Genetic and functional analyses of the *lgtH* gene, a member of the β -1,4-galactosyltransferase gene family in the genus *Neisseria*

Peixuan Zhu,[†] Robert A. Boykins and Chao-Ming Tsai

Correspondence

Peixuan Zhu
pzhu@creativmicrotech.com

Division of Bacterial, Parasitic and Allergenic Products, Center for Biologics Evaluation and Research, FDA, 8800 Rockville Pike, Bethesda, MD 20892, USA

Lipooligosaccharide (LOS) is a major virulence factor of the pathogenic *Neisseria*. Three galactosyltransferase genes, *lgtB*, *lgtE* and *lgtH*, responsible for the biosynthesis of LOS oligosaccharide chains, were analysed in five *Neisseria* species. The function of *lgtH* in *Neisseria meningitidis* 6275 was determined by mutagenesis and chemical characterization of the parent and mutant LOS chains. The chemical characterization included SDS-PAGE, immunoblot, hexose and mass spectrometry analyses. Compared with the parent LOS, the mutant LOS lacked galactose, and its oligosaccharide decreased by three or four sugar units in matrix-assisted laser desorption ionization (MALDI)-MS analysis. The results show that *lgtH* encodes a β -1,4-galactosyltransferase, and that the glucose moiety linked to heptose (Hep) in the α chain is the acceptor site in the biosynthesis of *Neisseria* LOS. To understand the sequence diversity and relationships of *lgtB*, *lgtE* and *lgtH*, the entire *lgt-1* locus was further sequenced in three *N. meningitidis* strains and three commensal *Neisseria* strains, and compared with the previously reported *lgt* genes from *Neisseria* species. Comparison of the protein sequences of the three enzymes LgtB, LgtE and LgtH showed a conserved N-terminal region, and a highly variable C-terminal region, suggesting functional constraint for substrate and acceptor specificity, respectively. The analyses of allelic variation and evolution of 23 *lgtB*, 12 *lgtE* and 14 *lgtH* sequences revealed a distinct evolutionary history of these genes in *Neisseria*. For example, the splits graph of *lgtE* displayed a network evolution, indicating frequent DNA recombination, whereas splits graphs of *lgtB* and *lgtH* displayed star-tree-like evolution, indicating the accumulation of point mutations. The data presented here represent examples of the evolution and variation of prokaryotic glycosyltransferase gene families. These imply the existence of multiple enzyme isoforms for biosynthesis of a great diversity of oligosaccharides in nature.

Received 6 July 2005

Revised 26 September 2005

Accepted 29 September 2005

[†]Present address: Creativ MicroTech, Inc., 11609 Lake Potomac Drive, Potomac, MD 20854, USA.

Abbreviations: ACTH, adrenocorticotrophic hormone; DHB, 2,5-dihydroxybenzoic acid; Gal, galactose; Glc, glucose; GlcN, glucosamine; GlcNAc, *N*-acetylglucosamine; Hep, heptose; Kdo, 2-keto-3-deoxyoctulosonic acid; LOS, lipooligosaccharide; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; NeuNAc, *N*-acetylneuramic acid; OS, oligosaccharide; PEA, phosphorylethanolamine.

The GenBank/EMBL/DDBJ accession numbers for the nucleotide sequences of the *lgt* genes reported in this paper are: strains 2001-083 and 2001-117, accession nos AY869717 and AY869718; strains 81186, 85071 and 87043, accession nos AY134876–AY134878.

Protein polymorphisms and physical properties for LgtB, LgtE and LgtH are shown in Supplementary Table S1, confirmation of the *lgtH* isogenic mutant in Supplementary Fig. S1, hexose composition analysis of oligosaccharide preparations from wild-type strain 6275 and isogenic mutant 6275 Δ *lgtH* in Supplementary Fig. S2, and detailed polymorphisms of 47 sequences of LgtB, LgtE and LgtH in Supplementary Fig. S3, available with the online version of this paper.

INTRODUCTION

Lipooligosaccharide (LOS) is a major virulence factor of two pathogenic *Neisseria* species, *Neisseria meningitidis* and *Neisseria gonorrhoeae* (Preston *et al.*, 1996; Kahler & Stephens, 1998). In contrast to the LPS in many enteric bacteria, *N. meningitidis* and *N. gonorrhoeae* LOS chains lack repeating O chains, and possess variable oligosaccharide (OS) chains. These OS chains are branched, and their branches are designated α , β and γ chains (Verheul *et al.*, 1993; Kahler & Stephens, 1998). The *Neisseria* LOS is heterogeneous, and the expression of LOS in the bacteria is subject to phase variation (Danaher *et al.*, 1995; Jennings *et al.*, 1995; Yang & Gotschlich, 1996; Burch *et al.*, 1997).

Multiple genes encoding glycosyltransferases are involved in the biosynthesis of *Neisseria* LOS (Fig. 1a). The genes responsible for the LOS chains are designated *lgtA*, *B*, *C*, *D*, *E*, *F*, *G*, *H* and *Z* (Gotschlich, 1994; Banerjee *et al.*, 1998;

Fig. 1. (a) Schematic presentation of *N. meningitidis* L3-immunotype LOS, and its biosynthetic genes (modified from Fig. 1 in Zhu *et al.*, 2002). The site of the *lgtH* gene (underlined) is indicated. The *lst* gene encodes an α -2,3-sialyltransferase for sialylation of LOS (Gilbert *et al.*, 1996). Gly is attached to the 7 position of HepII (Cox *et al.*, 2002). An immunotype L7 LOS is structurally very similar to L3 LOS, but without *N*-acetylneuramic acid (NeuNAc) (Kogan *et al.*, 1997). (b) Construction of the *lgtH* isogenic mutant. Type VII refers to the genetic organization at the *lgt-1* locus (Zhu *et al.*, 2002). The *lgtH* gene in *N. meningitidis* L3-LOS strain 6275 was inactivated by *BbsI* restriction digestion, followed by the insertion of a *HincII*-digested kanamycin-resistance cassette.

Zhu *et al.*, 2001). These *lgt* genes are found at three genetic loci (*lgt-1*, -2 and -3) in the bacterial chromosome. The *lgtA*, *B*, *C*, *D* and *E* genes at the *lgt-1* locus are responsible for the biosynthesis of the α chain. The *lgtF* and *rfaK* genes at the *lgt-2* locus are involved in the biosynthesis of α and γ chains, respectively (Kahler *et al.*, 1996a, b; van der Ley *et al.*, 1997). The *lgt-3* locus has an *lgtG* gene encoding α -1,3-glucosyltransferase for biosynthesis of the β chain (Banerjee *et al.*, 1998). The *lgtH* and *lgtZ* genes at the *lgt-1* locus in *N. meningitidis* are homologous to the *lgtB* and *lgtA* genes; however, their biological function remains to be determined (Zhu *et al.*, 2001).

