Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 342 (2007) 1275–1279
Carbohydrate RESEARCH

Note

Structural relation of the antigenic polysaccharides of *Escherichia coli* O40, *Shigella dysenteriae* type 9, and *E. coli* K47

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

[a]*TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*
[b]*Tianjin Key Laboratory for Microbial Functional Genomics, TEDA College, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*
[c]*N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation*

Received 16 January 2007; received in revised form 1 March 2007; accepted 5 March 2007
Available online 12 March 2007

Abstract—O-Polysaccharides were isolated from the lipopolysaccharides of *Escherichia coli* O40 and *Shigella dysenteriae* type 9 and studied by chemical analyses along with ^{1}H and ^{13}C NMR spectroscopy. The following new structure of the O-polysaccharide of *E. coli* O40 was established:

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

The O-polysaccharide structure of *S. dysenteriae* type 9 established earlier was revised and found to be identical to the reported structure of the capsular polysaccharide of *E. coli* K47 and to differ from that of the *E. coli* O40 polysaccharide in the presence of a 3,4-linked pyruvic acid acetal having the (*R*)-configuration (*R*Pyr):

→2)-β-D-Gal*p*3,4(*R*Pyr)-(1→4)-β-D-Man*p*-(1→4)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→

Keywords: *Escherichia coli*; *Shigella dysenteriae*; O-Antigen; Bacterial polysaccharide structure; Pyruvic acid acetal

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. Based on multilocus enzyme electrophoresis and analysis of housekeeping gene sequences, it was shown that *Shigella* species in reality are *E. coli* serotypes.[1] The O-antigen represents the polysaccharide chain (O-polysaccharide, OPS) of the lipopolysaccharide (LPS), which is usually built up of repeating units containing two to eight sugar residues and often also non-sugar substituents (e.g., amino acids, pyruvic acid acetals, lactic acid ethers, phosphate, *O*-acetyl groups, etc.). In this paper we present a new O-antigen structure of *E. coli* O40 and a revised structure of the O-antigen of *Shigella dysenteriae* type 9, which differs in the presence of a pyruvic acid acetal only.

The O-polysaccharide (OPS) of *E. coli* O40 was obtained by mild acid degradation of the LPS, isolated from dried bacterial cells by the phenol–water procedure, followed by separation of the high-molecular-mass material by GPC on Sephadex G-50. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the OPS revealed Man, Gal, and GlcN (from GlcNAc) in the ratios ~1:2:1. GLC analysis of the

* Corresponding author. Tel.: +7 495 9383613; fax: +7 495 1355328; e-mail: perepel@ioc.ac.ru

0008-6215/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2007.03.005

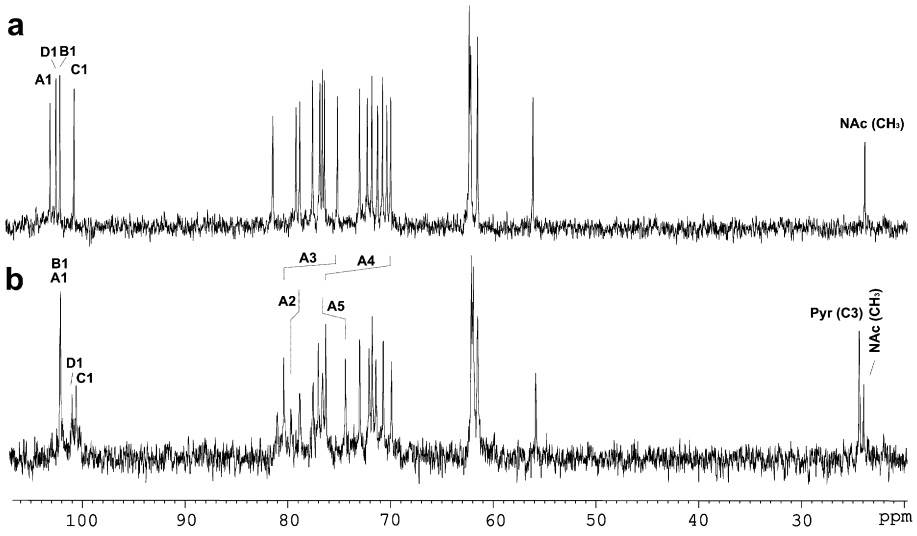

Figure 1. ^{13}C NMR spectrum of the OPS from *E. coli* O40 (a) and LPS_{NH_4OH} from *S. dysenteriae* type 9 (b). Arabic numerals refer to atoms in the sugar residues denoted by letters as shown in Table 1.

