

Note

## Structure of a teichoic acid-like O-polysaccharide of *Escherichia coli* O29

Andrei V. Perepelov,<sup>a,\*</sup> Quan Wang,<sup>b,c</sup> Sof'ya N. Senchenkova,<sup>a</sup> Sergei D. Shevelev,<sup>a</sup> Guang Zhao,<sup>b,c</sup> Alexander S. Shashkov,<sup>a</sup> Lu Feng,<sup>b,c</sup> Yuriy A. Knirel<sup>a</sup> and Lei Wang<sup>b,c</sup>

<sup>a</sup>*N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation*

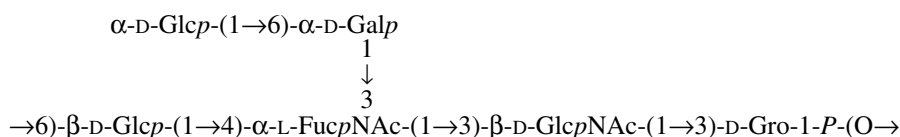
<sup>b</sup>*TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*

<sup>c</sup>*Tianjin Key Laboratory for Microbial Functional Genomics, TEDA College, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*

Received 10 April 2006; received in revised form 12 May 2006; accepted 17 May 2006

Available online 19 June 2006

**Abstract**—A teichoic acid-like O-polysaccharide was isolated by mild acid degradation of the lipopolysaccharide (LPS) of *Escherichia coli* O29. The O-polysaccharide and an oligosaccharide obtained by dephosphorylation of the O-polysaccharide were studied by sugar analysis along with <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy. The following structure of the branched oligosaccharide repeating unit, containing five monosaccharide residues and glycerol 1-phosphate (D-Gro-1-P), was established:



© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** *Escherichia coli*; O-Polysaccharide; O-Antigen; Bacterial polysaccharide structure; Glycerol phosphate; NMR

*Escherichia coli* is the predominant species in the human intestinal microflora and one of the most common causes of diarrhoeal diseases. *E. coli* strains are designated as O:K:H serotypes, where O is the O-antigen. It represents the polysaccharide chain (O-polysaccharide, OPS) of the lipopolysaccharide, which is usually built up of repeating units (O-units) containing two to seven sugar residues and often also non-sugar substituents (e.g., amino acids, pyruvic acid acetals, lactic acid ethers, phosphate, O-acetyl groups etc.). Each strain expresses only a particular O-antigen form, and the var-

iation of the O-antigen structure is thought to offer a selective advantage in the niche occupied. The *E. coli* species has 166 O-antigen forms, and the structures of only half of them have been elucidated. Now we present a new structure of a teichoic acid-like O-polysaccharide of *E. coli* O29:H10. For many years, the O29 serogroup was not characterized to be associated with any specific infections, but in 1979 some O29 strains were reported to be invasive<sup>1</sup> and able to produce a heat-stable enterotoxin.<sup>2</sup>

The OPS was obtained by mild acid degradation of the lipopolysaccharide isolated from dried cells of *E. coli* O29 by the phenol–water procedure.<sup>3</sup> It was separated from low-molecular-mass compounds, including

\* Corresponding author. Tel./fax: +7 495 1376148; e-mail: [perepel@ioc.ac.ru](mailto:perepel@ioc.ac.ru)

a core oligosaccharide, by GPC on Sephadex G-50. Sugar analysis after full acid hydrolysis of the OPS revealed Glc, Gal, GlcN, and FucN in the ratios 1.1:1.0:0.9:0.7, respectively. GLC analyses of the acetylated (*S*)-2-(+)-octyl glycosides demonstrated the *D* configuration of Glc, Gal, and GlcN and the *L* configuration of FucN.

