Genetic diversity of the C protein β -antigen gene and its upstream regions within clonally related groups of type Ia and Ib group B streptococci

Noriyuki Nagano,^{1,2} Yukiko Nagano,¹ Ryuichi Nakano,² Ryoichi Okamoto² and Matsuhisa Inoue²

Correspondence

Noriyuki Nagano

naganoynd3.dion.ne.jp

¹Medical Microbiology Laboratory, Funabashi Medical Center, 1-21-1 Kanasugi, Funabashi, Chiba 273-8588, Japan

²Department of Microbiology, School of Medicine and Environmental Infectious Disease, Graduate School of Medical Science, Kitasato University, Sagami-hara, Kanagawa, Japan

C protein β antigen (Bac), a surface protein of group B streptococci (GBS), is known to concurrently bind the Fc portion of IgA and factor H (FH). The authors' previous work has demonstrated that mRNA expression levels show diversity among clonally related strains containing genes (*bac*) encoding Bac, with high expression noted in invasive strains. In this study, the *bac* gene and upstream regions containing putative promoters, three ORFs and an IS1381 insertion sequence were characterized. Three invasive strains showed high *bac* expression levels and did not show any notable mutations except one strain producing Bac that was able to bind FH but not IgA. A deletion of 51 amino acid residues, including part of the Bac IgA-binding region, was identified and hypothesized to contribute to the loss of the IgA-binding ability of this strain. A vaginal strain that showed somewhat higher *bac* expression levels and produced Bac lacking immunoreactivity contained an 11 bp deletion, which generated a premature termination codon, in the region preceding the IgA-binding region. In another vaginal strain that did not express *bac*, disruption of the upstream ORFs of the sensor histidine kinase and DNA-binding response regulator, due to frameshift mutations, was noted although it is not known whether these proteins directly affect *bac* expression levels. An IS1381 insertion into the promoter region was found in another vaginal strain that showed low expression levels and produced Bac with a significantly larger proline-rich repeat region. These results demonstrate considerable genetic diversity of the *bac* and upstream regions of invasive and noninvasive GBS, which may contribute to the variability of *bac* expression levels among those strains.

Received 21 September 2005

Revised 12 December 2005

Accepted 12 December 2005

INTRODUCTION

Group B streptococci (GBS) are one of the most important causes of severe neonatal infections, including sepsis and meningitis. Several GBS surface proteins have been reported (Lindahl *et al.*, 2005) to contribute to interactions with epithelial cells (Baron *et al.*, 2004), extracellular matrices (Beckmann *et al.*, 2002; Spellerberg *et al.*, 1999), plasma

proteins (Schubert *et al.*, 2002), and immunocomponents (Bohnsack *et al.*, 1991; Hedén *et al.*, 1991; Jerlström *et al.*, 1991). The C protein β antigen (Bac) is able to simultaneously bind both the Fc portion of human IgA and factor H (FH) (Areschoug *et al.*, 2002b). IgA-binding surface proteins, Bac of GBS as well as M protein of *Streptococcus pyogenes*, have been shown to bind in the C₂-C₃ interdomain region of IgA, at sites also used by the human IgA receptor CD89, an important mediator of IgA effector function (Pleass *et al.*, 2001). Thus, the ability of Bac to bind IgA may allow GBS to evade immune responses that would normally be triggered by the binding of the IgA to CD89 (van Egmond *et al.*, 2000).

FH regulates the alternative complement pathway by acting as a cofactor for factor I in promoting the cleavage of surface-bound C3b to produce inactive iC3b, or by competing with factor B for binding to C3b, therefore disrupting the C3bBb complex (Pangburn *et al.*, 1977; Weiler *et al.*, 1976; Whaley &

Abbreviations: Bac, C protein β antigen; *bac*, gene encoding C protein β antigen; CSF, cerebrospinal fluid; Ct, cycle threshold; FH, factor H; GBS, group B streptococci.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AB121739 (for strain 5) and AB221536 (for strain 12).

A supplementary figure showing the amino acid sequence alignment of Bac from strains 22 and 5 and the sequences in GenBank accession numbers X59771 and X58470 is available with the online version of this paper.

Ruddy, 1976). Several important Gram-positive pathogens have exploited FH to avoid complement attack and phagocytosis (Blackmore *et al.*, 1998; Jarva *et al.*, 2002; Lindahl *et al.*, 2000; Ram *et al.*, 1998a, b). FH bound to Bac on the surface of GBS retains its ability to down-regulate complement activation, suggesting that Bac-expressing GBS may use bound FH to inhibit complement attack (Areschoug *et al.*, 2002b). Another possible virulence function of bacterial bound FH has been suggested in recent work which describes the binding of factor-H-like protein 1, representing a truncated form of FH, to fibronectin-binding surface protein that then promoted intracellular invasion by *S. pyogenes* (Pandiripally *et al.*, 2003). By binding two immunocomponents, IgA and FH, Bac may play an important role in infection by this organism, although conclusive evidence is not yet available. This is supported by our previous findings in which neonatal invasive strains from clonally related groups of serotypes Ia and Ib GBS expressed high levels of Bac (Nagano *et al.*, 2002). Recently, the surface protein PspC of *Streptococcus pneumoniae* has also been shown to bind both human secretory IgA and FH concurrently (Sandhya *et al.*, 2004).

