Group G Streptococcal M Protein Exhibits Structural Features Analogous to Those of Class I M Protein of Group A Streptococci

CARLEEN M. COLLINS,1 ALAN KIMURA,1 AND ALAN L. BISNO2,3,*

Departments of Microbiology and Immunology1 and Medicine,2 University of Miami School of Medicine, and Medical and Research Services,3 Miami VA Medical Center, Miami, Florida 33125

Received 23 April 1992/Accepted 23 June 1992

We have previously studied a collection of group G streptococcal strains isolated from bacteremic human infections and demonstrated that such strains resist phagocytosis by human polymorphonuclear leukocytes but are type specifically opsonized by homologous antiserum. We have now performed Southern hybridization analysis on genomic DNA from eight blood isolates. All eight isolates showed DNA homology to a group A emm24 gene probe. The M-protein gene of one of the isolates, strain 1750, has now been isolated. This gene (emmG1) encodes a polypeptide of 67 kDa (MG1) which is reactive with antibodies to the partially purified M protein of the parent strain. The predicted amino acid structure of MG1 demonstrates significant identity with the carboxy terminus (C, D, and anchor domains) of M6 and M24 but only limited identity with the amino terminus (variable portion) of these group A M proteins. Southern hybridization of genomic DNA of the eight group G blood isolates with an emmG1 gene probe indicated there were at least four emm alleles associated with these strains. These studies indicate that M proteins of group G streptococci, like those of group A, are genetically heterogeneous. Moreover, MG1 appears to conform to the recently proposed class I structure of M-protein molecules and thus shares certain distinct structural features with the M proteins of well-established rheumatogenic group A streptococcal serotypes. Further comparison of the structures of group G and group A M proteins of throat and skin isolates may cast light on those configurations of the M protein molecules which are and are not critical for the expression of rheumatogenicity.

Group G streptococci were first identified by Lancefield and Hare (26) over a half century ago. Although these organisms are part of the normal microbial flora of the skin, pharynx, vagina, and gastrointestinal tract of humans, there has been increasing recognition in recent years that group G streptococci may cause life-threatening human infections. These have included bacteremia, infective endocarditis, suppurative thrombophlebitis, arthritis, osteomyelitis, empyema, peritonitis, endometritis, and neonatal sepsis (1, 12, 13, 16, 18, 25, 29, 37-39).

Little is known of the mechanisms by which group G streptococci exert their virulence. The group A streptococcus (Streptococcus pyogenes) is by far the most extensively studied of the pathogenic beta-hemolytic streptococci because of its association with acute rheumatic fever. The principal virulence factor of S. pyogenes is a cell wall constituent known as M protein. M protein is antiphagocytic, and antibodies to specific M serotypes are generally accepted as being the basis of acquired human immunity to group A streptococcal infections.

We have previously identified surface protein constituents analogous to the M proteins of group A streptococci in group G streptococcal strains isolated from life-threatening bacterial human infections (10). These organisms multiplied luxuriantly when rotated in fresh human blood but were readily phagocytosed and killed in the presence of antisera raised against partially purified M-protein preparations from the homologous group G strain. Other investigators have demonstrated biological and immunochemical identity between M proteins of group A and group G streptococci (22) and homology between group G streptococcal DNA and group A streptococcal gene probes from types 6 (31) and 12 (33).

Recent evidence has linked rheumatogenicity of group A streptococci to pharyngeal infection with specific M-protein serotypes (6-8, 35) exhibiting certain defined structural characteristics (4, 5, 23). However, despite the fact that group G organisms colonize the throat, at times cause clinically evident pharyngitis, and produce M proteins, they have never been reported to cause acute rheumatic fever.

In this study, we have isolated the M-protein gene of bacteremic group G streptococcal strain 1750 (emmG1) and determined its DNA sequence. On the basis of this DNA sequence, we predict a structure for the M protein encoded by emmG1 (MG1) and compare this structure with that of M6, the M protein of a rheumatogenic group A streptococcal serotype.

