Molecular Analysis of Group A Streptococcus Type emm18 Isolates Temporally Associated with Acute Rheumatic Fever Outbreaks in Salt Lake City, Utah

James C. Smoot,1 E. Kent Korgenski,2 Judy A. Daly,2 L. George Veasy,2 and James M. Musser1*

Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840,1 and Primary Children’s Medical Center, Salt Lake City, Utah 841232

Received 17 October 2001/Returned for modification 3 December 2001/Accepted 10 January 2002

Acute rheumatic fever (ARF) and subsequent rheumatic heart disease are rare but serious sequelae of group A Streptococcus (GAS) infections in most western countries. Salt Lake City (SLC), Utah, and the surrounding intermountain region experienced a resurgence of ARF in 1985 which has persisted. The largest numbers of cases were encountered in 1985-1986 and in 1997-1998. Organisms with a mucoid colony phenotype when grown on blood agar plates were temporally associated with the higher incidence of ARF. To develop an understanding of the molecular population genetic structure of GAS strains associated with ARF in the SLC region, 964 mucoid and nonmucoid pharyngeal isolates recovered in SLC from 1984 to 1999 were studied by sequencing the emm gene. Isolates with an emm18 allele were further characterized by sequencing the spa, covR, and covS genes. Peak periods of ARF were associated with GAS isolates possessing an emm18 allele encoding the protein found in serotype M18 isolates. Among the serotype M18 isolates, the difference in the number of C repeats produced three size variants. Variation was limited in spa, a gene that encodes a streptococcal protective antigen, and covR and covS, genes that encode a two-component regulatory system that, when inactivated, results in a mucoid phenotype and enhanced virulence in mouse infection models. Pulsed-field gel electrophoresis showed a single restriction profile for serotype M18 organisms isolated during both peak periods of ARF. In SLC, the incidence of ARF coresured with the occurrence of GAS serotype M18 isolates that have very restricted genetic variation.

Acute rheumatic fever (ARF) is the leading cause of preventable pediatric heart disease globally. This devastating disease usually occurs in school-age children or young adults 2 to 3 weeks after pharyngeal infection with group A Streptococcus (GAS). Treatment of GAS pharyngitis with penicillin and related antimicrobial agents has produced a marked decline in the incidence of ARF in the United States and Europe since the 1960s. Despite this pronounced decrease, recent ARF outbreaks have occurred in distinct geographic areas and populations such as trainees in the armed forces (25). In addition, the incidence rate in developing countries remains high. For example, a recent analysis found that the lifetime cumulative incidence of ARF ranged from 2.8 to 5.7% in Australian Aboriginal communities (4). The importance of ARF and subsequent rheumatic heart disease as a global health problem and its ability to resurge in the United States for unknown reasons stress the need for increased study of the GAS strains and virulence factors that contribute to the pathogenesis of ARF in susceptible hosts.

GAS isolates have historically been characterized by the serologic reactivity of M protein, a major surface antigen that is antiphagocytic and that is an important protective antigen for some strains. M protein is composed of a hypervariable amino terminus, a central region with repeat regions, and a conserved carboxy terminus that is linked to the GAS cell membrane. The amino terminus consists of roughly 30 to 100 amino acid residues that give rise to more than 100 distinct M-protein serotypes (9, 13). The central region of M protein comprises amino acid repeat regions that vary in number and composition within and among M serotypes (13). Accumulated evidence suggests that M protein is associated with the pathogenesis of ARF as a consequence of immune responses directed against the repeat regions of M protein that cross-react with human proteins such as cardiac myosin (5, 28). It has been suggested that only certain M serotypes (referred to as class I) have this cross-reactive epitope (1, 2). GAS strains with the class I antigen tend to be associated nonrandomly with ARF episodes. This is consistent with the fact that despite the extensive genetic diversity in GAS, relatively few M-protein types have been repeatedly associated with ARF in the United States. This observation led in part to the theory that rheumatogenic strains of GAS have virulence factors that promote an autoimmune response in susceptible hosts. However, it is unknown if myosin-reactive antibodies directed against the class I epitope in the C-repeat region participate directly in the pathogenesis of ARF.

Accumulation of new insights into the molecular pathogenesis of ARF and rheumatic heart disease is restricted because most patients are culture negative when disease onset occurs. Hence, the organism responsible for initiating ARF usually is not known and must be inferred from study of GAS isolates cultured from epidemiologically associated individuals. The intermountain region of Salt Lake City, Utah, has experienced unusual and well-documented epidemiological features of

* 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.
ARF in recent years (30, 31). A recent resurgence of ARF in this region provided an opportunity to conduct a population-based study of GAS isolates temporally associated with an ARF outbreak. We used emm sequence analysis to study the population genetic structure of GAS strains cultured from patients with pharyngitis living in and around Salt Lake City and multilocus DNA sequencing and pulsed-field gel electrophoresis (PFGE) to show the limited diversity among GAS emm isolates collected during these two distinct ARF outbreaks.

