Structural Analysis of the Lipopolysaccharide from Nontypeable *Haemophilus influenzae* Strain R2846[†]

Susanna L. Lundström,*[,‡] Jianjun Li,[§] Mary E. Deadman,[||] Derek W. Hood,[||] E. Richard Moxon,[||] and Elke K. H. Schweda[‡]

Clinical Research Centre, Karolinska Institutet and University College of South Stockholm, Novum, S-14186 Huddinge, Sweden, Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada, K1A 0R6, and Molecular Infectious Diseases Group, University of Oxford, Department of Paediatrics, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, U.K.

Received December 21, 2007; Revised Manuscript Received February 27, 2008

ABSTRACT: We here report the lipopolysaccharide (LPS) structures expressed by nontypeable *Haemophilus influenzae* R2846, a strain whose complete genome sequence has recently been obtained. Results were obtained by using NMR techniques and ESI-MS on *O*-deacylated LPS and core oligosaccharide material (OS) as well as ESI-MS^n on permethylated dephosphorylated OS. A β-D-Glc*p*-(1→4)-D-α-D-Hep*p*-(1→6)-β-D-Glc*p*-(1→4) unit was found linked to the proximal heptose (HepI) of the conserved triheptosyl inner-core moiety, L-α-D-Hep*p*-(1→2)-[*P*Etn→6]-L-α-D-Hep*p*-(1→3)-L-α-D-Hep*p*-(1→5)-[*PP*Etn→4]-α-Kdo-(2→6)-lipid A. The β-D-Glc*p* (GlcI) linked to HepI was also branched with oligosaccharide extensions from O-4 and O-6. O-4 of GlcI was substituted with sialyllacto-*N*-neotetraose [α-Neu5Ac-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→] and the related structure [(*P*Etn→6)-α-D-Gal*p*NAc-(1→6)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→]. The distal heptose (HepIII) was substituted at O-2 by β-D-Gal. Phosphate, phosphoethanolamine, phosphocholine, acetate, and glycine were found to substitute the core oligosaccharide. Two heptosyltransferase genes, *losB1* and *losB2*, have been identified from the R2846 genome sequence and are candidates to add the noncore heptose to the LPS. Mutant strain R2846*losB1* did not show DD-heptose in the extension from HepI but still contained minor quantities of LD-heptose at the same position, indicating that the *losB1* gene is required to add DD-heptose to GlcI. The LPS from strain R2846*losB1*/*losB2* expressed no noncore heptose, consistent with *losB2* directing the addition of LD-heptose.

INTRODUCTION

Haemophilus influenzae (*Hi*) is a host-adapted bacterium that colonizes the respiratory tract of humans. The bacterium is an important cause of disease worldwide and exists in encapsulated (type a through f) and unencapsulated (nontypeable, NT) forms (*1, 2*). The potential of *Hi* to cause disease depends upon its surface-expressed carbohydrate antigens, capsular polysaccharide (*3*), and lipopolysaccharide (LPS[1]) (*4*). Structural studies of LPS from *Hi* have identified a conserved inner-core structure containing a glucose-substituted triheptosyl inner-core moiety L-α-D-Hep*p*-(1→2)-[*P*Etn→6]-L-α-D-Hep*p*-(1→3)-[β-D-Glc*p*-(1→4)]-L-α-D-Hep*p* linked to lipid A via 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo)-4-phosphate. This inner-core unit provides the template for attachment of oligosaccharides and noncarbohydrate substituents. Prominent noncarbohydrate substituents are free phosphate (*P*), phosphoethanolamine (*P*Etn), pyrophosphoethanolamine (*PP*Etn), phosphocholine (*P*Cho), acetate (Ac), and glycine (Gly) (*5*). The outer-core region of NT*Hi* LPS can mimic host glycolipids, and the expression of terminal epitopes is subject to high frequency phase variation, leading to heterogeneous LPS populations within a single strain. Phase variation is thought to provide an adaptive mechanism, which is advantageous for the survival of bacteria confronted by the differing microenvironments and immune responses of the host (*5*). Structural studies of LPS from various NT*Hi* strains have revealed that the β-D-Glc*p* of the inner-core moiety (GlcI) can have oligosaccharide (OS) extensions at both O-4 and O-6 in the same glycoform. The hexose linked to O-4 can be either a β-D-

[†] This project was supported by the Swedish Research Council (EKHS). D.W.H., M.E.D., and E.R.M. were supported by the Medical Research Council, U.K.

* Corresponding author. S. L. Lundström, Karolinska Institutet, Clinical Research Centre, NOVUM, S-141 86 Huddinge, Sweden. Tel: +46 8585 838 41. Fax: +46 8585 838 20. E-mail: Susanna.Lundstrom@ki.se.

[‡] Karolinska Institutet and University College of South Stockholm.

[§] National Research Council of Canada.

[||] University of Oxford.

[1] Abbreviations: Ac, acetate; *An*Kdo-ol, reduced anhydro Kdo; CE, capillary electrophoresis; DD-Hep, D-*glycero*-D-*manno*-heptose; GaT, [(PEtn→6)-α-D-Gal*p*NAc-(1→6)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→]; Hep, L-*glycero*-D-*manno*-heptose; Hex, hexose; HexNAc, *N*-acetylhexosamine; HPAEC-PAD, high performance anion exchange chromatography with pulsed amperometric detection; Kdo, 3-deoxy-D-*manno*-oct-2-ulosonic acid; lipid A-OH, *O*-deacylated lipid A; LPS, lipopolysaccharide; LPS-OH, *O*-deacylated lipopolysaccharide; MS^n, multiple step tandem mass spectrometry; Neu5Ac, *N*-acetyl neuraminic acid; NT*Hi*, nontypeable *Haemophilus influenzae*; OS, oligosaccharide; *P*Cho, phosphocholine; *P*Etn, phosphoethanolamine; *PP*Etn, pyrophosphoethanolamine; SiaT, sialyllacto-*N*-neotetraose; SRM, selective reaction monitoring; tHep, terminal heptose; tHex, terminal hexose.

10.1021/bi702510b CCC: $40.75
Published on Web 05/08/2008

Glc or a β-D-Gal, which can be further extended (*5*). O-6 of GlcI has been found substituted by heptose containing extensions β-D-Gal*p*-(1→4)-D-α-D-Hep*p*-(1→, β-D-Glc*p*-(1→4)-D-α-D-Hep*p*-(1→ and β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→6)-L-α-D-Hep*p*-(1→, or truncated versions thereof (*6–8*).

Hi strain Rd was the first organism for which the complete genome sequence was determined and published (*9*), and this has facilitated a comprehensive study of LPS biosynthetic genes and their roles in glycoform expression and virulence in several *Hi* strains (*5*). The genes involved in LPS biosynthesis have been investigated extensively in the type b strain Eagan and in the genome reference strain Rd (*10, 11*). Gene functions have been identified that are responsible for most of the steps in the biosynthesis of the OS portion of their LPS molecules (*5*). However, the genes responsible for incorporation of the oligosaccharides from GlcI containing a fourth heptose have not been characterized, likely because they are absent from the strain Rd genome sequence. The recent completion and release of further genome sequences for NT*Hi* strains (*12, 13*) offers further opportunity to identify novel LPS related genes. A candidate gene for a novel heptosyltransferase (LosB) was identified from the NT*Hi* strain R2846 genome sequence through homology to a gene in the related species *Haemophilus ducreyi* (*14*) that has recently been shown to direct the incorporation of a fourth heptose in the LPS of this species (*15*).

In this study, we present detailed structural studies of LPS from NT*Hi* R2846 and derived mutant strains. Strain R2846 expresses a very heterogeneous mixture of glycoforms. A major LPS glycoform contained a β-D-Glc*p*-(1→4)-D-α-D-Hep*p*-(1→6)-β-D-Glc*p*-(1→4) motif extending from HepI and a β-D-Gal*p*-(1→2) unit substituting the distal heptose (HepIII). Several noncarbohydrate substituents contributed to the structural diversity observed in the strain. Two heptosyltransferase candidate genes, *losB1* and *losB2*, have been identified from the R2846 genome sequence. Analysis of LPS from mutant strains R2846*losB1*, R2846*losB2*, and R2846*losB1/losB2* indicated that both *losB1* and *losB2* are involved in the addition of noncore heptose.

MATERIALS AND METHODS

Strains, Cultivation, and Preparation of Lipopolysaccharides. NT*Hi* strain R2846 is a clinical isolate originating from a patient with otitis media (*16*) and has been used in a genome sequencing project (GenBank NZ DQ007026). Strain R2846*lpsA* was made as described previously (*10, 17*), and details of the construction of the mutant strains R2846*losB1*, R2846*losB2*, and R2846*losB1/losB2* will be published elsewhere. Growth conditions of bacteria and extraction of LPS are described in Supporting Information.

Chromatography. Gel-permeation chromatography was performed on a Bio-Gel *P*4 column (2.5 × 80 cm, fraction range 800–4000 Da) as previously described (*18*). GLC was carried out on a Hewlett-Packard 6890 instrument with a DB-5 fused silica capillary column (30 m × 25 mm × 0.25 μm) as previously described (*18*). HPLC was performed on a 2690 Waters HPLC system (Waters, Milford, USA) using a microbore C-18 column (Phenomenex LUNA 5 μm C18(2)) at 21 °C. Gradient conditions are described in Supporting Information. High performance anion exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) was performed on a Dionex Series 4500i chromatography system (Dionex, Sunnyvale, USA) using a CarboPac PA20 column (3 × 150 mm). Details are given in Supporting Information.

Preparation of Oligosaccharides and Lipid A. (a) *O*-Deacylation: *O*-deacylated LPS (LPS-OH) was obtained by treatment of LPS with anhydrous hydrazine under mild conditions as previously described (*18*). *O*-deacylated OS material was obtained using 1 M NH_3 (*18*). (b) Mild acid hydrolyses: LPS-OH (0.1 mg) was treated with aqueous HCl (0.1 M, 80 °C, 1 h) to release Neu5Ac. The sample was evaporated, diluted in H_2O, and subjected to HPAEC-PAD. Reduced core OS was obtained after mild acid hydrolysis (1% acetic-acid, pH 3.1, 100 °C, 2 h) of LPS (25–50 mg) in the presence of borane-*N*-methylmorpholine complex. The insoluble lipid A part was separated from the OS by centrifugation. Following purification on a Biogel *P*4 column, OS fractions were obtained as follows: R2846 (OS1 to OS4), R2846*losB1* (OS1 to OS4), R2846*losB2* (OS1 to OS4), R2846*losB1/losB2* (OS1 to OS3), and R2846*lpsA* (OS1 to OS3). Details are given in Supporting Information. In a minor scale, 2 mg of LPS-OH was hydrolyzed using the acidic conditions described above but with the following reduction step with sodium borohydride in 1 M NH_3 solution (21 °C, 16 h). The oligosaccharide (OS′) obtained was separated from the lipid part by centrifugation. (c) Dephosphorylation of OS and purification of lipid A: The procedures are described in Supporting Information (*18–20*).

Analytical Methods. Sugars were identified as their alditol acetates or (+)-2-butyl-acetylated glycosides as previously described (*21, 22*). Following acetylation, dephosphorylation and permethylation analyses were accomplished on LPS-OH and on OS samples as previously described (*18*). The methylated compounds were recovered on a SepPak C18 cartridge and subjected to sugar analysis or to multiple step tandem HPLC-ESI-MS (HPLC-ESI-MS^n). For locating phosphate substituents, 2 mg OS was dephosphorylated after the permethylation step using 400 μL of 48% HF (4 °C, 48 h). After evaporation, ethylation was performed with iodoethane using the same conditions as those for permethylation followed by sugar analysis. The total content of fatty acids was analyzed as described previously (*20*).

Mass Spectrometry. GLC-MS was carried out with the Hewlett-Packard 6890 chromatograph (described above) connected to a Micromass quadrupole mass spectrometer. The relative proportions of the alditol acetates, partially methylated alditol acetates, partially ethylated and methylated alditol acetates, and fatty acids corresponded to the detector response of the GLC-MS. ESI-MS was performed with a VG Quattro mass spectrometer (Micromass, Manchester, UK) or on a Finnigan LCQ ion trap mass spectrometer in negative ion mode. Details are given in Supporting Information. ESI-MS^n experiments on permethylated OS samples were performed in the positive ion mode $[M + Na]^+$ on the Finnigan LCQ instrument coupled to the Waters HPLC system (described above). Capillary electrophoresis (CE)-ESI-MS/MS was carried out with a Crystal model 310 CE instrument (ATI, Unicam, Boston, MA) coupled to an API 3000 mass spectrometer (Perkin-Elmer/Sciex, Concord, ON) via a MicroIonspray interface as described previously (*23*).

