Genetic and functional analyses of the \textit{lgtH} gene, a member of the \(\beta\)-1,4-galactosyltransferase gene family in the genus \textit{Neisseria}

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

Lipooligosaccharide (LOS) is a major virulence factor of the pathogenic \textit{Neisseria}. Three galactosyltransferase genes, \textit{lgtB}, \textit{lgtE} and \textit{lgtH}, responsible for the biosynthesis of LOS oligosaccharide chains, were analysed in five \textit{Neisseria} species. The function of \textit{lgtH} in \textit{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 \textit{lgtH} encodes a \(\beta\)-1,4-galactosyltransferase, and that the glucose moiety linked to heptose (Hep) in the \(x\) chain is the acceptor site in the biosynthesis of \textit{Neisseria} LOS. To understand the sequence diversity and relationships of \textit{lgtB}, \textit{lgtE} and \textit{lgtH}, the entire \textit{lgt}-1 locus was further sequenced in three \textit{N. meningitidis} strains and three commensal \textit{Neisseria} strains, and compared with the previously reported \textit{lgt} genes from \textit{Neisseria} species. Comparison of the protein sequences of the three enzymes \textit{LgtB}, \textit{LgtE} and \textit{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 \textit{lgtB}, 12 \textit{lgtE} and 14 \textit{lgtH} sequences revealed a distinct evolutionary history of these genes in \textit{Neisseria}. For example, the splits graph of \textit{lgtE} displayed a network evolution, indicating frequent DNA recombination, whereas splits graphs of \textit{lgtB} and \textit{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.

**INTRODUCTION**

Lipooligosaccharide (LOS) is a major virulence factor of two pathogenic \textit{Neisseria} species, \textit{Neisseria meningitidis} and \textit{Neisseria gonorrhoeae} (Preston et al., 1996; Kahler & Stephens, 1998). In contrast to the LPS in many enteric bacteria, \textit{N. meningitidis} and \textit{N. gonorrhoeae} LOS chains lack repeating O chains, and possess variable oligosaccharide (OS) chains. These OS chains are branched, and their branches are designated \(x\), \(\beta\) and \(\gamma\) chains (Verheul et al., 1993; Kahler & Stephens, 1998). The \textit{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 \textit{Neisseria} LOS (Fig. 1a). The genes responsible for the LOS chains are designated \textit{LgtA}, \textit{B}, \textit{C}, \textit{D}, \textit{E}, \textit{F}, \textit{G}, \textit{H} and \textit{Z} (Gotschlich, 1994; Banerjee et al., 1998;...
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-I 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 strain 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 (2001-83, 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 2001-83 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:1.5.2, 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 (Helander et al., 1994). Splits decomposition analysis was 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′-GACAATCTATCGATTGTATCG-3′) and P102 (5′-GTCACAATCACTAACAAGC-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. 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 BamHI, 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′-GACAATCTATCGATTGTATCG-3′) and P102 (5′-GTCACAATCACTAACAAGC-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. PCRs were performed using primers P23 (5′-ATGGGACCATGTTATCG-3′), P26 (5′-CTATTCACAGGAGACAGC-3′), P36 (5′-GATGCGCTGCAACATGGG-3′), P37 (5′-TAGGGGCCGCTGCTAATG-3′) and P39 (5′-GGGTGTCGCTCQCGTTTGAAG-3′).

Isolation of LOS, SDS-PAGE and Western blotting. LOS samples were isolated from wild-type strains and mutant strains grown in Caftin 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 1 (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 vol 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 NaNH 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 (Mₐ 1297·5), adrenocorticotropic hormone (ACTH) 1–17 (Mₐ 2094·5), ACTH 18–39 (Mₐ 2466·7) and ACTH 7–38 (Mₐ 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 BbsI 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 6275ΔlgtH, was verified by Southern blot hybridization and PCR (Supplementary Fig. S1). The LOS from the isogenic mutant 6275ΔlgtH 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 6275ΔlgtH (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 6275ΔlgtH 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 x chain of L₃,7 LOS, and the M986 OP– variant had no Gal; the latter probably had the x chain terminated at the Glc moiety, similar to 6275ΔlgtH 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 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 x 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 x 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 6275ΔlgtH were analysed using MALDI-MS to determine the number of sugar units lost from the x 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.
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. 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).

