ISSN 0006-2979, Biochemistry (Moscow), 2008, Vol. 73, No. 4, pp. 406-410. © Pleiades Publishing, Ltd., 2008.
Published in Russian in Biokhimiya, 2008, Vol. 73, No. 4, pp. 498-503.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM07-237, November 11, 2007.

Structure and Characterization of the Gene Cluster of the O-Antigen of *Escherichia coli* O49 Containing 4,6-Dideoxy-4-[(*S*)-3-hydroxybutanoylamino]-D-glucose

A. V. Perepelov[1]*, Quan Wang[2,3], S. N. Senchenkova[1], S. D. Shevelev[1], A. S. Shashkov[1], Lu Feng[2,3], Y. A. Knirel[1], and Lei Wang[2,3]

[1]Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky pr. 47, 119991 Moscow, Russia; fax: (495) 137-6148; E-mail: perepel@ioc.ac.ru
[2]TEDA School of Biological Sciences and Biotechnology and [3]Tianjin Key Laboratory for Microbial Functional Genomics, TEDA College, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China; E-mail: wangquan@mail.nankai.edu.cn

Received July 11, 2007

Abstract—An O-polysaccharide was isolated by mild acid degradation of the lipopolysaccharide of enteropathogenic *Escherichia coli* O49 and studied by sugar analysis along with one- and two-dimensional ^{1}H- and ^{13}C-NMR spectroscopy. The following structure of the linear tetrasaccharide repeating unit of the O-polysaccharide was established:

→2)-α-D-Qui*p*4N(*S*3HOBut)-(1→4)-β-D-Gal*p*NAc-(1→4)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc6Ac-(1→ ,

where D-Qui4N(*S*3HOBut) stands for 4,6-dideoxy-4-[(*S*)-3-hydroxybutanoylamino]-D-glucose and O-acetylation of GlcNAc is partial (~30%). To our knowledge, no N-(3-hydroxybutanoyl) derivative of Qui4N has been hitherto found in bacterial polysaccharides. Gene functions of the O-antigen gene cluster of *E. coli* O49 were assigned by bioinformatics analysis and found to correspond to the O-polysaccharide structure. Two new genes were revealed and suggested to be responsible for synthesis and transfer of the 3-hydroxybutanoyl group.

DOI: 10.1134/S0006297908040044

Key words: *Escherichia coli*, lipopolysaccharide, bacterial polysaccharide structure, 3-hydroxybutyrate, 4-amino-4,6-dideoxy-D-glucose, O-antigen gene cluster

Escherichia coli is the predominant species in the human intestinal microflora and one of the most common causes of diarrheic diseases. This highly diverse bacterial group includes several pathogenic forms. *Escherichia coli* strains are usually classified by a combination of the O- and H- (and sometimes K-) antigens. Currently, 166 O-antigen forms of *E. coli* are known, and the structures of only half of them have been elucidated [1]. Now we report on a new O-antigen (O-polysaccharide, OPS) structure of *E. coli* O49 containing an unusual monosaccharide, 4,6-dideoxy-4-[(*S*)-3-hydroxybutanoylamino]-D-glucose (D-Qui4N(*S*3HOBut)). In addition, we characterized the O-antigen gene cluster of *E. coli* O49.

MATERIALS AND METHODS

Escherichia coli O49 type strain U12-41 (laboratory stock No. G1277) was obtained from the Institute of Medical and Veterinary Science (Adelaide, Australia). Sequencing of the chromosome region between *galF* and *gnd*, analysis of genes, and search databases for possible gene functions were performed as described [2].

Bacteria were grown to late log phase in 8 liters of Luria–Bertani liquid medium using a 10-liter 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 [3]. The

Abbreviations: DPS) O-deacetylated polysaccharide; 3HOBut) 3-hydroxybutyric acid; OPS) O-polysaccharide; Qui4N) 4-amino-4,6-dideoxyglucose; Rha) rhamnose.

* To whom correspondence should be addressed.

lipopolysaccharide in a yield of 7% was isolated from dried cells by the phenol–water method [4] and purified by precipitation of nucleic acids and proteins with aqueous 50% CCl_3CO_2H [5].