Table 1: Methylation Analysis Data Performed on Dephosphorylated LPS-OH, OS3, OS′ and OS1 Samples Derived from Strain R2846 and the Mutant Strains R2846*losB1*, R2846*losB2*, R2846*losB1/losB2*, and R2846*lpsA*

			relative detector response										
			R2846			R2846*losB2*		R2846*losB1*		R2846*losB1/losB2*		R2846*lpsA*	
methylated sugar[a]	linkage assignment	T_{gm}[b]	LPS-OH	OS3	OS1	LPS-OH	OS1	OS′	OS1	OS′	OS1	OS′	OS1
2,3,4,6-Me_4-Glc	D-Glc*p*-(1-	1.00	13	14	9	16	8	18	8	16	2	17	4
2,3,4,6-Me_4-Gal	D-Gal*p*-(1-	1.07	6	9	7	8	5	11	9	13	13	1	2
2,3,6-Me_3-Gal	-4)-D-Gal*p*-(1-	1.26	2		1	2	1	1	2	2	1	tr.[c]	tr.
2,3,6-Me_3-Glc	-4)-D-Glc*p*-(1-	1.27	1		1	1	1	3	8	5	6	1	tr.
2,4,6-Me_3-Gal	-3)-D-Gal*p*-(1-	1.31	4	2	7	7	4	4	7	7	14	2	8
2,3,4-Me_3-Glc	-6)-D-Glc*p*-(1-	1.32	16	11	9	12	9	4	2		1	20	9
2,3,4,6,7-Me_5-Hep	DD-Hep*p*-(1-	1.38			tr.	tr.							
2,3,4-Me_3-Gal	-6)-D-Gal*p*-(1-	1.43	tr.		5	1	5	2	7	tr.	13	tr.	4
2,3,4,6,7-Me_5-Hep	LD-Hep*p*-(1-	1.46	tr.		1	tr.		1	1		3	8	3
2,3-Me_2-Glc	-4,6)-D-Glc*p*-(1-	1.57	3	1	9	4	19	2	4		5	8	17
2,3,6,7-Me_4-Hep	-4)-DD-Hep*p*-(1-	1.62	17	19	16	18	10					21	14
2,3,6,7-Me_4-Hep	-4)-LD-Hep*p*-(1-	1.67	tr.		tr.			1	2			tr.	tr.
3,4,6,7-Me_4-Hep	-2)-LD-Hep*p*-(1-	1.69	10	22	14	9	14	24	18	18	13	2	4
2,6,7-Me_3-Hep	-3,4)-LD-Hep*p*-(1-	1.82	17	22	14	15	16	23	18	36	22	18	25
2,3,4,6-Me_4-GalN	D-GalNAc*p*-(1-	1.90	2		2	1	2	3	6	1	5	1	3
2,3,6-Me_3-GlcN	-4)-D-GlcNAc*p*-(1-	1.98	4		5	3	6	3	8	2	3	1	7
2,3,4-Me_3-GlcN	-6)-D-GlcNAc*p*-(1-	2.06	5			3							

[a] 2,3,4,6-Me_4-Glc represents 1,5-di-*O*-acetyl-2,3,4,6-tetra-*O*-methyl-D-glucitol-1-d_1, etc. [b] Retention times (T_{gm}) are reported relative to 2,3,4,6-Me_4-Glc. [c] tr., trace.

NMR Spectrometry. NMR spectra obtained from COSY, TOCSY, NOESY, and HMQC experiments of oligosaccharides and deacylated oligosaccharides were recorded in deuterium oxide at 25 °C on a JEOL JNM-ECP500 instrument. Details are given in Supporting Information.

RESULTS

NTHi R2846 Wild-Type and Mutant Strains. NT*Hi* strain R2846 is a clinical isolate taken from the middle ear of a patient with otitis media and has been used in a genome sequencing project. Knowledge of the complete genome sequence provides an excellent opportunity to correlate LPS structure and genetics for this strain and potentially identify new LPS-related genes. The repertoire of LPS-related genes identified for the R2846 genome sequence largely corresponded to that identified from the Rd genome sequence. However, of particular interest were two candidate genes, *losB1* and *losB2*, that might be required for the incorporation of a fourth heptose in the LPS outer core. R2846*losB1*, R2846*losB2*, and R2846*losB1/losB2* mutant strains were constructed. A strain mutated in the *lpsA* gene was also investigated (R2846*lpsA*); this gene controls extension from HepIII (*24*), and the LPS produced is less heterogeneous, enabling the more detailed examination of oligosaccharide extensions from the middle (HepII) and the proximal heptose (HepI) of the inner-core triheptosyl moiety. The wild-type and mutant strains were grown in liquid media, and LPS was isolated by the phenol/chloroform/light petroleum extraction method (*18*).

Characterization of LPS from Wild-Type Strain R2846 and Its Derived Major-Core Oligosaccharide. LPS was treated with anhydrous hydrazine under mild conditions to give water soluble *O*-deacylated material (LPS-OH), which was subjected to compositional and linkage analyses as well as ESI-MS. Compositional sugar analysis revealed D-glucose (Glc), D-galactose (Gal), 2-amino-2-deoxy-D-glucose (GlcN), D-*glycero*-D-*manno*-heptose (DD-Hep), and L-*glycero*-D-*manno*-heptose (Hep) in a ratio of 27:16:19:13:25 as identified by GLC-MS of their corresponding alditol acetate and (+)-2-butyl-glycoside derivatives.

LPS-OH was dephosphorylated with 48% hydrogen fluoride prior to methylation analysis. Significant amounts of terminal Glc, terminal Gal, 6-substituted Glc, 4-substituted DD-Hep, 2-substituted Hep, and 3,4-disubstituted Hep were detected (Table 1). The methylation analysis data was consistent with biantennary structures in NT*Hi* R2486 LPS containing the common inner-core element, L-α-D-Hep*p*-(1→2)-L-α-D-Hep*p*-(1→3)-[β-D-Glc*p*-(1→4)]-L-α-D-Hep*p*-(1→5)-α-Kdo*p* of *Hi* LPS but also DD-Hep-containing side chains.

LPS-OH contained sialic acid (Neu5Ac) as shown by HPAEC-PAD, following mild acid hydrolysis of the sample (*25*). The total amount of Neu5Ac was estimated to be 6 pmol (Neu5Ac)/μg (LPS-OH).

The ESI-MS spectrum of LPS-OH (negative mode) revealed abundant molecular peaks corresponding to doubly and triply deprotonated ions (Supporting Information, Table S1). The MS data indicated the presence of heterogeneous mixtures of glycoforms, consistent with each molecular species containing the conserved *P*Etn substituted triheptosyl inner-core moiety attached via a phosphorylated Kdo linked to the *O*-deacylated lipid A (lipid A-OH) (*26–28*).

Two major glycoforms were observed as doubly charged ions at *m*/*z* 1315.0 and 1376.5, corresponding to glycoforms with respective compositions $Hex_3 \cdot Hep_4 \cdot PEtn_{1-2} \cdot P \cdot Kdo \cdot$ lipidA-OH.

Mild acid hydrolysis of LPS with dilute aqueous acetic acid afforded insoluble lipid A and core oligosaccharide material. Purification by gel filtration chromatography resulted in oligosaccharide fractions OS1, OS2, OS3, and OS4. Sugar analyses performed on the OS fractions were consistent with the analysis performed on LPS-OH, but with a decrease of GlcN, confirming this sugar to be part of lipid A (data not shown).

Lipid A was confirmed to contain the conserved features as described previously for *Hi* (*20, 29*). Both 3-hydroxytetradecanoic- and tetradecanoic acid chains were detected in

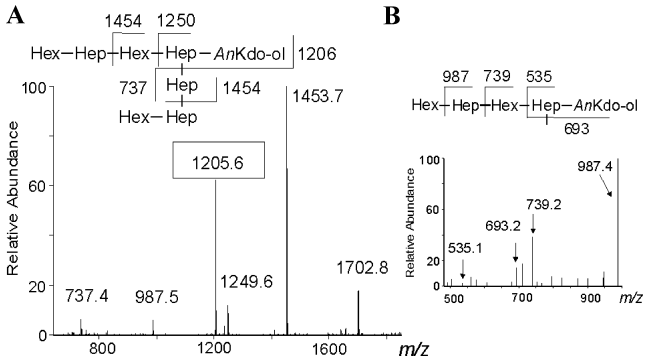


FIGURE 1: HPLC-ESI-MS^n analysis performed on the $Hex_3 \cdot Hep_4 \cdot$ *An*Kdo-ol glycoform of R2846*losB1* (identical to $Hex_3 \cdot Hep_4 \cdot$*An*Kdo-ol of R2846). (A) Product ion spectrum of $[M + Na]^+$ at *m/z* 1920.0 corresponding to the Hex3Hep4 glycoform. Proposed key fragments are indicated in the structure. (B) MS^3 of the ion at *m/z* 1205.6 from MS^2 of *m/z* 1920.0. Proposed key fragments are indicated in the structure.

the fatty acid analysis in agreement with the lipid A structure suggested by Helander et al. (*29*). Lipid A was investigated by ESI-MS^n in the negative mode as described earlier (*20*). Molecular ions at *m/z* 1745.0, 1716.3, 1534.9, and 1307.8 were fragmented in MS$^{2-4}$ experiments giving results identical to those obtained previously for other NT*Hi* strains (Table S2 and Figure S1) (*20*).

OS fractions were analyzed using ESI-MS (Table S3). The indicated high content of *O*-acetates and minor content of ester-linked glycine were confirmed by analyses performed on deacylated OS fractions in which *O*-acylated glycoforms disappeared (Table S4). The glycoform populations observed in OS1 and OS2 were more heterogeneous and of higher molecular mass than those observed in OS3 and OS4. The ESI-MS and 1D ^{1}H NMR spectra of OS3 and OS4 were virtually identical. Both fractions contained one major glycoform that in addition to the triheptosyl inner-core region contained three hexoses and one heptose (Hex3Hep4). In contrast, OS1 and OS2 contained glycoforms with additional hexoses and hexoseamines (HexNAc) and also glycoforms containing *P*, *P*Cho and additional *P*Etn substituents.

In agreement with the LPS-OH data, ESI-MS (negative mode) on *O*-deacylated oligosaccharide of the main fraction (OS3) indicated strong doubly and singly charged ions at *m/z* 799.0 and 1598.5 corresponding to a glycoform with the composition $Hex_3 \cdot Hep_4 \cdot$*P*Etn$\cdot$*An*Kdo-ol. Methylation analysis of OS3, showed a high content of terminal Glc, terminal Gal, 6-substituted Glc, 4-substituted DD-Hep, 2-substituted Hep, and 3,4-disubstituted Hep. In addition, minor amounts of 3-substituted Gal and 4,6-disubstituted Glc were observed (Table 1).

Sequence and branching details of the major Hex3Hep4 glycoform were obtained using HPLC-ESI-MS^n in the positive mode on dephosphorylated and permethylated OS3 (*18, 30*). The glycoform ($Hex_3 \cdot Hep_4 \cdot$*An*Kdo-ol) was observed as a singly charged sodium adduct ion at *m/z* 1919.8. MS^2 and subsequent MS^3 on the resulting product ions determined one isomer (Figure 1A and B). In the following, heptoses originating from the inner-core unit are referred to as HepI (proximal heptose), HepII (middle heptose), and HepIII (distal heptose), and noncore heptose is referred to as HepIV. Ions in the MS^2 spectrum at *m/z* 1453.7, 1249.5, and 1205.6 corresponded to the loss of tHex-Hep, tHex-HepIV-Hex, and tHex-HepIII-HepII, which indicated HepI to be substituted by a tHex-HepIV-Hex- unit and HepIII to be substituted by one hexose. The structure was confirmed in MS^3 experiments on *m/z* 1205.6, where the product ions at *m/z* 987.4 and 739.3 indicated the loss of tHex and tHex-HepIV, respectively.

