Differentiation between Two Biologically Distinct Classes of Group A Streptococci by Limited Substitutions of Amino Acids within the Shared Region of M Protein-like Molecules

By Debra E. Bessen and Vincent A. Fischetti

From the Laboratory of Bacteriology and Immunology, The Rockefeller University, New York, New York 10021

Summary

Group A streptococci can be categorized into two classes (I and II) based on immunodeterminants contained within a surface-exposed, conserved region (C repeat domain) of the major virulence factor, M protein. Previous studies have shown that several biological properties correlate strongly with streptococcal class, and thus, there is a strong impetus to precisely define the antigenic epitopes unique to class I and II M proteins. Using synthetic peptides, the binding sites of two class I-specific mAbs were mapped to distinct epitopes within the C repeat region of type 6 M protein (class I). A class II M protein-like gene (type 2) was cloned and sequenced, and the predicted amino acid sequence was compared for homology to class I and II molecules, whose sequences were previously reported. For a given C repeat block 35 amino acid residues in length, 20 residue positions were conserved among all sequences analyzed. Of the 15 variable amino acid positions, only four were class specific, and three of the four positions were localized in the area to which the class I-specific mAbs bound. The predicted secondary structures of class I and II C repeat blocks reveals that they are α-helical, except for a single area of disruption. In the class I molecules, the area of disruption corresponds to the class I-specific mAb binding sites. Importantly, the predicted conformational characteristics of this disruption differs for class I and II molecules. The data suggest that only limited changes in amino acid residues differentiate between class I and II molecules in the C repeat region. Therefore, selective (biological) pressures may have contributed to the evolution of these two classes of molecules.

Group A streptococci contain a family of surface proteins that are structurally similar in their COOH-terminal halves. Included in this family of closely related molecules are M proteins (1–5), an IgA receptor protein (Arp) (6), and an IgG receptor (protein H) (7). One notable feature of M and M-like protein molecules is that they contain tandemly arranged blocks of direct sequence repeats, of which there are one to three distinct sets per molecule. The conserved, COOH-terminal half of M and M-like molecules is comprised of a surface-exposed "C repeat region," which forms a distinct antigenic domain (8), followed by a cell-associated portion (1, 9) (for an example, see Fig. 5). Antigenic epitopes on the NH_2-terminal side of the C repeat region are far less conserved among molecules of the M protein family (8). A related IgG receptor, FcRA76, displays significant sequence homology to M and M-like molecules in its COOH-terminal region, but lacks the C repeat domain (10).

M protein itself exists as a α-helical coiled-coil dimer giving rise to a fibrillar structure (11) whose NH_2-terminal end is distal to the cell surface and displays extreme antigenic diversity (12). M proteins provide the basis for the major serologic typing scheme of group A streptococci (13). They are the major virulence factor present on most clinical isolates, and function by rendering the organism resistant to phagocytosis. M protein is defined by its function as an antiphagocytic surface factor that can be neutralized by type-specific antiserum (13), however, there is no known structural feature that is common to all M proteins but absent from M-like molecules. Because of the existence of multiple closely related genes, one must rely on the functional definition of M protein before concluding that a particular gene product is or is not M protein. For example, it remains to be established whether in some instances, type-specific antibodies are immunoreactive with antigens other than the antiphagocytic M molecule, or alternatively, whether some M proteins have IgG or IgA receptor activity in addition to their antiphagocytic property.

Two C repeat region-specific mAbs (10F5 and 10B6) directed to M6 protein have been used for immunochemical analysis of the C repeat region of >130 group A streptococcal isolates representing 50 different serotypes (8). The study revealed that organisms can be grouped into two classes based

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/90/12/1757/08 $2.00
Volume 172 December 1990 1757-1764

on the presence or absence of immunoreactivity with the C repeat region–specific mAbs. Class I streptococci are defined by their strong binding of one or both of the mAbs, whereas class II organisms fail to bind either of these two particular mAbs or bind only weakly (8). Of the isolates defined as class I, 90% bound both mAbs strongly, whereas the remaining 10% bound either 10F5 or 10B6. All class II organisms analyzed were devoid of 10B6 reactivity, and 21% displayed weak binding by 10F5. The immunoreactivity of the C repeat region–specific mAbs with a given streptococcal isolate may be attributed to either M or M-like protein molecules, or both if multiple gene products are expressed. Immunoblot analysis of crude extracts of streptococcal surface antigens demonstrate the presence of M or M-like molecules in class II isolates, by virtue of their immunoreactivity with anti–whole M molecule (8) (unpublished findings). For this reason, the lack of binding of two class I–specific mAbs to class II organisms led to the initial impression that most class II M protein molecules completely lacked the C repeat region. However, one fact that remained puzzling at the time was that a partial amino acid sequence of a C repeat region of a class II M protein (M49) displayed ~70% sequence homology with class I M proteins (8, 14). More recently, the complete amino acid sequence of an M49 molecule (5) has revealed a high degree of sequence homology with class I M proteins (1–4) in the C repeat region. Because type 49 is one of the few class II serotypes displaying weak reactivity with the class I–specific mAb, 10F5, it remained unclear as to whether the M49 protein sequence was intermediary to the more typical class I and II molecules.

