Escherichia coli O106, a new member of a group of enteric bacteria sharing an O-polysaccharide backbone structure

A. S. Shashkov,[a] S. N. Senchenkova,[a] O. I. Naumenko,[a,b] N. A. Kalinchuk,[a] A. V. Perepelov,[a]★ and Yu. A. Knirel[a]

[a]N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 47 Leninsky prosp., 119991 Moscow, Russian Federation. Fax: +7 (499) 137 6148. E-mail: andreivperepelov@gmail.com
[b]Higher Chemical College of the Russian Academy of Sciences, D. I. Mendeleev University of Chemical Technology of Russia, 9 Miusskaya pl., 125047 Moscow, Russian Federation

O-Polysaccharides (O-antigens) of a number of genetically related *Escherichia coli* O-serogroups (O17, O44, O73, O77, and O106) and *Salmonella enterica* O:6,14 possess an identical main chain composed of D-GlcNAc and D-Man residues and differ from each other by the absence or presence of glucose side chains at various positions. Using two-dimensional NMR spectroscopy, we established the structure of the O-polysaccharide of *E. coli* O106 having two glucose side chains in a hexasaccharide repeating unit.

Key words: *Escherichia coli*, O-polysaccharide, O-antigen, bacterial polysaccharide structure, glucosylation.

Escherichia coli is the predominant facultative anaerobe of the colonic flora of many mammals, including humans, and has both commensal and pathogenic forms. An O-polysaccharide (O-antigen) consisting of many oligosaccharide repeats (O-units) is a part of the lipopolysaccharide on the cell surface of Gram-negative bacteria. The O-antigen is one of most structurally variable cell constituents providing the basis for serotyping bacteria.

A number of *Escherichia coli* O-serogroups (O17, O44, O73, O77, and O106) and *Salmonella enterica* serogroup O:6,14 (H) called the O77 group[1] possess almost identical O-antigen biosynthesis gene clusters.[1—3] A comparison of sequences of *E. coli* O77 and *S. enterica* O:6,14 indicates the origin of the O-antigen gene clusters from a common ancestor followed by a divergence as the species diverged, rather than an independent acquisition or acquisition by one species from another.[1]

The O-polysaccharide of all bacteria of the O77 group share a main tetrasaccaride chain composed of one D-GlcNAc and three D-Man residues.[1,4—7] Biosynthesis of the main chain is driven by the O-antigen gene cluster located between conserved genes *galF* and *gnd* on the chromosome.[8] The O-units differ from each other in the absence (O77) or the presence of one (O17, O44, O:6,14) or two (O73) glucose side chains at various positions. These variations are due to glucosylation of the main chain encoded by genes located elsewhere in the genome, most likely, within putative prophages as demonstrated by a mutation test with *E. coli* O44.[1]

O-Polysaccharide structures have been elucidated for all bacteria of the O77 group[1,4—7] except for *E. coli* O106. In present work, we established the structure of this O-polysaccharide.

Experimental

Gel-permeation chromatography was carried out on a column (56×3.5 cm) packed with Sephadex G-50 Superfine in 0.1% AcOH; elution was monitored using a differential refractometer (Knauer, Germany). GLC analysis was performed on an Agilent 7820 chromatograph (Interlab, Russia) equipped with an HP-5 column (0.32 mm×30 m) (Agilent) using a temperature program of 160 °C (1 min) to 290 °C at a heating rate of 7 °C min^{-1}. NMR spectra were recorded on a Bruker Avance II instrument (Germany) at 20 °C. The chemical shifts are given in the δ scale relative to an internal standard sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (Sigma—Aldrich; δ_H 0, δ_C −1.6). Two-dimensional NMR spectra were obtained using standard Bruker software, Bruker TopSpin 2.1 program was employed to process the data. In TOCSY and ROESY experiments a mixing time of 150 ms were used.

The lipopolysaccharide was isolated from cells of *E. coli* O106 by phenol-water extraction[9] and purified as described earlier.[10] The O-polysaccharide was obtained by mild acid degradation of the lipopolysaccharide (2% AcOH, 100 °C, 2 h) followed by Sephadex G-50 gel-permeation chromatography.

