PROTEINS: Structure, Function, and Genetics 3:60–69 (1988)

Conformational Characteristics of the Complete Sequence of Group A Streptococcal M6 Protein

Vincent A. Fischetti,[1] David A.D. Parry,[2] Benes L. Trus,[3] Susan K. Hollingshead,[4] June R. Scott,[4] and Belur N. Manjula[1]

[1]*The Rockefeller University, New York, New York 10021;* [2]*Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand;* [3]*Computer Systems Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892;* [4]*Department of Microbiology, Emory University, Atlanta, Georgia 30322*

ABSTRACT **M protein is considered a virulence determinant on the streptococcal cell wall by virtue of its ability to allow the organism to resist attack by human neutrophils. The complete DNA sequence of the M6 gene from streptococcal strain D471 has allowed, for the first time, the study of the structural characteristics of the amino acid sequence of an entire M protein molecule. Predictive secondary structural analysis revealed that the majority of this fibrillar molecule exhibits strong alpha-helical potential and that, except for the ends, nonpolar residues in the central region of the molecule exhibit the 7-residue periodicity typical for coiled-coil proteins. Differences in this heptad pattern of nonpolar residues allow this central rod region to be divided into three subdomains which correlate essentially with the repeat regions A, B, and C/D in the M6 protein sequence. Alignment of the N-terminal half of the M6 sequence with PepM5, the N-terminal half of the M5 protein, revealed that 42% of the amino acids were identical. The majority of the identities were "core" nonpolar residues of the heptad periodicity which are necessary for the maintenance of the coiled coil. Thus, conservation of structure in a sequence-variable region of these molecules may be biologically significant. Results suggest that serologically different M proteins may be built according to a basic scheme: an extended central coiled-coil rod domain (which may vary in size among strains) flanked by functional end domains.**

Key words: coiled-coil, alpha-helix, antiphagocytic, heptad, antigenic variation, sequence repeats, cell wall protein, intermediate filaments, myosin, tropomyosin

INTRODUCTION

The ability of the group A streptococcus to cause human infection is primarily attributed to the M protein molecule, a fibrillar structure on the streptococcal surface.[1] Because of its antiphagocytic property, the M protein enables the organism to resist attack by human neutrophils.[2] Although type-specific antibodies to the M molecule allow the effective clearance of the invading streptococcus, the ability of the organism to vary the antigenic structure of this protein has likely been responsible for the survival of the streptococcus in nature.[3] In addition to the more than 80 distinct M serotypes that have been identified to date, a large number of streptococcal strains have been isolated for which typing sera are not as yet available (WHO meeting, Tokyo, Japan, 1985).[4]

Structural, immunochemical, and physicochemical studies have demonstrated that the M molecule is composed of two predominantly alpha-helical protein chains assembled in a coiled-coil fibrous conformation extending nearly 60 nm from the surface of the streptococcal cell.[1,5] Most of the available information concerning this structure for the M protein, however, is based on pepsin-derived fragments (PepM) comprising the N-terminal half of the molecule[1,5–7]—the segment most distal from the cell surface.[1] A detailed analysis of the protein sequence of the M5 pepsin fragment (PepM5) revealed that, but for a short N-terminal segment, this fragment has the structural features of alpha-helical coiled-coil proteins.[6,7] In addition, this half of the native M5 molecule may be divided into two distinct regions based on the distribution of nonpolar and charged residues.

The complete DNA sequence of the M6 structural gene, has, for the first time, allowed the characteristics of the amino acid sequence of an entire M molecule to be studied.[8] The sequence suggests the presence of a membrane anchor region at the C-terminus preceded by a region rich in proline and glycine residues that may be responsible for the stabilization of the protein in the streptococcal cell wall. The majority of the remainder of the M6 molecule (80%) is made up of three regions showing repeats both at the DNA and amino acid level.[8] Region A is composed of five tandem repeat blocks of 14 amino acids each, region B of five tandem repeats of 25 amino acids each, and region C of two non-tandem repeats of 27 amino acids each.[8] Sequence repeat

Received August 4, 1987; accepted October 5, 1987.

Address reprint requests to Vincent A. Fischetti, Ph.D., The Rockefeller University, 1230 York Avenue, New York, NY 10021.

Susan K. Hollingshead is now at the Department of Microbiology, University of Alabama, Birmingham, AL 35294.

regions have also been reported to occur in the partial amino acid sequence of the PepM24 protein[9,10] and the complete amino acid sequence of the M5 pepsin fragment,[7] M5 being the least repetitive of the three.

Besides antigenic variation, M protein also varies among strains in apparent molecular size.[11] M6 protein size mutants have been found at a frequency of about 1 in 2,000 colony-forming units (CFU) in a laboratory grown culture of strain D471.[12] Sequence analysis of the M gene in these size mutants revealed that the size variation is due to homologous recombination events between the reiterated blocks in repeat regions of the molecule.[13]

DNA hybridization analysis of M protein gene[14] combined with monoclonal antibody studies[15,16] revealed that regions within the C-terminal half of the M molecule are conserved among many M protein serotypes and those in the N-terminal region are more variable, the N-terminal 25% being hypervariable.[14,17] The conservative nature of the C-terminal region was further supported by DNA sequence analysis of the C-terminal region of the M gene from five different serotypes.[51]

Except for the physicochemical analysis of lysin-extracted M6 protein, which represents nearly the complete M6 molecule from the streptococcus,[1] no detailed analysis has previously been available regarding the structural characteristics of the C-terminal half or the complete M molecule.

In this report, we present a detailed conformational analysis of the complete sequence of the M6 protein from strain D471.[8] The N-terminal half of this molecule has also been compared to the corresponding region of the M5 protein from strain B788[6,7] to determine the common features that may play a role in the cross-reactive and antiphagocytic properties of the molecule. Because of the significant variability observed in both size and sequence among M proteins[11,13] the structural characteristics presented here for the M6 protein could serve as a prototype for comparisons with data generated from other M protein serotypes as well as strains of the same serotype.

MATERIALS AND METHODS

Streptococcal Strains

Group A streptococcal strains D471 (M type 6) and B788 (M type 5) are from the Rockefeller University collection.

Fourier Analysis

The Fourier periodicities in the amino acid sequence of the M6 protein[8] were calculated as previously described for the PepM5 protein.[6]

Computer Analyses

Secondary structural analysis was performed according to Robson,[18] since results for fibrous proteins appear to correspond more closely to that predicted by this algorithm.