The major Hex3Hep4 structure was established using detailed ^{1}H, ^{13}C, and ^{31}P NMR analyses. ^{1}H, ^{13}C, and ^{31}P NMR resonances were assigned using gradient chemical shift correlation techniques (COSY, TOCSY, and HMQC experiments). The chemical shift data is given in Table 2 and the TOCSY spectrum is shown in Figure 2A. Prior to NMR analyses, the sample was treated with 1 M NH_3 to remove *O*-acyl groups. In the spectra, each glycosyl residue was identified on the basis of spin connectivity pathways delineated in the ^{1}H chemical shift correlation maps, the chemical shift values, and the vicinal coupling constants. The chemical shift data were consistent with each sugar residue being present in the pyranosyl ring form. Further evidence for this conclusion was obtained from NOE data, which also served to confirm the anomeric configurations of the linkages and the monosaccharide sequence (Table 3) (*31*). The Hep ring systems were identified on the basis of the small $J_{1,2}$ values, and their α-configurations were confirmed by the occurrence of single intra-residue NOE between the respective H-1 and H-2 resonances. Occurrence of intra-residue NOE between the respective H-1, H-3, and H-5 resonances of GlcI, GlcII, and GalI indicated β-configuration of these residues. Several signals for methylene protons of *An*Kdo-ol were observed in the COSY and TOCSY spectra in the region δ 1.87 to 2.18. This is due to the fact that several anhydro forms of Kdo are formed during the hydrolysis by elimination of phosphate or pyrophosphoethanolamine from the C-4 position (*31*). ^{1}H−^{31}P NMR correlation experiments indicated correlations between the phosphomonoester of *P*Etn (δ_P –0.10) and the signals from H-6 of HepII (δ_H 4.57), which confirmed *P*Etn to be linked to O-6 of HepII of the inner-core unit.

In the ^{1}H NMR spectrum of OS3, anomeric resonances corresponding to the triheptosyl moiety (HepI-HepIII) were identified at δ 5.04–5.15, 5.68, and 4.99, respectively. Subspectra corresponding to the hexose residues and the noncore heptose were identified in the 2D COSY and TOCSY spectra at δ 4.50 (GlcI), 4.96 (HepIV), 4.56 (GlcII), and 4.40 (GalI). The downfield H-6 shifts of residue GlcI compared to GlcII (3.85/4.03 and 3.74/3.96, respectively) were consistent with GlcI being substituted in the O-6 position and GlcII being terminal. This was also confirmed in ^{1}H−^{13}C HMQC experiments in which the corresponding C-6 shifts were observed at δ 66.0 and 61.3. The chemical shift data indicated GalI to be terminal (Table 2) (*28*). The downfield-shifted H-4/C-4 (3.94/78.6) of HepIV indicated the noncore DD-Hep to be substituted in position O-4 in agreement with methylation analyses. Inter-residue NOE were observed between the proton pairs of HepIV H-1/GlcI H-6A and GlcII H-1/HepIV H-4, confirming a β-D-Glc*p*-(1→4)-D-α-D-Hep*p*-(1→6)-β-D-Glc*p*- unit extending from HepI (Table 3). Furthermore, inter-residue NOE between the proton pairs of GalI H-1/HepIII H-1/H2 confirmed GalI to be linked to position O-2 in HepIII. The structure of the major glycoform in strain R2846 is shown in Chart 1.

Table 2: ^{1}H and ^{13}C Chemical Shifts for the Major Hex3Hep4 and Hex2Hep3 Glycoforms, Derived from Strain R2846, and the Mutant Strains R2846*losB1*, R2846*losB2*, and R2846*losB1/losB2*[a,b,c]

Residue	Glycose unit	H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6$_{A/B}$/C-6	H-7$_{A/B}$/C-7
Major form; Hex3Hep4 (derived from OS3 of R2846 and R2846losB2)								
HepI	→3,4)-L-α-D-Hep*p*-(1→	5.04-5.15[d]	4.00-4.06[d]	4.01[e]	4.27[e]	-[f]	4.11[e]	-
		97.1	70.9	72.8	73.8		68.2	
HepII	→2,3)-L-α-D-Hep*p*-(1→	5.68	4.18	3.93	-	3.75[g]	4.57	3.72/3.90
	6 ↑ *P*Etn	99.2	79.6	69.3		71.6	75.3	62.4
HepIII	→2)-L-α-D-Hep*p*-(1→	4.99	4.14	4.00	-	-	-	-
		100.1	79.7	70.0				
GlcI	→6)-β-D-Glc*p*-(1→	4.50	3.40	3.44	3.54	3.59	3.85/4.03	
		103.6	74.1	77.2	70.6	74.1	66.0	
HepIV	→4)-D-α-D-Hep*p*-(1→	4.96	4.11	3.93	3.94[e]	4.18	-	-
		99.3	69.9	70.5	78.6	71.0		
GlcII	β-D-Glc*p*-(1→	4.56	3.31	3.55	3.41	3.55	3.74/3.96	
		103.2	73.8	76.0	70.2	76.9	61.3	
GalI	β-D-Gal*p*-(1→	4.40	3.59	3.70	3.92	3.75[e]	-	
		103.8	70.8	72.8	69.2	75.7		
*Hex2Hep3 (derived from OS3 of R2846*losB1 *and R2846*losB1/losB2*)*[h]								
HepIII	→2)-L-α-D-Hep*p*-(1→	5.12	4.22	3.96	-	-	-	-
		100.2	77.8	69.9				
GlcI	β-D-Glc*p*-(1→	4.53	3.37	3.46	3.46	3.46	3.80/4.03	
		103.3	74.2	76.7	70.5	76.7	61.5	
*P*Etn		4.15	3.30					
		62.5	40.5					

[a] Prior to NMR analyses the samples from OS3 were *O*-deacylated. Spectra were recorded in D_2O at 25 °C. Chemical shift values compared between the glycoforms could vary up to ± 0.01 ppm. Pairs of deoxy protons of reduced *An*Kdo-ol were identified in COSY and TOCSY spectra at δ 1.87 to 2.18. [b] OS fractions prior to *O*-acylation indicated Ac ($\delta_{H2/C2}$ 2.20/20.9) and Gly ($\delta_{H2/C2}$ 4.02/40.7). [c] Deacylated OS2 of R2846 indicated *P*Cho ($\delta_{H1/C1}$ 4.36/ 60.1, $\delta_{H2/C-2}$ 3.69/66.5 and $\delta_{methyl(H/C)}$ 3.23/54.5). 4,6-Disubstituted β-Glc (GlcI′) was identified by ($\delta_{H1/C1}$ 4.55/103.4, $\delta_{H4/C4}$ 3.78/78.7 and $\delta_{H6AB/C6}$ 3.87/4.16 and 68.3). Inter-residue NOE between HepIV H-1 and GlcI′ H-6 was detected. 4-Substituted β-GlcNAc was identified by $J_{1,2}$ (~8 Hz) and ($\delta_{H1/C1}$ 4.74/103.3, $\delta_{H2/C2}$ 3.83/55.7 and $\delta_{H4/C4}$3.74/82.5). No inter-residue NOE connections could be determined. *N*Ac ($\delta_{methyl(H/C)}$ 2.05/22.7) from GlcNAc and GalNAc was observed. [d] Several signals were observed for HepI due to heterogeneity in the *An*Kdo moiety. [e] Observed from intense NOE signals. [f] -, not determined. [g] Observed in COSY. [h] The anomeric signals of HepII and GalI occurred at δ 5.65 and 4.42.

Characterization of Glycose Extensions from O-4 Position of GlcI. Sialyllacto-*N*-neotetraose [α-Neu5Ac-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→] and the related structure [(*P*Etn→6)-α-D-Gal*p*NAc-(1→6)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→], here referred to as SiaT and GaT, respectively, have previously been observed linked to HepI in *Hi* strains (*32, 33*). These structural elements were also indicated in the LPS of strain R2846. Because of low abundance, the corresponding glycoforms were not detected in the full ESI-MS spectrum of LPS-OH. The presence of sialylated glycoforms, however, was confirmed in precursor ion monitoring tandem mass spectrometry experiments by scanning for loss of *m*/*z* 290 (Neu5Ac, negative ion mode) following CE-ESI-MS/MS. In the spectrum, quadruply and triply charged ions at *m*/*z* 862.0/1149.5 and 892.5/1190.0 corresponded to sialylated glycoforms with compositions Neu5Ac·HexNAc·Hex_5·Hep_4·*P*Etn_{1-2}·*P*·Kdo·lipidA-OH. Oligosaccharide samples of the leading fractions (OS1 and OS2) were investigated to determine the larger structures expressed. Since ketoses undergo hydrolysis under mild acidic conditions, sialic acid (Neu5Ac) residues were lost during delipidation.

ESI-MS on *O*-deacylated OS1 and OS2 of the wild-type strain (Table S4) were consistent with the data obtained on LPS-OH, revealing doubly deprotonated signals at *m*/*z* 1062.9 and 1102.9 corresponding to glycoforms with the respective compositions HexNAc·Hex_5·Hep_4·*P*Etn·P_{0-1}·*An*Kdo-ol. Furthermore, ions observed at *m*/*z* 1226.0 and 1266.0 indicated glycoforms with the respective compositions $HexNAc_2$·Hex_5·Hep_4·*P*Etn_2·P_{0-1}·*An*Kdo-ol.

Methylation analysis performed on OS1 (Table 1) showed high content of terminal Glc, terminal Gal, 3-substituted Gal, 6-substituted Glc, 6-substituted Gal, 4,6-disubstituted Glc,

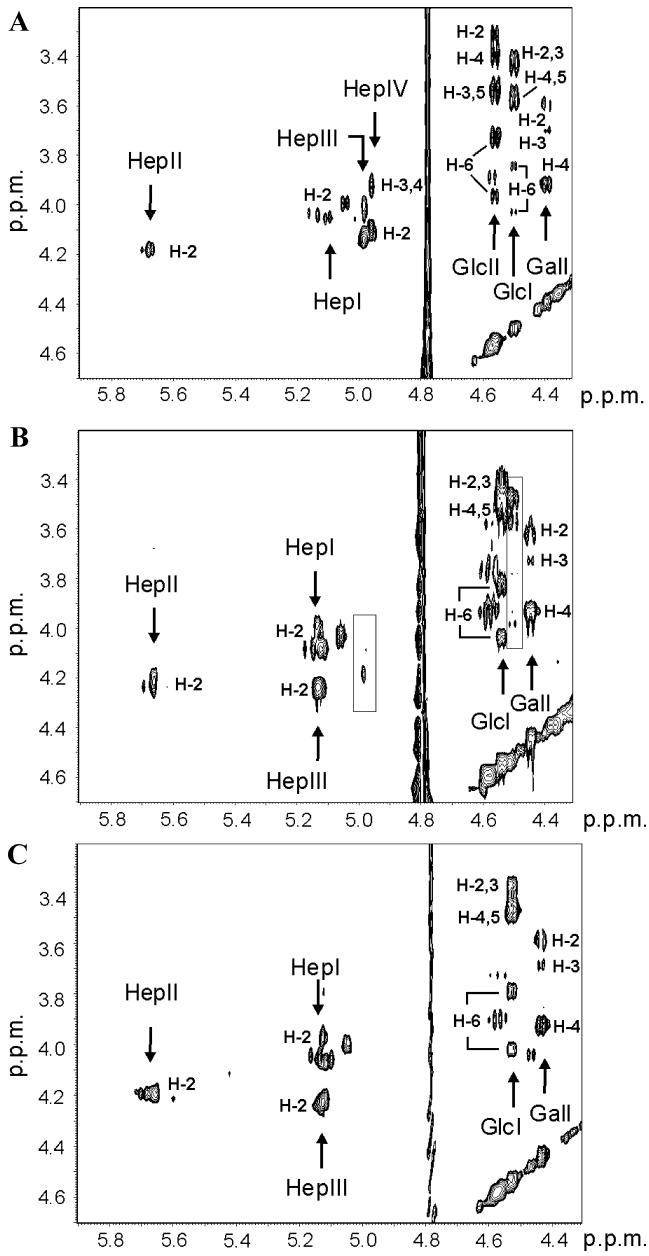


FIGURE 2: Selected region of phase sensitive TOCSY spectra (mixing time 180 ms) of the OS3 fractions of R2846 (A), R2846*losB1* (B), and R2846*losB1/losB2* (C). Cross-peaks of the major glycoform in the respective samples are labeled. Signals believed to originate from the Hex3Hep4 glycoform in R2846*losB1* (B) are marked out within the squares.

Table 3: NOE Data of OS3 Obtained from R2846 and R2846*losB2*[a]

anomeric proton of residue	observed proton: intra residue NOE	observed proton: inter residue NOE
HepI	H-2	H-5/H-7 Kdo
HepII	H-2	H-3 HepI, H-1 HepIII
HepIII	H-2	H-1 and H-2 HepII, H-1 GalI
GlcI	H-3 and H-5	H-4 and H-6 HepI
HepIV	H-2	H-6A GlcI
GlcII	H-3 and H-5	H-4 HepIV
GalI	H-3 and H-5	H-1 and H-2 HepIII

[a] Cross-peaks detected in the NOESY spectra were acquired with 250 ms spin lock time.

4-substituted DD-Hep, 2-substituted Hep, 3,4-disubstituted Hep, terminal GalN, and 4-substituted GlcN. In addition, minor amounts of 4-substituted Gal, 4-substituted Glc, terminal DD-Hep, terminal Hep, and 4-substituted Hep were detected.

