Activator role of the pneumococcal Mga-like virulence transcriptional regulator

Virtu Solano-Collado, Manuel Espinosa and Alicia Bravo*

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu, 9. E-28040 Madrid, Spain

Running title: Activator role of the Mga*Spn* regulator

* Corresponding author
Phone: (34) 918373112
Fax: (34) 915360432
E-mail: abravo@cib.csic.es

ABSTRACT

Global transcriptional regulators that respond to specific environmental signals are crucial in bacterial pathogenesis. In the case of the Gram-positive pathogen *Streptococcus pneumoniae* (the pneumococcus), the *sp1800* gene of the clinical isolate TIGR4 encodes a protein that exhibits homology to the Mga 'stand-alone' response regulator of the group A *Streptococcus*. Such a pneumococcal protein was shown to play a significant role in both nasopharyngeal colonization and development of pneumonia using murine infection models. Moreover, it was shown to repress the expression of several genes located within the *rlrA* pathogenicity islet. The pneumococcal R6 strain, which derives from the D39 clinical isolate, lacks the *rlrA* islet but has a gene (herein named *mgaSpn*) equivalent to the *sp1800* gene. In this work, and using *in vivo* approaches, we have identified the promoter of the *mgaSpn* gene (*Pmga*) and demonstrated that four neighbouring open reading frames of unknown function (*spr1623* to *spr1626*) constitute an operon. Transcription of this operon is under the control of two promoters (*P1623A* and *P1623B*) that are divergent from the *Pmga* promoter. Furthermore, we have shown that the Mga*Spn* protein activates the *P1623B* promoter *in vivo*. This activation requires sequences located around 50-120 nucleotides upstream of the *P1623B* transcription start site. By DNase I footprinting assays, we have also demonstrated that such a region includes an Mga*Spn* binding site. This is the first report on the activator role of the pneumococcal Mga-like protein.

INTRODUCTION

During infection, pathogenic bacteria must be able to survive in different niches of their hosts. This adaptation requires a coordinated regulation in the expression of many virulence and metabolic genes. Global transcriptional regulators that respond to specific environmental signals (response regulators) are key elements in such regulatory networks. One example is the Mga protein of the Gram-positive (G+) bacterium *Streptococcus pyogenes* (group A *Streptococcus*, GAS), which causes a broad spectrum of diseases in humans (3). To date, very little is known about how Mga (multiple gene regulator of GAS) is able to sense changes in the environment. However, its role in pathogenesis has been studied in detail (12, 23). During exponential growth, Mga activates directly the transcription of several virulence genes, including its own gene. These Mga-regulated genes encode proteins that enable the bacterium to colonize specific host tissues and evade the host immune response. In addition, a transcriptome analysis revealed that Mga activates or represses, likely in an indirect way, the expression of various genes involved in the transport and utilization of sugars and other metabolites (30). Homologues of Mga have been identified in several G+ pathogens, including *S. dysgalactiae* (5, 38), *S. pneumoniae* (9) and *Bacillus anthracis* (35).

S. pneumoniae (the pneumococcus) remains a main cause of morbidity and mortality worldwide. It usually resides in the nasopharynx of healthy individuals. However, when the immune system weakens, *S. pneumoniae* can cause serious diseases, such as pneumonia, meningitis and septicemia (15, 37). The genomic sequence of the TIGR4 strain (a serotype 4 clinical isolate) revealed that about 5% of

this genome is composed of insertion sequences that may contribute to genome rearrangements through uptake of foreign DNA (34). Signature-tagged mutagenesis experiments in TIGR4 led to the identification of several genes associated with virulence (8). One of them was the *sp1800* gene that is highly conserved in the pneumococcal strains whose genome has been totally or partially sequenced. The *sp1800* gene encodes a protein named MgrA (Mga-like repressor A) due to its homology to the Mga response regulator of GAS (9). MgrA (493 amino acids) was shown to play a significant role in both nasopharyngeal colonization and development of pneumonia in murine infection models (9). Furthermore, microarray experiments showed that MgrA is able to repress the expression of several genes located within the *rlrA* pathogenicity islet (9). Contrary to the *sp1800* gene, the *rlrA* islet has been found in a small number of pneumococcal strains (27), indicating that it might not be the main target of the MgrA regulator (9). This fact suggested that novel MgrA-regulated genes could be identified working with different pneumococcal strains and/or under different bacterial growth conditions. Indeed, there is evidence that some response regulators influence the transcriptional profile in a different manner depending on the bacterial strain and/or serotype (10, 11, 22, 30).

The genome of the pneumococcal R6 strain, which derives from the D39 clinical isolate (serotype 2), has been totally sequenced. Unlike the TIGR4 strain, R6 and D39 lack the *rlrA* pathogenicity islet (14, 20, 34). The *spr1622* gene (herein named *mgaSpn*) of the R6 strain encodes a protein (Mga*Spn*) that differs from the MgrA regulator in two amino acid residues. In the present work, we have identified the promoter of the *mgaSpn* gene (*Pmga*). Upstream of this promoter there are four open reading frames (ORFs) of unknown function (*spr1623* to *spr1626*) that are highly

conserved in the TIGR4 strain. We have demonstrated that these ORFs are transcribed into a single polycistronic mRNA molecule from two promoters (*P1623A* and *P1623B*) that are divergent from the *Pmga* promoter. Moreover, unlike previous studies in TIGR4 (9), we show here that Mga*Spn* activates the *P1623B* promoter *in vivo*. This activation requires sequences that are recognized by a His-tagged Mga*Spn* protein *in vitro*. Hence, our findings show, for the first time, that the pneumococcal Mga-like regulator has a positive effect on gene expression.

MATERIALS AND METHODS

Bacterial strains, oligonucleotides and plasmids. The *S. pneumoniae* R6 strain was used (14). To construct the R6Δ*mga* mutant strain, gene replacement by homologous recombination was carried out. A 1,165-bp DNA fragment that contained the pC194 *cat* gene (chloramphenicol resistance) (13) was flanked by R6 DNA sequences (543-bp and 605-bp, respectively). In the R6 genome, such DNA sequences are flanking the *mgaSpn* gene (promoter plus coding sequence). The *cat* cassette generated *in vitro* was used to transform competent R6 cells. Selection of transformants resistant to chloramphenicol (1.5 μg/ml) led to the isolation of the R6Δ*mga* strain. Dye-terminator sequencing at Secugen (CIB, Madrid) confirmed that R6Δ*mga* lacks the chromosomal region that spans the 1596826 and 1598431 coordinates.

Oligonucleotides used are listed in Table 1. Plasmids pAS, pAST and pAS-*T2T1rrnB* (herein named pAST2), which are based on the pMV158 replicon, were used (32). They carry the *tetL* gene (tetracycline resistance). Plasmid pDL287, a

122 pVA380-1 derivative that carries a kanamycin resistance gene, was also used (21).
123 To construct pAS-*Pmga*, a 170-bp region of the R6 genome (promoter *Pmga*) was
124 amplified by PCR using the PrSp1 and PrSp2 primers. The amplified DNA was
125 digested with *Bam*HI, and the 142-bp digestion product was inserted into the *Bam*HI
126 site of pAS. In pAS-*Pmga*, *gfp* expression is under the control of the *Pmga* promoter.
127 To construct pAST-*PAB* and pAST2-*Pmga*, a 333-bp region of the R6 DNA was
128 amplified with the PmgaSac and PABSac primers. After *Sac*I digestion, the 301-bp
129 restriction fragment (promoters *P1623A*, *P1623B* and *Pmga*) was cloned into the *Sac*I
130 site of pAST (pAST-*PAB*; *gfp* expression under the control of the *P1623A* and
131 *P1623B* promoters) and pAST2 (pAST2-*Pmga*; *gfp* expression under the control of
132 the *Pmga* promoter). To construct pAST-*PAB*Δ84, a 246-bp region of the R6 DNA
133 (promoters *P1623A* and *P1623B*), was amplified with the PABSac and PABΔ84Sac
134 primers. The PCR product was digested with *Sac*I, and the 216-bp restriction
135 fragment was inserted into the *Sac*I site of pAST. To construct pAST-*PAB*Δ153, a
136 177-bp region of the R6 DNA (promoters *P1623A* and *P1623B*), was amplified with
137 the PABSac and PABΔ153Sac primers. After *Sac*I digestion, the 146-bp restriction
138 fragment was cloned into the *Sac*I site of pAST. In pAST-*PAB*Δ84 and pAST-
139 *PAB*Δ153, *gfp* expression is under the control of the *P1623A* and *P1623B* promoters.
140 Construction of the pDL*PsulA*::*mga* plasmid was as follows: (i) amplification of a 189-
141 bp region of the R6 DNA (promoter *PsulA*) (32) using the PsulNde and PsulCla
142 primers. The PCR-synthesized DNA was digested with *Nde*I, generating the 172-bp
143 *PsulA* fragment, (ii) amplification of a 1,650-bp region of the R6 DNA (promoter-less
144 *mgaSpn* gene) using the mgaNde and mgaCla primers. After digestion with *Nde*I, the

145 1,636-bp restriction fragment was ligated to the 172-bp *PsulA* fragment (*PsulA*::*mga*
146 fusion gene), and (iii) amplification of the *PsulA*::*mga* fusion gene with the PsulCla
147 and mgaCla primers. After digestion with *Cla*I, the 1,777-bp restriction fragment was
148 cloned into the *Cla*I site of plasmid pDL287 (21).

149 **Growth and transformation of bacteria.** *S. pneumoniae* was grown in AGCH
150 medium (17, 32) supplemented with 0.2% yeast extract and 0.3% sucrose. For
151 plasmid-harbouring cells, the medium was supplemented with tetracycline (1 µg/ml)
152 and/or kanamycin (50 µg/ml). Experiments were performed at 37ºC. Procedures for
153 competence development and transformation of *S. pneumoniae* were reported (19).

154 **Isolation of DNA and RNA**. Genomic DNA from *S. pneumoniae* was prepared
155 as described (17). For small-scale preparations of plasmid DNA, the High Pure
156 Plasmid Isolation Kit (Roche Applied Science) was used (32). The Aurum Total RNA
157 Mini Kit (BioRad) was used to isolate total RNA from *S. pneumoniae*. Cells were
158 grown to an optical density at 650 nm (OD_{650}) of 0.3. Then, cultures were processed
159 as specified by the supplier, except that cells were resuspended in buffer L (50 mM
160 Tris-HCl, pH 7.6, 1 mM EDTA, 50 mM NaCl, 0.1% deoxycholate) and incubated at
161 30ºC for 10 min. DNA and RNA concentrations were determined using a NanoDrop
162 ND-1000 Spectrophotometer (BioRad).

163 **Polymerase chain reaction (PCR) conditions.** The Phusion High-Fidelity DNA
164 Polymerase (Finnzymes) and the Phusion HF Buffer were used. Reaction mixtures
165 (50 µl) contained 5-30 ng of template DNA, 20 pmol of each primer, 200 µM of each
166 dNTP and 1 unit of DNA polymerase. PCR conditions were reported previously (32).
167 PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN).