RESULTS

Secondary Structure Analysis

Predictive secondary structure analysis of the M6 sequence[8] by the algorithm of Robson[18] revealed that the majority of the M6 molecule exhibits strong alpha-helical potential (Fig. 1). A short N-terminal nonhelical segment is followed by an essentially alpha-helical region, interrupted from residues 34–77 by small segments of beta-sheet potential. The algorithm also predicts alpha-helical potential for region 78–364, with a small interruption at residues 282–289. Region 365–415 exhibits beta sheet, coil, and

```
1                   20                  40                  60
RVFPRGTVENPDKARELLNKYDVENSMLQANNDKLTTENNNLTDQNKNLTTENKNLTDQN
BBBBTCCCCCCHHHHHHHHHHHHHHBHHHHHHHHBBBBBTTCCHHTBBBBBBBTBHHHHHC

KNLTTENKNLTDQNKNLTTENKELKAEENRLTTENKGLTKKLSEAEEEAANKERENKEAI  120
BCBBBBTBHHHHHCBCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

GTLKKTLDETVKDKIAKEQESKETIGTLKKTLDETVKDKIAKEQESKETIGTLKKTLDET  180
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

VKDKIAKEQESKETIGTLKKILDETVKDKIAREQKSKQDIGALKQELAKKDEGNKVSEAS  240
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

RKGLRRDLDASREAKKQVEKDLANLTAELDKVKEEKQISDASRQGLRRDLDASREAKKQV  300
HHHBBBBHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCTTBBBBBHHHHHHHHHH

EKALEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEQLAKQAEELAKL  360
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

RAGKASDSQTPDAKPGNKVVPGKGQAPQAGTKPNQNKAPMKETKRQLPSTGETANPFFTA  420
HHHHCCTBBBBTCCTTBBBBBTBCBBBBBTCCCTTCCCHHHHHHBBCCCCCCCCCHHHHH

AALTVMATAGVAAVVKRKEEN 441
HHHHHHHHHHHHHHHHHHHHH
```

Fig. 1. Secondary structural characteristics of the complete M6 sequence as determined by the Robson algorithm.[18] Amino acids are indicated by single letter code. The characteristics assigned to each residue indicate the conformational potential of that residue. Assignments B, T, C, and H represent beta-sheet, beta-turn, random coil, and alpha-helix potential, respectively.

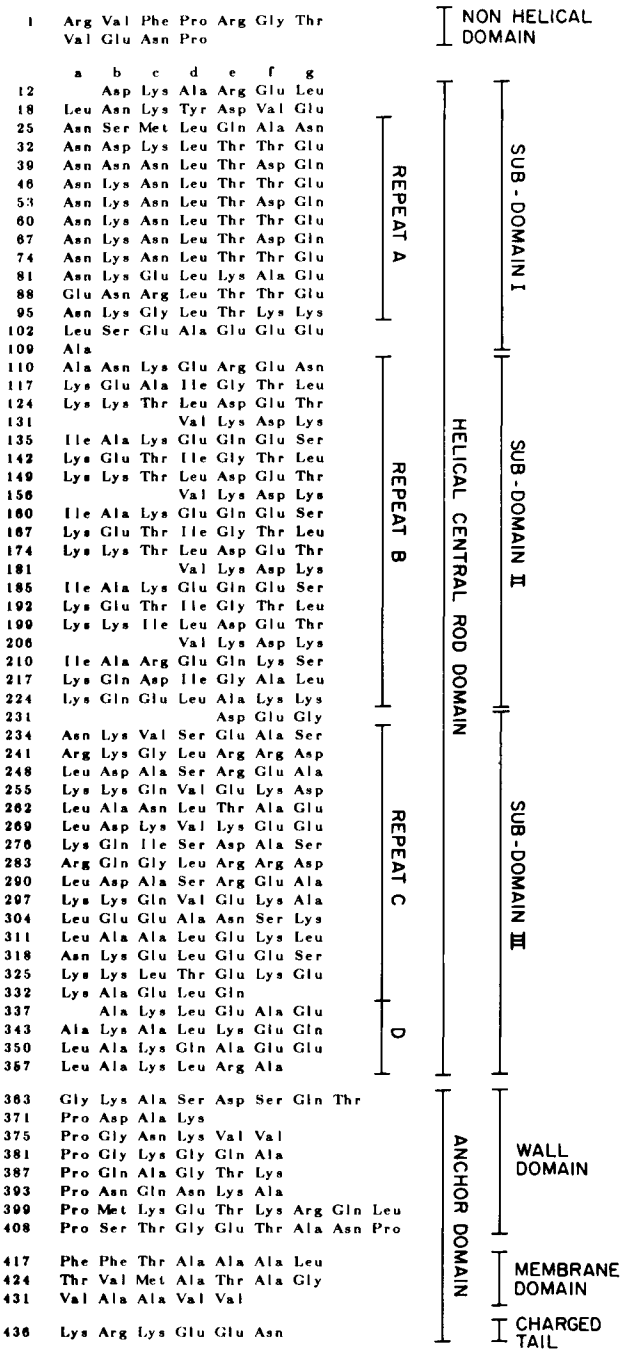

Fig. 2. Seven-residue periodicity and domain assignments in the complete M6 sequence. Arrangement of the M6 sequence is based on the position of amino acids in a seven-residue (heptad) repeat designated by letters "a"–"g." The periodicity begins at residue 12 of the mature protein and continues, with interruptions at residues 109, 131, 156, 181, 206, 231, and 337, through residue 362. Alignment from residue 363 to 416 is based essentially on the position of prolines in the sequence. However, no periodicity is present from residue 363 to the end. Sequence repeat regions A, B, and portions of C are designated based on the DNA analysis of Hollingshead et al.,[8] while portions of C and D are from this analysis. Three major domains are indicated (nonhelical, helical, and anchor) along with the positions of subdomains in the helical and anchor regions.

turn potential, with a short helical segment; and residues 416–441 again have alpha-helical potential.