At the *lgt-1* locus of *N. gonorrhoeae* strain F62, *lgtB*, *lgtC* and *lgtE* have been identified as galactosyltransferase genes (Gotschlich, 1994). The *lgtC* gene encodes α -1,4-galactosyltransferase, which catalyses the addition of galactose (Gal) to Gal for the biosynthesis of an alternative α chain. The sequence of *lgtC* is unrelated to others in the locus. However, the *lgtB* and *lgtE* genes are closely related in their sequences, and they have 76.2% homology in *N. gonorrhoeae* strain F62; both genes code for β -1,4-galactosyltransferase, but they add Gal to different acceptors for the biosynthesis of the α chain. *LgtB* is responsible for the addition of Gal to *N*-acetylglucosamine (GlcNAc), while *LgtE* is responsible

for the addition of Gal to glucose (Glc) (Gotschlich, 1994). The function of the orthologous *lgtB* and *lgtE* genes in *N. meningitidis* is similar to that in *N. gonorrhoeae* (Jennings *et al.*, 1995; Wakarchuk *et al.*, 1996). The *lgtH* gene was initially observed in *N. meningitidis* strains Z2491 and A1 (Parkhill *et al.*, 2000; Zhu *et al.*, 2001), and it has 80.4 and 72.8% homology to the *lgtB* and *lgtE* genes of *N. meningitidis* MC58 (Tettelin *et al.*, 2000), respectively. A further analysis has shown that 21/29 (72.4%) of *N. meningitidis* strains, including the seven LOS immunotypes L2, L3, L8, L9, L10, L11 and L12, have an *lgtH* gene (Zhu *et al.*, 2002). Interestingly, *lgtH* and *lgtE* are located at the same position of the *lgt-1* locus, and their presence is mutually exclusive (Zhu *et al.*, 2002). The nucleotide sequence of the *lgt-1* locus, which contains *lgtH*, from nine strains of *N. meningitidis* (6275, M986, BB305, A1, M120, Z2491, 7880, 7889 and 7897) has been determined (Parkhill *et al.*, 2000; Zhu *et al.*, 2001, 2002). Furthermore, the presence of *lgtH* is also observed in group C strains of *N. meningitidis* associated with both endemic and epidemic meningococcal disease in Canada (Tsang *et al.*, 2004), and in three species of commensal *Neisseria*, *Neisseria lactamica*, *Neisseria subflava* and *Neisseria polysacchareae*, by PCR and DNA hybridization (Zhu *et al.*, 2002), but the nucleotide sequences of *lgtH* in these strains have not been determined to date.

Study of the three closely related genes *lgtB*, *lgtE* and *lgtH* in *Neisseria* may present an opportunity for understanding the evolution and variation of prokaryotic glycosyltransferase gene families. Since the biological function of *lgtH* remains unclear, we first conducted a functional study of *lgtH* using mutagenesis as an approach, and then determined the nucleotide sequence of the *lgt-1* locus, which contains the *lgtA*, *lgtB* and *lgtH* genes, from three representative strains of *N. meningitidis* that were associated with endemic and epidemic meningococcal disease in Canada (Ashton *et al.*, 1991; Tsang *et al.*, 2004). In addition, the nucleotide sequences of the *lgt* genes in three representative commensal *Neisseria* strains of *N. lactamica*, *N. subflava* and *N. polysacchareae* were determined. Furthermore, the sequence data were compared to all sequences of *lgtB*, *lgtE* and *lgtH* available in the GenBank database. The distribution, variation and evolution of the three genes were evaluated.

METHODS

Bacterial strains, growth conditions and DNA isolation.

N. meningitidis LOS prototype strains 6275 (L3), M986 (L3,7), and a truncated-LOS variant M986 OP–, were obtained from the culture collection of Dr C. E. Frasch (Center for Biologics Evaluation and Research, FDA, Bethesda, MD, USA). Three strains of commensal *Neisseria* (81186, 85071, 87043), and three *N. meningitidis* strains (2001-083, 2001-114, 2001-117) were obtained from the culture collection of the National Microbiology Laboratory, Population and Public Health Branch, Health Canada, Canada (Tsang *et al.*, 2004). *N. meningitidis* strains 2001-083 and 2001-117, of phenotype ET-15, C:2a:1,7,1, were isolated from Quebec and Ontario, respectively; these two strains were associated with an outbreak of meningococcal disease in the Quebec area in 2001 (Tsang *et al.*, 2004). Strain 2001-114, of phenotype ET-15, C:2a:P1.5,2, was isolated in Ontario, and has been considered endemic in Canada since the 1980s (Ashton *et al.*, 1991). The bacterial strains were grown on Brain Heart Infusion (BHI) agar containing 1% horse serum at 37 °C in 5% CO₂ for 16 h. The chromosomal DNA was isolated as described by Sarkari *et al.* (1994).

PCR, DNA sequencing and computer analysis. The primers, PCR amplification, and DNA sequencing for the *lgt-1* locus were as previously reported (Zhu *et al.*, 2002). DNA sequences were analysed with the Genetics Computer Group package GCG10.2-Unix (University of Wisconsin, USA) and the molecular evolutionary genetics analysis software MEGA2.1 (Arizona State University, USA) (Devereux *et al.*, 1984; Kumar *et al.*, 1994). Splits decomposition analysis was performed using the program SplitsTree (<http://bibiserv.techfak.uni-bielefeld.de/splits/>) (Huson, 1998). Physical properties of proteins (*M_r*, pI and net charge) were calculated with Peptide-Sort (Genetics Computer Group), and structural properties were determined with PredictProtein (<http://www.embl-heidelberg.de/predictprotein/>). Protein motifs were searched with the program ScanProsite (<http://us.expasy.org/tools/scanprosite/>). Protein conserved domains were searched using the CD-Search program (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>).

Construction of the *lgtH* mutant. A fragment of 807 bp containing the entire coding region of *lgtH* was amplified from strain A1 of *N. meningitidis* by PCR using primers P23 (5'-ATGCAAAACAC-GTTATCAG-3') and P26 (5'-CTATTCAAACGGGACAACG-3'). After purification through a QIAquick spin-column (QIAGEN), the PCR product was cloned into a pCR T7/NT-TOPO plasmid

(Invitrogen). The *lgtH* insert in recombinant plasmid pEH1 was confirmed by nucleotide sequencing. The pEH1 plasmid was digested with the restriction endonuclease *Bbs*I, which cuts the *lgtH* gene at position 315 after the ATG start codon. The overhangs at the ends of the DNA fragments were filled in with T4 DNA polymerase (Promega), and the fragments were purified through a QIAquick spin-column, and then added to a *Hinc*II-digested kanamycin-resistance cassette from pUC4Kan (Amersham Pharmacia). The blunt ends were ligated using T4 DNA ligase (Promega). After transformation into *Escherichia coli* TOPO10F' cells, and selection on kanamycin-containing plates, colonies were examined by PCR for constructs with the kanamycin cassette in the same transcriptional orientation as the *lgtH* gene. One such plasmid construct, pEHK1, was linearized by digestion with *Eco*RI, and used to transform strain 6275 of *N. meningitidis* by electroporation, according to the method of A. Elizabeth (Elizabeth & Seifert, 2001). Cells were plated on BHI agar containing 50 µg kanamycin ml⁻¹, and transformants were selected.

