Cloning and Nucleotide Sequence of Type 3 M Protein Gene (emm3) Consisting of an N-Terminal Variable Portion and C-Terminal Conserved C Repeat Regions: Relation to Other Genes of *Streptococcus pyogenes*

Streptococcal Diseases Study Group (Chief: Kiyoshi HOSHINA)
Chihiro KATSUKAWA
Laboratory of Microbiology, Osaka Prefectural Institute of Public Health
(Received: November 12, 1993)
(Accepted: February 7, 1994)

Key words: *Streptococcus pyogenes*, M proteingenes, PCR, emm3, emm12

Abstract

The structural gene for type 3 M protein of *Streptococcus pyogenes,* which consists of an N-terminal variable portion and C-terminal conserved repeat regions, has been cloned by the polymerase chain reaction (PCR) with two primers (K-1 and K-2). They were selected from the best conserved region of the leader sequences and of the C-terminal portion near the Hexapeptide (LPSTGE) sequence found in different M proteins. From the nucleotide sequence of the product, 1645 base pairs were determined, including 32 amino acids of the leader sequences, the complete N-terminal variable region and the conserved C repeat regions. Analysis of the deduced amino acids of the sequence revealed the existence of two major repeat regions, the B and C repeat regions. Comparison of the C-repeat regions among M3 and other M proteins showed them to be more than 90% identical. The two B repeat blocks in M3 protein are also similar to those in M12 protein. Predictive secondary structure analysis of M3 protein reveals a strong alpha-helical potential. The algorithm also shows that the beta-sheet and turn potential for region 23–42 in M3 protein are similar to those for region 28–50 in M12 protein. The results indicate that M3 protein is closely related to M12 protein.

Introduction

Streptococcus pyogenes is responsible for a wide variety of human diseases, the most common of which are nasopharyngitis and impetigo[1]. Moreover, streptococcal pharyngeal infection in humans may develop into rheumatic fever or glomerulonephritis[1]. The principal virulence factor of *S. pyogenes* is a cell wall constituent known as M protein that gives the organism the ability to resist phagocytosis[2]. This virulence factor displays antigenic diversity within its amino terminal region[3]. The highly variable portions of M proteins form the basis of a serological typing scheme, and only antibodies directed to type-specific epitopes are capable of circumventing the antiphagocytic effect[2]. M protein is thought to inhibit alternative C3 convertase formation to restrict deposition of C3 on the streptococci and also to inhibit the classical C5 convertase formation in order to interfere with efficient complement receptor-mediated phagocytosis[4]. However, the relationship between M protein antigenic diversity and its antiphagocytic activity is not understood, and neither is the genetic basis for M protein antigenic diversity or the structural basis for

別刷請求先：（〒537）大阪市東成区中道1－3－69
大阪府立公衆衛生研究所　勝川　千尋

functions common to all M protein serotypes. In order to characterize them in more detail, the nucleotide sequences of genes encoding a number of different M protein serotypes need to be cloned, sequenced and compared. In this report, we describe the cloning and sequencing of emm3 from the *S. pyogenes* type 3 M protein gene which has not been known to date and compare this gene with others reported previously.

Materials and Methods

Bacterial strains, plasmid and media: *S. pyogenes,* the M^+ (type 3) strain C203 (ATCC 12384) was used in this study. *Escherichia coli* SJ2, harboring plasmid pJRS42.50, consists of an Xba I-Pvu II fragment including the emm6 gene from *S. pyogenes* D471 cloned into pUC19; it was a gift from Dr. June R. Scott[5]. *E. coli* JM109 was used as the recipient for plasmid transformation and for phage M13 propagation. The plasmid vector was pUC118 obtained from Takara Shuzo Co., Ltd, Kyoto, Japan. *S. pyogenes* was grown in Todd-Hewitt broth (Difco Laboratories, Detroit, MI, U.S.A.); *E. coli* strains were grown in LB broth.

Isolation of DNA: Chromosomal DNA from *S. pyogenes* strain C203 was prepared according to a procedure reported previously[6]. Briefly, the cultured strain C203 (500 ml) was centrifuged at 8,000 rpm for 30 min and the resulting pellet was lysed in 25 ml of 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 (TE), 1.25 ml of 10% SDS and 0.125 ml of a 20 mg/ml solution of proteinase K. The resulting mixture was incubated at 37°C for 45 min. To the lysate, 4.74 ml of 5 M NaCl was added with thorough mixing, and then 4 ml of CTAB/NaCl solution (10% hexadecyl trimethyl ammonium bromide in 0.7 M NaCl) was added to the mixed lysate and this mixture was incubated at 65°C for 20 min. The CTAB-treated lysate was extracted with an equal volume of phenol/chloroform/isoamyl alcohol to remove CTAB-protein/polysaccharide complexes. The aqueous phase was transferred to a fresh tube and then 0.6 volume of isopropanol was added. The resulting precipitate was washed once with 70% ethanol. The washed DNA was suspended at 100 ng of genomic DNA per ml in TE.

