Volume 16 Number 10 1988 Nucleic Acids Research

Comparison of the leader sequences of four group A streptococcal M protein genes

Elizabeth Haanes-Fritz*, Werner Kraus[1], Vickers Burdett[2], James B.Dale[1], Edwin H.Beachey[1] and Patrick Cleary

Department of Microbiology, University of Minnesota, Minneapolis, MN 55455, [1]Veterans Administration Medical Center and University of Tennessee, Memphis, TN 38104 and [2]Department of Microbiology, Duke University Medical Center, Durham, NC 27710, USA

Received December 23, 1988; Revised and Accepted April 14, 1988

ABSTRACT

The 5' portions and flanking sequences of genes encoding types 1, 12, 24, and 6 M proteins were compared. Although the DNA sequences encoding the amino-termini of the mature M proteins had no obvious similarity, upstream sequences, and those encoding the signal peptides (leader sequences) of the four M protein genes had considerable similarity. In general, the 5' ends of all the leader sequences were more conserved than the 3' ends, although the M6 and M24 leader sequences had identical 3' ends. Sequence similarity among the deduced amino acid sequences of the four signal peptides was more extensive than the corresponding DNA sequences. We found that strict DNA similarity among all four sequences extended only to the ends of the hydrophilic amino-terminal regions of the signal peptides, but that amino acid sequence conservation continued to the ends of the respective hydrophobic cores. With the exception of the M6 and M24 sequences, the regions adjacent to the signal peptidase cleavage sites were highly variable.

INTRODUCTION

A major virulence factor of Streptococcus pyogenes is the M protein. This antiphagocytic surface structure blocks opsonization of the streptococcal cells by the alternate complement pathway (1,2). Antibodies against M protein develop in infected individuals, and confer type specific immunity (3). This streptococcal species, however, includes 80 or more antigenically distinct, but functionally similar M protein serotypes, thus permitting repeated infections to occur (4).

M protein molecules are attached to the streptococcal cell surface by their carboxyl ends, with exposed amino-termini (5). Monoclonal antibody studies suggest that different M protein types share common epitopes near their carboxy-termini and exhibit antigenic variability in the more exposed amino-terminal regions (6). In addition, studies utilizing DNA probes from a cloned M6 gene suggest that the 3' ends of M protein genes are more conserved than the 5' ends (7,8).

At present, the origins of streptococcal M protein antigenic diversity are poorly understood. Intragenic recombination between repeated regions in various M6 genes, resulting in M6 proteins of various sizes has been suggested as one possible generator of antigenic diversity (9). Many M protein type-specific and opsonic determinants, however, are located in the non-repetitive amino-terminal regions (6,9), and hence they were probably not generated by the proposed intragenic recombination mechanism. We suggest, then, that additional, yet undefined mechanisms may also contribute to the large degree of M protein antigenic diversity.

In our attempts to define alternative mechanisms for the generation of M protein antigenic diversity, we are focusing on the possible involvement of regions outside of the M protein structural genes. Recent reports of the nucleotide sequences of the M6 (10), M12 (11), and M24 (12) genes, in addition to the partial M1 sequence presented here, provide an opportunity to compare the nucleotide sequences immediately flanking and within the 5' ends of these genes, including sequences encoding the putative signal peptides of each gene. Although the 5' sequences of the four mature M protein structural genes exhibited considerable variability, we found extensive similarities in the upstream flanking regions and in the sequences encoding the signal peptides of the four genes.

MATERIALS AND METHODS

Bacterial Strains and Growth Media.

Streptococcus pyogenes strain CS130 (original designation: T1/195/2) was used as the source of DNA for cloning of the M1 gene. Streptococci were routinely stored frozen at -80°C in Todd-Hewitt Broth (THB) supplemented with 2% neopeptone (Difco Laboratories, Detroit, MI). For DNA isolations, streptococci were grown in THB supplemented with 1% yeast extract (Difco Laboratories, Detroit, MI). Escherichia coli strains NM538 and NM539 (13) were used for the propagation of vector and recombinant bacteriophage lambda, respectively. In vitro packaging extracts were prepared from E. coli strains BHB2688 and BHB2690 by standard methods (14). E. coli strain JM83 (15) was used as the host for plasmid vectors, and strain JM109 (16) was used to propagate M13 recombinants. E. coli strains were grown either in LB medium (17), or in 2xYT medium (16).

