Structure and gene cluster of the O antigen of *Escherichia coli* L-19, a candidate for a new O-serogroup

Evelina L. Zdorovenko,1 Lyudmila D. Varbanets,2 Bin Liu,3 Olga A. Valueva,1† Quan Wang,3 Alexander S. Shashkov,1 Ekaterina G. Garkavaya,4 Oksana S. Brovarskaya,2 Lei Wang3 and Yuriy A. Knirel1

1N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninskii Prospekt 47, 119991 Moscow, Russia
2D. K. Zabolotny Institute of Microbiology and Virology, National Academy of Sciences of Ukraine, Akademik Zabolotny Street 154, 03680 Kyiv, GSP, Ukraine
3TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, 300457 Tianjin, PR China
4National Aviation University, Kosmonavt Komarov Prospekt 1, 03058 Kyiv, GSP, Ukraine

*Escherichia coli* L-19 isolated from a healthy individual did not agglutinate with any of 21 polyvalent antisera that cover 174 *E. coli* O-serogroups. The strain was studied in respect to the O-antigen (O-specific polysaccharide, OPS) structure and genetics. The LPS was isolated by phenol–water extraction of bacterial cells and cleaved by mild acid hydrolysis to yield the OPS. The OPS was studied by sugar and methylation analyses, along with 1D and 2D 1H and 13C NMR spectroscopy. The established structure of the linear tetrasaccharide repeating unit was found to be unique among known bacterial polysaccharide structures. A peculiar component of the L-19 OPS was an amide of glucuronic acid with 2-amino-1,3-propanediol (2-amino-2-deoxyglycerol) (GroN). The O-antigen gene cluster of L-19 between the conserved genes *galF* and *gnd* was sequenced, and gene functions were tentatively assigned by a comparison with sequences in the available databases and found to be in agreement with the OPS structure. Except for putative genes for synthesis and transfer of GroN, the sequences in the L-19 O-antigen gene cluster were little related to those of reference strains of the 174 known *E. coli* O-serogroups. The data obtained suggest that L-19 can be considered as a candidate for a new *E. coli* O-serogroup.

INTRODUCTION

*Escherichia coli*, the predominant facultative anaerobe of the colonic flora of many mammals including humans, is represented by both commensal and pathogenic forms. *E. coli* clones are normally classified by a combination of somatic (O), flagellar (H) and capsular (K) antigens. The O antigen (O-specific polysaccharide, OPS) is a part of the LPS present on the surface of Gram-negative bacteria, which consists of a number of repeats of an oligosaccharide unit (O-unit). The fine structure of the OPS defines the O specificity of bacteria, which is important for their niche adaptation. Many bacteria, including *E. coli*, are characterized by a high O-antigen diversity, which results from the variation in the type of sugars present, their arrangement, and the glycosidic linkages within and between the O-units. Genes involved in the biosynthesis of the O antigen are generally arranged in a cluster, which in most *E. coli* strains maps between the housekeeping genes *galF* and *gnd* (Bastin & Reeves, 1995).

Typing systems of medically important bacteria are necessary for identification of clinical isolates and epidemiological monitoring. Currently, 174 different *E. coli* O-antigen forms are known, designated O1 to O181 with the exception of some O-serogroups, which have been historically...
removed, and subtypes exist in several groups (Stenutz et al., 2006). Furthermore, O182–O187 are internationally recognized as new O-serogroups (http://www.ssi.dk/English/SSI%20Diagnostics/Products%20from%20SSI%20Diagnostics/ Antisera_antibodies/). In this work, we studied serologically, chemically and genetically the O antigen of an E. coli strain isolated from a healthy individual, and found it to be related to none of the E. coli clones characterized so far.

METHODS

Growth of bacteria and isolation of the LPS and OPS. E. coli L-19 was isolated from a healthy individual in the course of a preventive medical examination of a hospital staff member in Kiev, Ukraine. It did not agglutinate with any of 21 polyvalent antisera that cover the 174 E. coli O-serogroups. The strain was grown on solid meat–peptone agar medium at 28–30 °C for 24 h. Cells were separated by centrifugation (20 min, 3000 g), and dried with acetone and ether. LPS was isolated by the phenol–water procedure (Westphal & Jann, 1965), followed by removal of nucleic acids by ultracentrifugation at 105 000 g.

