Transcriptional Activation of \(sclA\) by Mga Requires a Distal Binding Site in \textit{Streptococcus pyogenes}

Audry C. Almengor and Kevin S. McIver*

Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas

Received 16 July 2004/Accepted 30 August 2004

\textit{Streptococcus pyogenes} (the group A streptococcus [GAS]) is a medically significant pathogen of humans, causing a range of diseases from pharyngitis to necrotizing fasciitis. Several important GAS virulence genes are under the control of a pleiotropic regulator called Mga, or the multiple gene regulator of GAS, including the gene encoding the streptococcal collagen-like protein, or \(sclA\). Analysis of the genome sequence upstream of \(sclA\) revealed two potential Mga-binding sites with homology to the published Mga-binding element, which were called \(P_{sclA-I}\) (distal) and \(P_{sclA-II}\) (proximal) based on their location relative to a predicted start of transcription. Primer extension was used to confirm that the Mga-dependent transcriptional start site for \(sclA\) was located adjacent to the proximal \(P_{sclA-II}\) binding site. By using overlapping \(P_{sclA}\) promoter probes and purified Mga-His fusion protein, it was shown by electrophoretic mobility shift assays that, unlike other Mga-regulated promoters, Mga binds only to a distal DNA-binding site (\(P_{sclA-I}\)). Binding of Mga to \(P_{sclA-I}\) could be competed with cold probes corresponding to known Mga-regulated promoters (\(P_{emm}\), \(P_{scpA}\), and \(P_{mga} \)) but not with a nonspecific probe or the proximal \(P_{sclA-II}\) fragment. With the use of a plasmid-based green fluorescent protein transcriptional reporter system, the full-length \(P_{sclA}\) was not sufficient to reproduce normal Mga-regulated activation. However, studies using a single-copy \(gusA\) transcriptional reporter system integrated at the native \(sclA\) chromosomal locus clearly demonstrated that the distal \(P_{sclA-I}\) binding site is required for Mga regulation. Therefore, \(P_{sclA}\) represents a new class of Mga-regulated promoters that requires a single distal binding site for activation.

A hallmark of the group A streptococcus (GAS; \textit{Streptococcus pyogenes}) is its ability to elicit a wide array of human diseases at varied locations throughout the human body. These syndromes may include relatively minor purulent infections that target either the upper respiratory tract or the skin as well as life-threatening invasive infections of deeper tissues, bloodstream, and organs (6, 30). GAS pathogenesis involves a number of different factors that allow the organism to survive within the human host. The initial stages of infection involve cell-associated factors that contribute to colonization, cell tropism, and immune evasion, including M proteins, capsule, and extracellular matrix binding proteins. Later stages often utilize secreted bacterial factors such as proteases, toxins, and other enzymes that allow the organism to spread and invade deeper tissues (6). In order to coordinate the expression of these different determinants with the changing environments that GAS encounters during an infection, the pathogen has developed multigene networks to regulate virulence.

One such regulator is the multiple virulence gene regulator of GAS known as Mga, which activates the transcription of genes encoding primarily surface-associated proteins involved in GAS pathogenesis. These factors include the M family of proteins (\(emm\), \(mrp\), \(arp\), and \(emm\)), C5a peptidase (\(scpA\)), fibrinectin-binding proteins (\(sop\) and \(fbx\)), and the secreted inhibitor of complement (\(sic\)) (4, 5, 12, 20). Mga is also required for its own positive regulation or autoactivation (19, 23). Environmental growth conditions such as increased \(CO_2\) and temperature have been shown to activate transcription of genes in the Mga regulon (3, 16, 26). Furthermore, expression of the Mga regulon is maximal during exponential phase and becomes transcriptionally silent as the bacteria enter stationary phase (18). This temporal regulation correlates with the requirement for Mga-regulated surface proteins during early stages of infection when GAS is actively growing but down regulates its expression during times of nutrient deprivation when the proteins may actually hinder the bacterium from spreading to new tissues. Thus, Mga controls its set of virulence factors in response to both environmental and temporal signals through an unknown mechanism.

Mga functions by binding directly to specific sequences found in the promoters of target genes, as demonstrated for \(emm\) (\(P_{emm}\)) and \(scpA\) (\(P_{scpA}\)) in the serotype M6 GAS strain JRS4 (15). An alignment of these two sites was used as the basis for a published consensus Mga-binding site (15). Binding of Mga in these two promoters occurs at similar locations consisting of a single site overlapping the −35 hexamer and just upstream of the start of transcription (15). Interestingly, autoregulation of \(mga\) transcription also involves binding of Mga within its own promoter (\(P_{mga}\)); however, Mga interacts with two separate binding sites located more than 100 bp upstream of the major P2 start site (19). Thus, based upon the variable location of binding sites in these three promoters, the mechanisms for Mga activation (\(P_{emm}\) and \(P_{scpA}\)) versus autoactivation (\(P_{mga}\)) may be quite different. Although all four of the identified Mga-binding sites are less than 50% identical at the DNA sequence level, recent studies have shown that the same two helix-turn-helix DNA-binding domains of Mga are responsible for recognizing each of these divergent sites (17). Furthermore, the ability of Mga to bind to these promoters...
This study

in vivo was shown to be essential for virulence gene activation (17). Exactly which nucleotides within a given binding site are required for Mga recognition and binding at these various promoters has not been established.

