Streptococcal M Protein: Molecular Design and Biological Behavior

VINCENT A. FISCHETTI

*The Rockefeller University, New York, New York 10021*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>286</td>
</tr>
<tr>
<td>ELECTRON MICROSCOPY</td>
<td>286</td>
</tr>
<tr>
<td>Thin Sections</td>
<td>286</td>
</tr>
<tr>
<td>Whole Streptococci</td>
<td>286</td>
</tr>
<tr>
<td>ISOLATION AND PURIFICATION</td>
<td>286</td>
</tr>
<tr>
<td>Pepsin, Detergent, and Lysin Extraction</td>
<td>287</td>
</tr>
<tr>
<td>STRUCTURAL ANALYSIS</td>
<td>288</td>
</tr>
<tr>
<td>Evidence for the Coiled-Coil Structure of M Protein</td>
<td>288</td>
</tr>
<tr>
<td>Structural Analysis of the Complete M Molecule</td>
<td>290</td>
</tr>
<tr>
<td>Basic Model</td>
<td>291</td>
</tr>
<tr>
<td>Analysis of the Wall-Associated Region</td>
<td>293</td>
</tr>
<tr>
<td>Domain Structure</td>
<td>293</td>
</tr>
<tr>
<td>IMMUNOCHEMISTRY</td>
<td>293</td>
</tr>
<tr>
<td>Cross-Reactions</td>
<td>294</td>
</tr>
<tr>
<td>Conserved, Variable, and Hypervariable Epitopes</td>
<td>294</td>
</tr>
<tr>
<td>EVIDENCE FOR TWO DISTINCT CLASSES OF M PROTEIN</td>
<td>294</td>
</tr>
<tr>
<td>MAP and Opacity Factor</td>
<td>294</td>
</tr>
<tr>
<td>MAP I Antigen and the M6 Protein C-Repeat Domain</td>
<td>295</td>
</tr>
<tr>
<td>C-Repeat Domain and Rheumatic Fever</td>
<td>295</td>
</tr>
<tr>
<td>SIZE VARIATION IN THE M MOLECULES</td>
<td>295</td>
</tr>
<tr>
<td>Isolation and Sequence Analysis of Size Mutants</td>
<td>295</td>
</tr>
<tr>
<td>ANTIGENIC VARIATION AND ITS RELATION TO SIZE VARIATION</td>
<td>297</td>
</tr>
<tr>
<td>Size Variation Leads to Antigenic Variation</td>
<td>297</td>
</tr>
<tr>
<td>Natural Selection of Size Mutants</td>
<td>297</td>
</tr>
<tr>
<td>Variation in the N-Terminal Nonrepeating Segment</td>
<td>298</td>
</tr>
<tr>
<td>COMPARISON OF M-PROTEIN SEQUENCES</td>
<td>298</td>
</tr>
<tr>
<td>Alignment of Known M-Protein Sequences</td>
<td>298</td>
</tr>
<tr>
<td>Alignment of the Variable Region of Two Different M Proteins</td>
<td>299</td>
</tr>
<tr>
<td>HOMOLOGY OF M PROTEIN WITH OTHER MOLECULES OF GRAM-POSITIVE BACTERIA</td>
<td>300</td>
</tr>
<tr>
<td>Group C and G Streptococci</td>
<td>300</td>
</tr>
<tr>
<td>Staphylococcal Protein A</td>
<td>300</td>
</tr>
<tr>
<td>STRUCTURAL RELATIONSHIP OF M PROTEIN WITH MAMMALIAN PROTEINS</td>
<td>301</td>
</tr>
<tr>
<td>Tropomyosin</td>
<td>301</td>
</tr>
<tr>
<td>Other Coiled-Coil Molecules</td>
<td>301</td>
</tr>
<tr>
<td>IMMUNOLOGICAL CROSS-REACTIVITY BETWEEN M PROTEIN AND MAMMALIAN PROTEINS</td>
<td>301</td>
</tr>
<tr>
<td>Heart-Reactive Antibodies</td>
<td>301</td>
</tr>
<tr>
<td>Immunodeterminants Shared with Myosin and Other Human Proteins</td>
<td>301</td>
</tr>
<tr>
<td>Renal Tissue-Associated Epitopes Shared with M Protein</td>
<td>301</td>
</tr>
<tr>
<td>Monoclonal Antibodies Cross-Reactive with Coiled-Coil Proteins</td>
<td>302</td>
</tr>
<tr>
<td>Conformational Determinants</td>
<td>302</td>
</tr>
<tr>
<td>THE emm GENE</td>
<td>303</td>
</tr>
<tr>
<td>Relationship of M-Protein Genes</td>
<td>303</td>
</tr>
<tr>
<td>Conversion of M⁻ to M⁺ Phenotype</td>
<td>303</td>
</tr>
<tr>
<td>Number of emm Gene Copies</td>
<td>303</td>
</tr>
<tr>
<td>Regulation of the emm Gene</td>
<td>304</td>
</tr>
<tr>
<td>HUMAN IMMUNE RESPONSE TO THE M MOLECULE</td>
<td>304</td>
</tr>
<tr>
<td>Immunodominant Region of the M Molecule</td>
<td>304</td>
</tr>
<tr>
<td>ANTIPHAGOCYTIC ACTIVITY OF M PROTEIN</td>
<td>304</td>
</tr>
<tr>
<td>N-Terminal Charge Domain</td>
<td>305</td>
</tr>
<tr>
<td>Effect on Phagocytic Cells</td>
<td>305</td>
</tr>
<tr>
<td>Complement</td>
<td>305</td>
</tr>
<tr>
<td>Factor H</td>
<td>305</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>306</td>
</tr>
<tr>
<td>ROLE OF M PROTEIN IN STREPTOCOCCAL ADHERENCE</td>
<td>306</td>
</tr>
<tr>
<td>M-PROTEIN VACCINES</td>
<td>306</td>
</tr>
<tr>
<td>Type-Specific Protection</td>
<td>306</td>
</tr>
</tbody>
</table>
INTRODUCTION

The group A streptococcus (Streptococcus pyogenes) is responsible for a number of suppurative human infections, of which acute pharyngitis and impetigo are the most common. As a consequence of ineffective antibiotic therapy or no therapy, as many as 3 to 5% of individuals who suffer a group A streptococcal pharyngeal infection may develop acute rheumatic fever (32, 204), a disease often resulting in cardiac damage. While not currently a major problem in developed countries, rheumatic fever is the leading cause of heart disease in school-age children in developing nations (1). For instance, by one estimate, over 6 million school-age children in India are afflicted with rheumatic heart disease (1). Acute glomerulonephritis, another sequela of group A streptococcal disease, is usually the consequence of infection by specific strains of streptococci (nephritogenic strains) which infect either the throat or skin (174). The ability of group A streptococci to persist in infected tissues is primarily due to the cell surface M protein, a molecule which confers to the streptococcus the ability to resist phagocytosis by polymorphonuclear leukocytes in the absence of type-specific antibodies. Resistance to a group A streptococcal infection appears to be related to the presence of type-specific antibodies to the M molecule (128, 129). Since there are >80 different serotypes of M protein (i.e., M5, M6, M24, etc.), an individual may become infected by more than one group A streptococcal type during a lifetime (129).

Streptococcal M protein was identified over 60 years ago by Rebecca Lancefield (126). A review by Lancefield in 1962 (129) clearly describes the studies carried out for nearly 35 years, defining this molecule as a major virulence factor for the streptococcus due to its antiphagocytic property. In 1974, a comprehensive review by Fox (81) delineated studies since the Lancefield review and underscored the knowledge to that time of the structure, function, and immunochromy of the M molecule.

The streptococcal M protein is now probably one of the best-defined molecules of the known bacterial virulence determinants. Its structure, function, immunochromy, and method of antigenic variation are unique among known virulence molecules and may serve as a model for certain microbial systems. To better appreciate the thinking which led to the current understanding of the M protein, whenever possible, attempts will be made to present this review in an historical perspective, with a chronological account of the studies over the years, emphasizing the time since 1974. While the antiphagocytic activity and genetics of the M molecule will be addressed in this review, the reader is also directed to recent reviews by Manjula (131) and Scott (J. R. Scott, in B. Igleswiski, ed., The Bacteria, Vol. XIII, in press) on these subjects.

In this review, the basic structure of the M molecule will be developed. Once established, it then becomes a simple matter to integrate this information into a better awareness of the biology of the M molecule and its relation to type specificity, antigenic diversity, and cross-reactivity. With these concepts understood, a strategy to control group A streptococcus-related diseases may become more apparent.

ELECTRON MICROSCOPY

Thin Sections

Thin-section electron micrographs identifying the M protein on the surface of group A streptococci were first published by Swanson et al. (193), using ferritin-labeled M-specific antibodies derived from type-specific rabbit sera. The data suggested that M protein appears as hairlike projections on the surface of the streptococcal cell wall (Fig. 1) (13, 169). M+ streptococci were found to be devoid of these surface projections (193).

Whole Streptococci

To better appreciate the features of the proteins on the surface of the streptococcal cell wall, organisms were prepared for electron microscopy by a variety of other techniques (68). Whole streptococci mounted on grids, critical-point dried, and viewed under a transmission electron microscope exhibit projections which appear to be the result of interactions among several adjacent M-protein fibrils forming tuftlike structures (Fig. 2a and b). Interactions between M molecules on adjacent streptococci, previously observed in thin sections (Fig. 1) (169), are also seen in these preparations. Whether these surface structures are composed exclusively of M protein or are the result of complexes of M protein and lipoteichoic acid (LTA) or other surface molecules was not determined. However, ferritin-labeled anti-M antibodies reveal that the extensions do contain M protein (Fig. 3). Surface replica preparations confirm the presence of these types of projections on the cell surface of whole organisms (Fig. 4). Streptococci treated with trypsin to remove the M protein were shown to have a smooth surface in all of these preparations (not shown).

In experiments designed to answer questions about the placement of M protein on the cell wall, Swanson et al. (193) found that, after trypsinization to remove existing M protein on living streptococci, newly synthesized M protein was first seen by electron microscopy on the cell in the position of the newly forming septum. In similar experiments designed after the classical experiments of Cole and Hahn and using fluorescein-labeled anti-M antibodies (44), trypsinized streptococci placed in fresh media for 10 min revealed M protein first at the newly forming septum (Fig. 5a). On organisms examined after extended incubation, M protein was not observed in the position of the old wall (Fig. 5b), suggesting that the M molecule is produced only where new cell wall is synthesized (193).

ISOLATION AND PURIFICATION

For many years the method used to prepare crude M protein for purification was based on the one developed by Lancefield (126) for the preparation of extracts for streptococcal typing (for references prior to 1974, see Fox [81] and Lancefield [129]). According to this procedure, organisms are suspended in dilute HCl at pH 2.0 and placed in a boiling-water bath for 10 min. After neutralization, the resulting extract is centrifuged and the supernatant is used as starting material for further characterization of the M mole-
cule. While the procedure effectively removes M protein from the cell, it is unlikely that the final purified product represents the native M-protein molecule (82). Recognizing this limitation, other procedures for extracting M protein from whole streptococci or their isolated cell walls were attempted. These included sonic oscillation (25, 157), alkali (47, 84), group C bacteriophage-associated lysis (43, 79), nonionic detergent (64, 65), nitrous acid (91), cyanogen bromide (202), and guanidine hydrochloride extraction (178), all of which were usually coupled to a particular purification scheme. However, in nearly all cases the final purified M protein exhibited extensive heterogeneity. These results led to the general conclusion at that time that M protein was composed of a heterogeneous group of molecules or of multiple subunits (82).

**Pepsin, Detergent, and Lysin Extraction**

The biochemical features of the M-protein molecule became more apparent with the development of an extraction procedure which specifically cleaves the M protein from the streptococcal surface. Pepsin digestion at a suboptimal pH of 5.8 releases a biologically active fragment of the M protein (termed PepM) from the cell wall (11, 46). After purification, the fragment is found to be homogeneous by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, strongly suggesting that the M molecule on the cell surface was also likely to be homogeneous, contrary to previous belief. Depending on the serotype, the size of the extracted PepM molecule varies from 20,000 to 40,000 in molecular weight (34, 116, 120, 133, 152, 169). When injected into rabbits, the pepsin-derived fragment elicits antibodies capable of initiating phagocytosis of streptococci of the homologous serotype. PepM can also absorb type-specific opsonic antibodies from serum (11, 137). Thus, the pepsin-derived M-protein fragment retains some of the biologically important determinants of the native M molecule present on the streptococcal surface. Pepsin extraction, therefore, proved to be one of the better methods for obtaining a workable fragment of the M protein and was used to provide the starting material in a variety of studies over the ensuing years (14, 18, 22, 117, 133, 137, 152, 158, 169). Although it was believed early on that pepsin removed the M protein in its entirety from the streptococcal surface (13), the nature of the complete molecule was not as yet understood.