**MATERIALS AND METHODS**

**Bacterial strains and media.** Streptococcal strains were isolated from cultures of human blood submitted to the Boston City Hospital microbiology laboratory and have been previously described (10, 16). Escherichia coli DH5α [F<sup>-</sup> endA1 hsdR17 (rK<sup>−</sup> mK<sup>−</sup>) supE44 thi-1 λ<sup>−</sup> recA1 gyrA96 relA1 d80 lacZΔM15] was the host for all recombinant plasmids used in this study. The plasmid vector was pGEM7Zf(−), pGEM3Zf(−), or pGEM3Zf(+) (Promega Corp., Madison, Wis.), pBR322 (11), or pUC19 (42).

E. coli organisms were grown in L broth or on L broth solidified with 1.5% agar. Media were supplemented with

---

* Corresponding author.
carbenicillin (100 µg/ml) when necessary. Streptococcal strains were grown either in Todd-Hewitt broth or on 5% sheep blood agar plates.

**Serum opacity factor assay.** The assay was performed by the method of Top and Wannamaker (36).

**N-terminal amino acid analysis of strain 1750 M protein.** Peptic extraction of M protein was performed by the method of Beachey et al. (3) as previously described (10). The fraction of the extract which precipitated at 30 to 60% ammonium sulfate was dialyzed in phosphate-buffered saline (pH 7.4), lyophilized, reconstituted, and applied to 10% polyacrylamide gels containing 0.1% sodium dodecyl sulfate (SDS). The peptides present on the SDS-polyacrylamide gels were transferred electrophoretically to nitrocellulose paper and stained with Coomassie brilliant blue. Three major bands were observed, as previously reported (10) and illustrated in Fig. 4. The amino acid sequence of the most slowly migrating band was determined in the laboratory of Keith Brew (Department of Microbiology, University of Miami School of Medicine) by Edman degradation, using an automated protein sequencer (Applied Biosystems).

**Western immunoblot analysis.** Western blot analysis using whole-cell lysates of E. coli was performed as described previously (24). The blot was reacted with a 1:200 dilution of rabbit antiserum raised against the 30 to 60% ammonium sulfate fraction of pepsin-extracted M protein of strain 1750. This antiserum is type specifically opsonic for strain 1750. Specifically bound antibodies were then detected by incubation with a 1:1,000 dilution of goat anti-rabbit immunoglobulin G-peroxidase conjugate (Sigma) followed by development in a solution of 4-chloro-1-naphthol (0.5 mg/ml; Sigma) plus hydrogen peroxide (0.01%).

**Nucleic acid isolation and restriction.** Plasmid DNA was isolated from E. coli by alkaline lysis (27). Streptococcal genomic DNA was isolated by the method of Cleary et al. (14). Restriction enzymes were purchased from either New England Biolabs, Inc., Beverly, Mass., or Bethesda Research Laboratories, Inc., Gaithersburg, Md. Enzymatic reaction conditions were those recommended by the manufacturer.

**Construction of recombinant plasmids.** For the construction of p1750-1, strain 1750 genomic DNA was digested to completion with 0.1 U of PvuII, and the 0.3-kb fragments were isolated after electrophoresis through a 1% low-melting-point agarose gel. The purified fragments were ligated, using T4 DNA ligase (Bethesda Research Laboratories), to PvuII-digested, alkaline phosphatase-treated pBR322. The resulting E. coli DH5α transformants were screened by colony hybridization for homology to the emm24 gene probe (for description, see below).

To construct p1750-7, 2.6-kb XbaI fragments of strain 1750 genomic DNA were isolated from low-melting-point agarose and ligated to XbaI-digested, alkaline phosphatase-treated pUC19. The recombinant plasmid containing the emm24-homologous XbaI fragment was termed p1750-6. This 2.6-kb XbaI fragment was isolated from p1750-6 and inserted downstream of the 0.3-kb PvuII-to-XbaI fragment from p1750-1 in pGEM7Zf(−). The resulting plasmid containing the XbaI fragment in the proper orientation was named p1750-7.

