Microbiology (2014), **160**, 2102–2107 **DOI** 10.1099/mic.0.080804-0

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 Wang[3] and Yuriy A. Knirel[1]

Correspondence
Evelina L. Zdorovenko
zdorovenkoe@mail.ru

[1]N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninskii Prospekt 47, 119991 Moscow, Russia

[2]D. K. Zabolotny Institute of Microbiology and Virology, National Academy of Sciences of Ukraine, Akademik Zabolotny Street 154, 03680 Kyiv, GSP, Ukraine

[3]TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, 300457 Tianjin, PR China

[4]National 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 ^{1}H and ^{13}C 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.

Received 23 May 2014
Accepted 1 July 2014

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

†Present address: DFG-Center for Regenerative Therapies Dresden, Fetscherstrasse 105, 01307 Dresden, Germany.

Abbreviations: COSY, correlation spectroscopy; GalA, galacturonic acid; GlcA, glucuronic acid; GlcN, glucosamine; GlcNAc, *N*-acetylglucosamine; GroN, 2-amino-1,3-propanediol (2-amino-2-deoxyglycerol); HMBC, heteronuclear multiple-bond correlation; OPS, O-specific polysaccharide (O-polysaccharide); ROESY, rotating-frame nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

The GenBank/EMBL/DDBJ accession number for the O-antigen gene cluster sequence of *E. coli* L-19 is KJ776793.

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%20Diagnostica/Products%20from%20SSI%20Diagnostica/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, 5000 ***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 CF_3CO_2H (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 Leontein & Lönngren (1993).

Methylation analysis. An OPS sample was methylated with CH_3I in DMSO in the presence of sodium methylsulfinylmethanide (Conrad, 1972). The methylated OPS was hydrolysed with 2 M CF_3CO_2H (120 °C, 2 h), and the partially methylated monosaccharides were conventionally reduced with $NaBH_4$, 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 deuterium-exchanged by freeze-drying from 99.9 % D_2O. NMR spectra were recorded for solutions in 99.95 % D_2O at 30 °C on a Bruker Avance II 600 MHz spectrometer with 5 mm broadband inverse probehead. Sodium 3-(trimethylsilyl)-propanoate-2,2,3,3-D_4 (δ_H 0) and acetone (δ_C 31.45) were used as internal standards for calibration. 2D NMR spectra were obtained using standard Bruker software, and the Bruker TopSpin 2.1 program was used to acquire and process the NMR data. Mixing times of 150 and 200 ms were used in 2D TOCSY (total correlation spectroscopy) and ROESY (rotating-frame nuclear Overhauser effect spectroscopy) experiments, respectively. A 2D $^1H,^{13}C$ 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 $J_{H,C}$ 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 IIx (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 *gnd*, 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 ultracentifugation. 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 ^{13}C NMR spectrum [the presence of signals for two carboxyl groups (C-6) δ 172.0 and 175.8] and the 2D $^1H,^{13}C$ 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 (GalA) by characteristic $^3J_{H,H}$ coupling constants estimated from the 2D $^1H,^1H$ COSY (correlation spectroscopy) and TOCSY spectra, particularly: $J_{3,4} \approx J_{4,5}$ ~9 Hz for GlcA; $J_{3,4}$ and $J_{4,5}$ <3 Hz for GalA.

In addition, the NMR spectra contained signals for 2-amino-1,3-propanediol (2-amino-2-deoxyglycerol) (GroN), including those for a nitrogen-bearing carbon (C-2) at δ 54.2 and two $HOCH_2$ groups (C-1 and C-3) at δ 61.9. Its position was determined by measuring NMR spectra in a 1 : 9 D_2O/H_2O 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 amide-linked 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 $LiBH_4$, a 4,6-disubstituted hexose derivative was detected also, which was evidently

Table 1. ^{1}H and ^{13}C NMR chemical shifts of the OPS in D_2O [δ (p.p.m.)]

