Carbohydrate Research 349 (2012) 86–89

Contents lists available at SciVerse ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Note

Structure and gene cluster of the O-antigen of *Escherichia coli* O41

Hongfei Zhu [a], Andrei V. Perepelov [b,*], Sof'ya N. Senchenkova [b], Alexander S. Shashkov [b], Lei Wang [a], Yuriy A. Knirel [b]

[a] TEDA School of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin, China
[b] N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation

ARTICLE INFO

Article history:
Received 11 November 2011
Received in revised form 7 December 2011
Accepted 9 December 2011
Available online 17 December 2011

Keywords:
Escherichia coli
Bacterial polysaccharide structure
O-antigen
Lipopolysaccharide
O-antigen gene cluster

ABSTRACT

The acidic O-polysaccharide (O-antigen) of *Escherichia coli* O41 was studied by sugar analysis along with 1D and 2D ^{1}H and ^{13}C NMR spectroscopy, and the following structure of the branched hexasaccharide repeating unit was established:

β-D-Glc*p*A
1
↓
4
→3)-α-D-Gal*p*-(1→3)-α-L-Fuc*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-α-L-Fuc*p*-(1→3)-β-D-Glc*p*NAc-(1→

This structure is unique among the known structures of bacterial polysaccharides. The O-antigen gene cluster of *E. coli* O41 was sequenced. The gene functions were tentatively assigned by a comparison with sequences in the available databases and found to be in full agreement with the *E. coli* O41 O-polysaccharide structure.

Escherichia coli clones, both commensal and pathogenic, are normally identified by a combination of their somatic (O), flagellar (H), and sometimes capsular (K) antigens.[1] The O-antigen (O-polysaccharide) consisting of a number of oligosaccharide repeats is an essential component of the lipopolysaccharide on the cell surface of Gram-negative bacteria and one of the most variable cell constituents. Up to now, more than 180 O-antigen forms have been recognized in *E. coli*, and a number of *E. coli* O-polysaccharide structures have been elucidated.[2]

The diversity of the O-antigen forms is almost entirely due to genetic variations in the O-antigen gene cluster, which is located between *galF* and *gnd* on the chromosome of *E. coli*. The following three groups of genes are identified in the cluster: (1) nucleotide sugar biosynthesis genes; (2) sugar transferase genes; and (3) O-antigen processing genes including those for O-antigen flippase (*wzx*) and polymerase (*wzy*).[1]

In this work, we established the O-polysaccharide structure of *E. coli* O41, which had been unknown earlier, and characterized the O-antigen gene cluster of this bacterium.

Structure elucidation of the O-polysaccharide. A high-molecular mass O-polysaccharide was obtained by mild acid degradation of the lipopolysaccharide isolated from dried bacterial cells by the phenol–water procedure. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the polysaccharide revealed Fuc, Gal, and GlcN in a ratio of ~2:1:1. GLC analysis of the acetylated (*S*)-2-octyl glycosides demonstrated the D configuration of Gal, GlcN, and GlcA (see below) and the L configuration of Fuc.

The ^{13}C NMR spectrum of the polysaccharide (Fig. 1) showed signals for six anomeric carbons in the region δ 100.1–104.9, three C-CH_2OH groups (C-6 of hexoses) at δ 61.2–63.0, two C-CH_3 groups (C-6 of two Fuc residues) at δ 16.5 and 16.6, two nitrogen-bearing carbons (C-2 of two GlcN residues) at δ 56.7 and 57.7, 22 oxygen-bearing non-anomeric sugar ring carbons in the region δ 67.6–81.7, one C-CO_2H group (C-6 of GlcA, see below) at δ 174.8 and two *N*-acetyl groups at δ 23.6, 23.7 (both CH_3), 176.1 and 176.3 (both CO). Accordingly, the ^{1}H NMR spectrum contained signals for six anomeric protons at δ 4.59–5.18, other sugar protons in the region δ 3.43–4.44, two C-CH_3 groups (H-6 of two Fuc residues) at δ 1.16 and 1.27 and two *N*-acetyl groups at δ 2.00 and 2.04. Therefore, the polysaccharide has a hexasaccharide repeat (O-unit) containing one residue each of D-Gal (denoted as unit **A**) and D-GlcA (unit **F**), two residues of D-GlcNAc (units **C** and **E**) and two residues of L-Fuc (units **B** and **D**).

The ^{1}H and ^{13}C NMR spectra of the polysaccharide were assigned using 2D homonuclear $^{1}H,^{1}H$ COSY, TOCSY, ROESY, and heteronuclear $^{1}H,^{13}C$ HSQC and HMBC experiments (Table 1).