168 **Primer extension of total RNA**. The ThermoScript Reverse Transcriptase
169 enzyme (Invitrogen) was used. Primers were ^{32}P-labeled at the 5´-end with [γ-^{32}P]-
170 ATP (3000 Ci/mmol; PerkinElmer) and T4 polynucleotide kinase (New England
171 Biolabs). Non-incorporated nucleotide was removed using MicroSpin G-25 columns
172 (GE Healthcare). In assays with non-radiolabeled primers, [α-^{32}P]-dATP (3000
173 Ci/mmol; Hartmann) was used in the extension reactions. Reaction mixtures (20 µl)
174 contained ~2 µg of total RNA and 1-2 pmol of primer. To anneal the primer with the
175 transcript, samples were incubated at 65ºC for 5 min. Extension reactions were
176 carried out at 58ºC for 60 min. After heating at 85ºC for 5 min, samples were ethanol-
177 precipitated, and dissolved in loading buffer (80% formamide, 1 mM EDTA, 10 mM
178 NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol). cDNA products were analyzed
179 by sequencing gel (8 M urea-6% polyacrylamide) electrophoresis. Dideoxy-mediated
180 chain-termination sequencing reactions were run in the same gel. Labeled products
181 were visualized using a Fujifilm Image Analyzer FLA-3000. The intensity of the bands
182 was quantified using the QuantityOne software (BioRad).
183 **RT-PCR assays**. For first-strand cDNA synthesis, 20 pmol of primer were
184 annealed to ~1.5 µg of total RNA. The mixture was incubated with 15 units of
185 ThermoScript Reverse Transcriptase (Invitrogen) at 55ºC for 45 min. PCRs were then
186 carried out using cDNA as template (10% of the first-strand reaction), 20 pmol of each
187 primer and the Phusion High-Fidelity DNA Polymerase (see PCR conditions). To rule
188 out the presence of genomic DNA in the RNA preparation, the same reactions were
189 performed in the absence of the reverse transcriptase. As positive control, PCRs
190 were performed with genomic DNA. PCR-products were analyzed by agarose (0.8%)

191 gel electrophoresis. Gels were stained with ethidium bromide and DNA was
192 visualized using a Gel-Doc system (Bio-Rad).

193 **Fluorescence assays**. Plasmid-carrying cells were grown to an OD_{650} of 0.3.
194 Fluorescence intensity was measured as reported (32) using a Thermo Scientific
195 Varioskan Flash instrument (excitation at 488 nm and emission at 515 nm). In each
196 case, three independent cultures were analyzed. The fluorescence corresponding to
197 200 µl of PBS buffer without cells was around 0.03 arbitrary units.

198 **Western blots**. Cells were grown to an OD_{650} of 0.3. The protocol used to
199 prepare whole-cell extracts was described previously (32). Total proteins were
200 separated by SDS-polyacrylamide (10%) gel electrophoresis. Proteins were
201 transferred electrophoretically to Immun-blot PVDF membranes (BioRad) using a Mini
202 Trans Blot (Bio-Rad) as reported (32). Membranes were probed with polyclonal
203 antibodies against His-tagged Mga*Spn*. Antigen-antibody complexes were detected
204 using anti-rabbit horseradish peroxidase-conjugated antibodies, the Immun-StarTM
205 HRP Substrate Kit (BioRad), and the Luminescent Image Analyzer LAS-3000 (Fujifilm
206 Life Science). The intensity of the bands was quantified using the QuantityOne
207 software (BioRad).

208 **Overproduction and purification of Mga*Spn*-His.** Gene *mgaSpn* was
209 engineered to encode a His-tagged Mga*Spn* protein (Mga*Spn*-His). A 1,512-bp region
210 of the R6 genome was amplified by PCR using the 1622Nde and 1622Xho-His
211 oligonucleotides, which include a single restriction site for *Nde*I and *Xho*I, respectively
212 (Table 1). The amplified DNA was digested with both enzymes, and the 1,481-bp
213 digestion product was cloned into the pET24b vector (Novagen), which enables a C-
214 terminal His_6-tag fusion. *E. coli* BL21(DE3) cells harboring plasmid pET24b-*mgaSpn*-

His were grown at 37ºC in tryptone-yeast extract medium containing kanamycin (30 μg/ml). When the culture reached an OD_{600} of 0.45, isopropyl-β-D-thiogalactopyranoside (IPTG) was added (1 mM). After 25 min, cells were incubated with rifampicin (200 μg/ml) for 60 min. Cells were then sedimented, washed twice with buffer V-His (10 mM Tris-HCl, pH 7.6, 5% glycerol, 300 mM NaCl, 5 mM β-mercaptoethanol), and stored at –80ºC. The cell pellet was concentrated (40x) in buffer V-His containing an EDTA-free protease inhibitor cocktail (Roche). Cells were disrupted by passage through a French pressure cell, and the whole-cell extract was centrifuged to remove cell debris. Imidazole (10 mM) was added to the clarified extract, which was loaded onto a HisTrap HP column (GE Healthcare) pre-equilibrated with buffer V-His containing 10 mM imidazole. After washing with the same buffer, Mga*Spn*-His was eluted with buffer V-His containing 250 mM imidazol. Fractions containing Mga*Spn*-His were identified by Coomassie-stained SDS-polyacrylamide (10%) gels, pooled, and dialyzed against buffer P (20 mM Tris-HCl, pH 7.6, 5% glycerol, 250 mM NaCl, 1 mM EDTA, 1 mM DTT). The protein preparation was concentrated by filtering through a 10 kDa cut-off membrane (Macrosep, Pall), loaded onto a HiLoad Superdex 200 gel-filtration column (Amersham), and subjected to fast-pressure liquid chromatography (FPLC; Biologic DuoFlow, BioRad). Fractions containing Mga*Spn*-His were pooled, concentrated and stored at -80ºC. Protein concentration was determined using a NanoDrop ND-1000 Spectrophotometer (BioRad).

DNase I footprinting assays. A 222-bp region of the R6 genome (coordinates 1598298-1598519) was amplified by PCR using the 1622I and 1622H primers. One of

the primers was previously ^{32}P-labeled at the 5´-end using [γ-^{32}P]-ATP (3,000 Ci/mmol; PerkinElmer) and T4 polynucleotide kinase. Reaction mixtures (10 μl) contained 30 mM Tris-HCl, pH 7.6, 1.2 mM DTT, 0.2 mM EDTA, 1 mM $CaCl_2$, 10 mM $MgCl_2$, 50 mM NaCl, 1% glycerol, 4 nM of the ^{32}P-labeled 222-bp DNA and different concentrations of Mga*Spn*-His. After 20 min at room temperature, 0.04 units of DNase I (Roche Applied Science) were added for 5 min at the same temperature. Reactions were stopped with 1 µl of 250 mM EDTA. Then, 4 µl of loading buffer (80% formamide, 1 mM EDTA, 10 mM NaOH, 0.1% bromophenol blue and 0.1% xylene cyanol) were added. Samples were heated at 95°C for 5 min, and loaded on 8 M urea-6% polyacrylamide gels. Dideoxy-mediated chain termination sequencing reactions using the 222-bp fragment and either the 1622I or the 1622H oligonucleotide were run in the same gel. Labeled products were visualized using a Fujifilm Image Analyzer FLA-3000 or by autoradiography. The intensity of the bands was quantified using the QuantityOne software (BioRad).

RESULTS

Transcription of the *mgaSpn* gene in pneumococcal R6 cells. The complete genome sequence of *S. pneumoniae* R6 has been published (14) (GenBank AE007317.1). The ATG codon at coordinate 1598270 is likely the translation start site of the *spr1622* gene (herein named *mgaSpn*), since it is preceded by a putative Shine-Dalgarno sequence (5´-AAAGAGAGAAAG-3´) (Fig.1) that complies with the reported consensus sequence for pneumococcus (http://www.changbioscience.com/biotoolkit2.html). Translation from this ATG codon

would produce a protein of 493 residues (Mga*Spn*). EMBOSS needle global sequence alignment (31) of Mga*Spn* and the Mga regulator (530 residues) encoded by the M6_Spy1720 gene of *S. pyogenes* MGAS10394 revealed a 42.6% of similarity and a 21.4% of identity.

To examine whether the *mgaSpn* gene was transcribed, RT-PCR experiments were carried out (Fig. 2). The 1622A oligonucleotide was used as primer for extension on total RNA isolated from R6 cells. The cDNA products were further amplified by PCR using either the 1622A and C1622D or the 1622A and 1622C primers. As controls, PCR reactions were performed using total RNA (negative control) or genomic DNA (positive control) as templates. With the 1622A and C1622D primers, a PCR-product that migrated at the position expected for a 1,023-bp DNA was synthesized. Such a product was not visualized in the negative control. With the 1622A and 1622C primers, no PCR-products were detected. However, a product with the mobility expected for a 1,221-bp fragment was synthesized in the positive control. Therefore, transcription of the *mgaSpn* gene was initiated at a site located downstream of coordinate 1598452. Sequence analysis of the region spanning this coordinate and the translation start codon of the *mgaSpn* gene revealed the existence of a putative promoter (herein named *Pmga*) (Fig. 1). It has a consensus –10 hexamer (5´-**TATAAT**-3´), a consensus –10 extension (5´-**TGTG**-3´) and shows a 3/6 match at the –35 hexamer (5´-A**TG**CT**A**-3´) (consensus residues are shown in bold). The –35 and –10 elements are separated by 16 nucleotides. The features of the *Pmga* promoter indicate that it would be likely recognized by a housekeeping σ factor similar to the *Escherichia coli* σ^{70} (7).

We next performed primer extension assays using RNA from R6 cells and the 1622D primer, which is complementary to the C1622D oligonucleotide (see Fig. 2). No cDNA products were detected (not shown), which indicated that the amount of *mgaSpn* transcripts in the RNA preparation was small. To amplify the signal, a 136-bp chromosomal region (coordinates 1598440 to 1598305), which contained the putative *Pmga* promoter, was inserted into the *Bam*HI site of the pAS vector (plasmid pAS-*Pmga*; Fig. 3). This site is located upstream of a promoter-less *gfp* allele that encodes a variant of the green fluorescent protein (32). The intensity of fluorescence in cultures of cells carrying the pAS-*Pmga* plasmid was slightly higher (1.5-fold) than in cultures of cells harbouring the pAS vector, indicating that the 136-bp region contained a promoter signal. Also, primer extension assays were performed using RNA from cells carrying plasmid pAS-*Pmga*. As primer, the INTgfp oligonucleotide, which anneals to *gfp* transcripts, was used (Fig. 3). Two cDNA products of 120 and 121 nucleotides were detected, indicating that transcription of the *gfp* gene started at a site located 7-8 nucleotides downstream of the –10 element of the *Pmga* promoter. Thus, the *Pmga* promoter was functional under our bacterial growth conditions.

