Carbohydrate Research 379 (2013) 51–54

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* O154

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

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

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 7 June 2013
Accepted 8 June 2013
Available online 18 June 2013

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

ABSTRACT

The O-polysaccharide (O-antigen) of *Escherichia coli* O154 was studied by sugar analysis along with 1D and 2D ^{1}H and ^{13}C NMR spectroscopy. The following structure of the branched pentasaccharide repeating unit was established:

α-D-Man*p*NAc
1
↓
3
→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-D-Gal*p*NAc-(1→

The O-antigen gene cluster of *E. coli* O154 was sequenced. The gene functions were tentatively assigned by comparison with sequences in the available databases and found to be in full agreement with the O-polysaccharide structure.

Escherichia coli clones including both commensal and pathogenic strains are normally identified by a combination of their somatic (O), flagellar (H), and sometimes capsular (K) antigens.[1] The O-antigen or O-polysaccharide (OPS), consisting of many repeats of an oligosaccharide (O-unit), is an essential component of the lipopolysaccharide on the surface of Gram-negative bacteria and one of the most variable cell constituents. Till now, 174 O-antigen forms have been recognized for *E. coli* clones.[2] Different O-antigen forms are almost entirely due to genetic variations in O-antigen gene clusters, which are located between *galF* and *gnd* in the chromosome in *E. coli*. The cluster usually contains genes involved in synthesis of nucleotide-activated sugars, which are specific components of the O-unit, sugar transfer, and O-antigen processing.[1] Studies of the O-antigen structure and genetics are helpful for development of diagnosis tools for *E. coli* strains useful for disease control and prevention.

Until recently, O-antigen structures have been established in about 120 *E. coli* O-serogroups (see *Escherichia coli* O-antigen database at http://www.casper.organ.su.se/ECODAB/ and bacterial carbohydrate structure database at http://csdb/glycoscience.ru/bacterial). In this work, we established the structure of the O-antigen of *E. coli* O154 and characterized the O-antigen gene cluster of this bacterium. Strains of serogroup O154 have been isolated from patients with diarrhea in Brazil, Myanmar, and Japan and identified as enteroadherent *E. coli* (EAEC).[3]

A high-molecular mass OPS 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 OPS revealed Rha, ManN and GalN in the ratio ~7:1:1 (detector response). GLC analyses of the acetylated (*S*)-2-(+)-octyl glycosides demonstrated the D configuration of GalN and the L configuration of Rha. The D configuration of ManN was established using the known regularities in the ^{13}C NMR chemical shifts.[4]

The ^{13}C NMR spectrum of the OPS (Fig. 1) showed signals for five anomeric carbons in the region δ 95.7–103.4, three CH_3–C groups (C-6 of Rha) at δ 18.0–18.3, two $HOCH_2$–C groups (C-6 of hexoses) at δ 61.5 and 62.1, two nitrogen-bearing carbons (C-2 of GalN and ManN) at δ 49.8 and 53.9, 18 oxygen-bearing non-anomeric sugar ring carbons in the region δ 67.8–79.8, and two *N*-acetyl groups at δ 23.2, 23.4 (both CH_3), 175.3 and 175.9 (both CO). Accordingly, the ^{1}H NMR spectrum of the OPS contained signals for five anomeric protons at δ 4.97–5.23, three methyl groups (H-6 of Rha) at 1.26–1.34, other sugar protons in the region δ 3.30–4.30, and two *N*-acetyl groups at δ 2.05. Therefore, the OPS has a pentasaccharide O-unit containing three residues of L-Rha

* 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 © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.carres.2013.06.004

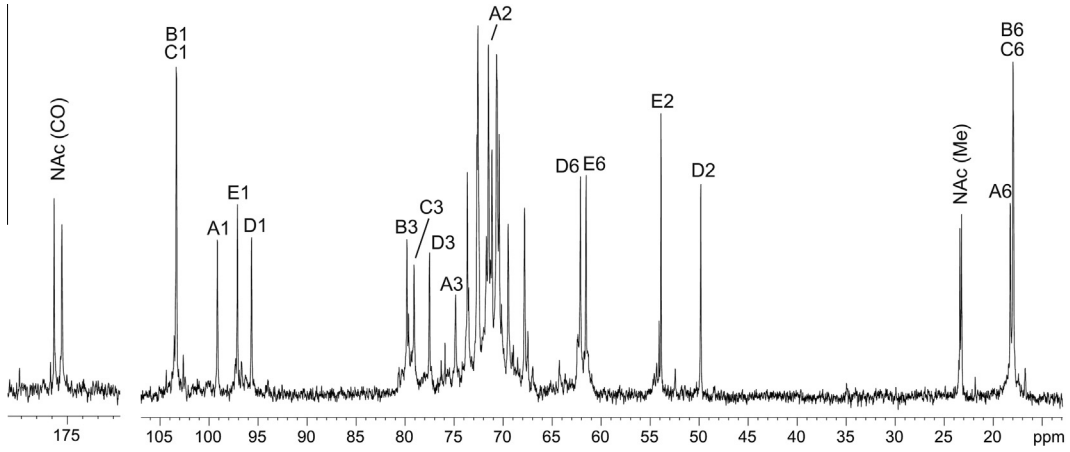

Figure 1. ^{13}C NMR spectrum of the OPS of *E. coli* O154. Numbers refer to carbons in sugar residues denoted as shown in Table 1.

(denoted as units **A**–**C**) and one residue each of D-GalNAc and D-ManNAc (units **D** and **E**, respectively).

