mM EDTA. DNA–protein complexes were separated at 150 V for 1 h, 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 non-specific unlabelled competitor DNA was also included in the mixture. When used, the protein amount of crude extract of IPTG-induced *E. coli* BL21(DE3) cells carrying either the pUA973 plasmid or the pET15b vector alone added to the binding mixture was about 500 ng. Protein concentrations were determined as described by Bradford (1976). All mobility shift assays were repeated a minimum of three times to assure reproducibility of the results.

RT-PCR analysis of *X. fastidiosa* gene expression. To determine the transcriptional organization of *X. fastidiosa lexA* and *recA* genes, reverse transcriptase (Roche) was used to generate cDNA by RT-PCR using total RNA from *X. fastidiosa* as a template and the pair of primers indicated in Table 2 and designed to amplify a PCR product of 355 bp if the *lexA* and *recA* genes constituted a single transcription unit. Total RNA from *X. fastidiosa* was prepared with Trizol as described by the supplier (Gibco) and treated with RNase-free DNase I (Roche) to assure the absence of contaminating DNA. RNA concentration and its integrity were determined by A_{260} measurements and 1 % formaldehyde-agarose gel electrophoresis, respectively (Sambrook *et al.*, 1992). In all RT-PCR experiments, the absence of contaminating DNA in RNA samples after RNase-free DNase I (Roche) treatment was confirmed by processing a duplicate of them in the same way but without reverse transcriptase addition.

Mitomycin C-mediated induction of several genes studied in this work was carried out by real-time quantitative RT-PCR analysis of total *X. fastidiosa* RNA with the LightCycler apparatus (Roche), using the LC-RNA master SYBR green I kit (Roche) and primers indicated in Table 2, following

Table 2. Oligonucleotide primers used in this work

EMSA, electrophoretic mobility shift assay.

Primer	Sequence (5′–3′)*	Posn†	Application
LexAup	GCTCAGTGTGATGCATCCATT	−273	Upper primer for cloning *X. fastidiosa lexA* gene
LexAdw	CAGCTTAAACTGCCAAGGCA	+726	Lower primer for cloning *X. fastidiosa lexA* gene
LexANdeI	*CATATG*AGTTTGAGCGATATTCAG	+1	Upper primer for cloning *X. fastidiosa lexA* gene in the pET15b vector
LexABamHI	*GGATCC*CAGCTTAAACTGCCAAGGCA	+726	Lower primer for cloning *X. fastidiosa lexA* gene in the pET15b vector
LexA+92	GTCTGCGAAGGAGAAACGCC	+92	Lower primer for cloning *X. fastidiosa lexA* promoter used as specific competitor in EMSA assays
LexA+92dig	DIG-GTCTGCGAAGGAGAAACGCC	+92	Lower primer used to obtain wild-type and mutagenized *lex*A probes (digoxigenin 5′-end labelled)
LexA-87	GCTGGATCTGGCAGTATGG	−87	Upper primer to obtain the LexA2 fragment
LexA-25	GTGCTTATTAGTAATAATAC	−25	Upper primer to obtain the LexA3 fragment
LexA-10	AATACTAATTATGAGTTTG	−10	Upper primer to obtain the LexA4 fragment
Mut1	GTGCTTAaccaTAATAATACTAATTATG	−25	Upper primer to obtain the mutagenized probe containing a change on the left motif of the proposed palindrome
Mut2	GTGCTTATTAGTAATAATAaaggTTATG	−25	Upper primer to obtain the mutagenized probe containing a change on the right motif of the proposed palindrome
Mut+ggg	GTGCTTATTAGTgggAATAATACTAATTATG	−25	Upper primer to obtain the mutagenized probe containing an insertion in the middle of the proposed palindrome
LexA1	GTGCTTcTTAGTAATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA2	GTGCTTAaTAGTAATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA3	GTGCTTATcAGTAATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA4	GTGCTTATTcGTAATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA5	GTGCTTATTAaTAATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA6	GTGCTTATTAGcAATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA7	GTGCTTATTAGTcATAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA8	GTGCTTATTAGTAcTAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA9	GTGCTTATTAGTAAcAATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA10	GTGCTTATTAGTAATcATACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA11	GTGCTTATTAGTAATAcTACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA12	GTGCTTATTAGTAATAAcACTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA13	5′-GTGCTTATTAGTAATAATcCTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA14	5′-GTGCTTATTAGTAATAATAaTAATTATGAG	−25	Upper primer to obtain single mutants in *lexA* promoter

Table 2 (*cont.*)

Primer	Sequence (5′–3′)*	Posn†	Application
LexA15	GTGCTTATTAGTAATAATACaAATTATGAGATATTCAGC	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA16	GTGCTTATTAGTAATAATACTgATTATGAGATATTCAGC	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA17	GTGCTTATTAGTAATAATACTAgTTATGAGATATTCAGC	−25	Upper primer to obtain single mutants in *lexA* promoter
LexA18	GTGCTTATTAGTAATAATACTAAgTATGAGATATTCAGC	−25	Upper primer to obtain single mutants in *lexA* promoter
RecAup	GACCGGATTCGTTTGCTACC	+522‡	Upper primer for cloning the upstream region of *X. fastidiosa recA* gene, to analyse transcriptional organization of *recA* and *lexA* genes, and for mRNA quantification assays
RecAdw	CCCTTGCCGAACTAATCTGG	+877‡	Lower primer for cloning the upstream region of *X. fastidiosa recA* gene, to analyse transcriptional organization of *recA* and *lexA* genes, and for mRNA quantification assays
RecAdw-dig	dig-CCCTTGCCGAACTAATCTGG	+877‡	Lower primer used to obtain *rec*A probes for EMSA assays (digoxigenin 5′-end labelled)
UvrAup	*GAATTC*TGAACAACGCGATATTGC	−387	Upper primer for cloning *X. fastidiosa uvrA* promoter for EMSA assays
UvrAdw	*GGATCC*GATTCTTAAGATTGTGTG	+53	Lower primer for cloning *X. fastidiosa uvrA* promoter for EMSA assays
UvrBup	*GAATTC*TCGCTAAGCAGGTGTTGC	−407	Upper primer for cloning *X. fastidiosa uvrB* promoter for EMSA assays
UvrBdw	*GGATCC*CCTGCTTCTCAGTGAGATG	+39	Lower primer for cloning *X. fastidiosa uvrB* promoter for EMSA assays
Ssbup	*GAATTC*CAGCTAACATCGTTAGC	−301	Upper primer for cloning *X. fastidiosa ssb* promoter for EMSA assays
Ssbdw	*GGATCC*GATCGTTCAAGCGGTTAC	+76	Lower primer for cloning *X. fastidiosa ssb* promoter for EMSA assays
DinGup	GAACTCGTCACCAATGATCC	−394	Upper primer for cloning *X. fastidiosa dinG* promoter for EMSA assays
DinGdw	GTTGTTCAAAAATCTCGCC	+75	Lower primer for cloning *X. fastidiosa dinG* promoter for EMSA assays
RecNup	ACACGTAGCCGACGAATATC	−420	Upper primer for cloning *X. fastidiosa recN* promoter for EMSA assays
RecNdw	GATGTTACGGACAACAGCG	+49	Lower primer for cloning *X. fastidiosa recN* promoter for EMSA assays
RuvAup	CCGCCATGTCTACCATGC	−526	Upper primer for cloning *X. fastidiosa ruvA* promoter for EMSA assays
RuvAdw	CTCCACACACATCAACCACG	+69	Lower primer for cloning *X. fastidiosa ruvA* promoter for EMSA assays
YigNup	CGTTATGCCGAATACTTTGC	−466	Upper primer for cloning *X. fastidiosa yigN* promoter for EMSA assays
YigNdw	GTCATGAAGCACCATTGAGC	+66	Lower primer for cloning *X. fastidiosa yigN* promoter for EMSA assays
Ftskup	TACTGGTTGAAGGTTTGC	−451	Upper primer for cloning *X. fastidiosa ftsK* promoter for EMSA assays
FstKdw	GCGATCATGATCAACACC	+58	Lower primer for cloning *X. fastidiosa ftsK* promoter for EMSA assays
XF2313up	TAACCAGATCATTGCCGTGG	−303	Upper primer for cloning *X. fastidiosa* XF2313 promoter for EMSA assays
XF2313dw	AAGCAGCCGCAGTGCGTCGC	+59	Lower primer for cloning *X. fastidiosa* XF2313 promoter for EMSA assays
XF1822up	CCAACAGTGCATATGTTGGC	+51	Upper primer for cloning *X. fastidiosa* XF1822 promoter for EMSA assays
XF1822dw	ATTCTGCTATTACGGTAGCC	−549	Lower primer for cloning *X. fastidiosa* XF1822 promoter for EMSA assays
XF1614up	TCTTCTAGTGGATTCTGTCC	−556	Upper primer for cloning *X. fastidiosa* XF1614 promoter for EMSA assays
XF1614dw	GCTCCAGTGCTTATAATCGG	−253	Lower primer for cloning *X. fastidiosa* XF1614 promoter for EMSA assays

Table 2 (*cont.*)

Primer	Sequence (5′–3′)*	Posn†	Application
XF1417up	TGCTTGGCGTTATCGTCACC	−126	Upper primer for cloning *X. fastidiosa* XF1417 promoter for EMSA assays
XF1417dw	AAAGGCCCTGTTCCAGCCTC	+180	Lower primer for cloning *X. fastidiosa* XF1417 promoter for EMSA assays
XF1271up	GTTCGATGTCGGAGGACGTC	−317	Upper primer for cloning *X. fastidiosa* XF1271 promoter for EMSA assays
XF1271dw	ACCATGCTCAGAATGATCGG	+32	Lower primer for cloning *X. fastidiosa* XF1271 promoter for EMSA assays
XF2297up	CTCTACGTAGTTGTACGC	−321	Upper primer for cloning *X. fastidiosa* XF2297 promoter for EMSA assays
XF2297dw	CGCCAATCAGTTAAACAC	+84	Lower primer for cloning *X. fastidiosa* XF2297 promoter for EMSA assays
XF2313/1	TTGTTGCGTCATCTGCGTCG	+5	Upper primer for cloning an internal region of *X. fastidiosa* XF2313 gene for mRNA quantification assays
XF2313/2	CTCTGTACGGCATCGGTGAG	+361	Lower primer for cloning an internal region of *X. fastidiosa* XF2313 gene for mRNA quantification assays
UvrA1	CGCGGACACACAATCTTAAG	+29	Upper primer for cloning an internal region of *X. fastidiosa uvrA* gene for mRNA quantification assays
UvrA3	TGCGCTCACGTACCACAG	+425	Lower primer for cloning an internal region of *X. fastidiosa uvrA* gene for mRNA quantification assays
UvrB1	TGAGCAGCGGCAATTGATCAG	+1	Upper primer for cloning an internal region of *X. fastidiosa uvrB* gene for mRNA quantification assays
UvrB2	GTCTGACCGGTAAGCGGATC	+202	Lower primer for cloning an internal region of *X. fastidiosa uvrB* gene for mRNA quantification assays
DinG1	TACCGGCACGTTCAACATGC	+63	Upper primer for cloning an internal region of *X. fastidiosa dinG* gene for mRNA quantification assays
DinG2	TAGATCGCGGTGGTAGAGCT	+431	Lower primer for cloning an internal region of *X. fastidiosa dinG* gene for mRNA quantification assays
RecN1	CAACATCACGATACACCACCC	+204	Upper primer for cloning an internal region of *X. fastidiosa recN* gene for mRNA quantification assays
RecN2	GATTCGCTCGGACATATCGC	+558	Lower primer for cloning an internal region of *X. fastidiosa recN* gene for mRNA quantification assays
Ssb1	TGGCCCGTGGTATCAATAAG	+7	Upper primer for cloning an internal region of *X. fastidiosa ssb* gene for mRNA quantification assays

Table 2 (*cont.*)

Primer	Sequence (5′–3′)*	Posn†	Application
Ssb2	TTGTCATAGCGGATGCTCCC	+263	Lower primer for cloning an internal region of *X. fastidiosa ssb* gene for mRNA quantification assays
YigN1	GTTTGGATGAGCTGACTGCC	+204	Upper primer for cloning an internal region of *X. fastidiosa yigN* gene for mRNA quantification assays
YigN2	AGACACCGTTGAGATCACCG	+518	Lower primer for cloning an internal region of *X. fastidiosa yigN* gene for mRNA quantification assays
FtsK1	AGGCTGGTGCTTGGATTGCG	+81	Upper primer for cloning an internal region of *X. fastidiosa ftsK* gene for mRNA quantification assays
FtsK2	AAGCACAGCACGACCAATGC	+415	Lower primer for cloning an internal region of *X. fastidiosa ftsK* gene for mRNA quantification assays
RuvA1	CTCGTGGTTGATGTGTGTGG	+50	Upper primer for cloning an internal region of *X. fastidiosa ruvA* gene for mRNA quantification assays
RuvA2	CTCAACTCCACCAGCATACG	+386	Lower primer for cloning an internal region of *X. fastidiosa ruvA* gene for mRNA quantification assays
Thr1	TGCTCAAGCACAAACACC	+1059	Upper primer for cloning an internal region of *X. fastidiosa* threonine synthase gene for mRNA quantification assays
Thr2	CTGAACCTCATTGTGCCC	+685	Lower primer for cloning an internal region of *X. fastidiosa* threonine synthase gene for mRNA quantification assays
Trp1	TGCACAGGCCTATGATGC	+246	Upper primer for cloning an internal region of *X. fastidiosa* tryptophan synthase gene for mRNA quantification assays
Trp2	ACCCAACAACTTCATCCG	+624	Lower primer for cloning an internal region of *X. fastidiosa* tryptophan synthase gene for mRNA quantification assays
rRNA1	ACGCTAATACCGCATACGACC	172439§	Upper primer for cloning an internal region of *X. fastidiosa* 16S rRNA gene for mRNA quantification assays (used as control)
rRNAdw2	GAAGTTAGCCGGTGCTTATTC	173003§	Lower primer for cloning an internal region of *X. fastidiosa* 16S rRNA gene for mRNA quantification assays

* When present, added restriction sites are shown in italics and introduced nucleotide changes are shown in lower case and underlined.
† Position of 5′ end of the oligonucleotide with respect to the proposed translational starting point of each *X. fastidiosa* gene.
‡ Position of 5′ end of the oligonucleotide with respect to the proposed translational starting point of *X. fastidiosa lexA* gene.
§ Position of 5′ end of the oligonucleotide in *X. fastidiosa* genome.

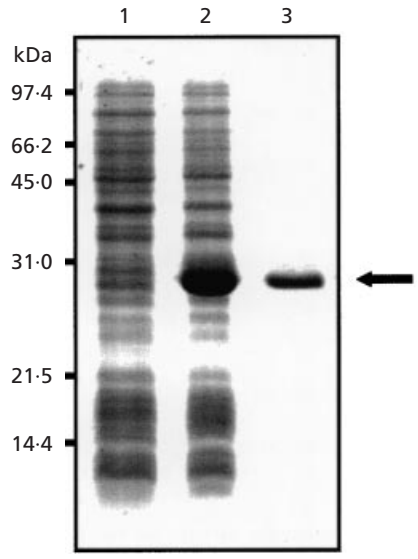


Fig. 1. Overproduction and purification of *X. fastidiosa* LexA in *E. coli*. Samples were analysed by SDS-PAGE (15% acrylamide). Lanes 1 and 2 contain crude extracts from *E. coli* BL21(DE3)/pUA973 cells in the absence or in the presence of IPTG, respectively. The purified His-tagged *X. fastidiosa* LexA protein after Co^{2+} -affinity chromatography is shown in lane 3. The molecular masses of protein markers are indicated on the left side.

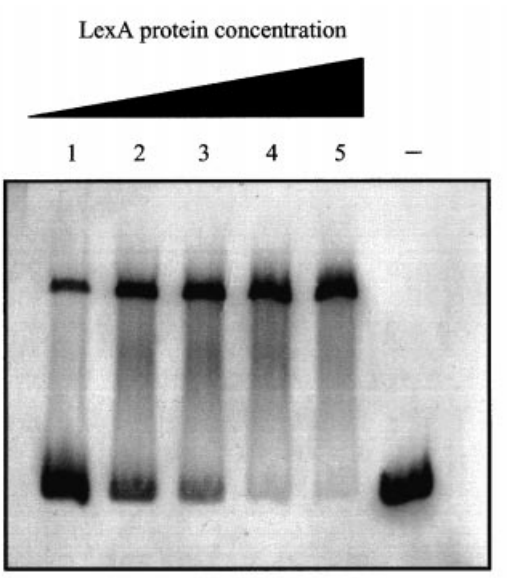


Fig. 2. Electrophoretic mobility of the DNA fragment containing the *X. fastidiosa lexA* promoter in presence of increasing concentrations of purified *X. fastidiosa* His_6–LexA protein. In each binding reaction (end volume of 20 µl), amounts of purified His_6–LexA protein used were 5 ng (lane 1), 10 ng (lane 2), 20 ng (lane 3), 40 ng (lane 4) and 80 ng (lane 5). The mobility of the *X. fastidiosa lexA* promoter in the absence of purified His_6–LexA protein is also shown as a control (–).

supplier's instructions. The concentration of total RNA of both treated and untreated cultures was adjusted to the same value. The amount of mRNA of each gene was determined by plotting it against a standard curve generated by the amplification of an internal fragment of the *X. fastidiosa* 16S rRNA with the appropriate primers indicated in Table 2. In all determinations, the amount of mRNA of *X. fastidiosa* XF1375 and XF2223 genes, encoding tryptophan synthase and threonine synthase, respectively, was also determined. These two genes were used as negative controls because their expression is not DNA-damage-inducible in *E. coli* (Courcelle *et al.*, 2001).

