Note

Structure and genetics of the O-antigen of *Escherichia coli* O169 related to the O-antigen of *Shigella boydii* type 6

Andrei V. Perepelov,[a,*] Alexander S. Shashkov,[a] Xi Guo,[b] Andrei V. Filatov,[a] Andrej Weintraub,[c] Göran Widmalm,[d] Yuriy A. Knirel,[a]

[a] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation*

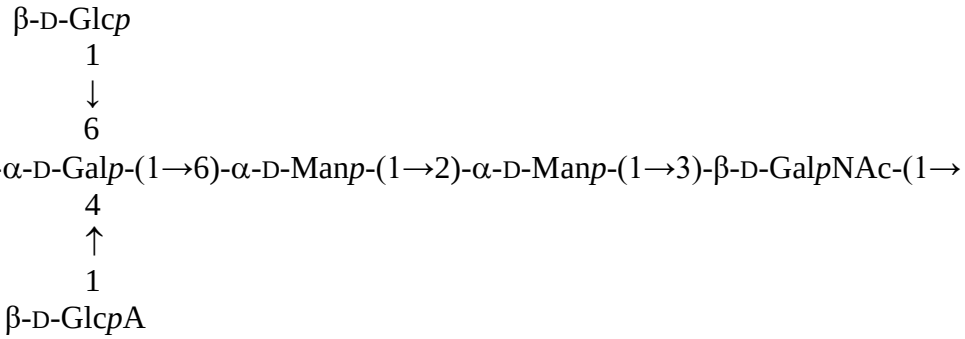
[b]*TEDA School of Biological Sciences and Biotechnology, Nankai University, TEDA, 300457 Tianjin, China*

[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

The O-polysaccharide (O-antigen) of *Escherichia coli* O169 was studied by sugar analysis along with 1D and 2D ^{1}H and ^{13}C NMR spectroscopy. The following structure of the branched hexasaccharide O-unit was established:



The O-polysaccharide of *E. coli* O169 differs from that of *Shigella boydii* type 6 only in the presence of a side-chain glucose residue. A comparison of the O-antigen biosynthesis gene clusters between the *galF* to *gnd* genes in the genomes of the two bacteria revealed their close relationship. The glycosyltransferase gene responsible for the formation of the β-D-Glc-(1→6)-α-D-Gal linkage in the O-antigen was identified in the clusters.

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

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

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

Escherichia coli is a clonal species, including both commensals and pathogens. The bacteria are normally identified by a combination of their O- and H- (and sometimes K-) antigens. Variations in types of sugars present, their arrangement and linkages make the O-antigen or O-polysaccharide (OPS) the most variable constituent on the cell surface and provide the basis for serotyping of the bacteria.[1] By now about 180 O-serotypes of *E. coli* have been recognized. Genes for O-antigen synthesis are normally located on the chromosome as an O-antigen gene cluster, and genetic variations in the cluster are the major basis for the diversity of the O-antigen forms. OPS structures for most serogroups have been elucidated (see *E. coli* O-antigen database at http://www.casper.organ.su.se/ECODAB/)[2] but several still remain to be established.

Serogroup O169 belongs to enterotoxigenic *E. coli* (ETEC) group, which has been recognized as a common cause of foodborne outbreaks.[3] *E. coli* O169 has notably become one of the most prevalent ETEC pathogens associated with foodborne outbreaks in many countries including the Republic of Korea,[4] Japan,[5]and the USA.[6] In this work, we established the OPS structure of *E. coli* O169 and compared the closely related O-antigen gene clusters of *E. coli* O169 and *Shigella boydii* type 6.

The OPS was obtained by mild acid degradation of the lipopolysaccharide of *E. coli* O169 and separated from lower molecular mass substances by GPC. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the OPS revealed Man, Glc, Gal, and GalN in the ratios ~2:1:0.5:0.5 (detector response). GLC analysis of the acetylated (*S*)-2-octyl glycosides demonstrated the D configuration of all constituent monosaccharides. Further NMR studies (see below) showed that the OPS also includes D-glucuronic acid (GlcA).