Degradation of the lipopolysaccharide (107 mg) was performed with aqueous 2% acetic acid (4 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 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). A high-molecular-mass OPS was obtained in a yield of 36% of the lipopolysaccharide weight.

The OPS (60 mg) was treated with aqueous 12.5% NH_4OH at 37°C for 16 h, and the 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 O-deacetylated polysaccharide (DPS, 42 mg).

The OPS 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 capillary column using a temperature gradient of 150 to 290°C at 5°C/min. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides as described [6, 7]. 3-Hydroxybutyric acid (3HOBut) was analyzed by GLC of the trifluoroacetylated (+)-2-octyl ester [8].

Prior to NMR measurements, samples were deuterium-exchanged by freeze-drying from D_2O and then examined as solutions in 99.96% D_2O or a 9 : 1 H_2O/D_2O mixture at 27°C. NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using internal acetone (δ_H 2.225 ppm, δ_C 31.45 ppm) as internal reference. 2D NMR spectra were obtained using standard Bruker software, and Bruker XWINNMR 2.6 program was used to acquire and process the NMR data. Mixing times of 200 and 100 msec were used in TOCSY and ROESY experiments, respectively.

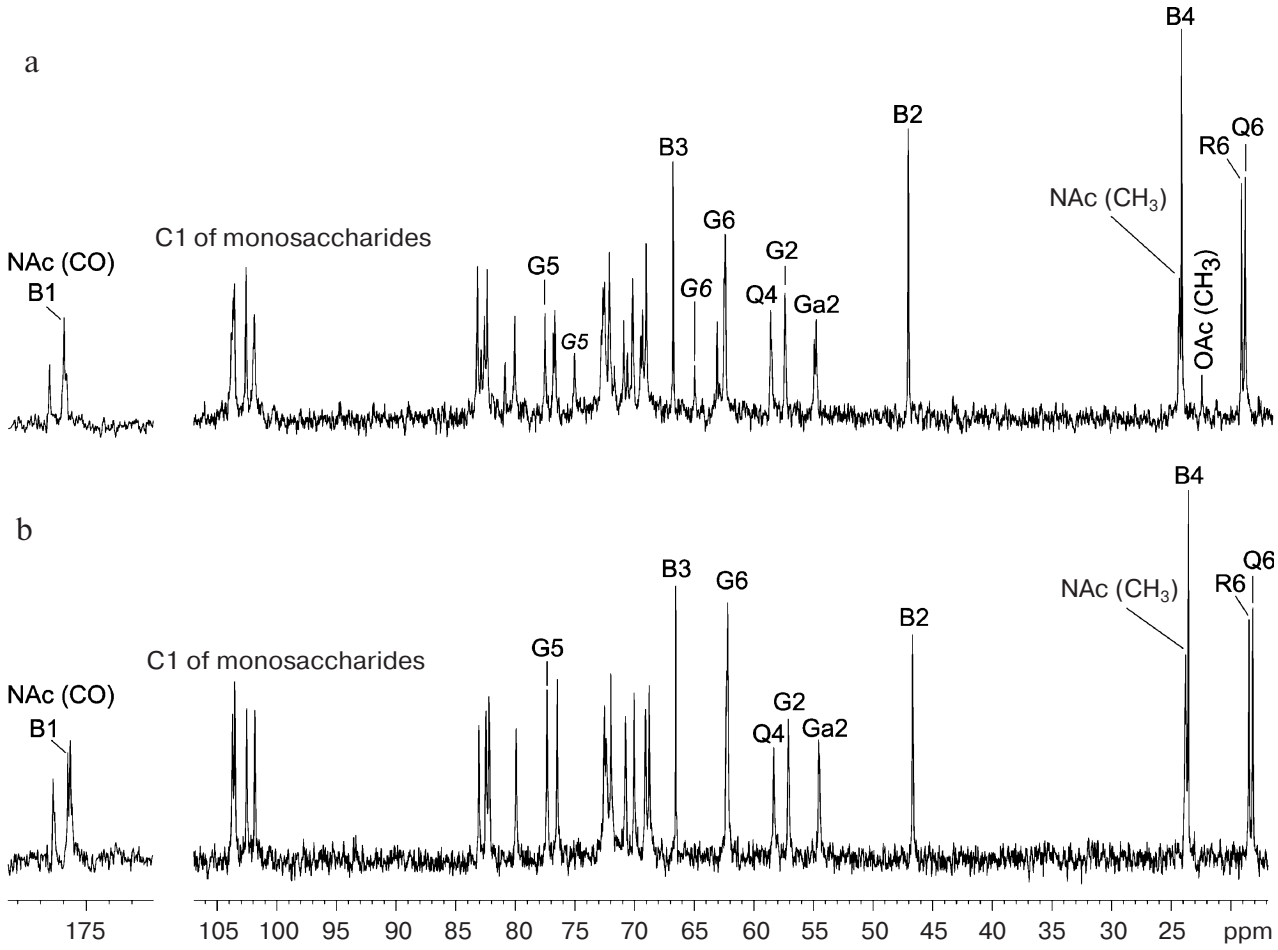
RESULTS AND DISCUSSION

The OPS was obtained by mild acid degradation of the lipopolysaccharide isolated from dried cells of *E. coli* O49 by the phenol–water procedure [4]. It was separated from low-molecular-mass compounds, including a core oligosaccharide, by gel-permeation chromatography on Sephadex G-50 (S). Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the OPS revealed rhamnose (Rha), GlcN, and GalN in the ratios ~1 : 1 : 1. GLC analysis of the acetylated (*S*)-2-octyl glycosides demonstrated the D configuration of both amino sugars and the L configuration of Rha. Yet another component of the OPS, Qui4N (4-amino-4,6-dideoxyglucose), was not detected in sugar analysis owing to its destruction during acid hydrolysis of the OPS. Its identity and absolute configuration were established by detailed analysis of the ^{13}C-NMR spectrum of the polysaccharide (see below).

The ^{13}C-NMR spectrum of the OPS (figure, panel (a)) contained signals having different intensities, most likely, owing to non-stoichiometric O-acetylation (there was a signal for CH_3 of an O-acetyl group at 21.8 ppm). Therefore, the OPS was O-deacetylated with aqueous ammonia and the resultant O-deacetylated polysaccharide (DPS) was studied by NMR spectroscopy.

The ^{13}C-NMR spectrum of the DPS (figure, panel (b)) showed signals for four anomeric carbons at 102.0-103.8 ppm, five *C*H_3-C groups (C6 of Rha and Qui4N at 18.3 and 18.6 ppm, C4 of 3OHBut (see below), and C2 of two N-acetyl groups at 23.6-23.9 ppm), two HO*C*H_2-C groups (C6 of GlcN and GalN at 62.3 and 62.4 ppm), three nitrogen-bearing carbons (C2 of GalN, GlcN and C4 of Qui4N at 54.7-58.4 ppm), 13 oxygen-bearing sugar ring carbons in the region 69.4-83.2 ppm, C2 and C3 of 3HOBut at 46.7 and 66.6 ppm, respectively, and three CO groups of N-acyl substituents at 175.8-176.6 ppm. Accordingly, the ^{1}H-NMR spectrum of the DPS contained signals for four anomeric protons at 4.75 (2 H), 4.85 and 5.14 ppm, three CH_3-C groups (H6 of Rha, Qui4N and H4 of 3HOBut at 1.12-1.30 ppm), H2 and H3 of 3HOBut at 2.47 and 4.19 ppm, respectively, two N-acetyl groups at 2.03 and 2.05 ppm, and other sugar signals at 3.49-4.15 ppm. Therefore, the DPS consists of tetrasaccharide repeating units containing one residue each of Qui4N, GlcN, GalN, and Rha, one 3HOBut, and two N-acetyl groups.

