Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 10060–10064, November 1992
Genetics

Large, identical, tandem repeating units in the C protein alpha antigen gene, *bca*, of group B streptococci

(molecular sequence data/*Streptococcus agalactiae*/bacterial antigens/amino acid sequence homology/repetitive sequences)

JAMES L. MICHEL*†, LAWRENCE C. MADOFF*‡, KRISTIN OLSON*, DAVID E. KLING*, DENNIS L. KASPER*‡, AND FREDERICK M. AUSUBEL§

*Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and ‡Division of Infectious Diseases, Beth Israel Hospital, Harvard Medical School, Boston, MA 02115; and §Department of Molecular Biology, Wellman-10, Massachusetts General Hospital, Fruit Street, Boston, MA 02114

Communicated by R. John Collier, July 6, 1992 (received for review May 2, 1992)

ABSTRACT **Group B *Streptococcus* (GBS) is the leading cause of neonatal sepsis and meningitis in the United States. The surface-associated C protein alpha antigen of GBS is thought to have a role in both virulence and immunity. We previously cloned the C protein alpha antigen structural gene (named *bca* for group B, C protein, alpha) into *Escherichia coli*. Western blots of both the native alpha antigen and the cloned gene product demonstrate a regularly laddered pattern of heterogeneous polypeptides. The nucleotide sequence of the *bca* locus reveals an open reading frame of 3060 nucleotides encoding a precursor protein of 108,705 Da. Cleavage of a putative signal sequence of 41 amino acids yields a mature protein of 104,106 Da. The 20,417-Da N-terminal region of the alpha antigen shows no homology to previously described protein sequences and is followed by a series of nine tandem repeating units that make up 74% of the mature protein. Each repeating unit is identical and consists of 82 amino acids with a molecular mass of 8665 Da, which is encoded by 246 nucleotides. The size of the repeating units corresponds to the observed size differences in the heterogeneous ladder of alpha C proteins expressed by GBS. The C-terminal region of the alpha antigen contains a membrane anchor domain motif that is shared by a number of Gram-positive surface proteins. The large region of identical repeating units in *bca* defines protective epitopes and may play a role in generating phenotypic and genotypic diversity of the alpha antigen.**

Streptococcus agalactiae [group B *Streptococcus* (GBS)] is an important pathogen in neonatal sepsis and meningitis, postpartum endometritis, and infections in adults, in particular in diabetics and immunocompromised hosts (1). The best-studied GBS virulence determinants are the type-specific capsular polysaccharides that are essential for pathogenesis (2, 3). The roles of GBS surface proteins in infection are less well understood (4, 5). The C proteins are surface-associated antigens expressed by most clinical isolates of capsular types Ia, Ib, and II and are thought to play a role in both virulence and immunity (6, 7). Two C protein antigens, alpha and beta, have been described biochemically and immunologically (5).

In 1975, Lancefield *et al.* (8) showed that antibodies raised to the C proteins in rabbits protected mice challenged with GBS bearing the C proteins. We described a monoclonal antibody to the alpha antigen (4G8) that induces opsonic killing of GBS and protects mice from lethal challenge with GBS (9). In addition, we cloned and expressed the genes encoding alpha and beta antigens in *Escherichia coli* and showed that antibodies raised to the clones of both alpha and beta encode different C proteins that define unique protective epitopes (10). The alpha and beta antigens are independently expressed and antigenically distinct proteins.

The C protein beta antigen that specifically binds to human serum IgA has been cloned (10, 11) and sequenced (12, 13). However, the role of the beta antigen and IgA binding in virulence is not known. Studies by Ferrieri and coworkers (14, 15) showed that C protein-bearing strains of GBS resist phagocytosis and inhibit intracellular killing. We found that opsonophagocytic killing in the presence of alpha antigen-specific monoclonal antibody (4G8) correlated directly with increasing molecular mass of the alpha antigen and with the quantity of alpha antigen expressed on the bacterial cell surface (7). GBS strains expressing the alpha antigen were resistant to killing by polymorphonuclear leukocytes in the absence of specific antibody; however, this resistance was not dependent on the size of the alpha antigen.

The completed nucleotide sequence of *bca* and flanking regions¶ reported here provides information regarding the size, structure, and composition of the alpha antigen gene. An interesting feature of both the native and the cloned gene products of the alpha antigen is that they exhibit protein heterogeneity by expressing a regularly repeating ladder of proteins differing by ≈8000 Da (9, 10). Since the protective monoclonal antibody 4G8 binds to the repeat region, this region defines a protective epitope (ref. 9; J.L.M. and L.C.M., unpublished data). Smaller tandemly repeated sequences encoding immunodominant epitopes have been reported in a number of pathogens but have not been associated with the protein heterogeneity seen in the alpha antigen (16–21). Though the maximum molecular size of the alpha antigen differs among strains of GBS, this protein heterogeneity is a constant feature (7).