In this work, we analysed gene sequences surrounding the *bac* gene, encoding Bac, of our representative GBS strains from colonized women and from invasive neonatal infections to investigate whether sequence variability exists among those strains with different Bac expression levels, and their association with FH-binding as well as with the previously described IgA-binding abilities. Moreover, we characterized the *bac* gene of a previous wild-type invasive strain (Nagano *et al.*, 2002) that produced a large amount of mature Bac lacking IgA-binding ability.

METHODS

Bacterial strains and growth conditions. A total of 19 GBS strains having 604 bp amplification fragments of the *bac* gene, previously reported (Nagano *et al.*, 2002), are listed in Table 1. Seven representative strains (strains 5, 6, 7, 13, 16, 22 and 23) selected on the basis of Bac production levels were used for genetic characterization. GBS strains were grown in Todd-Hewitt broth (BBL Microbiology Systems).

Immunoblotting. Extraction of cell-bound protein, SDS-PAGE, and electroblotting onto PVDF membranes (Clearblot P membranes; Atto) were performed as previously described (Nagano *et al.*, 2002). Protein samples (aliquots of 5 µg) were subjected to SDS-PAGE with 5–20% gradient gels. The membranes were incubated with human FH (10 µg ml⁻¹; Calbiochem). Bound ligands were detected by using goat anti-FH serum (1:1000; Calbiochem) and a peroxidase-labelled rabbit anti-goat immunoglobulin G (1:1000; Calbiochem). For colour development, 4-chloro-1-naphthol was used. The amounts of FH bound were analysed by densitometry (Scanning Imager PDSI and ImageQuaNT software; Amersham Pharmacia Biotech).

DNA manipulation and nucleotide sequencing. Genomic DNA was extracted by using a Wizard Genomic DNA Purification kit (Promega) and 20 U mutanolysin. Preliminary sequence data were obtained from GenBank (Hedén *et al.*, 1991; Jerlström *et al.*, 1991; accession number AY598359).

Table 1. Strains used in this study

Strain no.	Source*	Genotype
5	Woman, vagina	Ia-3
6	Woman, vagina	Ia-3
7	Woman, vagina	Ia-3
8	Woman, vagina	Ia-3
9	Woman, vagina	Ia-3
10	Infant, blood	Ia-3
11	Infant, CSF	Ia-3
12	Infant, CSF	Ia-3
13	Infant, CSF	Ia-3
14	Infant, CSF	Ia-3
15	Infant, CSF	Ia-3
16	Woman, vagina	Ib-1
17	Woman, vagina	Ib-1
18	Woman, vagina	Ib-1
19	Infant, blood	Ib-1
20	Infant, blood	Ib-1
21	Infant, blood	Ib-1
22	Infant, CSF	Ib-1
23	Infant, CSF	Ib-1

*CSF, cerebrospinal fluid.

PCR amplification of the *bac* gene and flanking regions was carried out with primer pair CB-SEQ (annealing temperature of 60 °C), 5'-CCTTTTATAGGACTGTAAAGCATCC-3' and 5'-GCAAAGAGAA-TATCGTCAACGTC-3', designed from published sequences. For strains 22 and 5, which showed interesting characteristics of Bac, amplified DNAs were purified with a Wizard SV Gel and PCR Clean-Up System (Promega), and cloned into a pGEM-T Easy vector (Promega) using *Escherichia coli* JM109 as the host strain. The plasmid DNAs were finally extracted and purified with a Wizard Plus SV Minipreps DNA Purification System (Promega). Sequences were determined by primer walking. Sequencing reactions were performed with a BigDye Terminator Cycle Sequencing Kit (version 1.1; Applied Biosystems) and the DNA sequences were resolved using an ABI PRISM model 3100 Genetic Analyser (Applied Biosystems). The sequences obtained were assembled into contigs with Sequencer (Gene Codes). For sequence confirmation, several PCR products, partially amplified with the genomic DNAs, were purified and subjected directly to sequence analyses.

The PCR products, amplified by using primer pair CB-SEQ, from strains 5 and 6 were purified and then subjected to restriction analysis using 10 U each of *Bgl*II or *Xba*I (Takara Bio) in a 20 µl volume for overnight reaction at 37 °C. A restriction analysis was performed in the same manner for strain 22 as a restriction profile control.

