Specific Amino Acids of the Glycosyltransferase LpsA Direct the Addition of Glucose or Galactose to the Terminal Inner Core Heptose of \textit{Haemophilus influenzae} Lipopolysaccharide via Alternative Linkages

Mary E. Deadman\textsuperscript{1,2}, Susanna L. Lundström\textsuperscript{3}, Elke K. H. Schweda\textsuperscript{3}, E. Richard Moxon\textsuperscript{1,4}, and Derek W. Hood\textsuperscript{1,4}

From the \textsuperscript{4}Department of Paediatrics, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, United Kingdom and the \textsuperscript{5}Clinical Research Centre, Karolinska Institutet and University College of South Stockholm, NOVUM, S-141 86 Huddinge, Sweden

Lipopolysaccharide is the major glycolipid of the cell wall of the bacterium \textit{Haemophilus influenzae}, a Gram-negative commensal and pathogen of humans. Lipopolysaccharide is both a virulence determinant and a target for host immune responses. Glycosyltransferases have high donor and acceptor substrate specificities that are generally limited to catalysis of one unique glycosidic linkage. The \textit{H. influenzae} glycosyltransferase LpsA is responsible for the addition of a hexose to the distal heptose of the lipopolysaccharide molecule and belongs to the glycosyltransferase family 25. The hexose added can be either glucose or galactose and linkage to the heptose can be either $\beta1-2$ or $\beta1-3$. Each \textit{H. influenzae} strain uniquely produces only one of the four possible combinations of linked sugar in its lipopolysaccharide. We show that, in any given strain, a specific allelic variant of LpsA directs the anomic linkage and the added hexose, glucose, or galactose. Site-directed mutagenesis of a single key amino acid at position 151 changed the hexose added \textit{in vivo} from glucose to galactose or \textit{vice versa}. By constructing chimeric \textit{lpsA} gene sequences, it was shown that the 3' end of the gene directs the anomic linkage ($\beta1-2$ or $\beta1-3$) of the added hexose. The \textit{LpsA} gene is the first known example where interstrain variation in lipopolysaccharide core structure is directed by the specific sequence of a genetic locus encoding enzymes directing one of four alternative possible sugar additions from the inner core.

\textit{Haemophilus influenzae} (Hi)\textsuperscript{3} is a host-adapted bacterium that regularly colonizes the respiratory tract of humans. Encapsulated strains can cause invasive disease, such as meningitis or septicaemia, whereas nonencapsulated, nontypeable (NTHi) strains commonly cause otitis media, sinusitis, and acute lower respiratory tract infections. Lipopolysaccharide (LPS) is a glycolipid that is a major component of the cell wall, functions as a virulence determinant (1), and is a target for host immune responses.

The structure of the LPS of Hi consists of a membrane-anchoring lipid A moiety, which is linked to a phosphorylated 2-keto-3-deoxyoctulosonic acid residue, which in turn is linked to a 1-glyco-$D$-manno-heptose (Hep) trisaccharide unit. The middle heptose (HepII) is substituted at the 6-position by a phosphoethanolamine (2). This forms the invariant inner core of the LPS molecule. The Hep residues each provide a point at which a hexose (Hex) can be added; these can in turn be further extended into oligosaccharide chains that form the outer core of the LPS molecule. The outer core is very heterogeneous both between and within strains. Substitution with other residues, such as phosphate, acetyl groups, glycine, and sialic acid can further add to the structural variability of the LPS molecule (Fig. 1).

Glycosyltransferases (GTs) (EC 2.4.1.-) are key enzymes involved in the synthesis of complex carbohydrates, including glycolipids, glycoproteins, and polysaccharides. They transfer an activated mono- or oligosaccharide residue to an acceptor molecule. Hundreds of independent glycosyltransferase functions are required to generate all known glycosidic linkages. Glycosyltransferases have high donor and acceptor substrate specificities and are in general limited to catalysis of one unique glycosidic linkage (3). It has yet to be determined how glycosyltransferases create these glycosidic linkages or how donor and acceptor specificity is determined (4).

Campbell et al. (5) introduced a classification system for glycosyltransferases to place them into families based on similarities in amino acid sequence. There were initially 27 GT families containing 600 sequences (5). Since then, the number of families has grown to 86, containing over 33,000 sequences (6). Unfortunately, the vast majority of sequences have been added solely on the basis of sequence similarity; very few of the GTs have been defined biochemically. To utilize this genomic resource to the fullest, it is essential to understand the sequences of the enzymes themselves and how these sequences
Functional Allelic Variants of the lpsA Glycosyltransferase

relate to an enzyme’s structure, mechanism, and specificity of action. Having a confirmed function for specific transferase sequences is essential in achieving this end.

The *Hi* glycosyltransferase LpsA has been shown to be responsible for addition of a Hex to the distal heptose (HepIII) of the inner core of the *Hi* LPS molecule (7) and belongs to the family 25 GTs (6, 8). More recently, it has been shown that the Hex can be either Glc or Gal; moreover, the linkage of the Hex to HepIII can be by a β1–2 or β1–3 linkage (9–11). Each *Hi* strain has been shown to produce only one of the four possible combinations of linked sugar in its LPS, suggesting that the addition might be directed by a specific sequence of the gene present in any given strain.

In this paper, we investigate the *lpsA* gene and its function in 28 *Hi* strains and conclude that variations in the sequence of the GT direct both the specificity of the linkage and the addition of hexose (glucose or galactose).

**EXPERIMENTAL PROCEDURES**

**Bacterial Strains and Culture Conditions**

The capsular and nontypeable *Hi* strains used in this study (Table 1) have been described previously (12, 13). *NTHi* strain 176 described in this study is a distinct clinical isolate from that described by Schweda et al. (11). The strain investigated in that study has been redesignated as *NTHi* 1292. *Hi* strains were grown at 37 °C in brain heart infusion broth supplemented with haemin (10 μg ml⁻¹), NAD (2 μg ml⁻¹), and, when appropriate, kanamycin (10 μg ml⁻¹). Brain heart infusion broth was supplemented with agar (1% w/v) and Levinthal’s reagent (14) for plate growth.

*Escherichia coli* strain DH5α was used to propagate most cloned DNA and was cultured at 37 °C in LB broth (15) supplemented, when appropriate, with agar (1.5% w/v) for growth on plates and kanamycin (50 μg ml⁻¹) or ampicillin (100 μg ml⁻¹) for selection of plasmid constructs.

**Plasmid Constructs for Transferring the lpsA Gene between Strains**

DNA-modifying enzymes were supplied by Roche Applied Science and used according to the manufacturer’s instructions. All oligonucleotide primers used in this study are listed in Table 2. Oligonucleotide primers were designed from the published *Hi* KW20 complete genome sequence (16) to amplify a region of DNA encompassing the *lpsA* (HI0765) gene by PCR. Primers 6018F (anneals in HI10763, nadB) and 6018G (anneals in HI0767, encoding a conserved hypothetical protein) amplified the region of DNA surrounding the *lpsA* gene from chromosomal DNA from strains RM118 (Rd), RM153 (Eagan), *NTHi* 486, *NTHi* 1292, and *NTHi* 176. PCR conditions were for 1-min periods of denaturation (94 °C), annealing (50 °C), and polymerization (72 °C) for 30 cycles. PCR products were cloned into plasmid pT7blue (Novagen) under conditions described by the supplier and transformed into *E. coli* strain DH5α (15).