acetylated (*S*)-2-octyl glycosides demonstrated the D configuration of all monosaccharides.

The ^{13}C NMR spectrum of the OPS (Fig. 1a) showed signals for four anomeric carbons in the region δ 100.8–103.2, four HO*C*H_2–C groups (C-6 of hexoses) at δ 61.5–62.4, one nitrogen-bearing carbon (C-2 of GlcNAc) at δ 56.1, and 15 oxygen-bearing non-anomeric sugar ring carbons in the region δ 70.0–81.5, as well as one *N*-acetyl group at δ 23.9 (CH_3) and 175.5 (CO). The absence of any signals in the region δ 82–88 demonstrated the pyran- ose form of all sugar residues. Accordingly, the ^{1}H NMR spectrum of the OPS contained signals for four anomeric protons at δ 4.47–4.88, other sugar protons in the region δ 3.42–4.22, and one *N*-acetyl group at δ 2.06. These data indicated that the OPS has a regular structure with a tetrasaccharide repeating unit containing one residue each of D-Man and D-GlcNAc and two residues of D-Gal.

Methylation analysis of the OPS by GLC of the partially methylated alditol acetates resulted in the identification of derivatives from one 2-substituted hexose and two 4-substituted hexose residues; no methylated derivative of GlcNAc was detected.

The ^{1}H and ^{13}C NMR spectra of the OPS (Table 1) were assigned using 2D ^{1}H,^{1}H COSY, TOCSY, ROESY, ^{1}H,^{13}C HSQC, HMQC-TOCSY, and HMBC experiments. The ^{1}H,^{1}H experiments revealed four sub-spectra for residues of β-Gal*p* (**A**), β-Man*p* (**B**), α-Gal*p* (**C**), and β-Glc*p*NAc (**D**) (the sugar units were denoted as **A**–**D** according to their sequence in the repeating unit, see below). The ^{13}C NMR signals for C-2 of unit **A**, C-4 of units **B** and **C**, and C-3 of unit **D** were shifted downfield to δ 77.6–81.5, as compared with their positions in the spectra of the corresponding non-substituted monosaccharides at δ 67.9–75.1.[2] These displacements were due to glycosylation effects and defined the modes of substitution of the monosaccharides, which were in agreement with the methylation analysis data for hexoses.

The HMBC spectrum showed cross-peaks between anomeric protons and non-anomeric carbons separated by two and three bonds (Fig. 2). Taking into account methylation analysis and ^{13}C NMR chemical shift data (see above), the cross-peaks at δ 5.40/81.5; 4.88/78.9; 4.84/77.6; and 4.47/79.2 were assigned to **C** H-1,**D** C-3; **D** H-1,**A** C-2; **B** H-1,**C** C-4; and **A** H-1,**B** C-4 transglycosidic correlations, respectively. A ROESY experiment confirmed these findings (data not shown) and, hence, the OPS of *E. coli* O40 is linear and has the following structure:

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

A **B** **C** **D**

This structure resembles much the backbone structure of the OPS of *S. dysenteriae* type 9, which was established as early as in 1978.[3] The two OPSs differ in the position of substitution of unit **A** and in the O-acetylation and acetalation with pyruvic acid of the type 9 OPS. As no unambiguous assignment of the NMR spectra of the type 9 OPS was performed at that time, we considered the published data not reliable

Table 1. ^{1}H and ^{13}C NMR data of polysaccharides from *E. coli* O40 and *S. dysenteriae* type 9 (δ, ppm)