Antigenic epitopes within the C repeat regions of the M protein family correlate with several biological properties of group A streptococci (8, 15). Class II organisms are producers of an apoproteinase termed opacity factor (OF)[1] (16); Class I streptococci fail to exhibit OF activity (8, 17). Perhaps the most important correlation between streptococcal class and pathogenicity is the most "rheumatogenic" M types are class I. In addition, most class I impetigo isolates display human IgG-Fc receptor activity, whereas class I organisms from nasopharyngeal sites tend to lack IgG-Fc receptors (8, 15). In contrast, nearly all class II isolates display IgG-Fc receptor activity regardless of tissue site of infection (15). The nonimmune binding of human IgA by group A streptococci appears to be a class II–specific property (15). Thus, the class of the C repeat domain appears to be fundamental in distinguishing between group A streptococci associated with particular diseases.

The strong correlations between streptococcal class and pathogenicity have provided the impetus to better define the class I–specific epitopes located within the C repeat region of M protein. To identify those amino acids responsible for the class I–specific epitopes, we cloned and sequenced a class II M protein–like gene (*emm*L2.1) derived from a strain that was completely devoid of immunoreactivity with the class I–specific mAbs, unlike *emm*49. We provide evidence that within a narrow zone of the C repeat region, there exists both shared and class-specific immunodeterminants. The predicted amino acid sequence of the M2 protein–like C repeat domain was compared with that of other class I and II molecules for homology. The findings suggest that only a small proportion of the amino acid substitutions within the C repeat region define class specificity. The secondary structure predicted by Garnier plots indicates a single disruption in the α-helix, corresponding to the class I–specific site. Importantly, the nature of the disruption in the helix differs for class I and II molecules.

[1] *Abbreviation used in this paper:* OF, opacity factor.

Materials and Methods

Bacterial Strains. Group A streptococci were from The Rockefeller University collection and the Institute of Hygiene and Epidemiology, Prague. In addition, strains whose *emm* or *arp* gene had been previously cloned and sequenced were obtained from the following sources: T5 Manfredo (M type 5) and A24 Vaughn (M24) (Dr. James Dale, VA Medical Center, Memphis, TN), AP4 (M4) (Dr. Gunnar Lindahl, University of Lund, Sweden), and CS24 (M12) (Dr. Patrick Cleary, University of Minnesota, Minneapolis, MN). In the case of CS24, the high M protein–producing, opaque colony variant (18) was selected for study. Additional strains include T2/44/Rb4 (M2), D471 (M6), B737/137/1 (M49, from which strain CS101 was derived), 29452 (M22), 29486 (M33), and 29586 (T type T8/25/Imp19). All strains were tested for OF production (8); strains of M types 2, 4, 22, and 49, and T8/25/Imp19 were OF positive (data not shown). In addition, all strains under study lacked the ability to bind mouse IgG by a nonimmune mechanism (8).

Antibody Absorption Assay. This assay was performed according to previously described methods (8, 19, 20). Heat-killed streptococci were tested at twofold dilutions, beginning at an OD_{650} equivalent to 10. mAbs 10F5 and 10B6 were kindly provided by Dr. Kevin Jones (The Rockefeller University), and have been previously described (8, 21).

Cloning and Sequencing of the emmL2.1 Gene. A genomic library of strain T2/44/Rb4 (M type 2, class II) was prepared from mechanically sheared DNA ligated to λ gt11 through EcoRI linkers, and the DNA was packaged in phage particles (22). Plaques produced in *Escherichia coli* Y1090 were screened first with affinity-purified anti-ColiM6 antibody (ColiM6, the product of the *emm*6.1 gene produced in *E. coli*; [23]), and secondly with M2 typing sera (13). Phage from immunoreactive plaques were used to produce lysogens in *E. coli* Y1089. After lysogen induction at 42°C, bacteriophage particles containing the *emm*L2.1 gene were purified on CsCl gradients, DNA was extracted from phage particles, and the cloned insert was excised by EcoRI digestion and purified by agarose gel electrophoresis. The purified restriction fragment was ligated to RF M13mp19 phage (24), and clones containing overlapping inserts of ssDNA were generated by T4 polymerase digestion (25), using the Cyclone I Biosystem (I.B.I., New Haven, CT). ssDNA was purified from the M13 phage particles and sequenced by the dideoxy chain termination method (26) using Sequenase (United States Biochemical Corp., Cleveland, OH). Strands in opposite orientations were sequenced.

Western Blot Immunoanalysis. SDS-PAGE and electrophoretic transfers were performed according to previously described methods (8). M protein was extracted from group A streptococci using the muralytic enzyme, bacteriophage lysin, according to previously described methods (27). The binding buffer for incubation of antibodies with blots consisted of 0.1 M Tris (pH 8.1), 0.4 M NaCl,

0.5% Tween 20. Affinity-purified antipeptide Igs were prepared from hyperimmune rabbit serum according to previously described methods (8, 19, 20, 28). Antiserum was raised in rabbits to synthetic peptides corresponding in sequence to residues 240–260 and 272–292 of M6 protein from strain D471 (Fig. 5), and to ColiM6. M2 typing sera was prepared as previously described (13). Antibodies to ColiM6 and the M2 typing sera were absorbed against lysates of λ gt11–infected *E. coli* Y1089. Alkaline phosphatase-conjugated secondary antibodies directed to rabbit and mouse IgG were used.

Competition ELISA with Synthetic Peptide Inhibitors. This assay was performed according to previously described methods (28). Synthetic peptides (19) (Fig. 5) were tested at twofold dilutions in 0.1-ml volume beginning at a concentration of 50 μg/ml.

Sequence Analyses. Amino acid sequence analysis for secondary structure by the Robson predictive algorithm (Garnier plots) (29), and construction of the DNA consensus sequence (Staden software), were conducted on The Rockefeller University UNIX system. The Protalign program was used for determining amino acid sequence homologies.