A 3.8 kb sequence containing a partial *bac* gene and the upstream region was analysed for strains 7, 13, 16, 22 and 23. Three primer pairs were designed based on published sequences so that an overlap in sequences was obtained among the PCR products: C-P (annealing temperature of 58.5 °C), 5'-TCACTTCCTTTTATAGGACTGT-3' and 5'-TTAGCTTCCGAATTATGTTTT-3'; H-S (annealing temperature of 57 °C), 5'-TTGCCTTTCTCTTTTCCT-3' and 5'-CAAACCAAC-TAGTTCTGGTAA-3'; and D-I (annealing temperature of 60 °C), 5'-CATCGTTTCAATATGGTCCT-3' and 5'-GTCAGCAAGTTTC-ATACACA-3'. PCR amplicons were TA cloned as described above and both strands were sequenced. Alternatively, direct sequencing of the PCR products was performed for confirmation.

Relative analysis of the copy number of the *bac* gene per genome was performed for 19 strains by using a real-time PCR assay. The primer pair R-C, 5'-TCCAGTATCAGTGGCAGATGTTTC-3' and 5'-TCCTCGGAACCTTGAGACAACAA-3', used for amplifying a 108 bp fragment of the *bac* gene, and the probe R-P, 5'-CCAGCTGTTTCTCGACCGGTTCAAC-3', labelled at the 5' end with FAM and at the 3' end with nonfluorescent Eclipse Quencher, were designed and obtained (Takara Bio). Real-time PCR was carried out using a Smart Cycler II system (Cepheid) and QuantiTect Probe Master Mix reagents (Qiagen). Each 25 μ l PCR mixture contained 12.5 μ l 2 $\times$ Master Mix, 200 nM of the probe, 500 nM of each of the primers and 1 μ l of the template. For template, three different concentrations of genomic DNA (1, 20 and 100 ng μ l⁻¹) were employed for all strains. PCRs were performed with the following parameters: 1 cycle of 15 min at 95 °C for HotStar Taq polymerase activation, 40 cycles of 15 s at 94 °C for denaturation and 60 s at 60 °C for annealing/extension. PCRs were run for each template concentration in triplicate. The cycle threshold (Ct) for each reaction was determined by using the software version 2.0 supplied with the system.

Bioinformatic analyses. BLAST-P and BLAST-N analyses were done using the NCBI web site (<http://www.ncbi.nlm.nih.gov/blast/>). Multiple sequence alignments of protein sequences were performed with CLUSTALW using the EBI web site (<http://www.ebi.ac.uk/services/index.html>) and the alignment was edited with BioEdit version 5.0.9 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred>) was used for prediction of protein secondary structures.

RESULTS

Binding of FH to Bac

Immunoblotting results for all 19 strains are shown in Fig. 1. For genotype Ia-3 strains, FH immunoreactivity was high in five strains from infant CSF (strains 11–15), but binding was not observed in strain 10, from infant blood. The binding was low or was not recognized in vaginal isolates of genotype Ia-3. For genotype Ib-1 strains, FH immunoreactivity was high in four strains from blood and CSF (strains 20 and 21, and 22 and 23, respectively), whereas it was low in strain 19 from blood. The binding was low or was not recognized in vaginal isolates of genotype Ib-1. All bands which were reactive with FH showed comparable bands that were predominantly reactive with anti-Bac monoclonal antibody (Naess *et al.*, 1991; kindly provided by L. Bevanger). A few bands that weakly reacted with anti-Bac monoclonal antibody, which were probably partially degraded products of the mature Bac (Russell-Jones & Gotschlich, 1984), were observed in strains 11–14 and 21–23, but were not in strains 15 and 20. Moreover, the binding levels of Bac to FH were found to be correlated with those to IgA among all strains examined except strain 22, which demonstrated FH binding but not IgA binding (Nagano *et al.*, 2002).

PCR, sequencing of the *bac* gene and flanking region

The primer pair CB-SEQ was designed to amplify the *bac* gene, 278 bp upstream including the *bac* putative promoter region, and 344 bp downstream. Amplicons of the predicted lengths of approximately 4.1 kb were obtained with three

Fig. 1. Western immunoblot analysis of Bac. The cell-bound proteins separated by SDS-PAGE were blotted onto a PVDF membrane and probed with human FH (FH). Levels of FH binding were shown by combining immunoblotting results of IgA and Bac monoclonal antibody (MAb) previously published by us (Nagano *et al.*, 2002). The numbers correspond to strain numbers: strains 5–15 from genotype Ia-3 and 16–23 from genotype Ib-1 (Table 1).

exceptions: an amplicon of approximately 3.9 kb was produced for strain 22, whereas amplicons of approximately 5.1 kb were produced for strains 5 and 6.

