INFECTION AND IMMUNITY, Dec. 1987, p. 3237–3239
0019-9567/87/123237-03$02.00/0

Vol. 55, No. 12

A Highly Conserved Region Present in Transcripts Encoding Heterologous M Proteins of Group A Streptococci

SUSAN K. HOLLINGSHEAD,[1]† VINCENT A. FISCHETTI,[2] AND JUNE R. SCOTT[1]*

Department of Microbiology and Immunology, Emory University Health Sciences Center, Atlanta, Georgia 30322,[1] and Rockefeller University, New York, New York 10021[2]

Received 20 July 1987/Accepted 28 August 1987

In group A streptococcal strains of 10 different serotypes, the sequence previously identified as homologous to the structural gene for type 6 M protein (*emm6*) was found to be transcribed. The conserved sequence, which shows greater than 95% homology among heterologous M proteins, was identified as the 3′ third of the gene.

Group A streptococci are important human pathogens, especially in developing countries where they are a major cause of cardiac damage in children (1). The M protein is a dimeric molecule attached to the surface of virulent group A streptococci by its carboxy terminus. Its presence allows bacterial cells to evade phagocytosis in human hosts, and production of opsonic antibodies directed against this protein results in immunity to infection (13). Group A streptococcal strains have been divided into more than 80 different M protein serotypes by using polyclonal antisera produced against whole bacteria and rendered specific by absorption against streptococci of several heterologous serotypes (14).

Despite the apparent type specificity of protective antibodies and the diversity of serotypes, the development of a vaccine to prevent streptococcal infection has focused on the possibility of using the M protein as an immunogen. For this vaccine approach, it may be critical to identify common determinants among heterologous M types and to learn about their prevalence. Common antigenic determinants on M proteins of different serotypes have been observed repeatedly (3, 5, 7, 11, 12, 16, 22). The possibility that all M types share common determinants was suggested by a recent DNA hybridization study (19). In this study, a DNA probe consisting of almost the entire *emm6* gene (encoding the M protein of serotype 6 from strain D471) hybridized with DNAs from strains of all 56 different M types tested and from four nontypable strains under conditions allowing 23% mismatch.

In other pathogens, for genes encoding surface antigens that generate protective immune responses in a host there are often additional chromosomal regions of homology that are not transcribed (pseudogenes) (20). These are used by the pathogen to create new antigenic variants of the surface protein. Because the M protein is also antigenically variable, it seemed possible that such unexpressed homologous regions existed in group A streptococci. We have previously shown (19) that in most group A streptococcal strains there is only a single chromosomal region homologous to *emm6*. However, it remained possible that this *emm6*-homologous gene was not expressed in streptococci producing M proteins of other serotypes and that the transcribed genes for other types of M protein are not homologous to *emm6*. The DNA hybridization experiments could not determine whether the homology detected was in a pseudogene or in an expressed *emm* gene in the strains of the different M serotypes. To determine unambiguously whether the *emm6*-homologous region is expressed and to learn the extent of sequence conservation of *emm* genes of different serotypes, the RNAs from streptococcal strains of 10 different serotypes and four nontypable strains were examined for homology to *emm6* DNA, and in five strains the homologous RNAs were sequenced.

Group A streptococcal strains of 10 different serotypes isolated from patients with different types of streptococcus-related disease (rheumatic fever, glomerulonephritis, scarlet fever, suppurative infection) were used. These strains included T1/195/2 (M1), 14RP81 (M1), T5B/126/4 (M5), 9RP11 (M12), LO1740 (M12), T12/126/4 (M12), C98/135/2 (M24), A586 (M28), 1GL130 (M30), 4GL130 (M30), 5GL130 (M30), 6GL130 (M30), D24/126/3 (M30), 1RP274 (M49), 3RSC17 (M49), A928/73/1 (M55), 7RP68 (M62), A956 (M62), and D458/68/2 (M62) from the Rockefeller University collection and strains SS266 (M19), SS479 (M19), SS400 (M19), and SS935 (M55) from Richard Facklam of the Centers for Disease Control, Atlanta, Ga. In addition, RNA was isolated from nontypable strains A486var, A922, J17A4, and J17F/123/4 from Rockefeller University. Strain T28/51/4B, which has no DNA homologous to the *emm6* gene probe (17, 18), served as a negative control. RNAs from two strains (T28/150A/5 [M28] and B737/137/2 [M49]) which have an *emm6*-homologous region in their genomic DNAs but apparently do not express M protein as determined in a bactericidal assay (13) were also examined.