Heptad Periodicity in the M6 Protein

Alpha-helical coiled-coil proteins are constructed from a repeating heptapeptide pattern (a–b–c–d–e–f–g)$_n$[19] in which residues in position "a" and "d" of the helix are generally nonpolar and form the "core" residues within the coiled coil. Residues in position "e" and "g" lie next to the "core" residues and are considered "inner" residues, while those in position "b," "c," and "f" are "outer."[19,20] The arrangement of the amino acid residues of the M6 protein within the repeating 7-residue pattern is shown in Figure 2. The 7-residue periodicity, which extends from amino acid 12 through 362, indicates the strong likelihood of a coiled-coil structure.[21] However, the continuity of the heptad pattern suffers several disruptions because of the insertion (or deletion) of up to three residues. Similar breaks in the periodicity have been noted previously in PepM5[6] as well as in other coiled-coil proteins such as myosin,[22,23] paramyosin,[24] intermediate filament proteins[25] and in the alpha, beta, and gamma chains of fibrinogen.[26]

The central rod domain of the M6 protein is made up of three distinct sub-domains, each characterized by differences in the heptad distribution of nonpolar residues. Two subdomains, A and B, differing in heptad pattern, were also seen in the N-terminal half of the M5 molecule (PepM5).[6]

The general characteristics of sub-domain I comprising residues 12–109 may be summarized as follows. Residues 12–24 are alpha-helix favoring, and exhibit the heptad periodicity characteristic of a coiled-coil protein.[21] Following this, region 25–99 contains a 14-residue repeat with a quasi-heptad substructure that is unusual in several ways. In particular, position "a" is 10/11 Asn and position "d" is 11/11 Leu. A similar distribution of Asn in the "a" position and Leu in the "d" position was observed earlier in domain A of the PepM5 protein.[6] The tendency towards placement of Asn in position "a" is also seen in the sequence of the PepM24 protein.[5] Thus, this feature appears to be common for streptococcal M proteins but is atypical of other coiled-coil molecules. Asn is not normally found in position "a" with such high frequency in the other coiled-coil proteins, and Leu has never before been shown to occur with such regularity either in position "a" or "d".[27] Although ionic interactions, which help stabilize the coiled-coil structure,[19,26,28] are few in number, overall it appears that the structure in region 25–99 will be a coiled-coil, but probably a flexible one. Electron microscope studies by Phillips et al.[1] support this conclusion. Region 100–109, is a short, alpha-helix-favoring coiled-coil structure in phase with the region preceeding it. Thus, the boundary between sub-domains I and II is defined by heptads containing nonpolar residues at both "a" and "d" positions.

Subdomain II, residues 110–230, has a 25-residue repeat which occurs nearly five times. If it is assumed that each repeat has the same conformation, then the continuity of the heptad structure is broken, since 25 is not a multiple of 7. This region shows four "stutters" at residues 131, 156, 181, and 206, with the "a" and "d" polarity reversing at each "stutter," i.e., a,d,a,d,d,a,d instead of the usual a,d,a,d,a,d,a, etc. In this subdomain, the "a" positions are essentially occupied by either Ile or Lys, with lysines occurring 2.5 times as often as Ile (10/15 Lys, 4/15 Ile). Lysine has been observed previously in the "a" positions in other fibrous proteins but rarely in "d".[27] Three fourths of the "d" positions are occupied by nonpolar residues, and the remaining one-fourth are occupied by Glu. Intramolecular ionic interactions do not provide appreciable extra stability to this segment. No long-range periods, other than that relating to the 25-residue repeat, have been observed in this region. Thus, like subdomain I, subdomain II of the M6 protein is likely to have a flexible coiled-coil structure.

The third coiled-coil domain, designated subdomain III, includes residues 231–362. Although not as identical as the repeats found in subdomains I and II, this domain has a 42-residue repeat that occurs about 2½ times within region 231–336. Also, the heptad phasing is continuous over the entire length of this region. The apolar residue content in positions "a" and "d" is higher than in the other coiled-coil subdomains. In addition, ionic interactions are appreciable (+10) for a parallel, in-register chain arrangement (i.e., 0.61 interactions per heptad pair).[28] Following this region, a break in heptad continuity occurs at residue 337, but a further short segment of coiled-coil structure is evident with a seven residue quasi-repeat (repeat D) having the consensus sequence LAKLXEE or LAKLXAE (where X is variable but 50% basic).

Fourier Analysis

Fourier analyses have been used previously to detect the regularities in the amino acid sequence of fibrous proteins.[26,29,30] For example, when tropomyosin, the archetypical alpha-helical coiled-coil protein, is analyzed by Fourier techniques, a very strong periodicity of 7 is seen with spacings of 7/2, 7/3, and 7 when the nonpolar residues are examined. Similar periodicities are found if one also examines the acidic or basic residues of tropomyosin. These seven residue repeats, together with the harmonics at 7/2 and 7/3 are indicative of a coiled-coil structure.

Similar Fourier analysis of the nonpolar, basic, or acidic residues in region 12–362 of the M6 sequence also revealed significant periodicities at 7/2 and 7/3 (Table I). In addition, asparagine also exhibited a periodicity at 7 within this region of the molecule. Other strong periodicities were also seen in the acidic and basic residues. For example, the acidic amino acids revealed a significant periodicity at 5, whereas for the basic residues it was at 5/2. A periodicity at 5/2 for basic amino acids was previously found in the PepM5 sequence.[6]

The periodicities seen for the nonpolar residues in repeat regions A, B, and C are directly related to the heptad substructures and reinforce the results of the Fourier analysis. The probabilities calculated for the periodicities observed in Table I are not quite as high as that seen for tropomyosin[30] (and our unpublished analysis), owing in part to minor irregularities in the heptad distribution of nonpolar residues in the M6 sequence. Nevertheless, the periodicities are statistically very significant and represent a characteristic pattern found in other alpha-helical coiled-coil structures.[23,27]

TABLE I. Significant Structural Periodicities Within the Rod Region of the M6 Sequence Revealed by Fourier Analysis*

Residue type	Period	Rational approximation	Probability†
Nonpolar			
Ala, Val, Leu,	4.2	42/10	0.2698×10^{-5}
Ile, Met, Tyr	3.54	7/2	0.6470×10^{-4}
	2.34	7/3	0.4235×10^{-3}
Asparagine	6.97	7	0.1606×10^{-6}
Basic			
Lys, Arg	3.53	7/2	0.1184×10^{-3}
	2.50	5/2	0.7064×10^{-3}
Acidic			
Glu, Asp	5.04	5	0.2962×10^{-5}
	3.53	7/2	0.4418×10^{-3}
	2.28	7/3	0.4638×10^{-3}

*Fourier analysis was performed on residues 12–362 of the M6 protein. Fourier methods evaluated selected types of amino acids for nonrandom periodicities. Only the strongest periodicities are shown.