Several glycoforms were observed by HPLC-ESI-MS^n on dephosphorylated and permethylated OS1 that were not detected in underivatized samples (Table 4). This is because of increased MS response obtained by permethylation in combination with added sodium acetate. In addition to a major ion at *m/z* 1919.8 corresponding to $Hex_3 \cdot Hep_4 \cdot$*An*Kdo-ol, the full scan mass spectrum showed sodiated singly positively charged adduct ions at *m/z* 2121.4, 2366.4, 2573.5, and 2817.8, corresponding to the glycoforms $HexNAc_{1-2} \cdot Hex_4 \cdot Hep_3 \cdot$*An*Kdo-ol and $HexNAc_{1-2} \cdot Hex_5 \cdot Hep_4 \cdot$*An*Kdo-ol, respectively. Also, ions at *m/z* 1672.4 ($Hex_3 \cdot Hep_3 \cdot$*An*Kdo-ol) as well as *m/z* 1715.8 and 2124.1 ($Hex_{2,4} \cdot Hep_4 \cdot$*An*Kdo-ol) were detected. Fragmenting the molecular ion at *m/z* 2573.5 (Figure 3A) revealed ions at *m/z* 2353.9, 2110.1, 1905.9, 2106.9, and 1859.3, resulting from the loss of tHex, tHex-HexNAc, tHex-HexNAc-Hex, tHex-Hep, and tHex-HepIII-HepII, respectively. MS^3 experiments on *m/z* 1859.3 (Figure 3B) confirmed that the first hexose substituting HepI is branched with both a t-Hex-HepIV- and a tHex-HexNAc-Hex- unit. The fragmentation pattern indicated ions at *m/z* 1642.1, 1395.8, 1192.2, and 1392.9 originating from the loss of tHex, tHex-HexNAc, tHex-HexNAc-Hex, and tHex-HepIV, respectively. In addition to the MS^3 experiment described above, further MS^{3-4} experiments were performed to confirm that the OS extension from HepI was not arranged in a linear fashion. As for the major form, HepIII was confirmed to be substituted by one hexose. The molecular ion at *m/z* 2817.8 corresponded to a glycoform with the composition $HexNAc_2 \cdot Hex_5 \cdot Hep_4 \cdot$*An*Kdo-ol. When this ion (Figure 3C) was further fragmented, the spectrum showed signals at *m/z* 2558.7, 2355.2, 2110.1, 1906.4, 2352.1, and 2103.9 resulting from the loss of tHexNAc, tHexNAc-Hex, tHexNAc-Hex-HexNAc, tHexNAc-Hex-HexNAc-Hex, tHex-Hep, and tHex-HepIII-HepII, respectively. MS^3 experiments confirmed the structure of this glycoform to be identical to the glycoform determined for the ion at *m/z* 2573.5 with the difference of containing a tHexNAc-Hex-HexNAc-Hex- instead of a tHex-HexNAc-Hex- element at the branched hexose linked to HepI. Figure 3D shows the MS^3 experiment on the fragment ion at *m/z* 2352.1 corresponding to the loss of tHex-Hep. The resulting spectrum indicated key fragments at *m/z* 1643.6, 1441.0, and 1638.0 corresponding to the loss of tHexNAc-Hex-HexNAc, tHexNAc-Hex-HexNAc-Hex, and tHex-HepIII-HepII, respectively. MS^2 and MS^3 experiments on the ions corresponding to glycoforms $HexNAc_{1-2} \cdot Hex_4 \cdot Hep_3 \cdot$*An*Kdo-ol (*m/z* 2121.4 and 2366.4) confirmed that these did not contain the tHex-HepIV- entity but still contained the same tHex-HexNAc-Hex- and tHexNAc-Hex-HexNAc-Hex- units connected to -Hex-HepI as described above.

In addition to the major $HexNAc \cdot Hex_4 \cdot Hep_3 \cdot$*An*Kdo-ol glycoform (*m/z* 2121.4) containing the tHex-HexNAc-Hex-Hex- unit from HepI, a different glycoform having a tHexNAc-Hex-Hex-Hex- unit from HepI was indicated in the MS^{2-3} spectra. Elongation with two hexoses from HepI was also detected in MS^{2-3} experiments on ions corresponding to glycoforms with compositions $Hex_3 \cdot Hep_3 \cdot$*An*Kdo-ol and $Hex_4 \cdot Hep_4 \cdot$*An*Kdo-ol (*m/z* 1672.4 and 2124.1). Because of heterogeneity and small abundance, the OS structures

Chart 1

β-D-Glc**II***p*-(1→4)-D-α-D-Hep**IV***p*-(1→6)-β-D-Glc**I***p*-(1→4)-L-α-D-Hep**I***p*-(1→5)-*An*Kdo-ol
3
↑
1
L-α-D-Hep**II***p*6←*P*Etn
2
↑
1
β-D-Gal**I***p*-(1→2)-L-α-D-Hep**III***p*

Table 4: Glycoform Structures Obtained from Strain R2846 and the Mutant Strains R2846*losB2*, R2846*losB1*, R2846*losB1/losB2*, and R2846*lpsA* Elucidated by HPLC-ESI-MS^n (Positive Mode) on Permethylated and Dephosphorylated OS1 Fractions[a]

Hex_q
|
Hep_p
|
$HexNAc_k$-Hex_l-$HexNAc_m$-Hex_n-Hex_o-Hep-*An*Kdo-ol
|
Hep
|
Hex_r-Hep

	relative glycoform abundance[b] (%)					structure								relative isomer abundance[c]				
glycoform	R2846 (A)	R2846*losB2* (B)	R2846*losB1* (C)	R2846*losB1/losB2* (D)	R2846*lpsA* (E)	k	l	m	n	o	p	q	r	(A)	(B)	(C)	(D)	(E)
Hex			tr.[d]	tr.		0	0	0	0	1	0	0	0			H	H	
Hex_2	tr.	tr.	44	73		0	0	0	0	1	0	0	1	H	H	H	H	
Hex_3	1		3	4		0	0	0	1	1	0	0	1	H		H	H	
						0	0	0	0	1	0	0	2			T[e]	T[e]	
Hex_4			tr.	1		0	0	0	2	1	0	0	1			H	H	
Hex·Hep	tr.	tr.			tr.	0	0	0	0	1	1	0	0	H	H			H
Hex_2·Hep	2	tr.	1		58	0	0	0	0	1	1	1	0	M	H	H		H
						0	0	0	0	1	1	0	1	M	L			[f]
Hex_3·Hep	72	71	4		4	0	0	0	0	1	1	1	1	H	H	H		-
						0	0	0	1	1	1	1	0	-	-	L		H
Hex_4·Hep	3	3				0	0	0	1	1	1	1	1	H	H			
HexNAc·Hex_2				tr.		0	0	1	1	1	0	0	0				H	
HexNAc·Hex_3			3	2		0	0	1	2	1	0	0	0			L	T	
						0	0	1	1	1	0	0	1			M	M	
						0	1	1	1	1	0	0	0			L	M	
HexNAc·Hex_4	1	2	17	11		0	1	1	1	1	0	0	1	H	H	H	H	
						0	0	1	2	1	0	0	1	T	T	T	T	
$HexNAc_2$·Hex_4	3	3	28	9		1	1	1	1	1	0	0	1	H	H	H	H	
HexNAc·Hex_3·Hep					2	0	0	1	1	1	1	1	0					H
HexNAc·Hex_4·Hep	tr.	2			14	0	1	1	1	1	1	1	0	nr.[g]	-			H
						0	1	1	1	1	1	0	1		M			-
						0	0	1	1	1	1	1	1		L			-
HexNAc·Hex_5·Hep	9	7	tr.			0	1	1	1	1	1	1	1	H	H	nr.		
						0	0	1	2	1	1	1	1	T	T			
$HexNAc_2$·Hex_4·Hep	tr.				22	1	1	1	1	1	1	1	0					H
$HexNAc_2$·Hex_5·Hep	9	12	tr.			1	1	1	1	1	1	1	1	H	H	H		

[a] Subscripts denoted by the letters k, l, m, n, o, p, q, and r refer to the number of glycose residues in the structure. All glycoforms contain Hep$_3$·*An*Kdo-ol. [b] Relative abundance expressed in percentage for each glycoform calculated from the detected molecular ions in the MS^1 spectrum. [c] Relative abundance for structural isomers of each glycoform calculated from the intensity of the product ions in the MS^2 spectrum and indicated as follows: H (high: over 80%), M (medium: 30–80%), L (low: 2–30%), and T (trace: < 2%). [d] tr., trace. [e] Indicated by ions in the MS^2 spectrum at *m*/*z* 1002 and 753 due to the loss of tHex-Hex-HepIII and tHex-Hex-HepIII-HepII from the molecular ion (*m*/*z* 1672). [f] -, not observed. [g] nr., not rationalized.

extending from the O-4 position of GlcI could not unambiguously be established by NMR. However, the combined data from precursor ion scan CE-ESI-MS and HPLC-ESI-MS^n analysis in addition to the significant amounts of t-GalNAc, 6-Gal, 4-GlcNAc, 3-Gal, and 4,6-Glc detected in methylation analysis makes it reasonable to suggest that strain R2846 expresses both extending SiaT- and GaT-epitopes from HepI (Chart 2).

Characterization of LPS from Strain R2846losB1. Compositional sugar analysis of LPS-OH from strain R2846*losB1* indicated Glc, Gal, GlcN, and Hep in the ratio of 29:24:18:30. In contrast to the wild-type strain, no DD-Hep was detected. Compositional analysis was also performed on dephosphorylated OS giving the same result (data not shown). The ESI-MS analysis of LPS-OH and the deacylated OS fractions (Tables S1 and S4) showed one major glycoform, indicative in LPS-OH from the doubly negatively charged ions at *m*/*z* 1137.9 and 1198.8 and in the OS samples from the singly negatively charged ion at *m*/*z* 1244.5, corresponding to the glycoforms Hex_2·Hep_3·*P*Etn_{1-2}·*P*·Kdo·lipidA-OH and Hex_2·Hep_3·*P*Etn·*An*Kdo-ol, respectively. Interestingly, a number of minor glycoforms containing four heptoses were observed in LPS-OH as doubly charged ions at *m*/*z* 1233.9 and 1376.5 (Hex_2·Hep_4·*P*Etn·*P*·Kdo·lipidA-OH and Hex_3·

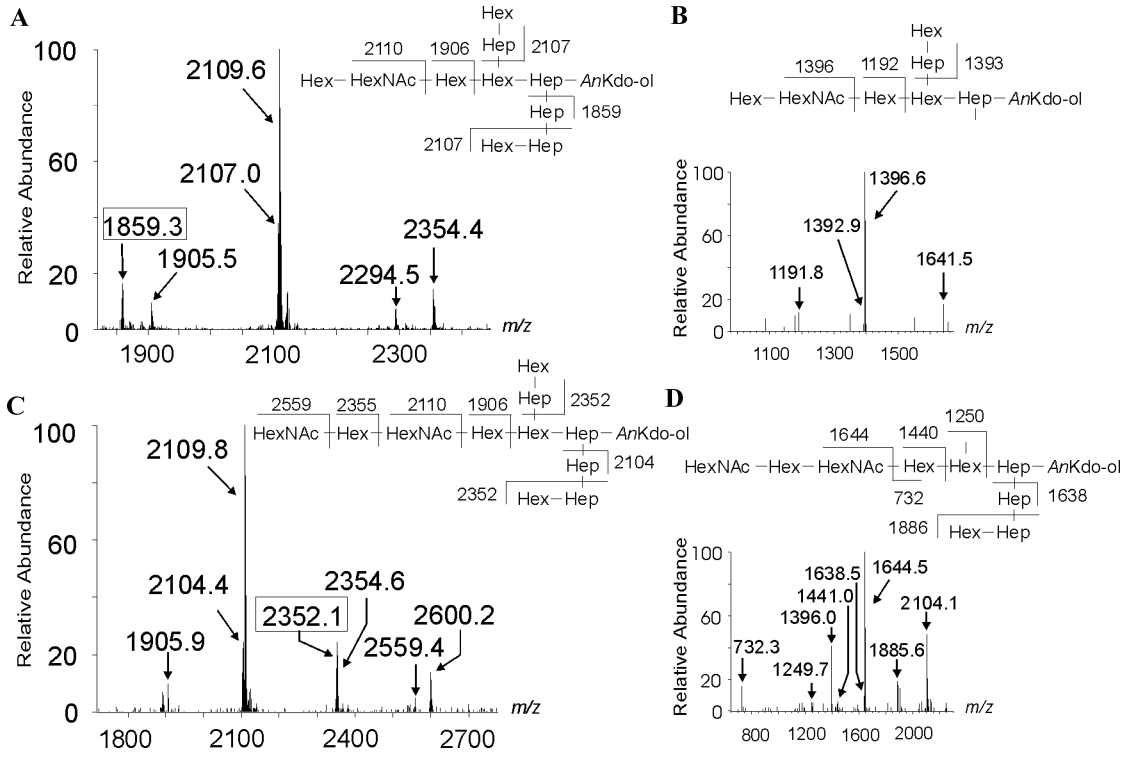
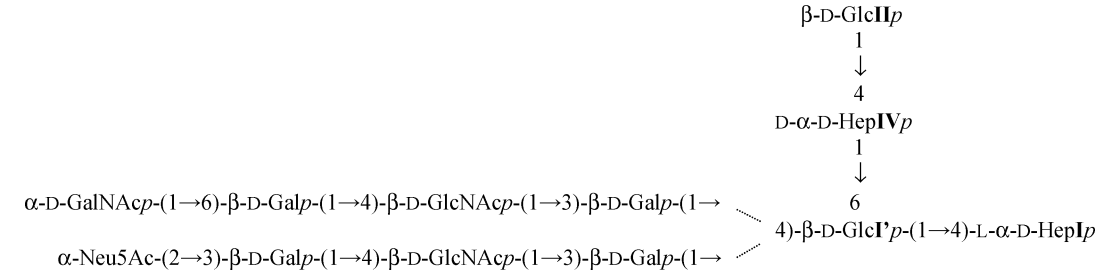


