Bioinformatic insights into the biosynthesis of the Group B carbohydrate in *Streptococcus agalactiae*

Iain C. Sutcliffe,¹ Gary W. Black¹ and Dean J. Harrington²

Correspondence

Iain Sutcliffe

iain.sutcliffe@unn.ac.uk

¹Northumbria University, Newcastle upon Tyne NE1 8ST, UK

²University of Bradford, West Yorkshire BD7 1DP, UK

Streptococcus agalactiae is a major human and animal pathogen, most notable as a cause of life-threatening disease in neonates. *S. agalactiae* is also called the Group B *Streptococcus* in reference to the diagnostically significant Lancefield Group B typing antigen. Although the structure of this complex carbohydrate antigen has been solved, little is known of its biosynthesis beyond the identification of a relevant locus in sequenced *S. agalactiae* genomes. Analysis of the sugar linkages present in the Group B carbohydrate (GBC) structure has allowed us to deduce the minimum enzymology required to complete its biosynthesis. Most of the enzymes required to complete this biosynthesis can be identified within the putative biosynthetic locus. Surprisingly, however, three crucial *N*-acetylglucosamine transferases and enzymes required for activated precursor synthesis are not apparently located in this locus. A model for GBC biosynthesis wherein the complete polymer is assembled at the cytoplasmic face of the plasma membrane before translocation to the cell surface is proposed. These analyses also suggest that GBC is the major teichoic acid-like polymer in the cell wall of *S. agalactiae*, whereas lipoteichoic acid is the dominant poly(glycerophosphate) antigen. Genomic analysis has allowed us to predict the pathway leading to the biosynthesis of GBC of *S. agalactiae*.

Received 31 October 2007

Revised 19 December 2007

Accepted 6 February 2008

INTRODUCTION

Streptococcus agalactiae is recognized as a leading aetiological agent of invasive neonatal disease worldwide (Henneke & Berner, 2006; Johri *et al.* 2006). In the USA, neonatal disease caused by *S. agalactiae* occurs in approximately 0.69 per 1000 live births and has an estimated mortality rate in neonates of between 4 and 6% (Brooks *et al.*, 2006; Johri *et al.* 2006). Despite varying strategies for disease management, European rates of neonatal disease are comparable e.g. 0.72 cases per 1000 live births in the UK (Heath *et al.*, 2004). The anomalously low incidence of *S. agalactiae* neonatal disease in the developing world may reflect problems in accurate data collection (Johri *et al.*, 2006) and high disease burdens have been reported recently in studies carried out in Jamaica and Malawi (Trotman & Bell, 2006; Gray *et al.*, 2007). *S. agalactiae* is also emerging as a cause of disease in the elderly (Edwards & Baker, 2005). In addition to its importance as a cause of human disease, *S. agalactiae* is also significant as a veterinary pathogen, notably causing mastitis (Lancefield, 1934; Bradley, 2002).

Abbreviations: GalT, galactosyltransferase; GBC, Group B carbohydrate; GlucPT, glucitol-phosphate transferase; NAG, *N*-acetylglucosamine; NAM, *N*-acetylmuramic acid; RhaT, rhamnosyltransferase.

A step-wise model for the biosynthesis of the *S. agalactiae* GBC, and three supplementary tables of data are available with the online version of this paper.

Beginning in the 1920s, the pioneering work of Rebecca Lancefield at the Rockefeller University (www.rockefeller.edu/vaf/rcbio.htm) led to the recognition that many streptococcal species could be typed on the basis of the carbohydrate-containing antigens in their cell walls. *S. agalactiae* was typed as the 'group B' haemolytic *Streptococcus* (Lancefield, 1934), hence the clinical usage of the name Group B *Streptococcus* as a synonym for this bacterium. The Lancefield typing system remains a cornerstone of the diagnosis of many streptococcal diseases, including the identification of *S. agalactiae* using rapid test kits for the detection of the Lancefield Group B carbohydrate (GBC; Greenberg *et al.*, 1995; Davies *et al.*, 2003).

Despite its diagnostic importance, GBC has received limited study in relation to its significance in the pathogenesis of *S. agalactiae* disease. Although GBC appears to be a proinflammatory component (Vallejo *et al.*, 1996), interpretation of studies of its immunobiology are in part hampered by the difficulty in purifying GBC away from contaminant cell-wall-derived materials as it is covalently attached to peptidoglycan (Deng *et al.*, 2000; Henneke & Berner, 2006). Studies have indicated that passively administered antibodies to GBC are not protective against infection in mice (Lancefield *et al.*, 1975) and that the overlying capsule may block access of opsonizing antibody to GBC (Marques *et al.*, 1994). Moreover, in humans, high levels of maternal antibodies to GBC do not always confer passive protection to neonates (Anthony

Fig. 1. Multiantennary structure of GBC. GBC is assembled from four component oligosaccharides (I–IV) into a branched structure that presents four main terminal branches. The structure is attached to the peptidoglycan by the stem oligosaccharide IV. P, Phosphate.

et al., 1985). Intriguingly a stable opaque variant strain of *S. agalactiae*, obtained by repeated subculture of a serotype III strain, has been reported to be deficient in antibody-detectable GBC (Pincus *et al.*, 1992). This strain exhibited impaired growth *in vitro* and formed long coccal chains due to aberrant septation. Subsequent studies revealed it to be less immunogenic, more readily killed by polymorphonuclear neutrophils and was significantly attenuated in a mouse model of infection (Pincus *et al.*, 1993).

In the present study we have used the availability of whole-genome sequences for *S. agalactiae* strains, in combination with a consideration of the detailed structure of GBC, to perform bioinformatic analyses of the pathway leading to GBC biosynthesis.

METHODS

Serotype V genome data were used as the prototypical *S. agalactiae* sequence (Tettelin *et al.*, 2002). Protein sequences for analysis were accessed via the UniProtKB/Swiss-Prot database (<http://us.expasy.org/sprot/>; UniProt Consortium, 2007).

These database entries were also used to access protein family data in the Pfam database (<http://pfam.sanger.ac.uk/>; Finn *et al.*, 2006), to access the Genome Proximity viewer and to access the GenBank entries for sequence and annotation data. Protein sequences were retrieved and homology searches performed using the NCBI BLAST server (www.ncbi.nlm.nih.gov/BLAST/; McGinnis & Madden, 2004). Specifically, searches were performed using either the BLASTP tool (unfiltered; *E* value cut-off typically set to 0.001) or using the Microbial Genomes tool to search for homologues in selected genomes. The latter searches were performed using the BLASTP tool, unfiltered with the *E* value cut-off typically set to 0.01. Functional information was also obtained from the CDD Conserved Domain search generated as part of the standard BLASTP output.

Proteins were analysed for membrane-spanning regions using a variety of topology prediction tools, including TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/; Krogh *et al.*, 2001), TMPred (www.ch.embnet.org/software/TMPRED_form.html; Hofmann & Stoffel, 1993) and Phobius (<http://phobius.cgb.ki.se/>; Käll *et al.*, 2004). The latter tool also allowed us to examine sequences for the presence of a signal peptide.