E. coli SJ2 cultured in LB broth containing 50 μg of ampicillin per ml was collected by centrifugation at 8,000 rpm for 30 min, and the plasmid in the bacteria was isolated with a Qiagen column (QIAGEN-tip 100) (Diagen Inc. Chatsworth, CA, U.S.A.), according to the protocol of the manufacturer. The isolated plasmid was digested with Msp I and Pvu II restriction enzymes. After electrophoresis of the restricted plasmid in a 0.8% gel, the fragment, which contained only the region encoding the M6 protein and lacked 32 bases at the 5′ end and 38 bases at the 3′ end of the gene, was purfied with a Gene Clean Kit (Bio 101, Inc., La Jolla, CA, U.S.A.).

Oligonucleotides: Two oligonucleotides were synthesized with a 380 B automatic DNA synthesizer (Applied Biosystems Inc., Foster City, CA, U.S.A.) and used as specific primers for the required extension and amplification reactions. They were 31-mer (specific primer K-1; 5′-CCGGGATCCTATTCGCTTA-GAAAATTAAAAA-3′) and 30-mer (specific primer K-2; 5′-CCGGTCGACAAGTTCTTCAGCTTGTT-TCGC-3′). The K-1 and K-2 primers contained Bam HI and Sal I recognition sequences at the 5′ end respectively. This made it easy to insert the amplified fragment into the vector.

Cloning of emm3 gene by the polymerase chain reaction: PCR was performed in a Hybaid Thermal Reactor (Hybaid Ltd., U.K.). The reaction mixture contained 10 μl of 100 ng/μl genomic DNA, 25 pmol of each phosphorylated primer, 0.2 mM deoxynucleotide triphosphate mix, 0.01% (w/v) gelatin, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$ and 2 U of Taq polymerase (Takara) in a total volume of 100 μl, and the reaction mixture was overlaid with 2 drops of mineral oil (Sigma Chemical Co., St. Louis, MO, U.S.A.). PCRs (30 cycles) were performed with each cycle consisting of denaturation (94°C, 1 min), annealing (55°C, 2 min) and extension (72°C, 2 min). The PCR product was precipitated with ethanol and digested with Bam HI and Sal I restriction enzymes. After electrophoresis, it was extracted from the preparative agarose gel and cloned into the cloning vector pUC118 which had been digested with Bam HI and Sal I.

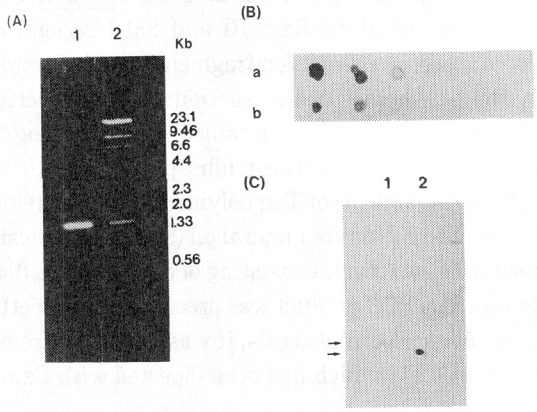
平成6年5月20日

DNA sequence: Plasmid pUC118 harboring the emm3 gene transformed in *E. coli* JM109 was grown in L-broth containing 50 μg of ampicillin per ml and purified with Econopack Q (Pharmacia LKB Biotechnology Inc., Piscataway, N.J., U.S.A.). The purified emm3 gene inserted into the vector was digested with Bam HI, Sal I and Bgl II (fragments of about 440 and 1030 base pairs), with Bam HI, Sal I and Sca I (fragments of about 880 and 590 base pairs) and with Bam HI, Sal I and Stu I (fragments of about 1200 and 270 base pairs). Each fragment was purified with a Sephaglas® Band Prep Kit (Pharmacia) and inserted into M13 phage. Single-strand DNA was purified from the supernatant of the cultured phage with a Sephaglas® Phage Prep Kit (Pharmacia). The purified single-strand DNA was sequenced with an Auto Read® Sequencing Kit (Pharmacia) by A.L.F. DNA Sequencer (Pharmacia).