Nonionic detergent was found to remove M protein from isolated streptococcal cell walls (70). The M molecule released by this method is seen by sodium dodecyl sulfate-polyacrylamide gel electrophoresis as several closely spaced protein bands which range from 15,000 to 35,000 in molecular weight. Like pepsin-extracted M protein, the detergent-extracted molecules are able to absorb opsonic antibodies from rabbit antiserum (70) and evoke type-specific opsonic antibodies when injected into rabbits (64). An examination of the physicochemical properties of these M molecules re-
vealed that they are not globular but asymmetrical, with a frictional ratio of 2.2 \( (70) \). Thus, the detergent-extracted M molecules, though not as homogeneous as the pepsin-derived protein, display many of the properties of PepM.

It was not until 1981, when M protein was purified from the extracellular supernatants of type 12 group A streptococcal L forms and protoplasts, that it became apparent that the size of the native M molecule could be as large as 58 kilodaltons (kDa) \( (199) \). Extraction of M protein from streptococci by solubilization of the cell wall with phage lysin (termed lysin, a muralytic enzyme produced during a group C streptococcal phage infection \( [69] \)) also yielded a large M molecule \( (43, 79) \); however, the association of this M protein with cell wall fragments, resulting in its high molecular weight, could not be ruled out.

**STRUCTURAL ANALYSIS**

**Evidence for the Coiled-Coil Structure of M Protein**

Though sequence and peptide data were limited up to 1979, the available information suggested that M molecules of heterologous serotypes differed in primary structure in spite of a common antiphagocytic activity \( (18, 65) \). How-
ever, despite these apparent differences, physical and chemical data accumulated to that time revealed that all of the M molecules examined shared certain noteworthy properties (81). Their extreme thermal stability (126), elongated shape (70, 84, 166), and appearance as a network of fibers on the cell surface (193) suggested that M molecules resemble previously characterized fibrous proteins. When the partial sequence of the PepM24 (terminology used for the M protein extracted from M24 streptococci with pepsin) protein (18) was compared with that of known proteins, two of the M24 cyanogen bromide peptides exhibited up to 40% identity with rabbit skeletal muscle tropomyosin (99). In addition, a hexapeptide, repeated five times in the PepM24 molecule, was identical to a sequence in tropomyosin. This was the first direct evidence that M protein may be structurally similar to tropomyosin, the prototype molecule for an α-helical coiled-coil fibrillar protein (145). The isoelectric point (a reflection of amino acid composition), solubility, heat stability, and stability in acid and base of tropomyosin were similar to those found for M protein, further confirming the chemical and structural relationship of these two molecules (99).

Additional biochemical support for the probable coiled-coil structure of M protein came when the partial sequences of the M5, M6, and M24 proteins were compared with that of tropomyosin (136). The sequences within these three molecules contained a repeating seven-residue periodicity of nonpolar amino acids, a basic characteristic of α-helical coiled-coil proteins like tropomyosin. A more detailed study of the structure and conformation of the PepM6 and lysin-extracted M6 proteins further established the close relationship between M protein and tropomyosin (169). In this analysis, ultracentrifugation studies revealed that the M molecules exist as stable dimers, and circular dichroism spectra indicated that the molecules are 70% α-helical. Electron microscopic images of platinum-shadowed M molecules were seen as rods in which the pepsin- and detergent-extracted forms were approximately half the length of the lysin-extracted protein. Measurements also revealed that the M molecules were similar in width to tropomyosin, and fiber X-ray diffraction indicated that a coiled-coil structure is present in the M molecule. Assessment of M protein on the streptococcal surface compared with isolated M molecules indicated that each fiber on the cell wall consists of a single dimeric M-protein molecule of about 50 to 60 nm in length (Fig. 1) and is not composed of multimers of smaller subunits as is the case for pili of gram-negative organisms (33). Though common in mammalian cells, the α-helical coiled-coil design is unique for bacterial surface molecules (41, 42).

Studies by Phillips et al. (169) indicated that purified M protein released from the streptococcal cell wall with phage lysin was similar to the size of the native cell-wall-bound M protein.
molecule. M proteins extracted with pepsin are almost half the size of the lysin-extracted molecule and are likely cleaved products derived from the native M molecule. This is further supported by electron micrographs of pepsin-treated M-bearing (M') streptococci depicting shorter surface fibers than untreated cells (13, 169). Thus, after pepsin digestion, roughly half of the M molecule remains associated with the streptococcal cell.

Structural Analysis of the Complete M Molecule

The completion of the PepM5 sequence (133) allowed for the first time a thorough analysis of this region of an M molecule (141), i.e., the distal half of the native protein (169). The results confirmed what previous partial sequence data had suggested: except for the N-terminal 11 amino acids, the remainder of the sequence exhibits a seven-residue periodicity in the distribution of nonpolar amino acids. In addition, it revealed that the N-terminal region of this fragment contains a significantly higher net negative charge than the C-terminal region. While sequence repeats were also identified in the PepM5 molecule (139), they were not as extensive as those found in the partial PepM24 sequence (18).

Until 1986, the majority of structural information was based on the pepsin fragment of the M molecule. No information, besides the physicochemical analysis on the lysin-extracted M protein (169), was available concerning the C-terminal half. The successful cloning of the M6 gene (emm-6) (180) resulted in the ability to determine the complete amino acid sequence of an entire M molecule (94). Like the partial PepM24 sequence (18, 19), the M6 sequence was composed of extensive amino acid repeats extending into the C-terminal half of the protein. Approximately 80% of the complete M6 molecule is composed of four sequence repeat regions (Fig. 6) (94). Regions A and B contain five direct tandem repeat blocks of 14 and 25 amino acids each, respectively. Each A repeat may be further divided into two seven-residue sub-blocks (not shown). The three central repeats in the A and B regions are identical, while the repeats at each end of the five repeat blocks diverge slightly. Region C consists of two and one-half tandem repeats of 42 amino acids each, which are not as conserved as the A and B repeats. Region D is composed of four seven-residue "quasirepeats" (77, 94). Recent reports show that amino acid sequence repeats also occur in the complete M24 (153),
M5 (148), and M12 (175) proteins, the M5 molecule being the least repetitive.

A C-terminal segment, composed of 19 hydrophobic amino acids, is predicted to have an α-helical conformation which would be sufficient in length to traverse the cytoplasmic membrane (77). An adjacent region, rich in proline and glycine, is predicted to be located within the cell wall peptidoglycan (161).

**Basic Model**

In general, α-helical coiled-coil proteins are constructed from a repeating seven-residue amino acid pattern (a-b-c-d-e-f-g)$_n$, in which residues in positions a and d are hydrophobic and form the “core” residues in the coiled-coil and the intervening residues are primarily helix promoting (145). The arrangement of the amino acids of the M6 protein within the seven-residue pattern is shown in Fig. 7. The seven-residue periodicity found from amino acids 12 through 362 strongly suggests that this region is in a coiled-coil conformation and forms the helical central rod region of the molecule. Discontinuities in the heptad pattern (seen especially in repeat B), which have been seen in the sequence of other M proteins (141) and other coiled-coil molecules (144, 163, 166), probably account for the flexibility of the M molecules observed in electron micrographs (Fig. 1). Based on these irregularities in the heptad pattern, the central rod region is divided into three subregions which correlate with the A-, B-, and C-repeat blocks (77).

Based on a detailed analysis of data of M-protein sequences, it was concluded that serologically different M proteins may be built according to a basic scheme: an extended central coiled-coil rod domain flanked by functional end domains (77). Because the α-helical coiled-coil is a structure capable of tolerating a vast number of amino acid sequence changes while preserving its conformation, a large...
FIG. 7. Complete M6 protein sequence arranged to highlight the seven-residue periodicity found in the helical central rod region. Region assignments are based on sequence and conformational analyses (77). Arrangement of the sequence is based on the position of amino acids in a seven-residue periodicity designated by letters a through g beginning at residue 12 and continues, with interruptions at residues 109, 131, 156, 181, 206, 231, and 337, through residue 362. Alignment from residues 363 to 416 is used essentially to highlight the regularity of the position of prolines in the sequence. No periodicity is found from residue 363 to the end. Three major regions are indicated (nonhelical, helical, and anchor). The region buried within the cell wall is delimited by the wall domain (161). The pepsin-susceptible site is between Ala-228 and Lys-229.

Because 3.6 residues are necessary to achieve one turn of the helix (102), hydrophobic residues at positions a and d in the heptad (Fig. 7) will form an inclined stripe around the helix. For thermodynamic stability, a dimeric, ropelike superstructure would be formed as a result of the internalization of the hydrophobic residues from two helical chains coiling around each other (75, 136, 145) (Fig. 8). The presence or absence of disruptions in the periodicity of hydrophobic residues at positions a and d of the heptad.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Amino Acid</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Arg</td>
<td>Val</td>
</tr>
<tr>
<td>2</td>
<td>Val</td>
<td>Phe</td>
</tr>
<tr>
<td>3</td>
<td>Phe</td>
<td>Pro</td>
</tr>
<tr>
<td>4</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>5</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>6</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>7</td>
<td>Thr</td>
<td>Ala</td>
</tr>
<tr>
<td>8</td>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>9</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>10</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>11</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>12</td>
<td>Val</td>
<td>Phe</td>
</tr>
<tr>
<td>13</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td>14</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>15</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>16</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>17</td>
<td>Lys</td>
<td>Val</td>
</tr>
<tr>
<td>18</td>
<td>Val</td>
<td>Thr</td>
</tr>
<tr>
<td>19</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>20</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>21</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>22</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>23</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>24</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>25</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>26</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>27</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>28</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>29</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>30</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>31</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>32</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>33</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>34</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>35</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>36</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>37</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>38</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>39</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>40</td>
<td>Thr</td>
<td>Val</td>
</tr>
<tr>
<td>41</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>42</td>
<td>Lys</td>
<td>Thr</td>
</tr>
</tbody>
</table>

FIG. 8. Proposed model of the M6 protein from strain D471 based on current sequence and structural data (77, 94). The coiled-coil rod region extends about 50 nm from cell wall with a short nonhelical domain at the N terminus distal from the cell surface. The proline-glycine-rich region of the molecule is located within the peptidoglycan, with a short segment of the coiled-coil rod region positioned in the group carbohydrate portion of the wall (161). The membrane anchor segment extends through the cell membrane, with the charged tail projecting into the cytoplasm. Regions containing conserved, variable, and hypervariable epitopes among heterologous M serotypes are indicated.
allows these regions to become more or less flexible (detailed in reference 77).

A short 11-residue nonhelical region is found at the N-terminal end of the coiled-coil rod region, which is distal to the cell surface (Fig. 8). In all M proteins examined thus far, except M49, which lacks this segment (116), the sequence of this region is distinct from one M type to another (94, 133, 148, 153, 175) but is nearly 100% identical within an M type (95). It is not certain whether the tertiary conformation of this N-terminal region is similar among the different M-protein types despite sequence variation. The N-terminal segment plays an important role in the biological activity of the molecule, since antibodies generated to this region effectively opsonize the streptococci of the M type from which it was derived (12, 15, 34, 55, 57, 106, 107) (see below). Thus, it seems likely that one purpose for the central rod region may be to act as a shaft to position the N-terminal region away from the cell surface (Fig. 8).

Beginning at the C-terminal end of the central rod region is a segment characterized by the presence of regularly spaced prolines and glycines (Fig. 7). This is followed by two phenylalanine residues beginning at residue 417 which signal the start of a 19-aminoc-acid stretch of hydrophobic residues terminated by six charged amino acids. It was suggested that the hydrophobic amino acids would serve as a membrane anchor and the adjacent proline-glycine region would be located in the peptidoglycan and act to stabilize the molecule within the cell wall (77). The charged tail, typical of a stop transfer signal seen in the membrane-associated segments of several membrane-bound proteins (178a), would be expected to be located on the cytoplasmic side of the membrane. Sequence comparison of the C-terminal region of the M6 protein with the C-terminal end of other cell wall proteins from gram-positive bacteria reveals significant homology (see below), suggesting that a common mechanism may exist for the attachment of proteins within the gram-positive cell wall.

Analysis of the Wall-Associated Region

From the translated deoxyribonucleic acid (DNA) sequence of the M6 C-terminal region, it was implied that this segment may be responsible for M-protein attachment (94); however, it was not directly determined. Experiments were therefore designed to isolate and identify the M-protein region embedded within the streptococcal cell wall. Trypsin was used to digest the exposed portion of M protein from the cell wall, leaving behind the wall-associated region. After the cell wall was solubilized with a muralytic enzyme, antibodies to synthetic peptides of sequences within the C-terminal region of the M molecule were used to identify the released wall-associated M-protein fragment. A 16-kDa fragment was identified as the smallest trypsin-resistant fragment of M protein. Because of its location within the cell wall matrix, this fragment was resistant to attack by trypsin despite the presence of several lysine and one arginine residue within the sequence. Amino acid composition of this fragment revealed that it was rich in proline and glycine but lacked the amino acids found in the membrane anchor. Sequence analysis of the purified fragment indicated that its N-terminal end was at about residue 300, thus localizing a small segment of the helical central rod region of the M molecule (residues 298 to 362) in addition to the proline-glycine region (residues 363 to 416), within the cell wall (Fig. 7).

Further evidence suggested that residues 298 to 362 of the helical rod region are located within the group carbohydrate moiety of the cell wall, localizing the proline-glycine region within the peptidoglycan. Since no detectable amino sugars were found with the wall-associated M-protein fragment, it is likely that the C-terminal region of the M6 molecule is intercalated within the group carbohydrate and cross-linked peptidoglycan and not covalently linked to them (161).