FIGURE 3: HPLC-ESI-MS^n analysis performed on the HexNAc$_{1-2}$·Hex$_5$·Hep$_4$·*An*Kdo-ol glycoforms of R2846*losB2* (identical to HexNAc$_{1-2}$·Hex$_5$·Hep$_4$·*An*Kdo-ol of R2846). (A) Product ion spectrum of $[M + Na]^+$ *m/z* 2573.1 corresponding to the HexNAcHex5Hep4 glycoform. Proposed key fragments are indicated in the structure. (B) MS^3 of the ion at *m/z* 1859.3 from MS^2 of *m/z* 2573.1. Proposed key fragments are indicated in the structure. (C) Product ion spectrum of $[M + Na]^+$ *m/z* 2818.1 corresponding to the HexNAc2Hex5Hep4 glycoform. Proposed key fragments are indicated in the structure. (D) MS^3 of the ion at *m/z* 2352.1 from MS^2 of *m/z* 2818.1. Proposed key fragments are indicated in the structure.

Chart 2



Hep$_4$·*P*Etn$_2$·*P*·Kdo·lipidA-OH) and in OS fractions as singly charged ions at *m/z* 1437.4 and 1598.5 (Hex$_{2-3}$·Hep$_4$·*P*Etn·*An*Kdo-ol). Since the R2846*losB1* mutant did not contain DD-Hep, the fourth heptose was concluded to be a noncore LD-Hep, which in the following will be referred to as HepIV$_{LD}$.

Methylation analysis performed on OS material that had not been fractionated by gel-permeation chromatography (OS′) showed major amounts of terminal Glc, terminal Gal, 2-substituted Hep, and 3,4-disubstituted Hep. In addition, 4-substituted Gal, 4-substituted Glc, 3-substituted Gal, 6-substituted Glc, 6-substituted Gal, terminal Hep, 4,6-disubstituted Glc, 4-substituted Hep, terminal GalN, and 4-substituted GlcN were detected (Table 1). The 4-substituted LD-Hep was concluded to originate from HepIV$_{LD}$. Methylation analysis performed on OS1 from R2846*losB1* was consistent with the analysis performed on OS′ but with an increase of 4-substituted Glc, 3-substituted Gal, 6-substituted Gal, 4,6-disubstituted Glc, terminal GalNAc, and 4-substituted GlcNAc.

Sequence and branching details of the major glycoform in OS1 were obtained using HPLC-ESI-MS^n in the positive mode on dephosphorylated and permethylated material (Table 4). In the full scan spectrum, ions were detected at *m/z* 1467.9, 1671.9, 1715.6, 1919.6, 1917.1, 2121.0, and 2366.2, corresponding to Hex$_{2-3}$·Hep$_3$·*An*Kdo-ol, Hex$_{2-3}$·Hep$_4$·*An*Kdo-ol, HexNAc·Hex$_{3-4}$·Hep$_3$·*An*Kdo-ol, and HexNAc$_2$·Hex$_4$·Hep$_3$·*An*Kdo-ol, respectively (Figure 4B). In addition, traces of two large glycoforms at *m/z* 2573 and 2818 corresponding to HexNAc$_{1-2}$·Hex$_5$·Hep$_4$·*An*Kdo-ol were observed using selective reaction monitoring (SRM). Only one isomer was found for the major Hex2Hep3 glycoform (*m/z* 1467.9). It was defined in MS^2 by the ions at *m/z* 1001.4 and 753.3, corresponding to the loss of tHex-HepIII and tHex-HepIII-HepII, respectively. HepI and HepIII were each substituted by one hexose, which was confirmed in MS^3 experiments (data not shown). The minor ions at *m/z* 1715.6 and 1919.6 as well as the detected glycoforms at *m/z* 2573 and 2818 with the compositions Hex$_{2-3}$·Hep$_4$·*An*Kdo-ol and HexNAc$_{1-2}$·Hex$_5$·Hep$_4$·*An*Kdo-ol corresponded to glycoforms containing a noncore heptose. MS^2 and MS^3 experiments confirmed the noncore heptose to be situated

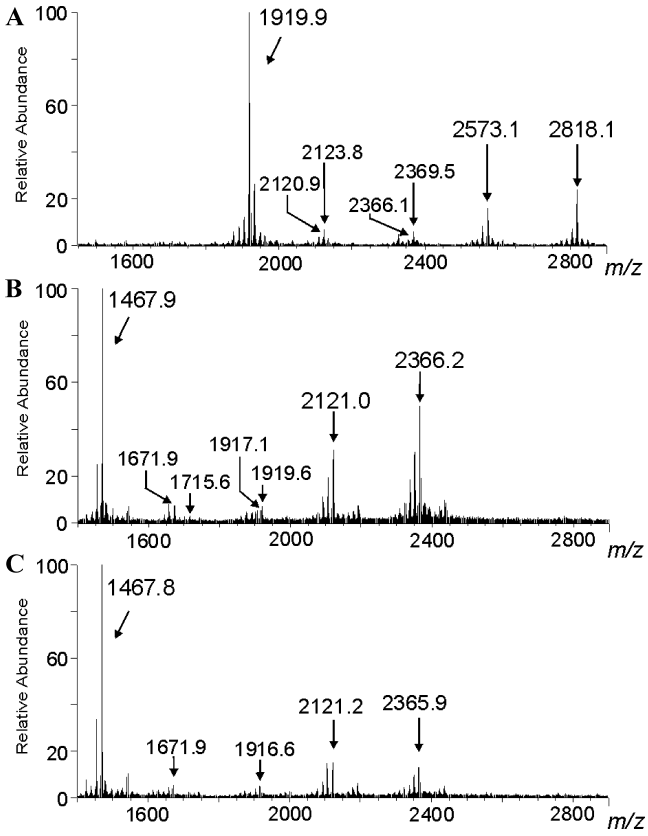


FIGURE 4: HPLC-ESI-MS full scan spectra (positive mode) on dephosphorylated and permethylated OS1 fractions of R2846*losB2* (A), R2846*losB1* (B), and R2846*losB1/losB2* (C). The singly charged sodiated ions observed are correspond to *m/z* 1468, $Hex_2 \cdot Hep_3 \cdot$*An*Kdo-ol; 1672, $Hex_3 \cdot Hep_3 \cdot$*An*Kdo-ol; 1716, $Hex_2 \cdot Hep_4 \cdot$*An*Kdo-ol; 1917, $HexNAc_1 \cdot Hex_3 \cdot Hep_3 \cdot$*An*Kdo-ol; 1920, $Hex_3 \cdot Hep_4 \cdot$*An*Kdo-ol; 2121, $HexNAc_1 \cdot Hex_4 \cdot Hep_3 \cdot$*An*Kdo-ol; 2124, $Hex_4 \cdot Hep_4 \cdot$*An*Kdo-ol; 2366, $HexNAc_2 \cdot Hex_4 \cdot Hep_3 \cdot$*An*Kdo-ol ; 2369, $HexNAc_1 \cdot Hex_4 \cdot Hep_4 \cdot$*An*Kdo-ol; 2573, $HexNAc_1 \cdot Hex_5 \cdot Hep_4 \cdot$*An*Kdo-ol; and 2818, $HexNAc_2 \cdot Hex_5 \cdot Hep_4 \cdot$*An*Kdo-ol.

at the same position as in the wild-type strain. Figure 1 shows ESI-$MS^{2,3}$ analysis of the ion at *m/z* 1919.6 ($Hex_3 \cdot Hep_4 \cdot$*An*Kdo-ol).

The predominant Hex2Hep3 glycoform was elucidated by NMR on deacylated OS3 (Table 2 and Figure 2B). In the TOCSY spectrum, resonances corresponding to GlcI and GalI were observed at $\delta_{H\text{-}1}$ 4.53 and 4.42, respectively. On the basis of the ^1H and ^{13}C chemical shift data, both residues were concluded to be terminal. Chemical shifts originating from the heptoses of the inner-core unit were similar to those assigned for the wild-type strain except for HepIII ($\delta_{H\text{-}1}$ 5.12; $\delta_{H\text{-}2}$ 4.22). Inter-residue NOE was observed between the proton pairs of GlcI H-1/HepI H-4/H-6 and GalI H-1/HepIII H-1/H2, which confirmed GlcI to substitute HepI at O-4 and GalI linked to O-2 of HepIII.

In addition to the major Hex2Hep3 glycoform, a cross-peak in the TOCSY spectrum between δ_{H-1} 4.96 and $\delta_{H\text{-}2}$ 4.09 was observed. It is reasonable to assume that this signal would originate from either $HepIV_{LD}$ or HepIII from the Hex3Hep4 glycoform. However, this could not be confirmed by NOESY experiments.

Characterization of LPS from Strain R2846losB2. Analyses of LPS from R2846*losB2* gave virtually the same results as the wild-type strain (Tables 1–4 and Supporting Information Tables S1, S3, and S4). In contrast to R2846, however, no traces of 4-substituted Hep could be detected in methylation analysis on OS1 derived from this strain, indicating that it did not express noncore LD-Hep. Figure 4A shows the full scan ESI-MS spectrum (positive mode) of dephosphorylated and permethylated OS1 derived from R2846*losB2*. Singly charged ions were observed at *m/z* 1919.9, 2123.8, 2120.9, 2366.1, 2369.5, 2573.1, and 2818.1, corresponding to $Hex_{3-4} \cdot Hep_4 \cdot$*An*Kdo-ol; $HexNAc_{1-2} \cdot Hex_4 \cdot Hep_3 \cdot$*An*Kdo-ol; $HexNAc \cdot Hex_{4-5} \cdot Hep_4 \cdot$*An*Kdo-ol; and $HexNAc_2 \cdot Hex_5 \cdot Hep_4 \cdot$*An*Kdo-ol, respectively. Figure 3 shows the ESI-$MS^{2,3}$ spectra on ions corresponding to the $HexNAc_{1-2} \cdot Hex_5 \cdot Hep_4 \cdot$*An*Kdo-ol glycoforms.

Characterization of LPS from Strain R2846losB1/losB2. Sugar analyses performed on LPS-OH from R2846*losB1/losB2* revealed the presence of Glc, Gal, GlcN, and Hep in the ratio of 26:25:9:41. The ESI-MS analysis (Tables S1, S3, and S4) of LPS-OH and the OS fractions showed the same major Hex2Hep3 glycoform as that present in R2846*losB1*. However, no glycoforms containing four heptoses were detected. This was also confirmed in HPLC-ESI-MS^n analysis (Table 4 and Figure 4C) on dephosphorylated and permethylated OS material. To unambiguously establish that no traces of glycoforms containing a noncore heptose were present, SRM was performed on ions calculated to contain four heptoses, (*m/z* 1512, 1716, 1920, 2124, 2165, 2369, 2573, 2614, and 2818, corresponding to $Hex_{1-4} \cdot Hep_4 \cdot$*An*Kdo-ol, $HexNAc_1 \cdot Hex_{3-5} \cdot Hep_4 \cdot$*An*Kdo-ol and $HexNAc_2 \cdot Hex_{4-5} \cdot Hep_4 \cdot$*An*Kdo-ol, respectively). Only noise was detected in the mass spectra obtained from those experiments.

