Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8661–8665, September 1992
Microbiology

Two major classes in the M protein family in group A streptococci

(repeat regions/antigenic variation/IgA/IgG/fibrinogen)

PAUL O'TOOLE*, LARS STENBERG, MARIANNE RISSLER, AND GUNNAR LINDAHL†

Department of Medical Microbiology, University of Lund, Sölvegatan 23, S-22362 Lund, Sweden

Communicated by Werner Arber, June 15, 1992 (received for review March 16, 1992)

ABSTRACT **The M protein family of molecules in the group A streptococcus comprises a number of cell surface proteins that interact with the immune system of the host. One of the proteins in this family is the IgA receptor Arp4, which has C repeats similar to those that characterize the known M proteins. The streptococcal strain expressing Arp4 also expresses a second immunoglobulin-binding protein, Mrp4, which is shown here to be encoded by a gene located immediately upstream of the gene for Arp4. In addition to binding IgG, Mrp4 also binds fibrinogen, a property ascribed to M proteins. DNA sequence analysis demonstrated that the Mrp4 protein indeed is a member of the M protein family, but it was unexpectedly found to have a type of repeat that is identical to the A repeat described for FcRA76, a partially sequenced streptococcal Fc receptor. Purified FcRA76 was shown to bind fibrinogen and IgG, like Mrp4. These data show that the known molecules in the M protein family can be divided into two classes, A and C, according to the type of repeat region found. Hybridization studies with a panel of clinical isolates indicate that many streptococcal strains express class A and class C proteins, whereas some strains express only class C proteins. Class A molecules show amino-terminal sequence variation, like class C molecules, which suggests that proteins of both classes are targets for the immune response.**

The group A streptococcus is a ubiquitous human pathogen that is the causative agent of various diseases, ranging from acute pharyngitis to rheumatic fever (1). Among the virulence factors of this bacterium, considerable attention has been devoted to the M protein, a fibrillar cell surface molecule with antiphagocytic function (2). Characterization of many streptococcal strains has shown that the M protein occurs in numerous antigenic variants, but the mechanism that generates the structural variation is poorly understood (3). The M protein is therefore an interesting model system not only for analyzing how microorganisms interfere with phagocytosis but also for studies of antigenic variation.

DNA sequence analysis has shown that all M proteins so far studied are structurally related and are therefore encoded by a family of genes (4–8). The regions of amino acid sequence homology in the proteins include the signal sequence, the C repeat region in the central part of the protein chain, and the carboxyl-terminal part. Several streptococcal immunoglobulin-binding proteins have also been shown to be members of the M protein family and to have C repeats (9–12), whereas two other group A streptococcal cell surface proteins, the T6 protein and the C5a peptidase, are not encoded by genes in this family (13, 14).

One of the immunoglobulin-binding proteins in the M protein family is Arp4, an IgA receptor expressed by a strain of serotype M4 (9). This M4 strain also expresses a second immunoglobulin-binding protein, Mrp4, encoded by a gene that is closely linked to the gene for Arp4 (15). The Mrp4 protein was first isolated as a fibrinogen-binding protein (16), but we subsequently showed that it also binds IgG (15). Since fibrinogen binding is a property ascribed to M proteins (17), Mrp4 was expected to be a member of the M protein family and to have the C repeat region, which is considered to be an essential feature of these molecules (18). The DNA sequence analysis described here shows that Mrp4 is indeed a member of the M protein family, but it was unexpectedly found to have a type of repeat, the A repeat, which is not homologous to the C repeat. Such A repeats have been found previously in FcRA76, a partially sequenced streptococcal Fc receptor with evolutionary relatedness to the M proteins (19). These and other data demonstrate that the known members of the M protein family can be divided into two major classes, A and C, characterized by distinct repeat sequences.‡

MATERIALS AND METHODS

Bacterial Strains, Plasmids, Bacteriophage, and Culture Conditions. The clinical isolates of group A streptococci and strains AP4 (type M4), AL168 (M22), AL368 (M28), and AW43 (M60) have been described (15). The M1 strain AP1 (11) was obtained from G. Kronvall (Karolinska Institute), the M6 strain D471 (4) was from V. Fischetti (Rockefeller University), and the M76 strain CS110 (19) was from C. Schalén (University of Lund). The *Escherichia coli* strain TB1 (20) was used as host for recombinant plasmids. The plasmid vectors pBR322 (21) and pUC18/19 (20) were employed. The plasmid pSIR2201 is a pUC18 recombinant directing expression of protein Sir22 (ref. 15; L.S., unpublished data). The λ EMBL3 clones expressing the Mrp4 protein have been described (15, 16), as has culture of bacterial strains and propagation of λ (21).

