INFECTION AND IMMUNITY, May 2005, p. 3096–3103 Vol. 73, No. 5
0019-9567/05/$08.00+0 doi:10.1128/IAI.73.5.3096–3103.2005

Structural and Genetic Diversity of Group B *Streptococcus* Capsular Polysaccharides

Michael J. Cieslewicz,[1,2]† Donald Chaffin,[3]† Gustavo Glusman,[4]† Dennis Kasper,[1,2,5] Anup Madan,[4] Stephani Rodrigues,[4] Jessica Fahey,[4] Michael R. Wessels,[1,2,6] and Craig E. Rubens[3]*

Channing Laboratory, Department of Medicine, Brigham and Women's Hospital,[1] Division of Infectious Diseases, Children's Hospital Boston,[6] and Department of Microbiology and Molecular Genetics,[5] Harvard Medical School,[2] Boston, Massachusetts, and Institute for Systems Biology[4] and Division of Infectious Diseases, Children's Hospital, University of Washington,[3] Seattle, Washington

Received 20 August 2004/Returned for modification 16 November 2004/Accepted 24 December 2004

Group B *Streptococcus* (GBS) is an important pathogen of neonates, pregnant women, and immunocompromised individuals. GBS isolates associated with human infection produce one of nine antigenically distinct capsular polysaccharides which are thought to play a key role in virulence. A comparison of GBS polysaccharide structures of all nine known GBS serotypes together with the predicted amino acid sequences of the proteins that direct their synthesis suggests that the evolution of serotype-specific capsular polysaccharides has proceeded through en bloc replacement of individual glycosyltransferase genes with DNA sequences that encode enzymes with new linkage specificities. We found striking heterogeneity in amino acid sequences of synthetic enzymes with very similar functions, an observation that supports horizontal gene transfer rather than stepwise mutagenesis as a mechanism for capsule variation. Eight of the nine serotypes appear to be closely related both structurally and genetically, whereas serotype VIII is more distantly related. This similarity in polysaccharide structure strongly suggests that the evolutionary pressure toward antigenic variation exerted by acquired immunity is counterbalanced by a survival advantage conferred by conserved structural motifs of the GBS polysaccharides.

Surface-associated polysaccharides are common features of both gram-positive and gram-negative bacteria. It is thought that microorganisms may have evolved extracellular polysaccharides for protection against both environmental and host factors that may be detrimental to their survival. In pathogenic bacteria, these extracellular polysaccharides may allow the organism to survive within the host by masking antigenic determinants associated with the bacterial surface (13), by mimicking host antigens (11, 34), or by interfering with complement-mediated killing (6, 18).

The gram-positive pathogen *Streptococcus agalactiae* (group B *Streptococcus*, or GBS) is the most common cause of bacterial sepsis and meningitis among newborn infants in the United States. Like many other pathogenic bacteria, clinical isolates of GBS elaborate capsular polysaccharides with varied repeating unit structures. The evolution by a pathogen of antigenically distinct capsular types is thought to be driven by selective pressure imposed by host immunity (23). In the case of *Streptococcus pneumoniae*, for example, more than 90 distinct capsular types are known, and the development of immunity to certain serotypes has been shown to create a niche for replacement by other serotypes in the host (23). However, selective pressure to evolve new capsular types may be counterbalanced by a specific virulence benefit conferred by the conservation of a particular polysaccharide structure. For example, although multiple capsular types of *Haemophilus influenzae* are known, strains of capsular type b have a much higher potential to cause invasive human infection than strains of other serotypes.

The degree of variation in capsular polysaccharide structures among clinical isolates of GBS falls between these extremes. Nine capsular types are known, all of which are associated with human infection. The repeating unit structures of the nine types are similar in their constituent monosaccharide compositions and certain structural motifs, yet they differ sufficiently to be antigenically distinct. This limited diversity of GBS capsular polysaccharide structures offers an unusual opportunity to reconstruct the development of capsule diversity in an important human pathogen. It is likely that the genetic mechanisms found to underlie capsule variation in GBS reflect the same biologic forces that have shaped the evolution of many encapsulated pathogens.

In many encapsulated bacteria, genes conserved across diverse capsular serotypes flank genes encoding enzymes unique to a specific capsular serotype (36). In GBS, a central group of genes for serotype-specific glycosyltransferases and polymerases is flanked on one side by genes encoding enzymes that synthesize and activate sialic acid, the terminal sugar on the side chain of all nine serotypes (4, 10, 27). Flanking the other side of the serotype-specific glycosyltransferase genes is a group of genes which are hypothesized to function in the export of the polysaccharide capsule (4). We now report the DNA sequences of the genes required for capsule biosynthesis for five GBS serotypes that have not been previously described. We analyzed the gene sequences of the serotype-specific glycosyltransferase regions of all nine GBS capsule serotypes to-

* Corresponding author. Mailing address: Division of Infectious Diseases, Children's Hospital, University of Washington, Seattle, WA 98109. Phone: (206) 987-2073. Fax: (206) 987-7311. E-mail: craig.rubens@seattlechildrens.org.

† M.J.C., D.C., and G.G. contributed equally to this work.

gether with their polysaccharide repeat unit (PRU) structures to reconstruct their evolutionary relationships. The conserved nature of the nine serotype repeating unit structures in GBS, combined with the genetic diversity of the polysaccharide capsule biosynthesis clusters, provides a unique model system with which both the genetic and structural evolution of PRU structures can be analyzed.

MATERIALS AND METHODS

DNA sequencing. The capsule clusters for serotypes II, VII, and VIII were PCR amplified from genomic DNAs with primers that correspond to the common regions flanking the central serotype-specific glycosyltransferase region. For these PCRs, Hi-Fidelity Supermix (Invitrogen, Carlsbad, Calif.) was used according to the manufacturer's specifications. PCR products were sonicated and end filled with T4 DNA polymerase to create blunt ends. The fragmented DNAs were isolated in 1.0% agarose gels, and DNAs migrating between 1.0 and 1.5 kb were isolated. After gel extraction, purified DNAs were ligated into pUC18 and transformed into XL-10 Gold competent cells. Ninety-six clones from each library were sequenced with both M13 forward and M13 reverse primers and BIG Dye Terminator version 2 chemistry. Sequences were resolved on an ABI 3700 sequencer, and chromatograms were processed by the Phred base caller and assembled by Phrap.

The assemblies were examined for low-quality regions and problem areas and were resolved by specific oligonucleotide-directed sequencing of the PCR products.

MacVector (Accelrys Inc., San Diego, Calif.) was used to evaluate nucleotide sequences for open reading frames encoding ≥100 amino acids. Translated open reading frames were searched against the nonredundant protein database by the use of BLASTP (1).

