Carbohydrate Research 388 (2014) 30–36

Contents lists available at ScienceDirect

Carbohydrate Research

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

Structure elucidation and gene cluster annotation of the O-antigen of *Escherichia coli* O39; application of anhydrous trifluoroacetic acid for selective cleavage of glycosidic linkages

Andrei V. Perepelov [a,*], Andrei V. Filatov [a], Quan Wang [b], Vyacheslav L. L'vov [c], Ye Qian [b], Alexander S. Shashkov [a], Lei Wang [b], Yuriy A. Knirel [a]

[a] N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, Russia
[b] TEDA School of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin, China
[c] Institute of Immunology, Federal Medical-Biological Agency, Moscow, Russia

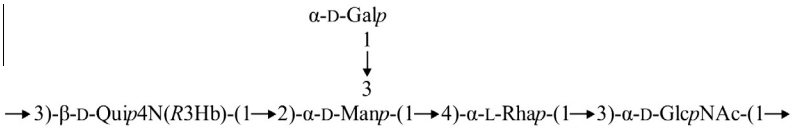
ARTICLE INFO

Article history:
Received 12 January 2014
Accepted 9 February 2014
Available online 18 February 2014

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

ABSTRACT

O-Polysaccharide (O-antigen) accompanied by a minor mannan was isolated from the lipopolysaccharide of *Escherichia coli* O39 and studied by component analyses, methylation, Smith degradation, mass spectrometry, and 1D and 2D NMR spectroscopy. In addition, a new approach, solvolysis with anhydrous trifluoroacetic acid, was applied to cleave selectively the rhamnosidic linkage. The following structure of the O-polysaccharide was established:



where D-Qui4N(*R*3Hb) indicates 4,6-dideoxy-4-[(*R*)-3-hydroxybutanoylamino]-D-glucose. The O-antigen gene cluster of *E. coli* O39 has been sequenced. The gene functions were tentatively assigned by a comparison with sequences in the available databases and found to be in agreement with the O-polysaccharide structure.

1. Introduction

Escherichia coli clones, both commensal and pathogenic, are normally identified by a combination of somatic (O), flagellar (H), and sometimes capsular (K) antigens.[1] The O-antigen is a polysaccharide chain (O-polysaccharide, OPS) of the lipopolysaccharide on the surface of Gram-negative bacteria. It consists of many repeats of an oligosaccharide (O-unit) and is the most variable cell constituent. Till now, 174 *E. coli* O-antigen forms have been recognized.[2] Different O-antigen forms are mainly due to genetic variations in the O-antigen gene cluster, which is located between housekeeping genes *galF* and *gnd* on the chromosome in *E. coli*. It contains genes for synthesis of nucleotide precursors of sugars that are specific O-unit components, genes that mediate sugar transfer in the O-unit assembly on a lipid carrier, and O-antigen processing genes, including *wzx* and *wzy* for O-unit translocation across membrane and polymerization, respectively.[1] Chemical, biochemical, and genetic studies of *E. coli* O-antigens are helpful for development of improved diagnostic tools necessary for disease prevention and control.

E. coli O39 strains have been identified as a Shiga toxin-producing *E. coli* (STEC) isolated from various foods including beef, hamburgers, and soft cheeses all over the world.[3–5] An *E. coli* O39 isolate, O39:NM, that does not fit into the existing scheme for classifying diarrheogenic *E. coli*, was also reported to cause a gastrointestinal illness outbreak in USA in 1991.[6] In this work, we established the OPS structure of *E. coli* O39 and characterized the O-antigen gene cluster of this bacterium.