Hybridization: DNA samples were denatured by heating for 10 min at 95°C and then chilled rapidly on ice for dot blot hybridization. DNA samples also were denatured in NaOH for Southern hybridization before application to membrane filters (Schleicher and Schuell, Dassel, Germany) which had been soaked in distilled water and then in 20 × SSC buffer (1 × SSC is 0.15 M sodium chloride, 0.15 M sodium citrate, pH 7.0). Hybridization of the restriction fragments was conducted by the method of Southern. The DNA samples were bound to a nitrocellulose membrane by baking for 2 hr in a vacuum at 80°C. The DNA bound to the filter was incubated for 2 hr at 68°C in prehybridization solution (0.25% powdered skim milk in 2 × SSC). Hybridization was carried out by overnight incubation at 65°C with labeled probe DNA diluted to a 5 ml final volume of prehybridization solution. The hybridized filters were washed twice in 2 × SSC, 0.1% SDS for 15 min at room temperature and twice at 68°C in 0.1 × SSC, 0.1% SDS for 15 min 68°C. Probes were labeled with digoxigenin-11-UTP (Boehringer Corp. Ltd., Sussex, U.K.) by random hexanucleotide primers according to the protocol supplied by the manufacturer. Digoxigenin-labeled probes were detected by using an anti-digoxigenin antibody alkaline phosphatase conjugate (Boehringer) and the substrates of BCIP (5-bromo-4-chloro-3-indolyl phosphate) and NBT (nitroblue tetrazolium).

Nucleotide sequence accession number. The 1465 base pair nucleotide sequence of emm3 gene is available from DDBJ, EMBL and GenBank Nucleotide Sequence Databases under accession number

Fig. 1 Analysis of the amplification product. (A) Ethidium bromide picture of the agarose gel (0.8%) of the amplified product. Lane 1, PCR product with K-1 and K-2 primers; lane 2, marker. (B) Dot blot hybridization analysis with the Msp I-Pvu II fragment of the emm6 gene probe. Serial five-fold dilutions of 6 ng of the Msp I-Pvu II fragment of the emm6 gene (lane a) and 5 ng of DNA amplified with K-1 and K-2 primers (lane b) were spotted onto nitrocellulose and probed with the digoxigenin-labeled Msp I-Pvu II fragment. (C) Southern blot analysis with the Msp I-Pvu II fragment of the emm6 probe. Lane 1, DNA amplified with K-1 and K-2 primers; lane 2, Msp I-Pvu II fragment of emm 6 gene. DNA hybridized with the probe is shown by the arrows.



D14415.

Results

Cloning and nucleotide sequence of the emm3 gene from *S. pyogenes* type 3 strain C203 with PCR: Much information is available on nucleotide sequences from various strains[7)~12)]. We tried to obtain the