Methylation analysis (Table 1) performed on OS′ material showed terminal Glc, terminal Gal, 4-substituted Gal, 4-substituted Glc, 3-substituted Gal, 2-substituted Hep, 3,4-disubstituted-Hep, terminal GalN, and 4-substituted GlcN in the relative amounts of 16:13:2:5:7:18:36:1:2. Traces of 6-substituted Gal were also detected.

NMR analysis on deacylated OS3 (Table 2; Figure 2C) revealed the same data as that previously described for the Hex2Hep3 glycoform in the R2846*losB1* mutant.

Characterization of LPS from Strain R2846lpsA. In agreement with previously investigated *lpsA* mutant strains, R2846*lpsA* did not express any elongation from HepIII (Tables 1 and 4, and Supporting Information Table S1). Compared to the wild-type strain, a decrease of Gal in sugar- and of t-Gal in methylation analysis was observed, as well as a significant increase of t-Hep. ESI-MS of LPS-OH revealed major doubly negatively charged ions at *m/z* 1233.9 and 1294.8, corresponding to the glycoforms $Hex_2 \cdot Hep_4 \cdot$*P*$Etn_{1-2} \cdot$*P*$\cdot$Kdo$\cdot$lipidA-OH. In sequence analysis on dephosphorylated and permethylated OS, no glycoforms were observed with substitution from HepIII. $MS^{2,3}$ analysis on the ion at *m/z* 1716.1 corresponding to the major glycoform at $Hex_2 \cdot Hep_4 \cdot$*An*Kdo-ol indicated HepIII to be terminal and HepI to be elongated by the tHex-HepIV-Hex- unit. Notable is an ion at *m/z* 2164.9 detected in the full scan spectrum corresponding to $HexNAc_1 \cdot Hex_3 \cdot Hep_4 \cdot$*An*Kdo-ol. MS^2 on this ion indicated a fragment ion at *m/z* 1654.9 corresponding to the loss of tHepIII-HepII. MS^3 on this ion resulted in the loss of *m/z* 1395.8 which corresponded to the loss of tHexNAc. Subsequent MS^4 on *m/z* 1395.8 gave fragments at *m/z* 1177.8, 929.6 and 1191.7 corresponding to the loss of tHex, tHex-HepIV-, and of two substituted hexose, respectively. These results indicated an isomer with no

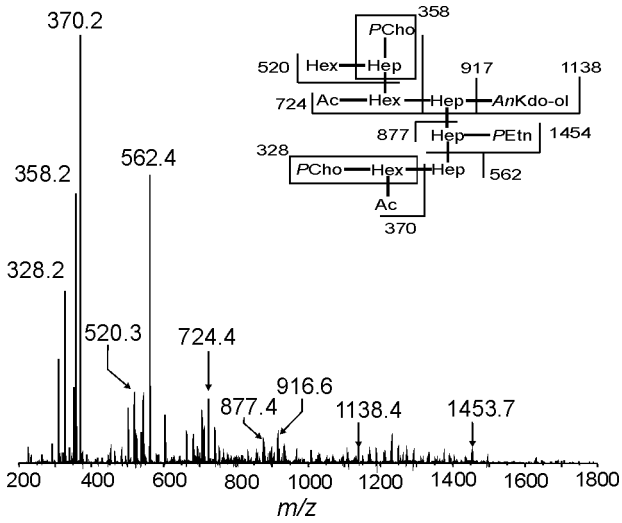


FIGURE 5: CE-ESI-MS/MS spectrum (positive mode) of the doubly charged ion at *m/z* 1008.4 corresponding to the composition $Ac_2 \cdot PCho_2 \cdot Hex_3 \cdot Hep_4 \cdot PEtn \cdot An$Kdo-ol. Proposed key fragments are indicated in the structure.

substitution from HepII and HepIII but with HepI substituted by a branched hexose elongated by a tHexNAc-Hex- subunit and a tHex-HepIV- subunit.

Noncarbohydrate Substituents in NTHi Strain R2486 and Mutant Strains R2846losB1 and R2846losB2. Information on the location of *P*Cho, *P*Etn, *P*, and Ac was provided by ESI tandem mass-spectrometry (MS/MS) in the positive mode following online separation by capillary electrophoresis (CE) of OS fractions derived from R2846, R2846*losB1*, and R2846*losB2* (*23*). The position of glycine could not be determined. Glycoforms containing *P* and *P*Cho substituents were predominantly detected in the leading fractions OS1 and OS2 (Tables S3 and S4). In OS1 of R2846, the doubly charged ion at *m/z* 1008.4 corresponded to the composition $Ac_2 \cdot PCho_2 \cdot Hex_3 \cdot Hep_4 \cdot PEtn \cdot An$Kdo-ol. MS/MS on this ion resulted, *inter alia*, in fragments at *m/z* 328.2, 358.2, 370.2, 724.4, 877.4, and 916.6 corresponding to the compositions *P*Cho·Hex, *P*Cho·Hep, Ac·*P*Cho·Hex, Ac·*P*Cho·Hex_2·Hep, Ac·*P*Cho·Hex·Hep_2·*P*Etn, and Ac·*P*Cho·Hex_2·Hep_2, respectively (Figure 5). Since there were no indications of an ion with the composition *P*Cho·Hep_2·*P*Etn, the intense ion at *m/z* 370.2 combined with the ion at *m/z* 877.4 indicated *P*Cho and Ac to be substituted to the hexose linked to HepIII. The ion at *m/z* 916.6 was fragmented, which resulted in ions at *m/z* 358.2, 562.3, and 754.4 corresponding to *P*Cho·Hep, Ac·*P*Cho·Hex·Hep, and Ac·*P*Cho·Hex·Hep_2, confirming *P*Cho to be substituted on HepIV and the hexose linked to HepI to be acetylated. A glycoform with the composition $Hex_3 \cdot Hep_4 \cdot PEtn \cdot P \cdot An$Kdo-ol (*m/z* 1681.0) observed in deacylated OS2 of R2846*losB2* was fragmented, giving rise to the ions at *m/z* 273.0, 435.2, and 750.4, corresponding to Hep·*P*, Hex·Hep·*P*, and Hex·Hep_2·*P*Etn·*P*, indicating that HepIII could be substituted with a phosphate.

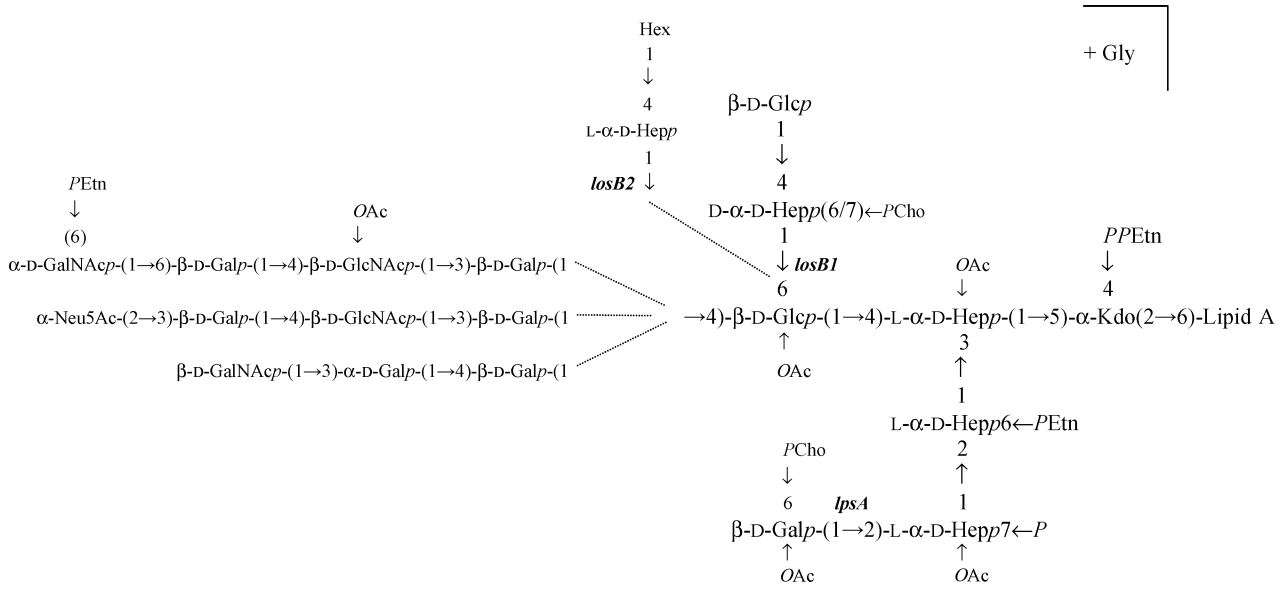
Information about the additional location of Ac was provided by CE-ESI-MS/MS in the positive mode as follows: in OS3 of R2846, a glycoform with the composition $Ac_2 \cdot Hex_3 \cdot Hep_4 \cdot PEtn \cdot An$Kdo-ol (*m/z* 1685.2) was fragmented, which resulted in a fragment ion at *m/z* 316.2, corresponding to HepII·*P*Etn. Significant ions at *m/z* 508.2, 550.3, 670.3, and 712.3 with the compositions $Hep_2 \cdot PEtn$, Ac·Hep_2·*P*Etn, Hex·Hep_2·*P*Etn, and Ac·Hex·Hep_2·*P*Etn, evidenced an acetylation site at HepIII. For OS1 of R2846*losB1*, CE-ESI-MS/MS on the ion at *m/z* 1288.7 corresponding to $Ac \cdot Hex_2 \cdot Hep_3 \cdot PEtn \cdot An$Kdo-ol indicated an additional acetylation site at HepI. In the spectrum, a minor fragment ion was detected at *m/z* 772.3 corresponding to a glycoform with the composition $Ac \cdot Hep_2 \cdot PEtn \cdot An$Kdo-ol. Since no ion with the composition Ac·Hep·*P*Etn was detected in the spectrum, the ion at *m/z* 772.3 indicated HepI to be acetylated. In OS1 of R2846*losB2*, the doubly charged ion at *m/z* 1270.2 corresponding to the composition $Ac_2 \cdot HexNAc_2 \cdot Hex_5 \cdot Hep_4 \cdot PEtn_2 \cdot An$Kdo-ol was further fragmented, giving ions at *m/z* 327.1, 489.4, 692.2, 734.5, 896.4, and 1058.5, corresponding to *P*Etn·HexNAc, *P*Etn·HexNAc·Hex, *P*Etn·HexNAc·Hex·HexNAc, *P*Etn·Ac·$HexNAc_2$·Hex and Ac·$HexNAc_2$·Hex_{2-3}·*P*Etn, respectively, indicating this element to contain both an acetate and a *P*Etn group. Since no fragments corresponding to Ac·HexNAc·*P*Etn or Ac·HexNAc·Hex·*P*Etn were observed, the Ac group is proposed to be situated on the 4- substituted GlcNAc unit of GaT.

Attempts were made to obtain the exact location sites of the *P*Cho and *P* substituents by a modification of methylation analysis (*28*). Following the methylation of OS1 of R2846, dephosphorylation, ethylation, hydrolysis, reduction, and acetylation were carried out, which afforded partially ethylated and methylated alditol acetates of originally phosphorylated sugars. Using this procedure, significant amounts of 1,5-di-*O*-acteyl-2,3,4-tri-*O*-methyl-6-*O*-ethyl-D-galactitol-1-d_1 were observed, evidencing *P*Cho to substitute the O-6 position of GalI. In addition, both 1,4,5-tri-*O*-acetyl-2,3,6-tri-*O*-methyl-7-*O*-ethyl-D-*glycero*-D-mannitol-1-d_1 and 1,4,5-tri-*O*-acetyl-2,3,7-tri-*O*-methyl-6-*O*-ethyl-D-*glycero*-D-mannitol-1-d_1 were detected, indicating *P*Cho to substitute either O-6 or O-7 of HepIV. It has been demonstrated that *P*Etn migrates between 6- and 7-positions in strong alkaline conditions (*34*), and it is therefore reasonable to assume that the same migration takes place for *P*Cho subunits. In the $^{31}P-^{1}H$ HMQC spectrum of OS2 of R2846, a strong signal assigned to a phosphate was observed at δ_H 3.91/δ_P 4.55 indicative of *P* linked to O-7 of HepIII (*34*). This was supported by the absence of derivatives corresponding to Hep residues phosphorylated at the 3- or 4-positions. Only 1,2,5-tri-*O*-acetyl-3,4,6-tri-*O*-methyl-7-*O*-ethyl-L-*glycero*-D-mannitol-1-d_1 and 1,2,5-tri-*O*-acetyl-3,4,7-tri-*O*-methyl-6-*O*-ethyl-L-*glycero*-D-mannitol-1-d_1 were detected in the modified methylation analysis, indicating a 2-substituted Hep with a phosphorylation site at the 7- and 6-positions. However, since HepII is substituted at O-6 with *P*Etn, this sugar will give the same alditol products.