Total cellular RNAs were extracted (15) from cultures of group A streptococcal strains grown in Todd-Hewitt yeast broth (4) plus 2 mM glycine to an optical density of 0.6 at 600 nm in a 1-cm cell. The RNA was pelleted by sedimentation through 5.7 M CsCl (6). Glyoxalated RNA was separated by electrophoresis and transferred to nylon filters (Northern [RNA] blot; 21). The probe used for detection of RNA homologous to the *emm6* gene (Fig. 1) consisted of a purified 1,255-base-pair *Nci*I-*Pvu*II fragment which contains most of the coding sequence for the *emm6* gene (32 base pairs of DNA at the 5′ end and 36 base pairs at the 3′ end of the coding sequence for the mature M6 protein are excluded from it). The fragment was purified from an agarose gel and radiolabeled by the method of Feinberg and Vogelstein (2). The negative control strain and the two strains which did not survive in the bactericidal test produced no RNA homologous to this *emm6* gene probe. Under conditions allowing 34% mismatch (37°C and 0.2× saline sodium citrate), this

* Corresponding author.
† Present address: Department of Microbiology, University Station, University of Alabama at Birmingham, Birmingham, AL 35294.

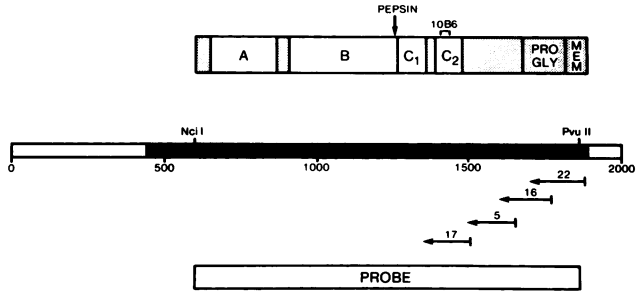

FIG. 1. Map of *emm*6. The box in the center of the figure represents the type 6 M protein gene from strain D471 whose sequence has been determined (8). The darkened area is the coding portion of the gene and includes the region that encodes the signal peptide not present in the mature protein. Within the mature M6 protein, there are several identifiable regions shown by the row of boxes at the top of the figure. Shaded boxes are nonduplicated protein sequences in M6, and open boxes marked A, B, C_1, and C_2 are repetitive. PRO GLY and MEM are proline-glycine-rich and membrane anchor segments of the M6 protein, respectively. The bracket under 10B6 marks the epitope region recognized by this monoclonal antibody and the overlapping region recognized by monoclonal antibody 10F5 (12), and the preferential site for pepsin digestion is indicated with a vertical arrow. In the lower half of the figure, the extent and direction of sequence analyses of RNA are shown by horizontal arrows, and the number above each arrow (17, 5, 16, and 22) indicates the name of the oligonucleotide used as a primer for the reaction. The fragment used as a probe in Northern gels is indicated at the bottom of the figure.

*Nci*I-*Pvu*II probe detected a single band in the other 27 group A streptococcal strains listed above. This demonstrated that the *emm*6-homologous region previously detected in the DNAs of heterologous strains is transcribed.

The region of homology within the *emm* gene has previously been localized to about the carboxy-terminal half of the M6 molecule both by using monoclonal antibodies and by DNA hybridization with smaller probes. Two different monoclonal antibodies (10B6 and 10F5) produced against the type 6 protein, which recognize M proteins of many different serotypes (12), react with overlapping epitope regions within the carboxy-terminal half of the M6 molecule (Fig. 1) (11). This is in agreement with DNA hybridization studies which showed that homology with other *emm* genes is greatest in the region that encodes this half of the protein (18). However, the antibody studies only localized two epitopes which, although broadly distributed among different M types, are not universally present among them. The carboxy-terminal DNA probe used in the hybridization studies, which reacted with all of the group A streptococcal strains tested, extended beyond the 3′ limit of *emm*6, so the homologous DNA might have been downstream of the gene encoding the M protein. Therefore, to determine the extent of homology in heterologous strains of types M5, M19, M24, M30, and M55, the sequence of the carboxy-terminal 33% of the mRNA was analyzed (Fig. 1). Synthetic oligonucleotide primers and reverse transcriptase were used for mRNA sequence analysis as described by Inoue and Cech (10) and Hollingshead et al. (9).

