

## Molecular Profiles of Group B Streptococcal Surface Protein Antigen Genes: Relationship to Molecular Serotypes

Fanrong Kong,<sup>1</sup> Sonia Gowan,<sup>2</sup> Diana Martin,<sup>2</sup> Gregory James,<sup>1</sup> and Gwendolyn L. Gilbert<sup>1\*</sup>

Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, Westmead, New South Wales, Australia,<sup>1</sup> and Institute of Environmental Science and Research, Porirua, Wellington, New Zealand<sup>2</sup>

Received 8 August 2001/Returned for modification 23 September 2001/Accepted 18 November 2001

**The study of surface protein antigens of group B streptococci (GBS) is important for understanding of the pathogenesis and epidemiology of infection, and several of these antigens have been proposed as components of GBS conjugate vaccines. In a previous study, we developed a novel PCR-and-sequencing system for identification of GBS serotypes and serosubtypes based on the capsular polysaccharide synthesis (*cps*) gene cluster. In this study, we used published sequences to develop PCR assays for identification of genes encoding GBS surface proteins including C alpha (*bca*), C alpha-like proteins 2 and 3 (*alp2* and *alp3*), Rib (*rib*), and C beta (*bac*). We showed that the prototype R reference strain, Prague 25/60, contained a novel alpha-like protein antigen gene (the proposed *alp4*), which presumably encodes an atypical, but antigenically similar, R-like protein. Initial evaluation of these gene-specific assays showed excellent specificity. By combining *cps* serotypes, serosubtypes, and surface protein gene profiles, we were able to divide 224 GBS isolates into 31 serovariants. GBS *bac*-positive strains could be further subtyped into 11 groups and 20 subgroups. Our results confirmed and extended reported associations between some *cps* serotypes and serosubtypes, on the one hand, and surface protein genes, on the other: serosubtypes III-1 and III-2 were associated with *rib*, serosubtype III-3 with *alp2*, serotype Ib with *bca* and *bac*, and serotype V with *alp3*. The associations between serotype Ia and *bca*, *bca* repetitive unit, and *bca* repetitive unit-like sequence-containing genes need to be studied further. These PCR-based methods will provide an alternative and objective tool for subtyping of GBS based on surface protein antigen genes.**

Group B streptococci (GBS)—*Streptococcus agalactiae*—are the commonest cause of neonatal and obstetric sepsis and an increasingly important cause of septicemia in the elderly and in immunocompromised patients (25). There are nine GBS capsular polysaccharide serotypes (based on the capsular polysaccharide synthesis [*cps*] gene cluster), which vary in their distribution among geographic areas, disease types, and patient age groups (5, 8). Capsular polysaccharides are important virulence factors and epidemiological markers and are the main components of conjugate vaccines. For studies of epidemiology and pathogenesis, it is important to identify as many phenotypic or molecular markers as possible to increase the discriminatory power of typing systems (6). In addition to capsular polysaccharide antigens, GBS surface protein antigens, which also contribute to the pathogenesis of GBS disease and induce protective immunity, are potentially useful markers (16). Their use in polysaccharide conjugate vaccines is under investigation (4, 17). Identification of surface protein antigens, combined with *cps* serotyping, allows subdivision of GBS strains into a large number of serovariants, which can facilitate epidemiological, pathogenetic, and other related studies of GBS infection (15).

The genes encoding the C alpha protein (*bca*), C alpha-like proteins 2 and 3 (*alp2* and *alp3*), and the Rib protein (*rib*) have been well studied, and their gene sequences have

been published in GenBank (16, 23, 29). They are members of a family of surface proteins containing repetitive elements, which produce variations in protein size and antigenicity (16, 29). The gene encoding the C beta or IgA binding protein (*bac*) has also been well described (7, 9). C alpha, C beta, and Rib proteins have all been proposed as potential vaccine components (17, 18, 22).

Numerous methods have been used to identify GBS surface antigens by using monoclonal (24) or polyclonal antibodies (3) or genes by hybridization with probes (27), PCR (19, 20, 21), and/or sequencing (16, 20). PCR-based methods are attractive because of their high discriminatory power and reproducibility (20). PCR methods to detect C alpha and C beta protein genes have been published (19, 20, 21), but the specificity, clinical application, and interpretation of these methods require further study. Specific PCR methods to identify genes encoding Rib and C alpha-like proteins 2 and 3, which are present in the more-virulent serotypes, III and V, have not yet been described (16, 29). Associations between *cps* serotypes and some protein antigens have been described (15, 27). They are likely to vary over time and in different populations and geographic locations (5, 8) and should be useful for studies of the epidemiology and pathogenesis of GBS infection.

In this study, we used published sequences of surface protein antigen genes, including *bca*, *bac*, *alp2*, *alp3*, and *rib* (7, 9, 16, 23, 29), to improve and/or develop protein gene-specific PCR assays. We used these assays to examine the distribution and variation of surface protein genes, and their associations with *cps* genes, in a large collection of GBS isolates collected over the past decade in Australia and New Zealand.

\* Corresponding author. Mailing address: Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Darcy Road, Westmead, New South Wales, 2145 Australia. Phone: (612) 9845 6255. Fax: (612) 9893 8659. E-mail: lyng@icpmr.wsahs.nsw.gov.au.

