Identification and Characterization of the scl Gene Encoding a Group A Streptococcus Extracellular Protein Virulence Factor with Similarity to Human Collagen

SLAWOMIR LUKOMSKI,† KAZUMITSU NAKASHIMA,† IMAN ABDI,† VINCENT J. CIPRIANO,† ROBIN M. IRELAND,‡ SEAN D. REID‡ GERALD G. ADAMS,‡ AND JAMES M. MUSSER*$

Department of Pathology, Baylor College of Medicine, Houston, Texas 77030,† and Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840‡

Received 19 April 2000/Returned for modification 31 May 2000/Accepted 6 September 2000

Group A Streptococcus (GAS) expresses cell surface proteins that mediate important biological functions such as resistance to phagocytosis, adherence to plasma and extracellular matrix proteins, and degradation of host proteins. An open reading frame encoding a protein of 348 amino acid residues was identified by analysis of the genome sequence available for a serotype M1 strain. The protein has an LPATGE sequence located near the carboxy terminus that matches the consensus sequence (LPXTG) present in many gram-positive cell wall-anchored molecules. Importantly, the central region of this protein contains 50 contiguous Gly-X-X triplet amino acid motifs characteristic of the structure of human collagen. The structural gene (designated scl for streptococcal collagen-like) was present in all 50 GAS isolates tested, which together express 21 different M protein types and represent the breadth of genomic diversity in the species. DNA sequence analysis of the gene in these 50 isolates found that the number of contiguous Gly-X-X motifs ranged from 14 in serotype M6 isolates to 62 in a serotype M41 organism. M1 and M18 organisms had the identical allele, which indicates very recent horizontal gene transfer. The gene was transcribed abundantly in the logarithmic but not stationary phase of growth, a result consistent with the occurrence of a DNA sequence with substantial homology with a consensus Scl binding site immediately upstream of the scl open reading frame. Two isogenic mutant M1 strains created by nonpolar mutagenesis of the scl structural gene were not attenuated for mouse virulence as assessed by intraperitoneal inoculation. In contrast, the isogenic mutant derivative made from the M1 strain representative of the subclone most frequently causing human infections was significantly less virulent when inoculated subcutaneously into mice. In addition, both isogenic mutant strains had significantly reduced adherence to human A549 epithelial cells grown in culture. These studies identify a new extracellular GAS virulence factor that is widely distributed in the species and participates in adherence to host cells and soft tissue pathology.

Group A Streptococcus (GAS) is a human pathogen that causes a wide variety of diseases. Although the molecular mechanisms of GAS pathogenesis are not fully understood it has been well documented that several cell surface proteins participate in host-pathogen interactions (21). Laboratory inactivation of many of these genes has been reported, and the resulting isogenic mutants have been found to be detrimentally affected in biomedically important properties such as resistance to phagocytosis (44), host cell internalization (26), and mouse virulence (7). Many of the proteins have structural features typical of the cell surface proteins of gram-positive bacteria (15), including a variable amino terminus, a central region with repeating sequences, and a cell-associated region containing a cell wall anchor motif with the amino acid sequence LPXTG(X).

Because bacterial cell surface components contribute to several phases of GAS pathogenesis, various molecular biology approaches have been used to identify previously undescribed genes encoding proteins that participate in interaction with the host (11, 48, 51). For example, a fibronectin-binding protein was found by screening a GAS expression library with antibody directed against the cell wall-associated region of streptococcal surface proteins (51). Inhibition of opsonic properties present in anti-GAS cell wall serum by fractionated streptococcal cell wall extracts allowed the identification of a new protective antigen (11). In addition, analysis of the GAS genome database permitted identification of a surface protein (designated GRAB) that regulates proteolysis at the bacterial cell surface by binding to host α2-macroglobulin (48).

Here, we report the identification and characterization of a gene encoding an extracellular protein that participates in adherence of the pathogen to host epithelial cells and contributes to virulence, as assessed by subcutaneous inoculation of mice. This gene was present in all 50 genetically diverse GAS isolates analyzed, which represent 21 distinct M protein serotypes. A noteworthy feature of the protein is the occurrence of extensive and contiguous Gly-X-X (where X is an undefined amino acid) amino acid repeats characteristic of human collagen. The structural gene (designated scl for streptococcal collagen-like), was transcribed abundantly in the exponential phase of growth. Isogenic strains in which the scl gene was inactivated by a nonpolar mutagenesis technique were significantly reduced in their ability to adhere to human epithelial cells grown in culture and were significantly less pathogenic in a mouse model of soft tissue infection.
Streptococcal collagen-like protein: a cell wall protein bearing structural and functional homology to collagen

I. Introduction

Streptococcal collagen-like protein is a cell wall protein that bears structural and functional homology to collagen. It is encoded by the \( \text{scI} \) gene in GAS strains.

II. Materials and Methods

A. Bacterial strains and growth.

56 GAS strains isolated from clinical sources were used. Growth conditions and DNA techniques were performed as described previously.

B. \( \text{scI} \) gene identification and expression

The \( \text{scI} \) gene was identified in GAS strains from GAS 6708 and MGAS 5005 with a nonpolymorphic specific resistance cassette designated \( \text{spc} \) (30). A 19.6-kb chromosomal fragment containing the \( \text{scI} \) gene and flanking regions was amplified from GAS 6708 with primers \( \text{scI-Smal} \) (5' TCTCCCGGGTTTCGCCTGAGCTC) and reverse primer \( \text{scI-Smal} \) (5' TCTCCCCGGTTCCGCTTGCTTTC) of the \( \text{scI} \) gene. This PCR fragment was digested with \( \text{PstI} \) and \( \text{Smal} \) to generate an 600-bp fragment, which was cloned into the \( \text{PstI} \)-\( \text{Smal} \) site of the plasmid DNA polymerase.

C. DNA hybridization and sequencing

The resulting plasmid was DNA polymerase. The resulting plasmid (pSL131) contained 200 bp of the 3' end coding region of \( \text{scI} \) and 400 bp of downstream noncoding DNA. A DNA fragment of 700 bp was generated from a nested PCR using primers \( \text{scI-SalI} \) and \( \text{scI-HindII} \). This fragment contained 300 bp of the 5' end coding region of \( \text{scI} \) and 400 bp of the upstream noncoding region and was cloned between the \( \text{SalI} \)-\( \text{HindIII} \) sites of pSL131, generating plasmid pSL132. Both \( \text{SalI} \) and \( \text{HindIII} \) sites were blunt-ended with the Klenow fragment DNA polymerase.

