Identification of a Family of Streptococcal Surface Proteins with Extremely Repetitive Structure*

(Received for publication, April 30, 1996, and in revised form, May 17, 1996)

Maria Wästfelt, Margaretha Stålhammar-Carlemalm, Anne-Marie Delisse, Teresa Cabezon, and Gunnar Lindahl†

From the †Department of Medical Microbiology, Lund University, Sölvegatan 23, S-223 62 Lund, Sweden and §SmithKline Beecham Biologicals, B-1330 Rixensart, Belgium

The group B Streptococcus (GBS) causes the majority of life-threatening bacterial infections in newborn children. Most GBS strains isolated from such infections express a surface protein, designated Rib, that confers protective immunity and therefore is of interest for analysis of pathogenetic mechanisms. Sequence analysis demonstrated that Rib has an exceptionally long signal peptide (55 amino acid residues) and 12 repeats (79 amino acid residues each) that account for >80% of the sequence of the mature protein. The repeats are identical even at the DNA level, indicating that an efficient mechanism operates to maintain a highly repetitive structure in Rib. The structure of Rib is similar to that of α, a previously characterized surface protein that is common among GBS strains lacking Rib. However, highly purified preparations of Rib and α did not cross-react immunologically, although the two proteins show extensive amino acid residue identity (47% in the repeat region). When analyzed in Western blots, Rib and α give rise to a regularly spaced ladder pattern, apparently due to hydrolysis of acid-labile Asp-Pro bonds in the repeats. We conclude that Rib and α are members of a novel family of streptococcal surface proteins with unusual repetitive structure.

The group B Streptococcus (GBS) is the major cause of life-threatening bacterial infections in the neonatal period. Many children are exposed to this bacterium at birth, when they may be colonized by GBS present in the vaginal flora of the mother. In most cases such colonization does not cause disease, but a minority of newborns fall seriously ill after birth due to invasive GBS infection. Other children are born ill, due to an infection starting during the later part of the pregnancy (1).

Immunity to GBS infection may be elicited both by the polysaccharide capsule (1, 2) and by different cell surface proteins (3–6). Detailed studies of the capsule have shown that it occurs in several different serotypes that vary in their importance for human disease (1, 2). Among the surface proteins that confer immunity, the first to be identified were two molecules designated α and β, which have been extensively characterized (9–14). However, α and β are almost never expressed by strains of capsular type III, which cause the majority of invasive infections (1). In contrast, the recently identified protein Rib was found to be expressed by most strains causing life-threatening infections, including almost all strains of serotype III. Thus, protein Rib is of considerable interest for the analysis of pathogenetic mechanisms in GBS infections and for vaccine development. This situation motivates detailed biochemical and immunological characterization of protein Rib.

Studies of protein Rib previously showed that it shares several properties with the α protein (6). Both of these proteins are resistant to trypsin and vary greatly in size between different isolates of GBS. Moreover, the NH2-terminal sequences of Rib and α were found to be related. These and other data suggested that the two proteins might be members of a family of proteins with related function. We have now sequenced the rib gene and compared it with the previously reported sequence of the α gene, which is known to be very repetitive (14). In addition, highly purified preparations of the Rib and α proteins have been characterized with regard to some immunological and biochemical properties. Our data show that Rib and α define a family of bacterial surface proteins with unique repetitive structure.

**Experimental Procedures**

Bacterial Strains and Cloning Vectors—The GBS strain BM110 (6) is a serotype III isolate obtained from Dr S. Mattingly (University of Texas, San Antonio, TX). Escherichia coli strain LE392 (Genofit, Geneva, Switzerland) was used as a host for the cloning vector pEMBL3 (Promega Co., Madison, WI). For subcloning, E. coli strain XL1-Blue (which is recA1) (Stratagene, La Jolla, CA) was used as a host for the cloning vector pGEM7Zf+ (Promega Co.), and the E. coli strain M103 (Amerham Corp.) was used as a host for the sequencing vectors M13mp18 or M13mp19 (Amerham Corp.). Standard techniques were used for work with E. coli and cloning vectors (15).