Southern blot hybridization and PCRs were performed to confirm the *lgtH* mutants. In the Southern blot hybridization assay, chromosomal DNA was isolated from the mutant and the wild-type strain. Chromosomal DNA was digested with *Bam*HI, subjected to electrophoresis on a 0.7% agarose gel, and transferred to a positively charged nylon membrane (Roche). The probe was generated from the puc4Kan template by PCR using primers P101 (5'-GACAATCTATCGATTGTATG-3') and P102 (5'-GTCCAACATCAATACAACCT-3'), and the DIG PCR Probe Synthesis Kit (Roche). Hybridization and washes were performed using the DIG Nucleic Acid Detection Kit (Roche), according to the manufacturer's instructions. PCR comparisons of the *lgt-1* locus in the mutant and the wild-type strain were performed using primers P23 (5'-ATGCAAAACCATGTTATCAG-3'), P26 (5'-CTATTCAAACGGGACAACG-3'), P36 (5'-GATCAGGCTGCCCAATGCGC-3'), P37 (5'-TAGCCGCGCTCGTGAATCAG-3') and P39 (5'-GCGTGTCCGTCGGTTTGAAG-3').

Isolation of LOS, SDS-PAGE and Western blotting. LOS samples were isolated from wild-type and mutant strains grown in Catlin medium (Tsai *et al.*, 1983) by the hot phenol/water extraction procedure (Westphal & Jann, 1965). The samples were further treated with 1 µg ml⁻¹ each of RNase A and DNase I (Sigma) at 37 °C for 30 min, and then digested with 1 mg proteinase K ml⁻¹ at 60 °C for 2 h. The samples were extracted with phenol/water again, and the LOS in the aqueous phase was precipitated by 2 vols acetone, and washed twice with 95% ethanol. Analysis of LOS by SDS-PAGE, silver staining and Western blotting was performed as described previously (Tsai & Frasch, 1982; Tsai *et al.*, 2002). Meningococcal mouse mAb 9-2-L3,7 for Western blotting was a generous gift from Dr W. D. Zollinger of the Walter Reed Army Institute of Research, Washington, DC, USA, and its binding specificity with LOS immunotypes has been described elsewhere (Scholten *et al.*, 1994).

Sugar composition analysis. LOS (2 mg ml⁻¹) in 1% acetic acid was hydrolysed at 100 °C in a water bath for 2 h to produce OS, 2-keto-3-deoxyoctulosonic acid (Kdo) and lipid A. The lipid A was removed by centrifugation at 500 000 g for 1 h using a Beckman Optima TLX ultracentrifuge. The supernatant containing the OS and Kdo was lyophilized to produce an OS preparation. The OS preparation was hydrolysed in 2 N trifluoroacetic acid (TFA) at 98 ± 2 °C for 6 h to produce monosaccharides. The TFA was removed by drying in a SpeedVac (ThermoSavant). The dried hydrolysate was dissolved in water, and hexoses in the sample were analysed by HPLC, using a Dionex PA1 column, and 14 mM NaOH as the column eluant (Hardy *et al.*, 1988).

MS analysis. LOS preparations were first O-deacylated by treatment with hydrazine at 37 °C for 30 min, followed by precipitation with cold acetone (Helander *et al.*, 1988). Matrix-assisted laser

desorption ionization time-of-flight (MALDI-TOF) MS analysis of the O-deacylated LOSs (Gibson *et al.*, 1997) was performed on a Voyager model DERP mass spectrometer (Applied Biosystems). Mass spectra were generated using a matrix of 2,5-dihydroxybenzoic acid (DHB). The matrix was prepared by dissolving 10 mg DHB in 1.0 ml HPLC-grade water. Samples were spotted on a stainless-steel target plate by adding 1.0 µl sample (1–10 mg ml⁻¹ in water) initially to the plate, followed by addition of 1.0 µl matrix. A typical spectrum was generated in the linear detector with delayed extraction using the negative ionization mode. The accelerating voltage was set at 20 kV, with the number of laser shots equal to 128. The grid voltage was 93.5%. The guide wire was set at 0.07%, and the extraction delay time was set at 150 ns; laser intensity was set at 1600–1850. External calibration was performed on the mass spectrometer using mass calibration standards (Applied Biosystems): angiotensin (M_{avg} 1297.5), adrenocorticotrophic hormone (ACTH) 1–17 (M_{avg} 2094.5), ACTH 18–39 (M_{avg} 2466.7) and ACTH 7–38 (M_{avg} 3660.2).

RESULTS

Mutagenesis of the *lgtH* gene results in altered LOS structure

In order to determine the role of the *lgtH* gene product, we constructed a knockout mutant of this gene by insertion of a kanamycin-resistance cassette into the *Bbs*I restriction site, which is located at position 315–318 after the ATG start codon of the 807 bp coding region (Fig. 1b). *N. meningitidis* LOS prototype strain 6275 (L3) was used for construction of the *lgtH* isogenic mutants. One mutant, designated 6275Δ*lgtH*, was verified by Southern blot hybridization and PCR (Supplementary Fig. S1). The LOS from the isogenic mutant 6275Δ*lgtH* was compared with that from wild-type strain 6275. Examination of silver-stained SDS-PAGE gels revealed that the LOS from the mutant strain increased its mobility compared with the wild-type LOS (Fig. 2a, lane 4), indicating a shortened sugar chain in the mutant LOS. A decrease in the molecular sizes of LOSs was observed: 4.1 kDa for wild-type strain 6275, and 3.6 kDa for the mutant 6275Δ*lgtH* (Fig. 2a, lanes 3 and 4). The doublet band of 6275 LOS showed the presence of both sialylated and non-sialylated forms, and that the heavy upper doublet was the sialylated species. Immunoblotting showed that the LOS from mutant 6275Δ*lgtH* lost its reactivity to the antibody to L3 LOS (Fig. 2b). These LOS alterations were similar to the LOSs from two other strains, M986 (L3,7) and its LOS-truncated variant M986 OP– (Quakyi *et al.*, 1997). M986, the wild-type strain, had a full-length α chain of L3,7 LOS, and the M986 OP– variant had no Gal; the latter probably had the α chain terminated at the Glc moiety, similar to 6275Δ*lgtH* LOS (Fig. 2, lanes 1 and 2).

Lack of Gal in the *lgtH* mutant LOS

The OS preparations from the LOSs of strain 6275 and its *lgtH* mutant were acid hydrolysed, and the hydrolysates were analysed for hexoses (Supplementary Fig. S2). The wild-type strain showed glucosamine (GlcN), Gal and Glc in a molar ratio of approximately 1.4:2:1. The expected molar ratio of GlcN, Gal and Glc is 2:2:1, based on the

Fig. 2. SDS-PAGE and Western blot analysis of LOS samples. (a) The silver-stained SDS-polyacrylamide gel contains LOS preparations from the L3,7 wild-type strain *N. meningitidis* M986 (lane 1), a truncated-LOS variant OP– of M986 (lane 2), the L3 wild-type strain 6275 (lane 3), and the isogenic mutant 6275Δ*lgtH* (lane 4). (b) The Western blot, containing the same LOS preparations as in (a), was probed with a monoclonal antibody, 9-2-L3,7. This antibody binds both L3 and L7 LOSs; the structure of L7 LOS is very similar to that of the L3 LOS in Fig. 1, but without NeuNAc (Kogan *et al.*, 1997).

reported 6275 OS structure (Yamasaki *et al.*, 1993). The GlcN content was lower than the expected value because the glycosidic linkage between GlcNAc and heptose (Hep)II in the γ chain (Fig. 1a) is rather resistant to acid hydrolysis, and GlcN is only partially released from HepII, as previously reported (Jennings *et al.*, 1980; Tsai *et al.*, 1983). By comparison, in the mutant OS, Gal was missing, but Glc was unchanged. GlcN, which was solely from GlcNAc in the γ chain, was greatly reduced for the reason mentioned above. The lack of Gal in the mutant LOS indicates that *lgtH* encodes a galactosyltransferase, and that the Glc in the α chain is the acceptor site for LgtH, because the Glc moiety is the attachment site for the first Gal in the stepwise extension of the α chain.