Proteins. The immunoglobulins and other serum proteins used were of human origin (15). The IgG-binding FcRA76 protein (19) was extracted from strain CS110 by treatment with mutanolysin and 3 M KCl, as described for the M22 strain AL168 (15), and purified by affinity chromatography on IgG-Sepharose, followed by gel filtration. The identity of the isolated FcRA76 protein was confirmed by showing that it reacted strongly with antiserum to the closely related Mrp4 protein (see below) but not with antiserum to protein Arp4.

DNA Preparations. DNA from λ clones and *E. coli* plasmid DNA were prepared according to standard methodology (21). Genomic DNA of group A streptococci was also prepared as described (22).

DNA Manipulations and Sequence Determination. Standard procedures were employed for plasmid cloning experiments (21). Double-stranded chain-termination DNA sequencing was performed with Sequenase (United States Biochemical) according to the manufacturer's instructions. Template plas-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: OF, opacity factor.

*Present address: Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, V8W 3P6, Canada.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M87831).

mid DNA was generated by the method of Misra (23), using exonuclease BAL-31 to generate ordered sets of deletions in cloned fragments. Alternatively, oligonucleotides complementary to sequenced regions of the cloned fragment were used as sequencing primers. Computer search for sequence homology was made in the GenBank and EMBL data bases. Analysis of DNA sequence data and homology searches were performed using the University of Wisconsin Genetics Computer Group package (24).

Hybridization Analysis. DNA samples for analysis by dot-blot hybridization were denatured and applied to nitrocellulose by standard methodology (21). The filters were probed with radiolabeled oligonucleotides, which were end-labeled with polynucleotide kinase and [γ-^{32}P]dATP. The oligonucleotides used for A repeats and C repeats had the sequences 5'-GATCTGCAAG(T/C)TAAGCTAGAT-3' and 5'-GC(A/G)TC(A/T)CG(T/C)GAAGCTAA(A/G)AAA-3', respectively. These sequences were selected on the basis of alignment of all known DNA sequences for molecules with A or C repeats.

Polymerase Chain Reaction (PCR). Amplification of DNA using the PCR was performed according to standard methodologies (25), using the thermostable polymerase AmpliTaq (Perkin–Elmer/Cetus). For amplification of the A repeat region of the *mrp60* gene, primers corresponding to residues 658–674 and 990–974 of the *mrp4* gene (Fig. 1) were employed, with an annealing temperature of 42°C. Thirty-six cycles were performed. The PCR product was made blunt-ended by filling in with Klenow polymerase and cloned into the *Sma* I site of pUC18 for DNA sequence analysis.

Other Methods. Proteins were radiolabeled with the Bolton–Hunter reagent (Amersham). The procedures for denaturing protein electrophoresis and dot-blot analysis have been detailed (15) and the antisera to the Arp4 and Mrp4 proteins have been described (16). Production of opacity factor (OF) was monitored as described (26).

RESULTS

Subcloning and Sequencing of the *mrp4* Gene. Using one of several previously described λ clones that express the Mrp4 protein (15, 16), we subcloned the entire insert as a 13-kilobase-pair (kbp) *Sal* I fragment into pBR322, to generate the recombinant plasmid pMRP4001. The *E. coli* strain harboring pMRP4001 had reduced growth rate and further attempts to subclone the *mrp4* gene were unsuccessful. These results suggested that *mrp4* or some closely linked sequence had a toxic effect on *E. coli*. Such a toxic effect had been suspected when the λ clones were first isolated, since these clones form small plaques (16). Several λ mutants forming normal-sized plaques were isolated from such small-plaque λ clones, and these mutants were shown to have lost the ability to express the *mrp4* gene (16). However, among 20 independent mutants we subsequently found 1 that still expressed fibrinogen binding, like the parental phage, and this phage was chosen for new subcloning experiments. As shown below, the mutation in this λ clone does not effect the *mrp4* gene.

The *mrp4* gene was subcloned from the λ mutant, using immunoblotting with anti-Mrp4 antibodies to detect expression of the gene. A 7.5-kbp fragment expressing Mrp4 was cloned into pUC19 to form the plasmid pMRP4005. *E. coli* cells harboring pMRP4005 grew normally and produced a protein that comigrated in SDS/PAGE with purified Mrp4 protein (15). Through a process of subcloning and BAL-31 deletion, we then isolated the plasmid pMRP4009, which directed expression of Mrp4 and contained 2.6 kbp of streptococcal DNA in pUC18.