Complete sequence analysis of the amplicon of strain 22, lacking IgA-binding ability, revealed a 3874 bp amplicon with a deleted 153 bp sequence between bp positions 1029 and 1030. Analysis of the deduced 1083 amino acid sequence of the Bac protein (see supplementary Fig. S1, available with the online version of this paper) indicated the presence of the motif MLKKIE which has previously been suggested to be essential for IgA binding (Jerlström *et al.*, 1996). However, 51 amino acids (position 214–264) starting from the twenty-fourth amino acid toward the C-terminal end from the motif were lost, and this caused a lack of the last 12 amino acid residues of the 73-residue IgA-binding region (Hedén *et al.*, 1991; Jerlström *et al.*, 1996). The deletion of these 51 amino acids caused a partial loss of α -helical structure of the region as defined by the software program

PSIPRED (data not shown). In addition to the missing 51 amino acid residues, a comparison of the deduced amino acid sequence of Bac with published sequences showed an Asp-7-Asn substitution and the presence of a shorter proline-rich repeat region consisting of 30 three-residue repeats in the C-terminal cell-wall-spanning domain, which was identical to that of the published sequence (Fig. S1).

For strain 5, nucleotide sequencing of the amplicon revealed that the fragment was 5122 bp in length and an 861 bp sequence was inserted between nucleotides 263 and 264, corresponding to 16 bp upstream of the *bac* gene translation start codon. This fragment, identified as IS1381 with minor variations from the published GBS-derived sequences (GenBank accession no. AY598359), contained a 14 bp terminal inverted-repeat sequence and 5 extra bases (TAATA) at the 5' inverted repeat end. It is noteworthy that this IS element was inserted in the opposite orientation from the *bac* gene transcriptional direction, where the original putative promoter regions (Jerlström *et al.*, 1991) were disrupted. The analysis of the deduced 1212 amino acid sequence of Bac revealed a significantly longer proline-rich repeat region of 168 amino acids (position 789 to 956) corresponding to 56 three-residue repeats in the C-terminal cell wall-spanning domain (Fig. S1). Additionally, two amino acid substitutions, Glu-307-Gly and Thr→Ala at the eleventh residue after the repeat region, were noted.

Restriction enzyme cleavage analysis of PCR products

Restriction endonuclease cleavage patterns of the 5.1 kb PCR products from strains 5 and 6 showed identical numbers and sizes of fragments. Three restriction fragments (3416, 923 and 783 bp) were generated by digestion using *Bgl*II, two (3602 and 1520 bp) using *Xba*I, and four (1896, 1520, 923 and 783 bp) using both. Fragment sizes were calculated on the basis of the restriction map generated from the nucleotide sequence of strain 5 (Fig. 2). A 923 bp *Bgl*II fragment was derived from a cleavage site located in the IS1381 element, whereas two *Bgl*II fragments (2588 and 845 bp) were obtained with the 3.9 kb PCR product from strain 22, which did not contain IS1381. Moreover, a 1520 bp *Xba*I fragment contained the whole proline-rich repeat region. These results indicated that the PCR products from strains 5 and 6 contained the same IS1381 location and proline-rich repeat region.

Upstream genetic environment of the *bac* gene

Nucleotide sequences of 1.0 kb partial *bac* genes and their 2.8 kb upstream regions were analysed in strains 7, 13, 16, 22 and 23 (Fig. 2). Sequences in putative *bac* promoter regions were identical among these five strains. Moreover, the same gene organization in the upstream region was observed in all five strains, in which three ORFs encoding a

Fig. 2. Schematic representation of *bac* genes and upstream regions. Genes are shown as open arrows to indicate their orientations: Bac, C protein β antigen; Hyp, hypothetical protein; His, sensor histidine kinase; DBP, DNA-binding response regulator; 1167, IS1167 putative transposase. The partially sequenced region of the *bac* gene is indicated as a solid box. The hatched arrows or box contain frameshift mutations and are disrupted by internal stop codons. For IS1381, left and right inverted repeats are indicated by open and solid triangles, respectively, and the directions of transcription of the transposase genes are indicated by arrows. Genes confirmed by PCR strategy are shown as dashed arrows or box. Restriction maps of the *bac* and flanking regions of strains 22 and 5 are shown. The diagram is not drawn to scale.

hypothetical protein with 89 % similarity to *sp2191* (GenBank accession no. AE007507), a sensor histidine kinase, a DNA-binding response regulator (GenBank accession no. AY598359), and an *IS1381* element were conserved in size and gene order. However, in strain 16, there was a one-nucleotide deletion found in the second ORF, and a one-nucleotide insertion found in the third ORF, which resulted in premature termination codons.

The deduced amino acid sequence of the partial *bac* genes showed signal sequences and an IgA-binding region, with Asn-87-Thr and Glu-160-Gly noted in strain 13 and Lys-128-Asn and Gln-142-Arg noted in strain 23. In strain 7, there was an 11 bp deletion in the region preceding the IgA-binding region, which generated a translational stop codon at six amino acids following the deletion. Strains 5 and 6 showed the same gene organizations observed in the above five strains.