**DNA sequence analysis.** DNA was sequenced by using the dideoxy method of Sanger et al. (30) as modified for use with Sequenase (U.S. Biochemical). Various p1750-7 subclones were constructed in pGEM7Zf(−), pGEM3Zf(−), or pGEM3Zf(+) (Promega). Single-stranded DNA templates derived from these subclones were sequenced by using oligonucleotide primers homologous to either the T7 or SP6 promoter encoded on the plasmids or primers designed to be internal to the cloned DNA fragment. Single-stranded DNA was prepared from inserts in pGEM vectors by using phage R408 as described by the manufacturer (Promega). [32P]dATP-labeled DNA fragments were separated by electrophoresis on 6% acrylamide gels and visualized by autoradiography.

**Southern hybridization.** Whole-cell streptococcal DNA digested with the indicated restriction enzymes was fractionated by agarose gel electrophoresis and transferred to nitrocellulose sheets (Schleicher & Schuell, Inc., Keene, N.H.) by the Southern technique (15, 34). After hybridization, the nitrocellulose sheets were washed in 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) with 0.1% SDS at 67°C. Probes were labeled with [32P]dCTP (Dupont, NEN Research Products, Boston, Mass.) by nick translation. Hybridization was revealed by autoradiography with XAR-Omat films (Eastman Kodak Co., Rochester, N.Y.) in the presence of an intensifying screen. To reprobe a blot, after autoradiography the nitrocellulose sheets were washed in 2.5 mM NaOH for 6 h at room temperature to remove the radioactive probe. Blots were then rinsed in 2× SSC and hybridized as described above.

The emm24 gene probe was isolated from pBR41-L3, a recombinant plasmid encoding the group A streptococcal M24 protein gene (emm24) (gift of R. W. Baird and J. Dale, University of Tennessee, Memphis). pBR41-L3 was digested with PvuII and HindIII, and the 372-bp fragment was isolated from low-melting-point agarose following electrophoresis.

**Nucleotide sequence accession number.** The GenBank accession number for the emmG1 sequence presented here is M95774.

**RESULTS**

Group G strains contain DNA homologous to group A M-protein genes. Previous studies on seven group G strains isolated from clinically severe bacteremic infections indicated that these strains expressed antiphagocytic surface proteins analogous to the M proteins from group A streptococci (14). We performed Southern hybridization analysis on genomic DNA from the seven strains plus one additional strain isolated from a human bacteremic infection. The DNA probe corresponded to the carboxy-terminal one-third of the group A M-protein gene emm24 (this probe will be referred to as the emm24 gene probe). This region is highly conserved in all M proteins studied thus far. Genomic DNA from all eight strains showed homology to the probe under stringent hybridization conditions (Fig. 1).

**Serum opacity factor.** All eight strains under study were opacity factor negative.

**Isolation of an M-protein gene from group G strain 1750.** An M-protein-encoding gene from strain 1750 was isolated by using the DNA homology seen between the emm24 gene probe and genomic DNA from group G strains. Group G strain 1750 genomic DNA was digested with various restriction endonucleases that have 6-bp recognition sites and hybridized to the emm24 gene probe. The probe recognized a 2-kb PvuII fragment and a 2.6-kb XbaI fragment of strain 1750 genomic DNA. PvuII fragments, 2 kb in size, were isolated from strain 1750 genomic DNA and ligated to Smal-digested pBR322, and the resulting recombinant plasmids were transformed into E. coli DH5α. One plasmid containing DNA homologous to the emm24 gene probe was
emmg1 encodes a group G strain 1750 immunoreactive polypeptide. Western analysis indicated that p1750-7 encodes polypeptides of approximately 60 to 68 kDa reactive with antibodies to the pepsin-extracted M-protein-like surface component from group G strain 1750 (Fig. 4, lane 5). (Presumably these various forms represent proteolytic degradation products of the same polypeptide.) There was no reactive material detected in any of the other E. coli extracts examined (Fig. 4, lanes 2 to 4).

The amino-terminal amino acid sequence of MG1. The amino-terminal sequence of the major polypeptide isolated from the pepsin-extracted surface material of strain 1750 was determined to confirm that emmg1 corresponded to our biologically active material. The sequence of the first 32 amino-terminal amino acids was Glu-Asn-Thr-Asp-Gln-Tyr-Tyr-Val-Lys-Lys. This sequence corresponds exactly with the predicted amino acid sequence of residues 42 to 53 of the emmg1-encoded polypeptide. The amino-terminal sequence at residue 42 suggested that the first 41 amino acids were removed from the mature protein, implicating the presence of a signal peptide.