Transcription of the *spr1623-spr1626* operon in pneumococcal R6 cells. Upstream of the *mgaSpn* gene there are four ORFs (*spr1623*-*spr1626*) that appeared to be organized in an operon (Fig. 1). The *mgaSpn* gene and the putative operon would be divergently transcribed. The ATG initiation codon of the *mgaSpn* gene (coordinate 1598270) and the ATG initiation codon of the *spr1623* ORF (coordinate 1598960) are separated by 689-bp. To investigate whether the putative operon was transcribed, RT-PCR assays were performed (Fig. 4). First, the 1623B oligonucleotide

was used as primer for extension on total RNA isolated from R6 cells. The resulting cDNA was amplified by PCR using either the 1623B and 1623C or the 1623B and 1623A primers. With the 1623B and 1623C primers, a PCR-product that migrated at the position expected for a 695-bp DNA was synthesized (Fig. 4). With the 1623B and 1623A primers, no PCR-products were visualized. Nevertheless, such primers amplified an 892-bp region using genomic DNA as template. Next, we performed RT-PCR assays using the 1626A primer for cDNA synthesis (Fig. 4). Amplification of the cDNA with the 1626A and 1623C primers generated a product that moved at the position expected for a 1,917-bp fragment. Collectively, these results indicated that the four ORFs (*spr1623*-*spr1626*) were transcribed into a polycistronic mRNA molecule from a site(s) located between coordinates 1598433 and 1598630. Sequence analysis of this region predicted a promoter sequence (herein named *P1623A*; Fig.1) that has a canonical –10 hexamer (5´-**TATAAT**-3´) and a near consensus –35 hexamer (5´-**TTGAC**T-3´). The spacing between both sequence elements is 17 nucleotides.

To analyze whether the *P1623A* promoter was functional *in vivo*, we performed primer extension assays using the PDA oligonucleotide (Fig. 5). Two cDNA products of 106 and 191 nucleotides were detected, which would correspond to transcription initiation events at coordinates 1598592 (*P1623A* promoter) and 1598507, respectively. This result indicated that the pneumococcal RNA polymerase recognized not only the *P1623A* promoter but also a promoter sequence (named *P1623B*) that has a consensus –10 hexamer (5´-**TATAAT**-3´) but lacks a –35 element (see Fig. 1). The functionality of the *P1623B* promoter was confirmed further by

primer extension using the PDB oligonucleotide (cDNA product of 60 nucleotides) (Fig. 5).

Mga*Spn* activates the *P1623B* promoter *in vivo*. To investigate whether Mga*Spn* influenced the activity of a particular promoter *in vivo*, we constructed a pneumococcal strain designed to overproduce Mga*Spn*. First, we constructed the *PsulA*::*mga* fusion gene, in which the *Pmga* promoter of the *mgaSpn* gene was replaced with the promoter of the pneumococcal *sulA* gene (*PsulA*) (18, 32). The fusion gene was then inserted into pDL287 (21) generating the pDL*PsulA*::*mga* recombinant. Compared to R6 plasmid-free cells (Fig. 6; lane 1) or R6 cells carrying pDL287 (not shown), the amount of Mga*Spn* increased ~8-fold in cells harbouring pDL*PsulA*::*mga* (Fig. 6; lane 3). By primer extension, we analyzed the effect of the Mga*Spn* overproduction on the activity of the *P1623A* and *P1623B* promoters located on the bacterial chromosome. We used a mix of oligonucleotides radioactively labeled at the 5' end: PDA (see Fig. 5) and PErpoE, which anneal to *spr1623* and *rpoE* transcripts, respectively. The *rpoE* gene (*spr0437* in the R6 genome) encodes the delta subunit of the RNA polymerase and was used as internal control. As shown in Figure 7, using RNA from R6/pDL287 cells (lane 1; low levels of Mga*Spn*), three cDNA products of 106-nt (*P1623A* promoter), 191-nt (*P1623B* promoter) and 231-nt (*PrpoE* promoter) were synthesized. Unlike the 231-nt product, the amount of the 106-nt and 191-nt cDNAs increased 2.6-fold and 4.5-fold, respectively, when RNA from R6/pDL*PsulA*::*mga* cells was used (lane 2; overproduction of Mga*Spn*). Therefore, overproduction of Mga*Spn* led to activation of promoters *P1623A* and

P1623B, although the effect appeared to be higher on the activity of promoter *P1623B*.

We next constructed an R6 derivative, named R6Δ*mga*, in which the chromosomal region spanning the coordinates 1596826 and 1598431 was replaced with the *cat* gene (chloramphenicol resistance) of plasmid pC194 (13). This mutant strain lacks the *mgaSpn* gene (including the *Pmga* promoter) but conserves the *P1623A* and *P1623B* promoter sequences (see Fig. 1). As expected, R6Δ*mga* cells did not synthesize Mga*Spn* (Fig. 6; lane 2). By primer extension, we examined the activity of the chromosomal *P1623A* and *P1623B* promoters in R6Δ*mga* cells carrying either pDL287 (absence of Mga*Spn*) (Fig. 7; lane 3) or pDL*PsulA*::*mga* (overproduction of Mga*Spn*) (lane 4). Again, a mix of the 5´-labeled PDA and PErpoE primers was used. Compared to R6/pDL287 cells (lane 1; low levels of Mga*Spn*), the 191-nt product (*P1623B*) was not detected in R6Δ*mga*/pDL287 cells (lane 3; absence of Mga*Spn*), although no changes were found in the amount of the 106-nt (*P1623A*) and 231-nt (*PrpoE*) cDNAs. Thus, in the absence of Mga*Spn* the activity of the *P1623B* promoter decreased without affecting the activity of the *P1623A* promoter. However, unexpectedly, the activity of the *P1623B* promoter on the R6Δ*mga* genome did not change in the presence of pDL*PsulA*::*mga* (lane 4; overproduction of Mga*Spn*). These results suggested that the genome of the R6Δ*mga* strain lacked not only the *mgaSpn* gene (including the *Pmga* promoter) but also a site required for Mga*Spn*-mediated activation of promoter *P1623B* (see below).

Mapping the site required for Mga*Spn*-mediated activation of the *P1623B* promoter. The promoter-probe vector pAST (32) (Fig. 8A) carries a multiple cloning

375 site between the *T1-T2* tandem terminators of the *E. coli rrnB* ribosomal RNA operon
376 and a promoter-less *gfp* allele. Moreover, the *T1-T2* terminators (*T1T2rrnB* region)
377 are located downstream of the *tetL* gene, which confers resistance to tetracycline
378 (19). Transcription of the *tetL* gene terminates efficiently at the *T1T2rrnB* region (32).
379 To delimit the site required for Mga*Spn*-mediated activation of the *P1623B* promoter,
380 a deletion analysis was carried out. Three chromosomal regions were inserted
381 independently into the *Sac*I site of pAST (Fig. 8A): (i) the *PAB* region (coordinates
382 1598304 to 1598600; pAST-*PAB*); (ii) the *PAB*Δ84 region (coordinates 1598388 to
383 1598600; pAST-*PAB*Δ84) and (iii) the *PAB*Δ153 region (coordinates 1598457 to
384 1598600; pAST-*PAB*Δ153). In these constructions, *gfp* expression was under the
385 control of both promoters, *P1623A* and *P1623B*. Thus, the promoter activity of each
386 chromosomal region was evaluated by fluorescence assays (Fig. 8A). The promoter
387 activity of the *PAB* and *PAB*Δ84 regions was 2-fold higher in R6 cells (low levels of
388 Mga*Spn*) than in R6Δ*mga* cells (absence of Mga*Spn*). However, the promoter activity
389 of the *PAB*Δ153 region was similar in both genetic backgrounds. These results
390 indicated that the region spanning the coordinates 1598388 and 1598457 contained
391 sequences that were required for Mga*Spn*-mediated activation of promoters *P1623A*
392 and/or *P1623B*.

393 The promoter activity of the *PAB*Δ84 and *PAB*Δ153 regions was further
394 examined by primer extension (Fig. 9). A mix of the 5´-labeled INTgfp and ASTtetL
395 primers was used. They anneal to *gfp* and *tetL* transcripts, respectively. The *tetL*
396 gene of pAST was used as internal control. Using RNA from R6 cells (low levels of
397 Mga*Spn*) harbouring pAST-*PAB*Δ84 (Fig. 9; lane 2), three cDNA products of 102-nt

(*PtetL* promoter), 111-nt (*P1623A* promoter) and 196-nt (*P1623B* promoter) were synthesized. Unlike the 102-nt and 111-nt cDNAs, the amount of the 196-nt cDNA decreased 5-fold when RNA from R6 cells harbouring pAST-*PAB*Δ153 was used (lane 3). Thus, in R6 cells (low levels of Mga*Spn*), deletion of the region that spans the 1598388 and 1598457 coordinates (see Fig. 8A) reduced the activity of the *P1623B* promoter but not the activity of the *P1623A* promoter. The specific decrease in the activity of promoter *P1623B* was also observed in R6Δ*mga* cells (absence of Mga*Spn*) carrying either *PAB*Δ84 (Fig. 9; lane 4) or *PAB*Δ153 (lane 5). These results demonstrated that Mga*Spn* was able to activate, directly or indirectly, the *P1623B* promoter *in vivo*. This activation required sequences located within the region 1598388-1598457 (see Fig.1 and Fig. 8A). Such a 70-bp region maps between the *Pmga* and *P1623B* divergent promoters, just 50-bp upstream of the *P1623B* transcription start site (coordinate 1598507).

Mga*Spn* does not influence the activity of promoter *Pmga in vivo*. Plasmid pAST2 (named pAS-*T2T1rrnB* in (32)) carries a multiple cloning site upstream of the promoter-less *gfp* gene (see Fig. 8B). Compared to the promoter-probe vector pAST, pAST2 carries the *T1-T2* terminators (*T1T2rrnB* region) inserted in the opposite orientation (*T2T1rrnB* region). The *T2T1rrnB* region functions as a transcriptional terminator signal, although it is not as efficient as the *T1T2rrnB* region (32). The site required for Mga*Spn*-mediated activation of the *P1623B* promoter is located around 80-150 nucleotides upstream of the *Pmga* transcription start site (coordinate 1598308). This fact suggested that Mga*Spn* might also influence the activity of the *Pmga* promoter *in vivo*. To test this hypothesis, the *PAB* chromosomal region

421 (coordinates 1598304-1598600), which carries the *P1623A, P1623B* and *Pmga*
422 promoters, was inserted into the *Sac*I site of pAST2 generating the pAST2-*Pmga*
423 recombinant (Fig. 8B). In this construction, *gfp* expression is under the control of the
424 *Pmga* promoter. Fluorescence assays showed that the activity of the *Pmga* promoter
425 located on pAST2-*Pmga* was similar in R6 and R6Δ*mga* cells. Therefore, Mga*Spn* did
426 not influence the activity of the *Pmga* promoter under our bacterial growth conditions.

