ISSN 0006-2979, Biochemistry (Moscow), 2009, Vol. 74, No. 4, pp. 416-420. © Pleiades Publishing, Ltd., 2009.
Published in Russian in Biokhimiya, 2009, Vol. 74, No. 4, pp. 515-521.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM08-270, February 8, 2009.

Structure of O-Antigen and Functional Characterization of O-Antigen Gene Cluster of *Salmonella enterica* O47 Containing Ribitol Phosphate and 2-Acetimidoylamino-2,6-dideoxy-L-galactose

A. V. Perepelov[1]*, B. Liu[2,3], S. N. Senchenkova[1], A. S. Shashkov[1], L. Feng[2,3], L. Wang[2,3], and Y. A. Knirel[1]

[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

Received August 21, 2008
Revision received October 16, 2008

Abstract—An O-polysaccharide was isolated by mild acid degradation of the lipopolysaccharide of *Salmonella enterica* O47 and studied by sugar analysis along with one- and two-dimensional ^{1}H- and ^{13}C-NMR spectroscopy. The following structure of the linear ribitol phosphate-containing repeating unit of the O-polysaccharide was established:

→2)-D-Ribitol-5-P-(O→6)-α-D-Gal*p*-(1→3)-α-L-Fuc*p*NAm-(1→3)-β-D-Glc*p*NAc-(1→,

where FucNAm stands for 2-acetimidoylamino-2,6-dideoxy-L-galactose. About 10% of Gal is O-acetylated at position 4 and another minor O-acetyl group is present at an undetermined position. Functions of the *S. enterica* O47 antigen biosynthetic genes were tentatively assigned by comparison with gene databases and found to be in agreement with the O-polysaccharide structure. A comparison of the O-antigen gene clusters of *S. enterica* O47 and *E. coli* O145 suggested their close evolutionary relationship.

DOI: 10.1134/S0006297909040099

Key words: *Salmonella enterica*, lipopolysaccharide, bacterial polysaccharide structure, acetimidoyl group, ribitol phosphate, O-antigen gene cluster

An estimated 1.4 million cases of salmonellosis and 556 (31%) estimated food-related deaths are attributed to *Salmonella* each year in the United States [1]. Serotyping of *Salmonella* based on surface O- and H-antigens has proven extremely useful for understanding the host range and disease spectrum of this pathogen and has been an invaluable typing method for epidemiological investigations. Currently, 46 *S. enterica* O-serogroups are described in the Kauffmann–White typing scheme [2]. So far, the O-antigen (O-polysaccharide) structures have been established for 22 of the O-serogroups. In this work, we established a new structure of *S. enterica* O47 O-polysaccharide, which contains 2-acetimidoylamino-2,6-dideoxy-L-galactose and ribitol phosphate, components that occur rarely in O-polysaccharides. Functional characterization of the O-antigen gene cluster of this bacterium is also reported.

MATERIALS AND METHODS

Salmonella enterica O47 type strain G1462 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 of databases for possible gene functions were performed as described [3].

Abbreviations: COSY, correlation spectroscopy; HSQC, heteronuclear single-quantum coherence; OPS, O-polysaccharide (O-antigen); ROESY, rotating-frame nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

* To whom correspondence should be addressed.

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 [4]. The lipopolysaccharide in a yield of ca. 10% was isolated from dried cells by the phenol–water method [5] and purified by precipitation of nucleic acids and proteins with aqueous 50% CCl_3CO_2H (till pH 2).

Lipopolysaccharide (120 mg) was degraded with aqueous 2% HOAc (4 ml) at 100°C until precipitation of lipid A. The precipitate was removed by centrifugation (13,000*g*, 20 min), and the supernatant was 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 O-polysaccharide (OPS) was obtained in a yield of 35% of the lipopolysaccharide weight. O-Deacetylation of the OPS was performed with 12.5% aqueous ammonia at 37°C for 16 h.

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].

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 30°C. NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using acetone (δ_H 2.225, δ_C 31.45) 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

O-Antigen structure elucidation. A high molecular mass O-polysaccharide (OPS) was obtained by mild acid degradation of the lipopolysaccharide isolated from dried cells of *S. enterica* O47 by the Westphal procedure. The OPS was separated from low-molecular mass compounds 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 Gal, FucN, GlcN, and ribitol. GLC analysis of the acetylated (*S*)-2-(+)-octyl glycosides demonstrated the D configuration of GlcN and Gal and the L configuration of FucN.