The nucleotide sequence of *bca* contains nine identical 246-nucleotide tandem repeating units. The estimated size of the peptide encoded by each of these repeats is 8665 Da and correlates with the intervals found in the heterogeneous laddering of the alpha antigen. The amino acid sequence derived from the DNA sequence revealed both significant homologies and important differences between the alpha antigen and other streptococcal proteins (12, 13, 19). The repeating units of the alpha antigen suggest possible mechanisms for phenotypic and genotypic variability and provide natural sites for gene rearrangements that could generate antigenic diversity.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, Transposons, and Media. GBS strain A909 (type 1a/$C_{\alpha,\beta}$) (8); *E. coli* strains MC1061 and

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: GBS, group B *Streptococcus*.

†To whom reprint requests should be addressed at: Channing Laboratory, 180 Longwood Avenue, Boston, MA 02115.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M97256).

DK1 (22), pCNB (23), DH5α (a derivative of DH1; GIBCO/BRL), and NK-8032; *E. coli* plasmids and clones pUC12, pUX12, and pJMS23; and the transposon Tn5seq1 have been described (10). The plasmid pGEM-7Zf(−) was purchased from Promega. Additional subclones of pJMS23 (pJMS23-1, -7, -9, and -10) are described below. Growth media for GBS and *E. coli* and antibiotics for selection have been described (10).

DNA Procedures and Nucleotide Sequencing Strategy. Standard procedures for the preparation of plasmid DNA, synthesis and purification of oligonucleotides, restriction endonuclease mapping, agarose gel electrophoresis, and Southern blot hybridization are from Ausubel *et al.* (22). Restriction endonucleases and other enzymes for manipulation of DNA (e.g., DNase, RNase, and ligase) were obtained from New England Biolabs and Boehringer Mannheim. Transposon mutagenesis utilized λTn5seq1 (24).

Nucleotide sequencing of double-stranded DNA used plasmids containing transposon Tn5seq1 insertions using primers to Sp6 or T7 promoters for bidirectional sequencing, synthetic oligonucleotide primers, and nested deletions using Erase-a-Base (Promega). A total of 12 primers were prepared to obtain the sequence in both directions for the areas of the gene flanking the repeat region. Sequencing of the region of repetitive DNA was completed with exonuclease III-generated nested deletions. All sequencing employed Sequenase, version 2, used according to manufacturer's specifications for double-stranded sequencing (United States Biochemical). Adenosine 5′-[α-[^{35}S]thio]triphosphate was obtained from Amersham. GeneAmp PCR kit with AmpliTaq polymerase was used according to manufacturer's instructions (Perkin–Elmer/Cetus).

Subclones pJMS23-1, pJMS23-7, and pJMS23-10 were prepared for transposon mutagenesis to target smaller regions within *bca* (10). Subclone pJMS23-1 contains a 5.9-kilobase *Hin*dIII fragment on pUX12, pJMS23-7 contains 2.8-kilobase *Alu* I fragment from pJMS23-1 ligated into the *Hin*cII site in the polylinker of pUC12, and pJMS23-10 is a *Bsa*B1/*Sma* I double restriction endonuclease digestion of pJMS23-7 that yielded a 2.3-kilobase insert containing the repeat region. For nested deletions, the *Alu* I fragment from pJMS23-1 was ligated into the *Sma* I site on pGEM-7Zf(−) to create pJMS23-9. Nested deletions, were constructed in the forward direction from the *Hin*dIII and *Nsi* I sites and in the reverse direction from *Eco*RI and *Sph* I sites. The sizes of the subclones, mutants, and deletions used for sequencing were confirmed by restriction endonuclease mapping and/or PCR with primers to pUC12 polylinker and Tn5seq1 (Sp6 and T7).

Data analysis used the Department of Molecular Biology computer at Massachusetts General Hospital (Boston) with Genetics Computer Group (Madison, WI) version 7 software and the Basic Local Alignment Search Tool (BLAST) network of the National Center for Biotechnology Information of the National Institutes of Health (Bethesda, MD).

Monoclonal Antibodies, SDS/PAGE, and Western Immunoblots. Extraction of GBS and *E. coli* proteins, SDS/PAGE, immunoblotting, and probing with the alpha antigen monoclonal antibody 4G8 have been described (7, 9, 10).

RESULTS

Nucleotide Sequence of *bca*. Subclones of pJMS23, which encodes the *bca* locus from GBS strain A909 (type Ia/C) and expresses the alpha antigen in *E. coli*, were used for determining the sequence of *bca* (10). As is often the case with Gram-positive genes cloned into *E. coli*, many of the subclones were unstable (25). This problem is compounded in *bca* by a large region of repetitive DNA that provides multiple, fixed sites for homologous recombination. To verify that pJMS23 encodes the complete native gene without deletions, Southern blots of genomic DNA from A909 were probed with gene fragments from the clone. There were no differences found in the restriction maps of *bca* between A909 and pJMS23 (data not shown). We obtained the complete nucleotide sequence of *bca* independently on both stands using three strategies: transposon mutagenesis with Tn5seq1, synthetic oligonucleotide primers, and exonuclease III nested deletions (Fig. 1).