The sequence obtained by using oligonucleotide 17 (Fig. 1) includes the region containing the epitope recognized by the two cross-reactive monoclonal antibodies 10B6 and 10F5 (11). Although these antibodies react with about 50% of the M proteins tested, including those from strains of the five serotypes we used in the current experiments (12), the sequence of the region with which they react (11) is completely identical to that of *emm*6 in only one of these five strains (Fig. 2). It is likely that the entire region defined by synthetic peptide analysis as reacting with these monoclonal antibodies is not required for their recognition. Because the substitutions in the epitope region in the other four strains are predominantly clustered in its amino-terminal half, the epitopes recognized by these monoclonal antibodies may be restricted to the carboxy-terminal half of the region delineated.

The carboxy-terminal region, whose sequence was determined by using oligonucleotides 16, 22, and 5 (Fig. 1), is highly conserved among the five strains of different M types examined (Fig. 2). This conserved area includes at least three sections of the M6 protein (Fig. 1): (i) a nonduplicated region of the protein carboxy terminal to three repetitive portions of the molecule, (ii) a proline-glycine-rich part of the M6 molecule (Fig. 1, PRO GLY) which appears to be associated with the streptococcal cell wall, and (iii) a hydrophobic portion of the protein which may be associated with the cell membrane (Fig. 1, MEM). Because all of these three common regions (carboxy terminal to residue 305) appear to be buried within the cell wall when the M protein is present on the surface of a streptococcus (V. Pancholi, and V. A. Fischetti, Abstr. Annu. Meet. Am. Soc. Microbiol. 1987, J6, p. 200; manuscript in preparation), a specific sequence within this part of the molecule may be required for attachment of the M protein to the cell. This requirement might result in evolutionary conservation of the carboxy-terminal part of the M protein shaded in Fig. 1. In addition, a short region external to the cell (part of the C_2 repeat) is highly conserved (Fig. 2). This suggests the possibility that its specific sequence is also critical for M protein function or that it is not freely available to the immune system because

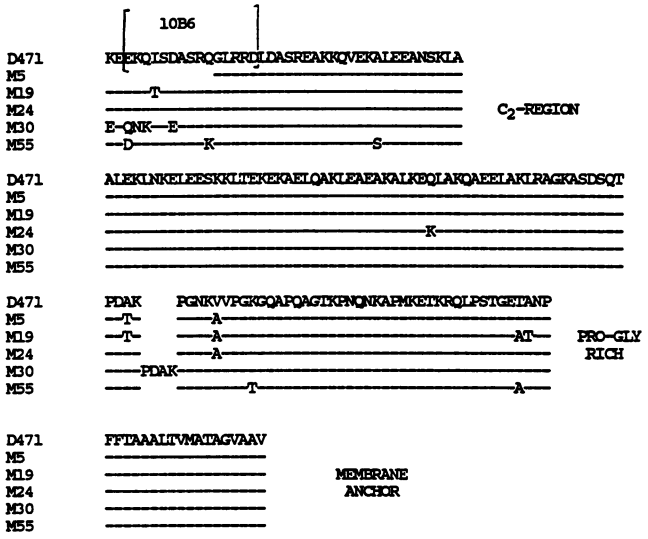

FIG. 2. The amino acid sequence of the carboxy-terminal part of five heterologous M proteins deduced from mRNA sequence analysis compared with the corresponding region of M6 strain D471. The RNAs were from strains T5B/126/4 (M5), SS266 (M19), C98/135/2 (M24), D24/126/3 (M30), and SS935 (M55). Lysine 273 of the mature protein is the first amino acid shown in the D471 sequence. Amino acids identical to those in M6 are shown as dashes, and amino acids that differ from M6 are delineated by letters of the single-letter amino acid code. Blank spaces indicate no amino acid, except for that in M5 near the top of the figure, which was not sequenced. The protein regions (C_2, PRO-GLY, etc.) on the right correspond to those shown in Fig. 1.

of its proximity to the cell surface and that it is thus less subject to selective pressure.