RESULTS AND DISCUSSION

X. fastidiosa LexA protein specifically binds to its own promoter

Gel mobility shift assays with the purified *X. fastidiosa* LexA were used to determine the binding ability of this protein to its own promoter. Addition of increasing concentrations of LexA to a fragment extending from −273 to +92 of the *X. fastidiosa lexA* gene promoter (with respect to its proposed translational start point) produces one retardation band whose intensity is directly related to the amount of protein used (Fig. 2). To more exactly locate the binding region, the *X. fastidiosa lexA* promoter was divided into four fragments (designated as LexA1, LexA2, LexA3 and LexA4) obtained by PCR amplification with the oligonucleotides listed in Table 2, DIG end-labelled and then used as probes in gel retardation assays (Fig. 3a). A stable DNA–protein complex was observed when fragments LexA1, LexA2 and LexA3 were incubated in the presence of purified *X. fastidiosa* LexA, but there was no change in the mobility of the LexA4 fragment under the same conditions (Fig. 3b). These data indicate that the region to which the LexA protein binds, or at least a portion of it, must lie between positions −25 and −10 of the *lexA* promoter. No DNA–protein complex was detected when a 355 bp fragment comprising 295 bp of the upstream region of *recA* gene was used as a probe (Fig. 3b). This is in agreement with the close proximity existing between the *X. fastidiosa lexA* and *recA* genes, which is only 184 bp (Fig. 3a) and with the fact that both constitute a single transcriptional unit in this organism (Fig. 4) as shown by RT-PCR analysis of the *X. fastidiosa* RNA with RecAup and RecAdw primers beginning 111 bp upstream and 60 bp downstream of the translational stop and starting codons of the *lexA* and *recA* genes, respectively (Fig. 3a). It is worth noting that this is the first bacterial species described so far in which *lexA* and *recA* genes constitute a single transcriptional unit.

The DNA–LexA complex was sensitive to competition by an excess of unlabelled LexA3 fragment, but not when non-specific DNA was used as competitor (Fig. 5a, lanes 3 and 4, respectively). The same DNA–LexA complex was obtained when, instead of purified LexA repressor, crude extracts from IPTG-induced *E. coli* BL21(DE3) cells overexpressing the *X. fastidiosa* LexA

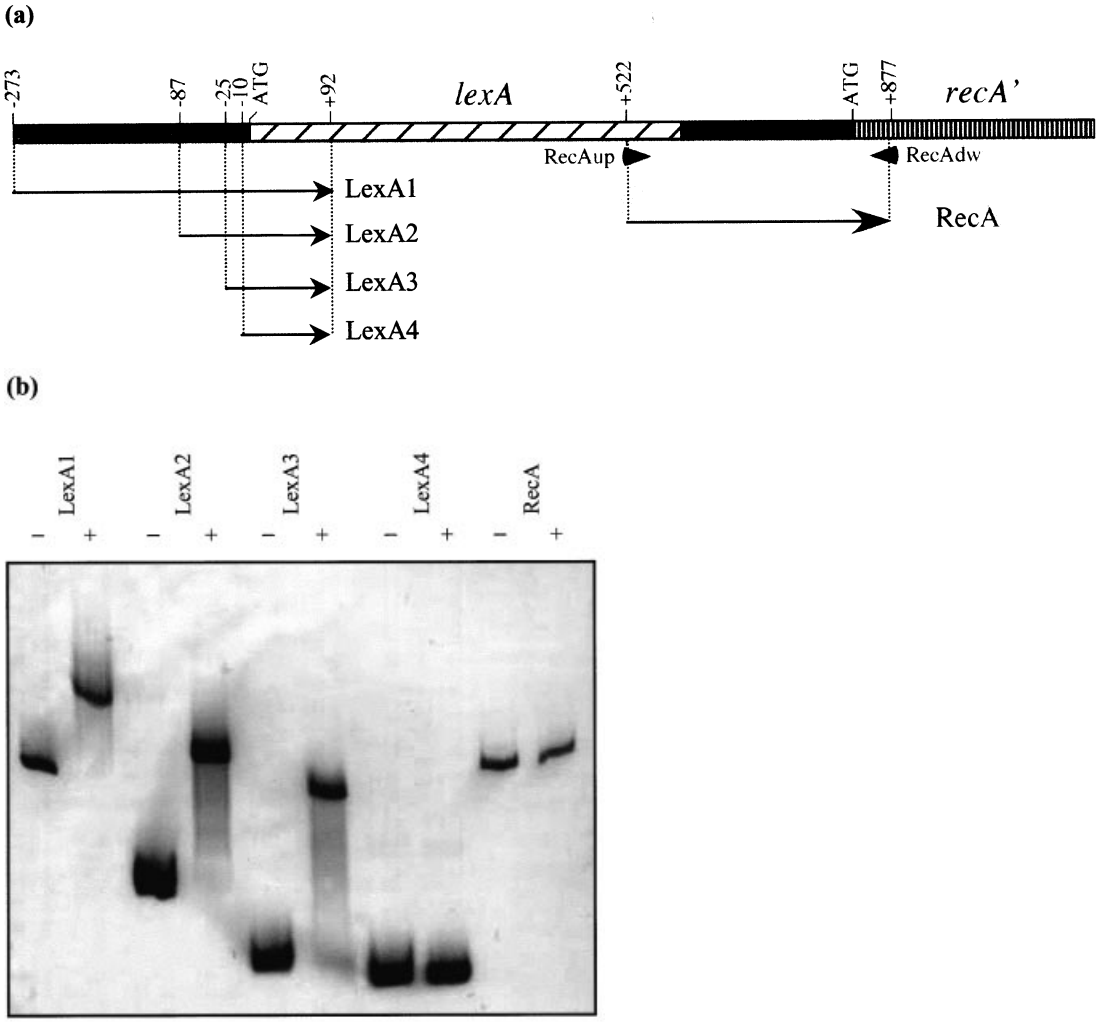


Fig. 3. (a) Diagram representing fragments amplified from the *X. fastidiosa lexA–recA* promoter used in gel retardation experiments. LexA1, LexA2, LexA3 and LexA4 fragments were generated by PCR amplification. The primers RecAup and RecAdw were used for amplification of the RecA DNA fragment and to analyse the transcriptional organization of the *lexA* and *recA* genes. In all cases, positions indicated refer to the predicted translational starting point of the *X. fastidiosa lexA* gene (Simpson *et al.*, 2000). (b) Electrophoretic mobility of LexA1, LexA2, LexA3, LexA4 and RecA fragments in the presence (+) and in the absence (−) of 40 ng purified *X. fastidiosa* LexA protein.

(Fig. 1, lane 2) were used (Fig. 5a, lane 5). Moreover, no DNA–LexA complex was detected when crude extracts of *E. coli* BL21(DE3) cells carrying the pET15b plasmid alone were employed in gel retardation assays with the LexA3 fragment (Fig. 5a, lane 6). These results unequivocally demonstrate that the complex was indeed formed by the specific binding of the X. *fastidiosa* LexA protein to the *lexA* promoter.

Mutational analysis of the *lexA* promoter

A search for potential binding motifs (i.e. either direct or inverted repeats) in the sequence between positions −25 and the proposed ATG translational starting codon was carried out. The palindrome TTAGTAATAATACTAA, centred at the −11 position, was detected in the LexA3 fragment. Then, mutations were introduced into the left (TTAG→ACCA) and the right (CTAA→AAGG) halves of this palindrome, and their effects on the mobility of the LexA3 fragment were analysed.

Results obtained indicated that no DNA–LexA complex was formed when probes containing substitutions in either the left or right halves of the palindrome were employed (Fig. 5b, lanes 3 and 5). Furthermore, a three-base insertion between the TTAG and CTAA motifs also abolished the gel retardation of the band (Fig. 5b, lane 4).

To determine which bases of this $TTAGN_8CTAA$ palindrome were directly involved in the LexA interaction with the DNA, the effect of a single substitution in each one of the nucleotides of the palindrome and its immediately surrounding bases was tested. Results obtained indicated that a single nucleotide

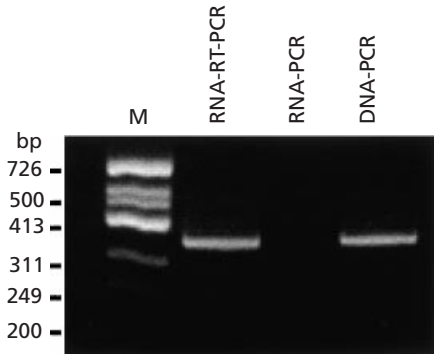


Fig. 4. RT-PCR analysis of transcripts of the *lexA–recA* operon in *X. fastidiosa* cells using RecAup and RecAdw primers in the presence of total RNA (RNA-RT-PCR). Results obtained when the PCR experiment is carried out with the same primers, but without reverse transcriptase, using either RNA (RNA-PCR) or DNA (DNA-PCR) are shown as controls. As indicated in Fig. 3a, the beginning of RecAup and RecAdw primers are 111 bp upstream and 60 bp downstream of the translational stop and starting codons of the *lexA* and *recA* genes, respectively. M, marker.

substitution in every position of the TTAG motif, as well as in any of the first three nucleotides (CTA) of the second motif of the palindrome, eliminates the DNA–LexA complex formation (Fig. 6). Additionally, mutagenesis of either T or A placed immediately upstream of the CTA trinucleotide also prevents DNA–LexA complex formation (Fig. 6). In accordance with these data, neither of the mutant fragments that do not form the DNA–LexA complex was able to abolish the mobility shift of the wild-type fragment originated in the presence of purified *X. fastidiosa* LexA protein when used in competition experiments (data not shown). These results demonstrate that binding of the *X. fastidiosa* LexA protein to the *lexA* promoter requires the TTA-GN_6TACTA imperfect palindrome. In agreement with the fact that *lexA* and *recA* genes are co-transcribed (Fig. 3b and Fig. 4), no motif presenting a sequence related with this TTAGN_6TACTA palindrome is found immediately upstream of the *X. fastidiosa recA* gene.

It must be noted that the LexA binding site present upstream of the *lexA–recA* operon is very close to the translational start codon of the *lexA* gene. This is not a usual situation since LexA binding motifs of *E. coli*, Gram-positive bacteria and alpha *Proteobacteria* are typically located about 25 nt upstream of their transcriptional starting points. However, some exceptions exist since *Mycobacterium tuberculosis* Rv3074 and Rv3776 genes also present a LexA binding sequence near their translational starting points (Davis *et al*., 2002).

LexA protein binds different gene promoters in *X. fastidiosa* and *E. coli*

As mentioned above, at least 40 genes are directly under the control of the LexA repressor in *E. coli* (Fernandez de Henestrosa *et al*., 2000; Courcelle *et al*., 2001). A

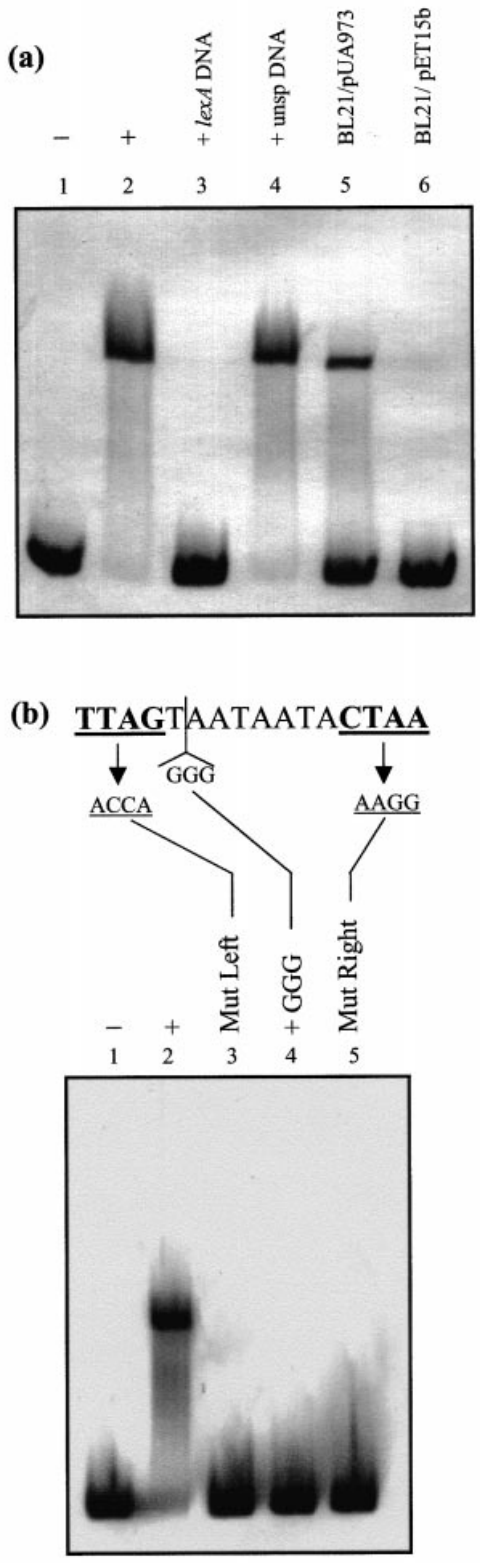


Fig. 5. (a) Effect of 300-fold molar unlabelled LexA1 fragment (lane 3) and pGEM-5Zf(+) plasmid DNA (lane 4) on the migration of the LexA3 fragment in the presence of 40 ng purified *X. fastidiosa* LexA protein. The electrophoretic mobility of the LexA3 fragment in the presence of crude extracts from IPTG-induced *E. coli* BL21(DE3) cells carrying either the pUA973 plasmid or the pET15b vector alone is also shown as a control

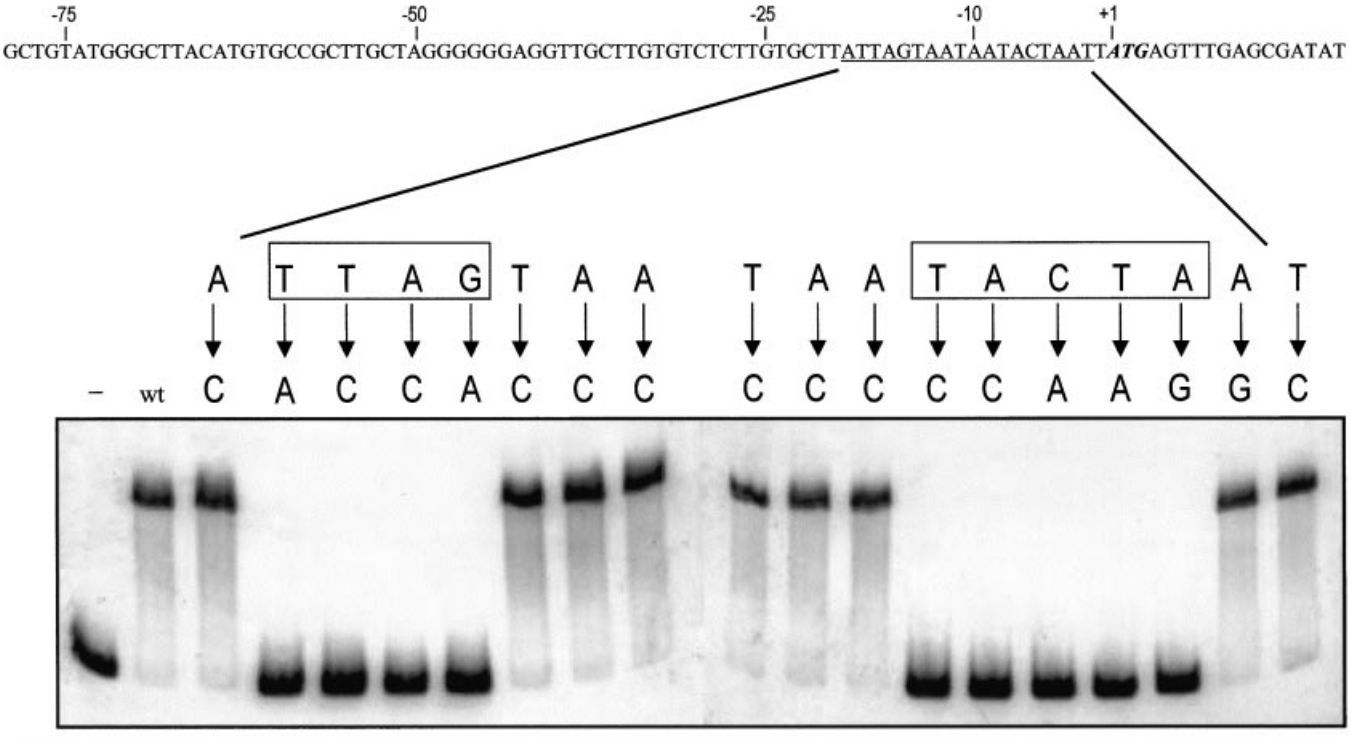


Fig. 6. Effect of single-nucleotide substitutions in the TTAG and CTAA motifs of the TTAGTAATAATACTAA palindrome as well as in its immediately surrounding bases in the electrophoretic mobility of the LexA3 fragment in the presence of 40 ng purified *X. fastidiosa* LexA protein. The mobility of the wild-type LexA3 fragment in absence (−) or in presence (wt) of the same amount of purified *X. fastidiosa* LexA protein is also presented as a control. The proposed translational starting codon (position +1) of the *X. fastidiosa lexA* gene (Simpson *et al.*, 2000) is in bold and italics and has been taken as the +1 position. Bases whose substitution abolish the LexA–DNA complex are boxed.