TABLE 1. Oligonucleotide primers used in this study

Primer reference	Target gene(s)	$T_m$ (°C) <sup>a</sup>	GenBank accession no.	Sequence <sup>b</sup>
IgAagGBS <sup>c</sup>	<i>bac</i>	73.8	X59771	<b>2663</b> GCGATTAAACAA CAA ACT ATT TTT GAT A TTG ACA ATG CAA <b>2702</b>
IgAS1 <sup>d</sup>	<i>bac</i>	72.8	X59771	<b>2765</b> GCT AAA TTT CAA AAA GGT CTA GAG ACA AAT ACG CCA <b>G2801</b>
IgAA1 <sup>d</sup>	<i>bac</i>	78.9	X59771	<b>3157</b> CCC ATC TGG TAA CTT CGG TGC ATC TGG AAG <b>C3127</b>
RigAagGBS <sup>c</sup>	<i>bac</i>	76.3	X59771	<b>3284</b> CAGCCAACCTCTTTC GTC GTT ACT TCC TTG AGA TGT AAC <b>3247</b>
GBS1360S <sup>e</sup>	<i>bac</i>	72.3	X59771	<b>1325</b> GTGAAATTGTAT AAG GCT ATG AGT GAG AGC TTG GAG <b>1360</b>
GBS1717S <sup>d</sup>	<i>bac</i>	75.0	X59771	<b>1685</b> ACA GTC ACA GCT AAA AGT GAT TCG AAG ACG ACG <b>1717</b>
GBS1937A <sup>e</sup>	<i>bac</i>	75.9	X59771	<b>1976</b> CCGTTTTAGAAATCTTT CTG CTC TGG TGT TTT AGG AAC TTG <b>1937</b>
BcaRUS <sup>f</sup>	<i>bca</i> repetitive unit	73.5	M97256	<b>769</b> GATAAATATGATCCAA CAG GAG GGG AAA CAA CAG TAC <b>805</b>
BcaRUA <sup>f</sup>	<i>bca</i> repetitive unit	77.2	M97256	<b>1003</b> CTGGTTTTGGTGTACAT GAA CCG TTA CTT CTA CTG TAT <b>CC963</b>
bcaS1 <sup>d</sup>	<i>bca</i> , <i>alp2</i> , <i>alp3</i>	71.7	M97256 and AF291065	<b>208/533</b> GGT AAT CTT AAT ATT TTT GAA GAG TCA ATA GTT GCT GCA TCT AC <b>251/576</b>
bcaS2 <sup>d</sup>	<i>bca</i> , <i>alp2</i> , <i>alp3</i>	78.0	M97256 and AF291065	<b>256/581</b> CCAGGGA GTG CAG CGA CCT TAA ATA CAA GCA TC <b>288/613</b>
balS <sup>d</sup>	<i>alp2</i> , <i>alp3</i>	73.8	AF291065	<b>677</b> GAT CCT CAA AAC CTC ATT GTA TTA AAT CCA TCA AGC TAT <b>TC717</b>
bcaA <sup>d</sup>	<i>bca</i>	74.2	M97256	<b>597</b> CGTTCTAAGT CTT CAA TCT TAT CCC TCA AGG TTG TTG <b>560</b>
balA <sup>d</sup>	<i>alp2</i> , <i>alp3</i>	73.6	AF291065	<b>978</b> CCA GTT AAG ACT TCA TCA CGA CTC CCA TCA <b>C948</b>
bal23S1 <sup>d</sup>	<i>alp2</i> , <i>alp3</i>	70.9	AF208158 and AF291065	<b>1093/1373</b> CAG ACT GTT AAA GTG GAT GAA GAT ATT ACC TTT ACG <b>G1129/1409</b>
bal23S2 <sup>d</sup>	<i>alp2</i> , <i>alp3</i>	72.9	AF208158 and AF291065	<b>1174/1454</b> CTT AAA GCT AAG TAT GAA AAT GAT ATC ATT GGA GCT CGT G <b>1213/1493</b>
bal2A1 <sup>d</sup>	<i>alp2</i>	78.3	AF208158	<b>1426</b> CGT GTT GTT CAA CAG TCC TAT GCT TAG CCT CTG GTG <b>1391</b>
bal2A2 <sup>d</sup>	<i>alp2</i>	70.8	AF208158	<b>1518</b> GGT ATC TGG TTT ATG ACC ATT TTT CCA GTT ATA CG <b>1484</b>
bal3A <sup>d</sup>	<i>alp3</i>	79.2	AF291065	<b>1693</b> GAC CGT TTG GTC CTT ACC TTT TGG TTC GTT GCT ATC <b>C1657</b>
ribS1 <sup>d,g</sup>	<i>rib</i>	65.2	U58333	<b>216</b> TAC AGA TAC TGT GTT TGC AGC TGA AG <b>241</b>
ribS2 <sup>d</sup>	<i>rib</i>	73.0	U58333	<b>238</b> GAAGTAATTTTCAG GAA GTG CTG TTA CGT TAA ACA CAA ATA TG <b>279</b>
ribA1 <sup>d</sup>	<i>rib</i>	78.8	U58333	<b>431</b> GAA GGT TGT GTG AAA TAA TTG CCG CCT TGC CTA ATG <b>396</b>
ribA2 <sup>d</sup>	<i>rib</i>	72.6	U58333	<b>462</b> AAT ACT AGC TGC ACC AAC AGT AGT CAA TTC AGA AGG <b>427</b>
ribA3 <sup>d,g</sup>	<i>rib</i>	61.3	U58333	<b>570</b> CAT CTA TTT TAT CTC TCA AAG CTG AAG <b>554</b>

<sup>a</sup> Primer  $T_m$  values were provided by the primer synthesizer (Sigma-Aldrich).