The complete nucleotide sequence of the *bca* locus and derived amino acid sequence for a single, large open reading frame are shown in Fig. 2. The structural gene consists of 3063 nucleotides, encodes 1020 amino acids, and has a calculated molecular mass of 108,705 Da. There is a prokaryotic promoter consensus sequence (TATAAT) upstream (at −10) from the initiating codon (27). There are no clear homologies in the −35 region assuming a spacing of 5–19 bases upstream from the −10 region (28). The probable ribosomal binding site flanking the 5′ end of *bca* is AGGAGA (29, 30). Downstream of the TAA termination codon are two regions with diad symmetry that could function as transcription terminators (31).

The derived amino acid sequence of the mature peptide of *bca* predicts a pK_a of 4.49, which is close to the values measured in our laboratory for both the native and the cloned C protein alpha antigen (L.C.M., unpublished data). The alpha antigen contains no cysteine and only a single methionine at the initiation codon. The alpha antigen is rich in proline (11% in the mature protein) but does not show the XPZ motif identified in the C protein beta antigen of GBS (12, 13) or the proline repeat motifs described in M protein of group A streptococci (32).

Deduced Signal Sequence of *bca* and Homologies. As a cell surface-associated protein, alpha antigen may use a signal sequence to be exported from the cytoplasm. A BLAST search identified five Gram-positive surface proteins with homology to the first 41 amino acids of the alpha antigen (Fig. 3*A*). Based on the pattern described for other Gram-positive signal sequences, it is likely that the first 41 amino acids of alpha antigen comprise a signal sequence (33, 34). There is a high proportion of arginine and lysine residues near the N terminal, followed by a hydrophobic region, a serine at position 36, and a valine at position 41. Other possibilities are cleavages after valine at position 54 or either of the alanine residues at positions 55 and 56 that follow a serine at position 52. Assuming that the signal sequence is cleaved following amino acid 41, the mature protein would contain 979 amino acids with a molecular mass of 104,106 Da. This suggests that the signal sequence is encoded by 123 nucleotides, making up 4% of the gene, and has a molecular mass of 4616 Da. Further support for a signal sequence of this size comes from Western blots comparing the sizes of the native and cloned alpha

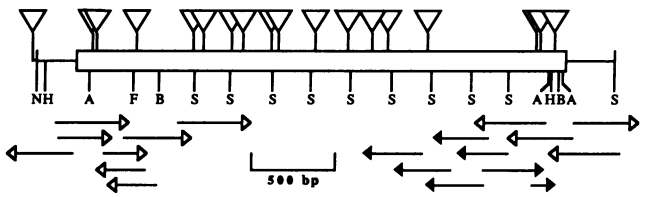

FIG. 1. Sequencing strategy and restriction endonuclease map of *bca*. The partial restriction endonuclease map encompasses the region of pJMS23 from an *Nde* I site to a *Sty* I site located at nucleotide 3594 for which the nucleotide sequence of *bca* and flanking region was determined. The open reading frame is illustrated by an open box. Transposon Tn5seq1 mutations (triangles) serve to prime nucleotide sequencing in both directions from each of the insertions. The regions of sequence obtained from oligonucleotide primers (open arrows) and the nested deletions (closed arrows) are also shown. Restriction endonuclease cleavage sites are abbreviated as follows: A, *Alu* I; B, *Bsm* I; F, *Fok* I; H, *Hin*cII; N, *Nde* I; S, *Sty* I. bp, base pairs.