Fig. 2 Nucleotide and deduced amino acid sequences of the emm3 gene. The DNA strand is located at 5′ to 3′ and its nucleotides are numbered above each line. Amino acid residues are presented as single letters below each line. B repeat blocks and C repeat blocks are indicated by underlining.

```
Signal peptide →
        10         20         30         40         50         60         70         80         90
TATTCGCTTAGAAAATTAAAAACAGGAACGGCTTCAGTAGCGGTTGCTTTGACAGTTTTAGGGACAGGACTGGTAGCAGGGCAGACAGTA
Y  S  L  R  K  L  K  T  G  T  A  S  V  A  V  A  L  T  V  L  G  T  G  L  V  A  G  Q  T  V
      Mature protein →
       100        110        120        130        140        150        160        170        180
AAGGCAGATGCTAGGAGTGTTAATGGAGAGTTTCCTAGACATGTTAAATTAAAAAATGAAATTGAGAACTTGTTAGATCAGGTTACACAA
K  A  D  A  R  S  V  N  G  E  F  P  R  H  V  K  L  K  N  E  I  E  N  L  L  D  Q  V  T  Q

       190        200        210        220        230        240        250        260        270
TTATATAATAAACATAATAGTAATTACCAACAATATAGTGCACAAGCTGGCAGACTTGACCTGAGACAAAAGGCTGAATATCTAAAAGGC
L  Y  N  K  H  N  S  N  Y  Q  Q  Y  S  A  Q  A  G  R  L  D  L  R  Q  K  A  E  Y  L  K  G

       280        290        300        310        320        330        340        350        360
CTTAATGATTGGGCTGAGAGGCTGTTACAAGAGTTAAATGGAGAAGATGTAAAAAAAGTTTTAGGTAAAGTGGCTTTTGAAAAAGATGAT
L  N  D  W  A  E  R  L  L  Q  E  L  N  G  E  D  V  K  K  V  L  G  K  V  A  F  E  K  D  D

       370        380        390        400        410        420        430        440        450
TTAGAAAAGGAGGTTAAAGAACTTAAGAAAAAAATAGACAAAAAGGAAAAGGAATATCAGGACTTAGATAAAGATTTTGACTTGGCCAAA
L  E  K  E  V  K  E  L  K  K  K  I  D  K  K  E  K  E  Y  Q  D  L  D  K  D  F  D  L  A  K

       460        470        480        490        500        510        520        530        540
CAGGGGTATGTTTTATCAGATAAAAGACATCAACAAGAACTAGAAGAAAAAGAAAAGAAAGTTACAGAAGCAACTGCTAAAGTTGGCCAG
Q  G  Y  V  L  S  D  K  R  H  Q  Q  E  L  E  E  K  E  K  K  V  T  E  A  T  A  K  V  G  Q

       550        560        570        580        590        600        610        620        630
ATTAGCGAAGAGCTAGAGACAGTTAAACAAAAAGTTGAAAGTACTATGCAAGATTTAACTGAAAAACAAAATCGTGTTTCTCAGTTAGAG
I  S  E  E  L  E  T  V  K  Q  K  V  E  S  T  M  Q  D  L  T  E  K  Q  N  R  V  S  Q  L  E

       640        650        660        670        680        690        700        710        720
CAAGAATTAGCTACTACTAAACAAAATGCTAAAGAAGATTTTGAACTAGCAGCATTAGCTAATGCCGCTGATAAACAAAAGTTAGAAGCT
Q  E  L  A  T  T  K  Q  N  A  K  E  D  F  E  L  A  A  L  A  N  A  A  D  K  Q  K  L  E  A

       730        740        750        760        770        780        790        800        810
AAGATTGCCGATTTAGAAACAAAACTAAAAGAGGCAAAGGAAGATTTTGAACTAGCAGCATTAGGTCACCAACATGCTCATAATGAGTAT
K  I  A  D  L  E  T  K  L  K  E  A  K  E  D  F  E  L  A  A  L  G  H  Q  H  A  H  N  E  Y
                                                  B1-repeat
       820        830        840        850        860        870        880        890        900
CAAGCAAAACTAGCAGAAAAAGATGATCAAATTAAACAACTAGAAGAGCAAAAACAAATCCTAGATGCTAGCCGTAAAGGTACAGCAAGA
Q  A  K  L  A  E  K  D  D  Q  I  K  Q  L  E  E  Q  K  Q  I  L  D  A  S  R  K  G  T  A  R
                                                                                  B2-repeat
       910        920        930        940        950        960        970        980        990
GACCTTGAAGCTGTTCGCCAAGCTAAAAAAGCTACGGAAGCTGAATTAAACAACCTCAAAGCAGAGCTTGCAAAAGTTACAGAACAAAAA
D  L  E  A  V  R  Q  A  K  K  A  T  E  A  E  L  N  N  L  K  A  E  L  A  K  V  T  E  Q  K

      1000       1010       1020       1030       1040       1050       1060       1070       1080
CAAATCTTAGATGCTAGTCGTAAAGGTACAGCAAGAGATCTTGAAGCAGTTCGCCAAGCAAAAGCACAAGTTGAAGCTGCTCTCAAACAA
Q  I  L  D  A  S  R  K  G  T  A  R  D  L  E  A  V  R  Q  A  K  A  Q  V  E  A  A  L  K  Q
                                                C1-repeat
      1090       1100       1110       1120       1130       1140       1150       1160       1170
CTTGAAGAACAAAACAGGATTTCAGAAGCAAGCCGTAAGGGTCTTCGCCGTGACTTGGACGCATCACGTGAAGCTAAGAAGCAAGTTGAA
L  E  E  Q  N  R  I  S  E  A  S  R  K  G  L  R  R  D  L  D  A  S  .R  E  A  K  K  Q  V  E
                                                                                  C2-repeat
      1180       1190       1200       1210       1220       1230       1240       1250       1260
AAAGATTTAGCAAACTTGACTGCTGAACTTGATAAGGTTAAAGAAGAAAAACAAATCTCAGACGCAAGCCGTCAAGGTCTTCGCCGTGAC
K  D  L  A  N  L  T  A  E  L  D  K  V  K  E  E  K  Q  I  S  D  A  S  R  Q  G  L  R  R  D

      1270       1280       1290       1300       1310       1320       1330       1340       1350
TTGGACGCATCACGTGAAGCTAAGAAACAAGTTGAAAAAGCTTTAGAAGAAGCAAACAGCAAATTAGCTGCTCTTGAAAAACTTAACAAA
L  D  A  S  R  E  A  K  K  Q  V  E  K  A  L  E  E  A  N  S  K  L  A  A  L  E  K  L  N  K

      1360       1370       1380       1390       1400       1410       1420       1430       1440
GAGCTTGAAGAAAGCAAGAAATTAACAGAAAAAGAAAAAGCTGAGCTACAAGCAAAACTTGAAGCAGAAGCAAAAGCACTCAAAGAACAA
E  L  E  E  S  K  K  L  T  E  K  E  K  A  E  L  Q  A  K  L  E  A  E  A  K  A  L  K  E  Q

      1450       1460 1465
TTAGCGAAACAAGCTGAAGAACTTG
L  A  K  Q  A  E  E  L
```

Fig. 3 Comparison of the leader sequences or of the deduced signal peptide of M types 3 and various M genes. (A) Leader sequences are subdivided according to structural domains of the resulting leader peptide. (B) Signal peptides are divided into basic (B), hydrophobic (H) and cleavage (C) regions. Unknown sequences and amino acids are indicated by asterisks.