The DNA sequences of the highly conserved 5′ and 3′ *cps* regions of serotypes IV, V, and VI were determined by the use of primers generated from the type III *cps* locus. Using these primers and a high-fidelity *Taq* polymerase (Roche Diagnostics, GmbH, Mannheim, Germany), we generated 0.5- to 2.0-kb PCR products. After treatment with exonuclease I and shrimp alkaline phosphatase (Promega, Madison, Wis.), the PCR products were directly sequenced by the use of internal primers, a BIG Dye sequencing kit, and an ABI 377 DNA sequencer. To determine the sequences of the nonhomologous glycosyltransferase regions of the serotypes, we performed long-range PCRs using primers generated from the 5′ and 3′ regions determined as described above. The PCR products were ligated into pGEM-T Easy (Promega), and the DNA inserts were sequenced via primer walking utilizing a BIG Dye sequencing kit. The sequences were assembled by the use of Sequencher (Gene Codes, Ann Arbor, Mich.), and open reading frames encoding >100 amino acids were found with the GCK program (SciQuest Inc., Research Triangle Park, N.C.). The translated open reading frames were screened for homologous sequences in GenBank by the use of BLASTP (1).

Gene names were assigned by using previously established nomenclature for polysaccharide-producing gram-positive bacteria (22, 36).

Sequence analysis. The complete sequences of all nine capsular serotypes were analyzed with the GESTALT Workbench program (8, 9). FASTY (25) was used to compare the genomic sequences to those of known capsular polysaccharide proteins. FASTY can model frameshifts in comparisons of DNAs with protein sequences, permitting reconstruction of the amino acid sequences produced prior to putative evolutionary frameshift mutations. Such conceptually translated sequences should give the most reliable results when attempting phylogenetic reconstructions.

The predicted protein sequences for the *cpsE*, *-F*, *-G*, *-H*, *-K*, *-L*, and *neuB* genes were aligned with ClustalW (12): an alignment was produced for each gene family, including available orthologs from all serotypes. In addition, a joint alignment of all CpsI, -J, -M, and -O genes, together with additional homologs extracted from public databases, was performed. Phylogenetic reconstruction was accomplished with the neighbor-joining algorithm (26) and qualified with 1,000 rounds of bootstrap analysis.

Nucleotide sequence accession numbers. The accession numbers for the serotype-specific sequences of the capsules of serotypes II, III, IV, V, VI, VII, and VIII are deposited in GenBank under accession numbers AY375362, AF163833, AF355776, AF349539, AF337958, AY376403, and AY375363, respectively.

RESULTS

The detailed structures of the PRUs of the nine GBS capsular serotypes have been described previously (5, 14–16, 19, 20, 29, 31, 33). The constituent monosaccharides included in the various capsular PRUs are conserved. Glucose (Glc*p*), galactose (Gal*p*), and *N*-acetylneuraminic acid (Neu*p*NAc) are found in all nine capsule serotypes. Except in types VI and VIII, *N*-acetylglucosamine (Glc*p*NAc) is also present. Rhamnose (Rha*p*) occurs only in the type VIII repeating unit.

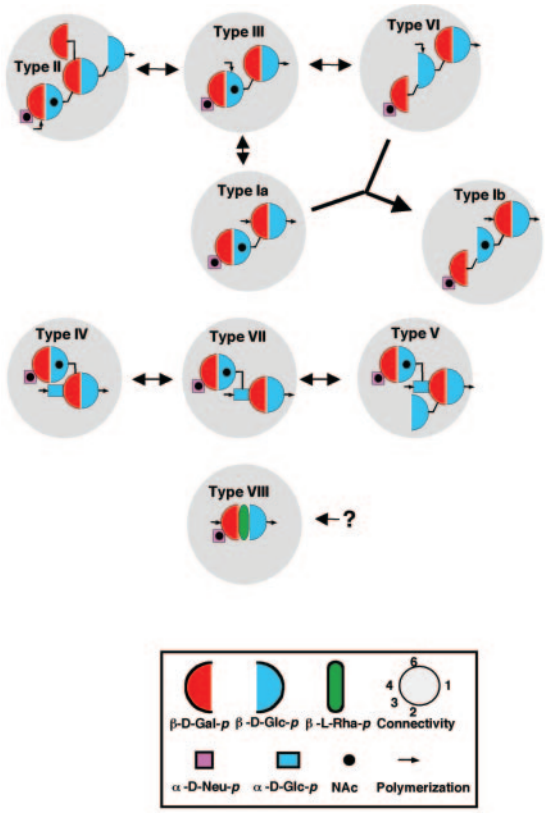

FIG. 1. PRU structures of the nine GBS serotypes. Arrows represent potential relationships between PRUs.

Analysis of recurring motifs within the PRU structures. To facilitate the analysis of the nine polysaccharide capsule repeating units and to elucidate pattern motifs and relationships, we redrew each repeating unit, with individual monosaccharides represented by simple shapes, the acetylation of sugars represented by a black dot within the shapes, and glycosidic linkages between individual monosaccharides represented by a circular assignment system (Fig. 1). While it is not meant to replace space-filling three-dimensional models of the PRUs, this simple method of drawing the PRU of each serotype highlights the structural motifs that are conserved between different serotypes.

An analysis of the nine capsule serotypes revealed two basic motifs. One motif, a disaccharide backbone of β-D-Gal*p*-(1→4)-β-D-Glc*p* (represented by a full circle, half red and half blue), is found in eight of the nine serotypes. A second motif, a variable trisaccharide side chain, consists of α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4) [or (1→3)]-β-D-Glc*p*NAc (or β-D-Glc*p*) and is represented in the figure by two half circles and a pink box. Serotype VIII is structurally unique in that it contains just a variable trisaccharide that is interrupted by a β-L-Rha*p*. The linkage between the variable trisaccharide and the disaccharide can further distinguish the PRUs. In five of the nine

serotypes (Ia, Ib, II, III, and VI), the variable trisaccharide and the disaccharide have a 1→3 linkage. In serotypes IV, V, and VII, an additional α-D-Glc*p* is linked 1→4 to the β-D-Gal*p* of the disaccharide and the 1→3 linkage between the disaccharide and the variable trisaccharide is altered to a 1→6 linkage.

The PRU structures of types Ia, Ib, II, III, and VI are all related by simple substitutions. The type Ia and III PRUs are nearly identical. Both contain the disaccharide and the variable trisaccharide connected by a 1→3 linkage (Fig. 1). The single difference between types Ia and III is the glycosidic bond that connects one repeating unit to the next: the type Ia units are linked 1→4 through the β-D-Gal*p* of the disaccharide, whereas the type III units are joined 1→6 through the β-D-Glc*p*NAc of the variable trisaccharide (Fig. 1).