†The accepted cutoff for significance was 0.05.

TABLE II. Percentage of Amino Acids in the Different Domains of the M6 Protein That Belong to "Core," "Inner," and "Outer" Positions

Residue type	Core "a,d"	Inner "e,g"	Outer "b,c,f"	Total
Subdomain I (region 12–109)				
Nonpolar*	17.35	1.02	4.08	22.45
Asparagine	10.20	1.02	9.18	20.41
Acidic†	1.02	10.20	9.18	20.41
Basic‡	—	3.06	12.25	15.31
Subdomain II (region 110–230)				
Nonpolar	15.70	4.96	5.79	26.45
Acidic	4.13	3.31	14.88	22.31
Basic	8.26	8.26	9.09	25.62
Subdomain III (region 231–362)				
Nonpolar	16.67	3.79	13.63	34.09
Asparagine	1.52	0.76	0.76	3.03
Acidic	—	12.88	10.61	23.48
Basic	5.30	6.06	12.88	24.24

*Nonpolar: Leu, Ile, Val, Tyr, Ala, Met.
†Acidic: Asp, Glu.
‡Basic: Lys, Arg.

Amino Acid Distribution Within the Heptad Positions of the M6 Rod region

The percentage of the different classes of amino acids in subdomains I, II, and III of the M6 protein that belong to the "core," "inner," and "outer" positions is presented in Table II. Nonpolar residues in the core "a" and "d" positions comprise about the same percentage in all three subdomains. However, the percentage of total nonpolar residues progressively increases from subdomain I to subdomain III.

Like domain A of the PepM5 protein,[6] subdomain I of the M6 molecule has a clear excess of negative charges. Unlike the PepM5 protein, the positive and negative charges in the "e" and "g" positions, which are largely responsible for any salt bridges occurring between chains,[19,28] are not equal. In M5, the excess in negative charge occurs in the outer "b,c,f" positions. It is probable that in M6 the negative charges in positions "e" and "g", which do not enter into interchain salt bridge formation, will be more exposed on the surface of the helix where they will be hydrated and effectively contribute a net negative charge on the outer surface of subdomain I. This common net negative charge region, which is in a segment of the molecule distal to the cell surface, is found in both M5 and M6 proteins. This conserved characteristic may, as previously suggested,[5,6,31] play a role in the antiphagocytic property of the M proteins.

Structural Features of the C-Terminal Region

Residues 363–416

As a result of the high content of proline and glycine residues (8 and 6, respectively), this region exhibits clear beta-turn and random coil potential (Fig. 1). The occurrence of proline residues at regular intervals (one every sixth residue repeated nearly 5 times and two separated by nine and eight residues) is a feature of particular interest and may point to the presence of a quasi-repeating structural motif within this segment. This region is located within the streptococcal cell wall (Pancholi and Fischetti, submitted) and will likely need to attain a specific conformation in order to intercalate within the cross-linked peptidoglycan and anchor the M molecule. However, its actual conformation is unknown at this time.

Residues 417–435

The 19 uncharged amino acids within this C-terminal portion of the sequence are predicted to have an alpha-helical conformation which would have a length exactly equal to that required to traverse the cytoplasmic membrane (~3 nm).[32] A hydropathicity analysis of this segment revealed a plot typical of a hydrophobic membrane anchor domain.[8] This membrane-spanning region probably functions to stop the transfer of the M molecule across the cytoplasmic membrane.[33–35]

Residues 436–441

The highly charged tail at the C-terminal end of the molecule is likely to protrude through the cell membrane to the cytoplasmic side. This has been seen previously in membrane-associated proteins; see for example, reference 36. In conjunction with the membrane anchor, this charged tail may also function to stop the transfer of the M protein across the membrane and thus act to further stabilize the M molecule in the membrane after biosynthesis.

Sequence Comparison With M5 Protein

Aligment of the PepM5 protein sequence,[7] comprising the N-terminal half of the native M5 molecule, with the corresponding region of the M6 sequence[8,17] revealed a number of identities. The first 85 residues of the M6 sequence showed 33 identities (39%) with the M5 molecule (Fig. 3). Maximal homology is observed between the two M molecules if a gap of 29 residues is introduced in the M5 sequence. This was necessary since both the pepsin-derived N-terminal fragments and the native M molecules from these two strains differ in size, M5 being smaller (V.A.F., unpublished data).[7,11] Because homologous recombination between repeated sequences generates size differences among M molecules,[11,13] the observed size difference between these two molecules is likely to result from such genetic events.

Fifty-five percent identity was observed between residues 115–228 of the M6 sequence and 85–197 of the PepM5 protein. This increases to 78% identity between residues 164–227 of M6 and 134–197 of M5. The latter segment (134–197 of M5) corresponds to the C-terminal half of subdomain II within the M6 rod region. Consistent with this extensive homology

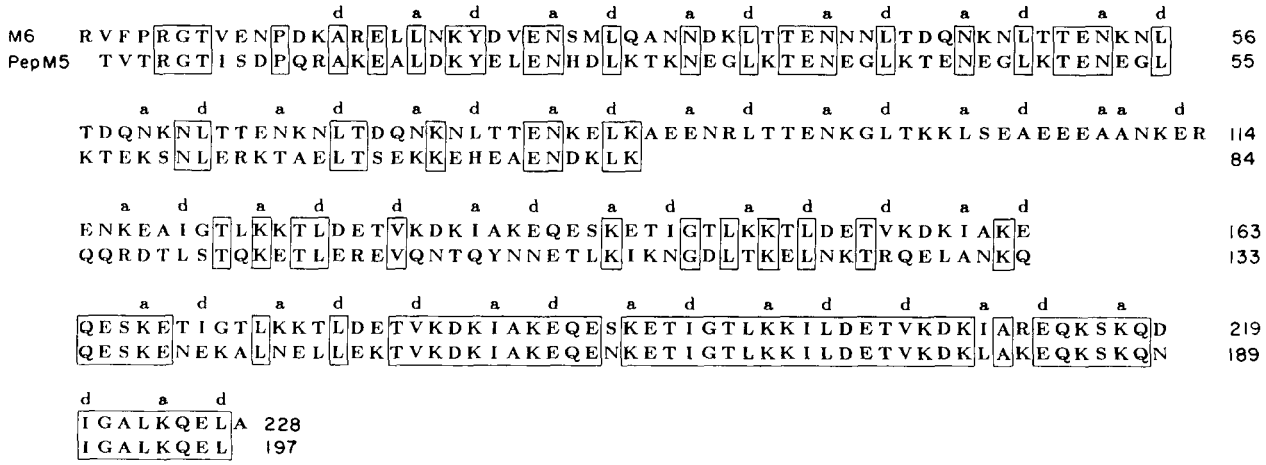

Fig. 3. Sequence comparison between the PepM5 protein[7] and the corresponding region of the M6 molecule (M6 N-terminal half).[8,17] Amino acids are indicated by single letter code. Identical residues are boxed. Lowercase "a" and "d" above the sequence designates the position of these residues within the heptad (a–b–c–d–e–f–g) based on the alignment in Figure 2. A 29-residue gap was introduced in the M5 sequence at residue 84 to maximize the number of identities (see Results for details).