Preparation of Bacteriophage and Chromosomal DNAs.

Streptococcal chromosomal DNA was prepared as previously described (18). Bacteriophage Lambda DNAs were prepared from plate lysates using the rapid Lambda DNA preparation method described by Maniatis (17), except that plate lysates were prepared on 150 mm petri dishes using LB medium plus 1.5% agarose. For nucleotide sequencing, M13 single stranded templates were prepared in JM109 using a scaled up version of the method described by R. Dale (19).

Molecular Cloning Procedures.

Partial MboI or Sau3A1 restriction fragments of 12-20 kilobases (kb) in length from the group A streptococcal strain CS130 were cloned into the bacteriophage Lambda replacement vector EMBL3 (Promega Biotec, Madison, WI) according to the methods of Frishauf, et al, (13). Bacteriophage plaques were screened by hybridization (20). The probe was pPC124, a plasmid containing a portion of the cloned M12 gene and upstream flanking regions (11,21). Plasmid DNA was labelled with [^{32}P]dCTP (New England Nuclear, Boston, MA) using a nick translation kit (Bethesda Research Laboratories, Gaithersburg, MD).

Recombinant phage from the M1 library hybridizing to the M12 probe were subjected to restriction analysis, and appropriate DNA fragments were subcloned into plasmids pUC9 (15), and pUC18 (16) for further mapping, and into the M13 vectors mp18 and mp19 for nucleotide sequencing (16). Enzymes were purchased from Bethesda Research Laboratories, and Boehringer Mannheim Biochemicals (Indianapolis, IN) and were used according to the manufacturers' specifications.

DNA Sequencing.

DNA sequencing was done by the dideoxy-nucleotide chain termination method of Sanger (22), using [^{35}S]dATP (New England Nuclear). Overlapping templates for sequencing were derived by the T4 polymerase deletion method of R. Dale

(19), using the "Cyclone" kit (IBI, New Haven, CT). Enzymes and reagents for the sequencing reactions were from the "Sequel" kit (IBI).

Computer Sequence Analyses.

Nucleotide and amino acid sequences were compared and aligned using the algorithms of Wilbur and Lipman (23), Lawrence and Goldman (submitted), and Altschul and Erickson (24), provided in the IntelliGenetics and MBIR software packages. Sequence editing and translation deductions were carried out on the software of David Mount.

Peptide Synthesis and Conjugation.

Based on the predicted amino acid sequence of the M1 protein determined in the present study, and by protein sequence analyses of isolated polypeptide fragments reported elsewhere (25), we synthesized a peptide containing residues 1-23 of the amino-terminus of the mature M1 protein. The peptide, with an additional carboxy-terminal cysteine residue, was synthesized by the solid phase method. The purity and composition of the peptides was assured by HPLC on Ultrasphere ODS2 (Whatman), quantitative amino acid analysis and automated Edman degradation to the pentultimate amino acid residue. The synthetic peptide designated S-M1(1-23)C was conjugated to Keyhole Limpet Hemocyanin (KLH).

Rabbit Immunization and Assays of Anti-M Protein Antibodies.

Three New Zealand White rabbits (2 kg) were immunized with a 50 nmol dose of S-M1(1-23)C-KLH conjugate emulsified in Complete Freund's Adjuvant, and immune rabbit sera were assayed for antibodies against M protein by enzyme linked imunosorbent assay (ELISA) and opsonophagocytosis assays, as described (25).

RESULTS

Cloning of the M1 Gene.

In preliminary Southern hybridization studies, we learned that plasmid pPC124, a 1.9 kb HaeIII fragment containing the proximal 5' end and upstream flanking region of the M12 gene (21,11), hybridized strongly to a single HaeIII fragment in genomic DNA of the M1 strain, CS130 (26). Assuming that the strong hybridization represented either the upstream portion of a common M protein locus shared by the two M types, or common sequences within the two M protein genes, we used plasmid pPC124 DNA as a hybridization probe to isolate the M1 gene from two separate CS130 chromosomal libraries constructed in lambda EMBL3. Nineteen independent recombinants that hybridized to the probe were isolated. Each positive clone was plaque purified three times and was rechecked at each purification for continued hybridization to the probe.

Mapping and Subcloning of the M1 Recombinants.