M protein is rarely detected in the spent growth media of streptococci. Therefore, the incorporation of this molecule into the cell wall and membrane appears to be closely regulated. It is postulated that the C-terminal membrane anchor may be necessary to keep the M molecule attached in the right orientation within the cell membrane until such time as the peptidoglycan is cross-linked (156) around the proline-glycine region. Recent studies (V. Pancholi and V. A. Fischetti, submitted for publication) have revealed that at pH 5.5 in the presence of 30% raffinose, the streptococcal cell wall may be removed with a muralytic enzyme, allowing nearly all the M protein to remain attached to the protoplast membrane. Shifting the protoplasts to pH 7.4 results in the release of the M molecule into the supernatant. A membrane-bound thiol enzyme was found to be responsible for the release. The biological significance of such an enzyme is currently speculative. Since the released form of the M molecule lacks the C-terminal 19 hydrophobic amino acids and charged tail (161), its attachment to the cell appears to be through the cytoplasmic membrane.

Domain Structure

Techniques using limited proteolysis have been effective in defining functional and structural domains in a variety of proteins (215), particularly immunoglobulins (88, 167, 185). Similarly, the PepM fragment of the M protein may be regarded as a functional domain and the pepsin-susceptible site may be seen as a hinge region of the molecule. This preferential site for pepsin at a suboptimal pH is found between the B- and C-repeat blocks in M5 (147), M6 (94), and M24 (153) proteins (Fig. 6) even though the B-repeat block in M24 is not homologous with that found in M5 and M6. PepM49, however, has a primary pepsin site after a C-repeat block and a secondary site between the B and C repeats (135). Thus, since the type-specific and antiphagocytic determinants are located within the PepM fragment, this half of the M molecule may be considered to be a functional domain of the protein. However, besides the wall and membrane anchor domains, the functional attributes of the C-terminal domain comprising the C- and D-repeat blocks have yet to be clearly defined.

Thus, the streptococcus uses the coiled-coil design to accomplish a variety of tasks. The rodlike conformation allows the functional antiphagocytic domain to be positioned away from the cell wall. The plasticity of the sequences necessary to form the coiled-coil structure allows the different domains of the molecule to be tailored for specific functions.

IMMUNOCHEMISTRY

The M-protein molecule is the basis of the Lancefield serotyping scheme of group A streptococci (126, 177, 195). The typing sera used for this purpose are a set of M-protein-specific rabbit sera prepared against whole organisms of specific serotypes. To remove cross-reactive antibodies, sera are extensively absorbed with organisms of selected heterologous serotypes. Nearly 80 to 85 different serotypes have now been identified, but there exist a large
number of strains in streptococcal collections for which typing sera are not as yet available. The identification of a specific strain as being a particular serotype requires that its M-protein extract react in a precipitation reaction with a specific typing serum. Therefore, these procedures tend to identify isolates which have not changed immunodeterminants responsible for the type-specific precipitation reaction with absorbed typing sera (177, 195).

Cross-Reactions

Immunochemical studies prior to 1985 were confined primarily to either acid- or detergent-extracted M protein or to the N-terminal pepsin fragment. Analysis of the reactivity of antibodies raised to detergent-extracted M6 protein with peptides representing nearly the complete M6 protein reveals that the detergent-derived M molecules (DetM) represent the N-terminal half of the native M6 protein (78; V. A. Fischetti, unpublished data) and thus are similar to the pepsin-derived molecule. It is speculated that the nonionic detergent activates an enzyme within the cell wall preparation which cleaves the M molecule in the region of the pepsin-susceptible site.

Acid extracts likely contain fragments of the M protein derived from both N- and C-terminal regions. Shared antigenic determinants among M serotypes have been described (81); however, based on the complexity of the acid M preparation, the nature of the observed cross-reactions was not clearly understood. One hypothesis was that cross-reactive determinants were actually cellular contaminants associated with or linked to the M protein (i.e., M-associated protein [MAP]) (211, 212), whereas others believed that some determinants of the M molecule were shared among certain types (80, 83).

It was clear that immunization of rabbits with the PepM fragment resulted in antibodies which were essentially type specific and opsonic and were free of extensive reactivity to non-type-specific epitopes found previously with acid-extracted M proteins (21). Peptide map analysis of M protein (derived from the N-terminal half of the molecule) from cross-reactive and non-cross-reactive serotypes revealed that cross-reactions correlated with the extent of structural similarity among the M molecules analyzed (65). Limited cross-reactions were observed when 125I-labeled M6 protein, representing the N-terminal half of the molecule, was used in a radioimmunoassay with unabsorbed rabbit antisera prepared to heterologous M serotypes (64). Using monoclonal antibodies, Dale and Bechey (52) found that some antibodies cross-reacted with purified PepM proteins from certain other serotypes; however, the location of these epitopes was not defined. Thus, it was clear that, while cross-reactive epitopes were present in the N-terminal half of the M molecules, they were apparently limited in scope.

Conserved, Variable, and Hypervariable Epitopes

Using polyclonal rabbit antisera to M6 protein, Manjula et al. (132) found cross-reactivity with the C-terminal region of the PepM5 molecule (located near the center of the native molecule), whereas no reactivity was observed with the N-terminal region of the PepM5 molecule. The results suggested that the N-terminal 25% of the native M molecule may represent a hypervariable region of M proteins. In support of this, Dale et al. (55, 57) established that synthetic peptides derived from the N-terminal sequence (residues 1 to 20) of the M5 molecule stimulated type-specific antibodies which did not cross-react with M protein from other serotypes. This was later proven also to be true for M6 (15, 106), M19 (34), and M1 (120) proteins.

Monoclonal antibodies directed to the N- and C-terminal halves of the native M6 protein were used by Jones et al. (109) in an attempt to identify the location of the epitopes accountable for limited and broad cross-reactions among M serotypes. Results revealed that an epitope responsible for cross-reactions among 5 of 56 serotypes was located in the N-terminal half of the molecule (within the B repeat) while a highly cross-reactive epitope (reactive with 20 of 56 serotypes) was located in the C-terminal half (within the C repeat) (Fig. 6). These studies were extended with two other monoclonal antibodies, one reacting with an epitope at the N-terminal end within the A-repeat region and one in the C-terminal half within the C-repeat (106, 108). The C-terminal antibody cross-reacted with 30 different serotype strains, while the new N-terminal monoclonal antibody was found to be type specific and reacted only with the M6 serotype strain. This monoclonal antibody was the only one of the four studied that was opsonic for M6 streptococci.

Thus, it is clear from a variety of immunological techniques that the C-terminal region of the M molecule is antigenically more conserved among different serotypes. As the N-terminal region is approached, the greater the antigenic variation (Fig. 8). Therefore, the basis of the M-protein typing scheme is now more clearly understood. In the preparation of typing sera, absorption of antistreptococcal antisera with heterologous strains was necessary to remove the antibodies reactive with the conserved regions of the M molecule, in addition to antibodies to other common antigens found on the cell surface. This resulted in what would be expected to be a monospecific antibody population, directed to the type-specific N-terminal region of the M molecule (Fischetti, unpublished data).

EVIDENCE FOR TWO DISTINCT CLASSES OF M PROTEIN

MAP andOpacity Factor

MAP were identified by Widdowson and co-workers in the 1970s (211, 213). The MAP antigens are present in acid extracts derived from streptococci which express M protein. These antigens lack type specificity but copurify with the type-specific M protein. At the time MAP antigens were characterized, other investigators had identified non-type-specific M antigens in acid extracts which are similar to MAP (83, 201). It was not until M protein was obtained by pepsin digestion of whole streptococci that the type-specific substance could be clearly separated from non-type-specific M antigens (46).

The MAP derived from one serotype shares antigenic determinants with MAP derived from many heterologous serotypes. Two serologically distinct MAP antigens, MAP I and MAP II, have been identified, and most isolates of a particular M serotype contain either MAP I or MAP II antigen on their surface (211, 214). Expression of MAP II is closely linked to production of opacity factor (OF), a lipoproteinase whose activity is detected by serum opalescence; streptococci bearing MAP I lack OF activity. The MAP I antigen is present on the surface of nearly all M serotypes associated with outbreaks of acute rheumatic fever. Thus, a link between MAP antigen and the virulence properties of group A streptococci is suggested. However, the biochemical composition of MAP I and II remained elusive.
MAP I Antigen and the M6 Protein C-Repeat Domain

In a recent antigenic analysis of surface-exposed M-protein epitopes, it became apparent that two basic M-protein structures exist (28). A panel of seven antibodies directed to well-defined epitopes located in the B-repeat, pepsin site, and C-repeat regions of M6 protein was tested for immunoreactivity to over 130 streptococcal isolates representing more than 60 different serotypes. Isolates of only four serotypes (M5, M14, M19, and M36) shared antigenic determinants located in the B-repeat and pepsin site regions of M6 protein. In contrast, about half of the serotypes tested displayed strong immunoreactivity with antibody probes directed to C-repeat region epitopes of M6 protein. There is a striking correlation between those serotypes possessing the C-repeat region epitopes of M6 and those previously described as having MAP I antigen (211). Furthermore, nearly all isolates which produce OF (and by definition express MAP II antigen) lack immunological cross-reactivity with antibodies to the C-repeat region. It is likely that the MAP I antigen represents the C-repeat region as defined for M6 protein. While it is possible that the MAP I antigen is located on another related surface molecule, the evidence that it is M protein is significant (28, 211, 213). In summary, the MAP I and MAP II M types closely parallel two major structural classes of M-protein molecules, termed classes I and II, respectively. Streptococci bearing class I M proteins share surface-exposed antigenic epitopes with the C-repeat region of M6 protein and fail to produce OF, whereas class II serotypes produce OF but are deficient in epitopes cross-reactive with the M6-like C-repeat region (28).

The close proximity of the highly conserved C-repeat region to the poorly homologous B-repeat and pepsin sites strongly suggests that the C-repeat region forms a distinct antigenic domain (28) (Fig. 6). In support of the immunochemical analysis, structural data suggest that the C-repeat antigenic domain may have unique conformational characteristics as well. The B-repeat region through residue 230 forms a more flexible coiled-coil structure than the C-repeat region beginning at residue 231 (77). In addition, amino acid sequence identity between maximally aligned sequences of M6 (94) and M12 (175) is <27% between residues 121 and 231, but abruptly increases to 97% complete identity between residues 232 and 298 (extending from the beginning of the first C-repeat subblock to the perimeter of the cell wall) (Fig. 7). Thus, the data strongly suggest that the amino-terminal end of the conserved class I M-protein antigenic domain lies at the beginning of the first C-repeat block (28).

Class I and II M proteins appear to lack immunological cross-reactivity in their surface-exposed regions; however, antigenic relatedness between class I and II M proteins can be demonstrated in their extracted forms (28). Whether the shared antigenic epitopes reside in buried portions of the M-protein molecule remains to be established. The degree of homology between M-protein genes of different serotypes has been addressed in nucleic hybridization studies. By using DNA probes corresponding in positions to the C-repeat region and to the cell wall and membrane anchor regions of M6, a class I M protein (147, 182) distinction can be made between class I and II serotypes (28). Hybridization to chromosomal DNA of most class II serotypes is observed only under conditions of low stringency, whereas class I serotypes hybridize under high-stringency conditions. These findings suggest that there are fundamental differences between class I and II M proteins at the DNA level (28).

C-Repeat Domain and Rheumatic Fever

Class I streptococcal serotypes are represented among both skin and nasopharyngeal isolates. Likewise, both throat and skin strains are found among the OF-producing class II serotypes. Thus, the conserved C-repeat domain does not appear to be involved in determining the tissue site of infection. In addition, serotypes associated with acute glomerulonephritis are represented by both class I and II streptococci. However, there is a strong correlation between certain serological M types and outbreaks of acute rheumatic fever (30). Of 16 streptococcal serotypes which have been associated with one or more major outbreaks of rheumatic fever, all but one type express the C-repeat region epitope which defines class I M protein (28). Rheumatic fever patients have high complement-fixing titers to MAP I antigen (211), and MAP I-specific antibodies are absorbed by human heart tissue (213). Thus, the surface-exposed conserved repeat domain of class I serotypes may be a virulence determinant for rheumatic fever (28).

SIZE VARIATION IN THE M MOLECULES

In this and following sections, a direct comparison of the known M-protein sequences is presented. But, to fully appreciate the sequence diversity found within M molecules, it is important to understand a recently observed property of the M-protein molecule, namely, size variation. By using a broadly cross-reactive monoclonal antibody as a probe, the size of lysin-extracted M protein from a number of streptococcal strains was examined by Western blot (immunoblot) (74). M protein derived from streptococcal isolates of 20 different serotypes exhibited variation in size ranging from 41 to 80 kDa, depending on the strain. Similar size variation is observed when different M6 streptococcal strains isolated over a period of 40 years are examined in this way. This variation in size may be explained by the observation that the M sequence contains extensive repeats at the DNA level (94). Long reiterated DNA sequences are likely to serve as substrates for recombinational events or for replicative slippage (62, 192), generating deletions and duplications within the M-protein gene and leading to the production of M proteins of different sizes.