The  $^{31}\text{P}$  NMR spectrum of the OPS showed a signal for one monophosphate group at  $\delta$  1.43. The OPS was dephosphorylated with 48% HF and the resultant oligosaccharide (OS) was isolated by GPC on TSK HW-40. Sugar analysis showed the 1.8:1.0:0.7:0.6 ratios of Glc, Gal, GlcN, and FucN, that is a significant increase of the glucose content in the hydrolysate compared with the OPS data. Therefore, it was suggested that the O-unit includes two glucose residues, one of which is phosphorylated.

The  $^{13}\text{C}$  NMR spectrum of the OS (Fig. 1A) showed the presence of five monosaccharide residues and one glycerol residue, as followed from the total number of signals, including those for five anomeric carbons at  $\delta$  98.3–104.3, one  $\text{CH}_3\text{--C}$  group (C-6 of FucN) at  $\delta$  16.9, and six  $\text{OCH}_2\text{--C}$  groups (C-6 of four monosaccharides, C-1 and C-3 of glycerol) at 62.0–63.9 and 72.3 (data of a DEPT experiment), 19 oxygen-bearing sugar ring carbons and C-2 of glycerol in the region  $\delta$  69.4–80.0, two nitrogen-bearing carbons (C-2 of GlcN and FucN) at  $\delta$  50.5 and 56.9, and two *N*-acetyl groups at  $\delta$  23.4, 23.7 (both  $\text{CH}_3$ ) and 175.5 (2CO). Accordingly, the  $^1\text{H}$  NMR spectrum of the OS contained signals for five anomeric protons at  $\delta$  4.46–5.11, one  $\text{CH}_3\text{--C}$  group (H-6 of FucN) at  $\delta$  1.31, two *N*-acetyl groups at  $\delta$  1.98 and 2.02, and other signals at  $\delta$  3.42–4.50. These

data indicate that the OS is a pentasaccharide containing two residues of Glc (Glc<sup>I</sup> and Glc<sup>II</sup>) and one residue each of Gal, GlcNAc, FucNAc, and glycerol (Gro).

The  $^1\text{H}$  NMR spectrum of the OS was assigned using 2D COSY, TOCSY, and ROESY experiments (Table 1). The TOCSY spectrum demonstrated correlations of H-1 with H-2–H-6 for the residues with the *gluco* configuration (Glc<sup>I</sup>, Glc<sup>II</sup>, and GlcNAc) and H-1 with H-2–H-4 for those with the *galacto* configuration (Gal and FucNAc). The remaining signals of the latter were assigned by correlations of H-4 with H-5 and H-6a for Gal and H-6 with H-4 and H-5 for FucNAc in the ROESY spectrum.

With the  $^1\text{H}$  NMR spectrum interpreted, all signals in the  $^{13}\text{C}$  NMR spectrum of the OS were assigned using H-detected 2D  $^1\text{H}$ ,  $^{13}\text{C}$  HSQC and HMQC-TOCSY experiments (Table 1). The GlcNAc and FucNAc residues were confirmed by correlations between protons at the nitrogen-linked carbons with the corresponding carbons at  $\delta$  4.45/50.5 and 3.87/56.9. Based on the remaining carbon signals, which belonged to glycerol, the proton signals of glycerol were assigned by H-1a,1b/C-1, H-2/C-2, and H-3a,3b/C-3 correlations in the HSQC spectrum (Table 1).

Relatively small  $J_{1,2}$  coupling constant values of  $\sim 3$  Hz indicated that Glc<sup>I</sup>, Gal, and FucNAc are  $\alpha$ -linked, whereas significantly larger  $J_{1,2}$  values of 7–8 Hz showed that Glc<sup>II</sup> and GlcNAc are  $\beta$ -linked. The pyranose form of all monosaccharide residues was inferred by the absence from the  $^{13}\text{C}$  NMR spectrum of any signals for non-anomeric sugar carbons at a field lower than  $\delta$  80.<sup>4</sup>