The DNA from each recombinant was mapped with a variety of enzymes, and the restriction digests were subjected to Southern hybridization with plasmid pPC124 probe DNA to identify the locations of the homologous regions. Inserts of the recombinants were of various lengths and were arranged in both orientations with respect to the Lambda arms, but all the inserts contained the same contiguous restriction fragments. The bacteriophage recombinant λ bfX-2 (Fig. 1A) was chosen for further study. The two SalI fragments of the λ bfX-2 insert were subcloned into the plasmid vector pUC9. The plasmid subclone selected by hybridization to pPC124 DNA was named pbf17. A SalI/EcoRI restriction fragment of plasmid pbf17 that hybridized to pPC124 DNA was subcloned into pUC18 to produce pbf82. Restriction maps of these two subclones are shown in Fig. 1B. In preliminary studies, we showed that lysates and sonicated extracts of strain JM83 containing plasmids pbf17 and pbf82 reacted with M1 antiserum by immunodiffusion analyses, suggesting that the region cloned indeed included a portion

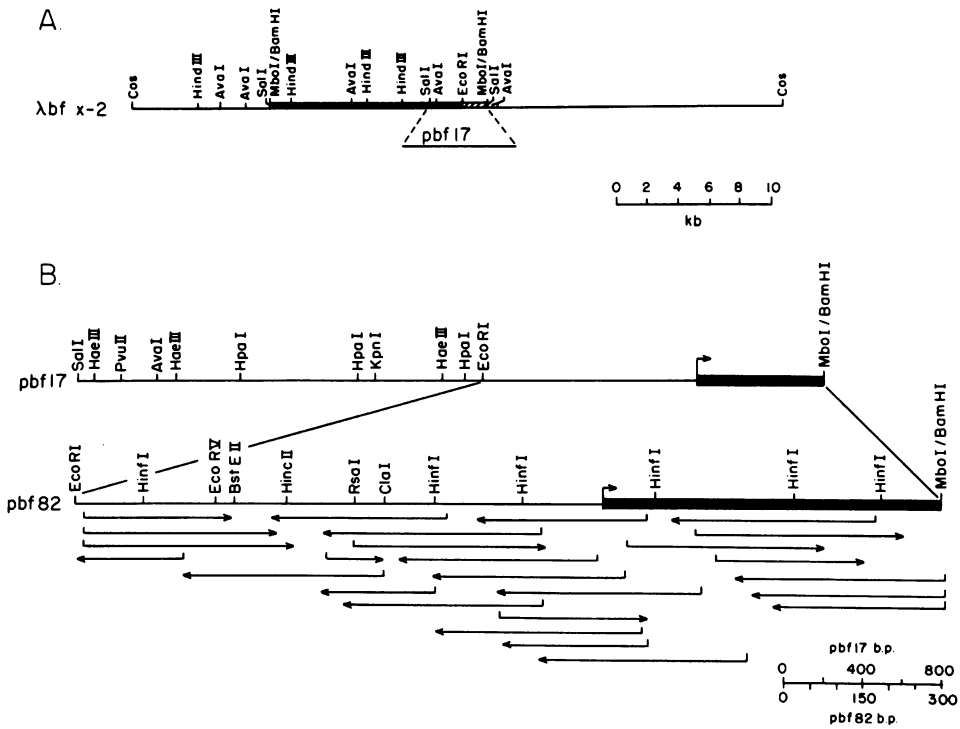

Fig. 1A
Restriction map of λ bfX-2. The thick line represents the streptococcal DNA inserted into the EMBL3 vector. The hatched portion of the insert shows the restriction fragment that hybridized to the pPC124 probe (see text). The portion that was subcloned into plasmid pbf17 (B) is indicated.

Fig. 1B
Restriction maps of the inserts from subclones JM83 (pbf17) and JM83 (pbf82). The M1 gene is represented on both maps by thick lines, with arrows showing the predicted direction of transcription. The strategy for nucleotide sequencing of pbf82 is shown by the arrows below the restriction map.

of the M1 gene (data not shown). A comparison of the restriction map of the pbf17 insert with that of the region upstream of the M12 gene revealed that many common fragments existed. The identity of these fragments was verified by DNA hybridization (26).

Nucleotide Sequencing of the M1 Gene Fragment.

