

Postgenomic Analysis of Four Novel Antigens of Group A *Streptococcus*: Growth Phase-Dependent Gene Transcription and Human Serologic Response

Sean D. Reid,¹ Nicole M. Green,¹ Gail L. Sylva,¹ Jovanka M. Voyich,¹ Elisha T. Stenseth,¹
Frank R. DeLeo,¹ Timothy Palzkill,² Donald E. Low,³ Harry R. Hill,⁴
and James M. Musser^{1*}

Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana¹; Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas²; Department of Microbiology, Mount Sinai Hospital and University of Toronto, Toronto, Canada³; and Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah⁴

Received 22 April 2002 /Accepted 11 June 2002

Analysis of three group A *Streptococcus* genomes (serotypes M1, M3, and M18) recently identified four previously undescribed genes that encode extracellular proteins. Each of these genes encode proteins with an LPXTG amino acid motif that covalently links many virulence factors produced by gram-positive bacteria to the cell surface. Western immunoblot analysis of serum samples obtained from 80 patients with invasive infections, noninvasive soft tissue infections, pharyngitis, and rheumatic fever indicated that these four proteins are expressed in vivo. However, the level of gene transcript and the time of maximal gene transcription varied in representative serotype M1, M3, and M18 strains. Surface expression of two proteins was confirmed by flow cytometry. Studies using a mouse infection model suggest that antibodies specific for one of the proteins (Spy0843) may contribute to a protective host immune response against a serotype M1 infection. These results are additional evidence that postgenomic strategies provide new ways to identify and investigate novel bacterial proteins that may participate in host-pathogen interactions or serve as targets for therapeutics research.

Group A *Streptococcus* (GAS) causes diverse human infections ranging from mild pharyngitis to severe disease, including toxic shock syndrome, necrotizing fasciitis, and rheumatic fever (6, 33). The pathogen is characterized by extensive allelic variation and production of many virulence factors (23, 31, 38). The incidence of GAS disease has increased since the 1980s, renewing interest in the mechanisms of pathogenesis and the development of new therapeutic agents (33). Protection against GAS infection is mediated primarily by antibodies to extracellular proteins that are secreted and freely diffusible or anchored to the bacterial cell wall (6). However, the identification of relatively few antigens that contribute to a protective immune response, coupled with genetic and serologic diversity, has limited understanding of GAS-host interactions and impeded development and licensure of a human vaccine.

Comparative genomics, proteomics, DNA microarray analysis, and other postgenomic strategies have provided many new avenues for investigating differences in pathogen phenotype, host specificity, and virulence determinants. Genes encoding proteins likely to be extracellular or displayed on the bacterial cell surface can now be identified by analysis of the genome sequence of the pathogen, facilitating rapid discovery of proteins that may interact with the host during natural infection (14, 21, 32). Analysis of genome sequence data can also assist

identification of proteins that may confer protective immunity against infection. For example, Pizza et al. (35) analyzed the genome sequence of a serogroup B *Neisseria meningitidis* strain and identified 570 open reading frames (ORFs) that were predicted to encode novel exported or surface-exposed proteins. The ORFs were cloned, and recombinant proteins were purified and used for immunologic studies. Seven proteins generated an antibody response that conferred complement-mediated bactericidal activity in a murine model of infection. Molecular population genetic analysis indicated that five of the seven proteins were conserved among 31 *N. meningitidis* strains representative of the species diversity found in natural populations. Taken together, the results have stimulated additional research into the utility of using these proteins as a meningococcal vaccine.

Analysis of the genomes of four GAS strains (serotypes M1, M3, M5, and M18) recently led to the discovery of four genes (*spy0747*, *spy0843*, *spy0872*, and *spy1972*) that encode novel extracellular proteins (Table 1) (37). The four proteins have conventional amino-terminal secretion signal sequences and have a carboxy-terminal LPXTG amino acid motif that covalently links many gram-positive bacterial virulence factors to the bacterial cell surface (13, 34, 37, 39). Sequencing and population genetic analysis of these four GAS genes in 37 strains revealed restricted allelic variation, indicating that the proteins are very well conserved in the species (37). Western immunoblot analysis conducted with acute- and convalescent-phase serum samples obtained from four patients with invasive infections indicated that all four of the recombinant proteins were reactive with one or more of the serum samples, consis-

* Corresponding author. Mailing address: Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 903 South 4th St., Hamilton, MT 59840. Phone: (406) 363-9315. Fax: (406) 363-9427. E-mail: jmusser@niaid.nih.gov.

TABLE 1. Chromosomal location and putative function of four GAS genes

Gene (<i>spy</i> no.) ^a	Chromosomal location ^b	Putative function of encoded protein ^c	LPXTG motif ^d
<i>spy0747</i>	607,542–610,274	Extracellular nuclease	LPKTG
<i>spy0843</i>	694,643–697,669	Cell surface protein	LPRTG
<i>spy0872</i>	718,489–720,501	2',3'-Cyclic-nucleotide 2'-phosphodiesterase	LPITG
<i>spy1972</i>	1,639,165–1,642,662	Pullulanase	LPKTG

^a *spy* no. refers to the corresponding numerical identifier of each ORF as listed in GenBank.

^b The chromosomal location of each gene in the sequenced GAS M1 strain ATCC 700294 (11).

^c Putative function as determined by BLAST analysis.

^d Amino acid motif used to anchor proteins to the cell wall in gram-positive organisms. Single-letter amino acid abbreviations are used (L, leucine; P, proline; K, lysine; R, arginine; I, isoleucine; T, threonine; G, glycine).

tent with the hypothesis that these proteins are produced during GAS infections (37). Taken together, these preliminary observations suggest that further analysis of these proteins is warranted.

The present study was undertaken to determine if these four proteins are expressed in multiple GAS disease types and to assess if variation exists in gene transcript level in different M types. Evidence was sought for *in vivo* expression of these proteins in infected humans by Western immunoblot analysis of sera obtained from 80 patients with invasive infections, noninvasive soft tissue infections, pharyngitis, and rheumatic fever. Inasmuch as there is considerable genetic variation among GAS strains, we analyzed the transcript level of the four genes at six time points throughout the growth cycle in three representative strains expressing serotype M1, M3, and M18 proteins. These strains were chosen for analysis because serotype M1 and M3 organisms commonly cause invasive infections worldwide, and M18 organisms have been implicated in pharyngitis and acute rheumatic fever (ARF) outbreaks in the United States (6, 33).

MATERIALS AND METHODS

Genes analyzed. The genes analyzed (*spy0747*, *spy0843*, *spy0872*, and *spy1972*) (Table 1) were identified in a recent study of the genomes of four GAS strains (serotypes M1, M3, M5, and M18) (37). Prior to that publication, none of the genes had been described.

