Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 342 (2007) 2676–2681
Carbohydrate RESEARCH

Note

Close relation of the O-polysaccharide structure of *Escherichia coli* O168 and revised structure of the O-polysaccharide of *Shigella dysenteriae* type 4

Andrei V. Perepelov,[a,*] Bin Liu,[b,c] Sof'ya N. Senchenkova,[a] Alexander S. Shashkov,[a] Lu Feng,[b,c] Yuriy A. Knirel[a] and Lei Wang[b,c]

[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, 23 HongDa Street, TEDA, Tianjin 300457, PR China*
[c]*Tianjin Key Laboratory for Microbial Functional Genomics, TEDA College, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*

Received 27 June 2007; accepted 16 August 2007
Available online 23 August 2007

Abstract—The O-polysaccharide was isolated from the lipopolysaccharide of *Escherichia coli* O168 and studied by chemical analyses and Smith degradation along with ^{1}H and ^{13}C NMR spectroscopies. The following structure of the branched pentasaccharide repeating unit of the O-polysaccharide was established:

α-L-Fuc*p*
1
↓
3
→4)-α-D-Glc*p*NAc6Ac-(1→4)-α-D-Glc*p*A-(1→3)-α-L-Fuc*p*-(1→3)-β-D-Glc*p*NAc-(1→

where 6-O-acetylation of GlcNAc is partial. Reinvestigation of the O-polysaccharide of *Shigella dysenteriae* type 4 established earlier showed it to have the same structure except for that the lateral Fuc residue is nonstoichiometrically O-acetylated at each position.

Keywords: *Escherichia coli*; *Shigella dysenteriae*; O-Antigen; Bacterial polysaccharide structure; Methodology

Escherichia coli is the predominant species in the human intestinal microflora and one of the most common causes of diarrhoeal diseases. *E. coli* serotypes are normally classified by a combination of their O and H (and sometimes K) antigens. For *Shigella*, only the O-antigen classification system is used, as they lack H and K antigens. *E. coli* and *Shigella* have long been known to be closely related, and the analysis of housekeeping gene sequences showed that most *Shigella* serotypes fall into three clusters within *E. coli*.[1] The O-antigen represents the polysaccharide chain (O-polysaccharide, OPS) of the lipopolysaccharide, which is usually built up of repeating units containing two to seven sugar residues and often also nonsugar substituents (e.g., amino acids, pyruvic acid acetals, lactic acid ethers, phosphate, *O*-acetyl groups, etc.).

Here we present a new structure of the O-polysaccharide of *E. coli* O168 and a revised structure of the O-polysaccharide of *Shigella dysenteriae* type 4, which differ only in the O-acetylation pattern.

The O-polysaccharide (OPS) of *E. coli* O168 was obtained by mild acid degradation of the lipopolysac-

* Corresponding author. Tel.: +7 095 9383613; fax: +7 095 1355328; e-mail: perepel@ioc.ac.ru

0008-6215/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.carres.2007.08.005

charide, which was isolated from dried cells by the phenol–water procedure, and separated from lower molecular mass substances by gel-permeation chromatography. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the OPS revealed Fuc and GlcN (from GlcNAc) in the ratios ~1:1. GLC analysis of the acetylated (*S*)-2-octyl glycosides demonstrated the D configuration of GlcN and L configuration of Fuc.

The ^{13}C NMR spectrum of the OPS (Fig. 1a) contained signals having different intensities, most likely, owing to nonstoichiometric O-acetylation (there was a signal for CH_3 of an *O*-acetyl group at δ 21.4). The ^{1}H NMR spectrum of the OPS showed signals for one *O*-acetyl and two *N*-acetyl groups at δ 2.17, 2.03, and 2.04 in the ratios 0.3:1:1, respectively.