The PRUs of types Ia and Ib differ only in the internal linkage of the variable trisaccharide, i.e., β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc for type Ia and β-D-Gal*p* (1→3)-β-D-Glc*p*NAc for type Ib (Fig. 1). Type VI, in contrast to type Ib, substitutes a β-D-Glc*p* for β-D-Glc*p*NAc in the variable trisaccharide (Fig. 1). However, the linkages between monosaccharides in the variable trisaccharides of types Ib and VI are the same. Structurally, it appears that type Ib is a combination of serotypes Ia and VI (Fig. 1). Linkages between individual PRUs are similarly conserved among these four serotypes: PRUs are linked β-D-Glc*p*-(1→4)-β-D-Gal*p* for types Ia and Ib, β-D-Glc*p*-(1→6)-β-D-Glc*p*NAc for type III, and β-D-Glc*p*-(1→6)-β-D-Glc*p* for type VI. Within the type II PRU, both the disaccharide and the variable trisaccharide are found in the type III repeating unit (Fig. 1). However, the type II repeating unit contains two additional sugars that are not found in type III: a β-D-Gal*p* linked 1→6 to the β-D-Gal*p* of the variable trisaccharide and a second β-D-Glc*p* linked 1→3 to the β-D-Glc*p* in the disaccharide (Fig. 1). Furthermore, type II PRUs are polymerized 1→2 to the β-D-Glc*p* of the variable trisaccharide.

Type IV, V, and VII PRU structures are interrelated. The PRUs of serotypes IV, V, and VII are all related by simple structural differences (Fig. 1). Type IV and VII have the same core group of sugars in their respective PRUs. In type IV, the trisaccharide side chain of Neu*p*NAc-Gal*p*-Glc*p*NAc is linked 1→6 to β-D-Gal of the disaccharide. However, in type VII, the same trisaccharide side chain is linked 1→6 to α-D-Glc. The addition of a single sugar, β-D-Glc*p*, differentiates type V from type VII (Fig. 1).

Sequence analysis of GBS capsule biosynthetic clusters. The DNA sequences of the variable serotype-specific regions from each of the nine GBS capsule clusters have been determined and have the following accession numbers: Ia, AB028896 (35); Ib, AB050723 (30); II, AY375362 (this study); III, AF163833 (2); IV, AF355776 (this study); V, AF349539 (28; this study); VI, AF337958 (this study); VII, AY376403 (this study); and VIII, AY375363 (this study) (Fig. 2A). Previously published studies have defined the functions of CpsI and CpsJ in the synthesis of type Ia and type Ib polysaccharide repeat units (30, 35). In addition, it has been shown that CpsH is required for the synthesis of linkages between the PRUs of types Ia and III (2). Using a combination of previously published data and BLAST homology searches, we assigned functions to the glycosyltransferases encoded by the variable gene regions of each of the nine serotypes (Fig. 3). For this study, we have used gene-naming conventions which were established previously for other polysaccharide-producing gram-positive bacteria (22, 36). To avoid confusion between gene name and serotype designations, we used Arabic numerals to identify serotype-specific genes. For example, *cps1aI* designates the *cpsI* gene from serotype Ia.

Despite the high degree of similarity between the repeating unit structures of the different capsular types, an analysis of the gene sequences revealed significant diversity at both the nucleotide and amino acid sequence levels (see Supplemental Table 1 at http://www.childrenshospital.org/cfapps/research/data_admin/Site381/mainpageS381P0.html). Furthermore, the patterns of similarity and divergence in the capsule gene sequences demonstrate striking parallels to the observed patterns of similarity and divergence in repeating unit structures of the polysaccharides. This sequence diversity in the glycosyltransferase region contrasts with the case of the common genes flanking either side of the variable trisaccharide regions (*cpsABCDEF* for types Ia to VII, *cpsABCD* for type VIII, and *cpsL neuBCDA*), which contain very little variability at the nucleotide level (Fig. 2A). A closer examination of the capsular biosynthetic cluster revealed two sections of *cpsG* in which the nucleotide sequences of serotypes Ia, Ib, II, III, IV, V, VI, and VII diverge (Fig. 2B, points 1 and 2). At position 236, serotypes II, III, and VI diverge from the other serotypes (Fig. 2B, point 1). At position 322, types Ia and Ib diverge from serotypes IV, V, and VII (Fig. 2B, point 2). Serotype VIII does not have a homolog of *cpsG*. The strict conservation of the amino-terminal region of CpsG suggests that the invariant structure of this portion of the protein is required for a non-serotype-specific function in the synthesis of GBS PRUs. However, the C-terminal region is more varied and may have a serotype-specific function. For the synthesis of the *S. pneumoniae* type 14 PRU, CpsG exhibits a 1→4 galactosyltransferase activity (21). However, the precise function of CpsG in the synthesis of GBS PRUs has not been demonstrated experimentally. It has been shown that CpsG is not required for the linkage specificity of individual PRUs in type Ia or III (2). The high degree of sequence similarity between the amino-terminal regions of the CpsG proteins from GBS and type 14 *S. pneumoniae* (>60%) (data not shown) suggests that CpsG functions as a 1→4 galactosyltransferase, attaching β-D-Gal*p* to β-D-Glc*p* in the backbone of the GBS PRUs of types Ia, Ib, and II to VII.

In contrast to *cpsG*, the *cpsK* genes from the nine serotypes are divergent at their 5′ ends and have several points where the nucleotide sequences converge (Fig. 2C, points 3 to 5). The *cpsK* genes from serotypes Ia, III, and V are similar (>98% similarity at the nucleotide level). In addition, *cpsK* from serotype VII is 88% identical to *cps1aK*. However, *cps1bK*, *cps4K*, and *cps6K* are divergent at their 5′ ends. Beginning at nucleotide position 328, *cps4K* becomes identical to *cpsK* from serotypes Ia, III, V, and VII (Fig. 2C, point 3). At nucleotide position 752, *cps1bK* and *cps6K* become nearly identical to the genes from the other serotypes (Fig. 2C, point 4). In contrast to *cpsK* from the other serotypes, *cps2K* is 100% identical to *cps1aK* over the first 199 nucleotides, at which point the nucleotide sequences diverge to less than statistical significance, as determined by a pairwise BLAST search. Yet the *cpsK* genes from serotypes Ia and II retain some primary amino acid sequence similarity. At nucleotide position 912, *cps2K* becomes identical to *cpsK* from serotypes Ia, Ib, III, IV, V, VI, and VII (Fig. 2C, point 5). The sialyltransferase gene of serotype VIII,