To disrupt the *lpsA* gene in plasmid constructs, cloned DNA was partially digested with restriction endonuclease EcoRI, which cuts within the *lpsA* gene but also in the plasmid, and then ligated with a kanamycin resistance cassette from pUC4kan (Amersham Biosciences), released by digestion with EcoRI. Following transformation into *E. coli*, the correct clones were confirmed by restriction endonuclease digestion and PCR amplification using the oligonucleotides 6018C (anneals 120 – 106 bp upstream of the *lpsA* initiation codon) and 6018A (anneals 84 – 103 bp downstream of the *lpsA* stop codon).

To construct plasmids containing intact *lpsA* genes linked to a kanamycin resistance marker, a cassette was inserted into DNA adjacent to *lpsA* for selection. Plasmids, derived from DNA amplified by PCR using primers 6018F/6018G from each strain, were digested with the restriction endonuclease SnaBl, which cuts in the HI0767 open reading frame, and then ligated with a kanamycin resistance cassette released from pUC4kan by digestion with HincII. Following transformation into *E. coli*, plasmid clones were confirmed by restriction endonuclease digestion and PCR amplification using the 6018A/6018C primer set. Plasmids containing the cloned *lpsA* and interrupted HI0767 genes were designated p10.6SA, p11.7SA, p176.2KanA, and p486.1Kan for selection. Plasmids, derived from DNA amplified by PCR using primers 6018F/6018G from each strain, were digested with the restriction endonuclease SnaBl, which cuts in the HI0767 open reading frame, and then ligated with a kanamycin resistance cassette released from pUC4kan by digestion with HincII. Following transformation into *E. coli*, plasmid clones were confirmed by restriction endonuclease digestion and PCR amplification using the 6018A/6018C primer set. Plasmids containing the cloned *lpsA* and interrupted HI0767 genes were designated p10.6SA, p11.7SA, p176.2KanA, and p486.1Kan for selection. Plasmids, derived from DNA amplified by PCR using primers 6018F/6018G from each strain, were digested with the restriction endonuclease SnaBl, which cuts in the HI0767 open reading frame, and then ligated with a kanamycin resistance cassette released from pUC4kan by digestion with HincII. Following transformation into *E. coli*, plasmid clones were confirmed by restriction endonuclease digestion and PCR amplification using the 6018A/6018C primer set. Plasmids containing the cloned *lpsA* and interrupted HI0767 genes were designated p10.6SA, p11.7SA, p176.2KanA, and p486.1Kan for selection. Plasmids, derived from DNA amplified by PCR using primers 6018F/6018G from each strain, were digested with the restriction endonuclease SnaBl, which cuts in the HI0767 open reading frame, and then ligated with a kanamycin resistance cassette released from pUC4kan by digestion with HincII. Following transformation into *E. coli*, plasmid clones were confirmed by restriction endonuclease digestion and PCR amplification using the 6018A/6018C primer set. Plasmids containing the cloned *lpsA* and interrupted HI0767 genes were designated p10.6SA, p11.7SA, p176.2KanA, and p486.1Kan for selection.

| Strain or plasmd name | Source |
|-----------------------|--------|
| Strains               |        |
| RM118 (Rd)            | Disease isolate, United States (43). Capsular b serotype |
| RM118 (Rd)            | Disease isolate, United States (43). Capsular b serotype |
| RM2104                | Disease isolate, United States (43). Capsular b serotype |
| RM71004               | Disease isolate, United States (43). Capsular b serotype |
| NTHi 162              | Finnish* |
| NTHi 176              | Finnish |
| NTHi 285              | Finnish |
| NTHi 375              | Finnish |
| NTHi 432              | Finnish |
| NTHi 477              | Finnish |
| NTHi 486              | Finnish |
| NTHi 667              | Finnish |
| NTHi 723              | Finnish |
| NTHi 981              | Finnish |
| NTHi 1008             | Finnish |
| NTHi 1124             | Finnish |
| NTHi 1158             | Finnish |
| NTHi 1159             | Finnish |
| NTHi 1180             | Finnish |
| NTHi 1181             | Finnish |
| NTHi 1200             | Finnish |
| NTHi 1207             | Finnish |
| NTHi 1209             | Finnish |
| NTHi 1231             | Finnish |
| NTHi 1232             | Finnish |
| NTHi 1233             | Finnish |
| NTHi 1247             | Finnish |
| NTHi 1268             | Finnish |
| NTHi 1292             | Finnish |
| Plasmids              |        |
| p10.6SA               | Hood et al. (9) |
| p11.7SA               | Hood et al. (9) |
| p176.2KanA            | This study |
| p486.1Kan             | This study |
| pC151TRd3             | This study |
| pT151C1762            | This study |
| pOClTRd3AIII          | This study |
| p486.1KanAIII         | This study |
| p5’486/3’Rd           | This study |
| p5’486/3’C151TRd      | This study |
| p5’Rd/3’486           | This study |

* Finnish, these strains were isolated as part of a series of studies relating to conjugate pneumococcal vaccines conducted by the Finnish Otitis Media Study Group during 1994–1995.
Functional Allelic Variants of the lpsA Glycosyltransferase

Table 2

Primer sequences used for PCR amplification and sequencing of the lpsA region (positions 827501–830354 in the Hi KW20 genome) with the relative position on the PCR fragment indicated.

| Primer name | Primer sequence | Position on fragment |
|-------------|----------------|----------------------|
| 6018A       | 5’-GCCTGGGGCAAATTTGGCG-3’ | 1959–1942            |
| 6016B       | 5’-AATGCAAAGGAGCGTG-3’ | 994–1011             |
| 6018C       | 5’-TTGAATACCTGTTGACAC-3’ | 965–983              |
| 6018D       | 5’-GCACATTTTUTGATACAG-3’ | 1104–1121            |
| 6018E       | 5’-TGTGGGCCTCGAATGTC-3’ | 1297–1304            |
| 6018F       | 5’-TCTCTGCATCTTCATC-3’ | 73–90                |
| 6018G       | 5’-AACACTCAAAGCTGATC-3’ | 2690–2673            |
| 6018H       | 5’-AGGAAAATCTGCTTACAC-3’ | 1185–1168            |
|               | 5’-ACTCTCGAGCAAGAATGG-3’ | 765–783              |
| RdC151Tfor   | 5’-GGTTAGCGGAATTTGG-3’ | 1531–1559            |
| RdC151Trev   | 5’-TGTGATTGAATAGCCAGC-3’ | 1559–1531            |
| RdC151Tfor   | 5’-GAAGATACTGATATATTGAAG-3’ | 1531–1559            |
| RdC151Trev   | 5’-GAACTTGCGTTGATTTCC-3’ | 1559–1531            |
| RdAflIIrev   | 5’-GCAAATGGCAAGATGTTC-3’ | 1287–1324            |
| RdAflIIrev   | 5’-TTGCGGTTGATTCACTTG-3’ | 1284–1287            |
| 486AflIIfor  | 5’-CTGAAGATGCTGCTTACAG-3’ | 1286–1323            |
| 486AflIIrev  | 5’-CCCATAAAACAGCATATTCG-3’ | 1323–1386            |
| lpsAF        | 5’-TTACGCCCGACATTATTG-3’ | 1205–1222            |
| lpsAFrev     | 5’-CAATAATGTCGGGCGTAATAG-3’ | 1222–1202           |
| lpsAN        | 5’-GATGGCAGGAAATCTG-3’ | 1533–1552            |
| lpsAO        | 5’-TTGCCCCGTATCTGACC-3’ | 1610–1627            |
| Rdspec1      | 5’-GCAAATGGCAAGATGTTC-3’ | 1453–1470            |
| Rdspec2      | 5’-GAACCTCTGCTATGTACC-3’ | 1779–1762            |
| 486spec1     | 5’-GAAAGATTGCTAATTTGGAG-3’ | 1426–1446           |
| 486idgo2     | 5’-TGTGATTGAAATGCGACG-3’ | 1557–1540            |
| 486idgo3     | 5’-GCTTTGGGGCGGAAATTTGG-3’ | 1784–1766           |

DNA Sequence Analysis of the lpsA Gene

Oligonucleotide primers 6018C and 6018A were used to specifically amplify the lpsA gene by PCR from the chromosomal DNA of the 28 Hi strains described in Table 1, under conditions described above. Sequencing of the PCR products was performed using dye terminator technology (Big Dye DNA sequencing kit; PE Applied Biosystems) and oligonucleotides listed in Table 2, with an ABI-Prism 377 autosequencer. DNA sequence was assembled and analyzed as described previously (13).