(A)

Basic region **Primer**

```
M3    **************************TATTCGCTTAGAAAATTAAAAACAGGAACG
M12   ATGGCTAAAAATACCACGAATAGACACTATTCGCTTAGAAAATTAAAAACAGGAACG
M1    ATGGCTAAAAATAACACGAATAGACACTATTCGCTTAGAAAATTAAAAACAGGAACG
M6    ATGGCTAAAAATAACACGAATAGACACTATTCGCTTAGAAAATTAAAAAAAGGTACT
M5    ATGGCTAGAGAAAATACCAATAAGCATTATTGGCTTAGAAAATTAAAAAAAGGCACT
M24   ATGACTAAAAACAACACGAATAGACACTATTCGCTTAGAAAATTAAAAACGGGAACG
M49   ATGGCTAGAAAAGATACGAATAAACAGTATTCGCTTAGAAAATTAAAAACAGGTACA
M2    ATGGCTAGAAAAGATACGAATAAACAGTATTCGCTTAGAAAATTAAAAACAGGTACA
```

Hydrophobic region

```
M3    GCTTCAGTAGCGGTTGCTTTGACAGTTTTAGGGACAGGACTGGTAGCA
M12   GCTTCAGTAGCGGTTGCTTTAACAGTCGTAGGAGCAGGGTTAGTAGCA
M1    GCTTCAGTAGCGGTAGCTTTGACTGTTTTAGGGGCAGGTTTTGCGAAT
M6    GCATCAGTAGCAGTGGCTTTGAGTGTAATAGGGGCAGGATTAGTTGTC
M5    GCATCAGTAGCAGTAGCTTTGAGTGTCTTAGGAGCAGGATTAGTTGTC
M24   GCTTCAGTAGCGGTAGCTTTGACAGTTTTAGGGGCAGGATTAGTTGTC
M49   GCATCCGTAGCGGTCGCTGTGGCTGTTTTAGGAGCAGGCTTTGCAAAC
M2    GCATCCGTAGCAGTCGCTGTGGCTGTTTTAGGAGCAGGCTTTGCAAAC
```

Cleavage region

```
M3    GGGCAGACAGTAAAGGCA
M12   GGGCAGACAGTAAGAGCA
M1    CAAACAGAGGTTAAGGCT
M6    AATACTAATGAAGTTAGTGCA
M5    AATACTAATGAAGTTAGTGCA
M24   AATACTAATGAAGTTAGTGCA
M49   CAAACAGAAGTTAAGGCT
M2    CAAACAACAGTTAAGGCG
```

(B)

```
      B                  H                C
M3  : *********YSLRKLKTGT  ASVAVALTVLGTGLVA  GQTVKA
M12 : MAKNTTNRHYSLRKLKTGT  ASVAVALTVVGAGLVA  GQTVRA
M1  : MAKNNTNRHYSLRKLKTGT  ASVAVALTVLGAGFAN  QTEVKA
M6  : MAKNNTNRHYSLRKLKKGT  ASVAVALSVIGAGLVV  NTNEVSA
M5  : MARENTNKHYWLRKLKKGT  ASVAVALSVLGAGLVV  NTNEVSA
M24 : MTKNNTNRHYSLRKLKTGT  ASVAVALTVLGAGLVV  NTNEVSA
M49 : MARKDTNKQYSLRKLKTGT  ASVAVAVAVLGAGFAN  QTEVKA
M2  : MARKDTNKQYSLRKLKTGT  ASVAVAVAVLGAGFAN  QTTVKA
```

Fig. 4 Comparison of C-repeat regions found among M3, M6, M5, M24 and M2 proteins. Identified amino acid residues with deduced M3 protein are indicated by colons.