TBLASTX search using each one of the encoded products of these *E. coli* LexA-regulated genes as a query has enabled us to demonstrate that, besides *lexA* and *recA*, only eight of them (*uvrA*, *uvrB*, *ssb*, *recN*, *ruvAB*, *dinG*, *ftsK* and *yigN*) are present in the genome of *X. fastidiosa*. Sequence analysis of the upstream regions of these eight *X. fastidiosa* genes revealed that none has a sequence related to the $TTAGN_6TACTA$ palindrome. In agreement with this fact, fragments containing 400 bp of the upstream region of these eight genes were not able to inhibit the DNA–*X. fastidiosa* LexA complex formation when used as unlabelled competitors in gel retardation experiments with the DIG-labelled LexA3 fragment as a probe (data not shown). Moreover, a search in the *X. fastidiosa* genome database using the Findpatterns Program of the Genetics Computer Group package (Devereux *et al.*, 1984) has revealed the presence of the $TTAGN_6TACTA$ palindrome upstream of six ORFs: *lexA–recA*, XF1271, XF1417, XF1614, XF1823 and XF2313. It was further verified by competitive gel retardation assays (Fig. 7) that the LexA repressor only binds to the promoter of XF2313 encoding a putative DNA-modification methylase (Simpson *et al.*, 2000). A similar case in which a putative LexA-binding sequence has been identified by *in silico* analysis, but which afterwards does not bind the LexA protein *in vitro*, has also been for the *E. coli dinJ* gene (Fernández de Henestrosa *et al.*, 2000). The reasons for this lack of binding have not been established although they could be attributed to additional factors such as the structure of the region or the secondary role of the neighbouring bases of the regulatory sequence, as has been demonstrated for several transcriptional regulators (Winterling *et al.*, 1998; Griffith & Wolf, 2001; Davis *et al.*, 2002). In agreement with this, the sequence between both motifs of the imperfect palindrome $TTAGN_6TACTA$ is very different in XF1271, XF1417, XF1614 and XF1823 than in XF2313 and *lexA* (Fig. 7). It must be noted that in the *X. fastidiosa* genome there is another ORF (called XF2297) which is almost identical to XF2313 although it is slightly shorter probably due to a frameshift near its N terminus, suggesting that XF2297 derives from XF2313.

(lanes 5 and 6, respectively). (b) Electrophoretic mobility of mutant derivatives of the LexA3 fragment presenting substitutions in either the TTAG (Mut Left), CTAA (Mut Right) motif of the TTAGTAATAATACTAA palindrome or an insertion of 3 nt (+GGG) in its internal region in the presence of 40 ng purified *X. fastidiosa* LexA protein. In both panels, the mobility of the LexA3 fragment in the absence or in the presence of the same amount of purified *X. fastidiosa* LexA protein (lanes 1 and 2, respectively) is also shown as control.

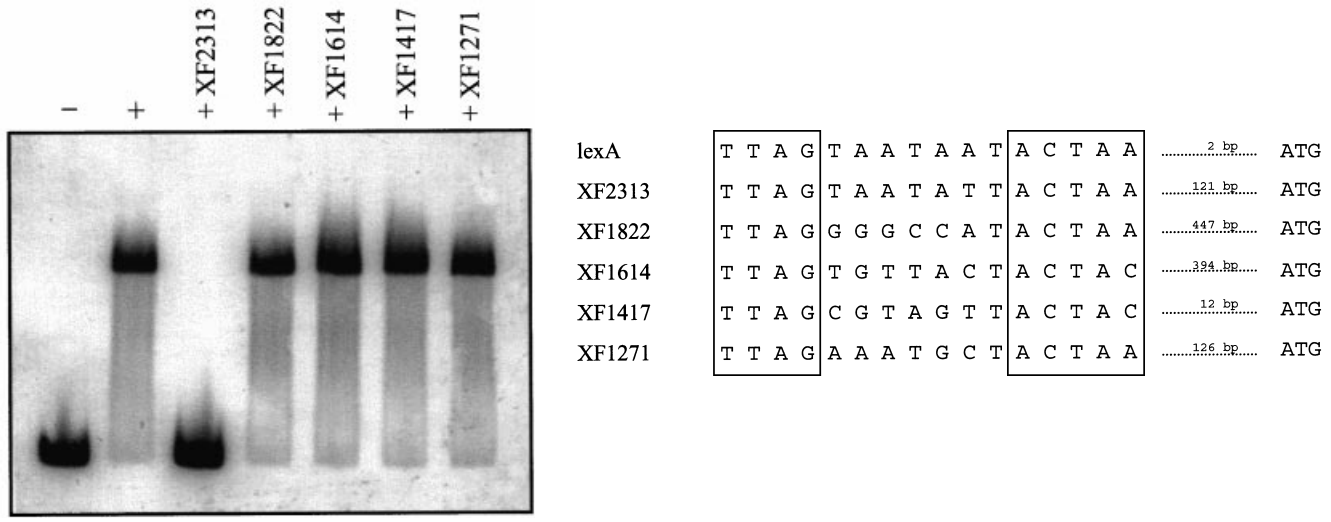


Fig. 7. Electrophoretic mobility of the LexA3 fragment in the presence of purified *X. fastidiosa* LexA protein and 300-fold molar excess of unlabelled DNA fragments containing about 400 bp of the upstream region of XF2313, XF1822, XF1614, XF1417 and XF1271 genes of *X. fastidiosa*. The mobility of the LexA fragment in the absence of any additional DNA (+) or purified *X. fastidiosa* LexA protein (−) is shown as a control. The sequence of the region between both palindromic halves (boxed) of the TTAGN$_6$TACTA motif, as well as the distance to their respective translational starting codons, are also shown for each fragment to illustrate differences existing among them.

Organism	Gene		Sequence											
X. fastidiosa CVC strain	*lexA*	TTA	**T**	**T**	**A**	**G**	T	A	A	T	A	A	**T**	**A**	**C**	**T**	**A**	ATTA
	XF2313	TAT	**T**	**T**	**A**	**G**	T	A	A	T	T	T	**T**	**A**	**C**	**T**	**A**	AATC
X. fastidiosa oleander strain	*lexA*	TTA	**T**	**T**	**A**	**G**	T	A	A	T	A	A	**T**	**A**	**C**	**T**	**A**	ATTA
	XF2313-like	TAT	**T**	**T**	**A**	**G**	T	A	A	A	A	T	**T**	**A**	**C**	**T**	**A**	AATA
X. fastidiosa almonder strain	*lexA*	TTA	**T**	**T**	**A**	**G**	T	A	A	T	A	A	**T**	**A**	**C**	**T**	**A**	ATTA
	XF2313-like	TAT	**T**	**T**	**A**	**G**	T	A	A	A	A	T	**T**	**A**	**C**	**T**	**A**	AATA
Xanthomonas citri	*lexA*	TCG	**T**	**T**	**A**	**G**	T	A	G	T	A	A	**T**	**A**	**C**	**T**	**A**	CTAA
Xanthomonas campestris	*lexA*	TCG	**T**	**T**	**A**	**G**	T	A	G	T	A	A	**T**	**A**	**C**	**T**	**A**	CTAA
Xanthomonas oryzae	*lexA*	TTA	**T**	**T**	**A**	**G**	T	A	A	T	A	A	**T**	**A**	**C**	**T**	**A**	AACA
Consensus sequence			**T**	**T**	**A**	**G**	T	A	R	W	A	W	**T**	**A**	**C**	**T**	**A**	

Fig. 8. Alignment of the sequence containing the TTAGN$_6$TACTA motif present upstream of several genes of members of the *Xanthomonas* and *Xylella* genera. The consensus sequence includes nucleotides conserved in more than the 20% of 9 sequences considered. The GenBank accession numbers of *Xanthomonas citri*, *Xan. campestris* and *Xan. oryzae lexA* genes are AF081945, AF399933 and AF399932, respectively. The upstream sequences of XF2313-like genes of *Xyl. fastidiosa* pv. *oleander* and *Xyl. fastidiosa* pv. *almonder* strains were obtained by TBLASTN analysis against their respective genomes (http://igweb.integratedgenomics.com/GOLD/prokaryagenomes.html) using the *Xyl. fastidiosa* XF2313 gene as a query. In agreement with the standard nomenclature R is A or G and W is A or T.

Nevertheless, a 600 bp fragment containing the upstream region of XF2297 neither presents a copy of the TTAGN$_6$TACTA motif nor binds the *X. fastidiosa* LexA protein (data not shown). The LexA-binding sequence was perhaps lost in the genetic rearrangement that produced XF2297 from XF2313.

Three proteins with a significant degree of identity (about 43%) to the putative *X. fastidiosa* DNA-modification methylase encoded by XF2313 have been detected in the *Streptomyces coelicolor* genome after a TBLASTX search in the GenBank database. However, upstream of these *S. coelicolor* genes a Gram-positive

LexA-binding sequence is not found, suggesting that the presence of a LexA-binding sequence upstream of orthologous genes encoding an XF2313-like protein is not a general phenomenon. The absence of a LexA box in the promoters of *E. coli* LexA-regulated orthologous genes has also been described for *Ps. aeruginosa* (*uvrA* and *uvrB*) and *M. tuberculosis* (*uvrA*, *recN*, *dinP* and *dinG*) (Rivera *et al*., 1996, 1997; Brooks *et al*., 2001).

Moreover, the $TTAGN_6TACTA$ palindrome is also present in an upstream region 169 nt long in the *Xanthomonas campestris lexA* gene, in which it has been shown that its own LexA protein binds, although the precise binding sequence has not been established (Yang *et al*., 2001). It must be noted that *Xyl. fastidiosa* and *Xan. campestris* LexA proteins present a high degree of identity in their N-terminal regions. According to this fact, the DNA binding sequence of both bacteria are expected to be very similar. Furthermore, the fact that the $TTAGN_6TACTA$ palindrome is perfectly conserved in the *lexA* promoters of *Xanthomonas oryzae*, *Xyl. fastidiosa* pv. *oleander* and *Xyl. fastidiosa* pv. *almonder*, as well as in the promoter of the XF2313-like genes of these two former species (Fig. 8), leads us to propose this sequence as the LexA binding site of the Order Xanthomonadales. The $TTAGN_6TACTA$ palindrome is the first LexA-binding sequence different from the *E. coli*-like one identified in the gamma subclass *Proteobacteria*.

X. fastidiosa lexA–recA and XF2313 transcriptional units are inducible by DNA damage

To further characterize the behaviour of *X. fastidiosa* genes binding to LexA protein, the effect of mitomycin C (20 μg ml^{-1}) on the expression of *lexA–recA* and XF2313 was analysed. Table 3 indicates that both transcriptional units are induced by DNA damage, whereas expression of two non-DNA-damage-related genes (those encoding tryptophan and threonine synthase) is not significantly stimulated by mitomycin C. It is worth noting that all LexA-regulated *E. coli* genes which are also present in *X. fastidiosa*, with the exception of *ssb*, are also DNA damage-inducible in this organism despite the fact they cannot bind to their own LexA protein (Table 3). A similar result has also been described for *M. tuberculosis* in which *uvrA* and *uvrB* are DNA damage-inducible but do not bind the LexA protein, whereas *recN*, *dinP* and *dinG* have their transcriptions neither stimulated by DNA injuries nor bind LexA (Brooks *et al*., 2001). The DNA damage-mediated transcription of genes that do not bind LexA protein in either *M. tuberculosis* or *X. fastidiosa* could be explained by two different mechanisms: i) the existence of a *lexA*-independent pathway of induction, or ii) the presence of an additional transcriptional regulatory factor which is under LexA control. Nevertheless, this last possibility seems to be unlikely since a gene encoding a putative transcriptional regulator, aside from *lexA*, presenting its own LexA-binding sequence has not been detected in either the *M. tuberculosis* or *X. fastidiosa* genomes, (Simpson *et al*., 2000; Cole *et al*., 1998). Additional investigations are necessary to better understand the molecular basis of this behaviour. Moreover, the fact that bacterial species belonging to phylogenetic groups as different as Gram-positive and the gamma subclass of *Proteobacteria* have DNA damage-inducible genes which are unable to bind the LexA repressor suggests that this phenomenon may be widespread in the bacterial world.

Table 3. Mitomycin C-mediated induction of genes whose promoters bind either the *X. fastidiosa* or *E. coli* LexA proteins

Gene	*X. fastidiosa* LexA binding	Induction factor*
lexA	+	3·91
XF2313	+	15·98
ssb	−	1·36
uvrA	−	2·27
uvrB	−	6·11
ftsK	−	4·3
recN	−	6·4
yigN	−	6·52
ruvAB	−	6·67
dinG	−	7·27
XF1375†	−	1·27
XF2223†	−	1·66

* The induction factor is the ratio of mRNA concentration of each gene of cells treated with mitomycin C to that of untreated cells. Values were calculated 16 h after addition of mitomycin C. For all cases, data presented are the means of three independent experiments (each in triplicate), and the single standard deviation of any value was never greater than 10%.

† XF1375 and XF2223 genes of *X. fastidiosa* encoding tryptophan synthase and threonine synthase, respectively, were used as negative controls.

The presented results clearly indicate that gene composition of the LexA network can have significant variations depending on the bacterial species. The reasons for these differences are not yet known, but, besides evolutionary divergences, they could probably be related to the specific ecological and physiological niches of each bacterial group. Nevertheless, at the moment there is not enough information to precisely establish this relationship. Further work is needed to identify new LexA-binding sequences in other bacterial phylogenetic groups to be able to determine the gene composition of their corresponding LexA regulons for the understanding of the variability of this important gene network.

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. Antonio R. Fernández de Henestrosa and Gerard Mazón were recipients of a postdoctoral contract and a predoctoral fellowship from the MCyT, respectively. We are deeply indebted to Dr Roger Woodgate for his generous gifts. We wish to acknowledge Joan Ruiz, Susana Escribano, Pilar Cortés, Rene R. Palma and Anelise G. Mariano for their excellent technical assistance and collaboration and to our English-teaching university colleague, Mr Chuck Simmons, for his help in the language revision and correction of this article.

REFERENCES

Bertrand-Burggraf, E., Hurstel, S., Daune, M. & Schnarr, M. (1987). Promoter properties and negative regulation of the *uvrA* gene by the LexA repressor and its amino-terminal binding domain. *J Mol Biol* **193**, 293–302.

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.

Brent, R. & Ptashne, M. (1981). Mechanism of action of the *lexA* gene product. *Proc Natl Acad Sci U S A* **78**, 4204–4208.

Brooks, P. C., Movahedzadeh, F. & Davis, E. O. (2001). Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*: apparent lack of correlation with LexA binding. *J Bacteriol* **183**, 4459–4467.

Cole, S. T., Brosch, R., Parkhill, J. & 36 other authors. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

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, J. M., French, W. J. & Schaad, N. (1981). Axenic culture of the bacteria associated with phony disease of peach and plum leaf scald. *Curr Microbiol* **6**, 309–314.

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.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.

Eisen, J. A. & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* **435**, 171–213.

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

Fernandez 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.

Griffith, K. L. & Wolf, R. E. (2001). Systematic mutagenesis of the DNA binding sites for SoxS in the *Escherichia coli zwf* and *fpr* promoters: identifying nucleotides required for DNA binding and transcription activation. *Mol Microbiol* **40**, 1141–1154.

Khil, P. P. & Camerini-Otero, R. D. (2002). Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. *Mol Microbiol* **44**, 89–105.

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

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. & Yanisch-Perron, C. R. (1981). Purified LexA protein is a repressor of the *recA* and *lexA* genes. *Proc Natl Acad Sci U S A* **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.

Makarova, K. S., Mironov, A. A. & Gelfand, M. S. (2001). Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome Biol* **2**, 131–138.

Monteiro, P. B., Teixeira, D. C., Palma, R. R., Garnier, M., Bové, J. M. & Renaudin, J. (2001). Stable transformation of the *Xylella fastidiosa* citrus variegated chlorosis strain with *oriC* plasmids. *Appl Environ Microbiol* **67**, 2263–2269.

Panina, E. M., Mironov, A. A. & Gelfand, M. S. (2001). Comparative analysis of Fur regulons in Gamma-proteobacteria. *Nucleic Acids Res* **29**, 5195–5206.

Rivera, E., Vila, L. & Barbé, J. (1996). The *uvrB* gene of *Pseudomonas aeruginosa* is not DNA damage inducible. *J Bacteriol* **178**, 5550–5554.

Rivera, E., Vila, L. & Barbé, J. (1997). Expression of the *Pseudomonas aeruginosa uvrA* gene is constitutive. *Mutat Res* **377**, 149–155.

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

Roy, S., Sahu, A. & Adhya, S. (2002). Evolution of DNA binding motifs and operators. *Gene* **285**, 169–173.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1992). *Molecular Cloning. A Laboratory Manual*, 2nd edn. 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 U S A* **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.

Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984). *Experiments With Gene Fusions*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Simpson, A. J. G., Reinach, F. C., Arruda, P. & 113 other authors (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151–159.

Tan, K., Moreno-Hagelsieb, G., Collado-Vives, J. & Stormo, G. D. (2001). A comparative genomics approach to prediction of new members of regulons. *Genome Res* **11**, 566–584.

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.

Voloshin, O. N., Ramirez, B. E., Bax, A. & Camerini-Otero, R. D. (2001). A model for the abrogation of the SOS response by an SOS protein: a negatively charged helix in DinI mimics DNA in its interaction with RecA. *Genes Dev* **15**, 415–427.

Walker, G. C. (1984). Mutagenesis and inducible responses to

deoxyribonucleic 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.

Yang, Y. C., Yang, M. K., Kuo, T. T. & Tu, J. (2001). Structural and functional characterization of the *lexA* gene of *Xanthomonas campestris* pathovar *citri*. *Mol Gen Genet* **265**, 316–326.

Received 4 June 2002; accepted 26 July 2002.

ARTICLE II

Mol Gen Genomics (2004) 271: 40–49
DOI 10.1007/s00438-003-0952-x

ORIGINAL PAPER

G. Mazón · J. M. Lucena · S. Campoy
A. R. Fernández de Henestrosa · P. Candau · J. Barbé

LexA-binding sequences in Gram-positive and cyanobacteria are closely related

Received: 30 June 2003 / Accepted: 29 October 2003 / Published online: 2 December 2003

Abstract The *lexA* gene of the cyanobacterium *Anabaena* sp. strain PCC7120 has been cloned by PCR amplification with primers designed after TBLASTN analysis of its genome sequence using the *Escherichia coli* LexA sequence as a probe. After over-expression in *E. coli* and subsequent purification, footprinting experiments demonstrated that the *Anabaena* LexA protein binds to the sequence TAGTACTAATGTTCTA, which is found upstream of its own coding gene. Directed mutagenesis and sequence comparison of promoters of other *Anabaena* genes, as well as those of several cyanobacteria, allowed us to define the motif RGTACNNNDGTWCB as the LexA box in this bacterial phylum. Substitution of a single nucleotide in this motif present in the *Anabena lexA* promoter is sufficient to enable it to bind the *Bacillus subtilis* LexA protein. These data indicate that Cyanobacteria and Gram-positive bacteria are phylogenetically closely related.