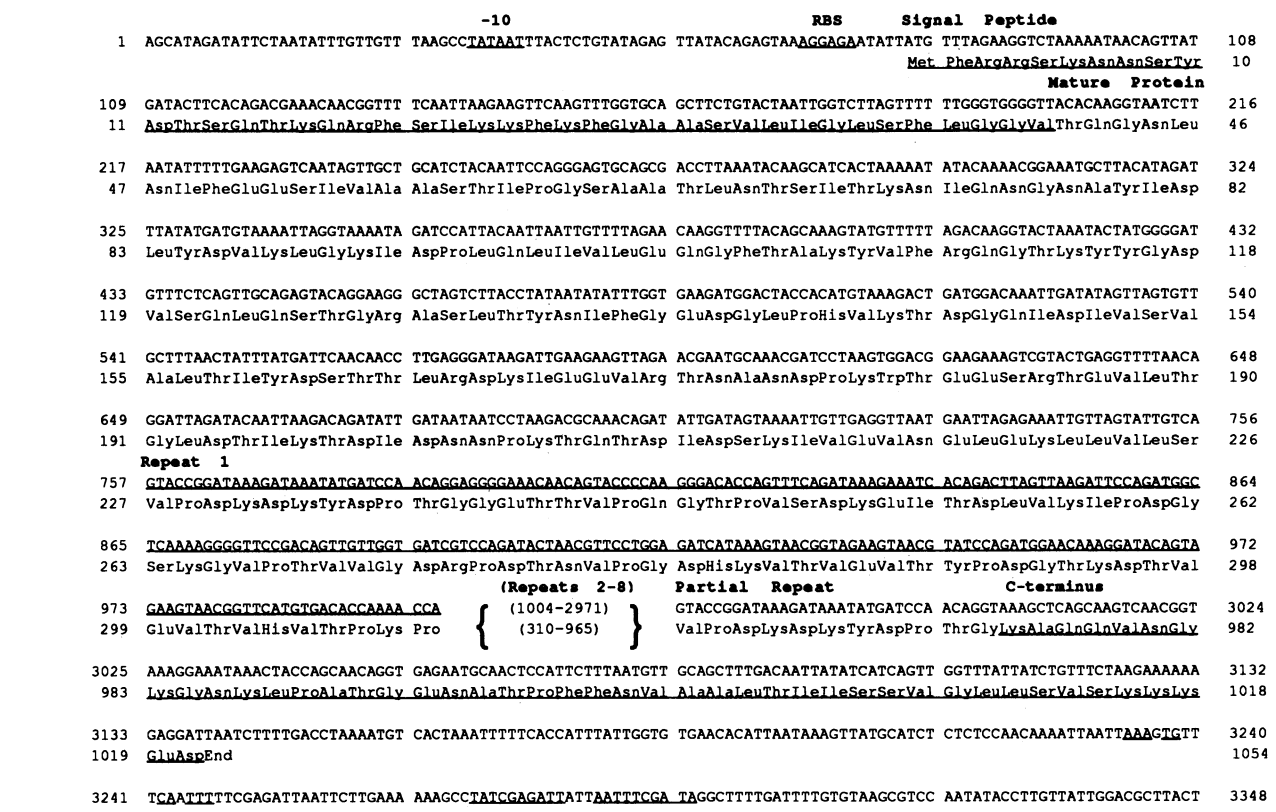

FIG. 2. Nucleotide and deduced amino acid sequences of *bca* and flanking regions. The DNA strand is shown 5′ to 3′, and nucleotides are listed on the upper line beginning 78 base pairs upstream from the open reading frame. The deduced amino acid sequence for the open reading frame is below the nucleic acid sequence. The G+C content of 40% and the codon usage are similar to other streptococcal genes (26). Highlighted features include the −10 (TATAAT) promoter consensus site, ribosomal binding site (RBS), signal sequence, repeat region 1, the C terminus, with the termination codon (TAA) at position 3161, and two regions of dyad symmetry that are potential transcriptional terminators.

antigens probed with the monoclonal antibody 4G8. As shown in Fig. 4, each of the steps of the alpha antigen protein ladder from clone pJMS23 is slightly larger than that of the native protein from GBS A909, which suggests that the signal sequence may not be processed in *E. coli* as it would be in GBS. The size difference is ≈4 kDa, which would correspond to a shorter (41 amino acids) rather than a larger (53–55 amino acids) signal sequence in *bca*.

Analysis of the N Terminus of *bca*. Following the putative signal sequence, there is a region of 185 amino acids before the repeated sequences. The N-terminal region contains 555 nucleotides, accounts for 18% of the gene, and encodes a polypeptide with a predicted molecular mass of 20,417 Da. A computer search comparing the primary nucleotide sequence and the derived amino acid sequence in all six reading frames of the N terminus of *bca* with sequences in GenBank and Swiss-Prot using the BLAST network of programs found no homologies, thus suggesting that this region of the gene is unlike any previously sequenced or described nucleic acid or amino acid sequence.

Repeating Unit Region of *bca*. Beginning at amino acid 679 of the DNA sequence, there are nine large tandem repeating units with identical nucleic acid and amino acid structures that encompass 74% of the gene. The size and repetitive nature of this region of *bca* are illustrated in Fig. 5. Each repeating unit consists of 246 nucleotides encoding 82 amino acids with a calculated molecular mass of 8665 Da. The entire repeat region contains 749 amino acids and consists of the nine identical repeating units and a partial repeating unit designated 9′. The calculated molecular mass of this region is 79,053 Da.

Until we understand the biological role and molecular mechanism of the repeating units, the determination of the beginning and end of the repeat is somewhat arbitrary. We chose to start from the N terminus, beginning with the first codon that was in the open reading frame. The repeating units could be defined as beginning out of frame or starting at the C-terminal side. BLAST computer searches for nucleic acid and derived amino acid homologies showed no significant matches for the repeat units. Therefore, these repeating units appear to be unique to the alpha antigen and are different in size and structure from those described for other streptococcal proteins (12, 13, 19, 35).