The ^{1}H- and ^{13}C-NMR spectra of the DPS were assigned using 2D COSY, TOCSY, ROESY, H-detected 2D ^{1}H,^{13}C HSQC, HMQC-TOCSY, and HMBC experiments (Tables 1 and 2). The TOCSY spectrum demonstrated correlations of H1 with H2-H6 for Qui4N, Rha, and GlcN and H2-H4 for GalN, which were assigned within each spin system using the COSY spectrum. The remaining GalN signals were assigned by correlations of H1 with H3 and H5 in the ROESY spectrum and C5 with H6a,6b in the HMQC-TOCSY spectrum. The *gluco* configuration of Qui4N was inferred based on characteristic $^3J_{H,H}$ coupling constants [9]. The amino sugars were confirmed by correlations between protons at the nitrogen-linked carbons and the corresponding carbons at 3.85/57.2 ppm for GlcN H2/C2, 3.92/54.7 ppm for GalN H2/C2, and 3.65/58.4 ppm for Qui4N H4/C4, as revealed by the HSQC experiment. The HMBC spectrum demonstrated cross-peaks between protons and carbons of 3HOBut at 1.23/46.7 and 1.23/66.6 ppm for H4/C2 and H4/C3; 2.47/175.9, 2.47/66.6, and 2.47/23.6 ppm for H2/C1, H2/C3, and H2/C4 correlations, respectively, which proved the structure of this N-acyl substituent.



^{13}C-NMR spectra of the O-polysaccharide (a) and O-deacetylated polysaccharide (b) from *E. coli* O49. Arabic numerals refer to carbon atoms in sugar residues denoted as follows: G, GlcNAc; Ga, GalNAc; *G*, GlcNAc6Ac; Q, Qui4NAcyl; R, Rha; B, 3HOBut

A relatively small $J_{1,2}$ coupling constant of ~3 Hz indicated that Qui4N is α-linked, whereas significantly larger $J_{1,2}$ values of 7-8 Hz showed that GlcN and GalN are β-linked. The α-configuration of Rha was inferred from the position of the C-5 signal at 68.9 ppm (compare the chemical shifts 70.0 and 73.2 ppm for C-5 in α-Rha*p* and β-Rha*p*, respectively [10]). The pyranose form of all monosaccharide residues was inferred by the absence from the ^{13}C-NMR spectrum of any signals for non-anomeric sugar carbons at a lower field than 83.2 ppm [11].

A relatively low-field positions of the signals for C2 of Qui4N, C3 of GlcN, C4 of GalN and C4 Rha at 80.2-83.2 ppm in the ^{13}C-NMR spectrum of the DPS, as compared with their positions in the spectra of the corresponding non-substituted monosaccharides [11, 12], demonstrated the glycosylation pattern in the repeating unit.

A two-dimensional ROESY experiment showed cross-peaks between the anomeric protons and protons at the linkage carbons at 5.14/4.02; 4.75/3.60; 4.85/3.63; and 4.75/3.73 ppm, which were assigned to interresidue Qui4N H1/GalN H4; GalN H1/Rha H4; Rha H1/GlcN H3; and GlcN H1/Qui4N H2 correlations, respectively. These data are in agreement with the positions of substitution of the monosaccharides revealed by the ^{13}C-NMR chemical shift data and established the sequence of the monosaccharide residues in the DPS. Relatively large glycosylation effects 7.5 and 9.3 ppm on C1 of GlcN and C2 of Qui4N, respectively, showed that both monosaccharide residues in the β-D-Glc*p*NAc-(1→2)-α-Qui*p*4NAcyl disaccharide fragment have the same absolute configuration (compare published data [13] 8.0 and 9.1 ppm for the same and 3.9 and 4.5 ppm for different absolute configurations, respectively). Hence, Qui4N has the D configuration.

The position of 3HOBut was determined using NMR spectroscopy of a DPS solution in a 9 : 1 H_2O/D_2O mixture, which enabled detection of nitrogen-linked protons. The ^{1}H-NMR spectrum showed the presence of three NH protons at 8.04, 8.08, and 8.30 ppm, which were assigned using a TOCSY experiment to GalN, Qui4N, and GlcN, respectively. The ROESY spectrum showed

Table 1. Data of the ^{1}H-NMR spectrum for the DPS of *E. coli* O49 (chemical shifts in ppm)

Residue	H1	H2	H3	H4	H5	H6 (6a; 6b)
→2)-α-D-Qui*p*4NAcyl-(1→	5.14	3.73	3.92	3.65	4.15	1.12
→4)-β-D-Gal*p*NAc-(1→	4.75	3.92*	3.76	4.02	3.71	3.88; 3.92
→4)-α-L-Rha*p*-(1→	4.85	3.73	3.83	3.60	4.01	1.30
→3)-β-D-Glc*p*NAc-(1→	4.75	3.85*	3.63	3.51	3.49	3.75; 3.95
*S*3HOBut		2.47	4.19	1.23		

* Signals for the N-acetyl groups are at 2.03 and 2.05 ppm.