MALDI-MS analysis of wild-type and *lgtH*-mutant LOS chains

Molecular masses of O-deacylated LOSs from wild-type strain 6275 and mutant 6275Δ*lgtH* were analysed using MALDI-MS to determine the number of sugar units lost from the α chain of LOS by *lgtH* mutation. Fig. 3(a) shows that the wild-type strain 6275 had two major pairs of molecular ions, a/A (3207.60/3330.56) and b/B (2916.61/3039.25); a molecular mass of 3208 has been reported

Fig. 3. MALDI-TOF analysis of O-deacylated LOS from (a) the wild-type strain 6275, and (b) the isogenic mutant 6275Δ*lgtH*.

elsewhere for this strain (Gibson *et al.*, 1993). The mass difference between the two pairs a/A and b/B is 291 (i.e. a–b and A–B), which corresponds to the mass of one sialic acid residue. Thus, the pair a/A represents sialylated species, and the pair b/B represents non-sialylated species. The mass difference within a pair (i.e. A–a or B–b) is 123, which corresponds to the mass of one phosphorylethanolamine (PEA) moiety. The *lgtH* mutant (Fig. 3b) had a much smaller major molecular ion, with a mass of 2388.77. The major glycoforms of the parent and mutant O-deacylated LOS chains are presented in Table 1, which includes the above observed molecular ion masses $[M-H]^-$, corresponding compositions, and calculated ion masses. Other less-abundant ions represent sodium adducts (+23) to the major ions, such as 3353.1 (3330.6 + Na⁺) and 3230.8 (3207.6 + Na⁺) in the parent strain, or loss of H₂O (–18) from ions, such as 2370.5 (2388.8 – H₂O) in the mutant. Cleavage of the labile glycosidic bond between Kdo and

lipid A during the ionization process resulted in OS and lipid A fragments. For the parent strain, the ions 1919.57, 2211.71 and 2254.61 were OS fragments, and 952.7 (Fig. 3, first ion peak on the left, unlabelled) and 1075.9 were lipid A fragments. The ion 952.7 corresponds to the di-glucosamine unit with two phosphate groups and two N-linked β-hydroxymyristic acids (Gibson *et al.*, 1997). The major lipid A species 1075.93 is one PEA (123) added to one of the phosphate groups of the lipid A with a mass of 953, probably at the di-glucosamine reducing end (Kulshin *et al.*, 1992). For the mutant, 952.2 and 1435.2 were lipid A and OS ions, respectively.

The hexose analysis in the preceding section showed that the OS of the *lgtH* mutant lacked Gal. The MALDI-MS analysis in Table 1 provides conclusive evidence that the mutant LOS terminated the α chain at the Glc moiety. Based on these results, and the 6275 OS structure in the literature

Table 1. Major molecular ions [M–H][–], and proposed compositions of O-deacylated LOS (O-dLOS) chains from *N. meningitidis* strains in MALDI-MS analysis

Average mass units of monosaccharides and other components for calculating molecular ions [M–H][–]: Gal, 162·14; GlcNAc, 203·19; NeuNAc, 291·26; PEA, 123·05; Hep, 192·17; Kdo, 220·18. Lipid A is O-deacylated, and contains two N-linked β-hydroxymyristic acids and two phosphates, with a total molecular mass of 953·01 (Gibson *et al.*, 1997).

Strain	O-dLOS*	[M–H] [–] observed (calculated)	Proposed composition
6275 wild-type	A	3330·6 (3329·9)	NeuNAc.Gal.GlcNAc.Gal.Glc.Hep ₂ .GlcNAc.PEA ₃ .Kdo ₂ .lipid A
	a	3207·6 (3206·9)	NeuNAc.Gal.GlcNAc.Gal.Glc.Hep ₂ .GlcNAc.PEA ₂ .Kdo ₂ .lipid A
	B	3039·3 (3038·6)	Gal.GlcNAc.Gal.Glc.Hep ₂ .GlcNAc.PEA ₃ .Kdo ₂ .lipid A
	b	2916·6 (2915·6)	Gal.GlcNAc.Gal.Glc.Hep ₂ .GlcNAc.PEA ₂ .Kdo ₂ .lipid A
6275Δ <i>lgtH</i>	C	2388·8 (2388·1)	Glc.Hep ₂ .GlcNAc.PEA ₂ .Kdo ₂ .lipid A

*The major O-dLOS chains in Fig. 3.

(Yamasaki *et al.*, 1993), we conclude that *lgtH* encodes a galactosyltransferase that transfers Gal from UDP-Gal to the Glc moiety in the α chain of LOS.

Sequencing the *lgt-1* locus from *N. meningitidis* and commensal *Neisseria*

It was recently reported that six *N. meningitidis* strains, associated with an outbreak of meningococcal disease in the Quebec area in 2001, have the *lgtA*, *lgtB* and *lgtH* genes at the *lgt-1* locus (LOS genotype VII) (Tsang *et al.*, 2004); these strains belong to the unique phenotypic variant ET-15, C:2a:1.7,1, which differs from another strain (ET-15, C:2a:P1.5,2) that has been considered endemic in Canada since the 1980s (Ashton *et al.*, 1991). To examine the genetic variation of the *lgt* genes in these meningococcal disease strains, the nucleotide sequences of the *lgt-1* locus were determined in three representative strains: 2001-083, 2001-117 (C:2a:1.7,1) and 2001-114 (C:2a:P1.5,2). The sequences of 3402 bp of the locus *lgt-1* from strains 2001-083 and 2001-114 were identical. The sequence of strain 2001-083 was deposited under GenBank accession number AY869717. Three ORFs were identified at the locus *lgt-1*. Two of the ORFs had 93·3 % and 93·9 % homology to *lgtA* and *lgtB* of gonococcal strain F62, respectively. One of the ORFs had 95·4 % homology to *lgtH* of meningococcal strain A1. The sequence of 3402 bp from the *lgt-1* locus in strain 2001-117 (GenBank accession no. AY869718) differed from that of strains 2001-083 and 2001-114 by one nucleotide in *lgtH* (C to T). This is a nonsynonymous mutation, which means that substitution in the DNA changes the amino acid sequence of resulting protein. This C to T mutation resulted in an amino acid substitution of LgtH from Thr in strain 2001-083 to Met in strain 2001-117.