C1-repeat regions

```
M3    NRISEASRKGLRRDLDASREAKKQVEKDLANLTAELDKVKE
      :  ::::::::::::::::::::::::::::::::::::::
M6    NKVSEASRKGLRRDLDASREAKKQVEKDLANLTAELDKVKE
      : :: ::::::::::::::::::: :     :
M5    NKISDASRKGLRRDLDASREAKKQLEAEHQKLEEQ
      : ::::::  :::::::::::::: :
M24   NKISEASRQSLRRDLDASREAKKQLE
        ::::::: ::::: ::: :::
M2     QISEASRKSLRRDLEASRAAKKD
```

C2-repeat regions

```
M3    EKQISDASRQGLRRDLDASREAKKQVEK
      ::::::::::::::::::::::::::::
M6    EKQISDASRQGLRRDLDASREAKKQVEKALEE
        :: ::: ::::::::::::::: :
M5     NKISEASRKGLRRDLDASREAKKQLEAEQQKLEEQ
       :: :::: ::::::::::::::::::
M24   NKISEASRQSLRRDLDASREAKKQVEKALEE
       ::: :::  ::::: ::: :::
M2     QISEASRKSLRRDLEASRAAKKD
```

emm3 gene from the C203 strain by PCR. We selected a pair of forward and reverse primers, one (K-1; 31 mer) from the best conserved portion in leader sequences of seven different strains, and the other (K-2; 30 mer) from the C-terminal conserved portion[5,7)~14)]. PCR was performed with these two primers. The amplified product showed a single band approximately 1.4-1.5 kilobases (Kb) in length in agarose gel electrophoresis (Fig. 1-A) and was hybridized with the digoxigenin-labeled Msp I-Pvu II fragment of the emm6 gene (Fig. 1-B and C).

The DNA sequence of 1465 Kb of the amplified product was determined with a DNA sequencer and is shown in Fig. 2. The sequence indicated that the oligonucleotide sequences of the two primers existed in the 5′ and 3′ end (forward primer, 5′ (CCGGGATCC) TATTCGCTTAGAAAATTAAAAA 3′; reverse primer, 5′ (CCGGTCGA) CAAGTTCTTCAGCTTGTTTCGC 3′), and also encoded 488 amino acids which contained a leader sequence (1-32; defect first 9 amino acids), B repeat region (B1, 286-309; B2, 328-351) and C repeat region (C1, 374-391; C2, 416-433).

Comparison of the amplified product with other M protein genes: When the amplified product was compared with previously reported M protein genes[7)~13)], the leader sequence or C repeat regions showed

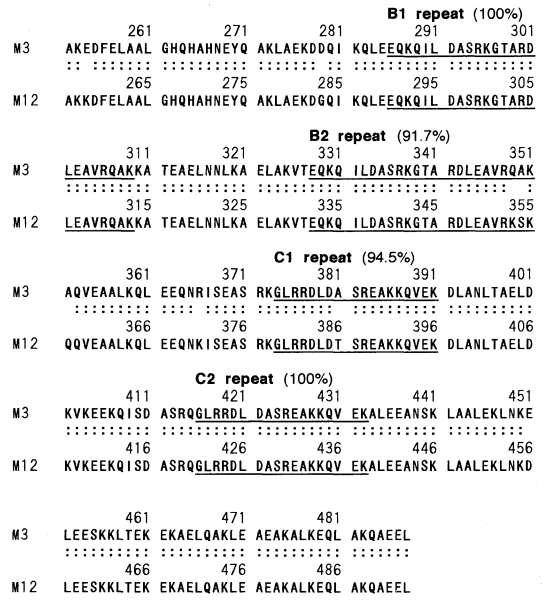


Fig. 5 Comparison of homologous regions in M3 and M12 proteins, and relationship between B and C repeats in the predicted amino acid sequences of M3 and M12 proteins. The residues from each protein being compared are indicated by the numbering system of Robbins et al.[12] for M12 protein. B repeat blocks and C repeat blocks are indicated by underlining, and identified amino acid residues are indicated by colons. The numbers enclosed in parentheses indicate % homology.

high homology with the other emm genes (Fig. 3 and 4). While the N-terminal amino acid portion of the amplified product was variable, it was found to be identical to 96 of 98 nucleotides downstream of the leader peptide sequence of another emm3 gene[15]. Therefore, the amplified product was characterized as the emm3 gene.

Comparison of homologous regions in M3 and M12 proteins: When the region of 252–488 deduced amino acids in M3 protein was compared with the region of 256–355 and 357–493 deduced amino acids in M12 protein, 96.6% homology was found between them (Fig. 5). Interestingly, the B repeat region showed high homology with only that of M12 protein (91.7 and 100%, respectively). However, the A repeat region in M12 protein was not present in the M3 protein. Predictive secondary structure analysis of M3 protein: From analysis of the predictive secondary structure of the amplified product by the algorithm of Robson[16], most of the product was found to exhibit strong alpha-helical potential. In addition, the beta-sheet and turn potential seen for region 23 to 42 in the M protein was similar to that seen for region 28 to 50 in M12 protein. The results suggest that M3 protein may be closely related to M12 protein.