Glycosyltransferase sequences for *S. agalactiae* were identified either by keyword searches within UniProtKB or from access to the CAZY genome browser (www.cazy.org/geno/geno_bacteria.html#S; Coutinho & Henrissat, 1999).

RESULTS AND DISCUSSION

Structure of GBC: definition as a 'teichoic-acid-like' polymer

The structure of GBC was elucidated by the elegant studies of Pritchard *et al.* (1984) and Jennings, Michon and co-workers (Michon *et al.*, 1987, 1988, 1991). Cumulatively, these studies defined a 'multi-antennary structure' (Figs 1 and 2) based on the arrangement of four component oligosaccharides into a progressively bifurcating structure that ultimately presents four main branches (Michon *et al.*,

Fig. 2. Fine structure of the component oligosaccharides of GBC. Oligosaccharide Unit IV provides the distinctive stem of the GBC structure. The branched structure is assembled from oligosaccharides I, II and III which are structurally inter-related. Notably, the oligosaccharide III structure is present within oligosaccharide II (grey text) whilst the oligosaccharide II structure is present within oligosaccharide I (boxed). Abbreviations used are as follows: Gal, galactose; NAG, *N*-acetylglucosamine; NAM, *N*-acetylmuramic acid; P, phosphate; Rha, rhamnose.

1988, 1991). Significantly, the component oligosaccharides are all interlinked by a phosphodiester linkage between a glucitol in one oligosaccharide and galactose in another (Figs 1 and 2). The presence of these glucitol-6-phosphate-derived linkages suggested that GBC was a teichoic-acid-like structure (Pritchard *et al.*, 1984; Michon *et al.*, 1987). The stereochemistry of the glucitol 6-phosphate linkage is directly comparable to that of glycerol 3-phosphate and ribitol 5-phosphate in typical wall teichoic acids (Neuhaus & Baddiley, 2003) and reflects the delivery of these units from CDP-alditol precursors (see below). Also consistent with this definition of GBC as the teichoic-acid-like polymer of *S. agalactiae* is the observation that the stem oligosaccharide IV (Figs 1 and 2) is covalently attached to *N*-acetylmuramic acid (NAM) residues of the cell-wall peptidoglycan (Deng *et al.*, 2000). Finally, it should be noted that, contrary to a previous preliminary report of a glycerophosphate teichoic acid (Goldschmidt & Panos, 1984), genomic analysis suggests *S. agalactiae* is unable to synthesize poly(glycerophosphate) teichoic acids as obvious orthologues of the crucial enzymes for glycerol 3-phosphate polymerization (TagB, TagD and TagF; Bhavsar & Brown, 2006) are not encoded in the *S. agalactiae* genome (our observations). This latter observation is of additional importance as it provides further evidence that lipoteichoic acid is the predominant poly(glycerophosphate) cell envelope polymer in *S. agalactiae* (Doran *et al.*, 2005; Henneke & Berner, 2006). We speculate that the poly(glycerophosphate) 'teichoic acid' previously identified by Goldschmidt & Panos (1984) may in fact have been deacylated lipoteichoic acid, a component identified by other authors (Erbing *et al.*, 1986; Mattingly & Johnston, 1987). We also note that products of the *dlt* (α -alanine-lipoteichoic) operon of *S. agalactiae* (Poyart *et al.*, 2001) most likely only alanylate the lipoteichoic acid since alanine has not been reported as a component of GBC (Pritchard *et al.*, 1984; Michon *et al.*, 1987, 1988) and it is unlikely that the *dlt* system is able to recognize the chemically distinct rhamnose-substituted glucitol-phosphate as a substrate.

Linkage analysis and the enzymology of GBC biosynthesis

The availability of whole-genome sequences for three strains of *S. agalactiae* (Tettelin *et al.*, 2002; Glaser *et al.*, 2002; Tettelin *et al.*, 2005) has provided valuable insight into the genetics of GBC biosynthesis as syntenous clusters of genes putatively encoding this pathway were identified in these genomes (Table 1; Fig. 3). All the proteins encoded at this locus are part of the GBS core genome defined by the sequencing of genomes for five other strains of *S. agalactiae* (Tettelin *et al.*, 2005). For clarity, proteins are referred to here using the annotation of Tettelin *et al.* (2002), i.e. prefixed 'SAG-'. The GBC locus is a series of closely linked (and in some cases overlapping) ORFs with a single annotated terminator (Glaser *et al.*, 2002; Fig. 3), suggesting possible transcription in a single unit.

Further details of the GBC biosynthetic path have yet to be reported. To establish whether this locus alone is sufficient for the biosynthesis of GBC, we have analysed the linkages present in the complete GBC structure (Fig. 2) and deduced the requisite minimal enzymology needed for its biosynthesis (Table 2). Even though GBC exhibits a comparatively complex branched structure (for example, there are 42 terminal α -rhamnose residues), the modular nature of GBC and the inter-related structures of oligosaccharides I, II and IV (Fig. 2) means that surprisingly few distinct transferase activities are theoretically needed to derive the structure (Table 2). The linkage analysis demands, as a minimum, the involvement of 11 transferase enzymes, i.e. five distinct α -rhamnosyltransferases (α -RhaTs), a single β -rhamnosyltransferase (β -RhaT), a single α -galactosyltransferase (α -GalT), a glucitol-phosphate transferase (GlucPT) and three *N*-acetylglucosamine (NAG)-transferases. With the notable exception of the NAG transferases (see below), candidates for all of the requisite transferases can be presumptively identified within the GBC locus (Table 1). Moreover, it is possible to deduce a complete pathway for GBC biosynthesis and assembly into the cell wall that can be accomplished in 127 steps (supplementary File 1, available with the online version of this paper). This biosynthetic path and the related enzymology are detailed in the following sections.

Initiation of GBC synthesis

Many cell-envelope polymers (most notably peptidoglycan, lipopolysaccharide and capsular polysaccharides) are synthesized, at least in part, at the cytoplasmic face of the plasma membrane as the requisite nucleotide-activated precursors are intracellular (Alaimo *et al.*, 2006; Bhavsar & Brown, 2006; Whitfield, 2006). These polymers or their repeat units are assembled on a polyisoprenol-phosphate carrier molecule (typically undecaprenylphosphate) and ultimately flipped to the extracytoplasmic face of the plasma membrane before their assembly and/or final localization. In Gram-positive bacteria, this model also holds true for the biosynthesis of poly(glycerophosphate) teichoic acids (Bhavsar & Brown, 2006; Damjanovic *et al.*, 2007) and various capsular polysaccharides (Bentley *et al.*, 2006). This strategy may facilitate co-ordination between peptidoglycan assembly and the attachment of polymers that are covalently linked to the glycan repeat units, which includes the capsular polysaccharide and GBC in *S. agalactiae* (Deng *et al.*, 2000). We therefore hypothesized that GBC would be biosynthesized attached to a polyisoprenol-phosphate, which was further supported by the identification of a gene encoding a putative flippase in the GBC locus (SAG1412, Table 1; see below) and the observation that analyses with various predictors of membrane protein topology predict that all the putative glycosyltransferases encoded in the GBC locus (Table 1) are cytoplasmic, although some may be peripheral membrane proteins. Thus the structure of oligosaccharide Unit IV in the GBC stem unit (Fig. 2) requires that the first step in

