Structural and genetic evidence that the *Escherichia coli* O148 O antigen is the precursor of the *Shigella dysenteriae* type 1 O antigen and identification of a glucosyltransferase gene

Lu Feng,1,2† Andrei V. Perepelov,3† Guang Zhao,1,2† Sergei D. Shevelev,3 Quan Wang,1,2 Sof’ya N. Senchenkova,3 Alexander S. Shashkov,3 Yunqi Geng,4 Peter R. Reeves,5 Yuriy A. Knirel3 and Lei Wang1,2

1TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China
2Tianjin Key Laboratory for Microbial Functional Genomics, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China
3N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation
4College of Life Sciences, Nankai University, Tianjin 300071, P. R. China
5School of Molecular and Microbial Biosciences (G08), University of Sydney, Sydney, NSW 2006, Australia

Correspondence Lei Wang wangleii@nankai.edu.cn

Received 10 August 2006
Revised 11 September 2006
Accepted 14 September 2006

**INTRODUCTION**

*Escherichia coli* serotypes are normally classified by a combination of their O and H (and sometimes K) antigens. For shigellae, only the O antigen classification system is used, as they lack H and K antigens. Based on multilocus enzyme electrophoresis (MLEE) and analysis of housekeeping gene sequences, it was shown that *Shigella* species in reality are *E. coli* serotypes. At least 166 O antigen types have been recognized in *E. coli* and 33 in *Shigella*. Thirteen of them were common to both based on cross-reactions summarized by Ewing (1986) and chemical structural data (Parolis et al., 1997), which also indicates the close relationship between *Shigella* and *E. coli*. It has been proposed that *Shigella* originated from *E. coli* in the Neolithic about 10 000 years ago when agriculture had been developed (Ochman et al., 1983; Pupo et al., 2000). The O antigen forms unique to *Shigella* were proposed to have been obtained since the *Shigella* mode of pathogenicity arose in *E. coli* (Wang et al., 2001b).

The O antigen, as part of the LPS in the outer membrane of Gram-negative bacteria, is a major target of both the immune system and bacteriophages, and plays an important role in the virulence of bacterial pathogens. For *Shigella dysenteriae* type 1 (O1) and *E. coli* O148, the O antigen structure is a tetrasaccharide repeating unit containing a galactosyltransferase gene. The O antigen of *S. dysenteriae* type 1 is a precursor of the O148 O antigen. The structure established for the tetrasaccharide repeating unit of the *E. coli* O148 O antigen is different from the structure reported earlier for *S. dysenteriae* type 1 by having a glucose (Glc) residue in place of a galactose (Gal) residue. The two bacteria also have the same genes for O antigen synthesis, with the same organization and high level of DNA identity, except that in *S. dysenteriae* type 1 *wbbG* is interrupted by a deletion, and a galactosyltransferase gene *wbbP* located on a plasmid is responsible for the transfer of galactose to make a novel antigenic epitope of the O antigen. The *S. dysenteriae* type 1 O antigen was reconstructed by replacing the *E. coli* O148 *wbbG* gene with the *wbbP* gene, and it had the LPS structure and antigenic properties of *S. dysenteriae* type 1, indicating that the *S. dysenteriae* type 1 O antigen evolved from that of *E. coli* O148. It was also confirmed that *wbbG* of *E. coli* O148 is a glucosyltransferase gene, and two serotype-specific genes of *E. coli* O148 and *S. dysenteriae* type 1 were identified.