D. Electrophoretic mobility of \( \text{scI} \) gene

E. DNA sequencing

F. Cloning experiments

G. Expression of \( \text{scI} \) gene

III. Results

A. DNA sequence analysis

B. DNA sequence analysis

C. DNA sequence analysis

D. DNA sequence analysis

E. DNA sequence analysis

F. DNA sequence analysis

G. DNA sequence analysis

H. DNA sequence analysis

I. DNA sequence analysis

IV. Discussion

A. Comparison of \( \text{scI} \) gene

B. Comparison of \( \text{scI} \) gene

C. Comparison of \( \text{scI} \) gene

D. Comparison of \( \text{scI} \) gene

E. Comparison of \( \text{scI} \) gene

F. Comparison of \( \text{scI} \) gene

V. Conclusion

A. Conclusion

B. Conclusion

C. Conclusion

D. Conclusion

E. Conclusion

F. Conclusion

G. Conclusion

H. Conclusion

V. Acknowledgments

A. Acknowledgments

B. Acknowledgments

C. Acknowledgments

D. Acknowledgments

E. Acknowledgments

F. Acknowledgments

G. Acknowledgments

H. Acknowledgments

I. Acknowledgments

References

A. References

B. References

C. References

D. References

E. References

F. References

G. References

H. References

I. References
FIG. 1. Nucleotide and amino acid sequence of the \( sc/ \) gene and Scl protein in serotype M1 GAS. (A) The \( sc/ \) coding sequence consists of 1,047 bp (nucleotides 77 to 1123) and encodes a polypeptide of 348 amino acids. The presumed \( sc/ \) promoter region contains a predicted ribosome-binding site (RBS) and \(-10\) and \(-35\) regions. A predicted transcription site, \(+1\), is denoted by a dot. A potential binding site for the streptococcal positive transcription regulator, Mga, overlapping with the \(-35\) region is shown. A potential transcription terminator (II) consisting of two inverted repeats was identified downstream of \( sc/ \) gene. The predicted ATG start codon and the TAA stop codon are shown in boldface type. SS, signal sequence showing the predicted cleavage site (arrowhead); V, V region of the mature Scl protein; CL, CL region consisting of 50 GXX triplets (boxed); L, L region with PGEKAPEKS repeats (underlined); WM, cell-associated Scl part containing a cell wall segment with the LPATGE motif (shaded), and cell membrane hydrophobic region with a charged tail. (B) Alignment of the consensus Mga binding site and putative promoter region of the \( sc/ \) gene. The consensus Mga binding site was identified on the basis of sequence alignments of several Mga-regulated genes (34). The essential bases are shown in boldface type, and essential sequences within the potential Mga binding site of the \( sc/ \) gene promoter which do not match with the consensus sequence are shown in lowercase type. Twenty-two of 29 (76\%) essential bases are identical. Underlined bases in the putative promoter region of the \( sc/ \) gene promoter denote a potential \(-35\) promoter region.
corresponding to the hydrophobic cell membrane domain of conserved cell anchor LPXTG(X) motif. The protein sequence brane domain (15). The cell-wall domain is located amino sequences with similarity to the query were identified. Project database (http//www.genome.ou.edu/strep.html), and strains was statistically significant.

The blot was developed with streptavidin-peroxidase conjugate and a chemiluminescence detection method was used to visualize hybridizing bands. A single transcript been deposited in GenBank under accession number 330137. Calculated with the wild-type and regulated with one within-subjects and one between-subjects factor. Statistical calculations were performed with Proc mixed software (SAS Institute Inc., Cary, N.C).

Nucleotide sequence accession number. The sequence reported here has been deposited in GenBank under accession number 330137.

RESULTS

Identification and analysis of the scl gene and inferred Scl protein in serotype M1 GAS. A common theme of the cell-associated carboxy-terminal part of surface proteins from gram-positive cocci is the presence of a charged tail of up to 7 amino acid residues preceded by a hydrophobic transmembrane domain (15). The cell-wall domain is located amino terminal to this hydrophobic segment and contains a highly conserved cell anchor LPXTG(X) motif. The protein sequence corresponding to the hydrophobic cell membrane domain of the cell surface protein M6 (FFTAALTVMATAGVAAVY) (14) was used to search the Streptococcal Genome Sequencing Project database (http//www.genome.ou.edu/strep.html), and several sequences with similarity to the query were identified. One of these ORFs (designated scl for streptococcal collagen-like) is shown in Fig. 1A. The scl ORF is 1,047 bp long (nucleotide 77 to 1123) and would encode a protein with 348 amino acid residues. A potential promoter located upstream of this ORF includes a −10 region (tataat; perfect match of the consensus sequence) and a −35 region (tttaca; five of six bases identical [in boldface type] to the consensus sequence ttgaca (52). A potential ribosome-binding site (gaaagaga; the consen-

sus sequence is taaggagg [identical residues shown in boldface type]) is located 9 nucleotides upstream from the ATG-Met start codon (20, 57).

The predicted molecular mass of the Scl protein is ~36.3 kDa. As expected for an exported GAS cell surface protein, the amino terminus (encompassing amino acid residues 1 to 37) (Fig. 1A) has structural features characteristic of signal sequences (42), such as a short amino-terminal hydrophilic region followed by a hydrophobic transmembrane segment. The signal peptidase cleavage site usually has a small amino acid residue, and Ala37 satisfies this criterion in Scl. The predicted molecular mass of the mature protein (residues 38 to 348) is ~32.4 kDa, and the predicted isoelectric point is 5.54. The amino-terminal segment of the inferred mature Scl protein (amino acid residues 38 to 105) is hydrophilic and has a predicted α-helical structure.