Relative copy number ratio of the *bac* gene per genome among strains

The Ct value, defined as the first cycle at which fluorescence is detected above the background, reflects the amount of template present initially. For each genomic DNA concentration used as template, the mean and standard deviations of Ct values for all 19 strains were comparable with those for individual strains; 24.4 ± 0.58 , 20.2 ± 0.38 and 18.0 ± 0.28 , for 1, 20 and 100 ng genomic DNA, respectively. Moreover, a linear correlation ($r^2 = 0.97$) was observed between Ct values and input concentration [log (nanograms)]. Taken together, these data indicated that equal copy number(s) of the *bac* gene were present in the genomic DNA regardless of strain.

DISCUSSION

The IgA-binding region was mapped to a 73-amino-acid region in the N-terminal portion of the Bac protein (Hedén *et al.*, 1991; Jerlström *et al.*, 1996), in which the motif MLKKIE was suggested to be responsible for the binding (Jerlström *et al.*, 1996). We have previously described a CSF-derived strain of serotype Ib that produced a remarkable amount of mature Bac that lacked IgA-binding ability (Nagano *et al.*, 2002); this absence of binding has not been reported in wild-type strains. Interestingly, complete sequence analysis of the *bac* gene of this strain revealed the presence of MLKKIE, whereas 51 amino acids (position 214 to 264) in the vicinity of the motif were lost. This deletion included the last 12 amino acid residues of the IgA-binding region. Alternatively, in preliminary experiments in which we treated Bac-producing bacterial cells with mutanolysin, we found through N-terminal sequencing that cleavage of Bac occurs reproducibly at least at two positions, between amino acids 217 and 218, and 382 and 383. Thus, cleaved fragments of approximately 122 kDa (containing the last 8 amino acid residues of the IgA-binding region) and 99 kDa (not containing the IgA-binding region) did not bind to IgA (data not shown). These results suggest that the last 12 amino acid residues (position 214 to 225) of the IgA-binding region

are necessary for Bac to maintain its functionality of IgA binding. The functional protein may need these missing residues in order to fold properly.

The binding levels of Bac to FH were found to correlate with binding levels to IgA among all strains examined except strain 22, which showed FH but not IgA binding. These observations are consistent with the non-overlapping binding sites for IgA and FH localized in the C-terminal half of the protein (Areschoug *et al.*, 2002b). Furthermore, pre-binding of Bac monoclonal antibodies did not affect levels of FH binding (data not shown).

PCR assay showed significantly longer amplicons from two vaginal strains, strains 5 and 6. In strain 5, there was an *IS1381* element with minor variations from the published GBS-derived sequences in the region immediately upstream (16 bp upstream of the translation start codon) of the *bac* gene, where the original putative promoter regions (Hedén *et al.*, 1991; Jerlström *et al.*, 1991) were disrupted. It is noteworthy that this IS element, which contained a 14 bp terminal inverted-repeat sequence and five extra bases (TAATA), was inserted in the opposite orientation from the transcriptional direction of the *bac* gene. Many IS elements have been shown to affect the expression of neighbouring genes by inserting upstream of those genes, resulting in new promoter sequences capable of driving their expression (Galas & Chandler, 1989; Mahillon & Chandler, 1998; Podglajen *et al.*, 1995). The *IS1381* contained in its 5' end the sequences TTGTAC and TAATAT (the last T from the original putative promoter region of the *bac* gene), which resemble the *E. coli* -35 and -10 consensus sequences, with a spacer of 23 bp. Thus, the newly inserted IS element may provide an alternative promoter for the transcription of the *bac* gene, with lower levels of mRNA transcripts being expressed (Nagano *et al.*, 2002). Several IS elements have been identified in GBS (Dmitriev *et al.*, 2003; Granlund *et al.*, 2001; Rubens *et al.*, 1989; Tamura *et al.*, 2000), and in some cases, with integration of these IS elements in virulence genes (Granlund *et al.*, 1998; Spellerberg *et al.*, 2000). In a study by Kong *et al.* (2002), integration of *IS1381* in the *bac* gene inactivated IgA-binding activity. Our analysis revealed that, among all strains including the above two vaginal strains, *IS1381* had integrated after the third upstream ORF nucleotide, corresponding to the integration site described in the literature (GenBank accession no. AY598359). *IS1381*, previously reported in *S. pneumoniae* (Sanchez-Beato *et al.*, 1997), has also been found in multiple copies in GBS (Tamura *et al.*, 2000). A BLAST search of our *IS1381* in the *Streptococcus agalactiae* A909 genome (<http://www.tigr.org/tdb>) identified at least six *IS1381* elements that were highly homologous. IS elements can contribute to the evolution of bacterial virulence and to the evolution of bacterial species by mediating genetic recombination (Mahillon & Chandler, 1998). Therefore it is noteworthy that two copies of *IS1381* were present in proximity to each other and to *bac*.