Residue	H-1	H-2	H-3	H-4	H-5	H-6a, 6b	NAc	NH*
→3)-β-D-Glc*p*A-(1→ **A**	4.63	3.63	3.61	3.58	3.91	–	–	–
→4)-β-D-Glc*p*-(1→ **B**	4.70	3.39	3.65	3.64	3.59	3.91, 4.02	–	–
→4)-β-D-Gal*p*A-(1→ **C**	4.81	3.60	3.68	4.46	4.77	–	–	–
→3)-β-D-Glc*p*NAc-(1→ **D**	4.94	3.88	3.83	3.50	3.47	3.75, 3.93	2.03	8.28
GroN	3.66, 3.71	4.04	3.66, 3.71	–	–	–	–	8.04
	C-1	**C-2**	**C-3**	**C-4**	**C-5**	**C-6**	**NAc**	
							CH_3	**CO**
→3)-β-D-Glc*p*A-(1→ **A**	102.6	71.7	80.2	73.1	76.2	172.0	–	–
→4)-β-D-Glc*p*-(1→ **B**	105.4	74.7	75.3	80.2	77.7	61.4	–	–
→4)-β-D-Gal*p*A-(1→ **C**	102.8	76.3	71.5	81.4	77.7	175.8	–	–
→3)-β-D-Glc*p*NAc-(1→ **D**	102.5	56.6	83.3	70.1	77.1	62.6	23.8	176.2
GroN	61.9	54.2	61.9	–	–	–	–	–

*Measured in a 9:1 H_2O/D_2O mixture.

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 δ 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 δ 2.03 and other protons at δ 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 δ 102.5–105.4, four $HOCH_2$-C groups (C-6 of GlcN and Glc, C-1 and C-3 of GroN) at δ 61.4–62.6, two CO_2H groups (C-6 of GlcA and GalA) at δ 172.0 and 175.8, two nitrogen-bearing carbons (C-2 of GlcN and GroN) at δ 54.2 and 56.6, other sugar ring carbons at δ 70.1–83.3, and one *N*-acetyl group at δ 23.8 (CH_3) and 176.2 (CO). The absence of signals from the region δ 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, $^{1}H,^{13}C$ heteronuclear HSQC (heteronuclear single-quantum coherence) and HMBC experiments. The β 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 α- and β-pyranosides (Bock & Pedersen, 1983).