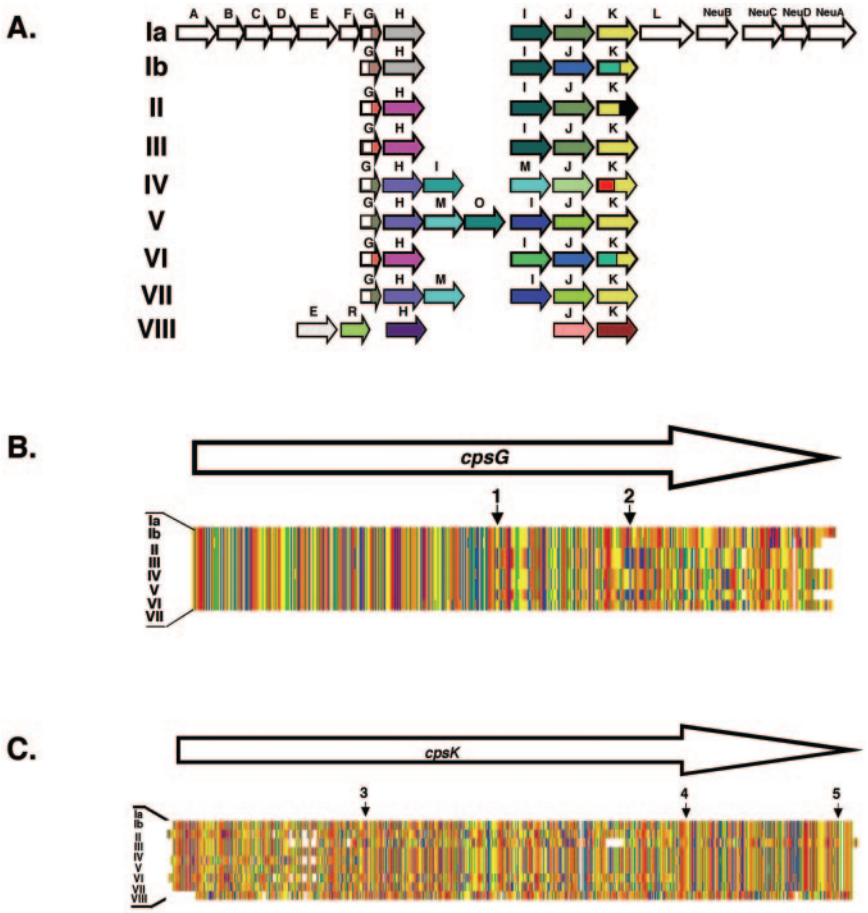

FIG. 2. Sequence analysis of all nine GBS capsular serotypes. (A) CpsA to -E and CpsL as well as NeuB, -D, -A, and -C are conserved in all nine serotypes and are shown only in the type Ia capsule cluster. The color inside each arrow indicates the degree of similarity of the amino acid sequence to those encoded by other open reading frames. A gap was introduced between CpsH and CpsI to permit an alignment of corresponding open reading frames. (B) Alignment of *cpsG* genes from serotypes Ia, Ib, II, III, IV, V, VI, and VII. (C) Alignment of *cpsK* genes from all nine serotypes. Each column of color in panels B and C represents 1 base of the nucleotide sequence. Red, thymine; green, guanine; yellow, adenine; blue, cytosine. Arrows in panels B and C denote regions where nucleotide sequences diverge or converge (see the text for more details).

cps8K, is not related to those of the other GBS serotypes, but it is homologous to lipooligosaccharide sialyltransferases from *Pasteurella multocida*, *H. influenzae*, *Neisseria meningitidis*, and *Neisseria gonorrhoeae*.

PRUs of types Ia, Ib, III, and VI have genetic diversity through horizontal gene transfer. An examination of the polysaccharide capsule clusters of types Ia, Ib, III, and VI showed that the genetic diversity among these serotypes mimics the structural relationship between PRUs. Types Ia and III differ significantly in only two genes, namely, the last 60 codons of *cpsG* and all of *cpsH* (Fig. 2; see Supplemental Table 1c at http://www.childrenshospital.org/cfapps/research/data_admin/Site381/mainpageS381P0.html). CpsH has been predicted to be a repeating unit polymerase (2). Genetically, the type Ia serotype may have arisen by horizontal gene transfer (HGT) when the segment of DNA encoding enzymes (CpsI, -J, and -K) that synthesize linkages common to both serotypes Ia and III recombined with a segment of DNA from an unknown source, replacing the 3′ terminus of *cps3G* and all of *cps3H* (Fig. 4). Alternatively, serotype III could have arisen from Ia by the same mechanism.

Two features distinguish the type III and type VI PRU structures (Fig. 1). The first difference is that β-D-Glc*p* in the type VI variable trisaccharide replaces β-D-Glc*p*NAc in the variable trisaccharide of type III. Cps1aI, which is identical to Cps3I, is a 1→3 *N*-acetylglucosaminyltransferase which functions to attach β-D-Glc*p*NAc in the type Ia variable trisaccharide to β-D-Gal*p* of the constant disaccharide (35). We predict that Cps6I, which is a paralog (59% identity and 84% similarity) of both Cps1aI and Cps3I, functions as a 1→3 glucosyltransferase, attaching the β-D-Glc*p* in the variable trisaccharide to the β-D-Gal*p* of the disaccharide. The second difference between the type III and VI PRUs is in the linkage between the sugars of the variable trisaccharides. Type III has a 1→4 linkage between β-D-Gal*p* and β-D-Glc*p*NAc of the variable trisaccharide, while type VI has a 1→3 linkage between β-D-Gal*p* and β-D-Glc*p* (Fig. 1). Cps3J is identical to Cps1aJ, which has been shown to be a 1→4 galactosyltransferase that transfers β-D-Gal*p* to β-D-Glc*p*NAc in the variable trisaccharide (35). We predict that Cps6J functions as a 1→3 galactosyltransferase, attaching β-D-Gal*p* to β-D-Glc*p*. The sequence divergence observed between Cps3J and Cps6J (16% identity and