DISCUSSION

We here report the lipopolysaccharide (LPS) structures expressed by nontypeable *Haemophilus influenzae* R2846, a strain for which the complete genome sequence has recently been obtained. The results are summarized in Chart 3. LPS from strain R2846 contains the conserved inner-core moiety L-α-D-Hep*p*-(1→2)-[*P*Etn→6]-L-α-D-Hep*p*-(1→3)-[β-D-Glc*p*-(1→4)]-L-α-D-Hep*p* linked to lipid A via Kdo-4-phosphate as every *Hi* strain investigated to date. In the predominantly

Chart 3



expressed glycoform, the β-D-Glc (GlcI) linked to the proximal heptose (HepI) is substituted at O-6 with the structural element [β-D-Glc*p*-(1→4)-D-α-D-Hep*p*-(1→] and as observed in several other *Hi* strains (*28, 35–37*), a β-D-Gal is linked to O-2 of the distal heptose (HepIII). Oligosaccharide extensions from O-6 of GlcI have previously been shown in strains having heptoses (L-α-D-Hep or D-α-D-Hep) in the outer core. Thus, extensions having β-D-Gal*p*-(1→4)-D-α-D-Hep*p*-(1→, β-D-Glc*p*-(1→4)-D-α-D-Hep*p*-(1→, or β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→6)-L-α-D-Hep*p*-(1→ and truncated versions thereof have been found at O-6 of GlcI (*6–8*).

Much of our current knowledge of the genetic basis of *Hi* LPS synthesis is based upon analysis of the Rd genome sequence (*10, 24, 38*), the first, and for a long time the only, *Hi* strain to be fully sequenced. NT*Hi* R2846 LPS contains structural features not present in the LPS of strain Rd. Thus, strain R2846 provides an excellent chance to investigate candidate genes required for the synthesis of LPS structures not present in strain Rd. In particular, the genes involved in the addition of the noncore heptose linked to O-6 of GlcI (*6–8*) have been unknown. Two genes, *losB1* and *losB2*, were identified as likely candidates to direct the addition of this noncore heptose, and in this study, we present detailed structural data consistent with both genes encoding heptosyltransferases. We conclude that the *losB1* gene in strain R2846 is responsible for adding the predominant DD-Hep to GlcI. When *losB1* was inactivated, DD-Hep was not expressed but minor quantities of LD-Hep were found in its place in the outer-core region. LPS from the R2846*losB1*/*losB2* mutant showed no noncore heptose, evidencing that the *losB2* gene is directing the addition of LD-Hep to the same position. From methylation analysis on R2846*losB1*, the LD-Hep was observed to be linked in the O-4 position, but the identity of the hexose substituting this residue could not be determined because of its limited abundance. Because of the lack of a reliable complementation system for *H. influenzae*, neither the *losB1* nor the *losB2* mutations have yet been complemented in the relevant mutant strains. However, the same phenotypes have been observed in multiple independent mutant strains made with different plasmid constructs for each gene, thus minimizing the chance that other factors relating to DNA transformation have influenced the LPS phenotype.

In the genome of strain R2846, each of the heptosyltransferase genes, *losB1* and *losB2*, is adjacent to genes *losA1* and *losA2*, respectively, encoding glycosyltransferases that are predicted to direct the substitution of hexoses to the noncore heptoses. Features of the LosA1 and LosA2 sequences are consistent with the sugar added being a glucose in each case (data not shown). As described previously (*12*), the presence of the gene pair *losA1* and *losB1* precludes the genes *lic2C* and *lic2B* being present. *Lic2C* is responsible for the addition of the first hexose to the middle heptose (HepII) in the LPS inner-core backbone (*38*). Its absence in strain R2846 concurs with our observations that no glycoforms containing chain extensions from HepII were present in this strain. The *lpsA* gene is required for the initiation of oligosaccharide extension from HepIII (*24*). Allelic variants of this gene direct the addition of either a β-D-Glc or a β-D-Gal in either (1→2) or (1→3) linkage (*39*). The *lpsA* sequence for strain R2846 would predict that HepIII is 2-substituted with a β-D-Gal (GalI), in agreement with our structural findings. Gene function was further confirmed by structural elucidation of LPS from a R2846*lpsA* mutant strain where HepIII was observed to be terminal in all glycoforms with no other significant changes observed in the LPS molecule.

In all *Hi* strains investigated to date that contain a β-D-Gal linked to HepIII, there is no evidence for this galactose being further extended with saccharides (*5*). Inconsistent with these results were the trace amounts of a Hex3Hep3 glycoform containing extension of two hexoses from HepIII that were detected in the HPLC-MS^n sequence analysis on OS1 of the R2846*losB1* and R2846*losB1*/*losB2* mutants; however, because of the limited amount of sample, these epitopes could not be elucidated further.

*P*Cho is an important structural feature of *Hi* LPS that contributes to the ability of the bacterium to colonize and persist within the human respiratory tract but that is detrimental to survival within the host vascular compartment (*40–42*). *P*Cho has been detected in a majority of *Hi* strains, typically at one

of several different positions (*5*). For NT*Hi* R2846, the bacterium elaborates LPS glycoforms that carry two *P*Cho substituents, a feature found for only a minority of strains (*37, 43*). The expression of *P*Cho is directed by the phase-variable *lic1* locus comprising four genes (*lic1A*−*lic1D*) of which *lic1D* determines the exact position of *P*Cho in the LPS structure (*35, 43*). In strain R2846, *P*Cho was observed linked to both O-6 of GalI and to HepIV. The substitution of *P*Cho to O-6 on a terminal β-Gal at HepIII has previously been observed in the type b strains RM7004 and Eagan as well as in nondisease-associated NT*Hi* strains 11 and 16 (*28, 35, 37*). The linkage of a *P*Cho to a noncore heptose on GlcI and the *lic1D* allele required for the synthesis of this epitope has recently been reported for NT*Hi* strain 1158 (*43*). Because of heterogeneity and the limited amount of sample, the exact position of *P*Cho on HepIV in strain R2846 was not unambiguously determined; however, the analysis on partially ethylated and methylated alditol acetates indicated *P*Cho to be linked to either O-6 or O-7 of HepIV. The *lic1* sequences in R2846 are equivalent to the *lic1* genes previously described that encode the addition of *P*Cho to HepIV and to a hexose linked to HepIII. In contrast to *P*Cho, *P* substitution has only been observed in *Hi* strain Eagan, where it was attached to O-4 of HepIII (*28*). In this study, CE-ESI-MS also evidenced *P* to be linked to HepIII; however, the proton-shift (δ 3.91), derived from the $^{1}H-^{31}P$ HMCQ spectrum of R2846 more likely originated from a H-7 than a H-4 shift of HepIII (*34*).

Sialic acid, which is a common and important feature of *Hi* LPS, has been shown to help the bacteria resist the killing effect of normal human serum (*44*). We have also demonstrated that host-derived Neu5Ac is incorporated in *Hi* LPS *in vivo* and that it is an essential virulence factor in experimental otitis media (*45*). Up to four sialyltransferase genes, *lic3A*, *lic3B*, *lsgB*, and *siaA* have been identified in any *Hi* strain (*46–48*). *lsgB* and *siaA* have been implicated in the sialylation of the lacto-*N*-neotetraose extension, SiaT from O-4 of HepI. The addition of these four sugar units follows a biosynthetic mechanism distinct from that of the rest of the LPS molecule and involves *en bloc* addition, similar to that seen in O-antigen biosynthesis (*33*). Both sialyllacto-*N*-neotetraose [α-Neu5Ac-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→] and the related structure [(*P*Etn→6)-α-D-Gal*p*NAc-(1→6)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-NAc-(1→3)-β-D-Gal*p*-(1→] were observed linked to O-4 of GlcI in strain R2846 (*32, 33*). Although the amounts were not of a quantity suited for NMR analysis, data obtained from methylation analysis combined with CE-ESI-MS and HPLC-MS^n clearly indicated the presence of both epitopes. However, the O-6 linkage of *P*Etn to GalNAc could not be confirmed. In addition to the SiaT and GaT epitopes, sequence analysis on HPLC-ESI-MS^n gave evidence for a tHexNAc-Hex-Hex- extension from GlcI and truncated versions thereof. Since traces of 4-substituted Gal were observed in the methylation analysis of both R2846 and the mutant strains, it is presumed that this epitope is from a globotetraose unit, [β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→], which is a common oligosaccharide epitope in several NT*Hi* strains (*5*). Expression of globotetraose and the truncated versions globoside [α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→] and lactose [β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→] involve the phase-variable expression of *lgtC* and *lic2A*, which are present in strain R2486.

Structural analysis of NT*Hi* strain R2846 indicated the bacterium to be heavily acetylated, and by the use of CE-ESI-MS, several acetylation sites could be determined, of which one is at a novel position. Acetylation on the GaT epitope, presumably on the 4)-β-D-Glc*p*NAc-(1 unit, have previously not been observed. In addition, acetates were seen linked to GalI, HepIII, HepI, and GlcI, positions that have been previously identified (*5*). The biological impact of complex *O*-acetylation patterns remains unknown, but recently, the gene encoding an *O*-acetyl-transferase, *oafA*, was studied (*49*). Inactivation of this gene led to the loss of one specific *O*-acetyl group in the LPS and an increase in the killing of the bacterium by normal human serum on the bacterium when compared to the wild-type strain.

The elucidation of the detailed structure of the LPS from strain R2846 and the concomitant genetic analysis based on the complete genome sequence have provided evidence for novel gene functions in NT*Hi* LPS synthesis. These analyses can be extended to investigate other structural features of NT*Hi* LPS where the genetic determinants are unclear, for example, the addition of the hexose(s) to HepIV. Also, we can investigate in more detail the genetic determinants for elements contributing to the extensive heterogeneity of R2846 LPS, such as the complex pattern of *O*-acetylation and the addition of *P* to HepIII. Findings from these studies can be extended to the LPS of other NT*Hi* strains where the detailed structure of the LPS glycoforms is known.

ACKNOWLEDGMENT

Alice L. Erwin and Arnold L. Smith are acknowledged for providing strain R2846 and for kindly sharing the annotated genome sequence in advance of publication. Mikael Engskog is acknowledged for PCP-extraction of the wild-type strain.

SUPPORTING INFORMATION AVAILABLE

Growth conditions of bacteria and LPS extraction, detailed descriptions of the experiments performed with HPLC and HPAEC-PAD, details regarding mild acid hydrolysis, dephosphorylation and purification of lipid A, parameters used in ESI-MS and NMR experiments, ESI-MS^n spectra (negative mode) performed on lipid A from R2846, ESI-MS data (negative ion mode) and proposed compositions for LPS-OH of strains R2846, R2846*losB1*, R2846*losB2*, R2846*losB1/losB2*, and R2846*lpsA*, lipid A structures of strain R2846 obtained after mild acid hydrolysis of LPS and observed by ESI-MS^n analysis (negative ion mode), ESI-MS data (negative ion mode) and proposed compositions for OS1–3 of strains R2846, R2846*losB1*, R2846*losB2*, and R2846*losB1/losB2*, and ESI-MS data (negative ion mode) and proposed compositions for deacylated OS1–3 of strains R2846, R2846*losB1*, R2846*losB2*, and R2846*losB1/losB2*. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Campagnari, A. A., Gupta, M. R., Dudas, K. C., Murphy, T. F., and Apicella, M. A. (1987) Antigenic diversity of lipooligosaccharides of nontypable *Haemophilus influenzae*. *Infect. Immun. 55*, 882–887.