* Corresponding author. Tel.: +7 499 1376148; fax: +7 499 1355328.
E-mail address: perepel@ioc.ac.ru (A.V. Perepelov).

0008-6215/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2011.12.009

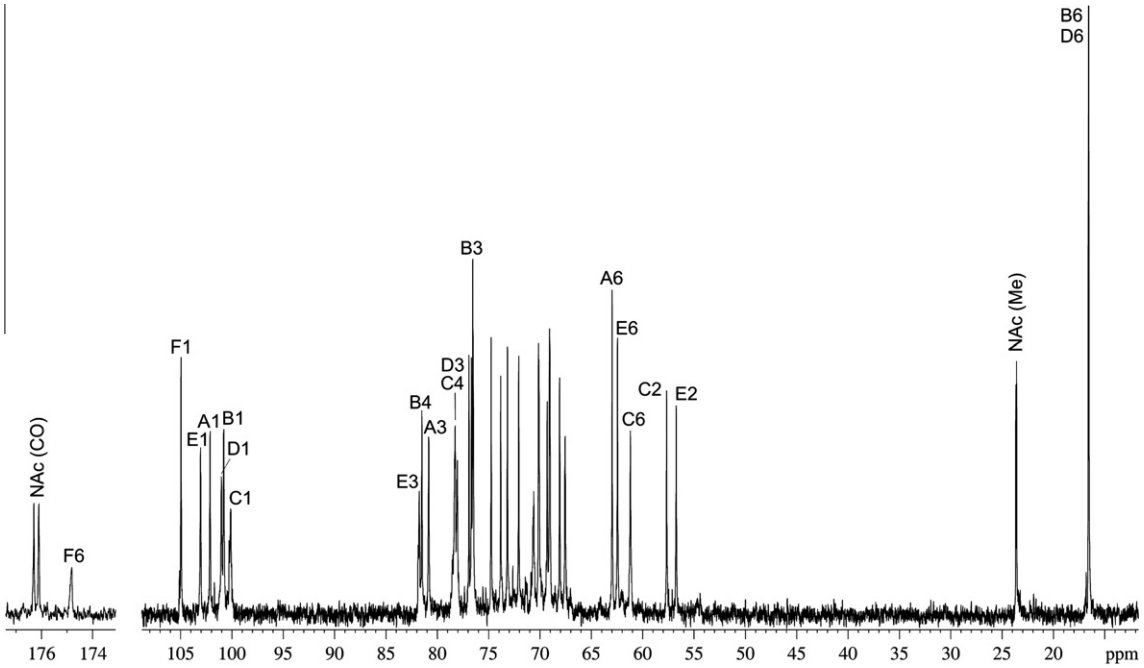

Figure 1. ^{13}C NMR spectrum of the O-polysaccharide of *E. coli* O41. Numbers refer to carbons in sugar residues denoted as shown in Table 1.

Table 1
^{1}H and ^{13}C NMR chemical shifts (δ, ppm) of the O-polysaccharide of *E. coli* O41

Sugar residue	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 (6a, 6b) C-6	NAc	
							CH_3	CO
→3)-α-D-Gal*p*-(1→	5.18	3.81	3.85	4.11	4.18	3.66; 3.71		
A	102.1	69.1	80.8	79.2	72.1	63.0		
→3,4)-α-L-Fuc*p*-(1→	5.00	4.04	3.98	4.15	4.44	1.27		
B	100.8	69.3	76.5	81.5	69.0	16.6		
→4)-β-D-Glc*p*NAc-(1→	4.68	3.74	3.68	3.60	3.54	3.84; 3.99	2.04	
C	100.1	57.7	73.8	78.2	76.7	61.2	23.6	176.1
→3)-α-L-Fuc*p*-(1→	5.04	3.80	3.99	3.88	4.29	1.16		
D	101.0	67.6	78.2	70.6	68.1	16.5		
→3)-β-D-Glc*p*NAc-(1→	4.85	3.89	3.69	3.50	3.64	3.75; 3.97	2.00	
E	103.0	56.7	81.7	70.2	76.9	62.5	23.7	176.3
β-D-Glc*p*A-(1→	4.59	3.43	3.54	3.63	3.77			
F	104.9	74.7	76.5	73.2	78.0	174.8		

Based on intra-residue H,H and H,C correlations and coupling constant values, spin systems were revealed for residues of **A–F**, all being in the pyranose form. A relatively large $J_{1,2}$ coupling constant value of ~7 Hz showed that units **C**, **E**, and **F** are β-linked, whereas significantly smaller values of <4 Hz indicated the α-linkage of units **A**, **B**, and **D**.