The ^{13}C-NMR spectrum of the OPS (Fig. 1) contained signals for six anomeric carbons in the region δ 96.7-105.3, five C-*C*H_2OH groups (C-6 of hexoses and GalN) at δ 62.8, 63.0, 63.2, 67.4, and 71.8 (data of attached-proton test), one nitrogen-bearing carbon (C-2 of GalN) at δ 53.2, one C-*C*O_2H group (C-6 of GlcA) at δ 176.6, and 23 oxygen-bearing sugar ring carbons in the region δ 65.7-81.3 as well as one *N*-acetyl group at δ 24.6 (CH_3) and 176.6 (CO). In the low-field region of the ^{1}H NMR spectrum, there were seven signals, including those for six anomeric protons at δ 4.46-5.13 and H-4 of a Gal residue at δ 4.47. The spectrum also contained signals for other sugar protons in the region δ 3.27-4.17 and one *N*-acetyl group at δ 2.02. These data indicated that the OPS has a hexasaccharide O-unit containing one residue each of D-Glc, D-Gal, D-GalNAc, D-GlcA, and two residues of D-Man.

The ^{1}H and ^{13}C NMR spectra of the OPS were assigned using 2D COSY, TOCSY, ^{1}H,^{13}C HSQC, HMBC, and HSQC-TOCSY experiments (Table 1). Based on intra-residue ^{1}H,^{1}H and ^{1}H,^{13}C correlations and coupling constant values estimated from the 2D NMR spectra, six spin

systems, including two systems each for *gluco*- (Glc and GlcA, denoted as units **A** and **B**), *galacto*- (Gal and GalNAc, units **C** and **F**) and *manno*-configurated sugars (two Man residues, units **D** and **E**) were recognized, all being in the pyranose form. The spin system for GalNAc was distinguished by a correlation between proton at the nitrogen-bearing carbon (H-2) and the corresponding carbon (C-2) at δ 4.04/53.2 in the HSQC spectrum. Unit **B** was identified as β-GlcA by a correlation of H-5 with C-6 (CO_2H) at δ 3.76/176.6 in the HMBC spectrum. A large $J_{1,2}$ coupling constant of ~7 Hz confirmed that units **A**, **B** and **F** are β-linked, whereas a significantly smaller value of ~3 Hz indicated the α-linkage of unit **C**. The position of signals for C-5 of units **D** and **E** at δ 72.7 and 75.6, respectively, indicated that both Man residues are α-linked (compare published data δ 73.34 and 77.00 for α- and β-Man*p*, respectively[7]).

The OPS was subjected to a Smith degradation and the products were fractionated by GPC. The main oligosaccharide product was studied by 1D and 2D NMR spectroscopy, including the full assignment of the ^{1}H and ^{13}C NMR signals (Fig. 1, Table 1). As a result, the following structure was elucidated:

β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→1)-Gro

F **C** **D'**

where Gro (unit **D'**) indicates glycerol derived from 6-substituted Man.

Subsequently, analysis of the OPS revealed relatively low-field positions of the signals for C-2 of unit **E**, C-3 of unit **F**, C-6 of unit **D**, C-3, 4, and 6 of unit **C** δ 80.7, 77.6, 67.4, 81.3, 78.3, and 71.8, as compared with their positions in the corresponding non-substituted monosaccharides at δ 71.69, 72.01, 61.99, 70.13, 70.28, and 62.04,[7] respectively, showed that the OPS is branched with unit **C** at the double-branching point and demonstrated the modes of glycosylation of the monosaccharide residues. In accordance with the side-chain position of Glc and GlcA, their C-2–C-6 chemical shifts were close to those in the corresponding non-substituted monosaccharides.[7]

A 2D ROESY experiment revealed strong interresidue cross-peaks at δ 4.46/3.86, 4.06; 4.94/4.47; 4.97/3.54; 5.06/4.01; 5.13/3.83, and 4.72/4.07, which were assigned to the following correlations between the anomeric protons and protons at the linkage carbons: **A** H-1/**C** H-6a, 6b; **B** H-1/**C** H-4; **C** H-1/**D** H-6a; **D** H-1/**E** H-2; **E** H-1/**F** H-3, and **F** H-1/**C** H-3, respectively. These data defined the monosaccharide sequence in the OPS, which was confirmed by a heteronuclear $^{1}H,^{13}C$ HMBC experiment, which showed correlations between the anomeric protons and linkage carbons (Fig. 2) and *vice versa*. Notably, the ^{13}C glycosylation shift of C-1 in residue **E** is small and the chemical shift of C-4 in residue **F** is shifted upfield compared to that of the monosaccharide, due to a γ-gauche effect,[8] consistent with the above deduced structural element

E-F, which also is present in the O-antigen from *E. coli* O175.[8] Therefore, the OPS of *E. coli* O169 has the structure shown in Chart 1.