Media, Chemicals, and Purified Proteins—GBS was grown in Todd-Hewitt broth, and E. coli was grown in LB broth at 37 °C. Ampicillin (50 μg/ml) and tetracycline (5 μg/ml) were added when appropriate. Restriction enzymes were purchased from Promega Co., New England Biolabs Inc. (Beverly, MA) or Boehringer Mannheim.

The Rib, α, and β proteins were purified from extracts of strains BM110, A909, and SB35, respectively, by a combination of ion exchange and molecular sieve chromatography (6), followed by a final step of hydroxyapatite chromatography for removal of small amounts of contaminating polysaccharides.3

DNA Sequencing and Sequence Analysis—DNA sequences were determined by the dideoxy chain termination method using [α-35S]dATP (Amersham Corp.) and Sequenase 2.0 (Amersham Corp.). Recombinant M13mp18 or M13mp19 phage DNA was used as template. M13 universal primer and –40 primer (Amersham Corp.) as well as custom made primers were used. The sequencing reaction products were resolved on

---

1 The abbreviations used are: GBS, group B Streptococcus; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; bp, base pairs; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

2 Larsson, C., Stålhammar-Carlemalm, M., and Lindahl, G., Infect. Immun., in press.
Figu...
made to clone the entire rib gene in E. coli. Because it seemed possible that Rib had a toxic effect on E. coli, the rib gene was cloned without the promoter and signal sequence regions. Appropriate fragments of chromosomal DNA from strain BM110 were cloned directly into the pGEM7Z(+) vector, generating clone pGRib116. Initial analysis of this clone showed that it contained a repeat region of the same size as the chromosomal rib gene. However, further analysis of pGRib116 indicated that the repeat region in this clone was highly unstable, although it was maintained under Rec+ conditions and not expressed (data not shown). Because the entire repeat region of the rib gene could not be stably maintained in E. coli, it was not possible to analyze the sequence of this region with standard methods.

To analyze the sequence of the repeat region, we chose to sequence individual repeats cloned at random. As described above, our analysis of the rib gene had indicated that all repeats contained a unique BglII site. We therefore cloned fragments obtained by BglII digestion of plasmid pGRib116, assuming that they would be representative of the whole repeat region. A total of 13 repeats were analyzed, and all of them were found to have identical nucleotide sequences. The conclusion that all repeats are identical was further supported by analysis of sequences at the extremities of the repeat region. The 5' half of the first repeat (up to the BglII site) and the 3' half of the last complete repeat (downstream from the BglII site) together formed a repeat whose nucleotide sequence was identical to that of repeats recovered after BglII digestion. In addition, the partial repeat (coding for 15 amino acid residues) had a nucleotide sequence identical to the corresponding region in the complete repeats.

Comparison between the Rib and α Proteins—Previous studies have shown that the α protein of GBS has a very repetitive structure with long repeats that are identical even at the DNA level (14), as reported here for protein Rib. As shown in Fig. 3, these two surface molecules of GBS exhibit extensive amino acid residue identity. The signal sequences show 80% residue identity and are unusually long: 55 residues in protein Rib (Fig. 1) and 56 residues in the α protein (6). In the non-repeated

![PCR analysis of the rib gene.](image)

**Fig. 2.** PCR analysis of the rib gene. PCR products were generated, as described under “Experimental Procedures,” from streptococcal strain BM110 DNA and from the plasmid clone pGRib105 using fivefold dilutions of the templates. Sizes (in bp) of the main PCR products are indicated. The PCR product of 3,400 bp corresponds to a rib gene with 12 complete repeats, and the PCR product of 2,700 bp corresponds to a rib gene with 9 complete repeats.