MATERIALS AND METHODS

Strain collections and sampling design. GAS pharyngeal isolates were obtained from two collections maintained at Primary Children’s Medical Center (PCMC), Salt Lake City. One collection with 3,545 isolates was obtained from May 1998 to September 1999 by population-based sampling of GAS pharyngeal isolates cultured from patients treated at 21 outpatient clinics of Intermountain Health Care (IHC). These 21 clinics are located in Salt Lake City and elsewhere in the state and provide health care to 65% of the Utah population and a small number of individuals living in southern Idaho. A second collection consisted of samples obtained from patients who were referred to PCMC or patients treated at a primary care family practice clinic affiliated with the University of Utah Medical School. In the aggregate, this collection contains isolates recovered between 1984 and 1999, but most isolates were obtained from 1985 to 1992 and 1997 to 1999. This collection is not population based. However, because patients with diverse socioeconomic backgrounds and of diverse ethnicities were treated at these localities, it is reasonable to believe that the culture collections represent strains circulating in the Salt Lake City area at the time that strains were collected.

On isolation the GAS colonies were visually examined and assigned to a mucoid or a nonmucoid category. The analyzed Salt Lake City (PCMC) isolates (collected from 1984 to 1999) contained 250 mucoid and 274 nonmucoid isolates. The analyzed IHC clinic isolates (collected in 1998 and 1999) contained 218 mucoid and 222 nonmucoid isolates. Assignment to a mucoid or a nonmucoid category was done with the understanding that a range of mucoid and nonmucoid phenotypes exists in each category. Due to a reported association of the mucoid colony phenotype with rheumatogenic potential (30), all mucoid GAS isolates in the collection were sampled by randomly selecting 5% of the isolates recovered in each year between 1984 and 1994. Five percent of the nonmucoid GAS isolates in the population-based collection were also randomly sampled.

DNA isolation and PCR amplification. GAS isolates were cultured overnight at 37°C in a 5% CO2 atmosphere on tryptic soy agar supplemented with 5% sheep blood. The isolates were subcultured on brain heart infusion agar and were grown overnight at 37°C in 5% CO2. Genomic DNA was isolated by enzymatic and chemical cell lysis and chemical purification as described elsewhere (26).

PCR amplification of the gene encoding M protein (emm) was done with previously described primers (26). PCR amplification of cosR, cosS, and S were done with primers with the following sequences: 5’-GTCTAGAGATATGAGATGAAATTC (primer cosR FOR), 5’-TAGATATGCTTTTGGTAAACG (primer cosR REV), 5’-TATATCCAAACAGTGCGTGG (primer cosS FOR), 5’-GATTACATATACCTGTAC (primer cosS REV), 5’-AGCAAGATATGCACTTA (primer spa FOR), and 5’-TGCTTGGTGTTAGAAGCAGTAG (primer spa REV).

PCR mixtures contained the following at the indicated final concentrations: reaction buffer (10 mM Tris-hydrochloride [pH 8.3], 50 mM KCl, 1.5 mM MgCl2, 10 ppm gelatin); 200 μM each dCTP, dGTP, dATP, and dTPP; 0.2 μM each forward and reverse primers; 0.04 U of Taq polymerase; and genomic DNA. All PCRs started with denaturation of the template DNA at 94°C for 5 min. PCR amplification of emm was conducted for 32 cycles of 94°C for 1 min, 55°C for 1 min and 3 s, and 72°C for 1 min and 45 s. PCR amplification of spa was conducted for 30 cycles of 94°C for 1 min, 55°C for 1 min and 15 s, and 72°C for 3 min and 45 s. PCR amplification of cosS was conducted in 30 cycles of 94°C for 1 min, 45°C for 1 min and 15 s, and 72°C for 3 min and 45 s. PCR amplification of cosR was conducted for 30 cycles of 94°C for 1 min, 55°C for 1 min and 3 s, and 72°C for 1 min and 15 s. Final extension for all except the emm PCR products was conducted at 72°C for 7 min; for the emm PCR products the extension time was 5 min. The PCR products were purified prior to sequencing with silica gel by the protocol of the manufacturer (Qiagen, Valencia, Calif.).