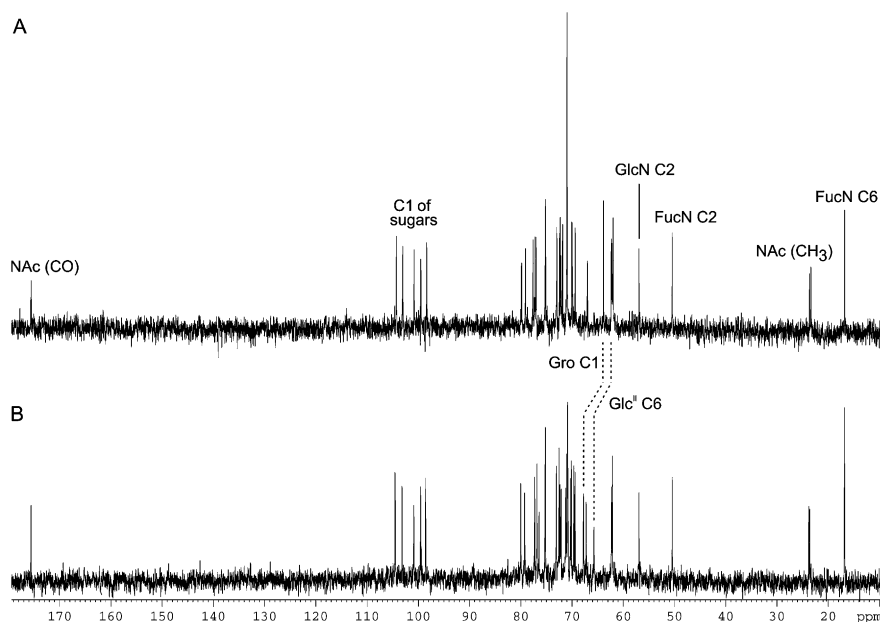


Figure 1.  $^{13}\text{C}$  NMR spectra of the OS (A) and OPS (B) from *E. coli* O29.

**Table 1.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data of the OS and OPS from *E. coli* O29 ( $\delta$ , ppm)

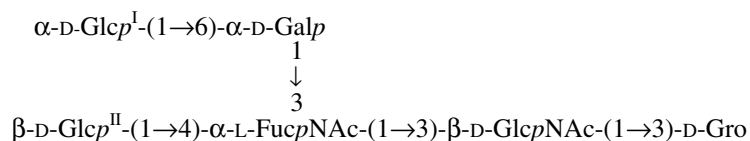
Sugar residue	Nucleus	1 (1a,1b)	2	3 (3a,3b)	4	5	6 (6a,6b)
<i>OS</i>							
$\alpha\text{-D-Glcp}^{\text{I}}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.92	3.58	3.55	3.45	3.80	3.76, 3.87
	$^{13}\text{C}$	98.3	72.5	75.8	71.0	73.0	62.0
$\rightarrow\text{6-}\alpha\text{-D-Galp}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	5.11	3.69	3.89	4.01	4.35	3.56, 3.89
	$^{13}\text{C}$	100.8	70.0	71.0	71.0	71.0	67.0
$\beta\text{-D-Glcp}^{\text{II}}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.63	3.42	3.51	3.45	3.48	3.76, 3.87
	$^{13}\text{C}$	104.3	75.2	77.1	71.0	77.3	62.4
$\rightarrow\text{3,4-}\alpha\text{-L-FucpNAc}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	5.00	4.45	4.06	4.24	4.50	1.31
	$^{13}\text{C}$	99.5	50.5	72.2	79.1	69.4	16.9
$\rightarrow\text{3-}\beta\text{-D-GlcpNAc}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.46	3.87	3.69	3.65	3.43	3.67, 3.94
	$^{13}\text{C}$	103.0	56.9	80.0	70.1	77.6	62.3
$\rightarrow\text{3-}\text{D-Gro}$	$^1\text{H}$	3.51, 3.60	3.83	3.60, 3.93			
	$^{13}\text{C}$	63.9	71.9	72.3			
<i>OPS</i>							
$\alpha\text{-D-Glcp}^{\text{I}}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.93	3.58	3.57	3.45	3.76	3.76, 3.86
	$^{13}\text{C}$	98.5	72.5	75.2	71.0	73.0	62.1
$\rightarrow\text{6-}\alpha\text{-D-Galp}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	5.09	3.69	3.89	4.09	4.32	3.65, 3.89
	$^{13}\text{C}$	100.8	69.6	70.8	71.2	71.0	67.2
$\rightarrow\text{6-}\beta\text{-D-Glcp}^{\text{II}}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.64	3.42	3.52	3.51	3.49	4.02, 4.09
	$^{13}\text{C}$	104.5	75.2	76.7	70.1	76.4	65.7
$\rightarrow\text{3,4-}\alpha\text{-L-FucpNAc}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.98	4.45	4.06	4.22	4.50	1.31
	$^{13}\text{C}$	99.5	50.4	72.1	79.2	69.4	16.8
$\rightarrow\text{3-}\beta\text{-D-GlcpNAc}\text{-(1}\rightarrow\text{)}$	$^1\text{H}$	4.49	3.87	3.69	3.65	3.47	3.76, 3.94
	$^{13}\text{C}$	103.1	56.9	79.9	70.1	77.3	62.3
$\rightarrow\text{3-}\text{D-Gro-1-}\text{P-(O}\rightarrow\text{)}$	$^1\text{H}$	3.85, 3.90	3.98	3.64, 3.97			
	$^{13}\text{C}$	67.7	70.8	72.3			