The insert from plasmid pbf82 was cloned into phage M13 and sequenced according to the strategy diagrammed in Fig. 1B. The resulting DNA sequence is shown in Fig. 2. A 624 base pair (bp) open reading frame starting at nucleotide 1028 was postulated to be the 5' end of the M1 gene. The absence of a translation termination codon in frame, however, indicated the original cloned insert of λ bfX-2 did not contain the complete M1 gene. Promoter regions resembling those previously published (10,11) precede the M1 gene (see Fig. 3), but our study did not include transcriptional analysis to determine which -35

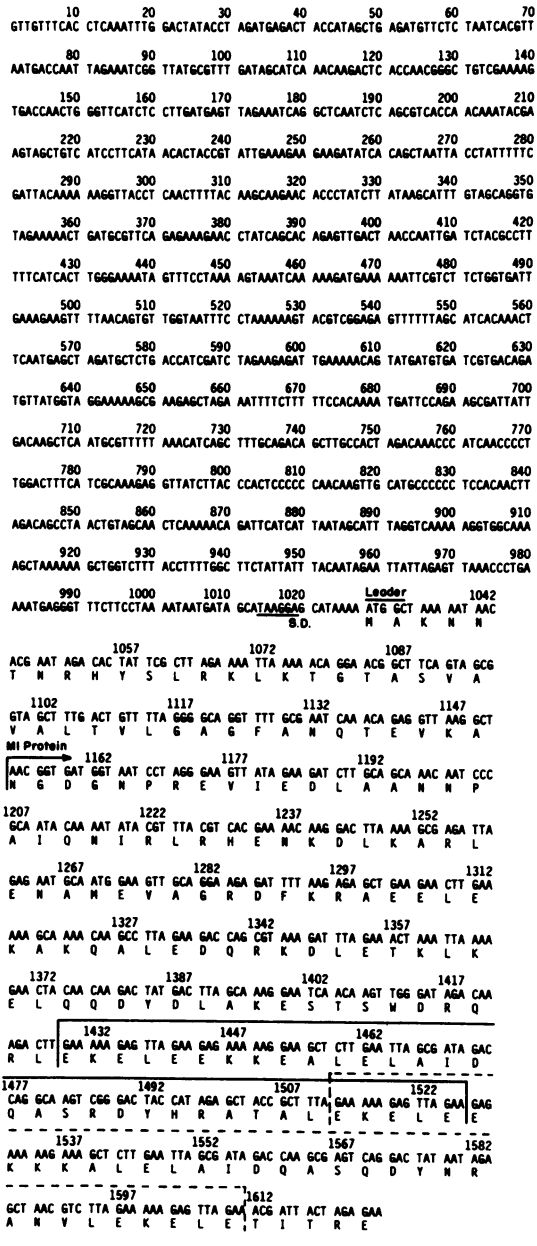

Fig. 2

Nucleotide sequence of pbf82 insert and deduced amino acid sequence of the amino-terminal portion of the M1 protein. The putative Shine Dalgarno sequence (S.D.) is indicated. A repeated region within the M1 gene is also shown.