Results

Immunoreactivity of Streptococci with the Class I–specific mAbs. Class I and II streptococci were compared for their ability to bind the two class I–specific mAbs by testing different concentrations of organisms for absorption of each mAb to surface-exposed antigens. The strains selected for study include six organisms whose sequence of an M or M-like molecule has been determined. One can determine the relative binding capacity of class I and II isolates by comparing the number of organisms required to bind 50% of a given antibody. Greater than 100-fold more class II organisms were required for 50% absorption of mAb 10F5 (Fig. 1 *A*). One exception was a type 24 strain, an atypical class I serotype in that it lacks the 10F5 site on its surface (8). Two class II strains (types 4 and 49) absorbed moderate amounts of 10F5, but only at very high concentrations of organisms. The mAb 10B6 binding capacities for class I and II organisms differed by at least 50-fold (Fig. 1 *B*).

Cloning and Sequencing of a Class II M Protein–like Gene. The overall aim of this study was to precisely identify the amino acids within the C repeat region that are responsible for distinguishing between class I and II streptococci. The amino acid sequences of several class I M proteins are known (1–4), and the complete sequence of a class II M49 protein has been recently reported (5). However, the type 49 strain from which the *emm*49 gene was derived is not entirely typical of a class II organism in that it displays some 10F5 reactivity (Fig. 1 *A*); the vast majority of class II isolates fail to bind 10F5 (8). Therefore, we cloned and sequenced the C repeat region of the 10F5- and 10B6-negative class II M protein-like molecule from type 2 strain T2/44/Rb4 (the *emm*L2.1 gene) (Fig. 1, *A* and *B*).

Phage from four plaques immunoreactive with M2 typing sera were used to produce lysogens in *E. coli* Y1089. Growing cultures were induced for synthesis of the cloned gene product, and whole cell lysates were subjected to Western blot immunoanalysis. The clones expressed a fusion product with

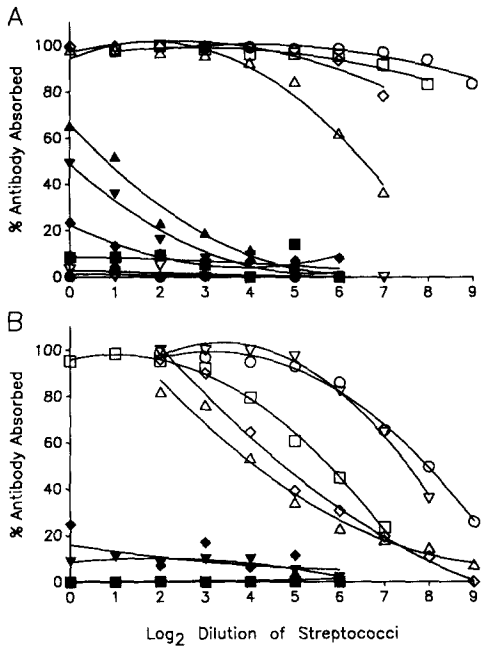

Figure 1. Absorption of class I–specific mAbs to class I and II streptococci. Streptococci were tested at twofold dilutions for absorption of 10F5 (*A*) and 10B6 (*B*). Class I streptococci (*open symbols*) are represented by types M5 (*circle*), M6 (*triangle*), M12 (*square*), M24 (*inverted triangle*), and M33 (*diamond*). Class II organisms (*closed symbols*) are represented by types M2 (*circle*), M4 (*triangle*), M22 (*square*), M49 (*inverted triangle*), and T8/25/Imp19 (*diamond*). (*B*) The symbols for M2, M4, and M22 overlap at 0 for all dilutions.

β-galactosidase, and displayed immunoreactivity with both anti-ColiM6 IgG and M2 typing serum (Fig. 2, *A* and *B*). In addition, immunoreactivity was observed with affinity-purified antibodies directed to synthetic peptides corresponding to sequences within the C repeat region of M6 protein (Fig. 2, *C* and *D*), but not with the class I–specific mAbs (Fig. 2, *E* and *F*). The size of the insert DNA from all four clones was ~1 kb (data not shown), and clone 9 was selected for DNA sequencing.

The nucleotide and predicted amino acid sequence of the C repeat region within the *emm*L2.1 gene is presented in Fig. 3. The M2 protein–like molecule contains four C repeat blocks, wherein each complete C block consists of a 23–amino acid residue C segment and a 12-residue spacer. Sequences immediately upstream from the C1 block are rich in glutamine and glutamic acid, and display strong homology with analogous positions within *emm*49 (5) and *arp*4 (6). The last repeat block (C4) contains the C segment only, and is followed by sequences that exhibit strong homology with all M and M-like molecules whose sequences are known (1–7) (data not shown). Because of its reactivity with the M2 typing serum and its high degree of sequence homology with other M proteins, we designate this gene as *emm*L2.1 (M2 protein like). Conclusive proof of whether this is, in fact, M2 protein will have to await functional tests for antiphagocytic activity.