Keywords Lex A box · SOS response · Cyanobacteria · Bacterial phylogeny

Communicated by R. Devoret

G. Mazón · A. R. Fernández de Henestrosa · J. Barbé (✉)
Departament de Genètica i Microbiologia,
Universitat Autònoma de Barcelona, Bellaterra,
08193 Barcelona, Spain

J. M. Lucena · P. Candau
Instituto de Bioquímica Vegetal y Fotosíntesis,
Universidad de Sevilla-CSIC, 41092 Sevilla, Spain

S. Campoy
Institut de Recerca i Tecnologia Agroalimentària (UAB-IRTA),
Centre de Recerca en Sanitat Animal (CReSA),
Universitat Autònoma de Barcelona, Bellaterra,
08193 Barcelona, Spain

Present address: A. R. Fernández de Henestrosa
Centro de Biología Molecular Severo Ochoa (CSIC), Cantoblanco,
28049 Madrid, Spain

Introduction

Cyanobacteria are photoautotrophic prokaryotes that are able to grow in a wide range of environments, including freshwater, marine and terrestrial habitats. Under these diverse growth conditions cyanobacterial cells are exposed to a wide variety of DNA-damaging agents, such as UV irradiation and environmental chemicals. Despite their permanent exposure to these deleterious conditions, little is known about the DNA repair mechanisms in cyanobacteria. *Escherichia coli* is the bacterial species which has been best studied from this point of view. In *E. coli*, many genes have been identified which encode proteins required to repair different types of DNA damage or guarantee cell survival when the lesions are not eliminated. Several of these genes constitute specific DNA repair networks. One such regulon is under the control of the *lexA* gene and is known as the *E. coli* SOS system (Walker 1984). This system is made up at least 40 genes (Fernández de Henestrosa et al. 2000), including *lexA* and *recA*, the latter being the other regulatory gene of the network. The *E. coli* LexA protein specifically recognizes and binds an imperfect 20-bp palindrome whose consensus sequence is $CTG(TA)_5$ CAG, which has been designated as the *E. coli* SOS box (Walker 1984). In vivo and in vitro studies have shown the RecA protein is activated when it binds single-stranded DNA regions generated by DNA damage-mediated inhibition of replication (Sassanfar and Roberts 1990). Once activated, RecA facilitates the autocatalytic cleavage of LexA, resulting in the expression of genes that are normally repressed by this protein (Walker 1984). Hydrolysis of the *E. coli* LexA is mediated by its Ser119 and Lys156 residues, and it occurs by a mechanism similar to that used during proteolysis by serine proteases (Luo et al. 2001). After DNA repair, the RecA protein is no longer activated, and the level of the LexA protein increases, allowing it to once again repress the genes which are under its negative control.

Among cyanobacteria, the presence of a *recA* -like gene has been demonstrated for *Anabaena* PCC7120, *Synechococcu* s PCC7002 and *Synechocystis* PCC6308 (Geoghegan et al. 1987; Murphy et al. 1987; Owttrim and Coleman 1987). However, there is no information available on the LexA regulon in this bacterial phylum. In this respect, significant variation exists with regard to the core genes of this regulon in the Domain Bacteria. Thus, some bacterial species do not have a *lexA* -like gene (*Aquifex aerolicus, Borrelia burgdorferi, Chlamydia pneumoniae, Mycoplasma pneumoniae, Campylobacter jejuni, Helicobacter pylori, Porphyromonas gingivalis* and *Chlorobium tepidum*), whereas in others in which such a gene occurs (*Deinococcus radiodurans, Geobacter sulfurreducens* and *Dehalococcoides ethenogenes*) the *recA* gene is not under its direct control (Narumi et al. 2001; Bonacossa de Almeida et al. 2002; Fernández de Henestrosa et al. 2002; Jara et al. 2003). Likewise, there is significant variability among the LexA-binding boxes in the different bacterial phylogenetic groups. Some of them, such as Gram-positive bacteria and Alpha Proteobacteria, are very homogenous, each class preserving the same LexA box: CGAACRNRYGTTYC, and $GTTCN_7GTTC$, respectively (Winterling et al. 1998; Tapias and Barbé 1999). In contrast, in other groups like the Gamma Proteobacteria, two different LexA boxes have been described: thus both the *E. coli* -like box and the sequence $TTAGN_6TACTA$ are found in Xanthomonadaceae (Campoy et al. 2002). Obviously, a negative regulator gene and its DNA recognition sequence must together follow a convergent evolutionary path if the control mechanism mediated by this gene is to be conserved. For this reason, the existence of significant variability in the sequence of the LexA-binding motif in a given bacterial class or phylum is a clear indication that that such a set is not a coherent phylogenetic group. Since the complete sequence of the *Anabaena* sp. PCC7120 genome is now available (http://www.kazusa.or.jp/cyano), in the present work we have isolated its *lexA* gene and, after purification of its product, the DNA sequence to which this LexA protein binds has been identified. The information obtained during this investigation has also allowed us to analyse the homogeneity of the LexA network in cyanobacteria, and to establish the relationship of its regulatory sequence to those of other bacterial phylogenetic groups.

Materials and methods

Bacterial strains, plasmids, oligonucleotides and DNA techniques

Bacterial strains and plasmids used in this work are listed in Table 1. *E. coli* strains were grown at 37°C in LB medium (Miller 1992). *Synechocystis* and *Anabaena* cells were phototrophically cultured in BG11 medium, as described by Chávez et al. (1995). Antibiotics were added to the cultures at the concentrations recommended by Sambrook and Russell (2001) . *E. coli* cells were transformed with plasmid DNA as described previously (Sambrook and Russell 2001). All restriction enzymes, oligonucleotide primers, T4 DNA ligase and polymerase, and the DIG-DNA Labelling and Detection Kit were purchased from Roche Diagnostics. DNA was isolated from *Synechocystis* and *Anabaena* cells as described by Chávez et al. (1995). The synthetic oligonucleotide primers used for PCR amplification are listed in Table 2. To facilitate subcloning of some DNA fragments obtained by PCR, specific restriction sites were incorporated into the oligonucleotide primers. Mutant derivatives of the *Anabaena lexA* promoter were constructed by PCR mutagenesis, using oligonucleotides carrying specific substitutions. The DNA sequences of all-mutagenized fragments were determined by the dideoxy method (Sanger et al. 1977) on an ALF Sequencer (Amersham Biosciences). In all cases the entire nucleotide sequence was determined on both DNA strands.

Table 1 Bacterial strains and plasmids used in this work

Strain/plasmid	Relevant features	Source/ reference
Anabaena sp. PCC7120	Wild type	Frías et al. (1997)
Synechocystis sp. PCC6803	Wild type	Chávez et al. (1995)
Escherichia coli		
DH5α	*supE44*F$^-$Δ*lacU169* (φ80 *lacZ* ΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Clontech
BL21(DE3) Codon Plus	F$^-$*ompT hsdS* $_B$ (r_B-r_B)*dcm gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^R]	Stratagene
Plasmids		
pGem-T	PCR cloning vector, Ap^R	Promega
pET15b	Overexpression vector carrying a N-terminal tag containing six histidine residues; Ap^R	Novagen
pUA1050	A pGemT derivative carrying a 839-bp PCR fragment amplified with LexA up Ab and LexA BamHI (dw) primers containing the promoter region and the coding region of *Anabaena lexA* gene	This work
pUA1051	A pGemT derivative carrying a 665-bp PCR fragment amplified with LexA NdeI and LexA BamHI primers containing the*Anabaena lexA* coding region	This work
pUA1052	A pET15b derivative carrying a 665-bp*Nde* I-*Bam*HI fragment containing the *Anabaena lexA* coding region	This work

Molecular cloning of the *Anabaena lexA* gene, and purification of its protein product

The *lexA* gene was amplified from the genomic DNA of the *Anabaena* sp. strain PCC7120 using the oligonucleotide primers LexA up and LexA BamHI (Table 2), corresponding to nucleotides −174 to −156 and +665 to +647 with respect to the proposed translational start point. The 839-bp PCR fragment so obtained was cloned into the pGEM-T vector, to generate the plasmid pUA1050. To confirm that no mutations had been introduced during the amplification reaction, the sequence of this fragment was determined directly. Plasmid pUA1050 was used as a template for PCR with the primers LexA NdeI and LexA BamHI (Table 2), and the fragment obtained was again cloned in pGEM-T. The putative

Table 2 Oligonucleotide primers used in this work

Primer	Sequence (5' to 3')	Position[c]	Application
LexA up Ab	GCTTGCTAGTATCTTAGG	–174	Upper primer used to clone the *Anabaena lexA* gene, and to obtain the FrgL1 probe for EMSAs and footprinting assays
LexA low Ab	GGATATATTCTGCCAGCC	+58	Lower primer used to clone the*Anabaena lexA* promoter for EMS competition assays and footprinting assays
LexA NdeI	*CATATG*GAACGCCTAACAGAAGC	+1	Upper primer used to clone the *Anabaena lexA* ORF in the pET15b vector
LexA BamHI	*GGATCC*AGGATGAGGGAAACAAGG	+665	Lower primer used to clone the *Anabaena lexA* gene
-74 lexA Ab	TTTATCTTGTTATTAGAG	–74	Upper primer to obtain FrgL2 probe for EMSA.
-53 lexA Ab	TTTTGTAGTACTAATGTTC	–53	Upper primer used to obtain FrgL3 as a probe for EMSAs
-34 lexA Ab	TAGGGACATTTATATTTATC	–34	Upper primer used to obtain the FrgL4 probe for EMSAs
DIG Low lexA Ab	DIG-TCATCTGACGAATTGAGG	+94	Lower primer used to obtain wild-type and mutagenized *lexA* probes (5' end-labeled with digoxigenin)
RecA up Ab	CACCTGGGGAATATTGAG	–149	Upper primer used to clone the*Anabaena recA* promoter for EMSAs
RecA dw Ab	ATGATTGCTCCTTTACCG	+92	Lower primer used to clone the*Anabaena recA* promoter for EMSAs
DIG RecA dw Ab	DIG-ATGATTGCTCCTTTACCG	+92	Lower primer used to clone the*Anabaena recA* promoter for EMSA (5'-end labeled with digoxigenin)
Mut Ab/Bs	TTTTGTAG<u>a</u>ACTAATGTTC	–53	Upper primer used to obtain the mutagenized *Anabaena lexA* probe containing a single substitution for EMSAs with *B. subtilis* LexA

[a]When present, added restriction sites are shown in *italics*
[b]Nucleotide changes are shown in *lowercase and underlined*
[c]Position of the 5' end of the oligonucleotide with respect to the proposed translational start point of each gene

ATG initiation triplet of LexA is part of the *Nde* I restriction site of the LexA NdeI primer, which enables IPTG-inducible over-expression of the *Anabaena* LexA protein. After digestion with *Nde* I and *Bam* HI, the *lexA* gene was cloned downstream of the T7 promoter in the pET15b expression vector (Novagen), which encodes an N-terminal tag containing six histidine residues, according to standard procedures (Sambrook and Russell 2001), and the ligation mix was transformed into competent DH5α cells. After confirming by DNA sequencing that no mutations had been introduced into the *lexA* gene contained in the pUA1052 plasmid, the plasmid was transformed into the *E. coli* BL21(DE3) Codon Plus strain for over-expression of the LexA protein. The LexA-His-tag fusion protein was purified using the Talon Metal Affinity Resin Kit (Clontech) as reported previously (Campoy et al. 2002).

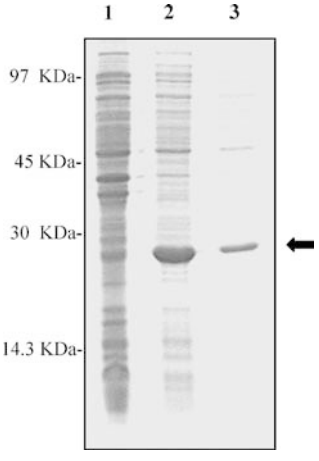


Fig. 1 Overproduction and purification of *Anabaena* LexA in *E. coli*. Lanes 1 and 2 contain crude extracts of *E. coli* BL21(DE3)/pUA1052 cells grown in the absence and presence of IPTG, respectively. The purified His-tagged *Anabaena* LexA protein is shown in lane 3. The molecular masses of protein markers are indicated on the *left*

The *Anabaena* LexA protein so obtained was more than 95% pure, as determined by Coomassie Blue staining of SDS-PAGE (15%) polyacrylamide gels (Fig. 1).

Mobility shift assays and DNase I footprinting

Anabaena LexA-DNA complexes were detected by electrophoresis mobility shift assays (EMSAs) using purified *Anabaena* LexA protein as described previously (Campoy et al. 2002). DNA probes were prepared by PCR amplification in which one of the primers was labelled at its 5′ end with digoxigenin (DIG) (Table 2). All EMSAs were performed a minimum of three times to ensure reproducibility of the results. Densitometric analysis of the retardation experiments was performed using the Kodak Digital Science 1Dprogram. The binding activities reported were calculated as described by Tapias and Barbe (1998) and are expressed as the percentage of the total labelled DNA that was bound by the target protein. For every experiment, wild-type DNA was included in the set of probes used, as an internal control. Binding of the wild-type *lexA* promoter fragment to the *Anabaena* LexA protein was taken as 100%, and the binding of this protein to mutant *lexA* promoters, or of LexA proteins from *Bacillus subtilis* and *D. ethenogenes* to corresponding DNA fragments, was expressed as a percentage of this value. The data presented are the means of three independent experiments, and values were reproducible to within ±10%. DNase I footprinting assays were performed using the ALF Sequencer (Amersham Biosciences) as described previously (Patzer and Hantke 2001; Campoy et al. 2003).

Results and discussion

Identification of the binding site for LexA from *Anabaena* PCC7120

Electrophoretic mobility shift assays (EMSAs) with the purified *Anabaena* PCC7120 LexA were used to deter-

mine the capacity of this protein to bind to the promoter of its own structural gene. The addition of increasing concentrations of LexA to a fragment extending from position −174 to +94 of the *Anabaena lexA* gene (with respect to its putative translational start point) produces one retarded band whose intensity is directly related to the amount of protein used (Fig. 2A). Furthermore, the formation of the *Anabaena lexA* promoter-LexA complex was sensitive to competition with an excess of the unlabelled FrgL1 fragment, but not by an excess of nonspecific DNA (plasmid pBSK+) (Fig. 2B). To localize the binding region, the *Anabaena lexA* promoter was divided into four fragments (designated as FrgL1–4) obtained by PCR amplification with the DIG-end

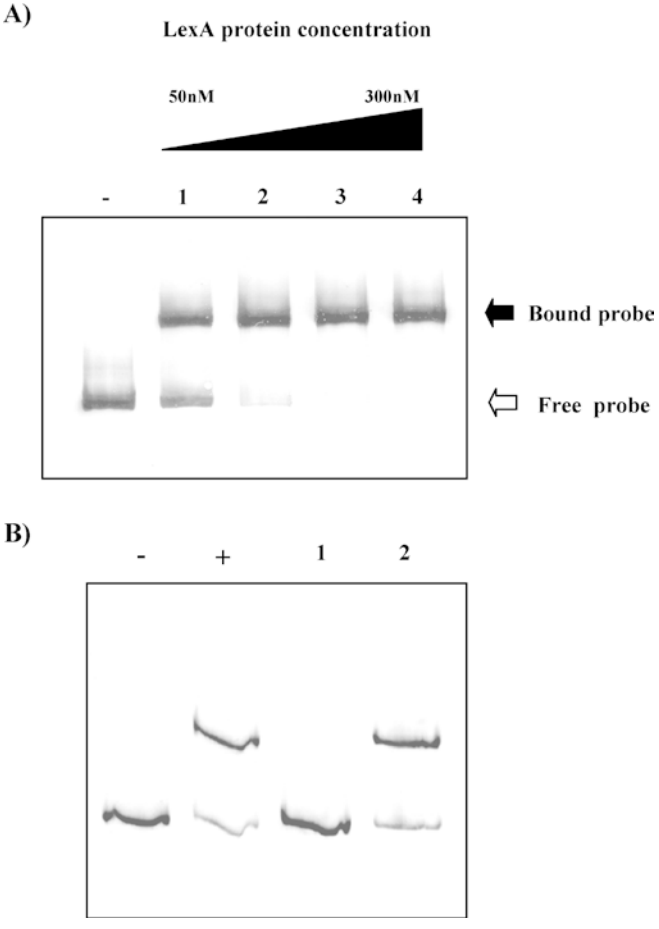


Fig. 2A, B Analysis of the electrophoretic mobility of a DNA fragment containing the *Anabaena lexA* promoter in the presence of increasing concentrations of purified *Anabaena* LexA protein. **A** In the binding reactions, the concentrations of purified LexA protein used were 50 nM (lane 1), 100 nM (lane 2), 150 nM (lane 3) and 300 nM (lane 4). The mobility of the *Anabaena lexA* promoter in the absence of purified LexA protein is also shown as a control in the *leftmost* lane. **B.** Effect of a 300-fold molar excess of unlabelled *Anabaena lexA* promoter (lane 1) or pBSK(+) plasmid DNA (lane 2) on the mobility of this DNA fragment in the presence of purified *Anabaena* LexA protein (150 nM). The electrophoretic mobility of the *Anabaena lexA* promoter in the absence (-) or in the presence (+) of 150 nM purified *Anabaena* LexA protein is also shown as a control

labelled oligonucleotides listed in Table 2 (Fig. 3). A stable DNA-protein complex was observed when either FrgL1, FrgL2 or FrgL3 was incubated with the *Anabaena* LexA protein, but the mobility of FrgL4 was unaffected under the same conditions (Fig. 3). These results indicate that the *Anabaena* LexA protein specifically binds to the promoter of the *lexA* gene in a region which includes the segment between positions –53 and –34, with respect to the translation start site.