Comparison of emmg1 with other M-protein genes. The GenBank data base was scanned for sequences homologous to that of emmg1 DNA. The emmg1 DNA sequence showed a high degree of identity to sequences of eight different streptococcal M-protein genes, including 67% nucleotide residue identity with emm12, encoding M protein type 12 from S. pyogenes; 61% nucleotide residue identity with emm24, encoding S. pyogenes M protein type 24; and 65% nucleotide residue identity with emm6, encoding S. pyogenes M protein type 6.

Comparison of the predicted amino acid sequence of MG1 with the amino acid sequences of various M proteins provided additional evidence that emmg1 encodes an M protein. There were two regions of homology between MG1 and the other M proteins. The first 41 amino acids of the predicted MG1 sequence have 74 and 76% amino acid residue identity with the amino acid signal peptides of M6 (Fig. 5A) and M24, respectively. Therefore, the first 41 amino acids of MG1 represent the signal peptide of the protein. This observation is consistent with the N-terminal amino acid sequence of mature MG1 originating at amino acid 42 (see above).

The second region of homology was the 208 amino acids at the carboxy terminus. This region of the predicted amino acid sequence of MG1 shows 94 and 86% identity with the carboxy-terminal amino acid regions of M6 (Fig. 5B to D) and M24, respectively. This terminal amino acid region encodes repeat C, repeat D, and the anchor domain of the M protein as defined by Hollingshead et al. (21).

In contrast to the signal peptide and the carboxy-terminal 208 amino acids comprising the C and D repeats and the anchor domain, the region of the protein composed of amino acids 42 to 384 (the start of the mature protein to the beginning of the C repeat) showed only limited amino acid identity with the corresponding regions from other M proteins. This region would encode the A- and B-repeat region of the protein. This region showed amino acid identities of 31% with M6 protein, 32% with M24 protein, 32% with M57 protein, and 35% with M5 protein. The identity seen in all cases was due to identity between the proteins in the
terminal 40 amino acids of this region. Therefore, it appears that the terminal 40 amino acids of the B repeat of these proteins are somewhat conserved among the different M types. There was only sporadic amino acid identity throughout the other portion of the region. Serotypic data indicate that MG1 is antigenically distinct from group A M-protein types (10). This difference in antigenicity can easily be attributed to differences in the A- and B-repeat regions.

MG1 contained a number of amino acid repeats in this variable region (e.g., ELDKKEQEL at positions 208 to 216, ELDQEKELET at positions 229 to 237, QKEKEL at positions 225 to 230, and QKEKEL at positions 232 to 237); however, MG1 did not contain the prominent tandem repeats found in the A- and B-repeat regions of M6 and M24.

As for other M proteins, the predicted structure of MG1 was predominantly alpha helical. By using the DELPHI protein secondary structure prediction and the amino acid sequence of mature MG1, the structure of native MG1 was estimated to be 83% alpha helix. The only portion of the molecule devoid of alpha-helical structure was the proline-glycine-rich wall domain. Fourier analysis of the predicted amino acid sequence of the mature protein indicated that the alpha-helical region had a strand periodicity of 1/6, strongly indicating a coiled-coil structure. M6 and other M-protein types have been shown to exist as coiled-coil structures.

Not all M-positive group G strains contain emmG1. Genomic DNA from the original group G strains, with the addition of the eighth isolate (strain 3125), was probed with the 0.8-kb HindIII-to-HindIII fragment from p750-7 (Fig. 6). This fragment codes for most of the variable region of MG1 and 68 amino acids of the C repeat. All eight strains contain DNA homologous to this emmG1 gene probe.