The ^{13}C-NMR spectrum of the OPS (Fig. 1) showed the presence of one ribitol and three monosaccharide residues, as followed from the total number of signals, including those for three anomeric carbons at δ 97.7-102.8, one *C*H_3–C group (C6 of FucN) at δ 16.8, four HO*C*H_2–C groups (C6 of Gal and GlcN, C1 and C5 of ribitol) at δ 62.0, 62.2, 66.4, and 67.3 (data of a DEPT experiment), two nitrogen-bearing carbons (C2 of FucN and GlcN) at δ 53.1 and 56.9, 13 oxygen-bearing sugar ring and ribitol carbons in the region δ 68.4-83.1, one N-acetyl group at δ 23.8 (CH_3) and 175.8 (CO), and one N-acetimidoyl group (Am) at δ 20.6 (CH_3) and 167.7 (C=N). In addition, there were minor signals (10% or less of the main signals) for O-acetyl groups at δ 21.6 and 21.8 (CH_3) and 174.9 (CO). Accordingly, the ^{1}H-NMR spectrum of the OPS (not shown) contained major signals for three anomeric protons at δ 4.58, 5.06, and 5.18, one

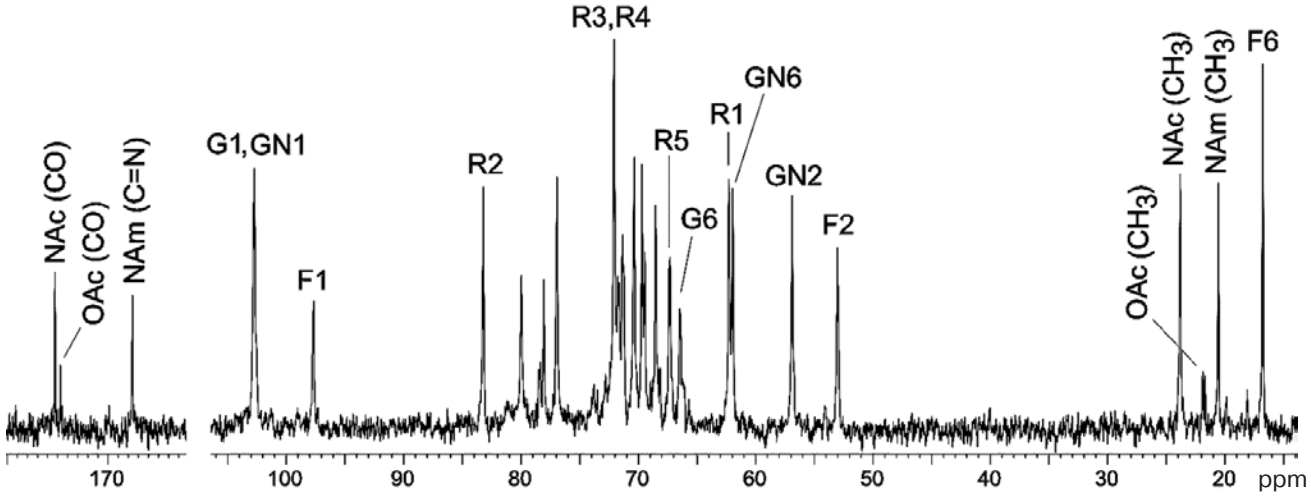

Fig. 1. ^{13}C-NMR spectrum of the O-polysaccharide from *S. enterica* O47. Arabic numerals refer to atoms in sugar residues denoted as follows: GN, GlcNAc; G, Gal; F, FucNAm; R, ribitol.

CH_3–C group (H6 of FucN) at δ 1.21, one N-acetyl group at δ 2.03, one N-acetimidoyl group at δ 2.26, and other proton signals at δ 3.49-4.45 as well as minor signals for O-acetyl groups at δ 2.16 and 2.18. The ^{31}P-NMR spectrum of the OPS showed a signal for one monophosphate group at δ 1.10. These data indicate that the OPS has a repeating unit containing one residue each of Gal, FucN, GlcN, ribitol, and phosphate group, one of the amino sugars bearing N-acetyl group and the other N-acetimidoyl group.