Table 2. Data of the ^{13}C-NMR spectrum for the DPS of *E. coli* O49 (chemical shifts in ppm)

Residue	C1	C2	C3	C4	C5	C6
→2)-α-D-Qui*p*4NAcyl-(1→	102.0	82.6	70.9	58.4	69.3	18.3
→4)-β-D-Gal*p*NAc-(1→	103.8	54.7*	72.4	80.2	76.6	62.4
→4)-α-L-Rha*p*-(1→	102.7	72.7	72.0	82.3	68.9	18.6
→3)-β-D-Glc*p*NAc-(1→	103.7	57.2*	83.2	70.2	77.5	62.3
*S*3HOBut	175.9	46.7	66.6	23.6		

* Signals for the N-acetyl groups are at 23.8, 23.9 (both CH_3), 175.8, and 176.6 (both CO) ppm.

correlations of NH protons of GalN and GlcN with CH_3 of the N-acetyl groups at 8.04/2.03 and 8.30/2.05 ppm, respectively, thus indicating N-acetylation of these monosaccharide residues. A cross-peak between NH of Qui4N and H2 of 3HOBut at 8.08/2.47 ppm confirmed N-acylation of Qui4N with 3HOBut. GLC of the trifluoroacetylated (+)-2-octyl ester of 3-hydroxybutyric acid, which was released by acid hydrolysis of the OPS, revealed (*S*)-3HOBut.

Position of the O-acetyl group was determined by a ^{1}H,^{13}C HMQC experiment on the initial OPS. As compared with the HMQC spectrum of the DPS, it showed a displacement of ~30% GlcNAc H6a,6b/C6 cross-peak from 3.75/62.3 and 3.95/62.3 to 4.34/64.7 and 4.49/64.7 ppm, respectively, which was due to a deshielding effect of the O-acetyl group and indicated partial O-acetylation of GlcNAc at position 6. This conclusion was confirmed by an upfield shift by 2.7 ppm of the signal for C5 of GlcNAc, which was caused by a β-effect of O-acetylation [14]. The data showed that the O-polysaccharide of *E. coli* O49 has the following structure:

→2)-α-D-Qui*p*4N(*S*3HOBut)-(1→

→4)-β-D-Gal*p*NAc-(1→4)-α-L-Rha*p*-(1→

→3)-β-D-Glc*p*NAc6Ac-(1→ .

A peculiar feature of the OPS is the presence of 4,6-dideoxy-[(*S*)-3-hydroxybutanoylamino]-D-glucose. Remarkably, Qui4N occurs often in bacterial polysaccharides carrying an unusual N-acyl group, such as formyl, malonyl, succinyl, N-acetylglycyl, N-acetylaspartyl, N-[(*R*)-3-hydroxybutanoyl]-L-alanyl, and 2,4-dihydroxy-3,3,4-trimethyl-5-oxoprolyl [12, 15, 16]. However, to the best of our knowledge, no N-[(*S*)-3-hydroxybutanoyl] derivative of Qui4N has been hitherto found in nature.

Aiming at understanding the evolutionary history of O-antigen gene clusters of enteric bacteria and in search for specific genes to be used in PCR-based assays for rapid identification and detection of strains, the O-antigen gene cluster of *E. coli* O49 was sequenced and analyzed. As in most other *E. coli* strains studied [2], this was found to map on the chromosome between the housekeeping genes *galF* and *gnd*. Sequencing revealed 13 open reading frames (*orfs*), which were assigned by a comparison with gene databases as follows: 1) *rmlA*, *rmlB*, *rmlC*, *rmlD*, and *vioA* (*orf1-orf5*) as genes for the synthesis of nucleotide diphosphate derivatives of L-Rha and D-Qui4N [17] as well as *gne* (*orf13*) as a gene for epimerization of D-GlcNAc to D-GalNAc [18]; 2) *orf9*, *orf11*, *orf12* as genes for assembling the O-antigen repeating unit by sequential transfer of the activated derivatives of L-Rha, D-GalNAc, and D-Qui4N(3HOBut) to D-GlcNAc on a lipid carrier; 3) *wzx* (*orf8*) and *wzy* (*orf10*) as genes of the O-antigen unit processing by the Wzy-dependent biosynthesis pathway, which encode flippase and polymerase, respectively [19]. These findings fully correspond to the established structure of the O-polysaccharide of *E. coli* O49. Two remaining genes (*orf6* and *orf7*) showed a similarity to acetyltransferase and hydratase genes, respectively, and were