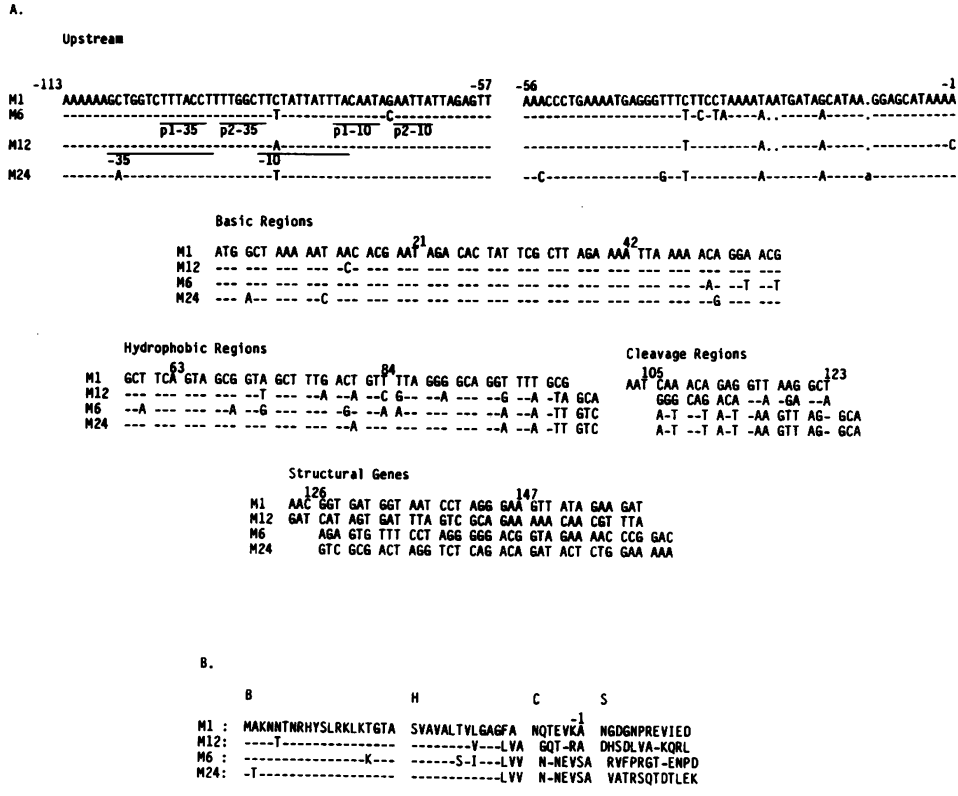

Fig. 3A
Comparison of the upstream and 5' nucleotide sequences of the M1, M12 (11), M6 (10), and M24 (12) genes. The published -35 and -10 promoter regions of the M12 and M6 genes are indicated. Nucleotides in the M12, M6 and M24 genes that are identical to those in the M1 gene are indicated by dashes, and gaps are indicated by dots. The leader sequences are subdivided according to the structural domains of the resulting leader peptides.

Fig. 3B
Comparison of the NH_2-terminal deduced amino acid sequences of M types 1, 12 (11), 6 (10) and 24 (12) precurser M proteins. Dashes in the latter three sequences indicate residues identical to those in the M1 amino acid sequence. The sequences are subdivided similarly to those in Fig. 3A. B: Basic regions, H: Hydrophobic regions, C: Cleavage regions, S: Mature proteins.

and -10 regions were actually utilized. A putative Shine-Dalgarno sequence, shown in Fig. 2 is identical to those described for the M12 and M6 genes. A 99bp tandemly repeated region was observed toward the 3' end of the open reading frame (Fig. 2). The repeats were flanked on each side by direct, identical, 15bp repeats which overlapped between the repeated regions. The open reading frame had no other significant repetitive regions, such as are present in the M6 gene (10). The deduced amino-terminal amino acid sequence of the M1 protein is also shown in Fig. 3. The sequence contained a putative 41

Type 1 M Protein Leader Sequence

```
        10         20         30         40         50         60         70         80
ATGGCTAAAA ATAACACGAA TAGACACTAT TCGCTTAGAA AATTAAAAAC AGGAACGGCT TCAGTAGCGG TAGCTTTGAC

        90        100        110        120
TGTTTTAGGG GCAGGTTTTG CGAATCAAAC AGAGGTTAAG GCT
```

Streptococcal Exotoxin Type A Leader Sequence (32)

```
        10         20         30         40         50         60         70         80
ATGGAAAACA ATAAAAAAGT ATTGAAGAAA ATGGTATTTT TTGTTTTAGT GACATTTCTT GGACTAACAA TCTCGCAAGA

        90
GGTATTTGCT
```

Streptolysin O Leader Sequence (33)

```
        10         20         30         40         50         60         70         80
ATGTCTAATA AAAAAACATT TAAAAAATAC AGTCGCGTCG CTGGGCTACT GACGGCAGCT CTTATCATTG GTAACCTTGT

        90
TACTGCTAAT GCTGAATCG
```

Streptokinase Leader Sequence (34)

```
        10         20         30         40         50         60         70
ATGAAAAATT ACTTATCTTT TGGGATGTTT GCACTGCTGT TTGCACTAAC ATTTGGAACA GTCAATTCTG TCCAAGCT
```

Protein G Leader Sequence (35)

```
        10         20         30         40         50         60         70         80
ATGGAAAAAG AAAAAAAGGT AAAATACTTT TTACGTAAAT CAGCTTTTGG GTTAGCATCC GTATCAGCTG CATTTTTAGT

        90
GGGATCAACG GTATTCGCT
```

Fig. 4
Comparison of the M1 leader sequence to other published streptococcal leader sequences. The references are given in parentheses.

amino acid leader peptide preceding the mature M protein sequence. A 33 residue tandem repeat that overlapped by 5 residues was identified toward the carboxyl end of the sequence. The deduced amino acid sequence corresponded to those determined by amino acid sequencing of three Type 1 M protein peptides extracted from whole streptococci with pepsin (pepM1) (25) and to the sequence reported by Moravek, et al. (27). To further confirm that the M1 protein predicted from the DNA sequence was correct, a peptide containing the 23 amino-terminal residues of the predicted M1 protein was synthesized. When covalently linked to KLH, the peptide evoked high titers of antibodies both against the synthetic peptide and against pepM1, and the immune sera type specifically enhanced the phagocytosis of type 1 streptococci rotated with fresh human blood (25).