<sup>b</sup> Numbers (in boldface) represent base positions at which primer sequences start and finish, taking the start point(s) of the corresponding GenBank accession number(s) as 1. Underlined sequences are bases added to modify previously published primers.

<sup>c</sup> From reference 21.

<sup>d</sup> Primers designed by us for this study.

<sup>e</sup> From reference 19.

<sup>f</sup> From reference 20.

<sup>g</sup> For sequencing use only; not entirely specific for the *rib* gene.

## MATERIALS AND METHODS

**GBS isolates, serotyping, and serosubtyping.** The isolates used in this study and the serotype and serosubtype identification methods have been described in detail elsewhere (14). Isolates included well-characterized reference panels kindly provided by Lawrence Paoletti, Channing Laboratory, Boston, Mass. (serotypes Ia to VIII; reference panel 1), and Diana Martin, Streptococcus Reference Laboratory, Institute of Environmental Science and Research, Porirua, Wellington, New Zealand (serotypes Ia to VI; reference panel 2), and 206 clinical isolates. All isolates were serotyped by conventional and molecular

methods, and some were serosubtyped by PCR and sequencing. Antisera used for serotyping were prepared against serotypes Ia, Ib, Ic, and II through VIII and against the R protein antigen. The prototype R reference strain Prague 25/60 was used to raise the R antiserum.

**Oligonucleotide primers.** Oligonucleotide primers used in this study, their target sites in the gene sequences, and their melting temperatures ( $T_m$ ) are shown in Table 1. The primers were synthesized according to our specifications by Sigma-Aldrich (Castle Hill, New South Wales, Australia). Six previously published oligonucleotide primers (19, 20, 21) and a series of new primers

TABLE 2. Specificity and expected lengths of amplicons of using different primer pairs

Primer pair <sup>a</sup>	Specificity	Length of amplicon (bp)	Protein gene profile code(s) <sup>c</sup>
IgAagGBS-RIgAagGBS	<i>bac</i>	532–838	B
IgAS1-IgAA1	<i>bac</i>	303–591	B
GBS1360S-GBS1937A	<i>bac</i>	652	B
GBS1717S-GBS1937A	<i>bac</i>	292	B
bcaS1-bcaA	5' end of <i>bca</i>	390	A
bcaS2-bcaA	5' end of <i>bca</i>	342	A
BcaRUS-BcaRUA	<i>bca</i> repetitive unit, <i>bca</i> repetitive unit-like region	235	a or as
bcaS1-balA	<i>alp2</i> , <i>alp3</i>	446	<i>alp2</i> or <i>alp3</i>
bcaS2-balA	<i>alp2</i> , <i>alp3</i>	398	<i>alp2</i> or <i>alp3</i>
balS-balA	<i>alp2</i> , <i>alp3</i>	302	<i>alp2</i> or <i>alp3</i>
bal23S1-bal2A1	<i>alp2</i>	334	<i>alp2</i>
bal23S2-bal2A1	<i>alp2</i>	253	<i>alp2</i>
bal23S1-bal2A2	<i>alp2</i>	426	<i>alp2</i>
bal23S2-bal2A2	<i>alp2</i>	345	<i>alp2</i>
bal23S1-bal3A	<i>alp3</i>	321	<i>alp3</i>
bal23S2-bal3A	<i>alp3</i>	240	<i>alp3</i>
ribS1-ribA3 <sup>b</sup>	<i>rib</i>	355	R or <i>alp4</i>
ribS2-ribA1	<i>rib</i>	194	R
ribS2-ribA2	<i>rib</i>	225	R
ribS2-ribA3	<i>rib</i>	333	R

<sup>a</sup> See Table 1 for primer sequences. For primer pair ribS1-ribA3, the annealing temperature used was 55 to 60°C. For three previously described primer pairs (unmodified), annealing temperatures were as described previously (19, 20, 21), and for all other new primer pairs, they were 68 to 72°C (according to the  $T_m$ ).

<sup>b</sup> For sequencing use only; not entirely specific for the *rib* gene (see the text for more detail).

<sup>c</sup> See Table 4 footnotes for explanation of protein gene profile codes.

designed by us were used to sequence parts of genes encoding GBS surface proteins and/or to specifically amplify these genes. All new primers, except two used only for sequencing *rib* and six previously published (unmodified), were designed with high  $T_m$  (>70°C) for use in rapid-cycle PCR (Table 1).

**DNA preparations and PCR.** DNA was prepared from GBS cultures (21), and PCR was performed as previously described (11, 12, 14). Denaturation, annealing, and elongation temperatures and times used were 96°C for 1 s, 45 to 72°C (according to the primer  $T_m$  or as previously described) for 1 s, and 74°C for 1 to 30 s (according to the length of amplicons), respectively, for 30 to 35 cycles, using a Perkin-Elmer Thermal Cycler 9600. Ten microliters of PCR products was analyzed by electrophoresis on 1.5% agarose gels, which were stained with 0.5  $\mu$ g of ethidium bromide ml<sup>-1</sup>. For detection and/or subtyping, the presence of PCR amplicons of the expected lengths, shown by UV transillumination, was accepted as positive. For sequencing, 40- $\mu$ l volumes of PCR products were further purified by the polyethylene glycol precipitation method (1).

