Note

Structure elucidation and biosynthesis gene cluster organization of the O-antigen of *Escherichia coli* O170

Alexander S. Shashkov,[a] Xi Guo,[b] Andrei V. Perepelov,[a,*] Andrej Weintraub,[c] Bin Liu,[b] Göran Widmalm,[d] 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, 300457 Tianjin, China*

[c]*Department of Laboratory Medicine, Division of Clinical Microbiology, Karolinska Institute, Karolinska University Hospital, S-141 86 Stockholm, Sweden*

[d]*Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden*

Abstract

Enterotoxigenic *Escherichia coli* are causative agents of diarrhea in humans as well as animals, and *E. coli* O170 belongs to this virotype. Upon mild acid degradation of the lipopolysaccharide of *Escherichia coli* O170, the branched O-polysaccharide chain was partially cleaved at β-D-glactofuranosidic linkages to give multiple products, including a linear tetrasaccharide and oligomers thereof. Studies of the acid degradation products and O-deacylated lipopolysaccharide by 1D and 2D ^{1}H and ^{13}C NMR spectroscopy enabled elucidation of the following O-polysaccharide structure:

→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*A-(1→3)-β-D-Gal*f*-(1→3)-β-D-Glc*p*NAc-(1→
3
↑
1
β-D-Gal*f*

Functions of genes in the O-antigen biosynthesis gene cluster were tentatively assigned and found to be in agreement with the O-polysaccharide structure.

* Corresponding author: Tel.: +7 499 1376148; fax: +7 499 1355328.

E-mail address: perepel@ioc.ac.ru (Andrei V. Perepelov).

Keywords**:** *Escherichia coli;* lipopolysaccharide; O-polysaccharide; bacterial polysaccharide structure; O-antigen gene cluster

Based on the O-antigens (or O-polysaccharides), about 180 O-serogroups of *Escherichia coli* have been recognized.[1] For most of them, O-polysaccharide structures have been elucidated (see *E. coli* O-antigen database at http://www.casper.organ.su.se/ECODAB/)[2] but several of them still remain to be established.

Strains of serogroup O170[3] belong to the virotype Enterotoxigenic *Escherichia coli* and cause diarrhea in dogs and lambs.[4,5] In this work, we have established the O-polysaccharide structure of *E. coli* O170 and found it to be in agreement with tentatively assigned functions of glycosyltransferases whose genes are present in the O-antigen biosynthesis gene cluster.

Lipopolysaccharide (LPS) was isolated from bacterial cells of *E. coli* O170 and hydrolyzed with mild acid to give a polysaccharide, which was isolated by GPC. Its analysis by ^{1}H and ^{13}C NMR spectroscopy showed an irregularity, which could be caused by cleavage of an acid-labile linkage(s). The polysaccharide was subjected to a prolonged hydrolysis, the products were fractionated by GPC and three fractions, I-III, were collected. Alternatively, the LPS was O-deacylated by treatment with aqueous ammonia. Analysis of sugar components after full acid hydrolysis of the O-deacylated LPS (LPS-OH) revealed D-Gal, D-GlcA, and D-GlcN.

The ^{13}C NMR spectrum of the LPS-OH (Fig. 1, top) showed signals of the O-polysaccharide chain, including signals for five anomeric carbons in the region δ 101.2-109.6, where two of which at δ 108.9 and 109.6 derived from furanosides.[6] There were also signals for four C-*C*H$_2$OH groups (C-6 of Gal and GlcN) at δ 61.2-64.3, two nitrogen-bearing carbons (C-2 of GlcN) at δ 56.4 and 56.6, 18 oxygen-bearing sugar ring carbons in the region δ 69.2-85.8, one CO$_2$H group (C-6 of GlcA) at δ 174.8, and two *N*-acetyl group at δ 23.4, 23.7 (both Me), 175.5 and 176.6 (both CO). The ^{1}H NMR spectrum showed signals for five anomeric protons at δ 4.58-5.16, other sugar protons in the region δ 3.36-4.21, and two *N*-acetyl groups at δ 2.02 and 2.03. Therefore, the O-antigen has a pentasaccharide repeating unit containing two residues each of D-Gal and D-GlcNAc, and one residue of D-GlcA.