The structure of the OPS of *E. coli* O169 differs from the known OPS structure of *Shigella boydii* type 6[9] in the presence of a side-chain glucose residue only (Chart 1).

The O-antigen gene cluster of *E. coli* O169 (GenBank accession number AB812069) includes nine genes, which all have the same transcriptional direction from *galF* to *gnd*, and three insertion sequence (IS) elements (Fig. 3). *manC* and *manB* are two genes that encode mannose-1-phosphate guanylyltransferase (ManC) and phosphomannomutase (ManB), respectively. Together with phosphomannose isomerase ManA, which is encoded by a gene located outside the O-antigen gene cluster, these two enzymes are involved in the synthesis of GDP-D-Man, the nucleotide precursor of D-Man that is present in the OPS. D-Glc and D-Gal are common sugars in bacteria, and genes for synthesis of their nucleotide precursors are usually located outside the O-antigen gene cluster.[10]

wecA gene for transfer of D-GlcNAc 1-phosphate from UDP-D-GlcNAc to the undecaprenyl phosphate (UndP) carrier to initiate the O-antigen synthesis is located outside the O-antigen gene cluster too. In *E. coli* O169, the first sugar of the O-unit is D-GalNAc, and *gnu* gene for the 4-epimerase that catalyzes conversion of UndP-D-GlcNAc to UndP-D-GalNAc,[11] is found upstream of *galF*. In accordance with the hexasaccharide O-unit, there are five glycosyltransferase genes: *orf1*, *orf2*, *orf4*, *orf5*, and *orf9*, in the O-antigen gene cluster. The presence of the flippase (Wzx) and O-antigen polymerase (Wzy) genes for translocation and polymerization of the O-unit, respectively, suggest that the O-antigen biosynthesis in *E. coli* O169 is mediated by the Wzx/Wzy-dependent process.

A comparison of the O-antigen gene clusters of *E. coli* O169 and *S. boydii* type 6 (GenBank accession number AF402314) shows that the first seven genes, including four glycosyltransferase genes, share 96% to 100% both DNA and protein identities (Fig. 3). Therefore, it is most likely that the fifth predicted glycosyltransferase gene, *orf9*, at the 3'-end of the O169 gene cluster is responsible for the formation of the β-D-Glc*p*-(1→6)-α-D-Gal*p* linkage, which is present in the *E. coli* O169-antigen only. The three IS elements are located between the *manB* and *wzx* genes in the O169-antigen gene cluster, and the same region in *S. boydii* type 6 includes an IS629 element, which interrupts the ribosyltransferase gene (*wbaM*).[10] These findings show a close relationship between the O-antigen gene clusters of *E. coli* O169 and *S. boydii* type 6 and suggest their origin from a common ancestor followed by different recombination events.

1. Experimental

1.1. Bacterial strain and isolation of the lipopolysaccharide

The strain of *E. coli* O169:H8:K- was obtained from the International Escherichia and Klebsiella Center (World Health Organization), Statens Serum Institute, Copenhagen, Denmark. The bacteria were grown and the LPS was isolated essentially as previously described.[8]

1.2. Isolation of the O-polysaccharide

Delipidation of the lipopolysaccharide (98 mg) was performed with aq 2% AcOH (6 mL) at 100 °C until precipitation of lipid A. The precipitate was removed by centrifugation (13,000 × *g*, 20 min), and the supernatant fractionated 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). A high-molecular-mass OPS was obtained in a yield of 32% of the lipopolysaccharide mass.

1.3. Chemical analyses

The O-polysaccharide 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 was determined by GLC of the acetylated (*S*)-2-octyl glycosides as described.[12]

1.4. Smith degradation

The OPS (10 mg) was oxidized with 0.1 M $NaIO_4$ (2 mL) in the dark for 72 h at 20 °C, reduced with an excess of $NaBH_4$ and desalted by dialysis against distilled water. The product was hydrolyzed with aq 2% HOAc for 2 h at 100 °C, and an oligosaccharide (3 mg) was isolated by GPC on TSK HW-40 (S) in aq 1% AcOH monitored as above.