Isolation and Sequence Analysis of Size Mutants

To test this hypothesis, it was necessary to isolate clonally related strains of M-protein size mutants in the laboratory, so that the parent and mutants could be compared. A Western blot system was devised so that M-protein size mutants could be detected and isolated if they comprised at least 5% of the culture. From type 6 strain D471, about 1 in 2,000 colony-forming units (CFU) were found to produce a smaller size M protein, a rate that seemed too high to be the result of spontaneous mutation (71). Analysis of the DNA sequence of the M6 genes of four spontaneous M-protein size mutants confirmed that they arose by recombinational events between reiterated DNA sequences within the structural gene encoding M protein (95). Each of the mutants bearing smaller-size M protein had a deletion of one or more of the A- or B-repeat blocks (Fig. 9).

Comparison of aligned sequences of the M6 size variants indicates that extensive changes can occur within this one strain. While the region on the C-terminal side of the last C-repeat sub-block remains highly conserved (95), the A- and B-repeat segments have undergone duplications and
deletions (Fig. 9). The size mutant derivative of the M6.1 parent, M6.2, deleted two A-repeat blocks, while mutant M6.3, a derivative of M6.2, was unchanged in the A-repeat blocks but is deleted for one B repeat. However, in spite of the extensive variations in the number of repeat blocks, these strains were still typed as M6.

Size variation also occurred in M6 strains isolated from single localized outbreaks of pharyngitis as well as in serial weekly isolates from the throats of individual patients in the 1940s, prior to penicillin therapy (Fig. 10) (74, 95). Based on DNA sequence analysis, the size variation of six clinical M6 isolates is the result of recombination events within the repeat sequences (95), as was found with the mutants isolated in the laboratory. For example, when compared with M6.1, M6 strain MO1015 producing M protein M6.8 (isolated during an M6 outbreak in the 1980s) has only one and one-half A-repeat segments (Fig. 9), whereas M6.9 (isolated in the 1940s) has five A repeats (two of which differ in size from the others) and an extra C-repeat segment (95) (data not shown). Sequence analysis of the M6 genes also indicated that the isolates from each clinical group are clonally related and not the result of separate acquisitions of unrelated M6 strains during the course of the infection (95). When the streptococcal strains isolated weekly from the throats of 15 patients were examined for the size of the M molecule (see Fig. 10 for an example of two such patients), 8 patients harbored streptococci which exhibited changes in M-protein size but not serotype during the course of the infection. The observation that size variants occur among clinical isolates suggests that the size mutant constituted the major organism in the streptococcal population at the time of isolation and suggests that a given size mutant has a selective advantage under clinical conditions.

Thus, it is quite clear that variations can occur in both the number and size of the repeat blocks within the M-protein sequence. Even within an outbreak resulting from a single infecting strain, the M protein from the progeny of the

FIG. 9. Alignment of M6 sequences of four different M6 strains (95). M6.1 from strain D471 is the prototype M6 protein. M6.2 is from a deletion mutant derivative of M6.1 in which two A-repeat blocks have been deleted. M6.3 is a deletion mutant derivative of M6.2 in which one B repeat has been removed. M6.8 from strain MO1015 is a natural M6 isolate from a pharyngitis outbreak in the 1980s. This strain differs from the prototype M6.1 in the nearly complete loss of A repeats. Region assignments are as those described in the legends to Fig. 6 and 7.

FIG. 10. Western blot of lysin extracts of serial streptococcal isolates from two patients with pharyngitis illustrating a change in the size of the M protein. Lanes represent extracts of weekly streptococcal isolates from the time the patients presented with pharyngitis (week 1). Seven weekly isolates of M17 streptococci were obtained without patient treatment in 1942. After week 3, patients are usually asymptomatic and are considered to be streptococcal carriers. The strains were grown and adjusted to comparable optical densities, and the M protein was extracted with phage lysis, which solubilizes the cell wall releasing M protein. After sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Western blot, the nitrocellulose was probed with a monoclonal antibody reactive with M protein. The blots represent the typical multiple banding pattern of M protein extracted by this method (74). It can be seen that in patient GL53 the size of the M protein produced by the streptococci was the same to the fifth weekly isolate. At week 6, the streptococci isolated from this individual synthesized a smaller M protein which changed again in week 7 (though all still typed as M17). In patient GL44, the size of the M17 molecule remained the same for 4 weeks, at which time the population of streptococci isolated produced a smaller M molecule (weeks 5 to 7). At week 7, the amount of M protein on these organisms was reduced, which is typical of organisms isolated in the carrier state.
Infecting streptococcus isolated from infected individuals exhibits size differences.

**ANTIGENIC VARIATION AND ITS RELATION TO SIZE VARIATION**

Antigenic diversity of surface antigens is a mechanism by which successful pathogens have been able to survive in the environment after infecting immunocompetent hosts (7, 45, 146, 194). In bacteria, the best-studied variable antigens are the gonococcal pilin (146, 194) and the *Borrelia* variable major protein (7). M protein differs from these other antigenically variable surface proteins in that they undergo gene conversion events involving extragenic recombination between multiple genes, whereas the single copy of *emm*-6 in streptococci (see below) must involve intragenic recombination events for changes to result.

Clues as to the mechanism of antigenic variation of the M protein are based on two previously mentioned observations: size variation of the M molecule (74) and the presence of repeat sequences within the M-protein gene (94). Based on these observations, it was hypothesized that changes in the amino acid sequence may arise as a result of recombination between inexact repeats. Because the end repeat blocks within the A- and B-repeat regions of the M6 sequence diverge slightly from the consensus sequence of the three internal blocks, recombination between one of these terminal blocks and any other central block can generate a new sequence. These changes may then be amplified by recombination events within these repeated sequences. Sequences involved in antigenic determinants may be duplicated or in some cases deleted by such recombination events (95). Thus, intragenic recombination events would be expected to accelerate the rate at which alterations accumulate in the M-protein gene. Such recombination events were found to have occurred in the A- and B-repeat regions of deletion mutants isolated in the laboratory from strain D471 (Fig. 9). In mutant M6.5, derived from M6.1, a deletion of the equivalent of two repeat blocks occurred from a recombination event beginning from the center of the divergent A1 repeat block to the center of the consensus A3 block. This resulted in a sequence change from an Asp-Lys (D-K) to an Asp-Asn (D-N) within this mutant (107) (Fig. 11).

**Size Variation Leads to Antigenic Variation**

To test the hypothesis that size changes in the *emm*-6 gene can lead to antigenic variation of the M protein, monoclonal and polyclonal antibodies directed to defined epitopes on the parental M6 molecule were used in competition assays with purified M protein isolated from the parental M6 strain and size mutants (107). The results revealed that antigenic changes occurred in epitopes located at or near the deletions within mutant M6 proteins, but not at more distant epitopes. The studies also revealed that an M6 monoclonal antibody opsonic for the parental M6 strain lost its ability to opsonize one of the M6 size mutants, supporting the view that size changes may also result in changes in antigenic determinants responsible for streptococcal survival in an immune environment. Thus, antibodies originally produced in response to sequences present in the parental M6 molecule were unable to bind to mutant M6 protein or opsonize certain M6 size mutants. It was discovered that, in some size mutants in which a direct sequence change had not occurred, changes in conformation were predominant (107). Therefore, a change in the size of the M molecule may be an attempt on the part of the streptococcus to evade immune clearance. While this may be true for changes in the opsonogenic A-repeat region, antibodies directed to the B- and C-repeat regions have been shown not to be opsonic for the streptococcus (106). Thus, the pressure for these regions to change may be at a different level, i.e., adherence and colonization.

**Natural Selection of Size Mutants**

Group A streptococcal pharyngitis is a self-limiting disease in which the production of opsonic antibodies to the M molecule results in the elimination of clinical illness (128). However, in some instances, after symptoms subside and in the presence of opsonic antibodies, the infected individual...
may carry the streptococcus in the throat for several months without clinical signs (termed carrier state) (86, 92, 111). It is clear that more than one osnonogenic epitope is present within the N-terminal region of the M molecule and antibodies to only one epitope are adequate for the phagocytosis of these organisms (15, 19, 106, 179). Whereas antigenic variation in the gonococcus and Borrelia surface proteins is the result of the exchange of large genetic blocks, in the streptococcus only a few amino acids are changed at a time. Therefore, M-protein variants created during a streptococcal infection may not be suitably altered in all osnonogenic epitopes to escape immune recognition and clearance. However, prior to clearance from the infective or carrier state, the streptococcus has the ability to infect other human hosts (203). Thus, M-protein mutants produced in the initial host during clinical illness or during the carrier state are able to be disseminated throughout the population, leading in time to the accumulation of changes within the M molecule. Consequently, through immunological pressure, the antigenic structure of the M protein is continuously evolving by the selection of organisms producing variant M molecules. This process is likely to result in a continuum of antigenically unique M molecules among the streptococci in the environment.

Variation in the N-Terminal Nonrepeating Segment

The N-terminal nonrepeating segment of the M molecule contains protective epitopes. While variation within the repeat segments may be attributed to intragenic recombination events, no information is currently available regarding the method by which sequence changes occur within the short nonrepeating segment. Sequence analysis of M6 from several strains isolated from the 1940s to the 1980s revealed that, while variation is seen in the A, B, and C repeats, the sequence from the N terminus to the first A repeat (about 30 amino acids depending on the serotype) has been conserved (95); S. Hollingshead, unpublished data). While current data suggest that no other homologous regions exist outside the emm-6 gene (182, 183), it has been suggested that this N-terminal segment could arise from recombination between extragenic non-omm-6 segments (148). Attempting to explain the mechanism of variation of this N-terminal region among M serotypes, Haanes-Fritz et al. (90) examined the DNA sequence of segments outside the emm-6 structural gene of M1, M6, M12, and M24. They found that, although the 5' sequences of the four mature M proteins exhibited considerable variability, extensive similarities were observed in the upstream flanking regions and signal peptides. The authors proposed that these conserved flanking regions may be involved in rare occurrences of homologous recombination with regions outside the M gene, resulting in the insertion of new variable segments. To date, however, the existence of these extragenic regions has not been established, which may be a consequence of the use of unsuitable probes.

An alternative interpretation of the apparent stability of N-terminal sequences within a serotype is based on the method used to type the strains used in sequence studies. As mentioned earlier, the M typing sera are highly absorbed, containing antibodies that are specific for a given serotype M protein; i.e., the typing sera define the serotype. Since the N-terminal end is the most variable region of the M molecules, antibodies in these typing sera are naturally directed to epitopes in the N-terminal extreme. Thus, when M6 strains are selected for M6 protein analysis, it stands to reason that they all will likely have a similar N-terminal sequence. An M6 strain which has diverged sufficiently in the N-terminal sequence to prevent the typing serum from reacting (a single amino acid change could accomplish this) would not be typed as M6 and, thus, would not be analyzed as an M6 protein. Since more than half of the streptococcal strains isolated in nature are classified as nontypable (173), it is conceivable that many of these organisms are derivatives of nontypable strains having sequence changes within the nonrepetitive N-terminal region. Since the sequences of only a few M molecules have been analyzed to date (none of which are derived from nontypable streptococci), it is difficult to determine which types, if any, are related to one another with respect to the sequence in the N-terminal extreme.

COMPARISON OF M-PROTEIN SEQUENCES

Alignment of Known M-Protein Sequences

In addition to the complete M6 sequence from strain D471 (94), sequence data are now available for a number of other streptococcal serotypes. The sequences are available for the complete M5 (strain Manfredo) (148), M24 (strain Vaughan) (153), and M12 (strain CS24) (175) proteins and the complete pepsin fragments of M5 (strain B788) (133) and M49 (strain B915) (116) and partial M1 (strain T1/195/2) (120, 152) and

FIG. 12. Alignment of the sequences of three different serotype M proteins (M6.1 [strain D471], M5 [strain Manfredo], and M24 [strain Vaughan]) (94, 148, 153). N terminal to the Pro-Gly region, each M protein differs in the size and number of repeat regions. M24 is lacking A-repeat sequences. Amino acid identity of \( >98\% \) is found within the Pro-Gly and membrane anchor regions among the three M proteins. Sequence variation increases as the N terminus is approached.
Therefore, the 

VOL. 

strains 

residue 

insertion 

insertion 

in 

greatly 

and 

nonrepetitive region 

these residues in 

the MS 

sequences 

among 

one-third 

C-terminal 

(SS400), M24 (strain 

(unpublished strain) (34) proteins. 

For instance, 

than 

M5 and 

M6 and M24 

serotypes 

in 

addition 

number and size 

serotypes 

in 

sequence; 

however, 

seen 

still exist 

(116). 

The homology of 

sequence 

A-repeat region of 

M5 (54% identity) (Fig. 12). The homology of 

sequence 

wall), and, depending 

on 

mutations 

on 

conserved region (>98% identity) is found C terminal to the 

D repeat (located within the cell wall), and, depending on 

the 

serotypes being compared, sequence variation increases 

among these types as the N terminus is approached (Fig. 8). 