Identification of the Leader Sequences in the M1 Gene and Three Other M Protein Genes.

The DNA sequences and deduced amino acid sequences of the extreme 5' portions of the M1, M6, M12 and M24 genes, and flanking upstream sequences, are compared in Figs. 3A and 3B. Each sequence had at its beginning a region encoding a putative signal peptide. The deduced signal peptides adhered

strongly to the characteristics ascribed to bacterial signal peptides (28,29). Each had a positively charged hydrophilic amino-terminus, a hydrophobic core, and a more polar cleavage region. The boundaries of these regions, as shown in Fig. 3, were assigned by previously described conventions (30). The putative signal peptidase cleavage points of the M1, M12, and M24 peptides, as shown in Fig. 3B, were assigned by the "-1,-3 rule" (31), and the cleavage point of the M6 peptide was previously published (10). The cleavage point in the M1 protein is further substantiated by the amino terminal sequence analysis of the mature M protein fragments extracted from streptococcal cells with pepsin. The sequence of the largest extracted peptide begins immediately following the predicted cleavage site (25).

Sequence similarities upstream of the 4 M protein genes.

All available sequences flanking the 5' ends of the M6, M12, and M24 genes were compared to the sequence upstream of the M1 gene, by the Lawrence and Goldman algorithm. All the compared sequences had large domains of homology, with standard deviation (S.D.) scores of >30 (data not shown). An S.D. score of >6 indicates a high probability that the sequence similarity did not occur by chance. Indeed, the entire 1027bp of sequence preceding the M1 gene is almost identical to the corresponding sequence preceding the M12 gene. The four upstream sequences were also aligned using the SS2 algorithm with a gap penalty of 2.5 and an incremental penalty for bases in gaps of 0.5 (24). The results of the SS2 alignment in the region immediately preceding the M protein genes is shown in Fig. 3.A.

Similarity of the Leader Sequences of the Four M Protein Genes

Except for a single base substitution at position 14 in the M12 sequence, and two substitutions at positions 4 and 12 in the M24 sequence, the four M protein leader sequences were entirely identical up to base 49. A larger number of base substitutions were evident in the four leader sequences from base 50 to base 102. From base 103 to the ends of the respective leader sequences, the M1, M12, and M6 sequences were highly dissimilar and remained so well into the respective structural genes (Fig. 3A and data not shown). Interestingly, even though the M6 and M24 signal sequences differed from each other both in the 5' portions of the leader sequences and in the mature protein coding regions, they were identical from base 86 to the respective 3' ends of the leader sequences.

In order to ensure that the sequence similarity of the four M protein leader sequences was not just a common feature of leader sequences from streptococci, we compared the M1 leader sequence with the available DNA sequences encoding other streptococcal signal peptides (32-35) which are shown in Fig. 4. When compared with the M1 leader sequence by the Lawrence & Goldman algorithm, the four other streptococcal leader sequences had S.D. scores <2.5, indicating that no statistically significant homology domains existed. In addition, all the bacterial and bacteriophage DNA sequences in the Genbank database were searched for possible similarity to the M1 leader sequence using a gap penalty of 4, a word length of 4 and a window size of 20 (23). With the exception of the M6 leader sequence present in this database, none of the sequences had similarity scores indicating non-random similarity.

Similarities Between the Four Deduced M Protein Signal Peptides.

The deduced amino acid sequences of the four signal peptides were similar to a greater extent than the DNA sequences. Only 5 of the first 32 amino acid positions contained substitutions, even though 15 of the first 32 codons contained base substitutions. Among the 5 positions with amino acid substitutions, the Thr/Ser substitution at position 27 and the Leu/Val/Ile substitutions at position 29 were relatively conservative. In light of the importance of basic residues in the function of signal peptides (28,29), it is interesting

to note the additional lysine residue at position 17 in the M6 peptide. From amino acid 35 of the signal peptides to within the amino-termini of the respective mature M proteins, the M1, M12, and M6 sequences were highly variable. The M6 and M24 signal peptide sequences were identical from residue 28 to the respective cleavage points, but their mature amino-terminal regions were also highly variable. Even though the cleavage regions of the four signal peptides were non-homologous, the four peptides had identical cleavage consensus sequences, with the -1 residue being alanine and the -3 residue being valine in each case. The amino acids in the -2 position, however, were not conserved.