DNA sequencing and analysis. DNA sequencing reactions were done with BigDye terminator chemistry with 1/10 volume of 1× Ready-PCR terminator Ready-Reaction Mix concentrations, and under standard thermocycler conditions (Perkin-Elmer Applied Biosystems, Foster City, Calif.). The samples were purified with AUTOPREP 96-well plates as described by the manufacturer (Princeton Separations, Adelphia, N.J.) and were analyzed with a PE3700 capillary electrophoresis sequencer (Perkin-Elmer Applied Biosystems). Sequence data were analyzed with Macintosh versions of Sequencher software (version 3.1; Gene Code, Chicago, Ill.) and Lasergene (DNASTar, Madison, Wis.) software.

PFGE. PFGE was performed with a CHEF DR II apparatus (Bio-Rad, Mississauga, Ontario, Canada). Chromosomal DNA was digested with Smal by a modified method described elsewhere (24). Modifications included supplementation of the lysis buffer with 20 μg of mutanolysin (Sigma Chemical Co., Mississauga, Ontario, Canada) per ml and reduction of the lysis time from overnight to 2 to 5 h. The electrophoretic parameters were as follows: pulse times, 5 to 60 s; temperature, 12°C; and 200 V for 20 h.

Statistical analysis. Test if the frequency of occurrence of isolates with emm18 among mucoid GAS isolates in the PCMC collection was independent of peak ARF events, contingency tables and chi-square tests were produced with S-plus 2000 software (MathSoft, Inc., Seattle, Wash.).

RESULTS

Overall descriptive epidemiology of ARF. Cases of ARF in the Salt Lake City area increased in the mid-1990s and persisted at a relatively high rate through the mid-1990s (30, 31). The number of ARF cases in the mid-1990s was similar to that in the previous 5 years, with an average of 26 new ARF cases diagnosed annually at PCMC (Fig. 1). ARF cases resurfaced in the late 1990s, with a 100% increase observed in the winter of 1997-1998. Commensurate with the 1997-1998 increase in ARF, an increase in the number of mucoid GAS isolates also occurred. For example, in the first 6 months of the population-based survey (June to December 1998), mucoid GAS isolates represented 10.9% of the total GAS population collected from patients with pharyngitis. Subsequently, the proportion of mucoid GAS pharyngeal isolates decreased to 1.0% of the total by September 1999 (data not shown). The increased prevalence of mucoid GAS isolates among patients with pharyngitis was also associated with the outbreak of ARF in 1985 and 1986.

emm gene sequence analysis. To determine the emm types of the GAS isolates circulating among Salt Lake City area patients during the period of the study, the region of the emm gene encoding the hypervariable (type-specific) amino terminus of M protein was sequenced for 964 mucoid and nonmucoid isolates. Among the GAS isolates in the PCMC collection, the proportion with the emm18 allele was higher among mucoid isolates collected during years with peaks in the numbers of ARF cases (1998 and 1995) than among isolates collected during all other years during which samples were obtained (χ² = 36.9, P = 1.22 × 10⁻⁶) (Fig. 2A). In addition, GAS isolates with the emm18 allele dominated the population of mucoid GAS isolates collected at IHC clinics in 1998 and 1999 (Fig. 3A). Twenty-four emm types were detected among nonmucoid isolates collected in Salt Lake City between 1984 and 1994 (Fig. 2B), and 18 emm types were detected among nonmucoid isolates collected at IHC clinics in 1998 and 1999 (Fig. 3B). No one specific emm type was dominant among the nonmucoid isolates (Fig. 2B and 3B). Rather, a complex array of strains was observed, and although the distribution of the emm types varies by year, the more frequently isolated strains during ARF outbreaks were also detected during years with a reduced incidence of ARF. For instance, strains with the emm3,
emm12, and sta207 alleles were common in 1985 through 1990 (Fig. 2B); and strains with emm2 and emm77 alleles, which were common in 1998 and 1999 (Fig. 3B), were also common from 1987 to 1990. These findings indicate that GAS isolates with the emm18 allele were the organisms that reemerged in the Salt Lake City area during the two peak periods of ARF.

**Variation in the emm18 gene.** No allelic variation in the 5′/H11032 nucleotide sequence of emm18 was detected in Salt Lake City emm18 isolates, although three distinct size variants of emm18 were identified by PCR (Fig. 4A). To determine the source of size variation in emm18, the entire emm18 genes of 33 isolates were sequenced. Variation in the number of 126-bp DNA segments encoding the C-repeat region produced the observed size differences (Fig. 4B). Among the organisms with an emm18 allele, this segment of DNA was either absent (size variant c) or present in one or two copies (variants a and b, respectively). Most of the isolates (80 to 85% per year) had the a variant of emm18. Organisms with the b variant of emm18 comprised a stable and minor proportion of the population (12 to 15% per year); organisms with the c variant of emm18 were rare, and annually, the proportion of c variants never exceeded 3% of the organisms cultured.