*Shigella dysenteriae* type 1 is the most virulent serotype of *Shigella*. Enterotoxigenic *Escherichia coli* O148 is pathogenic and can cause diarrhea. The following structure was established for the tetrasaccharide repeating unit of the *E. coli* O148 O antigen: \(\beta\)-D-Galp-(1→3)-\(\alpha\)-D-GlcNAc-(1→2)-\(\alpha\)-L-Rhap-(1→3)-\(\beta\)-L-Rhap-(1→2)-\(\beta\)-D-Glc-(1→3)-\(\beta\)-D-GlcNAP-(1→). This differs from the structure reported earlier for *S. dysenteriae* type 1 by having a glucose (Glc) residue in place of a galactose (Gal) residue. The two bacteria also have the same genes for O antigen synthesis, with the same organization and high level of DNA identity, except that in *S. dysenteriae* type 1 *wbbG* is interrupted by a deletion, and a galactosyltransferase gene *wbbP* located on a plasmid is responsible for the transfer of galactose to make a novel antigenic epitope of the O antigen. The *S. dysenteriae* type 1 O antigen was reconstructed by replacing the *E. coli* O148 *wbbG* gene with the *wbbP* gene, and it had the LPS structure and antigenic properties of *S. dysenteriae* type 1, indicating that the *S. dysenteriae* type 1 O antigen evolved from that of *E. coli* O148. It was also confirmed that *wbbG* of *E. coli* O148 is a glucosyltransferase gene, and two serotype-specific genes of *E. coli* O148 and *S. dysenteriae* type 1 were identified.
role in the bacterium–host interplay. It is one of the most variable cell constituents and also plays an important role in virulence. Genes for O antigen synthesis are normally clustered between galF and gnd in the chromosome of E. coli.

Of Shigella serotypes, S. dysenteriae type 1 was the first described, and it attracts special attention for frequency of epidemics, the severity of symptoms, high attack rate, high case-fatality rate, and various complications (Bennish et al., 1990). It was responsible for large dysentery epidemics in Guatemala and other parts of Central America (Mendizabal-Morris et al., 1971), Bangladesh (Ronsmans et al., 1988), Zaire (Group, 1995), Kenya (Iijima et al., 1995), and recently West Africa (Guerin et al., 2003) and India (Pazhani et al., 2004). The O antigen of S. dysenteriae type 1 is essential for virulence, and there is indirect evidence that antibodies against the O antigen are protective (Passwell et al., 2001; Robbins et al., 1992; Taylor et al., 1993). The O antigen consists of tetrasaccharide repeating units of the following structure: →3)-α-L-Rhap(→3)-α-L-Rhap(→3)-α-D-Galp(→3)-α-D-GlcPβNAc(→(Dmitriev et al., 1976). The enzymes for the biosynthesis of O antigen are encoded by an O antigen gene cluster and the wbbP gene (previously known as rfpB), located on the chromosome near the his locus and a 9 kb plasmid pHW400, respectively (Gohmann et al., 1994).

The wbbP gene was shown to encode a membrane-bound galactosyltransferase, adding the second sugar of the structure shown above. A number of E. coli and Shigella O antigen gene clusters have been cloned and characterized (Reeves & Wang, 2002), with Shigella sonnei the only other example with genes essential for O antigen synthesis being on a plasmid (Shepherd et al., 2000). The presence of this essential gene on a plasmid indicated that the galactose component of the S. dysenteriae type 1 O antigen was gained recently.

E. coli O148::H28 is one of the most common causes of diarrhoea in children in developing countries as well as in travellers to these areas (Black, 1986). In this study, the E. coli O148 O antigen repeating unit was shown to differ from that of S. dysenteriae type 1 only by the presence of a glucose residue in place of the galactose residue. The genes and organization of the O antigen gene cluster were shown to be the same in both bacteria, except that wbbG is interrupted by a deletion in S. dysenteriae type 1, which together with the presence in S. dysenteriae type 1 of wbbP accounts for the difference in structure. After deleting wbbG in E. coli O148 and cloning the S. dysenteriae type 1 wbbP gene into the O antigen deficient mutant strain, the construct strain was shown to have the same O antigen structure as S. dysenteriae type 1. Therefore, wbbG was shown to be responsible for transferring glucose to N-acetylglucosamine (GlcnAc) in E. coli O148 and it also appears that S. dysenteriae type 1 gained its current O antigen gene cluster from E. coli O148. Genes specific to E. coli O148 and S. dysenteriae type 1 were identified by screening strains representing 186 E. coli serotypes including Shigella.

**METHODS**

**Bacterial strains and growth conditions.** All strains were grown on liquid or solid Luria–Bertani medium. The E. coli O148::H28 type strain E 519-66 (Ewing, 1986) and S. dysenteriae type 1 type strain LSPQ3472 (Pupo et al., 2000) were from the Institute of Medical and Veterinary Science, Adelaide, Australia. All other Shigella and E. coli strains used were previously described (Feng et al., 2004). E. coli DH5α was used as a host strain for plasmids throughout the study. When required, antibiotics were included at the following concentrations: ampicillin 100 μg ml⁻¹ and chloramphenicol 10 μg ml⁻¹.