DISCUSSION

In this study we compared the upstream and leader sequences of type M1 protein to those of three other M proteins. Although we found that the 5' end of the mature M1 coding region had no similarity with the 5' ends of the mature M12, M24 or M6 coding regions, the four genes were preceded by highly similar upstream and leader sequences. As shown in Fig. 3, the basic regions of the leader sequences were highly conserved at the DNA level, and the hydrophobic regions, although more variant at the DNA level, were highly conserved at the amino acid level.

The observed correlation between levels of sequence homology and the structural domains of the signal peptides may be important. The respective functions of the basic amino-terminus and the hydrophobic core regions might have provided selective pressure for particular nucleotide or amino acid sequences. For example, it is known that the amino-terminal region of signal peptides is important in the coupling of translation and transport of exported proteins, and mRNA structure is suggested to have a role in this coupling (36,37). Hence, conservation of nucleotide sequence might be important in the N-terminal region. In this respect, detailed analyses of the various M protein signal peptides might provide valuable information on how the fine structure of signal peptides affects their function in protein transport.

Moreover, the upstream and leader sequence nucleotide similarities observed between the four M protein genes may have a role in the generation of M protein antigenic diversity. These 5' regions of similarity, together with the conserved 3' ends of M protein genes (7), flank the highly variable and non-repetitive 5' ends of the M protein structural genes. These conserved flanking regions might be involved in rare occurrences of homologous recombination with regions outside of the M protein genes, resulting in the insertion of new variable regions. A recombination system of this sort is involved in pilin antigenic variation in Neisseria gonorrhoeae (38,39). The gonococcal chromosome contains silent pilin loci, containing multiple variable regions flanked by short stretches of conserved DNA. These variable regions are recombined into the pilin expression loci, resulting in variant antigens.

Silent M protein gene loci have not yet been defined in the streptococcal chromosome. Kehoe, et al, however, do find multiple regions within the genome of an M5 strain that hybridize to M5 gene probes (40). Scott, et al, on the other hand, claim that besides the resident M protein genes, no additional regions similar to M6 gene probes were found in the various streptococcal genomes they tested by DNA hybridization (7). Regions of DNA similarity required for a possible exchange of DNA might not be extensive enough in some strains to be detected by conventional nick translated DNA hybridization probes. The use of specific oligonucleotide probes of the sequences immediately flanking the variable regions of M protein genes might be useful in detecting possible conserved sequences elsewhere in the group A streptococcal chromosome. In addition, new variable regions could be acquired from exogenous

sources through transduction or transposition. Clearly, then, even if extragenic regions of similarity are not found in the chromosomes of group A streptococci, acquisition of new variable regions from outside of the actual M protein genes by homologous recombination cannot be ruled out as a possible mechanism for the generation of antigenic diversity in the non-repetitive N-terminal regions of streptococcal M proteins.

ACKNOWLEDGEMENTS

We thank Edna Chiang and Andrea Mouw for expert technical assistance, Mark Garrett for assistance in rabbit immunization and bleeding, and Gregory Bohach for a critical review of this manuscript. DNA analyses were carried out with the support of the University of Minnesota College of Biological Sciences Molecular Biology Computer Facility.

This work was supported by Public Health Service grants AI-16722 (P.C.), AI-15619 (V.B.), and AI-10085 (E.H.B.) from the National Institute of Allergy and Infectious Diseases.