Site-directed Mutagenesis of lpsA

Single Amino Acid Changes—Site-specific changes in single amino acids of the encoded protein were introduced into the cloned lpsA gene using the QuikChange (Stratagene) protocol and oligonucleotide primers containing a specific sequence mismatch (underlined in Table 2). Briefly, RdC151Tfor and RdC151Trev (reverse and complement of RdC151Tfor primer) and 176T151Cfor and 176T151Crev (reverse and complement of 176T151Cfor primer) were used to amplify the plasmids p10.6SA (Rd) and pC151TRd3 with no resulting change in the encoded amino acid sequence. Oligonucleotide primers RdAflIIfor and RdAflIIrev (reverse and complement of RdAflIIfor primer), which anneal 207–237 bp downstream of the Rd lpsA initiation codon, and 486AflIIfor and 486AflIIrev (reverse and complement of 486AflIIfor primer), which anneal 206–236 bp downstream of the NTHi 486 lpsA initiation codon containing a specific sequence mismatch (underlined in Table 2) were designed to create a unique AflII restriction site (boldface type in Table 2) in the cloned lpsA genes. The primers were used to amplify the plasmids p10.6SA (Rd), pC151TRd3, or p486.1KanB by PCR. Following processing and transformation, clones p10.6SAfII, pC151TRd3fII, and p486.1KanBfII were digested by restriction endonucleases AflII and SacII, and the fragments were run on a 1% agarose gel and then purified (Qiagen) to obtain two fragments, each containing part of the lpsA region from each strain. The 495-bp SacII/AflII fragment containing the 5’ portion of the lpsA gene was denoted 5’, and the 6261-bp AflII/SacII fragment containing the 3’ portion of the lpsA gene plus the remainder of the insert and the vector was denoted 3’. The fragments were then ligated to produce chimeric plasmids p5’/3’ Rd, p5’/3’C151 TRd, and p5’/3’486 and then transformed into E. coli. Plasmids purified from the resulting transformants were checked by PCR amplification with specific primers and by DNA sequencing to confirm that the constructs were correct.

Construction of H. influenzae Mutant and Allelic Exchange Strains

Plasmids containing disrupted lpsA genes were linearized by digestion with XbaI and then used to transform Hi strains.
Functional Allelic Variants of the \textit{lpsA} Glycosyltransferase

Rd, Eagan, NTHi 486, and NTHi 176 by the MIV procedure (17). Cells were plated upon brain heart infusion-kanamycin, and then transformants were confirmed as \textit{lpsA} mutants by PCR using the 6018A/6018C primer set and by Southern analysis (15). Plasmids containing \textit{lpsA} and a kanamycin resistance cassette in HI0767 were digested with BamHI and XbaI to release the insert from the vector and then used for transformation of the appropriate \textit{H. influenzae} strains by the MIV procedure. The chimeric plasmids were similarly digested and used to transform the \textit{Hi} strains Rd, Eagan, and NTHi 486 only. Allelic exchange in transformants was confirmed by PCR amplification and DNA sequencing as described above.

**Analysis of LPS by Electrophoresis**

LPS was prepared from bacterial lysates and analyzed by T-SDS-PAGE as previously described (12).

**Structural Analysis of Purified LPS**

Each LPS was structurally investigated using chemical, NMR, and mass spectrometry methods following a protocol outlined previously (18, 19). Briefly, cells from 5-liter batch cultures (5 × 1 liter) were harvested after overnight growth, and LPS was extracted from lyophilized bacteria. O-Deacylated LPS (LPS-OH) was prepared by mild hydrazinolysis. Core oligosaccharide fractions (OS) were obtained following mild acid hydrolysis of LPS. Electrospray ionization mass spectrometry (ESI-MS) on LPS-OH and OS samples was recorded on a VG Quattro triple quadrupole mass spectrometer (Micromass, Manchester, UK) in the negative ion mode. Multiple step tandem ESI-MS (ESI-MS\(^n\)) experiments on permethylated OS samples were performed in the positive ion mode on a Finnigan LCQ iontrap mass spectrometer (Finnigan-MAT, San Jose, CA). Sugars were identified as their alditol acetates using authentic standards. Methylation analysis was accomplished on dephosphorylated OS obtained after treatment with 48% hydrogen fluoride. NMR spectra were recorded for OS samples in deuterium oxide (D\(_2\)O) at 25 °C. Spectra were acquired on a JEOL 500-MHz spectrometer using standard pulse sequences as previously described. For interresidue correlation, two-dimensional NOE spectroscopy experiments with a mixing time of 200 or 250 ms were used. TOCSY experiments were done using mixing times of 50 and 180 ms.

**RESULTS**

**Revised Assignment of Translational Start Codons of \textit{H. influenzae} LpsA in Rd and 27 Other Strains**

The gene \textit{lpsA} is required to add the first hexose sugar (Glc or Gal) to the third hexose (HepIII) of the conserved triheptose backbone of \textit{Hi} LPS (Fig. 1). Using PCR amplification, it was shown that sequences homologous to \textit{lpsA} were present in each of 28 distinct \textit{H. influenzae} isolates that included both encapsulated and capsule-deficient (nontypeable) strains (data not shown). We sequenced \textit{lpsA} from each of the strains (GenBank\textsuperscript{TM} DQ647395–DQ647422) and aligned the deduced amino acid sequences (Fig. 2). When the strain Rd LpsA sequence was compared with the closest 48 matches obtained by searching with the BLASTP algorithm, it was atypical. Based on the annotation of \textit{lpsA} (HI076S) from the KW20 genome sequence (16), in which GTG (Val) was proposed as the translational initiation codon, the predicted open reading frame
would encode 282 amino acids. In contrast, in the matching glycosyltransferases in GT family 25, translation is initiated with an ATG (Met) initiation codon, located 26 amino acids downstream of the suggested Val translational start of HI0765 and resulting in an open reading frame of 256 amino acids. If Val were to be the initiation codon, the proposed lpsA open reading frames in 11 of the 28 strains would be truncated due to the presence of termination codons. However, if Met is proposed as the initiation codon, the prediction would be that only three strains (NTHi 285, 477, and 981) would have truncated lpsA reading frames. We analyzed the structure of the LPS obtained from NTHi 285,4 477,5 and 981 (18), and, consistent with our prediction of Met as the initiation codon, these three strains are the only three of the 28 where there was no addition of a Hex to HepIII. Based on this evidence, we have assigned Met as amino acid 1 and the Val (the original suggested initiation codon) as −26 (Fig. 2).