The ^{1}H and ^{13}C NMR spectra of the O-polysaccharide of the LPS-OH were assigned using 2D COSY, TOCSY, ^{1}H,^{13}C-HSQC, and HSQC-TOCSY experiments (Table 1). Based on ^{13}C NMR chemical shifts, intra-residue ^{1}H,^{1}H and ^{1}H,^{13}C correlations, and $^{3}J_{H,H}$ coupling constant values estimated from the 2D NMR spectra, five spin-systems were identified, including two each for β-Gal*f* (units **B** and **E**) and β-Glc*p*NAc (units **A** and **D**), and one for β-Glc*p*A (unit **C**). The

spin systems for the GlcNAc residues were recognized by correlations between protons at the nitrogen-bearing carbons (H-2) at δ 3.82 and the corresponding carbons (C-2) at δ 56.4 and 56.6 in the $^{1}H,^{13}C$ HSQC spectrum. The ^{1}H NMR chemical shifts δ 4.58-4.60 and relatively large $J_{1,2}$ coupling constants ~8 Hz confirmed that units **A**, **C**, and **D** are β-linked, and the C-1 chemical shifts δ 108.9 and 109.6 of units **B** and **E** indicated β-linkages of both Gal*f* residues.

A 2D $^{1}H,^{13}C$ HMBC experiment with the LPS-OH (Fig. 2) revealed interresidue cross-peaks at δ 4.58/75.2, 5.05/82.1, 4.60/85.8, 4.58/81.2, and 5.16/77.4, which were assigned to the following correlations between the anomeric protons and linkage carbons: **A** H-1/**D** C-4, **B** H-1/**A** C-3, **C** H-1/**B** C-3, **D** H-1/**C** C-4, and **E** H-1/**D** C-3, respectively. Interresidue **B** H-1/**B** C-4 and **E** H-1/**E** C-4 correlations at δ 5.05/83.4 and 5.16/85.1 confirmed the furanosidic form of these sugar residues.[7,8] These data were in agreement with downfield displacements of the signals for the linkage carbons[9] of units **A**–**D** (Table 1) as compared with their positions in the corresponding non-substituted monosaccharides.[6,10] Therefore, the O-polysaccharide has structure **1** shown in Chart 1.

Similar NMR spectroscopic studies of fractions I-III (the ^{13}C NMR spectra are shown in Fig. 1, middle and bottom; for assignments of the ^{1}H and ^{13}C NMR chemical shifts see Table 1) enabled elucidation of their structures shown in Chart 1. None of the compounds contained unit whereas the other sugar residues were present. As a result of cleavage of unit **E** from position 3 of unit **D**, the chemical shifts of unit **D** changed significantly as compared to the spectra of **1**. Particularly, the signal for C-3 shifted upfield from δ 77.4 to 73.4–73.5 and that for C-4 shifted downfield from δ 75.2 to 80.2-80.3.

Fraction III contained oligosaccharide **2**. Unit **A** that was 3-substituted in **1**, became terminal in **2** as followed from an upfield displacement of the signal for C-3 from δ 82.1 to 74.7. Unit **B** converted from the β-linked furanosidic form in **1** to the reducing pyranosidic form (**Bα** and **Bβ**) in **2** (Table 1). Therefore, oligosaccharide **2** is a linear tetrasaccharide with the structure shown in Chart 1.

Fractions I and II were found to include oligomers of **2**, designated as **3**. NMR spectra of **3** showed major signals for the interior repeating unit, chemical shifts of units **A**–**C** being similar to those in **1** and of unit **D** similar to **2** (Table 1). Minor signals belonged to the terminal sugar residues (Glc*p*NAc **A** and Gal*p* **Bα** and **Bβ**), and their chemical shifts were essentially identical to those in **2**. Therefore, oligosaccharides **3** have the structure shown in Chart 1. As judged by relative intensities of the NMR signals of the terminal and interior monosaccharides, fractions I and II oligosaccharides **3** included on the average 4 and 2.5 (**2** and **3**) repeating units, respectively.