Recently, a gene encoding a cell-wall-anchored streptococcal collagen-like protein (sclA or sclI) was identified in all serotypes of GAS (13, 27). A mutant of the newly described surface protein was found to be defective for adherence to lung epithelial cells, but not pharyngeal cells, and was attenuated for virulence in a mouse model of soft-tissue pathology (13, 27). Expression of sclA was found to be defective for adherence during exponential growth as expected (13, 14, 27). Expression of sclA was found to be regulated in a serotype M1 GAS, and maximal expression occurred during exponential growth as expected (13, 14, 27). With the use of the published consensus Mga-binding site (15) to search the promoter region upstream of sclA, a potential Mga-binding site was found overlapping the −35 hexamer of a suspected start of transcription (13, 27) in a location similar to that observed for Pemn and PscpA. In this study, we identify a second distal binding site, and we investigate whether Mga plays a direct role in transcriptional activation at the sclA promoter.

**MATERIALS AND METHODS**

**Bacterial strains and media.** *S. pyogenes* (GAS) SF370 (31) represents an available serotype M1 GAS genome sequence (8). JRS4 is a streptococcal-resistan derivative of serotype M6 strain D471 (29). JRS519 (mga-10) is an Mga derivative of JRS4 (18). MGAS166 is a streptococcal-resistant M1 strain (22). *Escherichia coli* strain DH5a (pBR322, pSC101 gyrA endA1 relA1) was used as the host for plasmid constructions. The T7 polymerase *E. coli* strain BL21 (DE3) [F ompT hsdSB (rB - mB) gal dcm met] was used for protein expression and purification.

All *E. coli* strains were grown in Luria-Bertani broth. GAS was cultured in Todd-Hewitt medium supplemented with 0.2% yeast extract (THY), and growth was assayed by optical density with a Klett-Summmerson photoelectric colorimeter with the A filter. Antibiotics were used at the concentrations indicated: ampicillin at 100 μg/ml for *E. coli*, spectinomycin at 100 μg/ml for both *E. coli* and GAS, kanamycin at 50 μg/ml for *E. coli* and 300 μg/ml for GAS, and erythromycin at 500 μg/ml for *E. coli* and 1.0 μg/ml for GAS.

**DNA manipulations.** Plasmid DNA was isolated from *E. coli* by using either the Wizard Miniprep system (Promega) or Maxi/Midi Prep purification systems (Qiagen). DNA fragments were isolated from agarose gels by using the QIAquick gel extraction kit (Qiagen). GAS chromosomal DNA was isolated using the FastDNA kit and a FastPrep cell disruptor (Bio 101, Inc.). PCR for cloning and promoter probes was performed using Pfura high-fidelity DNA polymerase (Invitrogen), and reaction mixtures were purified using the QIAquick PCR purification system. PCR for diagnostic assays was performed using Vent DNA polymerase (New England Biolabs). DNA sequencing was performed either using the Excel II cycle sequencing kit (Epicentre, Inc.) or through the automated sequencing core facility in the McDermott Center at University of Texas Southwestern Medical Center.

**Construction of the Mga•M1 strains** KSMS165-L and MGAS166,165-L. To inactivate mga in SF370 and MGAS166, a suicide plasmid was constructed as follows: a 983-bp PCR fragment containing an internal fragment of mga was amplified from M1 SF370 genomic DNA by using primers OYR-4 and DivMgaL6 (Table 1). The resulting fragment was cloned blunt into an SmaI-digested suicide vector, pCIV2 (23), to produce pKSM165-L. Serotype M1 SF370 and MGAS166 were transformed with 200 to 300 μg of pKSM165-L, and chromosomal integrations were selected for by growth on THYA plates containing kanamycin at 50 μg/ml. Mutants were verified by PCR analysis of the mga locus and slot blot analysis of mga transcripts (data not shown).

**Northern blot analysis.** Total RNA was isolated from samples in late exponential phase (65 Klett units) by using the FastRNA kit and a FastPrep cell disruptor as previously described (17). Northern blot assays of total RNA were performed using the NorthernMax kit (Ambion). Briefly, either 1 (emm) or 5 (sclA) μg of total RNA was separated on a Smal-digested suicide vector, pCIV2 (23), to produce pKSM165-L. Serotype M1 SF370 and MGAS166 were transformed with 200 to 300 μg of pKSM165-L, and chromosomal integrations were selected for by growth on THYA plates containing kanamycin at 37°C. Mutants were verified by PCR analysis of the mga locus and slot blot analysis of mga transcripts (data not shown).

