Carbohydrate Research 420 (2016) 1–5

Contents lists available at ScienceDirect

Carbohydrate Research

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

Structure and gene cluster of the o–antigen of *Escherichia coli* o96

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

[a] N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russia
[b] TEDA Institute of Biological Sciences and Biotechnology, Nankai University, TEDA, 300457 Tianjin, China

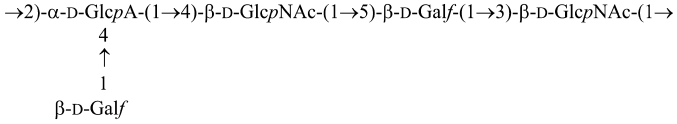
ARTICLE INFO

Article history:
Received 14 September 2015
Accepted 12 November 2015
Available online 23 November 2015

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

ABSTRACT

Mild acid degradation of the lipopolysaccharide of *Escherichia coli* O96 afforded a mixture of two polysaccharides. The following structure of the pentasaccharide repeating unit of the major polymer was established by sugar analysis, Smith degradation, and ^{1}H and ^{13}C NMR spectroscopy:



The O-antigen gene cluster of *E. coli* O96 between conserved *galF* and *gnd* genes was found to be consistent with this structure, and hence, the major polysaccharide represents the O96-antigen. The O96-antigen structure and gene cluster are similar to those of *E. coli* O170, and two proteins encoded in the gene clusters of both bacteria were putatively assigned a function of galactofuranosyltransferases. The minor polymer has the same structure as a peptidoglycan-related polysaccharide reported earlier in *Providencia alcalifeciens* O45 and several other O-serogoups of this species (Ovchinnikova OG, Liu B, Kocharova NA, Shashkov AS, Kondakova AN, Siwinska M, Feng L, Rozalski A, Wang L, Knirel YA. *Biochemistry (Moscow)* 2012;**77**:609-15)

→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc3(*R*lac-LAla)-(1→

where *R*lac-LAla indicates (*R*)-1-[(*S*)-1-carboxyethylaminocarbonyl]ethyl.

1. Introduction

Escherichia coli is the predominant facultative anaerobe of the colonic flora of many mammals, including humans, and has both commensal and pathogenic forms.[1] The O-polysaccharide chain of the lipopolysaccharide (LPS) called O-antigen is the most variable cell component as, being exposed on the cell surface, it is subject to intense selection by the host immune system and bacteriophages. O-antigen structure variations provide the basis for serotyping of bacterial strains. Genes involved in the O-antigen biosynthesis are generally combined in a cluster, which in *E. coli* and related bacteria is usually located between conserved *galF* and *gnd* genes.[2] Up to date, 184 O serogroups of *E. coli* have been internationally recognized, with all O-antigen gene clusters being sequenced.[3] In this work, chemical structure of the O96-polysaccharide was established and the O-antigen gene cluster of *E. coli* O96 was analyzed and found to be consistent with the O-polysaccharide structure.

Mild hydrolysis with dilute acetic acid of the LPS isolated from *E. coli* O96 cells by phenol-water extraction afforded a mixture of polysaccharides. Sugar analysis using GLC of the alditol acetates after full acid hydrolysis revealed Gal and GlcNAc. Further studies showed that GlcA also was present. GLC analysis of the acetylated glycosides with (+)-2-octanol indicated that all monosaccharides have the D configuration.

The ^{13}C NMR spectrum of the isolated material (Fig. 1) showed two series of signals having different intensities, suggesting the presence of two polysaccharides. The series for the major polysaccharide (PS-1) includes signals for five anomeric carbons at δ 99.2–109.4, four $HOCH_2$-C groups at δ 62.0-64.02, two nitrogen-bearing carbons (C-2 of GlcNAc) at δ 56.4 and 57.1, other sugar ring carbons at δ 69.6-84.6, two *N*-acetyl groups (Me at δ 23.5 and 23.6), and CO of acetyl and carboxyl groups at δ 175.1-176.0. The ^{1}H NMR spectrum of PS-1 showed signals for anomeric protons in the region at δ 4.68-5.70, two *N*-acetyl groups at δ 2.04 (2 Me), and other protons at δ 3.47-4.26.

* Corresponding author. N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russia. Tel.: +7 499 1376148; fax: +7 499 1355328.
E-mail address: perepel@ioc.ac.ru (A.V. Perepelov).