The *mrp4* gene in pMRP4009 was found to be truncated and part of the sequence was therefore determined from pMRP4005. The nucleotide sequence of the *mrp4* gene, which consists of 388 codons, is shown in Fig. 1.

```
               AACTGACCTTTACCTTTTGGCTTTTTTTATTTAGAATAATTTTATTGGAGAGATGCTTAATAATTTAAGCACAATTCTTAGAAATTGAGAAATAAGGAGTAAACA   105
                  -35   -35                  -10                                                                 RBS

ATGTCTAAAACAAATCCAAACAAACTCTATTCACTGAGAAAGTTAAAAACAGGTACTGCATCAGTAGCAGTAGATTTGACAGTTTTGGGGACTGGATTAGCAAACACAACTGATGTAAAG   225
MetSerLysThrAsnProAsnLysLeuTyrSerLeuArgLysLeuLysThrGlyThrAlaSerValAlaValAspLeuThrValLeuGlyThrGlyLeuAlaAsnThrThrAspValLys    -2

GCTGAGAGTCGTCGTTATCAGGCACCTCCTCGTGTGTTACTGCAAGGCAAAGAAGCTAACAAAGTATTCGAAGAGCGCAAAGCCTTGGAAAAACAAGCACGTGATTTGGGTGACACTATT   345
AlaGluSerArgArgTyrGlnAlaProProArgValLeuLeuGlnGlyLysGluAlaAsnLysValPheGluGluArgLysAlaLeuGluLysGlnAlaArgAspLeuGlyAspThrIle    39
 +1

AACCACATGTCACAAACCATTAGCGAGCAAAGCCGCAAGATTGCAGCACTAAAGTCTGAAGCAGAACTTAAAAACCAACAAGCTCTTGAAGCTTTAAACAATAAAAACAAGCAAATCTCA   465
AsnHisMetSerGlnThrIleSerGluGlnSerArgLysIleAlaAlaLeuLysSerGluAlaGluLeuLysAsnGlnGlnAlaLeuGluAlaLeuAsnAsnLysAsnLysGlnIleSer    79

GATTTAACCAACGAAAACGCACAGTTAAAAGAAGCCATTGAAGGTTATGTGCAAACTATCCAAAACGCTAGTCGTGAAATCGCAGCAAAACAACAAGAACTTGCAGCTGCAAAAAGCCAG   585
AspLeuThrAsnGluAsnAlaGlnLeuLysGluAlaIleGluGlyTyrValGlnThrIleGlnAsnAlaSerArgGluIleAlaAlaLysGlnGlnGluLeuAlaAlaAlaLysSerGln   119

TTAGAGGCAAAAAATGCTGAGATTGAGGCATTGAAACAACAAGATGCCTCTAAGACTGAGGAAATTGCTAAATTGCAATCAGAAGCAGCAACTTTAGAAAACCTCCTAGGTTCAGCTAAG   705
LeuGluAlaLysAsnAlaGluIleGluAlaLeuLysGlnGlnAspAlaSerLysThrGluGluIleAlaLysLeuGlnSerGluAlaAlaThrLeuGluAsnLeuLeuGlySerAlaLys   159
                                                   └→A1

CGTGAGTTGACTGAATTGCAAGCTAAGCTAGATACAGCAACTGCTGAAAAAGCAAAACTAGAATCACAAGTAACAACCTTAGAAAACCTCCTAGGTTCAGCTAAGCGTGAATTGACTGAT   825
ArgGluLeuThrGluLeuGlnAlaLysLeuAspThrAlaThrAlaGluLysAlaLysLeuGluSerGlnValThrThrLeuGluAsnLeuLeuGlySerAlaLysArgGluLeuThrAsp   199
                                     └→A2

CTGCAAGCTAAGCTAGATGCAGCTAACGCTGAAAAAGAAAAGCTCCAATCACAAGCAGCAACCCTAGAAAAACAACTAGAAGCAACTAAAAAAGAGTTAGCTGATTTACAGGCTAAATTA   945
LeuGlnAlaLysLeuAspAlaAlaAsnAlaGluLysGluLysLeuGlnSerGlnAlaAlaThrLeuGluLysGlnLeuGluAlaThrLysLysGluLeuAlaAspLeuGlnAlaLysLeu   239
                        └→A3

GCAGCAACCAACCAAGAAAAAGAAAAGTTAGAAGCTGAAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAGCTTGCTAAGCTAAAAGCAGATAAAGCTTCAGGAGCTCAA   1065
AlaAlaThrAsnGlnGluLysGluLysLeuGluAlaGluAlaLysAlaLeuLysGluGlnLeuAlaLysGlnAlaGluGluLeuAlaLysLeuLysAlaAspLysAlaSerGlyAlaGln   279
     ──┘                 └→22R                                                       ──┘

AAACCAGATACTAAACCTGGCAATAAAGAGGTTCCAACAAGACCGTCACAAACAAGAACAAACACTAATAAAGCTCCTATGGCTCAAACAAAGAGACAATTACCGTCAACAGGCGAAGAA   1185
LysProAspThrLysProGlyAsnLysGluValProThrArgProSerGlnThrArgThrAsnThrAsnLysAlaProMetAlaGlnThrLysArgGlnLeuProSerThrGlyGluGlu   319

ACAACCAACCCATTCTTCACTGCAGCAGCATTGACAGTGATCGCATCTGCAGGCGTACTTGCCCTAAAACGCAAAGAAGAAAACTAAGTCCAACCCACATTATCTTTTCTAGCCCAAGAA   1305
ThrThrAsnProPhePheThrAlaAlaAlaLeuThrValIleAlaSerAlaGlyValLeuAlaLeuLysArgLysGluGluAsn                                         347
         └→M                                                  ──┘

AAAAACAAAAAAAGAGGAAGCCCCTTCCTCTTTTTTTGAACGGTTAAACAGCAAAAAGGTCAAAAAGGCACTAAAGTCTCAAAAACCTGGTCTTT                         1400
```

