JOURNAL OF BACTERIOLOGY, Mar. 2002, p. 1335–1339 Vol. 184, No. 5
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.5.1335–1339.2002

Trinucleotide GAA Repeats Dictate pMGA Gene Expression in *Mycoplasma gallisepticum* by Affecting Spacing between Flanking Regions

Li Liu,[1] Victor S. Panangala,[2] and Kevin Dybvig[1]*

Department of Genomics and Pathobiology, University of Alabama at Birmingham, Birmingham, Alabama 35294,[1] and Department of Pathobiology, College of Veterinary Medicine, Auburn University, Auburn, Alabama 36849[2]

Received 18 July 2001/Accepted 2 November 2001

The pMGA genes of the avian respiratory pathogen *Mycoplasma gallisepticum* encode a family of hemagglutinins that are subject to phase variation. A trinucleotide GAA repeat region is located upstream of the pMGA transcription start site. The length of the repeat region varies at a high frequency due to changes in the number of repeat units. Previous studies have shown that pMGA genes are transcribed when 12 GAA repeats are present but are not transcribed when the number of repeats is not 12. To further analyze the mechanism of gene regulation, the pMGA promoter region was modified either by deleting the nucleotides 5′ of the GAA repeats or by inserting linkers of 10 or 12 bp at a position 3′ of the repeats. The modified promoter region was fused to a promoterless *lacZ* gene and transformed into *M. gallisepticum* by using transposon Tn*4001* as a vector. Transformants and successive generations of progeny were analyzed with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) to monitor β-galactosidase activity. For the transformants of *M. gallisepticum* containing the reporter with deletion of nucleotides 5′ of the GAA repeats, GAA-dependent pMGA gene regulation was abolished. For the transformants containing the reporter with an addition of 10- or 12-bp linkers, *lacZ* was expressed only when eight GAA repeats were present. These data indicate that the nucleotides 5′ of the GAA repeats as well as the spacing between the GAA repeats and sequences downstream (3′) of the repeats are important for pMGA gene expression.

Many pathogenic mycoplasmas undergo high-frequency changes in surface protein production, presumably to enable colonization and persistence in a variety of host tissues (2, 10, 11). The pMGA family of hemagglutinin (adhesin) proteins of *Mycoplasma gallisepticum* are likely to play an important role in colonization and chronicity of respiratory disease in the avian host (4). Previous studies have shown that the number of pMGA genes among strains of *M. gallisepticum* ranges from about 32 to 70 (1), with only one gene being predominantly expressed in any given strain (5, 8, 9). The growth of *M. gallisepticum* in the presence of antibody specific for the predominant pMGA protein that is produced in strain S6 (pMGA1.1) led to a reversible shift in the cell population such that cells that produced the alternative protein pMGA1.9 became dominant (9). From experimental infection of chickens, it has been inferred that cells switch from producing pMGA1.1 to producing alternative pMGA proteins in vivo (4). The switching in pMGA protein production occurred early in infection, suggesting that factors in addition to antibodies can provide selective pressure affecting the cell population.

Each pMGA gene has a tandem GAA repeat region, the 3′ end of which is located about 53 nucleotides upstream of the transcription start site. Switching of pMGA gene expression is associated with changes in the number of trinucleotide repeats. pMGA genes that are expressed have 12 GAA repeats. Genes that have 5, 9 to 11, 14 to 18, 20, and 24 GAA repeats have been documented as being transcriptionally silent (3, 4, 7, 9). Thus, 12 GAA repeats seem to be required for gene expression. The length (number of repeats) of the GAA repeat region increases and decreases at a high frequency, presumably as a result of slipped-strand mispairing (6), accounting for pMGA phenotypic switching.

We described previously the construction and use of a pMGA-*lacZ* reporter gene to study pMGA gene expression based on blue/white color selection of *M. gallisepticum* colonies on agar supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-Gal). LacZ activity was monitored over several successive generations, and the promoter region of the reporter genes of individual progeny clones was sequenced and analyzed. The only difference between the promoter regions of $LacZ^+$ and $LacZ^-$ colonies was in the number of GAA repeats. Colonies that expressed pMGA-*lacZ* had exactly 12 copies of the GAA repeat, while those that lacked *lacZ* expression had more or fewer than 12 repeats (7).