**Limited sequence variations in covR, covS, and spa in emm18 isolates.** Sequence analysis of three additional genes (covR, covS, and spa) was done to further characterize the molecular population genetics of the emm18 organisms. These genes were chosen for analysis because their products either were members of a two-component regulatory system that controls expression of virulence factors such as the capsule, exotoxins, and extracellular proteases (covR and covS) or confer protective immunity (spa) in animal models (6, 10, 15, 21). In addition, inactivation of the covRS system has been linked to a hypervirulent phenotype in a mouse infection model (15) and enhanced resistance to opsonophagocytosis (21). Sequence analysis of 150 emm18 GAS isolates found that covR, covS, and spa were highly conserved among these isolates. No mutations in covR, covS, and spa were detected in isolates from the PCMC collection (n = 80 isolates). One nonsynonymous mutation in covR, two nonsynonymous mutations in covS, and three nonsynonymous mutations in spa occurred in isolates from the IHC clinic collection (n = 70 isolates). In addition, the noncoding region located 5′ to covR in serotype M18 isolates compared to its sequence in a nonmucoid serotype M1 isolate had two purine transition mutations located between the promoter and ribosome binding sites (15).

**PFGE characterization of GAS isolates with emm18.** PFGE analysis was used to determine if major structural differences were present in the chromosomes of isolates with the two common emm18 size variants (designated variants a and b). Three emm18 isolates obtained from ARF patients and seven emm18 isolates representing the two major size variants detected during both peaks of ARF were analyzed. Digestion of chromosomal DNA with SmalI produced identical patterns for all isolates and an estimated genome size of 1.9 Mbp. No differences in the banding patterns of isolates with the two emm18 size variants were identified (data not shown).

**DISCUSSION**

In this study we determined the molecular population genetics of GAS isolates causing pharyngitis in an area that has had repeated resurgences of ARF cases. The analysis discovered that emm18 was the most abundant genotype among mucoid GAS isolates causing pharyngitis during the peaks of
ARF incidence. The coresurgence of isolates with the emm18 allele and ARF cases suggests that when a cohort of susceptible hosts reaches a critical density, this strain successfully disseminates and infects these hosts, resulting in an increased number of ARF cases.

One hypothesis that may explain the repeated resurgence of GAS M18 strains in the Salt Lake City area is that escape mutants of GAS emm18 arise by immune selection and subsequently disseminate among susceptible hosts. However, we found little evidence to support this idea. Structural variants of M18 were identified; but these variants never increased significantly in frequency in the study population, and no variation was observed in spa. Hence, structural variation in neither M18 nor Spa contributed to changes in the frequency of M18 strains in Salt Lake City during peak ARF events. We note that similar conclusions were reported for M1 epidemics in Finland (16). The occurrence of the c variant during the years between peak ARF events suggests that it is a more important member of the population during those years, but the small number of isolates available during those years precludes formal testing. However, this hypothesis is important and should be considered in future strain surveillance studies.

The pathogenic and immunologic properties of M protein have been extensively studied (5, 13). Size variations in M protein were observed among isolates of a serotype M6 strain passaged in vitro and were reported in GAS serotype M6 isolates that were collected over a 40-year period (11, 12). Size variation in M6 is due to recombination events in DNA repeat regions that occur spontaneously in vitro at a rate of 1/1,000 (18, 19). The frequency of occurrence of M-protein size variants in natural populations of GAS isolates is unknown. However, our data indicate that size variation occurs in M18 protein and that variation is the result of differences in the number of repeats in the C region. The occurrence of the a and b size variants among isolates from both peak ARF events indicates that these two size variants are common in GAS emm18 isolates from the Salt Lake City population, whereas the rare

![Figure 2](image2.png)

**FIG. 2.** Distribution of mucoid and nonmucoid GAS isolates in Salt Lake City from 1984 through 1991 and 1997 through 1999. (A) Mucoid GAS isolates from this collection were dominated by serotype M18 strains during peak incidences of ARF. The frequencies of GAS isolates with the emm18 allele are shown separately (blue bars) for each year. Stacked bars show the frequencies of all other emm types within a given year. The inset is the color code for all remaining emm types presenting a mucoid phenotype identified. Sample sizes are given for each year. (B) Three-dimensional plot showing the distributions of nonmucoid GAS isolates from this collection from 1984 through 1994. The inset is the color code for the years during which isolates were sampled. Samples were normalized across years during which isolates were sampled.