Discussion

DNA sequence analysis has made it clear that all M proteins studied to date[7)~12)] are structurally related and are therefore encoded by a family of genes. The regions of amino acid sequence homology in the protein include the signal sequences, the C repeat region in the central part of the protein chain and the carboxyl-terminal part. On the basis of available genetic information, we cloned the emm3 gene by using PCR and sequenced its DNA. The primers prepared originated from the best conserved leader sequence or the C-terminal portion of the emm genes[7)~12)]. The amplified product was hybridized with an emm6 gene probe (Fig. 1). Sequence analysis of the amplified product identified sequences complementary to both oligonucleotide primers (Fig. 2). In addition, the product had both B and C repeat regions. By comparing the amplified product with known emm genes, we found not only that the N-terminal portion is very variable, but also that the C-terminal region is conserved[7)~12),14)].

Several streptococcal immunoglobulin-binding proteins have also been characterized as members of the M protein family and as having C repeats[17,18]. Our selected primers have similar regions which can be

平成6年5月20日

screened. However, our product amplified by PCR from the type 3 strain C203 of *S. pyogenes* was a single DNA and had no homology with amino terminal regions of the streptococcal immunoglobulin-binding proteins. Furthermore, 96 of 98 nucleotides downstream of the leader peptide sequence of the amplified product were found to be identical to the corresponding sequence of another emm3 gene of type 3 M strain 3-3/317[15]. The evidence shows that the amplified product is the emm3 gene.

B and C repeat blocks that exist in M3 protein are similar to those in M12 protein (Fig. 5). Furthermore, predictive secondary structure analysis of M3 protein revealed that the majority of the products exhibit strong alpha-helical potential as found with other M protein structures[19]. The algorithm also showed that region 23–42 exhibits beta-sheet and turn potential with a pattern similar to that for region 28–50 found by predictive secondary analysis of M12 protein.

S. pyogenes can be divided into two major classes on the basis of their immune reactivity with monoclonal antibodies (mAbs) directed against epitopes which lie within the conserved half of M proteins[20]. Class I serotype are defined as those which bind their mAbs, whereas class II isolates do not. Mainly the class I-specific mAb binding sites map to a region of C repeats within M proteins. Inasmuch as the C repeat region of our emm 3 gene represents more than 90% homology with the known emm genes, it belongs to the class I serotype. This agrees with the report of Bessen et al. who decided that M type 3 *S. pyogenes* had a class I protein[20]. Furthermore, we found similarity between emm 3 and emm 12 genes in their B repeat regions and predictive secondary structure. Thus, there may exist a subclass of class I M proteins. Bessen et al.[20] discriminated between serotypes sharing both B and C repeat region epitopes and those sharing only C repeat region epitopes by using only antibody probes directed to antigenic sites within the B and C repeat regions of the M protein molecules in class I serotype. We suggest that M3 and M12 proteins belong to a subclass of class I M proteins.

Acknowledgements

I am grateful to Dr. Kyonsgu Hong of the Department of Bacteriology, Osaka University Medical School, Dr. Ikuya Yano of the Department of Bacteriology, Osaka City University Medical School, Dr. Masanao Makino and Dr. Yasuhiko Suzuki of Osaka Prefectural Institute of Public Health for technical advice and helpful discussions. In addition, I thank Dr. June R. Scott (Emory University) for providing the pUC19: M6 plasmid and Dr. Tatsuya Tanaka of the Central Laboratoey for Research and Education, Osaka University, Faculty of Medicine, for synthesizing the oligonucleotides. This work was supported by a grant-in-aid from the Ministry of Education, Science and Culture of Japan (C63570192).

References

1) Bisno, A.: Group A streptococcal infections and acute rheumatic fever. N. Engl. J. Med. 325: 783—793, 1991.
2) Lancefield, R.C.: Current knowledge of the type-specific M antigens of group A streptococci. J. Immunol. 89: 307—313, 1962.
3) Beachey, E.H., Seyer, J.M., Dale, J.B., Simpson, W.A. & Kang, A.H.: Type-specific protective immunity evoked by synthetic peptide of *Streptococcus pyogenes* M protein. Nature (London) 292: 457—459, 1981.
4) Hong, K., Kinoshita, T., Takeda, J., Kozono, H., Pramoonjago, P., Kim, Y.U. & Inoue, K.: Inhibition of the alternative C3 convertase and classical C5 convertase of complement by group A streptococcal M protein. Infect. Immun. 58: 2535—2541, 1990.
5) Scott, J.R., Pulliam, W.M., Hollingshead, S.K. & Fischetti, V.A.: Relationship of M protein genes in group A streptococci. Proc. Natl. Acad. Sci. USA 82: 1822—1826, 1985.
6) Murry, M.G. & Thompson, W.F.: Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Res. 8: 4321—4325, 1988.
7) Bessen, D.E. & Fischetti, V.A.: Nucleotide sequence of two adjacent M or M-like protein genes of group A streptococci: Different RNA transcript levels and identification of a unique immunoglobulin A-binding protein. Infect. Immun. 60: 124—135, 1992.
8) Haanes, E.J. & Cleary, P.P.: Identification of a divergent M protein gene and an M protein-related gene family in