In the present study, the promoter region of the pMGA-*lacZ* reporter was modified to further investigate pMGA gene expression. When nucleotides upstream of the GAA repeats were deleted, GAA-dependent pMGA gene expression was abolished. When nucleotides (10- or 12-bp linkers) were inserted immediately downstream of the GAA repeats, gene expression was retained only if there was a reduction in the number of GAA repeats to eight. Taken together, these findings indicate that sequences both upstream and downstream of the GAA repeats are important for pMGA gene regulation and that the number of repeats affects expression by altering the spacing between the flanking sequences.

* Corresponding author. Mailing address: Department of Genomics and Pathobiology, Volker Hall, Room 223, University of Alabama at Birmingham, Birmingham, AL 35294-0019. Phone: (205) 934-9327. Fax: (205) 975-4418. E-mail: dybvig@uab.edu.

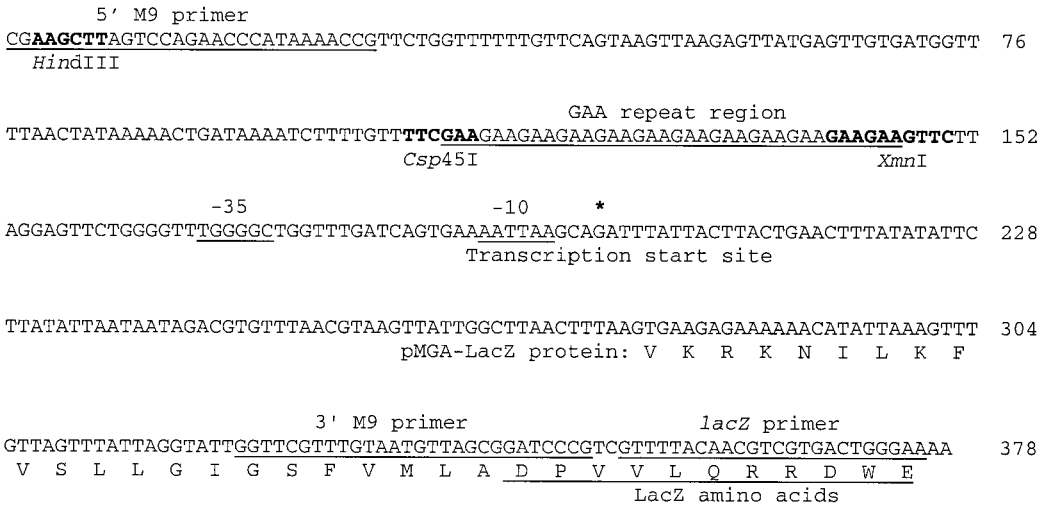

FIG. 1. Nucleotide sequence of the 5′ end region of the unmodified pMGA-*lacZ* fusion gene. Sequences corresponding to the oligonucleotide primers used for PCR amplification and DNA sequencing are underlined. Also underlined are the GAA repeats, the amino acids encoded by *lacZ*, and the −10 and −35 regions of the promoter. The putative transcription start site (asterisk) was identified based on sequence similarity to the previously determined start site of pMGA1.1 (5). Restriction enzyme recognition sites are in bold.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *M. gallisepticum* PG31 (ATCC 19610) was propagated in modified Frey broth or agar as described elsewhere (7). Plasmid constructions were performed in *Escherichia coli* strain INVαF' (Invitrogen, Carlsbad, Calif.) grown in Luria-Bertani medium.

Modifications in the pMGA-*lacZ* reporter. *E. coli* plasmid pCR2.1-M9.lacZ containing the pMGA-*lacZ* reporter inserted in the cloning vector pCR2.1 has been described previously (7). The reporter consists of a 336-bp pMGA gene fragment extending from 104 bp 5′ of the GAA repeats into the region coding for the first 22 amino acids of the pMGA protein that is fused in frame with *lacZ* (Fig. 1). *Hin*dIII and *Csp*45I cleave plasmid pCR2.1-M9.lacZ at single sites within the pMGA-*lacZ* reporter (Fig. 1 and 2). To excise the nucleotides upstream of the GAA repeats, pCR2.1-M9.lacZ was digested with *Hin*dIII and *Csp*45I. The ends of the reporter-containing *Hin*dIII-*Csp*45I DNA fragment were made flush by incubation with T4 DNA polymerase, and *Sph*I linkers were attached. The fragment was then inserted into the transposon Tn*4001* portion of plasmid pISM2062.S, which is a modified version of pISM2062 in which the *Sma*I cloning site was replaced with *Sph*I (7). Plasmid constructs were screened by restriction mapping to identify those that had the modified reporter oriented in Tn*4001* as shown in Fig. 2. The resulting plasmid is referred to as pISM-M9.lacZ.del.