The sequence to which the *Anabaena* LexA protein binds was localized precisely through footprinting experiments with the FrgL1 fragment. The results obtained showed that a core region of 16 nt (TAGTACTAATGTTCTA) was protected when both *lexA* coding and noncoding strands were analyzed (Fig. 4). Visual inspection of the upstream region of the *recA* gene of *Anabaena* PCC7120 also revealed the presence of the sequence TAGTATATCTGTTCTA (Fig. 5A), which is practically identical to that which is protected in a footprinting experiment with the *lexA* promoter (Fig. 4). This fact suggested that the *Anabaena* PCC7120 *recA* promoter can also bind the endogenous LexA protein. To confirm this, an EMSA experiment was performed using the *recA* promoter as either a direct or competitive probe. Figure 5B shows that the *recA* promoter is indeed able to bind the LexA protein specifically. Furthermore, formation of the LexA-*recA* labelled promoter complex is abolished when an unlabelled *lexA* promoter fragment is included in the binding reaction (Fig. 5B). These results demonstrate that the *Anabaena* PCC7120 LexA protein binds to a motif that is common to the *recA* and *lexA* promoters.

To confirm the footprinting results, as well as to more precisely define the LexA-binding motif in *Anabaena* PCC7120, the effect of systematic mutagenesis of each one of the nucleotides in the TAGTACTAATGTTCTA sequence upon the electrophoretic mobility of the FrgL3 fragment was analysed in EMSA experiments. Figure 6 demonstrates that some of these bases seem not to be involved in the recognition of the *Anabaena* PCC7120 LexA protein. Based on the effects of mutations at each site in the complete sequence, the consensus sequence RGTACNNNDGTWCB was derived. Mutations at each of the defined positions in this sequence reduced the binding of LexA by more than 75% with respect to the wild-type version.

Presence of the *Anabaena* PCC7120 LexA box in other cyanobacteria

Once the *Anabaena* PCC7120 LexA box had been identified, a search was carried out with the RCGScanner program (Erill et al. 2003), which uses a recursive method to identify a given nucleotide sequence as well as its neighboring ORFs, in the genome sequence of this organism in order to find genes with which this LexA-binding sequence was associated. Besides *lexA* and *recA*, another four genes were found (*uvrA, ssb, all4790* and

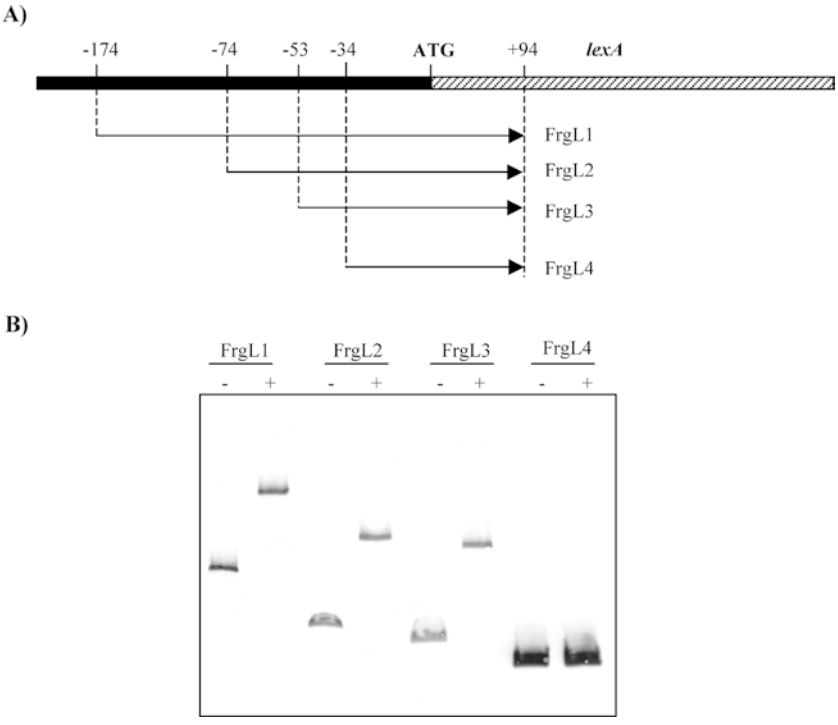


Fig. 3A, B Analysis of the binding of defined fragments of the *Anabaena lexA* promoter to purified *Anabaena* LexA protein. **A** Diagram depicting the fragments amplified from the *Anabaena lexA* promoter for use in gel retardation experiments. FrgL1, FrgL2, FrgL3 and FrgL4 were generated by PCR-amplification with appropriate oligonucleotide primers. In all cases, positions are indicated with reference to the translational start point of the *Anabaena lexA* gene. **B** Electrophoretic mobility of FrgL1, FrgL2, FrgL3 and FrgL4 fragments after incubation in the absence (-) and presence (+) of 150 nM purified *Anabaena* LexA protein

alr4905) which harbored a putative *Anabaena* -like LexA box upstream of their respective coding regions (Table 3). All these sequences were able to prevent complex formation between the *Anabaena* LexA protein and its promoter when used as unlabelled competitors in EMSA experiments (Fig. 7). These data imply that the core of the *Anabaena* LexA regulon includes some genes which are common to the core set in *E. coli* (*recA*, *uvrA* and *ssb*). However, and despite the fact that other genes which are under the direct control of the LexA protein in *E. coli* are also present in *Anabaena* (*uvrB, ruvA, sulA, ftsK* and *recN*), none of these has the *Anabaena* LexA box in its upstream region. This absence of LexA boxes in the promoters of genes which belong to the LexA regulon in *E. coli* has also been reported for other bacterial species, such as *D. ethenogenes, Mycobacterium tuberculosis* and several Gamma Proteobacteia (Campoy et al. 2002; Davis et al. 2002; Fernández de Henestrosa et al. 2002; Erill et al. 2003). The reasons for this difference in the composition of the core set of genes regulated by LexA have not yet been elucidated, although they may be related either to the environmental characteristics of each one of these organisms or to particular aspects of their evolution. It is worth noting that in the cyanobacterium *Nostoc punctiforme* the *Anabaena* LexA box is found upstream of five genes (*lexA, recA, uvrA, ssb* and *npun* 5119) whose homologs also bind the LexA protein in *Anabaena* (Table 3). This indicates that in both cyanobacteria the gene core of the LexA regulon are practically identical. However, in *Synechocystis* PCC6803, an *Anabaena* -like LexA box is only found upstream of the *lexA* gene, whereas it is absent in all the other genes which display this motif in *Anabaena* and *N. punctiforme*. Moreover, the LexA box present upstream of the *Synechocystis lexA* gene is functional, because EMSA experiments have demonstrated that it binds to the *Anabaena* LexA protein (data not shown).

Alignment of the sequences of all the *Anabaena* -like LexA boxes found upstream of all genes listed in Table 3 reveals the presence of the common motif AGTACNNNNGTTC, which is compatible with the LexA binding site of *Anabaena* (RGTACNNNDGTWCB) defined by directed mutagenesis (Fig. 6). Because no complete chromosomal sequence is yet available for any other cyanobacteria, it was not possible to search for the presence of this motif in their genomes. Nevertheless, using TBLASTN to search the sequence database maintained by the Kazusa Research Institute (http://www.kazusa.or.jp/cyano/blast.html), the presence of a *lexA* gene in several cyanobacteria has been confirmed. Alignment of the sequences of their protein products reveals a high degree of similarity among their N-terminal ends, in which the helix-turn-helix motif responsible for recognition of the LexA box (Luo et al. 2001) is localized (Fig. 8). In agreement with this striking similarity, it is worth noting that an *Anabaena* -like LexA-binding site is also found upstream of the *lexA* genes of *Synechococcus* sp. WH8102, *Prochlorococcus marinus* MIT9313 and *Prochlorococcus pastoris* CCMP1378 (data not shown). This emphasizes that the RGTACNNNDGTWCB motif can be considered as the cyanobacterial LexA box. Moreover, comparison of all of these LexA protein sequences reveals that the *Synechocystis* PCC6803 LexA shows some differences in the residues involved in the autohydrolysis process. Thus, the Ala residue of the Ala-Gly bond at which the hydrolysis of LexA is expected to take place has been replaced by a Gly, and the distance between this Gly-Gly bond and the other reactive residues (Ser and Lys) is modified in the *Synechocystis* PCC6803 LexA protein (Fig. 8). In this context,

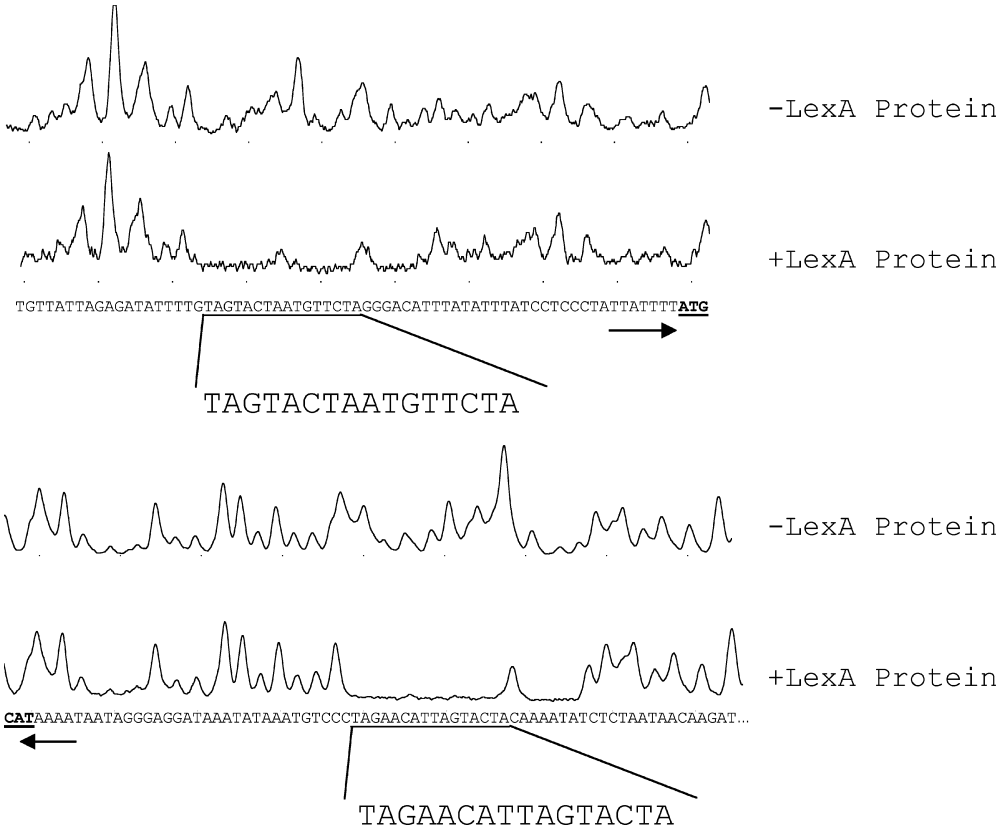


Fig. 4 DNase I footprinting assays carried out with coding and non-coding Cy5-labelled strands of the DNA fragment containing the *Anabaena lexA* promoter, obtained by PCR using the primer pair UpLexAAb/LowLexAAb, in the absence or presence of 200 nM LexA protein purified from this organism. The *arrows* indicate the translational direction of each strand. For these, the translational starting codon is shown in *bold and underlined*

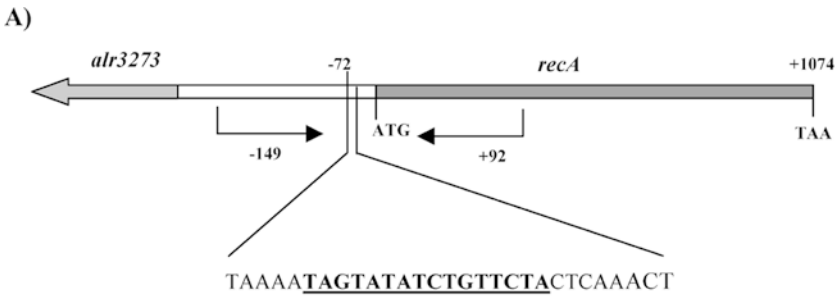
it is tempting to speculate that these modifications may perhaps negatively affect the autocatalytic cleavage of this LexA repressor, and that this is the reason why no other genes in *Synechocystis* PCC6803 have the cyanobacterial LexA box. In fact, the reduction in the number of genes directly regulated by LexA in *Synechocystis* PCC6803, in comparison to *Anabaena* and *N. punctiforme*, is compatible with this hypothesis. In the promoters of these *Synechocystis* PCC6803 genes, a progressive mutational process could have occurred up to the point at which the LexA-binding sequence was definitively lost. A similar process seems to have affected one of the two *recA* genes in the bacterium *Myxococcus xanthus* (Campoy et al. 2003).

Alteration of a single base in the *Anabaena* LexA-binding sequence confers the ability to bind the *B. subtilis* LexA protein

A detailed analysis of the cyanobacterial LexA box (RGTACNNNDGTWCB) reveals a close relationship with the Gram-positive LexA binding site (CGAACRNRYGTTYC), implying that the introduction of slight changes in the cyanobacterial LexA box could enable it to bind the LexA protein of *B. subtilis*. To test this hypothesis, the T in the GTAC tetranucleotide of the cyanobacterial LexA box present upstream of the *Anabaena lexA* gene was replaced by an A (GTAC → GAAC). Figure 9 shows that this single substitution is indeed sufficient to enable the *Anabaena lexA* promoter to bind the *B. subtilis* LexA protein effectively.

It has recently been reported that the green-non-sulphur bacterium *D. ethenogenes*, which is a Gram-negative organism, has the same LexA box as Gram-positive bacteria (Fernández de Henestrosa et al. 2002). In agreement with this, the *D. ethenogenes* LexA protein binds more tightly to the mutant (GAAC) *Anabaena lexA* promoter than to the wild-type sequence (Fig. 9). It should also be noted that the *Anabaena* LexA protein is itself able to bind to the mutant *lexA* promoter fragment, although with a lower efficiency (Fig. 9).

The N-terminal domain of the *E. coli* LexA has three α helices, which are necessary for its binding to DNA (Fogh et al. 1994; Knegtel et al. 1995; Luo et al. 2001). The characterization of mutants in which some of the amino acids in these regions have been altered suggests that the first two α helices are needed for maintenance of the overall structure of the active molecule. Moreover, residues S39, N41, A42, E44 and E45 in the third α helix of the *E. coli* LexA have been proposed to particip‘ate in its specific interaction with the CTGT motif in the LexA box of this organism (Knegtel et al. 1995). An alignment



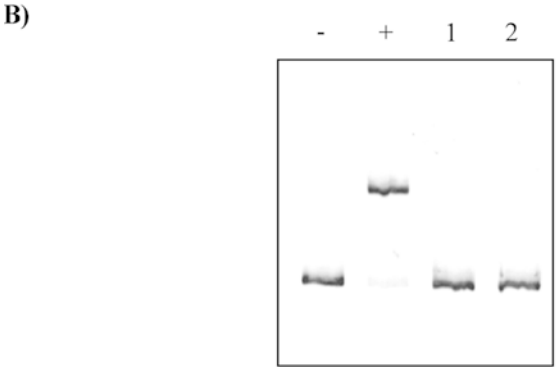


Fig. 5A, B Analysis of the binding of the LexA protein from *Anabaena* to the chromosomal region containing the *recA* gene of *Anabaena*. **A** Schematic diagram of the *recA* region. The 16-bp region found upstream of the *recA* gene whose sequence is practically identical to the motif in the *lexA* promoter that is protected from DNase I by LexA (see Fig. 4) is shown in *bold and underlined*. Positions are indicated with reference to the translational start point of the *Anabaena recA* gene. **B** Electrophoretic mobility of the *Anabaena recA* promoter in the absence (-) or presence (+) of 150 nM purified *Anabaena* LexA protein. The effect of added unlabelled *lexA* (lane 1) or *recA* promoter fragment (lane 2) on the electrophoretic mobility of the labelled *Anabaena recA* promoter in the presence of 150 nM purified *Anabaena* LexA protein is shown

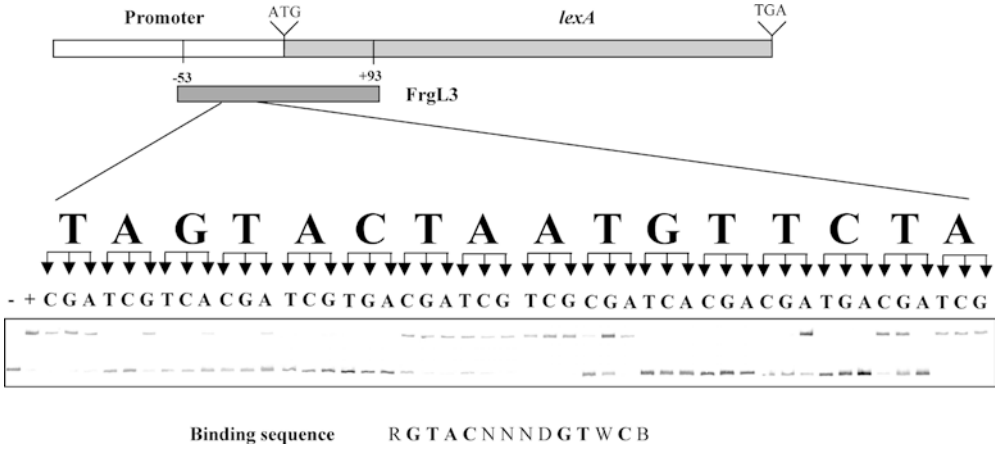


Fig. 6 Effect of single-nucleotide substitutions in the TAGTACTAATGTTCTA sequence on the electrophoretic mobility of the FrgL3 fragment in the presence of 150 nM purified *Anabaena* LexA. The mobility of the wild-type FrgL3 fragment (wt) in the absence (-) or presence (+) of LexA is also shown. The binding sequence is defined as those positions at which alteration of the wild-type sequence causes a ≥75% decrease in the binding ratio, relative to the wild-type value. The binding ratio of the LexA protein was determined in each case as described in Materials and methods. Data are the means of three independent experiments, and are reproducible to within ±10%

of the LexA proteins of *E. coli, B. subtilis, D. ethenogenes* and *Anabaena* (Fig. 10) indicates the presence in all four proteins of these three α helices, and reveals that the Ser at residue 39 in the *E. coli* protein is totally conserved. This fact strongly suggests that this residue is unlikely to be involved in the specific recognition of the LexA box. Furthermore, in the LexA proteins of *B. subtilis* and *D. ethenogenes*, the amino acids at positions corresponding to N41 and A42 in the *E. coli* LexA are strongly conserved (Fig. 10). Moreover, the level of conservation of this same region decreases when the LexA proteins of *Anabaena*, *B. subtilis* and *D. ethenogenes* are compared. This is in agreement with the fact that the LexA boxes of Gram-positive and cyanobacteria are related, but they are not the same (Fig. 9). However, it should be mentioned that the amino acids corresponding to E44 and E45 in the *E. coli* protein are not conserved between *B. subtilis, D. ethenogenes* and *Anabaena* (Fig. 10). This suggests that the role of these residues in the interaction of the LexA proteins of these three bacteria with their own LexA boxes may not be as critical as in the case of *E. coli*. Furthermore, the conservation of *E. coli* residues L47, L50, G54 and I56 in the