DNA sequence analysis. Chromosomal DNA was isolated with the Puregene DNA Isolation kit (Gentra Systems). DNA sequencing primers were designed on the basis of M1, M3, and M18 genome data (11) (Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health). Sequence data obtained from both DNA strands with an Applied Biosystems 3700 automated sequencer were assembled with DNASTar (DNASTar, Inc.). Multiple-sequence alignment of the inferred amino acid sequences was performed with CLUSTALW (version 1.8) (42), and analysis of sequence polymorphisms was performed with MEGA 2.1 (<http://www.megasoftware.net/>).

Gene cloning and expression of recombinant GST fusion proteins. Cloning primers were designed on the basis of M1 genome data (11), and MGAS5005 (serotype M1) was used as the source strain for DNA. Full-length genes, minus the region encoding the putative signal secretion sequence, were cloned with the Univector plasmid fusion system (UPS) (26). The method employs bacteriophage P1 *cre-lox* site-specific recombination to catalyze *in vitro* plasmid fusion between the Univector containing the gene of interest, and a host vector (pHB2-GST) containing a GST tag (26). Briefly, each gene was cloned into the pUNI-D vector in the same frame as the *loxP* site using vaccinia virus topoisomerase I-based cloning (40). For these experiments, the pUNI-D vector adapted with topoisomerase I was obtained from Invitrogen, Inc., and PCR products of the genes of interest were inserted into the plasmid according to the manufacturer's instructions. The pUNI-D clones were converted to glutathione *S*-transferase (GST) fusions by combining 0.4 μ g of the pUNI-D clone with pHB2-GST DNA in 20 μ l of 1 \times buffer S (50 mM Tris-Cl [pH 7.5], 10 mM MgCl₂, 30 mM NaCl, and 0.1 mg of bovine serum albumin per ml) on ice. Immediately after addition

of 0.1 to 0.2 μ g of GST-Cre enzyme, the reaction mixture was incubated at 37°C for 20 min followed by 5 min at 65°C to inactivate the GST-Cre enzyme. Recombination products of the UPS were selected on Luria-Bertani agar containing 50 μ g of kanamycin per ml after transformation of *Escherichia coli* DH5 α by standard methods. Clones were sequenced to rule out the possibility of spurious mutations. To assess protein production, recombinant *E. coli* DH5 α strains were grown at 37°C in 10 ml of Luria-Bertani broth supplemented with 50 μ g of kanamycin per ml. Cultures were induced at an A_{600} of 0.5 with 0.2 to 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) and grown overnight at 25°C. Cells were pelleted by centrifugation, lysed, and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Western immunoblot analysis. A 1:20 dilution of each *E. coli* lysate containing recombinant protein was analyzed by SDS-PAGE, transferred to a nitrocellulose membrane (Millipore), and probed with patient sera. The sera studied included the following: convalescent-phase serum samples collected from 9 patients with pharyngitis, paired acute- and convalescent-phase serum samples obtained from 27 patients with invasive GAS infections, paired acute- and convalescent-phase serum samples collected from four patients with superficial skin infections, and convalescent-phase serum samples obtained from 40 patients with a history of ARF. Convalescent-phase sera were collected approximately 3 weeks postinfection. In some cases, sera obtained from patients with a history of ARF were collected several years after the last documented presentation with ARF symptoms.

Recombinant proteins were transferred with a Bio-Rad semidry transfer chamber (Bio-Rad Laboratories) for 60 min at 15 V. Following transfer, the membrane was treated with a 5% (wt/vol) solution of dehydrated milk in blocking buffer (100 mM Tris-HCl [pH 7.4] and 150 mM NaCl) for 1 h. Primary antibody (patient serum) was added to the blocking reagent, and the membrane was incubated for 1 h. The patient sera was used at a dilution of either 1:500 or 1:1,000, depending on serologic reactivity. Goat anti-human affinity-purified immunoglobulin G (IgG) (Bio-Rad) was used as the secondary antibody. Signal detection was conducted with SuperSignal West Pico chemiluminescent substrate (Pierce).

TaqMan real-time reverse transcriptase PCR analysis. Cultures of representative GAS serotype M1 (MGAS5005), serotype M3 (MGAS315), and serotype M18 (MGAS8232) strains were grown in THY medium (37) overnight at 37°C (5% CO₂). A 100- μ l aliquot of each culture was added to 50 ml of THY medium, incubated at 37°C (5% CO₂), and harvested at six time points (A_{600} = 0.05, 0.1, 0.2, 0.4, 0.6, and 0.8) throughout the growth cycle (see Fig. 2). Total RNA was isolated at each time point as previously described (4).

The primers and probes used for each gene were designed on the basis of gene regions that were identical in all three strains. TaqMan assays were performed with an ABI 7700 instrument (PE Applied Biosystems) as described by Chaussee et al. (4). Briefly, reverse transcriptase PCR was performed with the TaqMan One-Step RT-PCR Master Mix Reagents Kit (PE Applied Biosystems) as described by the manufacturer. The amplification profile used was as follows: 1 cycle at 48°C for 30 min, 1 cycle at 95°C for 10 min, 40 cycles at 95°C for 15 s, and 60°C for 1 min. The critical threshold cycle (Ct) is defined as the cycle at which fluorescence becomes detectable above background levels and is inversely proportional to the logarithm of the initial concentration of template. A standard curve was plotted for each reaction with Ct values obtained from amplification of known quantities of genomic DNA isolated from strain MGAS5005. The standard curves were used to transform Ct values of the experimental samples to the relative number of DNA molecules. The quantity of cDNA for each experimental gene was normalized to the quantity of the constitutively transcribed control gene (*gyrA*) in each sample. Specific transcript levels were expressed as fold

difference compared to *gyrA* or the fold difference between the conditions compared (4).

Purification of recombinant Spy0843 and Spy1972. Purified recombinant Spy0843 and Spy1972 were obtained with the B-PER GST Fusion Protein Purification Kit (Pierce) with slight modifications. One milliliter of resin was used for every 100 ml of bacterial culture. Benzonuclease (1 μ l) (Novagen) was added for every ml of B-PER Reagent used to resuspend the cell pellet, and phenylmethylsulfonyl fluoride was added to a final concentration of 25 mM.