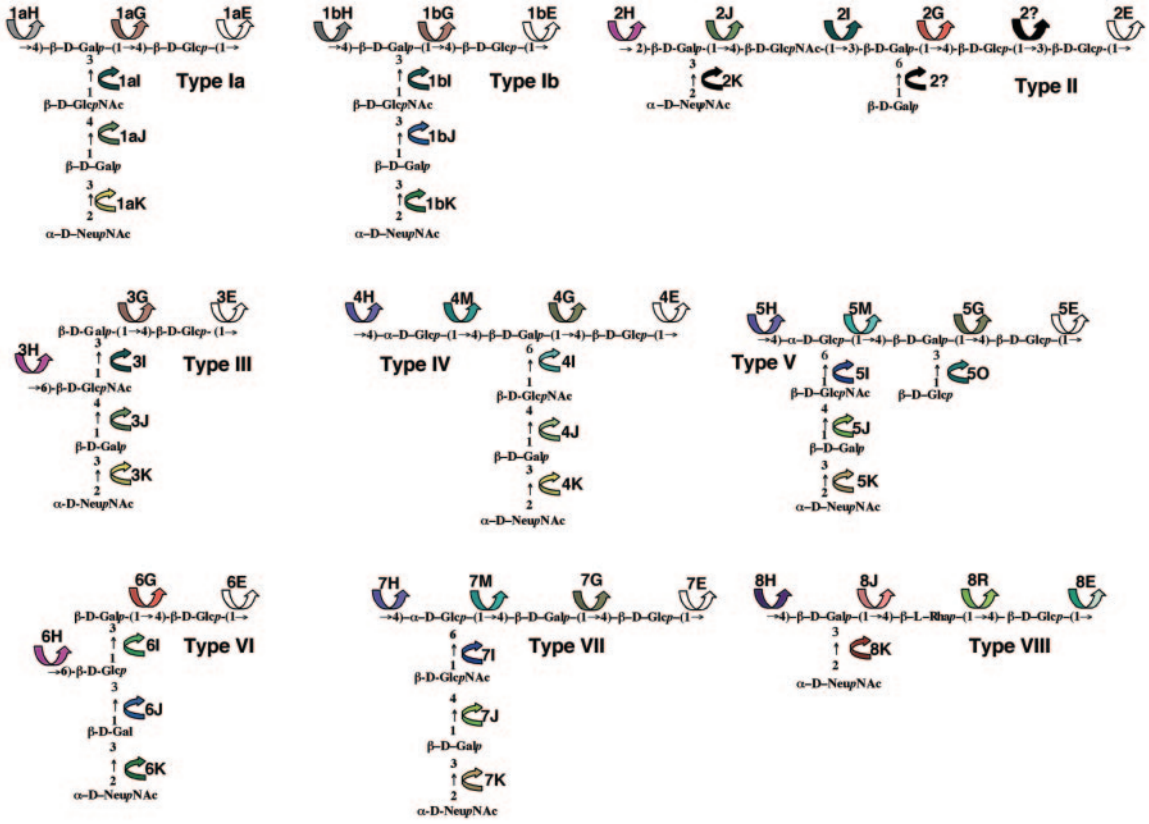

FIG. 3. Assignment of gene function on the basis of previously published results and sequence comparisons. The colors of the arrows have the same meanings as those in Fig. 2. Question marks denote two required glycosyltransferases with proposed functions. However, the genes encoding these enzymes have not yet been identified.

50% similarity) most likely results in both a new recipient sugar substrate for the galactosyltransferase and an altered linkage between β-D-Gal*p* and β-D-Glc*p* (Fig. 3). Cps3K has been shown to be a sialyltransferase (3). We predict that Cps6K also functions as a sialyltransferase, attaching α-D-Neu*p*NAc to β-D-Gal*p* in the variable trisaccharide. We hypothesize that the structural divergence of the type VI polysaccharide compared to the type III polysaccharide is most likely due to HGT of a segment of DNA encoding CpsJ and part of CpsK, which are enzymes that synthesize linkages unique to type VI, with DNA encoding CpsG, -H, and -I from type III, which are enzymes that synthesize linkages common to serotypes III and VI (Fig. 4).

Structurally and genetically, the type Ib variable trisaccharide has features which are found in the variable trisaccharides of both type Ia and type VI. The Ib variable trisaccharide has the same monosaccharides as Ia, but the sugars are linked as in type VI (Fig. 1). Genetically, it appears that Cps1aH (repeating unit polymerase [2]) and Cps1aI (1→3 *N*-acetylglucosaminyltransferase [35]) recombined with Cps6J and Cps6K (Fig. 4). Cps1bJ has 1→3 galactosyltransferase activity (30), which accounts for the β-D-Gal*p*-1→3-β-D-Glc*p* in the variable trisaccharide of type VI. Cps6J and Cps1bJ are 53% identical and 83% similar, with large blocks of identities occurring in the amino terminal two-thirds of the aligned proteins (data not shown). The lack of amino acid identity in the carboxyl-terminal third of the two proteins suggests that this region may encode specificity for the acceptor molecules, since the acceptor molecule for Cps1bJ is GlcNAc and that for Cps6J is Glc (Fig. 3). It is intriguing to speculate that when the proposed recombination occurred between types Ia and VI, the acceptor molecule for both enzymes was Glc, and that subsequent genetic drift or selective pressure caused a shift from Glc to GlcNAc as an acceptor molecule for serotype Cps1bJ.

Type II PRU is genetically related to type III PRU. Surprisingly, the presumed PRU polymerase, Cps2H, is identical to Cps3H, although the linkages between the type II and III PRUs are distinct (Fig. 1). Furthermore, the type II PRU includes six internal glycosidic linkages, but only four putative glycosyltransferase genes are found in the capsule locus (Fig. 2A). These discrepancies suggest that the PRU polymerase,

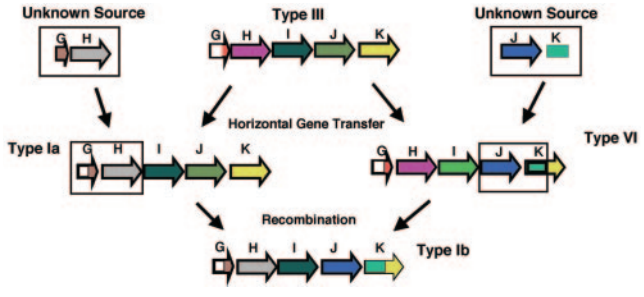

FIG. 4. Model of genetic recombination among serotypes Ia, Ib, III, and VI. Boxed genes denote sections that have recombined.

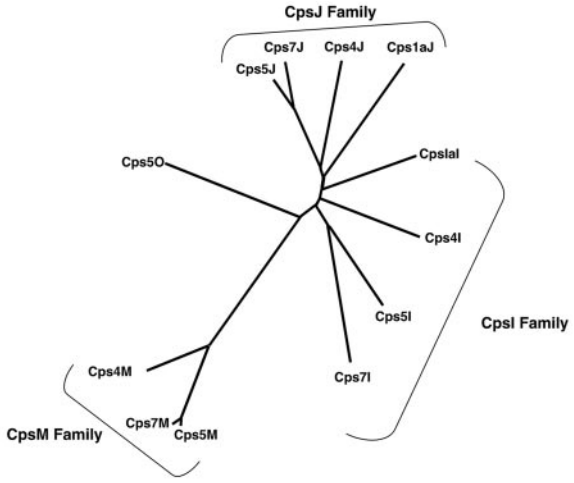

FIG. 5. Phylogenetic reconstruction of glycosyltransferases from serotypes Ia, IV, V, and VII.

CpsH, and some glycosyltransferases may have dual functions or that additional capsule synthesis genes may be located elsewhere in the type II GBS genome. It is also possible that the reported structure of the type II polysaccharide is in error.