**Sequencing.** To confirm the specificity of newly designed or modified primer pairs, we sequenced 10, 13, and 10 selected amplicons produced by bcaS1-bcaA (targeting the 5' end of *bca*), ribS1-ribA3 (targeting *rib*), and GBS1360S-GBS1937A (targeting *bac*), respectively, from the two panels of reference strains and 31 randomly selected clinical isolates. All amplicons of primer pairs bcaS1-balA (targeting *alp2* and *alp3*), bal23S1-bal2A2 (targeting *alp2*), and IgAagGBS-RIgAagGBS (targeting *bac*) from any of the 224 isolates were sequenced.

PCR products were sequenced using Applied Biosystems Taq DyeDeoxy terminator cycle-sequencing kits according to standard protocols. The corresponding amplification primers or inner primers were used as the sequencing primers.

**Database similarity searching and sequence comparison.** Databases were searched for sequence similarity by using the FastA program in the SeqSearch program group. Sequences were compared using the Bestfit and Gap programs in the Comparison program group. The Translate program in the Translation program group was used to translate from DNA sequences to amino acid sequences. All programs are provided in WebANGIS (Australian National Genomic Information Service), version 3.

**Surface protein gene profile codes.** Each isolate was given a protein gene profile code according to positive PCR results using various primer pairs, as shown in Table 2.

**Nucleotide sequence accession numbers.** The sequences generated during this study have appeared in GenBank with the following accession numbers: AF367974 (partial *bac* sequence, with an insertion sequence, IS1381, from one isolate), AF362685 to AF362704 (partial *bac* sequences for all *bac*-positive isolates), and AF373214 (partial proposed new alpha-like protein 4 gene [*alp4*] for reference strain Prague 25/60, an R protein standard strain).

Previously published gene sequences used in this study and their GenBank accession numbers are as follows: M97256 (*bca*), X58470 and X59771 (*bac*),

U58333 (*rib*), AF208158 (*alp2*), AF291065 to AF291072 (*alp3*), and AF064785 (IS1381).

## RESULTS

**PCR results.** With few exceptions, all primer pairs produced amplicons of the predicted lengths from isolates giving positive results (Table 2). Figure 1 shows a representative gel. The exceptions included one isolate that was positive by PCR using primer pairs GBS1360S-GBS1937A and GBS1717S-GBS1937A (both targeting *bac*) but produced amplicons significantly longer than those of other *bac*-positive isolates. Sequencing showed that the amplicon contained the insertion sequence IS1381 with minor variations from the published sequences (28). The amplicons produced using primers IgAagGBS-RIgAagGBS and IgAS1-IgAA1 (also targeting *bac*) varied in length (2) and were sequenced for further subtyping (see below and Table 3).

### Evaluation of the protein gene-specific primer pairs by di-

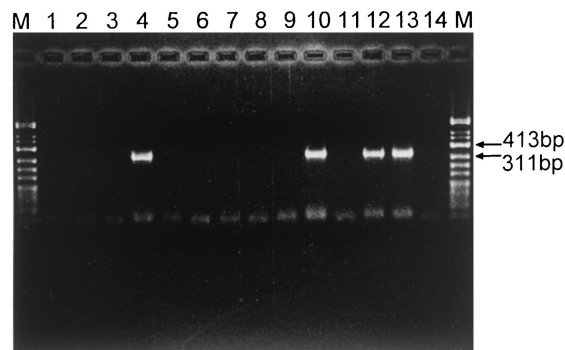


FIG. 1. Results of PCR amplification using the primer pair bcaS2-bcaA (targeting and specific for the 5' end of *bca*). Lanes M, molecular weight marker  $\phi$ X174 DNA/*Hin*II; lanes 1 to 7, seven serotype Ia isolates, one of which (lane 4) was positive for the 5' end of *bca*; lanes 8 to 14, seven serotype II isolates, three of which (lanes 10, 12, and 13) were positive for the 5' end of *bca*.

TABLE 3. Genetic groups and subgroups of *bac* (C beta protein gene) based on amplicon length (using primers IgAagGBS and RIgAagGBS) and sequence heterogeneity

Group or subgroup ( <i>n</i> )	Amplicon length (bp)	GenBank accession no.	No. of sites different from main group <sup>a</sup>	Molecular serotype(s) or serosubtypes <sup>b</sup> ( <i>n</i> )
B1 (19)	532	AF362685 <sup>c</sup>		Ib (17), II (2)
B1a (1)	532	AF362686	1 (B1)	Ib
B2 (3)	550	AF362687		Ib, II, III-4
B3 (2)	586	AF362688		Ib (2)
B3a (1)	586	AF362689	4 (B3)	V
B3b (1)	586	AF362690	21 (B3)	VI
B3c (1)	586	AF362691	24 (B3)	Ib
B4 (8)	604	AF362692		Ib (4), II (4)
B4a (1)	604	AF362693	1 (B4)	II
B4b (2)	604	AF362694	2 (B4)	Ib (2)
B5 (2)	622	AF362695 <sup>d</sup>		Ia, VI
B5a (1)	622	AF362696	2 (B5)	Ia
B6 (1)	640	AF362697		Ib
B7 (1)	658	AF362698		Ib
B7a (1)	658	AF362699	34 (B7)	VI
B8 (1)	712	AF362700		Ib
B9 (2)	748	AF362701		II (2)
B9a (1)	748	AF362702	13 (B9)	Ib
B10 (2)	820	AF362703		Ib (2)
B11 (1)	838	AF362704		Ib

<sup>a</sup> The main group is shown in parentheses.