427 **Mga*Spn*-His binds to a site located upstream of promoter *P1623B*.** To
428 determine whether Mga*Spn* was able to interact with the *P1623B* promoter region, we
429 performed DNase I footprinting assays with a His-tagged Mga*Spn* protein (Mga*Spn*-
430 His). This variant of Mga*Spn* carries six additional His residues at the C-terminal end.
431 We used a 222-bp DNA fragment (coordinates 1598298-1598519 of R6) that
432 contained the 70-bp region (1598388-1598457) known to be required for Mga*Spn*-
433 mediated activation of the *P1623B* promoter (Fig. 10C). Such a DNA fragment was
434 radioactively labeled either at the 5' end of the coding strand (Fig. 10A) or at the 5'
435 end of the non-coding strand (Fig. 10B). On the coding strand and in the presence of
436 Mga*Spn*-His, the region spanning the –52 and –90 positions relative to the
437 transcription start site of the *P1623B* promoter was protected against DNase I
438 digestion. On the non-coding strand, changes in the DNase I sensitivity (diminished
439 cleavages) were observed from –57 to –79 and from –83 to –102. Moreover, the –82
440 and –104 positions were slightly more sensitive to DNase I cleavage. We conclude
441 that Mga*Spn* interacts with sequences located between the positions –52 and –102
442 relative to the *P1623B* transcription start site (Fig. 10C). Such sequences are
443 included within the region shown to be required for Mga*Spn*-mediated activation of

444 the *P1623B* promoter. Hence, the Mga*Spn* regulator activates directly the *P1623B*
445 promoter.
446
447 **DISCUSSION**

448 In pathogenic bacteria, global transcriptional regulators whose activity and/or
449 intracellular concentration change in response to external stimuli are crucial during
450 the infection process. Some of these response regulators are associated to a
451 membrane-bound sensor histidine kinase, the so-called two-component signal
452 transduction systems (1, 29). Also, various 'stand-alone' response regulators, whose
453 sensory elements remain unidentified, have been implicated in the regulation of
454 virulence gene expression (23). To this class of global regulators belongs the Mga
455 protein of GAS (12). In the G+ bacterium *S. pneumoniae*, several two-component
456 systems are known to contribute to its virulence, although some of them to different
457 extent depending on the strain and/or infection model used (26). Moreover, signature-
458 tagged mutagenesis in the pneumococcal TIGR4 strain revealed that other putative
459 transcriptional regulators might control the expression of specific virulence genes (8).
460 It was the case of the *sp1800* gene product, an Mga orthologue that was shown to act
461 as a repressor of the *rlrA* pathogenicity islet (9). Here, we have performed a
462 transcriptional analysis of the region that spans the 1596789 and 1600589
463 coordinates of the pneumococcal R6 genome (14). Such a region contains the
464 *spr1622* gene (*mgaSpn* in this work), which is equivalent to the *sp1800* gene of the
465 TIGR4 strain, and four divergent ORFs (*spr1623*-*spr1626*) that are highly conserved
466 in TIGR4. We have identified the promoter of the *mgaSpn* gene (*Pmga*) and

467 demonstrated that the four ORFs constitute an operon that is transcribed from two
468 promoters (*P1623A* and *P1632B*). Furthermore, we have shown, for the first time, that
469 the pneumococcal Mga-like protein is able to act as a transcriptional activator: (i)
470 Mga*Spn* activates the *P1623B* promoter *in vivo* and, therefore, the expression of the
471 *spr1623-spr1626* operon, (ii) this activation requires sequences located upstream of
472 the *P1623B* promoter, and (iii) Mga*Spn* interacts with such sequences *in vitro*.
473 Hemsley *et al.* (9) reported that the *sp1800* gene product of the TIGR4 strain did not
474 affect transcription of the neighbouring cluster of genes. This discrepancy with our
475 results might be due to the use of different pneumococcal strains and/or to the use of
476 different bacterial growth conditions.

477 The pneumococcal Mga*Spn* regulator is homologous (42.6% similarity and
478 21.4% identity) to the Mga regulator of GAS. In Mga, two helix-turn-helix domains
479 were mapped near the N-terminus of the protein. Both of them are known to be
480 required for DNA binding and transcriptional activation (24, 36). In addition, Mga is
481 known to bind to regions located upstream of its target promoters. The position of the
482 Mga binding site with respect to the start of transcription varies among the promoters
483 tested (12). In the case of the Mga*Spn* regulator, analysis of its amino acid sequence
484 using the Pfam database (4) revealed that it also has two putative DNA-binding
485 domains within the N-terminal region, the so-called HTH_Mga (residues 6-65) and
486 Mga (residues 71-158) domains. Moreover, we have shown that Mga*Spn* interacts
487 with sequences located upstream of the *P1623B* promoter. Such sequences are also
488 needed for Mga*Spn*-mediated activation of the *P1623B* promoter. Thus, the Mga and
489 Mga*Spn* regulators might have similar DNA binding properties.

Using the BLASTN 2.2.25+ nucleotide sequence alignment program (40), we have analyzed the region that spans the *P1623A* and *Pmga* promoters of the R6 genome (see Fig. 1), which includes promoter *P1623B* and the Mga*Spn* binding site identified in this work. This analysis has revealed that such a region is identical in ten pneumococcal strains whose genome has been totally sequenced (R6, D39**,** TIGR4, G54, JJA, P1031, TCH8431/19A, CGSP14, Taiwan19F-14 and 70585). Furthermore, according to protein sequence database similarity searches, Mga*Spn* is highly conserved among the above pneumococcal strains. Compared to R6 and D39, Mga*Spn* has only two amino acid changes (I309M and V358I) in the TIGR4, G54, JJA, P1031, TCH8431/19A and CGSP14 strains, three amino acid changes (C280Y, I309M and V358I) in the Taiwan19F-14 strain, and three amino acid changes (I309M, V358I and P450S) in the 70585 strain.

The pneumococcal R6 strain has an additional *mga*-like gene (*spr1404*). It is adjacent to the divergent *spr1403* gene that encodes a collagen-like protein (PclA) (28). Both genes are absent in TIGR4 (2). Although the *spr1404* gene product has homology to the Mga*Spn* (59.3% similarity) and Mga (40% similarity) regulators, Paterson *et al*. (28) reported that single deletion mutants lacking either *spr1404* or *spr1403* were not attenuated in a mouse model of invasive pneumoniae. Thus, as pointed out by the authors, further work is required to elucidate whether the *spr1404* gene has a significant role in pathogenesis.

The function of the *spr1623-spr1626* operon is unknown. It encodes products of 188 (*spr1623*; hypothetical protein), 56 (*spr1624*; putative lipoprotein), 202 (*spr1625*; putative general stress protein 24) and 67 (*spr1626*; hypothetical protein) residues, respectively. According to the Protein Clusters database (16), the *spr1624*, *spr1625*

514 and *spr1626* products are identical in the ten pneumococcal strains mentioned above.
515 However, unlike R6 and D39, the operon of the other strains has an additional ORF
516 (named *sp1801* in TIGR4). It is located upstream of the equivalent *spr1623* ORF and
517 would encode a product of 54 residues (hypothetical protein, putative
518 transglycosylase associated protein). The absence of this ORF in R6 (and D39) is
519 due to the deletion of one nucleotide between coordinates 1598751 and 1598752,
520 which results in a truncated ORF that would encode a product of 20 amino acids.
521 Several observations suggest that the *spr1623-spr1626* operon might play a role
522 in virulence. First, the product (202 amino acids) of the *spr1625* gene has homology
523 (69% similarity) to the product of the *Enterococcus faecalis gls24* gene (EF0080 in
524 strain V583; 186 amino acids), which was shown to be a general stress-inducible
525 gene involved in bile-salts resistance (6). Also, it was shown to be important for
526 virulence using a mouse peritonitis model (33). Second, Hemsley *et al.* (9) reported
527 the characterization of a TIGR4 mutant strain (STM206) that carries a transposon
528 inserted ~300 bp upstream of the predicted translation start codon of the *sp1800*
529 gene. This mutant strain was attenuated for both nasopharyngeal carriage and lung
530 infection in murine models. Moreover, it was much more affected in virulence than a
531 mutant strain (AC1272) that carries a transposon inserted into the coding sequence of
532 the *sp1800* gene (*mgaSpn* gene in R6). Now, we have shown that the distance
533 between the translation start codon of the *mgaSpn* gene (coordinate 1598270) and
534 the transcription initiation sites of the *spr1623-spr1626* operon is 323 (from the
535 *P1623A* promoter; coordinate 1598592) and 238 nucleotides (from the *P1623B*
536 promoter; coordinate 1598507), respectively. Hence, the transposon in strain

537 STM206 could be affecting the expression of the *sp1801-sp1805* operon (*spr1623-*
538 *spr1626* operon in R6). If this were the case, the attenuation phenotype of the
539 STM206 mutant strain (Hemsley *et al.*, 2003) would indicate an important role of the
540 operon in pneumococcal virulence.

541 In summary, this is the first report on the activator role of the pneumococcal
542 Mga-like regulator (Mga*Spn*). This regulatory protein activates the transcription of a
543 four-gene operon from a site located upstream of the target promoter.

544

545

546

547 **ACKNOWLEDGMENTS**

548 Thanks are due to Lorena Rodríguez for her excellent technical assistance.

549 Work supported by grants CSD2008-00013-INTERMODS to M.E. and BFU2009-
550 11868 to A.B. from the Spanish Ministry of Science and Innovation, and grant PIE-
551 201020E030 to A.B. from the Spanish National Research Council. V.S.-C was
552 recipient of a fellowship (BES-2007-17086) from the Spanish Ministry of Science and
553 Innovation.

554

REFERENCES

1. **Beier, D., and R. Gross.** 2006. Regulation of bacterial virulence by two-component systems. Curr. Opin. Microbiol. **9:**143-152.
2. **Brückner, R., M. Nuhn, P. Reichmann, B. Weber, and R. Hakenbeck.** 2004. Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. Int. J. Med. Microbiol. **294:**157-168.
3. **Cunningham, M. W.** 2000. Pathogenesis of group A streptococcal infections. Clin. Microbiol. Rev. **13:**470-511.
4. **Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H.-R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. L. Sonnhammer, and A. Bateman.** 2008. The Pfam protein families database. Nucl. Acids Res. **36:**D281-288.
5. **Geyer, A., and K. H. Schmidt.** 2000. Genetic organisation of the M protein region in human isolates of group C and G streptococci: two types of multigene regulator-like (mgrC) regions. Mol. Gen. Genet. **262:**965-976.
6. **Giard, J. C., A. Rince, H. Capiaux, Y. Auffray, and A. Hartke.** 2000. Inactivation of the stress- and starvation-inducible *gls24* operon has a pleiotrophic effect on cell morphology, stress sensitivity, and gene expression in *Enterococcus faecalis*. J. Bacteriol. **182:**4512-4520.
7. **Haugen, S. P., W. Ross, and R. L. Gourse.** 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat. Rev. Microbiol. **6:**507-516.
8. **Hava, D. L., and A. Camilli.** 2002. Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. Mol. Microbiol. **45:**1389-1406.

9. **Hemsley, C., E. Joyce, D. L. Hava, A. Kawale, and A. Camilli.** 2003. MgrA, an orthologue of Mga, acts as a transcriptional repressor of the genes within the *rlrA* pathogenicity Islet in *Streptococcus pneumoniae*. J. Bacteriol. **185:**6640-6647.

10. **Hendriksen, W. T., H. J. Bootsma, A. van Diepen, S. Estevão, O. P. Kuipers, R. de Groot, and P. W. M. Hermans.** 2009. Strain-specific impact of PsaR of *Streptococcus pneumoniae* on global gene expression and virulence. Microbiology **155:**1569-1579.

11. **Hendriksen, W. T., N. Silva, H. J. Bootsma, C. E. Blue, G. K. Paterson, A. R. Kerr, A. de Jong, O. P. Kuipers, P. W. Hermans, and T. J. Mitchell.** 2007. Regulation of gene expression in *Streptococcus pneumoniae* by response regulator 09 is strain dependent. J. Bacteriol. **189:**1382-1389.

12. **Hondorp, E. R., and K. S. McIver.** 2007. The Mga virulence regulon: infection where the grass is greener. Mol. Microbiol. **66:**1056-1065.

13. **Horinouchi, S., and B. Weisblum.** 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. J. Bacteriol. **150:**815-825.