FIG. 1. Nucleotide sequence and derived amino acid sequence of the *mrp4* gene. Possible −35 and −10 promoter sequences, a possible ribosome binding site (RBS), and a potential transcription termination signal are underlined. Numbering of residues in Mrp4 begins with the first amino acid in the processed protein, labeled +1 (15). The three A repeats, a 22-residue sequence (22R) common to molecules in the M protein family, and the putative membrane anchor region (M) are indicated. The last 10 residues of this DNA sequence correspond to the first 10 of the published nucleotide sequence for *arp4* (9).

The *mrp4* gene that had been sequenced originated from a λ mutant, as described above, and it therefore seemed possible that we had sequenced a mutated *mrp4* gene. To investigate this possibility, primers based upon this *mrp4* sequence were used for determining the corresponding sequence from pMRP4001, in which the insert was derived directly from the parental small-plaque λ clone. The sequence in the *mrp4* structural gene and flanking regions was identical to that shown in Fig. 1. Preliminary data indicate that the λ mutant is affected in a region upstream of *mrp4*.

The *mrp4* and *arp4* genes are known to be closely linked (15, 16). The sequencing of *mrp4* revealed that part of the coding sequence and all of the downstream flanking sequence are identical to the sequence upstream of the *arp4* gene (ref. 9 and L.-O. Hedén, personal communication). The *mrp4* and *arp4* genes are therefore adjacent in the streptococcal chromosome, with an intergenic distance of 229 nucleotides.

Analysis of the Amino Acid Sequence of the Mrp4 Protein. Comparison of the amino acid sequence deduced from the *mrp4* gene with the amino-terminal sequence of the processed Mrp4 protein (15) indicates the presence of a 41-residue signal peptide. This signal peptide shows 62–93% residue identity with the known signal peptides of M proteins and immunoglobulin-binding proteins in group A streptococci. The processed form of Mrp4, 347 residues, has a calculated molecular weight of 37,987, in good agreement with SDS/PAGE analysis (15). The deduced amino acid composition of the Mrp4 protein shows a high content of alanine, glutamic acid, lysine, and leucine, as reported for protein Arp4 and M proteins (3, 9).

The Mrp4 protein shows organizational features that characterize molecules in the M protein family (3, 10). The amino terminus includes a typical signal sequence. Three repeats, each comprising 35 amino acid residues, are found in the central part of the sequence (Fig. 1). These repeats, designated A1–A3, show 54–77% amino acid residue identity with each other. Surprisingly, the A repeats in Mrp4 show negligible homology with the C repeats common to all M proteins sequenced to date, but, as discussed further below, are identical to the A1–A3 repeats described for the IgG-binding FcRA76 protein (19).

Seven amino acid residues downstream from the third repeat, a 22-residue sequence begins that is present in all group A streptococcal M proteins and immunoglobulin receptors sequenced to date. It is conserved to 100% in Arp4, Arp60, FcRA76, protein H, M5, M6, M12, and M49 and there is a single divergent residue in the corresponding region of the M24 molecule (4–11, 19). This 22-residue motif is also present in two Arp-like molecules expressed by a strain of type M2 (12), but these two molecules share three residues that diverge from the otherwise perfect consensus for this region.

At the carboxyl-terminal end of the Mrp4 polypeptide, there is a putative membrane anchor region (3) that shows 75% residue identity with corresponding regions in all known M proteins and immunoglobulin receptors from the group A streptococcus.