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 a 150-ms duration of spin-lock time in 2D ROESY experiments. $^1H,^{13}C$ HMBC experiments were optimized for a coupling constant $J_{H,C}$ of 8 Hz.

Acknowledgements

This work was supported by the Russian Science Foundation (Project 14-14-01042) for Y. A. K., 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. (621-2013-4859).

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**:341–347.
3. Dalton CB, Mintz ED, Wells JG, Bopp CA, Tauxe RV. *Epidemiol Infect* 1999;**123**:9–16.
4. Cho SH, Kim J, Oh KH, Hu JK, Seo J, Oh SS, Hur MJ, Choi YH, Youn SK, Chung GT, Choe YJ. *Epidemiol Infect* 2014;**142**:616–23.
5. Harada T, Itoh K, Yamaguchi Y, Hirai Y, Kanki M, Kawatsu K, Seto K, Taguchi M, Kumeda Y. *Jpn J Infect Dis* 2013;**66**:530–3.
6. Beatty ME, Bopp CA, Wells JG, Greene KD, Puhr ND, Mintz ED. *Emerg Infect Dis* 2004;**10**:518–21.
7. Jansson P-E, Kenne L, Widmalm G. *Carbohydr Res* 1989;**188**:169–91.
8. Svensson MV, Weintraub A, Widmalm G. *Carbohydr Res* 2011;**346**:449–453.
9. Senchenkova SN, Feng L, Yang J, Shashkov AS, Cheng J, Liu D, Knirel YA, Reeves P, Jin Q, Ye Q, Wang L. *J Bacteriol* 2005;**187**:2551–4.
10. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Wang Q, Reeves PR, Wang L. *FEMS Microbiol Rev* 2008;**32**:627–53.
11. Cunneen MM, Liu B, Wang L, Reeves PR. *PLoS One* 2013;**14**:e67646.
12. Leontein K, Lönngren J. *Methods Carbohydr Chem* 1993;**9**:87–9.