Sugar residue	Nucleus	1	2	3	4	5	6 (6a, 6b)	NAc (CH_3, CO)	Pyr (CH_3, C, CO)
OPS from E. coli O40									
→2)-β-D-Gal*p*-(1→	^{1}H	4.47	3.72	3.72	3.87	3.71	3.78; 3.78		
A	^{13}C	103.2	78.9[a]	75.2[a]	70.4	76.7	62.4		
→4)-β-D-Man*p*-(1→	^{1}H	4.84	4.20	3.76	3.74	3.52	3.85; 4.16		
B	^{13}C	102.2	71.3	73.0	79.2	76.5	62.2		
→4)-α-D-Gal*p*-(1→	^{1}H	5.40	3.88	3.89	4.22	3.90	3.68; 3.84		
C	^{13}C	100.8	70.0	70.8	77.6	71.8	61.5		
→3)-β-D-Glc*p*NAc-(1→	^{1}H	4.88	3.81	3.74	3.69	3.42	3.74; 3.91	2.06	
D	^{13}C	102.6	56.1	81.5	72.3	76.9	62.3	23.9, 175.5	
LPS_{NH_4OH} from S. dysenteriae type 9									
→2)-β-D-Gal*p*3,4*R*Pyr-(1→	^{1}H	4.56	3.70	4.26	4.14	4.01	3.80, 3.82		1.57
A	^{13}C	102.3	79.7	80.3	76.4	74.4	62.2		24.4, 108.7, 177.0
→4)-β-D-Man*p*-(1→	^{1}H	4.79	5.15	3.72	3.72	3.47	3.76, 4.01		
B	^{13}C	102.3	71.4	73.0	78.9	76.3	61.9		
→4)-α-D-Gal*p*-(1→	^{1}H	5.40	3.83	3.86	4.18	3.84	3.64, 3.79		
C	^{13}C	100.7	69.1	70.7	77.6	71.7	61.5		
→3)-β-D-Glc*p*NAc-(1→	^{1}H	4.88	3.79	3.71	3.67	3.42	3.73, 3.85	2.04	
D	^{13}C	101.1	55.7	81.0	72.1	77.0	62.0	23.8, 175.9	

[a] Assignment for C-2 and C-3 of unit **A** could not be performed in a straightforward way owing to a coincidence of H-2 and H-3 resonances at δ 3.72 but taking into account methylation analysis data.

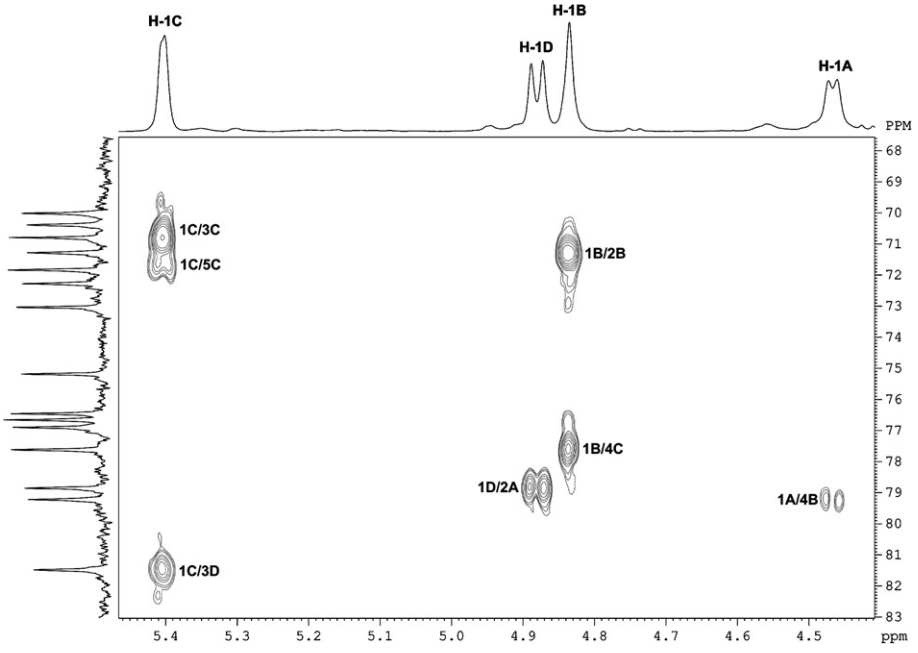

Figure 2. A part of the HMBC spectrum of the OPS from *E. coli* O40. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are displayed along the horizontal and vertical axes, respectively. Arabic numerals before and after oblique stroke refer to protons and carbons, respectively, in the sugar residues denoted by letters as shown in Table 1.