*Sal*I linkers were inserted into pCR2.1-M9.lacZ at the *Xmn*I site located immediately downstream of the GAA repeats. The plasmid has a second *Xmn*I site in the ampicillin resistance determinant. Therefore, a partial digestion of pCR2.1-M9.lacZ with *Xmn*I was performed. After *Xmn*I digestion, the samples were subjected to agarose gel electrophoresis, and full-length, linear pCR2.1-M9.lacZ molecules were isolated from the gel. *Sal*I linkers of either 10 or 12 bp were ligated to the linear plasmid. After digestion with *Sal*I to generate cohesive ends, the plasmid molecules were circularized by ligation and transformed into *E. coli*. The restriction map of the plasmid isolated from the transformants was determined to identify plasmids in which the *Xmn*I site at the GAA repeats had been converted to *Sal*I. The *Sal*I-modified reporter was excised from the pCR2.1 vector by digestion with *Hin*dIII, the ends of the liberated DNA fragment were made flush by digestion with T4 DNA polymerase, and *Sph*I linkers were attached. As described above, the reporter gene was inserted into the Tn*4001* portion of pISM2062.S, screening for plasmid constructs that had the reporter oriented as depicted in Fig. 2. The resulting plasmids are referred to as pISM-M9.lacZ.ins10 (10-bp linker) and pISM-M9.lacZ.ins12 (12-bp linker).

Transformation of *M. gallisepticum* and pedigree analysis by blue/white color selection. The plasmids used in this study do not replicate in *M. gallisepticum*. Plasmids containing the wild-type gene or one of the modified pMGA-*lacZ* genes in pISM2062.S were transformed into *M. gallisepticum* by electroporation as described previously, resulting in transformants in which Tn*4001*, containing the pMGA-*lacZ* reporter, had transposed into the mycoplasmal chromosome (7). Transformants were selected in medium supplemented with gentamicin (50 μg/ml) and X-Gal (160 μg/ml) to monitor pMGA-*lacZ* expression on the basis of blue/white color selection. Incubation of *M. gallisepticum* in 10% CO_2 improved the quality of the blue/white color. Pedigree analysis was performed by filter cloning procedures as described previously (7).

DNA analysis of subclones. The promoter region of the reporter gene from representative blue and white *M. gallisepticum* colonies was amplified by PCR and sequenced. For the wild-type reporter and for the modified reporters containing *Sal*I linkers, the primers used for PCR amplification (5′ M9 and the *lacZ* primers) and DNA sequencing (3′ M9 primer) were as depicted in Fig. 1 and as described previously for amplification and sequence analysis of the wild-type reporter (7). For analysis of the promoter region of the modified reporter in pISM-M9.lacZ.del, a new primer was required for PCR amplification because the binding site for the 5′ M9 primer had been deleted. The replacement primer 5′-TAGCCATTTCGTTGACTTTGTCGG-3′ binds to Tn*4001* sequences located 282 bp 5′ of the GAA repeats. The PCR products were directly sequenced at the Iowa State University DNA Sequencing and Synthesis Facility, Ames. Sequence analysis was performed with the Sequencher (version 3.0; Gene Codes Corporation, Ann Arbor, Mich.) and MacVector (version 6.5.3; Oxford Molecular Group Inc., Beaverton, Oreg.) software packages. The *lacZ* probe and Southern analysis procedures for monitoring transposition of Tn*4001* in *M. gallisepticum* were as described previously (7).