<table>
<thead>
<tr>
<th>Strain</th>
<th>O-dLOS*</th>
<th>[M−H] observed (calculated)</th>
<th>Proposed composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>6275lgtH</td>
<td>C</td>
<td>2388-8 (2388-1)</td>
<td>Glc.Hep₂.GlcNAc.PEA₂.Kdo₂.lipid A</td>
</tr>
</tbody>
</table>

*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:1.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, 2001-114 and 2001-117 (GenBank accession no. AY869718). 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 LtgH 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),
whereas an interconnecting network structure was observed in the splits graph of the \textit{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 \textit{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 (Arg$^{14}$-Arg$^{15}$, Phe$^{29}$-Gln$^{30}$-Phe$^{31}$-Phe$^{32}$-Asp$^{33}$, Glu$^{88}$-Asp$^{89}$-Asp$^{90}$). A DXD motif conserved in many GT families (Saxena et al., 1995; Busch et al., 1998; Wiggins & Munro, 1998) was observed (Asp$^{112}$-Xaa$^{113}$-Asp$^{114}$). 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 Coilscan 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 glycosyltransferase 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

---

**Table 2. Allelic variation of the \textit{lgtB}, \textit{lgtE} and \textit{lgtH} genes**

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>\textit{lgtB} CDS*</th>
<th>Allele†</th>
<th>\textit{lgtE} CDS</th>
<th>Allele</th>
<th>\textit{lgtH} CDS</th>
<th>Allele</th>
<th>GenBank accession number (reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{N. meningitidis}</td>
<td>126E</td>
<td>–</td>
<td>–</td>
<td>843</td>
<td>5</td>
<td>–</td>
<td>–</td>
<td>U65788 (Jennings et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>35E</td>
<td>840</td>
<td>4</td>
<td>843</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>AF470655 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>6275</td>
<td>828</td>
<td>8</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>6</td>
<td>AF470656 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>MC58</td>
<td>828</td>
<td>12</td>
<td>831</td>
<td>11</td>
<td>–</td>
<td>–</td>
<td>U25839 (Jennings et al., 1995)</td>
</tr>
<tr>
<td></td>
<td>M986</td>
<td>828</td>
<td>8</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>5</td>
<td>AF470657 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>BB305</td>
<td>828</td>
<td>11</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>4</td>
<td>AU039684 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>89I</td>
<td>828</td>
<td>16</td>
<td>843</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>AF470658 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>M981</td>
<td>828</td>
<td>13</td>
<td>843</td>
<td>8</td>
<td>–</td>
<td>–</td>
<td>AF470659 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>M992</td>
<td>828</td>
<td>15</td>
<td>843</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>AF470660 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>6155</td>
<td>828</td>
<td>10</td>
<td>843</td>
<td>7</td>
<td>–</td>
<td>–</td>
<td>AF470661 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>M978</td>
<td>828</td>
<td>9</td>
<td>843</td>
<td>9</td>
<td>–</td>
<td>–</td>
<td>AF355195 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>A1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>2</td>
<td>AF355194 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>M120</td>
<td>840</td>
<td>7</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>7</td>
<td>AF470662 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>Z2491</td>
<td>840</td>
<td>5</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>1</td>
<td>AL162753 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>7880</td>
<td>840</td>
<td>7</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>8</td>
<td>AF470663 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>7889</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>3</td>
<td>AF470664 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>7897</td>
<td>840</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>3</td>
<td>AF470665 (Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>2001-083</td>
<td>840</td>
<td>20</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>12</td>
<td>AY869717 (this study)</td>
</tr>
<tr>
<td></td>
<td>2001-117</td>
<td>840</td>
<td>20</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>13</td>
<td>AY869718 (this study)</td>
</tr>
<tr>
<td>\textit{N. gonorrhoeae}</td>
<td>F62</td>
<td>840</td>
<td>1</td>
<td>843</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>U14554 (Gotschlich, 1994)</td>
</tr>
<tr>
<td></td>
<td>FA1090</td>
<td>840</td>
<td>3</td>
<td>843</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>AE004969 (Lewis et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>1291</td>
<td>839</td>
<td>2</td>
<td>843</td>
<td>2</td>
<td>–</td>
<td>–</td>
<td>AF112135 (Harvey et al., 2000)</td>
</tr>
<tr>
<td>\textit{N. subflava}</td>
<td>Ns44</td>
<td>828</td>
<td>14</td>
<td>843</td>
<td>10</td>
<td>–</td>
<td>–</td>
<td>AF240672 (Arking et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>85071</td>
<td>840</td>
<td>19</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>11</td>
<td>AY134877 (this study)</td>
</tr>
<tr>
<td>\textit{N. lactamica}</td>
<td>81186</td>
<td>840</td>
<td>17</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>9</td>
<td>AY134876 (this study)</td>
</tr>
<tr>
<td>\textit{N. polysaccharea}</td>
<td>87043</td>
<td>840</td>
<td>18</td>
<td>–</td>
<td>–</td>
<td>807</td>
<td>10</td>
<td>AY134878 (this study)</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>23</td>
<td>12</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Length of the coding region (bp).
†Allele number.
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 6275ΔlgtH LOS is truncated. The slightly slower mobility of the 6275ΔlgtH LOS on SDS-PAGE suggests that the overall negative charges of 6275ΔlgtH 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 6275ΔlgtH also altered the antibody-binding specificity. Monosaccharide composition analysis showed that LOS from mutant 6275ΔlgtH 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 6275ΔlgtH. 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.
LgtB, LgtE and LgtH; three of these features are common motifs in the GT25 family of \( \beta \)-1,4-galactosyltransferases, in which four conserved residues (Asp\(^{33} \), Glu\(^{88} \), Asp\(^{89} \) and Asp\(^{90} \) in LgtB, LgtE and LgtH) have been predicted elsewhere to be potential catalytic residues (Heinrichs et al., 1998). A putative DXD motif surrounded by five hydrophobic residues was also observed in LgtB, LgtE and LgtH. The DXD motif has been found in many families of glycosyltransferases, both in prokaryotes and eukaryotes, and these motifs have been proposed to employ the coordination of a divalent cation in binding the UDP moiety (Saxena et al., 1995; Busch et al., 1998; Wiggins & Munro, 1998).