To understand the sequence diversity of the *lgtH* gene in commensal *Neisseria* species, the nucleotide sequences of the *lgt-1* locus were determined in three representative commensal *Neisseria* strains: *N. lactamica* 81186, *N. subflava* 85071 and *N. polysacchareae* 87043. The three sequences were deposited under GenBank accession nos AY134876–AY134878. Based on the eight types of genetic organization

reported at the *lgt-1* locus (Zhu *et al.*, 2002), strains 81186 and 85071 belong to LOS genotype II (*lgtA*, *B*, *C*, *D* and *H*), and strain 87043 belongs to genotype VII (*lgtA*, *B* and *H*).

Allelic variation of the *lgtB*, *lgtE* and *lgtH* genes in *Neisseria*

Each ORF comprising a distinctive nucleotide sequence was assigned an arbitrary allele designation for the *lgt* genes (Zhu *et al.*, 2002). The sizes and allelic variations are summarized in Table 2. Among the 23 *lgtB* genes sequenced, 20 distinct alleles were distinguished. Based on the size of the coding region, the *lgtB* genes were divided into two groups, with lengths of 828 and 840 bp. The 828 bp group (alleles 8, 9, 10, 11, 12, 13, 14, 15 and 16) had a deletion of 12 nucleotides (TAATCGGCAAGA) at 17 bp before the TAA stop codon. The *lgtB* gene from *N. gonorrhoeae* strain 1291 was an exception, with 839 bp resulting from the deletion of one cytidine at position 744. In 12 *lgtE* sequences, 11 distinct alleles were observed. All *lgtE* genes had the same size of 843 bp, except that the gene of *N. meningitidis* MC58 had 831 bp. Of 14 *lgtH* genes, 13 distinct alleles were identified; no size variation was observed in the 14 *lgtH* genes. Totals of 184 (21·9 %), 122 (14·5 %) and 88 (10·9 %) polymorphic sites were observed for the aligned 23 *lgtB*, 12 *lgtE* and 14 *lgtH* sequences, respectively.

Splits decomposition analysis of *lgtB*, *lgtE* and *lgtH*

Splits decomposition analysis was used to investigate the influence of recombination on the evolution of each gene. Splits decomposition constructs an interconnecting network structure when recombination plays a role in the evolutionary history. The genetic distances for *lgtB*, *lgtE* and *lgtH* were calculated by using the Kimura two-parameter method, as implemented in MEGA (Kumar *et al.*, 1994), and a splits decomposition analysis was performed using the SplitsTree program (Huson, 1998). The splits graphs from the *lgtB*, *lgtE* and *lgtH* genes are shown in Fig. 4. A star-like tree structure was observed in the splits graph of the *lgtB* and *lgtH* gene, showing no recombination signal (Fig. 4a, c),

Table 2. Allelic variation of the *lgtB*, *lgtE* and *lgtH* genes

Species	Strain	<i>lgtB</i>		<i>lgtE</i>		<i>lgtH</i>		GenBank accession number (reference)
		CDS*	Allele†	CDS	Allele	CDS	Allele	
<i>N. meningitidis</i>	126E	—	—	843	5	—	—	U65788 (Jennings <i>et al.</i> , 1999)
	35E	840	4	843	4	—	—	AF470655 (Zhu <i>et al.</i> , 2002)
	6275	828	8	—	—	807	6	AF470656 (Zhu <i>et al.</i> , 2002)
	MC58	828	12	831	11	—	—	U25839 (Jennings <i>et al.</i> , 1995)
	M986	828	8	—	—	807	5	AF470657 (Zhu <i>et al.</i> , 2002)
	BB305	828	11	—	—	807	4	AY039684 (Zhu <i>et al.</i> , 2002)
	89I	828	16	843	6	—	—	AF470658 (Zhu <i>et al.</i> , 2002)
	M981	828	13	843	8	—	—	AF470659 (Zhu <i>et al.</i> , 2002)
	M992	828	15	843	6	—	—	AF470660 (Zhu <i>et al.</i> , 2002)
	6155	828	10	843	7	—	—	AF470661 (Zhu <i>et al.</i> , 2002)
	M978	828	9	843	9	—	—	AF355193 (Zhu <i>et al.</i> , 2002)
	A1	—	—	—	—	807	2	AF355194 (Zhu <i>et al.</i> , 2002)
	M120	840	7	—	—	807	7	AF470662 (Zhu <i>et al.</i> , 2002)
	Z2491	840	5	—	—	807	1	AL162753 (Zhu <i>et al.</i> , 2002)
	7880	840	7	—	—	807	8	AF470663 (Zhu <i>et al.</i> , 2002)
	7889	—	—	—	—	807	3	AF470664 (Zhu <i>et al.</i> , 2002)
	7897	840	6	—	—	807	3	AF470665 (Zhu <i>et al.</i> , 2002)
	2001-083	840	20	—	—	807	12	AY869717 (this study)
	2001-117	840	20	—	—	807	13	AY869718 (this study)
<i>N. gonorrhoeae</i>	F62	840	1	843	1	—	—	U14554 (Gotschlich, 1994)
	FA1090	840	3	843	3	—	—	AE004969 (Lewis <i>et al.</i> , 2003)
	1291	839	2	843	2	—	—	AF121135 (Harvey <i>et al.</i> , 2000)
<i>N. subflava</i>	Ns44	828	14	843	10	—	—	AF240672 (Arking <i>et al.</i> , 2001)
	85071	840	19	—	—	807	11	AY134877 (this study)
<i>N. lactamica</i>	81186	840	17	—	—	807	9	AY134876 (this study)
<i>N. polysacchara</i>	87043	840	18	—	—	807	10	AY134878 (this study)
Total	26	23		12		14		

*Length of the coding region (bp).

†Allele number.

whereas an interconnecting network structure was observed in the splits graph of the *lgtE* gene (Fig. 4b), indicating that homologous recombination occurred.

Protein variation of LgtB, LgtE and LgtH

A multiple alignment was performed on the 47 deduced protein sequences, including 12 LgtE, 14 LgtH and 21 LgtB sequences. The conserved domains and polymorphic positions are indicated on a representative alignment of three sequences of LgtB, LgtE and LgtH from *N. meningitidis* strains MC58 and 6275 (Fig. 5). Three conserved motifs that have previously been reported as consensus sequence features of the GT25 family of proteins (Heinrichs *et al.*, 1998) were highlighted in the alignment (Arg¹⁴-Arg¹⁵, Phe²⁹-Gln³⁰-Phe³¹-Phe³²-Asp³³, Glu⁸⁸-Asp⁸⁹-Asp⁹⁰). A DXD motif conserved in many GT families (Saxena *et al.*, 1995; Busch *et al.*, 1998; Wiggins & Munro, 1998) was observed (Asp¹¹²-Xaa¹¹³-Asp¹¹⁴). This pair of aspartates was surrounded by hydrophobic residues; however, the four continuing hydrophobic residues were at the C-terminal

side in LgtB, LgtE and LgtH, rather than at the N-terminal side in other GT families. In addition, all the LgtH proteins had an identical length of 268 aa, whereas the LgtB and LgtE proteins had length variations resulting from insertions or deletions (Fig. 5). A coiled-coil segment was predicted in the C-terminal region of LgtH at positions 230–260, but not in LgtB or LgtE, by using the Lupas CoilsScan method (Lupas *et al.*, 1991; Lupas 1996). The protein polymorphisms and physical properties for LgtB, LgtE and LgtH are summarized in Supplementary Table S1. Polymorphism details for 47 sequences are presented in Supplementary Fig. S3.