The OPS was obtained by degradation of the LPS precipitate with aqueous 2 % HOAc for 3 h at 100 °C, followed by gel-permeation chromatography on a column (70 × 3.0 cm) of Sephadex G-50 Superfine (Amersham Biosciences), using 0.05 M pyridinium acetate pH 4.5 as eluent and monitoring by the phenol–sulfuric acid reaction.

Sugar analyses. Hydrolysis of the OPS was performed with 2 M CF3CO2H at 120 °C, 2 h; the monosaccharides were analysed by GLC as the alditol acetates (Sawardeker et al., 1965) on a Maestro (Agilent 7820) chromatograph (Interlab) equipped with an HP-5 column (0.32 mm × 30 m) using a temperature programme of 160 °C (1 min) to 290 °C at 7 °C min−1. The absolute configuration of glucose and glucosamine (GlcN) was determined by GLC of the acetylated (S)-2-octyl glycosides as described by Leontiev & Lönngren (1993).

Methylation analysis. An OPS sample was methylated with CH3I in DMSO in the presence of sodium methylsulfinylmethanide (Conrad, 1972). The methylated OPS was hydrolysed with 2 M CF3CO2H at 120 °C, 2 h, and the partially methylated monosaccharides were conventionally reduced with NaBH4, acetylated and analysed by GLC–MS on an Agilent MSD 5975C instrument equipped with an HP-5ms column using a temperature gradient of 150 °C (3 min) to 320 °C at 5 °C min−1.

NMR spectroscopy. Samples were methylated by freeze-drying from 99.95% D2O. NMR spectra were recorded for solutions in internal standards for calibration. 2D NMR spectra were obtained by acquiring and processing the NMR data. Mixing times of 150 ms were used in 2D TOCSY (total correlation spectroscopy) and ROESY (rotating-frame nuclear Overhauser effect spectroscopy) experiments, respectively. A 2D 1H,13C HMBC (heteronuclear multiple-bond correlation) experiment was recorded with a 60 ms delay for evolution of long-range couplings to optimize the spectrum for coupling constant J1,H C 8 Hz.

Sequencing and analysis of genes. Chromosomal DNA was prepared using DNA preparation kits from Promega. A genomic library containing 3 kb inserts for illumina paired-end sequencing was prepared using the protocols provided by Illumina, and a Solexa Genome Analyser Ix (Illumina) was used to sequence the sample to a depth of about 100-fold coverage. The reads obtained were assembled using the de novo genome-assembly program Velvet, to generate a multi-contig draft genome. The sequence of the O-antigen gene cluster, located between galF and gnu, was retrieved from the genome, and was analysed using computer programs as described by Feng et al. (2004).

Serotyping. Serotyping was performed as described by Orskov & Orskov (1984) using antisera produced by Tianjin Biochip. L-19 was tested with 21 sets of polyvalent antisera that together contain antisera against all 174 E. coli O-serogroups. L-19 also was tested against an individual antiserum to E. coli O143 included in one of the sets.

RESULTS

Structure elucidation of the OPS

The LPS was isolated from bacterial cells of E. coli L-19 by the Westphal procedure and purified by ultracentrifugation. Mild acid degradation of the LPS with 2 % acetic acid afforded high-molecular-mass OPS, which was isolated by gel-permeation chromatography on Sephadex G-50 Superfine.

Sugar analysis of the OPS by GLC of the alditol acetates derived after full acid hydrolysis revealed glucose, galactose and GlcN in the ratio 5:1:2 (detector response). Further NMR spectroscopic studies showed that, opposite to the other monosaccharides, galactose displayed no major signals in the NMR spectra; hence, it does not enter into the repeating unit and, most likely, is a component of the LPS core (Holst, 1999) linked to the isolated OPS. GLC of the acetylated (S)-2-octyl glycosides showed that the glucose and GlcN have the D configuration.