between M6 and M5 is the finding of immunological crossreaction between these regions of the two M molecules.[15–17] In all, 42% identity is observed between these two segments which represent the N-terminal halves of the native M5 and M6 proteins.

Based on previous studies[14,15,17] and this analysis, residues 1–163 of M6 and 1–133 of M5 may be considered the hypervariable regions of these M molecules. Although the number of identities within these hypervariable regions is far less than for the region C-terminal to these segments, 49% of the identities (22/45) are those that occupy positions "a" and "d" within the repeating heptads (Fig. 3).[5,27] Therefore, despite considerable sequence dissimilarity between the two M molecules within this hypervariable region, identities are preserved in critical positions required for the maintenance of the coiled-coil structure of the M protein. As previously suggested,[6,37] the maintenance of the coiled-coil rod domain appears to be a biologically significant characteristic of the M molecule.

Conformational Changes in M Protein Deletion Mutants

Mutants have been isolated in vitro from M6 strain D471 which, through homologous recombination, have deleted segments within the "A" repeat region of the M gene (located within subdomain I).[12,13] For example, DNA sequence analysis of deletion mutant D471d113 revealed the deletion of the equivalent of two "A" repeat blocks (beginning from the center of block A0 to the center of block A2) resulting in hybrid block A0/A2 and thus a shortened subdomain I (Fig. 4).[13] Mutant D471d112 also was deleted for two "A" repeat blocks, but in this case, two complete identical repeat blocks were removed.[13] Since the central blocks A1, A2, and A3 are identical, it is uncertain whether A1 and A2 or A2 and A3 were deleted. In this example it is assumed that A1 and A2 were deleted. Conformational analysis of the sequence of these new M6 molecules indicates significant changes in the helical potential within the "A" repeat region when compared to the D471 parental strain (Fig. 4). Such differ-

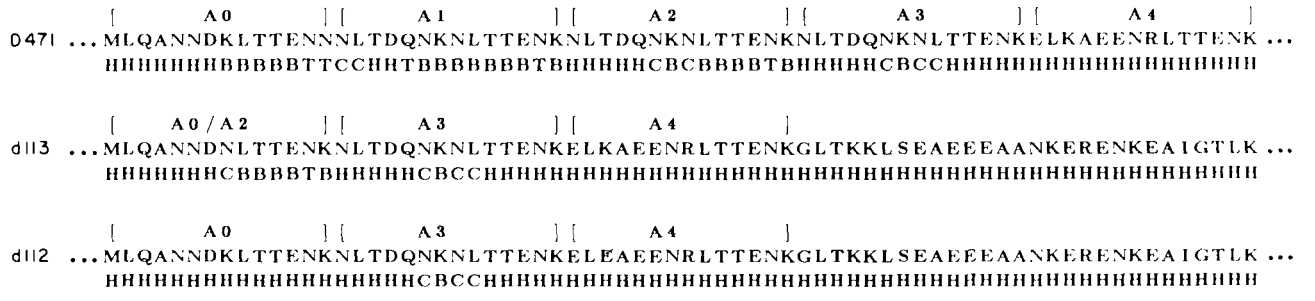

Fig. 4. Secondary structure analysis of the "A" sequence repeat region of M6 strain D471 and deletion mutants D471d113 (d113) and D471d112 (d112) by the Robson predictive algorithm.[18] Letters designating the secondary structure of the amino acids are the same as in Figure 1. Square brackets above the sequence enclose repeat blocks A0, A1, A2, A3, and A4.[18] A deletion in D471 of the region beginning from the center of repeat A0 to the center of A2 by homologous recombination resulted in the hybrid block A0/A2 in mutant d113.[13] Mutant d112 was derived from D471 by a deletion of the complete A1 and A2 repeat blocks. The sequence begins at methionine-27 in all three analyses and continues for 70 residues.

ences in the M protein conformation in these mutants are likely to result in changes in immunodeterminants within or adjacent to the altered region when compared to the parental strain.

Model

A model has been proposed for the complete M6 molecule based on the above data (Fig. 5). No specific conformation has been assigned to the N-terminal 11 residues. This nonhelical terminal domain is located prior to the coiled-coil central rod region which is composed of three alpha-helical subdomains totaling 351 amino acid residues (region 12–362) and calculated to be ~52 nm in length (assuming a 0.149 nm rise/residue).[38] This is close to the 50 nm observed for the length of platinum shadowed M6 molecules extracted from the cell with phage lysin.[1] Unlike tropomyosin, which is nearly 100% alpha-helical coiled-coil and has no distortion of the nonpolar periodicity, the breaks observed in the heptad phasing of the M6 protein sequence would probably allow the molecule to be flexible at these regions of irregularity. In addition, the unusual occurrence of Asn in the heptad "a" position within the N-terminal 25% of the molecule might prevent or modify optimal packing of the amino acid side chains between the two protein chains within this segment. However, measurements of platinum-shadowed native M6 molecules have revealed that this region remains in a rod-like confirmation.[1] The three subdomains of the central rod region correlate well with the position of sequence repeat regions A, B, and C/D, respectively, in this M6 protein[8] (Fig. 2).

Adjacent to the C-terminal end of the central rod region is the cell anchor domain of the molecule. This is composed of a 54 amino acid wall anchor subdomain, an adjacent 19-residue membrane anchor subdomain, and a 6-residue charged tail at the C-terminus.