A noteworthy structural feature of Scl is a collagen-like region consisting of 50 Gly-X-X (GXX) triplet repeats that encompass amino acid residues 106 to 255 (designated CL, for collagen-like region) (47). The GXX sequence in collagen forms a right-handed triple helix with three molecular chains. The inferred bacterial cell-associated part of Scl (designated the WM region, encoded by amino acid residues 267 to 348) has an apparent cell wall domain with a cell anchor motif (LFATGE) and a hydrophobic transmembrane segment followed by a charged tail at the carboxyl terminus. The CL and WM regions are connected with a linker sequence (L) consisting of repeats with a PGEKAPEKS core sequence. In M1 GAS there are two of these repeats.

scl gene expression in M1 GAS strains. Northern blot analysis was performed to determine if the scl gene was transcribed (Fig. 2A). Total RNA was isolated from strain MGAS 6708 grown to logarithmic (OD_{600} ~ 0.5) or stationary (OD_{600} ~ 0.8 and overnight) phase and probed with the entire scl gene
amplified by PCR. A single transcript of ~1.2 kb corresponding to the predicted length of a single-gene \( scl \) transcript was identified. The same results were obtained with strain MGAS 5005. A predicted +1 transcription start site is shown in Fig. 1A. In addition, a potential transcription terminator with two inverted repeat sequences is present 40 bp downstream from the TAA stop codon. When these upstream and downstream putative regulatory regions are considered, the calculated 1,193-bp size of the \( scl \) transcript agrees with the size of the hybridizing band identified by Northern analysis (Fig. 2A).

Only weak \( scl \) transcription occurred in the early stationary phase of growth, and no \( scl \)-specific mRNA was detected in the overnight cultures of both M1 strains tested. Together, this pattern of \( scl \) transcription is characteristic of genes regulated by Mga (multiple gene regulator), a positive transcriptional regulator of several virulence genes (4, 6, 43). In this regard, we note that upstream of the \( scl \) ORF there is a DNA sequence homologous to the consensus Mga binding site (34) (Fig. 1B). This potential binding site for Mga overlaps with the predicted -35 promoter region, a molecular arrangement consistent with the characteristics of Mga binding sites (43). Northern blot analysis of isogenic M1 strains (wild type and \( mga \) mutant) confirmed our prediction that \( scl \) is under \( mga \) control (unpublished data).

The presence of the Scl protein in the cell wall-associated fraction was confirmed by analysis of two genetically distinct GAS strains harvested in the exponential phase of growth (Fig. 2B). Both wild-type MGAS 5005 and MGAS 6708 strains expressed a protein with an approximate molecular mass of 49 kDa that reacted with rabbit antisera raised against an M1-specific Scl peptide. Although the predicted molecular mass of the mature Scl protein is ~32 kDa, it is a common observation that cell surface proteins from gram-positive bacteria migrate aberrantly slowly in SDS-polyacrylamide electrophoresis gels (1, 27, 48). It is also well known that proteins with repetitive motifs tend to migrate aberrantly. Taken together, the data indicate that the \( scl \) gene is transcribed and translated by GAS serotype M1 strains.

**scl gene variation among genetically diverse GAS strains.**

Inasmuch as GAS isolates can have extensive chromosomal structural heterogeneity, we next screened 50 genetically diverse and epidemiologically unrelated strains by PCR with \( scl \)-specific primers to assess the distribution of the \( scl \) gene. A PCR product was obtained from all 50 strains, but considerable size variation was identified between strains expressing different M protein serotypes (Fig. 3A). In addition, \( scl \) size variation was identified among serotype M12, M28, and M56 organisms (Fig. 3B).

The \( scl \) gene in these 50 strains was sequenced to determine the molecular basis of size variation. No allelic variation was found among strains expressing serotype M1 (\( n = 7 \) isolates), M2 (\( n = 2 \)), M3 (\( n = 5 \)), M6 (\( n = 4 \)), or M18 (\( n = 3 \)). Three mature Scl variants each were identified among six M12 isolates and four M28 GAS strains (Table 1). In addition, two Scl variants each were found among M52 (\( n = 3 \)), M56 (\( n = 2 \)), and M77 (\( n = 2 \)) organisms. Two Scl proteins from M49 GAS strains differed by one amino acid residue within the signal peptide. \( scl \) gene variation among isolates of serotypes M4, M9, M22, M41, M55, M57, M75, M76, ST2035, and ST2967 could not be assessed, because only one isolate of each of these M types was sequenced.

When comparing the 37 distinct Scl variants identified, well-conserved and highly polymorphic regions were found. The amino-terminal region of the mature Scl proteins (amino acids 38 to 105 [Fig. 1A]) was highly polymorphic among these strains. For purposes of description and discussion, the amino-terminal portion of the mature Scl protein is designated as the V (variable) region. When comparing different M types, the V region was polymorphic in number of amino acids and primary sequence (Table 1). The CL region, located adjacent to the V region, also was highly variable in length as a consequence of the different number of GXX motifs. The CL region in Scl from all M1 strains (Fig. 1A) has 50 GXX triplets. However, 14 GXX motifs were identified in the four M6 organisms studied, whereas one M41 isolate had 62 GXX motifs. The GXX motifs present in the 50 GAS isolates also differed in primary amino acid sequence. For example, 21 distinct GXX amino acid motif sequences were identified among the 50 GXX triplets in the CL region of the M1 strains. In the aggregate, 50 distinct GXX amino acid motif sequences were identified in the CL regions of the 50 strains studied. Thus, the CL region was characterized by variation in number and primary amino acid sequence of the GXX amino acid motifs, thereby resulting in a highly polymorphic segment of Scl.