As judged by the 13C NMR spectrum [the presence of signals for two carboxyl groups (C-6) δ 172.0 and 175.8] and the 2D 1H,13C HMBC spectrum, which showed correlations between C-6 and H-5, the OPS contained two hexuronic acids, which are not detectable in GLC analysis of the alditol acetates. They were identified as glucuronic acid (GlcA) and galacturonic acid (GaA) by characteristic J1H,C 3J1J2 coupling constants estimated from the 2D 1H,1H COSY (correlation spectroscopy) and TOCSY spectra, particularly: J3,4 ≈ J4,5 ~9 Hz for GlcA; J3,4 and J4,5 <3 Hz for GaA.

In addition, the NMR spectra contained signals for 2-amino-1,3-propanediol (2-amino-2-deoxyglucitol) (GroN), including those for a nitrogen-bearing carbon (C-2) at δ 54.2 and two HOCH2 groups (C-1 and C-3) at δ 61.9. Its position was determined by measuring NMR spectra in a 1:9 D2O/H2O mixture, which enabled detection of nitrogen-linked protons. Signals for two NH protons were observed at δ 8.28 and 8.04, and assigned to GlcN and GroN, respectively, using 1H,1H COSY and TOCSY spectra. A 2D ROESY experiment revealed correlations between NH of GroN and H-5 and H-4 of GlcA, and, hence, GroN is amidelinked to the carboxyl group of GlcA.

In methylation analysis of the OPS, a derivative of a 4-substituted hexose was identified by GLC–MS of partially methylated alditol acetates. After reduction of the methylated polysaccharide with LiBH4, a 4,6-disubstituted hexose derivative was detected also, which was evidently
derived from GalA. No derivative of GlcA was detected as it was amidated with GroN and was not carboxyl-reduced under the conditions used.

The $^1$H NMR spectrum of the OPS showed six signals in the low-field region $\delta$ 4.63–4.94, which were assigned to four anomeric protons, as well as H-4 and H-5 of GalA. In addition, the spectrum contained signals for one N-acetyl group at $\delta$ 2.03 and other protons at $\delta$ 3.39–4.04 (Table 1).

The $^{13}$C NMR spectrum of the OPS (Fig. 1, Table 1) showed signals for four anomeric carbons at $\delta$ 102.5–105.4, four HOCH$_2$-C groups (C-6 of GlcN and Glc, C-1 and C-3 of GroN) at $\delta$ 61.4–62.6, two CO$_2$H groups (C-6 of GlcA and GalA) at $\delta$ 172.0 and 175.8, two nitrogen-bearing carbons (C-2 of GlcN and GroN) at $\delta$ 54.2 and 56.6, other sugar ring carbons at $\delta$ 70.1–83.3, and one N-acetyl group at $\delta$ 23.8 (CH$_3$) and 176.2 (CO). The absence of signals from the region $\delta$ 82–88 showed that all monosaccharides occurred in the pyranose form.

The $^1$H and $^{13}$C NMR spectra of the OPS were assigned (Table 1), and spin systems for four monosaccharides [Glc, GlcA, GalA and GlcNAc (N-acetylglucosamine)] and GroN were identified using 2D $^1$H,$^1$H COSY, TOCSY, ROESY, $^1$H,$^1$C heteronuclear HSQC (heteronuclear single-quantum coherence) and HMBC experiments. The $\beta$ configuration of all sugar residues was established by relatively large $J_{1,2}$ values $>7$ Hz and C-5 chemical shifts compared with published data of the corresponding $\alpha$- and $\beta$-pyranosides (Bock & Pedersen, 1983).