DISCUSSION

From current information on the sequence of the N-terminal half of M5,[6,7] the partial sequence of M24[9,10] and the complete M6 sequence,[8] it is likely that streptococcal M proteins from different serotypes may be constructed according to a common plan. A large number of M protein molecules, which differ in amino acid sequence but maintain a common conformation, may be assembled if the organism adheres to some basic rules. The amino acid sequence within the central rod domain of the molecule is formed by a series of heptad repeats of the kind $(a\text{–}b\text{–}c\text{–}d\text{–}e\text{–}f\text{–}g)_n$ where positions "a" and "d" are usually hydrophobic and the intervening residues are primarily helix promoting. Although the sequence of the heptad may differ between M types, and even among strains of the same type, the presence of helix-promoting residues will allow the different molecules to attain a similar alpha-helical conformation. Since 3.6 residues are re-

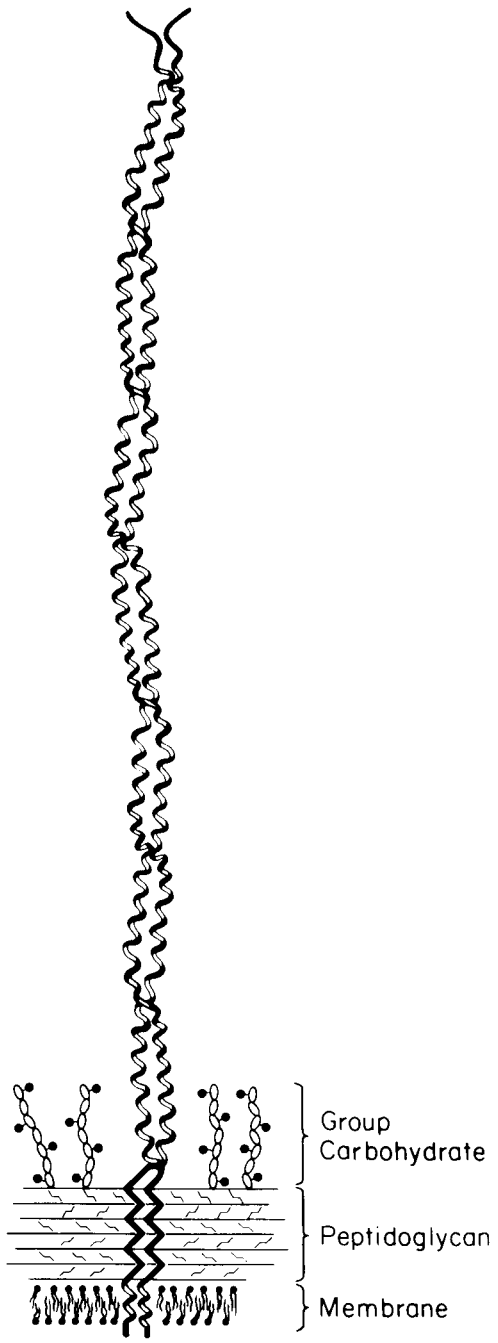

Fig. 5. Model of the complete M6 protein from strain D471 based on both previous data[8] and this analysis. In this proportional representation of the M molecule and its relationship with the streptococcal cell, the alpha-helical coiled-coil rod region extends about 50 nm from the surface of the cell with the short nonhelical N-terminal domain distal to the cell surface. The proline-glycine-rich region of the protein is intercalated within the cross-linked peptidoglycan and the membrane anchor segment extends through the cell membrane. The short charged tail at the C-terminus protrudes through the membrane into the cytoplasm. Evidence suggests that a short segment of the rod region is embedded within the group carbohydrate portion of the cell wall[44] (Pancholi and Fischetti, submitted) which is composed of a rhamnose backbone (open circle) and N-acetylglucosamine branches (closed circles).[50]

quired to make one turn of the helix,[38] the hydrophobic residues at positions "a" and "d" on the heptad would form an inclined stripe around the helix. For thermodynamic stability, a coiled-coil rope would be formed by the internalization of the hydrophobic residues from the two helical chains coiling around each other.[5,19,37] The presence or absence of disruptions in the periodicity of hydrophobics at positions "a" and "d" of the heptad would allow these regions to become more or less flexible.

A short 11-residue nonhelical domain is found at the N-terminal end of the coiled-coil rod region which is distal to the cell surface (Fig. 2). In all M proteins examined thus far, the sequence of this region is distinct from one M type to another,[5,7] but seems identical within an M type.[13] However, the conformation within this region appears to differ from the adjacent helical segment. It is not certain whether, despite sequence variation, the actual conformation within this N-terminal region is similar among the different M protein types. This N-terminal segment seems to play an important role in the biological activity of the molecule, since antibodies generated to this segment will effectively opsonize the stretococcus of the M type from which it was derived.[39–42,52] The type-specific dependence of opsonic antibodies[2] is consistent with the observation that this N-terminal segment is located within the hypervariable region of the M molecule.[16,17] It seems likely that the central rod domain may function as a shaft to position the N-terminal domain away from the cell surface.

The C-termimal end, which also differs considerably in structure from the central coiled-coil domain, functions to anchor the M protein to the streptococcus within both the cell wall and membrane. The amino acid sequence of this region is conserved among strains of different M protein types [51] as well as in surface proteins from other gram positive organisms, i.e., protein A from *Staphylococcus aureus*[8] and protein G from group G streptococci.[43] As previously suggested,[8] this domain may serve to anchor molecules like M protein in the gram positive cell wall (Pancholi and Fischetti, submitted). We speculate that since M protein is synthesized at the septum,[44] where the cell wall is least cross-linked,[45] the membrane anchor may function to stop the transfer of the M protein during its maturation. The peptidoglycan is then cross-linked around this proline/glycine region thus anchoring the M molecule in the cell wall (Pancholi and Fischetti, submitted).

The M protein molecules from different strains of any one serotype are very heterogeneous in size.[11] We have shown that in M6 this results from frequent homologous recombination events (about 1 in 2,000 CFU's) within the extensive DNA repeats[12,13] which compose most of the central rod region of the molecule (Fig. 2).[8] When the recombination events generate deletions, they not only produce changes in the size of the M6 molecule, but, because the external repeats within a repeat block are nonidentical, recombinations involving these repeats may also produce changes in the amino acid sequence, as is the case for mutant d113.[13] The conformational analysis presented above (Fig. 4) demonstrates that concomitant with changes in the size of the M molecule two of the deletion events studied also produced changes in the helical potential in the region of the protein where the deletion occurred. Each M6 strain may vary with respect to the number of repeat regions and therefore would deviate from one another with respect to the specific structure and conformation within these areas of the molecule.[13] Currently, no data exist describing the influence that such changes in conformation could have on the biological or immunological activities of the M molecule. One could envision, however, that if the influence is substantial, this could translate to changes in specific immunodeterminants (conformational or linear) as well as to the antiphagocytic property of the molecule. Studies are in progress to examine these questions in more detail.