DISCUSSION

The glycosyltransfer reaction is considered to be one of the most important biological functions. Hundreds of different glycosyltransferases (EC 2.4.x.y) are involved in the biosynthesis of OS and polysaccharide. Based on sequence similarity, an initial 553 NDP-sugar hexosyltransferases have been classified into 26 families (Campbell *et al.*, 1997). The glycosyltransferases that aid in the production of OS

chains in *N. gonorrhoeae* and *N. meningitidis* LOS belong to four families: family 1 (LgtG), family 2 (LgtA and LgtD), family 8 (LgtC) and family 25 (LgtB and LgtE) (Campbell *et al.*, 1997). A survey of more than 7200 related sequences has updated and extended this classification system of glycosyltransferases to at least 65 families (GT1-65) (Coutinho *et al.*, 2003). Currently, the GT25 family (formerly family 25) includes more than 100 sequences in the database encoding β -1,4-galactosyltransferase, which is responsible for the biosynthesis of LPS. This report adds a new member, *Neisseria* LgtH, to the GT25 family, based on the relationships of sequence homology and biological function.

Mutagenesis of the *lgtH* gene in *N. meningitidis* strain 6275 resulted in an altered LOS structure similar to that of the M986 OP- variant (Quakyi *et al.*, 1997), demonstrating that the α chain of mutant 6275 Δ *lgtH* LOS is truncated. The slightly slower mobility of the 6275 Δ *lgtH* LOS on SDS-PAGE suggests that the overall negative charges of 6275 Δ *lgtH* and M986 OP- LOSs, contributed by phosphate or PEA substitution in the lipid A, or at the heptose of the inner core, are different. Immunoblotting assay with an L3 antibody showed that the LOS of mutant 6275 Δ *lgtH* also altered the antibody-binding specificity. Monosaccharide composition analysis showed that LOS from mutant 6275- Δ *lgtH* does not contain any Gal, but does contain Glc. Further comparisons using MALDI-MS of O-deacylated LOSs confirm the loss of two Gal and one GlcNAc in mutant 6275 Δ *lgtH*. We have, therefore, concluded that the *lgtH* gene encoding a β -1,4-galactosyltransferase is responsible for addition of Gal to the α chain. This is similar to the function of the *lgtE* gene that has been described in a number of reports (Gotschlich, 1994; Erwin *et al.*, 1996; Jennings *et al.*, 1999).

Sequence comparisons of LgtB, LgtE and LgtH reveal that these enzymes are typical two-domain molecules. They have remarkably similar sequences at the N terminus, and exhibit extensive diversity at the C terminus. The conserved region is found to be the N-terminal 174 codon region, from the initiating Met to Pro at position 174. The remainder of the sequence is the C-terminal variable region, which is composed of variable amino acids and deletions. These observations are consistent with the structural studies of glycosyltransferases from different bacterial sources, which show that one domain is responsible for binding donor sugar molecules, and that the other domain is responsible for binding acceptor sugar molecules, or for membrane attachment and protein stabilization (Saxena *et al.*, 1995; Persson *et al.*, 2001; Hu *et al.*, 2003). Four consensus sequence features were found in the N-terminal region of

Fig. 4. Splits decomposition analysis of distinct alleles of three glycosyltransferase genes responsible for biosynthesis of LOS in *Neisseria* species. (a) Twenty alleles of *lgtB*, (b) 11 alleles of *lgtE*, (c) 13 alleles of *lgtH*. The numbers refer to the allele number (Table 2). The splits graph of *lgtE* shows an interconnected network structure, suggesting DNA recombination.

in order to provide a target for inhibitor design. The crystal structures of 18 glycosyltransferases from different GT families have been determined, and these have been summarized in a three-dimensional glycosyltransferase database (Breton *et al.*, 2005). Currently, the crystal structure of β -1,4-galactosyltransferase from the GT25 family is not available. Insights into the sequence conservations and variations of the LgtB, E and H enzymes are likely to increase our understanding of the amino acids important in producing the functional variants, and help in the subdivision of *Neisseria* β -1,4-galactosyltransferases.

There are multiple *lgt* genes in the *Neisseria* genomes responsible for biosynthesis of LOS chains. The distribution, variation and expression of these genes play a critical role in the antigenic diversity of *Neisseria* LOSs. In this study, we analysed the sequence variations and evolution of three paralogous genes at the *lgt-1* locus, i.e. *lgtB*, *lgtE* and *lgtH*, in five *Neisseria* species. The coding regions of *lgtB* and *lgtE* exhibit both length and sequence variations, but that of *lgtH* has an identical length among different alleles. The *lgtH* gene is more conserved than the other two, suggesting that it was probably recently imported into *Neisseria* through horizontal gene transfer from another bacterial species. No sequence variation was observed in the *lgt* genes of the most recent outbreak strain (C:2a:P1.7,1) and an endemic strain (C:2a:P1.5,2). This suggests that most Canadian serogroup C strains have a conserved phenotype and genotype of LOS, though they changed serosubtype due to *porA* mutations. Interestingly, all six Canadian strains of C:2a:P1.7,1 show a uniform LOS pattern in SDS-PAGE analysis (Tsang *et al.*, 2004), except for one strain, 2001-117, which we tested recently (data not shown). Therefore, we sequenced the same region of *lgt-1* in strain 2001-117, and revealed a nonsynonymous mutation in the *lgtH* gene. This mutation resulted in an amino acid substitution from a Thr to a Met, which indicates a possible genetic basis for the altered LOS pattern. Since the mutation occurred at the N-terminal conserved region of LgtH, we considered the possibility that it affects the binding activity of the sugar donor for β -1,4-galactosyltransferase, and therefore changes the LOS structure. It is worthy of further study to see whether substitution of a single amino acid at a specific position can affect the substrate-binding activity of the β -1,4-galactosyltransferase.

Although the *lgtE* and *lgtH* genes occupy the same genomic position at the *lgt-1* locus, they are mutually exclusive in a single genome (Zhu *et al.*, 2002). Most *lgtE* genes have been found in pathogenic *Neisseria* species, the only one exception being that in *N. subflava* (Arking *et al.*, 2001; Zhu *et al.*, 2002). In particular, the *lgtE* gene has been associated with *N. gonorrhoeae*, i.e. all 51 gonococcal strains tested to date possess an *lgtE* gene (Zhu *et al.*, 2002). In this study, splits graphs were constructed using the split decomposition method to analyse and visualize pairwise distance data between sequences by using the program SplitsTree (Huson, 1998). This method does not make the *a priori* assumption that the sequences have a tree-like structure, which might

conflict with phylogenetic signals in the data, such as evidence of recombination. Instead, it can be used to visualize the generation of interconnected network structures. Splits analysis showed substantial differences in the *lgtB*, *lgtE* and *lgtH* genes. The splits graph obtained from the *lgtE* gene presents a network-like structure. This indicates the presence of homoplasies, which are substitutions occurring at an identical position, but caused by separate events (Maynard Smith & Smith, 1998) that are probably evolved from intragenic recombination. However, the splits graphs obtained from the *lgtB* and *lgtH* genes display a star-like structure consisting of a single origin in the centre of the graph, from which a single branch radiates, indicating no recombination signal. Clearly, these *lgt* genes have different evolutionary histories. Mutation and recombination events have created the sequence diversity of the *lgt* genes, and these events are subject to natural selection. Although it is likely that all *lgt* genes have a similar probability of divergence, the selection of new *lgt* variants also depends on the function of the gene, and the interaction of its product with other cellular components. For example, the sequence variations may provide selective advantages and disadvantages in enzyme activities, binding specificities, and the subcellular localization of Lgt proteins. It seems that the functional consequences play an important role in *lgt* evolution by selecting new variants, and shaping the sequence diversities.