**Construction of a random DNase I shotgun bank.** Chromosomal DNA was prepared as previously described (Bastin & Reeves, 1995). Primers wl-1098 (5′-ATTGGTAGCTTAAAGCAGG-GCCGTAGCGT-3′) and wl-1524 (5′-TAGCGCCGTGNCCGTG-GATTAAATCGC-3′), based on the JUMPStart and gnd gene, respectively, were used to amplify the E. coli O148 O antigen gene cluster by using the Expand Long Template PCR system (Roche). PCR amplifications were in a final volume of 50 μl containing 1 μl purified DNA, 10 mM dNTP mix and 50 pmol each primer. The PCR conditions were as follows: 1 cycle 94°C for 2 min, 30 cycles 94°C for 15 s, 60°C for 30 s, 68°C for 15 min, final extension 68°C for 5 min. To limit any PCR errors, five individual PCR products were pooled. The PCR products were digested with DNase I, and the resulting DNA fragments were cloned into pGEM-T Easy to produce a bank by using the method described previously (Wang & Reeves, 1998).

**Sequencing and analysis.** The plasmid DNA template used for sequencing was prepared by the method of Sambrook et al. (1989). Sequencing was carried out with an ABI PRISM 3730 automated DNA sequencer using ABI BigDye terminator chemistry. The Staden package (Staden, 1996) and the Artemis program (Rutherford et al., 2000) were used for sequence assembly and gene annotation, respectively. The BLOCKMAKER program (Henikoff et al., 1995) was used for searching conserved motifs. BLAST and PSI-BLAST (Altschul et al., 1997) were used for searching databases, including GenBank and the Pfam protein motif database (Bateman et al., 2002), for possible gene functions. The algorithm of Eisenberg et al. (1984) was used to identify potential transmembrane segments. Sequence alignment and comparison were done with the CLUSTALW program (Thompson et al., 1994).

**Construction of an E. coli O148 wbbG knockout mutant.** The wbbG gene was replaced by a chloramphenicol acetyltransferase (CAT) gene by using the Red recombination system of phage lambda (Datsenko & Wanner, 2000; Yu et al., 2000). The CAT gene was PCR amplified from plasmid pK223–2 (Pharmacia) by using primers binding to the 5′ and 3′ ends of the gene, and each primer carried 40 bp based on the E. coli O148 DNA which flanks wbbG: wl-1333 (5′-ATAATCATATCTGGTTGATGCGCTGAAA-GTGCTGGGTTCTTCAAGGATCAAGAATA-C3′) and wl-1334 (5′-TTCAACCATGATATCTTGTGAAATGCAT-3′). The wbbG::cat PCR product was gel-purified using the UNIQ-10 Gel Extraction Kit (Sangon, Shanghai, China) and transformed into an E. coli O148 strain carrying pKD20. Chloramphenicol-resistant transformants were selected after induction of the Red genes by the protocol described by Datsenko & Wanner (2000). Gene replacement was verified by PCR and sequencing using the primer pairs wl-1337 (5′-CGATGTGGTTACGGATCC-3′)/wl-1335 (5′-TCTTCGAGCTGGATGCGT-3′) and wl-1336 (5′-GGATAGTGCTAACCCCTTGTTT-3′)/wl-1338 (5′-AGACCGTATAACCGACGCA-C3′), which were located in cat and wbbG flanking genes. One positive colony was retained as E. coli O148 strain H1229.
Cloning of the *S. dysenteriae* type 1 *wbbP* gene into *E. coli* O148 derivative H1229. The *wbbP* gene and its promoter were PCR-amplified from the *S. dysenteriae* type 1 plasmid using primer pair wl-4109 (5′-ACTGGAATTCAATATTATTCT-TATCAA-3′) and wl-2152 (5′-AGTTGGATCGGAATTCGACA-GCCGC-3′) bearing EcoRI and BamHI restriction sites (not present elsewhere in the amplification product). The PCR product was digested with EcoRI and BamHI and ligated into the vector pUC19 digested using the same two enzymes. The plasmid plW1059 was transformed into *E. coli* O148 strain H1229 and a positive colony was retained as *E. coli* O148 strain H1230.