The basic structure of the M6 protein presented here is remarkably close to that of eucaryotic intermediate filaments, whose elemental structure is also based on a coiled-coil central rod domain with structurally different end domains.[25,46] In addition, like M protein, intermediate filaments differ immunologically as well as in size and composition depending on their source.[46] Because of its coiled-coil conformation, M protein also resembles other coiled-coil molecules such as myosin and tropomyosin both in basic conformation and in primary sequence. Up to 50% sequence identity was observed between M6 and segments of tropomyosin[5] (V.A.F., unpublished data) and nearly 23% identity with myosin heavy chain in a 105 amino acid overlap (V.A.F., unpublished data). Similar homology and structural similarity was also observed between myosin and the PepM5 protein.[47] In addition, up to 18% sequence identity in a 223 amino acid overlap of the M6 sequence was also observed with the intermediate filament protein, type I keratin (V.A.F., unpublished data).

The implications of the resemblance of M protein to regions of mammalian proteins are emphasized by the immunological cross-reaction observed between M5 protein and myosin.[48,49] While the pathological implications of the close similarity of this streptococcal surface protein with a number of mammalian structural analogues has not been fully explored, it is difficult to imagine that no relationship exists.

ACKNOWLEDGMENTS

This work was supported by Research Grants AI11822 (to V.A.F.), AI20723 (to J.R.S and V.A.F), and HL36025 (to B.N.M) from the National Institutes of Health. One of us (D.A.D.P) wishes to acknowledge the contribution of Dr. A.C. Steven, National Insti-

tute of Arthritis and Musculo-Skeletal and Skin Diseases, who arranged a short-term visit to NIH under the Courtesy Associateship Program and which allowed part of the research work described here to be undertaken.