14. **Hoskins, J., W. E. Alborn, Jr., J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D.-J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P.-M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, Jr.,**

P. L. Skatrud, and J. I. Glass. 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J. Bacteriol. **183:**5709-5717.

15. **Kadioglu, A., J. N. Weiser, J. C. Paton, and P. W. Andrew.** 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. Nat. Rev. Microbiol. **6:**288-301.

16. **Klimke, W., R. Agarwala, A. Badretdin, S. Chetvernin, S. Ciufo, B. Fedorov, B. Kiryutin, K. O'Neill, W. Resch, S. Resenchuk, S. Schafer, I. Tolstoy, and T. Tatusova.** 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucl. Acids Res. **37:**D216-223.

17. **Lacks, S. A.** 1966. Integration efficiency and genetic recombination in pneumococcal transformation. Genetics **53:**207-235.

18. **Lacks, S. A., B. Greenberg, and P. López.** 1995. A cluster of four genes encoding enzymes for five steps in the folate biosynthetic pathway of *Streptococcus pneumoniae*. J. Bacteriol. **177:**66-74.

19. **Lacks, S. A., P. López, B. Greenberg, and M. Espinosa.** 1986. Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. J. Mol. Biol. **192:**753-765.

20. **Lanie, J. A., W. L. Ng, K. M. Kazmierczak, T. M. Andrzejewski, T. M. Davidsen, K. J. Wayne, H. Tettelin, J. I. Glass, and M. E. Winkler.** 2007. Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. J. Bacteriol. **189:**38-51.

21. **LeBlanc, D. J., Y. Y. Chen, and L. N. Lee.** 1993. Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380-1.

626 Plasmid **30:**296-302.
627 22. **McCluskey, J., J. Hinds, S. Husain, A. Witney, and T. J. Mitchell.** 2004. A
628 two-component system that controls the expression of pneumococcal surface
629 antigen A (PsaA) and regulates virulence and resistance to oxidative stress in
630 *Streptococcus pneumoniae*. Mol. Microbiol. **51:**1661-1675.
631 23. **McIver, K. S.** 2009. Stand-alone response regulators controlling global virulence
632 networks in *Streptococcus pyogenes*. Collin, M., Schuch, R. (eds). Bacterial
633 Sensing and Signaling. Contrib. Microbiol. Basel, Karger **16:**103-119.
634 24. **McIver, K. S., and R. L. Myles.** 2002. Two DNA-binding domains of Mga are
635 required for virulence gene activation in the group A streptococcus. Mol.
636 Microbiol. **43:**1591-1601.
637 25. **Miller, W. G., and S. E. Lindow.** 1997. An improved GFP cloning cassette
638 designed for prokaryotic transcriptional fusions. Gene **191:**149-153.
639 26. **Paterson, G. K., C. E. Blue, and T. J. Mitchell.** 2006. Role of two-component
640 systems in the virulence of *Streptococcus pneumoniae*. J. Med. Microbiol.
641 **55:**355-363.
642 27. **Paterson, G. K., and T. J. Mitchell.** 2006. The role of *Streptococcus*
643 *pneumoniae* sortase A in colonisation and pathogenesis. Microbes Infect. **8:**145-
644 153.
645 28. **Paterson, G. K., L. Nieminen, J. M. Jefferies, and T. J. Mitchell.** 2008. PclA,
646 a pneumococcal collagen-like protein with selected strain distribution,
647 contributes to adherence and invasion of host cells. FEMS Microbiol. Lett.
648 **285:**170-176.
649 29. **Perry, J., K. Koteva, and G. Wright.** 2011. Receptor domains of two-

component signal transduction systems. Mol. Biosyst. **7:**1388-1398.

30. **Ribardo, D. A., and K. S. McIver.** 2006. Defining the Mga regulon: comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus. Mol. Microbiol. **62:**491-508.

31. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16:**276-277.

32. **Ruiz-Cruz, S., V. Solano-Collado, M. Espinosa, and A. Bravo.** 2010. Novel plasmid-based genetic tools for the study of promoters and terminators in *Streptococcus pneumoniae* and *Enterococcus faecalis*. J. Microbiol. Methods **83:**156-163.

33. **Teng, F., E. C. Nannini, and B. E. Murray.** 2005. Importance of *gls24* in virulence and stress response of *Enterococcus faecalis* and use of the Gls24 protein as a possible immunotherapy target. J. Infect. Dis. **191:**472-480.

34. **Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser.** 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science **293:**498-506.

35. **Tsvetanova, B., A. C. Wilson, C. Bongiorni, C. Chiang, J. A. Hoch, and M. Perego.** 2007. Opposing effects of histidine phosphorylation regulate the AtxA

virulence transcription factor in *Bacillus anthracis*. Mol. Microbiol. **63:**644-655.

36. **Vahling, C. M., and K. S. McIver.** 2006. Domains required for transcriptional activation show conservation in the Mga family of virulence gene regulators. J. Bacteriol. **188:**863-873.

37. **van der Poll, T., and S. M. Opal.** 2009. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet **374:**1543-1556.

38. **Vasi, J., L. Frykberg, L. E. Carlsson, M. Lindberg, and B. Guss.** 2000. M-like proteins of *Streptococcus dysgalactiae*. Infect. Immun. **68:**294-302.

39. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103-119.

40. **Zhang, Z., S. Schwartz, L. Wagner, and W. Miller.** 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. **7:**203-214.

FIGURE LEGENDS

FIG. 1. Genetic map of the region spanning the 1596789 and 1600589 coordinates of the R6 genome (14). Gene *spr1622* has been named *mgaSpn* in this work. For each ORF, the coordinates of the predicted start and stop codons are indicated. The nucleotide sequence of the region spanning the start codon of the *mgaSpn* gene (coordinate 1598270) and the start codon of the *spr1623* ORF (coordinate 1598960) is shown. The putative Shine-Dalgarno sequence (SD) of *mgaSpn* is indicated. The main sequence elements (-35 box, -10 box and extended –10 box) of the promoters identified in this work (*Pmga*, *P1623B* and *P1623A*), as well as the transcription start sites (+1 position), are indicated.

FIG. 2. Transcription of *mgaSpn in vivo*. RT-PCR assays were performed using RNA from R6 cells. The position of the oligonucleotides used (1622A, 1622C and C1622D) is shown. RT-PCR reactions (lane R) were subjected to agarose (0.8%) gel electrophoresis. RT-PCR reactions without adding the reverse transcriptase were performed as negative control (lane N). The size of PCR-amplified DNA fragments (1 and 2) using genomic DNA as template (lane P, positive control) is indicated. The size (in bp) of DNA fragments (lane M) used as molecular weight markers (HyperLadder I, Bioline) is indicated on the right of the gel.

FIG. 3. Promoter *Pmga* is functional *in vivo*. Primer extension reactions were carried out on total RNA isolated from R6 cells harbouring plasmid pAS-*Pmga*. The *gfp* gene carries translation initiation signals optimized for prokaryotes (SD) (25). The *tetL* gene confers resistance to tetracycline. The main sequence elements of the *Pmga* promoter (grey boxes) and the ATG initiation codon of the *gfp* gene (black box) are

711 indicated. *Bam*HI sites are underlined. The asterisks indicate the 3´-end of the cDNA
712 products synthesized using the INTgfp primer. The size of the cDNA products (lane
713 R) is indicated in nucleotides on the right of the gel. Dideoxy-mediated chain-
714 termination sequencing reactions using pLS1 DNA (19) and the F-pLS1 primer (5´-
715 TGCTGGCAGGCACTGGC-3´; coordinates 802-818 of pLS1) were used as DNA size
716 markers (lanes A, C, G, T).

717 **FIG. 4.** The *spr1623-spr1626* ORFs constitute an operon. RT-PCR assays were
718 performed using RNA from R6 cells. The position of the oligonucleotides used
719 (1623A, 1623B, 1623C and 1626A) is shown. RT-PCR reactions (lane R) were
720 analyzed by agarose (0.8%) gel electrophoresis. As negative control (lane N), RT-
721 PCR reactions were carried out without adding the reverse transcriptase. The size of
722 PCR-amplified DNA fragments (1, 2 and 3) using genomic DNA as template (lane P,
723 positive control) is indicated. Lane M: DNA fragments used as molecular weight
724 markers (in bp) (HyperLadder I, Bioline).

725 **FIG. 5.** The *spr1623-spr1626* operon is transcribed from promoters *P1623A* and
726 *P1623B*. Primer extension assays were performed using RNA from R6 cells and the
727 PDA (left gel) or PDB (right gel) primers. The size of the cDNA products (lane R) is
728 indicated in nucleotides on the right of the gels. A, C, G, T sequence ladders were
729 used as DNA size markers. They were prepared using M13mp18 DNA (39) and the –
730 40 M13 primer (5´-GTTTTCCCAGTCACGAC-3´) (left gel) or a PCR-amplified DNA
731 fragment from the *E. faecalis* V583 genome and the Rev primer (5´-
732 GATTTCTTCAATTTGTTCCATC-3´) (right gel). The asterisks in the scheme below
733 the gels indicate the transcription start sites identified in this study.

734 **FIG. 6.** Detection of Mga*Spn* in pneumococcal cell extracts by Western blotting using
735 polyclonal antibodies against His-tagged Mga*Spn*. Total proteins from R6 cells (lane
736 1), R6Δ*mga* cells (lane 2) and pDL*PsulA::mga*-carrying cells (lane 3) were separated
737 by SDS-PAGE. His-tagged Mga*Spn* protein (6 ng) (lane 4) and pre-stained proteins
738 (Invitrogen) (not shown) were run in the same gel.

739 **FIG. 7.** Mga*Spn* mediates activation of the *P1623B* promoter. Primer extension
740 reactions were carried out using total RNA from R6/pDL287 (lanes 1, 5 and 6);
741 R6/pDL*PsulA::mga* (lane 2); R6Δ*mga*/pDL287 (lane 3) or R6Δ*mga*/pDL*PsulA::mga*
742 (lane 4) cells. 5´-labeled oligonucleotides were used as primers: a mix of the PDA and
743 PErpoE primers (lanes 1, 2, 3 and 4), primer PDA (lane 5) or primer PErpoE (lane 6).
744 The size of the cDNA products is indicated in nucleotides on the left of the gel: 106-nt
745 (*P1623A* promoter), 191-nt (*P1623B* promoter) and 231-nt (*PrpoE* promoter).
746 Sequence ladders were used as DNA size markers (lanes A, C, G, T). They were
747 prepared using a PCR-amplified DNA fragment from the *E. faecalis* V583 genome
748 and the Fw primer (5´-CGTTTGAGCAATATAATCGTTTG-3´).

749 **FIG. 8.** Fluorescence assays**.** (A) Activity of the *P1623A* (*PA*) and *P1623B* (*PB*)
750 promoters. The promoter-probe vector pAST was described (32). The position of the
751 *tetL* (tetracycline resistance) and *gfp* (green fluorescence protein) genes is indicated.
752 The *T1T2* box represents the tandem terminators *T1* and *T2* of the *E. coli rrnB*
753 ribosomal RNA operon. Grey boxes represent DNA fragments from the R6 genome.
754 (B) Activity of the *Pmga* promoter. Plasmid pAST2 was described (named pAS-
755 *T2T1rrnB* in Ruiz-Cruz *et al*. (32)). Compared to pAST, it carries the *T1T2rrnB* region
756 inserted in the opposite orientation (box *T2T1*). The position of promoter *Pmga* is

757 shown. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture
758 (OD_{650}=0.3). In each case, three independent cultures were analyzed.