Affinity-purified anti-Spy0843 and anti-Spy1972 antibodies. Purified recombinant Spy0843 and Spy1972 were supplied to Bethyl Laboratories for producing purified antibodies. Rabbits were immunized or given booster doses of 100 μ g of antigen every 2 weeks for 2 months and then immunized or given booster doses once a month. The first immunization contained antigen and complete Freund's adjuvant at a 1:1 ratio. All subsequent immunizations contained antigen and incomplete Freund's adjuvant at a 1:1 ratio. The animals were bled every 2 weeks after the third immunization (5 weeks). Hyperimmune serum samples from rabbits were processed over a GST column to absorb anti-GST antibodies. The serum was then passed over an immunosorbent (Spy0843 or Spy1972 peptide linked to agarose using a cyanogen bromide method) to capture antibodies specific for the protein. Antibodies were filtered in a sterile manner and supplied in a neutral buffer with sodium azide (0.1%) as an antimicrobial agent.

Expression of Spy0843 and Spy1972 on the GAS cell surface. Surface expression of Spy0843 and Spy1972 was analyzed with a FACScaliber flow cytometer (Becton Dickinson) using the affinity-purified Spy0843- and Spy1972-specific antibodies. Purified rabbit IgG raised against an irrelevant protein antigen was used as a control for nonspecific antibody binding. Briefly, GAS strain MGAS5005 (serotype M1) was grown to late exponential phase ($A_{600} = 0.7$) in THY medium, harvested by centrifugation, washed twice in Dulbecco's phosphate-buffered saline (DPBS) (pH 7.2), and resuspended in DPBS at 10^8 CFU/ml. Anti-Spy1972 and anti-Spy0843 antibodies were added to 100 μ l of bacterial suspension at a final concentration of 1 μ g/100 μ l and incubated for 30 min on ice. Samples were washed with DPBS containing 1% goat serum and stained with phycoerythrin-conjugated donkey anti-rabbit IgG (Jackson ImmunoResearch) (1:400 dilution) for 30 min on ice prior to flow cytometry.

Active immunization of mice with purified Spy0843 and Spy1972. Four groups of 25 female Swiss CD-1 outbred mice were inoculated with either PBS or 5, 10, or 20 μ g of Spy0843 or Spy1972 in PBS subcutaneously on day 1. Booster doses of the same treatments were administered on days 14 and 28. The adjuvant monophosphoryl lipid A (Corixa, Hamilton, Mont.) was added to all doses. Mice were challenged intraperitoneally with 1.82×10^8 CFU (0.2 ml) of strain MGAS5005 (serotype M1) on day 35. Bacteria were harvested when maximal transcription of the gene encoding the protein of interest was reached as indicated by TaqMan assays (see Results). The number of CFU inoculated per mouse was verified by colony counts. Mice were monitored every 3 h for 7 days, mortality was recorded, and Kaplan-Meier survival curves were plotted (JMP statistical software). The Wilcoxon test was used to test for statistical differences in survival.

RESULTS

Allelic variation and sequence polymorphism. All four genes (*spy0747*, *spy0843*, *spy0872*, and *spy1972*) were sequenced in a serotype M1, M3, and M18 strain (MGAS5005, MGAS315, and MGAS8232, respectively). Each strain had a unique allele of each gene. The *spy1972* gene had the highest level of nucleotide divergence, with an average of 26.1 nucleotide substitutions per 3,495 bp. Variation at the amino acid level was low among the four proteins, ranging from an average of 5.1 amino acid replacements per 910 sites in Spy0747 to an average of 11.6 amino acid replacements per 1,165 sites in Spy1972. Overall, the level of polymorphism in each gene sequence was equivalent to the levels identified in seven GAS housekeeping genes (9), suggesting that selective pressure is acting to maintain protein structure and function. This is in contrast to many GAS virulence factors that have increased levels of polymorphism (e.g., *emm*, *sic*, *scl*, and *speB*) likely in response to selective pressure from the host immune system (38). The *spy0747*, *spy0843*, and *spy0872* genes had an average adenosine

thymidine (AT) content of 63%, a level consistent with a genome-wide calculation of GAS AT% (61.5%) (11). In contrast, the AT content of *spy1972* was lower (56.9%).

Reactivity of human sera with recombinant proteins. A recent study suggested that the four proteins investigated here were produced in humans infected with GAS (37). The preliminary study was limited to the Western immunoblot analysis of sera collected from seven healthy individuals and four patients with invasive GAS infections. Hence, it is not known if antibodies specific for each protein are generated in response to several distinct infection types. To investigate this issue, Western immunoblot analysis was conducted with paired acute- and convalescent-phase serum samples collected from 27 patients with invasive GAS infections (culture positive from a normally sterile site) and from 4 patients with superficial skin infections (noninvasive) (Table 2). The infecting strains represented 19 distinct M protein serotypes (Table 2).

Spy0747 (putative extracellular nuclease), Spy0843 (putative cell surface protein), and Spy1972 (putative pullulanase) reacted with convalescent-phase sera from 81% of patients with invasive infections and all convalescent-phase sera obtained from patients with superficial skin infections. In contrast, Spy0872 (putative 2',3'-cyclic-nucleotide 2'-phosphodiesterase) was reactive with only 41% of convalescent-phase sera collected from patients with invasive infections and 50% of sera from patients with superficial infections. Recombinant Spy0872 was produced at the same level as that of the other four recombinant proteins in *E. coli*, indicating that the relatively low reactivity was not due to lack of recombinant protein synthesis (data not shown).

Each of the proteins reacted with $\leq 19.4\%$ of the acute-phase serum samples ($n = 31$) isolated from patients with invasive disease (data not shown). This is not unexpected, given the variation in the length of time from the onset of infection to the time the patient seeks treatment. Importantly, when reactivity with acute-phase sera was observed, the reactivity of the convalescent-phase serum sample was far more intense, a result consistent with recent exposure to a given protein (Fig. 1).

Inasmuch as the complement of genes expressed during GAS pharyngitis or infections leading to ARF may differ from invasive infections, Western immunoblot analysis of convalescent-phase sera obtained from 9 patients with recent pharyngitis and 40 individuals with a history of ARF was conducted (Table 3). The serologic reactivity among these 49 patients was closely similar to that for the 31 subjects with invasive or superficial skin infections (Tables 2 and 3). Spy0747, Spy0843, and Spy1972 reacted with 89, 78, and 78%, respectively, of sera obtained after episodes of pharyngitis. Similarly, Spy0747, Spy0843, and Spy1972 reacted with the sera obtained from $\geq 85\%$ of individuals with ARF. Spy0872 was far less reactive ($< 25\%$ with sera from individuals with pharyngitis or ARF) (Table 3).