![Comparison of the Rib and α proteins.](image)

**Fig. 3.** Comparison of the Rib and α proteins. A, alignment (16) of the amino acid sequences of Rib from strain BM110 and α from strain A909 (14). The two vertical arrows indicate the ends of the signal sequences (6). The repeat regions are shown in the shaded box. Only one full repeat from each protein is shown, followed by the partial repeat. B, overall structure of Rib from strain BM110 and α from strain A909 and degree of amino acid residue identity between different regions of the proteins. S, signal peptide; N, NH2-terminal region; R, one repeat; P, partial repeat; C, COOH-terminal region. The number of amino acids in each region is indicated. The Rib protein has 12 repeats of 79 amino acids, and the α protein has 9 repeats of 82 amino acids.
NH₂-terminal parts of the mature proteins (174 and 170 residues, respectively), the degree of residue identity is 61%. The repeats (79 and 82 residues, respectively) show a somewhat lower degree of residue identity, 47%. The short COOH-terminal regions of the two proteins are almost identical and have the characteristics of cell wall attachment regions in surface proteins of Gram-positive bacteria, including an LPXTG sequence (19).

The Rib and α proteins have an unusually high content of Asp, Val, Thr, Pro, and Lys, which together account for about 60% of the amino acid residues in each protein. Computer-assisted analysis indicated that the Rib and α proteins are highly acidic, with isoelectric points of 4.3 and 4.5, respectively. Analysis of the protein sequences by protein structure algorithms (Ref. 16 and the GeneWorks program) predicted a high β-sheet content in each protein, including the repeat regions.

Immunological Relationship between the Rib and α Proteins—The Rib and α proteins were previously found to be immunologically unrelated, when analyzed with specific rabbit antisera in Western blots and Dot-blot (6). However, the extensive sequence homology between the two proteins suggested that a cross-reactivity might be detected if more sensitive tests were performed (Fig. 4). The reactivity between Rib, immobilized in microtiter plates, and anti-Rib serum was inhibited by pure protein Rib, but the addition of the α protein did not cause any inhibition even when a large excess was added (Fig. 4A). Similarly, the reaction between α and anti-α serum was inhibited by purified α protein but not by protein Rib (Fig. 4B). These results indicate that the large majority of antibodies directed against Rib or α completely lack reactivity for the heterologous antigen.

Aberrant Migration Behavior of the Rib and α Proteins in SDS-PAGE—An unusual feature of Rib and α is their behavior in SDS-PAGE gels, where the apparent molecular mass of each protein was found to vary depending on the acrylamide concentration of the gel (Fig. 5A). At an acrylamide concentration of 5% the major polypeptide species in the Rib and α protein preparations migrated at positions corresponding to molecular masses of about 178 and 166 kDa, respectively (Fig. 5B), but at an acrylamide concentration of 10% the apparent molecular masses were approximately 107 and 111 kDa, respectively (Fig. 5C). According to the deduced amino acid sequences, the predicted molecular masses of the mature Rib and α proteins are 123 and 103 kDa, respectively. Unlike Rib and α, the group B streptococcal β protein, an IgA-binding surface protein that is structurally unrelated to the Rib and α proteins and lacks long repeats (11, 12), had the same apparent molecular mass in the different SDS-PAGE gels (Fig. 5).

Analysis of Ladder Patterns Generated by the Rib and α Proteins in SDS-PAGE: Evidence for Hydrolysis of Acid-labile Asp-Pro Bonds—It has previously been reported that bacterial extracts containing the α protein give rise to a regular ladder pattern in immunoblotting experiments, indicating that the α protein is size heterogeneous (20). Interestingly, the distance between the ladder steps was found to correspond to one repeat, suggesting that the different molecular species in the ladder represented polypeptides with different number of repeats (14). A similar ladder pattern was also observed in Western blots of the Rib protein (6). It was suggested that this size heterogeneity could be the result of early termination of translation, RNA-mediated self cleavage, acid hydrolysis, or protease activity (14). A repetitive protein from the salivary glands of Chironomus tentans has also been shown to form a regular ladder pattern in Western blots, and it was suggested that the heterogeneity reflects a degradation that occurs naturally in the salivary glands (21). It was therefore of interest to analyze the mechanism that generates such ladder patterns.