RESULTS AND DISCUSSION

Nucleotides upstream of the GAA repeats are required for pMGA-*lacZ* gene expression. *M. gallisepticum* was transformed with pISM-M9.lacZ.del and assayed for LacZ activity. All of the initial transformants were $LacZ^-$. Five transformants were selected for further study. Essentially 100% of the progeny from three of the transformants were $LacZ^-$, but about 5% of the progeny from the other two transformants were $LacZ^+$. The nucleotide sequence of the promoter region of the pMGA *lacZ* reporter from both $LacZ^-$ and $LacZ^+$ clones was determined. No nucleotide differences were detected, and 12 tandem GAA repeats were present. Because the LacZ phenotype was independent of the sequence of the promoter region, we examined the pMGA-*lacZ* reporter by Southern analysis to determine whether a change in phenotype might correlate with transposition of Tn*4001* to an alternative chromosomal site.

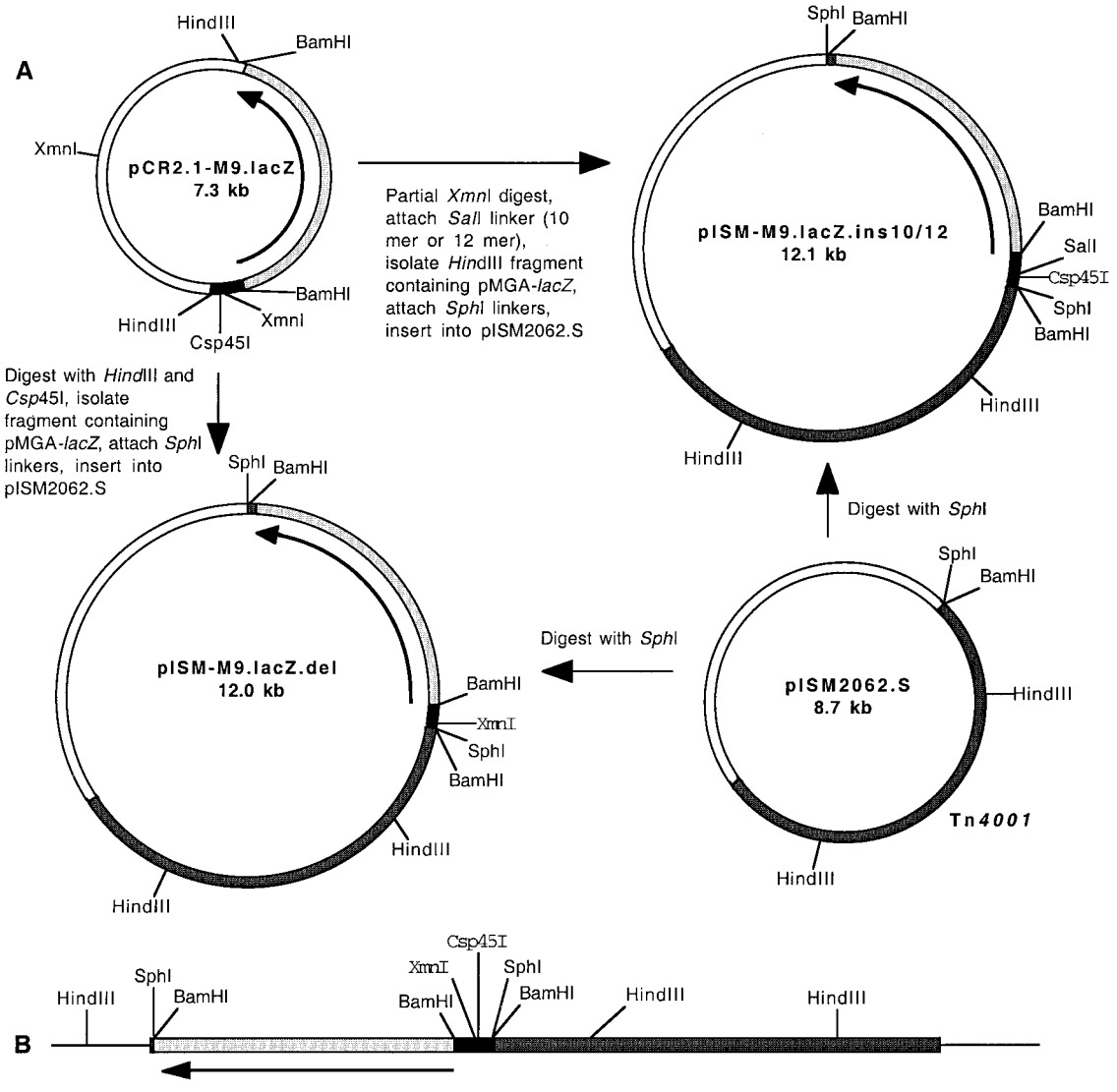

FIG. 2. Schematic diagram of the construction of the modified pMGA-*lacZ* reporters and their insertion into the chromosome of *M. gallisepticum*. (A) Deletion of nucleotides to generate pISM-M9.lacZ.del and insertion of *Sal*I linkers to generate pISM-M9.lacZ.ins10/12. White regions denote sequences derived from *E. coli* plasmid vectors. Regions in black, dark gray, and light gray denote the pMGA gene fragment, sequences originating from Tn*4001*, and *lacZ*, respectively. Arrows in plasmids show the direction of transcription. (B) A representation of Tn*4001* containing pMGA-*lacZ* inserted into the chromosome of *M. gallisepticum.