Strong sequence homology is observed among these three 

types within the C-repeat region; however, this homology is 

greatly reduced in the B-repeat region and nearly lost in the 

A-repeat region. The extent of homology depends on the 

serotypes being compared, some types being more related 

than others. For instance, the C-repeat blocks between M6 

and M5 and between M6 and M24 are about 80% identical; 

however, the B blocks in M6 are not very homologous to 
those in M24 (18% identity) but are significantly homologous 
to those in M5 (54% identity) (Fig. 12). The homology of the 

B blocks between M5 and M24 is <10%. In the M24 

sequence, all of the A repeats have been deleted, leaving an 

extended nonrepetitive region N terminal to the B repeat. A 

similar pattern is seen in the sequence of the pepn fragment 
of the M49 sequence; however, in this sequence, remnants of 
an A-repeat region still exist (not shown) (116).

The preceding discussion serves to illustrate the differences in the sequence and repeat blocks for the M protein from these specific serotype strains. M proteins from different strains of the same serotype are likely to differ in size from the prototype strain and will exhibit differences in sequence as well as immunodeterminants depending on the nature of the size variation. For instance, the major difference between the N-terminal half of M5 from strain B788 (112) and M5 from strain Manfredo (148) is a 14-residue insertion in the A-repeat region of strain B788 and a 25-residue insertion in the B-repeat region of strain Manfredo. Therefore, strain designations are critical when comparisons 

are to be made between different M proteins. Although very few M sequences are currently known, it is clear that certain serotypes are more closely related than others based on the degree of homology within the B- and C-repeat segments (28, 108, 109). Perhaps, as more data are gathered on other M proteins, the evolution of the M serotypes will become more apparent.

Alignment of the Variable Region of Two Different M Proteins

The sequence of PepM5 protein from strain B788 (133), comprising the N-terminal half of the native M5 molecule (to the C-terminal end of the B-repeat region), was compared with the corresponding region of the M6.1 sequence (from strain D471) (77, 94). Maximal homology is achieved if a gap of 29 residues is introduced in the M5 sequence. This is necessary since the M molecules from these two strains differ in size, M5 being smaller (Fig. 13) (74). The first 85 residues of the M6 sequence show 33 identities (39%) with the PepM5 molecule. Homology of 56% is observed between residues 115 to 228 of the M6 sequence and 85 to 197 of PepM5. This increases to 77% identity between residues 164 to 228 of M6 and 134 to 197 of M5. In all, 43% identity is observed between these two segments.

Although the number of identities within the hypervariable region of these two molecules (residues 1 to 163 of M6 and 1 to 133 of M5) is far less than those in the region C terminal to these segments, half (22 of 47) of the identities are either asparagine or hydrophobic amino acids that occupy core residues at positions a and d of the heptad periodicity (Fig. 7). It is clear that hydrophobic residues in these positions are required for the maintenance of the coiled-coil structure of the M protein (136, 164). Therefore, in spite of considerable sequence dissimilarity between these two M molecules within the hypervariable region, identities are preserved in critical structural positions. Thus, the maintenance of the coiled-coil rod region appears to be a biologically significant characteristic of the M molecule (76, 77, 141).
HOMOLOGY OF M PROTEIN WITH OTHER MOLECULES OF GRAM-POSITIVE BACTERIA

Group C and G Streptococci

M protein or M-like proteins of group A streptococci have been reported in group B (142), C (142, 217), E (58), and G (142, 143) streptococci. These conclusions are based either on serological cross-reactions with typing sera developed to M protein from group A organisms or on the ability of these organisms to resist phagocytosis. DNA hybridization analysis was performed to define better the relationship of the M protein from group A streptococci with that of other streptococcal groups (183). By using nearly the complete M6 protein gene as a probe, homologous DNA is found in group C and G streptococci but not other groups (183). Group C streptococci are primarily animal pathogens but have been implicated in serious human infections (151). Group G organisms are found as part of the normal flora of the skin, upper respiratory tract, and female genital tract. However, these organisms are responsible for an increasing number of human diseases (4, 6, 142).

Monoclonal antibodies prepared to the M6 protein from group A streptococci cross-react with a protein found on group G clinical isolates (105). More than half of the group G strains examined are resistant to phagocytosis in human blood. When surface protein is extracted from group G cells with the muriarly enzyme mutanolysin and reacted in Western blots with M-protein-specific monoclonal antibodies, the protein pattern resembles that seen with M protein from group A organisms. A highly purified, pepsin-derivated protein fragment from a group G strain was capable of inducing antibodies in rabbits which opsonized group G cells in a phagocytosis assay (105). In a separate study, group G streptococcal clinical isolates were found to survive in human blood and shown to exhibit surface fibers by electron microscopy. Rabbit antisera to either whole group G streptococci or partially purified pepsin extracts were also opsonic for homologous isolates (31). Using DNA probes representing the C-terminal and leader sequences of the M12 protein gene, Simpson et al. (187) found that DNA from group G strains which infect humans is related to the group A streptococcal M-protein DNA. This relationship was not seen in DNA from group G strains which cause animal diseases. Thus, M proteins on group A and G streptococci, especially group G organisms involved in human disease, appear to be functionally and structurally similar.

A second protein located on the surface of group G streptococci (protein G) has sequence homology with group A M protein. Protein G is both an immunoglobulin G (IgG)-binding protein that binds all four classes of human IgG and animal IgG classes and an albumin-binding protein (2). Analysis of the DNA sequence reveals that the C-terminal end of protein G is highly homologous to the M-protein C-terminal membrane anchor and a small portion of the wall domain, with 56% identical residues (61, 159) (Fig. 14). Although there is some sequence homology between the remaining portion of protein G and M protein, the most striking feature lies in the presence of sequence repeat blocks. Like M protein, the repeat blocks divide protein G into distinct domains. A repeat domain located within the C-terminal half of the molecule, proximal to the cell wall, is responsible for the IgG-binding activity of protein G (89), while the albumin-binding activity is located in the N-terminal repeat domain (3). Although protein G is rich in a-helix-forming residues, a detailed analysis concerning its conformation as a coiled-coil structure has not as yet been performed.

S. equi, a Lancefield group C organism, is the causative agent of "strangles," a disease of horses resembling human pharyngitis. Reports of an M-like molecule with the ability to render these streptococci resistant to phagocytosis are well documented (189, 196, 217). The interesting characteristic of the M or M-like protein on S. equi is the lack of different serotypes (8, 9). For instance, Bazeley reports that an antiserum prepared against one strain of S. equi protected mice from challenge by each of 32 different strains isolated from various parts of Australia (8). Thus, antigenic diversity does not appear to exist in this group C antiphagocytic molecule as it does in the group A M protein, even though many of the characteristics of the cloned S. equi M-like protein are shared with group A M protein (85).

Staphylococcal Protein A

Protein A is an IgG-binding protein from staphylococci which also has features in common with streptococcal M protein. Like protein G, the region of sequence homology with M protein is located at the C terminus of protein A, where it exhibits 22% identity with M protein (89a) and 35% with protein G (Fig. 14) (61, 159). The homology among these three proteins is even more striking when conservation of charge or hydrophobicity is considered. The conservation of sequence and chemical properties of amino acids within the C-terminal region suggests that there likely exists a common mechanism by which protein A, protein G, and M protein are anchored to the gram-positive cell wall. Although protein A does not exhibit a-helical coiled-coil potential (159), the repetitive nature of the sequence (198) is reminiscent of the repeats found in both M protein (94) and protein G (61). While serological differences between protein A molecules on various staphylococcal strains has not yet been reported, the presence of size variation among clinical isolates has recently been described (38).
TABLE 1. Sequence homology of M protein with mammalian proteins

<table>
<thead>
<tr>
<th>M protein</th>
<th>Mammalian protein</th>
<th>% Identity (amino acid overlap)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>M24</td>
<td>Myosin heavy chain</td>
<td>34 (44)</td>
</tr>
<tr>
<td></td>
<td>(rabbit cardiac)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tropomyosin (human cardiac)</td>
<td>33 (42)</td>
</tr>
<tr>
<td></td>
<td>Keratin (type I, bovine)</td>
<td>40 (30)</td>
</tr>
<tr>
<td></td>
<td>Lamin (human)</td>
<td>36 (45)</td>
</tr>
<tr>
<td>M5</td>
<td>Tropomyosin (human fibroblast)</td>
<td>32 (28)</td>
</tr>
<tr>
<td></td>
<td>Myosin heavy chain</td>
<td>39 (41)</td>
</tr>
<tr>
<td></td>
<td>(rabbit cardiac)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keratin (type I, bovine)</td>
<td>43 (30)</td>
</tr>
<tr>
<td></td>
<td>Tropomyosin (human cardiac)</td>
<td>33 (45)</td>
</tr>
<tr>
<td>M6</td>
<td>Myosin (rat cardiac)</td>
<td>38 (50)</td>
</tr>
<tr>
<td></td>
<td>Keratin (type I, bovine)</td>
<td>44 (32)</td>
</tr>
<tr>
<td></td>
<td>Tropomyosin (human cardiac)</td>
<td>38 (50)</td>
</tr>
<tr>
<td>M12</td>
<td>Keratin (type I, human)</td>
<td>41 (32)</td>
</tr>
<tr>
<td></td>
<td>Myosin heavy chain</td>
<td>31 (32)</td>
</tr>
</tbody>
</table>

* Identity based on fastp analysis of known sequences (130) and inspection.

STRUCTURAL RELATIONSHIP OF M PROTEIN WITH MAMMALIAN PROTEINS

Tropomyosin

M protein resembles muscle α-tropomyosin on the basis of physicochemical properties and sequence homologies (99). As described previously, N-terminal sequences of certain M24 peptides have up to 40% identity with tropomyosin, and a hexapeptide of tropomyosin appears five times in the M24 protein (99). In subsequent studies, homologies with tropomyosin were also observed with other M sequences, namely, M5 and M6 (136). Many of the observed identities correspond to the core hydrophobic amino acids in positions a and d in the seven-residue repeat pattern necessary to maintain the coiled-coil (refer to Fig. 7 and section on structure) (77, 136, 164). However, in certain regions, identities are also found in externally positioned amino acids. It was suggested that the correlation between sequences in M protein and tropomyosin provides direct evidence, at a molecular level, of a structural similarity between a streptococcal and a mammalian protein, and this similarity may be relevant in the pathogenesis of streptococcal diseases (99, 130).

Other Coiled-Coil Molecules

An analysis of the PepM5 sequence revealed that the periodicity and sequence of this M protein more closely resemble the coiled-coil rod region of myosin than tropomyosin (138, 141). Sequence comparison of the complete M5 (148), M6 (77, 94), M12 (175), and M24 (153) proteins with known protein sequences in a data base reveals the best homology to be primarily with myosin and tropomyosin (Table 1). Other coiled-coil proteins such as keratin and lamin also show significant homology with M proteins.

The basic structure of the M6 protein (77) is remarkably close to that of eukaryotic intermediate filaments (keratin, desmin, and vimentin), whose elemental structures are also based on a coiled-coil central rod region with structurally different end regions (166, 190). Its superstructure is composed of two to three α-helices wrapped around each other, and, like M protein, intermediate filaments from different sources vary immunologically as well as in size and composition (190). While it would be tempting to speculate on the possible role of these homologies in the pathogenesis of certain streptococcus-related diseases, it would be best to await the results of future experiments.

IMMUNOLOGICAL CROSS-REACTIVITY BETWEEN M PROTEIN AND MAMMALIAN PROTEINS

Heart-Reactive Antibodies

Cross-reactions have been observed between streptococcal components and mammalian tissue. Kaplan and Meyersian (113) described a component of M5 and M19 streptococcal cell walls which induces antibodies reactive with muscle tissue and correlates with M protein. Zabriskie and Freimer (219) observed that purified streptococcal membranes evoked cross-reactive antibodies which reacted with human cardiac muscle sarcolemma. While subsequent attempts were made to identify these cross-reactive antigens from the streptococcus (200), they have not been characterized sufficiently to determine their identity.

Immunodeterminants Shared with Myosin and Other Human Proteins

Dale and Beachey (54) discovered that rabbits immunized with PepM5 protein produced antibodies which cross-react with myosin but not tropomyosin or actin. By enzyme-linked immunosorbent assay, the cross-reactive epitope is also found in the M6 and M19 proteins and is exposed on the surface of the streptococcus. Furthermore, these investigators found that the M5 molecule contains at least three epitopes which stimulate antibodies to human sarcolemmal membranes and are partially shared with M6 and M19 proteins (53). One of the myosin cross-reactive epitopes was localized to amino acids 84 to 116 of the M5 sequence (56).

More recently, Sargent et al. (179), using antibodies to eight synthetic peptides representing the entire PepM5 protein, determined that peptide 164-197 stimulates antibodies in rabbits which react with isolated cardiac sarcolemmal membranes but not with purified myosin. The reactivity is localized to a 40-kDa protein in the sarcolemmal membrane. The results indicate that the epitope located in the 164-197 sequence is shared with M6 but not M24 or M19. This sequence is located close to the center of the native M5 molecule (within the B-repeat region), N terminal to the pepsin-susceptible site. This segment is highly conserved between the M5 and M6 proteins (77) (Fig. 13) but not M24.