2. Results and discussion

2.1. Structure elucidation of the O-polysaccharide

A high-molecular mass polysaccharide was obtained by mild acid degradation of the lipopolysaccharide isolated from bacterial

* Corresponding author. 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.2014.02.013
0008-6215/

cells by the phenol–water procedure. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the polysaccharide revealed Rha, Man, Gal, and GlcNAc in the ratios ~1:3:1:0.5 (detector response), respectively. GLC analysis of the acetylated (*S*)-2-octyl glycosides demonstrated the D configuration of Man, Gal, and GlcN, and the L configuration of Rha. Yet another polysaccharide component, 4,6-dideoxy-4-[(*R*)-3-hydroxybutanoylamino]-D-glucose [D-Qui4N(*R*3Hb)], was not detected in sugar analysis but was identified by further NMR spectroscopic studies. The D configuration of its sugar moiety was established by analysis of ^{13}C NMR glycosylation shifts (see below), and the *R* configuration of 3-hydroxybutanoic acid was determined by GLC of the trifluoroacetylated (*S*)-2-octyl ester obtained from the polysaccharide hydrolysate.

Methylation analysis of the polysaccharide resulted in identification of the major alditol acetates derived from 2,3,4,6-tetra-*O*-methylhexose (from terminal Gal), 4,6-di-*O*-methylhexose (from a 2,3-substituted hexose, evidently Man), and 6-deoxy-2,3-di-*O*-methylhexose (from 4-substituted Rha), as well as 2-deoxy-4,6-di-*O*-methyl-2-(*N*-methyl)acetamidohexose (from 3-substituted GlcNAc). Therefore, the OPS is branched with Man at the branching point and Gal as the terminal monosaccharide of the side chain. Again, no Qui4N derivative was detected in methylation analysis.

Identification of only one major methylated mannose derivative but a high content of Man detected in sugar analysis (three times as much as Gal) suggested the presence of mannan contamination. Indeed, minor peaks from terminal, 2-substituted, 3-substituted, 6-substituted, and 2,6-disubstituted hexoses were identified in methylation analysis. Most from them are typical of bacterial mannans, whereas terminal and 3-substituted hexoses could originate from the lipopolysaccharide core that was attached to the OPS.

The ^{1}H NMR and ^{13}C NMR (Fig. 1) spectra of the polysaccharide contained signals of different intensities, evidently owing to the mannan contamination of the OPS. The major series in the ^{13}C NMR spectrum included signals for three C-CH_2OH groups (C-6 of hexoses) at δ 61.3–62.6, four C-CH_3 groups (C-6 of Rha and Qui4N, C-2 of NAc, and C-4 of 3Hb) at δ 18.0–23.5, two nitrogen-bearing carbons (C-2 of GlcN and C-4 of Qui4N) at δ 54.2 and 57.4, C-2 and C-3 of 3Hb at 46.2 and 66.2 ppm, respectively, and two CO groups of the *N*-acyl substituents at δ 174.9 and 175.5; signals for anomeric carbons were at δ 98.5–103.6 and other sugar ring carbons in the region δ 66.9–82.9. The ^{1}H NMR spectrum of the polysaccharide showed, inter alia, major signals for three CH_3-C groups (H-6 of Rha and Qui4N and H-4 of 3Hb) at δ 1.16–1.24, H-2 of 3Hb at δ 2.34, and NAc at δ 2.07.

Assignment of the NMR spectra of the OPS was complicated by the mannan admixture. Whereas acidic polysaccharides can be easily purified by anion-exchange gel chromatography, this was not the case of the polysaccharides from *E. coli* O39 as both were neutral. To solve the problem the OPS was cleaved selectively by two methods.

Smith degradation of the polysaccharide mixture, including periodate oxidation, borohydride reduction, and mild acid hydrolysis, afforded an oligosaccharide glycoside (OS-1) with the 1-deoxyerythritol (1dEry-ol) aglycon derived evidently from 4-substituted Rha. It was isolated by GPC on TSK HW-40 (S) and studied by NMR spectroscopy and ESI MS.