* Thin lines denote mycoplasmal chromosomal DNA flanking the transposon. Shaded regions are as in panel A. The leftmost *Hin*dIII site is the mycoplasmal *Hin*dIII site most proximal to the left end of the transposon. The *Xmn*I site in the pMGA promoter region is replaced with *Sal*I in the transposon of pISM-M9.lacZ.ins10 and pISM-M9.lacZ.ins12. The *Csp*45I site and upstream pMGA sequences were deleted in the transposon of pISM-M9.lacZ.del.

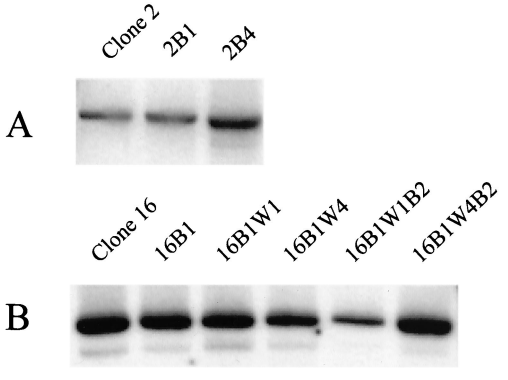

FIG. 3. Southern analysis of mycoplasmal genomic DNA digested with *Hin*dIII and probed with *lacZ*. Controls showing that the *lacZ* probe does not hybridize with DNA from *M. gallisepticum* strain PG31 that lacks the pMGA-*lacZ* reporter were previously described (7). (A) Clone 2 is a $LacZ^-$ transformant of *M. gallisepticum* transformed with pISM-M9.lacZ.del. Clones 2B1 and 2B4 are $LacZ^+$ progeny of clone 2. (B) Clone 16 is a $LacZ^-$ transformant of *M. gallisepticum* transformed with pISM-M9.lacZ.ins.10. The lineage and phenotype of $LacZ^+$ and $LacZ^-$ derivatives of clone 16 (clones 16B1, 16B1W1, 16B1W4, 16B1W1B2, and 16B1W4B2) are provided in Fig. 4. As determined by comparison to the mobility of radiolabeled *Hin*dIII fragments of bacteriophage lambda DNA, the hybridizing DNA fragments from clones 2 and 16 are about 15 and 5 kb, respectively. Similar experiments (data not shown) indicate that transposition of Tn*4001* also did not occur during switching of LacZ production in *M. gallisepticum* transformed with pISM-M9.lacZ.ins.12.