The lpsA gene in strain NTHi 285 has a 5-base pair insertion localized to between 5 and 29 base pairs into the reading frame. This predicts that the downstream sequence would be out of frame with the start codon and would result in a truncated open reading frame of only 14 amino acids. The insertion results in three tandem copies of a perfect 5’-ACAGT pentanucleotide repeat immediately followed by two imperfect copies, both of which have one base pair that varies from the repeat motif (Fig. 3).

Considering next the lpsA genes of NTHi strains 477 and 981, each has a 57-base pair deletion located 282 base pairs downstream of the proposed ATG (Met) start. This deletion results in the loss of 19 amino acids, including two of three amino acids that make up the DxD motif of GTs (20–22). Based on crystal structures of several GTs, it has been proposed that this DXD region binds Mn²⁺ and forms the catalytic domain for the donor sugar (23). Thus, the loss of this functional domain and the remaining 17 amino acids would explain the lack of GT activity (no additions to HepIII) in strains 477 and 981.

The only strain that does not have an ATG at positions 1–3 of its lpsA coding sequence is NTHi 723, which has a G to T transversion at the third base pair of the proposed ATG initiation codon. Structural analysis of the LPS of strain 723 has shown that a Glc is added to HepIII, but at relatively low levels (24). Based upon bacterial codon usage tables, ATT (Ile) has been proposed to act as an alternative Met initiation codon (available on the World Wide Web at www.ncbi.nlm.nih.gov/Taxonomy/); this may initiate translation, but relatively inefficiently.

Upstream and downstream of the proposed reading frame there are two regions (−75 to −45 bp upstream of the initiation codon and +2 to +45 bp downstream of the termination codon) of inverted repeats that would form strong stem-loop terminating structures. These repeat regions both contain tandem inverted copies of the Haemophilus 9-bp core uptake sequence (5’-AAGTGCAGG) (25).

4 E. K. Schweda, unpublished results.
5 J. C. Richards, unpublished results.
Functional Allelic Variants of the lpsA Glycosyltransferase

| NTHi 285 | ATGC | ACAGT | ACAGT | ACAGT | ACAGC | ACAAT | TATGTTTATAGTTTTA
|----------|------|-------|-------|-------|-------|-------|------------------|
| NTHi 486 | ATGC | ATAAAT | GCAGC | TCAGC | ACAGC | ACAAT | TATGTTTATACGTTAA

FIGURE 3. Alignment of the DNA sequences from the beginning of the lpsA open reading frame from strains NTHi 285, NTHi 486, and Rd. The 5-bp insertion in the NTHi 285 sequence that results in a truncated translated amino acid sequence is shown. The three perfect and two imperfect repeats are underlined, with the imperfect nucleotide in boldface type.

Alleric Variations of lpsA Are Associated with the Addition of either Glucose or Galactose through Different Anomeric Linkages

The most striking observation from the alignment of the lpsA sequences was that they represent two families. One major group (22 strains) had limited polymorphisms between sequences when compared with the KW20 genome sequence (designated Rd-like; 94–99% identity to Rd). The other six strains resembled the NTHi 486 lpsA gene sequence (designated 486-like; 95% identity to 486); five strains, including NTHi 285, are at least 99% identical to each other. The NTHi 486 sequence has regions of homology to both families; 76% overall identity to Rd lpsA and 95% identity to the other five strains, a mosaic sequence pattern that may reflect interstrain allelic exchange.

In the LPS of strain Rd, Glc is added via a β-1→2 linkage to HepIII (26), but in NTHi 486, Glc is added through a β-1→3 linkage to the same residue (10). Analysis of the detailed LPS structure (detailed below) of strains containing 486-like lpsA sequences showed that each of them added hexose to HepIII via a β-1→3 linkage, whereas those that were Rd-like used a β-1→2 linkage (18, 27, 28).

When the Rd and NTHi 486 nucleotide sequences were compared, the regions were identical at both the 5’ (bp 1→160) and 3’ (bp 748→768) ends. In one small region (bp 161→210) and one larger one (bp 336→747), the sequence homology was only 65%. These regions flanked a more conserved (86%) region of DNA (bp 211→335) (data not shown).

To investigate whether or not the specific sequence of the lpsA gene was sufficient to direct the addition of either a Glc or Gal to the HepIII via a β-1→2 or β-1→3 linkage, we constructed variant strains in which the allelic lpsA gene sequences were exchanged. We used cloned lpsA genes from strains Rd or Eagan (where the hexose is β-1→2-linked) and strain NTHi 486 (where the hexose is β-1→3-linked) as donor DNA to transform each of the respective wild-type strains and thereby obtain the reciprocal linkages through allelic exchange. With the exception of a variant of strain Eagan containing the lpsA gene sequence from strain 486, transformants appropriate to and representative of each of the variant lpsA alleles were obtained and confirmed by PCR amplification and DNA sequencing.

Structural Characterization of LPS from Transformant Strains

Analysis of LPS by T-SDS-PAGE indicated changes in fractionation patterns of the transformants when compared with gel profiles of the parent strains (Fig. 4). The LPS profiles visible on the gels represent complete or partial extensions of HepIII (possibly due to enzyme efficiency) and the remaining LPS structure. We used methylation analysis and NMR experiments to identify the hexose sugar and its linkage to HepIII (Tables 3–5). In 1H NMR spectra, the anomeric proton resonances from the three Hep residues were typically recognized between 4.9 and 6.00 ppm. As described earlier, the anomeric protons of HepI and HepIIl have similar chemical shifts occurring between 4.9 and 5.2 ppm, whereas that from HepIIl occurs further downfield between 5.5 and 5.9 ppm (19). Interresidue NOE connectivities between proton pairs HepII H-1/HepIIl H-1 were used as diagnostic markers to confirm the chemical shift of H-1 of HepIIl. NOE connectivities from the anomeric proton of a particular sugar (Hex) differ significantly on whether the 2- or 3-position of HepIIl is substituted. A cross-peak between HepII H-1/HepIIl H-1 is strongly indicative of 2-substitution. In the case of 3-substitution, this cross-peak is absent, but NOE connectivities between Hex H-1/HepIIl H-2,H-3 are observed (Fig. 5, Table 4). Every strain was investigated thoroughly and included the use of ESI-MS® on permethylated OS to confirm sequences and branching patterns (18). However, here results focusing on the HepIIl region of the LPS are presented, since they are the ones that differ significantly from those of the parent strains and confirm the proposed specificity of lpsA in the respective strains.

Structural Analyses of LPS from Strains with Exchanged lpsA Genes

O-2 to O-3 Exchange—In strain Rd, HepIIl is substituted at O-2 with globotetraose (β-D-GalNAc-(1→3)-α-D-Galp-(1→4)-β-D-Galp-(1→4)-β-D-Glcp) and truncated versions thereof (26). HepI is substituted at the O-4 position with PCho-(6)-β-D-Glcp. Methylation analysis of dephosphorylated OS derived from the transformant Rd lpsA486 (strain Rd with an exchange in the lpsA gene to that of strain 486) revealed, inter alia, significant amounts of 3-substituted Hep (see Tables 3 and 5) that are not observed in the wild-type strain. This indicated HepIIl to be substituted at O-3. The anomeric proton corresponding to a →4) β-D-Glcp residue was observed at δ 4.74 in the 1H NMR spectrum. This proton showed NOE connectivities to H-2 and H-3 of HepIIl (δ 4.28 and 4.14, respectively; Fig. 6B) in agreement with 3-substitution (10).