Table 3 LexA boxes associated with various genes in cyanobacteria

Species	Gene	Position[a]	Sequence
Anabaena sp. PCC7120	*lexA*	–35	TAGTACTAATGTTCTA
	recA	–53	TAGTATATCTGTTCTA
	uvrA	–71	CAGTACTATTGTTCTA
	ssb	–15	AAGTACTTATGTACTA
	alr4905 (homologous to two component *his* regulator)	–99	TAGTTCTCATGTACTA
	alr4790	–32	GCGTACATTTGTACCA
Nostoc punctiforme ATCC29133	*lexA*	–32	TAGTACTAATGTTCTA
	recA	–61	TAGTATATCTGTTCTA
	uvrA	–70	CAGTACGATTGTTCTA
	npun5119 (homologous to*alr4790*)	–35	TCGTACATTTGTACTA
	npun2201 (homologous to*ssb*)	–15	AAGTACACCTGTACTG
Synechocystis sp. PCC6803	*lexA*	–28	TAGTCCTAGAGTCCTA
Consensus sequence[b]			T**AGTAC**TWAT**GTTC**TA
In vitro binding sequence[c]			N**RGTAC**NNND**GTWC**BN

[a]Position of the end of the LexA box with respect to the proposed translational start point
[b]Consensus derived from the alignment of the LexA box motifs present upstream of the different genes
[c]Binding sequence defined on the basis of in vitro mutagenesis studies on the *Anabaena* sp. PCC7120*lexA* promoter. In accordance with the standard nomenclature, R = A or G, D = A or G or T, W = A or T and B = C or G or T

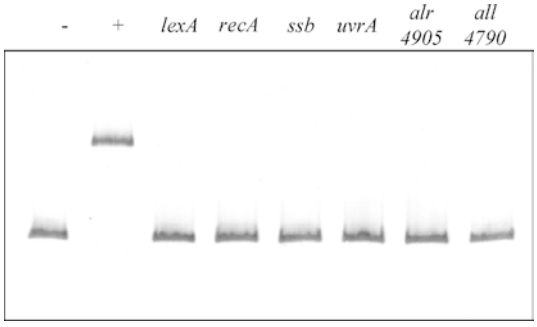


Fig. 7 Electrophoretic mobility of the FrgL3 fragment in the presence of 150 nM *Anabaena* LexA protein and a 300-fold molar excess of unlabelled fragments comprising about 400 bp of the upstream regions of the genes *ssb, uvrA, alr4905* and *allr4790* of *Anabaena*. As positive controls, the effect of unlabelled *lexA* and *recA* promoter fragments on the mobility of the labelled FrgL3 fragment in the presence of the same amount of LexA protein is presented. The mobility of the FrgL3 fragment in the absence of any additional DNA (+) or purified LexA protein (-) is also shown

Fig. 8 CLUSTAL W alignment of LexA protein sequences from *Nostoc puctiforme* (Np), *Anabaena* sp strain PCC7120 (Ab), *Synechococcus* sp WH8102 (Scc), *Prochlorococcus marinus* MIT9313 (Pm), *Prochlorococcus pastoris* CCMP1378 (Pp) and *Synechocystis* sp PCC6803 (Sy). The conserved Ala-Gly bond at which the LexA protein should be hydrolyzed, and the Ser and Lys residues involved in this process, are indicated by the *arrows*. The modifications in these residues or their positions found in the *Synechocystis* LexA protein are indicated in *bold and boxed*. The sequences of several of the LexA proteins were obtained by TBLASTN search in the http://www.kazusa.or.jp/cyano/blast.html database using the *Anabaena* LexA sequence as a query

four LexA proteins would indicate that these positions are not involved in the specificity of DNA recognition, but rather in the maintenance of the structure of the third helix.

Finally, all of these data indicate that Cyanobacteria and Gram-positive phyla are closely related. In this respect, our results are in concordance with a recently proposed hypothesis regarding the branching order of the bacterial evolutionary tree, which has been suggested on the basis of comparative in silico analysis of either insertions or deletions of specific amino acids in particular proteins (Gupta and Griffiths 2002). According to this hypothesis, Gram-positive bacteria branched from the common ancestor before green non-sulphur bacteria, which also branched immediately before cyanobacteria. Our data on the differential affinity of the *B. subtilis, D. ethenogenes* and *Anabaena* LexA proteins for either the wild-type or the modified *Anabaena lexA* promoter are in agreement with this prediction. It is worth noting that, to our knowledge, this work is the first study in which the functional identification of a regulatory sequence has allowed elucidation of the order of appearance of the branches of a bacterial phylogenetic group. Determination of new bacterial LexA boxes may thus be very useful in understanding the evolutionary trajectories of other phyla in the Domain Bacteria.

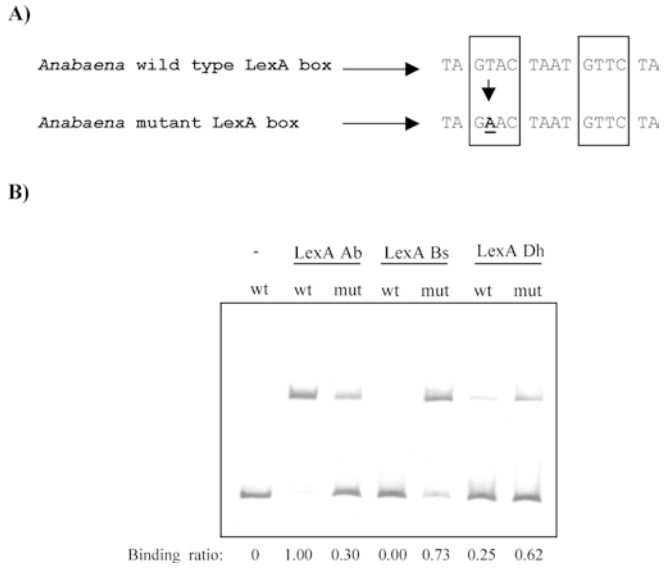


Fig. 9A, B Generation (**A**) and analysis (**B**) of a putative *B. subtilis* LexA-binding site in the *Anabaena lexA* promoter. **A** A single-point mutation was introduced into the *Anabaena* LexA-binding sequence, and wild-type and mutant fragments were analyzed for the ability to bind LexA proteins from *Anabaena* and *B. subtilis*. **B**. Electrophoretic mobility of the wild-type (wt) and mutant (mut) *Anabaena lexA* promoter following incubation in the absence (-) or presence of purified LexA proteins (150 nM) from *Anabaena* (Ab), *B. subtilis* (Bs) or *D. ethenogenes* (Dh). The binding ratio of each LexA protein to either the wild-type or the mutant *Anabena lexA* promoter is indicated, and was determined as described in Materials and methods. Data are the means of three independent experiments, and are reproducible to ±10%

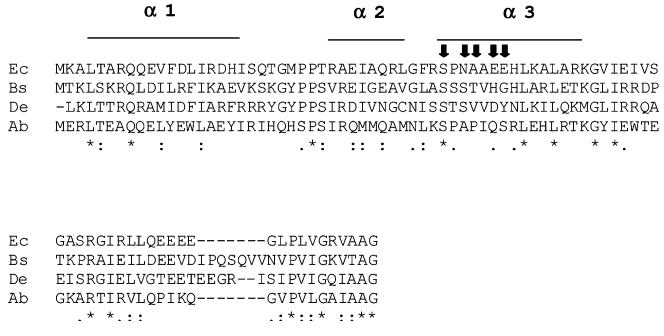


Fig. 10 CLUSTAL W alignment of the N-terminal domains of the LexA proteins of *E. coli* (Ec), *B. subtilis* (Bs), *D. ethenogenes* (De) and *Anabaena* sp. PCC7120 (Ab). The three predicted α helices in these four LexA proteins are indicated (α1, α2 and α3). The residues of the α3 helix of the *E. coli* LexA protein, which has been shown to be involved in the specific recognition of the LexA box (S39, N41, A42, E44 and E45), are marked with an *arrow*

Acknowledgements This work was funded by Grants BMC2001-2065 from the Ministerio de Ciencia y Tecnología (MCyT) de España, 2001SGR-206 from the Departament d'Universitats, Recerca i Societat de la Informació (DURSI) de la Generalitat de Catalunya and CVI-112 from the Junta de Andalucía. Antonio R. Fernández de Henestrosa and Gerard Mazón were recipients of a postdoctoral contract and a predoctoral fellowship from the MCyT, respectively. We are deeply indebted to Dr. Roger Woodgate for his generous gift of the *B. subtilis* LexA protein. We also wish to acknowledge Dr. Pilar Cortés, Joan Ruiz and Susana Escribano for their excellent technical assistance. We are grateful to Prof. R. S. Gupta and Prof. R. Guerrero for their critical reading of the manuscript

References

Bonacossa de Almeida C, Coste G, Sommer S, Bailone A (2002) Quantification of RecA protein in *Deinococcus radiodurans* reveals involvement of RecA, but not LexA, in its regulation. Mol Genet Genomics 268:28–41

Campoy S, Mazón G, Fernández de Henestrosa AR, Llagostera M, Brant-Monteiro P, Barbé J (2002) A new regulatory DNA motif of the gamma subclass *Proteobacteria*: identification of the LexA protein binding site of the plant pathogen *Xylella fastidiosa*. Microbiology 148:3583–3597

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

Chávez S, Reyes JC, Chauvat F, Florencio FJ, Candau P (1995) The NADP-glutamate dehydrogenase of the cyanobacterium *Synechocystis* PCC6803: cloning, transcriptional analysis and disruption of the *gdhA* gene. Plant Mol Biol 28:173–188

Davis EO, Dullaghan EM, Rand L (2002) Definition of the mycobacterial SOS box and its use to identify LexA-regulated genes in *Mycobacterium tuberculosi* s. J Bacteriol 184:3287–3295

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

Fernández de Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, 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 AR, Cuñé J, Erill I, Magnuson JK, 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 RH, 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

Frías JE, Flores E, Herrero A (1997) Nitrate assimilation gene cluster from the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC7120. J Bacteriol 179:477–486

Geoghegan CM, Houghton JA (1987) Molecular cloning and isolation of a cyanobacterial gene which increases the UV and methyl methanesulphonate survival of *recA* strains of *Escherichia coli* K12. J Gen Microbiol 133:119–126

Gupta RS, Griffiths E (2002) Critical issues in bacterial phylogeny. Theor Popul Biol 61:423–434

Jara M, Núñez C, Campoy S, Fernández de Henestrosa AR, Lovley DR, Barbé J (2003) *Geobacter sulfurreducens* has two autoregulated *lexA* genes whose products do not bind the *recA* promoter: differing responses of *lexA* and *recA* to DNA damage. J Bacteriol 185:2493–2502

Knegtel RMA, Fogh RH, 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

Luo Y, Pfuetzner RA, Mosimann S, Paetzel M, Frey EA, Cherney M, Kim B, Little JW, Strynadka NC (2001) Crystal structure of LexA: a conformational switch for regulation of self-cleavage. Cell 106:585–594

Miller JH (1992) A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Murphy RC, Bryant DA, Porter RD, Tandeau de Marsac N (1987) Molecular cloning and characterization of the *recA* gene from the cyanobacterium *Synechoccus* sp. strain PCC7002. J Bacteriol 169:2739–2747

Narumi I, Satoh K, Kikuchi M, Funayama T, Yanagisawa T, Kobayashi Y, Watanabe H, Yamamoto K (2001) The LexA protein from *Deinococcus radiodurans* is not involved in RecA induction following gamma irradiation. J Bacteriol 183:6951–6956

Owttrim GW, Coleman JR (1987) Molecular cloning of a *recA* -like gene from the cyanobacterium *Anabaena variabilis*. J Bacteriol 169:1824–1829

Patzer SI, 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

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual (3rd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

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

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

Tapias A, Barbé J (1998) Mutational analysis of the *Rhizobium etli recA* operator. J Bacteriol 180:6325–6331

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

Walker GC (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol Rev 48:60–93

Winterling KW, Chafin D, Hayes JJ, Sun J, Levine AS, Yasbin RE, Woodgate R (1998) The *Bacillus subtilis* DinR binding site: redefinition of the consensus sequence. J Bacteriol 180:2201–2211

ARTICLE III

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 two different perfect palindromes were carried out. The binding ratio of the LexA protein to both the TGCAC-N4-GTGCA palindrome and the wild type sequence was practically the same. In contrast, there was a substantial reduction (about 75%) in the LexA binding ratio to the TGCCC-N4-GGGCA palindrome. This demonstrates 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 LexA proteins from several bacterial groups was performed. The results obtained (Fig. 4), together with phylogenetic trees including the λ CI repressor as an outgroup, 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 lateral gene transfer (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* (TGCACAAAAGTTCA)*, uvrA* (TATTCAAATGTTCA), *ssb* (TGCCTCCTCGAGCA) and *ruvAB* (AGCTCAAAGGCGCA) 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

The phylogenetic tree shown in 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 derive from the LexA of 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. 7a) 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 Proteobacteria 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* (Fig. 7b) 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 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 deoxyribonucleic 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. 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. 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. 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. 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 shown 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 binding 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. Experimentally confirmed LexA binding sequences of *B. subtilis* (Bsu), *E. coli* (Eco) and *R. sphaeroides* (Rps) are shown for comparison.

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.

Table 1. Bacterial strains and plasmids used in this work

Strain / plasmid	Relevant features	Source or reference
Fibrobacter succinogenes		
ATCC19169	Wild type strain	ATCC
E. coli		
DH5α	*supE4* Δ*lacU169* (Ø80 *lacZ*Δ M15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Clontech
BL21(λDE3) Codon Plus	F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *dcm*$^+$ Tc^R *gal λ (DE3) endA* Hte [*argU ileY leuW* Cam^R]	Stratagene
Plasmids		
pGEM-T	PCR cloning vector, Ap^R	Promega
pGEX-4T1	GST fusion expression vector, Ap^R	Amersham Phamacia Biotech
pUA1033	pGEMT derivative carrying a 954 bp-PCR fragment containing the promoter and the coding region of the *F.succinogenes lexA* gene	This work
pUA1037	pGEMT derivative carrying a 678 bp-PCR fragment containing the *F. succinogenes lexA* gene	This work
pUA1038	pGEX-4T1 derivative, carrying a *GST-lexA* fusion created by cloning the 678 bp *Eco*RI-*Sal*I fragment from pUA1037	This work

Table 2. Oligonucleotide primers used in this work.

EMSA, electrophoretic mobility shift assay.

Primer	Sequence (5'-3') [a]	Posn [b]	Application
LexA Up	TGCGGGGCGGCTCTTTGTGTATCGGGG	-276	Upper primer for cloning the *F. succinogenes lexA* gene and its promoter
LexA Dw	*GTCGAC*TTAGTTGACTTTTCTCATGACGCCG	+678	Lower primer for cloning the *F. succinogenes lexA* gene and its promoter. Contains a *Sal*I restriction site
LexA Low DIG	DIG-AGGGCTGCAAGGATAGAACGCACACCG	+169	Lower primer used to obtain wild type and mutagenized *lexA* probes (digoxigenin 5'-end labelled)
LexA Low	AGGGCTGCAAGGATAGAACGCACACCG	+169	Lower primer used to obtain unlabelled *lexA* probes
LexA EcoRI	*GAATTC*ATGGAAAATACAAATGAAAAACGC	+1	Upper primer for cloning the *F. succinogenes lexA* gene in pGEX-4T1. Contains an *Eco*RI restriction site
LexA Fb 154	CCTATTTTTGAACATTACG	-154	Upper primer to obtain a 323 bp *lexA* promoter fragment for EMSA
LexA Fb 72	TGACATTTGCCCAGTTGTGC	-72	Upper primer to obtain a 241 bp *lexA* promoter fragment for EMSA
LexA Fb 57	TGTGCATTATATTTTATAC	-57	Upper primer to obtain a 226 bp *lexA* promoter fragment for EMSA
LexA Fb 47	ATTTTATACATAGAAATTCC	-47	Upper primer to obtain a 216 bp *lexA* promoter fragment for EMSA
LexA Fb 37	TAGAAATTCCAGAACCC	-37	Upper primer to obtain a 206 bp *lexA* promoter fragment for EMSA
LexA FOOT	TTCCATGGTTTGCCTCTC	+6	Lower primer to obtain a 152 bp *lexA* promoter fragment to perform footprinting analyses
Mut1 Fb	TGACAT**g**TGCCCAGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut2 Fb	TGACATT**c**GCCCAGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut3 Fb	TGACATTT**a**CCCAGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut4 Fb	TGACATTTG**a**CCAGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut5 Fb	TGACATTTGC**t**CAGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut6 Fb	TGACATTTGCC**a**AGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut7 Fb	TGACATTTGCCC**g**GTTGTGCA	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut8 Fb	TGACATTTGCCCA**a**GTTGTGCA	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA

Table 2. (continued).