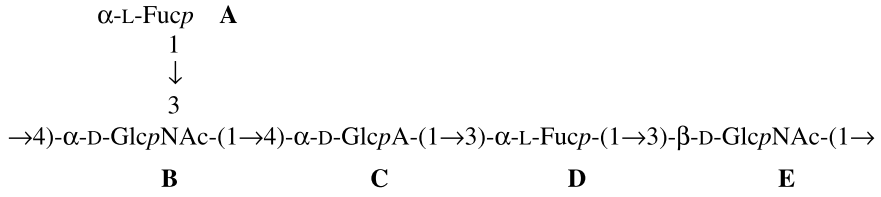
Therefore, the OPS was subjected to O-deacetylation, and the resultant O-deacetylated polysaccharide (DPS) was found to be regular. Its ^{13}C NMR spectrum (Fig. 1b) showed signals for five anomeric carbons in the region δ 99.6–102.0, two HO*C*H_2–C groups (C-6 of GlcN) at δ 60.8–63.2, two nitrogen-bearing carbons (C-2 of two GlcN) at δ 55.3 and 57.0, two *C*H_3–C groups (C-6 of two Fuc) at δ 16.5 and 16.9, one C–*C*O_2H group (C-6 of GlcA) at δ 175.7 (data of an HMBC experiment), and 18 oxygen-bearing sugar ring carbons in the region δ 68.0–81.9 as well as two *N*-acetyl groups at δ 23.5, 23.7 (both CH_3), 175.7 and 176.2 (both CO). The absence of signals in the region δ 82–88 demonstrated the pyranose form of all sugar residues. Accordingly, in the low-field region of the ^{1}H NMR spectrum, there were six signals, including those for five anomeric protons at δ 4.55–5.17 and H-5 of one of the Fuc residues at δ 4.72. The spectrum also contained signals for two methyl groups (H-6 of two Fuc) at δ 1.14 and 1.25, other sugar protons in the region δ 3.31–4.31, and two *N*-acetyl groups at δ 2.02 and 2.04. These data indicated that the DPS has a pentasaccharide repeating unit containing one residue of D-GlcA and two residues each of D-GlcNAc and L-Fuc.

The ^{1}H and ^{13}C NMR spectra of the DPS were assigned (Table 1) using 2D ^{1}H,^{1}H COSY, TOCSY, ROESY, and ^{1}H,^{13}C HSQC experiments. The ^{1}H,^{1}H experiments revealed four spin systems for two α-Fuc residues (units **A** and **D**), α-GlcNAc (unit **B**), β-GlcNAc (unit **E**), and α-GlcA (**C**) (the sugars are denoted as **A**–**E** according to their sequence in the repeating unit, see below). The anomeric configurations of the monosaccharides were determined based on the $J_{1,2}$ coupling constant values.[2]

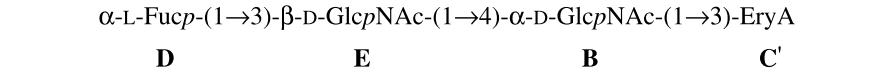
The ^{13}C NMR signals for C-3 and C-4 of unit **B**, C-3 of units **D** and **E**, and C-4 of unit **C** were shifted downfield to δ 73.3–81.9, as compared with their positions in the spectra of the corresponding nonsubstituted monosaccharides at δ 71.7–74.8.[3,4] These displacements were due to glycosylation effects[3] and defined the substitution pattern in the repeating unit.

An ^{1}H,^{13}C HMBC spectrum of the DPS showed cross-peaks between protons and carbons separated by two and three bonds. Taking into account methylation analysis and ^{13}C NMR chemical shift data (see above), cross-peaks at δ 5.17/79.2; 5.17/74.6; 5.11/73.3; and 4.97/81.9 were assigned to **B** H-1,**C** C-4; **C** H-1, **D** C-4; **A** H-1,**B** C-3; and **D** H-1,**E** C-3 correlations between anomeric protons and linkage carbons, respectively. A ROESY experiment confirmed this assignment and revealed the missing **E**,**B** correlation, namely an **E** H-1,**B** H-4 cross-peak at δ 4.55/3.87.

