Carbohydrate Research 346 (2011) 828–832

Contents lists available at ScienceDirect

Carbohydrate Research

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

O-antigen structure and gene clusters of *Escherichia coli* O51 and *Salmonella enterica* O57; another instance of identical O-antigens in the two species

Andrei V. Perepelov [a,*], Bin Liu [b,c], Sof'ya N. Senchenkova [a], Dan Guo [b,c], Sergei D. Shevelev [a], Lu Feng [b,c], Alexander S. Shashkov [a], Lei Wang [b,c], 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 300457, PR China
[c] Tianjin Key Laboratory of Microbial Functional Genomics, Nankai University, TEDA, Tianjin 300457, PR China

ARTICLE INFO

Article history:
Received 7 December 2010
Received in revised form 17 February 2011
Accepted 21 February 2011
Available online 25 February 2011

Keywords:
Salmonella enterica
Escherichia coli
O-polysaccharide structure
O-antigen gene cluster

ABSTRACT

The O-polysaccharides were released by mild acid hydrolysis from the lipopolysaccharides of *Escherichia coli* O51 and *Salmonella enterica* O57 and found to possess the same structure, which was established by sugar analysis and 1D and 2D NMR spectroscopy:

β-D-Glc*p*NAc
1
↓
2
→3)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→4)-α-D-Glc*p*-(1→3)-β-D-Gal*p*NAc-(1→

The O-antigen gene clusters of *E. coli* O51 and *S. enterica* O57 were sequenced and found to contain the same genes with a high-level similarity. All genes expected for the synthesis of the O-antigen were identified based on their similarity to genes from available databases.

1. Introduction

The O-antigen is a polysaccharide chain of the lipopolysaccharide, a major component of the outer membrane of Gram-negative bacteria. Most often, the O-antigen consists of a number of repeats of an oligosaccharide (O-unit). The O-antigen is one of the most variable constituents on the cell surface due to variation in the types of sugars present, their arrangement within the O-unit and the linkages within and between O-units. *Salmonella enterica* is recognized as a major pathogen of both animals and humans, and in many countries, it is the leading cause of outbreaks of food-borne infections. *Escherichia coli* includes both commensal and pathogenic strains. The O-antigen variation is the basis for bacterial serotyping, and at least 180 O-serogroups in *E. coli* and 46 in *S. enterica* have been recognized.

In *E. coli* and *S. enterica*, genes involved in O-antigen synthesis are usually clustered between *galF* and *gnd* on the chromosome. They are classified into three groups: nucleotide-sugar biosynthesis genes, genes for glycosyltransferases, and O-unit processing genes *wzx* and *wzy* encoding O-unit flippase and O-antigen polymerase, respectively. Genetic variations in O-antigen gene clusters contribute major differences to diverse O-antigen forms.

Previous studies showed that four pairs of *E. coli* and *S. enterica* O-serogroups, including *E. coli* O111 and *S. enterica* O35, *E. coli* O55 and *S. enterica* O50, *E. coli* O157 and *S. enterica* O30, and *E. coli* O77 and *S. enterica* O:6,14 (H), contain identical or almost identical O-antigens.[1,2] The O-antigen gene clusters from each pair of the *E. coli* and *S. enterica* serogroups share obvious similarity (70–74%), which is close to the lower end of the range for housekeeping genes. This finding indicates that the related O-antigen gene clusters may originate from a common ancestor, and then diverged at a higher rate than housekeeping genes under consistent selection pressure.[1] Recently, more closely related *E. coli* and *S. enterica* O-serogroups have been recognized (Ref. 3 and references cited thereof).

In this work, we found a new pair of identical *E. coli* and *S. enterica* O-serogroups, which are characterized by the same O-antigen structure and a high-level similarity of O-antigen gene clusters.

* 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.2011.02.020

2. Results and discussion

The lipopolysaccharides were isolated from cells of *E. coli* O51 and *S. enterica* O57 by the phenol–water procedure, and the O-polysaccharides were released by mild acid degradation and isolated by GPC on Sephadex G-50. The O-polysaccharides from both species showed essentially identical ^{13}C NMR spectra and, hence, they were concluded to have the same structure. Further studies were performed with the polysaccharide of *E. coli* O51.

The ^{13}C NMR spectrum of the polysaccharide (Fig. 1) indicated a pentasaccharide O-unit (there were signals for five anomeric carbons at δ 95.5–104.5), including two *N*-acetylamino sugars (signals for two nitrogen-bearing carbons at δ 52.4 and 57.0, and two *N*-acetyl groups: CH_3 at δ 23.7 and 23.9 and CO at δ 175.3 and 176.2) and two 6-deoxy sugars (signals for CH_3-C at δ 17.9 and 18.3). The ^{1}H NMR spectrum showed, inter alia, signals for five anomeric protons at δ 4.67–5.12, two CH_3–C groups of 6-deoxy sugars at δ 1.26, and two *N*-acetyl groups at δ 2.00 and 2.04 (Table 1).

Sugar analysis by GLC of the acetylated alditols derived after full acid hydrolysis of the polysaccharide revealed rhamnose, glucose, GlcN, and GalN in the ratio of 2.2:1.6:0.7:1.0 (detector response). Determination of the absolute configurations by GLC of the (*S*)-2-octyl glycosides showed that rhamnose is L and the other monosaccharides are D. Methylation analysis of the polysaccharide, including GLC–MS analysis of the partially methylated alditol acetates, demonstrated major derivatives from 2-substituted and 2,3-disubstituted rhamnose, 4-substituted glucose, non-substituted GlcN and 3-substituted GalN. In addition, there were minor derivatives from 3-substituted rhamnose, terminal, 2-substituted, and 3-substituted hexoses.

