INFECTION AND IMMUNITY, June 1996, p. 2193–2200 Vol. 64, No. 6
0019-9567/96/$04.00+0

DNA Sequencing and Gene Expression of the *emm* Gene Cluster in an M50 Group A Streptococcus Strain Virulent for Mice

DER-LI YUNG AND SUSAN K. HOLLINGSHEAD*

Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama

Received 22 May 1995/Returned for modification 20 July 1995/Accepted 29 March 1996

The strain B514, an M serotype 50 strain, is capable of causing a natural upper respiratory infection leading to death in mice, as reported by Hook et al. in 1960 (E. W. Hook, R. R. Wagner, and R. C. Lancefield, Am. J. Hyg. 72:111–119, 1960). Thus, this strain was of interest for use in developing an animal model for group A streptococcal colonization and disease. The *emm* gene cluster for this strain was examined by PCR mapping and found to contain three *emm* family genes and cluster pattern 5. PCR-generated fragments corresponding to the SF4 (*mrp50*), SF2 (*emmL50*), and SF3 (*enn50*) genes were cloned and the entire gene cluster was sequenced. The gene cluster has greater than 97% DNA identity to previously sequenced regions of the gene cluster of the M2 strain T2/44/RB4 if two small divergent regions that encode the mature amino terminus of the SF-2 and SF-3 gene products are not included. If expressed, the genes encode proteins which bind human immunoglobulin G (Mrp50 and EmmL50) or immunoglobulin A (Enn50). However, in isolates taken directly after passage in mice, the surface proteins arising from these genes were barely detectable. The transcription of each gene in the B514 strain was investigated by Northern (RNA) hybridization, and mRNA transcripts were detected and quantitated relative to those of the *recA* gene, a housekeeping gene. Transcription of all three *emm* family genes was found to be over 30-fold attenuated relative to transcription of the same genes in strain T2/44/RB4. This suggests that the positive regulator, Mga, either is not expressed in this strain or has a different requirement for activation; it also suggests that the capsule may be sufficient to inhibit phagocytosis under these circumstances.

Group A streptococci (GAS) are important human pathogens capable of causing a wide variety of infections, the most common of which are nasopharyngitis and impetigo. Of greater concern are the invasive infections, such as necrotizing fasciitis, pyomyositis, and toxic shock syndrome, which appear to have increased in incidence over the past decade. Rheumatic fever and acute glomerulonephritis are also poorly understood sequelae of GAS infection. Understanding the complex pathogenesis of the different group A streptococcal infections is important and will require knowledge of the functions of many related and unrelated virulence factors. Among these factors are a family of proteins which form dimeric fibrils projected from their carboxy termini which are embedded in the wall of the bacterial cell surface (33).

Within this family are the M proteins, defined as such because of their antiphagocytic function and thought to contain epitopes defining the antigenically variable determinants of serotype specificity in their amino termini (24, 25). Recent molecular analyses have revealed over 100 distinct *emm* family genes which encode proteins that are highly similar to the M protein, most of which do not have antiphagocytic properties but which do have other virulence-related phenotypes (18). For example, the genes *arp4* (27), *enn2* (4), *fcrA76* (13), and *mrp4* (44) encode immunoglobulin-binding proteins. Additional binding properties associated with members of this family are fibrinogen, plasminogen, factor H, albumin, and C4b binding protein (1, 22, 38, 46, 49).

All of the proteins encoded by *emm* family genes have highly similar sites towards their carboxy termini, and by alignment of these carboxy-terminal gene regions, four divergent forms within this gene family, called subfamilies (SF1 to SF4), have been identified (16). Individual genes in each subfamily may be mapped on the streptococcal chromosome by PCR amplification with subfamily-specific primers. One to three distinct copies of *emm* family genes are found to make up an *emm* gene cluster in any one GAS strain, but each strain has a different *emm* cluster. The *emm* cluster is located downstream from a transcriptional regulatory gene (*mga*, previously termed *vir* or *mry*) (7, 39) and upstream of the *scpA* gene, which encodes a C5a peptidase (8). Mga is a positive transcription factor for the genes in the *emm* gene cluster and the *scpA* gene (9, 32). At least five different chromosomal patterns of *emm* gene clusters have been found by examination of over 300 diverse GAS strains (16, 20), and there are significant associations of these patterns with tissue site of isolation and with streptococcal class (6, 17).

To understand the roles of M and M-like proteins in disease pathogenesis, it is necessary to construct and compare isogenic strains that differ in their expression of specific M and M-like proteins in animal models of disease. There are several complications when interpreting the results from animal models. First, because group A streptococci are primarily human pathogens, the capacity for animal models to reflect the human situation is limited to the capacity of individual strains to colonize and cause disease in rodents. Second, there are often three M or M-like proteins in group A streptococcal isolates

* Corresponding author. Mailing address: Dept. of Microbiology, BBRB 406/6, University of Alabama at Birmingham, Birmingham, AL 35294. Phone: (205) 934-0570. Fax: (205) 975-5479. Electronic mail address: hollingshead@orion.cmc.uab.edu.

TABLE 1. Oligonucleotides used for PCR mapping, cloning, and probes

Primer	Sequence	Description[a]
For PCR mapping		
SF3-F	5′-GGTAGAGCTGCTCAAACAGCTACAAGACCT-3′	2409–2438 *emmL2.2* (4)
SF2-F	5′-AACGCTAAAGTAGCCCCACAAGCTAACCGT-3′	1076–1105 *emmL2.1* (4)
SF4-F	5′-CCAACAAGACCATCACAAAC-3′	1318–1388 *fcrA76* (13)
SF2-LDR	5′-AATCTGCAGTATTCGCTTAGAAAATTAAAA-3′	47–68 *emmL2.1* (4)
SF4-LDR	5′-GAAATCCAAACAAGCACTACCTACTG-3′	224–249 *fcrA76* (13)
DP-2	5′-ATCCCTAATAGTCGCTTTTGAGG-3′	2091–2062 *emm6.1*, 674–645 *scpA* (8, 18)
SF2-R	5′-GTTAGCTTGTGGGGCTACTT-3′	1101–1076 *emmL2.1* (4)
SF3-R	5′-GCTGTTTGAGCAGCTCTACC-3′	2429–2409 *emmL2.2* (4)
CW-1R	5′-GAATGGGTTAGCTGTTTC-3′	Composite from *emm* gene alignment
For cloning and probes		
UP-8	5′-TGAAAACAGCTCAAAAAAACTGACC-3′	Pair with IG5-R for cloning *mrp50* (SF4)
IG5-R	5′-GGCTAGAAAAGATAGTGTGGGTTG-3′	Pair with UP-8 for cloning *mrp50* (SF4)
SF2-LDR	5′-AATCTGCAGTATTCGCTTAGAAAATTAAAA-3′	Pair with OM2-4 for cloning *emmL50* (SF2) and with OM2-3 for cloning *enn50* (SF3)
LDRALL	5′-GGATCCCCGGGCATCCGTAGCAGTCGCT-3′	Probe of SF4 gene with 2127-R
2127R	5′-TTCTTGGTTGGTTGCTGCTAATT-3′	Probe of SF4 gene with LDRALL
915-R	5′-GTTCTTGATAACGTTTTTCTACTTCTCG-3′	Probe of SF2 gene with SF2-LDR
Enn50-F	5′-TGGAACTTCTGTAAATAATGG-3′	Probe of SF3 gene with Enn50-F
245-R	5′-TTCTTTAGTAGTCTTAGCTAAAGTTGT-3′	Probe of SF3 gene with 245-R