O-3 to O-2 Exchange—In NTHi strain 486, HepIIl is substituted at O-3 with sialyllactose [α-Neu5Ac-(2→3)-β-D-Galp-(1→4)-β-D-Glc] and truncated versions thereof (10). HepI is


Functional Allelic Variants of the \textit{lpsA} Glycosyltransferase

substituted at O-4 with a $\beta$-\textit{d}-Glc\text{}p residue, and HepII is substituted at O-3 with P\textit{Cho}→(6)-$\alpha$-\textit{d}-Glc\text{}p. Methylation analysis of dephosphorylated OS derived from 486lpsARd (strain 486 with an exchange in the \textit{lpsA} gene to that of strain Rd) showed, \textit{inter alia}, significant amounts of terminal Hep (see Tables 3 and 5) and no 3-substituted Hep, thus indicating that HepIII is terminal. This was confirmed by ESI-MS and ESI-MS$^\ddagger$ analyses on LPS-OH and permethylated OS, respectively (data not shown). In the ESI-MS spectrum of LPS-OH (negative mode), major triply charged ions at \textit{m}/\textit{z} 812.2 and 853.3 together with their quadruply and pentuply charged counterparts (\textit{m}/\textit{z} 609.6/487.2 and 640.2/511.8, respectively) corresponded to glycoforms with respective compositions P\textit{ChoHex}_3\text{Hep}_3\text{P\textit{Etn}}_{1,2}\text{PKdolipidA-OH} (where \textit{P\textit{Etn}} represents phosphoethanolamine and Kdo represents 2-keto-3-deoxyoctulosonic acid). ESI-MS$^\ddagger$ on dephosphorylated and permethylated OS allowed the detection of minor amounts of Hex3 and Hex4 glycoforms. However, sequence analyses by MS$^\ddagger$ on their corresponding ions revealed only trace amounts of isomeric glycoforms where HepIII was substituted. NMR signals due to the hexose linked to HepIII thus could not be observed due to low abundance, and the proposed change of the attachment site from O-3 to O-2 was not confirmed.

\textit{O-2 Glc to O-2 Gal Exchange}—Preliminary structural data on LPS from \textit{NTHi} strain 176 have shown HepIII to be substituted by $\beta$-\textit{d}-Galp-(1→ at O-2.$^4$ Methylation analysis of dephosphorylated OS derived from Rd\textit{lpsA}176 (strain Rd with an exchange in the \textit{lpsA} gene to that of strain 176) revealed, \textit{inter alia}, significant amounts of terminal Gal (see Tables 3 and 5). Only traces of 4-substituted Gal, 4-substituted Glc, and 3-substituted Gal were observed, indicating that the terminal Gal is substituted directly to HepIII. The anomeric proton corresponding to the $\beta$-\textit{d}-Galp residue was observed at $\delta$ 4.60 in the $^1$H NMR spectra (Fig. 6A). This proton showed NOE connectivities to H-1 and H-2 of HepII (\textit{\delta} 5.11 and 4.25, respectively), in agreement with there being 2-substitution (19).

\textit{O-3 Glc to O-2 Gal Exchange}—Another strain having a terminal Gal linked to O-2 of HepIII is \textit{Hi} type b strain Eagan (29). Methylation analysis of dephosphorylated OS derived from 486lpsAEa (strain 486 with an exchange in the \textit{lpsA} gene to that of strain Eagan) revealed, \textit{inter alia}, significant amounts of terminal Gal (see Tables 3 and 5). In the $^1$H NMR spectrum, the anomeric proton corresponding to a $\beta$-\textit{d}-Galp residue was observed at $\delta$ 4.36. This proton showed NOE connectivities to H-1 and H-2 of HepII (\textit{\delta} 5.12 and 4.01, respectively; Fig. 5A), in agreement with 2-substitution. It is noteworthy that \textit{NTHi} strain 486 could be transformed to have HepIII O-2 substituted with Gal but not with Glc.

\textbf{Structural Analyses of LPS from Strains with \textit{lpsA} Genes Altered at Amino Acid Position 151 Changing Substitution from Gal to Glc or Vice Versa}

In previous work by this group (9), where heterologous \textit{lpsA} genes were transferred between \textit{Hi} strains with $\beta_1$–2-linked hexoses to HepIII, it was demonstrated that in strain Rd, the specific \textit{lpsA} sequence from Eagan or Rd directed the incorporation of a Gal or Glc, respectively, to HepIII. The \textit{lpsA} sequences from \textit{Hi} strains investigated here were compared...
with sequences obtained for glycosyltransferases across a wide range of bacterial and animal species (data not shown). It was noted that the presence of the amino acid threonine at an equivalent position (amino acid 151) in sequences where the gene function was known and not simply predicted was invariably associated with the addition of a Gal, regardless of whether the linkage was \( \beta1-2 \) or \( \beta1-3 \). Glc was added when a cysteine, alanine, or methionine was present at the same position.

The translated \( \text{lpsA} \) sequence from two of the \( \beta1-2 \)-linked strains, Rd and NTHi 176, differ from each other at only three amino acids but direct the attachment of Glc or Gal, respectively. One of these three differences was at position 151. To determine if the amino acid at this position was indeed crucial in directing the addition of either a Gal or Glc, we altered this residue in the Rd and NTHi 176 \( \text{lpsA} \) sequences by site-directed mutagenesis.

The cloned \( \text{lpsA} \) gene from the two strains was mutated such that Thr\(^{151} \) in the \( \text{lpsA} \) sequence in NTHi 176 was changed to a cysteine (Cys\(^{151} \)), and conversely the Cys\(^{151} \) in the Rd \( \text{lpsA} \) gene was changed to Thr\(^{151} \). Each altered \( \text{lpsA} \) gene (T\(_{151}\)C176 or C\(_{151}\)T176) was transformed into both the homologous and heterologous strain and also into strain Eagan to replace the existing \( \text{lpsA} \) gene by reciprocal exchange. The LPS from each strain was then investigated for structural changes as described above.

### Table 4

| Sample         | HepII | HepIII | Glc H1 | Gal H1 | Interresidue NOE from HepII, H-1 to | Interresidue NOE from HepIII, H-1 to |
|----------------|-------|--------|--------|--------|-------------------------------------|-------------------------------------|
| RdpsAC151TRd   | H1    | H2     | H1     | H2     | H3                                  | H3                                  |
| Rd's RD3486    | 5.80  | 4.28   | 5.23   | 4.28   | 4.14                                | 4.74                                |
| RdpsA176       | 5.64  | 4.33   | 5.11   | 4.25   | 3.96                                | 4.60                                |
| RdpsAC151TRd   | 5.64  | 4.33   | 5.11   | 4.25   | 3.96                                | 4.60                                |
| Rd's 486/3C176 | 5.64  | 4.35   | 5.12   | 4.25   | 3.92                                | 4.69                                |
| RdpsAT151C176  | 5.64  | 4.35   | 5.12   | 4.25   | 3.92                                | 4.69                                |
| 4806psARd      | 5.83  | 4.28   | 5.03   | 4.00   | 3.79                                | 4.36                                |
| 4865/486/3Rd   | 5.83  | 4.28   | 5.03   | 4.00   | 3.79                                | 4.36                                |
| 4806psAEa      | 5.72  | 4.27   | 5.12   | 4.01   | 3.77                                | 4.36                                |
| 4865/486/3C176 | 5.72  | 4.27   | 5.12   | 4.01   | 3.77                                | 4.36                                |
| EaganlpsAT151C176 | 5.70 | 4.30   | 5.13   | 4.04   | 3.73                                | 4.42                                |
| EaganlpsAC151TRd | 5.71 | 4.32   | 5.13   | 4.04   | 3.81                                | 4.43                                |

- Glc to Gal Exchange—Structural analyses of RdpsAC151TRd (strain Rd with an altered \( \text{lpsA} \) gene C\(_{151}\)T from Rd) gave virtually identical results as obtained for RdpsA176 (see above; Tables 3–5, Fig. 6A), evidence that HepIII was substituted by \( \beta\text{D-Galp} \) at the O-2 position.