Analysis of the sequences of the Rib and α proteins suggested that the ladder pattern might be due to hydrolysis of Asp-Pro bonds, which are found in the repeats of both proteins (Fig. 6D). It is known that such bonds are sensitive to acid hydrolysis (22). To analyze whether acid-labile sites are responsible for the ladder pattern, purified preparations of the Rib and α proteins were first analyzed under standard conditions (Fig. 6A). Under these conditions, the ladder pattern was seen in gels but not in stained gels, indicating that only a small fraction of the purified proteins were of lower molecular weight and gave rise to the ladder (Fig. 6A). Next, the purified Rib and α proteins were incubated at pH 4.0 at 37°C for 16 h before analysis. The resulting preparations were either boiled directly in sample buffer or neutralized before boiling in sample buffer. When these preparations were analyzed by SDS-PAGE, the analysis showed that distinct ladder patterns, readily detectable also in stained gels, were formed when the proteins have been boiled for 5 min in sample buffer at acidic pH (Fig. 6B). However, only a minor degradation was detected in the samples that had been neutralized before the analysis (data not
fragments as well as internal peptides generated by hydrolysis of Asp-Pro sites in the repeats (7.2- and 1-kDa peptides from the repeats of protein Rib and an 8.7-kDa peptide from the repeats of the α protein). Surprisingly, neither COOH-terminal fragments nor internal peptides were found, indicating that these peptides had been further degraded or lost during the analysis (Fig. 6C). Interestingly, the ladder pattern formed by the salivary gland protein from C. tentans also showed the absence of internal peptides corresponding to single repeats (21).

**DISCUSSION**

In this report we describe the nucleotide sequence of the streptococcal surface protein Rib and show that Rib together with the α protein of GBS define a novel family of bacterial surface proteins. The most remarkable feature of Rib and α is the presence of an extensive repeat region, in which the repeats are identical even at the DNA level, as first reported for the α protein (14). The presence of this repeat region can explain why the Rib and α proteins expressed by different clinical isolates of GBS show size variation (6, 20), which could arise through recombination or slipped strand mispairing during DNA replication. If size variation occurred frequently, it might also explain why the repeats are identical, because frequent expansions and contractions of the repeat region could eliminate divergent repeats that arise through mutation. Such frequent variation in size has been reported for a gene with short identical repeats in Ureaplasma urealyticum (23). However, protein Rib expressed by a given clinical isolate of GBS shows great stability in size, at least under laboratory conditions (6). This finding indicates that the size variation between Rib molecules expressed by different strains of GBS is due to rare events and makes it difficult to explain the identity between the repeats through a model with frequent expansions and contractions in the repeat region. In addition, expansions and contractions would be expected to generate divergent repeats at the extremities of the repeat region, as described for the M6 protein of Streptococcus pyogenes (24). However, all repeats within the rib gene, and also within the α gene, are apparently identical. A possible alternative explanation for the identity of the repeats could be the existence of base pairing within RNA secondary structures, which might limit mutation frequencies (25). However, it is clear that mutations can accumulate in the repeat region, because the repeats of Rib and α must have evolved from a common predecessor (Fig. 3). Taken together, these data suggest that any mutation that occurs in a repeat is either rapidly eliminated or rapidly spread to the other repeats, by unknown mechanisms.

With regard to long repeats that are identical also at the nucleotide level, it should be noted that such repeats have also been described for genes encoding human apolipoprotein(a) (26) and the salivary protein from C. tentans referred to above (21). In addition, several genes of the mucin family have highly conserved repeat regions (27), and the gene for porcine submaxillary gland apomucin contains long repeats that are identical also at the nucleotide level (28). For the Rib and α proteins, it seems possible that the repeat regions confer unusual physico-chemical properties on the proteins and contribute to their protease resistance (6, 13) and aberrant migration behavior in SDS-PAGE (Fig. 5).