[a] For all PCR mapping primers except CW-1R, the description includes position (in base pairs), gene, and (parenthetically) applicable reference(s).

from both humans and mice, rather than a single protein, so evaluation of virulence contributions depends upon knowledge of the entire cluster. Third, variable gene expression of the different M and M-like proteins in the same strain has been reported, so monitoring expression is important. For these reasons we investigated the virulence gene characteristics of a natural group A streptococcal strain which is virulent for mice.

The M serotype 50 strain B514 was first identified as the cause of several outbreaks of natural infections in mice (21). This strain was and is highly mucoid on blood agar plates, indicating that it is encapsulated. This strain continues to be more virulent for mice than other GAS strains; when administered intranasally, strain B514 had a 100% lethal dose of 10^5 CFU, as compared with $>10^7$ CFU for five other GAS strains (26), and as little as 20 CFU can give an intense inflammatory response in mouse lymphoid tissue (50, 51). The reasons for the extreme virulence of the B514 strain for mice are not understood. The B514 strain has also been used to study the protective effects of vaccination (26, 45) and group A streptococcal pneumonia (23). Therefore, this strain is of continuing interest for studies of streptococcal pathogenesis. To begin examining the role of multiple phenotypes associated with M and M-like proteins in pathogenesis, the *emm* cluster of this strain was mapped by PCR, individual genes of the cluster were cloned, the gene cluster was sequenced, and the expression of each gene was examined by quantitative Northern (RNA) hybridization analysis.

MATERIALS AND METHODS

Bacterial strains and *emm* gene family nomenclature. Group A streptococcal strain B514 is the serotype 50 typing strain, and T2/44/RB4 is the serotype 2 typing strain (40). T2/44/RB4 was received from D. Bessen (Yale University, New Haven, Conn.). A streptomycin-resistant spontaneous variant of B514, recently tested in a mouse model of virulence, was received from M. Caparon (Washington University, St. Louis, Mo.) and is designated herein as B514-Sm1. B514-Sm2 is the same strain after an additional passage in laboratory medium.

The regulatory gene upstream of the *emm* gene cluster was previously called either *mry* or *vir*. It is now called *mga* (43). All homologues of *emm* are referred to as *emm* family genes, but the specific designation *emm* is reserved for genes encoding proteins with known antiphagocytic functions. Therefore, the functionally neutral terms *mrp*, *emmL*, and *enn* are used for the SF4, SF2, and SF3 genes, respectively, in this manuscript.

PCR and gene cloning. B514 chromosomal DNA was isolated by a microwave method described previously (20). PCR was used to map the gene cluster with subfamily-specific primers, to generate fragments for the cloning of individual genes, and to generate fragments for gene-specific DNA probes for individual genes. All of the oligonucleotide primers used in this study are listed in Table 1, and an overview map of all the PCR fragments is included in Fig. 1. PCR reactions designed to map the genes in the cluster were carried out as described by Hollingshead et al. (20); a standard PCR mixture of 50 μl containing 2.5 mM $MgCl_2$, 200 μM each deoxynucleoside triphosphate, 50 pmol of each primer, and 2.5 U of *Taq* DNA polymerase was used for all other PCRs, with a cycle of 1 min at 95°C, 1 min at 62°C, and 5 min at 72°C, repeated 35 times. For cloning, PCR products were first purified from low-melting-temperature agarose gels (SeaPlaque agarose; FMC) by GeneClean (BIO 101, Inc., La Jolla, Calif.) and ligated to the pCR vector (Invitrogen Co.) before transformation into *Escherichia coli* INVaF′. White colonies were picked from Luria agar plates containing 50 μg of ampicillin per ml, and the plasmid DNA was isolated to screen for plasmids containing the DNA insert of the correct size. The plasmids pUAB021, pUAB022, and pUAB023 contain the SF4, SF2, and SF3 genes, respectively, from the *emm* gene cluster of B514 that were cloned in the pCR vectors (Fig. 1).

DNA sequencing. Plasmid DNAs from pUAB021, pUAB022, and pUAB023 were sequenced by the dideoxy chain termination method (41) following procedures described in the Sequenase manual (version 1.0; United States Biochemical Corp.) with the primers listed in Table 1. The regions between the cloned genes were sequenced from additional PCR fragments with an Applied Biosciences automated DNA sequencer.

Surface protein analysis and mouse passage. Strain B514 was passaged in mice intraperitoneally by inoculating either 10^6 or 10^4 bacteria. If death due to sepsis had not occurred by 48 h, mice were killed at that time, and individual organs were then dissected and homogenized. The homogenates were plated on sheep blood agar plates, and individual bacterial colonies from each organ were chosen for further analysis. CNBr-extracted surface proteins were analyzed by the method of Raeder et al. (36).

RNA purification. B514 cultures were grown in 37°C standing cultures in Todd-Hewitt yeast broth plus 20 mM glycine in a 5% CO_2–20% O_2 atmosphere. Total cellular RNA was isolated from streptococcal cultures at optical densities at 600 nm between 0.6 and 1.0. Total cellular RNA was purified following centrifugation on CsCl gradients and stored in the presence of vanadyl ribonucleosides as previously described (19).