<sup>b</sup> See Table 4 for further details of *cps* serotype and serosubtype relationships with protein antigen genes.

<sup>c</sup> The AF362685 sequence is identical with the corresponding portion of X58470.

<sup>d</sup> The AF362695 sequence is identical with the corresponding portion of X59771.

**rect sequencing of PCR amplicons.** All 10 amplicons of primer pair *bcaS1-bcaA* and 12 of 13 amplicons (except that from strain Prague 25/60 [see below]) of primer pair *ribS1-ribA3* were identical with the corresponding portions of the gene sequences in GenBank (M97256 [*bca*] and U58333 [*rib*], respectively). Four of 10 amplicons of primer pair GBS1360S-GBS1937A (targeting *bac*) were identical with the corresponding gene sequence in GenBank (X58470, X59771). A single point mutation (A to G at position 1441 of X59771) was found in the remaining six *bac*-positive amplicons, including the one that contained the insertion sequence *IS1381* (see above and GenBank accession number AF367974).

Fifty isolates produced amplicons with primer pair *bcaS1-balA*. The sequences of 9 were identical with the corresponding portion of the published sequence of *alp2* (AF208158), and those of 41 were identical with that of *alp3* (AF291065). There are two consistent heterogeneity sites between *alp2* and *alp3* in the sequences of *bcaS1-balA* amplicons, which can be used to distinguish them, in addition to *alp2*- and *alp3*-specific PCR. All nine amplicons obtained with primer pair *bal23S1-bal2A2* were identical with the corresponding portion of the *alp2* sequence in GenBank (AF208158). Primer pair IgAagGBS-RIgAagGBS identified *bac* in 52 isolates. There was considerable sequence variation, which allowed separation of *bac*-positive isolates into 11 groups and 20 subgroups based on amplicon length and sequence heterogeneity, respectively (Table 3). The groups contained small numbers (1 to 5) of isolates except for B1 (20 isolates, 2 subgroups) and B4 (11 isolates, 3 subgroups). In general, the presence or absence of short repetitive sequences was responsible for differences in amplicon length (2, 9).

**Further confirmation of specificity of surface protein gene-specific primer pairs.** To confirm primer pair specificity, we compared the results of PCR using the primer pairs we had designed or modified for *bac* PCR with those of PCR using

previously published primer pairs (19, 21) and found 100% correlation.

The previously reported nonspecificity of the published primer pair *bcaRUS-bcaRUA* (targeting the *bca* repetitive unit) was confirmed (20). When these primers were used, all nine *alp2*-positive (*BcaS1-BcaA*-negative) isolates and 53 isolates which were PCR negative with primer pairs *bcaS1-bcaA* and *bcaS2-bcaA* (targeting the 5' end of *bca*) and with primer pairs *bal23S1-bal2A2* and *bal23S2-bal2A1* (targeting the 5' end of *alp2*) produced amplicons. Our sequencing showed that *bca* and *alp2* have significant homology in the regions targeted by *bcaRUS-bcaRUA*, allowing amplicon formation from *alp2*-positive strains (16, 20). These false positive results could be due to the presence of other C alpha-like protein genes, containing regions homologous with the *bca* repetitive unit (*bca* repetitive unit-like sequence).

We also showed that the results of PCR using two or more primer pairs that we had designed for individual genes (*rib*, *alp2*, and *alp3*) correlated well, supporting the specificity of each set. The only exception, as mentioned above, was *ribS1-ribA3*, which produced a nonspecific amplicon from 1 of 224 isolates tested.

**Prague 25/60 contains another new alpha-like surface protein antigen gene, *alp4* (proposed).** Strain Prague 25/60 (which is used to raise the R antiserum), in reference panel 2, produced an amplicon with primer pair *ribS1-ribA3* but not with *ribS2-ribA1*, *ribS2-ribA2*, or *ribS2-ribA3*. It was therefore assumed not to contain *rib*, although the amplicon sequence showed considerable homology with *rib* and other members of the family of surface proteins (see below). This isolate was the only one, of 224 tested, for which PCRs were negative using *ribS2-ribA1* and *ribS2-ribA2* but positive using *ribS1-ribA3*. The latter primer pair is, then, not entirely specific for *rib* and was therefore used only for sequencing.

TABLE 4. Relationships between GBS protein gene profiles and capsular polysaccharide synthesis (*cps*) gene molecular serotypes and serosubtypes

Serotype or serosubtype <sup>a</sup> (n)	No. of isolates with surface protein gene profile <sup>b</sup> :									
	None	Aa	AaB	R	alp3	a	as	alp2as	RB	alp4a
Ia (43)			2			35	3	3		
Ib (37)		1	35		1					
II (29)		3	10	8	2	5				1 <sup>c</sup>
III-1 (30)				30						
III-2 (22)				22						
III-3 (5)								5		
III-4 (3)			1		1			1		
IV (9)				1		8				
V (38)	1			1	35				1	
VI (5)		1	3			1				
VII (1)					1					
VIII (2)	1				1					
Total (224)	2	5	51	62	41	49	3	9	1	1

<sup>a</sup> See the text for an explanation of *cps* serosubtypes (14).