Synthetic peptides representing the N-terminal 20 amino acids of M5 (57) and M6 (15) proteins were found to evoke type-specific protective but not heart-reactive antibodies. These studies suggested that, by using sequences from the N-terminal end of the M molecule, heart-reactive epitopes might be avoided. However, subsequent experiments (34) revealed that PepM19 and synthetic peptides representing residues 1 to 24 of M19 evoked both opsonic antibodies and heart-reactive antibodies to a sarcolemmal protein, indicating that epitopes stimulating autoimmune may also be localized within the type-specific N-terminal region.

Renal Tissue-Associated Epitopes Shared with M Protein

Sera from patients with poststreptococcal glomerulonephritis are shown to have antibodies against basement membrane antigens (63). The antigens involved in these reactions are glomerular heparin-sulfate proteoglycan and basement
membrane type IV collagen and laminin. It is speculated that the cross-reactivity may be related to structural similarity between the streptococcal M protein and basement membrane antigens such as laminin. More recently, it has been shown that monoclonal antibodies prepared to renal cortex cross-react with M proteins from type 6 and 12 streptococci (87). These studies were extended to type 1 M protein, in which antisera raised in rabbits to a synthetic peptide representing the N-terminal 1 to 26 amino acids and a similar peptide lacking residues 20 to 22 both were able to opsonize M1 streptococci and react with human renal glomeruli by immunofluorescence (119). The cross-reactive epitope was localized to the tetrapeptide Ile-Arg-Leu-Arg located at residues 23 to 26 in the M1 peptide, while the region responsible for eliciting the opsonic antibodies was residues 1 to 20. The Ile-Arg-Leu-Arg sequence has not been found in other M-protein sequences such as M5, M6, and M24, which are not considered nephritogenic serotypes. However, it was also not part of the M49 sequence, a nephritis-associated serotype (116). The authors state that this result indicates that the observed tetrapeptide may not be the only sequence defining a renal autoimmune epitope (119).

Monoclonal Antibodies Cross-Reactive with Coiled-Coil Proteins

Cunningham and Russell (49) prepared monoclonal antibodies to streptococcal membrane from M5 streptococci which cross-react with detergent-extracted human heart sarcolemmal antigens. The monoclonal antibodies react more strongly with heart antigens than with kidney or skeletal muscle antigens. Some of the monoclonal antibodies are directed to skeletal muscle myosin, and one is reactive with ventricular myosin. The myosin-reactive antibody could be absorbed with either myosin or whole streptococci (121). Two monoclonal antibodies reactive with M5 protein and a streptococcal membrane protein were found to be polyspecific. One reacts with DNA, polyinosinic acid, polydeoxythymidilic acid, and cardiolipin, and its reactivity with the cytoskeleton could be blocked with antivimentin (50); the other reacts with the cytoskeleton and is inhibited by the α-helical proteins myosin, actin, and keratin. Further characterization of these monoclonal antibodies revealed that they are specific for epitopes in the heavy chain of myosin: one for light meromyosin and the other for both light and heavy meromyosin. Both of these myosin fragments contain a portion of the coiled-coil rod region. The authors speculate that the induction of antibody-producing clones against myosin by a streptococcal antigen may explain at least some of the cross-reactive responses seen during streptococcal sequelae. Recently, Fenderson et al. (62a) described the production of monoclonal antibodies which react with M5 and M6 proteins and human and rabbit tropomyosin. At least one of these cross-reactive monoclonal antibodies also exhibits strong reactivity with human heart tissue by immunofluorescence.

Conformational Determinants

Since many of the M-protein cross-reactive antibodies (48, 50, 56, 119, 121) also react with other coiled-coil proteins (i.e., myosin, tropomyosin, keratin, vimentin, etc.), it is likely that in several cases the immunological recognition is based on a common conformational determinant. Sequence alignments between M protein and tropomyosin (Fig. 15), as well as between other cross-reactive coiled-coil molecules, do not reveal a common linear sequence (Fischetti, unpublished data). Identities are usually distributed among several nonadjacent amino acids within a localized region of these molecules. This is not surprising based on the molecular constraints imposed on a coiled-coil structure, i.e., where every first and fourth amino acid in the heptad repeat are
internalized in the formation of the coiled-coil, leaving the intervening residues in external positions (145). For example, an analysis of the position of the identities within the seven-residue repeats found in an alignment between M6 protein and tropomyosin, reveals that they appear primarily in positions b, c, g, e, and f, all of which are located nearly contiguous to one another on the outer surface of the coiled-coil structure (Fig. 15). Thus, it is possible that the spatial distribution of these residues in two conformationally similar molecules could form the basis of a common immunodeterminant.

Consequently, since streptococci are capable of varying the M-protein sequence (95, 107), certain sequences found within this changing molecule are likely to be homologous with sequences located in relatively stable mammalian coiled-coil proteins. Thus, as seen with the analysis of the M5, M6, M12, and M24 proteins, each shares homology with myosin, tropomyosin, and other coiled-coil proteins to differing degrees as well as to different regions of the respective molecules (Table 1). While some homologous regions may form cross-reactive determinants, the significance of these determinants in the pathogenesis of streptococcal diseases is just beginning to be understood.

THE enm GENE

As noted previously, no information regarding the C-terminal half of the M molecule was available until the gene for M6 protein was first cloned from streptococcal strain D471 (emm-6.1) into Escherichia coli (94, 180). The M protein produced in E. coli (ColiM6) was shown to react with both monoclonal and polyclonal antibodies prepared to the streptococcus-derived M6 protein. The cloned molecule accumulates in the periplasmic space of E. coli, and no M protein is detected on the surface of E. coli or in the growth supernatant (73). A small amount of M protein is present in the cytoplasm, where its molecular size is larger than the periplasmic form. This is consistent with the idea that the larger molecule retains the hydrophobic leader sequence prior to export to the periplasm.

The M-protein genes from M5 (115), M12 (175), and M24 (153) streptococcal serotypes have also been cloned and shown to produce immunologically active M proteins. On sodium dodecyl sulfate-polyacrylamide gel electrophoresis, the cloned M6 molecule appears as a close triplet of bands, a characteristic also seen with the cloned M5 protein (171). The reason for the multiple banding pattern is not entirely understood, but seems to be associated with the C-terminal region, since pepsin cleavage of ColiM6 results in a C-terminal fragment which retains the triple banded structure (72) while the N-terminal half (or PepM) is homogeneous (21, 34, 72, 137).

Relationship of M-Protein Genes

The relationship of the emm-6 gene in M6 streptococci to emm genes in other streptococcal serotypes was determined by hybridization experiments, using the emm-6.1 gene as a probe (183). The results revealed that DNA from all of the 56 different serotypes tested contain homologous emm genes. Among the group A streptococci tested in these studies were several which were functionally M+ (i.e., those phagocytized in human blood). An emm gene could be identified in all but one M− strain. The reason for the M− phenotype in emm-containing strains is not known, but may reflect a defect at the level of translation or transcription of M protein.

Localization of the region in the emm gene responsible for hybridization with all tested serotypes was accomplished by hybridizing DNA from 10 different serotypes with smaller probes derived from the emm-6.1 gene. This investigation revealed that the 3′ region of the emm-6.1 gene is conserved among these serotypes while the 5′ segment is variable (182). A similar result was also reported when DNA probes corresponding to the 5′ and 3′ ends of the emm-5 gene (their snap5 notation for the gene) were used in hybridization studies with DNA from heterologous M serotypes (114). These findings corroborated earlier results of studies which used monoclonal antibodies with defined epitopes to identify the variable and conserved segments of the M molecule (109).

Ribonucleic acid sequence analysis of emm gene transcripts primed with specific oligonucleotide probes from emm-6.1 was used to determine the degree of homology within the 3′ end of the emm-5, emm-19, emm-24, emm-30, and emm-35 genes (96). The results revealed that the C-terminal one-third of these genes are >95% identical to emm-6.1. This was later confirmed with the reported sequence from other strains of M5 (148) and M24 (153) proteins.

Conversion of M− to M+ Phenotype

DNA hybridization studies demonstrated that M− strain T28/51/4 lacked an emm gene (183), and electron microscopy revealed that this strain has no surface fibrils associated with the presence of M protein (193). In addition, this strain is easily phagocytized in normal human blood (193). With a specially constructed shuttle vector, experiments were designed to move the cloned emm-6.1 gene into this emm − deficient streptococcus (181). The resultant transconjugant was converted to M +, resistant to phagocytosis in human blood, and opsonized by anti-M6 antiserum. Immunofluorescence studies revealed that M6 protein was located on the cell surface of the transconjugant. When inoculated into rabbits, the transconjugant induced opsonic antibodies to M6 streptococci. In a similar set of experiments, the emm-5 gene was transformed into S. sanguis, and M5 protein was demonstrated on the cell surface by immunofluorescence (171). Although producing less M protein than authentic M5 streptococci, these organisms removed opsonic antibodies from anti-M5 serum, indicating that the antiphagocytic determinants are expressed on the surface of S. sanguis. Thus, the emm-6.1 and emm-5 genes contain all information necessary to convert an M − streptococcus to a biologically and immunologically M+ organism, further emphasizing the role played by M protein in the antiphagocytic capacity of streptococci.

Number of emm Gene Copies

As stated earlier, one copy of the emm gene is detected in an M6 strain (183) and in the chromosome of nine different serotypes as determined by Southern hybridization with M6 gene probes (182). This supports the hypothesis that recombination leading to size and antigenic variation is an intragenic event. However, in contrast, two bands were found to hybridize with a digest of the M5 chromosomal DNA. This result may be due to restriction site heterogeneity for this M5 strain or the presence of two emm-5 gene copies in the M5 organisms. The latter possibility is supported by Kehoe et al. (115), who identified two copies of an emm-5 gene in a strain different from the one used in the studies by Scott et al. (182). Since these were the only two reported studies of
emm-5 genes, it is not certain whether the presence of two emm genes in M5 streptococci is the rule or whether these strains are unusual.

Regulation of the emm Gene

Lancefield observed that streptococcal strains maintained in the laboratory for extended periods of time (several weeks to a few months) lose their ability to survive in human blood (129). Quantitative measurements of the M-protein content suggest that these streptococci do not contain sufficient M protein for survival (24, 129). Passage of these organisms through mice or human blood results in the surviving streptococci producing higher quantities of M protein (24), suggesting that a regulatory mechanism may be responsible for M-protein expression.

Cleary et al. (40) found that M12 strain CS44 produced M mutants at high frequency which correlated with a variation in their M-protein synthesis. Spanier et al. (188) isolated two M mutants from this population, which, by Southern blot analysis, were found to contain a small deletion in a region upstream from the M12 structural gene promoter. In a recent analysis of these strains and derivatives thereof, it was determined that the deletion was about 500 base pairs upstream of the emm-12 gene (186). Although this characteristic could not be confirmed in other M serotypes, these results suggest that a positive regulator may be located upstream from the M-protein structural gene.

A positive regulator for the production of M protein was also suggested by the experiments of Scott et al. (181), in which a plasmid containing the emm-6.1 gene and its promoter was transferred to a strain lacking an emm gene (T28/51/4). Even though the plasmid was present in a high copy number, the production of M6 protein was about 30- to 50-fold less than in the parental M6 strain from which the emm-6.1 gene was derived. They suggested that the emm-deleted strain may also be deleted for a positive regulator necessary for wild-type expression. By using Tn916 mutagenesis in D471, the strain from which the emm-6.1 gene was derived, a mutant was found which produced 50-fold less M protein than the parental strain. By Southern blot analysis, it was shown that the Tn insertion was about 2 kilobases upstream from the emm-6.1 structural gene. Northern (ribonuclease acid) blot analysis revealed that the mutant did not produce any detectable emm-6.1-specific messenger ribonucleic acid. Thus, the authors named the gene responsible for this effect mry, for M-protein ribonuclease acid yield.

In support of the hypothesis that the emm-deleted strain T28/51/4 was unable to sustain full emm expression, Southern blot hybridization revealed that this strain was in fact deleted for the mry gene (35). Thus, a positive regulator appears to be necessary for high M-protein expression. Whether environmental factors play a role in this expression or whether, like certain positive regulators in other pathogens (149, 206), mry controls other virulence determinants in the streptococcus requires further study.

HUMAN IMMUNE RESPONSE TO THE M MOLECULE

In earlier experiments designed to examine the human immune response to the M protein as a result of a streptococcal infection, it was clear that, once infected with a specific streptococcal serotype, an individual remained protected from subsequent infection by strains of that serotype but was susceptible to other streptococcal types (see review by Lancefield [129]). This limited protection was based on the development of type-specific opsonic antibodies to the M molecule which were found to persist years after a streptococcal infection (128). However, after such an infection, measurable amounts of opsonic antibodies are usually delayed when compared with the response to other streptococcal antigens (124, 125, 176). In more recent studies with competitive inhibition assays with purified M protein, it was discovered that type-specific opsonic sera contained antibodies directed to a majority of the immunodeterminants on the M molecule while nonopsonic sera did not (64). Thus, it was suggested that, for some reason, opsonic antibodies are produced late during an immune response to the M protein.