The hybridization profiles (Fig. 6) indicate that the emm genes of group G strains 1750 and 1737 share restriction sites and most probably represent the same M-protein type. Group G strains 1434, 1395, and 1454 appear to contain emm genes similar to each other but distinct from the emm genes of other strains. Group G strains 3125 and 1618 appear to contain homologous emm genes, and strain 1714 appears to have a unique profile. The hybridization profiles suggest that there are at least four distinct emm alleles associated with group G streptococci. These results are in accord with opsonophagocytic studies (9) indicating shared opsonic determinants between strains 1750 and 1737 and between strains 1395 and 1434.
### DISCUSSION

While previous studies from our laboratories and others (10, 22, 28, 31-33) indicate that group G streptococci encode and express M proteins, emmG1 is the first reported group G streptococcal M-protein gene to be cloned and sequenced. emmG1 was isolated on the basis of its homology to the terminal region of emm24, which encodes a portion of the C repeat, the D repeat, and the anchor domain of M24. There are a number of streptococcal proteins that contain a similar carboxy-terminal region, including M proteins, the immunoglobulin A-binding protein (20), and the immunoglobulin G-binding protein of group G streptococcus (protein G) (17). The nucleotide sequence of emmG1, along with initial hybridization data obtained by using a protein G gene probe (a generous gift of Genex Corp., Gaithersburg, Md.), indicated that we have not isolated the gene encoding protein G. We believe that emmG1 encodes an M protein, on the basis of (i) the immunoreactivity of the emmG1 protein product with type-specific opsonizing antibodies raised to MG1 and (ii) the predicted structure of the emmG1-encoded protein. The predicted amino acid sequence of MG1 indicates it to be a mostly alpha-helical protein (83%) with a periodicity suggestive of a coiled-coil structure. These two findings are characteristic of M proteins.
Southern hybridization using two different M-protein gene probes demonstrated that there are at least four alleles of the M-protein gene in group G streptococci. Thus, it is likely that there are multiple distinct group G M-protein serotypes. This observation is supported by functional studies of opsonophagocytosis (9), which confirm the existence of more than one group G M-protein serotype. It is not surprising that there are distinct group G M-protein serotypes, since there are more than 80 different serotypes of group A M-proteins. Further studies are needed to delineate the number of group G M-protein serotypes and to determine whether, similar to group A streptococci, certain group G serotypes are associated with specific host colonization sites.

The structural characteristics of group A streptococcal M protein have been the object of intense investigative scrutiny in recent years. This increased interest has been spurred in large measure by a reassessment of the association of M serotype and nonsuppurative streptococcal sequelae. It is now clear that strains of particular M serotypes are preferentially associated with acute rheumatic fever. That is to say, there exist rheumatogenic and nonrheumatogenic M types, just as there are nephritogenic and nonnephritogenic ones (6–8, 35). Moreover, distinct structural differences between the M proteins epidemiologically associated with rheumatic fever and poststreptococcal glomerulonephritis have been described (23).

Widdowson et al. (41) recognized antigenic differences in certain non-type-specific moieties (2) of M protein which they termed M-associated proteins (MAP) (40). On the basis of these distinctions, they classified the most common M serotypes into two groups, MAP I and MAP II. MAP I included those serotypes classically associated with acute pharyngitis, including most of the highly immunogenic, opacity factor-negative types epidemiologically associated with rheumatic fever. Certain serotypes gave variable results and could not be definitely assigned an MAP type.

The significance of these earlier studies has now been placed in clearer perspective by utilization of modern molecular biologic techniques to elucidate M-protein structure. M-protein molecules share tandemly arranged blocks of repeating sequences. The C-terminal portion of the M-protein molecule is highly conserved among different serotypes, while the N terminus is much more variable from type to type (19). Bessen et al. (4, 5) have proposed the existence of two major structural classes of M proteins. This proposal is based on analysis of epitopes within the C-repeat region of M-protein molecules representative of a number of distinct serotypes. The C-repeat region is located adjacent to the carboxy-terminal side of the pepsin cleavage site, and class

**FIG. 6.** Evidence that group G streptococcal strains contain distinct emm genes. Genomic DNA from eight group G streptococcal strains was digested with HindIII and hybridized under stringent conditions to the 0.8-kb HindIII fragment from pl1750-7. This probe encodes most of the variable region of strain 1750 emm. Lanes: 1, HindIII-digested pBR41-L3 (emm24-encoding plasmid); 2, group G strain 1750; 3, group G strain 1737; 4, group G strain 1434; 5, group G strain 1395; 6, group G strain 3125; 7, group G strain 1618; 8, group G strain 1454; 9, group G strain 1714; 10, streptococcal group B genomic DNA; 11, group A streptococcal genomic DNA from an emm24-encoding strain.
I molecules share a surface-exposed domain within this region. The serotype distribution and biologic characteristics of streptococcal strains within classes I and II of Bessen et al. (4, 5) conform closely to the earlier MAP I and MAP II scheme, suggesting that MAP may represent a product of the conserved region of the M-protein gene.