TaqMan analysis of gene transcription. The variation in serologic reactivity led us to investigate if strain differences existed in the level of transcription of each gene. Inasmuch as considerable allelic and chromosomal variation exists in natural populations of GAS (38), strain variation in gene transcription may contribute to the altered levels of serologic response to some of the proteins investigated. Representative strains of

TABLE 2. Reactivity of four recombinant GAS proteins with human convalescent sera obtained from patients with GAS disease

Patient	emm sequence type ^a	Disease type ^b	Reactivity ^c			
			Spy0747	Spy0843	Spy0872	Spy1972
1	emm1	Invasive	+	+	+	+
2	emm1	Invasive	+	+	+	+
3	emm1	Invasive	+	+	+	+
4	emm3.2	Invasive	+	+	+	+
5	emm5.8193b	Invasive	+	+		+
6	emm11	Invasive	+	+	+	
7	emm13w	Invasive	+	+	+	
8	emm13w	Invasive				+
9	emm13w	Invasive	+			+
10	emm22.2	Invasive	+	+		+
11	emm28	Invasive	+	+		+
12	emm28	Invasive	+	+		+
13	emm28	Invasive	+	+		+
14	emm28	Invasive	+	+	+	+
15	emm28	Invasive	+	+		+
16	emm28	Invasive	+			+
17	emm36	Invasive	+		+	+
18	emm58	Invasive	+	+		
19	emm58	Invasive	+	+		+
20	emm75	Invasive	+	+	+	+
21	emm89	Invasive	+	+		+
22	emm89	Invasive	+			+
23	stns14x.2	Invasive	+	+		+
24	St833.1	Invasive	+	+	+	
25	St2917	Invasive		+		+
26	St2967	Invasive	+	+	+	
27	St3757	Invasive				+
28	emm12	Noninvasive	+	+		+
29	emm44/emm61	Noninvasive	+	+	+	+
30	emm75	Noninvasive	+	+	+	+
31	St6735	Noninvasive	+	+		+

^a Nucleotide sequence of the emm allele in the infecting GAS strain. Alleles were assigned on the basis of reference alleles (9).

^b An infection was classified as invasive on the basis of a positive culture of GAS from a normally sterile site. Each of the noninvasive cases was a superficial skin infection.

^c Reactivity of four recombinant GAS proteins (Spy0747, Spy0843, Spy0872, and Spy1972) with human convalescent sera from patients with GAS disease. +, reactive. The percentages of serum samples reactive with the recombinant GAS proteins were 87, 81, 42, and 84 for Spy0747, Spy0843, Spy0872, and Spy1972, respectively.

serotype M1 (MGAS5005), M3 (MGAS315), and M18 (MGAS8232) were studied. Genetic population analysis has indicated that serotype M1 and M3 strains are the most common causes of invasive GAS infections worldwide in most case studies (33), and M18 organisms have been implicated in many pharyngitis episodes and outbreaks of ARF in the United States (33). Equal amounts of total RNA isolated at six time points throughout the growth curve ($A_{600} = 0.05, 0.1, 0.2, 0.4, 0.6,$ and 0.8) were used. Two important observations were made. First, the analysis discovered that the time of maximal gene transcript level differed among the three strains studied (Fig. 2). For example, the transcript level of the M18 allele of *spy0747* was highest in the mid-exponential phase of growth ($A_{600} = 0.4$), whereas the transcript level of the M1 and the M3 allele was highest in early stationary phase ($A_{600} = 0.8$) (Fig. 2).

Second, the level of gene transcript also varied among the three strains. For example, the transcript level of *spy0843* in the M3 strain was higher than the M1 or M18 strain at all but the earliest time point ($A_{600} = 0.05$). Similarly, the transcript level of *spy0872* was very similar at each time point until stationary phase ($A_{600} = 0.8$), at which point the transcript level in the M1 strain increased > 2.5 -fold and the transcript levels in the M3 and M18 strains decreased (Fig. 2). Transcript levels of all

three alleles of *spy1972* were low until $A_{600} = 0.6$ when the transcript level increased ≥ 35 -fold (Fig. 2). Interestingly, the transcript level of *spy1972* in the M1 strain remained greatly elevated at $A_{600} = 0.8$, whereas the transcript level of the alleles in the M3 and M18 strains decreased to near-initial levels.

Purification of recombinant Spy0843 and Spy1972. Spy0843 and Spy1972 were overexpressed in soluble form in *E. coli* BL21(DE3) and purified to apparent homogeneity (Fig. 3). Amino-terminal amino acid sequencing confirmed the identity of each protein (data not shown). We were unable to purify Spy0747 because the recombinant protein was insoluble. The subsequent purification of Spy0872 was not pursued due to the lack of seroreactivity with the majority of serum samples examined and the apparent low level of protein expression throughout growth in three GAS serotypes.

In vivo production of Spy0843 and Spy1972 by GAS. The levels of anti-Spy0843 and anti-Spy1972 antibodies in acute- and convalescent-phase sera recently isolated from 10 patients with pharyngitis were measured by enzyme-linked immunosorbent assay. Nine of 10 patients had significantly higher levels of anti-Spy0843 antibodies in convalescent-phase sera than in acute-phase sera ($P = 0.003$) (data not shown). Similarly, 7 of 10 patients had significantly higher levels of anti-Spy1972 an-

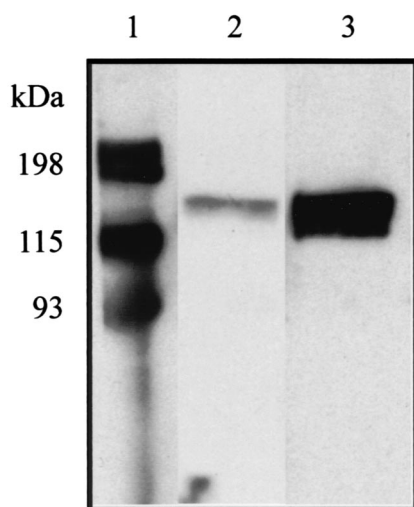


FIG. 1. Western immunoblot showing typical reactivity of paired acute- and convalescent-phase serum samples with purified recombinant proteins. Acute-phase serum was collected when a patient first sought treatment. Convalescent-phase serum was collected approximately 3 weeks postinfection. The immunoblot is representative of results obtained for each of the four group A *Streptococcus* proteins (reactivity with Spy1972 is shown). A 1:20 dilution of *E. coli* lysate containing recombinant Spy1972 and a 1:500 dilution of patient serum was used. Lane 1, molecular mass markers; lane 2, reactive acute-phase serum sample; lane 3, reactive, convalescent-phase serum. The increased reactivity with the convalescent-phase serum sample is consistent with recent exposure to Spy1972.

tibodies in convalescent-phase sera than in acute-phase sera ($P = 0.036$) (data not shown). These results confirm the Western immunoblot analysis that Spy0843 and Spy1972 are produced in humans infected with GAS.

Detection of Spy0843 and Spy1972 on the surface of GAS.