The Rib and α proteins form a regular ladder pattern in Western blots, where the distance between the steps in the ladder was found to correspond to the size of one repeat (6, 14, 20). A similar ladder pattern was reported for the repetitive salivary glycoprotein from C. tentans (21). It was assumed that this size heterogeneity represents a true heterogeneity in the proteins, and various mechanisms have been proposed that could generate the ladder pattern (14, 21). However, the anal-

**Figure 6. Analysis of ladder patterns formed by the Rib and α proteins in SDS-PAGE.**

A, Western blot analysis of purified preparations of the Rib, α, and β proteins under standard conditions using specific rabbit antisera. Molecular mass markers are in kDa. B, proteins adjusted to pH 4.0 and then boiled with sample buffer for 5 min. Stained gel, 10% acrylamide. C, proteins adjusted to pH 4.0 and then boiled with sample buffer for 15 min. Stained Tricine gel, 16.5% acrylamide. In gels B and C, molecular mass markers (in kDa) are included in the figure. D, overall structure of the mature Rib and α proteins. Amino-terminal sequences and putative acid-sensitive Asp-Pro (DP) sites are indicated. The bars denoted a–d show possible structures for the fragments indicated in B and C. N, NH2-terminal non-repeated region; R, one repeat.

Thus, the ladder patterns were largely due to fragmentation during boiling in non-neutralized sample buffer (Fig. 6B). The Rib and α proteins were further degraded when the samples were boiled at acidic pH for a longer period (15 min), as detected in a stained Tricine gel (Fig. 6C). In contrast, the group B streptococcal β protein, which does not contain Asp-Pro sequences (11, 12), was not degraded at acidic pH (Figs. 6, B and C). The repeats in the Rib protein contain two Asp-Pro sites (Fig. 6D), which may explain why this protein gives rise to doublet bands (Fig. 6B).

To further analyze the formation of the ladder, bands generated by the Rib and α proteins at acidic pH were subjected to NH2-terminal sequence analysis. Bands analyzed included those labeled a–d in Fig. 6, B and C, as well as polypeptides of higher molecular weight. All bands analyzed had sequences identical to the NH2-terminal sequences of the mature proteins, i.e. AEVIS for the Rib protein and ST1PG for the α protein (Fig. 6D). These data may be explained by assuming that acid hydrolysis occurred at all Asp-Pro sites in the Rib and α proteins, except the most NH2-terminally located site in each protein. Cleavage at this site would have given rise to a short NH2-terminal fragment, which was not detected.

Although the data reported above suggest that the ladder pattern observed for the Rib and α proteins is generated by cleavage of Asp-Pro bonds, cleavage of such bonds would be expected to generate both NH2-terminal and COOH-terminal sequences. However, all bands analyzed had sequences containing only an NH2-terminal fragment, which was not detected. Therefore, further studies are needed to determine the exact mechanism of ladder formation.
ysis of the Rib and α proteins reported here suggests that the apparent size heterogeneity may at least partially be due to hydrolysis, during the analysis, of acid-labile Asp-Pro bonds in the repeats. In agreement with this explanation, Asp-Pro bonds are present not only in the repeats of the Rib and α proteins but also in the C. tentans protein (21). Thus, it remains an open question whether these proteins really are size heterogeneous in vivo. However, it should be noted that the normal habitat of GBS is the human vagina, where the normal pH is less than 4.5 (29), a condition that could favor some hydrolysis of Asp-Pro bonds also in vivo and cause release of biologically active polypeptides from the bacterial cell.

The Rib and α proteins show 61% residue identity in the unique NH2-terminal regions and 47% identity in the repeat regions (Fig. 3). Within these regions, shorter sequences with even higher degree of homology are also present, underlining the close evolutionary relationship between Rib and α. It was therefore surprising that experiments reported previously had indicated that the two proteins are immunologically unrelated (6). However, the inhibition experiments reported here confirmed the results previously obtained (Fig. 4). This result suggests that the sequences that are similar in Rib and α are not immunogenic or represent epitopes that are hidden in the intact proteins.

In summary, the Rib and α proteins of the group B streptococcus have been shown to be members of a family of bacterial surface proteins with remarkable repetitive structure. Further characterization of this protein family is of interest for studies of genes with highly repetitive sequence, which are common both in bacteria and in man. In addition, the knowledge that is now available about Rib and α will permit studies of their biological function and definition of protective epitopes. Such studies are of interest with regard to the mechanisms of pathogenesis used by Gram-positive bacteria and may also contribute to the development of a protein vaccine against GBS disease.