The structures of **2** and **3** are in agreement with the structure of the O-polysaccharide **1** and showed that these oligosaccharides resulted from full cleavage of the side-chain Gal*f* (unit **E**) and full (**2**) or partial (**3**) splitting of the Gal*f* (unit **B**) linkage in the main chain.

The O-antigen gene cluster of *E. coli* O170 between conserved genes *galF* and *gnd* (GenBank accession number AB812070)[11] includes seven genes, which all have the same transcriptional direction from *galF* to *gnd*. D-Glc, D-Gal, and D-GlcNAc are common sugars in bacteria, and genes for synthesis of their nucleotide precursors are usually located outside the O-antigen gene cluster. In accordance with the occurrence in the O-polysaccharide of Gal*f* (units **B** and **E**), gene *glf* for conversion of UDP-D-Gal*p* into UDP-D-Gal*f* was present. There were four glycosyltransferase genes for the O-unit assembly and O-antigen processing genes, including *wzx* for flippase responsible for translocation of the lipid-linked O-unit through the inner membrane and *wzy* for O-antigen polymerase that mediates O-unit polymerization. Gene *ugd* coding for dehydrogenase that converts UDP-D-Glc*p* to UDP-D-Glc*p*A is located between genes *gnd* and *wzz* that encodes the O-antigen chain length regulator. Therefore, the O170 polysaccharide structure is consistent with putatively assigned functions of genes in the O-antigen gene cluster of this serotype.

1. Experimental

1.1. Bacterial strain and isolation of the lipopolysaccharide

E. coli O170 strain CCUG 36538 was obtained from the Culture Collection, University of Göteborg, Sweden. The bacteria were grown and the LPS was isolated essentially as described.[12]

1.2. Degradation of the lipopolysaccharide

Mild acid hydrolysis of the LPS (40 mg) was performed with aq 2% AcOH (2 mL) at 100 °C until precipitation of a lipid (1 h). A polysaccharide (19 mg) was isolated from the supernatant by GPC on a column (56 × 2.6 cm) of Sephadex G-50 Superfine (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer pH 4.5 monitored by a differential refractometer (Knauer, Germany). It was further hydrolyzed for 6 h under the same conditions, and fractions I (2.0 mg), II (4.4 mg) and III (2.9 mg) were isolated by GPC on a column (80 × 1.6 cm) of TSK HW-40 (S) in aq 1% AcOH monitored as above.

An LPS sample (45 mg) was treated with aq 25% ammonia (2 mL) at 37 °C for 8 h and O-deacylated LPS (LPS-OH) (23 mg) was isolated from the supernatant by GPC on Sephadex G-50 Superfine as described above.

1.3. Chemical analyses

A sample of LPS-OH was hydrolyzed with 2 M CF_3CO_2H (120 °C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on a Maestro 7820 GC instrument (Interlab, Russia) equipped with a HP-5ms column using a temperature program of 160 °C (1 min) to 290 °C at 7 °C min^{-1}. The absolute configuration of the monosaccharides, including detection of GlcA, was determined by GLC of the acetylated (*S*)-2-octyl glycosides as described.[13]

1.5. NMR spectroscopy

Samples were deuterium-exchanged by freeze-drying from 99.9 % D_2O and then examined as solutions in 99.95 % D_2O. NMR spectra were recorded on a Bruker Avance II 600 MHz spectrometer (Germany) at 30 °C. ^1H and ^{13}C NMR chemical shifts were measured with internal sodium 3-(trimethylsilyl)propanoate-2,2,3,3-d_4 (δ_H 0, δ_C −1.6) as reference. 2D NMR spectra were obtained using standard Bruker software, and Bruker TopSpin 2.1 program was used to acquire and process the NMR data. A mixing time of 100 ms was used in 2D TOCSY experiments, and ^1H,^{13}C HMBC experiments were optimized for a coupling constant $J_{H,C}$ of 8 Hz.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (project 14-04-01709_a), the International Science & Technology Cooperation Program of China (2012DFG31680 and 2013DFR30640), the National Key Program for Infectious Diseases of China (2013ZX10004216-001-001 and 2013ZX10004221-003), National Natural Science Foundation of China (NSFC) Program (31371259 and 81471904), Research Project of Chinese Ministry of Education (NO.113015A) for X. G., and the Swedish Research Council for G. W.