Table 1. *S. agalactiae* GBC biosynthesis loci identified from whole-genome sequencing

Protein designation in strain:			Annotation and predicted function
2603/V	NEM316	A909	
SAG1409	GBS1479	Absent	RogB . Not part of core genome: annotated as containing an authentic frame shift in strain 2603/V and absent from strain A909. Marks the boundary of the <i>S. agalactiae</i> GBC locus.
SAG1410	GBS1480	SAK_1445	Glycosyl transferase, group 1 family protein . Putative α -GalT using UDP-Gal substrate.
SAG1411	GBS1481	SAK_1446	Glycosyl transferase, group 2 family protein . Putative α -RhaT. 115/282 (48 %) amino acid identity to <i>S. thermophilus</i> EpsQ putative α -RhaT (Bourgoin <i>et al.</i> , 1999). Homologous to other putative α -RhaTs, e.g. WelH of <i>Lactobacillus rhamnosus</i> .
SAG1412	GBS1482	SAK_1447	Polysaccharide biosynthesis protein . 149/427 (34 %) amino acid identity with <i>S. pneumoniae</i> Wzx and 98/408 (24 %) amino acid identity with <i>E. coli</i> RfbX (Wzx) flippases.
SAG1413	GBS1483	SAK_1448	Membrane protein, putative . Conserved hypothetical of restricted distribution in streptococci and lactococci.
SAG1414	GBS1484	SAK_1449	Glycosyl transferase, group 2 family protein . Putative α -RhaT. 145/306 (47 %) amino acid identity to SAG1415. 79/298 (26 %) amino acid identity with <i>S. mutans</i> RgpB α -RhaT (Yamashita <i>et al.</i> , 1998).
SAG1415	GBS1485	SAK_1450	Glycosyl transferase, group 2 family protein . Putative α -RhaT. 145/306 (47 %) amino acid identity to SAG1414.
SAG1416	GBS1486	SAK_1451	Nucleotide sugar dehydratase, putative . 129/359 (35 %) amino acid identity to a putative Mnp2 (mannitol) dehydratase/epimerase in the capsule biosynthesis loci of <i>S. pneumoniae</i> strains 1936/39, 198/71 and 7765/43 (Bentley <i>et al.</i> , 2006).
SAG1417	GBS1487	SAK_1452	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase . 142/237 (59 %) amino acid identity with putative IspD-family (mannitol) transferase in <i>S. pneumoniae</i> strain 1936/39 capsule locus (Bentley <i>et al.</i> , 2006). 72/234 (30 %) amino acid identity with <i>H. influenzae</i> Bcs1 CDP-ribitol pyrophosphorylase (IspD) domain (Follens <i>et al.</i> , 1999).
SAG1418	GBS1488	SAK_1453	Polysaccharide biosynthesis protein, putative . Member of the LicD family. 85/274 (31 %) amino acid identity with <i>S. pneumoniae</i> LicD1 (Zhang <i>et al.</i> , 1999). 79/270 (29 %) amino acid identity with <i>H. influenzae</i> LicD (Lysenko <i>et al.</i> , 2000).
SAG1419	GBS1489	SAK_1454	Lipoprotein, putative . 117/464 (25 %) amino acid identity with <i>L. lactis</i> subsp. <i>cremoris</i> LACR_0221 (only significant homologue).
SAG1420	GBS1490	SAK_1455	Hypothetical protein . Uncharacterized conserved hypothetical integral membrane protein of restricted distribution.
SAG1421	GBS1491	SAK_1456	Glycosyl transferase, group 2 family protein . Putative α -RhaT. 107/233 (45 %) amino acid identity with <i>S. pyogenes</i> Spy_0794.
SAG1422	GBS1492	SAK_1457	Glycosyl transferase, group 2 family protein . Putative α -RhaT. 181/315 (57 %) amino acid identity with <i>S. mutans</i> RgpB RhaT (Yamashita <i>et al.</i> , 1998).
SAG1423	GBS1493	SAK_1458	Glycosyl transferase, putative . 234/362 (64 %) amino acid identity with RgpA, involved in initiating rhamnose-glucose polysaccharide synthesis in <i>S. mutans</i> (Yamashita <i>et al.</i> , 1998; Shibata <i>et al.</i> , 2002). 185/388 (47 %) amino acid identity with WchF <i>Streptococcus oralis</i> β -RhaT (Yoshida <i>et al.</i> , 2006). 180/388 (46 %) amino acid identity to the putative β -RhaT in the unique type VIII <i>S. agalactiae</i> capsule locus (Cieslewicz <i>et al.</i> , 2005).
SAG1424	GBS1494	SAK_1459	dTDP-4-dehydrorhamnose reductase RfbD . 247/284 (86 %) amino acid identity with RmlD dTDP-4-keto-L-rhamnose reductase <i>S. mutans</i> (Tsukioka <i>et al.</i> , 1997).
SAG1425	GBS1495	SAK_1460	Conserved hypothetical protein . DUF59 (PF01883) family protein. Probable boundary of locus.

GBC biosynthesis would be transfer of a NAG unit from UDP-NAG onto the polyisoprenol-phosphate carrier molecule. However, we were unable to identify an appropriate NAG transferase in the GBC locus, but instead noted that SAG0140 is significantly homologous to TagO [135/324 (41 %) amino acid identity], which carries out the identical reaction in teichoic acid synthesis in *Bacillus subtilis* (Soldo *et al.*, 2002; Bhavsar & Brown, 2006; D'Elia *et al.*, 2006). SAG0140 is also highly homologous to RgpG, which carries out the comparable reaction initiating rhamnose-glucose polysaccharide (RGP) synthesis in *Streptococcus mutans* (Yamashita *et al.*, 1999; Shibata *et al.*, 2002) and WecA, which initiates LPS and

enterobacterial common antigen biosynthesis in *Escherichia coli* (Lehrer *et al.*, 2007). Indeed, the predicted topology of SAG0140 matches that determined for WecA (Lehrer *et al.*, 2007) and key amino acid sequence motifs are conserved. Thus we propose that SAG0140 initiates GBC biosynthesis on a polyisoprenol-phosphate lipid carrier.