Identification of Class-specific Amino Acid Substitutions by Sequence Homologies. The predicted amino acid sequences of

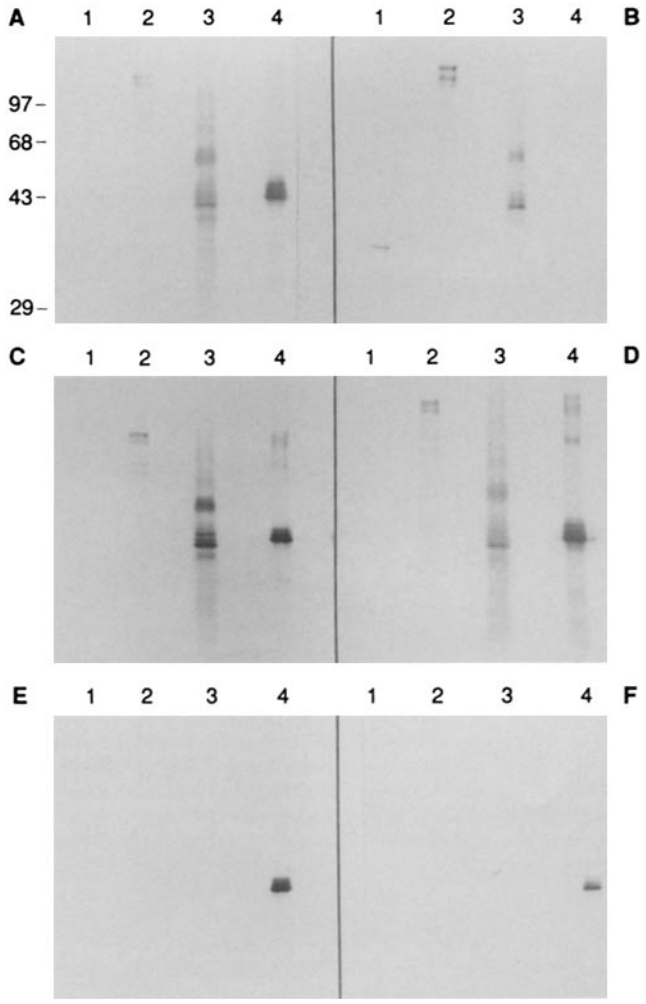

Figure 2. Western immunoblot analysis of lysogens. (Lane *1*) Whole cell lysates of *E. coli* Y1089 infected with λ gt11 and induced with IPTG; (lane *2*) as lane *1* except that λ gt11 contained the streptococcal insert DNA (clone 9); (lane *3*) crude lysin extract of parent strain T2/44/Rb4; (lane *4*) purified ColiM6. Blots were tested for immunoreactivity with the following antibodies: (*A*) anti-ColiM6; (*B*) M2 typing sera; (*C*) anti-pep(272–292); (*D*) anti-pep(240-260); (*E*) 10F5; (*F*) 10B6. Nonimmune rabbit serum present in excess failed to detect any material (data not shown). 1/10 the antigen load was used in *A* (for ColiM6 only) and in *B* (for T2/44/Rb4 lysin extract only).

```
GAAGTAGAAAAACGTTATCAAGAACAACTCCAAAAACAACAACAATTAGAAAAAGAAAAG
GLUVALGLULYSARGTYRGLNGLUGLNLEUGLNLYSGLNGLNGLNLEUGLULYSGLULYS
                          10                                  20

C1
CAAATCTCAGAAGCTAGCCGTAAGAGCCTAAGGCGTGACCTTGAAGCTTCTCGTGCAGCT
GLNILESERGLUALASERARGLYSSERLEUARGARGASPLEUGLUALASERARGALAALA
                          30                                  40

      C1|S1                                   S1|C2
AAAAAAGACCTTGAAGCTGAGCACCAAAAACTCAAAGAGGAAAAACAAATCTCAGAAGCA
LYSLYSASPLEUGLUALAGLUHISGLNLYSLEULYSGLUGLULYSGLNILESERGLUALA
                          50                                  60

                                                  C2|S2
AGCCGTAAGAGCCTAAGGCGTGACCTTGAAGCTTCTCGTGCAGCTAAAAAAGACCTTGAA
SERARGLYSSERLEUARGARGASPLEUGLUALASERARGALAALALYSLYSASPLEUGLU
                          70                                  80

                      S2|C3
GCTGAGCACCAAAAACTCAAAGAGGAAAAACAAATCTCAGAAGCAAGCCGTCAAGGTCTA
ALAGLUHISGLNLYSLEULYSGLUGLULYSGLNILESERGLUALASERARGGLNGLYLEU
                          90                                  100

                                              C3|S3
AGCCGTGACCTTGAAGCTTCTCGTGCAGCTAAAAAAGACCTTGAAGCTGAGCACCAAAAA
SERARGASPLEUGLUALASERARGALAALALYSLYSASPLEUGLUALAGLUHISGLNLYS
                          110                                 120

            S3|C4
CTCAAAGAGGAAAAACAAATCTCAGAAGCAAGCCGTCAAGGTCTAAGCCGTGACCTTGAA
LEULYSGLUGLULYSGLNILESERGLUALASERARGGLNGLYLEUSERARGASPLEUGLU
                          130                                 140

                      C4
GCGTCTCGCGAAGCTAAGAAAAAAGTAGAAGCAGACTTAGCCGAAGCAAATAGCAAACTT
ALASERARGGLUALALYSLYSLYSVALGLUALAASPLEUALAGLUALAASNSERLYSLEU
                          150                                 160

CAAGCCCTTGAAAAACTAAACAAA
GLNALALEUGLULYSLEUASNLYS
```

Figure 3. Nucleotide and predicted amino acid sequences of the C repeat region of the *emm*L2.1 gene. Numbers refer only to nucleotide sequence shown, and do not reflect the entire *emm*L2.1 gene. The sequences of the four C repeat blocks and 60 sequences flanking each side are presented. These sequence data are available from EMBL/GenBank/DDJB under accession number X56398.

the C repeat regions of seven class I and II molecules (M and the M-like Arp proteins) are presented in Fig. 4. The sequences shown are those that are currently available for analysis; they are not necessarily representative of the most common class I and II C repeat regions (8), because 10% of class I isolates completely lack binding by one of the class I–specific mAbs (e.g., M24), and one-fifth of class II strains display weak 10F5 reactivity (e.g., M4 and M49; Fig. 1 *A*). Bearing in mind the limitations of mAb immunoreactivity, the sequences presented allow one to draw several conclusions regarding class-specific amino acid residues.