Figures

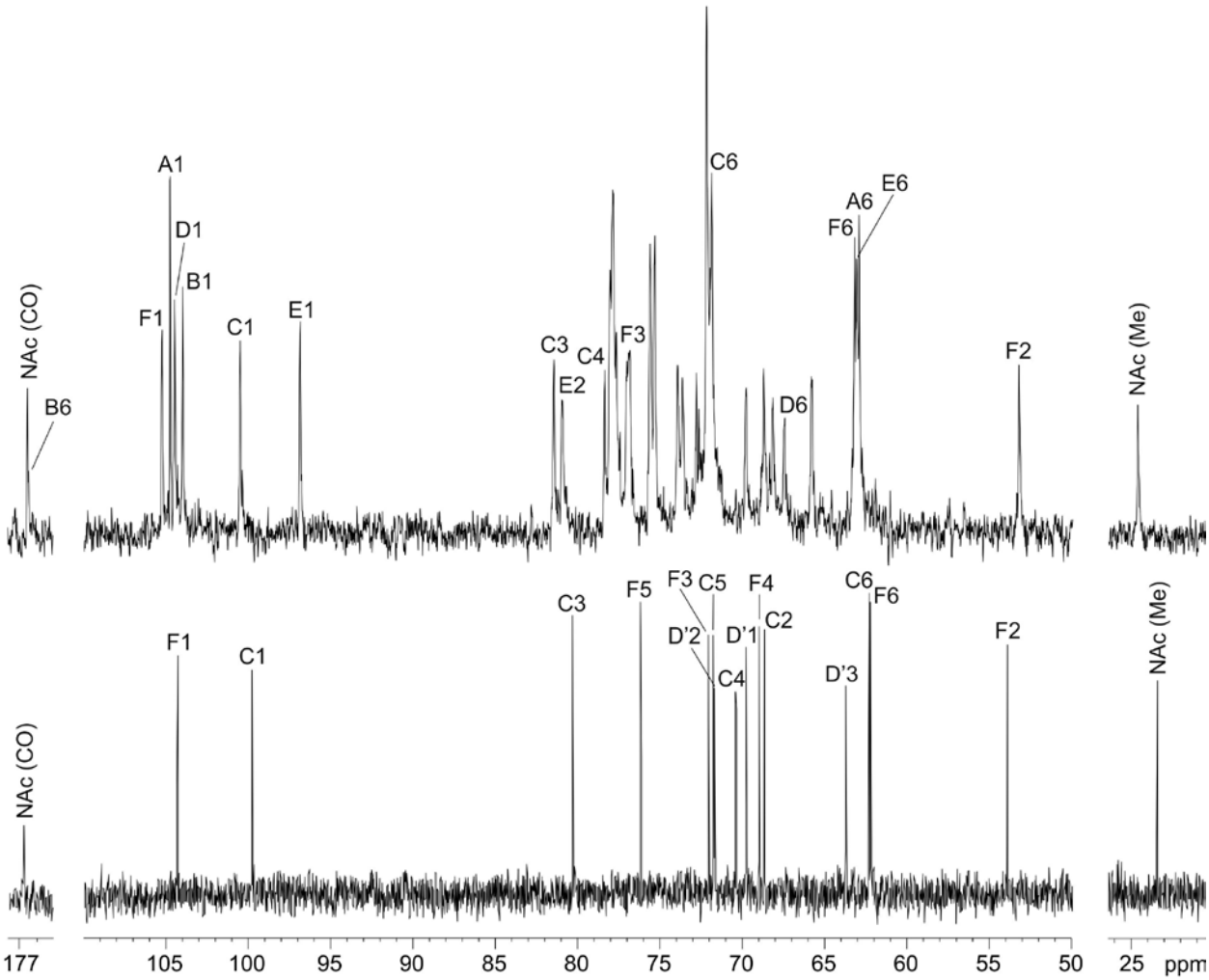


Fig. 1. ^{13}C-NMR spectra of the OPS (top) and oligosaccharide derived by Smith degradation of the OPS (bottom) from *E. coli* O169. Arabic numerals refer to carbons in sugar residues denoted by letters as shown in Table 1 and Chart 1.