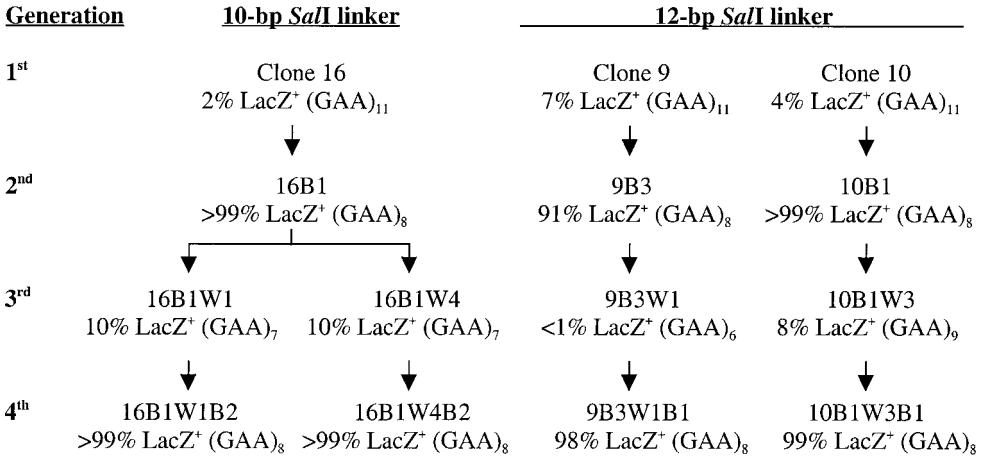

FIG. 4. Pedigree analysis of *M. gallisepticum* containing the pMGA-*lacZ* reporter modified by insertion of 10- and 12-bp linkers. For each of the subclones, the percentage of progeny that were scored as blue, lacZ$^+$ colonies is indicated.

The *lacZ* probe hybridized with a *Hin*dIII fragment of the same size from both LacZ$^-$ and LacZ$^+$ clones (Fig. 3A). The available data do not explain why some clones are LacZ$^+$, but it is apparent that deletion of pMGA sequences upstream of the GAA repeats abolished GAA-dependent gene regulation.

Effect of adding *Sal*I linkers 3′ of the GAA repeats. Transformants of *M. gallisepticum* that contain the pMGA -*lacZ* reporter modified by the addition of 10- or 12-bp *Sal*I linkers were monitored on X-Gal for β-galactosidase production. The initial transformants were LacZ$^-$, but second-generation LacZ$^+$ progeny were isolated at a high frequency. From the LacZ$^+$ progeny, third-generation LacZ$^-$ progeny were isolated, which in turn were subcloned to obtain LacZ$^+$ fourth-generation progeny. The percentage of blue (LacZ$^+$) progeny colonies for each of these successive generations of subclones is summarized in Fig. 4, revealing the phase-variable production of LacZ as described previously for the wild-type pMGA-*lacZ* reporter (7). Southern analysis of mycoplasmal DNA probed with *lacZ* revealed that switching of LacZ production did not correlate with transposition of Tn*4001* to alternative sites in the chromosome (Fig. 3B).

The promoter region of the *Sal*I-modified reporter from representative LacZ$^+$ and LacZ$^-$ subclones was PCR amplified and sequenced to determine whether switching in LacZ production correlated with changes in the number of GAA repeats. The insertion of the linkers into the *Xmn*I site disrupted the terminal GAA repeat, resulting in 11 tandem repeats followed by a GA dinucleotide, the linker, and the final A nucleotide of the 12th trinucleotide repeat (Fig. 5). Thus, the initial LacZ$^-$ transformants had 11 tandem GAA repeats. The numbers of tandem GAA repeats in the LacZ$^+$ and LacZ$^-$ progeny that were analyzed in successive generations are given in Fig. 4. The second-generation LacZ$^+$ progeny all had eight tandem GAA repeats, regardless of whether the linker was 10 or 12 bp. No third-generation LacZ$^-$ subclones had eight GAA repeats, but all fourth-generation LacZ$^+$ subclones once again had eight repeats. Thus, in the *Sal*I-modified reporters, eight tandem GAA repeats were associated with gene expression.

For LacZ$^+$ clones containing the 10-bp *Sal*I linker, eight GAA repeats restores the pMGA promoter region to match as closely as possible the length of the promoter region found in LacZ$^+$ clones containing the unmodified promoter (12 GAA repeats). The 10 bp that was added by addition of the linker was compensated by the deletion of three GAA repeats (9 bp), resulting in a net gain of a single base pair (Fig. 5). This result strongly suggests that the length of the GAA repeat region is critical for gene expression.