Immunodominant Region of the M Molecule

To define further the development of an immune response to the total M6 protein, synthetic peptides representing 82% of the native molecule were examined for reactivity with rabbit and human sera by kinetic enzyme-linked immunosorbent assay (78). The results reveal that rabbits immunized with the native M protein or whole streptococci respond by reacting first and predominantly to one of the three sequence repeat regions of the molecule, namely, the B repeat. Antibodies to this region have recently been shown to be nonopsonic (106). Antibodies to peptides representing the variable N-terminal and adjacent A-repeat regions appear only when opsonic antibodies are detected in the serum. Antibodies to peptides located within the conserved C-terminal half of the molecule are limited even after several immunizations. By the same assay, an examination of human sera revealed that those sera opsonic for M6 streptococci contain antibodies reactive predominantly to the N-terminal and A-repeat regions. Nonopsonic human sera to M6 streptococci show low reactivity to most peptides. By Western blot analysis of the N- and C-terminal halves of the M6 molecule, all human sera tested contain antibodies to the conserved C-terminal half, while only sera opsonic for M6 streptococci react with the N-terminal variable half. The authors suggest that antibodies to the C-terminal half are primarily directed to conformational determinants and thus may not react with small synthetic peptides (78).

The immunodominant nature of the B-repeat region could explain the observed delay in the appearance of opsonic antibodies in human sera after exposure to M-protein-positive streptococci (124, 125, 176). It is possible that an exaggerated response to the nonopsonogenic B repeat may in some way affect the response to the adjacent regions. It also explains the observed presence of high levels of binding antibodies to the M molecule in the absence of opsonic activity. That M6 antibodies to the N-terminal and A-repeat segments were elevated in the M6 opsonic human sera is in keeping with the observation of Lancefield (128) that lifelong type-specific immunity usually occurs after a streptococcal infection.

ANTIPHAGOCYTIC ACTIVITY OF M PROTEIN

The capacity of the group A streptococcus to cause disease may be attributed partially to its ability to resist phagocytosis by human neutrophils. This capacity can easily be demonstrated in the laboratory, using a bactericidal or opsonophagocytosis assay developed by Lancefield (127) in which a few log-phase M streptococci are added to a small amount of heparinized human blood from a donor lacking antibodies to the streptococcal serotype being tested. The tube is sealed and rotated end over end at 37°C to allow
maximal contact between phagocytes and streptococci. At the end of 3 h, the contents of the tube are analyzed for surviving streptococci. In this assay, the M⁺ streptococcus grows unabated, whereas streptococci with insufficient quantities of M protein are easily cleared from the system. In the case of an M⁻ organism, several thousand may be added to this system and be phagocytized effectively, while <10 streptococcal chains of an M⁺ strain are adequate to allow growth to several hundred chains at the end of the incubation period. The addition to the assay tubes of rabbit antiserum, prepared against either purified M protein or whole streptococci of a specific serotype, neutralizes the antiphagocytic activity of the M molecule, allowing streptococci of that same serotype to be phagocytized. Antiser prepared to other streptococcal serotypes which cross-react with the M protein of the organisms in the assay have no opsonic effect on these streptococci.

A different assay has been used by Beachey and coworkers (21) in which streptococci are mixed with human blood much the same as described in the Lancefield assay (127); however, a higher bacterial inoculum is used. After 3-h rotation, a smear is made of the blood and the phagocytes are observed under the microscope and scored for their association with streptococci. This assay does not distinguish between live attached organisms and ingested killed streptococci. In many cases, however, good correlation is seen with the Lancefield method (127); nevertheless, a different parameter is likely being measured by this assay (i.e., at the level of attachment and ingestion).

N-Terminal Charge Domain

The mechanism by which the M molecule exerts its antiphagocytic effect has been elusive since its discovery nearly 60 years ago. Despite sequence variation within the N-terminal hydropervious region of the M molecules, a common feature found in the sequence of M5 (139), M6 (77), M24 (153), and M49 (116) is a concentration of acidic amino acids within this segment. These residues occur predominantly in external positions within the coiled-coil structure (77). Some of these amino acids likely enter into salt bridge formation between the two M-protein chains and function to stabilize the coiled-coil (165), while others become hydrated and contribute to the net negative charge in this region. While the exact function of this conserved charge domain is not known, it has been suggested that it may contribute to the antiphagocytic property of the molecule (66, 136, 139) by hampering contact between the streptococcus and the phagocyte by electrostatic repulsion. In support of this idea, it was found that, when opsonic IgG was fractionated into acidic and basic components, the opsonic activity was localized in the IgG population that was most basic in charge while nonopsonic binding antibodies were located in the other charged IgG populations (66). This suggests that the basic antibodies bind to the more acidic regions of the M molecule (i.e., the N-terminal region) and thereby neutralize the negative charge.

Effect on Phagocytic Cells

In attempts to determine whether the M protein on the whole streptococcus impairs the phagocytic capacity of neutrophils, M⁺ streptococci were rendered nonviable with mitomycin and used at 150-fold excess over phagocytes in a bactericidal assay containing a small number of viable phagocytosis-susceptible streptococci (140). The results revealed that the phagocytes could effectively seek out and destroy the relatively few susceptible streptococci in the presence of the excess M⁺ cells. Thus, the nonviable M⁺ streptococci did not interfere with the phagocytic activity of the neutrophils. When these experiments were repeated with live streptomycin-susceptible M⁺ and streptomycin-resistant M⁻ streptococci, to select for M⁻ survivors on streptomycin plates, similar results were obtained (Fischetti, unpublished data). This helped to rule out the possible effects of mitomycin or nonviable organisms in the previous experiments. Hence, in the presence of M⁺ streptococci, the phagocytes are not preoccupied by attempting to phagocytize the M⁺ cells: they simply ignore them.

Complement

The alternative complement pathway is considered to be an important part of the nonspecific defense system against infection (155). The deposition of a large number of complement component C3b fragments on the surface of microbial particles, termed opsonization, results in the uptake of these particles by phagocytic cells bearing C3b receptors. Several lines of evidence have indicated that M protein exerts its antiphagocytic effect by interfering with the alternative complement pathway. It has been shown that M⁻ streptococci are quickly phagocytized after being opsonized via the alternative pathway while M⁺ streptococci resist phagocytosis in the absence of type-specific antibodies (29, 168). That M⁺ streptococci fail to bind effectively to phagocytic cells suggests that C3b was either insufficiently deposited on the streptococcal surface or inaccessible to the C3b receptors of the phagocyte. When Jacks-Weis et al. (103) quantitated the amount of C3 deposited on the streptococcal surface, they found that M⁻ organisms bound up to four times more C3 than M⁺ cells. They concluded that, while reduced, the amount of C3 deposited on the M⁻ cell should be sufficient for phagocytosis and, thus, could not be the sole reason for the antiphagocytic effect of the M molecule.

Factor H

The deposition of C3b on the surface of particles is amplified by particle-bound C3 convertase, also termed C3b,Bb, which is a labile enzyme complex that activates C3 (154). Specific inhibitors such as decay-accelerating factor and factor H prevent C3 activation by rendering C3b susceptible to inactivation by factor I. Factor H is a 150-kDa β-globulin present in serum (500 µg/ml) and functions to control fluid-phase as well as surface-bound C3b (205). While the studies of Jacks-Weis et al. (103) demonstrate that M⁺ streptococci bind less C3b than M⁻ cells, the deposition on M⁻ cells was found to be uneven compared with the more uniform distribution of C3b on M⁺ organisms. Horstmann et al. (98), using purified complement components instead of serum, did not find any difference in either the deposition or the inactivation of C3b or the dissociation of the C3b,Bb complex on the surface of M⁺ and M⁻ streptococci when complement regulatory proteins were omitted. In fact, both strains exhibited a spontaneous decay of the C3b,Bb complex similar to zymosan-bound C3b,Bb which served as a control. However, when purified factor H was added to streptococcus-bound C3b, the cofactor activity of factor H for the cleavage of C3b by factor I was six to eight times higher when C3b was bound to M⁺ streptococci instead of M⁻ organisms. M⁺ streptococci were found to bind factor H selectively. Furthermore, M⁺ streptococci
five different serotypes, as well as purified M protein, were found to bind factor H with dissociation constants as high as $6 \times 10^{-7}$ M, emphasizing the functional significance of this binding reaction.

It was proposed that the antiphagocytic action of the M protein may be explained by this finding. When a streptococcus contacts serum, the M protein specifically binds factor H to its surface. Any C3b fixed in the vicinity of this complex is controlled by factor H, which inhibits or reverses the formation of C3b.Bb complexes and serves as a cofactor in the conversion of C3b to iC3b by factor I. Thus, the low numbers of C3b molecules deposited on the M$^+$ streptococcus will prevent C3b-dependent phagocytosis. The regulation of complement deposition by binding factor H to a bacterial surface appears to be a novel route by which a pathogen is able to evade alternative pathway activation and phagocytic clearance.

**Fibrinogen**

Kantor (110) found that a component in partially purified acid extracts of group A streptococci precipitated fibrinogen. In a number of studies, the active component was subsequently found to be M protein (101, 110, 152, 207, 209). Similar binding characteristics for fibrinogen were found for other streptococcal groups (C and G), suggesting that M-protein analogs are also present on these organisms (122). In attempts to examine the role of this interaction with regard to the antiphagocytic action of the M protein, Whittack and Beachey (207) measured the association of streptococci with polymorphonuclear leukocytes in the presence of serum and plasma. In this study, bound and ingested streptococci were not differentiated. Their results indicate that fibrinogen blocks the association of streptococci with the polymorphonuclear leukocytes. Immunofluorescence analysis with anti-C3 revealed that streptococci pretreated with fresh serum bound this complement component evenly on their surfaces whereas streptococci treated with plasma exhibited limited complement deposition. The authors suggested that the restricted deposition of C3 prevents complement-mediated opsonization of these organisms. This finding differed from that of Jacks-Weis et al. (103), who detected an uneven deposition of C3 on streptococci pretreated with nonimmune serum. Chhatwal et al. (39) also demonstrated, by immunofluorescence, that fibrinogen inhibited C3 fixation on the surface of group A, B, C, and G streptococci.

Subsequent experiments by Whittack et al. (210) were designed to examine the influence of fibrinogen on opsonization of streptococci by type-specific and cross-reactive antiserum. Rabbit antiserum against PepM protein of the homologous type overcame the antiosipnic effect of the fibrinogen. Using cross-reactive sera, these authors found that, in the two M proteins (M5 and M6) studied in detail, antibodies inhibiting fibrinogen binding to M protein accounted for a large part of the cross-reacting anti-M antibody in the sera. In all, these studies indicate that fibrinogen binds to both type-specific and cross-reactive regions of the M molecule. To explain the common antiphagocytic function of M proteins from different serotypes, it was suggested that the fibrinogen binding site may be conserved among different serotypes. That the CR3 receptor on human polymorphonuclear leukocytes binds both complement C3b and fibrinogen (218) may explain why fibrinogen-coated streptococci (208) adhere to but are not ingested by these phagocytic cells. Thus, since the assay used in many of these studies measures phagocyte association, the role of fibrinogen in streptococcal clearance is not completely defined.

R. D. Hortsman and V. A. Fischetti (manuscript in preparation) examined the effect of fibrinogen-bound M protein on factor H binding and on the regulation of C3b on the streptococcal surface. Results revealed that, while the presence of fibrinogen on the streptococcal surface slightly reduced the binding of factor H to M protein, it did not interfere with complement inhibition. They found that C3b is equally well regulated by factor H when it is bound to fibrinogen. Thus, factor H reduces the deposition of C3b on the streptococcal surface with or without fibrinogen bound to the M protein.

**ROLE OF M PROTEIN IN STREPTOCOCCAL ADHERENCE**

Group A streptococci colonize the nasopharyngeal mucosa of humans, leading to an inflammatory response followed by clinical illness or an asymptomatic carrier state. It was first shown by Ellen and Gibbons (66) that streptococci bearing the M-protein molecule on their surface can adhere better to epithelial cells in vitro than M-deficient organisms. There is also strong evidence that LTA secreted from the streptococcus mediates direct contact with the epithelial cell surface. M protein or another surface molecule might facilitate this process by anchoring and orienting the LTA adhesin so that the lipid moiety is able to make contact with the epithelial cell (20). Studies also suggest that M-positive streptococci can attach to epithelial cells by a mechanism independent of LTA. When M$^+$ and M$^-$ streptococci are tested for their binding capacity to human pharyngeal, buccal, and tongue epithelial cells, the M$^+$ bacteria attach in significantly higher numbers to the pharyngeal than to the buccal or tongue cells (197). The M$^-$ organisms bind in equivalent low numbers to all three oral cell types. While M$^+$ streptococci could be inhibited from binding to pharyngeal cells by using the complete M6 protein molecule as well as LTA, 1,000-fold more LTA than M protein is required to attain equivalent inhibition.

M protein binds specifically to F actin, forming regular parallel bundles of actin filaments in solution (36). In addition to isolated M protein, whole M$^+$ streptococci also bind specifically to F-actin filaments. While M protein does not compete with tropomyosin for binding to actin, it does compete with myosin subfragment 1 as a result of this interaction. While the biological implication of this binding is unclear at this time, it is becoming more apparent that actin is present on the surface of a number of cell types such as fibroblasts (184), lymphocytes (160), and human B cells (5). Therefore, the interaction of streptococcal M protein with surface actin on a host cell may be a possible event. Thus, while there may be multiple mechanisms by which streptococci can attach to epithelial cells, it seems apparent that M protein provides the organism with an adherence advantage; consequently, it is considered to be an attachment factor.