759 **FIG. 9.** Genomic region needed for Mga*Spn*-mediated activation of the *P1623B*
760 promoter. Primer extension reactions were carried out using total RNA from
761 R6/pAST-*PAB*Δ84 (lanes 1 and 2), R6/pAST-*PAB*Δ153 (lane 3), R6Δ*mga*/pAST-
762 *PAB*Δ84 (lane 4) or R6Δ*mga*/pAST-*PAB*Δ153 (lane 5) cells. 5´-labeled
763 oligonucleotides were used as primers: a mix of the INTgfp and ASTtetL primers
764 (lanes 2, 3, 4 and 5) or the ASTtetL primer (lane 1). The size of the cDNA products is
765 indicated in nucleotides on the left of the gel: 102-nt (*PtetL* promoter), 111-nt (*P1623A*
766 promoter) and 196-nt (*P1623B* promoter). Dideoxy-mediated chain-termination
767 sequencing reactions using pAST DNA and the INTgfp primer were run in the same
768 gel (lanes A, C, G, T).

769 **FIG. 10.** DNase I footprints of Mga*Spn*-His-DNA complexes. The 222-bp DNA
770 fragment (coordinates 1598298-1598519) was labeled at the 5' end of either the
771 coding (A) or the non-coding (B) strand. The labeled DNA (4 nM) was incubated with
772 the indicated concentrations of Mga*Spn*-His. Dideoxy-mediated chain termination
773 sequencing reactions were run in the same gel (lanes A, C, G, T). Densitometer
774 scans corresponding to DNA without protein (grey line) and DNA with protein (black
775 line; 240 nM in (A) and 200 nM in (B)) are shown. The Mga*Spn*-His-protected regions
776 are indicated with brackets. Arrowheads indicate positions that are slightly more
777 sensitive to DNase I cleavage. The indicated positions are relative to the transcription
778 start site of the *P1623B* promoter. (C) Nucleotide sequence of the region that spans
779 the coordinates 1598509 and 1598380 of the R6 genome. It includes the transcription

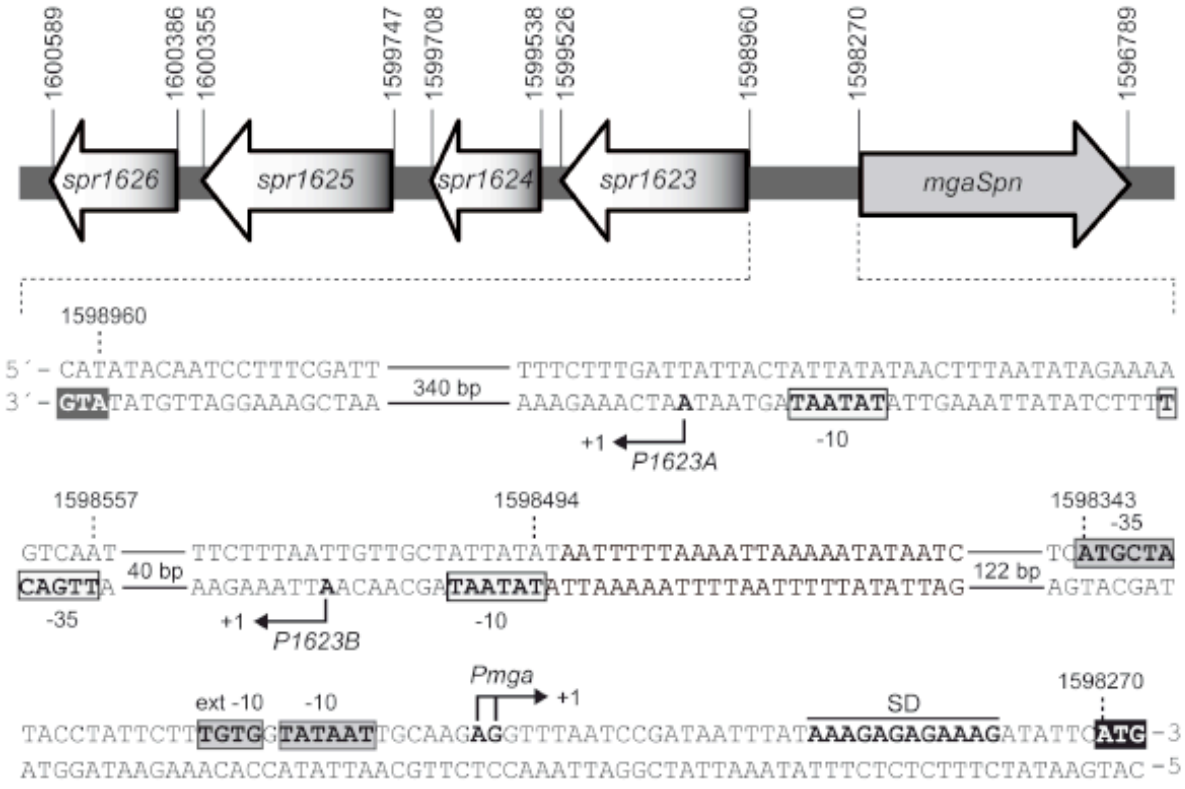
780 start site of the *P1623B* promoter (coordinate 1598507), the region required for
781 Mga*Spn*-mediated activation of the *P1623B* promoter (1598457-1598388) and the
782 site recognized by Mga*Spn*-His (brackets).
783
784

785

786 TABLE 1. Oligonucleotides used in this work.

Name	Sequence[a] (5´ to 3´)
1622A	AGTTCCTGATTGTATTCCCT
1622C	GATTCTGTATTCACGCCCTC
1622D	TTCTAATTGCCTATGACTTTTTTTAG
C1622D	CTAAAAAAAGTCATAGGCAATTAGA
INTgfp	CATCACCATCTAATTCAACAAG
PErpoE	GCCCAGCAAATACTTCTAATTCC
ASTtetL	GAGGGCAGACGTAGTTTATAGGG
1623A	GAGGGCGTGAATACAGAATC
1623B	CGTAAATTTACATGAACAGTTGGG
1623C	GGAGGGTAGGCAGTGTTGTGATC
1626A	GCACCTTCTACAGCGTCTTTAGCG
PDA	GTGATTTTACCTGCCAAGAGACC
PDB	GAAAAGTCAATTATTTCGATTG
PrSp1	ATAAATTATC**GGATCC**AACCTCTTGC
PrSp2	GAATTTGATTCT**GGATCC**ACGCCCTC
PmgaSac	CTTTATAAATTAT**GAGCTC**AAACCTCTTGC
PABSac	ATATCAAAAAATC**GAGCTC**TTTGATTATTAC
PABΔ84Sac	ATTTCGTATAA**GAGCTC**TACGGAGACAATATA
PABΔ153Sac	GAATACAGAATC**GAGCTC**AAGTCTAAAG
PsulNde	CAAGGATTTTCAT**CATATG**ATTTTTC
PsulCla	ACTGATTGTTA**ATCGAT**TTGCTTTCTGT
mgaNde	TGCAAGAGGTTT**CATATG**ATAATTTATAAAG
mgaCla	GTACATTTTTCTTA**ATCGAT**TGAAGGTCTTTTC
1622Nde	GAGAGAAAGATA**CATATG**AGAGATTTA
1622Xho-His	TTTTGTTATTTT**CTCGAG**CTCATCTAATCG
1622H	CGGATTAAACCTCTTGCAATTATACC
1622I	CAAATTCTTTAATTGTTGCTATTA

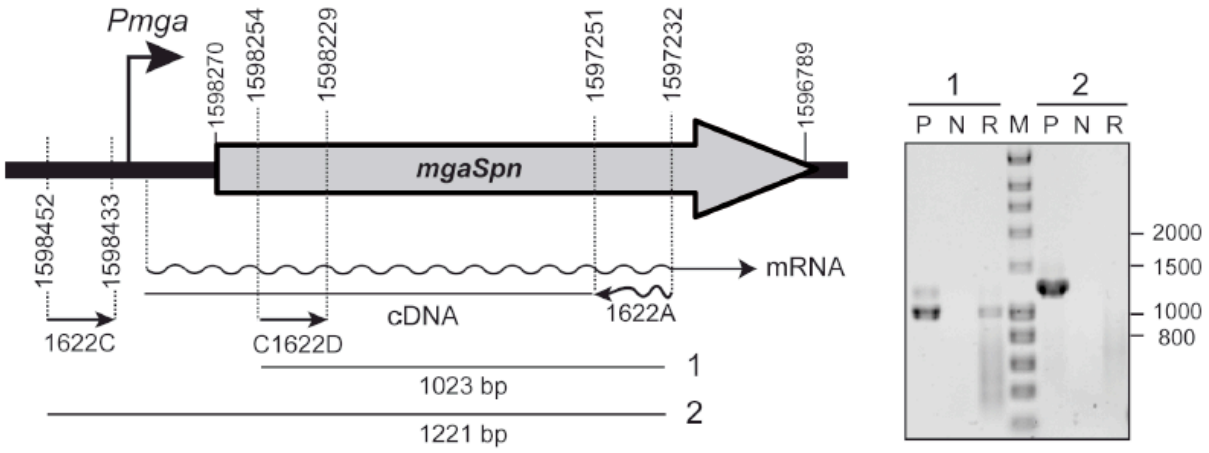
787 [a]Base changes that generate restriction sites (in bold) are underlined.