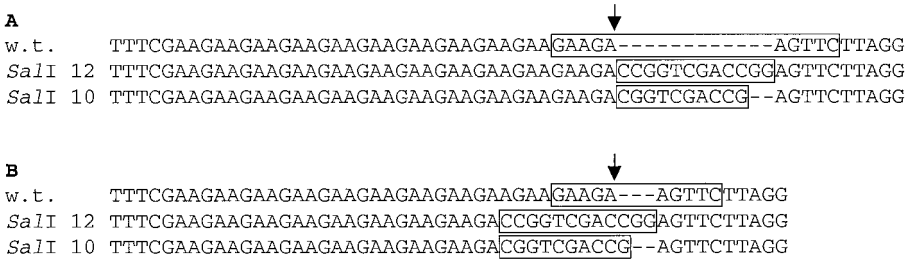

FIG. 5. Sequence alignment of the pMGA GAA repeat region of the unmodified promoter (w.t.) and promoters modified by the addition of 10-bp (*Sal*I 10) and 12-bp (*Sal*I 12) linkers. For the unmodified promoter, the *Xmn*I site is boxed, with an arrow showing the site of *Xmn*I cleavage. For the *Sal*I-modified promoters, the *Sal*I linker is boxed. Dashes refer to gaps in the nucleotide sequence introduced to realign the sequences 3′ of the *Xmn*I site or linker. (A) Alignment of the unmodified sequence to that of the original *Sal*I-modified sequences that had 12 GAA repeats and gave rise to LacZ$^-$ transformants; (B) alignment of sequences from LacZ$^+$ clones.

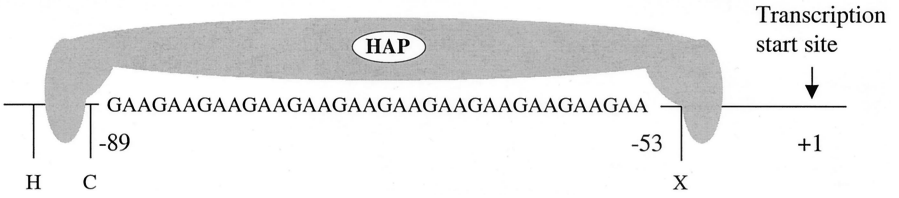

FIG. 6. Schematic illustration of the binding of HAP to sequences upstream and downstream of 12 GAA repeats. Numbers are distances (in nucleotides) from the indicated position to the transcription start site. Restriction enzyme sites: H, *Hin*dIII; C, *Csp*45I; X, *Xmn*I.

The results obtained for $LacZ^+$ clones containing the 12-bp linker were something of a surprise. It had been anticipated that $LacZ^+$ clones would have had seven GAA repeats, which would have restored the length of the promoter region to precisely that of the unmodified promoter. However, the insertion of the 12-bp linker was compensated by the loss of only three GAA repeats, resulting in a net gain of 3 bp (Fig. 5). Clones containing the 12-bp linker that had seven GAA repeats were not isolated in this study. Therefore, it is not known whether the phenotype of such clones would be $LacZ^-$ or $LacZ^+$. A finding of $LacZ^+$ clones with $(GAA)_7$ would indicate some degree of flexibility in the number of repeats required for gene expression. A finding of $LacZ^-$ $(GAA)_7$ clones would indicate that the overall length of the repeats is not the only consideration. The 12-bp *Sal*I linker might have introduced structural changes to the pMGA promoter region independent of length, perhaps as a consequence of the linker consisting of a palindrome of high G+C content.

Because the $LacZ^+$ subclones that contained the 12-bp linker had one more GAA repeat than had been anticipated, experiments were undertaken to determine whether the wild-type pMGA promoter might have more flexibility than previously realized with regard to the number of GAA repeats required for gene expression. Specifically, whether gene expression would occur when 13 GAA repeats were present was unknown. From a $LacZ^+$ clone with the wild-type promoter sequence and 12 GAA repeats, 42 independent $LacZ^-$ progeny were selected. The promoter region of each of these progeny was PCR amplified, and the nucleotide sequence was determined. Seven of the $LacZ^-$ subclones had 13 GAA repeats. The other $LacZ^-$ subclones had 14, 11, 10, 8, or 6 repeats. Thus, the wild-type promoter seems to strictly require 12 GAA repeats for gene expression, and 13 repeats, in particular, does not promote expression.