**TABLE 1. PCR primers used in this study**

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Sequence (5′→3′)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>emm</td>
<td>OM6-16</td>
<td>GTTTCCTCATTGTTGTC</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>OM6-35</td>
<td>AACAGCAAAATAGCTGTC</td>
<td>17</td>
</tr>
<tr>
<td>gusA</td>
<td>gusA-PE</td>
<td>GTTGGGCTTTCACAGGACG</td>
<td>This study</td>
</tr>
<tr>
<td>mga</td>
<td>DivMgaL6</td>
<td>CTTTAAAGAAACATTITTC</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>MgaPetI_Nde</td>
<td>ggcccatATGTATGAATGAATTGTTT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>MgaPetII_Xho</td>
<td>aacctcagATGGTGGAGGGGG</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>OYR-4</td>
<td>GATCCCATACCAATTTT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>rpsL-1</td>
<td>GAATGTAGATGCCTACAATTAACCA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>rpsL-2</td>
<td>GTGCGCCACGAAACGATATG</td>
<td>This study</td>
</tr>
<tr>
<td>nrII</td>
<td>nrII-PE</td>
<td>GGTATACTGAGGCTAAGTG</td>
<td>This study</td>
</tr>
<tr>
<td>Pemn6</td>
<td>Pemn-L1</td>
<td>GCATGATCCATCAGAAGACTA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>Pemn-R1</td>
<td>GGGGCTCAGATATGGTATCCTGTTATT</td>
<td>This study</td>
</tr>
<tr>
<td>Pemn1</td>
<td>Pemn-L1</td>
<td>AAGCGAATTATTGACAAGCTC</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>Pemn-R1</td>
<td>AGTCAAAGCTACCGCTACTG</td>
<td>This study</td>
</tr>
<tr>
<td>Pmga</td>
<td>OYR-25</td>
<td>GATTTGATCATACAAAGCTC</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>OYI-25</td>
<td>TACCATATATACCCTTC</td>
<td>This study</td>
</tr>
<tr>
<td>PscL7</td>
<td>PscL4-L</td>
<td>TTAAAGAAGAGATTGCAAGGAT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-R</td>
<td>ATGGTGCTTGTGATGCAACA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-L2</td>
<td>TGTTAATCTGTAATGCTGTC</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-R2</td>
<td>GGGGATCGAGGACTTTT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-L4</td>
<td>AGCAGTCAAAAAGTAACA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-R4</td>
<td>TGATCCCTTGGTCCTGTT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-L5</td>
<td>ACAGCTATGATCATTAGTG</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-R5</td>
<td>GATTATGTTTGGTATTGTGC</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PscL4-L6</td>
<td>GTGATCCCAAGCCCCCAA</td>
<td>This study</td>
</tr>
<tr>
<td>PscpA</td>
<td>CS-L1</td>
<td>AAGAGTATGATAGAGGCTACAA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>CS-R1</td>
<td>GGCAGATGGAAGCCATGTT</td>
<td>This study</td>
</tr>
<tr>
<td>sclA</td>
<td>scl4-F2</td>
<td>AGCTGATGCAAAGGAGCTAAAAATG</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>scl4-R1</td>
<td>GATTTCCTCATTGTTGTC</td>
<td>This study</td>
</tr>
</tbody>
</table>

* Underlined nucleotides are restriction sites, and lowercase nucleotides are anchor sequences introduced into the primer.
50°C for 15 min. Blots were visualized by exposure to a phosphorimaging cassette for 2 h. Probes were PCR amplified from a serotype M1 strain by using the primers listed in Table 1 for *emm* and *sclA*.

**Primer extension analysis.** Total RNA was extracted from samples in late exponential phase as described previously (17) by using the FastRNA kit and a FastPrep cell disruptor. Primer extensions were performed on 25 μg of total RNA as described previously (19) with the primers *P*~sclA*-L and mnl-PE (Table 1). The primer extension products were run on a 6% denaturing polyacrylamide gel (Amresco), and gels were processed as described above. Sequence was generated using labeled *P*~sclA*-R and mnl-PE primers (Table 1) on a PCR product amplified from SF370 genomic DNA (Table 1; *P*~sclA*-L2 and P*~sclA*-R) as described above.

**Expression and purification of Mga-His protein from *E. coli*.

A carboxy-terminal fusion of six histidines to the serotype M6 Mga was constructed as follows: a 1.4-kb PCR fragment containing the mna open reading frame was amplified from GAS JRS4 genomic DNA by using primers MgaPetI-Nde and MgaSalI-His. The primer extension products were run on a 6% denaturing polyacrylamide gel, and gels were processed as described above. Sequence was generated using labeled *P*~sclA*-R and mnl-PE primers (Table 1) on a PCR product amplified from SF370 genomic DNA (Table 1; *P*~sclA*-L2 and P*~sclA*-R) as described above.