Northern hybridization. Total cellular RNA (20 μg) from each strain was separated by electrophoresis on formaldehyde-containing agarose gels (28) and transferred to nylon membranes by capillary transfer. The probes used for Northern hybridization were PCR-generated fragments of each of the three genes in the *emm* gene cluster of B514 that include only the gene-specific regions of each gene. These gene-specific probes, shown in Fig. 1, included bp 181 to 1091 of the SF4 gene (*mrpA50*), bp 1643 to 2036 of the SF2 gene (*emmL50*), and bp 3215 to 3472 of the SF3 gene (*enn50*). A *recA* gene probe of 314 bp was previously cloned from GAS strain D471 as described by Dybvig et al. (10). All DNA fragments used for probes were purified from low-gelling-temperature agarose and then radiolabeled with ^{32}P by random primer labeling (Boehringer Mannheim Corp.). The probes had specific activities of 2×10^6 to 5×10^6 cpm/pmol, and they were

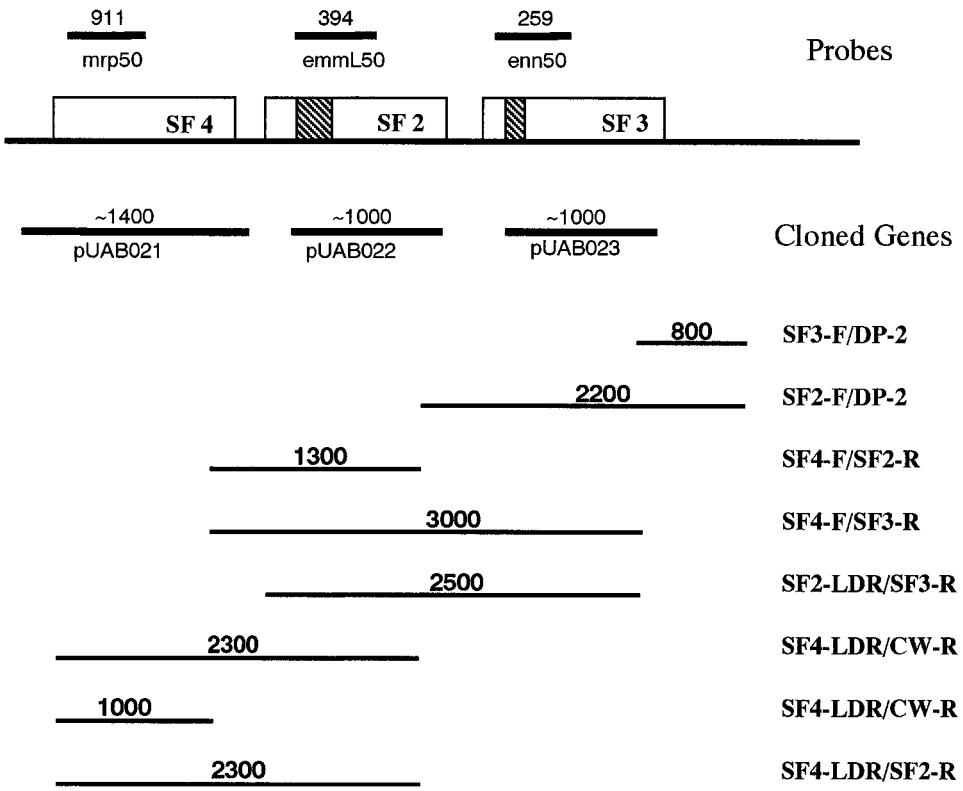

FIG. 1. Chromosomal map of *emm* gene cluster of serotype M50 strain B514. The numbers indicate sizes in base pairs. PCR amplification with the subfamily-specific primers shown in the bottom half of the figure was used to map three genes: an *mrp50* (SF4) gene, an *emmL50* (SF2) gene, and an *enn50* (SF3) gene. Primers for the amplification products are listed to the right of the lines indicating the PCR products. Gene-specific PCR fragments that were cloned for nucleotide sequencing and smaller fragments that were used as DNA probes are indicated below and above the genes, respectively. The two divergent regions in the *emmL50* and *enn50* genes (i.e., regions that are significantly different from T2/44/Rb4 regions) are indicated by the two small hatched boxes.

used at concentrations of 10 pmol/ml. Northern hybridizations were performed in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–50% formamide at 42°C. Hybridization was visualized by exposure to XAR-1 X-ray film.

Quantitation of transcript levels. Radioactive bands on the blots were quantitated by using a Molecular Dynamics PhosphorImager. All blots that were to be compared were exposed overnight for equivalent time periods on the same screen, and counts were then taken. The internal radiolabeling method used for the probes allowed the detection of small quantities of transcript (100-fold below the detection level for the *recA* transcript). The *recA* gene transcript was used as an internal control in order to compare the relative transcript levels of different genes in the same strain and to roughly compare transcript levels of the same gene in different strains.

First, quantitation controls were performed to examine potential signal variability due to loading volumes, transfer procedures, or stripping of blots for reprobing. There was no loss of signal with stripping of blots, and there was less than a twofold difference in duplicate samples. For experimental runs, three identical blots were first probed with the three *emm*-specific probes; then the blots were stripped and reprobed with the *recA*-specific probe. Ratios of each *emm* gene transcript to the *recA* transcript from the same lane and the same blot were then calculated for reporting. In comparing the PhosphorImager counts of different transcripts, corrections for differing probe length were included. Any differences greater than 10-fold remaining after a correction for probe length were considered significant.

Nucleotide sequence accession number. The DNA sequence of the *emm* gene cluster for serotype M50 has been submitted to GenBank, and the nucleotide accession number is U520008.

RESULTS

PCR mapping of the gene cluster and cloning of individual genes. PCRs were carried out with subfamily-specific primer pairs to examine the structure and order of the *emm* gene cluster of M50 strain B514 as described previously (20). Figure 1 shows the approximate sizes and locations of amplified PCR products of the B514 chromosomal DNA, on the basis of which the strain was shown to have chromosomal pattern 5. The order of genes in B514 is SF4 gene-SF2 gene-SF3 gene, and they are located between the *mga* and *scpA* genes. In order to sequence the individual genes from strain B514, PCR products that encompassed most of each gene were individually cloned in the plasmids pUAB021, pUAB022, and pUAB023 (Fig. 1).

DNA sequence. Figure 2 shows the DNA sequence of the entire *emm* gene cluster of B514. The gene cluster of serotype M50 strain B514 is over 97% identical to the gene cluster of serotype M2 strain T2/44/RB4 with the exception of two divergent regions characteristic of mosaic genes. The two small divergent regions of the M50 and M2 strains (hatched boxes on the map in Fig. 1) are present near the amino termini, immediately after the signal peptide, in both the SF2 and the SF3 genes. In the M2 strain, both the SF2 gene (*emmL2.1*) and the SF3 gene (*emmL2.2*) were previously characterized for immunoglobulin-binding characteristics and completely sequenced (4) and the SF4 gene was partially sequenced (5).