C-Terminal Anchor of *bca* and Homologies. Following the repeating units is a small C-terminal region containing 148 nucleotides and making up 4.4% of the gene. This region encodes 45 amino acids with a calculated molecular mass of 4672 Da. A BLAST search for amino acid homologies identified a class of Gram-positive surface proteins with a common membrane anchor motif (Fig. 3*B*), including the M proteins of group A *Streptococcus* and IgG binding proteins from both group A and group G *Streptococcus* (36). The amino acid composition at the C terminus is characteristic of the peptide membrane anchor, including a hydrophilic stretch with lysine before the LPXTGE motif (Fig. 3*B*) (32). This is followed by hydrophobic region with the consensus PPFFXXAA, where X designates a hydrophobic amino acid. Finally, there is a hydrophilic tail ending in aspartic acid that presumably extends into the cytoplasm of the cell.

Analysis of the Nucleotide Sequence and the Deduced Alpha Antigen Protein. Fig. 5 illustrates four distinct regions within the open reading frame of *bca* as determined from the nucleotide and derived amino acid sequences. A hydrophobicity plot of the amino acid sequence shows that the putative signal sequence has a short, hydrophilic N terminus, followed by a hydrophobic stretch, and ending in a hydrophilic region, whereas the C-peptide membrane anchor has a hydrophobic wall-spanning domain and a small hydrophilic tail (37, 38).

A

1 MFRRSKNNSYDTSQTKQRFSIKKFK FGAASVLIGL SFLGGV TQGNLNIFEESIVAA
2 MFKSNYERKMRYSIRK FSVGVASVAVRSLFMG SVAHA
3 MARQQTKKNYSLRKLK TGTASVAVALTVLGAGFA NQTEVRA
4 MTKNNTNRHYSLRKLK TGTASVAVALTVLGAGLVV NTNEVSA
5 MAKNNTNRHYSLRKLK TGTASVAVALTVLGAGFA NQTEVKA
6 MAKNNTNRHYSLRKLK TGTASVAVALTVLGAGFA NQTEVKANGDGNPREV

B

1 KAQQVNGKGNK LPATGE NAT PFFNVAAL TIISSVGLLSV SKKKE D
2 NKAPMKETKRQ LPYTGV TAN PFFTAAAL TVMATAGVAAVV KRKEE N
3 RPSQNKGMRSQ LPSTGE AAN PFFTAAAA TVMVSAGMLAL KRKEE N
4 AKKEDAKKAET LPTTGE GSN PFFTAAAL AVMAGAGALAVA SKRKE D
5 AKKDDAKKAET LPTTGE GSN PFFTAAAL AVMAGAGALAVA SKRKE D
6 SRSAMTQQKRT LPSTGE TAN PFFTAAAA TVMVSAGMLAL KRKEE N
7 NKAPMKETKRQ LPSTGE TAN PFFTAAAL TVM AAA
8 KGNPTSTTEKK LPYTGV ASN LVLEIMGLLGLIGTSFIAM KRRKS

FIG. 3. Homologies to the putative signal sequence and C-terminal membrane anchor of the C protein alpha antigen. (*A*) The N terminus of the C protein alpha antigen is shown on the top line (sequence 1) and is compared with the following Gram-positive signal sequences (accession codes are listed for each of the sequences): 2, the C protein beta antigen (S15330; STRBAGBA) and four M proteins of group A *Streptococcus*; 3, ennX (STRENNX); 4, emm24 (STREMM24); 5, M1 (S00767); 6, S01260. Lysine (K) and arginine (R) residues preceding the underlined hydrophobic stretch are in boldface type, as are serine (S) and threonine (T) residues preceding the probable signal cleavage sites. The probable cleavage site for the alpha signal is following the valine at position 41; however, alternate cleavage sites exist at positions 53–56. (*B*) The C terminus of the C protein alpha antigen is shown on the top line (sequence 1) and compared with the following Gram-positive membrane anchor peptides: 2, M5 (A28616), M6 (A26297), and M24 (A28549); 3, ennX (STREENX); 4, S00128, STRPROTG, and A26314; 5, spg (A24496); 6, arp4 (S05568) and emm49 (STRM49NX, STRMM24); 7, emm12 (STR12M). M5, M6, M24, emm12, emm49, and ennX are all M proteins; arp4 is a binding protein of group A *Streptococcus*. S00128, STRPPROTG, spg, and A26314 are IgG binding proteins of group G *Streptococcus*. Sequence 8 illustrates the membrane anchor for the beta antigen, which lacks the PPFFXXAA motif. Highlighted areas include lysine residues (K) preceding the LPXTGE motif (boxed), the hydrophobic region (underlined) with the PPFFXXAA consensus (boxed and underlined), and the terminal amino acid aspartic acid (D) or asparagine (N).

