Carbohydrate Research 430 (2016) 24–28

Contents lists available at ScienceDirect

Carbohydrate Research

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

Note

Structure and gene cluster of the O-antigen of *Escherichia coli* O156 containing a pyruvic acid acetal

Zhifeng Duan [a], Sof'ya N. Senchenkova [b], Xi Guo [a], Andrei V. Perepelov [b], Alexander S. Shashkov [b], Bin Liu [a], Yuriy A. Knirel [b,*]

[a] *Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, TEDA Institute of Biological Sciences and Biotechnology, Nankai University, TEDA, 300457 Tianjin, China*
[b] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation*

ARTICLE INFO

Article history:
Received 5 April 2016
Received in revised form 26 April 2016
Accepted 27 April 2016
Available online 29 April 2016

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

ABSTRACT

The lipopolysaccharide of *Escherichia coli* O156 was degraded under mild acidic and alkaline conditions and the resulting polysaccharides were studied by sugar analysis and ^{1}H and ^{13}C NMR spectroscopy. The following structure of the pentasaccharide repeating unit of the O-polysaccharide was established:

→4)-α-L-Fuc*p*-(1→6)-α-D-Gal*p*-(1→3)-α-L-Fuc*p*-(1→3)-β-D-Glc*p*NAc-(1→
3
↑
1
α-D-Gal*p*4,6*R*pyr

where *R*pyr indicates *R*-configurated pyruvic acid acetal. Minor O-acetyl groups also were present and tentatively localized on the Gal residues. The gene cluster for biosynthesis of the O-antigen of *E. coli* O156 was analyzed and shown to be consistent with the O-polysaccharide structure.

Escherichia coli is the predominant facultative anaerobe of the colonic flora of many mammals, including humans, and has both commensal and pathogenic forms. Pathogenic strains causing diarrhea and other disorders belong to 6 categories: enteropathogenic (EPEC), enterotoxigenic (ETEC), enteroaggregative (EAEC), enterohemorrhagic (EHEC), enteroinvasive (EIEC), and diffusely adherent (DAEC) *E. coli*.[1] O-polysaccharide (O-antigen) is a part of the lipopolysaccharide (LPS) on the outer membrane of gram-negative bacteria and consists usually of oligosaccharide repeats (O-units) containing two to eight residues from a broad range of common and rarely occurring sugars and their derivatives. The O-antigen is the most variable cell constituent with variations in the types of sugars present, their arrangement, and the linkages within and between O-units.[2]

Genes involved in the O-antigen synthesis are generally combined in a cluster; in *E. coli* it is usually localized between conserved *galF* and *gnd* genes. Most of the genes fall into one of three major classes: sugar nucleotide synthesis genes, glycosyl transferase genes, and O-unit processing genes. Genes for synthesis of sugars that are also present in other bacterial structures or involved in metabolism usually map at other loci. Most variations among diverse O-antigen forms are due to polymorphism of the O-antigen gene cluster.[3]

Currently, >180 O-serogroups of *E. coli* are internationally recognized. Sequences of their O-antigen gene clusters have been reported[4] and >140 O-antigen structures have been determined (http://nevyn.organ.su.se/ECODAB/). In this work, we report the structure of the O-polysaccharide of *E. coli* O156, which belongs to Shiga toxin-producing *E. coli* (STEC) and has been reported to be associated with EHEC-related disease in humans.[5] In addition, the O156-antigen gene cluster was analyzed and putative gene functions were assigned based on their homology to genes in available databases.

Structure elucidation of the O-polysaccharide. Mild acid degradation of LPS obtained from *E. coli* O156 cells by phenol–water extraction afforded a polysaccharide (PS) isolated by GPC on Sephadex G-50. NMR analysis revealed an irregularity in the PS structure, owing to the presence of O-acetyl groups (δ_H 2.18–2.19, δ_C 21.6–21.8 for Me) and pyruvic acid acetal (δ_H 1.50, δ_C 26.3 for Me) in non-stoichiometric amounts. Therefore, LPS was treated with alkali under mild conditions to yield an O-deacylated lipopolysaccharide (LPS_{OH}), with O-polysaccharide chain having a regular structure. Therefore, pyruvic acid was partially cleaved in the course of mild acid hydrolysis of LPS. Indeed, a longer hydrolysis under the same acidic conditions (4 h versus 1 h) resulted in a modified polysaccharide

* Corresponding author. N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation. Tel.: +7 499 1376148; fax: +7 499 1355328.
E-mail address: yknirel@gmail.com (Y.A. Knirel).