The SF4 gene (*mrp50*) encodes a protein with 97% identity to FcrA76 which binds the Fc region of human immunoglobulin G1 (IgG1), IgG2, and IgG4 proteins. Like other SF4 genes, instead of having central C repeat regions, the *mrp50* gene has three A repeat regions which have been characterized in *fcrA76* and *mrp4* genes as being the binding site for the IgG Fc (13, 31). Therefore, the *mrp50* gene encodes an IgG-binding protein. Protein expressed from this gene in *E. coli* binds human IgG when expressed and purified in vitro (data not shown).

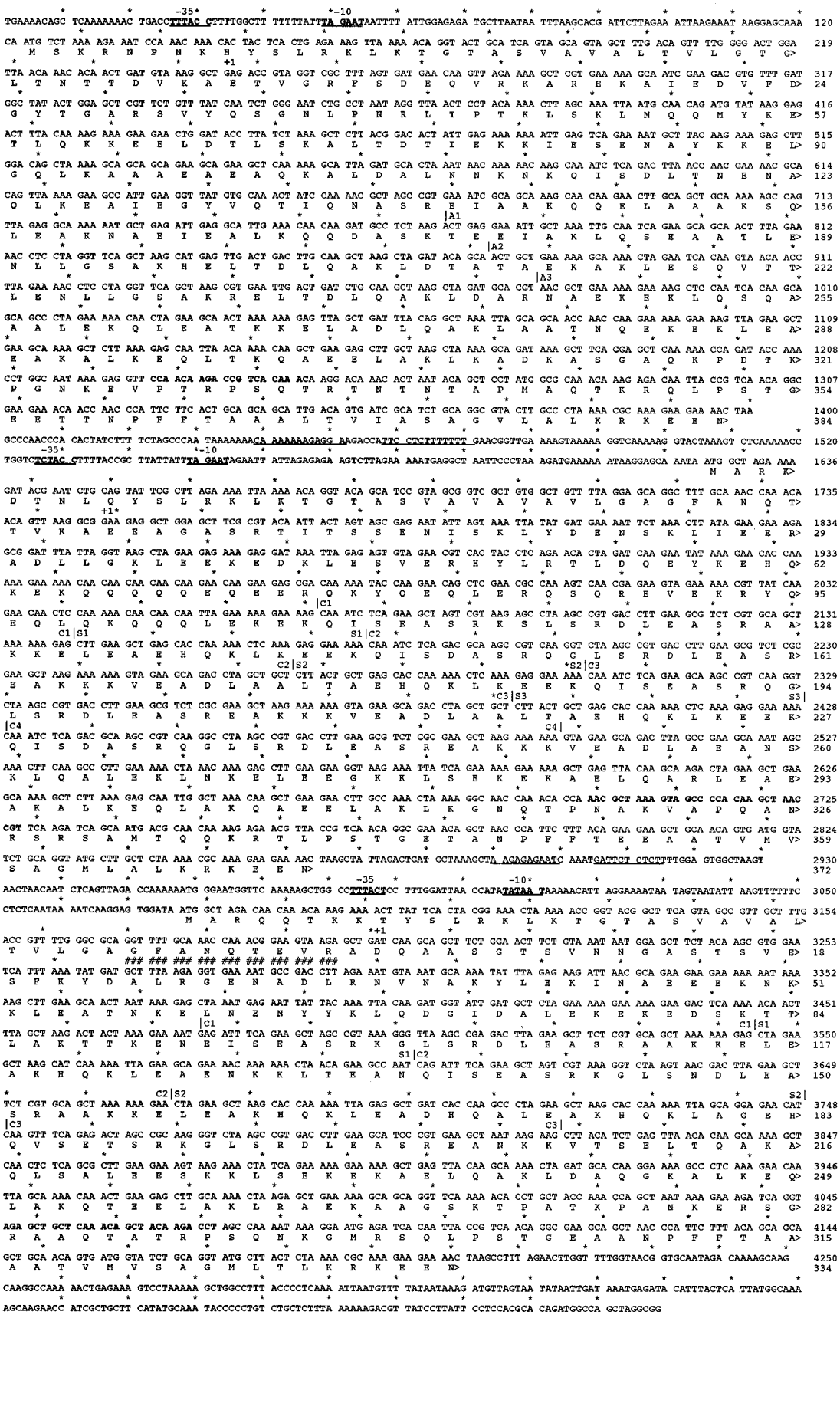

FIG. 2. Nucleotide and deduced amino acid sequences of *emm* gene cluster in M50 serotype strain B514. The DNA sequence was determined by dideoxy sequence analysis with the clones and PCR-generated fragments shown in Fig. 1. Putative promoter sequences (−35 and −10) and ribosomal binding sites are indicated by boldface underlining. The inverted repeats that may serve as potential transcriptional terminator structures are indicated by nonbold underlining. Amino acid residue numbers are given, beginning with the first residue of the processed form, on the basis of sequence similarity with the leader sequences of other M and M-like protein genes. The *mrp50* (SF4) gene, as indicated here, contains A repeats (A). C repeats (C) are present in *emmL50* and *enn50* genes. The intervening spacers (S) are indicated. The pound signs are directly above the IgA binding site in the *enn50* (SF3) gene.

The SF2 gene, *emmL50*, also encodes a potential IgG-binding protein. Its sequence is highly similar (97% identity) to that of the gene *emmL2.1*, in the entire coding region with the exception of a small region encoding amino acids 1 to 81 of the mature protein (bp 1749 to 1990). The remaining portion of the mature protein contains regions homologous to protH (SF1), *emm49* (SF2), and *emmL2.1* (also SF2), all of which bind to the IgG Fc region (4, 11, 12). Mutations in the homologous region in recombinant molecules of protH and *emmL2.1* affect IgG binding (6). A recombinant *emm50* gene product expressed in vitro was also shown to bind human IgG (data not shown).

The 5′ end of the SF3 gene of strain B514 contains a domain (amino acids 24 to 32 [ALRGENADLR]) identical to that required for IgA binding in *emmL2.2* (SF3) (2) and similar to that in the genes *arp4* and *arp60* (SF2) (14), which are also human IgA-binding proteins. Again, protein from this gene expressed in vitro was found to bind human IgA (data not shown), although there is a published report that strain B514 does not bind IgA (42).