Results and Discussion

Monosaccharide analysis by GLC of the alditol acetates[11] derived after acid hydrolysis of the O-polysaccharide

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 8, pp. 1538—1541, August, 2018.
1066-5285/18/6708-1538

Table 1. ^{1}H (600 MHz) and ^{13}C (125 MHz) NMR chemical shifts (δ) of the O-polysaccharide from *E. coli* O106 measured 99.95% D_2O at 55 °C*

Monosaccharide residue	Nucleus	1	2	3	4	5	6 (6a, 6b)
→3,6)-α-D-Glc*p*NAc-(1→	^{1}H	4.85	4.11	3.93	3.65	3.92	3.78, 4.00
(**A**)	^{13}C	98.6	54.2	81.9	69.7	71.7	67.0
→2,4)-β-D-Man*p*-(1→	^{1}H	4.78	3.98	3.98	3.83	3.54	3.74, 3.85
(**B**)	^{13}C	101.2	76.6	75.7	77.0	76.6	62.3
→2)-α-D-Man*p*-(1→	^{1}H	5.33	4.09	4.02	3.76	3.96	3.80, 3.95
(**C**)	^{13}C	100.6	79.0	71.5	68.1	73.2	62.2
→6)-α-D-Man*p*-(1→	^{1}H	5.07	4.11	3.85	3.97	3.78	3.53, 4.13
(**D**)	^{13}C	103.8	71.3	72.1	67.3	72.9	66.5
α-D-Glc*p*-(1→	^{1}H	4.97	3.56	3.73	3.43	3.71	3.77, 3.85
(**E**)	^{13}C	99.1	72.9	74.5	71.0	73.2	62.0
α-D-Glc*p*-(1→	^{1}H	5.29	3.55	3.67	3.40	3.72	3.75, 3.85
(**F**)	^{13}C	101.4	73.2	74.3	70.9	74.1	62.1

* Chemical shifts for the *N*-acetyl group are δ_H 2.05, δ_C 23.4 (Me) and 175.0 (CO).

(2 *M* CF_3CO_2H, 120 °C, 2 h) revealed Glc, Man, and GlcNAc in the ratio ~2 : 3 : 0.6 (detector response). The ^{1}H and ^{13}C NMR spectra of the O-polysaccharide contained the signals for six anomeric atoms at δ_H 4.78—5.33 and δ_C 98.6—103.8 (Table 1), and, consequently, the O-polysaccharide has a hexasaccharide O-unit. This O-unit includes one GlcNAc residue, which was confirmed by the signals for a nitrogen-linked carbon atom C(2) at δ_C 54.2 and an *N*-acetyl group at δ_H 2.05, δ_C 23.4 (Me) and 175.0 (CO).

The NMR spectra were assigned using 2D experiments including ^{1}H—^{1}H COSY, ^{1}H—^{1}H TOCSY, and ^{1}H—^{13}C