- Glc to Glc Exchange—The results for RdpsAT151C176 (strain Rd with an altered \( \text{lpsA} \) gene T\(_{151}\)C from 176) were similar to those of the wild-type strain with a predominating Hex3 glycoform. The anomeric proton corresponding to a \( \rightarrow-4\)\( \beta\text{-D-Galp} \) residue was observed at \( \delta \) 4.69 in the \( ^1\text{H} \) NMR spectrum. This proton showed NOE connectivities to H-1 and H-2 of HepIII (\( \delta \) 5.12 and 4.25, respectively) in agreement with 2-substitution.

- Gal to Glc Exchange—In Hi type b strain Eagan, HepIII is substituted at O-2 with a \( \beta\text{-D-Galp} \) residue. Hep II is substituted at O-4 with a \( \beta\text{-D-Glc} \) residue, and Hep II is substituted at O-3 with \( \alpha\text{-D-Galp}-(1\rightarrow4)\text{-D-Galp}-(1\rightarrow4)\text{-D-Glc}-(1\rightarrow4)\text{-D-Glc} \) and truncated versions thereof (29). Methylation analysis of dephosphorylated OS derived from EaganlpsAT151C176 (strain Eagan with an altered \( \text{lpsA} \) gene T151C from 176) revealed, \( \text{inter alia} \), only minor amounts of terminal Gal (see Tables 3 and 5). This indicated that HepIII was substituted by a \( \beta\text{-D-Glc} \) residue and not by \( \beta\text{-D-Galp} \). The anomeric proton corresponding to a \( \beta\text{-D-Glc} \) residue was observed at \( \delta \) 4.42 in the \( ^1\text{H} \) NMR spectrum. This proton showed NOE connectivities to H-1 and H-2 of HepIII (\( \delta \) 5.13 and 4.04, respectively) in agreement with 2-substitution.

- Gal to Gal Exchange—The results from structural analyses of LPS from EaganlpsAC151TRd (strain Eagan with an altered \( \text{lpsA} \) gene C151T from Rd) were virtually identical to those of the wild-type strain. Methylation analysis and NOE data are given in Tables 3 and 4, respectively. We found that \( \beta\text{-D-Galp} \) was attached to the O-2 position of HepIII; since NOE connectivities between the anomeric proton of a \( \beta\text{-D-Galp} \) residue at \( \delta \) 4.38 and H-1 and H-2 of HepIII (\( \delta \) 5.11 and 4.04, respectively) were observed.

Together, the findings show that indeed a single amino acid, that at position 151 in the \( \text{lpsA} \) gene, is responsible for specifically directing the addition of either a Glc or a Gal to HepIII in Hi LPS. Interestingly, when the initial allelic exchanges of \( \text{lpsA} \) were performed between strains Rd and Eagan (9), it was dem-
onstrated that the Rd gene when transformed into Eagan was less able to incorporate a hexose, in this case a Glc, than the native gene. Mutation of this Rd gene to encode for the addition of Gal, C151TRd gene sequence, appears to now allow equally efficient hexose incorporation into Eagan as the native Eagan lpsA gene. Furthermore, when Glc is being expressed in Eagan LPS from the T151C176 allelic replacement, the Glc is added efficiently. The three amino acid differences between the Rd and NTi 176 sequences are at amino acid position 25, where Rd has a glutamic acid (Glu25) and NTi 176 has an aspartic acid (Asp25), the known change at position 151, plus a change at position 218 where Rd has a glutamine (Gln218) but NTHi 176 has a lysine (Lys218). Eagan also has the Glu25, but only Rd, NTHi 1209, and NTHi 1233 and the truncated NTHi 477 and NTHi 981 have a Gln218. Gln has a polar but uncharged R group, whereas Lys has a positively charged R group; this may be an explanation for the reduced activity of the wild-type Rd-derived protein in strain Eagan.

Structural Analyses of LPS from Strains with Exchanged Chimeric lpsA Genes

The specificity for the addition of either Glc or Gal in the sequence of the glycosyltransferase correlates to a single amino acid change; however, the sequence differences between β1–2
and β1–3 linkage-related LpsA enzymes is much greater. In an initial investigation to determine whether the 5′ or 3′ block of divergent sequence or both direct the specificity of linkage, we constructed chimeric genes. Due to a lack of suitable restriction endonuclease digestion sites for chimeric gene construction, enabling the two regions of most divergent lpsA sequence to be mixed, a unique AflII site was introduced into the area of the gene with 85% homology between the β1–2- and β1–3-specific forms. A single nucleotide was altered by site-directed mutagenesis without changing the encoded amino acid.
Plasmids containing chimeric lpsA genes, comprising the 5′ portion of one strain and the 3′ portion of a second, were constructed to give plasmids p5′Rd/3′486, p5′486/3′Rd, and p5′486/3′C151Rd. Once transformed into Hi strains Rd, Eagan, and NTHi 486, LPS from the resulting allelic replacement transformants was investigated as described above (see Fig. 4).

O-2 Glc to O-2 Gal Chimeric Exchange—Structural analyses of LPS from strain Rd5′486/3′C151Rd gave similar results as obtained for Rd/lpsA176 and Rd/lpsAC151Rd (see Tables 3–5) and demonstrated a β-d-Galp residue linked to O-2 of HepIII. 

O-2 Glc to O-3 Glc Chimeric Exchange—Structural analyses of Rd5′Rd/3′486 gave similar results as obtained for Rd/lpsA486. Fig. 6B shows the TOCSY spectrum of its oligosaccharide, which was virtually identical with that from Rd/lpsA486. The results provided evidence that HepIII was substituted by β-d-Glc at the O-3 position.

O-3 Glc to O-2 Glc Chimeric Exchange—Structural analyses of LPS from strain 4865/486/3′Rd gave similar results as obtained for 486/lpsARd. Even in this strain, substitution of HepIII was too minor, and no evidence for a Glc residue substituting O-2 Glc to O-2 Gal Chimeric Exchange—Structural analyses of Rd5′Rd/3′486 gave similar results as obtained for Rd/lpsA486. Fig. 6C shows the TOCSY spectrum of its oligosaccharide. Methylation analysis and NOE data indicated that HepIII was substituted by β-d-Galp at the O-2 position. Together, the findings show that the 3′ portion of the gene is primarily responsible for determining the linkage of the hexose to HepIII.

DISCUSSION

Based on sequence similarity, the LpsA enzyme of Hi belongs to the GT-A superfamily of glycosyltransferases. Three superfamilies, GT-A, GT-B, and GT-C, each contain 75% of known GTs, and the remainder lack the structural similarity to be included in the three major classification groups. The essential characteristics of a GT-A superfamily glycosyltransferase are (i) a nucleotide diphosphosugar transferase that transfers an activated monosaccharide residue from UDP-Gal or UDP-Glc to an acceptor molecule (in this instance the LPS molecule) and (ii) possession of a DxD motif, a conserved amino acid sequence that binds Mn2+ and forms the catalytic domain for the donor sugar (20–22). The transferases are assigned to the superfamilies based upon the folding structures of the protein, with the GT-A and GT-B superfamilies both having Rossman-like folds (30). The Hi LpsA enzyme uses an inverting catalytic mechanism in which the α-linked nucleotide diphosphosugar is the donor substrate from which the sugar is transferred to the LPS core to form a β-linked product, and a single nucleophilic substitution is sufficient to produce this β-configuration (31).