Analysis of the secondary structure of Mrp4 predicts that the protein is almost totally α-helical, like M proteins (3). Another similarity to M proteins and immunoglobulin receptors is a seven-residue periodicity in the distribution of nonpolar amino acids (3, 9, 19), which is found in the region between residues 78 and 273 of Mrp4.

Sequence Homology Between Mrp4 and FcRA76. As described above, three regions in Mrp4 have significant homology with all known members of the M protein family: the signal peptide, the conserved 22-residue region, and the carboxyl-terminal region (which includes the membrane anchor). This homology is exemplified in Fig. 2 by the alignment of Mrp4 and Arp4, which are expressed by the same strain. A more dramatic homology was observed between Mrp4 and the partially sequenced FcRA76 protein (19). Allowing minimum gaps, the overall protein sequences are 78% identical, while being 93% identical in the signal peptides and 99% identical in the carboxyl-terminal two-thirds (Fig. 2). The latter encompasses the three A repeats. The only region of dissimilarity is the amino-terminal region of the processed polypeptides, comprising 53 residues of Mrp4 and 92 residues of FcRA76. At the DNA level, the homology displayed by the

```
Arp4    MARKDTNKQYSLRKLKTGTASVAVAVAVLGAGFANQTEVKAAEIKK-PQADSAWNWPKEYNALLKENEELKV------EREKYLSYADDKEKDPQYRALMGENQDLRKREGQYQDKIEEL   71
Mrp4    MSKTNPNKLYSLRKLKTGTASVAVDLTVLGTGLANTTDVKAES---------RR-----------YQAPPRVLLQGKEANKV----------------FEERKALEKQARDLGDTINHM   42
FcRA76  MSKRNPNKHYSLRKLKTGTASVAVALTVLGTGLANTTDVKAETVGRFSDEQVRKAREKAIEDVFDGYTGARSVYQSGNLPNRLTPTKLSKLMQQMYKETLQKKEELDTLSKALTHTIEKK   79

Arp4    EKERKEKQERQEQLERQYQIEADKHYQEQQKKHQQEQQQLEAEKQKLAKDKQISDASRQGLSRDLEASR----AAKKELEAEHQKLKEEKQISDASRQGLSRD---------LEASREA   168
Mrp4    SQTISEQSRKIAALKSEAELKNQQALEALNNKNKQISD-LTNENAQLKEAIEGYVQTIQNASREIAAKQQELAAAKSQLEAKNAEIEALKQQDASKTEEIAKLQSEAATLENLLGSAKRE   161
FcRA76  IESENAYKKELGQLKAAAEAEAQKALDALNNKNKQISD-LTNENAQLKEAIEGYVQTIQNASREIAAKQQELAAAKSQLEAKNAEIEALKQQDASKTEEIAKLQSEAATLENLLGSAKRE   198

Arp4    KKKVEADLAALTAEHQKLKE-----DKQISDASRQ--GLSRDLEASREAKKKVEADLAEANSKLQALEKLNKELEEGKKLSEKEKAELQARLEAEAKALKEQLAKQAEELAKLK-----G   300
Mrp4    LTELQAKLDTATAEKAKLESQVTTLENLLGSAKRELTDLQAKLDAANAEKEKLQSQAATLEKQLEATKKELADLQA--KLAATNQE--KEKLEAEAKALKEQLAKQAEELAKLKADKASG   277
FcrA76  LTDLQAKLDTATAEKAKLESQVTTLENLLGSAKRELTDLQAKLDAANAEKEKLQSQAAALEKQLEATKKELADLQA--KLAATNQE--KEKLEAEAKALKEQLAKQAEELAKLKADKASG   314

Arp4    NQTPNAKVA----------PQANRSRSAMTQQKRTLPSTGE-TANPFFTAAAATVMVSAGMLALKRKEEN   345
Mrp4    AQKPDTKPGNKEVPTRPSQTRTNTNKAPMAQTKRQLPSTGEETTNPFFTAAALTVIASAGVLALKRKEEN   347
FcrA76  AQKPDTKPGNKEVPTRPSQTRTNTNKAPMAQTKRQLPSTGEETTNPFFTA--------------------   364
```

FIG. 2. Triple alignment of Arp4, Mrp4, and the available sequence of FcRA76. Gaps are indicated by empty spaces; residue identities are indicated by colons. Arp4 and FcRA76 were individually aligned with Mrp4 using the parameters K-tuple size 1, gap penalty 2, and window size 10. Slightly variant alignments are also possible. Numbering of the amino acids, indicated to the right, begins with the first residue after removal of the signal peptide. Three regions in which all three proteins have significant homology, with each other and with group A streptococcal M proteins and immunoglobulin receptors, are boxed: the signal peptide, a 22-residue sequence on the carboxyl-terminal side of the repeat region, and the carboxyl-terminal region.