The native alpha antigen demonstrates a ladder of polypeptides at regularly repeating intervals that is also seen with the cloned gene product (Fig. 4). The size of the individual repeats in *bca* could code for a polypeptide of 8665 Da, which corresponds to the size differences in the protein ladder. To look at possible mechanisms generating protein heterogeneity, *bca* nucleotide and derived RNA and protein sequences were surveyed. Analysis of the nucleotide sequence of *bca* failed to show codons within the repeat regions that could cause early termination of translation. In addition, the amino acid sequence of the repeat region was screened with the Genetics Computer Group program for potential sites for proteolytic cleavage. A unique site within each repeat was sensitive to pH 2.5, represented by aspartic acid followed by proline. However, these sites were also found in the N terminus. Although the alpha antigen is relatively resistant to trypsin, there were numerous potential trypsin cleavage sites found in the sequence. Finally, modeling of RNA sequence and tertiary structure failed to identify regions within the repeats that might be involved with RNA-mediated self-cleavage.

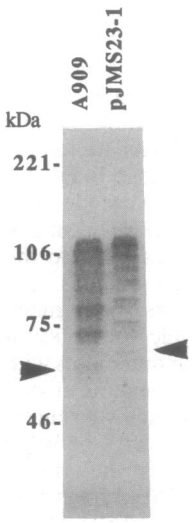

FIG. 4. Comparison of the cloned and native gene products of *bca*. Surface proteins of the A909 strain of group B *Streptococcus* (type 1a/C) and C protein alpha antigen clone pJMS23-1 were analyzed by SDS/PAGE and Western blotting and were probed with the alpha antigen-specific monoclonal antibody 4G8. Arrowheads illustrate an example of the differences between proteins. Molecular mass markers (in kDa) are shown to the left.

DISCUSSION

Two biological properties identified for the alpha antigen of GBS are the ability to resist opsonophagocytosis in the absence of specific antibody and the expression of epitopes that elicit protective antibodies (7, 8, 14, 15). Analysis of the sequence of the alpha antigen shows four distinct structural domains. The putative N-terminal signal sequence and the C-terminal membrane anchor support the hypothesis that the alpha antigen is a surface-associated membrane protein. These properties, along with the repeating unit motif, are shared by a number of Gram-positive proteins that are thought to be involved in the pathogenesis of bacterial infections (19).

The alpha antigen sequence identified a region of large, identical, tandem repeats composing 74% of the gene and demonstrating no homology to previously described protein or nucleic acids sequences. However, a number of virulence-associated proteins contain multiple repetitive elements. The M protein of group A *Streptococcus*, which is antiphagocytic, carries protective epitopes, and displays variability in antigen size and presentation, contains two extended tandem repeat regions and one nontandem repeat region occupying nearly two-thirds of the gene (17, 39, 40). The individual repeats are smaller in M protein than in the alpha antigen and range from 21 to 81 base pairs. In addition, there is divergence between the repeating units at the ends of the repeat region, while those in the middle are nearly identical. Pneumococcal surface protein A contains a region containing up to 10 repetitive segments of 20 amino acids each (35). Both M protein and pneumococcal surface protein A demonstrate antigenic variability and changes in protein/gene size thought to be mediated by repetitive DNA sequences in their structural genes (19, 35, 40). Other Gram-positive genes with repetitive motifs

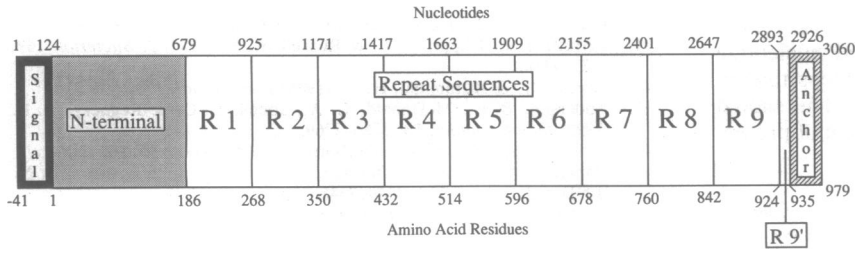

FIG. 5. Schematic of the open reading frame of *bca*. Summary of the structural features of the open reading frame of the C protein alpha antigen based on analysis of the amino acid sequence deduced from the nucleotide sequence of *bca*. The numbers above the boxes indicate the nucleotide position, and the numbers below are the amino acid residues of the mature protein within the open reading frame.

include the glycotransferase genes from *Streptococcus sobrinus* and *Streptococcus mutans* (41, 42). Immunodominant epitopes associated with repetitive sequences have been identified in a number of other pathogens including *Rickettsia rickettsii, Trichomonas vaginalis*, and *Clostridium difficile* (20, 43, 44). The repeats found in alpha antigens are unique for three reasons: (*i*) They are larger than those found for other Gram-positive surface proteins. (*ii*) They are identical at the nucleic acid level and do not diverge. (*iii*) The size of protein encoded by the repeating units corresponds to the laddering seen in the native and cloned alpha antigens.