Primer	Sequence (5'-3') [a]	Posn [b]	Application
Mut9 Fb	TGACATTTGCCCAG**g**TGTGCA	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut10 Fb	TGACATTTGCCCAGT**c**GTGCATTATATTT	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut11 Fb	TGACATTTGCCCAGTT**a**TGCATTATATTT	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut12 Fb	TGACATTTGCCCAGTTG**c**GCATTATATTT	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut13 Fb	TGACATTTGCCCAGTTGT**a**CATTATATTT	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut14 Fb	TGACATTTGCCCAGTTGTG**t**ATTATATTTTATAC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut15 Fb	TGACATTTGCCCAGTTGTGC**g**TTATATTTTATAC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut16 Fb	TGACATTTGCCCAGTTGTGCA**g**TATATTTTATAC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut5 A Fb	TGACATTTGC**a**CAGTTGTGC	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
Mut12 G Fb	TGACATTTGCCCAGTTG**g**GCATTATATTT	-72	Upper primer to obtain single mutants in *lexA* promoter for EMSA
RecA Up Fb	AAATCGAGGCCAAATCCGAG	-143	Upper primer used to clone the *F.succinogenes recA* promoter for EMSA
RecA dw Fb	TGCTTTTGCTTTTTCGCTGG	+57	Lower primer used to clone the *F.succinogenes recA* promoter for EMSA
UvrA up Fb	ATCCTATGACCAGTGTGG	-155	Upper primer used to clone the *F.succinogenes uvrA* promoter for EMSA
UvrA dw Fb	GTGTTCATGAGCATCGCG	+39	Lower primer used to clone the *F.succinogenes uvrA* promoter for EMSA
Ssb up Fb	CTTCTAGCAGTGTTCCGG	-401	Upper primer used to clone the *F.succinogenes ssb* promoter for EMSA
Ssb dw Fb	TTACCGATGAGCATAACC	+35	Lower primer used to clone the *F.succinogenes ssb* promoter for EMSA
RuvAB up Fb	AACTGCGGATTATAGACC	-141	Upper primer used to clone the *F.succinogenes ruvAB* promoter for EMSA
RuvAB dw Fb	TCCACAACCACAAACGTG	+59	Lower primer used to clone the *F.succinogenes ruvAB* promoter for EMSA
TrpE up Fb	AACTCCCCCTTGCCATGTG	-255	Upper primer used to clone the *F.succinogenes trpE* promoter for EMSA
TrpE dw Fb	GGAGAGCGACGTAAATGC	+73	Lower primer used to clone the *F.succinogenes trpE* promoter for EMSA
Mut Bs-Fb	TGACATT**c**G**aa**CAGTTGT**t**C**g**TTATATTTTA	-72	Upper primer used to obtain a *F.succinogenes lexA* promoter derivative with the *B.subtilis lexA* box for EMSA

Table 2. (continued).

Primer	Sequence (5'-3') [a]	Posn [b]	Application
Mut Fb-Mx	TGACAT**c**TGCCCAGTTGT**t**CA**g**TATATTTTATAC	-72	Upper primer used to obtain a *F.succinogenes lexA* promoter derivative with the *M.xanthus lexA* box for EMSA
Mut Fb-Ec	TGACAT**c**TG**ta**CAGTTGT**a**CA**g**TATATTTTATAC	-72	Upper primer used to obtain a *F.succinogenes lexA* promoter derivative with the *E.coli lexA* box for EMSA
Wt Ab 53	TTTTGTAGTACTAATGTTC	-53 [c]	Upper primer used to clone the *Anabaena lexA* promoter for EMSA
Mut Ab-Rs 1	TTTTGTAGTAC**ctg**TAATGTTCTAGG	-53 [c]	Upper primer used to obtain an *Anabaena lexA* promoter derivative with the *lexA* box of *R.sphaeroides* for EMSA
Mut Ab-Rs 2	TTTTGTAGT**t**C**ctg**TAATGTTCTAGG	-53 [c]	Upper primer used to obtain an *Anabaena lexA* promoter derivative with the *lexA* box of *R.sphaeroides* for EMSA

[a] When present, added restriction sites are shown in italics and introduced nucleotide changes are shown in lowercase and underlined
[b] Position of 5'-end of the oligonucleotide with respect to the proposed translational starting point of each *F.succinogenes* gene
[c] Position of 5'-end of the oligonucleotide with respect to the proposed translational starting point of *Anabaena lexA* gene

Fig. 1

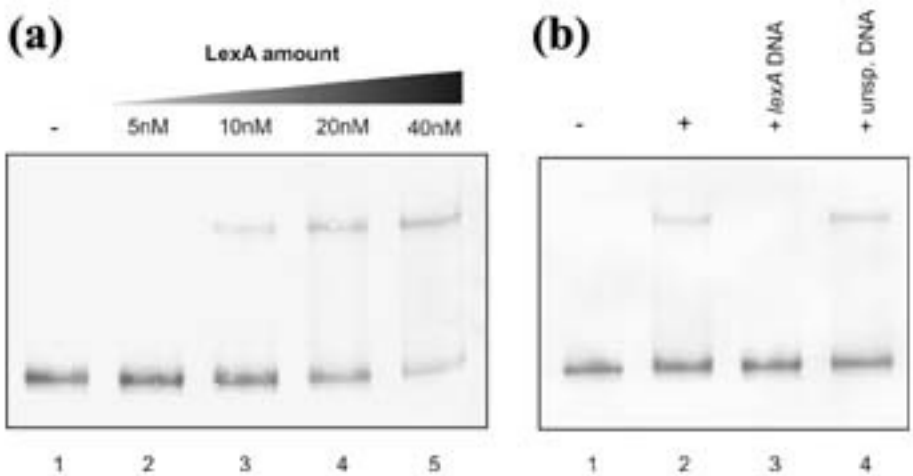

(a)
LexA amount
-
5nM
10nM
20nM
40nM
1
2
3
4
5
(b)
-
+
+ lexA DNA
+ unsp. DNA
1
2
3
4

Fig. 2

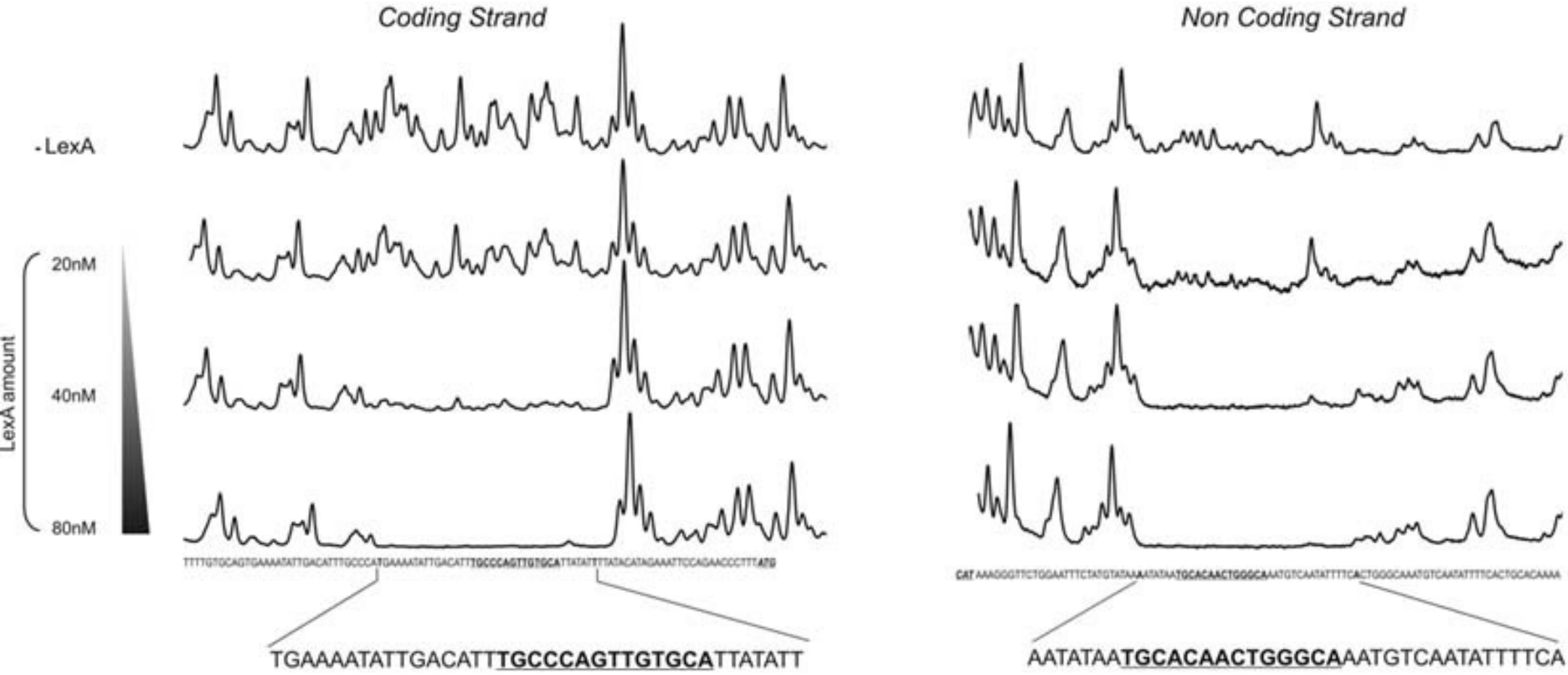

Coding Strand
Non Coding Strand
- LexA
20nM
40nM
80nM
LexA amount
TGAAAATATTGACATTTGCCCAGTTGTGCATTATATT
AATATAATGCACAACTGGGCAAATGTCAATATTTTCA

Fig. 3

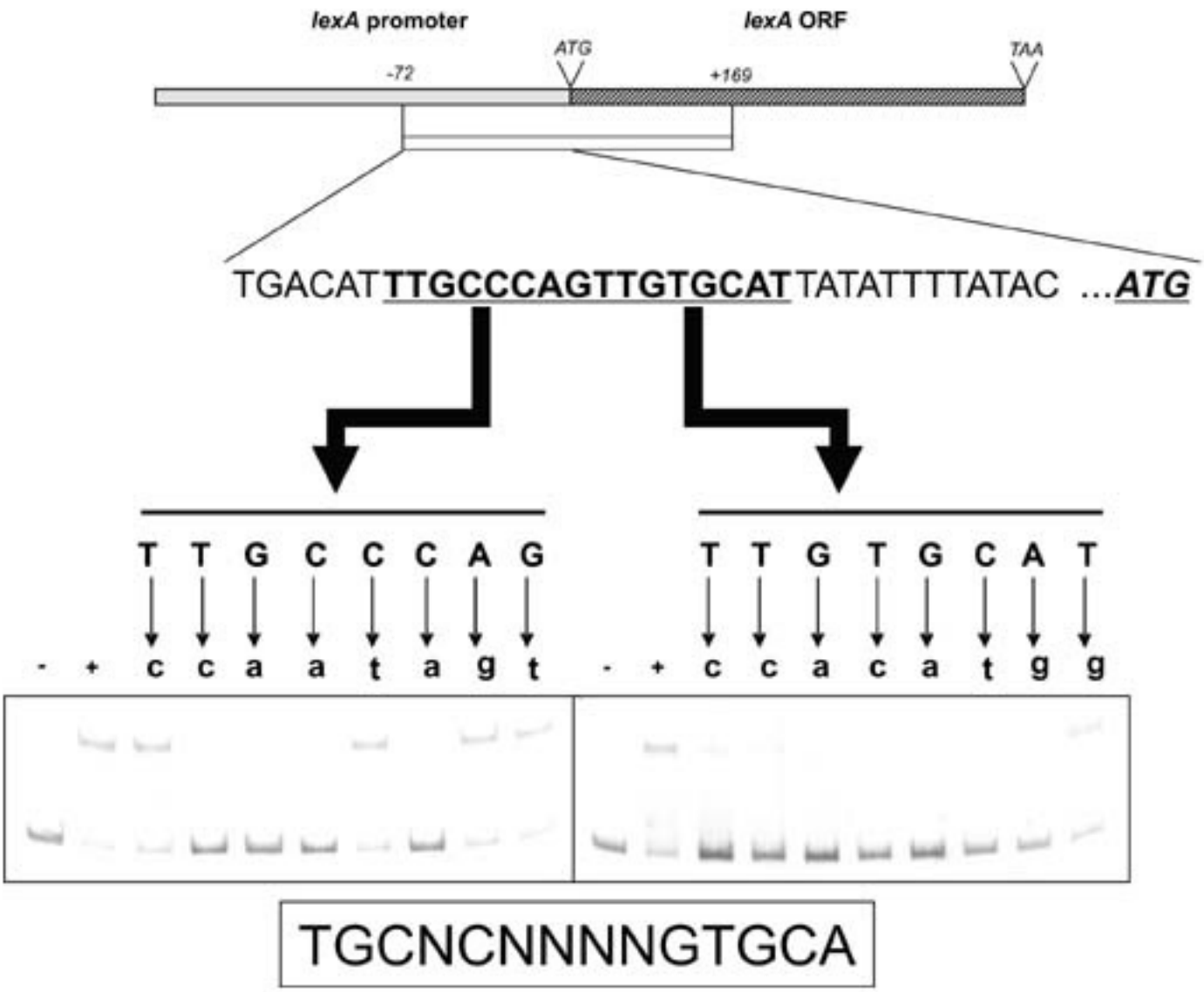

lexA promoter
lexA ORF
ATG
TAA
-72
+169
TGACATTTGCCCAGTTGTGCATTATATTTTATAC ...ATG
T T G C C C A G
- + c c a a t a g t
T T G T G C A T
- + c c a c a t g g
TGCNCNNNNGTGCA

Fig. 4

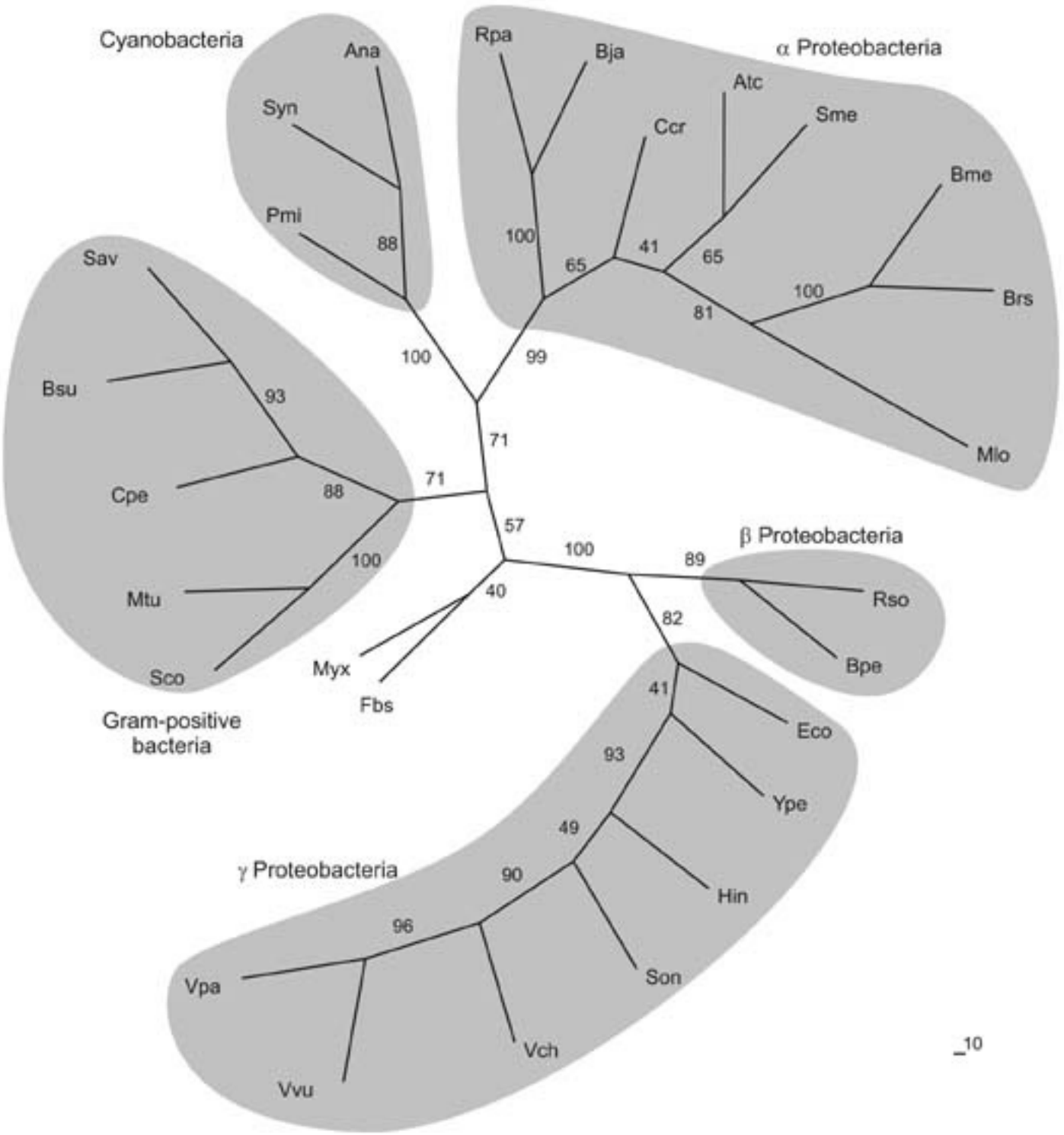

Cyanobacteria
Ana
Syn
Pmi
88
α Proteobacteria
Rpa
Bja
Atc
Sme
Ccr
Bme
100
65
41
65
100
Brs
81
Mlo
Sav
Bsu
93
Cpe
88
71
100
Mtu
Sco
Gram-positive bacteria
100
99
71
57
100
89
40
82
β Proteobacteria
Rso
Bpe
Myx
Fbs
41
Eco
93
Ype
49
γ Proteobacteria
90
Hin
96
Son
Vpa
Vch
Vvu
10