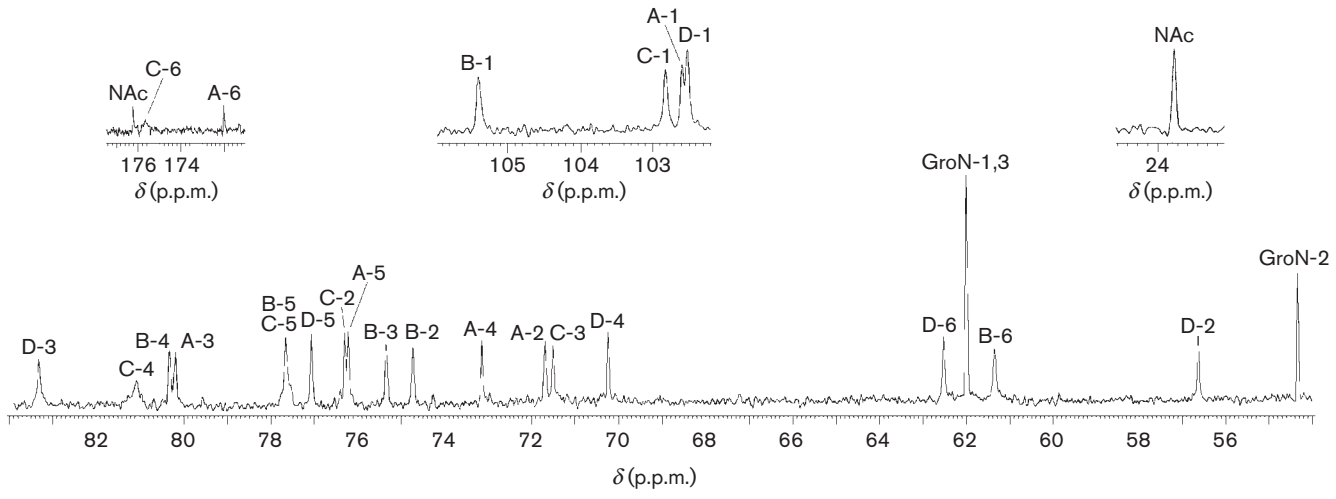

Fig. 1. ^{13}C NMR spectrum of the OPS of *E. coli* L-19. Numbers refer to carbons in GroN and sugar residues that are denoted as follows: A, β-Glc*p*A; B, β-Glc*p*; C, β-Gal*p*A; D, β-GlcNAc.

Table 2. Correlations for H-1 and C-1 in the 2D ROESY and $^{1}H,^{13}C$ HMBC spectra of the OPS

Sugar abbreviations: A, β-Glc*p*A; B, β-Glc*p*; C, β-Gal*p*A; D, β-GlcNAc.

Anomeric atom in sugar residue (δ)	Correlations to atom in sugar residue (δ)	
	ROESY	HMBC
A H-1 (4.63)	B H-3 (3.65), A H-3 (3.61), A H-5 (3.91)	B C-3 (75.3)
A C-1 (102.6)		A H-2 (3.63), A H-3 (3.61), A H-5 (3.91)
B H-1 (4.70)	C H-4 (4.46), B H-3 (3.65), B H-5 (3.59)	C C-4 (81.4)
B C-1 (105.4)		C H-4 (4.46), B H-2 (3.39)
C H-1 (4.81)	D H-3 (3.83), C H-3 (3.68), C H-5 (4.77)	D C-3 (83.3)
C C-1 (102.8)		D H-3 (3.83), C H-2 (3.60), C H-5 (4.77)
D H-1 (4.94)	A H-3 (3.61), D H-3 (3.83), D H-5 (3.47)	A C-3 (80.2)
D C-1 (102.5)		A H-3 (3.61), D H-2 (3.88)

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 $^{1}H,^{13}C$ 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 ^{13}C NMR chemical shifts. Relatively large negative β-effects (−1.7 and −1.1 p.p.m.) on C-3 of D-Glc*p* and C-4 of D-Glc*p*NAc caused by their glycosylation at position 4 or 3 by β-Glc*p*A or β-Gal*p*A, 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 % identitiy/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^{-53}) 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^{-14}); it shares 51 % similarity to the putative Wzx 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 Wzy proteins; it shares 56 % similarity with the Wzy protein of *Salmonella enterica* O59. Therefore, *orf4* and *orf6* were proposed to encode

E. coli L-19

→ 3)-β-D-Glc*p*A(GroN)-(1→4)-β-D-Glc*p*-(1→4)-β-D-Gal*p*A-(1→3)-β-D-Glc*p*NAc-(1→

A B C D

E. coli O143/*S. boydii* type 8

→ 2)-β-D-Gal*p*A(GroN)3,4Ac-(1→3)-α-D-Gal*p*NAc-(1→4)-β-D-Glc*p*A-(1→3)-β-D-Glc*p*NAc-(1→

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.