Genetic diversity of serotypes IV, V, and VII. The genetic loci required for the synthesis of the PRUs of types IV, V, and VII contain several conserved genes as well as a family of paralogous genes (Fig. 2A). Three genes—*cpsG*, encoding a putative glycosyltransferase (21); *cpsH*, encoding a subunit polymerase (2); and *cpsK*, encoding a sialyltransferase (3) (Fig. 2A)—are conserved in the serotype-specific regions of these three serotypes. Between *cpsK* and *cpsH*, there is a family of paralogous genes (*cps4I*, *-4J*, *-4 M*, *-5I*, *-5J*, *-5 M*, *-5O*, *-7I*, *-7J*, and *-7 M*) that are predicted to encode 10 glycosyltransferases. Multiple sequence alignment and phylogenetic reconstruction of the predicted amino acid sequences of these 10 glycosyltransferases and of Cps1aI and Cps1aJ revealed three lineages (Fig. 5). The CpsJ lineage consists of Cps1aJ, Cps4J, Cps5J, and Cps7J. Cps1aJ is predicted to be a β-D-galactosyltransferase which attaches β-D-Gal 1→4 to β-D-GlcNAc in the variable trisaccharide (35). This linkage in the variable trisaccharide is common to serotypes Ia, IV, V, and VII. Therefore, the close branching of Cps1aJ, -4J, -5J, and -7J was an expected outcome of the phylogenetic reconstruction. A second lineage, the CpsI lineage, consists of Cps1aI, -4I, -5I, and -7I. Cps1aI is a β-D-glucosaminyltransferase attaching β-D-GlcNAc 1→3 of the variable trisaccharide to β-D-Gal of the disaccharide. In serotypes IV, V, and VII, the corresponding β-D-GlcNAc of the variable trisaccharide is linked 1→6 to β-D-Gal, in contrast to the 1→3 linkage in type Ia, and in types V and VII the acceptor molecule is an α-D-Glc, in contrast to the β-D-Gal in types Ia and IV. These differences in glycosyltransferase functions correlate with the independent branching of Cps4I, -5I, and -7I. Serotypes IV, V, and VII all contain an additional monosaccharide, α-D-Glc. The distant relationship between Cps4 M, 5 M, and 7 M and the CpsI and CpsJ families suggests that CpsM is the (1→4) α-D-glucosyltransferase attaching α-D-Glc to the β-D-Gal found in the disaccharide (Fig. 3). This linkage is unique to serotypes IV, V, and VII. Finally, Cps5O branches near the CpsI family. However, it is unclear if Cps5O is in a family of its own or is a distant member of the CpsI family. Although not shown here, Cps1bI, -2I, -3I, and -6I and Cps1bJ, -2J, -3J, and -6J group phylogenetically with the CpsI and CpsJ families, respectively (data not shown).

The simplest model to account for the genetic diversity among serotypes IV, V, and VII is that segments of DNA, including *cps4G-cps4J*, *cps7G-cps7J*, and *cps5G-cps5J*, were each acquired by GBS in separate HGT events. This model accounts for the low level of similarity between CpsH proteins from serotypes IV, V, and VII (<50%) (see Supplemental Table 1c at http://www.childrenshospital.org/cfapps/research/data_admin/Site381/mainpageS381P0.html) and any of the other GBS serotypes. However, a closer examination of the phylogenetic reconstruction reveals that Cps5 M and -7 M branch together, as do Cps5I and -7I and Cps5J and -7J (Fig. 5), suggesting that they are more closely related to each other than to glycosyltransferases in serotype IV. A second model to account for these data is that *cps4G-4I* and *cps7G-7I* were acquired in separate HGT events by GBS. Subsequently, Cps5O was either acquired by HGT into the type VII capsule cluster or *cps7I* was duplicated and acquired a new enzyme activity. The second model takes into account both the lack of similarity between CpsG and -H of serotypes IV, V, and VII and those of serotypes Ia, Ib, II, III, and VI and the close branching of Cps5 M and -7 M and of Cps5I and -7I.

The type VIII capsule biosynthesis gene cluster is genetically unique to GBS. As predicted by a comparison of the polysaccharide structures, the type VIII biosynthesis cluster is unique to GBS (Fig. 2). Unlike the other serotypes, it contains Rha*p*. In addition, the type VIII-specific glycosyltransferases and polymerases (Cps8E, Cps8H, Cps8J, and Cps8K) appear to be only distantly related to the glycosyltransferases of the other eight serotypes (see Supplemental Tables 1a and 1b at http://www.childrenshospital.org/cfapps/research/data_admin/Site381/mainpageS381P0.html). Orthologs of these open reading frames are found in type 23f *S. pneumoniae*, which elaborates a Rha*p*-containing capsular polysaccharide (17). An examination of the PRUs of GBS type VIII and *S. pneumoniae* type 23f revealed that they are structurally similar, as both have an identical backbone of (1→4)-β-D-Gal-(1→4)-β-L-Rha*p*-(1→4)-β-D-Glc*p*. However, like the other GBS serotypes, type VIII has α-D-Neu*p*NAc attached to β-D-Gal*p*, in contrast to type 23f, which has α-L-Rha*p* attached to β-D-Gal*p*. In addition, type 23f has a glycerol phosphate attached to β-D-Gal*p*. The combination of the structural and genetic similarities between these two serotypes suggests that GBS acquired a DNA segment encoding Cps8ERHJ by HGT. However, it is not clear whether this acquisition was between type VIII and type 23f or whether both GBS and *S. pneumoniae* acquired this DNA segment from another source.

Phylogenetic analysis of sialyltransferases. Cps3K is a sialyltransferase that links α-D-Neu*p*NAc to β-D-Gal in the variable trisaccharide (Fig. 3) (3). Despite the strict conservation of the α-D-Neu*p*NAc-(2→3)-β-D-Gal motif in all GBS serotypes, multiple sequence alignment and phylogenetic reconstruction of the sialyltransferases (Cps1aK, -1bK, -2K, -3K, -4K, -5K, -6K, -7K, and -8K) revealed significant diversity among them (Fig. 6; see Supplemental Table 1c at http://www.childrenshospital.org/cfapps/research/data_admin/Site381/mainpageS381P0.html). The sialyltransferases cluster into dis-

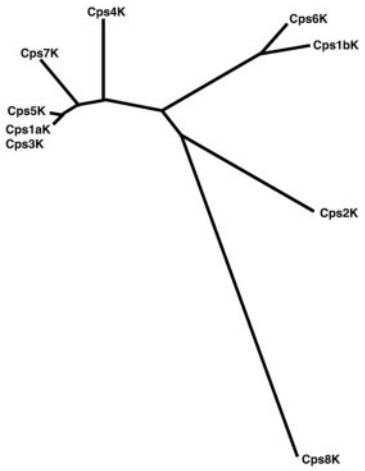

FIG. 6. Phylogenetic reconstruction of CpsK genes of all nine serotypes.

tinct groups that correlate with the linkage between the β-D-Gal and β-D-Glc*p*NAc moieties in the variable trisaccharide. Types Ia, III, IV, V, and VII contain a 1→4 linkage between β-D-Gal and β-D-Glc*p*NAc, and CpsKs from these five serotypes all branch to a common root. Types Ib and VI, which have a 1→3 linkage between β-D-Gal and β-D-Glc*p*NAc, have CpsKs that group together. Types II and VIII, which both have structurally unique PRUs among the nine capsule serotypes (Fig. 1), have sialyltransferase genes that are distantly related to the other CpsKs (Fig. 6). This pattern suggests that the specificities of the sialyltransferases depend not only on recognition of the conserved galactose residue, but also on the other sugar(s) to which the galactose is linked.