Each complete C repeat block consists of a 23-residue C segment and a 12- or 19-residue spacer (Fig. 4). If one considers the 35 residues that comprise the 23-mer C segment and the 12-mer spacer, 20 of the 35 positions are largely invariant (*boxed*). Of the 15 variant positions, only four are involved in class-specific changes (the atypical, 10F5-deficient M24 protein is excluded from this aspect of the overall analysis [Fig. 1 *A*] [8]). The change in Asp-15 to Glu is class specific, as is Gln-23 substituted for amino acid X (usually Asp or Lys). In addition, class II organisms differ from class I in that each C repeat unit has either Gly-9 or Arg-11 changed to Ser, or both changes. The data suggest that despite a moderate measure of variability in the C repeat region, only a small proportion of the variable amino acids are related to class-specific changes.

Identification of Class-specific Amino Acid Substitutions Involved in Binding by the Class I–specific mAbs. A series of overlapping peptides were synthesized that correspond in sequence to the C repeat region of M6 protein (Fig. 5). The peptides were tested for their ability to inhibit the binding of 10F5 and 10B6 to ColiM6 antigen by ELISA (Table 1). The data indicate that pep(240–260) is a strong inhibitor of both 10F5 and 10B6 binding, whereas pep(272–292) strongly inhibits

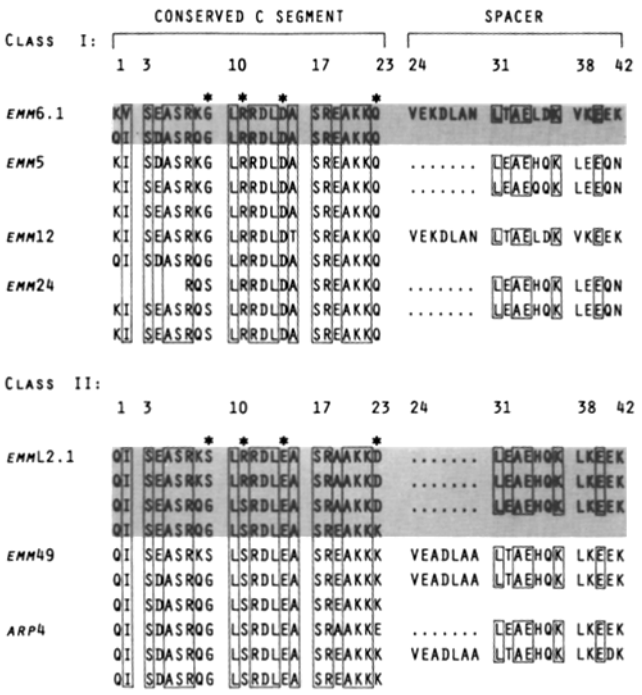

Figure 4. Sequence homologies between the repeat blocks of class I and II molecules. We offer a modified perspective of the C repeat region, which better illustrates the major points of this report. C repeat regions are depicted as 23-residue C segments (positions 1–23) separated by 12- or 19-residue spacer sequences (positions 31–42 or 24–42, respectively). One C segment and the spacer that follows constitute a complete C block (either 35 or 42 residues in length), whereas the last C block is incomplete, containing the C segment only; immediately COOH terminal to the last C block, there is a high degree of sequence homology among all M types studied. The amino acid sequences of the C repeat regions of three class I (M6, M5, M12), one 10F5-deficient class I (M24), and three class II (M2-like, M49, Arp4) molecules are presented. The amino acid residues presented for each molecule are as follows: M6, 235–299 (1); M5, 258–350 (3); M12, 2386–2581 (nucleotide positions) (2); M24, 311–397 (4); M2-like, 21–148 (this study); M49, 509–825 (nucleotide positions) (5); Arp4, 124–223 (6). The 20 invariant positions within the 35-mer block only (positions 1–23 and 31–42) are boxed; a position is considered invariant if all C blocks shown contain the same residue, allowing for one exception. The class-specific positions are indicated (*). The C repeat positions (1–42) of synthetic peptides are indicated in Fig. 5. Residues are grouped in a manner as to highlight the seven-residue periodicity characteristic of coiled-coil molecules (12).

10F5 binding only. Thus, 10F5 and 10B6 have distinct binding specificities. It should be noted that pep(240–260) and pep(272–292) contain partial repetitive blocks that display extensive sequence homology to each other (Fig. 5). Previous studies have been shown that a peptide fragment generated by V8 protease digestion of ColiM6, corresponding to amino acid residues 275–295, inhibited both 10F5 and 10B6 binding (21). Thus, it appears that residues 293–295 (repeated at positions 251–253) are critical for 10B6 binding.

Based on the immunoreactivities of the type 2 and 6 gene products (Fig. 2), it is possible to relate amino acid sequences to antigenic epitopes (Fig. 4). Both the type 2 streptococcus and the *emm*L2.1 gene product completely lack 10F5 and 10B6

Table 1. *Competitive Inhibition of Binding by Class I–specific mAbs to M6 Protein by Synthetic Peptides*

Peptide	10B6	10F5
	μg/ml	
223–244	>50	>50
240–260	2.85×10^{-3}	1.17×10^{-4}
248–269	>50	>50
256–277	>50	14.9
272–292	>50	5.87×10^{-5}

Figures represent the amount of peptide needed for 50% inhibition.

reactivity. By examining the sequence homologies between the M6 and M2-like proteins in the regions to which the class I–specific mAbs map, it becomes apparent that the variant position 15, and either position 9 or 11, are critical for mAb binding.