CNBr-extracted surface proteins. The method of Raeder et al. (36) was used to examine the expression of surface proteins in B514. This procedure preferentially releases the *emm* gene family proteins in all streptococcal strains in which it has been examined (36). CNBr extraction of B514 resulted in very little surface protein relative to the amounts obtained from extraction of other strains even upon concentration (see Fig. 4).

RNA transcripts. Transcription of each of the three genes in the *emm* gene cluster of B514 was analyzed by Northern hybridization gels. PCR fragments from the 5′ end region of each gene were used as the *mrp50*, *emmL50*, and *enn50* probes (Fig. 1); limiting probes to the 5′ end of each gene makes them gene specific and allows the differentiation of the transcripts of the three separate family members in the *emm* gene cluster. A transcript of the *recA* gene of *Streptococcus pyogenes* was used as an internal marker for standardization of the loading quantities of RNA samples.

Figure 3 displays the autoradiographs of individual gene transcripts of the *emm* gene cluster of B514. Monocistronic RNA transcripts of each of the three genes in the *emm* gene cluster were detected, with sizes appropriate for previously described transcriptional start and stop sites of these genes (15, 34). Duplicate cultures and RNA preparations of the mouse-virulent strain B514 showed the same levels of each relative transcript. The transcripts of the *mrp50* and *emmL50* genes were detected in quantities similar to that of a *recA* transcript for a housekeeping gene, but the *enn50* gene transcript level was nearly 50-fold lower and barely detectable above the background. The very low level of the *enn50* transcript in B514 might explain the inability of Schalen to detect cell surface binding of human IgA in B514 (42).

In order to quantitate the relative levels of expression of these three genes in strains B514 and T2/44/RB4, ratios of counts detected for *emm* transcripts to counts detected for *recA* transcripts on the same gels were calculated (Table 2). Rough estimates of the relative transcript levels for genes in

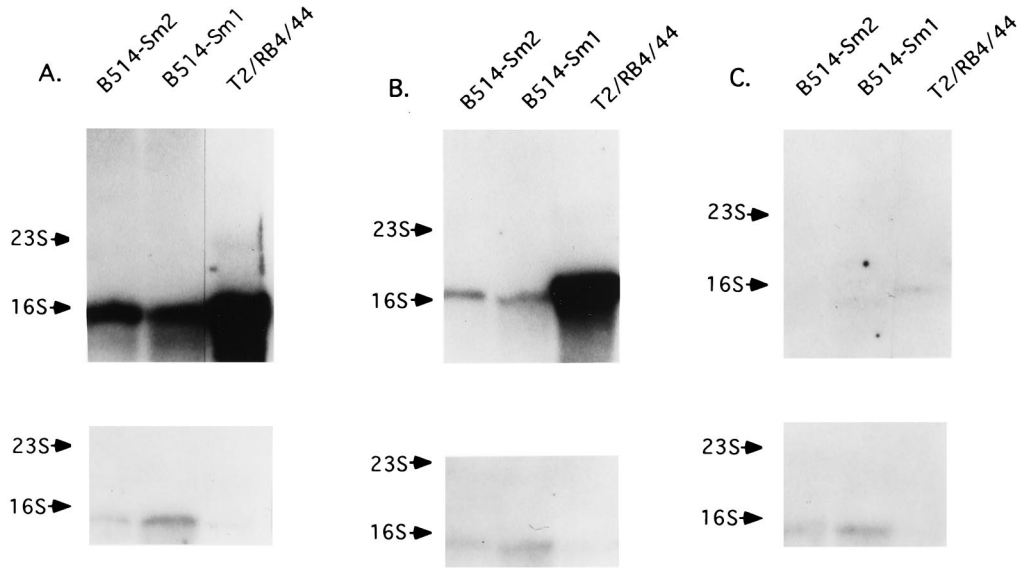

FIG. 3. Northern hybridization to show relative transcript levels in T2/44/RB4 (M2) and two isolates of strain B514 (M50-Sm1 and -Sm2). All strains had been recently passaged in mice, and RNA was extracted from cells grown in vitro under Mga-positive conditions. In this experiment, DNA probes specific for the *mrp50*, *emmL50*, and *enn50* genes included only the region that encodes the 5′ mature end of each protein (shown in Fig. 1). (A) Transcripts detected with the *mrp50* probe (SF4); (B) transcripts detected with the *emmL50* probe (SF2); (C) transcripts detected with the *enn50* probe (SF3). Each lane contains 20 μg of total cellular RNA from each culture. The exposure time was 4 h (A and B) and 48 h (C). The positions of two ribosomal RNAs are shown on the left. The bottom half of each panel shows the *recA* transcript detected on the same blot, used as the internal standard.

TABLE 2. Ratios of counts of specific *emm* gene transcripts to those of *recA* control transcript[a]

Gene probe	Transcript ratio in strain	
	B514 (M50)	T2/44/RB4 (M2)
SF4 gene (*mrp50*)	0.7	32
SF2 gene (*emmL50*)	0.6	59
SF3 gene (*enn50*)	0.02	2

[a] Corrections were made for variations in probe length.

the *emm* cluster of each strain can be made from the values in Table 2. Thus, the ratio of *mrp50* transcripts to *emmL50* transcripts in strain B514 was roughly 1:1, while the ratios of both *mrp50* and *emmL50* transcripts to those of the *enn50* gene in B514 were approximately 35:1. In strain T2/44/RB4, the ratio of transcripts within the gene cluster (*mrp*:*emmL*:*enn*) was similar to the internal ratio in B514 (Table 2), but T2/44/RB4 differed from B514 in that it had much more abundant levels of all three *emm* family transcripts in cultures grown under identical laboratory conditions (Fig. 3).

Analysis of *emm* family gene expression after passage in mice. To determine if the very low levels of gene expression were significantly different from the gene expression during passage in mice, individual B514 isolates from dying mice were examined. Bacteria were introduced to mice by the intraperitoneal route, and individual isolates from dissemination sites such as the spleen, lung, and brain were examined for surface proteins by the CNBr extraction method. All of the individual B514 isolates from mice had undetectable surface protein expression compared with that of two other group A strains (Fig. 4). Gene cluster expression in fresh passage isolates was indistinguishable from that in the isolate in which expression was initially quantitated. We infer that expression of this gene cluster may not be critical for virulence in this model system and that phase variation to an "on" state is not the usual outcome of passage of this strain in mice. Additionally, all isolates obtained after passage in mice were highly mucoid on blood agar plates, suggesting high-level expression of the capsule in B514 isolates in vivo as well as in vitro.