The carboxy-terminal end of the CL region (the segment located closer to the cell wall) was less polymorphic than the amino-terminal end of the CL region located more distant from the cell wall anchor. Between the CL region and the conserved LPATGE cell wall anchor is a segment of the protein referred to as the L region. When comparing the sc-
TABLE 1. Size variation in the scl gene GAS strains expressing 21 M types

<table>
<thead>
<tr>
<th>M type</th>
<th>No. of strains sequenced</th>
<th>scl pattern (no. of strains)</th>
<th>No. of amino acids in V region</th>
<th>No. of GXY motifs in CL region</th>
<th>No. of PGEKAPEKS repeats in L region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>67</td>
<td>60</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>85</td>
<td>46</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>70</td>
<td>10, 25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>70</td>
<td>42</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>70</td>
<td>14</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>69</td>
<td>41</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>75</td>
<td>38</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>67</td>
<td>50</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>72</td>
<td>45</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>75</td>
<td>48</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>83</td>
<td>62</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>49&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2</td>
<td>71</td>
<td>46</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>3</td>
<td>68</td>
<td>60</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>1</td>
<td>68</td>
<td>59</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>2</td>
<td>71</td>
<td>55</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>1</td>
<td>71</td>
<td>38</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>1</td>
<td>75</td>
<td>30</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>1</td>
<td>69</td>
<td>44</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>2</td>
<td>67</td>
<td>52</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
<td>55</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
<td>47</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>ST2035</td>
<td>1</td>
<td>85</td>
<td>43</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>ST2967</td>
<td>1</td>
<td>66</td>
<td>45</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Refers to PCR product size and DNA sequence data.
<sup>b</sup>The scl gene in M3 isolates has a mutation which creates an in-frame TAA stop codon within the 11th GXX triplet motif.
<sup>c</sup>The two isolates of M49 differed at amino acid position 7 (H or Y) in the signal sequence.

sequences from the 50 strains analyzed, the L region was composed of two to seven conserved repeats, each containing an identical 9-amino-acid core sequence (PGEKAPEKS) (Table 1).

Two observations warrant special comment. First, serotype M1 and M18 strains had the identical scl allele, a result indicative of very recent horizontal gene transfer. Second, serotype M3 organisms had a nucleotide substitution creating a stop codon that would truncate the Scl protein at amino acid 139 of the Scl protein. Truncation of Scl at this location would result in an exported protein with an inferred molecular mass of 11 kDa, containing 10 GXX triplets.

Nonpolar inactivation of the scl gene in serotype M1 GAS.

To facilitate assessment of the role of Scl in host-pathogen interaction, isogenic nonpolar scl mutant derivatives were constructed from serotype MJ strains MGAS 6708 and MGAS 5005. We focused the analysis on M1 strains because this is the most common M type recovered from invasive infections in many case series. MGAS 6708, a speA-Iacking strain, is identical to the strain whose chromosome was recently sequenced (59), and MGAS 5005 (speA positive) is representative of organisms causing most cases of human invasive disease and pharyngitis (24, 36). Two M1 organisms were used because considerable evidence has accrued in recent years demonstrating that chromosomal variation occurs among M1 strains (23, 36). Hence, not all M1 strains are genetically equivalent. In addition, although a chromosomal sequence is available for MGAS 6708, this strain is not genetically representative of the vast majority of M1 organisms recovered from invasive and pharyngitis cases worldwide (23).

Each strain yielded numerous spectinomycin-resistant colonies after electroporation with suicide plasmid pSL134 (Fig. 4A). One scl mutant derivative of MGAS 6708 and one of MGAS 5005 were selected for further studies after analysis by PCR, Southern hybridization, and sequencing documented that the correct molecular genetic constructs had been generated (data not shown). These genetic analyses found that the mutants of the scl gene in both strains were identical and contained the expected deletion of nucleotides 114 to 1082 and replacement with the spc2 nonpolar cassette (Fig. 4A). Also as expected, scl-specific transcripts were not detected in the mutant strains during exponential growth, a time when abundant scl-specific mRNA is synthesized by the parental organisms (Fig. 4B). These results were further supported by the lack of the Scl protein in the periplasmic fraction of the scl mutants, whereas the parental wild-type isolates produced Scl (Fig. 4C). The isogenic mutants had the same colony morphology and identical growth curves as the parental isolates (data not shown).

Scl and GAS virulence in mice.

Genetic inactivation of expression of several GAS extracellular products has been reported to significantly decrease virulence in mice (7, 32, 35). In addition, we have shown in the present study that the scl gene was transcribed in the phase of growth when several known GAS virulence factors are made. As a consequence, we compared the ability of the wild-type parental organisms and isogenic mutant strains to cause mouse death after i.p. inoculation.

Previous experiments demonstrated that i.p. inoculation of 2 x 10<sup>7</sup> CFU of wild-type MGAS 5005 (speA positive) caused the death of 90% of mice after 5 days (data not shown). To determine the inoculum of MGAS 6708 (speA laking) re-
FIG 4. scl gene expression in wild-type and mutant M1 GAS. (A) Schematic representation of suicide plasmid pSL134 used for nonpolar inactivation of the scl gene in serotype M1 GAS. Most of the scl gene sequence (969 of 1,047 bp) was replaced with a nonpolar spectinomycin resistance cassette (spc2) inserted in frame. To enhance the likelihood of homologous recombination, ~400 bp of DNA upstream and downstream of the scl ORF also was cloned. pSL134 is a suicide plasmid that cannot replicate in GAS. Ss, signal sequence; V, variable region; CL, collagen-like region; WM, cell wall and cell membrane region; tt, transcription terminator; MeS, multiple cloning site. (B) Northern blot analysis of the total RNA from logarithmic-phase GAS cultures. Samples (15 μg) were blotted and hybridized with scl- and recA-specific biotinylated DNA probes and detected by chemiluminescence. The scl-specific transcript is detected in the wild-type (wt) but not mutant sample. Control hybridization with the recA probe showed a positive signal for both the wild-type and scl mutants. Identical results were obtained for the MGAS 6708 isogenic pair (data not shown). (C) Western blot analysis of the cell wall-associated protein fractions obtained from the wild-type and scl mutant strains. Scl is detected in the samples prepared from the wild-type but not mutant strains.
required to cause equivalent mortality, mice were inoculated i.p. with twofold serial dilutions ranging from $10^7$ to $10^9$ CFU. Interestingly, at the highest inoculum used ($10^9$ CFU), only 70% of mice died (Fig. 5A), which means that MGAS 6708 is significantly less virulent than MGAS 5005. This result is consistent with the rarity of recovery of this M1 subclone from human infections (23).