*To whom correspondence should be addressed

REFERENCES

1. Jacks-Weis, J., Kim, Y., and Cleary, P.P. (1982) J. Immunol. **128**, 1897-1902.
2. Whitnack, E., and Beachey, E.H. (1985) J. Exp. Med. **162**, 1983-1997.
3. Lancefield, R.C. (1959) J. Exp. Med. **110**, 271-292.
4. Lancefield, R.C. (1962) J. Immunol. **89**, 307-313.
5. Phillips, G.N., Flicker, P.F., Cohen, C., Manjula, B.N., and Fischetti, V.A. (1981) Proc. Natl. Acad. Sci. USA **78**, 4889-4693.
6. Jones, K.H., Manjula, B.N., Johnston, K.H., Hollingshead, S.K., Scott, J.R. and Fischetti, V.A. (1985) J. Exp. Med. **161**, 623-628.
7. Scott, J.R., Hollingshead, S.K., and Fischetti, V.A. (1986) Infect. Immun. **52**, 609-612.
8. Hollingshead, S.K., Fischetti, V.A., and Scott, J.R. (1987) Infect. Immun. **55**, 3237-3239.
9. Hollingshead, S.K., Fischetti, V.A., and Scott, J.R. (1987) Mol. Gen. Genet. **207**, 196-203.
10. Hollingshead, S.K., Fischetti, V.A. and Scott, J.R. (1986) J. Biol. Chem. **261**,1677-1686.
11. Robbins, J., Spanier, J., Jones, S., Simpson, W. and Cleary, P., (1987) J. Bact. **169**, 5633-5640.
12. Mouw, A., Beachey, E. and Burdette, V., (1987) J. Bact. **170**, 676-684.
13. Frischauf, A.M., Lehrach, H., Poustka, A., and Murray, N., (1983) J. Mol. Biol. **170**, 827-842.
14. Hohn, B. (1979) Methods in Enzymol. **68**, 299-309.
15. Vieira, J., and Messing, J. (1982) Gene. **19**, 259-268.
16. Yanisch-Perron, C., Vieira, J., and Messing J., (1985) Gene. **33**, 103-119.
17. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. Skjold, S., Quie, P., Fries, L., Barnham, M., and Cleary, P., (1987) J. Infect. Dis. **155**, 1145-1150.
19. Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985) Plasmid **13**, 31-40.
20. Benton, W.D., and Davis, R.W. (1977) Science **196**, 180-182.
21. Spanier, J.G., Jones, S.J.C., and Cleary, P.P. (1984) Science **225**, 935-938.
22. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.

23. Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA **80**, 726-730.
24. Altschul, S.F., and Erickson, B.W. (1986) Bull. of Math. Biol. **48**, 603-616.
25. Krause, W., Haanes-Fritz, E., Cleary, P., Seyer, J., Dale, J., and Beachey, E. (1987) J. Immunol. **139**, 3084-3090.
26. Haanes-Fritz, E., Robbins, J. and Cleary, P. (1987) In Ferretti, J., and Curtiss III, R. (eds), Streptococcal Genetics, ASM, Washington D.C., pp.109-111.
27. Moravek, L., Kuhnemund, O., Havlicek, J., Kopecky, P., and Pavlik, M. (1986) FEBS Letter **208**, 435-438.
28. Silhavy, T.J., Benson, S.A. and Emr, S.D. (1983) Microbiol. Rev. **47**, 313-344.
29. Oliver, D. (1985) Ann. Rev. Microbiol. **39**, 615-648.
30. Von Heijne, G. (1985) J. Mol. Biol. **184**, 99-105.
31. Von Heijne, G. (1983) Eur. J. Biochem. **133**, 17-21.
32. Weeks, C.R., and Ferretti, J.J. (1986) Infect. Immun. **52**, 144-150.
33. Kehoe, M.A., Miller, L., Walker, J.A., and Boulnois, G.J. (1987) Infect. Immun. **55**, 3238-3232.
34. Malke, H., Roe, B., and Ferretti, J.J. (1985) Gene. **34**, 357-362.
35. Fahnestock, S.R., Alexander, P., Nalge, J., and Filpula, D., (1986) J. Bacteriol. **167**, 870-880.
36. Hall, M.N., Gabay, J., Debarbouille, M., and Schwartz, M. (1982) Nature **295**, 616-618.
37. Schwartz, M., Roa, M., and Debarbouille, M. (1981) Proc. Natl. Acad. Sci. USA **78**, 2937-2941.
38. Hagblom, P., Segal, E., Billyard, E., and So, M. (1985) Nature **315**, 156-158.
39. Segal, E., Hagblom, P., Seifert, H.S., and So, M. (1986) Proc. Natl. Acad. Sci. USA **83**,2177-2181.
40. Kehoe, M.A., Poirier, T.P., Beachey, E.H., and Timmis, K.N. (1985) Infect. Immun. **48**, 190-197.