Therefore, the DPS is branched and has the following structure:



In order to confirm this structure, the DPS was subjected to Smith degradation. Studies of the resultant oligosaccharide by methylation analysis and 1D and 2D NMR spectroscopies as described above for the DPS, including the full assignment of ^{1}H and ^{13}C NMR signals (Table 1), enabled elucidation of the following structure:



where EryA (**C**′) is erythronic acid derived from 4-substituted GlcA (**C**).

Methylation analysis of the DPS using GLC–MS of the partially methylated alditol acetates resulted in identification of derivatives from terminal and 3-substituted Fuc, 3-substituted and 3,4-disubstituted GlcNAc in nearly equal quantities.

Position of the *O*-acetyl group was determined by an ^{1}H,^{13}C HSQC experiment on the OPS. As compared to the HSQC spectrum of the DPS, about one third of the H-6a,6b/C-6 cross-peaks of unit **B** shifted from δ 3.71, 3.83/60.8 to 4.12, 4.48/63.7.

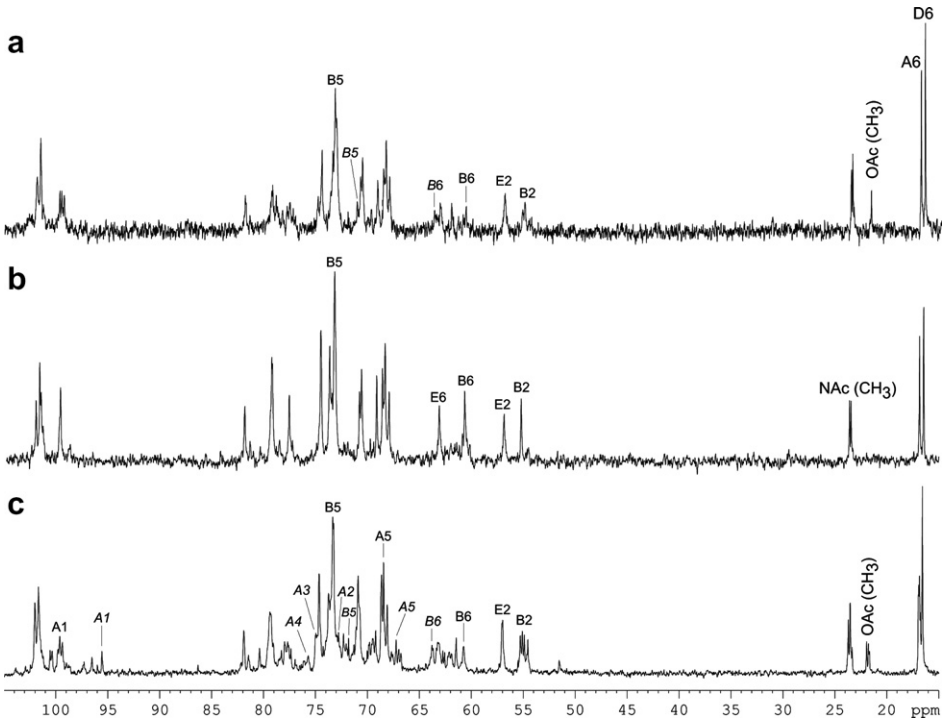


Figure 1. ^{13}C NMR spectra of the OPS of *E. coli* O168 (a), DPS of *E. coli* O168 (b), and OPS of *S. dysenteriae* type 4 (c). Arabic numerals refer to atoms in the sugar residues denoted by letters as shown in Table 1.