Table 3. Characteristics of the ORFs in the O-antigen gene cluster of *E. coli* L-19

ORF no.	Position of gene	G+C content (mol%)	Conserved domain(s)	Similar protein(s), strain(s) (GenBank accession no.)	% Identical / % similar (no. of aa overlap)	Putative function of protein
1	1622–2722	32.3	Aminotransferase class I and II (PF00155) *e* value = 8.3×10^{-53}	WfdP, *S. boydii* type 8 (ACA24747)	43/63 (365)	Aminotransferase
2	2759–3448	29.1		WfdQ, *S. boydii* type 8 (ACA24748)	30/54 (223)	Unknown
3	3402–4316	28.3		WfdR, *S. boydii* type 8 (ACA24749)	45/65 (280)	Transferase
4	4313–5533	29.7	Polysaccharide biosynthesis protein (PF01943) *e* value = 5.8×10^{-14}	O-antigen transporter, *Grimontia* sp. AK16 (WP_002535587)	28/51 (398)	O-antigen flippase
5	5548–6417	31.7	Glycosyltransferase group 2 (PF00535) *e* value = 6.1×10^{-19}	Glycosyltransferase, *Halomonas titanicae* (WP_009286559.1)	44/60 (290)	Glycosyltransferase
6	6420–7658	30.4		O-antigen polymerase, *Salmonella enterica* O59 (AFW04847)	30/56 (175)	O-antigen polymerase
7	7645–8556	32.0	Glycosyltransferase group 2 (PF00535) *e* value = 2.6×10^{-28}	Glycosyltransferase, *Bacillus halodurans* C-125 (NP_244580)	34/56 (301)	Glycosyltransferase
8	8599–9861	29.6		WbgV, *Shigella sonnei* (AAG17422)	36/55 (419)	Unknown
9	9873–10682	27.5	Glycosyltransferase group 2 (PF00535) *e* value = 2.3×10^{-6}	Glycosyltransferase, *Photobacterium profundum* (WP_006232917)	53/71 (270)	Glycosyltransferase

O-antigen flippase and O-antigen polymerase, respectively. Orf5, Orf7 and Orf9 belong to glycosyltransferase group 2 family (PF00155, probability values 6.1×10^{-19}, 2.6×10^{-28} and 2.3×10^{-6}, respectively). Therefore, *orf5*, *orf7* and *orf9* are putative glycosyltransferase genes involved in the assembly of the L-19 O-unit. No function could be assigned to Orf8 based on similarity to genes in available databases.

Genes for synthesis of the nucleotide precursors of GlcNAc and Glc, which are common sugars in bacteria, are located outside O-antigen gene clusters. UDP-GlcA is synthesized from UDP-Glc by UDP-glucose 6-dehydrogenase (Ugd), and UDP-GalA from UDP-GlcA by a 4-epimerase (Gla). Both *ugd* and *gla* map outside of the O-antigen gene cluster too, usually between *gnd* and *his* operons in *E. coli* and *Shigella* sp. (Liu *et al.*, 2008). Finally, *wecA*, which mediates initiation of the O-antigen synthesis by transfer of GlcNAc-1-P from UDP-GlcNAc to undecaprenylphosphate, is located in the gene cluster of the enterobacterial common antigen and is not duplicated in the O-antigen gene cluster (Reeves & Wang, 2002).

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 OPSs (see the Bacterial Carbohydrate Structure Database at http://csdb.glycoscience.ru/bacterial), including those of *E. coli* O143 (Landersjö *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/

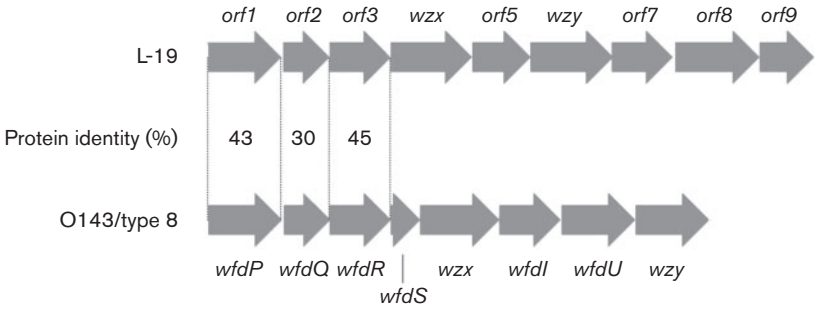

Fig. 3. Organization of the O-antigen gene clusters of *E. coli* L-19 (this work) and *E. coli* O143/*S. boydii* type 8 (adapted from Liu *et al.*, 2008). For putative gene functions of L-19 see Table 3.