Electroporation. All electroporations were carried out in a Bio-Rad GenePulsar apparatus using 0.2 cm gap cuvettes. Preparation of *E. coli* electrocompetent cells and electroporation were performed using standard protocols (Ausubel et al., 1999). The *E. coli* cultures electroporated with the *wbbG*::cat PCR product were incubated at 37 °C for 3 h before plating on selective plates.

Serological tests. The production of *S. dysenteriae* type 1 O antigen was determined by slide agglutination with *S. dysenteriae* type 1 type-specific antiserum. A fresh overnight bacterial culture in liquid Luria–Bertani medium was used as the antigen for slide agglutination tests.

Bacterial cultivation and extraction of the O antigen. LPS of *E. coli* O148 and *E. coli* O148 strain H1230 (1047 and 737 mg, respectively) was extracted from dried cells (12.6 and 8.9 g, respectively) by the phenol/water method (Wang et al., 2001a) and purified by precipitation of nucleic acids and proteins with aqueous 50% trichloroacetic acid as described (Wang et al., 1998). Delipidation of LPS of *E. coli* O148 and mutant strain H1230 (70 and 160 mg, respectively) was performed with aqueous 2% acetic acid (6 ml) at 100 °C until lipid A precipitation. The precipitate was removed by centrifugation (13 000 × g, 20 min), and the supernatant was fractionated on a column (56 × 2.6 cm) of Sephadex G-50 (S) (Amersham Biosciences) in 0.05 M pyridinium acetate buffer pH 4.5 monitored using a differential refractometer (Knauer). High-molecular-mass polysaccharides were obtained in yields of 40% and 35%, respectively, of the LPS weight.

Chemical analyses of the O antigen. The polysaccharides were hydrolysed with 2 M CF₂CO₂H (120 °C, 2 h) and sugars were identified by GLC of the alditol acetates on a Hewlett Packard 5890 chromatograph equipped with an Ultra-2 column (Supelco) using a temperature gradient of 160 °C to 290 °C at 3 °C min⁻¹. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (−)-2-octyl glycosides as described by Leontein & Lonnfreg (1993).

NMR spectroscopy analyses of the O antigen. Samples were deuterium-exchanged by freeze-drying twice from D₂O and then examined as solutions in 99.96% D₂O at 27 °C. NMR spectra were recorded on a Bruker DRX-500 spectrometer using internal acetone (δH 2.225, δC 31.45) as reference. Two-dimensional NMR spectra were obtained using standard Bruker software, and the Bruker XWINNMR 2.6 program was used to acquire and process the NMR data. Mixing times of 200 and 100 ms were used in total correlation spectroscopy (TOCSY) and rotating Overhauser effect spectroscopy (ROESY) experiments, respectively.

PCR specificity assay. Chromosomal DNA was prepared from 186 *E. coli* strains, including *Shigella* strains of different O antigen serotypes. The quality of DNA was examined by PCR amplification of the *mdh* gene (encoding malate dehydrogenase) and present as a housekeeping gene in *E. coli* by using primers as described previously (Pupo et al., 1997). A total of 14 pools of DNA were made, each containing DNA from 15–20 *E. coli* or *Shigella* O serotypes as described previously (Feng et al., 2004). The pools were screened by PCR by using primers based on *E. coli* O148 and *S. dysenteriae* type 1 specific genes wxz [wl-2412 (5′-TTCCCGTTATAGTTTTC-3′)/wl-2413 (5′-TTGGCCCAACATGAC-3′)]; wl-2414 (5′-CAGAACA-CAATGCTAAAC-3′)/wl-2415 (5′-TGAAATAATGCGAGTAG-3′)] and wzy [wl-2416 (5′-GTTCATATGCAATGCG-3′)/wl-2417 (5′- GATAAACGACGTCCA-3′)]; wl-2418 (5′-GGATGGAAAGG- TTGTGT-3′)/wl-2419 (5′-GGGACGCTACCTTCTT-3′)]. Each PCR was carried out in a 25 μl (total volume) mixture, and 15 μl of the mixture was loaded on an agarose gel to check for amplified DNA.