Synthesis of the oligosaccharide components of GBC and their assembly into the full polymer

We predict that the cytoplasmic enzymology of GBC synthesis dictates that biosynthesis of GBC proceeds from

Fig. 3. Genetic organization of the GBC locus within the genome of *S. agalactiae* strain 2603/V. Gene organization was obtained from the JGI Integrated Microbial Genomes gene neighbourhood tool (<http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>) and annotated with relevant gene names. The single terminator (indicated by T) was identified from the annotation of the genome sequences of strain NEM316 (Glaser *et al.*, 2002). Note that the 5' boundary of the GBC locus is a probable pseudogene (*sag1409 pg*; *rogB*) in strain 2603/V. *RogB* is a regulator that is part of the recently defined ~11 kb 'Pilus island 2a' present in most strains studied, but is absent from strain A909 (Gutekunst *et al.*, 2003; Rosini *et al.*, 2006). The 3' boundary of the GBC locus is discussed in the text.

the initiation step described above directly through to the full polymer, i.e. without separate synthesis, translocation and subsequent assembly of the distinct oligosaccharide intermediates. Following the initiation of synthesis on the polyisoprenol-phosphate carrier, the next sequence of reactions would be the sequential action of the transferases required to build up oligosaccharide units IV and I in the GBC stem. Thus the second reaction would be the formation of the only β -rhamnose linkage in the polymer. We predict that the β -RhaT responsible is SAG1423, which exhibits significant homologies to other β -RhaT enzymes (Table 1), most notably RgpA which is the first RhaT in RGP synthesis in *S. mutans* (Shibata *et al.*, 2002). Other homologues include WchF, which is proposed to carry out a comparable reaction in streptococcal coaggregation receptor polysaccharide biosynthesis (Yoshida *et al.*,

2006) and the Cps8R putative transferase for the incorporation of β -rhamnose into the type VIII capsular polysaccharide, which is notable as the only *S. agalactiae* serotype to contain β -rhamnose within its polysaccharide repeating unit (Kogan *et al.*, 1996; Cieslewicz *et al.*, 2005). Consistent with the initiation of oligosaccharide Unit IV synthesis in close proximity to the cytoplasmic face of the plasma membrane, a hydrophobic region in the N-terminal part of the SAG1423 protein may facilitate its location as a peripheral membrane protein (data not shown).

The next step in GBC synthesis is the formation of the three α -rhamnose 1-3 rhamnose linkages which are also exclusive to the base of oligosaccharide Unit IV. The α -RhaT responsible for these three steps is likely to be SAG1422, which exhibits very high sequence homology to

Table 2. GBC sugar linkage analysis based on the structure determined by Michon *et al.* (1991)

Linkage formed	Context (oligosaccharide structural unit)	Enzyme needed
NAG-P-DolP linkage formation	Not present in final structure	Transferase for NAG transfer to the lipid carrier for initiation of GBC assembly (TagO homologue)
L-Rha β 1-4 D-NAG	IV	β -RhaT utilizing NAG C4 as acceptor
L-Rha α 1-3 D-Gal	I, II, IV	α -RhaT utilizing galactose C3 as acceptor
L-Rha α 1-2 L-Rha	I, II, III	α -RhaT utilizing rhamnose C2 as acceptor
L-Rha α 1-3 L-Rha	IV	α -RhaT utilizing rhamnose C3 as acceptor
L-Rha α 1-3 Glucitol	I, II, III	α -RhaT utilizing glucitol C3 as acceptor
L-Rha α 1-1 Glucitol	I, II, III	α -RhaT utilizing glucitol C1 as acceptor
Glucitol-6'-P-6 Gal (phosphodiester) linkage	I-IV, I-I, II-I, II-II, III-II	CDP-glucitol GlucPT utilizing galactose C6 as acceptor
D-Gal α 1-3 D-NAG	I, II, IV	α -D-GalT
D-NAG β 1-3 L-Rha	IV	β -NAG transferase utilizing rhamnose C3 as acceptor
D-NAG β 1-4 L-Rha	I, II	β -NAG transferase utilizing rhamnose C4 as acceptor
NAG-P-NAM linkage formation	IV-muramate	Transferase for GBC transfer from the lipid carrier onto the peptidoglycan

RgpB, the second RhaT required for RGP synthesis in *S. mutans* (Shibata *et al.*, 2002). SAG1422 also has a short hydrophobic stretch of amino acids (data not shown) and thus may be a peripheral membrane protein, consistent with its utilization of a small lipid-anchored oligosaccharide substrate. Thus SAG1422 and SAG1423 are considered orthologues of *S. mutans* RgpB and RgpA, respectively.

The synthesis of oligosaccharide Unit IV can then be completed by the action of an NAG transferase, α -GalT and an α -rhamnose 1-3 galactose α -RhaT. As SAG1410 is the only Pfam family PF00534 glycosyltransferase encoded in the GBC locus, we propose that this is the α -GalT using UDP-galactose as substrate (Table 1). The assignment of the NAG transferase and the α -RhaT are discussed below.

Following the synthesis of oligosaccharide Unit IV, the synthesis of the Unit I oligosaccharide is predicted to begin with the transfer of glucitol-phosphate to form a phosphodiester linkage to the C6 of the galactose in Unit IV. As all the oligosaccharide units in GBC are linked by this same linkage type (Michon *et al.*, 1988), a single GlucPT can be considered sufficient for the formation of the phosphodiester cross-links throughout the GBC structure. In teichoic acid biosynthesis, alditol-phosphates are derived from CDP-alditol precursors (Neuhaus & Baddiley, 2003; Bhavsar & Brown, 2006). Thus it is significant to note that SAG1417 encodes a putative alditol-phosphate cytidyltransferase (Table 1) which could synthesize CDP-glucitol from CTP and glucitol-phosphate (see below). Given the above-mentioned absence of TagB or TagF homologues in the *S. agalactiae* genomes, we predict that the GlucPT using this CDP-glucitol is likely to be SAG1418, as this protein is a member of the LicD family (Table 1). LicD family proteins are strong candidates for the diphosphonucleoside choline transferases of *Streptococcus pneumoniae* and *Haemophilus influenzae*, which transfer phosphorylcholine from CDP-choline to the pneumococcal teichoic acid and *H. influenzae* LPS, respectively (Zhang *et al.*, 1999; Lysenko *et al.*, 2000). Thus the similarity of the CDP-activated substrate and the phosphoryl group transfer reaction in the activities of LicD family proteins makes SAG1418 a strong candidate for the GlucPT.

The transfer of glucitol-phosphate to oligosaccharide Unit IV should allow Unit I biosynthesis to proceed via the action of two further α -RhaT enzymes, which transfer rhamnose to the C3 and C1 positions of glucitol. A third α -RhaT can then add two α 1-2-linked rhamnose to the latter of these residues. Subsequently, a single NAG transferase is proposed to add β 1-4 NAG to both the terminal and penultimate rhamnose of the growing Unit I oligosaccharide. This creates a crucial branch point from which each NAG can be modified with an α 1-3-linked galactose by the same α -GalT used in Unit IV biosynthesis. Similarly, each galactose can then be modified with an α 1-3-linked rhamnose by the same α -RhaT as used in oligosaccharide Unit IV biosynthesis, completing the synthesis of the branched Unit I structure (Fig. 2). Significantly, the Unit I

oligosaccharide synthesized presents two α 1-3-linked galactose residues. Transfer of glucitol-phosphate to each of these galactose residues by the SAG1418 protein would then create the first bifurcation in the GBC structure, with another Unit I oligosaccharide then being built up on each glucitol, by the same reaction series as described above. Furthermore, each of these Unit I oligosaccharides synthesized would again present two α 1-3-linked galactose residues: each of these could receive a glucitol-phosphate, thereby presenting the start of each terminal branch of the tetra-antennary structure upon which oligosaccharides II and then III can be assembled. Important here is the recognition that the oligosaccharide II structure is contained within Unit I and that the oligosaccharide III structure is contained within oligosaccharide units I and II (Fig. 2). Thus, the full branched structure of GBC can be proposed to be completed without the involvement of any additional transferase enzymes beyond those used to complete oligosaccharide units I and IV, i.e. through the co-ordination of 11 different transferase activities in 125 steps (supplementary File 1).