DISCUSSION

The present investigation provides evidence that capsular polysaccharide diversity in GBS has been driven by HGT, by which new segments of DNA from non-GBS sources have integrated into the GBS capsule cluster, and by interserotype recombination events, by which gene segments have been exchanged between GBS capsule serotypes. The acquisition of new genes from other species might be considered unlikely, as GBS is not known to be naturally competent for the acquisition of foreign DNA. However, analyses of the GBS genome and comparative genome hybridization studies have revealed a surprising degree of genetic heterogeneity among GBS isolates (7, 28). Regions of diversity are often flanked by transposons, insertion sequences, or bacteriophage remnants that probably represent vestiges of mobile genetic elements that mediated the acquisition of DNA from other species and the transfer of genes among GBS strains (28).

It has been suggested that variation in capsular polysaccharide structures may confer a survival advantage for encapsulated bacteria in the face of acquired immune responses by the host. For GBS, the capsular polysaccharide plays a central role in resistance to complement-mediated opsonophagocytic killing by human blood leukocytes. However, the capsule also serves as a target for specific antibodies that can defeat the antiphagocytic properties of the capsule, thereby conferring on the host protective immunity to GBS infection. Such serotype-specific immune pressure may have contributed to the emergence of new capsule types.

The restricted diversity of GBS capsule structures compared with that for *S. pneumoniae* might reflect, in part, the greater capacity of the latter species to acquire foreign DNA through natural competence. It is also possible that the conservation of particular structural motifs of the GBS capsular polysaccharides confers a survival advantage in human and/or animal hosts. The presence of terminal sialic acid residues on the GBS type III polysaccharide has been shown to inhibit the activation of the alternative complement pathway, thereby conferring bacterial resistance to phagocytosis in nonimmune plasma (6, 24). The striking conservation of α-D-Neu*p*NAc (2→3)-β-D-Gal*p* among all known GBS capsular types suggests that this structural element is central to the antiphagocytic function of all GBS capsular polysaccharides. This conclusion is further supported by the diversity among the sialyltransferases of the various capsular types. The diversity of sialyltransferase sequences parallels the structural variation in closely linked sugars and supports the hypothesis that the maintenance of a functional sialyltransferase is critical to GBS survival, even as other elements of the polysaccharide structures evolve. Indeed, the selective pressure of α-D-Neu*p*NAc (2→3)-β-D-Gal*p* is such that the distantly related serotype VIII retains this structural feature.

Host immunity may have shaped the evolution of the GBS capsular polysaccharides not only by the selection of antigenically distinct variants but also by the recognition of nonself carbohydrate epitopes. Together, the conserved GBS disaccharide and the variable trisaccharide constitute an oligosaccharide motif which is very similar to that displayed by the sialyl-Lewis antigen and related carbohydrate epitopes on human glycoproteins and glycolipids. It has been proposed that this structural similarity to host oligosaccharides represents molecular mimicry by GBS (11). Human antibodies to the GBS type III polysaccharide recognize a polysaccharide chain length-dependent epitope; this observation is consistent with the hypothesis that the simple oligosaccharide epitope of the repeating unit is seen as a self-antigen (32, 33). The restricted structural diversity of the GBS polysaccharides may reflect an adaptive advantage of capsules that mimic host antigens, thus limiting immune recognition to more complex epitopes. It is intriguing that the type V polysaccharide, predicted by the present analysis to have evolved most recently, evokes antibodies that are predominantly of the immunoglobulin M isotype in humans (C. J. Baker and D. L. Kasper, unpublished results). Relatively weak class switching to immunoglobulin G might reflect antigenic relatedness of a type V-specific structural motif (such as linkage of the trisaccharide side chain to a backbone glucose residue) to self-antigens.

Taken together, these analyses imply that diversity in GBS capsular polysaccharide structures has arisen through the introduction of novel DNA sequences and genetic recombination. Counterbalancing the adaptive advantage of antigenic diversity are equally strong selective forces that have maintained particular structural motifs that shield the organisms from immune surveillance mechanisms. The result of these opposing effects is a repertoire of capsular polysaccharides that display conserved motifs which are central to the biologic func-

tions of the GBS capsule and that have sufficient structural variation for antigenic uniqueness.

ACKNOWLEDGMENTS

We thank Sheng-Mei Jiang, Anders Håkansson, and Karen Puopolo for helpful discussions and critical readings of the manuscript.