REFERENCES

1. Phillips, G.N., Flicker, P.F., Cohen, C., Manjula, B.N., Fischetti, V.A. Streptococcal M protein: Alpha-helical coiled-coil structure and arrangement on the cell surface. Proc. Natl. Acad. Sci. U.S.A. 78:4689–4693, 1981.
2. Lancefield, R.C. Current knowledge of the type specific M antigens of group A streptococci. J. Immunol. 89:307–313, 1962.
3. Lancefield, R.C. Persistence of type specific antibodies in man following infection with group A streptococci. J. Exp. Med. 110:271–292, 1959.
4. Quinn, R.W., Vander Zwaag, R., Lowry, P.N. Acquisition of a group of A streptococcal M protein antibodies. Pediatr. Infect. Dis. 4:374–378, 1985.
5. Manjula, B.N. and Fischetti, V.A. Tropomyosin-like seven residue periodicity in three immunologically distinct streptococcal M proteins and its implication for the antiphagocytic property of the molecule. J. Exp. Med. 151:695–708, 1980.
6. Manjula, B.N., Trus, B.L., Fischetti, V.A. Presence of two distinct regions in the coiled-coil structure of the streptococcal Pep M5 protein: Relationship to mammalian coiled-coil proteins and implications to its biological properties. Proc. Natl. Acad. Sci. U.S.A. 82:1064–1068, 1985.
7. Manjula, B.N., Acharya, A.S., Mische, S.M., Fairwell, T., Fischetti, V.A. The complete amino acid sequence of a biologically active 197-residue fragment of M protein from type 5 group A streptococci. J. Biol. Chem. 259:3686–3693, 1984.
8. Hollingshead, S.K., Fischetti, V.A., Scott, J.R. Complete nucleotide sequence of type 6 M protein of the group A streptococcus: Repetitive structure and membrane anchor. J. Biol. Chem. 261:1677–1686, 1986.
9. Beachey, E.H., Seyer, J.M., Kang, A.H. Repeating covalent structure and streptococcal M protein. Proc. Natl. Acad. Sci. U.S.A. 75:3163–3167, 1978.
10. Beachey, E.H., Seyer, J.M., Dale, J.B., Hasty, D.L. Repeating covalent structure and protective immunogenicity of native and synthetic polypeptide fragments of type 24 streptococcal M protein. J. Biol. Chem. 258:13250–13257, 1983.
11. Fischetti, V.A., Jones, K.F., Scott, J.R. Size variation of the M protein in group A streptococci. J. Exp. Med. 161:1384–1401, 1985.
12. Fischetti, V.A., Jarymowycz M., Jones, K.F., Scott, J.R. Streptococcal M protein size mutants occur at high frequency within a single strain. J. Exp. Med. 164:971–980, 1986.
13. Hollingshead, S.K., Fischetti, V.A., Scott, J.R. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. 207:196–203, 1987.
14. Scott, J.R., Hollingshead, S.K., Fischetti, V.A. Homologous regions within M protein genes in group A streptococci of different serotypes. Infect. Immun. 52:609–612, 1986.
15. Jones, K.F., Manjula, B.N., Johnston, K.H. Hollingshead, S.K., Scott, J.R., Fischetti, V.A. Location of variable and conserved epitopes among the multiple serotypes a streptococcal M protein. J. Exp. Med. 161:623–628, 1985.
16. Jones, K.F., Khan, S.A., Erickson, B.W., Hollingshead, S.K., Scott, J.R., Fischetti, V.A. Immunochemical localization and amino acid sequence of cross-reactive epitopes within the group A streptococcal M6 protein. J. Exp. Med. 164:1226–1238, 1986.
17. Manjula, B.N., Acharya, A.S., Fairwell, T., Fischetti, V.A. Antigenic domains of the streptococcal Pep M5 protein: Localization of epitopes crossreactive with type 6 M protein and identification of a hypervariable region of the M molecule. J. Exp. Med. 163:129–138, 1986.
18. Garnier, J., Osguthorpe, D.J., Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97–120, 1978.
19. McLachlan, A.D., and Stewart, M. Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. J. Mol. Biol. 98:293–308, 1975.
20. Crick, F.H.C. The packing of alpha-helices: Simple coiled-coil. Acta Cryst. 6:689–697, 1953.
21. Cohen, C., and Parry, D.A.D. Alpha-helical coiled coils—a widespread motif in proteins. T.I.B.S. 11:245–248, 1986.
22. Parry, D.A.D. Structure of rabbit skeletal myosin: An analysis of the amino acid sequences of two fragments from the rod region. J. Mol Biol. 153:459–464, 1981.
23. McLachlan, A.D. and Karn, J. Periodic features in the amino acid sequence of nematode myosin rod. J. Mol. Biol. 164:605–626, 1983.
24. Cohen, C., Lanar, D.E., Parry, D.A.D. Amino acid sequence and structural repeats in the schistosome paramyosin match those of myosin. Biosci. Rep. 7:11–16, 1987.
25. Parry, D.A.D. and Fraser, R.D.B. Intermediate filament structure: 1 Analysis of IF protein sequence data. Int. J. Biol. Macromol. 7:203–213, 1985.
26. Parry, D.A.D. Determination of structural information from the amino acid sequences of fibrous proteins. In: "Fibrous Proteins: Scientific, Industrial and Medical Aspects." New York: Academic Press, 1979:393–427.
27. Parry, D.A.D. Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coil on other proteins. Biosci. Rep. 2:1017–1024, 1982.
28. Parry, D.A.D., Crewther, W.G., Fraser, R.D.B., MacRae, T.P. Structure of alpha keratin: Structural implications of the amino acid sequences of the type I and type II chain segments. J. Mol. Biol. 113:449, 1977.
29. Fraser, R.D. and MacRae, T.P. The molecular structure of feather keratin. In: "Proceedings of the 16th International Ornithological Congress." Canberra: Australian Acad. Sci., 1976: 443–451.
30. McLachlan, A.D. and Stewart, M. The 14-fold periodicity in alpha-tropomyosin in the interaction with actin. J. Mol. Biol. 103:271–298, 1976.
31. Fischetti, V.A. Requirements for the opsonic activity of human IgG directed to type 6 group A streptococci: Net basic charge and intact Fc region. J. Immunol. 130:896–902, 1983.
32. Perutz, M.F. New X-ray evidence on the configuration of polypeptide chains. Nature 167:1053–1054, 1951.
33. Waxman, D.J. and Strominger, J.L. Limited proteolysis of penicillin-sensitive D-alanine carboxypeptidase purified from *Bacillus subtilis* membranes: Active water-soluble fragments generated by cleavage of a C-terminal membrane anchor. J. Biol. Chem. 256:20259–2066, 1981.
34. Boeke, J.D. and Model, P. A procaryotic membrane anchor sequence: Carboxy-terminus of bacteriophage f1 gene III retain it in the membrane. Proc. Natl. Acad. Sci. U.S.A. 79:5200–5204, 1982.
35. Yost, C.S., Hedgpath, J. Lingappa, V.R. Stop transfer sequence confers predicatable transmembrane orientation of a previously secreted protein in cell free systems. Cell 34:759–766, 1983.
36. Garoff, H., Frishauf, A., Simons, K., Lehrach, H., Delius, H. Nucleotide sequence of cDNA coding for semliki forest membrane glycoproteins. Nature 288:236–241, 1980.
37. Fischetti, V.A. and Manjula, B.N. Biologic and immunologic implications of the structural relationship between streptococcal M protein and mammalian tropomyosin. In: "Seminars in Infectious Disease, Vol. IV, Bacterial Vaccines." Robbins, J.B., Hill, J.C., and Sadoff, J.C. eds. New York: Thieme-Stratton, Inc., 1982: 411–418.
38. Huxley, H.E. and Perutz, M.F. New X-ray evidence on the configuration of polypeptide chains. Nature 167:1053–1054, 1953.
39. Beachey, E.H., Seyer, J.M., Dale, J.B., Simpson, W.A., Kang, A.H. Type-specific protective immunity evoked by synthetic peptide of Streptococcus pyogenes M protein. Nature 292:457–459, 1981.
40. Dale, J.B., Seyer, J.M., Beachey, E.H. Type-specific immunogenicity of a chemically synthesized peptide fragment of type 5 streptococcal M protein. J. Exp. Med. 158:1727–1732, 1983.
41. Dale, J.B. and Beachey, E.H. Localization of protective epitopes of the amino terminus of type 5 streptococcal M protein. J. Exp. Med. 163:1191–1202, 1986.
42. Beachey, E.H. and Seyer, J.M. Protective and nonprotective

epitopes of chemically synthesized peptides of the NH2-terminal region of type 6 streptococcal M protein. J. Immunol. 136:2287–2292, 1986.
43. Fahnstock, S.R., Alexander, P., Nagle, P., Filpula, D. Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. 167:870–880, 1986.
44. Swanson, J., Hsu, K.C., Gotschlich, E.C. Electron microscopic studies on streptococci. I.M. antigen. J. Exp. Med. 130:1063–1091, 1969.
45. Fan, P.F. and Beckman, B.E. Structural difference between wall from hemispherical caps and partial septa of *Bacillus subtilis*. J. Bacteriol. 114:790–797, 1973.
46. Steinert, P.M., Stevens, A.C., Roop, D.R. The molecular biology of intermediate filaments. Cell 42:411–419, 1985.
47. Manjula, B.N. and Fischetti, V.A. Sequence homology of group A streptococcal PepM5 protein with other coiled-coil proteins. Biochem. Biophys. Res. Commun. 140:684–690, 1986.
48. Cunningham, M.W., Hall, N.K., Krisher, K.K., Spanier, A.M. A study of anti-group a streptococcal monoclonal antibodies cross reactive with myosin. J. Immunol. 136:293–299, 1986.
49. Dale, J.M., and Beachey, E.H. Sequence of myosin cross reactive epitopes of streptococcal M protein. J. Exp. Med. 164:1785–1790, 1986.
50. Coligan, J.E., Kindt, T.J., Krause, R.M. Structure of the streptococcal groups A, A-variant and C carbohydrates. Immunochemistry 15:755–760, 1978.
51. Hollingshead, S.K., V.A. Fischetti, and J.R. Scott. A highly conserved region present in transcripts encoding heterologous M proteins of group A streptococci. Infect. Immun. 55:3237–3239, 1987.
52. Jones, K.F. and V.A. Fischetti. Importance of the location of antibody binding on the M6 protein for opsonization and phagocytosis of group A M6 streptococci. J. Exp. Med. (In press), 1987.