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

Sugar residue	Nucleus	1	2	3	4	5	6
E. coli O168 DPS							
α-L-Fuc*p*-(1→	^{1}H	5.11	3.68	3.95	3.79	4.72	1.25
A	^{13}C	99.7	69.2	70.7	73.3	68.0	16.9
→3,4)-α-D-Glc*p*NAc-(1→	^{1}H	5.17	4.13	3.91	3.87	3.72	3.71; 3.83
B	^{13}C	99.6	55.3	73.3	79.3	73.3	60.8
→4)-α-D-Glc*p*A-(1→	^{1}H	5.17	3.57	3.73	3.73	4.18	
C	^{13}C	102.0	73.3	74.6	79.3	73.7	175.7
→3)-α-L-Fuc*p*-(1→	^{1}H	4.97	3.88	3.89	3.91	4.31	1.14
D	^{13}C	101.6	68.7	74.6	73.7	68.4	16.5
→3)-β-D-Glc*p*NAc-(1→	^{1}H	4.55	3.80	3.63	3.31	3.44	3.59; 3.96
E	^{13}C	101.5	57.0	81.9		77.6	63.2
Smith-degraded oligosaccharide							
α-L-Fuc*p*-(1→	^{1}H	5.01	3.72	3.84	3.81	4.34	1.17
D	^{13}C	101.5	69.4	71.0	73.3	68.4	16.6
→3)-β-D-Glc*p*NAc-(1→	^{1}H	4.64	3.89	3.70	3.57	3.55	3.77; 3.95
E	^{13}C	102.4	56.8	81.7	69.9	77.3	62.1
→4)-α-D-Glc*p*NAc-(1→	^{1}H	5.10	3.96	3.93	3.67	3.93	
B	^{13}C	99.0	54.8	71.0	80.9	72.1	61.4
→3)-EryA	^{1}H		4.25	3.98	3.70; 3.75		
C′	^{13}C	178.9	74.8	82.8	61.9		

The chemical shifts for the *N*-acetyl groups are δ_H 2.02 and 2.04; δ_C 23.5, 23.7 (Me) and 175.7, 176.2 (CO) in the DPS; δ_H 2.04 and 2.06; δ_C 23.3, 23.7 (Me) and 175.7, 176.2 (CO) in the Smith-degraded oligosaccharide.

This displacement was due to a deshielding effect of the *O*-acetyl group[5] and indicated partial O-acetylation of unit **B** at position 6. The O-acetylation pattern was confirmed by an upfield shift by 2.3 ppm of a part of the C-5 signal of unit **B** (β-effects of O-acetylation[5]).

Therefore, the O-polysaccharide of *E. coli* O168 has the following structure:

α-L-Fuc*p*
1
↓
3
→4)-α-D-Glc*p*NAc6Ac$_{0.3}$-(1→4)-α-D-Glc*p*A-(1→3)-α-L-Fuc*p*-(1→3)-β-D-Glc*p*NAc-(1→

This structure resembles that of *S. dysenteriae* type 4 reported in 1977.[6] Having a doubt about the reliability of the *S. dysenteriae* type 4 structure, which was established before the time when shift-correlated NMR spectroscopy was introduced to structural studies of polysaccharides, we reinvestigated it using essentially the same approach as in studies of the OPS of *E. coli* O168. The OPS isolated by mild acid degradation of the lipopolysaccharide was found to be extensively and nonstoichiometrically O-acetylated (there were signals for multiple *O*-acetyl groups at δ 21.7–21.9 in the ^{13}C NMR spectrum (Fig. 1c), *O*-acetyl and *N*-acetyl groups at δ 1.97–2.20 in the ^{1}H NMR spectrum). The OPS was O-deacetylated by treatment with aqueous ammonia, and the ^{13}C NMR spectrum of the resultant DPS was found to be identical to that of *E. coli* O168. The identity of the two DPS was additionally confirmed by the assignment of the ^{1}H NMR spectrum and Smith degradation of the DPS of *S. dysenteriae* type 4.