The spin systems for units **C** and **E** were distinguished by a correlation between proton at the nitrogen-bearing carbons (H-2) and the corresponding carbons (C-2) at δ 3.74/57.7 and 3.89/56.7, respectively. Unit **F** was identified as GlcA based on correlations of H-1 with H-2–H-5 in the TOCSY spectrum and H-5 with C-6 at δ 3.77/174.8 in the HMBC spectrum.

The signals for C-3 of units **A, D** and **E**, C-4 of unit **C**, and both C-3 and C-4 of unit **B** were shifted downfield as compared with their positions in the corresponding non-substituted monosaccharides.[3] These data demonstrated a branching character of the polysaccharide chain and defined the glycosylation pattern in the O-unit.

The ROESY spectrum of the polysaccharide showed the following correlations between anomeric protons and protons at the linkage carbons: **A** H-1, **B** H-3; **B** H-1, **C** H-4; **C** H-1, **D** H-3; **D** H-1, **E** H-3, **E** H-1, **A** H-3 and **F** H-1, **B** H-4 at δ 5.18/3.98; 5.00/3.60, 4.68/3.99; 5.04/3.69; 4.85/3.85, and 4.59/4.15, respectively. The monosaccharide sequence thus determined was confirmed by a heteronuclear ^{1}H,^{13}C HMBC experiment, which showed correlations between the anomeric protons and linkage carbons and vice

β-D-Glc*p*A **F**
1
↓
4
→3)-α-D-Gal*p*-(1→3)-α-L-Fuc*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-α-L-Fuc*p*-(1→3)-β-D-Glc*p*NAc-(1→
A **B** **C** **D** **E**

versa (data not shown).Therefore, the O-polysaccharide of *E. coli* O41 has the structure shown above. To our knowledge, this structure is unique among the known bacterial polysaccharide structures.

Characterization of the O-antigen gene cluster. The O-antigen gene cluster of *E. coli* O41 was found between the housekeeping genes *galF* and *gnd*. A DNA sequence of 15,358 bp was obtained, which contains 12 open reading frames (ORFs) with transcription direction from *galF* to *gnd* (Fig. 2). All genes were assigned functions based on their similarities to genes from the available databases (Table 2). The gene *wecA* that is responsible for the transfer of GlcNAc-1-phosphate to an undecaprenol phosphate (UndP) carrier to initiate the O-unit synthesis is located outside the O-antigen gene cluster.[4] Genes involved in the biosynthesis of common sugar nucleotide precursors such as UDP-D-GlcNAc and UDP-D-Gal are not in the cluster too, and the gene *ugd* responsible for the synthesis of UDP-D-GlcA is located between *gnd* and *hisI* near the O-antigen gene cluster.[5]

Proteins encoded by *orf7-10* and *orf12* share high level of identity to many known Gmd, Fcl, Gmm, ManC, and ManB proteins (53–90%), respectively. The corresponding genes including *gmd*, *fcl*, *gmm*, *manC*, and *manB* have been identified to be responsible for the biosynthesis of GDP-L-Fuc.[6,7] Therefore, *orf7-10* and *orf12* were concluded to encode the enzymes for the synthesis of GDP-L-Fuc and named accordingly.

Orf2, 3, 5, and 11 share 28–69% identities to glycosyltransferases of other origins and belonged to glycosyltransferase family 1 (Pf00534) or family 2 (PF00535). Therefore, *orf2*, *3*, *5*, and *11* were proposed to encode glycosyltransferases and named *wfcV*, *wfcW*, *wfcX*, and *wfcZ*, respectively.

Orf1 and 4 are the only two proteins with predicted membrane segments. Orf1 has 10 predicted transmembrane segments, which is a typical topological character of Wzx proteins.[8] It also shares 32% identity to the Wzx protein of *Yersinia enterocolitica* O8. Orf4 has 9 predicted transmembrane segments and a periplasmic loop of 52 amino acid residues, which is a typical topological character of Wzy proteins.[9] Therefore, *orf1* and *orf4* were assigned functions of the O-antigen flippase and polymerase genes, respectively, and named accordingly.