References

1. Reeves PR, Wang L. *Curr Top Microbiol Immunol* 2002;**264**:109–35.
2. Rojas-Macias MA, Ståhle J, Lütteke T, Widmalm G. *Glycobiology* 2015;**25**:341–7.
3. Ørskov J, Ørskov F, Rowe B. *Acta Path Microbial. Immunol Scand. Sect. B* 1984;**92**:189-193.
4. Holland RE, Walker RD, Sriranganathan N, Wilson RA, Ruhl DC, *Vet Microbiol* 1999;**70**:261-268.
5. Bandyopadhyay S, Mahanti A, Samanta I, Dutta TK, Ghosh MK, Bera AK, Bandyopadhyay S, Bhattacharya D *Trop Anim Health Prod* 2011;**43**:705-710.
6. Bock K, Pedersen C. *Adv Carbohydr Chem Biochem* 1983;**41**:26-66.
7. Linnerborg M, Wollin R, Widmalm G. *Eur J Biochem* 1997;**246**:565-573.
8. Svensson MV, Zhang X, Huttunen E, Widmalm G. *Biomacromolecules* 2011;12:2496-2501.
9. Söderman P, Jansson P-E, Widmalm G. *J Chem Soc, Perkin Trans 2* 1998: 639-648
10. Jansson P.-E., Kenne L, Widmalm G. *Carbohydr Res* 1989;**188**:169-191.
11. Iguchi A, Iyoda S, Kikuchi T, Ogura Y, Katsura K, Ohnishi M, Hayashi T, Thomson NR. *DNA Res.*, 2014;**22**:105-7.
12. Svensson MV, Weintraub A, Widmalm G. *Carbohydr Res* 2011;**346**:449–453.
13. Leontein K, Lönngren J. *Methods Carbohydr Chem* 1993;**9**:87–9.