We next inoculated mice i.p. with the wild-type and scl mutant strains ($2 \times 10^7$ CFU was used for the MGAS 5005 strain pair and $10^9$ CFU was used for the MGAS 6708 strain pair). No significant difference in mouse mortality was observed for either pair of isogenic GAS strains ($P < 0.539$ (log-rank test) for the MGAS 5005 strain pair; $P < 0.821$ (log-rank test) for the MGAS 6708 strain pair) (Fig. 5B and C). The same results were obtained when the i.p. inoculation experiments were repeated (data not shown). Hence, we conclude that Scl does not participate significantly in virulence in this mouse model.

Certain extracellular GAS products also have been reported to participate in the pathogenesis of invasive skin disease when mice are injected s.c. with GAS (2, 28, 31). We next investigated the effect of inactivation of the scl gene on mouse mortality and skin pathology after s.c. inoculation with the wild-type MGAS 5005 and isogenic mutant derivative. The MGAS 6708 strain pair was not studied in this model due to its relative lack of virulence in the i.p. model. There was no significant difference between MGAS 5005 mutant and wild-type strains in mouse mortality ($P < 0.3279$ (log-rank test)). Mice inoculated with both wild-type and mutant strains developed skin lesions that healed over 3 weeks of the experiment (Table 2). Similarly, after the initial weight loss, mice in both experimental groups regained weight. When weight loss and severity of the skin lesions in two experimental groups were compared on days 2 and 4 after the inoculation, the differences were not statistically significant (t test), probably due to the large variation within animal groups. However, when those parameters were compared over the time of the experiment, the differences in both weight loss ($P < 0.0001$) and the severity of soft-tissue pathology (area, $P < 0.0001$; volume, $P < 0.0001$) between mice injected with the wild-type GAS and scl isogenic mutant were statistically significant (mixed model repeated measures). Hence, inactivation of the scl gene was associated with a decreased morbidity in this model of GAS pathogenesis.

**Scl and GAS adherence to human epithelial cells.** Attachment of GAS to host cells via adhesins is considered to be an important step in successful colonization. We hypothesized that Scl may act as an adhesin as a consequence of its extracellular location and collagen-like structure. To test this hypothesis, we compared the ability of wild-type MGAS 5005 and MGAS 6708 and their isogenic mutant derivatives to adhere to human A549 epithelial cells grown in vitro. There was no significant difference in the number of wild-type MGAS 5005 and MGAS 6708 that adhered to the A549 epithelial cells and their isogenic mutant derivatives to adhere to human A549 epithelial cells grown in vitro. There was no significant difference in the number of wild-type MGAS 5005 and MGAS 6708 that adhered to the A549 epithelial cells. In contrast, compared to the wild-type parental organisms, significantly fewer isogenic mutant bacteria adhered to the host cells (Fig. 6). The growth indices in the tissue culture medium were identical for the parental and mutant bacterial strains (data not shown). Taken together, the data are consistent with the hypothesis that the lack of Scl expression caused by nonpolar inactivation of the scl gene results in a detrimental effect on GAS adherence to host cells in vitro.

**Molecular evolutionary genetic analysis.** The observations that Scl is an extracellular protein potentially subjected to selective pressure in the host suggested that molecular evolutionary genetic analysis would provide insight into the processes shaping variation at the scl locus. Multiple-sequence

---

**TABLE 2. Skin pathology in mice infected s.c. with the wild-type and scl isogenic mutant GAS strains**

<table>
<thead>
<tr>
<th>Animal group infected with GAS</th>
<th>Severity of skin lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean area (cm²) (range) at:</td>
</tr>
<tr>
<td></td>
<td>48 h</td>
</tr>
<tr>
<td>5005 wild type</td>
<td>152 (42-549)</td>
</tr>
<tr>
<td>5005 scl</td>
<td>110 (31-234)</td>
</tr>
</tbody>
</table>
alignment of 50 inferred Scl amino acid sequences by the Clustal W method required introduction of numerous alignment gaps, particularly in the V and CL domains. To determine if the level of selective constraint varies in different regions of Scl, we calculated the proportion of nonsynonymous and synonymous nucleotide substitutions (\(p_s\) and \(p_s\), respectively) for the four Scl domains among the unique \(scl\) alleles (the L region was analyzed together with the CL region). The proportions, calculated as \(p_s - p_s\) ± standard error, for the following domains were as indicated: signal sequence, \(-0.0999 ± 0.0243\); V region, \(-0.1785 ± 0.0304\); CL region (the L region was analyzed together with the CL region), \(-0.1887 ± 0.0261\); wall membrane region, \(-0.0151 ± 0.0092\); overall, \(-0.0819 ± 0.0087\). The difference, \(p_s - p_s\), is a measure of the degree of selective constraint. The more negative the value, the less the contribution of amino acid replacements and the greater the contribution of synonymous (silent, not resulting in amino acid replacements) nucleotide substitutions. A difference of zero indicates selectively neutral variation, where the per site rates of \(p_s\) and \(p_s\) are equal. A positive difference (amino acid replacements exceed silent substitutions) suggests the action of diversifying (positive) selection. In general, the degree of selective constraint is low for \(scl\) as a whole, with the value of \(p_s - p_s\) approaching zero. The level of constraint is not uniform, however. The region of highest constraint spans the V and CL domains, with the two values differing by 0.01. In contrast, the values for the V and CL domains differ from the signal sequence and wall membrane domains by an average of 0.1261. The pattern of nonsynonymous and synonymous changes across the gene suggests the locus is responding to selective pressure, and the elevated constraint of the V and CL domains demonstrates the functional importance of this region.

DISCUSSION

Bacterial pathogenesis is a complex process that can involve attachment, colonization, internalization by epithelial and endothelial cells, local tissue destruction, and dissemination to distant anatomic sites. GAS express virulence factors that mediate one or more of these processes, including cell surface adhesins that participate in attachment to host cells (21). In this study we identified a new streptococcal cell surface molecule with the following characteristics: (i) the structural gene is commonly found in GAS strains and is expressed in a growth phase-dependent fashion, (ii) the protein contains a large region composed of GXX repeats with structural homology to human collagen, (iii) the protein participates in GAS attachment to human epithelial cells, and (iv) the protein participates in soft-tissue pathogenesis in the mouse.