Streptococcus pyogenes serotype 49. J. Bacteriol. 171: 6397—6408, 1989.
9) Hollingshead, S.K., Fischetti, V.A. & Scott, J.R.: Complete nucleotide sequence of type 6 M protein of the group A streptococcus: Repetitive structure and membrane anchor. J. Biol. Chem. 261: 1677—1686, 1986.
10) Miller, L., Gray, L., Beachey, E. & Kehoe, M.: Antigenic variation among group A streptococcal M proteins: Nucleotide sequence of the serotype 5 M protein gene and its relationship with genes encoding type 6 and 24 M proteins. J. Biol. Chem. 263: 5668—5673, 1988.
11) Mouw, A.R., Beachey, E.H. & Burdett, V.: Molecular evolution of streptococcal M protein: Cloning and nucleotide sequence of type 24 M protein gene and relation to other genes of *Streptococcus pyogenes*. J. Bacteriol. 170: 676—684, 1988.
12) Robbins, J.C., Spanier, J.G., Jones, S.J., Simpson, W.J. & Cleary, P.P.: *Streptococcus pyogenes* type 12 M protein gene regulation by upstream sequences, J. Bacteriol. 169: 5633—5640, 1987.
13) Haanes-Fritz, E., Kraus, W., Burdett, V., Dale, J.B. & Beachey, E.H.: Comparison of the leader sequences of four group A streptococcal M protein genes. Nucleic Acids Res. 16: 4667—4677, 1988.
14) Hollingshead, S.K., Fischetti, V.A. & Scott, J.R.: A highly conserved region present in transcripts encoding heterologous M proteins of group A streptococci. Infect. Immun. 55: 3237—3239, 1987.
15) Podbielski, A., Baird, R. & Kaufhold, A.: The group A streptococcal M-type 3 protein gene exhibits a C terminus typical for class I M proteins. Med. Microbiol. Immunol. 181: 209—213, 1992.
16) Garnier, J., Osguthorpe, D.J. & Robson, B.: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97—120, 1978.
17) Gomi, H., Hozumi, T., Hattori, S., Tagawa, C., Kishimoto, F. & Björck, L.: The gene sequence and some properties of protein H. J. Immunol. 144: 4046—4052, 1990.
18) Heath, D.G. & Cleary, P.P.: Fc-receptor and M-protein genes of group A streptococci are products of gene duplication. Proc. Natl. Acad. Sci. USA 86: 4741—4745, 1989.
19) Phillips, G.N., Flicker, P.F., Cohen, C., Manjula, B.N. & Fischetti, V.A.: Streptococcal M protein: α-helical coiled-coil structure and arrangement on the cell surface. Proc. Natl. Acad. Sci. USA 78: 4689—4693, 1981.
20) Bessen, D., Jones, K.F. & Fischetti, V.A.: Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. J. Exp. Med. 169: 269—283, 1989.

PCR法を用いてクローニングしたA群溶血レンサ球菌M3蛋白遺伝子の解析および他菌型との比較

（レンサ球菌感染症研究会：会長　保科　清）
大阪府立公衆衛生研究所
勝　川　千　尋

（平成5年11月12日受付）
（平成6年2月7日受理）

要　旨

A群溶血レンサ球菌M3蛋白遺伝子のN末端の高度型特異領域からC末端の保存領域までの部分をPCR法を用いてクローニングを行った．

各菌型に共通なN末端のリーダーシクエンス部分とC末端の保存領域部分をプライマーとして用い，1465bpの遺伝子配列を決定し，他のM蛋白遺伝子と比較検討した．その結果，N末端側の100塩基から750塩基の範囲にM3型特異的な領域を見いだすことができた．また，アミノ酸配列を検討したところ，2つの繰り返し配列を見いだした(Bリピートおよび Cリピート)．Cリピートは現在知られている他のM蛋白遺伝子の塩基配列と非常に高い相同性を示した．これに対して，Bリピートは M12蛋白遺伝子のBリピート配列とのみ高い相同性を示し，また二次構造の解析結果でもこの2菌型は構造が類似していた．これらの結果よりM3蛋白とM12蛋白は遺伝学的に非常に近い蛋白であることが示唆された．

平成6年5月20日