The finding of large tandem repeating units raises many questions about the genotypic and phenotypic variability of the alpha antigen. When probed on Western immunoblots with the 4G8 monoclonal antibody, both the native and the cloned alpha antigen display a regular ladder of proteins varying by ≈8 kDa, and the size of the alpha antigen varies between strains (7). Restriction endonuclease mapping of the original alpha antigen clone pJMS23 showed multiple *Sty* I fragments of ≈270 base pairs (10). We found that strain A909 contains only one copy of *bca* and speculated that these fragments may be responsible for the protein heterogeneity. The nucleotide sequence confirms the repetitive nature of the gene but does not identify the mechanism of protein laddering.

Since multiple protein sizes are seen in both native and cloned backgrounds and since there is no evidence for a gene family, we postulate that laddering results from a mechanism common to both *E. coli* and GBS and/or is mediated by a property specific to the alpha antigen. Western blots on Tn*5* transposon insertion mutations within the repeat region still show laddering, which demonstrates that the C terminus is not required for heterogeneity, suggesting that either the N-terminal or repeat region determines laddering. It is not yet clear whether the mechanism functions at the level of transcription, during translation, or by posttranslational modification.

Studies of the alpha antigen among GBS isolates using a monoclonal antibody showed that the maximum molecular size of the alpha antigen is constant for a given isolate but varies widely among different isolates (7). The tandem repeating units could provide convenient fixed recombination sites for deletion or duplication of the repeat region. Deletion would reduce the size of the gene and might occur during DNA replication by unequal crossover or mispaired template slippage, which would occur in frame (45). Duplication of DNA could be a mechanism to amplify mutations within a repeat and create antigenic diversity. However, we have no evidence that the variation in the protein size of the alpha antigen is accompanied by antigenic diversity and the expression of different protective or opsonic epitopes.

The nine complete tandem repeats in the alpha antigen from A909 are identical at the nucleic acid level, which demonstrates a highly conserved structure. This suggests that the duplication causing the repeats is a recent event, that there are properties internal to the repeats that maintain their integrity, or that their structure is essential for the gene. Southern blots of genomic DNA from alpha antigen-bearing strains of GBS probed with alpha antigen-specific DNA show variability in gene size among strains (J.L.M., unpublished data). To look at the mechanism of genotypic diversity among strains, it will be necessary to clone and sequence *bca* from other phenotypic variants and to determine the phylogenetic relationships among C protein-bearing strains of GBS. The cloning, immunological and biochemical characterization, and nucleotide sequence analysis of the C proteins are initial steps toward understanding their biological roles (5, 10–12, 46). This should lead to the identification, isolation, and expression of protective epitopes that could be vaccine candidates.

We are grateful to Dr. Joanna B. Goldberg, Dr. Mike Cherry, Bryce Beseth, Sarah Olken, Jaylyn Olivo, and Julie McCoy. This research was supported by Public Health Service Grants AI28500, AI00981, and AI23339 from the National Institute of Allergy and Infectious Diseases and by a grant from Hoechst AG to Massachusetts General Hospital.