REFERENCES

1. Baker, C. J., and Edwards, M. S. (1995) in Infectious Diseases of the Fetus and Newborn Infant (Remington, J. S., and Klein, J. O., eds) pp. 980-1054, W. B. Saunders Co., Philadelphia

2. Wessels, M. R., Padgett, L. C., Kasper, D. L., DiFabio, J. L., Michon, F., Holme, K., and Jennings, H. J. (1990) J. Clin. Invest. 86, 1428-1433

3. Lancefield, R. C., McGarty, M., and Everly, W. N. (1975) J. Exp. Med. 142, 165-179

4. Bevanger, L., and Naess, A. I. (1985) Acta Pathol. Microbiol. Immunol. Scand. Sect. B 93, 121–124

5. Michel, J. L., Madoff, L. C., Kling, D. E., Kasper, D. L., and Ausubel, F. M. (1991) Infect. Immun. 59, 2023-2028

6. Ståhlin-Marxen, M., Stenberg, L., and Lindahl, G. (1993) J. Exp. Med. 177, 1593–1603

7. Wilkinson, H. W., and Eagon, R. G. (1971) Infect. Immun. 4, 596–604

8. Bevanger, L., and Maeland, J. A. (1979) Acta Pathol. Microbiol. Immunol. Scand. Sect. B 87, 51–54

9. Russell-Jones, G. J., Gotschlich, E. C., and Blake, M. S. (1984) J. Exp. Med. 160, 1467–1475

10. Lindahl, G., Åkerström, B., and Faerman, L. (1991) Eur. J. Immunol. 20, 2241-2247

11. Hedén, L. O., Frithz, E., and Lindahl, G. (1991) Eur. J. Immunol. 21, 1481-1490

12. Jerström, P. G., Chhatwal, G. S., and Timmis, K. N. (1991) Mol. Microbiol. 5, 843–849

13. Michel, J. L., Madoff, L. C., Kling, D. E., Kasper, D. L., and Ausubel, F. M. (1991) Genetics and Molecular Biology of Streptococci, Lactococci, and Enterococci (Dunny, G. M., Cleary, P. P., and McKay, L. L., eds) pp. 214–218, American Society for Microbiology, Washington, DC

14. Michel, J. L., Madoff, L. C., Olson, K., Kling, D. E., Kasper, D. L., and Ausubel, F. M. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 10060-10064

15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

16. Genetics Computer Group (1994) Program Manual for the GCG Package, Version 8, University of Wisconsin, Madison, WI

17. Schägger, H., and von Jagow, G. (1987) Anal. Biochem. 166, 368–379

18. Musser, J. M., Mattingly, S. J., Quentin, R., Goudeau, A., and Selander, R. K. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4731–4735

19. Schneewind, O., Mihaylova-Petkov, D., and Model, P. (1993) EMBO J. 12, 4013–4021

20. Madoff, L. C., Hori, S., Michel, J. L., Baker, C. J., and Kasper, D. L. (1991) Infect. Immun. 59, 2638–2644

21. Galli, J., and Wieslander, L. (1993) J. Clin. Invest. 92, 11888–11893

22. Landon, M. (1977) Methods Enzymol. 47, 145–149

23. Zheng, X., Teng, L. J., Watson, H. L., Glass, J. I., Blanchard, A., and Cassell, G. H. (1991) Infect. Immun. 63, 891–898

24. Hulinghead, S. K., Fischetti, V. A., and Scott, J. R. (1986) J. Biol. Chem. 261, 1677–1686

25. Mita, K., Ichimura, S., Zama, M., and James, T. C. (1988) Acta Pathol. Microbiol. Immunol. Scand. Sect. B 87, 51–54

26. McLean, J. W., Tomlinson, J. E., Kuang, W. J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M., and Lawn, R. M. (1987) Nature 330, 132–137

27. Strous, G. J., and Dekker, J. (1992) Crit. Rev. Biochem. Mol. Biol. 27, 57–92

28. Timpte, C. S., Eckhardt, A. E., Abernethy, J. L., and Hill, R. L. (1988) J. Biol. Chem. 263, 1081–1088

29. Paavonen, J. (1983) Scand. J. Infect. Dis. Suppl. 40, 31–35