788

789 Figure 1

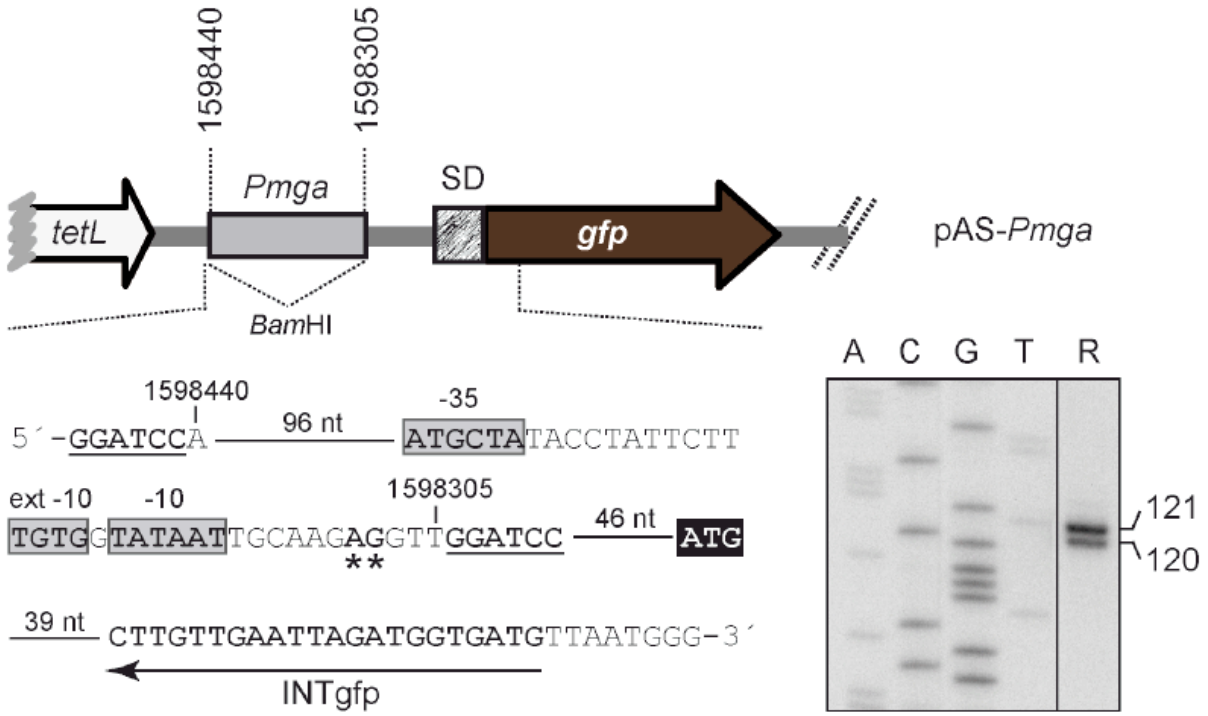
790



791

792 Figure 2

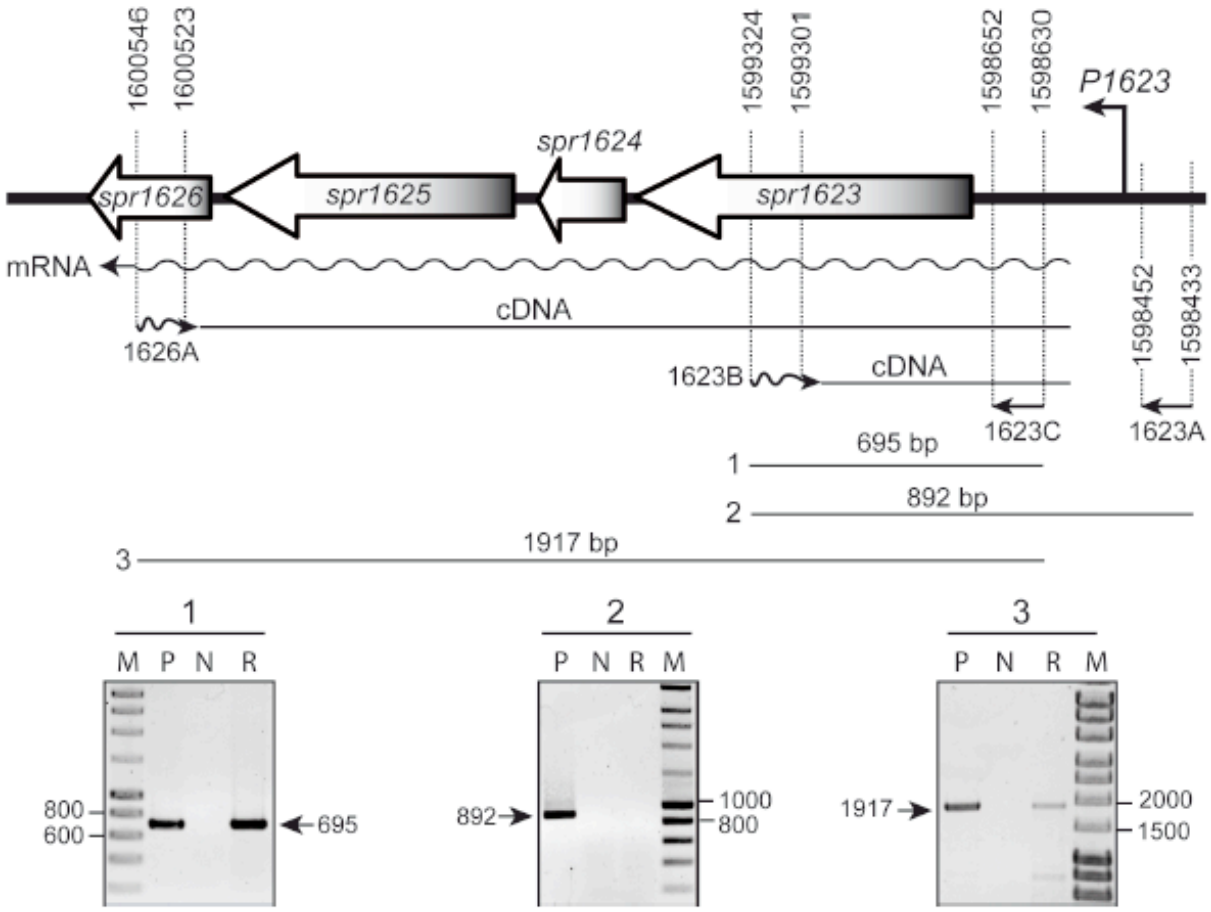
793



794

795 Figure 3

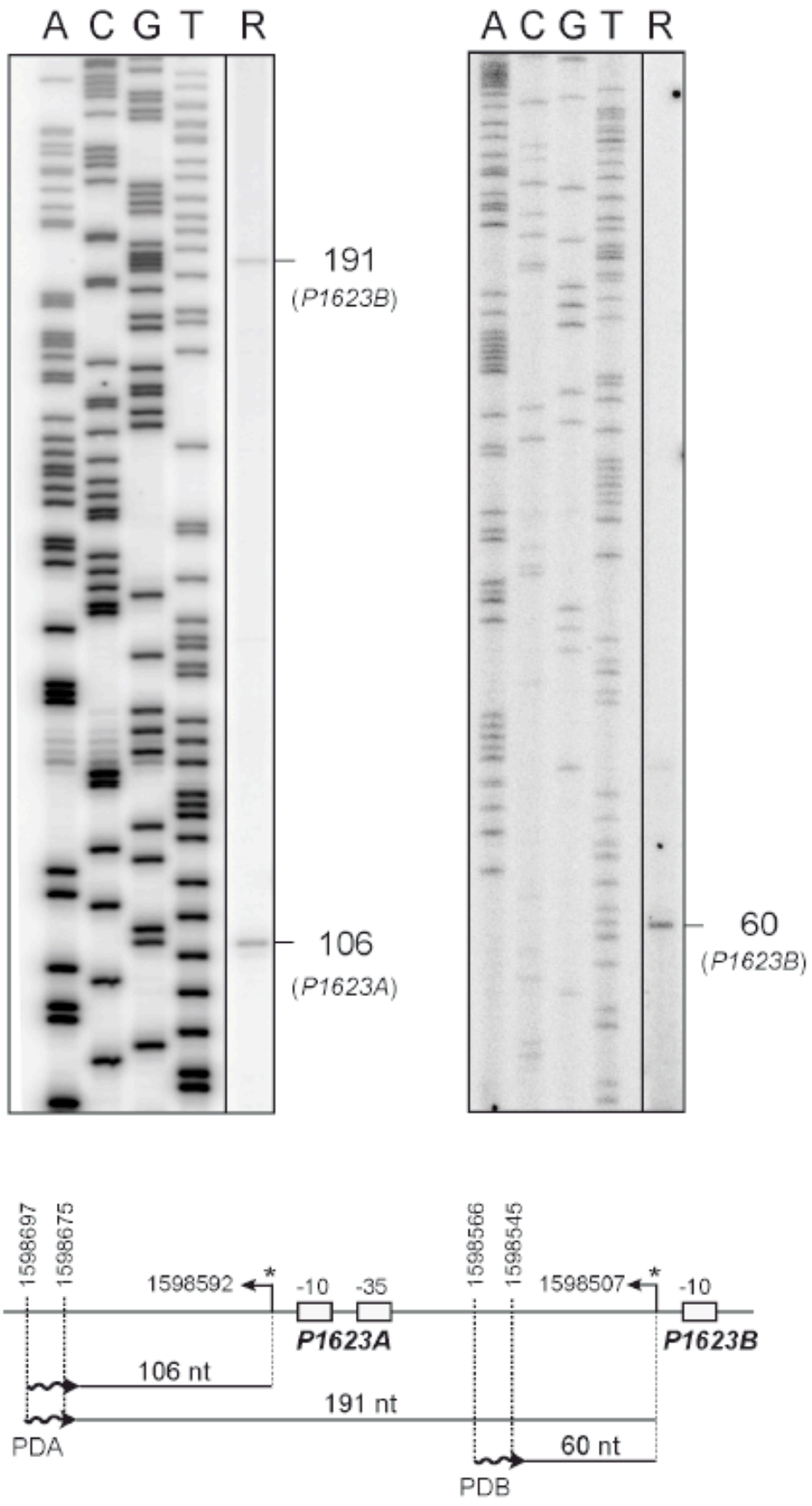
796



797

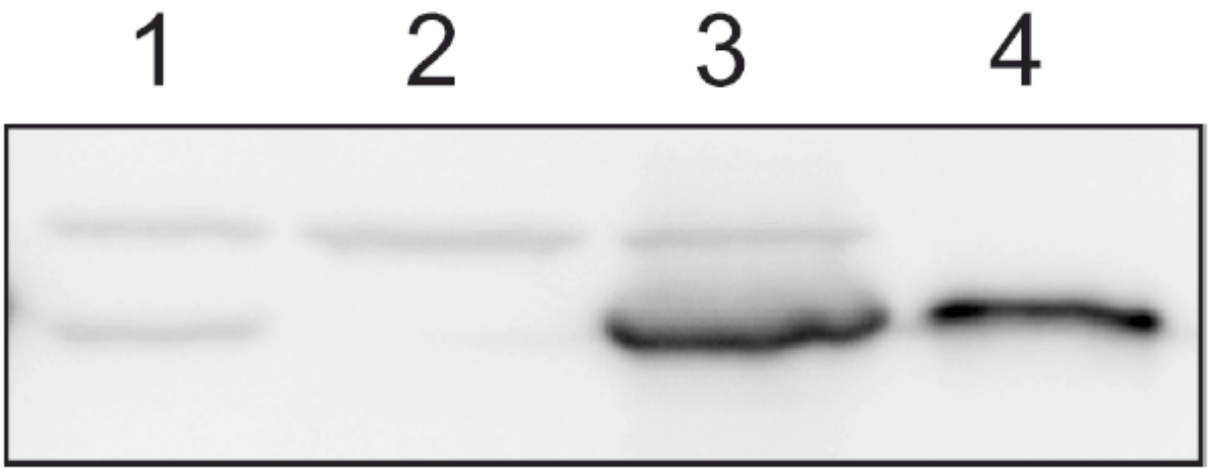
798 Figure 4

799


A C G T R
191
(P1623B)
106
(P1623A)
A C G T R
60
(P1623B)
1598697
1598675
1598592
-10
-35
P1623A
106 nt
191 nt
PDA
1598566
1598545
1598507
-10
P1623B
60 nt
PDB