We propose the existence of a protein, hemagglutinin activator protein (HAP), that serves as a regulatory protein(s) that binds to the pMGA promoter region and stimulates transcription if 12 GAA repeats are present. Although negative regulators of pMGA transcription may also exist, they are not required to explain the data. Most pMGA gene family members in any given strain of *M. gallisepticum* are not transcribed and have a number of GAA repeats other than 12. Accordingly, it is our view that the pMGA promoter is insufficient to stimulate transcription in the absence of HAP. The pMGA genes lack a strong consensus sequence −10 and −35 promoter region upstream of the transcription start site (Fig. 1), which supports the contention that accessory factors are required. Because deletion of nucleotides upstream of the GAA repeats abolished GAA-dependent gene expression, we propose that one HAP-binding site is within these deleted sequences, as shown in Fig. 6. According to our model, a second HAP-binding site is located 3′ of the GAA repeats, and changes in the number of repeats affect transcription by varying the spacing between the two HAP-binding sites. The GAA repeats might also serve as a HAP-binding site, which as discussed previously (7) may involve a triple helical structure with an unpaired single strand. Proteins that regulate transcription initiation in *Mycoplasma* species have not been reported. Thus, interactions between HAP and the pMGA genes would serve as a model for studying the regulation of mycoplasmal gene expression.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI41113 from the National Institutes of Health.

We thank Warren Simmons for technical assistance.

REFERENCES

1. **Baseggio, N., M. D. Glew, P. F. Markham, K. G. Whithear, and G. F. Browning.** 1996. Size and genomic location of the pMGA multigene family of *Mycoplasma gallisepticum*. Microbiology **142:**1429–1435.
2. **Dybvig, K., and L. L. Voelker.** 1996. Molecular biology of mycoplasmas. Annu. Rev. Microbiol. **50:**25–57.
3. **Glew, M. D., N. Baseggio, P. F. Markham, G. F. Browning, and I. D. Walker.** 1998. Expression of the pMGA genes of *Mycoplasma gallisepticum* is controlled by variation in the GAA trinucleotide repeat lengths within the 5′ noncoding regions. Infect. Immun. **66:**5833–5841.
4. **Glew, M. D., G. F. Browning, P. F. Markham, and I. D. Walker.** 2000. pMGA phenotypic variation in *Mycoplasma gallisepticum* occurs in vivo and is mediated by trinucleotide repeat length variation. Infect. Immun. **68:**6027–6033.
5. **Glew, M. D., P. F. Markham, G. F. Browning, and I. D. Walker.** 1995. Expression studies on four members of the pMGA multigene family in *Mycoplasma gallisepticum* S6. Microbiology **141:**3005–3014.
6. **Levinson, G., and G. A. Gutman.** 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:**203–221.
7. **Liu, L., K. Dybvig, V. S. Panangala, V. L. van Santen, and C. T. French.** 2000. GAA trinucleotide repeat region regulates the M9/pMGA gene expression in *Mycoplasma gallisepticum*. Infect. Immun. **68:**871–876.
8. **Liu, L., D. M. Payne, V. L. van Santen, K. Dybvig, and V. S. Panangala.** 1998. A protein associated with monoclonal antibody-mediated agglutination of *Mycoplasma gallisepticum* is a member of the pMGA family. Infect. Immun. **66:**5570–5575.
9. **Markham, P. F., M. D. Glew, G. F. Browning, K. G. Whithear, and I. D. Walker.** 1998. Expression of two members of the pMGA gene family of *Mycoplasma gallisepticum* oscillates and is influenced by pMGA-specific antibodies. Infect. Immun. **66:**2845–2853.
10. **Razin, S., D. Yogev, and Y. Naot.** 1998. Molecular biology and pathogenicity of mycoplasmas. Microbiol. Mol. Biol. Rev. **62:**1094–1156.
11. **Wise, K. S., D. Yogev, and R. Rosengarten.** 1992. Antigenic variation, p. 473–489. *In* J. Maniloff, R. N. McElhaney, L. R. Finch, and J. B. Baseman (ed.), Mycoplasmas: molecular biology and pathogenesis. American Society for Microbiology, Washington, D.C.