Interestingly, distinctive sequence differences do exist at the C-terminal region of \textit{Neisseria} \( \beta \)-1,4-galactosyltransferases. First, a coiled-coil segment was found in the C-terminal region of LgtH, but not in that of LgtB or LgtE. \textit{Neisseria} glycosyltransferases for the biosynthesis of LOS are associated with the inner membrane, possibly through interaction of a basic C-terminal domain with membrane phospholipids (Wakarchuk et al., 1996, 1998; Persson et al., 2001). The terminal 50 residues of LgtB and LgtC have been proposed to be involved in enzyme attachments to the surface of the bacterial membrane (Wakarchuk et al., 1998), and the coiled-coil segment of LgtH is also located in this proposed region; however, it is unclear whether the presence of a coiled-coil segment leads to the specific location or membrane association of LgtH when this enzyme is compared with LgtB and LgtE. In addition, LgtH has an acidic Glu at the C terminus, but LgtE has a basic Lys. All the LgtH proteins also possess a 12 aa deletion at the C-terminal region; one of these deletions is at the ompT cleavage site (Wakarchuk et al., 1998) within the region for membrane attachment, and others are upstream of the region (Fig. 5). The biological impact of these sequence differences is not known.

The genes encoding glycosyltransferases have been recently referred to as a family of segmentally variable genes that contain highly variable regions (>70 aa) and well-conserved regions (Zheng et al., 2004). These genes are necessary for micro-organisms to deal with host–pathogen interactions or intracellular responses, and they exhibit great variability in the acceptor-binding domains, and conservation in donor-binding domains (Zheng et al., 2004). For most glycosyltransferases found in pathogenic micro-organisms, detailed functions of acceptor-binding domains remain unknown. The major challenge is to identify acceptor-binding specificity and the essential amino acid motif involved in recognition of glycosyl transfer or membrane association,
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 lgtH 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 (369YF6D018551). 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


Galactosyltransferase in Neisseria