The ^{1}H- and ^{13}C-NMR spectra of the OPS were assigned using two-dimensional COSY, TOCSY, ROESY, ^{1}H,^{13}C HSQC, and ^{1}H,^{31}P HMQC experiments (Tables 1 and 2). The TOCSY spectrum demonstrated correlations of H1 with H2-H6 for GlcN, H2-H5 for Gal, and H2-H4 for FucN, which were assigned within each spin system using the COSY spectrum. The remaining Gal and FucN signals were assigned by H5/H6 correlations in the COSY spectrum. The amino sugars were confirmed by correlations in the HSQC spectrum between protons at the nitrogen-linked carbons and the corresponding carbons (H2/C2) at δ 3.84/56.9 and 4.05/53.1 for GlcN and FucN, respectively. Cross-peaks at δ 3.76, 3.88/62.2; 3.84/83.1; 3.83/72.2; 3.79/71.2; and 3.92, 4.00/67.3 in the HSQC spectrum were assigned to H1/C1-H5/C5 correlations of ribitol.

Relatively small $J_{1,2}$ coupling constants of ~3 Hz indicated that Gal and FucN are α-linked, whereas a significantly larger $J_{1,2}$ value of ~7 Hz showed that GlcN is β-linked. The pyranose form of all monosaccharide residues was inferred by characteristic coupling constants between neighboring protons and confirmed by the absence from the ^{13}C-NMR spectrum of any signals for non-anomeric sugar carbons at a lower field than δ 83.1 [8].

Relatively low-field positions in the ^{13}C-NMR spectrum of the OPS of the signals for C6 of Gal, C3 of FucN and GlcN, and C2 and C5 of ribitol at δ 66.4, 78.1, 80.0, 83.1, and 67.3, respectively, as compared with their positions in the corresponding non-substituted monosaccharides and ribitol [8, 9], demonstrated the modes of substitution (glycosylation or phosphorylation) of the monomers. The ^{1}H,^{31}P HMQC spectrum of the OPS showed a correlation of the phosphate group with H6a,6b of Gal at δ 1.10/3.98, 4.08, and H5a,5b of ribitol at δ 1.10/3.92, 4.00, thus indicating that Gal and ribitol are interlinked by phosphorylation at positions 6 and 5, respectively.

A two-dimensional ROESY experiment showed inter-residue cross-peaks between the anomeric protons and protons at the linkage carbons at δ 5.06/4.05; 5.18/3.71; and 4.58/3.84, which were assigned to Gal H1/FucN H3; FucN H1/GlcN H3; and GlcN H1/ribitol H2 correlations, respectively. These data are in agreement with the glycosylation pattern revealed by the ^{13}C-NMR chemical shift data (see above) and established the

Table 1. Data of the ^{1}H-NMR spectrum of the non-O-acetylated O-unit of the O-polysaccharide of *S. enterica* O47 (δ, ppm). Chemical shift for the N-acetyl group is δ 2.03 and for the N-acetimidoyl group is δ 2.26

Residue	H1 (1a, 1b)	H2	H3	H4	H5 (5a, 5b)	H6 (6a, 6b)
→2)-D-Ribitol-5-P→	3.76, 3.88	3.84	3.83	3.79	3.92, 4.00	
→6)-α-D-Gal*p*-(1→	5.06	3.79	3.86	4.02	4.22	3.98, 4.08
→3)-α-L-Fuc*p*NAm-(1→	5.18	4.05	4.05	4.10	4.45	1.21
→3)-β-D-Glc*p*NAc-(1→	4.58	3.95	3.71	3.57	3.49	3.78, 3.93

Table 2. Data of the ^{13}C-NMR spectrum of the non-O-acetylated O-unit of the O-polysaccharide of *S. enterica* O47 (δ, ppm). Chemical shifts for the N-acetyl group are δ 23.8 (CH_3) and 175.8 (CO), and the N-acetimidoyl group δ 20.6 (CH_3) and 167.7 (C=N)

Residue	C1	C2	C3	C4	C5	C6
→2)-D-Ribitol-5-P→	62.2	83.1	72.2	71.2	67.3	
→6)-α-D-Gal*p*-(1→	102.8	69.4	70.3	70.5	71.7	66.4
→3)-α-L-Fuc*p*NAm-(1→	97.7	53.1	78.1	72.0	68.4	16.8
→3)-β-D-Glc*p*NAc-(1→	102.7	56.9	80.0	69.5	76.9	62.0

sequence of the monosaccharide residues in the repeating unit of the OPS.