DISCUSSION

Strain B514 is important for mouse models of group A streptococcal infection because it causes natural infections in mice. On the basis of the DNA sequence of the *emm* gene cluster in strain B514, it is evident that there are two genes that encode probable IgG-binding proteins and one gene that encodes a probable IgA-binding protein. Binding studies of proteins produced in vitro from cloned genes showed the predicted binding capabilities (data not shown); however, our studies show that these genes are not expressed in the streptococcal host. By PCR mapping, the *emm* gene cluster was characterized as having chromosomal pattern 5, which has the SF4, SF2, and SF3 genes (in that order) located between the *mga* and *scpA* genes. This pattern is positively associated with the ability to produce a serum opacity factor (SOF) that causes mammalian sera to become opalescent (3, 20). In our collection, 89% of pattern 5 strains are SOF^+ while only 2% of non-pattern 5 strains are SOF^+. B514 is SOF^-, even though it is a pattern 5 strain, and it fails to produce an SOF.

The sequence of the M50 gene cluster was >97% identical to that of the gene cluster of a serotype M2 strain, T2/44/RB4, except for two small regions near the amino termini of the SF2 gene (81 amino acids) (Fig. 2) and the SF3 gene (23 amino acids) (Fig. 2). One of these two heterogeneous regions may define the M50 type-specific domain of B514, differentiating M50 strains from M2 strains in the Lancefield serotyping system for group A streptococci. Although serotype M2 strains are frequently reported to be associated with human infections, M2 strains are not virulent for mice unless they have been passaged several times through mice to select for virulent variants. In contrast, the serotype M50 strains are not usually found among human isolates and are among the very few group A streptococci originating from multiple mouse colony outbreaks (21).

Low transcript levels for the entire cluster in strain B514. Although transcripts of all three genes in strain B514 could be detected, quantitation studies revealed at least 30-fold less transcript in B514 than in strain T2/44/RB4 (Table 2). The low expression levels were initially surprising because laboratory passage of the strain had been minimized after recent testing for full virulence in mice and care was taken to use growth conditions under which maximum expression of *emm* genes has been obtained. Low levels of expression were consistently found for multiple RNA preparations made from the same strains. The low transcript levels did seem to explain our inability to detect surface proteins in B514 by a cyanogen bromide extraction procedure (Fig. 4). In most streptococcal strains, two or more proteins are extracted by this procedure, but with B514 very little, if any, protein is extracted.

The very low level of transcription of all three genes in the M50 cluster may represent the uninduced levels of transcription for these genes. These genes are positively regulated by the protein Mga (previously called Mry or Vir) (7), and the regulation involves binding of Mga protein to conserved sites present near the −35 region of the promoters of these genes (29, 34). All of the sequences of the promoter regions are identical in strains B514 and T2/44/RB4 (Fig. 2), so regulation site differences are unlikely to explain the difference in expression between these two strains. Thus, the extremely low level of expression in B514 may indicate that the *mga* gene is either defective or only active under conditions that differ from those required by most streptococcal strains. Footprinting experiments have identified an Mga-binding site for the *emm6* and *scpA* genes located near the −35 region of only one of the two promoters for these genes identified by primer extension experiments (37). It is possible that the basal level of transcrip-

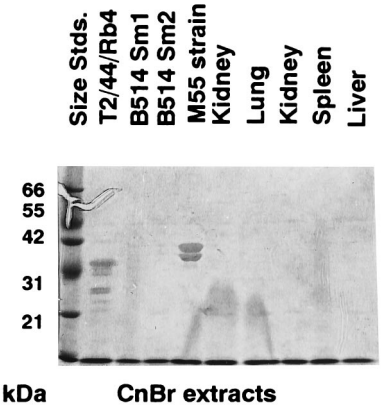

FIG. 4. CNBr extracts of surface proteins from B514-Sm1, from control strains, and from individual isolates obtained immediately after passage in mice. Lane 1, size standards; lanes 2 and 5, protein from an M2 strain and an M55 strain, respectively, isolated from mice (positive controls for strains expressing *emm* loci proteins); lanes 6 to 10, CNBr-extracted proteins from isolates obtained immediately after passage in mice. Stds., standards.

tion found in the B514 strain results in transcripts originating from promoters that are not under the control of the Mga gene product. A defect in Mga expression in this strain may also explain why the strain is SOF^- even though it is a pattern 5 strain. DNA sequencing of the *sor* gene locus revealed a promoter sequence that matches the consensus sequence for Mga binding (29, 37).

The SF3 gene is poorly expressed relative to other genes. As has been previously reported for the genes in this regulon (4, 34), there was a large difference in the relative transcript levels for different genes in the *emm* gene cluster of a specific strain. For example, the SF2 transcript was 30-fold more abundant than the SF3 gene transcript in both the T2/44/RB4 and B514 strains (Table 2). This quantitative estimate of difference in transcription levels is consistent with the over 32-fold difference seen by Bessen and Fischetti (4) and the 16-fold difference for analogous transcripts in strain CS101 seen by Podbielski et al. (34) as well as with additional reports of low-level expression of a gene in this position (48). This could be the result of differential regulation of individual promoters or of differences in promoter strength.

A recent publication highlighted a role for group A streptococcal capsules in pharyngeal colonization of mice (47). The B514 strain is highly mucoid and encapsulated. If the roles in pathogenesis for the capsule and the *emm* gene cluster are overlapping, encapsulation may be important for the virulence of strain B514 for mice in light of the poor expression of *emm* genes in this strain. Whether these roles are equivalent to the roles in human infections is very difficult to address, but strains of certain M serotypes are frequently heavily encapsulated.

Switching from high to low expression of M proteins is referred to as phase variation, and the molecular mechanism by which phase variation occurs is unknown (39). To further test whether in vivo expression of the gene cluster was likely to be much higher in strain B514, passages in mice were performed and expression of this gene cluster in multiple isolates sampled from body sites of dissemination was examined. The low-level-expression or "off" state of this gene cluster in the mouse isolates appeared to be the usual state (Fig. 4).