In summary, the biological function of the *lgtH* gene was determined as β -1,4-galactosyltransferase. Three *Neisseria* β -1,4-galactosyltransferases, i.e. LgtB, LgtE and LgtH, have been shown to have strong conservation in their N-terminal regions, suggesting functional constraint for sugar donor molecules, but they display remarkable diversity in the C-terminal regions, indicating acceptor specificity for transfer of a sugar moiety, or other biological functions. Though the *lgtB*, *lgtE* and *lgtH* genes are located at a single genomic locus, they have distinct evolutionary histories. The data presented here provide useful information for understanding the distribution, variation and evolution of the GT25 gene family, which is responsible for biosynthesis of LOS in bacterial populations.

ACKNOWLEDGEMENTS

P.Z. was supported by an NIH Fogarty Postdoctoral Fellowship (369VFFD018551). We thank George Kao for technical assistance, Carl E. Frasch and Raymond S. W. Tsang for providing the strains of *N. meningitidis* and commensal *Neisseria*, Michael J. Klutch for DNA sequencing, and Daniel Huson for helpful comments.

REFERENCES

- Amado, M., Almeida, R., Schwientek, T. & Clausen, H. (1999). Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim Biophys Acta* **1473**, 35–53.
- Arking, D., Tong, Y. & Stein, D. C. (2001). Analysis of lipooligosaccharide biosynthesis in the *Neisseriaceae*. *J Bacteriol* **183**, 934–941.

- Ashton, F. E., Ryan, J. A., Borczyk, A., Caugant, D. A., Mancino, L. & Huang, D. (1991). Emergence of a virulent clone of *Neisseria meningitidis* serotype 2a that is associated with meningococcal group C disease in Canada. *J Clin Microbiol* **29**, 2489–2493.
- Banerjee, A., Wang, R., Uljon, S. N., Rice, P. A., Gotschlich, E. C. & Stein, D. C. (1998). Identification of the gene (*lgtG*) encoding the lipooligosaccharide β chain synthesizing glucosyl transferase from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A* **95**, 10872–10877.
- Breton, C., Snajdrova, L., Jeanneau, C., Koea, J. & Imberty, A. (2005). Structures and mechanisms of glycosyltransferases superfamily. *Glycobiology* (in press).
- Burch, C. L., Danaher, R. J. & Stein, D. C. (1997). Antigenic variation in *Neisseria gonorrhoeae*: production of multiple lipooligosaccharides. *J Bacteriol* **179**, 982–986.
- Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D. & Aktories, K. (1998). A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. *J Biol Chem* **273**, 19566–19572.
- Campbell, J. A., Davies, G. J., Bulone, V. & Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J* **326**, 929–939.
- Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328**, 307–317.
- Cox, A. D., Li, J. & Richards, J. C. (2002). Identification and localization of glycine in the inner core lipopolysaccharide of *Neisseria meningitidis*. *Eur J Biochem* **269**, 4169–4175.
- Danaher, R. J., Levin, J. C., Arking, D., Burch, C. L., Sandlin, R. & Stein, D. C. (1995). Genetic basis of *Neisseria gonorrhoeae* lipooligosaccharide antigenic variation. *J Bacteriol* **177**, 7275–7279.
- Devereux, J., Haerberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.
- Elizabeth, A. S. & Seifert, H. S. (2001). The *recX* gene potentiates homologous recombination in *Neisseria gonorrhoeae*. *Mol Microbiol* **40**, 1301–1310.
- Erwin, A. L., Haynes, P. A., Rice, P. A. & Gotschlich, E. C. (1996). Conservation of the lipooligosaccharide synthesis locus *lgt* among strains of *Neisseria gonorrhoeae*: requirement for *lgtE* in synthesis of the 2C7 epitope and of the β chain of strain 15253. *J Exp Med* **184**, 1233–1241.
- Gibson, B. W., Melaugh, W., Phillips, N. J., Apicella, M. A., Campagnari, A. A. & Griffiss, J. M. (1993). Investigation of the structural heterogeneity of lipooligosaccharides of pathogenic *Haemophilus* and *Neisseria* species and of R-type lipopolysaccharides from *Salmonella typhimurium* by electrospray mass spectrometry. *J Bacteriol* **175**, 2702–2712.
- Gibson, B. W., Engstrom, J. J., John, C. M., Hines, W. & Falick, A. M. (1997). Characterization of bacterial lipooligosaccharides by delayed extraction matrix-assisted laser ionization time-of-flight mass spectrometry. *J Am Soc Mass Spectrom* **8**, 645–658.
- Gilbert, M., Watson, D. C., Cunningham, A. M., Jennings, M. P., Young, N. M. & Wakarchuk, W. W. (1996). Cloning of the lipooligosaccharide α -2,3-sialyltransferase from the bacterial pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *J Biol Chem* **271**, 28271–28276.
- Gotschlich, E. C. (1994). Genetic locus for the biosynthesis of the variable portion of *Neisseria gonorrhoeae* lipooligosaccharide. *J Exp Med* **180**, 2181–2190.
- Hardy, M. R., Townsend, R. R. & Lee, Y. C. (1988). Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection. *Anal Biochem* **170**, 54–62.
- Harvey, H. A., Porat, N., Campbell, C. A., Jennings, M., Gibson, B. W., Phillips, N. J., Apicella, M. A. & Blake, M. S. (2000). Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm. *Mol Microbiol* **36**, 1059–1070.
- Heinrichs, D. E., Yethon, J. A., Amor, P. A. & Whitfield, C. (1998). The assembly system for the outer core portion of R1- and R4-type lipopolysaccharides of *Escherichia coli*. The R1 core-specific beta-glucosyltransferase provides a novel attachment site for O-polysaccharides. *J Biol Chem* **273**, 29497–29505.
- Helander, I. M., Lindner, B., Brade, H., Altmann, K., Lindberg, A. A., Rietschel, E. T. & Zähringe, U. (1988). Chemical structure of *Haemophilus influenzae* strain I-69Rd/b+. *Eur J Biochem* **177**, 483–492.
- Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M. & Walker, S. (2003). Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci U S A* **100**, 845–849.
- Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.
- Jennings, H. J., Bhattacharjee, A. K., Kenne, L., Kenny, C. P. & Calver, G. (1980). The R-type lipopolysaccharides of *Neisseria meningitidis*. *Can J Biochem* **52**, 128–136.
- Jennings, M. P., Hood, D. W., Peak, I. R. A., Virji, M. & Moxon, E. R. (1995). Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. *Mol Microbiol* **18**, 729–740.
- Jennings, M. P., Srikhanta, Y. N., Moxon, E. R., Kramer, M., Poolman, J. T., Kuipers, B. & van der Ley, P. (1999). The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in *Neisseria meningitidis*. *Microbiology* **145**, 3013–3021.
- Kahler, C. M. & Stephens, D. S. (1998). Genetic bases for biosynthesis, structure and function of meningococcal lipooligosaccharide. *Crit Rev Microbiol* **24**, 281–334.
- Kahler, C. M., Carlson, R. W., Rahman, M. M., Martin, L. E. & Stephens, D. S. (1996a). Two glycosyltransferase genes, *lgtF* and *rfaK*, constitute the lipooligosaccharide *ice* (inner core extension) biosynthesis operon of *Neisseria meningitidis*. *J Bacteriol* **178**, 6677–6684.
- Kahler, C. M., Carlson, R. W., Rahman, M. M., Martin, L. E. & Stephens, D. S. (1996b). Inner core biosynthesis of lipooligosaccharide (LOS) in *Neisseria meningitidis* serogroup B: identification and role in LOS assembly of the α 1,2 N-acetylglucosamine transferase (RfaK). *J Bacteriol* **178**, 1265–1273.
- Kogan, G., Uhrin, D., Brisson, J. R. & Jennings, H. J. (1997). Structural basis of the *Neisseria meningitidis* immunotypes including the L4 and L7 immunotypes. *Carbohydr Res* **298**, 191–199.
- Kulshin, V. A., Zähringer, U., Lindner, B., Frasch, C. E., Tsai, C. M., Dmitriev, B. A. & Rietschel, E. T. (1992). Structural characterization of the lipid A component of pathogenic *Neisseria meningitidis*. *J Bacteriol* **174**, 1793–1800.
- Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci* **10**, 189–191.
- Lewis, L. A., Gillasp, A. F., McLaughlin, R. E. & 22 other authors (2003). The gonococcal genome sequencing project (<http://www.genome.ou.edu/gono.html>).
- Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem Sci* **21**, 375–382.
- Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164.