The presence of Spy0843 and Spy1972 on the surface of live GAS was confirmed by flow cytometry (Fig. 4). The ability of purified anti-Spy0843 and anti-Spy1972 antibodies to bind to GAS was demonstrated by increased fluorescence compared to that with control antibody (Fig. 4). Each antibody recognized its respective antigen on the surface of GAS, as demonstrated by a substantial shift in fluorescence over control antibody (Fig. 4). Western immunoblot analysis confirmed that the respective antibodies were not cross-reactive (data not shown).

Active immunization of mice with purified Spy0843 and Spy1972. Immunization with 20 μg of purified Spy0843 conferred significant protection ($P = 0.0387$ by the Wilcoxon test) against lethal challenge with the virulent strain MGAS5005 with twice as many mice surviving until the 120-h endpoint

(compared to nonimmunized mice). Mice immunized with 20 μg of purified Spy0843 had a mean survival time of 41.2 ± 8.8 h compared to nonimmunized mice, which had a mean survival time of 32.8 ± 6.5 h. Lower dosages of Spy0843 did not confer protection. Immunization with up to 20 μg of Spy1972 did not confer protection against lethal challenge.

DISCUSSION

Abundant chromosomal, allelic, and serologic diversity in GAS has hindered vaccine development, understanding of pathogen-host interactions, and differences in strain behavior. For more than 40 years, it has been known that patients with GAS infections produce antibodies to a large number of extracellular proteins (16–20, 24), but most of these antigens have not been identified. Availability of genome sequences has permitted virulence factors and therapeutic agent candidates to be identified very rapidly by comparative genomics and other postgenomic methods (2, 3, 14, 21, 24, 32, 35, 37). The studies reported here were stimulated by a recent analysis of four genomes of GAS which led to the discovery of four previously undescribed genes encoding extracellular proteins (37). Three observations suggested that further investigation of the four proteins was warranted. First, none of the proteins was hypervariable in the 37 strains studied (12 M serotypes), and few had long regions of increased amino acid variation. Thus, if one or more of these proteins is able to stimulate production of protective antibodies, they are likely to react with the cognate proteins produced by a broad range of common GAS serotypes. Second, all four of the genes were transcribed and preliminary Western immunoblot results suggested that the proteins are produced in infected hosts. Third, several of the proteins are homologous to virulence factors produced by other bacteria, suggesting that inhibition of these molecules may be therapeutic (37).

A common strategy used to identify proteins for vaccine-related research is to select antigens on the basis of immunologic data obtained from patients with diverse infection types (14). One important limitation of our initial study (37) was that only four serum samples obtained from patients with invasive disease were examined. Inasmuch as the complement of proteins made during various types of GAS infections (pharyngitis, cellulitis, necrotizing fasciitis, etc.) may differ, we sought evidence that each protein was produced in many patients with multiple infection types. Western immunoblot analysis confirmed that all four proteins (Spy0747, Spy0843, Spy0872, and Spy1972) are made during the course of distinct GAS infections. This result suggests that if antibodies against one or more of the proteins contribute to protection against infection, protection may extend to different GAS infection types.

The frequency of reactivity of Spy0872 with sera isolated from human patients was considerably less than those of the other three proteins. However, all serum samples obtained from patients infected by a serotype M1 organism were reactive with recombinant Spy0872 (Table 2). Consistent with the serologic data, only the serotype M1 strain had an increase in the level of *spy0872* transcript throughout the growth cycle (Fig. 2). Thus, the low level of serologic reactivity may be due to relatively minimal transcription of the *spy0872* gene by non-M1 serotypes. This is an important new finding. Spy0872

TABLE 3. Immunoreactivity of recombinant GAS proteins with human sera obtained from patients with GAS infections

Patient	% Immunoreactivity ^a			
	Spy0747	Spy0843	Spy0872	Spy1972
Pharyngitis ($n = 9$)	89	78	22	78
ARF ($n = 40$)	85	90	15	88

^a Immunoreactivity of recombinant GAS proteins Spy0747, Spy0843, Spy0872, and Spy1972 with convalescent-phase sera isolated from patients with recent episodes of GAS pharyngitis or with a history of ARF.