The *emmL50* gene was found to have a mosaic structure. The heterogeneous 81-bp region that was different from the *emmL2.1* sequence was instead very similar (>80% identity) to the published sequence of *enn5.6183.2*, which is from a serotype 5 strain (48). The *enn5.6183.2* gene is in a subfamily distinct from that of *emmL2.1* or *emmL50*; genes in different subfamilies show an average divergence of 17 to 27% while genes in the same subfamily show an average divergence of 3%. The M5 strain from which the *enn5.6183.2* gene was isolated has *emm* gene cluster pattern 2, containing two SF1 genes and none from the other subfamilies. The cluster patterns are also thought to have resulted from evolution of this multigene regulon. The mosaicism of the *emmL50* gene could be explained if a recombination event occurred between an M2-like strain and a strain carrying an *enn5.6183.2*-like gene. This could result in most of the *emm50* gene being almost identical to an *emm2.1* gene while another part of the gene is almost identical to the *enn5*-like gene. The difference in chromosomal patterns (or gene context) suggests that past recombination events involved horizontal transfer between strains that were genetically quite distant from each other, at least with respect to this particular segment of the chromosome. Evidence for the contribution of horizontal transfer events in the evolution of the *emm* gene family has recently been discussed (5, 17, 30, 35, 48).

ACKNOWLEDGMENTS

This work was supported by a grant-in-aid from the American Heart Association to S.K.H. Computer support was provided in part by a grant to the Center for AIDS Research (P30 AI27767) at the University of Alabama at Birmingham.

We are indebted to Terri L. Readdy and to Debra Bessen for discussions throughout the course of this work and to Linda Husmann and June Scott for sharing unpublished data regarding the B514 strain animal models.

REFERENCES

1. **Berge, A., and U. Sjobring.** 1993. PAM, a novel plasminogen-binding protein from Streptococcus pyogenes. J. Biol. Chem. **268:**25417–25424.
2. **Bessen, D. E.** 1994. Localization of immunoglobulin A-binding sites within M or M-like proteins of group A streptococci. Infect. Immun. **62:**1968–1974.
3. **Bessen, D. E., and V. A. Fischetti.** 1990. Differentiation between two biologically distinct classes of group A streptococci by limited substitutions of amino acids within the shared region of M protein-like molecules. J. Exp. Med. **172:**1757–1764.
4. **Bessen, D. E., and V. A. Fischetti.** 1992. Nucleotide sequences of two adjacent M or M-like protein genes of group A streptococci: different RNA transcript levels and identification of a unique immunoglobulin A-binding protein. Infect. Immun. **60:**124–135.
5. **Bessen, D. E., and S. K. Hollingshead.** 1994. Allelic polymorphism of *emm* loci provides evidence for horizontal gene spread in group A streptococci. Proc. Natl. Acad. Sci. USA **91:**3280–3284.
6. **Bessen, D. E., and S. K. Hollingshead.** 1995. Horizontal transfer and mosaic-like *emm* gene structures in group A streptococci, p. 169–173. *In* J. J. Feretti, M. S. Gilmore, T. R. Klaenhammer, and J. Brown (ed.), Genetics of streptococci, enterococci and lactococci. Karger, Basel.
7. **Caparon, M. G., and J. R. Scott.** 1987. Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. Proc. Natl. Acad. Sci. USA **84:**8677–8681.
8. **Chen, C. C., and P. P. Cleary.** 1990. Complete nucleotide sequence of the streptococcal C5a peptidase gene of Streptococcus pyogenes. J. Biol. Chem. **265:**3161–3167.
9. **Cleary, P. P., D. LaPenta, D. Heath, E. J. Haanes, and C. Chen.** 1991. A virulence regulon in *Streptococcus pyogenes*, p. 147–151. *In* G. M. Dunny, P. P. Cleary, and L. L. McKay (ed.), Genetics and molecular biology of streptococci, lactococci, and enterococci. American Society for Microbiology, Washington, D.C.
10. **Dybvig, K., S. K. Hollingshead, D. G. Heath, D. B. Clewell, F. Sun, and A. Woodard.** 1992. Degenerate oligonucleotide primers for enzymatic amplification of *recA* sequences from gram-positive bacteria and mycoplasmas. J. Bacteriol. **174:**2729–2732.
11. **Frick, I. M., P. Akesson, J. Cooney, U. Sjobring, K.-H. Schmidt, H. Gomi, S. Hattori, C. Tagawa, F. Kishimoto, and L. Bjorck.** 1994. Protein H—a surface protein of *Streptococcus pyogenes* with separate binding sites for IgG and albumin. Mol. Microbiol. **12:**143–151.
12. **Haanes, E. J., and P. P. Cleary.** 1989. Identification of a divergent M protein gene and an M protein-related gene family in *Streptococcus pyogenes* serotype 49. J. Bacteriol. **171:**6397–6408.
13. **Heath, D. G., and P. P. Cleary.** 1989. Fc-receptor and M-protein genes of group A streptococci are products of gene duplication. Proc. Natl. Acad. Sci. USA **86:**4741–4745.
14. **Heden, L., and G. Lindahl.** 1993. Conserved and variable regions in protein Arp, the IgA receptor of *Streptococcus pyogenes*. J. Gen. Microbiol. **139:**2067–2074.
15. **Hollingshead, S. K.** 1987. Nucleotide sequences that signal the initiation of transcription for the gene encoding type 6 M protein in *Streptococcus pyogenes*, p. 98–100. *In* J. J. Feretti and R. Curtiss III (ed.), Streptococcal genetics. American Society for Microbiology, Washington, D.C.
16. **Hollingshead, S. K., J. Arnold, T. L. Readdy, and D. E. Bessen.** 1994. Molecular evolution of a multigene family in group A streptococci. Mol. Biol. Evol. **11:**208–219.
17. **Hollingshead, S. K., and D. E. Bessen.** 1995. Evolution of the *emm* gene family: virulence gene clusters in group A streptococci, p. 163–169. *In* J. J. Feretti, M. S. Gilmore, T. R. Klaenhammer, and J. Brown (ed.), Genetics of streptococci, enterococci and lactococci. Karger, Basel.
18. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1986. Complete nucleotide sequence of type 6 M protein of the group A streptococcus. J. Biol. Chem. **261:**1677–1686.
19. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207:**196–203.
20. **Hollingshead, S. K., T. L. Readdy, D. L. Yung, and D. E. Bessen.** 1993. Structural heterogeneity of the *emm* gene cluster in group A streptococci. Mol. Microbiol. **8:**707–717.
21. **Hook, E. W., R. R. Wagner, and R. C. Lancefield.** 1960. An epizootic in Swiss