The chemical shifts for the *N*-acetyl groups are  $\delta_{\text{H}}$  1.98 and 2.02;  $\delta_{\text{C}}$  23.4, 23.7 (both Me), and 175.5 (2CO).

Relatively low-field positions of the signals for C-6 of Gal, C-3 of GlcNAc, C-3 and C-4 of FucNAc, and C-1 (C-3) of Gro at  $\delta$  67.0, 80.0, 72.2, 79.1, and 72.3, respectively, in the  $^{13}\text{C}$  NMR spectrum of the OS, as compared with their positions in the corresponding non-substituted monosaccharides and glycerol,<sup>4,5</sup> demonstrated the modes of monosaccharide glycosylation. Particularly, these data showed that OS is branched with the FucNAc residue locating at the branching point. The chemical shifts of C-2–C-5 of Glc<sup>I</sup> and Glc<sup>II</sup> in

glycosylation pattern revealed by the  $^{13}\text{C}$  NMR chemical shift data and established the sequence of the monosaccharide residues in the OS.

The absolute configuration of glycerol was determined using the 2,2,6,6-tetramethyl-1-piperidinyloxy radical (TEMPO)-mediated oxidation of OS,<sup>6</sup> which resulted in *D*-glyceric acid identified by GLC of the acetylated (*S*)-2-octyl ester. Therefore, the OS contains a 3-substituted *D*-glycerol residue and, hence, has the following structure:



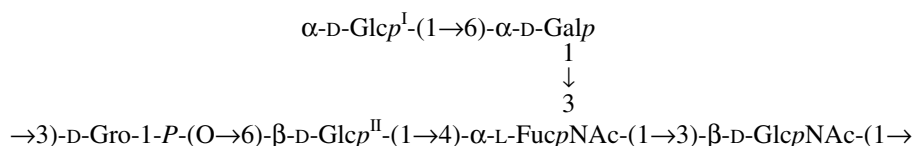
the OS were essentially the same as in free  $\alpha$ - and  $\beta$ -glucopyranoses, respectively,<sup>5</sup> thus indicating the terminal position of these residues at the non-reducing ends.