No function of Orf6 was revealed by searching available databases. However, as the O-antigen contains six sugars but only four possible glycosyltransferase genes were found in the O-antigen

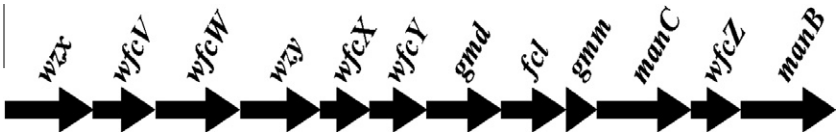

Figure 2. Schematic of the *E. coli* O41 O-antigen gene cluster.

Table 2
Characteristics of ORFs in the *E. coli* O41 O-antigen gene cluster

Orf No.	Gene name	Position of gene	G+C content (%)	Conserved domain(s)	Similar protein(s), strain(s) (Genbank Accession No.)	%Identical / %similar (No. of aa overlap)	Putative function of protein
1	*wzx*	1114..2433	32.0	Polysaccharide biosynthesis protein (PF01943) *E* value = $6 \times e^{-25}$	Wzx *Y. enterocolitica* type 0:8 (AAC60766)	32/56 (426)	O-antigen flippase
2	*wfcV*	2433..3362	30.6		Glycosyl transferase, group 1/2 family protein *G. sulfurreducens* PCA (AAR36415)	28/52 (274)	Glycosyl transferase
3	*wfcW*	3374..4630	32.8	Glycosyl transferases group 1 (PF00534) *E* value = $1.4 \times e^{-27}$	Putative glycosyltransferase *B. thetaiotaomicron* VPI-5482 (AAO78052)	33/54 (416)	Glycosyltransferase
4	*wzy*	4649..5845	28.1		NADH dehydrogenase subunit 2 *D. simulans* (AAF77291)	26/47 (215)	O-antigen polymerase
5	*wfcX*	5842..6591	28.5	Glycosyl transferase group 2 (PF00535) *E* value = $2.1 \times e^{-23}$	Putative glycosyl transferase *B. fragilis* (AAD40722)	31/53 (218)	Glycosyltransferase
6	*wfcY*	6588..7445	28.6				Glycosyltransferase
7	*gmd*	7447..8565	42.1	NAD dependent epimerase/dehydratase family (PF01370) *E* value = $4.9 \times e^{-18}$	Gmd *Y. enterocolitica* (type 0:8) (AAC60773)	84/93 (271)	GDP-mannose-4,6-dehydratase
8	*fcl*	8569..9534	33.3	NAD dependent epimerase/dehydratase family (PF01370) *E* value = $1.3 \times e^{-06}$	GDP-L-fucose synthetase *Y. pseudotuberculosis* (type O:1b) (CAB63301)	76/85 (320)	GDP-L-fucose synthetase
9	*gmm*	9537..9998	37.2	NUDIX domain (PF00293) *E* value = $2.3 \times e^{-19}$	GDP-mannose mannosyl hydrolase *S. typhimurium* (AAG24815)	53/68 (148)	GDP-mannose mannosyl hydrolase
10	*manC*	10004..11410	42.1	Nucleotidyl transferase (PF00483) *E* value = $6.9 \times e^{-127}$	GDP-mannose pyrophosphorylase *E. coli* (AAG41753)	81/91(468)	GDP-mannose pyrophosphorylase
11	*wfcZ*	11410..12154	34.3	Glycosyl transferase group 2 (PF00535) *E* value = $2.3 \times e^{-33}$	Glycosyltransferase-like protein *Y. pseudotuberculosis* (type O:1b) (CAB63303.1)	67/81 (241)	Glycosyltransferase
12	*manB*	12163..13587	38.3	Phosphoglucomutase/phospho-mannomutase (PF02879) *E* value = $2.9 \times e^{-35}$	Phosphomannomutase *S. boydii* (AAL27334)	90/94 (473)	Phosphomannomutase

gene cluster, Orf6 was proposed to be a glycosyltransferase gene and named *wfcY*.

Therefore, the functions assigned tentatively to the genes in the O-antigen gene cluster of *E. coli* O41 correspond well with the O-antigen structure established in this work.

1. Experimental

1.1. Bacterial strain and isolation of the lipopolysaccharide

E. coli O41 type strain (laboratory stock number G3080) was obtained from the Institute of Medical and Veterinary Science (Adelaide, Australia). Bacteria were grown to late log phase in 8 L of Luria–Bertani broth using a 10-L BIOSTAT C-10 fermentor (B. Braun Biotech Int., Germany) under constant aeration at 37 °C and pH 7.0. Bacterial cells were washed and dried as described.[10]

The lipopolysaccharide was isolated in a yield 6.1% from dried cells by the phenol–water method,[11] the crude extract was dialyzed without separation of the layers and freed from nucleic acids and proteins by treatment with 50% aq CCl_3CO_2H to pH 2 at 4 °C. The supernatant was dialyzed and lyophilized.