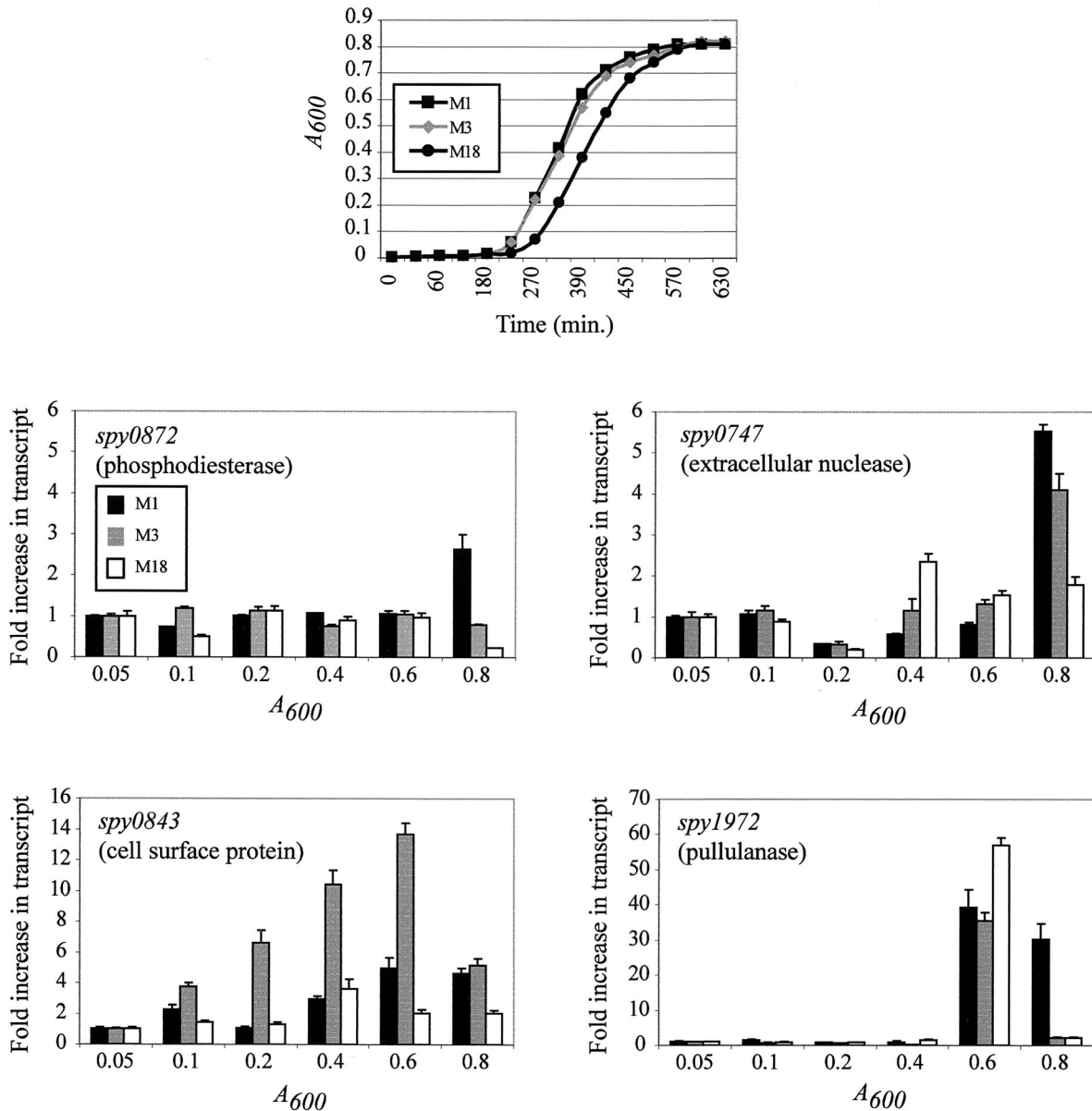


FIG. 2. Fold increase in gene transcript levels assessed by TaqMan assays at six time points throughout the growth cycle. Cultures of representative group A *Streptococcus* serotype M1 (MGAS5005), serotype M3 (MGAS315), and serotype M18 (MGAS8232) strains grown overnight were diluted 1:500 in 50 ml of THY medium, incubated at 37°C (5% CO₂), and harvested at six time points (A_{600} = 0.05, 0.1, 0.2, 0.4, 0.6, and 0.8) throughout the growth cycle. cDNA prepared from strains of serotype M1 (MGAS5005), M3 (MGAS315), and M18 (MGAS8232) was measured for transcripts of *spy0747* (putative extracellular nuclease), *spy0843* (putative cell surface protein), *spy0872* (putative 2',3'-cyclic-nucleotide 2'-phosphodiesterase), and *spy1972* (putative pullulanase). All measurements were normalized to the *gyrA* transcript as described in Materials and Methods. Values are expressed as the fold increase in transcript compared to the transcript level at an A_{600} of 0.05. The data represent values obtained with at least two independently isolated RNA samples analyzed in triplicate.

may not be suitable for further vaccine-related studies due to a lack of seroreactivity with the majority of serum samples examined and the apparent low level of protein expression throughout growth in three GAS serotypes (Tables 2 and 3 and Fig. 2).

For unknown reasons, serotype M1 organisms cause many more GAS infections than the other serotypes do (33). One observation of note is that only serum samples obtained from

patients infected with a serotype M1 organism consistently had antibodies to all four proteins. Very low levels of amino acid variation in all four proteins argue against the possibility that the absence of serologic reactivity with these proteins was caused by strain-specific variation in amino acid sequence. In contrast, TaqMan assay results clearly indicated that the levels of transcripts produced can vary by serotype. It is possible that decreased levels of transcription by some strains led to very low

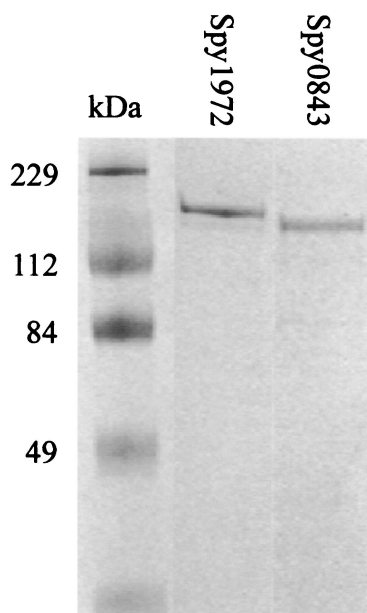


FIG. 3. SDS-PAGE showing purified recombinant Spy1972 and Spy0843 (recombinant Spy1972 and Spy0843 are GST fusion proteins). The gel was stained with Coomassie brilliant blue. Lane 1, molecular mass marker; lane 2, purified recombinant Spy1972; lane 3, purified recombinant Spy0843.

levels of protein synthesis and antibody production, and hence, lack of reactivity in the Western immunoblot analysis. Perhaps M1 organisms are adept at causing infection due to higher levels of gene transcription and virulence factor production, on average, than those of many other GAS strains.

TaqMan analysis has proved to be a very rapid method of gaining insight into GAS gene regulation (4, 37, 41). Analysis of isogenic mutant strains by TaqMan assays (37) has implicated the GAS virulence gene regulators *covR* (negative regulator of several GAS virulence factors, including cysteine protease and capsule) (10, 25) and *mga* (positive regulator of M protein, C5a peptidase, streptococcal inhibitor of complement,

and streptococcal collagen-like protein 1) (27, 29, 30) in the control of expression of several of the genes studied here. For example, transcription of *spy0747* and *spy0843* was up regulated in the absence of the CovR repressor, whereas expression of *spy1972* was down regulated in the absence of the positive regulator encoded by *mga* (37). In contrast, a recent microarray and TaqMan analysis of differential gene expression in response to alteration in growth temperature did not implicate temperature in the regulation of these four genes (41). It will be important in future studies to analyze the pattern of gene transcription in experimentally infected animals and human patients with naturally acquired disease.

Several recent studies have suggested that postgenomic strategies to triage the analysis of protein vaccine candidates identified by comparative genomics could benefit by taking into consideration the degree of amino acid conservation (1, 35, 37, 38, 43). Use of an antigen that is conserved among all or most strains of a species is likely to reduce the possibility of variability in serologic reactivity, and subsequent immunologic protection, due to strain-specific differences in protein sequence. The same line of reasoning applies to gene transcription. Our studies indicate that the maximal transcript level and the time of maximal expression differed among the four genes for each of the three strains (serotypes M1, M3, and M18) examined. This implies that although a protein may be conserved and capable of interacting with the host during infection, variation in the level and timing of gene expression and subsequent protein synthesis among strains may fundamentally alter the role of the protein in host-pathogen interactions. An integrated strategy for identifying new vaccine candidates that incorporates analysis of variation in gene transcription (TaqMan analysis) and amino acid composition (comparative sequencing) may assist identification of constitutively expressed genes which encode highly conserved protein antigens that stimulate a relatively consistent immunologic response in the host. Moreover, this approach would likely identify strains that have significant variation in gene transcription or amino acid composition of the antigen. These strains could be very useful for heterologous challenge studies.