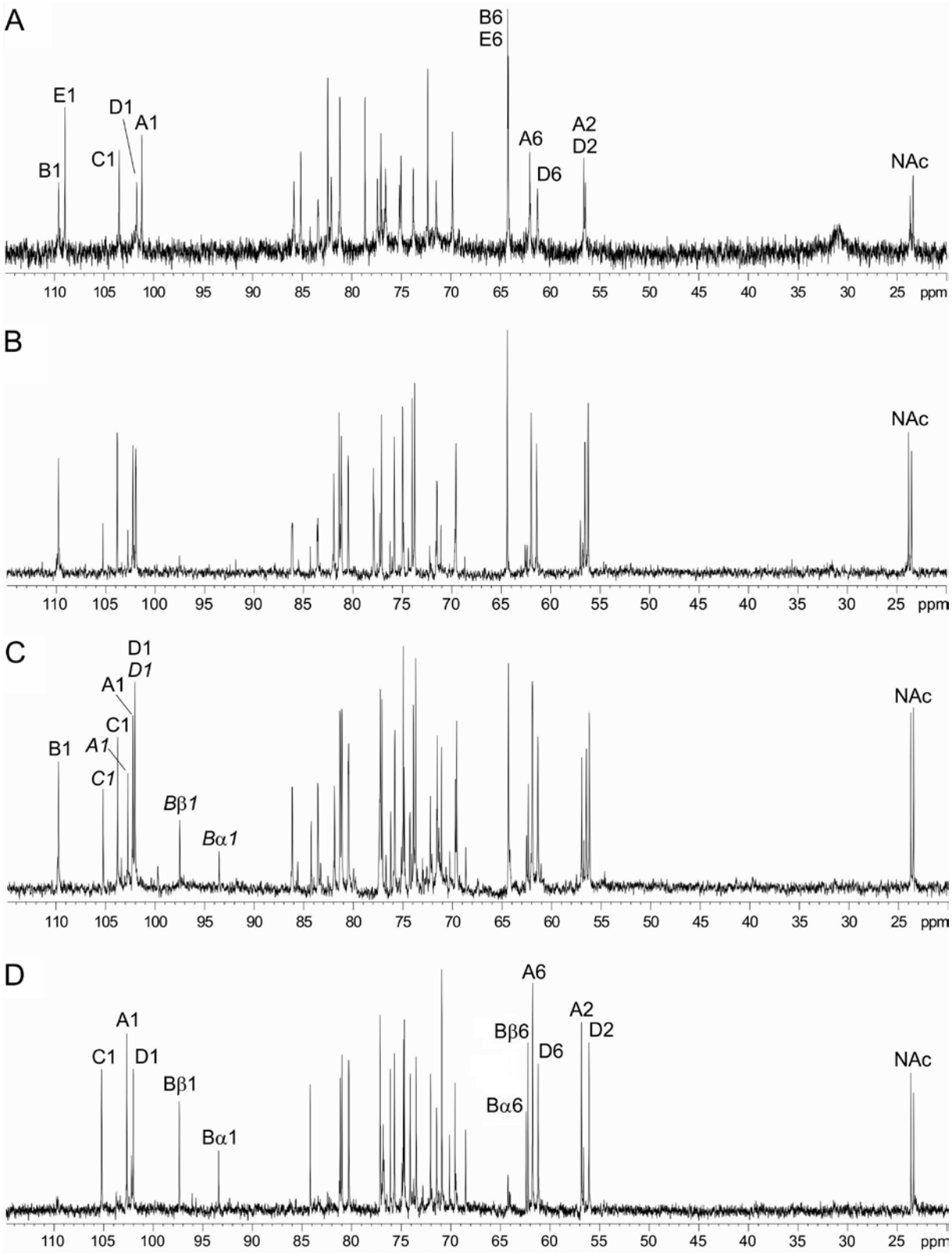

Fig. 1. ^{13}C-NMR spectra of the LPS-OH (A), fractions I and II oligosaccharides **3** (B and C, respectively), and fraction III oligosaccharide **2** (D) from *E. coli* O170. Signals for CO groups are not shown. Arabic numerals refer to carbons in sugar residues denoted by letters as shown in Table 1 and Chart 1. In panel C, peak annotations for sugar anomeric carbons in terminal repeats are italicized.