Other methods. Membrane preparation, SDS-PAGE, and silver staining for visualizing the LPS were carried out as described by Wang & Reeves (1994).

RESULTS AND DISCUSSION

Structure of the *E. coli* O148 O antigen

The O antigen was obtained by mild acid degradation of the corresponding LPS isolated from dried cells by the phenol/water procedure (Wang et al., 2001a). Sugar analysis after full acid hydrolysis of the polysaccharide revealed rhamnose (Rha), glucose and glucosamine (GlcN) in the ratios 1:2:1:0.2. GLC of the acetylated glycosides with (S)-2-octanol showed that glucosamine and glucose have the D configuration while rhamnose has the L configuration.

The 13C NMR spectrum of the polysaccharide (Fig. 1, top) contained signals for four anemic carbons at δ 95.6–103.4, one nitrogen-bearing carbon (C-2 of GlcN) at δ 53.4, two non-substituted HOCH₂-C groups at δ 61.7 (double signal), two CH₂-C groups (C-6 of Rha) at δ 18.1 and 18.3, one N-acetyl group at δ 23.6 (CH₃) and 175.5 (CO), and 15 other carbons in the region δ 68.0–80.0. Accordingly, the 1H NMR spectrum of the polysaccharide contained signals for four anomic protons at δ 5.02–5.53, one N-acetyl group at δ 2.07 and two methyl groups of rhamnose at δ 1.29 and 1.33, as well as signals of the other sugar protons. Therefore, the polysaccharide is regular and has a repeating unit containing one glucose, one glucosamine and two rhamnose residues.

The signals in the 1H and 13C NMR spectra of the polysaccharide were assigned using two-dimensional correlation spectroscopy, TOCSY, H-detected 1H,13C heteronuclear single quantum correlation (HSQC) and heteronuclear multiple band correlation (HMBC) experiments (Tables 1 and 2). Based on the coupling constant values estimated from the two-dimensional spectra, the spin systems were assigned to two rhamnose residues (Rha' and Rha") and one residue each of GlcNAc and Glc. The spin system for GlcNAc was recognized by a correlation between the nitrogen-bearing carbon (C-2) and H-2 at δ 53.4/4.12 in the 1H,13C HSQC spectrum.

The J₁₂ coupling constant values of ~ 3 Hz indicated that Glc and GlcNAc are α-linked, and the position of the signals for C-5 at δ 70.8 indicated that Rha' and Rha" are α-linked too (compare published data: Jansson et al., 1989). The pyranose form of all monosaccharide residues was demonstrated by the absence from the 13C NMR spectrum of any
signals for non-anomeric sugar carbons at a lower field than δ 80 (Bock & Pedersen, 1983).

Relatively low-field positions of the signals for C-3 of RhaI and RhaII, C-2 of Glc and C-3 of GlcNAc at δ 76.4, 79.5, 78.0 and 77.1, respectively, as compared with their positions in the corresponding non-substituted monosaccharides (Lipkind et al., 1988), demonstrated the modes of glycosylation of the monosaccharides in the repeating unit.

The sequence of the sugar residues was determined by a 1H,1H ROESY experiment, which revealed strong inter-residue RhaI H-1/RhaII H-3; RhaII H-1/Glc H-2; Glc H-1/GlcNAc H-3 and GlcNAc H-1/RhaI H-3 cross-peaks at δ 125-MHz 13C NMR spectra of the O antigens of E. coli O148 (A) and construct strain H1230 (B). Hex stands for glucose in E. coli O148 or galactose in H1230. The most significant chemical shift differences between the spectra are observed for C-2 and C-3 of Hex.