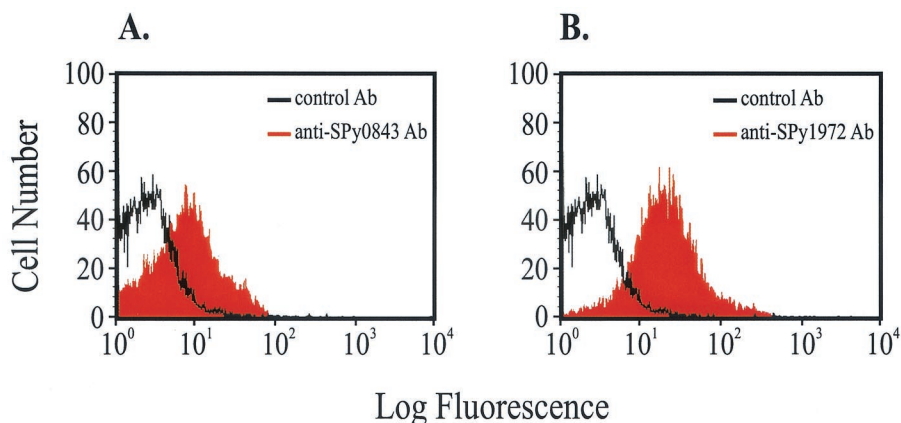


FIG. 4. Detection of Spy0843 (A) and Spy1972 (B) on the GAS cell surface by flow cytometry. Control antibody (unshaded histograms) and purified anti-Spy0843 and anti-Spy1972 (shaded histograms) were incubated with strain MGAS5005 (serotype M1) and analyzed by flow cytometry. The unshaded histograms represent the level of nonspecific binding of the control antibody to MGAS5005. The shaded histograms indicate increased fluorescence due to the binding of anti-Spy0843 or anti-Spy1972 antibody to the MGAS5005 cell surface.

Each of the genes studied encoded a protein with a carboxy-terminal LPXTG cell wall anchor motif. More than 50 extracellular proteins with this motif in gram-positive bacteria have been described, and many of these proteins are virulence factors (12). For example, in GAS this amino acid motif is present in M protein, M-like proteins, C5a peptidase, GRAB protein, serum opacity factor, a fibronectin-binding protein, streptococcal protective antigen, and two collagen-like proteins (5, 8, 12, 27, 28, 36). All of these proteins are known to be accessible on the GAS cell surface. We confirmed that the LPXTG motif-containing proteins Spy0843 and Spy1972 are also displayed on the GAS cell surface. Importantly, all five GAS cell surface proteins with an LPXTG anchor motif or closely related amino acid sequence that have been extensively studied (M protein, C5a peptidase, serum opacity factor, a fibronectin-binding protein, and streptococcal protective antigen) are virulence factors and contribute to protective immunity in mouse models (7, 8, 15, 22). Here we show that immunization with purified recombinant Spy0843 can confer protection against intraperitoneal challenge in a mouse model of GAS infection. This suggests that antibodies specific for Spy0843 may contribute to a protective host immune response. Additional work will be needed to investigate this issue in other models of animal infection.

In summary, our studies provide additional evidence that new therapeutic agent candidates can be identified and characterized rapidly by postgenomic strategies. The results also highlight the advantages of integrating TaqMan transcription analysis with Western immunoblot results to determine if a protein is made in infected hosts and transcribed in strains that express different M-protein serotypes. Analyses of the roles of these four proteins in host-pathogen interactions and of their ability to stimulate protective immunity in animal models may provide new avenues for control of GAS disease.

ACKNOWLEDGMENTS

We thank Robert L. Cole, Benfang Lei, and Robin Ireland for technical assistance and Nancy P. Hoe for critical reading of the manuscript.