Translocation of GBC to the cell envelope and cross-linkage to the cell wall

Following the synthesis of GBC, the polysaccharide needs to be translocated to the plasma membrane surface and correctly localized within the cell envelope. Significantly, SAG1412 belongs to the Wzx (RfbX) family of flippases involved in capsule and lipopolysaccharide O-antigen biosynthesis (Liu *et al.*, 1996; Marolda *et al.*, 2006; Whitfield, 2006), members of which are also predicted to translocate pneumococcal teichoic acid (Damjanovic *et al.*, 2007) and most capsular polysaccharides of the different *S. pneumoniae* serotypes (Bentley *et al.*, 2006). The presence of NAG as the reducing terminal sugar is apparently important for the activity of Wzx flippases (Alaimo *et al.*, 2006) and thus it is significant that this sugar is present at the terminal structure of Unit IV (Fig. 2). Three other genes in the GBC locus encode membrane-associated proteins that can be speculated to participate in the translocation process. SAG1413 is a putative integral membrane protein of unknown function. Topology prediction suggests this protein has 11 membrane-spanning domains and an extracytoplasmic loop near its N terminus. Homologues of SAG1413 are only found in selected streptococci and lactococci. Likewise, SAG1420 is a small conserved hypothetical protein of unknown function and is predicted to be an integral membrane protein with three membrane-spanning domains and a short C-terminal cytoplasmic tail. It is tempting to speculate that these two integral membrane proteins act as accessory proteins to the SAG1412 flippase. Finally, a putative lipoprotein with a possible Ca^{2+} -binding motif (SAG1419; Sutcliffe & Harrington, 2004) is encoded in the GBC locus. As lipoproteins in Gram-positive bacteria are tethered to the outer leaflet of the plasma membrane by their lipid anchor (Sutcliffe & Harrington, 2004), the positioning of this

lipoprotein at the membrane-wall interface means it can be speculated that this protein may have some role in the transfer of GBC from its lipid carrier onto the cell wall. Consistent with this, a lipoprotein-deficient GBS *lgt* mutant, which lacks the ability to post-translationally lipid modify proteins, exhibits reduced levels of GBC in its cell wall (B. A. Bray, I. C. Sutcliffe & D. J. Harrington, unpublished data).

SAG1419 is notable in having only a single homologue identifiable by BLASTP analysis, the LACR_0221/LLMG_0226 proteins of *Lactococcus lactis* subsp. *cremoris* strains (Makarova *et al.*, 2006; Wegmann *et al.*, 2007). These lactococcal proteins are encoded in a genomic region of *L. lactis* subsp. *cremoris* that also contains homologues of SAG1413 (LACR_0222/LLMG_0227) and SAG1420 (LACR_0214/LLMG_0219), enzymes for the synthesis of dTDP-rhamnose and multiple glycosyltransferases (supplementary Table S1, available with the online version of this paper), suggesting a comparable locus for the biosynthesis of a lactococcal cell-envelope polysaccharide. Several of the glycosyltransferases in this *L. lactis* subsp. *cremoris* cluster have a distinctive G + C content, suggesting their acquisition by horizontal gene transfer (Wegmann *et al.*, 2007). Analysis of the DNA of the strain 2603/V GBC locus suggests that its G + C content (about 33.8 mol%) is only marginally distinct from that of the genome as a whole (35.6 mol%), although SAG1419 and SAG1420 exhibit notably lower G + C contents (supplementary Table S1).

Following translocation, GBC is transferred from its polyisoprenol-lipid carrier and attached to the C6 of NAM residues in the glycan backbone of the peptidoglycan (Deng *et al.*, 2000). The biochemistry and genetics of this stage in teichoic acid biosynthesis are as yet unknown (Bhavsar & Brown, 2006), although it is reasonable to speculate that this step must be co-ordinated with biosynthesis of the glycan chains in peptidoglycan. The restricted species distribution of homologues of SAG1413, SAG1419 and SAG1420 suggests these are unlikely candidates for this role. Consequently, it can be suggested that the relevant transferase machinery is encoded outside the GBC locus.

Assignment of the putative α -RhaT proteins encoded in the GBC locus

Five distinct α -RhaT activities are needed to allow GBC synthesis (Table 2). We propose that SAG1411, SAG1414, SAG1415, SAG1421 and SAG1422 are these transferases, based on their homologies to other glycosyltransferases (Table 1) and as all belong to Pfam family PF00535. All of the L-RhaTs involved in GBC synthesis are predicted to use dTDP-rhamnose as the rhamnose donor substrate (see below). SAG1422 is confidently predicted to participate in the synthesis of oligosaccharide Unit IV (see above). However, the assignments of the other four RhaTs remain rather speculative. SAG1421 may be the Rha α 1-2 Rha RhaT and it is noted that homologues are present in the

loci for rhamnan-based wall polysaccharides of *S. mutans*, *Streptococcus pyogenes* and *Streptococcus equi* (data not shown). SAG1411 may transfer the α 1-2-linked rhamnose to the galactose in units I, II and IV.

The SAG1414 and SAG1415 proteins share 47% amino acid sequence identity (supplementary Table S2, available with the online version of this paper), but SAG1414 is distinguished from SAG1415 by a weakly hydrophobic region in the C-terminal third of the protein sequence (data not shown). Two of the α -RhaTs needed for GBC synthesis are required to transfer rhamnose to a glucitol acceptor (Table 2). Depending on the precise timing of the reactions (supplementary File 1) one of the acceptors will be glucitol-phosphate, whereas the other enzyme will have to recognize a rhamnose-substituted glucitol-phosphate acceptor. We propose that the related putative α -RhaTs (SAG1414 and SAG1415) carry out these two reactions. However, it is not possible at this stage to predict which of these proteins performs which reaction.

The 'missing' NAG-transferases

Consistent with the observations of Glaser *et al.* (2002) we noted that the GBC locus only contained seven putative glycosyltransferases and, as described above, the putative functions of these glycosyltransferases (along with SAG1418) can account for the formation of many of the requisite linkages in the GBC structure. However, none of these putative glycosyltransferases are predicted to be the required β -NAG transferases and so these must be encoded elsewhere in the *S. agalactiae* genome. As described above, the β -NAG transferase initiating Unit IV synthesis can be confidently predicted to be SAG0140. The identity of the two other requisite β -NAG transferases remains debatable, although many putative glycosyltransferases can be identified from the *S. agalactiae* genome annotation and/or the CAZY website (supplementary Table S3, available with the online version of this paper). Many of these can be assigned roles in other processes, such as capsular biosynthesis and peptidoglycan biosynthesis, including three glycosyltransferases that are likely to be part of a locus for the secretion and glycosylation of the large serine-rich putative glycoprotein SAG1462 (analogous to the secretion and glycosylation of GspB in *Streptococcus gordonii*; Takamatsu *et al.*, 2004).