The \(scl\) gene is common among GAS strains and is expressed in the logarithmic phase of growth. GAS is composed of a heterogeneous array of chromosomal genotypes. Many genes encoding cell surface or other extracellular proteins have considerable allelic variation (25, 37), and many are not present in all GAS isolates. We found that the \(scl\) gene is widely distributed among 50 GAS strains that together represent the breadth of genomic diversity in the species. However, extensive size and protein sequence variation was identified (Fig. 7). The comparison of sequence data indicated that the amino-terminal signal sequence and the carboxy-terminal cell wall and membrane regions were largely conserved, whereas the central V and CL regions varied between the strains. This result suggested that the Scl protein might be under positive selection for amino acid replacements. To test this hypothesis, we measured the level of selective constraint, an indicator of the functional importance of the molecule and the degree of selective pressure present (38). Overall, constraint is low, with the value of \(p_s - p_s\) approaching zero. However, the degree of constraint is not uniform, for the central region of the gene (encoding V and CL) possesses an elevated negative value of \(p_s - p_s\). Variation of \(p_s - p_s\) across a protein has been used by Reid et al. (49) to assess the relative functional importance of specific domains in an extracellular antigen of pathogenic Escherichia coli. The high level of codon bias present in the CL region (characterized by an increased proline and lysine content) and the requirement that glycine occupy every third residue demonstrate an increased level of selective constraint as indicated by the change in \(p_s - p_s\). Presumably, the maintenance of the GXX motif in Scl is essential for the formation of a structure potentially required for interaction with host molecules such as extracellular matrix proteins and integrins. Hence, the level of constraint in the CL region restricts the nature of the changes such that the GXX motif is maintained. The molecular nature and biological importance of the constraint acting on the V region is yet to be elucidated.

The \(scl\) gene is abundantly transcribed in logarithmic growth but not in the stationary growth phase. Temporal variation in the expression of proven and presumed GAS virulence factors has been well described. Several complex gene regulatory systems are known, including the Mga regulon (6, 43, 56); the CovR-CovS two-component regulatory system (13, 22, 29); a positive regulator, RoIA (16); and the negative regulator Nra (46). Genes controlled by Mga are expressed in the exponential phase of growth in a process that involves binding of Mga to promoter regions. Mga binding sites generally overlap with the -35 regulatory sequence (34, 43). Our analysis identified a sequence in the presumed \(scl\) promoter with homology with the consensus Mga binding site sequence (34) (Fig. 1B). Taken together, these data indicate that the \(scl\) gene is present in most if not all GAS strains and is coexpressed with other virulence factors under Mga regulation.

\(scl\) contains a unique GXX triplet repeat region. The Scl protein contains a region consisting of extensive GXX triplet repeats characteristic of human collagen. CL repeats are exceedingly rare in prokaryotic pathogens. Interestingly, a region

FIG. 6. Adherence of isogenic M1 GAS strains to cultured human epithelial cells. Mid-log phase cultures of the wild-type strains (open bars) and \(scl\)-inactivated mutants (filled bars) were incubated with A549 epithelial cells for 1 h. Unattached bacteria were removed by washing. The attached GAS were stained and counted microscopically. The results are presented as mean numbers of GAS cells attached to each A549 cell ± standard error of the mean (error bars). The difference in adherence between the wild type and mutant derivatives was significant by \(t\)-test for both MGAS 5005 and MGAS 6708 strain pairs.

The proportions, calculated as \(p_s - p_s\) ± standard error, for the following domains were as indicated: signal sequence, \(-0.0999 ± 0.0243\); V region, \(-0.1785 ± 0.0304\); CL region (the L region was analyzed together with the CL region), \(-0.1887 ± 0.0261\); wall membrane region, \(-0.0151 ± 0.0092\); overall, \(-0.0819 ± 0.0087\). The difference, \(p_s - p_s\), is a measure of the degree of selective constraint. The more negative the value, the less the contribution of amino acid replacements and the greater the contribution of synonymous (silent, not resulting in amino acid replacements) nucleotide substitutions. A difference of zero indicates selectively neutral variation, where the per site rates of \(p_s\) and \(p_s\) are equal. A positive difference (amino acid replacements exceed silent substitutions) suggests the action of diversifying (positive) selection. In general, the degree of selective constraint is low for \(scl\) as a whole, with the value of \(p_s - p_s\) approaching zero. The level of constraint is not uniform, however. The region of highest constraint spans the V and CL domains, with the two values differing by 0.01. In contrast, the values for the V and CL domains differ from the signal sequence and wall membrane domains by an average of 0.1261. The pattern of nonsynonymous and synonymous changes across the gene suggests the locus is responding to selective pressure, and the elevated constraint of the V and CL domains demonstrates the functional importance of this region.

DISCUSSION

Bacterial pathogenesis is a complex process that can involve attachment, colonization, internalization by epithelial and endothelial cells, local tissue destruction, and dissemination to distant anatomic sites. GAS express virulence factors that mediate one or more of these processes, including cell surface adhesins that participate in attachment to host cells (21). In this study we identified a new streptococcal cell surface molecule with the following characteristics: (i) the structural gene is commonly found in GAS strains and is expressed in a growth phase-dependent fashion, (ii) the protein contains a large region composed of GXX repeats with structural homology to human collagen, (iii) the protein participates in GAS attachment to human epithelial cells, and (iv) the protein participates in soft-tissue pathogenesis in the mouse.

The \(scl\) gene is common among GAS strains and is expressed in the logarithmic phase of growth. GAS is composed of a heterogeneous array of chromosomal genotypes. Many genes encoding cell surface or other extracellular proteins have considerable allelic variation (25, 37), and many are not present in all GAS isolates. We found that the \(scl\) gene is widely distributed among 50 GAS strains that together represent the breadth of genomic diversity in the species. However, extensive size and protein sequence variation was identified (Fig. 7). The comparison of sequence data indicated that the amino-terminal signal sequence and the carboxy-terminal cell wall and membrane regions were largely conserved, whereas the central V and CL regions varied between the strains. This result suggested that the Scl protein might be under positive selection for amino acid replacements. To test this hypothesis, we measured the level of selective constraint, an indicator of the functional importance of the molecule and the degree of selective pressure present (38). Overall, constraint is low, with the value of \(p_s - p_s\) approaching zero. However, the degree of constraint is not uniform, for the central region of the gene (encoding V and CL) possesses an elevated negative value of \(p_s - p_s\). Variation of \(p_s - p_s\) across a protein has been used by Reid et al. (49) to assess the relative functional importance of specific domains in an extracellular antigen of pathogenic Escherichia coli. The high level of codon bias present in the CL region (characterized by an increased proline and lysine content) and the requirement that glycine occupy every third residue demonstrate an increased level of selective constraint as indicated by the change in \(p_s - p_s\). Presumably, the maintenance of the GXX motif in Scl is essential for the formation of a structure potentially required for interaction with host molecules such as extracellular matrix proteins and integrins. Hence, the level of constraint in the CL region restricts the nature of the changes such that the GXX motif is maintained. The molecular nature and biological importance of the constraint acting on the V region is yet to be elucidated.