Fig. 5

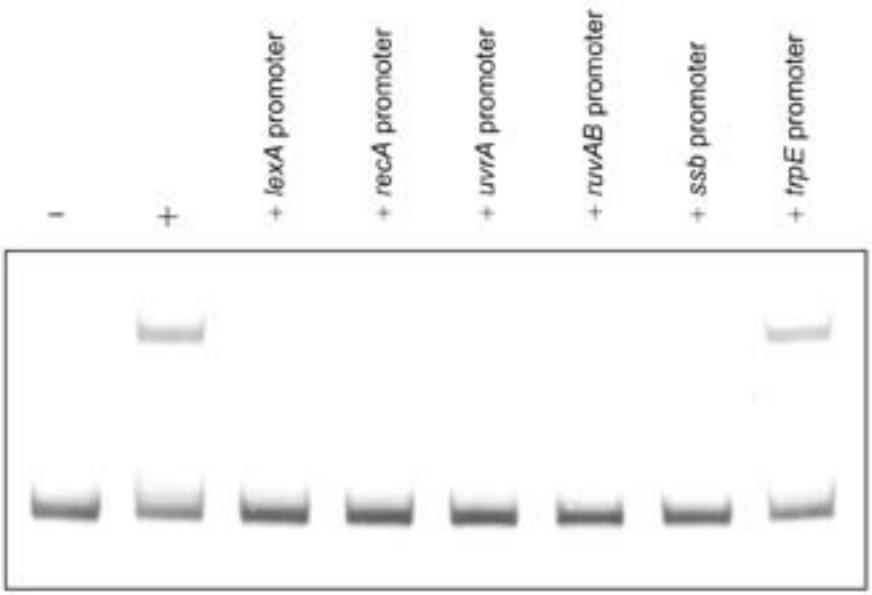

-
+
+ lexA promoter
+ recA promoter
+ uvrA promoter
+ ruvAB promoter
+ ssb promoter
+ trpE promoter

Fig. 6

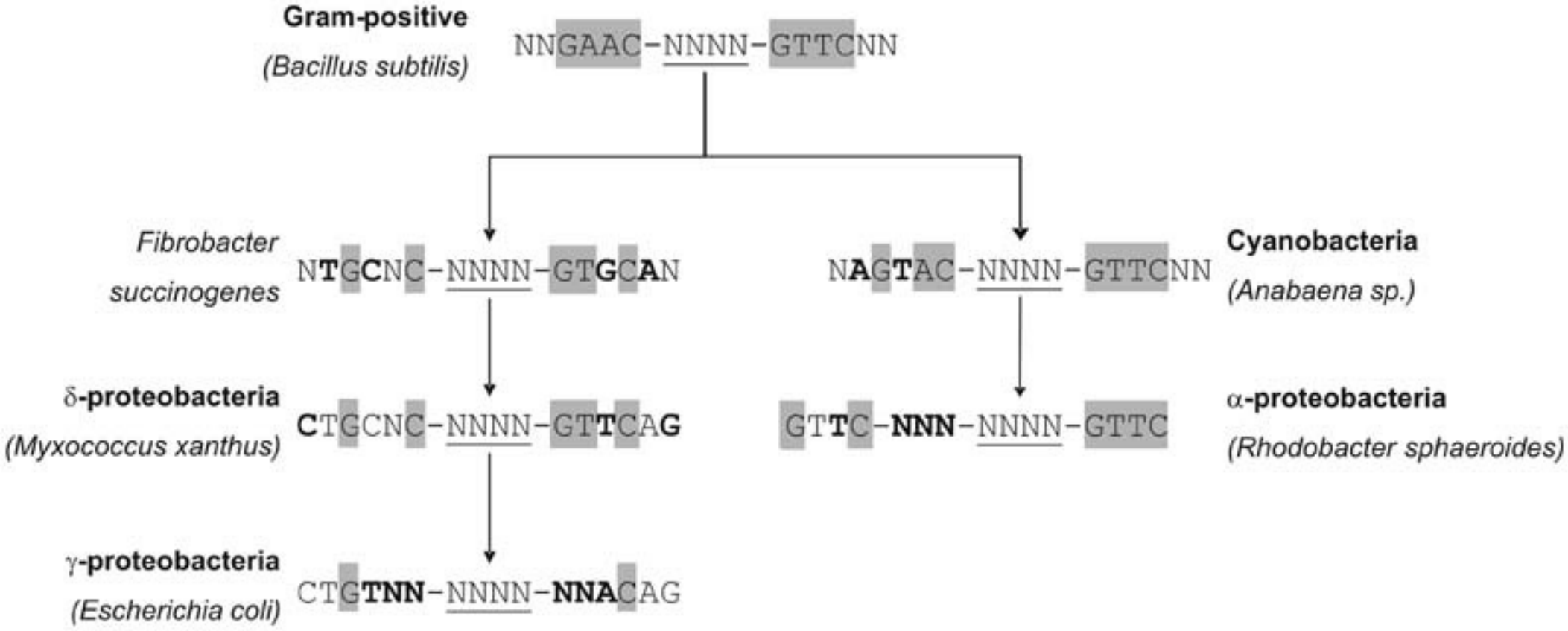
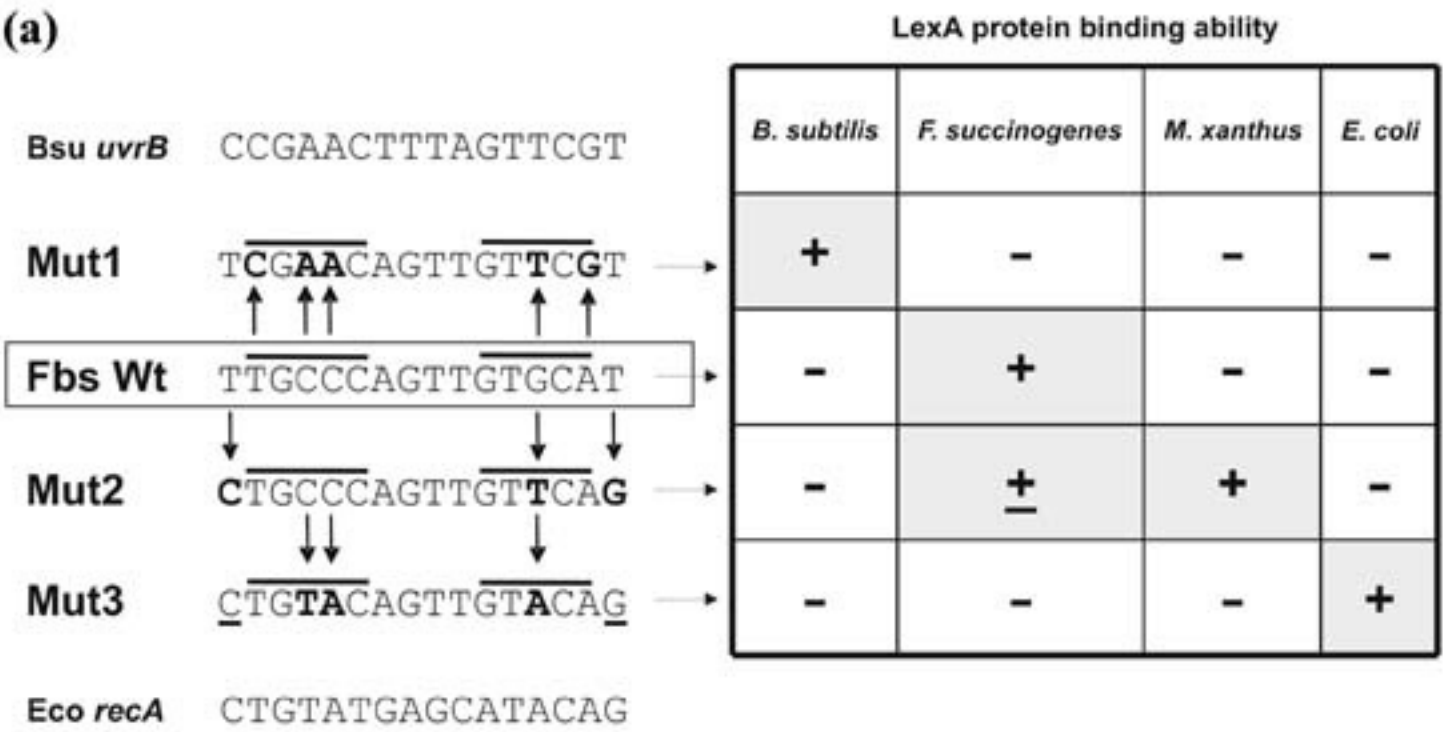

Gram-positive
(Bacillus subtilis)
NNGAAC-NNNN-GTTCNN
Fibrobacter succinogenes
NTGCNC-NNNN-GTGCAN
Cyanobacteria
(Anabaena sp.)
NAGTAC-NNNN-GTTCNN
δ-proteobacteria
(Myxococcus xanthus)
CTGCNC-NNNN-GTTCAG
α-proteobacteria
(Rhodobacter sphaeroides)
GTTC-NNN-NNNN-GTTC
γ-proteobacteria
(Escherichia coli)
CTGTNN-NNNN-NNACAG

Fig. 7


(a)
LexA protein binding ability
Bsu uvrB CCGAACTTTAGTTCGT
Mut1 TCGAACAGTTGTTCGT
Fbs Wt TTGCCCAGTTGTGCAT
Mut2 CTGCCCAGTTGTTCAG
Mut3 CTGTACAGTTGTACAG
Eco recA CTGTATGAGCATACAG
B. subtilis
F. succinogenes
M. xanthus
E. coli
+ - - -
- + - -
- ± + -
- - - +

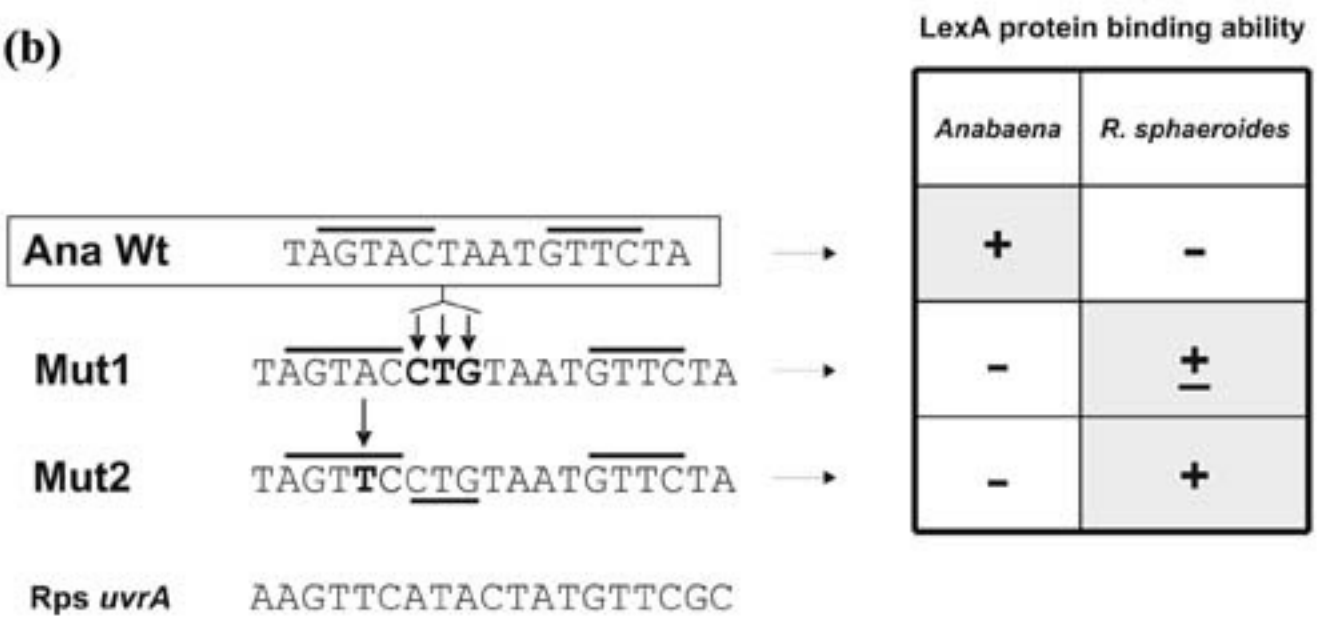

(b)
LexA protein binding ability
Ana Wt TAGTACTAATGTTCTA
Mut1 TAGTACCTGTAATGTTCTA
Mut2 TAGTTCCTGTAATGTTCTA
Rps uvrA AAGTTCATACTATGTTCGC
Anabaena
R. sphaeroides
+ -
- ±
- +

Fig. 8

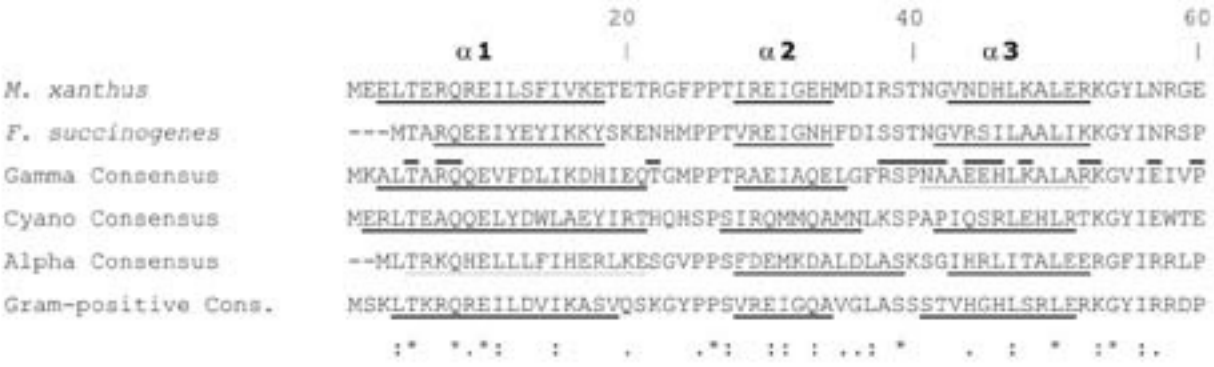
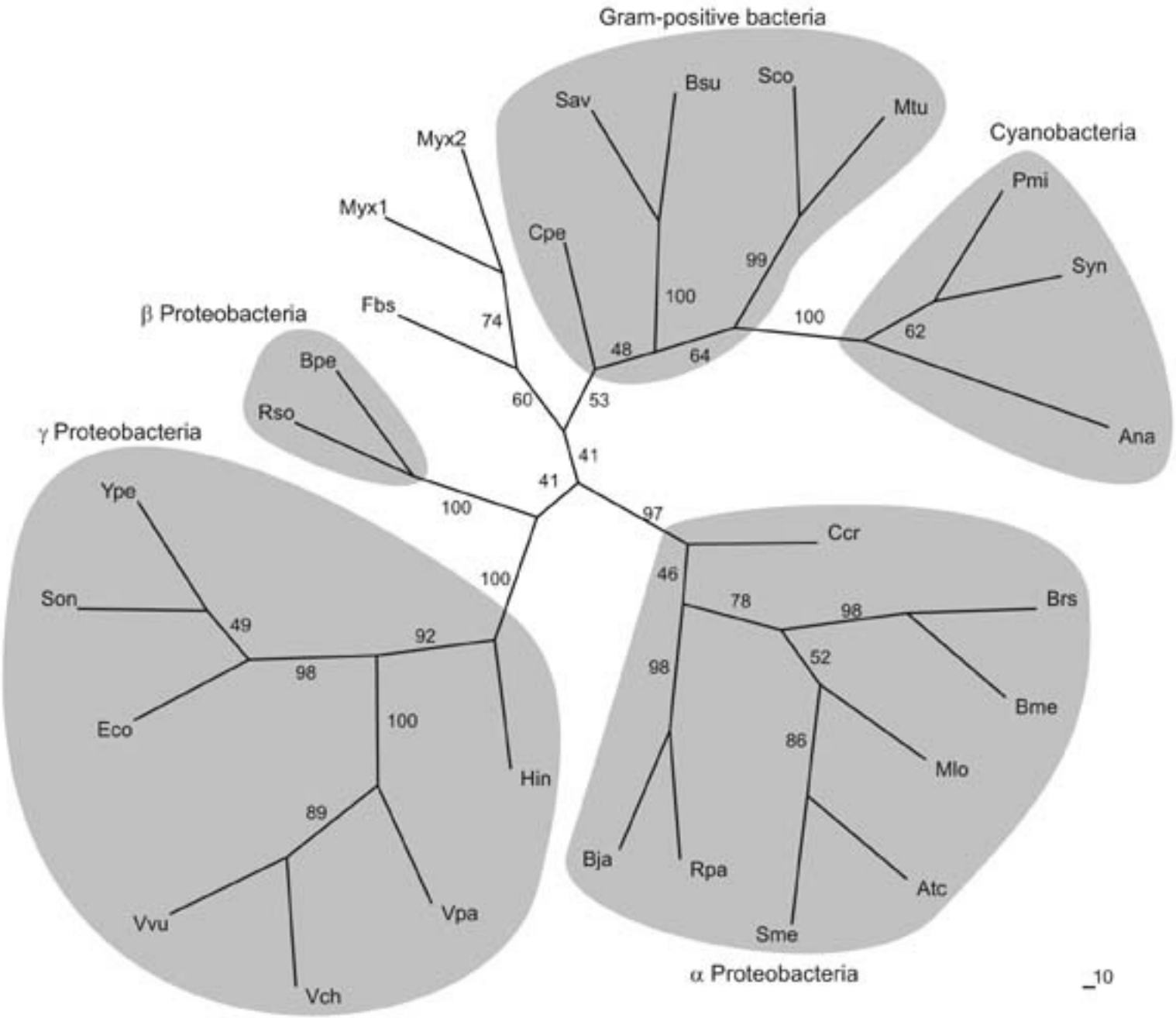

20
40
60
α1
α2
α3
M. xanthus MEELTERQREILSFIVKETETRGFPPTIREIGEHMDIRSTNGVNDHLKALERKGYLNRGE
F. succinogenes ---MTARQEEIYEYIKKYSKENHMPPTVREIGNHFDISSTNGVRSILAALIKKGYINRSP
Gamma Consensus MKALTARQQEVFDLIKDHIEQTGMPPTRAEIAQELGFRSPNAAEEHLKALARKGVIEIVP
Cyano Consensus MERLTEAQQELYDWLAEYIRTHQHSPSIRQMMQAMNLKSPAPIQSRLEHLRTKGYIEWTE
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
74
99
100
48
64
100
62
60
53
41
41
100
97
100
46
78
98
49
92
98
52
98
100
86
89
_10