O-Acetylation sites in *S. dysenteriae* type 4 were determined by a comparison of the ^{1}H,^{13}C HSQC spectra of the OPS and DPS. It showed down-field displacements due to deshielding effects of the *O*-acetyl groups of about half H-6a,6b/C-6 cross-peaks of α-GlcNAc (unit **B)** from δ 3.72, 3.83/60.8 to 4.12, 4.48/63.7 and about 15% H-2/C-2; H-3/C-3; and H-4/C-4 cross-peaks of lateral Fuc (unit **A**) from δ 3.68/69.2; 3.95/70.7; and 3.79/73.3 to 4.87/72.5; 5.03/75.2; and 5.19/76.1, respectively. These data indicated partial O-acetylation of unit **B** at position 6 and unit **A** at positions 2, 3, and 4.

Based on these data, it was concluded that the O-polysaccharide of *S. dysenteriae* type 4 has the following structure:

α-L-Fuc*p*Ac$_{0.45}$
1
↓
3
→4)-α-D-Glc*p*NAc6Ac$_{0.5}$-(1→4)-α-D-Glc*p*A-(1→3)-α-L-Fuc*p*-(1→3)-β-D-Glc*p*NAc-(1→

S. dysenteriae type 4

where nonstoichiometric O-acetylation occurs on about half disubstituted GlcNAc at position 6 and lateral Fuc at each position. Earlier, a random nonstoichiometric O-acetylation of a lateral 6-deoxy-L-talose residue has been reported in the O-polysaccharide of *Aeromonas hydrophila* O:34.[7]

Therefore, the O-polysaccharide structure of *S. dysenteriae* type 4 reported earlier[6] is revised in respect to the anomeric configuration and positions of glycosylation of GlcNAc at the branching point; in addition, the O-acetylation pattern previously unknown is now defined. The revised OPS structure of *S. dysenteriae* type 4 differs from that of *E. coli* O168 in a higher degree of O-acetylation of GlcNAc (50% vs 30%) and random O-acetylation of lateral Fuc. Characterization of the O-antigen gene clusters of the two bacteria will be published elsewhere. Remarkably, the O-polysaccharides of *E. coli* O168 and *S. dysenteriae* type 4 are not only closely related to each other but also similar to the O-polysaccharide of *E. coli* O159 having the following structure:[8]

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

E. coli O159

1. Experimental

1.1. Bacterial strain, isolation and degradation of lipopolysaccharides

E. coli O168 type strain G1090 and *S. dysenteriae* type 4 strain G1021 were obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia (IMVS). Bacteria were grown to late log phase in 8 L of LB using a 10-L fermentor (BIOSTAT C-10, B. Braun Biotech International, Germany) under constant aeration at 37 °C and pH 7.0. Bacterial cells were washed and dried as described.[9]

Lipopolysaccharide samples of *E. coli* O168 and *S. dysenteriae* type 4 (600 and 430 mg) were isolated from dried cells (7 and 4.5 g) by the phenol–water method[10] and purified by precipitation of nucleic acids and proteins by adding aq 50% trichloroacetic acid at 4 °C.

Delipidation of *E. coli* O168 and *S. dysenteriae* type 4 lipopolysaccharide samples (125 and 134 mg) was performed with aq 2% HOAc (4 mL) at 100 °C until precipitation of lipid A (~6 h). The precipitate was removed by centrifugation (13,000*g*, 20 min) and the supernatant fractionated by gel-permeation chromatography on a column (56 × 2.6 cm) of Sephadex G-50 (S) (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer pH 4.5, monitored by a differential refractometer (Knauer, Germany). High-molecular-mass OPS was obtained in yields of 40% and 18% of the lipopolysaccharide weight, respectively.

1.2. O-Deacetylation of O-polysaccharides

The OPS of *E. coli* O168 and *S. dysenteriae* type 4 (50 and 24 mg) was treated with aq 12.5% ammonia at 37 °C for 16 h, ammonia was removed with a stream of air, the remaining solution was desalted on a column (90 × 2.5 cm) of TSK HW-40 (S) (Merck, Germany) in water and freeze-dried to give the DPS (40 and 20 mg, respectively).