enough and reinvestigated the OPS structure in this work.

For this purpose, the OPS was isolated from the *S. dysenteriae* type 9 LPS by a prolonged mild acid hydrolysis to remove the pyruvic acid acetal. The ^{13}C NMR spectrum of the deacetalated polysaccharide was essentially identical to that of *E. coli* O40. Further studies using 2D $^{1}H,^{1}H$ shift-correlated spectroscopy, as described above, confirmed that the *S. dysenteriae* type 9 OPS backbone has the same structure as the OPS of *E. coli* O40. Hence, the position of substitution of unit **A** should be revised from 3 to 2. In contrast to the earlier observation,[3] no O-acetylation was observed even when mild acid hydrolysis of the LPS was performed for a short time. This discrepancy could be due to differences between two *S. dysenteriae* type 9 strains studied or between conditions of their cultivation.

To determine the position of the pyruvic acid acetal, the type 9 LPS was O-deacylated by treatment with aq ammonia. The ^{13}C NMR spectrum (Fig. 1b) of the resultant polymer (LPS_{NH_4OH}) contained, inter alia, signals for pyruvic acid acetal (Pyr) at δ 24.4 (CH_3), 108.7 (acetal carbon), and 177.0 (CO_2H). The ^1H NMR spectrum of LPS_{NH_4OH} showed a signal for a CH_3–C group (H-3 of Pyr) at δ 1.57.

A ROESY experiment with LPS_{NH_4OH} showed Pyr H-3,Gal **A** H-3,4 cross-peaks, thus indicating the location of the acetal group at positions 3 and 4 of unit **A**. This conclusion was confirmed by downfield displacements of both C-3 and C-4 signals of Gal **A** from δ 75.2 and 70.4 in the deacetalated polysaccharide to δ 80.3 and 76.4 in LPS_{NH_4OH}. Therefore, the OPS structure of *S. dysenteriae* type 9 should be revised also in respect to the location of the pyruvic acid acetal (from positions 4 and 6 to positions 3 and 4 of Gal **A**) as shown below. The (*R*)-configuration of the acetal was inferred by ^{13}C NMR chemical shifts[4,5] and a strong cross-peak between Pyr H-3 and Gal **A** H-2 in the ROESY spectrum.[5]

→2)-β-D-Gal*p*3,4*R*Pyr-(1→4)-β-D-Man*p*-(1→4)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→

A **B** **C** **D**

Comparison of the O-antigen gene clusters of *E. coli* O40 and *S. dysenteriae* type 9 revealed the same organization and a high level of similarity except for that the former has an IS3 element inserted between the genes for O-antigen polymerase and pyruvate transferase. It appears that, as a result of the insertion, the pyruvate transferase gene is not expressed and this accounts for the lack of the pyruvic acid acetal from the *E. coli* O40 antigen. These data will be reported in detail elsewhere.

Remarkably, the revised structure of *S. dysenteriae* type 9 OPS is identical to the structure of the capsular polysaccharide of *E. coli* K47,[5] which, therefore, should have the same or closely related gene cluster. It can be suggested that the gene cluster for the O40 and type 9 O-antigens and K47 polysaccharide has been assembled in one of the strains or in a common ancestor and then spread to the other strains in the period when the currently observed diversity pattern of bacterial polysaccharide antigens formed.