The \(scl\) gene is abundantly transcribed in logarithmic growth but not in the stationary growth phase. Temporal variation in the expression of proven and presumed GAS virulence factors has been well described. Several complex gene regulatory circuits are known, including the Mga regulon (6, 43, 56); the CovR-CovS two-component regulatory system (13, 22, 29); a positive regulator, RoIA (16); and the negative regulator Nra (46). Genes controlled by Mga are expressed in the exponential phase of growth in a process that involves binding of Mga to promoter regions. Mga binding sites generally overlap with the -35 regulatory sequence (34, 43). Our analysis identified a sequence in the presumed \(scl\) promoter with homology with the consensus Mga binding site sequence (34) (Fig. 1B). Taken together, these data indicate that the \(scl\) gene is present in most if not all GAS strains and is coexpressed with other virulence factors under Mga regulation.

\(scl\) contains a unique GXX triplet repeat region. The Scl protein contains a region consisting of extensive GXX triplet repeats characteristic of human collagen. CL repeats are exceedingly rare in prokaryotic pathogens. Interestingly, a region
of 30 amino acid with 10 GXX repeats was described previously in the GAS bacteriophage-encoded hyaluronidase (33). In addition, Erickson et al. (12) reported that a platelet aggregation-associated protein of Streptococcus sanguis has a sequence (PGEQGPK) with two GXX repeats. However, the roles of these proteins in virulence were not assessed with isogenic mutant strains and model systems. In human collagen, three chains entwine to form a right-handed superhelix (47). Three characteristics of the GXX region in Scl suggest that it is capable of forming a similar quaternary structure. First, Scl fulfills the absolute requirement that glycine occupy every third residue (only glycine can fit into the center of the triple helix without distorting it). Second, on the basis of computer modeling, the GXX region of Scl is predicted to form a coiled structure. Third, lysines and prolines essential for triple-helix stabilization together comprise ~31% (percent by frequency) of the nonglycine residues. Scl and host interactions. Our studies indicated that Scl contributes significantly to adherence of GAS to human epithelial cells grown in culture, thereby adding to the relatively large number of surface-associated bacterial molecules that contribute to this process (21). The scl gene is transcribed in the logarithmic phase of growth, suggesting that the adhesin is made early in the course of host-pathogen interaction. Other molecules that contribute to host adherence that are made early in the growth phase include M protein (5), M-like proteins (17), and serum opacity factor (7). Among nonprotein streptococcal adhesins, hyaluronic acid capsule (55) and lipoteichoic acid (21, 40) also participate in adherence of GAS. Although our studies show that Scl participates in adherence in vitro, we did not address the molecular mechanism whereby Scl contributes to GAS-host cell interaction. Inasmuch as many eukaryotic cells bind collagen through receptors expressed on cell surfaces (3, 61, 62), Scl may participate in colonization of the host by docking GAS to host cell receptors (60). One biochemically important example of a collagen ligand is integrins, a class of transmembrane proteins located on the surface of mammalian cells that participate in cell-cell adhesion and attachment to the extracellular matrix (53). Interactions between GAS products and integrins via the RGD sequence have recently been reported (9, 41, 58). Integrins also bind to collagen through a different mechanism based on the triple helical structure of collagen (50). Since all 50 strains of 21 different M types had the scl gene, this mechanism could be widespread in GAS. This type of interaction between the CL region of the Scl protein and extracellular matrix and integrins could promote adherence of GAS to host tissue and trigger host-cell signal transduction.

Our data also showed that Scl contributes to soft tissue pathogenesis in a mouse model of inoculation, but they did not reveal a molecular mechanism to explain this observation. We favor the idea that the Scl protein is not directly toxic to the host but, rather, participates in soft tissue disease as a consequence of delayed clearance by the host. Under this hypothesis, Scl-positive organisms are cleared more slowly than the mutant because wild-type bacteria are not recognized as for-

<table>
<thead>
<tr>
<th>M 28 (I)</th>
<th>M 1</th>
<th>M 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLTSKHNKLNLWNYGLTSA</td>
<td>MLTSKHNKLNLWNYGLTSA</td>
<td>MLTSKHNKLNLWNYGLTSA</td>
</tr>
<tr>
<td>AAVLLATGQGASSYVA</td>
<td>AAVLLATGQGASSYVA</td>
<td>AAVLLATGQGASSYVA</td>
</tr>
<tr>
<td>EVSTAAAMKRRQEQKRTKE</td>
<td>EVSTAAAMKRRQEQKRTKE</td>
<td>EVSTAAAMKRRQEQKRTKE</td>
</tr>
<tr>
<td>EKKVTRFRRERFSSWFG</td>
<td>EKKVTRFRRERFSSWFG</td>
<td>EKKVTRFRRERFSSWFG</td>
</tr>
<tr>
<td>RYNNISVYIADLSQTSRQL</td>
<td>RYNNISVYIADLSQTSRQL</td>
<td>RYNNISVYIADLSQTSRQL</td>
</tr>
<tr>
<td>AADVEEVEVAMGAK</td>
<td>AADVEEVEVAMGAK</td>
<td>AADVEEVEVAMGAK</td>
</tr>
<tr>
<td>TTAGVLAVTKRKENN</td>
<td>TTAGVLAVTKRKENN</td>
<td>TTAGVLAVTKRKENN</td>
</tr>
<tr>
<td>PFFTAAAVAVM</td>
<td>PFFTAAAVAVM</td>
<td>PFFTAAAVAVM</td>
</tr>
<tr>
<td>EVSSTTMSTSRRGGERETGAQ</td>
<td>EVSSTTMSTSRRGGERETGAQ</td>
<td>EVSSTTMSTSRRGGERETGAQ</td>
</tr>
<tr>
<td>GKDGAPGEKGEKGDRGETGAQGPVGPQGEAGK</td>
<td>GKDGAPGEKGEKGDRGETGAQGPVGPQGEAGK</td>
<td>GKDGAPGEKGEKGDRGETGAQGPVGPQGEAGK</td>
</tr>
<tr>
<td>GDIPGATGAQGPQGEAGK</td>
<td>GDIPGATGAQGPQGEAGK</td>
<td>GDIPGATGAQGPQGEAGK</td>
</tr>
<tr>
<td>PFFTAAAVAVM</td>
<td>PFFTAAAVAVM</td>
<td>PFFTAAAVAVM</td>
</tr>
<tr>
<td>TTAGVLAVTKRKENN</td>
<td>TTAGVLAVTKRKENN</td>
<td>TTAGVLAVTKRKENN</td>
</tr>
</tbody>
</table>