A 2D ROESY experiment revealed interresidue cross-peaks between the anomeric protons and protons at the linkage carbons at  $\delta$  4.92/3.56, 3.89; 5.11/4.06; 5.00/3.69; 4.46/3.60, 3.93; and 4.63/4.24, which were assigned to Glc<sup>I</sup> H-1, Gal H-6a,6b; Gal H-1, FucNAc H-3; FucNAc H-1, GlcNAc H-3; GlcNAc H-1, Gro H-1a,1b(H-3a,3b); and Glc<sup>II</sup> H-1, FucNAc H-4 correlations, respectively (Table 1). These data are in agreement with the

In order to determine the mode of interlinkage of the O-units, the  $^1\text{H}$  and  $^{13}\text{C}$  (Fig. 1B) NMR spectra of the OPS were assigned using 2D COSY, TOCSY, and  $^1\text{H}$ ,  $^{13}\text{C}$  HSQC experiments as described above for the OS, and the assignments are given in Table 1. The  $^1\text{H}$ ,  $^{31}\text{P}$  HMQC spectrum of the OPS showed a correlation of the phosphate group with H-6a,6b of Glc<sup>II</sup> at  $\delta$  1.43/4.02, 4.09 and H-1a,1b of Gro at  $\delta$  1.43/3.85, 3.90. Downfield displacements of the signals for C-6 of Glc<sup>II</sup> and C-1 of Gro from  $\delta$  62.4 and 63.9 in the  $^{13}\text{C}$  NMR spectrum of OS to  $\delta$  65.7 and 67.7 in the spectrum of the OPS were in agreement with phosphorylation of

these residues at positions 6 and 1, respectively. The monosaccharide sequence in the O-unit was further confirmed by 2D  $^1\text{H}$ ,  $^1\text{H}$  ROESY and  $^1\text{H}$ ,  $^{13}\text{C}$  HMBC experiments with the OPS (data not shown).

On the basis of the data obtained, it was concluded that the O-polysaccharide of *E. coli* O29 has the following structure:



Biosynthesis of O-antigens is initiated by transfer of the first monosaccharide of the O-unit to a lipid carrier.<sup>7</sup> In all reported cases (e.g., Refs. 8 and 9), when present, a 2-acetamido-2-deoxy-D-hexose is the first monosaccharide of trisaccharide and higher O-units of heteropolysaccharides. Therefore, most likely, D-GlcNAc is the first sugar residue in the *E. coli* O29 O-unit and the structure shown above represents the true, so-called biological repeating unit.

The *E. coli* O29 O-polysaccharide structure is remarkably similar to that of *Proteus vulgaris* O12,<sup>10</sup> the only difference being the presence of a 4-O-acetylated  $\alpha\text{-D-GalpNAc}$  residue in the latter instead of the  $\alpha\text{-D-Galp}$  residue in the former.

## 1. Experimental

### 1.1. Bacterial strain and isolation of the lipopolysaccharide

*E. coli* O29 type strain was obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia (IMVS). Bacteria were grown to late log phase in 8 L of LB using a 10-L fermentor (BIOSTAT C-10, B. Braun Biotech International, Germany) under constant aeration at 37 °C and pH 7.0. Bacterial cells were washed and dried as described.<sup>11</sup>

The lipopolysaccharide (980 mg) was isolated from dried cells (13.4 g) by the phenol–water method<sup>3</sup> and purified by precipitation of nucleic acids and proteins with aq 50% trichloroacetic acid.

### 1.2. Degradation of the lipopolysaccharide

Delipidation of the lipopolysaccharide (104 mg) was performed with aq 2% HOAc (6 mL) at 100 °C until precipitation of lipid A. The precipitate was removed by centrifugation (13,000g, 20 min), and the supernatant fractionated by GPC on a column (56 × 2.6 cm) of Sephadex G-50 (S) (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer, pH 4.5, monitored

by a differential refractometer (Knauer, Germany). A high-molecular-mass OPS was obtained in a yield of 45% of the LPS weight.

The OPS (30 mg) was dephosphorylated with aq 48% HF (0.2 mL, 4 °C, 20 h), and the OS (14 mg) was isolated by GPC on a column (80 × 2.5 cm) of TSK HW-40 (S) in water monitored as above.