<sup>b</sup> Protein antigen gene profile codes are as follows: A, 5' end of *bca* positive; a or as, *bca* repetitive unit or *bca* repetitive unit-like region positive, with multiple- or single-band amplicons, respectively; B, *bac* positive; R, *rib* positive; alp2, *alp2* positive; alp3, *alp3* positive.

<sup>c</sup> This isolate was the atypical reference strain Prague 25/60, which was nonserotypeable using capsular polysaccharide antisera but was assigned to molecular serotype II. It expresses a protein that is antigenically similar to R (and is used to generate an R antiserum), for which we propose the name "alpha-like protein 4," encoded by the "*alp4*" gene.

Sequencing of the Prague 25/60 ribS1-ribA3 amplicon showed considerable homology with other members of the surface protein gene family defined by *bca-rib*. The homologous regions were located at the 5' ends of the genes (for *bca*, the positions are in the region between nucleotides 251 and 559, and for the corresponding amino acid sequence, they are between positions 58 and 160 of the sequence identified by GenBank accession number M97256). The ratios of similarity (excluding the primer sequences) to DNA sequences of *bca*, *rib*, *alp2*, and *alp3* were 66.7, 64.4, 62.5, and 62.5%, respectively. For the corresponding amino acid sequences, similarity ratios were 62.1, 60.2, 57.3, and 57.3%, respectively. Since this amplicon sequence is most similar to that of *bca*, which encodes C alpha, the prototype of the surface protein family, we propose that the gene be named *alp4* (C alpha-like protein antigen 4 gene). Cloning and sequencing of the whole gene were beyond the scope of this study.

**Surface protein gene profiles.** For each GBS surface protein gene (except the *bca* repetitive unit and *bca* repetitive unit-like region), we selected two primer pairs to identify and characterize it by PCR. Four common profiles accounted for 203 of 224 (90.6%) isolates: R (62 isolates), AaB (51 isolates), a (49 isolates), and alp3 (41 isolates) (see Table 4). Only two isolates contained no surface protein gene markers. All but one isolate with the *bac* gene (B) also had *bca* with its repetitive unit (Aa); one had *rib* (R). All alp2 isolates contained single *bca* repetitive unit-like sequences (as). A, R, alp2, alp3, and the proposed new protein type alp4 were all mutually exclusive. Sixty-two of 63 isolates with *rib* (R) and 41 of 41 isolates with *alp3* had no other protein antigen gene markers.

**Relationship between surface protein antigen gene profiles and *cps* serotypes and serosubtypes.** Development of the molecular serotype (MS) identification method and comparison with conventional serotyping (CS) have been described elsewhere (14). A *cps* MS was assigned to all isolates, and the results correlated with CS results except for 19 of 224 isolates that were nontypeable by use of antisera. Relationships be-

tween surface protein gene profiles and *cps* molecular serotypes are summarized in Table 4.

The following strong associations were confirmed or demonstrated: MS Ia with *bca* repetitive unit or *bca* repetitive unit-like sequence (most with profile a), MS III-1 and III-2 with *rib*, MS III-3 with *alp2*, MS Ib with *bca* and *bac*, and MS V with *alp3*. MS II showed the most varied surface protein gene profiles. However, the relationships were not absolute, and different combinations of polysaccharide *cps* serotypes and protein gene profiles produced 31 serovariants, or 51 when *bac* (B) subgroups were considered.

**Relationship between surface protein antigens and protein gene profiles.** Based on CS, 33 isolates (belonging to CS Ia/c, Ib/c, IIc, IIB, IIIc, or IIIb) reacted with the C antiserum. The surface protein gene profiles of all of these isolates included *bca* (A) and/or *bca* repetitive unit-related (a or as) markers as follows: Aa (3 isolates), AaB (18 isolates), a (11 isolates), and alp2as (1 isolate). Twenty-nine isolates reacted with the R antiserum; of these, 22 contained *rib* and 6 contained *alp3*. The remaining isolate was Prague 25/60 (the reference strain used to raise the R protein antiserum), which contained the new presumed alpha-like protein 4 gene, *alp4* (see above).

## DISCUSSION

In our previous study, all the isolates used in the present study were serotyped by conventional and molecular methods that identified their *cps* serotypes and, in some cases, serosubtypes (14). In this study we developed PCR-based methods to identify GBS surface protein genes and further characterize these isolates. Using the published *bac* sequence, we modified *bac*-specific primers (19, 21) and designed new primers, with high  $T_m$  (>70°C), suitable for rapid-cycle PCR (10, 13) and targeting all major surface protein genes.

As previously reported, a published PCR primer pair targeting the *bca* repetitive unit (at the 3' end of *bca*) was not entirely specific for *bca* (20). We designed two new primer pairs tar-

getting the 5' end of *bca* in order to improve the specificity. However, very few serotype Ia strains gave positive results with these two primer pairs, whereas all were PCR positive using a primer pair targeting the *bca* repetitive unit (20). These results were consistent with a previous report (6) that a probe targeting the 5' end of *bca* hybridized with only one of nine serotype Ia strains, whereas a large *bca* probe, including the tandem repeat region, hybridized with all nine. Further study is required to define the sequences and specificities of different portions of *bca* and their effects, if any, on the structure and functions of C alpha and related proteins.