Based on genetic and structural analysis of LPS from 28 strains, the major finding we report here is that Hi possesses at least four GT-A allelic enzyme variants of LpsA, each of which has different specificities. Each allele determines the addition of either Glc or Gal and whether the anomeric linkage is β1–2 or β1–3. The strains used for this analysis were selected because we have shown previously that they are representative of the genetic diversity of more than 400 Hi strains, a collection that includes both encapsulated and capsule-deficient (nontypeable) isolates collected over a 35-year period from diverse geographic regions (13). Inspection of the dendrogram of these Hi isolates shows that the addition of either Gal or Glc to the third heptose (HepIII) of the conserved core LPS backbone falls almost perfectly into two clades (data not shown), the exceptions being NTHi strain 162 and the two closely related type b strains (Eagan and RM7004). Because recombination is common in natural populations of Hi, the phylogenetic signals observed from inspection of the dendrogram (based on ribotyping) are relatively weak, so that the observed clustering suggests that the GT activities specifying the addition of one of two hexoses probably resulted from the relatively recent acquisition of a point mutation. In contrast, both the 21 strains in which the hexose is linked β1–2 and the seven strains in which the hexose is β1–3 linked are scattered across the dendrogram. This suggests that the evolution of sequences specifying the two distinct anomeric linkages occurred earlier in the life history of the species, and there has therefore been sufficient recombination to obscure clustering of strains possessing the distinct β1–2 and β1–3 alleles. The location of a DNA uptake sequence (32), that increases the efficiency of acquiring DNA, immediately adjacent to each end of the LpsA gene may have facilitated interstrain recombination between allelic variants, thus giving rise to the mosaic sequence observed in strain 486.

The amino acid (position 151) that directs this hexose addition, located in the 3′ region of the gene, is the second of a run of five amino acids, four of which are conserved in a region of DNA that has only 65% identity (corresponding to nucleotides 336–747). Comparison of our LpsA sequences with other glycosyltransferases of known function in the public data bases shows an equivalent amino acid to the Thr at position 151 in glycosyltransferases that are specific for Gal addition across a range of species (including E. coli, Neisseria meningitidis, Helicobacter pylori, Salmonella typhimurium, and Haemophilus ducreyi). Furthermore, an example of a single amino acid directing sugar specificity is seen in some of the glycosyltransferases involved in plant secondary metabolism. Kubo et al. (33) observed that the last amino acid residue in a glycosyltransferase-specific conserved region showed specificity to the hexose added to the substrate, such that a glutamine was always associated with the glycosyltransferases, whereas a histidine was specific for the galactosyltransferases.

The specific sequence(s) that directs the addition of the hexose via a β1–2 or β1–3 linkage has not been defined. The lack of proven function of a majority of GTs generally precludes the simple assignment of linkage types to the GTs in the data base by simple sequence analysis, although prediction is now possible within Hi. However, despite not knowing the precise location within the 3′ sequence of the amino acid(s) that specifies the anomeric linkage, both the sugar substitution and linkage type can be accurately predicted from the sequence alone. There is therefore no requirement for detailed structural analysis of the LPS, a testimony to the power of establishing good gene sequence/structure/function relationships.

Thus, given that any Hi strain possesses one of four alleles that
Functional Allelic Variants of the lpsA Glycosyltransferase

direct the addition of either Glc or Gal in either a β1–2 or β1–3 anomic linkage, and based on the CAZy glycosyltransferase nomenclature (6) (Carbohydrate-Active Enzymes data base available on the World Wide Web at www.cazy.org/), we propose the following: that the designation of the allelic variant protein originally identified in strain Rd be LpsA, but that the other alleles be called LpsB, directing an addition of Gal in a β1–2 linkage; LpsC, directing an addition of Glc in a β1–3 linkage; and LpsD, directing an addition of Gal in a β1–3 linkage, respectively. The genes should correspondingly be named lpsA, lpsB, lpsC, and lpsD.

One of the remarkable features of Hi is the extraordinary diversity of its LPS glycoforms. Both intrastrain (phenotypic variation arising in progeny derived from a founding bacterial cell) and interstrain LPS variation (the individual differences between the glycoforms characteristic of distinct strains) have been well documented. Intrastrain diversity is largely the result of the phase variation of particular genes, those involved in the biosynthesis of core sugars and also nonsugar modifications, such as phosphorylcholine (34) and acylation (35). Interastrain variation is known to reflect in part the presence or absence of particular biosynthetic genes, for example lex2 (36) and lic2 (37). In addition, high molecular weight glycoforms are a characteristic of some but not all Hi strains (38). Allelic variation, now well documented here for LpsA, is yet another important mechanism contributing to the heterogeneity of LPS glycoforms observed in different strains. A further example of allelic variation in Hi, associated with differences in biosynthetic functions, is afforded by the distinct sialyltransferases Lic3A and Lic3B,6 which share 91% amino acid identity. However, in this latter example of allelic variation, each of these genes can be found in the same strain, whereas for LpsA, only one variant is found in a given strain.

There is at least one other example of single amino acid substitutions directing the specificity of a protein’s function in LPS biosynthesis. In N. meningitidis, a sialyltransferase, Lst, normally adds sialic acid in an α2-3 linkage to a galactose in one strain, whereas in another strain it adds the sialic acid through an α2-6 linkage. In the α2-3-linked strain, Lst has a glycine at amino acid position 168, whereas the α2-6-linked Lst has an isoleucine. The α2-6-linked enzyme can make both linkages in vitro using a synthetic acceptor, whereas the α2-3-linked enzyme only uses the α2-3 linkage. Surprisingly, the α2-3-linked enzyme is bifunctional in vivo, but the α2-6 linkage is seen at much lower levels than the α2-3 linkage (39).

The variations in the LpsA to LpsD alleles of Hi have functional consequences that go beyond the specificity of the hexose and the anomeric linkage. In the case of lpsA, the addition of Gal as compared with Glc to Hep III apparently precludes the addition of further sugars (11). As a consequence, there is an absence of the lactose acceptor that is required for the addition of α2-3-linked sialic acid. The importance of this is that terminal sialic acid on the LPS confers relative serum resistance in vitro (12) and enhanced virulence in vivo (40). In general, sialic acid is also thought to contribute to structures that mimic human epitopes (41), and its presence can down-regulate the host-microbial interactions between other bacterial cell surface structures (e.g. outer membrane proteins) and host receptors. Thus, allelic variations in LpsA to LpsD GTs are important in determining the commensal and virulence behavior of Hi. Nor should the variations in LpsA to LpsD be viewed in isolation. Analogous to the Lps alleles, lic1D genes from different strains of Hi also contain allelic variations where their exchange between different strains (e.g. Rd and Eagan) results in the addition of phosphorylcholine to different locations on the LPS molecule. One allelic variant in Lic1D resulted in the addition of phosphorylcholine to hexose off Hep1, whereas another variant resulted in its addition to the chain-terminating Gal attached to HepII (42). The presence of phosphorylcholine in different molecular environments was associated with changes in serum resistance in vitro (42). It can be seen that the combinations of lic1D and specific lpsA genes might have a huge influence on the survival or dissemination of a strain of Hi in different environments in the host.

