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

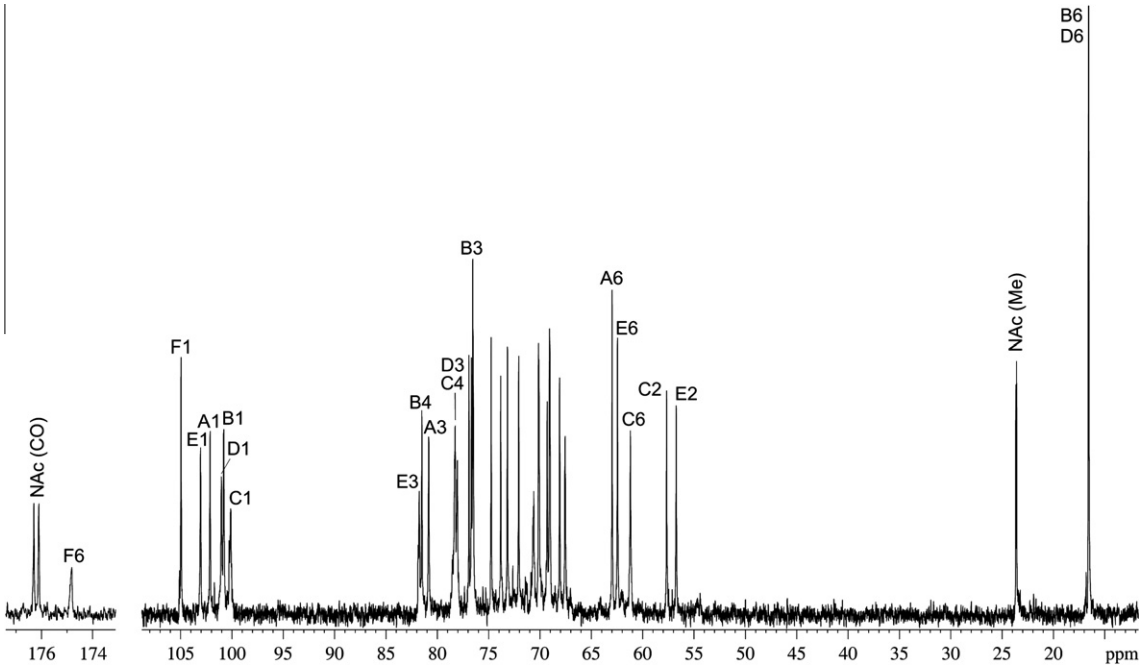

Figure 1. ^{13}C NMR spectrum of the O-polysaccharide of *E. coli* O41. Numbers refer to carbons in sugar residues denoted as shown in Table 1.

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**