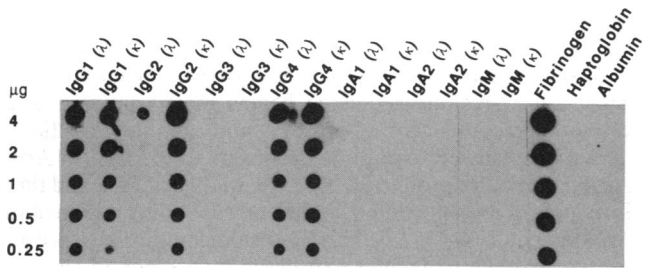

FIG. 3. Dot-blot analysis of the FcRA76 protein. The indicated amounts of human serum proteins, diluted in phosphate-buffered saline, were applied in aliquots of 100 μl. After blocking, the membrane was probed with ^{125}I-labeled FcRA76, washed, and subjected to autoradiography.

coding sequences is reinforced by almost complete identity of the available upstream flanking regions (99% identity).

Mrp4 and FcRA76 Proteins Have Similar Binding Properties. Since the Mrp4 and FcRA76 proteins exhibit extensive sequence homology, it seemed possible that they would have similar binding properties. A dot-blot analysis (Fig. 3) demonstrated that FcRA76 indeed has binding properties similar to those of the Mrp4 protein (15) and binds fibrinogen in addition to IgG1, IgG2, and IgG4 proteins.

Distribution of the A and C Repeat Sequences in Group A Streptococci. Three strains of M types 22, 28, and 60 express a Mrp protein that has binding properties similar to the Mrp4 protein (15). It therefore seemed likely that expression of a protein characterized by the A repeats found in Mrp4 and FcRA76 might be a widespread property among group A streptococci. Two approaches were taken to investigate this hypothesis. First, a DNA sequence corresponding to most of the A repeat region of Mrp4 was amplified by PCR and cloned from a λ clone known to harbor the *mrp60* gene (15). The DNA sequence of this region was almost identical to the corresponding sequence of *mrp4*, with a total of five divergent nucleotides in the 330-residue-long region amplified. This result reinforces the correlation between the expression of a Mrp protein and presence of the A repeat.

As a second approach, the distribution of nucleotide sequences encoding A repeats and C repeats was studied by DNA hybridization analysis of different streptococcal strains, using oligonucleotide probes specific for each type of repeat sequence. Seven reference strains of known M type and a selection of 38 clinical isolates were analyzed. Data for 16 of the clinical isolates and for the reference strains are shown in Fig. 4. The clinical isolates were chosen to represent four different immunoglobulin-binding patterns previously defined (15, 27). The results clearly indicated that the C repeat sequence is present in all strains tested, whereas the A repeat could be detected in those reference strains that are known to express a protein similar to Mrp4 and in 22 of the 38 clinical isolates. The gene for protein Sir, an immunoglobulin receptor with recently described properties (15), was also shown to have C repeats. Presence of the A repeat sequence correlated precisely with expression of OF.