**EMSAs.** Primer extension analysis. Anti-Mga-His immunoblot assays were performed as described previously (17). Briefly, purified Mga-His proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis as described below.

**Western blot analysis.** Anti-Mga-His immunoblot assays were performed as described previously (17). Briefly, purified Mga-His proteins were separated by SDS–10% PAGE. Proteins were transferred to nitrocellulose membranes and reacted either with the anti-Mga-pep2 antisera (17) at 1:250 dilution for 2 h at RT or with the anti-His monoclonal antiserum (Novagen) at a 1:2,000 dilution for 2 h at RT. Blots containing GAS whole-cell extracts were reacted with anti-rat immunoglobulin G horseradish peroxidase-conjugated secondary antibody (Sigma) or anti-mouse immunoglobulin G horseradish peroxidase-conjugated secondary antibody (Chemicon International), respectively; developed using the Renaissance chemiluminescence system (New England Nuclear); and visualized on autoradiography film (Kodak).

EMSAs. Promoter probes were generated by PCR amplification with either serotype M1 strain SF370 or serotype M6 strain JRS4 chromosomal DNA and the relevant primer pairs (*P*~sclA*-L, *P*~sclA*-R, *P*~sclA*-L5, and *P*~sclA*-R4) listed in Table 1. PCR fragments were end labeled with [γ-32P]ATP with the use of T4 polynucleotide kinase (New England Biolabs). Labeled fragments were excised from a 5% polyacrylamide gel, extracted by crush and soak elution, and purified using the QIAquick PCR purification system. EMSAs was performed as described previously (15). Briefly, a constant amount of labeled promoter probe DNA (ca. 1 to 5 ng) and increasing amounts of purified Mga-His (1 to 6 μg) were used in each reaction. Competition assays were performed by addition of 750 ng of unlabeled promoter probes to binding reaction mixtures. Gels were dried under vacuum at 80°C for 1 h and exposed overnight either to film or to a phosphorimaging screen. Phosphor screens were scanned using a Storm 820 scanner (Amersham Biosciences), and resulting data were analyzed with the ImageQuant analysis software (version 5.2).

**Construction and use of a green fluorescent protein (GFP)-based transcriptional reporter plasmid.** The plasmid pGreeTIR contains a promoterless gfp cassette modified for maximal solubility and fluorescence in bacteria (21). An 854-bp EcoRI fragment was amplified containing the promoterless gfp and its ribosomal binding site was cloned into the EcoRI-digested GAS shuttle vector pRS525 (18) to produce the plasmid pKSM123. A 854-bp fragment of pUCHKm24 (22), possessing tkm2 and T4 terminators, was gel purified and cloned at the 3′ end of gfp in the EcoRV-digested pKSM123 to form the transcriptional reporter plasmid pKSM410. Blunt PCR products corresponding to Peum1 (pKSM425), a 500-bp *P*~sclA*-L-R fragment (pKSM412), and a 235-bp *P*~sclA*-R fragment (pKSM414) were amplified from the serotype M1 GAS strain SF370 genomic DNA and cloned into XhoI-digested (blunted) pKSM410 (primers shown in Table 1). Plasmids were introduced into the serotype M1 strains GAS166 (Mga+) and GAS166.165-L (Mga−) by electroporation as described above.

GAS strains containing GFP reporter plasmids were grown to late logarithmic phase (65 Kleit units), pelleted, and resuspended in phosphate-buffered saline. Aliquots (200 μl) in duplicate were transferred to a 96-well black microtiter plate with a clear bottom (Costar) along with a phosphate-buffered saline control. Samples were assayed using a FLUOSTar Optima (BMG Lab Technology) fluorometer by using an excitation filter at 485 nm and an emission filter at 510 nm. Relative fluorescence units for each sample were expressed as fold above the level of the promoterless control strain containing pKSM410.

**Construction and use of a chromosomal GusA-based transcriptional reporter.** A 257-bp fragment of *gusA* containing *gusA*-binding site II was PCR amplified from SF370 by using the *P*~gusA*-L5-R4 primer pair (Table 1) and ligated into the EcoRV-digested pBluestar II KS− (Stratagene) to form pKSM417. A 619-bp fragment of *gusA* containing sequence upstream of *gusA*-binding site II was PCR amplified from SF370 by using the *P*~gusA*-L5-R4 primer pair (Table 1) and cloned into the SmaI-digested pKSM417 to form pKSM418. To form the plasmid pKSM442, an 890-bp fragment of pKSM418 was PCR amplified using the *P*~gusA*-L5-R5 primer pair (Table 1) and ligated into the SmaI-digested pKSM418 reporter plasmid pKSM440 (23). The 854-bp *gusA* reporter cassette was excised by XhoI digestion, gel purified, and cloned into the EcoRV site of the temperature-sensitive pRS233 (25) to form the pKSM-Δ site gusA plasmid pKSM443.