suggested to be responsible for the synthesis and transfer of the 3-hydroxybutanoyl group on Qui4N. The exact functions of these genes and D-Qui4N(*S*3HOBut) biosynthesis pathway remain to be determined.

This work was supported by the Russian Foundation for Basic Research (05-04-48992 and 05-04-39015), 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), NSFC Key Program (30530010), and funds from the Science and Technology Committee of Tianjin City (05YFGZG04700).

REFERENCES

1. Stenutz, R., Weintraub, A., and Widmalm, G. (2006) *FEMS Microbiol. Rev.*, **30**, 382-403.
2. Feng, L., Senchenkova, S. N., Yang, J., Shashkov, A. S., Tao, J., Guo, H., Cheng, J., Ren, Y., Knirel, Y. A., Reeves, P., and Wang, L. (2004) *J. Bacteriol.*, **186**, 4510-4519.
3. Robbins, P. W., and Uchida, T. (1962) *Biochemistry*, **1**, 323-335.
4. Westphal, O., and Jann, K. (1965) *Meth. Carbohydr. Chem.*, **5**, 83-91.
5. Zych, K., Toukach, F. V., Arbatsky, N. P., Kolodziejska, K., Senchenkova, S. N., Shashkov, A. S., Knirel, Y. A., and Sidorczyk, Z. (2001) *Eur. J. Biochem.*, **268**, 4346-4351.
6. Leontein, K., Lindberg, B., and Lonngren, J. (1978) *Carbohydr. Res.*, **62**, 359-362.
7. Gerwig, G. J., Kamerling, J. P., and Vliegenthart, J. F. G. (1979) *Carbohydr. Res.*, **77**, 1-7.
8. Shashkov, A. S., Paramonov, N. A., Veremeychenko, S. N., Grosskurth, H., Zdorovenko, G. M., Knirel, Y. A., and Kochetkov, N. K. (1998) *Carbohydr. Res.*, **396**, 297-303.
9. Altona, C., and Haasnoot, C. A. G. (1980) *Org. Magn. Reson.*, **13**, 417-429.
10. Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V., and Kochetkov, N. K. (1988) *Carbohydr. Res.*, **175**, 59-75.
11. Bock, K., and Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem.*, **41**, 27-66.
12. Perepelov, A. V., Senchenkova, S. N., Shashkov, A. S., Torzewska, A., Moll, H., Bartodziejska, B., Rozalski, A., Zahringer, U., and Knirel, Y. A. (2001) *Carbohydr. Res.*, **331**, 195-202.
13. Shashkov, A. S., Lipkind, G. M., Knirel, Y. A., and Kochetkov, N. K. (1988) *Magn. Reson. Chem.*, **26**, 735-747.
14. Jansson, P.-E., Kenne, L., and Schweda, E. (1987) *J. Chem. Soc. Perkin Trans.*, **1**, 377-383.
15. Kondakova, A. N., Lindner, B., Fudala, R., Senchenkova, S. N., Moll, H., Shashkov, A. S., Kaca, W., Zahringer, U., and Knirel, Y. A. (2004) *Biochemistry (Moscow)*, **69**, 1034-1043.
16. Bacterial Carbohydrate Structure Database. Online: http://www.glyco.ac.ru
17. Samuel, G., and Reeves, P. (2003) *Carbohydr. Res.*, **338**, 2503-2519.
18. Wang, L., Huskic, S., Cisterne, A., Rothemund, D., and Reeves, P. R. (2002) *J. Bacteriol.*, **184**, 2620-2625.
19. Raetz, C. R. H., and Whitfield, C. (2002) *Annu. Rev. Biochem.*, **71**, 635-700.