The 10F5-reactive peptides, pep(240–260) and pep(272–292), have positions 6–16 of M6 protein in common, with the exception of a single substitution at position 8 (Gln or Lys) (Fig. 4). The first two C blocks of the M2-like protein differ from the M6 sequence within the 10F5 binding region at position 9 (Gly in type 6, compared with Ser in type 2). Although the third and fourth C blocks of the M2-like protein have Gly-9, similar to the M6 protein, these two blocks have a Ser in place of the class I–specific Arg-11. In addition, the type 2 protein has a Glu instead of Asp-15. The 10F5-deficient type 24 organism has an M protein that differs from the M6 sequence in this region at Ser-9 only, adding further support that Gly-9 is a critical residue for the 10F5 epitope. The low level binding of 10F5 by type 4 and 49 organisms (Fig. 1 *A*) may be due to low affinity sites on the *emm*49 and *arp*4 gene products, or due to other M-like proteins. In summary, the data strongly suggest that substitutions in Gly-9 → Ser or Arg-11 → Ser (perhaps in combination with Asp-15 → Glu) result in loss of the 10F5 binding site.

Based on peptide inhibition experiments (Table 1) (21), the 10B6 mAb epitope lies within positions 6–19 of M6 protein. Since pep(272–292) (which ends at position 16) was noninhibitory for 10B6 binding, whereas the 275–295 fragment was inhibitory (21), it seems likely that the class I–specific Asp-15 is critical for 10B6 binding. Because type 24 organisms bound 10B6 strongly (Fig. 1 *B*) it is highly improbable that a change in Gly-9 by itself can affect 10B6 reactivity. However, Gly-9 or Arg-11 might influence 10B6 binding in combination with Asp-15. Interestingly, the critical change at position 15 from Asp to Glu is a conserved substitution, with preservation of an acidic amino acid residue. In summary, of the four class-specific positions [9, 11, 15, 23] within the C repeat region, the three residues at positions 9, 11, and 15 appear to be involved in determining the 10B6 and 10F5 binding sites.

Predicted Secondary Structure of Class I and II C Repeat Regions. The Robson algorithm for predicting secondary

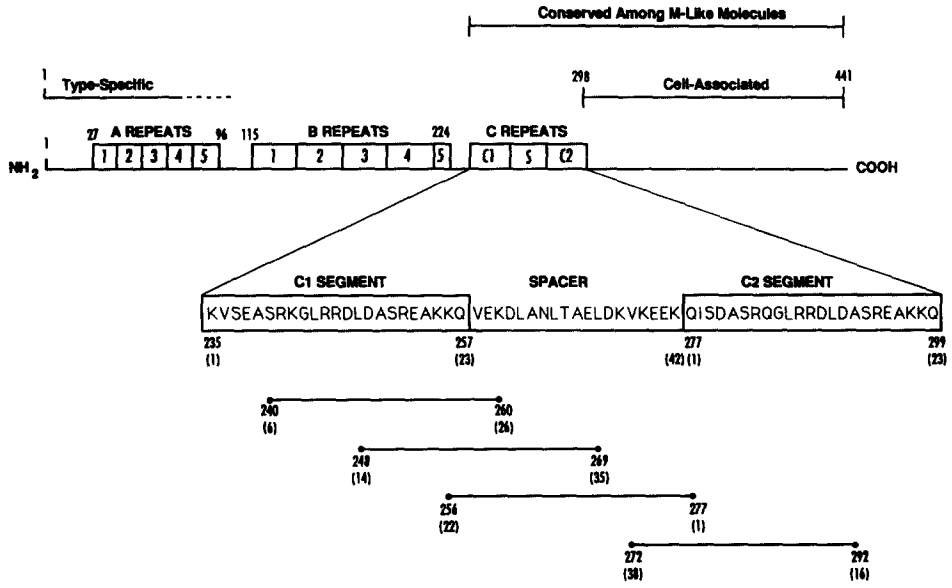

Figure 5. Structure of the M6 protein of strain D471. The C repeat region of M6 protein is depicted as 23-residue segments (C1 and C2; positions 1–23) separated by a 19-residue spacer sequence (S1; position 24–42). C1 and S1 constitute a complete C block, whereas the last C block (C2) is incomplete, containing the 23-residue segment only. Synthetic peptides corresponding to sequences within the C repeat region (used for competitive inhibition studies [Table 1] and antipeptide antibodies [Fig. 2]) are indicated; amino acid residue numbers and C repeat positions (parentheses) are shown. Amino acid residues are based on the sequence of the mature M6 protein of strain D471 (1).

structure (Garnier plots) (29) was used for the seven class I and II molecules described above (Fig. 6). For all of the molecules under study, most of each C repeat unit was α-helical. However, there was a single area of disruption in the α-helix, corresponding approximately to positions 7 to 13 (*boxed*). The class I molecules (with the exception of 10F5-negative M24 protein) have predicted β-turns and β-sheet in the disruption, with the basic structure TTTCSST or TTTSSST. In contrast, the disruption in the helix of class II molecules is largely the result of random coil potential and is devoid of β-sheet, having the structure TTCCCCC occurring at least once in each molecule, and in some blocks is as short as TTC (e.g., M2-like protein). Therefore, only in the region to which the class I–specific mAbs map is there a different predicted secondary structure for class I and II molecules.