Table 1. 1H NMR data (δ, p.p.m.) of the O antigens

<table>
<thead>
<tr>
<th>Sugar residue</th>
<th>H-1</th>
<th>H-2</th>
<th>H-3</th>
<th>H-4</th>
<th>H-5</th>
<th>H-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli O148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→3)-α-L-RhapL-(1→</td>
<td>5.09</td>
<td>4.22</td>
<td>3.92</td>
<td>3.55</td>
<td>3.87</td>
<td>1.33</td>
</tr>
<tr>
<td>→3)-α-L-RhapII-(1→</td>
<td>5.08</td>
<td>4.15</td>
<td>3.86</td>
<td>3.55</td>
<td>3.87</td>
<td>1.29</td>
</tr>
<tr>
<td>→2)-α-D-Glcp-(1→</td>
<td>5.53</td>
<td>3.64</td>
<td>3.75</td>
<td>3.48</td>
<td>3.63</td>
<td>3.81</td>
</tr>
<tr>
<td>→3)-α-D-GlcpNAc-(1→</td>
<td>5.02</td>
<td>4.12</td>
<td>4.04</td>
<td>3.78</td>
<td>4.04</td>
<td>3.81</td>
</tr>
<tr>
<td>Construct strain H1230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→3)-α-L-RhapL-(1→</td>
<td>5.10</td>
<td>4.22</td>
<td>3.93</td>
<td>3.55</td>
<td>3.90</td>
<td>1.33</td>
</tr>
<tr>
<td>→3)-α-L-RhapII-(1→</td>
<td>5.05</td>
<td>4.16</td>
<td>3.87</td>
<td>3.56</td>
<td>3.90</td>
<td>1.29</td>
</tr>
<tr>
<td>→2)-α-D-Galp-(1→</td>
<td>5.59</td>
<td>3.93</td>
<td>3.88</td>
<td>4.01</td>
<td>3.92</td>
<td>3.77</td>
</tr>
<tr>
<td>→3)-α-D-GlcpNAc-(1→</td>
<td>5.03</td>
<td>4.13</td>
<td>4.08</td>
<td>3.78</td>
<td>4.20</td>
<td>3.82</td>
</tr>
</tbody>
</table>

Table 2. 13C NMR data (δ, p.p.m.) of the O antigens

<table>
<thead>
<tr>
<th>Sugar residue</th>
<th>C-1</th>
<th>C-2</th>
<th>C-3</th>
<th>C-4</th>
<th>C-5</th>
<th>C-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli O148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→3)-α-L-RhapL-(1→</td>
<td>103.4</td>
<td>68.2</td>
<td>76.4</td>
<td>71.8</td>
<td>70.8</td>
<td>18.3</td>
</tr>
<tr>
<td>→3)-α-L-RhapII-(1→</td>
<td>102.8</td>
<td>71.1</td>
<td>79.5</td>
<td>72.7</td>
<td>70.8</td>
<td>18.1</td>
</tr>
<tr>
<td>→2)-α-D-Glcp-(1→</td>
<td>99.2</td>
<td>78.0</td>
<td>73.8</td>
<td>70.6</td>
<td>73.7</td>
<td>61.7</td>
</tr>
<tr>
<td>→3)-α-D-GlcpNAc-(1→</td>
<td>95.6</td>
<td>53.4</td>
<td>77.1</td>
<td>72.5</td>
<td>73.2</td>
<td>61.7</td>
</tr>
<tr>
<td>Construct strain H1230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→3)-α-L-RhapL-(1→</td>
<td>103.4</td>
<td>68.1</td>
<td>76.4</td>
<td>71.8</td>
<td>70.8</td>
<td>18.3</td>
</tr>
<tr>
<td>→3)-α-L-RhapII-(1→</td>
<td>102.8</td>
<td>71.0</td>
<td>79.5</td>
<td>72.7</td>
<td>70.6</td>
<td>18.1</td>
</tr>
<tr>
<td>→2)-α-D-Galp-(1→</td>
<td>99.1</td>
<td>75.2</td>
<td>70.5</td>
<td>70.9</td>
<td>72.3</td>
<td>62.2</td>
</tr>
<tr>
<td>→3)-α-D-GlcpNAc-(1→</td>
<td>95.6</td>
<td>53.4</td>
<td>76.3</td>
<td>72.6</td>
<td>73.3</td>
<td>61.6</td>
</tr>
</tbody>
</table>
\[ \text{ORF} \text{ 9 was confirmed to be the glucosyltransferase gene (see below) and named as } wbbG. \]

The O antigen gene clusters of \textit{E. coli} O148 and \textit{S. dysenteriae} type 1 have the same genes in the same order with DNA identity ranging from 89.8 % to 99.5 % (Fig. 3). However, \textit{wbb}\textit{G} in \textit{S. dysenteriae} type 1 has a deletion of 22 bases flanked by 8 bp direct repeat sequences (Fig. 4), which generated a frame-shift and stop codon. ORF 9 was not assigned a function in \textit{S. dysenteriae} type 1 (Klena & Schnaitman, 1993).