**RESULTS**

Identification of two potential Mga-binding sites upstream of *sclA*. To identify new Mga-regulated virulence genes in GAS, a homology search of the serotype M1 SF370 genome database (AE004092) was performed using the published Mga consensus binding sequence (15). As expected, binding sites were found within promoters of known Mga-regulated genes, including *emm*, *scpA*, and *sic* (data not shown). A putative Mga-binding site was also identified centered 51 bp upstream...
of the Mga-regulated sclA (Fig. 1A; PsclA-II) and corresponds to a site described previously (13, 27). The site shares 66% identity to the Mga consensus sequence at conserved sites and was located in PsclA overlapping the −35 hexamer of a predicted start of sclA transcription (Fig. 1A; PsclA-II), identical to the established Mga-binding sites for Pemm and PscpA (15). Interestingly, a second site showing 72% identity to the Mga consensus sequence (Fig. 1A; PsclA-I) was identified centered 153 bp upstream from PsclA. Scanning of the 3′ sequence adjacent to PsclA-I did not reveal an obvious second start of transcription, suggesting that Mga bound at this potential element may target the downstream transcriptional start site next to PsclA-II (Fig. 2B). Furthermore, sequences that are 100% identical to the conserved consensus nucleotides for both sites (Fig. 1A) are found upstream of sclA in strains representing a wide array of GAS serotypes, including M2, M3, M6, M12, M18, M28, M49, M52, M56, and M77 (data not shown). Thus, PsclA appears to possess two potential Mga-binding elements arranged in a pattern that is unique to this Mga-regulated promoter and is conserved across GAS serotypes. Previous studies found that expression of sclA is Mga regulated in the M1 strains AP1 (27) and JRS301 (14). In order to confirm that sclA is also Mga regulated in the M1 SF370 and MGAS166 strains used in this study, a Northern analysis was performed on wild-type and mga-inactivated versions of each strain on RNA isolated at the exponential phase of growth (Fig. 1B). As a control, Northern analysis was also performed on the established Mga-regulated gene emm (Fig. 1B). As expected, transcript levels for both emm and sclA are dramatically lower in the Mga−/H11002 versions of SF370 and MGAS166 than in their wild-type counterparts, confirming Mga regulation of sclA in these M1 strains.

Identification of the Mga-regulated start of transcription for sclA. Since the predicted location of the Mga-regulated start site for sclA expression has not been experimentally verified, primer extension analysis was performed on total RNA isolated from both wild-type (SF370) and mga−-inactivated (KSM165-L) serotype M1 GAS strains by using an antisense primer, PsclA-R (Table 1), located in the 5′ end of sclA. A primer extension product was clearly observed in a background producing Mga (Fig. 2A) that corresponded exactly to the start of transcription predicted in the literature (13, 27). However, the product was clearly absent in the mga−-inactivated KSM165-L (Fig. 2A). As a control for the primer extension analysis, the transcriptional start for the divergently transcribed nrdI gene found directly upstream of sclA was mapped in both strains.
tested (Fig. 2A). Interestingly, although the \textit{nrdI} start site is located within the putative distal P\textsubscript{sclA}-I Mga-binding element, there was no alteration in \textit{nrdI} transcription either in the presence or in the absence of Mga. These data indicate that \textit{sclA} contains a single Mga-regulated start of transcription that is located directly downstream of the potential P\textsubscript{sclA}-II Mga-binding site (Fig. 2B).

\textbf{Mga-His purified from \textit{E. coli}} binds to the distal P\textsubscript{sclA}-I site but not the promoter-proximal P\textsubscript{sclA}-II site.\textbf{.} To determine whether Mga directly interacts with the two potential binding elements within P\textsubscript{sclA}, EMSAs were performed on overlapping PCR-generated promoter probes encompassing the P\textsubscript{sclA} region (Fig. 3A). Previous EMSA experiments have successfully utilized an MBP-Mga protein generated from a serotype M6 GAS strain (15, 17, 19); however, this chimera is much larger in size than Mga alone (105 versus 62 kDa) and exhibits only 60% of wild-type Mga activity when expressed in vivo (18). In order to bypass these limitations, a carboxy-terminal fusion of six histidines to the full-length Mga from the M6 GAS strain JRS4 (Mga-His; 63 kDa) was constructed and purified.
from E. coli lysates as described in Materials and Methods. Purified Mga-His showed DNA-binding activity to known binding sites in vitro comparable to that of MBP-Mga and was capable of wild-type transcriptional activation of Mga-regulated genes in vivo (data not shown).