The longer *bac* gene from strain 5 also contained extra nucleotides in the repeat region located in the C-terminal

cell-wall-spanning domain compared to the original sequences. This region, designated the XPZ region by Hedén *et al.* (1991) or the Wr region by Jerlström *et al.* (1991), is characterized by proline-rich sequences with highly periodic, three-residue tandem repeats. A high degree of genetic variability within this region has been observed among different Bacs (Berner *et al.*, 2002; Kong *et al.*, 2002), yet sequence periodicity remains strictly conserved (Areschoug *et al.*, 2002a). The repeat region of 168 amino acids corresponding to 56 three-residue repeats, found in the larger Bac, is the largest among the sizes that have been reported (Areschoug *et al.*, 2002a; Berner *et al.*, 2002). It has been shown that the proline-rich repeat region is not required for FH binding (Areschoug *et al.*, 2002b), and the functions of this region having size polymorphism have remained unclear. However, because this region is exposed on the bacterial cell surface (Areschoug *et al.*, 2002a), it may contribute to interactions with host proteins.

In strains 5 and 6, PCR amplicons shared restriction enzyme patterns and molecular size. Concurrent findings of IS1381 and the largest repeated region may be not uncommon since these two strains were derived from unrelated healthy women.

We have previously reported that the level of expression of *bac* mRNA varied among strains in spite of each having *bac* genes (Nagano *et al.*, 2002). In strain 7, an 11 bp deletion in the region preceding the IgA-binding region generated a translational stop codon at six amino acids following the deletion. This strain did not show Bac immunoreactivity, and presence of Bac protein in the culture supernatant could not be detected; therefore, an incomplete and non-functional Bac may be produced by this strain. Nevertheless, the mRNA transcript level for this strain was somewhat higher than those from strains lacking Bac production (Nagano *et al.*, 2002). In many organisms, including bacteria, it has often been found that mutations that generate premature termination codons in coding sequences are associated with reduced levels of the corresponding mRNA (Harries *et al.*, 2004; Maquat, 1995; Peltz *et al.*, 1994; Ruizchevarria *et al.*, 1996). It is possible that the mRNA transcript levels observed in strain 7 might be the result of a decrease in the original mRNA level due to a frameshift-derived premature termination codon.

In conclusion, this study documents a considerable genetic diversity of the *bac* gene and upstream regions among invasive and noninvasive GBS, which may contribute to the variability of *bac* expression levels among those strains. Our findings included IS element insertions into *bac* promoter regions, and an 11 bp deletion in the *bac* gene that disrupted the protein by an internal stop codon. Disruption of the sensor histidine kinase and DNA-binding response regulator by frameshift mutations was also noted although it is not known whether such proteins affect *bac* expression. Moreover, we identified that the 153 bp deletion of *bac* gene contributed to the loss of IgA-binding ability observed in a wild-type invasive strain.

The finding that a wild-type invasive strain had lost IgA-binding ability challenges the hypothesis that IgA binding contributes to virulence. However, it is not known at which stage(s) of infection the IgA or FH binding contributes. On mucosal surfaces, which are the major potential invasion sites, secretory IgA, as the main antibody in secretions, provides a first line of defence against pathogenic microorganisms (Kerr, 1990). So interaction of Bac with IgA may be critical for the survival of GBS on the mucosal epithelium through evasion of IgA-mediated clearance. In animal studies, the risk of GBS meningitis has been correlated with the magnitude and duration of bacteraemia (Ferrieri *et al.*, 1980). Since the strain without IgA-binding ability was derived from infant CSF, one possible role of the FH-binding Bac may be survival in the bloodstream and tropism of the organism for the blood-brain barrier. The role of IgA and FH binding to Bac in invasiveness needs to be clarified in further studies.

ACKNOWLEDGEMENTS

We are grateful to T. Sasahara and Y. Sato, Department of Microbiology, School of Medicine, Kitasato University, for helpful discussions. Y. Saito, Medical Microbiology Laboratory, Funabashi Medical Center, is acknowledged for technical assistance.