As the two unaccounted-for β -NAG transferases required for GBC synthesis both form similar linkages (β -NAG 1-3 L-rhamnose in Unit IV and β -NAG 1-4 L-rhamnose in units I and II), and both presumably use UDP-NAG as substrate, it is probable that these β -NAG transferases will be both homologous to each other and possibly linked in terms of genome proximity. Notably, two pairs of homologous putative glycosyltransferases (SAG1459/SAG1460 and SAG2060/SAG2061) fit these criteria. SAG1459 and SAG1460 are both predicted to be cytoplasmic proteins, whereas SAG2060 is a predicted integral membrane protein and SAG2061 is a predicted peripheral membrane protein.

Precursors needed for GBC synthesis

The synthesis of GBC requires a supply of appropriately activated-sugar precursors. Examination of the GBC locus suggests that rhamnose is activated as dTDP-L-rhamnose because SAG1424 encodes a putative dTDP-4-keto-L-rhamnose reductase (RmlD). This enzyme carries out the last reaction of the four-step pathway by which D-glucose 1-phosphate can be converted to dTDP-L-rhamnose (Tsukioka *et al.*, 1997). The enzymes for the preceding three steps of this pathway (RmlA, RmlB and RmlC) are not encoded in the GBC locus. However, orthologues of these enzymes are encoded by SAG1200 (RmlA), SAG1198 (RmlB) and SAG1199 (RmlC). The *rmlCBA* locus is part of the core genome and thus a complete pathway for the synthesis of dTDP-L-rhamnose can be reconstructed for *S. agalactiae*.

The substrates for the other glycosyltransferase are predicted to be UDP-activated: UDP-galactose for the α -GalT (SAG1410) would be obtained from UDP-glucose by the action of the GalE UDP-glucose epimerase (SAG1923). Likewise, the UDP-NAG for the NAG transferases would be available from general metabolism (notably the peptidoglycan biosynthesis pathway).

The remaining precursor required for GBC synthesis would be a source of activated glucitol-phosphate. As noted above, comparisons with teichoic acid structure and biosynthesis suggest that the activated glucitol-phosphate is CDP-glucitol which we predict is synthesized by SAG1416 and SAG1417. The putative alditol-phosphate cytidylyltransferase SAG1417 can be predicted to synthesize a CDP-alditol from CTP and an alditol-phosphate (Table 1). However, it seems probable that the alditol-phosphate utilized is not glucitol-phosphate as SAG1416 encodes a putative nucleotide sugar epimerase/dehydratase (Pfam family PF01370). Thus we predict that SAG1417 synthesizes a CDP-alditol intermediate which is converted to CDP-glucitol by SAG1416. This proposal is also consistent with the apparent absence in *S. agalactiae* of a glucitol phosphotransferase (PTS) uptake system as an obvious source of glucitol-phosphate (Barabote & Saier, 2005; our observations). The core *S. agalactiae* genome includes a putative PTS system (Tettelin *et al.*, 2002; SAG1805 enzyme IIC and/or SAG1933–SAG1935) for the uptake of galactitol as galactitol-phosphate, suggesting this may be the alditol-phosphate utilized.

The boundary of the GBC locus

The final protein that may belong to the GBC locus is SAG1425. This protein belongs to Pfam family PF01883 DUF59, members of which are of unknown function but broad distribution (Almeida *et al.*, 2005). As we cannot attribute an obvious role in GBC biosynthesis to this protein, it probably represents the boundary of the GBC locus. Whilst we note that DUF59 family proteins are located proximal to putative wall polysaccharide biosynthesis loci in *S. mutans*, *S.*

pyogenes and *Streptococcus thermophilus*, this association is not consistently maintained with regard to DUF59 family proteins in other genomes [e.g. *Lactococcus lactis* (supplementary Table S1), *Streptococcus suis*] and other DUF59 proteins are associated with chromosomal loci associated with sulfur metabolism (Almeida *et al.*, 2005).

Conclusions

The present study provides a bioinformatic overview of the pathway leading to GBC synthesis in *S. agalactiae*. Re-evaluation of both the structure and the biosynthetic pathway allows us to conclude that GBC is the predominant teichoic-acid-like cell-wall polymer of *S. agalactiae* and to propose that poly(glycerophosphate) teichoic acids are likely to be absent from this organism. Further studies of GBC are important given its status as the definitive diagnostic antigen for the identification of *S. agalactiae*, especially as there is renewed interest in streptococcal type antigens as putative protective antigens (Michon *et al.*, 2005; Sabharwal *et al.*, 2006).

Importantly, in the present study we have been able to show that the GBC locus in the genomes of *S. agalactiae* strains (Table 1; Tettelin *et al.*, 2002; Glaser *et al.*, 2002; Tettelin *et al.*, 2005) can be considered necessary but not sufficient for the synthesis of GBC. Understanding the genetic and biochemical basis of this process is important as the key enzymes may provide targets for novel therapeutic and prophylactic strategies targeted very specifically at *S. agalactiae*. Moreover, GBC is an attractive target as it is present in all strains of *S. agalactiae* and does not share the inherent variation of the capsular serotypes (Cieslewicz *et al.*, 2005) or the antigenic variation of many surface proteins (Lindahl *et al.*, 2005). Bioinformatic prediction of the steps that culminate in GBC synthesis, as with any *in silico* analysis, has inherent limitations, but nonetheless provides a meaningful and testable model for this biosynthetic pathway. It is worth emphasizing, however, that definitive identification of the requisite NAG transferases (including the validation of SAG0140) must ultimately be verified experimentally. Likewise, as in teichoic acid biosynthesis, it will be important to identify the transferase enzymes which attach GBC to peptidoglycan.

Finally, given the energetic demand evidently involved in synthesizing GBC, we agree with Bhavsar & Brown (2006) that by 'far the most puzzling aspect of anionic polymer biosynthesis is the function of these abundant structures'. We hope that the hypothesis presented herein for the route to GBC biosynthesis and incorporation into the cell envelope will prompt further investigations of its contribution to the pathogenicity of *S. agalactiae*.

REFERENCES

Alaimo, C., Catrein, I., Morf, L., Marolda, C. L., Callewaert, N., Valvano, M. A., Feldman, M. F. & Aebi, M. (2006). Two distinct but

interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J* 25, 967–976.

Almeida, M. S., Herrmann, T., Peti, W., Wilson, I. A. & Wüthrich, K. (2005). NMR structure of the conserved hypothetical protein TM0487 from *Thermotoga maritima*: implications for 216 homologous DUF59 proteins. *Protein Sci* 14, 2880–2886.

Anthony, B. F., Concepcion, N. F. & Concepcion, K. F. (1985). Human-antibody to the group-specific polysaccharide of Group-B *Streptococcus*. *J Infect Dis* 151, 221–226.

Barabote, R. D. & Saier, M. H. (2005). Comparative genomic analysis of the bacterial phosphotransferase system. *Microbiol Mol Biol Rev* 69, 608–634.

Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitz, E., Collins, M., Donohoe, K., Harris, D., Murphy, L. & other authors (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2, e31.

Bhavsar, A. P. & Brown, E. D. (2006). Cell wall assembly in *Bacillus subtilis*: how spirals and spaces challenge paradigms. *Mol Microbiol* 60, 1077–1090.

Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B. & Guedon, G. (1999). Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene* 233, 151–161.

Bradley, A. J. (2002). Bovine mastitis: an evolving disease. *Vet J* 164, 116–124.

Brooks, S., Apostol, M., Nadle, J., Wymore, K., Haubert, N., Burnite, S., Daniels, A., Hadler, J. L., Farley, M. M. & other authors (2006). Early-onset and late-onset neonatal Group B streptococcal disease – United States, 1996–2004. *JAMA* 295, 1371–1372.

Cieslewicz, M. J., Chaffin, D., Glusman, G., Kasper, D., Madan, A., Rodrigues, S., Fahey, J., Wessels, M. R. & Rubens, C. E. (2005). Structural and genetic diversity of Group B *Streptococcus* capsular polysaccharides. *Infect Immun* 73, 3096–3103.

Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering*, pp. 3–12. Edited by H. J. Gilbert, G. Davies, B. Henrissat & B. Svensson. Cambridge: The Royal Society of Chemistry.

D'Elia, M. A., Millar, K. E., Beveridge, T. J. & Brown, E. D. (2006). Wall teichoic acid polymers are dispensable for cell viability in *Bacillus subtilis*. *J Bacteriol* 188, 8313–8316.

Damjanovic, M., Kharat, A. S., Eberhardt, A., Tomasz, A. & Vollmer, W. (2007). The essential *tacF* gene is responsible for the choline-dependent growth phenotype of *Streptococcus pneumoniae*. *J Bacteriol* 189, 7105–7111.

Davies, S., Gear, J. E., Mason, C. M. & McIntyre, S. M. (2003). *Streptococcus* grouping latex kits: evaluation of five commercially available examples. *Br J Biomed Sci* 60, 136–140.

Deng, L. Y., Kasper, D. L., Krick, T. P. & Wessels, M. R. (2000). Characterization of the linkage between the type III capsular polysaccharide and the bacterial cell wall of Group B *Streptococcus*. *J Biol Chem* 275, 7497–7504.

Doran, K. S., Engelson, E. J., Khosravi, A., Maisey, H. C., Fedtke, I., Equils, O., Michelsen, K. S., Arditi, M., Peschel, A. & Nizet, V. (2005). Blood–brain barrier invasion by Group B *Streptococcus* depends upon proper cell-surface anchoring of lipoteichoic acid. *J Clin Invest* 115, 2499–2507.

Edwards, M. S. & Baker, C. J. (2005). Group B streptococcal infections in elderly adults. *Clin Infect Dis* 41, 839–847.

Erbing, B., Kenne, L., Lindberg, B., Helting, T. & Hammerschmid, F. (1986). Structural studies of a teichoic acid from *Streptococcus agalactiae* Type III. *Carbohydr Res* 156, 145–155.

Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A. & other authors (2006). Pfam: clans, web tools and services. *Nucleic Acids Res* 34, D247–D251.

Follens, A., Veiga-Da-Cunha, M., Merckx, R., Van Schaftingen, E. & Van Eldere, J. (1999). *acsI* of *Haemophilus influenzae* type a capsulation locus region II encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol pyrophosphorylase. *J Bacteriol* 181, 2001–2007.

Glaser, P., Rusniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., Zouine, M., Couve, E., Lalioui, L. & other authors (2002). Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol* 45, 1499–1513.

Goldschmidt, J. C. & Panos, C. (1984). Teichoic acids of *Streptococcus agalactiae*: chemistry, cytotoxicity, and effect on bacterial adherence to human cells in tissue culture. *Infect Immun* 43, 670–677.

Gray, K. J., Bennett, S. L., French, N., Phiri, A. J. & Graham, S. M. (2007). Invasive group B streptococcal infection in infants, Malawi. *Emerg Infect Dis* 13, 223–229.

Greenberg, D. N., Ascher, D. P., Yoder, B. A., Hensley, D. M., Heiman, H. S. & Keith, J. F. (1995). Sensitivity and specificity of rapid diagnostic tests for detection of Group B streptococcal antigen in bacteremic neonates. *J Clin Microbiol* 33, 193–198.

Gutekunst, H., Eikmanns, B. J. & Reinscheid, D. J. (2003). Analysis of *RogB*-controlled virulence mechanisms and gene expression in *Streptococcus agalactiae*. *Infect Immun* 71, 5056–5064.

Heath, P. T., Balfour, G., Weisner, A. M., Esfratiou, A., Lamagni, T. L., Tighe, H., O'Connell, L. A. F., Cafferkey, M., Verlander, N. Q. & other authors (2004). Group B streptococcal disease in UK and Irish infants younger than 90 days. *Lancet* 363, 292–294.

Henneke, P. & Berner, R. (2006). Interaction of neonatal phagocytes with Group B *Streptococcus*: recognition and response. *Infect Immun* 74, 3085–3095.

Hofmann, K. & Stoffel, W. (1993). TMbase – a database of membrane spanning protein segments. *Biol Chem Hoppe Seyler* 374, 166.

Johri, A. K., Paoletti, L. C., Glaser, P., Dua, M., Sharma, P. K., Grandi, G. & Rappuoli, R. (2006). Group B *Streptococcus*: global incidence and vaccine development. *Nat Rev Microbiol* 4, 932–942.

Käll, L., Krogh, A. & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338, 1027–1036.

Kogan, G., Uhrin, D., Brisson, J. R., Paoletti, L. C., Blodgett, A. E., Kasper, D. L. & Jennings, H. J. (1996). Structural and immunochemical characterization of the type VIII Group B *Streptococcus* capsular polysaccharide. *J Biol Chem* 271, 8786–8790.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305, 567–580.

Lancefield, R. C. (1934). A serological differentiation of specific types of bovine hemolytic streptococci (Group B). *J Exp Med* 59, 441–458.

Lancefield, R. C., McCarty, M. & Everly, W. N. (1975). Multiple mouse-protective antibodies directed against group B streptococci. Special reference to antibodies effective against protein antigens. *J Exp Med* 142, 165.

Lehrer, J., Vigeant, K. A., Tatar, L. D. & Valvano, M. A. (2007). Functional characterization and membrane topology of *Escherichia coli* WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. *J Bacteriol* 189, 2618–2628.

Lindahl, G., Stålhammar-Carlemalm, M. & Areschoug, T. (2005). Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. *Clin Microbiol Rev* 18, 102–127.