The first sugar in the O antigen repeat unit of \textit{E. coli} O148 and \textit{S. dysenteriae} type 1

The \textit{E. coli} O148 gene cluster does not include a gene for an initial sugar phosphate transferase. However, the structure includes GlcNAc, and we assume that, as shown for several such O antigens, GlcNAc is the initial sugar, with GlcNAc-P transferred to undecaprenol phosphate by WecA, encoded by the \textit{wecA} gene of the enterobacterial common antigen gene cluster. We note that early workers on the genetics of the related \textit{S. dysenteriae} type 1 suggested that the first sugar of the repeat unit was galactose (Sturm \textit{et al.}, 1986), based essentially on the claim that the presence of the \textit{wbbP} gene conferred addition of galactose to the \textit{E. coli} LPS core. However, given the current understanding of the role of \textit{wecA} in \textit{E. coli} and \textit{Shigella} O antigen synthesis, we consider that conclusion most improbable, and have shown GlcNAc as first sugar of the repeat unit in Fig. 2.
Construction of a mimic of the *S. dysenteriae* type 1 O antigen from *E. coli* O148 and identification of the glucosyltransferase gene

A *wbbG* knockout mutant of *E. coli* O148 was constructed by replacing it with a CAT gene. This mutant, named H1229, produced no O antigen, whereas the wild-type strain E 519-66 produced normal LPS (Fig. 5). The *wbbG* gene and its own promoter were cloned from *S. dysenteriae* type 1 into pUC19, and the resultant plasmid pLW1059 was electroporated into H1229 to give strain H1230. Strain H1230 produced normal LPS with polymeric O antigen (Fig. 5), showing that the plasmid pLW1059 could complement the mutation in strain H1229. Slide agglutination tests showed that strain H1230 reacted with *S. dysenteriae* type 1.

### Table 3. Genes of the *E. coli* O148 O antigen gene cluster with selected properties of the genes and encoded proteins

<table>
<thead>
<tr>
<th>Gene</th>
<th>G+C content (mol%)</th>
<th>No. of residues</th>
<th>Conserved domain(s)</th>
<th>Putative function</th>
<th>Related protein</th>
<th>%Identical aa/ %Similar aa (no. of aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>rmlB</em></td>
<td>42.5</td>
<td>361</td>
<td>NAD-dependent epimerase/dehydratase family</td>
<td>dTDP-glucose 4,6-dehydratase</td>
<td>RmlB <em>Shigella boydii</em> (AAL27347)</td>
<td>97/98 (361)</td>
</tr>
<tr>
<td><em>rmlD</em></td>
<td>46.9</td>
<td>299</td>
<td>RmlD substrate-binding domain</td>
<td>dTDP-6-deoxy-l-mannose dehydrogenase</td>
<td>RmlD <em>Shigella boydii</em> (AAL27348)</td>
<td>94/97 (299)</td>
</tr>
<tr>
<td><em>rmlA</em></td>
<td>43.1</td>
<td>292</td>
<td>Nucleotidyl transferase</td>
<td>Glucose-1-phosphate thymidylyltransferase</td>
<td>RmlA <em>Shigella boydii</em> (AAL27349)</td>
<td>97/98 (289)</td>
</tr>
<tr>
<td><em>rmlC</em></td>
<td>35.7</td>
<td>185</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase</td>
<td>RmlC <em>Shigella boydii</em> (AAL27325)</td>
<td>81/88 (179)</td>
</tr>
<tr>
<td><em>wzx</em></td>
<td>31.8</td>
<td>396</td>
<td>Polysaccharide biosynthesis protein</td>
<td>O antigen transporter</td>
<td>RfbX <em>Shigella dysenteriae</em> type 1 (AAA16934)</td>
<td>97/98 (396)</td>
</tr>
<tr>
<td><em>wzy</em></td>
<td>31.1</td>
<td>380</td>
<td>None</td>
<td>O antigen polymerase</td>
<td>Rfc <em>Shigella dysenteriae</em> type 1 (AAA16935)</td>
<td>98/98 (380)</td>
</tr>
<tr>
<td><em>wbbR</em></td>
<td>30.0</td>
<td>282</td>
<td>Glycosyltransferase</td>
<td>Rhamnosyltransferase</td>
<td>RfbR <em>Shigella dysenteriae</em> type 1 (AAA16936)</td>
<td>98/98 (282)</td>
</tr>
<tr>
<td><em>wbbQ</em></td>
<td>30.0</td>
<td>303</td>
<td>Glycosyltransferase</td>
<td>Rhamnosyltransferase</td>
<td>RfbQ <em>Shigella dysenteriae</em> type 1 (AAA16937)</td>
<td>99/99 (303)</td>
</tr>
<tr>
<td><em>wbbG</em></td>
<td>31.0</td>
<td>363</td>
<td>Glycosyltransferases group 1</td>
<td>Galactosyltransferase</td>
<td>WbsH <em>Escherichia coli</em> (AAA37690)</td>
<td>33/50 (380)</td>
</tr>
</tbody>
</table>