**Table 1.** $^1$H and $^{13}$C NMR chemical shifts of the OPS in D$_2$O [$\delta$ (p.p.m.)]

<table>
<thead>
<tr>
<th>Residue</th>
<th>H-1</th>
<th>H-2</th>
<th>H-3</th>
<th>H-4</th>
<th>H-5</th>
<th>H-6a, 6b</th>
<th>NAc</th>
<th>NH*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$3$\rightarrow$-$\beta$-D-GlcPA-(1→ A)</td>
<td>4.63</td>
<td>3.63</td>
<td>3.61</td>
<td>3.58</td>
<td>3.91</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\rightarrow$4$\rightarrow$-$\beta$-D-GlcP-(1→ B)</td>
<td>4.70</td>
<td>3.39</td>
<td>3.65</td>
<td>3.64</td>
<td>3.59</td>
<td>3.91, 4.02</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\rightarrow$4$\rightarrow$-$\beta$-D-GalPA-(1→ C)</td>
<td>4.81</td>
<td>3.60</td>
<td>3.68</td>
<td>4.46</td>
<td>4.77</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\rightarrow$3$\rightarrow$-$\beta$-D-GlcpNAc-(1→ D)</td>
<td>4.94</td>
<td>3.88</td>
<td>3.83</td>
<td>3.50</td>
<td>3.47</td>
<td>3.75, 3.93</td>
<td>2.03</td>
<td>8.28</td>
</tr>
<tr>
<td>GroN</td>
<td>3.66</td>
<td>3.71</td>
<td>4.04</td>
<td>3.66, 3.71</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Measured in a 9:1 H$_2$O/D$_2$O mixture.

**Fig. 1.** $^{13}$C NMR spectrum of the OPS of *E. coli* L-19. Numbers refer to carbons in GroN and sugar residues that are denoted as follows: A, $\beta$-GlcPA; B, $\beta$-GlcP; C, $\beta$-GalPA; D, $\beta$-GlcNAc.
Table 2. Correlations for H-1 and C-1 in the 2D ROESY and 1H,13C HMBC spectra of the OPS

<table>
<thead>
<tr>
<th>Anomeric atom in sugar residue (α)</th>
<th>ROESY</th>
<th>Correlations to atom in sugar residue (δ)</th>
<th>HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A H-1 (4.63)</td>
<td>B H-3 (3.65), A H-3 (3.61), A H-5 (3.91)</td>
<td>B C-3 (75.3)</td>
<td></td>
</tr>
<tr>
<td>B H-1 (4.70)</td>
<td>C H-4 (4.46), B H-3 (3.65), B H-5 (3.59)</td>
<td>C C-4 (81.4)</td>
<td></td>
</tr>
<tr>
<td>B C-1 (105.4)</td>
<td>C H-4 (4.46), B H-2 (3.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C H-1 (4.81)</td>
<td>D H-3 (3.83), C H-3 (3.68), C H-5 (4.77)</td>
<td>D C-3 (83.3)</td>
<td></td>
</tr>
<tr>
<td>C C-1 (102.8)</td>
<td>D H-3 (3.83), C H-2 (3.60), C H-5 (4.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D H-1 (4.94)</td>
<td>A H-3 (3.61), D H-3 (3.83), D H-5 (3.47)</td>
<td>A C-3 (80.2)</td>
<td></td>
</tr>
<tr>
<td>D C-1 (102.5)</td>
<td>A H-3 (3.61), D H-2 (3.88)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Positions of substitution of the monosaccharides were determined by downfield displacements of the signals for C-3 of GlcNAc and GlcA, and C-4 of Glc and GalA (Table 1), as compared with their positions in the corresponding unsubstituted monosaccharides (Bock & Pedersen, 1983). Glycosylation of GlcNAc and GlcA at position 3 and Glc and GalA at position 4 was confirmed by inter-residue correlations in the 2D 1H,13C HMBC and ROESY spectra (Table 2), which also defined the monosaccharide sequence in the repeating unit.