REFERENCES

- Areschoug, T., Linse, S., Ståhlhammar-Carlemalm, M., Hedén, L.-O. & Lindahl, G. (2002a). A proline-rich region with a highly periodic sequence in streptococcal β protein adopts the polyproline II structure and is exposed on the bacterial surface. *J Bacteriol* **184**, 6376–6383.
- Areschoug, T., Ståhlhammar-Carlemalm, M., Karlsson, I. & Lindahl, G. (2002b). Streptococcal β protein has separate binding sites for human factor H and IgA-Fc. *J Biol Chem* **277**, 12642–12648.
- Baron, M. J., Bolduc, G. R., Goldberg, M. B., Auperin, T. C. & Madoff, L. C. (2004). Alpha C protein of group B *Streptococcus* binds host cell surface glycosaminoglycan and enters cells by an actin-dependent mechanism. *J Biol Chem* **279**, 24714–24723.
- Beckmann, C., Waggoner, J. D., Harris, T. O., Tamura, G. S. & Rubens, C. E. (2002). Identification of novel adhesins from group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding. *Infect Immun* **70**, 2869–2876.
- Berner, R., Ruess, M., Bereswill, S. & Brandis, M. (2002). Polymorphisms in the cell wall-spanning domain of the C protein β -antigen in clinical *Streptococcus agalactiae* isolates are caused by genetic instability of repeating DNA sequences. *Pediatr Res* **51**, 106–111.
- Blackmore, T. K., Fischetti, V. A., Higgs, N. H., Sadlon, T. A., Ward, H. M. & Gordon, D. L. (1998). M protein of the group A streptococcus binds to the seventh short consensus repeat of human complement factor H. *Infect Immun* **66**, 1427–1431.
- Bohnsack, J. F., Mollison, K. W., Buko, A. M., Ashworth, J. C. & Hill, H. R. (1991). Group B streptococci inactivate complement component C5a by enzymic cleavage at the C-terminus. *Biochem J* **273**, 635–640.
- Dmitriev, A., Yang, M., Shakleina, E., Tkacikova, L., Suvorov, A., Mikula, I. & Yang, Y. H. (2003). The presence of insertion elements IS1548 and IS861 in group B streptococci. *Folia Microbiol* **48**, 105–110.