REFERENCES

- Bolduc, G. R., V. Bouchet, R. Z. Jiang, J. Geisselsoder, Q. C. Truong-Bolduc, P. A. Rice, S. I. Pelton, and R. Goldstein. 2000. Variability of outer membrane protein P1 and its evaluation as a vaccine candidate against experimental otitis media due to nontypeable *Haemophilus influenzae*: an unambiguous, multifaceted approach. *Infect. Immun.* **68**:4505–4517.
- Chakravarti, D. N., M. J. Fiske, L. D. Fletcher, and R. J. Zagursky. 2000. Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* **19**:601–612.
- Chakravarti, D. N., M. J. Fiske, L. D. Fletcher, and R. J. Zagursky. 2000. Mining genomes and mapping proteomes: identification and characterization of protein subunit vaccines. *Dev. Biol.* **103**:81–90.
- Chaussee, M. S., R. O. Watson, J. C. Smoot, and J. M. Musser. 2001. Identification of Rgg-regulated exoproteins of *Streptococcus pyogenes*. *Infect. Immun.* **69**:822–831.
- Courtney, H. S., D. L. Hasty, Y. Li, H. C. Chiang, J. L. Thacker, and J. B. Dale. 1999. Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of M type 2 *Streptococcus pyogenes*. *Mol. Microbiol.* **32**:89–98.
- Cunningham, M. W. 2000. Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.* **13**:470–511.
- Dale, J. B. 1999. Group A streptococcal vaccines. *Infect. Dis. Clin. N. Am.* **13**:227–243.
- Dale, J. B., E. Y. Chiang, S. Liu, H. S. Courtney, and D. L. Hasty. 1999. New protective antigen of group A streptococci. *J. Clin. Invest.* **103**:1261–1268.
- Enright, M. C., B. G. Spratt, A. Kalia, J. H. Cross, and D. E. Bessen. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. *Infect. Immun.* **69**:2416–2427.
- Federle, M. J., K. S. McIver, and J. R. Scott. 1999. A response regulator that represses transcription of several virulence operons in the group A *Streptococcus*. *J. Bacteriol.* **181**:3649–3657.
- Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, G. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najjar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe, and R. McLaughlin. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. USA* **98**:4658–4663.
- Fischetti, V. A. 2000. Surface proteins on gram-positive bacteria, p. 11–24. *In* V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (ed.), *Gram-positive pathogens*. ASM Press, Washington, D.C.
- Fischetti, V. A., V. Pancholi, and O. Schneewind. 1990. Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci. *Mol. Microbiol.* **4**:1603–1605.
- Grandi, G. 2001. Antibacterial vaccine design using genomics and proteomics. *Trends Biotechnol.* **19**:181–188.
- Guzman, C. A., S. R. Talay, G. Molinari, E. Medina, and G. S. Chhatwal. 1999. Protective immune response against *Streptococcus pyogenes* in mice after intranasal vaccination with the fibronectin-binding protein SfbI. *J. Infect. Dis.* **179**:901–906.
- Halbert, S. P. 1958. The use of precipitin analysis in agar for the study of human streptococcal infections. III. The purification of some of the antigens detected by these methods. *J. Exp. Med.* **108**:385–410.
- Halbert, S. P., and T. Auerbach. 1960. The use of precipitin analysis in agar for the study of human streptococcal infections. IV. Further observations on the purification of group A extracellular antigens. *J. Exp. Med.* **113**:131–157.
- Halbert, S. P., and S. L. Keatinge. 1961. The analysis of streptococcal infections. VI. Immunoelectrophoretic observations on extracellular antigens detectable with human antibodies. *J. Exp. Med.* **113**:1013–1028.
- Halbert, S. P., L. Swick, and C. Sonn. 1955. The use of precipitin analysis in agar for the study of human streptococcal infections. I. Oudin technique. *J. Exp. Med.* **101**:539–556.
- Halbert, S. P., L. Swick, and C. Sonn. 1955. The use of precipitin analysis in agar for the study of human streptococcal infections. II. Ouchterlony and Oakley techniques. *J. Exp. Med.* **101**:557–575.
- Janulczyk, R., and M. Rasmussen. 2001. Improved pattern for genome-based screening identifies novel cell wall-attached proteins in gram-positive bacteria. *Infect. Immun.* **69**:4019–4026.
- Ji, Y., B. Carlson, A. Kondagunta, and P. P. Cleary. 1997. Intranasal immunization with C5a peptidase prevents nasopharyngeal colonization of mice by the group A *Streptococcus*. *Infect. Immun.* **65**:2080–2087.
- Kehoe, M. A., V. Kapur, A. M. Whatmore, and J. M. Musser. 1996. Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology. *Trends Microbiol.* **4**:436–443.
- Lei, B., S. Mackie, S. Lukomski, and J. M. Musser. 2000. Identification and immunogenicity of group A *Streptococcus* culture supernatant proteins. *Infect. Immun.* **68**:6807–6818.
- Levin, J. C., and M. R. Wessels. 1998. Identification of *csrR/csrS*, a genetic locus that regulates hyaluronic acid capsule synthesis in group A *Streptococcus*. *Mol. Microbiol.* **30**:209–219.
- Liu, Q., M. Z. Li, D. Leibham, D. Cortez, and S. J. Elledge. 1998. The Univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**:1300–1309.
- Lukomski, S., K. Nakashima, I. Abdi, V. J. Cipriano, R. M. Ireland, S. D. Reid, G. G. Adams, and J. M. Musser. 2000. Identification and characterization of the *scl* gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. *Infect. Immun.* **68**:6542–6553.
- Lukomski, S., K. Nakashima, I. Abdi, V. J. Cipriano, B. J. Shelvin, E. A. Graviss, and J. M. Musser. 2001. Identification and characterization of a second extracellular collagen-like protein made by group A *Streptococcus*: control of production at the level of translation. *Infect. Immun.* **69**:1729–1738.
- McIver, K. S., and J. R. Scott. 1997. Role of *mga* in growth phase regulation of virulence genes of the group A *Streptococcus*. *J. Bacteriol.* **179**:5178–5187.
- McIver, K. S., A. S. Thurman, and J. R. Scott. 1999. Regulation of *mga* transcription in the group A *Streptococcus*: specific binding of *mga* within its own promoter and evidence for a negative regulator. *J. Bacteriol.* **181**:5373–5383.
- Musser, J. M., A. R. Hauser, M. H. Kim, P. M. Schlievert, K. Nelson, and R. K. Selander. 1991. *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. *Proc. Natl. Acad. Sci. USA* **88**:2668–2672.
- Musser, J. M., and S. L. Kaplan. 2001. *Pneumococcus* research transformed. *N. Engl. J. Med.* **345**:1206–1207.
- Musser, J. M., and R. M. Krause. 1998. The revival of group A streptococcal diseases, with a commentary on staphylococcal toxic shock syndrome, p. 185–218. *In* R. M. Krause (ed.), *Emerging infections*. Academic Press, New York, N.Y.
- Navarre, W. W., and O. Schneewind. 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.* **63**:174–229.

35. **Pizza, M., V. Scarlato, V. Masignani, M. M. Giuliani, B. Arico, M. Comanducci, G. T. Jennings, L. Baldi, E. Bartolini, B. Capecchi, C. L. Galeotti, E. Luzzi, R. Manetti, E. Marchetti, M. Mora, S. Nuti, G. Ratti, L. Santini, S. Savino, M. Scarselli, E. Storni, P. Zuo, M. Broecker, E. Hundt, B. Knapp, E. Blair, T. Mason, H. Tettelin, D. W. Hood, A. C. Jeffries, N. J. Saunders, D. M. Granoff, J. C. Venter, E. R. Moxon, G. Grandi, and R. Rappuoli.** 2000. Identification of vaccine candidates against serogroup B *Meningococcus* by whole-genome sequencing. *Science* **287**:1816–1820.
36. **Rasmussen, M., H. P. Muller, and L. Bjorck.** 1999. Protein GRAB of *Streptococcus pyogenes* regulates proteolysis at the bacterial surface by binding α 2-macroglobulin. *J. Biol. Chem.* **274**:15336–15344.
37. **Reid, S. D., N. M. Green, J. K. Buss, B. Lei, and J. M. Musser.** 2001. Multilocus analysis of extracellular putative virulence proteins made by group A *Streptococcus*: population genetics, human serologic response, and gene transcription. *Proc. Natl. Acad. Sci. USA* **98**:7552–7557.
38. **Reid, S. D., N. P. Hoe, L. M. Smoot, and J. M. Musser.** 2001. Group A *Streptococcus*: allelic variation, population genetics, and host-pathogen interactions. *J. Clin. Investig.* **107**:393–399.
39. **Schneewind, O., P. Model, and V. A. Fischetti.** 1992. Sorting of protein A to the staphylococcal cell wall. *Cell* **70**:267–281.
40. **Shuman, S.** 1994. Novel approach to molecular cloning and polynucleotide synthesis using *Vaccinia* DNA topoisomerase. *J. Biol. Chem.* **269**:32678–32684.
41. **Smoot, L. M., J. C. Smoot, M. R. Graham, G. A. Somerville, D. E. Sturdevant, C. A. Migliaccio, G. L. Sylva, and J. M. Musser.** 2001. Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. *Proc. Natl. Acad. Sci. USA* **98**:10416–10421.
42. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
43. **Wizemann, T. M., J. H. Heinrichs, J. E. Adamou, A. L. Erwin, C. Kunsch, G. H. Choi, S. C. Barash, C. A. Rosen, H. R. Masure, E. Tuomanen, A. Gayle, Y. A. Brewah, W. Walsh, P. Barren, R. Lathigra, M. Hanson, S. Langermann, S. Johnson, and S. Koenig.** 2001. Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect. Immun.* **69**:1593–1598.