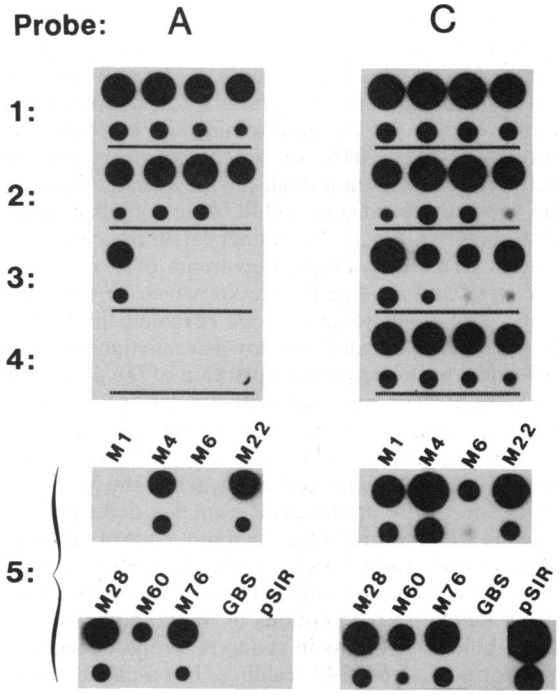

FIG. 4. Dot-blot hybridization of genomic DNA from group A streptococci with oligonucleotide probes specific for the A and C repeats. For each DNA sample, 1500 ng (top) and 150 ng (bottom) were applied to duplicate filters that were then hybridized with the A repeat and C repeat probes as indicated. Groups 1–4, DNA from clinical isolates representing different immunoglobulin-binding patterns previously defined (15, 27). Group 1, Four strains that bind IgA and IgG well, like the M22 and M28 reference strains; group 2, four strains that bind IgA well but IgG weakly, like the M4 and M60 strains; group 3, four strains that bind only IgG, like the M1 and M76 strains; group 4, four strains that do not bind immunoglobulin, like the M6 strain. Group 5, DNA from seven reference strains of indicated M types; GBS, DNA from a group B streptococcus; pSIR, a recombinant plasmid expressing protein Sir22.

DISCUSSION

The group A streptococcal strain of type M4 studied here expresses two cell surface proteins with related properties and of similar size, the Arp4 and Mrp4 proteins (9, 15, 16). The present work demonstrates that these two proteins are encoded by adjacent genes, with the *mrp4* gene upstream and an intergenic spacing of 229 nucleotides. The structure of the two proteins is schematically illustrated in Fig. 5. Protein Arp4 is known to be structurally similar to all known M proteins and to have the typical C repeats (9), and the properties of Mrp4 led us to expect that this protein would have a similar structure. The lack of C repeats in Mrp4 was therefore surprising, as was the similarity to the partially sequenced FcRA76 protein (19). Although the A repeat found in these two proteins does not show significant homology to the C repeat, Mrp4 and FcRA76 can be classified as members of the M protein family by virtue of significant homology in other regions of the

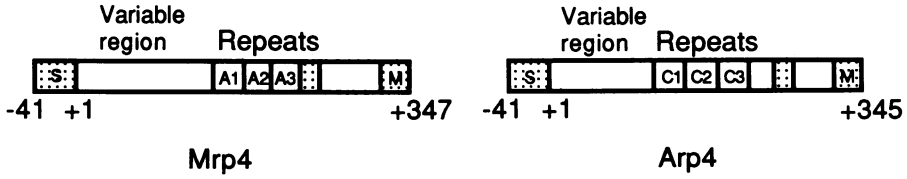

FIG. 5. Schematic representation of Mrp4 and Arp4, two cell surface proteins expressed by strain AP4. These molecules are members of class A and class C, respectively, in the M protein family. Shaded areas indicate regions of extensive sequence homology to all known proteins in the M protein family. The numbering of amino acids begins with the first residue in the processed proteins, labeled +1. In the variable region molecules of the same class vary in sequence. S, signal sequence; M, membrane anchor region.

molecules (Fig. 2). In Mrp4, these regions include the signal sequence, a 22-residue sequence located on the carboxyl-terminal side of the A repeat region, and the carboxyl-terminal region. In FcRA76 the carboxyl-terminal region is truncated, but the similarities to M proteins (19) and to Arp4 (Fig. 2) clearly show that it is a member of the M protein family. Repeats of type A are probably also present in the Mrp60 protein, which suggests that this molecule is a member of the M protein family, like Mrp4 and FcRA76. We conclude from these data that the known members of the M protein family can be divided into two classes, A and C, characterized by distinct repeat regions. The molecules in class C can be further divided into two subclasses, CI and CII, based on the fine structure of the C repeat region (3, 18). It should also be noted that the six group A streptococcal immunoglobulin-binding proteins that have been sequenced (refs. 9–12 and 19, and this report) all are members of the M protein family and can be assigned either to class A or to class C.

The expression of a class A molecule and a class C molecule by a single strain is not a unique property of the M4 strain described above, since the strain of type M60 studied by us also expresses two such proteins, the Mrp60 and Arp60 proteins (ref. 10 and this report). Furthermore, two strains of serotypes M22 and M28 express a protein of the Mrp type, which probably has A repeats, and a protein designated Sir (15), which has C repeats (Fig. 4). The hybridization analysis with DNA from different clinical isolates also indicates that many streptococcal strains have genes encoding a class A molecule and a class C molecule. However, some strains probably express only class C molecules (Fig. 4). Such strains express either a single gene or two adjacent genes with C repeats (3, 12). It should also be noted that the presence of A repeats correlates with the ability to produce OF. The division of group A streptococci into two classes, OF^+ and OF^- strains (26), therefore parallels our division of such strains into those that express class A and class C molecules and those that express only class C molecules.