Signals in the ^{1}H and ^{13}C NMR spectra of the OPS were assigned using 2D ^{1}H,^{1}H COSY, TOCSY, ROESY, ^{1}H,^{13}C HSQC, HSQC–TOCSY, and HMBC experiments (Table 1). Based on intra-residue H,H and H,C correlations and coupling constant values, spin systems were assigned to residues **A**–**E**, all being in the pyranose form. Positions at δ 70.4–70.7 of the signals for C-5 indicated that units **A**–**C** are α-linked (compare published data δ 70.0 and 73.2 for α- and β-Rha*p*, respectively[4]). Likewise, the position of C-5 at δ 72.7 and 73.7 demonstrated the α configuration of units **D** and **E** (compare published data δ 71.7 and 76.4 for α- and β-GalNAc, δ 73.3 and 77.6 for α- and β-ManNAc, respectively[4]).

The spin systems for **D** and **E** were distinguished by a correlation between proton at the nitrogen-bearing carbon (H-2) and the corresponding carbon (C-2) in the ^{1}H,^{13}C HSQC spectrum and assigned based on C-2 chemical shifts of δ 49.8 and 53.9 (compare published data δ 51.5 for α-GalNAc and δ 54.5 for α-ManNAc[4]). The signals for C-2 and C-3 of unit **A** and C-3 of units **B**–**D** were shifted significantly downfield, as compared with their positions in the corresponding non-substituted monosaccharides,[4] whereas the chemical shifts for C-2,3,4,6 of unit **E** differed insignificantly. These data demonstrated the branching character of the OPS with unit **A** at the branching point and unit **E** at the terminal position in the side chain; the other monosaccharide residues are monosubstituted at position 3.

The ROESY spectrum of the OPS showed the following correlations between anomeric protons and protons at the linkage carbons: **A** H-1,**B** H-3; **B** H-1,**C** H-3; **C** H-1,**D** H-3; **D** H-1,**A** H-2 and **E** H-1,**A** H-3 at δ 5.23/3.89, 5.00/3.87, 4.97/4.00, 5.09/4.30 and 5.05/4.05, respectively. The monosaccharide sequence thus determined was confirmed by a heteronuclear ^{1}H,^{13}C HMBC experiment, which showed correlations between anomeric protons and linkage carbons and vice versa (Fig. 2).

Therefore, the OPS of *E. coli* O154 has the following structure, which, to our knowledge, is unique among the known bacterial polysaccharide structures:

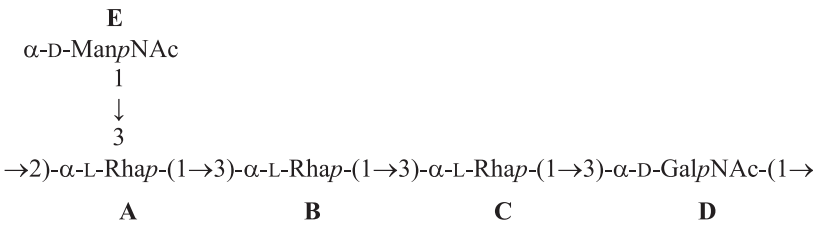

Table 1
^{1}H and ^{13}C NMR chemical shifts (δ, ppm) of the OPS of *E. coli* O154

Sugar unit	H_1	H_2	H_3	H_4	H_5	H_6 (6a, 6b)
	C1	C2	C3	C4	C5	C6
→2,3)-α-L-Rha*p*-(1→	5.23	4.30	4.05	3.7	3.89	1.34
A	99.1	71.5	74.8	71.7	70.7	18.3
→3)-α-L-Rha*p*-(1→	5.00	4.16	3.89	3.55	3.80	1.28
B	103.4	71.2	79.8	72.6	70.4	18.0
→3)-α-L-Rha*p*-(1→	4.97	3.92	3.87	3.55	3.88	1.26
C	103.3	71.5	79.1	72.6	70.7	18.0
→3)-α-D-Gal*p*NAc-(1→	5.09	4.35	4.00	4.14	3.91	3.67; 3.75
D	95.7	49.8	77.5	69.5	72.7	62.1
α-D-Man*p*NAc-(1→	5.05	4.37	4.06	3.65	3.97	4.85; 4.85
E	97.1	53.9	70.7	67.8	73.7	61.5

The chemical shifts for the *N*-acetyl group are δ_H 2.05 (2 Me); δ_C 23.2, 23.4 (Me) and 175.3, 175.9 (CO).

The O-antigen gene cluster of *E. coli* O154 was found between the housekeeping genes *galF* and *gnd*. A DNA sequence of 13,635 bp was obtained, which contains 10 genes with transcription direction from *galF* to *gnd* (Fig. 3). Their functions were assigned based on similarities to genes from available databases (Table 2). Synthesis of O-units containing GlcNAc as the first sugar is initiated by transfer of GlcNAc 1-phosphate from UDP-GlcNAc to the undecaprenol phosphate lipid carrier (Und-PP) catalyzed by glycosylphosphatetransferase WecA.[5] An epimerase called Gnu is necessary for conversion of GlcNAc-P-P-Und to GalNAc-P-P-Und in synthesis of the O-units that begin with GalNAc.[6] Both *wecA* and *gnu*, as well as genes involved in synthesis of common sugar nucleotide precursors, such as UDP-D-GlcNAc, are located outside the O-antigen gene cluster.

Orf1-4 shared high level identities to many known RmlB, D, A and C proteins (73–98%). The *rmlBDAC* gene set responsible for the synthesis of dTDP-L-Rha has been well characterized.[7–9] Orf7