800 Figure 5

801



802

803 Figure 6

804

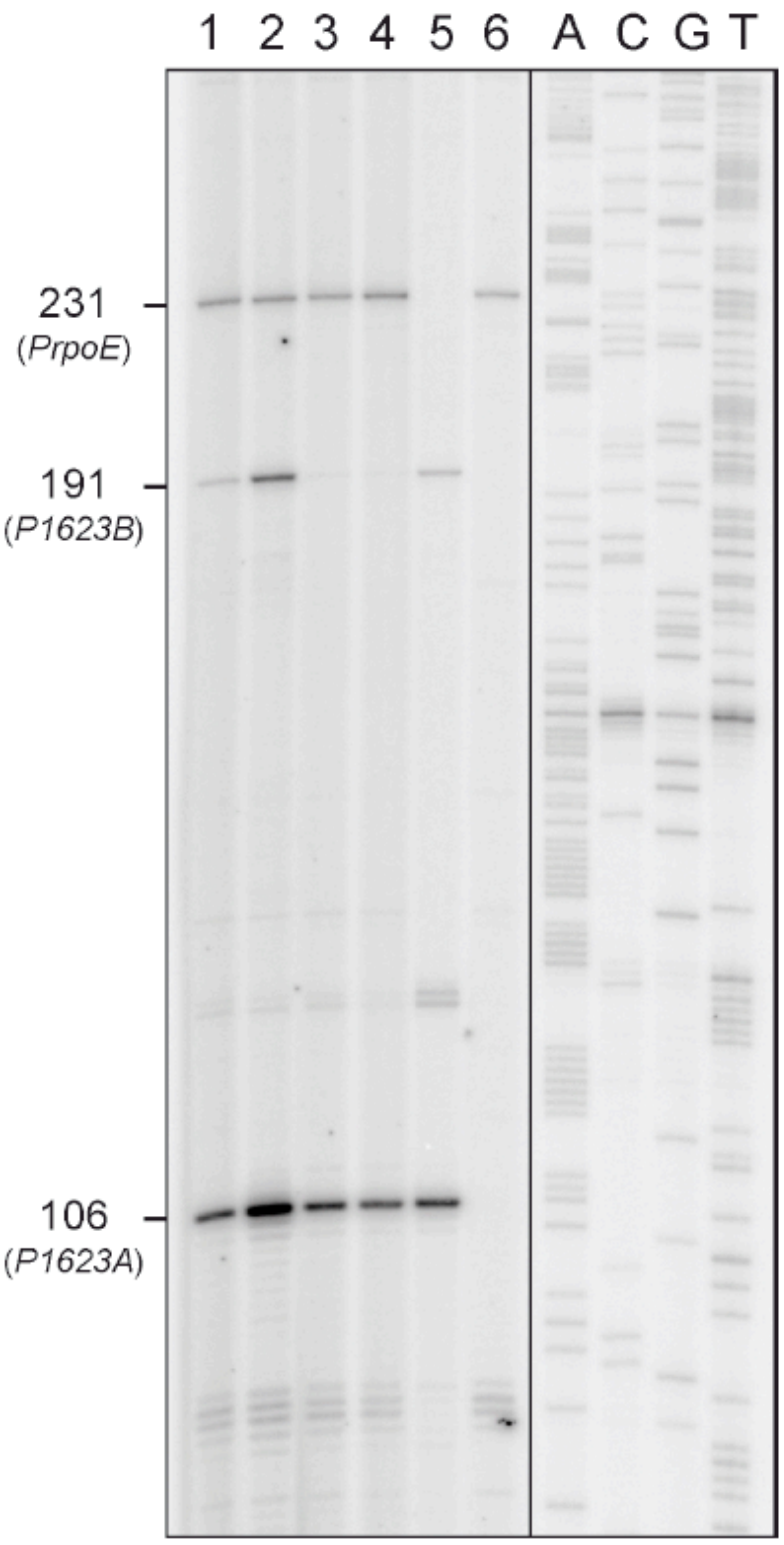
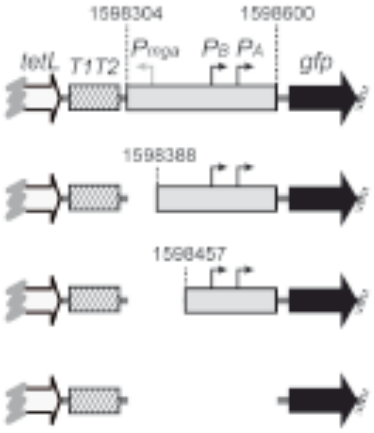


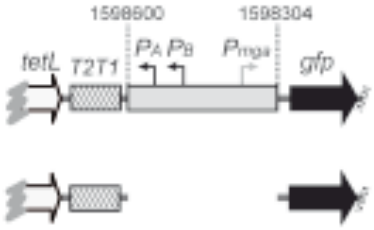
Figure 7

A



	Intensity of fluorescence	
	R6	R6Δ*mga*
pAST-*PAB*	10.75 ± 0.75	5.64 ± 0.38
pAST-*PAB*Δ84	13.03 ± 1.37	6.36 ± 0.24
pAST-*PAB*Δ153	7.26 ± 0.86	6.74 ± 0.08
pAST	0.23 ± 0.13	0.25 ± 0.04

B



	R6	R6Δ*mga*
pAST2-*Pmga*	2.17 ± 0.07	2.21 ± 0.29
pAST2	1.09 ± 0.07	0.89 ± 0.06

807

808 Figure 8

809

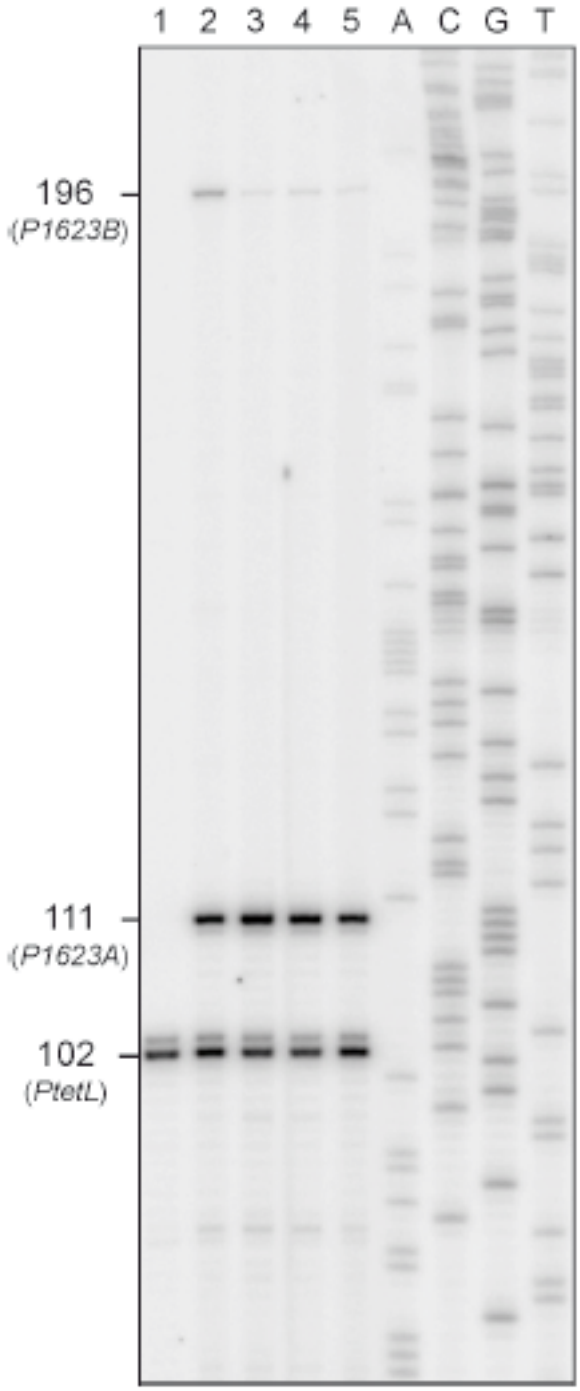
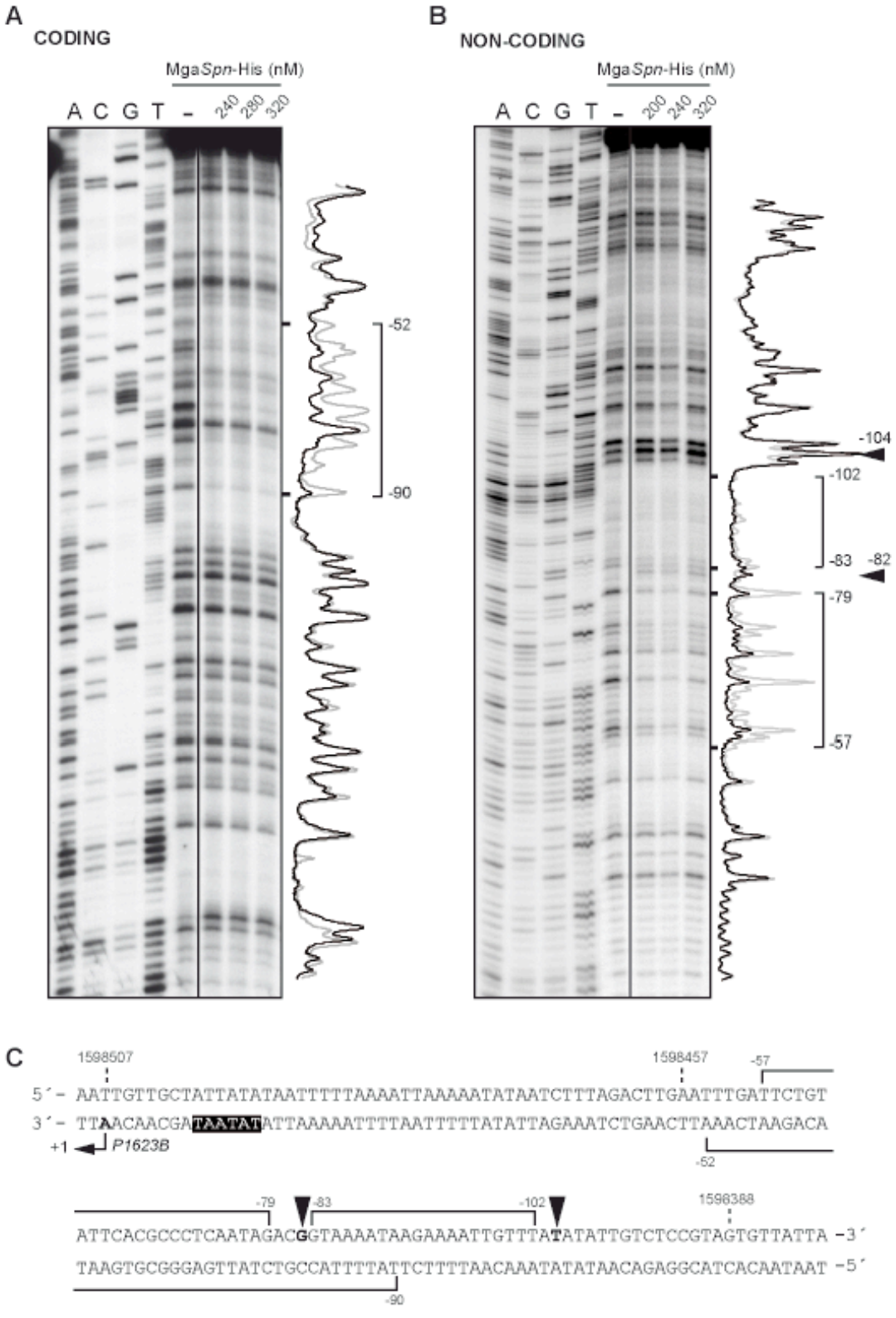


Figure 9



813

814 Figure 10