- Liu, D., Cole, R. A. & Reeves, P. R. (1996). An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol* **178**, 2102–2107.
- Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M. & Weiser, J. N. (2000). The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C-reactive protein-mediated killing. *Mol Microbiol* **35**, 234–245.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V. & other authors (2006). Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* **103**, 15611–15616.
- Marolda, C. L., Tatar, L. D., Alaimo, C., Aebi, M. & Valvano, M. A. (2006). Interplay of the Wzx translocase and the corresponding polymerase and chain length regulator proteins in the translocation and periplasmic assembly of lipopolysaccharide O antigen. *J Bacteriol* **188**, 5124–5135.
- Marques, M. B., Kasper, D. L., Schroff, A., Michon, F., Jennings, H. J. & Wessels, M. R. (1994). Functional activity of antibodies to the group B polysaccharide of Group B streptococci elicited by a polysaccharide-protein conjugate vaccine. *Infect Immun* **62**, 1593–1599.
- Mattingly, S. J. & Johnston, B. P. (1987). Comparative analysis of the localisation of lipoteichoic acid in *Streptococcus agalactiae* and *Streptococcus pyogenes*. *Infect Immun* **55**, 2383–2386.
- McGinnis, S. & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20–W25.
- Michon, F., Katzenellenbogen, E., Kasper, D. L. & Jennings, H. J. (1987). Structure of the complex group-specific polysaccharide of group B *Streptococcus*. *Biochemistry* **26**, 476–486.
- Michon, F., Brisson, J. R., Dell, A., Kasper, D. L. & Jennings, H. J. (1988). Multiantennary group-specific polysaccharide of Group B *Streptococcus*. *Biochemistry* **27**, 5341–5351.
- Michon, F., Chalifour, R., Feldman, R., Wessels, M., Kasper, D. L., Gamian, A., Pozsgay, V. & Jennings, H. J. (1991). The α -L-(1→2)-trirhamnopyranoside epitope on the group-specific polysaccharide of Group B Streptococci. *Infect Immun* **59**, 1690–1696.
- Michon, F., Moore, S. L., Kim, J., Blake, M. S., Auzanneau, F. I., Johnston, B. D., Johnson, M. A. & Pinto, B. M. (2005). Doubly branched hexasaccharide epitope on the cell wall polysaccharide of group A streptococci recognized by human and rabbit antisera. *Infect Immun* **73**, 6383–6389.
- Neuhaus, F. C. & Baddiley, J. (2003). A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in Gram-positive bacteria. *Microbiol Mol Biol Rev* **67**, 686–723.
- Pincus, S. H., Cole, R. L., Wessels, M. R., Corwin, M. D., Kamangasollo, E., Hayes, S. F., Cieplak, W. & Swanson, J. (1992). Group B Streptococcal opacity variants. *J Bacteriol* **174**, 3739–3749.
- Pincus, S. H., Cole, R. L., Kamangasollo, E. & Fischer, S. H. (1993). Interaction of Group B Streptococcal opacity variants with the host-defense system. *Infect Immun* **61**, 3761–3768.
- Poyart, C., Lamy, M. C., Boumaila, C., Fiedler, F. & Trieu-Cuot, P. (2001). Regulation of D-alanyl-lipoteichoic acid biosynthesis in *Streptococcus agalactiae* involves a novel two-component regulatory system. *J Bacteriol* **183**, 6324–6334.
- Pritchard, D. G., Gray, B. M. & Dillon, H. C. (1984). Characterization of the group-specific polysaccharide of group B *Streptococcus*. *Arch Biochem Biophys* **235**, 385–392.
- Rosini, R., Rinaudo, C. D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzi, C. & other authors (2006). Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. *Mol Microbiol* **61**, 126–141.
- Sabharwal, H., Michon, F., Nelson, D., Dong, W. L., Fuchs, K., Manjarrez, R. C., Sarkar, A., Uitz, C., Viteri-Jackson, A. & other authors (2006). Group A *Streptococcus* (GAS) carbohydrate as an immunogen for protection against GAS infection. *J Infect Dis* **193**, 129–135.
- Shibata, Y., Yamashita, Y., Ozaki, K., Nakano, Y. & Koga, T. (2002). Expression and characterization of streptococcal *rgp* genes required for rhamnan synthesis in *Escherichia coli*. *Infect Immun* **70**, 2891–2898.
- Soldo, B., Lazarevic, V. & Karamata, D. (2002). *tagO* is involved in the synthesis of all anionic cell-wall polymers in *Bacillus subtilis* 168. *Microbiology* **148**, 2079–2087.
- Sutcliffe, I. C. & Harrington, D. J. (2004). Putative lipoproteins of *Streptococcus agalactiae* identified by bioinformatic genome analysis. *Antonie van Leeuwenhoek* **85**, 305–315.
- Takamatsu, D., Bensing, B. A. & Sullam, P. M. (2004). Four proteins encoded in the *gspB-secY2A2* operon of *Streptococcus gordonii* mediate the intracellular glycosylation of the platelet-binding protein GspB. *J Bacteriol* **186**, 7100–7111.
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Eisen, J. A., Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. E., Margarit, I. & other authors (2002). Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* **99**, 12391–12396.
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L. & other authors (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome’. *Proc Natl Acad Sci U S A* **102**, 13950–13955.
- Trotman, H. & Bell, Y. (2006). Neonatal group B streptococcal infection at the University Hospital of the West Indies, Jamaica: a 10-year experience. *Ann Trop Paediatr* **26**, 53–57.
- Tsukioka, Y., Yamashita, Y., Nakano, Y., Oho, T. & Koga, T. (1997). Identification of a fourth gene involved in dTDP-rhamnose synthesis in *Streptococcus mutans*. *J Bacteriol* **179**, 4411–4414.
- UniProt Consortium (2007). The universal protein resource (UniProt). *Nucleic Acids Res* **35**, D193–D197.
- Vallejo, J. G., Baker, C. J. & Edwards, M. S. (1996). Role of the bacterial cell wall and capsule in induction of tumor necrosis factor alpha by Type III Group B streptococci. *Infect Immun* **64**, 5042–5046.
- Wegmann, U., O’Connell-Motherwy, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., Ventura, M., Goesmann, A., Gasson, M. J. & other authors (2007). Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J Bacteriol* **189**, 3256–3270.
- Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**, 39–68.
- Yamashita, Y., Tsukioka, Y., Tomihisa, K., Nakano, Y. & Koga, T. (1998). Genes involved in cell wall localization and side chain formation of rhamnose-glucose polysaccharide in *Streptococcus mutans*. *J Bacteriol* **180**, 5803–5807.
- Yamashita, Y., Shibata, Y., Nakano, Y., Tsuda, H., Kido, N., Ohta, M. & Koga, T. (1999). A novel gene required for rhamnose-glucose polysaccharide synthesis in *Streptococcus mutans*. *J Bacteriol* **181**, 6556–6559.
- Yoshida, Y., Ganguly, S., Bush, C. A. & Cisar, J. O. (2006). Molecular basis of L-rhamnose branch formation in streptococcal coaggregation receptor polysaccharides. *J Bacteriol* **188**, 4125–4130.
- Zhang, J. R., Idanpaan-Heikkilä, I., Fischer, W. & Tuomanen, E. I. (1999). Pneumococcal *licD2* gene is involved in phosphorylcholine metabolism. *Mol Microbiol* **31**, 1477–1488.

Edited by: D. W. Ussery