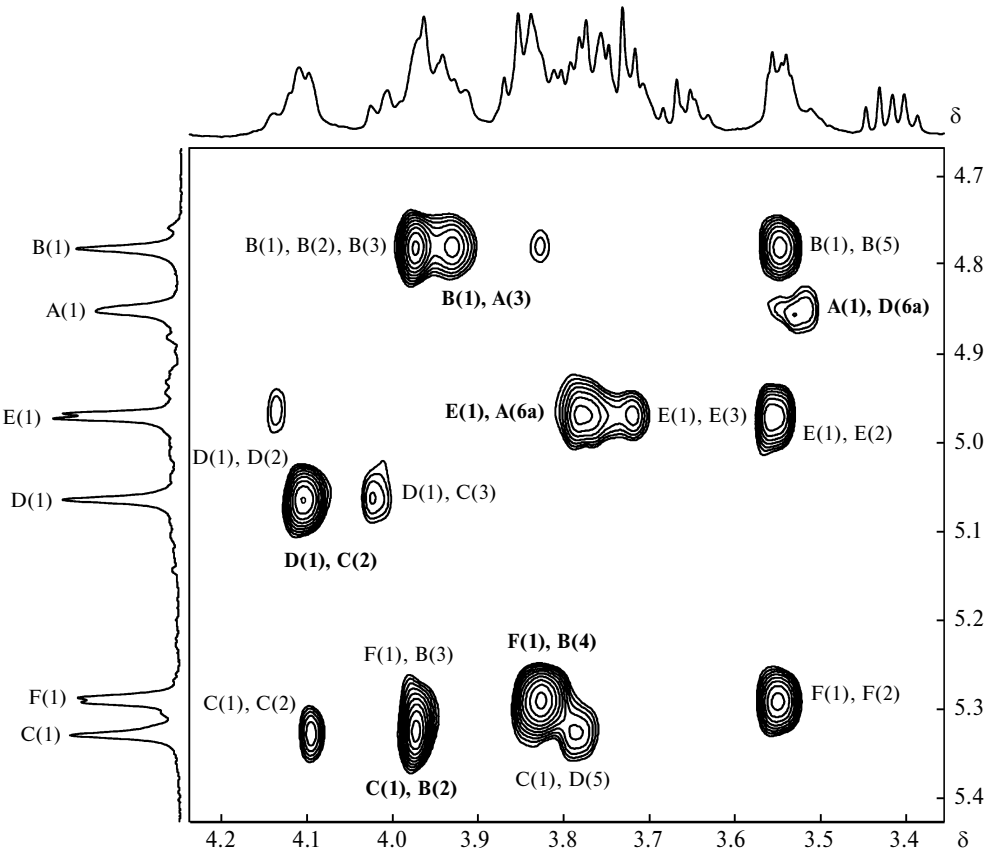

Fig. 1. Part of a 2D ^{1}H—^{1}H ROESY spectrum of the O-polysaccharide from *E. coli* O106. The corresponding parts of the ^{1}H NMR spectrum are displayed along the axes. Numbers refer to protons in sugar residues denoted by letters as shown in Table 1. Transglycosidic cross-peaks are annotated in bold face.

Table 2. Correlations for H(1) in the 2D ^{1}H—^{1}H ROESY spectrum of the O-polysaccharide from *E. coli* O106

Anomeric proton in monosaccharide residue (δ)	Correlations to protons in monosaccharide residues (δ)
A (4.85)	**D** H(6a) (3.53, weak)
B (4.78)	**A** H(3) (3.93), **B** H(2) (3.98), H(3) (3.98), H(4) (3.83, weak), H(5) (3.54)
C (5.33)	**B** H(2) (3.98), **C** H(2) (4.09, weak), **D** H(5) (3.78)
D (5.07)	**C** H(2) (4.09), H(3) (4.02, weak), **D** H(2) (4.11)
E (4.97)	**A** H(6a) (3.78), **E** H(2) (3.56), H(3) (3.73)
F (5.29)	**B** H(4) (3.83), H(3) (3.98), **F** H(2) (3.55)

HSQC (see Table 1). Based on characteristic $^3J_{H,H}$ coupling constants, spin systems of α-Glc*p*NAc (unit **A**), three Man residues (units **B**—**D**), and two α-Glc*p* residues (units **E** and **F**) were identified. The α configuration of units **C** and **D** and the β configuration of unit **B** were inferred by the C(5) chemical shifts of δ 73.2, 72.9, and 76.6, respectively (*cf.* Ref. 12: δ 73.7 and 77.4 for α-Man*p* and β-Man*p*, respectively).

Downfield shift of the signals for the linkage carbons including the C(2) atom of unit **C**, the C(6) atom of unit **D**, the C(2) and C(4) atoms of unit **B**, and the C(3) and C(6) atoms of unit **A** (see Table 1) as compared with their positions in the spectra of the corresponding non-substituted monosaccharides[12] revealed the substitution pattern in the O-unit. The chemical shifts of units **E** and **F** were close to those in non-substituted α-Glc*p* and, therefore, these units occupied the terminal position in the side chains. The sequence of the monosaccharide residues was determined using a 2D ^{1}H—^{1}H ROESY experiment, which showed correlations between anomeric protons and protons at the linkage carbons (Fig. 1, Table 2).