**FIG. 7.** Schematic representation of variation in the Scl protein sequence among three different M-type GAS strains. The amino acid sequence of the Scl protein from M1 (source strain), M6, and M28 (of pattern I [Table 1]) GAS strains is shown. The length of each Scl domain among all 50 GAS strains tested is shown as the number of amino acid residues. Glycine residues within the CL region are shown in boldface type. Polymorphic sites within the signal sequence and linker region are in boldface type and italicized. The signal sequence consists of 37 amino acids in all strains tested. In addition to the G31 V substitution in the Scl signal sequence of the M6 strain, the following amino acid replacements were found in other GAS strains: H7Y, L9I, L12Q, and T19I. The alanine (A256 in M1 strain) residue (shown of 30 amino acid with 10 GXX repeats was described previously in the GAS bacteriophage-encoded hyaluronidase (33). In addition, Erickson et al. (12) reported that a platelet aggregation-associated protein of Streptococcus sanguis has a sequence (PGEQGPK) with two GXX repeats. However, the roles of these proteins in virulence were not assessed with isogenic mutant strains and model systems. In human collagen, three chains entwine to form a right-handed superhelix (47). Three characteristics of the GXX region in Scl suggest that it is capable of forming a similar quaternary structure. First, Scl fulfills the absolute requirement that glycine occupy every third residue (only glycine can fit into the center of the triple helix without distorting it). Second, on the basis of computer modeling, the GXX region of Scl is predicted to form a coiled structure. Third, lysines and prolines essential for triple-helix stabilization together comprise ~31% (percent by frequency) of the nonglycine residues. Scl and host interactions. Our studies indicated that Scl contributes significantly to adherence of GAS to human epithelial cells grown in culture, thereby adding to the relatively large number of surface-associated bacterial molecules that contribute to this process (21). The scl gene is transcribed in the logarithmic phase of growth, suggesting that the adhesin is made early in the course of host-pathogen interaction. Other molecules that contribute to host adherence that are made early in the growth phase include M protein (5), M-like proteins (17), and serum opacity factor (7). Among nonprotein streptococcal adhesins, hyaluronic acid capsule (55) and lipoteichoic acid (21, 40) also participate in adherence of GAS. Although our studies show that Scl participates in adherence in vitro, we did not address the molecular mechanism whereby Scl contributes to GAS-host cell interaction. Inasmuch as many eukaryotic cells bind collagen through receptors expressed on cell surfaces (3, 61, 62), Scl may participate in colonization of the host by docking GAS to host cell receptors (60). One biochemically important example of a collagen ligand is integrins, a class of transmembrane proteins located on the surface of mammalian cells that participate in cell-cell adhesion and attachment to the extracellular matrix (53). Interactions between GAS products and integrins via the RGD sequence have recently been reported (9, 41, 58). Integrins also bind to collagen through a different mechanism based on the triple helical structure of collagen (50). Since all 50 strains of 21 different M types had the scl gene, this mechanism could be widespread in GAS. This type of interaction between the CL region of the Scl protein and extracellular matrix and integrins could promote adherence of GAS to host tissue and trigger host-cell signal transduction.

Our data also showed that Scl contributes to soft tissue pathogenesis in a mouse model of inoculation, but they did not reveal a molecular mechanism to explain this observation. We favor the idea that the Scl protein is not directly toxic to the host but, rather, participates in soft tissue disease as a consequence of delayed clearance by the host. Under this hypothesis, Scl-positive organisms are cleared more slowly than the mutant because wild-type bacteria are not recognized as for-
cign as rapidly as mutant cells. An alternative hypothesis is that wild-type organisms are more resistant to phagocytosis and subsequent killing by host cells.

Autoimmune issues. Infection with GAS can result in sequelae such as acute rheumatic fever, rheumatic heart disease, glomerulonephritis, and other manifestations. It is widely believed that some of these postinfection sequelae have an autoimmune component, and evidence has been presented to support this idea (10). There has also been considerable speculation that GAS-induced autoantibodies play a role in many other human diseases, although conclusive evidence has not been obtained (18). The common theme in these and other autoimmune diseases including rheumatoid arthritis and systemic lupus erythematosus is the pathology of collagen within diseased sites and the presence of anticollagen antibodies in patient sera (19). Moreover, autoimmune diseases can be induced in experimental animals after exposure to collagen (8). The discovery of a widely distributed extracellular GAS protein with structural features similar to human collagen may add another dimension to GAS-induced autoimmunity considerations. In addition, our study found that the scl gene in M3 organisms had a mutation that introduced a stop codon. This polymorphism would create a truncated protein that lacks the cell wall anchor motif and, hence, would be secreted free into the extracellular milieu. It is possible that the Scl protein made by the M3 strains would interact with the host in a fundamentally different fashion than GAS cell wall-anchored Scl. Additional experiments will be required to test this idea.

ACKNOWLEDGMENTS

We thank N. P. Hoe for helpful discussion.

This research was supported by Public Health Service grant AI-33119 to J.M.M., and by funds from the Moran Foundation to S.L.

REFERENCES


Editor: E. I. Tuomanen