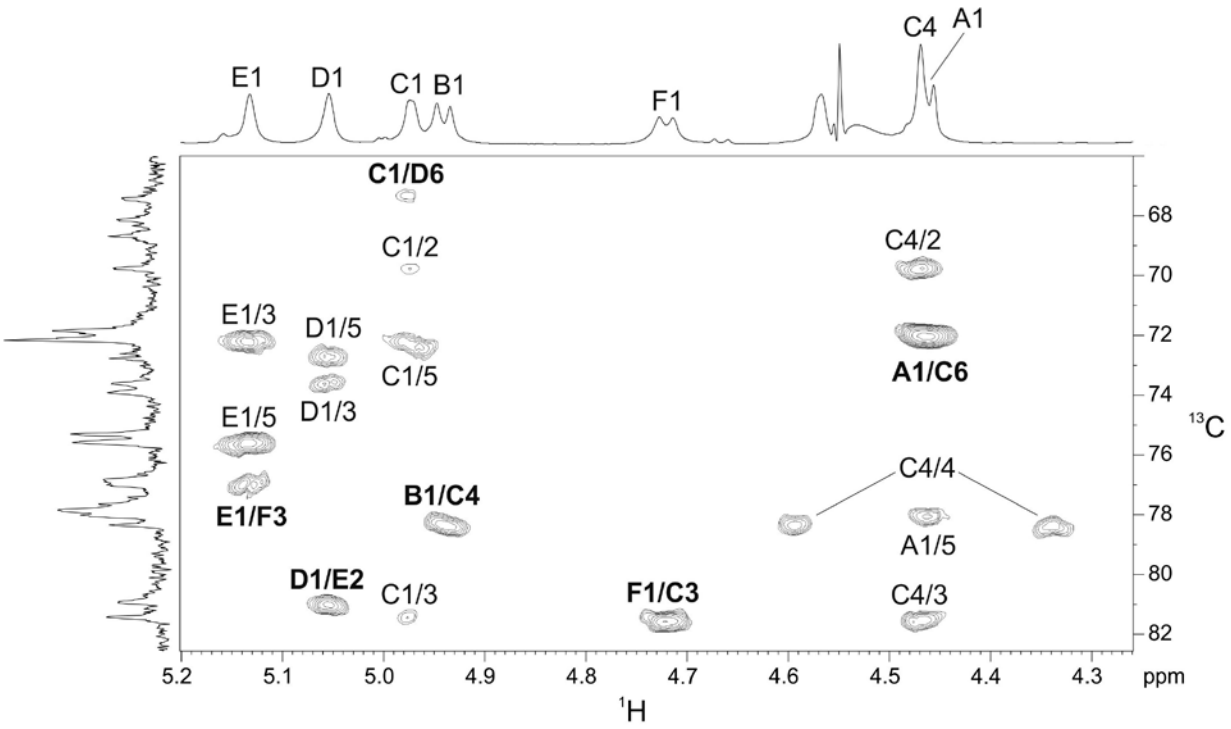


Fig. 2. Part of a $^{1}H,^{13}C$ HMBC spectrum of the OPS from *E. coli* O169. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are displayed along the horizontal and vertical axes, respectively. Arabic numerals before and after oblique stroke refer to protons and carbons, respectively; sugar residues are denoted by letters as shown in Table 1. Interresidue cross-peaks are annotated in bold face. Change the ^{1}H and ^{13}C labels so that ^{1}H will be below the proton chemical shift scale and that ^{13}C will be to the right of the carbon chemical shift scale.

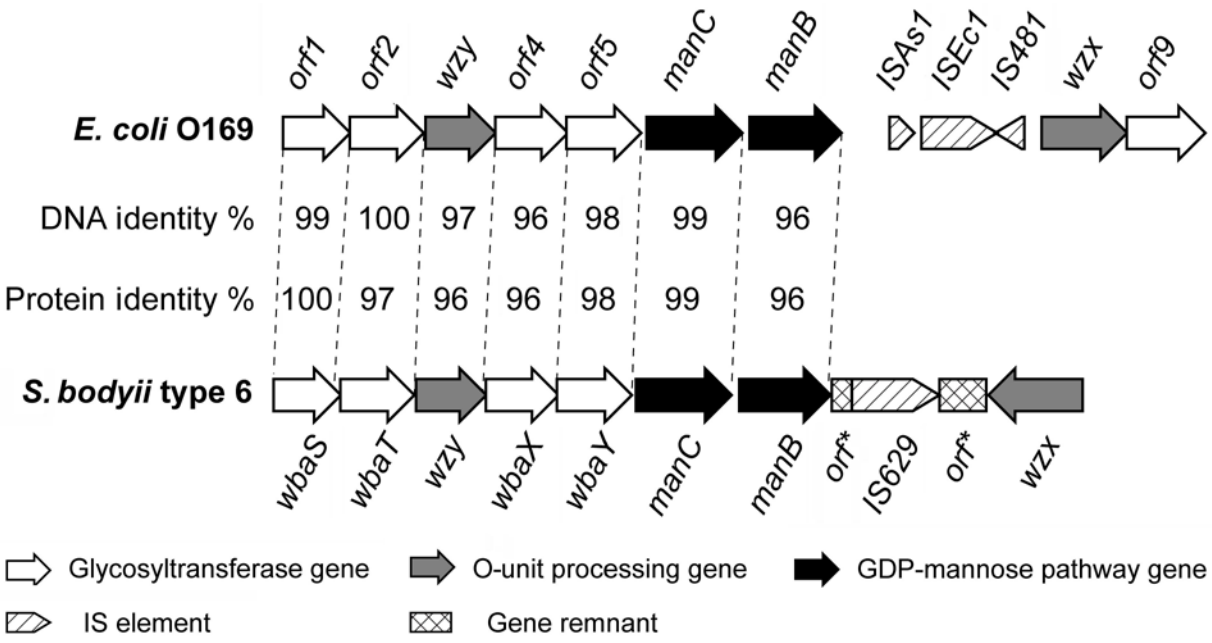


Fig. 3. Comparison of the O-antigen gene clusters of *E. coli* O169 and *S. boydii* type 6. Dotted lines indicate homologous genes between the two strains.