All three promoter probes that contained the distal PsclA-I binding site demonstrated a slower migration consistent with binding when incubated in increasing amounts of the purified Mga-His protein (Fig. 3B; L2-R, L2-R2, and L4-R). In contrast, no detectable binding of Mga-His was observed to a probe that lacked the distal site I but possessed the proximal PsclA-II binding element and the sclA start of transcription (see above) at any of the concentrations tested (Fig. 3B; L-R). Furthermore, no additional binding sites were found in the region located upstream of PsclA-I (Fig. 3B; L2-R4). Thus, Mga appears to show in vitro DNA binding only to the distal PsclA-I binding site situated more than 100 bp upstream of the Mga-regulated start of transcription.

Specific binding of Mga-His to PsclA-I in vitro. The specificity of Mga-His binding to the PsclA-I element was investi-
Mga may bind to the proximal P_{sclA}-II site but with markedly lower affinity than that observed for the distal P_{sclA}-I site.

Mga-dependent regulation of P_{sclA} is not observed if located on a multicopy plasmid. To investigate the role of Mga binding to the upstream P_{sclA}-I site in regulation of P_{sclA} expression, a transcriptional reporter system based on GFP was developed. Use of GFP as a reporter allows direct quantification of promoter activity via levels of fluorescence without the need to lyse the bacteria or add reagents. Unfortunately, attempts to generate single-copy promoter fusions of the strong P_{emm} promoter to gfp in the chromosome of GAS did not allow detection over background levels (data not shown). Therefore, a multicopy GFP transcriptional reporter plasmid called pKSM410 was constructed that is capable of replicating in GAS as well as E. coli (see Materials and Methods).

Promoter fragments corresponding to P_{emm} (pKSM425), the entire P_{sclA}-ndrl intergenic region (pKSM412), and P_{sclA} lacking the distal P_{sclA}-I Mga-binding site (pKSM414) were cloned in front of the promoterless gfp in pKSM410 (Fig. 5A) and introduced into both wild-type MGAS166 and mga-inactivated MGAS166.165-L serotype M1 GAS. Strains were grown to late logarithmic phase, and the levels of fluorescence were quantified using a fluorometer as described in Materials and Methods. As expected for an Mga-regulated promoter, the control P_{emm} construct pKSM425 showed extremely high expression in MGAS166 but was reduced 7.5-fold in MGAS166.165-L (Fig. 5B; pKSM425). However, the full-length P_{sclA}-L2-R construct pKSM412, which has both the P_{sclA}-I binding site and the Mga-regulated start of transcription, demonstrated very little promoter activity over background and was not affected by the presence or absence of Mga (Fig. 5B; pKSM412). Likewise, the transcriptional start site without the distal P_{sclA}-I binding site showed little change in activity in the presence or absence of Mga (Fig. 5B; pKSM414). Therefore, the presence of the upstream Mga-binding site P_{sclA}-I in the P_{sclA} promoter was not sufficient to show Mga-regulated gfp expression in vivo and suggests that additional cis sequences not included in this construct are required.

**P_{sclA}-I Mga-binding site is essential for Mga regulation of P_{sclA} at its native chromosomal location.** Due to the lack of Mga-dependent regulation of P_{sclA} in a plasmid system, analysis of the promoter at its native chromosomal location was undertaken. Initial attempts to introduce a deletion of the P_{sclA}-I binding site into the MGAS166 chromosome via allelic exchange were unsuccessful, possibly due to the concomitant deletion of the divergently transcribed ndrl promoter (Fig. 2B; data not shown). Therefore, a merodiploid approach was used to allow the temperature-sensitive integration of single-copy P_{sclA}-gusA reporter fusions into the native chromosomal location while retaining a second wild-type promoter expressing both ndrl and P_{sclA} (Fig. 6A). Merodiploid insertion-duplication mutants containing either wild-type P_{sclA} (KSM421) or the P_{sclA}-I deletion allele (KSM423) fused to a promoterless gusA were constructed in MGAS166 (Mga+) as well as MGAS166.165-L (Mga-), and GusA liquid assays were performed on exponential-phase cultures as described in Materials and Methods. The wild-type KSM421 exhibited strong Mga-regulated GusA activity, whereas the P_{sclA}-I deletion strain KSM423 showed GusA activity at background levels regardless of the presence or absence of Mga (Fig. 6B). These
results support in vivo the results of the in vitro EMSAs and indicate an essential role for the distal PsclA-I binding site in the Mga-dependent expression of sclA.

**DISCUSSION**

**What constitutes an Mga-binding site?** Since every Mga-regulated gene described to date encodes a protein involved in GAS pathogenesis, the ability to identify potential Mga-regulated genes through an Mga-binding site in their promoter would be highly advantageous. The published consensus Mga-binding element was constructed using known binding sites (Pemm and Psca from serotype M6) as well as predicted sites from sequenced promoters corresponding to emm, mpr, arp, enn, and scpA from various other GAS serotypes (15). The resulting 45-bp consensus element contains three regions of conserved nucleotides totaling 29 nucleotides separated by several stretches of irrelevant nucleotides (Fig. 1A). However, this consensus was not able to predict the location of the two Mga-binding sites found in PmgA (19), indicating that Mga can interact with diverse sequences and that any true consensus must reflect these various sites. Since Mga appears to use the same two helix-turn-helix domains of the protein for binding to all of the known promoter sites (17), there likely exists a core consensus binding element that is shared by all sites.