![Figure 3](image3.png)

**FIG. 3.** Distribution of GAS isolates collected at IHC clinics in greater Utah during 1998 and 1999. (A) Mucoid GAS isolates from this collection were dominated by isolates with the emm18 allele. The inset is the frequency distributions for all remaining emm types presenting a mucoid phenotype identified (NT, nontypeable). (B) The distribution of emm types among nonmucoid GAS isolates was more uniform during the same period.
proteins that resemble human collagen. Lukomski et al. (22, 23) reported that GAS has two surface cell-surface-displayed proteins, in addition to M protein, have pathogenesis were recently characterized. For example, GAS plex set of GAS virulence factors and host genetics. In this logical response to an antecedent GAS infection. It is generally stimulation that results in disease. However, it is a useful pa-
inflammatory source of autoimmune evidence exists that it is the de
Although the C-repeat region occurs in several M-protein se-
immunoglobulin G to the class I epitope (3), suggesting recent
some ARF patients from Salt Lake City have elevated levels of
acid sequence of the M18 C-repeat regions of the Salt Lake
referred to as class I and class II (1, 2). The predicted amino
protein has been used to divide GAS isolates into two categories,
appearance of the c variant indicates that it is a transient member of the population.
Variation in the structure of the C-repeat region of M protein has been used to divide GAS isolates into two categories, referred to as class I and class II (1, 2). The predicted amino acid sequence of the M18 C-repeat regions of the Salt Lake City GAS isolates indicates that it is a class I M protein, which is in agreement with epitope mapping results (1). The sera of some ARF patients from Salt Lake City have elevated levels of immunoglobulin G to the class I epitope (3), suggesting recent infection with an isolate of GAS expressing the class I epitope. Although the C-repeat region occurs in several M-protein serotypes that are epidemiologically associated with ARF, no evidence exists that it is the definitive source of autoimmune stimulation that results in disease. However, it is a useful parameter for epidemiological characterization of the immunological response to an antecedent GAS infection. It is generally accepted that ARF results from interactions between a complex set of GAS virulence factors and host genetics. In this regard, several GAS genes that may have critical roles in pathogenesis were recently characterized. For example, GAS cell-surface-displayed proteins, in addition to M protein, have cross-reactive properties with host proteins (14, 20), and Lukomski et al. (22, 23) reported that GAS has two surface proteins that resemble human collagen.

The covR and covS genes encode a two-component regulatory system that controls expression of several GAS virulence factors, including the capsule. Inactivation of this regulatory system results in increased production of the capsule and several other proven or putative virulence factors and a hyper-
virulent phenotype in mice (7). However, we found no evidence that nonsynonymous mutations in covR or covS in M18 isolates contributed significantly to the population biology of these strains in the Salt Lake City area. Hence, if altered expression of genes controlled by covR or covS contributes significantly to differences in human-GAS interactions in vivo, it does so by a mechanism yet to be determined. The importance of two single-nucleotide polymorphisms in the upstream regulatory region of covR to autoregulation remains to be investigated, but in M18 strains, these polymorphisms may enhance capsule synthesis. Although the present analysis included only isolates recovered from patients with pharyngitis, we note that Hoe et al. (17) recently reported that pharyngitis and invasive strains of serotype M1 GAS from Finland also lacked sequence variation in these two genes.

GAS emm18 is an important member of the GAS population circulating in the Salt Lake City and intermountain region of the United States. This organism is strongly clonal, based on multilocus typing methods (8, 27), our sequencing data, PFGE analyses, and microarray analysis (29). Although several M-protein serotypes have been historically associated with ARF, the contemporary association between GAS emm18 organisms and ARF in Salt Lake City and other localities in the United States (25) points to the importance of this organism, which should be evaluated in future studies of human-GAS interactions and the pathogenesis of ARF.

ACKNOWLEDGMENTS

We thank Daniel Allen (PCMC), Gail Sylva (Rocky Mountain Laboratories), and Mengyao Liu (Rocky Mountain Laboratories) for technical assistance. We are indebted to James Dale (Veteran’s Hospital, Memphis, Tenn.) for the spa sequence and spa-specific primers and Alison McGeer and Barbara Willey (Mt. Sinai Hospital, Toronto, Ontario, Canada) for PFGE analysis. This study was supported in part at PCMC by grants from the Margolis and Thrasher Foundations.

REFERENCES

2. Bessen, D. E., and V. A. Fischetti. 1990. Differentiation between two biologi-

10. Federle, M. J., K. S. McVey, and J. R. Scott. 1999. A response regulator that represses transcription of several virulence operons in the group A Strepto-