Thus, the O-polysaccharide of *E coli* O106 has the structure shown in Table 3. This structure is unique among known bacterial polysaccharide structures (see Bacterial Carbohydrate Structure Database at http://csdb.glycoscience.ru/bacterial/). As indicated above, another O-polysaccharide from the group O77 that of *E. coli* O73, also has two α-Glcp side chains but they occupy different positions in the O-unit. Table 3 shows all O-polysaccharide structures of the group O77 that are closely related structurally and genetically. Distribution of the corresponding strains into different O-serogroups in the classification scheme of *E. coli* despite that their O-polysaccharides have the same main chain is evidently due to a significant contribution into the serospecificity of the bacteria of epitopes associated with the side-chain α-Glc*p* residues.

The authors are grateful to Bin Liu and Lei Wang (Nankai University, Tianjin, China) for providing the bacterial mass.

This work was financially supported by the Russian Science Foundation (Project No. 14-14-01042_P).

References

1. W. Wang, A. V. Perepelov, L. Feng, S. D. Shevelev, Q. Wang, S. N. Senchenkova, W. Han, Y. Li, A. S. Shashkov, Y. A. Knirel, P. R. Reeves, L. Wang, *Microbiology,* 2007, **153**, 2159.
2. A. Iguchi, S. Iyoda, T. Kikuchi, Y. Ogura, K. Katsura, M. Ohnishi, T. Hayashi, N. R. Thomson, *DNA Res.*, 2014, **22**, 101.
3. R. S. Coimbra, F. Grimont, P. Lenormand, P. Burguière, L. Beutin, P. A. D. Grimont, *Res. Microbiol.,* 2000, **151**, 639.
4. H. Masoud, M. B. Perry, *Biochem. Cell Biol.*, 1996, **74**, 241.
5. M. Staaf, G. Widmalm, A. Weintraub, J. P. Nataro, *Eur. J. Biochem.*, 1995, **233**, 473.

Table 3. Related structures of the O-polysaccharides from *E. coli* and *S. enterica* group O77

O-Serogroup	O-Polysaccharide structure	Reference
Escherichia coli		
O17	→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→ α-D-Glc*p*-(1→6)⌋	4
O44	→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→ α-D-Glc*p*-(1→4)⌋	5
O73	→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→ α-D-Glc*p*-(1→4)⌋ ⌊(3←1)-α-D-Glc*p*	3
O77	→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→	6
O106	→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→ α-D-Glc*p*-(1→4)⌋ ⌊(6←1)-α-D-Glc*p*	Present work
Salmonella enterica		
O:6,14	→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→ α-D-Glc*p*-(1→3)⌋	7

6. H. Yildirim, A. Weintraub, G. Widmalm, *Carbohydr. Res.*, 2001, **333**, 179.
7. J. L. Di Fabio, J.-R. Brisson, M. B. Perry, *Carbohydr. Res.*, 1988, **179**, 233.
8. G. Samuel, P. R. Reeves, *Carbohydr. Res.*, 2003, **338**, 2503.
9. O. Westphal, K. Jann, *Methods Carbohydr. Chem.*, 1965, **5**, 83.
10. S. N. Senchenkova, G. Xu, Y. Qi, A. V. Perepelov, A. S. Shashkov, B. Liu, Y. A. Knirel, *Russ. Chem. Bull.*, 2017, **66**, 1304.
11. J. S. Sawardeker, J. H. Sloneker, A. Jeanes, *Anal. Chem.*, 1965, **37**, 1602.
12. G. M. Lipkind, A. S. Shashkov, Y. A. Knirel, E. V. Vinogradov, N. K. Kochetkov, *Carbohydr. Res.*, 1988, **175**, 59.

Received May 1, 2018;
in revised form May 5, 2018;
accepted May 22, 2018