**Fig. 4.** Sequence comparison of part of the *E. coli* O148 *wbbG* gene (top) and the corresponding sequence in *S. dysenteriae* type 1 (bottom). The sequences in the frame are direct repeats in *E. coli* O148 thought to have been involved in generating the deletion in *S. dysenteriae* type 1.
**S. dysenteriae type 1 gained its O antigen genes from E. coli O148**

The above data suggest that the *S. dysenteriae* type 1 O antigen cluster is derived from the *E. coli O148* gene cluster by loss of the proposed glucosyltransferase gene *wbbG* and gain of a plasmid-borne galactosyltransferase. The *S. dysenteriae* type 1 O antigen is one of those not otherwise known to occur in *E. coli* but is shown here to be derived from the *E. coli O148* O antigen.

Although the exact origin of the galactosyltransferase gene is unknown, it may well have been transferred from another species in the human enteric tract on the plasmid, which presumably arose by transfer of a galactosyltransferase gene from a typical polysaccharide gene cluster to a plasmid. It is most probable that the plasmid was gained before the inactivation of the glucosyltransferase gene *wbbG*, as with neither *wbbG* or *wbbP*, the O unit could not be completed and the LPS would lack O antigen.

MLEE and *mdh* sequence data show that *E. coli O148* and *S. dysenteriae* type 1 are not very closely related (Pupo et al., 1997). However, the *E. coli O148* O antigen is found in association with several H antigens and further work would be needed to determine if *S. dysenteriae* type 1 was derived from a different *E. coli O148* serotype.

**Identification of E. coli O148 and S. dysenteriae type 1 specific genes**

PCR assays based on specific genes have been developed for many pathogenic *E. coli* strains such as *E. coli O157* and *O111* (Wang et al., 1998; Wang & Reeves, 1998). Two pairs of primers designed for each of *wzx* and *wzy* from *E. coli O148* were shown to be specific to *E. coli O148* and *S. dysenteriae* type 1 by screening against 186 O serotypes of *E. coli* and *Shigella* strains by PCR. It would also be possible to distinguish the two closely related O antigens by use of primers for the plasmid-encoded *wbbP* gene found only in *S. dysenteriae* type 1.

**Concluding remarks**

In summary, the *E. coli O148* O antigen structure and gene cluster were elucidated. These data showed that *E. coli O148* has a similar O antigen and gene cluster to *S. dysenteriae* type 1, and a mimic of *S. dysenteriae* type 1 O antigen was constructed from *E. coli O148*. Based on these results, it is concluded that *S. dysenteriae* type 1 gained its O antigen from *E. coli O148*. In addition, a glucosyltransferase gene of *E. coli O148* was identified and two genes were confirmed to be serotype-specific to *E. coli O148* and *S. dysenteriae* type 1.
ACKNOWLEDGEMENTS

This work was supported by the NSFC (National Natural Science Foundation of China) Key Programs (30530010, 20536040), NSFC General Programs (30370023, 30370339), Tianjin Municipal Fund for Science and Technology Innovation (05FZZDHS0000), the Russian Foundation for Basic Research (05-04-48992) and the Council on Grants of the President of the Russian Federation (MK-1597.2005.3).

REFERENCES


Edited by: N. J. High