1.3. Chemical analyses

The OPS of *E. coli* O168 was hydrolyzed with 2 M CF_3CO_2H (120 °C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on a Hewlett–Packard 5890 chromatograph (USA) equipped with an Ultra-1 column using a temperature gradient of 160–290 °C at 10 °C min^{-1}. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides as described.[11,12]

1.4. Smith degradation

The DPS of *E. coli* O168 and *S. dysenteriae* type 4 (38 and 18 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 products were hydrolyzed with aq 2% HOAc for 2 h at 100 °C and oligosaccharide fractions (9 and 6 mg) were isolated by GPC on TSK HW-40 (S) in aq 1% AcOH. The oligosaccharides were treated with aq 12.5% ammonia at 37 °C for 16 h to destroy an EryA lactone followed by desalting on a column (90 × 2.5 cm) of TSK HW-40 (S) in aq 1% AcOH and freeze-dried to the target products (6 and 4 mg, respectively).

1.5. Methylation analysis

Methylation of the OPS of *E. coli* O168 and OS of *S. dysenteria* type 4 (1 mg each) was performed with CH_3I in dimethyl sulfoxide in the presence of sodium methylsulfinylmethanide.[13] Partially methylated monosaccharides were derived by hydrolysis under the same conditions as in sugar analysis, reduced with $NaDH_4$, acetylated and analyzed by GLC–MS on a Hewlett–Packard HP 5989A instrument equipped with a 30-m HP-5ms column (Hewlett–Packard) under the same chromatographic conditions as in GLC.

1.6. NMR spectroscopy

NMR spectra were recorded for solutions in 99.96% D_2O at 30 °C on a Bruker DRX-500 spectrometer (Germany) using XWINNMR software on SGI Indy/Irix 5.3. Internal TSP (δ_H 0.0) and acetone (δ_C 31.45) were used as references. Mixing times of 200 and 100 ms were used in TOCSY and ROESY experiments, respectively.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (05-04-48992), the Russian Science Support Foundation, the Council on Grants at the President of the Russian Federation for Support of Young Russian Scientists (MK-157.2007.4 to A.V.P.), Tianjin Municipal Special Fund for Science and Technology Innovation Grant 05FZZDSH00800, the NSFC Key Programs Grants 30530010 and 20536040, NSFC General Program Grant 30670038, and the National 863 program of China Grants 2006AA020703 and 2006AA06Z409.

References

1. Pupo, G. M.; Lan, R.; Reeves, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10567–10572.
2. Altona, C.; Haasnoot, C. A. G. *Org. Magn. Reson.* **1980**, *13*, 417–429.
3. Lipkind, G. M.; Shashkov, A. S.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K. *Carbohydr. Res.* **1988**, *175*, 59–75.
4. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *188*, 169–191.
5. Jansson, P.-E.; Kenne, L.; Schweda, E. *J. Chem. Soc., Perkin Trans. 1* **1987**, 377–383.
6. Dmitriev, B. A.; L'vov, V. L.; Kochetkov, N. K.; Hofman, I. L. *Bioorg. Khim.* **1977**, *3*, 1226–1233.
7. Knirel, Y. A.; Shashkov, A. S.; Senchenkova, S. N.; Merino, S.; Tomas, J. M. *Carbohydr. Res.* **2002**, *337*, 1381–1386.

8. Linnerborg, M.; Weintraub, A.; Widmalm, G. *Eur. J. Biochem.* **1999**, *266*, 246–251.
9. Robbins, P. W.; Uchida, T. *Biochemistry* **1962**, *1*, 323–335.
10. Westphal, O.; Jann, K. *Methods Carbohydr. Chem.* **1965**, *5*, 83–91.
11. Leontein, K.; Lindberg, B.; Lönngren, J. *Carbohydr. Res.* **1978**, *62*, 359–362.
12. Gerwig, G. J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1979**, *77*, 1–7.
13. Conrad, H. E. *Methods Carbohydr. Chem.* **1972**, *6*, 361–364.