**M-PROTEIN VACCINES**

**Type-Specific Protection**

Since the observation by Lancefield that antibodies to the M protein have the capacity to opsonize streptococci in preparation for phagocytic clearance (126), the M protein has been a prime candidate for a vaccine to prevent group A streptococcal infections. Fox has reviewed the trials and tribulations of the early attempts in M-protein vaccine development (81). Since the early 1970s, few human trials have
been attempted (recently reviewed [D. Bessen and V. A. Fischetti, in M. M. Levine and G. Wood, ed., The Molecular Approach to New and Improved Vaccines, in press]). This is partially based on previous problems with hypersensitivity reactions found with the acid-extracted M-protein preparations (81); essentially, only type-specific protection was observed. In addition, attempts to separate heterologous protein contaminants from the type-specific determinants were disappointing. Except for one investigation (22), all vaccine development since the 1970s has been based on animal studies and on an analysis of the immune response to a variety of M-protein preparations with and without adjuvants (10, 26, 27, 37, 78, 104, 216) and in combination with other antigens (37).

In 1979, Beachey et al. (22) used pepsin-extracted M24 protein (the N-terminal half of the native M molecule) to immunize human volunteers. This highly purified preparation was found to be free of non-type-specific reactivity. Unlike earlier acid-extracted products (81), skin delayed-type hypersensitivity tests in 37 adults proved to be negative with this PepM protein. Immunization with alum-precipitated PepM24 protein led to the development of type-specific opsonic antibodies in 10 of 12 volunteers and the formation of positive delayed cutaneous hypersensitivity to PepM24 in 11 individuals. None of the volunteers developed heat-reactive antibodies as determined by immunofluorescence. These studies clearly indicated that M-protein vaccines free of sensitizing antigens could be produced, but also illustrated the type specificity of the immune response.

Encouraged by these studies, Beachey and co-workers embarked on a systematic approach to develop methods by which a type-specific epitope-based vaccine could be developed to protect against streptococcal disease. It was quickly learned that peptides representing the first 20 or so N-terminal amino acids of M24 protein evoked type-specific opsonic antibodies to M24 streptococci (17). Experiments with synthetic peptides of the N terminus of M1, M5, M6, and M19 proteins resulted in the same conclusions (12, 15, 23, 34, 107, 179). When a hybrid peptide was synthesized from the sequence of the M24 and M5 proteins and injected into animals in complete Freund adjuvant, opsonic antibodies were produced to both M5 and M24 streptococci (12). Opsonic antibodies to three M proteins were obtained when the N-terminal sequences of three M-protein sequences (M5-M6-M24) were synthesized in tandem and injected into rabbits (16).

Attenuated Salmonella typhimurium has also been used effectively for delivering the M-protein antigen in a mouse model (174). The cloned gene for the complete M5 protein was introduced into the arsA strain of S. typhimurium by transformation with a plasmid vector. The transformed salmonella expressed the M5 protein in the cytoplasm and had the capacity to induce serum and salivary IgA, IgG, and IgM to M5 protein when fed orally to BALB/c mice. The orally immunized mice were protected from challenge by M5 but not heterologous M24 streptococci.

An important factor to consider for the development of a type-specific epitope-based vaccine is the potential of the streptococcus to generate new M serotypes by changing the amino-terminal portion of M protein. In the type 6 M-protein molecule of strain D471, type-specific opsonizing antibodies are directed against epitopes located both at the amino-terminal end (residues 1 through 21) and within the A-repeat block (106), which begins at amino acid 27 and continues to residue 96 (77, 94). As mentioned previously, high-frequency intragenic recombinational events within the A-repeat block region can lead to a significant loss in the opsonizing ability of monospecific antibody (107). However, to date little information is available on mechanisms for generating antigenic variation at the extreme amino-terminal nonrepeat region. A second major concern for type-specific epitope-based vaccines centers on the antigenic diversity of the more than 80 serological types of M protein. Certain serotypes are more likely than others to be associated with severe cases of pharyngitis or outbreaks of rheumatic fever; thus, the number of distinct type-specific determinants incorporated into a vaccine can be limited to some extent (28). A type-specific streptococcal vaccine against upper respiratory infection and acute rheumatic fever would require a multivalent antigen corresponding to stable immunodeterminants of those serotypes which together account for the majority of nasopharyngeal isolates.

A clear understanding of the molecular basis of antigenic variation within the hypervariable nonrepeating segment of the M protein should enable investigators to make decisions on how best to circumvent the evasive mechanisms used by group A streptococci.

Non-Type-Specific Protection

The incidence of group A streptococcal respiratory infection rises sharply at age 4, peaks at age 6, and rapidly declines above age 10, reaching adult levels by 18 years (32). At its peak incidence, 50% of children between the ages of 5 and 7 suffer from streptococcal infection each year. The decreased incidence of streptococcal pharyngitis in adults might be explained by an age-related host factor. Alternatively, protective antibodies directed to antigens common to all group A streptococcal serotypes might arise as a consequence of multiple infections experienced during childhood resulting in an elevated response to conserved protective epitopes of the M molecule. In exploring the second hypothesis, it was found that sera of most adults examined have a strong antibody response to the conserved regions of M6 protein (27a, 78). While antibodies directed to conserved epitopes fail to neutralize the antiphagocytic property of M protein (106), they might afford protection by an alternative mechanism.

Kurl et al. (123) found that whole-cell vaccines consisting of type 12, 18, 30, 49, 50, or 55 streptococci were equally as efficient at protecting mice immunized intranasally from lethal infection after challenge with M50 streptococci. The authors suggest that local immunity against M50 streptococci could be achieved by a nonspecific mechanism; however, the molecular components of the streptococcus involved in eliciting this protection were not defined.

It should be emphasized at this juncture that humans are the only reservoir for group A streptococci, and except for M type 50, which is naturally virulent for mice (97), all other serotypes need to be adapted to the animal being considered as a possible model for streptococcal disease. For instance, group A streptococci isolated fresh from human infection are not virulent for mice, and nearly 10⁷ CFU need to be administered intraperitoneally to cause death within 1 to 2 days (129). However, after such a passage in the mouse (usually several passages are required), a strain is isolated which will cause death with a dose of <10 CFU. For some group A strains, a mouse-virulent organism may not be isolated. Although the amount of M protein increases after mouse passage, its role in mouse virulence is not completely defined (93). However, it is clear that immunization with M protein or passively administered type-specific M-protein
antibodies will protect mice from lethal streptococcal challenge with the homologous serotype (59).

The C-repeating region of MG protein is highly conserved among streptococci of many distinct serological types (27a, 28, 106, 108, 109). Monoclonal antibodies directed to determinants localized to the C-repeating region of M6 protein bind to the surface of whole streptococci which represent more than half of the serotypes examined (108, 109). Despite these findings, Miller et al. (147) suggest that antibodies to conserved M-protein epitopes located on the C-terminal side of the peptidase cleavage site do not bind to intact streptococci and, thus, are of no value in a vaccine. Their findings are based on antibodies prepared to a synthetic peptide whose sequence is located within the conserved region immediately C-terminal to the peptidase site. Perhaps the conformation of the peptide used to raise antibodies is not sufficiently homologous to the same region in the native molecule to enable the antibodies to react with the intact molecule. This is supported by their finding that, after cleavage of the native molecule with peptidase (resulting in a conformational change in the M molecule [107]), the peptide-specific antibodies become reactive.

In view of overwhelming evidence for the presence of exposed conserved epitopes on the native M molecule (28, 106, 108, 109), it should be possible to generate antibodies reactive to the majority of serological types by using only a few distinct antigens for immunization. If a conserved-region vaccine is to be effective in preventing nasopharyngeal infection by group A streptococci, it may require stimulation of the secretory immune response.

Secretory IgA is a major first-line defense against bacterial infections of the mucosal epithelium (118), and it has been demonstrated that M-protein-specific secretory IgA can provide significant protection against streptococcal infection in a mouse model (26, 27a, 67). Passive immunization at the mucosa allowed for a more precise evaluation of the role played by secretory IgA in preventing streptococcal infection at a mucosal site (26). It was found that affinity-purified human anti-M-protein secretory IgA administered intranasally protected mice against systemic infection after intranasal challenge with group A streptococci. In contrast, anti-M-protein-specific IgG administered intranasally had no protective effect at this site, although it was highly opsonic and promoted streptococcal phagocytosis in whole blood.

An oral or intranasal route of immunization has been used for several group A streptococcal vaccines, with the intent of evoking a strong secretory IgA response (27, 51, 123, 170, 172). To determine whether the conserved exposed epitopes of M protein influence the course of mucosal colonization by group A streptococci, peptides corresponding to these regions were used as immunogens in a mouse model (27). Synthetic peptides corresponding to residues 216 to 235, 248 to 269, and 275 to 284 of the M6 sequence (Fig. 7) were covalently linked to the mucosal adjuvant cholera toxin B subunit and administered intranasally to the mice. About 1 month later, animals were challenged intranasally with live streptococci and pharyngeal colonization was monitored for 15 days. It was found that mice immunized with the cholera toxin B subunit-peptide complex showed a significant reduction in colonization compared with mice receiving cholera toxin B subunit alone (Fig. 16). Thus, despite the fact that conserved-region peptides are unable to evoke an opsonic antibody response (106), these peptides have the capacity to induce antibodies capable of influencing the colonization of group A streptococci at the nasopharyngeal mucosa in this model system.

The successful cloning of the emm-6.1 gene into vaccinia virus and its high expression in mammalian cells infected with the recombinant vaccinia virus may prove to be a powerful vector for delivery of the M molecule to mucosal surfaces (100). In an effort to express only the conserved-region epitopes located in the C-terminal half of the molecule, genetic engineering methods were used to truncate the N-terminal half of the M6 gene and recombine the C-terminal gene fragment into the vaccinia genome (VV:M6'). Western blot analysis revealed that the cells infected with the VV:M6' recombinant produce a molecule of about 30 kDa that is reactive with a monoclonal antibody specific for the
C-terminal half of the M protein (V. A. Fischetti, W. M. Hodges, and D. E. Hruby, Science, in press). Mice were immunized intranasally with either the VV:M6' recombinant or wild-type vaccinia virus and challenged intranasally and orally 1 month later with 5 × 10⁶ streptococci. Pharyngeal cultures taken up to 10 days postchallenge revealed that VV:M6'-immunized animals differed significantly from controls. Of the VV:M6'-immunized animals, only 16% of 152 total swabs taken were streptococcus positive, with 10 (6%) yielding >100 CFU, whereas 69% of 115 swabs were positive in the wild-type group and 40 (35%) displayed >100 CFU. On average, >70% of the animals immunized with wild-type virus were culture positive for group A streptococci at every swab day up to 10 days after challenge (Table 2). This is compared with <30% colonization of mice immunized with the VV:M6' recombinant. These results confirm and extend the previous studies that used synthetic peptides from conserved regions of the M molecule to protect against streptococcal colonization (27).

The achievement of non-type-specific protection in animal models (27, 123, 216) and the evidence that protection can be mediated by nonopsonic antibody (26) hold promise for an alternative to type-specific M-protein vaccines. The greatest challenge will lie in identifying those common epitopes, which elicit the broadest protection without stimulating antibodies cross-reactive with host tissue.

CONCLUDING REMARKS

The group A streptococcus utilizes a coiled-coil molecular design for the M protein, a pattern commonly found in eucaryotes and viruses (41, 42) which has not been thoroughly explored in bacteria. This type of structure offers a number of distinct advantages to the organism. For instance, the rodlike coiled-coil structure allows functional domains to be located sufficiently away from the difficulties inherent in microbial surfaces which in many cases actively stimulate the alternative complement pathway. Because the coiled-coil design may tolerate a large number of sequence changes, yet maintain a common conformation, the molecule has the capacity to change its immunodeterminants readily, allowing the organism to avoid or delay immune recognition. Furthermore, the rodlike motif coupled with the plasticity of the sequence allows this type of structure to be tailored for several specific functions. This is accomplished by dividing the molecule into discrete functional domains directed by immunological and colonization pressures. Because of these wide-ranging characteristics, the coiled-coil pattern is likely to be found as commonly in bacteria as it is in eucaryotes.

ACKNOWLEDGMENTS

For their help in many of the studies on the structure, function, and genetics of the M molecule, I acknowledge all of my collaborators over the years. Discussions with Maclyn McCarty, John Zabriskie, Emil Gotschlich, and the late Rebecca Lancefield have helped shape my current understanding of the group A streptococcus. Their time and patience are greatly appreciated. I thank Debra Bessen, Emil Gotschlich, Kevin Jones, Belur Manjula, Vijaykumar Pancholi, Olaf Schnieewind, and June Scott for critical review and suggestions for the manuscript.

This work was supported by Public Health Service grant AI11822 from the National Institutes of Health.

LITERATURE CITED


137. Rotta, R. C. Krause, R. C. Lancefield, W. Everly, and H.