With the known sugar sequence, the absolute configurations of GlcA and GalA were determined by known effects of glycosylation on 13C NMR chemical shifts. Relatively large negative β-effects (−1.7 and −1.1 p.p.m.) on C-3 of β-GlcP and C-4 of β-GlcPNAc caused by their glycosylation at position 4 or 3 by β-GlcPα or β-GalPα, respectively, indicated the same absolute configuration of the linked monosaccharides, i.e. the D configuration of GlcA and GalA (in case of the different configurations, the β-effects would be about zero or positive) (Shashkov et al., 1988; Lipkind et al., 1988). Therefore, the OPS of E. coli L-19 has the structure shown in Fig. 2.

Genetics of the O-antigen biosynthesis

Nine ORFs were found in the O-antigen gene cluster of L-19, which is located between galF and gnd; all ORFs had the same transcriptional direction from galF to gnd (Table 3, Fig. 3). orf1–orf3 of L-19 share 43/63 %, 30/54 % and 45/ 65 % identity/similarity, respectively, to the first three genes of the O-antigen gene cluster shared by E. coli O143 and Shigella boydii type 8 (Liu et al., 2008). Orf1 belongs to aminotransferase class I and II family (PF00155, probability value 8.3 × 10−33) and is a homologue of WfdP, a putative aminotransferase of E. coli O143/S. boydii type 8. Orf3 is similar to WfdR, which is a putative transferase also involved in the synthesis of E. coli O143/S. boydii type 8 O antigen. As the O antigens of both these bacteria (Landersjö et al., 1996; L'vov et al., 1983) and E. coli L-19 (this work) contain GroN, orf1 and orf3 were proposed to encode enzymes responsible for synthesis of GroN and its transfer to GalA (in O143 and type 8) or GlcA (in L-19). Orf2 is a homologue of WfdQ of E. coli O143/S. boydii type 8, whose function has not been assigned.

Orf4 and Orf6 are the only two proteins with predicted transmembrane segments. Orf4 has 12 well-proportioned transmembrane segments and belongs to a flippase protein family (PF01943, probability value 5.8 × 10−46); it shares 51 % similarity to the putative Wzy protein of Grimontia sp. AK16. Orf6 has 10 predicted transmembrane segments with a large periplasmic loop of 69 aa residues, which is a typical topological character of Wzx proteins; it shares 56 % similarity with the Wzx protein of Salmonella enterica O59. Therefore, orf4 and orf6 were proposed to encode

\[ \text{E. coli L-19} \]

\[ \rightarrow 3')β-\text{d-GlcP}(-\text{GroN})-(1 \rightarrow 4')\beta-\text{d-GlcP}-(1 \rightarrow 4')β-\text{d-GalP}-(1 \rightarrow 3')\beta-\text{d-GlcPNAc}-(1 \rightarrow 3')β-\text{d-GlcPNAc}-(1 \rightarrow \]

\[ \text{E. coli O143/S. boydii type 8} \]

\[ \rightarrow 2')β-\text{d-GalP}(-\text{GroN})3,4\text{Ac}-(1 \rightarrow 3')\alpha-\text{d-GalPNAc}-(1 \rightarrow 4')β-\text{d-GlcP}-(1 \rightarrow 3')β-\text{d-GlcPNAc}-(1 \rightarrow \]

Fig. 2. Structures of the O-antigens of E. coli L-19 (this work) and E. coli O143/S. boydii type 8 (Landersjö et al., 1996; L'vov et al., 1983). A-D indicate monosaccharide residues.

http://mic.sgmjournals.org 2105

On: Tue, 08 Jan 2019 05:26:41
Table 3. Characteristics of the ORFs in the O-antigen gene cluster of E. coli