2. Murphy, T. F., and Apicella, M. A. (1987) Nontypable *Haemophilus influenzae*: a review of clinical aspects, surface antigens, and the human immune response to infection. *Rev. Infect. Dis. 9*, 1–15.
3. Anderson, P., Johnston, R. B., Jr., and and Smith, D. H. (1972) Human serum activities against *Haemophilus influenzae*, type b. *J. Clin. Invest. 51*, 31–38.
4. Zwahlen, A., Rubin, L. G., and Moxon, E. R. (1986) Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. *Microb. Pathog. 1*, 465–473.
5. Schweda, E. K., Richards, J. C., Hood, D. W., and Moxon, E. R. (2007) Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*: Implication in virulence. *Int. J. Med. Microbiol. 297*, 297–306.
6. Rahman, M. M., Gu, X. X., Tsai, C. M., Kolli, V. S., and Carlson, R. W. (1999) The structural heterogeneity of the lipooligosaccharide (LOS) expressed by pathogenic non-typeable *Haemophilus influenzae* strain NTHi 9274. *Glycobiology. 9*, 1371–1380.
7. Månsson, M., Hood, D. W., Moxon, E. R., and Schweda, E. K. (2003) Structural diversity in lipopolysaccharide expression in nontypeable *Haemophilus influenzae*. Identification of L-*glycero*-D-*manno*-heptose in the outer-core region in three clinical isolates. *Eur. J. Biochem. 270*, 610–624.
8. Månsson, M., Hood, D. W., Moxon, E. R., and Schweda, E. K. (2003) Structural characterization of a novel branching pattern in the lipopolysaccharide from nontypeable *Haemophilus influenzae*. *Eur. J. Biochem. 270*, 2979–2991.
9. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science. 269*, 496–512.
10. Hood, D. W., Deadman, M. E., Allen, T., Masoud, H., Martin, A., Brisson, J. R., Fleischmann, R., Venter, J. C., Richards, J. C., and Moxon, E. R. (1996) Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharide biosynthesis. *Mol. Microbiol. 22*, 951–965.
11. Hood, D. W., Deadman, M. E., Jennings, M. P., Bisercic, M., Fleischmann, R. D., Venter, J. C., and Moxon, E. R. (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. U.S.A. 93*, 11121–11125.
12. Erwin, A. L., Nelson, K. L., Mhlanga-Mutangadura, T., Bonthuis, P. J., Geelhood, J. L., Morlin, G., Unrath, W. C., Campos, J., Crook, D. W., Farley, M. M., Henderson, F. W., Jacobs, R. F., Muhlemann, K., Satola, S. W., van Alphen, L., Golomb, M., and Smith, A. L. (2005) Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. *Infect. Immun. 73*, 5853–5863.
13. Harrison, A., Dyer, D. W., Gillaspy, A., Ray, W. C., Mungur, R., Carson, M. B., Zhong, H., Gipson, J., Gipson, M., Johnson, L. S., Lewis, L., Bakaletz, L. O., and Munson, R. S., Jr. (2005) Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. *J. Bacteriol. 187*, 4627–4636.
14. Tullius, M. V., Phillips, N. J., Scheffler, N. K., Samuels, N. M., Munson Jr, R. S., Jr., Hansen, E. J., Stevens-Riley, M., Campagnari, A. A., and Gibson, B. W. (2002) The *lbgAB* gene cluster of *Haemophilus ducreyi* encodes a beta-1,4-galactosyltransferase and an alpha-1,6-DD-heptosyltransferase involved in lipooligosaccharide biosynthesis. *Infect. Immun. 70*, 2853–2861.
15. Post, D. M., Munson, R. S., Jr., Baker, B., Zhong, H., Bozue, J. A., and Gibson, B. W. (2007) Identification of genes involved in the expression of atypical lipooligosaccharide structures from a second class of *Haemophilus ducreyi*. *Infect. Immun. 75*, 113–121.
16. Barenkamp, S. J., and Leininger, E. (1992) Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. *Infect. Immun. 60*, 1302–1313.
17. Herriott, R. M., Meyer, E. M., and Vogt, M. (1970) Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. *J. Bacteriol. 101*, 517–524.
18. Lundström, S. L., Twelkmeyer, B., Sagemark, M. K., Li, J., Richards, J. C., Hood, D. W., Moxon, E. R., and Schweda, E. K. (2007) Novel globoside-like oligosaccharide expression patterns in nontypeable *Haemophilus influenzae* lipopolysaccharide. *FEBS. J. 274*, 4886–4903.
19. Kussak, A., and Weintraub, A. (2002) Quadrupole ion-trap mass spectrometry to locate fatty acids on lipid A from Gram-negative bacteria. *Anal. Biochem. 307*, 131–137.
20. Mikhail, I., Yildirim, H. H., Lindahl, E. C., and Schweda, E. K. (2005) Structural characterization of lipid A from nontypeable and type f *Haemophilus influenzae*: variability of fatty acid substitution. *Anal. Biochem. 340*, 303–316.
21. Sawardeker, J. S., Sloneker, J. H., and Jeanes, A. (1965) Quantitative determination of monosaccharides as their alditol acetates by gas liquid chromatography. *Anal. Chem. 37*, 1602–1604.
22. Gerwig, G. J., Kamerling, J. P., and Vliegenthart, J. F. (1979) Determination of the absolute configuration of mono-saccharides in complex carbohydrates by capillary G.L.C. *Carbohydr. Res. 77*, 1–7.
23. Li, J., Bauer, S. H., Månsson, M., Moxon, E. R., Richards, J. C., and Schweda, E. K. (2001) Glycine is a common substituent of the inner core in *Haemophilus influenzae* lipopolysaccharide. *Glycobiology. 11*, 1009–1015.
24. Hood, D. W., Cox, A. D., Wakarchuk, W. W., Schur, M., Schweda, E. K., Walsh, S. L., Deadman, M. E., Martin, A., Moxon, E. R., and Richards, J. C. (2001) Genetic basis for expression of the major globotetraose-containing lipopolysaccharide from H. influenzae strain Rd (RM118). *Glycobiology. 11*, 957–967.
25. Bauer, S. H., Månsson, M., Hood, D. W., Richards, J. C., Moxon, E. R., and Schweda, E. K. (2001) A rapid and sensitive procedure for determination of 5-*N*-acetyl neuraminic acid in lipopolysaccharides of *Haemophilus influenzae*: a survey of 24 non-typeable *H. influenzae* strains. *Carbohydr. Res. 335*, 251–260.
26. Risberg, A., Masoud, H., Martin, A., Richards, J. C., Moxon, E. R., and Schweda, E. K. (1999) Structural analysis of the lipopolysaccharide oligosaccharide epitopes expressed by a capsule-deficient strain of *Haemophilus influenzae* Rd. *Eur. J. Biochem. 261*, 171–180.
27. Risberg, A., Alvelius, G., and Schweda, E. K. (1999) Structural analysis of the lipopolysaccharide oligosaccharide epitopes expressed by *Haemophilus influenzae* strain RM.118-26. *Eur. J. Biochem. 265*, 1067–1074.
28. Masoud, H., Moxon, E. R., Martin, A., Krajcarski, D., and Richards, J. C. (1997) Structure of the variable and conserved lipopolysaccharide oligosaccharide epitopes expressed by *Haemophilus influenzae* serotype b strain Eagan. *Biochemistry. 36*, 2091–2103.
29. Helander, I. M., Lindner, B., Brade, H., Altmann, K., Lindberg, A. A., Rietschel, E. T., and Zähringer, U. (1988) Chemical structure of the lipopolysaccharide of *Haemophilus influenzae* strain I-69 Rd−/b+. Description of a novel deep-rough chemotype. *Eur. J. Biochem. 177*, 483–492.
30. Reinhold, V. N., Reinhold, B. B., and Costello, C. E. (1995) Carbohydrate molecular weight profiling, sequence, linkage, and branching data: ES-MS and CID. *Anal. Chem. 67*, 1772–1784.
31. Schweda, E. K., Hegedus, O. E., Borrelli, S., Lindberg, A. A., Weiser, J. N., Maskell, D. J., and Moxon, E. R. (1993) Structural studies of the saccharide part of the cell envelope lipopolysaccharide from *Haemophilus influenzae* strain AH1–3 (lic3+). *Carbohydr. Res. 246*, 319–330.
32. Cox, A. D., Hood, D. W., Martin, A., Makepeace, K. M., Deadman, M. E., Li, J., Brisson, J. R., Moxon, E. R., and Richards, J. C. (2002) Identification and structural characterization of a sialylated lacto-*N*-neotetraose structure in the lipopolysaccharide of *Haemophilus influenzae*. *Eur. J. Biochem. 269*, 4009–4019.
33. Hood, D. W., Randle, G., Cox, A. D., Makepeace, K., Li, J., Schweda, E. K., Richards, J. C., and Moxon, E. R. (2004) Biosynthesis of cryptic lipopolysaccharide glycoforms in *Haemophilus influenzae* involves a mechanism similar to that required for O-antigen synthesis. *J. Bacteriol. 186*, 7429–7439.
34. Stewart, A., Bernlind, C., Martin, A., Oscarson, S., Richards, J. C., and Schweda, E. K. (1998) Studies of alkaline mediated phosphate migration in synthetic phosphoethanolamine L-*glycero*-D-*manno*-heptoside derivatives. *Carbohydr. Res. 313*, 193–202.
35. Schweda, E. K., Brisson, J. R., Alvelius, G., Martin, A., Weiser, J. N., Hood, D. W., Moxon, E. R., and Richards, J. C. (2000) Characterization of the phosphocholine-substituted oligosaccharide in lipopolysaccharides of type b *Haemophilus influenzae*. *Eur. J. Biochem. 267*, 3902–3913.
36. Cox, A. D., Masoud, H., Thibault, P., Brisson, J. R., van der Zwan, M., Perry, M. B., and Richards, J. C. (2001) Structural analysis of the lipopolysaccharide from the nontypable *Haemophilus influenzae* strain SB 33. *Eur. J. Biochem. 268*, 5278–5286.
37. Landerholm, M. K., Li, J., Richards, J. C., Hood, D. W., Moxon, E. R., and Schweda, E. K. (2004) Characterization of novel

structural features in the lipopolysaccharide of nondisease associated nontypeable *Haemophilus influenzae*. *Eur. J. Biochem. 271*, 941–953.

38. Hood, D. W., Deadman, M. E., Cox, A. D., Makepeace, K., Martin, A., Richards, J. C., and Moxon, E. R. (2004) Three genes, *lgtF*, *lic2C* and *lpsA*, have a primary role in determining the pattern of oligosaccharide extension from the inner core of *Haemophilus influenzae* LPS. *Microbiology. 150*, 2089–2097.
39. Deadman, M. E., Lundström, S. L., Schweda, E. K., Moxon, E. R., and Hood, D. W. (2006) Specific amino acids of the glycosyltransferase LpsA direct the addition of glucose or galactose to the terminal inner core heptose of *Haemophilus influenzae* lipopolysaccharide via alternative linkages. *J. Biol. Chem. 281*, 29455–29467.
40. Weiser, J. N., Pan, N., McGowan, K. L., Musher, D., Martin, A., and Richards, J. (1998) Phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. *J. Exp. Med. 187*, 631–640.
41. Swords, W. E., Buscher, B. A., Ver Steeg Ii, K., Preston, A., Nichols, W. A., Weiser, J. N., Gibson, B. W., and Apicella, M. A. (2000) Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol. Microbiol. 37*, 13–27.
42. Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M., and Weiser, J. N. (2000) The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C-reactive protein-mediated killing. *Mol. Microbiol. 35*, 234–245.
43. Fox, K. L., Li, J., Schweda, E. K., Vitiazeva, V., Makepeace, K., Jennings, M. P., Moxon, E. R., and Hood, D. W. (2008) Duplicate copies of Lic1 direct the addition of multiple phosphocholine residues in the lipopolysaccharide of *Haemophilus influenzae*. *Infect. Immun. 76*, 588–600.
44. Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C., and Moxon, E. R. (1999) Sialic acid in the lipopolysaccharide of *Haemophilus influenzae*: strain distribution, influence on serum resistance and structural characterization. *Mol. Microbiol. 33*, 679–692.
45. Bouchet, V., Hood, D. W., Li, J., Brisson, J. R., Randle, G. A., Martin, A., Li, Z., Goldstein, R., Schweda, E. K., Pelton, S. I., Richards, J. C., and Moxon, E. R. (2003) Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc. Natl. Acad. Sci. U.S.A. 100*, 8898–8903.
46. Hood, D. W., Cox, A. D., Gilbert, M., Makepeace, K., Walsh, S., Deadman, M. E., Cody, A., Martin, A., Månsson, M., Schweda, E. K., Brisson, J. R., Richards, J. C., Moxon, E. R., and Wakarchuk, W. W. (2001) Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from *Haemophilus influenzae*. *Mol. Microbiol. 39*, 341–350.
47. Jones, P. A., Samuels, N. M., Phillips, N. J., Munson, R. S., Jr., Bozue, J. A., Arseneau, J. A., Nichols, W. A., Zaleski, A., Gibson, B. W., and Apicella, M. A. (2002) *Haemophilus influenzae* type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation. *J. Biol. Chem. 277*, 14598–14611.
48. Fox, K. L., Cox, A. D., Gilbert, M., Wakarchuk, W. W., Li, J., Makepeace, K., Richards, J. C., Moxon, E. R., and Hood, D. W. (2006) Identification of a bifunctional lipopolysaccharide sialyltransferase in *Haemophilus influenzae*: incorporation of disialic acid. *J. Biol. Chem. 281*, 40024–40032.
49. Fox, K. L., Yildirim, H. H., Deadman, M. E., Schweda, E. K., Moxon, E. R., and Hood, D. W. (2005) Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in *Haemophilus influenzae*, identification of a gene for adding *O*-acetyl groups. *Mol. Microbiol. 58*, 207–216.

BI702510B