1. Baker, C. J. & Edwards, M. S. (1990) in *Infectious Diseases of the Fetus and Newborn Infant*, eds. Remington, J. S. & Klein, J. O. (Saunders, Philadelphia), pp. 742–811.
2. Rubens, C. E., Wessels, M. R., Heggen, L. M. & Kasper, D. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7208–7212.
3. Wessels, M. R., Rubens, C. E., Benedi, V. J. & Kasper, D. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8983–8987.
4. Ferrieri, P. (1988) *Rev. Infect. Dis.* S363–S366.
5. Michel, J. L., Madoff, L. C., Kling, D. K., Kasper, D. L. & Ausubel, F. M. (1991) in *Genetics and Molecular Biology of Streptococci, Lactococci, and Enterococci*, eds. Dunny, G. M., Cleary, P. C. & McKay, L. L. (Am. Soc. Microbiol., Washington), pp. 214–218.
6. Johnson, D. R. & Ferrieri, P. (1984) *J. Clin. Microbiol.* **19**, 506–510.
7. Madoff, L. C., Hori, S., Michel, J. L., Baker, C. J. & Kasper, D. L. (1991) *Infect. Immun.* **59**, 2638–2644.
8. Lancefield, R. C., McCarty, M. & Everly, W. N. (1975) *J. Exp. Med.* **142**, 165–179.
9. Madoff, L. C., Michel, J. L. & Kasper, D. L. (1991) *Infect. Immun.* **59**, 204–210.
10. Michel, J. L., Madoff, L. C., Kling, D. E., Kasper, D. L. & Ausubel, F. M. (1991) *Infect. Immun.* **59**, 2023–2028.
11. Cleat, P. H. & Timmis, K. N. (1987) *Infect. Immun.* **55**, 1151–1155.
12. Heden, L.-O., Frihtz, E. & Lindahl, G. (1991) *Eur. J. Immunol.* **21**, 1481–1490.
13. Jerlstrom, P. G., Chhatwal, G. S. & Timmis, K. N. (1991) *Mol. Microbiol.* **5**, 843–849.
14. Payne, N. R. & Ferrieri, P. (1985) *J. Infect. Dis.* **151**, 672–681.
15. Payne, N. R., Kim, Y. K. & Ferrieri, P. (1987) *Infect. Immun.* **55**, 1243–1251.
16. Enea, V., Ellis, J., Zavala, F., Arnot, D. E., Asavanich, A., Masuda, A., Quakyi, I. & Nussenzweig, R. S. (1984) *Science* **225**, 628–630.
17. Fischetti, V. A., Jones, K. F., Hollingshead, S. K. & Scott, J. R. (1988) *Rev. Infect. Dis.* S356–S359.
18. Pereira, M. E., Mejia, J. S., Ortega, B. E., Matzilevich, D. & Prioli, R. P. (1991) *J. Exp. Med.* **174**, 179–191.
19. Fischetti, V. A., Pancholi, V. & Schneewind, O. (1991) in *Genetics and Molecular Biology of Streptococci, Lactococci, and Enterococci*, eds. Dunny, G. M., Cleary, P. C. & McKay, L. L. (Am. Soc. Microbiol., Washington), pp. 290–294.
20. Dailey, D. C. & Alderete, J. F. (1991) *Infect. Immun.* **59**, 2083–2088.
21. vonEichel-Streiber, C. & Sauerborn, M. (1990) *Gene* **96**, 107–113.
22. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1990) *Current Protocols in Molecular Biology* (Wiley, New York).
23. Lopilato, J., Bortner, S. & Beckwith, J. (1986) *Mol. Gen. Genet.* **205**, 285–290.
24. Nag, D. K., Huang, H. V. & Berg, D. E. (1988) *Gene* **64**, 135–145.
25. Schneewind, O., Friedrich, K. & Lutticken, R. (1988) *Infect. Immun.* **56**, 2174–2179.
26. Hollingshead, S. K., Fischetti, V. A. & Scott, J. R. (1986) *J. Biol. Chem.* **261**, 1677–1686.
27. Doi, R. H. & Wang, L. F. (1986) *Microbiol. Rev.* **50**, 227–243.
28. Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acids Res.* **11**, 2237–2255.
29. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
30. Gold, L., Prinbnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stromo, G. (1981) *Annu. Rev. Microbiol.* **35**, 365–403.
31. Brendel, V. & Trifonov, E. N. (1984) *Nucleic Acids Res.* **12**, 4411–4427.
32. Fischetti, V. A., Pancholi, V. & Schneewind, O. (1990) *Mol. Microbiol.* **4**, 1603–1605.
33. vonHeijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
34. vonHeijne, G. & Abrahmsen, L. (1989) *FEBS Lett.* **244**, 439–446.
35. Yother, J. & Briles, D. E. (1992) *J. Bacteriol.* **174**, 601–609.
36. Wren, B. W. (1991) *Mol. Microbiol.* **5**, 797–803.
37. Engelman, D. M., Steitz, T. A. & Goldman, A. (1986) *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353.
38. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
39. Hollingshead, S. K., Fischetti, V. A. & Scott, J. R. (1986) *J. Biol. Chem.* **261**, 1677–1686.
40. Haanes, E. J. & Cleary, P. P. (1989) *J. Bacteriol.* **171**, 6397–6408.
41. Ferretti, J. J., Gilpin, M. L. & Russell, R. R. (1987) *J. Bacteriol.* **169**, 4271–4278.
42. Shiroza, T. & Kuramitsu, H. K. (1988) *J. Bacteriol.* **170**, 810–816.
43. Anderson, B. E., McDonald, G. A., Jones, D. C. & Regnery, R. L. (1990) *Infect. Immun.* **58**, 2760–2769.
44. vonEichel-Streiber, C. & Sauerborn, M. (1990) *Gene* **96**, 107–113.
45. Harayama, S., Rekik, M., Bairoch, A., Neidle, E. L. & Ornston, L. N. (1991) *J. Bacteriol.* **173**, 7540–7548.
46. Lindahl, G., Akerstrom, B., Vaerman, J. P. & Stenberg, L. (1990) *Eur. J. Immunol.* **20**, 2241–2247.