### 1.3. Chemical analyses

The OS and OPS were hydrolyzed with 2 M  $\text{CF}_3\text{CO}_2\text{H}$  (120 °C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on a Hewlett-Packard 5890 chromatograph (USA) equipped with an Ultra-1 column using a temperature gradient of 150–290 °C at 5 °C min<sup>−1</sup>. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides as described.<sup>12,13</sup> The absolute configuration of glycerol was determined by oxidation of the OS as described<sup>6</sup> followed by identification of D-glyceric acid by GLC of the acetylated (*S*)-2-octyl ester using a temperature program of 150–230 °C at 5 °C min<sup>−1</sup>.

### 1.4. NMR spectroscopy

Samples were deuterium-exchanged by freeze-drying twice from D<sub>2</sub>O and then examined as solutions in 99.96% D<sub>2</sub>O at 30 °C. NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using internal acetone ( $\delta_{\text{H}}$  2.225,  $\delta_{\text{C}}$  31.45) and external aq 85% H<sub>3</sub>PO<sub>4</sub> ( $\delta_{\text{P}}$  0) as references. 2D NMR spectra were obtained using standard Bruker software, and Bruker xwinnmr 2.6 program was used to acquire and process the NMR data. Mixing times of 200 and 100 ms were used in TOCSY and ROESY experiments, respectively.

### Acknowledgments

This work was supported by the Russian Foundation for Basic Research (05-04-48992 to A.V.P. and 05-04-39015 to Y.A.K.), the Russian Science Support Foundation, the Council on Grants at the President of the Russian Federation for Support of Young Russian Scientists (MK-1597.2005.3 to A.V.P.), NSFC Program (30370023, 30370339, 30530010), and funds from the Science and Technology Committee of Tianjin City to L.W. and L.F.

## References

1. Toledo, M. R. F.; Reis, M. H. L.; Almeida, R. G.; Trabulsi, L. R. *J. Clin. Microbiol.* **1979**, *9*, 288–289.
2. Merson, M. H.; Orskov, F.; Orskov, I.; Sack, R. B.; Huq, I.; Koster, F. T. *Infect. Immun.* **1979**, *23*, 325–329.
3. Westphal, O.; Jann, K. *Methods Carbohydr. Chem.* **1965**, *5*, 83–91.
4. Bock, K.; Pedersen, C. *Adv. Carbohydr. Chem. Biochem.* **1983**, *41*, 27–66.
5. Lipkind, G. M.; Shashkov, A. S.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K. *Carbohydr. Res.* **1988**, *175*, 59–75.
6. Rundlöf, T.; Widmalm, G. *Anal. Biochem.* **1996**, *243*, 228–233.
7. Raetz, C. R. H.; Whitfield, C. *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
8. Katzenellenbogen, E.; Kocharova, N. A.; Zatonsky, G. V.; Shashkov, A. S.; Bogulska, M.; Knirel, Y. A. *FEMS Immunol. Med. Microbiol.* **2005**, *45*, 269–278.
9. Bystrova, O. V.; Knirel, Y. A.; Lindner, B.; Kocharova, N. A.; Zahringer, U.; Pier, G. B. *FEMS Immunol. Med. Microbiol.* **2006**, *46*, 85–99.
10. Perepelov, A. V.; Torzewska, A.; Shashkov, A. S.; Senchenkova, S. N.; Rozalski, A.; Knirel, Y. A. *Eur. J. Biochem.* **2000**, *267*, 788–793.
11. Robbins, P. W.; Uchida, T. *Biochemistry* **1962**, *1*, 323–335.
12. Leontin, K.; Lindberg, B.; Lönngren, J. *Carbohydr. Res.* **1978**, *62*, 359–362.
13. Gerwig, G. J.; Kamerling, J. P.; Vliegthart, J. F. G. *Carbohydr. Res.* **1979**, *77*, 1–7.