1. Experimental

1.1. Bacterial strains and isolation of lipopolysaccharides

E. coli O40 type strain H316 and *S. dysenteriae* type 9 (G1274) were obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia (IMVS). Bacteria were grown to late log phase in 8 L of Luria broth 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.[6]

LPSs of *E. coli* O40 and *S. dysenteriae* type 9 were isolated in yields 11.7% and 9.2% from dried cells (12.8 and 8.7 g, respectively) by the phenol–water method[7] and purified by precipitation of nucleic acids and proteins by adding of aq 50% trichloroacetic acid.

1.2. Degradation of lipopolysaccharides

Delipidation of the LPSs of *E. coli* O40 and *S. dysenteriae* type 9 (120 and 130 mg, respectively) was performed with aq 2% HOAc (6 mL) at 100 °C until precipitation of lipid A (ca. 3 h). 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 (S) (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer, pH 4.5, monitored by a differential refractometer (Knauer, Germany). High-molecular-mass OPSs were obtained in a yield of ~40% of the LPS weight.

Treatment of the OPS of *S. dysenteriae* type 9 (30 mg) with aq 2% HOAc (105 °C, 15 h) followed by GPC on a column (90 × 2.5 cm) of TSK HW-40 (S) (Merck, Germany) in water gave a deacetalated polysaccharide (10 mg).

The LPS of *S. dysenteriae* type 9 (80 mg) was treated with aq 12.5% ammonia at 37 °C for 16 h, the solution was desalted on Sephadex G-50 (S) as described above and freeze-dried to give an O-deacylated polymer (LPS_{NH_4OH}, 55 mg).

1.3. Chemical analyses

The OPS of *E. coli* 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 equipped with an Ultra-1 column using a temperature gradient of 160–290 °C at 10 °C min^{-1}. The absolute configuration of the monosaccharides was determined by GLC of the acetylated (*S*)-2-octyl glycosides.[8,9]

1.4. Methylation analysis

Methylation of the OPS was performed with CH_3I in dimethyl sulfoxide in the presence of sodium methylsulfin-

ylmethanide.[10] Partially methylated monosaccharides were derived by hydrolysis under the same conditions as in sugar analysis, reduced with $NaBD_4$, 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.

1.5. NMR spectroscopy

NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) equipped with a SGI Indy/Irix 5.3 workstation and XWINNMR software. Samples were freeze-dried from 99.9% D_2O and studied as solutions in 99.96% D_2O at 30 °C, using internal TSP (δ_H 0.0) and acetone (δ_C 31.45) as references. A spin-lock time of 30 ms was used in TOCSY and HMQC-TOCSY experiments and a mixing time of 100 ms in a ROESY experiment. HMBC experiment was optimized for the coupling constant 8 Hz.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (Project 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-1597.2005.3 to A.V.P.), NSFC General Program (30370339, 30370023), NSFC Key Program (30530010), and Tianjin Municipal Fund for Science and Technology Innovation (05FZZDSH00800).

References

1. Pupo, G. M.; Lan, R.; Reeves, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10567–10572.
2. Lipkind, G. M.; Shashkov, A. S.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K. *Carbohydr. Res.* **1988**, *175*, 59–75.
3. Dmitriev, B. A.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K.; Hofman, I. L. *Bioorg. Khim.* **1978**, *4*, 40–46.
4. Garegg, P. J.; Jansson, P.-E.; Lindberg, B.; Lindh, F.; Lönngren, J.; Kvarnström, I.; Nimmich, W. *Carbohydr. Res.* **1980**, *78*, 127–132.
5. De Bruin, A. H.; Parolis, H.; Parolis, L. A. S. *Carbohydr. Res.* **1992**, *233*, 195–204.
6. Robbins, P. W.; Uchida, T. *Biochemistry* **1962**, *1*, 323–335.
7. Westphal, O.; Jann, K. *Methods Carbohydr. Chem.* **1965**, *5*, 83–91.
8. Leontein, K.; Lindberg, B.; Lönngren, J. *Carbohydr. Res.* **1978**, *62*, 359–362.
9. Gerwig, G. J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1979**, *77*, 1–7.
10. Conrad, H. E. *Methods Carbohydr. Chem.* **1972**, *6*, 361–364.