http://dx.doi.org/10.1016/j.carres.2015.11.005
0008-6215/

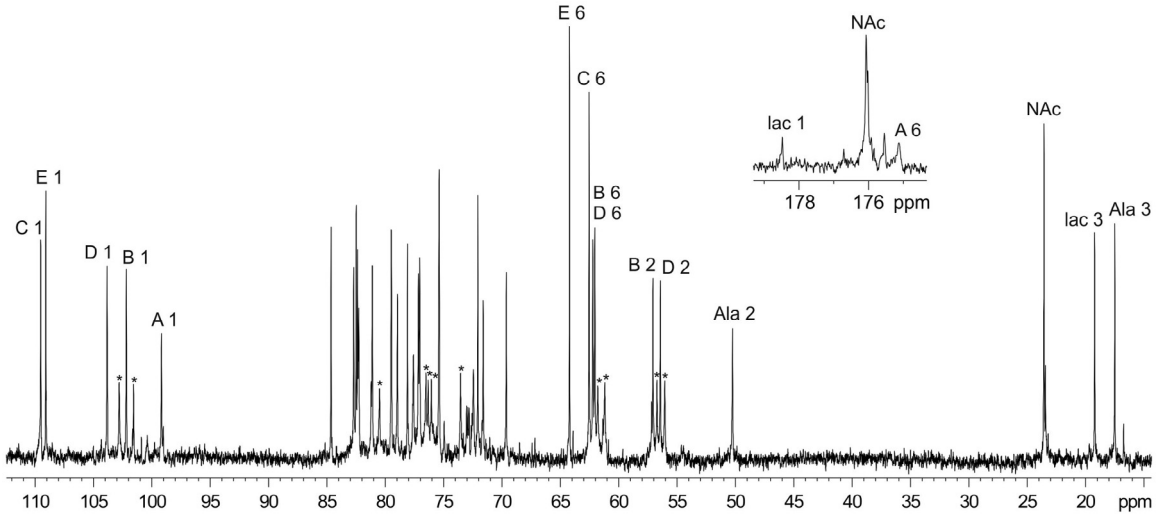


Fig. 1. ^{13}C NMR spectrum of the polysaccharides from *E. coli* O96. Arabic numerals refer to carbons in sugar residues denoted by letters as shown in Table 1 and Chart 1. Signals for a minor peptidoglycan-related polysaccharide are shown by asterisk.

The ^{1}H and ^{13}C NMR spectra of PS-1 were assigned using 2D ^{1}H,^{1}H COSY, TOCSY, ROESY, 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 β-Gal*f* (**C** and **E**) and β-Glc*p*NAc (**B** and **D**) and one residue of Glc*p*A (**A**) were recognized. The GlcNAc residues were confirmed by correlations between protons at the nitrogen-bearing carbons (H-2) and the corresponding carbons (C-2) at δ 3.71/57.1 and 3.83/56.4 in the ^{1}H,^{13}C HSQC, and GlcA was confirmed by a correlation between H-5 and C-6 (CO_2H) at δ 4.15/175.1 in the ^{1}H,^{13}C HMBC spectrum.

Downfield displacements of the signals for C-2 and C-4 of unit **A**, C–3 of units **D**, C-4 of unit **B** and C-5 of unit **C** to δ 77.6-82.3, respectively, as compared with their positions in the spectra of the corresponding non-substituted monosaccharides,[4] demonstrated the glycosylation pattern in the repeating unit. The 2D ROESY spectrum (Fig. 2) showed correlations between anomeric protons and protons at the linkage carbons at δ 5.70/3.72, 4.68/3.96, 5.02/3.63, 4.74/3.75, and 5.00/3.67, which were assigned as follows taking into account the ^{13}C NMR chemical shift data: **A** H-1/**B** H-4, **B** H-1/**C** H-5, **C** H-1/**D** H-3, **D** H-1/**A** H-2, and **E** H-1/**A** H-4, respectively. The

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

Sugar residue	Nucleus	1	2	3	4	5	6
O-polysaccharide[a]							
→2,4)-α-D-Glc*p*A-(1→	^{1}H	5.70	3.75	3.77	3.67	4.15	
A	^{13}C	99.2	81.1	71.6	79.5	72.4	175.1
→4)-β-D-Glc*p*NAc-(1→	^{1}H	4.68	3.71	3.88	3.72	3.58	3.76, 3.91
B	^{13}C	102.2	57.1	75.4	77.6	75.4	62.0
→5)-β-D-Gal*f*-(1→	^{1}H	5.02	3.99	4.26	4.14	3.96	3.69, 3.69
C	^{13}C	109.4	82.4	77.2	82.7	79.0	62.5
→3)-β-D-Glc*p*NAc-(1→	^{1}H	4.74	3.83	3.63	3.49	3.47	3.76, 3.93
D	^{13}C	103.8	56.4	82.3	69.6	77.0	62.2
β-D-Gal*f*-(1→	^{1}H	5.00	4.08	4.05	4.09	3.82	3.65, 3.68
E	^{13}C	109.1	82.5	78.1	84.6	72.1	64.2
Peptidoglycan-related polysaccharide[b]							
→4)-β-D-Glc*p*NAc-(1→	^{1}H	4.50	3.76	3.68	3.54	3.46	3.69, 3.83
F	^{13}C	101.6	56.7	73.6	81.1	76.0	61.8
→3,4)-β-D-Glc*p*NAc-(1→	^{1}H	4.52	3.78	3.66	3.85	3.50	3.69, 3.88
G	^{13}C	102.8	56.0	80.5	76.3	76.5	61.2
*R*lac-	^{1}H		4.40	1.37			
	^{13}C	178.5	79.5	19.2			
L-Ala	^{1}H		4.27	1.42			
	^{13}C	176.0	50.3	17.5			
Oligosaccharide **1**[a]							
β-D-Glc*p*NAc-(1→	^{1}H	4.68	3.73	3.55	3.46	3.44	3.77, 3.92
D	^{13}C	104.2	56.8	74.8	70.7	77.1	61.8[c]
→2)-α-D-Glc*p*A-(1→	^{1}H	5.72	3.69	3.72	3.51	4.01	
A	^{13}C	99.2	81.3	72.7	73.2	73.4	
→4)-β-D-Glc*p*NAc-(1→	^{1}H	4.64	3.74	3.88	3.73	3.58	3.77, 3.92
B	^{13}C	102.0	56.8	75.2	77.2	75.3	61.9[c]
→2)-Thr-ol-(1→	^{1}H	3.64,3.72	3.83	3.74	3.63,3.74		
C'	^{13}C	61.9	80.6	71.7	63.4		

Chemical shifts for the *N*-acetyl groups are δ_C 23.3-23.4 (both Me), 175.5-176.1 (both CO). [a] δ_H 2.04-2.05. [b] δ_H 1.98 and 2.03.
[c] Assignment could be interchanged.