1.2. Isolation of the O-polysaccharide

Delipidation of the lipopolysaccharide (100 mg) was performed with 2% aq HOAc at 100 °C until precipitation of lipid A. The precipitate was removed by centrifugation (13,000×*g*, 20 min), and the supernatant was fractionated by GPC on a column (56 × 2.6 cm) of Sephadex G-50 Superfine (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer pH 5.5, monitored with a differential refractometer (Knauer, Germany). A high-molecular-mass polysaccharide was obtained in a yield of 33% of the lipopolysaccharide mass.

1.3. Monosaccharide analysis

A polysaccharide sample (0.5 mg) was hydrolyzed with 2 M CF_3CO_2H (120 °C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on an Agilent 7820A chromatograph (USA) equipped with an HP-5 column (0.32 mm × 30 m) and a temperature program of 160 (1 min) to 290 °C at 7 °C min^{-1}. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides as described.[12]

1.4. NMR spectroscopy

A polysaccharide sample was deuterium-exchanged by freeze-drying from 99.9% D_2O and then examined as a solution in 99.95% D_2O. NMR spectra were recorded on a Bruker Avance II 600 spectrometer (Germany) at 30 °C using internal TSP (δ_H 0) and acetone (δ_C 31.45) as references. 2D NMR spectra were obtained using standard Bruker software, and Bruker TopSpin 2.1 program was used to acquire and process the NMR data. A mixing time of 100 and 150 ms was used in TOCSY and ROESY experiments, respectively.

1.5. Sequencing and analysis of genes

Chromosomal DNA was prepared as described previously.[13] The primers #1523 and #1524 based on the housekeeping genes *galF* and *gnd*,[14] respectively, were used to amplify the O-antigen gene cluster. The PCR cycles used were as follows: denaturation at 94 °C for 10 s, annealing at 60 °C for 30 s, and extension at 68 °C for 15 min. A shotgun bank was constructed as described.[15] Sequencing was carried out using an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA), and sequence data were analyzed using computer programs as described.[16]

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (projects 11-04-91173-a and 11-04-01020-a), the National Science Foundation of China (30900255 and 31111120026), Tianjin Research Program of Application Foundation and Advanced Technology (10JCYBJC10100).

References

1. Reeves, P. P.; Wang, L. *Curr. Top. Microbiol. Immunol.* **2002**, *264*, 109–135.
2. Stenutz, R.; Weintraub, A.; Widmalm, G. *FEMS Microbiol. Rev.* **2006**, *30*, 382–403.
3. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *188*, 169–191.
4. Alexander, D. C.; Valvano, M. A. *J. Bacteriol.* **1994**, *176*, 7079–7084.
5. Samuel, G.; Reeves, P. R. *Carbohydr. Res.* **2003**, *338*, 2503–2519.
6. Frick, D. N.; Townsend, B. D.; Bessman, M. J. *J. Biol. Chem.* **1995**, *270*, 24086–24091.
7. Ginsburg, V. *J. Biol. Chem.* **1961**, *236*, 2389–2393.
8. Liu, D.; Cole, R. A.; Reeves, P. R. *J. Bacteriol.* **1996**, *178*, 2102–2107.
9. Daniels, C.; Vindurampulle, C.; Morona, R. *Mol. Microbiol.* **1998**, *28*, 1211–1222.
10. Robbins, P. W.; Uchida, T. *Biochemistry* **1962**, *1*, 323–335.
11. Westphal, O.; Jann, K. *Methods Carbohydr. Chem.* **1965**, *5*, 83–91.
12. Leontein, K.; Lönngren, J. *Methods Carbohydr. Chem.* **1993**, *9*, 87–89.
13. Bastin, D. A.; Reeves, P. R. *Gene* **1995**, *164*, 17–23.
14. Feng, L.; Wang, W.; Tao, J.; Guo, H.; Krause, G.; Beutin, L.; Wang, L. *J. Clin. Microbiol.* **2004**, *42*, 3799–3804.
15. Wang, L.; Reeves, P. R. *Infect. Immun.* **1998**, *66*, 3545–3551.
16. Feng, L.; Senchenkova, S. N.; Yang, J.; Shashkov, A. S.; Tao, J.; Guo, H.; Zhao, G.; Knirel, Y. A.; Reeves, P.; Wang, L. *J. Bacteriol.* **2004**, *182*, 383–392.