PCR primers specific for *rib*, *alp2*, and *alp3* have not been described previously. The primer pairs we designed mainly targeted the 5' ends of the genes and were chosen after comparison of their heterogeneity with related gene sequences (16, 29). We designed two or more primer pairs for each gene in order to check primer specificity by comparing the results of different PCRs targeting the same genes. Protein gene profiles *alp2* and *alp3* were distinguished on the basis of the *alp2*- and *alp3*-specific PCR and/or two sequence heterogeneity sites in the amplicons of *bcaS1-balA* or *bcaS2-balA*.

To confirm the specificities of our primers, we used them to examine two reference panels and selected GBS isolates. The longest amplicons produced by PCR for each gene were sequenced in order to provide maximal sequence information and ensure that the inner primers were not located at strain heterogeneity sites. Our sequencing results confirmed the specificities of the primers. Two pairs of primers for each gene were compared, with similar results. Finally, six gene- or region-specific primer pairs (including the one targeting the *bca* repetitive unit) were used to define protein antigen gene profiles for all 224 isolates.

The study showed that only one member of the surface protein gene family containing repetitive sequences—*rib*, *bca*, *alp2*, or *alp3* (or, presumably, *alp4*)—was present in any single isolate (15, 16). However, all isolates containing *bac*, which is not a member of the surface protein gene family containing repetitive sequences, also contained either *bca* (51 of 52 isolates) or *rib* (1 of 52 isolates) (15).

The C beta protein gene, *bac*, was present in 23% of isolates, a proportion similar to those (19 to 22%) previously reported (2). In common with others, we found variations in the *bac* (2) amplicons due to variable small internal repetitive sequences (2, 9) that, unlike those of the *bca-rib* family, were irregular. Their role is not clear, but they are potentially useful molecular markers for epidemiological studies (2, 7).

Our study confirmed previously reported relationships between *cps* serotypes and surface protein gene profiles (16). For example, some serotype III isolates (our MS III-1 and III-2) were closely associated with *rib* (26), and others (our MS III-3) were closely associated with *alp2* (16). Serotype Ib was associated with *bca* and *bac* (15), and serotype V was associated with *alp3* (16). However, as the relationship was not absolute, different combinations of *cps* serotypes and protein gene profiles identified many serovariants, which will be useful in epidemiological studies and in the formulation of conjugate vaccines (15, 16). Based on PCR only, we were able to divide our 224 isolates into 31 serovariants based on *bac* (B) groups or into 51 serovariants, based on subgroups. Theoretically, there are likely to be additional serovariants.

Comparison of protein antigen (C and R proteins) serotyping results with protein gene profiles showed that the presence of the gene does not necessarily indicate the expression of the corresponding protein. This is one reason for discrepancies between genetic and serotyping results; another is that C and R protein antisera are not entirely specific (16). Our analysis showed that reaction with the C antiserum generally correlated with the presence of genes encoding C alpha (*bca*) or alpha-like protein 2 (*alp2*). Reaction with the R antiserum correlated with the presence of genes encoding Rib protein (*rib*), alpha-like protein 3 (*alp3*), or the presumed new, rare alpha-like protein 4 (*alp4*) (found in this study). Apparently, antigenic cross-reactivity does not necessarily reflect genetic similarity, since the *alp4* sequence studied was more similar to that of *bca* than to that of *rib* or *alp3*. More-extensive analysis of these genes and the relationships between the proteins they encode is required.

These methods will be useful in further studies of the effects of various antigen profiles on virulence and to further define the genealogy of GBS serotypes and various subtypes.

#### ACKNOWLEDGMENTS

We thank Ansuia Sharma, Rebecca Hoile, Leanne Montgomery, and David Smith for their help in culturing GBS strains, Moana Ngatai and Julie Morgan for serotyping of some isolates, and Mark Wheeler for valuable help in sequencing.

#### REFERENCES

- Ahmet, Z., P. Stanier, D. Harvey, and D. Holt. 1999. New PCR primers for the sensitive detection and specific identification of group B beta-hemolytic streptococci in cerebrospinal fluid. *Mol. Cell. Probes* **13**:349–357.
- Berner, R., A. Bender, C. Rensing, J. Forster, and M. Brandis. 1999. Low prevalence of the immunoglobulin-A-binding beta antigen of the C protein among *Streptococcus agalactiae* isolates causing neonatal sepsis. *Eur. J. Clin. Microbiol. Infect. Dis.* **18**:545–550.
- Bevanger, L., A. I. Kvam, and J. A. Maeland. 1995. A *Streptococcus agalactiae* R protein analysed by polyclonal and monoclonal antibodies. *APMIS* **103**:731–736.
- Gravekamp, C., D. L. Kasper, L. C. Paoletti, and L. C. Madoff. 1999. Alpha C protein as a carrier for type III capsular polysaccharide and as a protective protein in group B streptococcal vaccines. *Infect. Immun.* **67**:2491–2496.
- Harrison, L. H., J. A. Elliott, D. M. Dwyer, J. P. Libonati, P. Ferrieri, L. Billmann, and A. Schuchat. 1998. Serotype distribution of invasive group B streptococcal isolates in Maryland: implications for vaccine formulation. *Maryland Emerging Infections Program. J. Infect. Dis.* **177**:998–1002.
- Hauge, M., C. Jespersgaard, K. Poulsen, and M. Kilian. 1996. Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease. *Infect. Immun.* **64**:919–925.
- Heden, L. O., E. Frithz, and G. Lindahl. 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.
- Hickman, M. E., M. A. Rench, P. Ferrieri, and C. J. Baker. 1999. Changing epidemiology of group B streptococcal colonization. *Pediatrics* **104**:203–209.
- Jerlstrom, P. G., G. S. Chhatwal, and K. N. Timmis. 1991. The IgA-binding beta 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.
- Ke, D., C. Menard, F. J. Picard, M. Boissinot, M. Ouellette, P. H. Roy, and M. G. Bergeron. 2000. Development of conventional and real-time PCR assays for the rapid detection of group B streptococci. *Clin. Chem.* **46**:324–331.
- Kong, F., X. Zhu, W. Wang, X. Zhou, S. Gordon, and G. L. Gilbert. 1999. Comparative analysis and serovar-specific identification of multiple-banded antigen genes of *Ureaplasma urealyticum* biovar 1. *J. Clin. Microbiol.* **37**:538–543.
- Kong, F., G. James, Z. Ma, S. Gordon, B. Wang, and G. L. Gilbert. 1999. Phylogenetic analysis of *Ureaplasma urealyticum*—support for the establishment of a new species, *Ureaplasma parvum*. *Int. J. Syst. Bacteriol.* **49**:1879–1889.