Although the two sites identified within PsclA shared significant identity (72 and 66%, respectively) to the nucleotides in the published consensus, we have shown here that Mga was able to bind only the distal PsclA-I site in vitro (Fig. 2 to 4). In fact, PsclA-I contains only four nucleotides identical to conserved consensus sequences that are lacking in PsclA-II (Fig. 1A, asterisks). Thus, it is likely that one or more of these nucleotides are critical for the in vitro interaction of Mga and provide excellent targets for mutational analysis of Mga-binding sites. Attempts to introduce both Pmga and the PsclA-I binding site into the Mga-binding consensus resulted in a very nonspecific sequence consisting of two short stretches of primarily adenine nucleotides (data not shown). Therefore, future experiments will focus on changing the nonbinding PsclA-II sequence to resemble PsclA-I to allow for a gain-of-function phenotype and identification of those nucleotides that are essential for binding. These types of directed biochemical and mutational analyses of Mga-binding sites should provide a
more reliable way than sequence alignment to establish the essential core Mga-binding sequence.

Interestingly, when Mga is purified from GAS compared to being purified from *E. coli*, binding to PsclA-I is noticeably enhanced (Fig. 4). In addition, Mga purified from *E. coli* is still sufficient to activate transcription of Mga-regulated promoters in an in vitro transcription assay (A. C. Almengor and K. S. McIver, unpublished results), but not as effectively as Mga purified from GAS. This difference may indicate several possibilities: that an additional factor present only in GAS copurifies with Mga and augments its activity, that a specific modification of Mga occurs in GAS and not *E. coli*, or simply that the recombinant expression of Mga in a heterologous host is not optimal for its activity (e.g., improper folding). Regardless of the mechanisms involved, these results support the importance of using Mga purified from GAS in future experiments dealing with its activity.

**PsclA represents a new category of Mga-regulated promoters.** Mga is a DNA-binding protein that functions through direct interaction at binding sites within its target promoters. Based upon the biochemical analysis of three Mga-regulated promoters from the serotype M6 strain JRS4 (15, 19), the
location of Mga binding at these promoters falls into two types. In the case of Pemm and PscepA, Mga binds to a single site centered at −51 bp from the start of transcription and juxtaposed with the −35 hexamer. This places Mga in a position similar to class I transcriptional activators that stabilize RNA polymerase through direct contact with the carboxy-terminal domain of the α subunit and increase initiation (11, 15). In contrast, Pmgα contains two Mga-binding sites situated further upstream (−100 bp and −181 bp) from the P2 start of transcription, and both of these sites are required for the autoregulation of the downstream start site (19). Currently, it is not clear by what mechanism Mga is able to activate transcription initiation at such a distance.

In this study, we have shown that Mga binds specifically within the promoter of the Mga-regulated sclA, which encodes a streptococcal collagen-like protein involved in GAS virulence. Although the sclA promoter (PscL-α) contains two potential binding sites based upon homology to the published consensus sequence, Mga interacts only with a single binding site (PscL-α-I) located upstream (−168 bp) of the Mga-regulated start of transcription (Fig. 2 to 4). Additionally, the distal PscL-α-I was found to be necessary for activation of sclA transcription in vivo (Fig. 6). The location of PscL-α-I is somewhat similar to the most distal Pmgα-Mga-binding site (−181 bp) and may suggest some commonality of activation between these two promoters. Activators that bind at a distance will often use DNA looping or bending to bring the regulator into physical contact with the transcriptional machinery (2, 10). On the other hand, the fact that PscL-α uses only a single binding site suggests that PscL-α represents a unique type of Mga-regulated promoter. Unpublished work from our laboratory indicates that an Mga-regulated promoter (Pssof) from a class II strain of GAS also possesses a single distal Mga-binding site of this type.

Based on the differences observed in the location of Mga-binding sites with respect to the start of transcription in Mga-regulated promoters, we propose that the promoters be grouped into three categories: a single proximal binding site (category A, Pemm and PscepA), a single distal binding site (category B, PscLα and Pssof), and two distal binding sites (category C, Pmgα). At some time in the past, the second putative PscLα binding site (PscLα-II) may have been fully functional but is no longer utilized to regulate sclA expression. Alternatively, PscLα-II may be functional for binding only under very specific temporal or environmental conditions in vivo. Evidence of the possible functionality of PscLα-II comes from our observation of barely detectable Mga binding to this site upon significant overexpression of the EMSA, several orders of magnitude below that seen at PscLα-I (data not shown). Therefore, PscLα may have once represented an Mga-regulated promoter containing both a proximal (category A) and a distal (category B) binding site. However, PscLα has evolved such that it no longer requires the proximal site for Mga-specific activation.