This work was supported in part by the Charles Hood Foundation (M.J.C.), by NIH grant AI42940 to M.W., and by NIH grant AI22498 to C.E.R.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**217–221.
2. **Chaffin, D. O., S. B. Beres, H. H. Yim, and C. E. Rubens.** 2000. The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon. J. Bacteriol. **182:**4466–4477.
3. **Chaffin, D. O., K. McKinnon, and C. E. Rubens.** 2002. CpsK of *Streptococcus agalactiae* exhibits alpha2,3-sialyltransferase activity in *Haemophilus ducreyi.* Mol. Microbiol. **45:**109–122.
4. **Cieslewicz, M. J., D. L. Kasper, Y. Wang, and M. R. Wessels.** 2001. Functional analysis in type Ia group B *Streptococcus* of a cluster of genes involved in extracellular polysaccharide production by diverse species of streptococci. J. Biol. Chem. **276:**139–146.
5. **Di Fabio, J. L., F. Michon, J.-R. Brisson, and H. J. Jennings.** 1989. Structure of the capsular polysaccharide antigen of type IV group B *Streptococcus.* Can. J. Chem. **67:**877–882.
6. **Edwards, M. S., D. L. Kasper, H. J. Jennings, C. J. Baker, and A. Nicholson-Weller.** 1982. Capsular sialic acid prevents activation of the alternative complement pathway by type III, group B streptococci. J. Immunol. **128:**1278–1283.
7. **Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couvé, L. Lalioui, C. Poyart, P. Trieu-Cuot, and F. Kunst.** 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol. Microbiol. **45:**1499–1513.
8. **Glusman, G., and D. Lancet.** 2000. GESTALT: a workbench for automatic integration and visualization of large-scale genomic sequence analyses. Bioinformatics **16:**482–483.
9. **Glusman, G., and D. Lancet.** 2001. Visualizing large-scale genomic sequences. IEEE Eng. Med. Biol. Mag. **20:**49–54.
10. **Haft, R. F., M. R. Wessels, M. F. Mebane, N. Conaty, and C. E. Rubens.** 1996. Characterization of *cpsF* and its product CMP-*N*-acetylneuraminic acid synthetase, a group B streptococcal enzyme that can function in K1 capsular polysaccharide biosynthesis in *Escherichia coli.* Mol. Microbiol. **3:**555–563.
11. **Hayrinen, J., D. Bitter-Suermann, and J. Finne.** 1989. Interaction of meningococcal group B monoclonal antibody and its Fab fragment with alpha 2-8-linked sialic acid polymers: requirement of a long oligosaccharide segment for binding. Mol. Immunol. **26:**523–529.
12. **Higgins, D. G., J. D. Thompson, and T. J. Gibson.** 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. **266:**383–402.
13. **Howard, C. J., and A. A. Glynn.** 1971. The virulence for mice of strains of *Escherichia coli* related to the effects of K antigens on their resistance to phagocytosis and killing by complement. Immunology **20:**767–777.
14. **Jennings, H. J., E. Katzenellenbogen, C. Lugowski, and D. L. Kasper.** 1983. Structure of native polysaccharide antigens of type Ia and type Ib group B *Streptococcus.* Biochemistry **22:**1258–1264.
15. **Jennings, H. J., C. Lugowski, and D. L. Kasper.** 1981. Conformational aspects critical to the immunospecificity of the type III group B streptococcal polysaccharide. Biochemistry **20:**4511–4518.
16. **Jennings, H. J., K.-G. Rosell, E. Katzenellenbogen, and D. L. Kasper.** 1983. Structural determination of the capsular polysaccharide antigen of type II group B streptococci. J. Biol. Chem. **258:**1793–1798.
17. **Jones, C.** 1985. Identification of the tetrasaccharide repeating-unit of the *Streptococcus pneumoniae* type 23 polysaccharide by high-field proton N.M.R. spectroscopy. Carbohydr. Res. **139:**75–83.
18. **Kim, K. S., J. H. Kang, and A. S. Cross.** 1986. The role of capsular antigens in serum resistance and *in vivo* virulence of *Escherichia coli.* FEMS Microbiol. Lett. **35:**275–278.
19. **Kogan, G., J.-R. Brisson, D. L. Kasper, C. von Hunolstein, G. Orefici, and H. J. Jennings.** 1995. Structural elucidation of the novel type VII group B *Streptococcus* capsular polysaccharide by high resolution NMR spectroscopy. Carbohydr. Res. **277:**1–9.
20. **Kogan, G., D. Uhrín, J.-R. Brisson, L. C. Paoletti, A. E. Blodgett, D. L. Kasper, and H. J. Jennings.** 1996. Structural and immunochemical characterization of the type VIII group B *Streptococcus* capsular polysaccharide. J. Biol. Chem. **271:**8786–8790.
21. **Kolkman, M. A. B., B. A. M. van der Zeijst, and P. J. M. Nuijten.** 1997. Functional analysis of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* serotype 14. J. Biol. Chem. **272:**19502–19508.
22. **Lee, J. C., and C. Y. Lee.** 1999. Capsular polysaccharides of *Staphylococcus aureus*, p. 185–205. *In* J. B. Goldberg (ed.), Genetics of bacterial polysaccharides. CRC Press, Boca Raton, Fla.
23. **Lipsitch, M.** 1999. Bacterial vaccines and serotype replacement: lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae.* Emerg. Infect. Dis. **3:**336–345.
24. **Marques, M. B., D. L. Kasper, M. K. Pangburn, and M. R. Wessels.** 1992. Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. Infect. Immun. **60:**3986–3993.
25. **Pearson, W. R., T. Wood, Z. Zhang, and W. Miller.** 1997. Comparison of DNA sequences with protein sequences. Genomics **46:**24–36.
26. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
27. **Suryanti, V., A. Nelson, and A. Berry.** 2003. Cloning, over-expression, purification, and characterisation of *N*-acetylneuraminate synthase from *Streptococcus agalactiae.* Protein Expr. Purif. **27:**346–356.
28. **Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser.** 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae.* Proc. Natl. Acad. Sci. USA **99:**12391–12396.
29. **von Hunolstein, C., S. D. Ascenzi, B. Wagner, J. Jelinkova, G. Alfarone, S. Recchia, M. Waner, and G. Orefici.** 1993. Immunochemistry of capsular type polysaccharide and virulence properties of type VI *Streptococcus agalactiae* (group B *Streptococcus).* Infect. Immun. **61:**1272–1280.
30. **Watanabe, M., K. Miyake, K. Yanae, Y. Kataoka, S. Koizumi, T. Endo, A. Ozaki, and S. Iimima.** 2002. Molecular characterization of a novel β1,3-galactosyltransferase for capsular polysaccharide synthesis by *Streptococcus agalactiae* type Ib. J. Biochem. **131:**183–191.
31. **Wessels, M. R., J. L. DiFabio, V.-J. Benedi, D. L. Kasper, F. Michon, J.-R. Brisson, J. Jelinkova, and H. J. Jennings.** 1991. Structural determination and immunochemical characterization of the type V group B *Streptococcus agalactiae* capsular polysaccharide. J. Biol. Chem. **266:**6714–6719.
32. **Wessels, M. R., A. Munoz, and D. L. Kasper.** 1987. A model of high-affinity antibody binding to type III group B *Streptococcus* capsular polysaccharide. Proc. Natl. Acad. Sci. USA **24:**9170–9174.
33. **Wessels, M. R., V. Pozsgay, D. L. Kasper, and H. J. Jennings.** 1987. Structure and immunochemistry of an oligosaccharide repeating unit of the capsular polysaccharide of type III group B *Streptococcus.* J. Biol. Chem. **262:**8262–8267.
34. **Wyle, F. A., M. S. Artenstein, B. L. Brandt, E. C. Tramont, D. L. Kasper, P. L. Altieri, S. L. Berman, and J. P. Lowenthal.** 1973. Immunologic response of man to group B meningococcal polysaccharide vaccines. J. Infect. Dis. **126:**514–521.
35. **Yamamoto, S., K. Miyake, Y. Koike, M. Watanabe, Y. Machida, M. Ohta, and S. Iijima.** 1999. Molecular characterization of type-specific capsular polysaccharide biosynthesis genes of *Streptococcus agalactiae* type Ia. J. Bacteriol. **181:**5176–5184.
36. **Yother, J.** 1999. Common themes in the genetics of streptococcal capsular polysaccharides, p. 161–184. *In* J. B. Goldberg (ed.), Genetics of bacterial polysaccharides. CRC Press, Boca Raton, Fla.

Editor: V. J. DiRita