<table>
<thead>
<tr>
<th>ORF no.</th>
<th>Position of gene (G+C content)</th>
<th>G+C content (%)</th>
<th>Similar protein(s), strain(s) (GenBank accession no.)</th>
<th>% Identical / % similar (no. of aa overlap)</th>
<th>Putative function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1622–2722 32.3</td>
<td></td>
<td>Aminotransferase class I and II (PF00155) type 8 (ACA24747) 43/63 (365)</td>
<td>Aminotransferase</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2759–3448 29.1</td>
<td></td>
<td>WfdQ, S. boydii type 8 (ACA24748) 30/54 (223)</td>
<td>Unknown</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4313–5533 29.7</td>
<td></td>
<td>Polysaccharide biosynthesis protein 2 (PF00355) AK16 (WP_002535587) e-value 6.1</td>
<td>Unknown</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5548–6417 31.7</td>
<td></td>
<td>Glycosyltransferase group 2 (PF00535) O-antigen polymerase</td>
<td>Unknown</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6420–7658 30.4</td>
<td></td>
<td>Glycosyltransferase group 2 (PF00355) O-antigen polymerase</td>
<td>Unknown</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7645–8556 30.4</td>
<td></td>
<td>Glycosyltransferase group 2 (PF00355) O-antigen polymerase</td>
<td>Unknown</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8599–9861 27.5</td>
<td></td>
<td>Glycosyltransferase group 2 (PF00355) O-antigen polymerase</td>
<td>Unknown</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9873–10682 27.5</td>
<td></td>
<td>Glycosyltransferase group 2 (PF00355) O-antigen polymerase</td>
<td>Unknown</td>
<td>9</td>
</tr>
</tbody>
</table>

DISCUSSION

The OPS of E. coli L-19 has a linear tetrasaccharide repeating unit containing two hexuronic acid residues. One of them, GaLA, has a free carboxyl group and confers acidic properties on the OPS, and the other, GlcA, occurs as an amide with GroN. To our knowledge, the latter component has not been hitherto reported in OPSs but in capsular polysaccharides of Campylobacter jejuni (Karlyshev et al., 2005), whereas amide of GaLA with GroN is more common and has been found in a number of OSPs (see the Bacterial Carbohydrate Structure Database at http://csdb.glycoscience.ru/bacterial), including those of E. coli O143 (Landersjø et al., 1996) and S. boydii type 8 (L'vov et al., 1983). The OPSs of these bacteria have an identical structure (Fig. 2) that shows both similarities with and differences from that of E. coli L-19.

All in all, the OPS structure of E. coli L-19 established in this work is unique among known bacterial polysaccharide structures. This itself does not mean that no existing O-serogroup shares this structure as OPS structures have been elucidated only in ~70% of O-serogroups (see http://www.casper.organ.su.se/ECODAB/ and http://csdb.glycoscience.ru/bacterial). However, the O-antigen gene clusters have been sequenced in all 168 E. coli O-serogroups that existed until recently, 101 of them being reported and 67 from our unpublished data. None of the sequences shows marked homology with the E. coli L-19 sequences, except for putative genes for the synthesis and transfer of GroN, which are homologous in E. coli L-19 and E. coli O143/S. boydii type 8 (Fig. 3). Moreover, E. coli L-19 was not recognized by any of the 21 polyvalent antisera that cover 174 E. coli O-serogroups, nor by individual antiserum against the most closely related E. coli O143.

Recently, additional E. coli O-serogroups, O182–O187, have been internationally recognized (http://www.ssi.dk/English/
SSD%20Diagnostica/Products%20from%20SSD%20Diagnostica/ Antisera_antibodies/), though they have not been published. No data on their OPS structures or O-antigen gene cluster sequences are available, and it is not excluded that \textit{E. coli} L-19 falls in one of these serogroups. However, if the O182–O187 O-serogroups are finally approved and further studies show that L-19 does not belong to any of them, this strain must be classified into a new \textit{E. coli} O-serogroup.

**ACKNOWLEDGEMENTS**

E. L. Z., A. S. S. and Y. A. K. were supported by the Russian Science Foundation (14-14-01042); B. L., Q. W. and L. W. were supported by the National Key Program for Infectious Diseases of China (2013ZX10004216-001-001), National 973 Program of China grants (2013CB733904 and 2011CB504900), the International Science & Technology Cooperation Program of China (2012DFG31680 and 2013DEFR36040), the National Natural Science Foundation of China (NSFC) Program (31030002, 31371259 and 31270003) and a research project of the Chinese Ministry of Education (113015A).

**REFERENCES**


Edited by: V. Ejsink