We thank Jane Rhatigan for expert technical assistance. We also thank Richard Facklam for providing strains and Barbara Froehlich for advice on RNA procedures.

This work was supported by Public Health Service grant AI20723 to J.R.S. and V.A.F. from the National Institutes of Health.

LITERATURE CITED

1. **El Kholy, A. J., J. Rotta, L. W. Wannamaker, T. Strasser, B. Bytchenko, W. Ferreira, L. Houang, and E. Kiisberg.** 1978. Recent advances in rheumatic fever control and future prospects: a WHO memorandum. Bull. W.H.O. **56**:887–912.
2. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
3. **Fischetti, V. A.** 1977. Streptococcal M protein extracted by nonionic detergent. II. Analysis of the antibody response to the multiple antigenic determinants of the M protein molecule. J. Exp. Med. **146**:1108–1123.
4. **Fischetti, V. A., K. F. Jones, and J. R. Scott.** 1985. Size variation of the M protein in group A streptococci. J. Exp. Med. **161**: 1383–1401.
5. **Fox, E. N., and M. K. Wittner.** 1968. Antigenicity of the M proteins of group A hemolytic streptococci. IV. Cross-reactivity between serotypes. J. Immunol. **100**:39–45.
6. **Glisin, V., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry **13**: 2633–2637.
7. **Hirst, G. K., and R. C. Lancefield.** 1939. Antigenic properties of the type-specific substance derived from group A hemolytic streptococci. J. Exp. Med. **69**:425–445.
8. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1986. Complete nucleotide sequence of type 6 M protein of the group A streptococcus: repetitive structure and membrane anchor. J. Biol. Chem. **261**:1677–1686.
9. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207**:196–203.
10. **Inoue, T., and T. R. Cech.** 1985. Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. Proc. Natl. Acad. Sci. USA **82**:648–652.
11. **Jones, K. F., S. A. Khan, B. W. Erickson, S. K. Hollingshead, J. R. Scott, and V. A. Fischetti.** 1986. Immunochemical localization and amino acid sequences of cross-reactive epitopes within the group A streptococcal M6 protein. J. Exp. Med. **164**:1226–1238.
12. **Jones, K. F., B. N. Manjula, K. H. Johnston, S. K. Hollingshead, J. R. Scott, and V. A. Fischetti.** 1985. The location of variable and conserved epitopes among the multiple serotypes of streptococcal M protein. J. Exp. Med. **161**:623–628.
13. **Lancefield, R. C.** 1959. Persistence of type specific antibodies in man following infection with group A streptococci. J. Exp. Med. **110**:271–292.
14. **Lancefield, R. C.** 1962. Current knowledge of type-specific M antigens of group A streptococci. J. Immunol. **89**:307–313.
15. **McDonald, P. M., E. Kutter, and G. Mosig.** 1984. Regulation of a bacteriophage T4 late gene, *SOC*, which maps in an early region. Genetics **106**:17–27.
16. **Maxted, W. R., and J. P. Widdowson.** 1972. The protein antigens of group A streptococci, p. 251–266. *In* L. W. Wannamaker and J. M. Matsen (ed.), Streptococci and streptococcal diseases. Academic Press, Inc., New York.
17. **Scott, J. R., P. C. Guenthner, L. M. Malone, and V. A. Fischetti.** 1986. Conversion of an M^- group A streptococcus to M^+ by transfer of a plasmid containing an M6 gene. J. Exp. Med. **164**:1641–1651.
18. **Scott, J. R., S. K. Hollingshead, and V. A. Fischetti.** 1986. Homologous regions within M protein genes in group A streptococci of different serotypes. Infect. Immun. **52**:609–612.
19. **Scott, J. R., W. M. Pulliam, S. K. Hollingshead, and V. A. Fischetti.** 1985. Relationship of M protein genes in group A streptococci. Proc. Natl. Acad. Sci. USA **82**:1822–1826.
20. **So, M.** 1987. Pilus of *Neisseria gonorrhoeae*: phase and antigenic variation, p. 401–417. *In* M. Inouye, (ed.), Bacterial and outer membranes as model systems. John Wiley & Sons, Inc., New York.
21. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **9**:5201–5205.
22. **Widdowson, J. P., W. R. Maxted, and A. M. Pinney.** 1971. An M associated protein antigen (MAP) of group A streptococci. J. Hyg. **69**:553–564.