http://dx.doi.org/10.1016/j.carres.2016.04.025
0008-6215/

Table 1
^{1}H and ^{13}C NMR chemical shifts (δ, ppm)

Sugar residue	Nucleus	1	2	3	4	5	6
LPS_{OH} (O-polysaccharide chain)							
→3)-β-D-Glc*p*NAc-(1→[a]	^{1}H	4.63	3.84	3.69	3.52	3.39	3.67, 3.87
A	^{13}C	102.1	57.1	81.3	70.0	77.5	62.3
→3)-α-L-Fuc*p*-(1→	^{1}H	5.02	3.91	3.90	3.95	4.33	1.17
B	^{13}C	100.7	68.4	79.7	72.9	68.1	16.6
→6)-α-D-Gal*p*-(1→	^{1}H	5.19	3.83	3.94	4.05	4.25	3.57, 3.85
C	^{13}C	101.8	69.9.0	70.5	70.5	70.7	67.6
→3,4)-α-L-Fuc*p*-(1→	^{1}H	4.88	4.02	3.96	4.03	4.06	1.24
D	^{13}C	99.8	69.3	75.4	78.9	68.3	17.1
α-D-Gal*p*4,6*R*pyr-(1→	^{1}H	5.32	3.91	4.02	4.26	3.96	3.90, 4.04
E	^{13}C	101.5	69.6	69.1	72.9	63.94	66.3
DPS_{OH}							
→3)-β-D-Glc*p*NAc-(1→[b]	^{1}H	4.66	3.91[c]	3.70	3.59	3.46	3.76, 3.91
A	^{13}C	102.4	57.1[b]	81.2	69.8	77.4	62.2
→3)-α-L-Fuc*p*-(1→	^{1}H	5.04	3.91	3.91	3.98	4.36	1.17
B	^{13}C	101.0	68.4	79.4	72.9	68.1	16.6
→6)-α-D-Gal*p*-(1→	^{1}H	4.90	3.83	3.96	4.06	4.27	3.57, 3.87
C	^{13}C	99.9	70.0	70.5	70.5	70.7	67.3
→3,4)-α-L-Fuc*p*-(1→	^{1}H	5.21	4.05	4.04	4.14	4.10	1.27
D	^{13}C	101.8	69.3	74.7	79.4	68.5	17.2
α-D-Gal*p*-(1→	^{1}H	5.31	3.81	3.96	4.02	4.21	3.73, 3.78
E	^{13}C	101.0	69.9	70.7	70.8	72.7	62.7

Additional chemical shifts for the N-acetyl group are [a]δ_H 2.01, δ_C 23.6 (CH_3) and 175.5 (CO); for pyr δ_H 1.50, δ_C 26.3 (C-3), 100.9 (C-2) and 175.7 (C-1); [b]δ_H 2.04, δ_C 23.5 (CH_3) and 175.7 (CO).

(DPS) that contained only a trace amount of pyruvic acid. Mild alkaline hydrolysis of DPS resulted in an O-deacetylated polysaccharide (DPS_{OH}).

The ^{13}C NMR spectrum of LPS_{OH} showed signals for five monosaccharide residues, including five anomeric atoms at δ_H 4.63–5.32 and δ_C 99.8–102.1; two CH_3-C groups (C-6 of Fuc) at δ_H 1.17 and 1.24, δ_C 16.6 and 17.1; three OCH_2-C groups (C-6 of hexoses) at δ_C 62.3 (non-substituted), 66.3 and 67.6 (both O-substituted); one nitrogen-bearing carbon (C-2 of GlcNAc) at δ_C 57.1; one acetal-linked pyruvic acid residue (pyr) at δ_H 1.50, δ_C 26.3 (C-3, Me), 100.9 (C-2) and 175.7 (C-1, CO_2H); one N-acetyl group at δ_H 2.01, δ_C 23.6 (Me) and 175.5 (CO); and other signals at δ_H 3.39–4.33 and δ_C 63.9–81.3 (Table 1). Signals for OCH_2-C and CO groups were assigned using multiplicity-edited $^{1}H,^{13}C$ HSQC and $^{1}H,^{13}C$ HMBC experiments, respectively.

Sugar analysis of LPS_{OH} using GLC of the alditol acetates after full acid hydrolysis revealed Fuc, Gal, and GlcNAc in the ratios 1.2 : 1.0 : 0.4. GLC analysis of the acetylated glycosides with (+)-2-octanol indicated that Fuc has the L configuration and Gal and GlcNAc have the D configuration.