- Ferrieri, P., Burke, B. & Nelson, J. (1980). Production of bacteremia and meningitis in infant rats with group B streptococcal serotypes. *Infect Immun* 27, 1023–1032.
- Galas, D. J. & Chandler, M. (1989). Bacterial insertion sequences. In *Mobile DNA*, pp. 109–162. Edited by D. E. Berg & M. M. How. Washington, DC: American Society for Microbiology.
- Granlund, M., Oberg, L., Sellin, M. & Norgren, M. (1998). Identification of a novel insertion element, IS1548, in group B streptococci, predominantly in strains causing endocarditis. *J Infect Dis* 177, 967–976.
- Granlund, M., Michel, F. & Norgren, M. (2001). Mutually exclusive distribution of IS1548 and GBSI1, an active group II intron identified in human isolates of group B streptococci. *J Bacteriol* 183, 2560–2569.
- Harries, L. W., Hattersley, A. T. & Ellard, S. (2004). Messenger RNA transcripts of the hepatocyte nuclear factor-1 α gene containing premature termination codons are subject to nonsense-mediated decay. *Diabetes* 53, 500–504.
- Hedén, L.-O., Frithz, E. & Lindahl, G. (1991). Molecular characterization of an IgA receptor from group B streptococci: sequence of the gene, identification of a proline-rich region with unique structure and isolation of N-terminal fragments with IgA-binding capacity. *Eur J Immunol* 21, 1481–1490.
- Jarva, H., Janulczyk, R., Hellwage, J., Zipfel, P. F., Björck, J. & Meri, S. (2002). *Streptococcus pneumoniae* evades complement attack and opsonophagocytosis by expressing the *pspC* locus-encoded Hic protein that binds to short consensus repeats 8–11 of factor H. *J Immunol* 168, 1886–1894.
- Jerlström, P. G., Chhatwal, G. S. & Timmis, K. N. (1991). The IgA-binding β antigen of the c protein complex of group B streptococci: sequence determination of its gene and detection of two binding regions. *Mol Microbiol* 5, 843–849.
- Jerlström, P. G., Talay, S. R., Valentin-Weigand, P., Timmis, K. N. & Chhatwal, G. S. (1996). Identification of an immunoglobulin A binding motif located in the β -antigen of the c protein complex of group B streptococci. *Infect Immun* 64, 2787–2793.
- Kerr, M. A. (1990). The structure and function of human IgA. *Biochem J* 271, 285–296.
- Kong, F., Gowan, S., Martin, D., James, G. & Gilbert, G. L. (2002). Molecular profiles of group B streptococcal surface protein antigen genes: relationship to molecular serotypes. *J Clin Microbiol* 40, 620–626.
- Lindahl, G., Sjöbring, U. & Johnsson, E. (2000). Human complement regulators: a major target for pathogenic microorganisms. *Curr Opin Immunol* 12, 44–51.
- Lindahl, G., Ståhlhammar-Carlemalm, M. & Areschoug, T. (2005). Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. *Clin Microbiol Rev* 18, 102–127.
- Mahillon, J. & Chandler, M. (1998). Insertion sequences. *Microbiol Mol Biol Rev* 62, 725–774.
- Maquat, L. E. (1995). When cells stop making sense: effects of nonsense codons on RNA-metabolism in vertebrate cells. *RNA* 1, 453–465.
- Naess, A. I., Bevanger, L., Iversen, O.-J. & Maeland, J. A. (1991). Evaluation of monoclonal antibodies in serovar classification of group B streptococci (GBS). *Acta Pathol Microbiol Immunol Scand* 103, 731–736.
- Nagano, N., Nagano, Y. & Taguchi, F. (2002). High expression of a C protein β antigen gene among invasive strains from certain clonally related groups of type Ia and Ib group B streptococci. *Infect Immun* 70, 4643–4649.
- Pandiripally, V., Wei, L., Skerka, C., Zipfel, P. F. & Cue, D. (2003). Recruitment of complement factor H-like protein 1 promotes intracellular invasion by group A streptococci. *Infect Immun* 71, 7119–7128.
- Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. (1977). Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta 1H for cleavage of C3b and C4b in solution. *J Exp Med* 146, 257–270.
- Peltz, S. W., He, F., Welch, E. & Jacobson, A. (1994). Nonsense-mediated messenger-RNA decay in yeast. *Prog Nucleic Acid Res Mol Biol* 47, 271–298.
- Pleass, R. J., Areschoug, T., Lindahl, G. & Woof, J. M. (2001). Streptococcal IgA-binding proteins bind in the C α 2-C α 3 interdomain region and inhibit binding of IgA to human CD89. *J Biol Chem* 276, 8197–8204.
- Podglajen, I., Breuil, J., Casin, I. & Collatz, E. (1995). Genotypic identification of two groups within the species *Bacteroides fragilis* by ribotyping and by analysis of PCR-generated fragment patterns and insertion sequence content. *J Bacteriol* 177, 5270–5275.
- Ram, S., McQuillen, D. P., Gulati, S., Elkins, C., Pangburn, M. K. & Rice, P. A. (1998a). Binding of complement factor H to loop 5 of porin protein 1A: a molecular mechanism of serum resistance of nonsialylated *Neisseria gonorrhoeae*. *J Exp Med* 188, 671–680.
- Ram, S., Sharma, A. K., Simpson, S. D., Gulati, S., McQuillen, D. P., Pangburn, M. K. & Rice, P. A. (1998b). A novel sialic acid binding site on factor H mediates serum resistance of sialylated *Neisseria gonorrhoeae*. *J Exp Med* 187, 743–752.
- Rubens, C. E., Heggen, L. M. & Kuypers, J. M. (1989). IS861, a group B streptococcal insertion sequence related to IS150 and IS3 of *Escherichia coli*. *J Bacteriol* 171, 5531–5535.
- Ruizechevarria, M. J., Czapinski, K. & Peltz, S. W. (1996). Making sense of nonsense in yeast. *Trends Biochem Sci* 21, 433–438.
- Russell-Jones, G. J. & Gotschlich, E. C. (1984). Identification of protein antigens of group B streptococci, with special reference to the Ibc antigens. *J Exp Med* 160, 1476–1484.
- Sanchez-Beato, A. R., Garcia, E., Lopez, R. & Garcia, J. L. (1997). Identification and characterization of IS1381, a new insertion sequence in *Streptococcus pneumoniae*. *J Bacteriol* 179, 2459–2463.
- Sandhya, D., Carmicle, S., Hammerschmidt, S., Pangburn, M. K. & McDaniel, L. S. (2004). Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H1. *J Immunol* 173, 471–477.
- Schubert, A., Zakikhany, K., Schreiner, M., Frank, R., Spellerberg, B., Eikmanns, B. J. & Reinscheid, D. J. (2002). A fibrinogen receptor from group B *Streptococcus* interacts with fibrinogen by repetitive units with novel ligand binding sites. *Mol Microbiol* 46, 557–569.
- Spellerberg, B., Rozdzinski, E., Martin, S., Weber-Heynemann, J., Schnitzler, N., Lütticken, R. & Podbielski, A. (1999). Lmb, a protein with similarities to the Lral adhesin family, mediates attachment of *Streptococcus agalactiae* to human laminin. *Infect Immun* 67, 871–878.
- Spellerberg, B., Martin, S., Franken, C., Berner, C. & Lütticken, R. (2000). Identification of a novel insertion sequence element in *Streptococcus agalactiae*. *Gene* 241, 51–56.
- Tamura, G. S., Herndon, M., Prezekwas, J., Rubens, C. E., Ferrieri, P. & Hillier, S. L. (2000). Analysis of restriction fragment length polymorphisms of the insertion sequence IS1381 in group B streptococci. *J Infect Dis* 181, 364–368.
- van Egmond, M., van Garderen, E., van Spriël, A. B., Damen, C. A., van Amersfoort, E. S., van Zandbergen, G., van Hattum, J., Kuiper, J. & van de Winkel, J. G. (2000). Fc α RI-positive liver Kupffer cells: reappraisal of the function of immunoglobulin A in immunity. *Nat Med* 6, 680–685.

Weiler, J. M., Daha, M. R., Austen, K. F. & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein beta 1H. *Proc Natl Acad Sci U S A* **73**, 3268–3272.

Whaley, K. & Ruddy, S. (1976). Modulation of the alternative complement pathway by beta 1H globulin. *J Exp Med* **144**, 1147–1163.