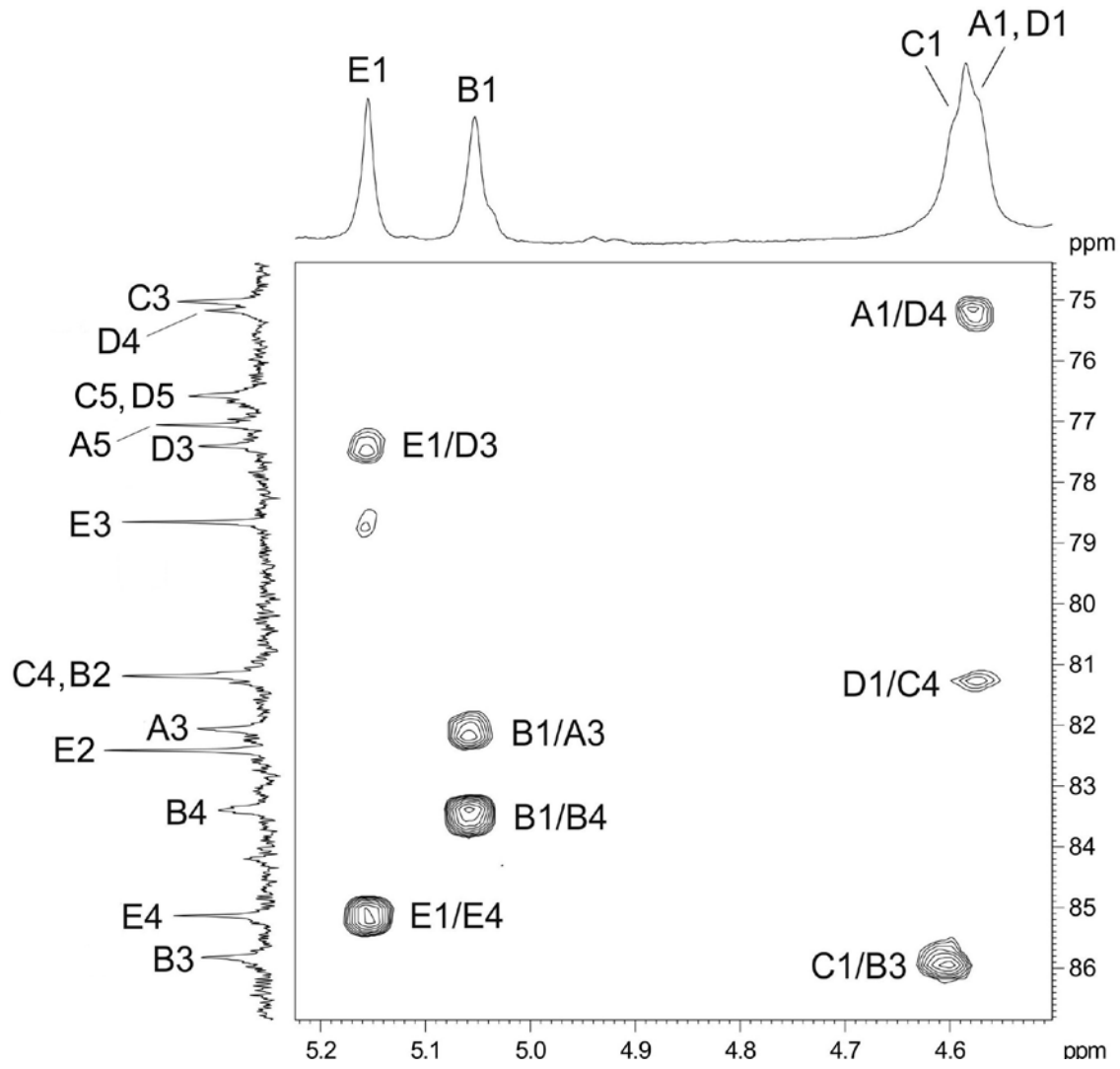

Fig. 2. Part of a $^{1}H,^{13}C$ HMBC spectrum of the LPS-OH from *E. coli* O170. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are displayed along the horizontal and vertical axes, respectively. Arabic numerals before and after oblique stroke refer to protons and carbons, respectively; sugar residues are denoted by letters as shown in Table 1. Interresidue cross-peaks are annotated in bold face.

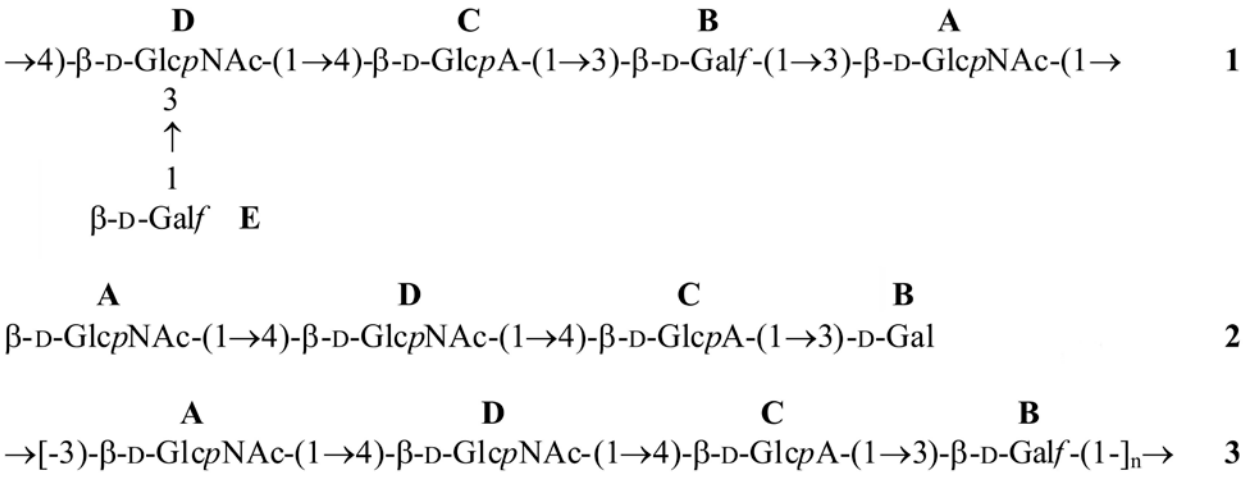

Chart 1. Structures of the O-polysaccharide **1**, fraction III oligosaccharide **2**, and fractions I and II oligosaccharides **3** from *E. coli* O170. In **3**, n = 2-4.