Notwithstanding the distinctive repeat sequence, proteins in class A have several properties that characterize molecules in the M protein family, including seven-residue periodicity in the distribution of nonpolar amino acids, overall organization, and the ability to bind fibrinogen and/or immunoglobulin. Furthermore, class A molecules exhibit amino-terminal sequence variation, like class C molecules (3, 10), as shown by the alignment of Mrp4 and FcRA76 (Fig. 2) and by a comparison of the previously determined amino-terminal sequences of Mrp4 and Mrp60 (15). This observation suggests that molecules of both classes are targets for the immune response (28).

In summary, all available evidence indicates that the known members of the M protein family can be divided into two major classes, A and C, characterized by distinct repeat sequences. We believe that this finding may be significant for understanding the pathogenic mechanisms in group A streptococcal infections.

We thank Drs. Hans Jörnvall and Tomas Lindahl for reading the manuscript and Annika Andersson for valuable assistance. Monoclonal human IgG proteins of different subclasses were generously provided by Dr. F. Skvaril, World Health Organization/International Union of Immunological Societies Immunoglobulin Subcommittee (Berne, Switzerland). This work was supported by grants to G.L. and P.O. from the Swedish Medical Research Council and the Erik Philip-Sörensen Foundation and to G.L. from the Medical Faculty of the University of Lund, the Royal Physiographic Society in Lund, the Foundations of Crafoord, Kock, and Österlund, HighTech Receptor, Inc., and by King Gustaf V's 80-Year Foundation.

1. Bisno, A. L. (1991) *N. Engl. J. Med.* **325,** 783–793.
2. Lancefield, R. C. (1962) *J. Immunol.* **89,** 307–313.
3. Fischetti, V. A. (1989) *Clin. Microbiol. Rev.* **2,** 285–314.
4. Hollingshead, S. K., Fischetti, V. A. & Scott, J. R. (1986) *J. Biol. Chem.* **261,** 1677–1686.
5. Robbins, J. C., Spanier, J. G., Jones, S. J., Simpson, W. J. & Cleary, P. P. (1987) *J. Bacteriol.* **169,** 5633–5640.
6. Miller, L., Gray, L., Beachey, E. & Kehoe, M. (1988) *J. Biol. Chem.* **263,** 5668–5673.
7. Mouw, A. R., Beachey, E. H. & Burdett, V. (1988) *J. Bacteriol.* **170,** 676–684.
8. Haanes, E. J. & Cleary, P. P. (1989) *J. Bacteriol.* **171,** 6397–6408.
9. Frithz, E., Hedén, L. & Lindahl, G. (1989) *Mol. Microbiol.* **3,** 1111–1119.
10. Lindahl, G., Åkerström, B., Stenberg, L., Frithz, E. & Hedén, L. (1991) in *Genetics and Molecular Biology of Streptococci, Lactococci, and Enterococci*, eds. Dunny, G. M., Cleary, P. P. & McKay, L. L. (Am. Soc. Microbiol., Washington), pp. 155–159.
11. Gomi, H., Hozumi, T., Hattori, S., Tagawa, C., Kishimoto, F. & Björck, L. (1990) *J. Immunol.* **144,** 4046–4052.
12. Bessen, D. E. & Fischetti, V. A. (1992) *Infect. Immun.* **60,** 124–135.
13. Schneewind, O., Jones, K. F. & Fischetti, V. A. (1990) *J. Bacteriol.* **172,** 3310–3317.
14. Chen, C. C. & Cleary, P. P. (1990) *J. Biol. Chem.* **265,** 3161–3167.
15. Stenberg, L., O'Toole, P. & Lindahl, G. (1992) *Mol. Microbiol.* **6,** 1185–1194.
16. Lindahl, G. (1989) *Mol. Gen. Genet.* **216,** 372–379.
17. Whitnack, E. & Beachey, E. H. (1985) *J. Bacteriol.* **164,** 350–358.
18. Bessen, D. E. & Fischetti, V. A. (1990) *J. Exp. Med.* **172,** 1757–1764.
19. Heath, D. G. & Cleary, P. P. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4741–4745.
20. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
22. Spanier, J. G. & Cleary, P. P. (1984) *Virology* **130,** 514–522.
23. Misra, T. K. (1985) *Gene* **34,** 263–268.
24. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
25. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic, New York).
26. Maxted, W. R., Widdowson, J. P., Fraser, C. A. M., Ball, L. C. & Bassett, D. C. J. (1973) *J. Med. Microbiol.* **6,** 83–90.
27. Lindahl, G. & Stenberg, L. (1990) *Epidemiol. Infect.* **105,** 87–93.
28. Borst, P. (1991) in *Immunoparasitology Today*, eds. Ash, C. & Gallagher, R. B. (Elsevier, Cambridge, U.K.), pp. 29–33.