This study elucidates unique gene/function relationship for glycosyltransferases, whereby variation of sequence at the same genetic location in different strains directs the addition of one of four sugar/linkage combinations in LPS. At least part of the sequence is conserved in each gene, and evidence suggests that recombination has occurred probably both at the inter- and intraspecies level. This novel mechanism of increasing the structural heterogeneity underlies the considerable effort that this bacterium has made to maximize glycoform diversity of LPS within the species.

Acknowledgments—We thank Juhani Eskola and the Otitis Media Study Group (National Public Health Institute, Finland) for the H. influenzae OM isolates. We acknowledge the technical expertise of Kate Fox, Gaynor Jenkins, and Kathy Makepeace.

REFERENCES

1. Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996) Crit. Rev. Microbiol. 22, 139–180
2. Schweda, E. K., Hegedus, O. E., Borrelli, S., Lindberg, A. A., Weiser, J. N., Maskell, D. J., and Moxon, E. R. (1993) Carbohydr. Res. 246, 319–330
3. Amado, M., Almeida, R., Schwientek, T., and Clausen, H. (1999) Biochim. Biophys. Acta 1473, 35–53
4. Lazarus, B. D., Milland, J., Ramsland, P. A., Moultouis, E., and Sandrin, M. S. (2002) Glycobiology 12, 793–802
5. Campbell, J. A., Davies, G. J., Bulone, V. V., and Henrissat, B. (1997) Biochem. J. 326, 929–939
6. Coutinho, P. M., and Henrissat, B. (1999) in Recent Advances in Carbohydrate Bioengineering (Gilbert, H. J., Davies, G., Henrissat, B., Svensson, B., eds) pp. 3–12, The Royal Society of Chemistry, Cambridge, UK
7. 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) Mol. Microbiol. 22, 951–965
8. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) J. Mol. Biol. 328, 307–317
9. Hood, D. W., Deadman, M. E., Cox, A. D., Makepeace, K., Martin, A., Richards, J. C., and Moxon, E. R. (2004) Microbiology 150, 2089–2097
10. Månsson, M., Bauer, S. H., Hood, D. W., Richards, J. C., Moxon, E. R., and Schweda, E. K. (2001) Eur. J. Biochem. 268, 2148–2159
11. Schweda, E. K., Li, J., Moxon, E. R., and Richards, J. C. (2002) Carbohydr. Res. 337, 409–420
12. Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C., and Moxon, E. R. (1999) Mol. Microbiol. 33, 6

6 K. L. Fox, A. D. Cox, M. Gilbert, W. W. Wakarchuk, J. Li, J. C. Richards, E. R. Moxon, D. W. Hood, submitted for publication.
679–692
13. Cody, A. J., Field, D., Feil, E. J., Stringer, S., Deadman, M. E., Tsolaki, A. G., Gratz, B., Bouhet, V., Goldstein, R., Hood, D. W., and Moxon, E. R. (2003) Infect. Genet. Evol. 3, 57–66
14. Alexander, H. E. (1965) Bacterial and Fungal Infections of Man (Dubos, R. H., and Hirsch, J. G., eds) pp. 724–741, Pitman Medical Publishing, London
15. Sambor, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
16. 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., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritsch, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) Science 269, 71–83
17. Herriot, R. M., Meyer, E. M., and Vogt, M. J. (1970) J. Bacteriol. 101, 517–524
18. Månsson, M., Hood, D. W., Moxon, E. R., and Schweda, E. K. (2003) Eur. J. Biochem. 270, 2979–2991
19. Schweda, E. K., and Richards, J. C. (2003) Methods Mol. Med. 71, 161–183
20. Wiggins, C. A., and Munro, S. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 7945–7950
21. Shibayama, K., Ohsuka, S., Tanaka, T., Arakawa, Y., and Ohta, M. (1998) J. Bacteriol. 180, 5313–5318
22. Tarbouriech, N., Charnock, S. J., and Davies, G. J. (2001) J. Mol. Biol. 314, 655–661
23. Gastinel, L. N., Cambillau, C., and Bourne, Y. (1999) EMBO J. 18, 3546–3557
24. Yildirim, H. H., Li, J., Richards, J. C., Hood, D. W., Moxon, E. R., and Schweda, E. K. (2005) Carbohydr. Res. 340, 2598–2611
25. Kroll, J. S., Wilks, K. E., Farrant, J. L., and Langford, P. R. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 12381–12385
26. Risberg, A., Masoud, H., Martin, A., Richards, J. C., Moxon, E. R., and Schweda, E. K. (1999) Eur. J. Biochem. 266, 171–180
27. Månsson, M., Hood, D. W., Li, J., Richards, J. C., Moxon, E. R., and Schweda, E. K. (2002) Eur. J. Biochem. 269, 808–818
28. Schweda, E. K., Landerholm, M. K., Li, J., Moxon, E. R., and Richards, J. C. (2003) Carbohydr. Res. 338, 2731–2744
29. Masoud, H., Moxon, E. R., Martin, A., Krajcarski, D., and Richards, J. C. (1997) Biochemistry 36, 2091–2103
30. Liu, J., and Mushegian, A. (2003) Protein Sci. 12, 1418–1431
31. Saxena, I. M., Brown, R. M., Jr., Fevre, M., Geremia, R. A., and Henrissat, B. (1995) J. Bacteriol. 177, 1419–1424
32. Smith, H. O., Gwinn, M. L., and Salzberg, S. L. (1999) Res. Microbiol. 150, 605–616
33. Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004) Arch. Biochem. Biophys. 429, 198–203
34. Weiser, J. N., Shchepetov, M., and Chong, S. T. (1997) Infect. Immun. 65, 945–950
35. Fox, K. L., Yildirim, H. H., Deadman, M. E., Schweda, E. K., Moxon, E. R., and Hood, D. W. (2005) Mol. Microbiol. 58, 207–216
36. Griffin, R., Cox, A. D., Makepeace, K., Richards, J. C, Moxon, E. R., and Hood, D. W. (2003) Microbiology 149, 3165–3175
37. Richards, J. C., Cox, A. D., Schweda, E. K., Martin, A., Hood, D. W., and Moxon, E. R. (2001) Adv. Exp. Med. Biol. 491, 515–524
38. Hood, D. W., Randle, G., Cox, A. D., Makepeace, K., Li, J., Schweda, E. K., Richards, J. C., and Moxon, E. R. (2004) J. Bacteriol. 186, 7429–7439
39. Warkaas, W. W., Watson, D., St. Michael, F., Li, J., Wu, Y., Briss, J. R., Young, N. M., and Gilbert, M. (2001) J. Biol. Chem. 276, 12785–12790
40. Bouchet, V., Hood, D. W., Li, J., Briss, 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) Proc. Natl. Acad. Sci. U. S. A. 100, 8898–8903
41. Morin, A., Prendergast, M. M., and Appelmelk, B. J. (1996) FEMS Immunol. Med. Microbiol. 16, 105–115
42. Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M., and Weiser, J. N. (2000) Mol. Microbiol. 35, 234–245
43. Anderson, P., Johnston, R. B., Jr., and Smith, D. H. (1972) J. Clin. Invest. 51, 31–38
44. Alexander, H. E., and Leidy, G. (1951) J. Infect. Genet. Evol. 148, 71–81

Functional Allelic Variants of the IpaA Glycosyltransferase