Table 1. ^{1}H and ^{13}C NMR data (δ, ppm).

Sugar residue	Nucleus	1	2	3	4	5	6
LPS-OH (O-polysaccharide **1**)[a]							
→3)-β-D-GlcpNAc-(1→	^{1}H	4.58	3.82	3.69	3.45	3.42	3.74, 3.90
A	^{13}C	101.2	56.4[b]	82.1	69.2	77.1	62.1
→3)-β-D-Gal*f*-(1→	^{1}H	5.05	4.18	4.20	4.18	3.87	3.60, 3.63
B	^{13}C	109.6	81.2	85.8	83.4	71.5	64.2[c]
→4)-β-D-Glc*p*A-(1→	^{1}H	4.60	3.36	3.61	3.75	3.83	
C	^{13}C	103.5	73.8	75.0	81.2	76.6	174.8
→3,4)-β-D-Glc*p*NAc-(1→	^{1}H	4.58	3.82	3.81	3.82	3.51	3.72, 3.90
D	^{13}C	101.7	56.6[b]	77.4	75.2	76.7	61.3
β-D-Gal*f*-(1→	^{1}H	5.16	4.01	4.01	4.21	3.81	3.67, 3.74
E	^{13}C	108.9	82.4	78.7	85.1	72.3	64.3[c]
Fraction I oligosaccharide **3** (interior repeat)[d]							
→3)-β-D-GlcpNAc-(1→	^{1}H	4.61	3.85	3.68	3.50	3.50	3.74, 3.91
A	^{13}C	102.1	56.4	81.7	69.4	77.0	61.8
→3)-β-D-Gal*f*-(1→	^{1}H	5.06	4.18	4.20	4.18	3.87	3.61, 3.63
B	^{13}C	109.7	81.2	86.2	83.4	71.4	64.2
→4)-β-D-Glc*p*A-(1→	^{1}H	4.63	3.36	3.64	3.77	3.91	
C	^{13}C	103.7	73.6	74.8	81.1	75.8	175.6
→4)-β-D-Glc*p*NAc-(1→	^{1}H	4.52	3.72	3.68	3.64	3.53	3.66, 3.85
D	^{13}C	102.3	56.1	73.4	80.2	75.6	61.2
Fraction III oligosaccharide **2**[e]							
→3)-α-D-Gal*p*	^{1}H	5.21	3.89	3.89	4.13	4.93	3.65
Bα	^{13}C	93.4	68.5	81.2	70.2	71.4	62.2
→3)-β-D-Gal*p*	^{1}H	4.56	3.56	3.71	4.07	3.63	3.68
Bβ	^{13}C	97.4	72.1	84.2	69.6	76.1	62.4
→4)-β-D-Glc*p*A-(1→	^{1}H	4.63[f]	3.39	3.57	3.71	3.74	
C	^{13}C	105.2	74.1	74.8	81.0	76.8	174.9
→4)-β-D-Glc*p*NAc-(1→	^{1}H	4.46	4.67	3.62	3.58	3.47	3.61, 3.80
D	^{13}C	102.2	56.1	73.5	80.3	75.7	61.2
β-D-GlcpNAc-(1→	^{1}H	4.52	4.68	3.50	3.41	3.43	3.64, 3.86
A	^{13}C	102.7	56.8	74.7	70.9	77.1	61.8

[a,d,e]Chemical shifts for the *N*-acetyl groups are [a]δ_H 2.02 and 2.03, δ_C 23.4, 23.7 (both Me), 175.5 and 175.6 (both CO); [d]δ_H 2.02 and 2.07, δ_C 23.3, 23.6 (both Me), 175.7 and 176.0 (both CO); [e]δ_H 1.97 and 2.00, δ_C 23.3, 23.6 (both Me), 175.9 and 176.1 (both CO).

[b,c]Assignment could be interchanged.

[f]Linked to α-D-Gal*p*; δ 4.62 when linked to β-D-Gal*p*.