Mga regulation of PscLα occurs at its native locus but is not observed on a multicopy plasmid. Expression of sclA has been shown previously to be strongly activated at the level of transcription in wild-type M1 GAS strains but not in isogenic strains lacking Mga (Fig. 1B) (14, 27). Our primer extension data confirm these results, showing a strong start of transcription in total RNA isolated from wild-type M1 strain SF370 but not in the mga-inactivated KSM165-L (Fig. 2A). Thus, Mga appears to control sclA by activating transcriptional initiation at PscLα. The ability of Mga to bind to the promoter-distal PscLα-I suggests strongly that interaction at this site is directly involved in the observed activation of the downstream transcriptional start site. Further, the presence of PscLα-I at its native position in the GAS chromosome is essential to the activation of sclA transcription in vivo with a GusA reporter system (Fig. 6). However, when a large (500-bp) region of PscLα, including the PscLα-I binding site and the downstream start site, was inserted in a transcriptional reporter plasmid based on a promoterless gfp, it did not exhibit detectable Mga-specific activation, nor did it demonstrate significant promoter activity over background levels of the system (Fig. 5). In contrast, a control plasmid containing the category A Pemm-GFP produced very strong Mga-specific activation in the GFP reporter system. There are several possible explanations for why Mga regulation of sclA occurs only in a single copy at its native locus and not on a multicopy plasmid. First, one or more cis-acting elements located either further upstream or within sclA itself may be required for Mga regulation. Alternatively, there may be a gene dosage effect observed in the plasmid system, resulting in the titration of a factor necessary for Mga regulation to occur. Finally, DNA looping may be required to bring Mga in contact with the RNA polymerase, and this event may not be occurring properly in our plasmid system.

Upstream of sclA in the SF370 M1 genome is the divergently transcribed nrdI gene, which is predicted to encode a putative ribonucleotide reductase (Fig. 2B). As an internal control for primer extension of PscLα, we mapped the start of transcription for nrdI and found that it was located in the center of the PscLα-I Mga-binding site on the antisense strand. Although this might be expected to result in the Mga-specific repression of nrdI transcription, a primer extension product was clearly observed in SF370, an Mga’ strain (Fig. 2A). In fact, the presence of the nrdI promoter in the middle of the distal Mga-binding site may have hindered our attempts to delete PscLα-I at its native site in the chromosome (data not shown), suggesting the importance of nrdI in growth under the conditions used in this study. This suggests an even more complex interplay among Mga, PscLα, and the overlapping PnrdI regions that must be further explored.

In order to assess the importance of the distal Mga-binding site of PscLα while avoiding the deletion of PnrdI, single-copy PscLα-gusA reporter fusions were integrated into the native chromosomal location such that a second wild-type promoter expressing both nrdI and sclA was preserved in each case. These integrations clearly showed that Mga-regulated activity requires the distal Mga-binding site (Fig. 6). As stated above, several reasons for Mga regulation of sclA only at its native locus can be proposed, including DNA topology and gene dosage from a multicopy plasmid. These possibilities might explain why a category A Mga-regulated promoter (Pemm), which presumably does not require DNA looping, shows significant Mga-specific activity in the plasmid-based GFP reporter whereas a category B PscLα does not function in the same way. However, it is clear that the ability of Mga to regulate PscLα from a distal binding site is more complex than activation from a proximal site and may require additional factors.

In conclusion, we have demonstrated that Mga binds to a
single distal binding site within the *sclA* promoter but not to a second putative site located proximal to the Mga-regulated start of transcription. Further, *PslA*-I is required for in vivo transcriptional activation of *sclA*. Notably, an alignment of sequences for *PslA*-I found in multiple serotypes of GAS showed no differences in key conserved nucleotides. Therefore, *PslA* and the role of Mga in its regulation appear to be conserved among most serotypes of GAS. Analysis of this novel Mga-regulated promoter should help build our overall knowledge of the molecular mechanisms by which Mga controls virulence in GAS.

ACKNOWLEDGMENTS

We thank Deborah Ribardo, Cheryl Vahlng, and Temekka Leday for critical reading of the manuscript; Steve Lindow for providing the pGreenTIR plasmid; Sławomir Łukomski for unpublished *PslA* promoter sequences; and Rhonda Myles for technical assistance. This work was supported by grants from the American Heart Association, Texas Affiliate (beginning-grant-in-aid no. 0060073Y to K.S.M.) and the National Institutes of Health (NIH AI-47928 to K.S.M.). A.C.A. is supported in part by an NIH/NIAID Molecular Microbiology training grant (ST32 AI07520) and an NIH/NIAID research supplement for underrepresented minorities (RSUM AI-47928-S).

REFERENCES