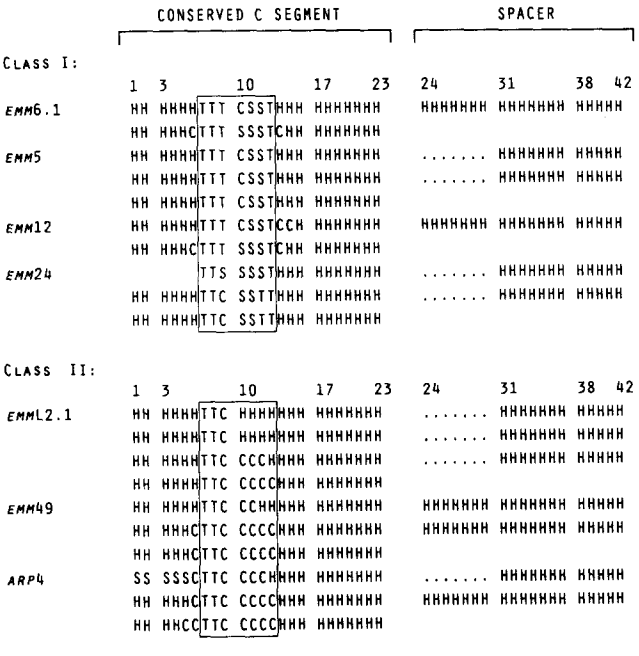

Figure 6. Garnier plots of the class I and II C repeat regions. Algorithms for predicted secondary structures were performed on the entire sequence of each molecule, but only the C repeat region is shown here. For M2-like protein, Garnier plots were performed on the sequence presented in Fig. 3. Each position is assigned a single structure: α-helix (*H*), β-sheet (*S*), β-turn (*T*), and random coil (*C*).

Discussion

The C repeat domain appears to be fundamental in distinguishing between group A streptococci of differing pathogenic properties (8, 15). In this report, we propose to define the molecules of the "M protein family" as those with closely related COOH-terminal halves, including a C repeat domain. Through knowledge of the differences between class I and II C repeat regions, we hope to gain further insight on the distinct biological functions of these two classes of streptococci. The data indicate that only 20 residues within the 35-mer C repeat block are invariant among the class I and II molecules analyzed, and furthermore, the degree of sequence homology among class I C repeat regions approximates the homology between class I vs. II. Only 4 of the 15 variant positions are involved in class-specific changes, and three of the four map to the class I–specific mAb epitopes. Thus, limited amino acid changes are responsible for distinguishing between class I and II C repeat regions.

The class I–specific mAb binding sites correspond to the single break in the α-helix predicted by Garnier plots. The disruption in the α-helix may represent an immunodominant region, due to greater flexibility and accessibility of the polypeptide chain. In general, predictive algorithms for secondary structure cannot be relied upon heavily, although the Robson-Garnier analyses are well suited for molecules rich in α-helical

content (29, 30). In the case of the C repeat domain, the single disruption in the predicted α-helical structure was located precisely at the positions responsible for both the class-specific amino acid changes and class I–specific mAb binding sites. Most importantly, the content of β-turn, β-sheet, and random coil at this location differed for the class I and II molecules. The coincidence of the disruption in the helix with the class-specific epitopes adds a large measure of significance to the predicted secondary structures.

How class I and II M proteins evolved is not readily apparent. The small number of class-specific amino acids within the C repeat domain, a region that contains far more non–class-associated substitutions, suggests that the class-specific changes are fairly stable. This, in turn, raises the possibility that there exists a selective pressure that acts to preserve the class-specific determinants. Therefore, the class-specific sites would not simply be a marker but rather, they would represent a region having a class-specific, biological function. Current studies indicate that the complement control protein, factor H, binds to the C repeat region of M6 protein (Fischetti et al., manuscript in preparation); whether the class I or II C repeat domains differ in their factor H binding capacity is not yet known. In so far as the biological functions that correlate with streptococcal class (rheumatogenicity, OF production, and IgA- and IgG-Fc receptor activities [8, 15], there is no established role for the C repeat domain.

Intragenic recombination occurs between repeat blocks of a given M protein molecule at an estimated frequency of 1 in 2,000 colony-forming units (31). Whether intergenic recombination occurs within the COOH-terminal halves of M and M-like molecules remains to be established. Within a single molecule, the C repeat region can be found in combination with antiphagocytic, IgG, or IgA receptor activities. One interesting possibility is that either class I or II C repeat domains occur only in combination with particular NH_2 termini where, at least in the M protein molecule, the type-specific and antiphagocytic properties reside. Thus, each M protein–like molecule may consist of a unique patchwork of distinct domains that ultimately confers a unique pathogenic potential to the organism.

We thank Dr. Kevin Jones for providing mAbs 10F5 and 10B6. We also thank those investigators who provided strains, Dr. Emil Gotschlich for technical advice, and Christopher Dickson and Frank Butera for their technical assistance.

This work was supported by a Grant-in-Aid from the American Heart Association (New York City affiliate) and U.S. Public Health Service grant AI-28944 (D. Bessen), and by a grant from Institut Merieux and U.S. Public Health Service grant AI-11822 (V.A. Fischetti).

Address correspondence to Debra E. Bessen, Box 276, The Rockefeller University, 1230 York Avenue, New York, NY 10021.

Received for publication 25 August 1990.