The polysaccharide was analyzed using 1D ^{1}H,^{1}H TOCSY and ROESY, 2D ^{1}H,^{1}H COSY, ^{1}H,^{13}C HSQC, and ^{1}H,^{13}C HMBC experiments. Five sugar spin systems were revealed and assigned by the connectivity patterns and coupling constant values to two residues of rhamnose (Rha^{I} and Rha^{II}) and one residue each of Glc, GlcNAc, and GalNAc, all being in the pyranose form (Table 1). A $J_{1,2}$ coupling constant value of <3 Hz indicated that Glc is α-linked, whereas relatively large $J_{1,2}$ values of 7–8 Hz showed that GlcNAc and GalNAc are β-linked. The α-configuration of both Rha residues was inferred by the C-5 chemical shifts of δ 70.5–71.0 (compare published data[4] δ 70.0 for α-Rha*p* and δ 73.2 for β-Rha*p*).

The substitution pattern of the monosaccharides was confirmed by significant downfield displacements to δ 78.4–82.1, due to glycosylation, of the signals for C-4 of Glc and Rha^{I}, C-2, and C-3 of Rha^{II}, as compared with their positions at δ 71.3–72.7 in the corresponding non-substituted monosaccharides.[4] A lower glycosylation shift from δ 72.4 to 76.2 observed for C-3 of GalNAc is characteristic for α-1,3-linked disaccharides with the *galacto* configuration of the glycosylated monosaccharide.[5] These data are in agreement with methylation analysis data (see above) and showed that the polysaccharide is branched with a 2,3-disubstituted Rha residue at the branching point. The chemical shifts for C-2 to C-6 of GlcNAc were close to those of non-substituted β-Glc*p*NAc[4] and thus confirmed the terminal position of this residue in the side chain.

The monosaccharide sequence in the repeating unit was established by 1D ^{1}H,^{1}H ROESY experiments with selective excitation

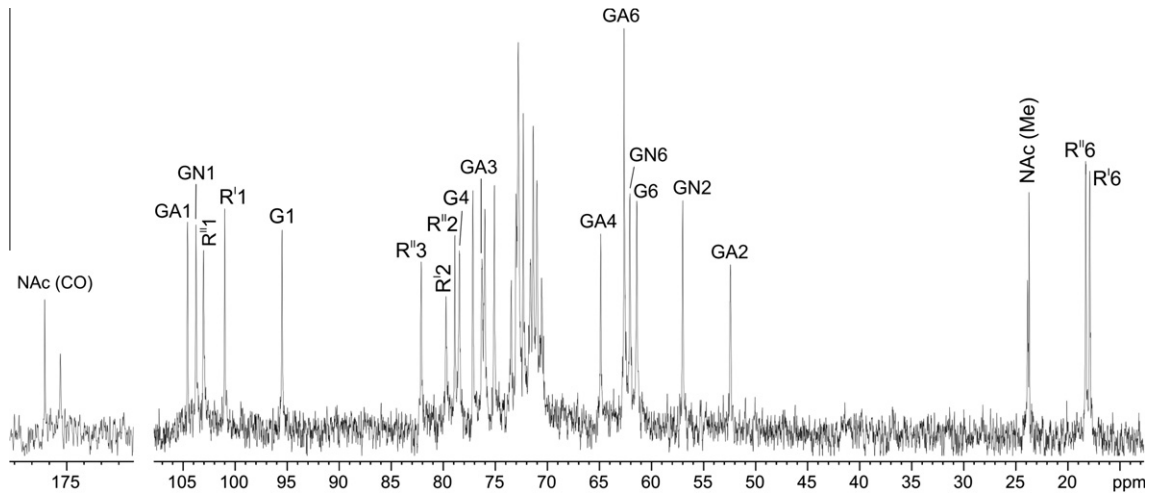

Figure 1. ^{13}C NMR spectrum of the O-polysaccharide of *E. coli* O51. Numbers refer to carbons in sugar residues denoted as follows: G, Glc; GN, GlcN; GA, GalN; R^{I}, Rha^{I}; R^{II}, Rha^{II}.

Table 1
^{1}H and ^{13}C NMR chemical shifts of the O-polysaccharide of *E. coli* O51 (δ, ppm)

Residue	Nucleus	Monosaccharide						NAc	
		1	2	3	4	5	6	CH_3	CO
→3)-β-D-Gal*p*NAc-(1→	^{1}H	4.67	4.15	3.79	4.21	3.69	3.83, 3.83	2.00	
	^{13}C	104.5	52.4	76.2	64.8	76.0	62.6	23.7	175.3
→4)-α-D-Glc*p*-(1→	^{1}H	5.12	3.55	3.67	3.54	3.67	3.72, 3.83		
	^{13}C	95.5	72.8[a]	72.8[a]	78.4	73.0[a]	61.4		
→2)-α-L-Rha*p*^{I}-(1→	^{1}H	4.92	3.93	3.81	3.46	4.03	1.26		
	^{13}C	101.0	79.7	71.6	73.4	70.5	17.9		
→2,3)-α-L-Rha*p*^{II}-(1→	^{1}H	5.07	4.41	3.82	3.34	3.69	1.26		
	^{13}C	103.0	78.9	82.1	72.3	71.0	18.3		
β-D-Glc*p*NAc-(1→	^{1}H	4.88	3.72	3.57	3.44	3.43	3.74, 3.88	2.04	
	^{13}C	103.7	57.0	75.1	71.3	77.1	62.0	23.9	176.2

[a] Assignment could be interchanged.