SSI%20Diagnostica/Products%20from%20SSI%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 *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 *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 2013DFR30640), 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

Bastin, D. A. & Reeves, P. R. (1995). Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. *Gene* **164**, 17–23.

Bock, K. & Pedersen, C. (1983). Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Adv Carbohydr Chem Biochem* **41**, 27–66.

Conrad, H. E. (1972). Methylation of carbohydrates with methyl iodide in dimethyl sulfoxide in the presence of methylsulfinyl-anion. *Methods Carbohydr Chem* **6**, 361–364.

Feng, L., Senchenkova, S. N., Yang, J., Shashkov, A. S., Tao, J., Guo, H., Zhao, G., Knirel, Y. A., Reeves, P. & Wang, L. (2004). Structural and genetic characterization of the *Shigella boydii* type 13 O antigen. *J Bacteriol* **186**, 383–392.

Holst, O. (1999). Chemical structure of the core region of lipopolysaccharides. In *Endotoxin in Health and Disease*, pp. 115–154. Edited by H. Brade, S. M. Opal, S. N. Vogel & D. C. Morrison. New York: Marcel Dekker.

Karlyshev, A. V., Champion, O. L., Churcher, C., Brisson, J. R., Jarrell, H. C., Gilbert, M., Brochu, D., St Michael, F., Li, J. & other authors (2005). Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. *Mol Microbiol* **55**, 90–103.

L'vov, V. L., Tokhtamysheva, N. V., Shashkov, A. S., Dmitriev, B. A. & Kochetkov, N. K. (1983). [Bacterial antigenic polysaccharides. 12. Structure and ^{13}C NMR spectrum of the polysaccharide chain of *Shigella boydii* type 8 lipopolysaccharide]. *Bioorg Khim* **9**, 60–73 (in Russian).

Landersjö, C., Weintraub, A. & Widmalm, G. (1996). Structure determination of the O-antigen polysaccharide from the enteroinvasive *Escherichia coli* (EIEC) O143 by component analysis and NMR spectroscopy. *Carbohydr Res* **291**, 209–216.

Leontein, K. & Lönngren, J. (1993). Determination of the absolute configuration of sugars by gas-liquid chromatography of their acetylated 2-octyl glycosides. *Methods Carbohydr Chem* **9**, 87–89.

Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V. & Kochetkov, N. K. (1988). A computer-assisted structural analysis of regular polysaccharides on the basis of ^{13}C-n.m.r. data. *Carbohydr Res* **175**, 59–75.

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R. & Wang, L. (2008). Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev* **32**, 627–653.

Orskov, F. & Orskov, I. (1984). Serotyping of *Escherichia coli*. *Methods Microbiol* **14**, 43–112.

Reeves, P. P. & Wang, L. (2002). Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* **264**, 109–135.

Sawardeker, J. S., Sloneker, J. H. & Jeanes, A. (1965). Quantitative determination of monosaccharides as their alditol acetates by gas liquid chromatography. *Anal Chem* **37**, 1602–1604.

Shashkov, A. S., Lipkind, G. M., Knirel, Y. A. & Kochetkov, N. K. (1988). Stereochemical factors determining the effects of glycosylation on the ^{13}C chemical shifts in carbohydrates. *Magn Reson Chem* **26**, 735–747.

Stenutz, R., Weintraub, A. & Widmalm, G. (2006). The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev* **30**, 382–403.

Westphal, O. & Jann, K. (1965). Bacterial lipopolysaccharides. Extraction with phenol-water and further applications of the procedure. *Methods Carbohydr Chem* **5**, 83–91.

Edited by: V. Eijsink