The ^{1}H and ^{13}C NMR spectra of LPS_{OH} were assigned using 2D $^{1}H,^{1}H$ COSY, TOCSY, and $^{1}H,^{13}C$ HSQC experiments (Table 1). Based on $^{3}J_{H,H}$ coupling constants and ^{1}H and ^{13}CNMR chemical shifts, spin systems for two residues each of α-Gal*p* (units **C** and **E**) and α-Fuc*p* (units **B** and **D**), and one residue of β-Glc*p*NAc (unit **A**) were identified. The GlcNAc residue was confirmed by correlations between the proton at the nitrogen-bearing carbon (H-2) and the corresponding carbon (C-2) at δ 3.84/57.1 in the $^{1}H,^{13}C$ HSQC spectrum.

Linkage and sequence analysis of LPS_{OH} was performed using the 2D $^{1}H,^{1}H$ ROESY (Fig. 1) and $^{1}H,^{13}C$ HMBC (Fig. 2) experiments, which showed correlations between anomeric protons and protons at the linkage carbons (ROESY) or between anomeric protons and linkage carbons and vice versa (HMBC) (Table 2). The sugar substitution pattern in the O-polysaccharide repeat thus defined (Chart 1) was confirmed by significantly downfield displacements due to glycosylation of the signals for C-3 of units **A** and **B**, C–3 and C-4 of unit **D**, and C-6 of unit **C** (Table 1), as compared with their positions in the spectra of the corresponding non-substituted monosaccharides.[6]

Similarly, a structure of DPS_{OH} (Chart 1) was established by 2D NMR spectroscopy (for assigned ^{1}H and ^{13}C NMR chemical shifts of DPS_{OH}, see Table 1) and found to be in full agreement with the data of LPS_{OH}. A comparison of the ^{13}C NMR and 2D $^{1}H,^{13}C$ HSQC spectra of DPS_{OH} and LPS_{OH} revealed downfield displacements of the signals for C-4 and C-6 of unit **E** from δ 70.8 and 62.7 in DPS_{OH} to δ 72.9 and 66.3 in LPS_{OH} and an upfield displacement of the signal for C-5 of unit **E** from δ 72.7 to δ 63.9. This finding indicated that pyr is localized at positions 4 and 6 of unit **E**. The chemical shifts of the methyl group δ_H 1.50 and δ_C 26.3 showed that pyr has the *R* configuration.[7]

Therefore, the polysaccharide chain in LPS_{OH} includes D-Gal*p*4,6*R*pyr and has the structure shown in Chart 1. Earlier, D-Glc*p*NAc4,6*S*pyr has been found in the O-polysaccharides of *E. coli* O112ac[8] and O149,[9] whereas D-Gal*p*4,6*R*pyr is identified in *E. coli* for the first time.

A comparison of the NMR spectra of DPS_{OH} and DPS revealed a displacement of a minor part of the H-1/C-1 cross-peak from δ 5.29/101.0 in the former to δ 5.34/98.6 in the latter. This displacement was evidently due to a β-effect of O-acetylation and showed that one of the O-acetyl groups occurs at position 2 of unit **E**. As judged by relative intensities of the signals for H-1 of the O-acetylated and non-acetylated unit **E** in DPS, the degree of 2-O-acetylation was ~35%. Positions of other minor O-acetyl groups could not be unambiguously determined by NMR spectroscopy. However, signals for units **A**, **B**, and **D** were essentially the same in the spectra of DPS_{OH} and DPS, and hence these units were not O-acetylated. Therefore, the minor O-acetyl groups could be localized on a Gal residue(s) at position 3 (unit **C** or/and **E**) or/and position 4 (unit **C**).

Characterization of the O-antigen gene cluster. The O-antigen gene cluster of *E. coli* O156 between the housekeeping genes *galF* and *gnd* has been sequenced (GenBank accession number AB812065).[4] It contains 14 ORFs having the same transcription direction from *galF* to *gnd* (Fig. 3). D-GlcNAc and D-Gal that are contained in the O156-polysaccharide are common sugars in bacteria, and genes for synthesis of their nucleotide precursors are usually located outside the O-antigen gene cluster.[10] Hence, no such genes were found between *galF* and *gnd* in O156 too. The biosynthesis pathway of GDP-L-Fuc, the nucleotide precursor of L-Fuc, has been