The classification proposed by Bessen et al. (4) was initially based on the differential immunoreactivity of various M proteins with two C-repeat-region-specific monoclonal antibodies. Subsequently, Bessen and Fischetti (5) compared in detail the amino acid sequences of the C-repeat regions of class I and class II isolates. They determined that the C repeats consist of 35 amino acids, of which 20 are highly conserved. Of the 15 variable residues, they found only 4 to be class specific. These workers found, however, as had Widdowson et al. (40, 41) (see above), that these distinctions were not absolute. One-fifth of class II isolates reacted weakly with one of the monoclonal antibodies used to define class I. Moreover, M protein of type 49, categorized as class II by its reactivity with the monoclonal antibodies, was found to share a high degree of sequence homology with the C-repeat region of class I M proteins.

Despite these inconsistencies, it seems clear that the M-protein serotypes most strongly epidemiologically associated with rheumatic fever (i.e., types 1, 3, 5, 6, 14, 19, 24, and 29) (6) all appear to have class I structures. This fact has raised the possibility that the class I configuration is intimately involved in the pathogenesis of rheumatic fever. However, the epidemiologic and structural data do not match perfectly, particularly in the case of M type 12 (class I but believed to be nonrheumatogenic). Thus, the predicted structure of an M protein derived from a streptococcal serogroup (group G) which has never been demonstrated to cause rheumatic fever is of considerable interest.

The homology of the C-repeat region of MG1 with that of the M proteins of type 6 (Fig. 5B) and other class I molecules appears to categorize MG1 as belonging to this class, as delineated by the criteria of Bessen et al. (4, 5). The first two blocks of the C-repeat region (residues 386 to 469) contain all 40 of the invariant amino acid residues (underlined in Fig. 5B) and 7 of the 8 class I-specific amino acid residues (indicated by triangles in Fig. 5B). Only the glutamic acid residue at position 442 is suggestive of a class II molecule. The fact that strain 1750 is opacity factor negative is further evidence of its class I structure (4). Moreover, we have now succeeded in cloning emm genes from two additional strains in our collection (1395 and 1454), and the predicted amino acid sequences of the M proteins from both strains include all eight of the class I-specific residues (29a).

Strain 1750 was isolated from an intravenous drug abuser with soft tissue infection and bacteremia, a circumstance which is in accord with the observation that the M proteins of streptococci causing the most severe forms of localized or invasive human infection possess class I structure. The link between this configuration and rheumatogenicity remains to be defined, however. Our studies of strain 1750, a virulent streptococcus belonging to a serogroup not known to initiate rheumatic fever, suggest that not all M-positive, class I streptococci possess rheumatogenic potential. Thus, if the class I structure is a prerequisite for rheumatogenicity, it is not likely in itself to be sufficient for this purpose.

It should be borne in mind, however, that 1750, as well as the other invasive strains used in our previous and current investigations, initiated disease from a cutaneous or soft tissue site (16). Further analysis of pharyngeal strains of group G streptococci will be instructive, because only pharyngeal infections with group G streptococci trigger rheumatic fever. Such analysis is complicated, however, by the fact that most throat isolates of group G streptococci are felt to be saprophytes rather than true pathogens. Nevertheless, additional studies along these lines with carefully selected clinical isolates are planned.

ACKNOWLEDGMENTS

This work was supported by a merit review program from the Department of Veterans Affairs.

The group G strains utilized in this study were provided by Donald E. Craven. We gratefully acknowledge the advice and assistance of Kent Miller, Richard Cowman, and Keith Brew. We thank Lillian Chiu and Heike Laman for expert technical assistance and Edith Decker and Gladys Dunning for invaluable secretarial and editorial assistance.

REFERENCES

16. Craven, D. E., A. I. Rixinger, A. L. Bisno, T. A. Goularte, and...