- Maynard Smith, J. & Smith, N. H. (1998). Detecting recombination from gene trees. *Mol Biol Evol* 15, 590–599.
- Parkhill, J., Achtman, M., James, K. D. & 25 other authors (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404, 502–506.
- Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. (2001). Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat Struct Biol* 8, 166–175.
- Preston, A., Mandrell, R. E., Gibson, B. W. & Apicella, M. A. (1996). The lipooligosaccharides of pathogenic gram-negative bacteria. *Crit Rev Microbiol* 22, 139–180.
- Quakyi, E. K., Hochstein, H. D. & Tsai, C. M. (1997). Modulation of the biological activities of meningococcal endotoxins by association with outer membrane proteins is not inevitably linked to toxicity. *Infect Immun* 65, 1972–1979.
- Sarkari, J., Pandit, N., Moxon, E. R. & Achtman, M. (1994). Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of promoter containing polycytidine. *Mol Microbiol* 13, 207–217.
- Saxena, I. M., Brown, R. M., Jr, Fevre, M., Geremia, R. A. & Henrissat, B. (1995). Multidomain architecture of beta-glycosyl transferases: implications for mechanism of action. *J Bacteriol* 177, 1419–1424.
- Scholten, R. J., Kuipers, B., Valkenburg, H. A., Dankert, J., Zollinger, W. D. & Poolman, J. T. (1994). Lipo-oligosaccharide immunotyping of *Neisseria meningitidis* by a whole-cell ELISA with monoclonal antibodies. *J Med Microbiol* 41, 236–243.
- Tettelin, H., Saunders, N. J., Heidelberg, J. & 39 other authors (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 278, 1809–1815.
- Tsai, C. M. & Frasch, C. E. (1982). A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Anal Biochem* 119, 115–119.
- Tsai, C. M., Boykins, R. & Frasch, C. E. (1983). Heterogeneity and variation among *Neisseria meningitidis* lipopolysaccharides. *J Bacteriol* 155, 498–504.
- Tsai, C. M., Kao, G. & Zhu, P. (2002). Influence of the length of the lipooligosaccharide alpha chain on its sialylation in *Neisseria meningitidis*. *Infect Immun* 70, 407–411.
- Tsang, R. S., Tsai, C. M., Zhu, P., Ringuette, L., Lorange, M. & Law, D. K. (2004). Phenotypic and genetic characterization of a unique variant of serogroup C ET-15 meningococci (with the antigenic formula C:2a:P1.7,1) causing invasive meningococcal disease in Quebec, Canada. *J Clin Microbiol* 42, 1460–1465.
- van der Ley, P., Kramer, M., Martin, A., Richards, J. C. & Poolman, J. T. (1997). Analysis of the *icsBA* locus required for biosynthesis of the inner core region from *Neisseria meningitidis* lipopolysaccharide. *FEMS Microbiol Lett* 146, 247–253.
- Verheul, A. F. M., Snippe, H. & Poolman, J. T. (1993). Meningococcal lipopolysaccharides: virulence factor and potential vaccine component. *Microbiol Rev* 57, 34–39.
- Wakarchuk, W., Martin, A., Jennings, M. P., Moxon, E. R. & Richards, J. C. (1996). Functional relationships of the genetic locus encoding the glycosyltransferase enzymes involved in expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. *J Biol Chem* 271, 19166–19173.
- Wakarchuk, W. W., Cunningham, A., Watson, D. C. & Young, N. M. (1998). Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen *Neisseria meningitidis*. *Protein Eng* 11, 295–302.
- Westphal, O. & Jann, K. (1965). Bacterial lipopolysaccharides. Extraction with phenol-water and further application of the procedure. *Methods Carbohydr Chem* 5, 83–91.
- Wiggins, C. A. & Munro, S. (1998). Activity of the yeast MNN1 α -1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc Natl Acad Sci U S A* 95, 7945–7950.
- Yamasaki, R., Griffiss, J. M., Quinn, K. P. & Mandrell, R. E. (1993). Neuraminic acid is alpha 2→3 linked in the lipooligosaccharide of *Neisseria meningitidis* serogroup B strain 6275. *J Bacteriol* 175, 4565–4568.
- Yang, Q. L. & Gotschlich, E. C. (1996). Variation of gonococcal lipooligosaccharide structure is due to alterations in poly-G tracts in *lgt* genes encoding glycosyltransferases. *J Exp Med* 183, 323–327.
- Zheng, Y., Roberts, R. J. & Kasif, S. (2004). Segmentally variable genes: a new perspective on adaptation. *PLoS Biol* 2, 452–464.
- Zhu, P., Klutch, M. J. & Tsai, C. M. (2001). Genetic analysis of conservation and variation of lipooligosaccharide expression in two L8-immunotype strains of *Neisseria meningitidis*. *FEMS Microbiol Lett* 203, 173–177.
- Zhu, P., Klutch, M. J., Bash, M. J., Tsang, R. S. W., Ng, L. K. & Tsai, C. M. (2002). Genetic diversity of three *lgt* loci for biosynthesis of lipooligosaccharide (LOS) in *Neisseria* species. *Microbiology* 148, 1833–1844.