References

1. Hollingshead, S.K., V.A. Fischetti, and J.R. Scott. 1986. Complete nucleotide sequence of type 6 M protein of the group A streptococcus: repetitive structure and membrane anchor. *J. Biol. Chem.* 261:1677.
2. Robbins, J.C., J.G. Spanier, S.J. Jones, W.J. Simpson, and P.P. Cleary. 1987. *Streptococcus pyogenes* type 12 M protein regulation by upstream sequences. *J. Bacteriol.* 169:5633.
3. Mouw, A.R., E.H. Beachey, and V. Burdett. 1988. Molecular evolution of streptococcal M protein: cloning and nucleotide sequence of type 24 M protein gene and relation to other genes of *Streptococcus pyogenes*. *J. Bacteriol.* 170:676.
4. Miller, L., L. Gray, E.H. Beachey, and M.A. Kehoe. 1988. Antigenic variation among group A streptococcal M proteins: nucleotide sequence of the serotype 5 M protein gene and its relationship with genes encoding types 6 and 24 M proteins. *J. Biol. Chem.* 263:5668.
5. Haanes, E.J., and P.P. Cleary. 1989. Identification of a divergent M protein gene and an M protein related gene family in serotype 49 *Streptococcus pyogenes*. *J. Bacteriol.* 171:6397.
6. Frithz, E., L.-O. Heden, and G. Lindahl. 1989. Extensive sequence homology between IgA receptor and M protein in *Streptococcus pyogenes*. *Mol. Microbiol.* 3:1111.
7. Gomi, H., T. Hozumi, S. Hattori, C. Tagawa, F. Kishimoto, and L. Bjorck. 1990. The gene sequence and some properties of protein H. *J. Immunol.* 144:4046.
8. Bessen, D., K.F. Jones, and V.A. Fischetti. 1989. Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. *J. Exp. Med.* 169:269.
9. Pancholi, V., and V.A. Fischetti. 1988. Isolation and characterization of the cell-associated region of group A streptococcal M6 protein. *J. Bacteriol.* 170:2618.
10. Heath, D.G., and P.P. Cleary. 1989. Fc-receptor and M protein genes of group A streptococci are products of gene duplication. *Proc. Natl. Acad. Sci. USA.* 86:4741.
11. Phillips, G.N., P.F. Flicker, C. Cohen, B.N. Manjula, and V.A. Fischetti. 1981. Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc. Natl. Acad. Sci. USA.* 78:4689.
12. Fischetti, V.A. 1989. Streptococcal M protein: molecular design and biological behavior. *Clin. Microbiol. Rev.* 2:285.
13. Lancefield, R.C. 1962. Current knowledge of the type specific M antigens of group A streptococci. *J. Immunol.* 89:307.

14. Khandke, K.M., T. Fairwell, A.S. Acharya, B.L. Trus, and B.N. Manjula. 1988. Complete amino acid sequence of streptococcal PepM49 protein, a nephritis associated serotype: conserved conformational design among sequentially distinct M protein serotypes. *J. Biol. Chem.* 263:5075.
15. Bessen, D., and V.A. Fischetti. 1990. A human IgG receptor of group A streptococci is associated with tissue site of infection and streptococcal class. *J. Infect. Dis.* 161:747.
16. Saravani, G.A., and D.R. Martin. 1990. Opacity factor from group A streptococci is an apoproteinase. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett.* 68:35.
17. Widdowson, J.P. 1980. The M-associated protein antigens of group A streptococci. *In* Streptococcal Diseases and the Immune Response. S.E. Read and J.B. Zabriskie, editors. Academic Press, New York. 125–147.
18. Simpson, W.J., and P.P. Cleary. 1987. Expression of M type 12 protein by a group A streptococcus exhibits phaselike variation: evidence for coregulation of colony opacity determinants and M protein. *Infect. Immun.* 55:2448.
19. Bessen, D., and V.A. Fischetti. 1990. Synthetic peptide vaccine against mucosal colonization by group A streptococci. I. Protection against a heterologous M serotype with shared C repeat region epitopes. *J. Immunol.* 145:1251.
20. Bessen, D., and V.A. Fischetti. 1988. Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. *Infect. Immun.* 56:2666.
21. Jones, K.F., S.A. Khan, B.W. Erickson, S.K. Hollingshead, J.R. Scott, and V.A. Fischetti. 1986. Immunochemical localization and amino acid sequence of cross-reactive epitopes within the group A streptococcal M6 protein. *J. Exp. Med.* 164:1226.
22. Young, R.A., and R.W. Davis. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. *Science (Wash. DC).* 222:778.
23. Fischetti, V.A., K.F. Jones, B.N. Manjula, and J.R. Scott. 1984. Streptococcal M6 protein expressed in *Escherichia coli*. Localization, purification, and comparison with streptococcal-derived M protein. *J. Exp. Med.* 159:1083.
24. Messing, J. 1983. New M13 vectors for cloning. *Methods Enz.* 101:20.
25. Dale, R.M.K., B.A. McClure, and J.P. Houchins. 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. *Plasmid.* 13:31.
26. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
27. Fischetti, V.A., K.F. Jones, and J.R. Scott. 1985. Size variation of the M protein in group A streptococci. *J. Exp. Med.* 161:1384.
28. Jones, K.F., and V.A. Fischetti. 1988. The importance of the location of antibody binding on the M6 protein for opsonization and phagocytosis of group A M6 streptococci. *J. Exp. Med.* 167:1114.
29. Garnier, J., D.J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97.
30. Cohen, C., and D.A.D. Parry. 1990. Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. *Proteins Struct. Funct. Genet.* 7:1.
31. Hollingshead, S.K., V.A. Fischetti, and J.R. Scott. 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol. Gen. Genet.* 207:196.