13. Kong, F., S. Gordon, and G. L. Gilbert. 2000. Rapid-cycle PCR for detection and typing of *Mycoplasma pneumoniae* in clinical specimens. *J. Clin. Microbiol.* **38**:4256–4259.
14. Kong, F., S. Gowan, D. Martin, G. James, and G. L. Gilbert. 2002. Serotype identification of group B streptococci by PCR and sequencing. *J. Clin. Microbiol.* **40**:216–226.
15. Kvam, A. I., A. Efstratiou, L. Bevanger, B. D. Cookson, I. F. Marticorena, R. C. George, and J. A. Maeland. 1995. Distribution of serovariants of group B streptococci in isolates from England and Norway. *J. Med. Microbiol.* **42**:246–250.
16. Lachenaier, C. S., R. Creti, J. L. Michel, and L. C. Madoff. 2000. Mosaicism in the alpha-like protein genes of group B streptococci. *Proc. Natl. Acad. Sci. USA* **97**:9630–9635.
17. Larsson, C., M. Stålhammar-Carlemalm, and G. Lindahl. 1996. Experimental vaccination against group B streptococcus, an encapsulated bacterium, with highly purified preparations of cell surface proteins Rib and  $\alpha$ . *Infect. Immun.* **64**:3518–3523.
18. Madoff, L. C., J. L. Michel, E. W. Gong, A. K. Rodewald, and D. L. Kasper. 1992. Protection of neonatal mice from group B streptococcal infection by maternal immunization with beta C protein. *Infect. Immun.* **60**:4989–4994.
19. Maeland, J. A., O. G. Brakstad, L. Bevanger, and A. I. Kvam. 1997. *Streptococcus agalactiae* beta gene and gene product variations. *J. Med. Microbiol.* **46**:999–1005.
20. Maeland, J. A., O. G. Brakstad, L. Bevanger, and S. Krokstad. 2000. Distribution and expression of *bca*, the gene encoding the c alpha protein, by *Streptococcus agalactiae*. *J. Med. Microbiol.* **49**:193–198.
21. Mawn, J. A., A. J. Simpson, and S. R. Heard. 1993. Detection of the C protein gene among group B streptococci using PCR. *J. Clin. Pathol.* **46**:633–636.
22. Michel, J. L., L. C. Madoff, D. E. Kling, D. L. Kasper, and F. M. Ausubel. 1991. Cloned alpha and beta C-protein antigens of group B streptococci elicit protective immunity. *Infect. Immun.* **59**:2023–2028.
23. Michel, J. L., L. C. Madoff, K. Olson, D. E. Kling, D. L. Kasper, and F. M. Ausubel. 1992. Large, identical, tandem repeating units in the C protein alpha antigen gene, *bca*, of group B streptococci. *Proc. Natl. Acad. Sci. USA* **89**:10060–10064.
24. Moyo, S. R., J. A. Maeland, and L. Bevanger. 1999. Comparison of three different methods in monoclonal antibody-based detection of *Streptococcus agalactiae* protein serotype markers. *APMIS* **107**:263–269.
25. Schuchat, A. 1999. Group B streptococcus. *Lancet* **353**:51–56.
26. Stalhammar-Carlemalm, M., L. Stenberg, and G. Lindahl. 1993. Protein rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. *J. Exp. Med.* **177**:1593–1603.
27. Suvorov, A., A. Dmitriev, I. Ustinovitch, C. Schalen, and A. A. Totolian. 1997. Molecular analysis of clinical group B streptococcal strains by use of alpha and beta gene probes. *FEMS Immunol. Med. Microbiol.* **17**:149–154.
28. Tamura, G. S., M. Herndon, J. Przekwas, C. E. Rubens, P. Ferrieri, and S. L. Hillier. 2000. Analysis of restriction fragment length polymorphisms of the insertion sequence *IS1381* in group B streptococci. *J. Infect. Dis.* **181**:364–368.
29. Wastfelt, M., M. Stalhammar-Carlemalm, A. M. Delisse, T. Cabezon, and G. Lindahl. 1996. Identification of a family of streptococcal surface proteins with extremely repetitive structure. *J. Biol. Chem.* **271**:18892–18897.