mice caused by a group A streptococcus, newly designated type 50. Am. J. Hyg. **72:**111–119.
22. **Horstmann, R. D., H. J. Sievertsen, J. Knobloch, and V. A. Fischetti.** 1988. Antiphagocytic activity of streptococcal M protein: selective binding of complement control factor H. Proc. Natl. Acad. Sci. USA **85:**1657–1661.
23. **Husmann, L., and J. R. Scott.** Personal communication.
24. **Jones, F. K., and V. A. Fischetti.** 1988. The importance of the location of antibody binding on the M6 protein for opsonization and phagocytosis of group A M6 streptococci. J. Exp. Med. **167:**1114–1123.
25. **Kraus, W., E. Haanes-Fritz, P. P. Cleary, J. M. Seyer, J. B. Dale, and E. H. Beachey.** 1987. Sequence and type-specific immunogenicity of the amino-terminal region of type 1 streptococcal M protein. J. Immunol. **139:**3084–3090.
26. **Kurl, D. N., A. Stjernquist-Desatnik, C. Schalen, and P. Christensen.** 1985. Induction of local immunity to group A streptococci type M50 in mice by non-type-specific mechanisms. Acta Pathol. Microbiol. Immunol. Scand. Sect. B Microbiol. **93:**401–405.
27. **Lindahl, G.** 1989. Cell surface proteins of a group A streptococcus type M4: the IgA receptor and a receptor related to M proteins are coded for by closely linked genes. Mol. Gen. Genet. **216:**372–379.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **McIver, K. S., A. S. Heath, B. D. Green, and J. R. Scott.** 1995. Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A streptococcus. J. Bacteriol. **177:**6619–6624.
30. **Musser, J. M., V. Kapur, J. Szeto, X. Pan, D. S. Swanson, and D. R. Martin.** 1995. Genetic diversity and relationships among *Streptococcus pyogenes* strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. Infect. Immun. **63:**994–1003.
31. **O'Toole, P., L. Stenberg, M. Rissler, and G. Lindahl.** 1992. Two major classes in the M protein family in group A streptococci. Proc. Natl. Acad. Sci. USA **89:**8661–8665.
32. **Perez-Casal, J., M. G. Caparon, and J. R. Scott.** 1991. Mry, a *trans*-acting positive regulator of the M protein gene of *Streptococcus pyogenes* with similarity to the receptor proteins of two-component regulatory systems. J. Bacteriol. **173:**2617–2624.
33. **Phillips, G. N., P. F. Flicker, C. Cohen, B. N. Manjula, and V. A. Fischetti.** 1981. Streptococcal M protein: alpha-helical coiled coil structure and arrangement on the cell surface. Proc. Natl. Acad. Sci. USA **78:**4689–4693.
34. **Podbielski, A., A. Flosdorff, and J. Weber-Heynemann.** 1995. The group A streptococcal *virR49* gene controls expression of four structural *vir* regulon genes. Infect. Immun. **63:**9–20.
35. **Podbielski, A., B. Krebs, and A. Kaufhold.** 1994. Genetic variability of the *emm*-related genes of the large *vir* regulon of group A streptococci: potential intra- and intergenomic recombination events. Mol. Gen. Genet. **243:**691–699.
36. **Raeder, R., R. A. Otten, L. Chamberlin, and M. D. P. Boyle.** 1992. Functional and serological analysis of type II immunoglobulin G-binding proteins expressed by pathogenic group A streptococci. J. Clin. Microbiol. **30:**3074–3081.
37. **Rakonjac, J. V., J. C. Robbins, and V. A. Fischetti.** 1995. DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. Infect. Immun. **63:**622–631.
38. **Retnoningrum, D. S., and P. P. Cleary.** 1994. M12 protein from *Streptococcus pyogenes* is a receptor for immunoglobulin G3 and human albumin. Infect. Immun. **62:**2387–2394.
39. **Robbins, J. C., J. G. Spanier, S. J. Jones, W. J. Simpson, and P. P. Cleary.** 1987. *Streptococcus pyogenes* type 12 M protein gene regulation by upstream sequences. J. Bacteriol. **169:**5633–5640.
40. **Rotta, J., R. M. Krause, R. C. Lancefield, W. Everly, and H. Lackland.** 1971. New approaches for the laboratory recognition of M types of group A streptococci. J. Exp. Med. **134:**1298–1315.
41. **Sanger, F., S. Niklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
42. **Schalen, C.** 1993. Prevalence of IgA receptors in clinical isolates of *Streptococcus pyogenes* and *Streptococcus agalactiae*: serologic distinction between the receptors by blocking antibodies. FEMS Immunol. Med. Microbiol. **7:**39–46.
43. **Scott, J. R., P. P. Cleary, M. G. Caparon, M. Kehoe, L. Heden, J. M. Musser, S. K. Hollingshead, and A. Podbielski.** 1995. New name for the positive regulator of the M protein of group A Streptococcus. Mol. Microbiol. **17:**799.
44. **Stenberg, L., P. O'Toole, and G. Lindahl.** 1992. Many group A streptococcal strains express two different immunoglobulin-binding proteins, encoded by closely linked genes: characterization of the proteins expressed by four strains of different M-type. Mol. Microbiol. **6:**1185–1194.
45. **Stjernquist-Desatnik, A., D. N. Kurl, C. Schalen, and P. Christensen.** 1990. Protective effect of heterologous gram-positive vaccine against lethal upper respiratory tract infection with type M50 group A streptococci in mice. Vaccine **8:**150–152.
46. **Thern, A., L. Stenberg, B. Dahlbäck, and G. Lindahl.** 1995. Ig-binding surface proteins of Streptococcus pyogenes also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. J. Immunol. **154:**375–386.
47. **Wessels, M. R., and M. S. Bronze.** 1994. Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. Proc. Natl. Acad. Sci. USA **91:**12238–12242.
48. **Whatmore, A. M., and M. A. Kehoe.** 1994. Horizontal gene transfer in the evolution of group A streptococcal *emm*-like genes: gene mosaics and variation in Vir regulons. Mol. Microbiol. **11:**363–374.
49. **Whitnack, E., J. B. Dale, and E. H. Beachey.** 1984. Common protective antigens of group A streptococcal M proteins masked by fibrinogen. J. Exp. Med. **159:**1201–1212.
50. **Wildfeuer, A., B. Heymer, W. Schachenmayr, and O. Haferkamp.** 1975. Morphological and immunological characteristics of Streptococcus pyogenes, group A, type 50. Med. Microbiol. Immunol. **161:**193–201.
51. **Wildfeuer, A., B. Heymer, H. Schafer, and O. Haferkamp.** 1978. The cervical lymph nodes in Streptococcus pyogenes, group A, type 50, infection in mice. Med. Microbiol. Immunol. **165:**129–137.

Editor: V. A. Fischetti