The ^{1}H NMR and ^{13}C NMR (Fig. 2) spectra of OS-1 were assigned (Table 1) using 2D homonuclear ^{1}H,^{1}H COSY, TOCSY, ROESY, and heteronuclear ^{1}H,^{13}C HSQC and HMBC experiments. OS-1 was found to contain one residue each of GlcN, Qui4N, and Man called units **A**–**C**, respectively, as well as 1dEry-ol and both *N*-acyl substituents, including NAc and 3Hb groups. The amino sugars were confirmed by correlations of protons at the nitrogen-linked carbons with the corresponding carbons, namely C-2 of GlcN at δ 3.88/54.8 and C-4 of Qui4N at δ 3.79/57.7. The composition of OS-1 was confirmed by the negative ion ESI mass spectrum, which revealed a singly charged $[M-H]^-$ ion at *m*/*z* 701.2981 (the calculated molecular mass of OS-1 is 702.3059 Da).

A relatively large $J_{1,2}$ coupling constant of ~7 Hz and a relatively small $J_{1,2}$ value of <4 Hz indicated that Qui4N was β-linked and GlcN was α-linked, respectively. The C-5 chemical shift δ 74.1 demonstrated the α configuration of Man (compare published data[7] δ

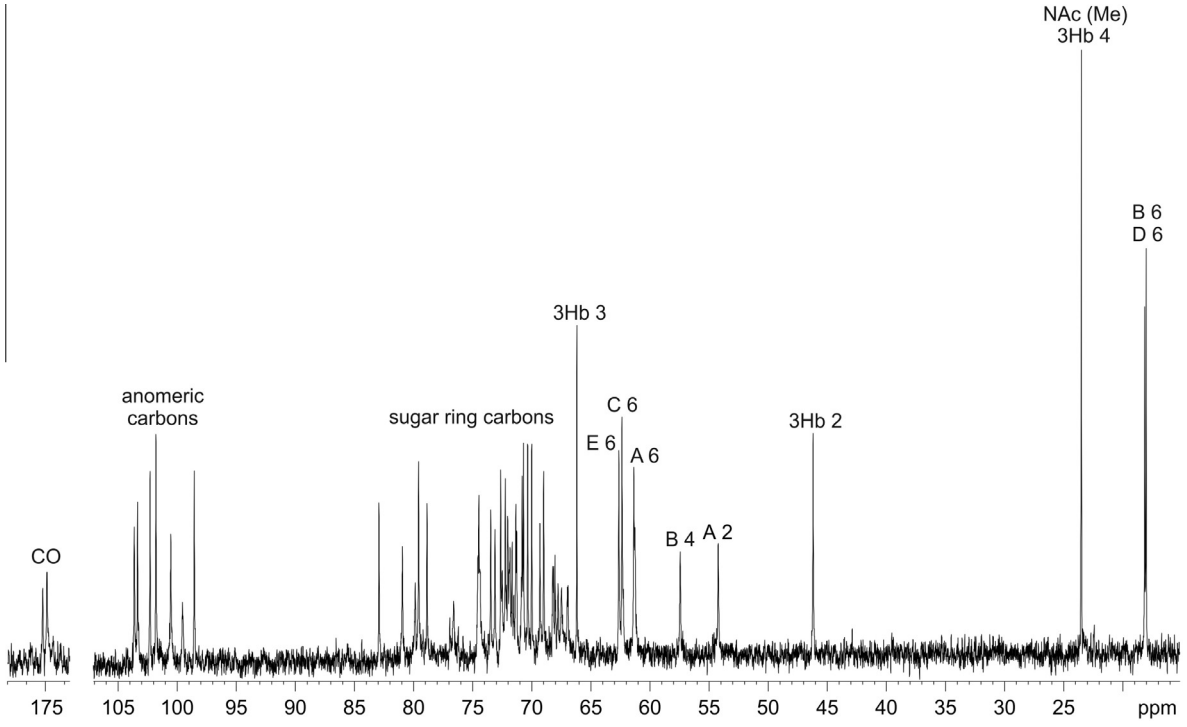


Figure 1. ^{13}C NMR spectrum of the OPS/mannan mixture from *E. coli* O39.