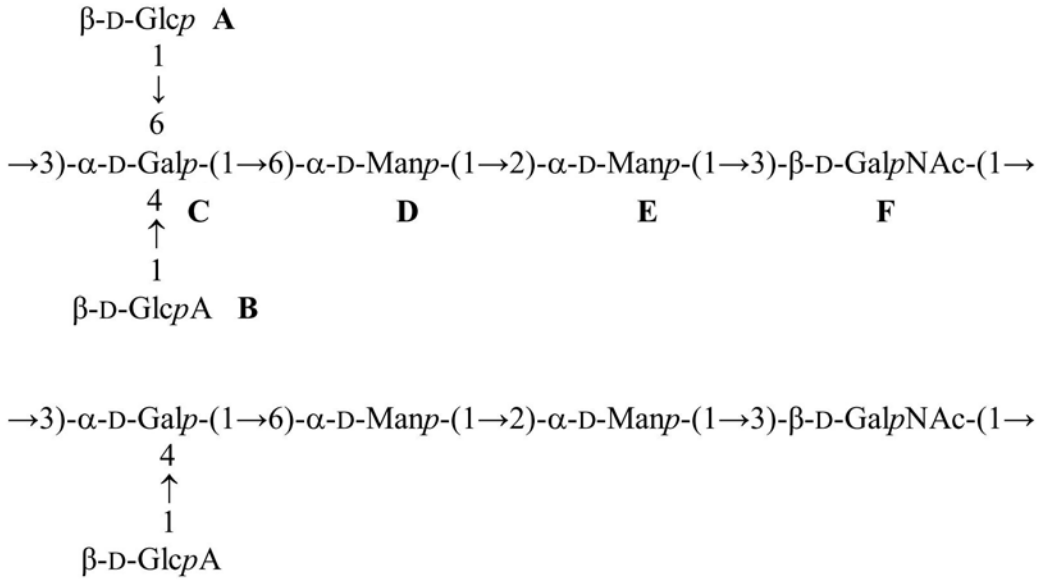


Chart 1. Related structures of the O-polysaccharides of *E. coli* O169 (top) and *S. boydii* type 6 (bottom).

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

Sugar residue	Nucleus	1	2	3	4	5	6
OPS from *E. coli* O169							
β-D-Glc*p*-(1→	^{1}H	4.46	3.27	3.48	3.38	3.42	3.79, 3.90
A	^{13}C	104.7	75.3	77.8	71.8	78.0	62.8
β-D-Glc*p*A-(1→	^{1}H	4.94	3.30	3.55	3.53	3.76	
B	^{13}C	103.9	75.6	77.7	73.9	77.6	176.6
→3,4,6)-α-D-Gal*p*-(1→	^{1}H	4.97	4.04	4.07	4.47	4.15	3.86, 4.06
C	^{13}C	100.4	69.7	81.3	78.3	72.1	71.8
→6)-α-D-Man*p*-(1→	^{1}H	5.06	4.08	3.87	3.95	3.83	3.54, 4.17
D	^{13}C	104.3	72.2	73.5	68.1	72.7	67.4
→2)-α-D-Man*p*-(1→	^{1}H	5.13	4.01	3.84	3.72	3.55	3.75, 3.84
E	^{13}C	96.7	80.7	72.5	68.4	75.6	63.0
→3)-β-D-Gal*p*NAc-(1→	^{1}H	4.72	4.04	3.83	4.13	3.64	3.79, 3.79
F	^{13}C	105.3	53.2	77.6	65.7	76.8	63.2
Oligosaccharide derived by Smith degradation of the OPS							
β-D-Gal*p*NAc-(1→	^{1}H	4.63	3.94	3.75	3.94	3.68	3.77, 3.80
F	^{13}C	104.3	53.8	72.0	68.9	76.1	62.1
→3)-α-D-Gal*p*-(1→	^{1}H	4.92	3.89	3.96	4.22	3.95	3.78, 3.78
C	^{13}C	99.7	68.6	80.3	70.4	71.7	62.2
→1)-Gro	^{1}H	3.58, 3.77	3.95	3.65, 3.69			
D'	^{13}C	69.7	71.6	63.6			

The chemical shifts for the *N*-acetyl group are δ_H 2.02, δ_C 24.6 (Me) and 176.6 (CO) in the OPS; δ_H 2.04, δ_C 23.4 (Me) and 176.6 (CO) in the oligosaccharide.