The distribution of the N-acyl substituents was determined using NMR experiments with an OPS 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 four NH protons at δ 8.47, 8.55, 8.79, and 8.86. Using a TOCSY experiment, the NH protons at δ 8.55 and 8.86 were assigned to FucN and GlcN, respectively. The NOESY spectrum showed correlations of the NH proton of FucN with CH_3C=N of the N-acetimidoyl group at δ 8.86/2.26 and the NH proton of GlcN with CH_3CO of the N-acetyl group at δ 8.55/2.03, thus indicating N-acetylation of GlcN and N-acetimidoylation of FucN. Correlations in the NOESY spectrum of the NH protons of the N-acetimidoyl group at δ 8.47 and 8.79 with H2 of FucN at δ 4.05 confirmed the presence of FucNAm.

O-Deacetylation of the OPS with aqueous ammonia was accompanied by conversion of the N-acetimidoyl group into an N-acetyl group. This caused a significant change in the ^{13}C-NMR chemical shift of the C2 signal of FucN from δ 53.1 to 49.7, which further confirmed location of the N-acetimidoyl group on FucN. A comparison of the ^{1}H,^{13}C HSQC spectra of the initial and O-deacetylated OPS showed a significant downfield displacement of a minor part (ca. 10%) of the H4/C4 cross-peak of Gal from δ_H/δ_C 4.02/70.5 to 5.38/72.8 (assignment of the signal at δ 5.38 to H4 of Gal was performed using COSY and TOCSY experiments). This downfield shift in both dimensions was due to a deshielding effect of the O-acetyl group and indicated non-stoichiometric O-acetylation of the Gal residue at position 4. The position of the other minor O-acetyl group was not determined owing to its too low content. Therefore, the O-polysaccharide of *S. enterica* O47 has the following structure, where the degree of O-acetylation of Gal is ca. 10%:

→2)-D-Ribitol-5-P-(O→6)-α-D-Gal*p*4Ac-
(1→3)-α-L-Fuc*p*NAm-(1→3)-β-D-Glc*p*NAc-(1→.

As in a number of *E. coli* O-antigens studied earlier [10], GlcNAc seems to be the first sugar of the biological O-unit, whose transfer to a lipid carrier initiates the O-antigen biosynthesis [11]. Interestingly, the *S. enterica* O47 polysaccharide shares a disaccharide fragment with the O-polysaccharides of *S. enterica* sv. Toucra O48, *S. enterica* ssp. *arizonae* O21, and *Escherichia coli* O145 [12] (see below).

O-Antigen gene cluster characterization. The O-antigen gene cluster of *S. enterica* O47 was found between housekeeping genes *galF* and *gnd*, and was sequenced. Thirteen open reading frames (*orfs*) excluding *galF* and *gnd* were identified, all of which have the same transcriptional direction from *galF* to *gnd*. Functions were assigned to most *orfs* based on their similarity to related genes from the available databases and taking into account the *S. enterica* O47 antigen structure. As genes for the synthesis of common sugars (GlcNAc and Gal in case of *S. enterica* O47) are located outside the O-antigen gene cluster [13], only genes for the synthesis of UDP-L-FucNAm and CDP-ribitol were expected in the cluster.

Orf1 shares 45% identity to Acs1 of *Haemophilus influenzae*. Acs1 was identified as a bifunctional enzyme that converts ribulose 5-phosphate into D-ribitol 5-phosphate and further into CDP-ribitol, which is known as the activated precursor form for incorporation of ribitol 5-phosphate into the capsular polysaccharide [14]. Therefore, *orf1* gene was proposed to be responsible for the synthesis of CDP-ribitol. Orf9, Orf10, and Orf11 share 92, 78, and 94% identity to FnlA, FnlB, and FnlC, respectively, of the *Escherichia coli* O145 antigen gene cluster, which have been identified as enzymes of the UDP-L-FucNAc biosynthesis pathway [15]. Therefore, we proposed that Orf9-11 in *S. enterica* O47 are responsible for the synthesis of UDP-L-FucNAc, which is an intermediate in the synthesis of the L-FucNAm.

Orf6 share 85% identity to WbuX of *E. coli* O145, which has been identified as aminotransferase for synthesis of L-FucNAm from L-FucNAc [12, 16]. Therefore, *orf6* was proposed to be the aminotransferase gene involved in the synthesis of L-FucNAm. Orf7 and Orf8 share 83 and 88% identity to WbuY and WbuZ, respectively, of *E. coli* O145 antigen gene cluster. It has been proposed that they both are related with the synthesis of N-acetimidoyl group: WbuZ acts as a glutaminase and, with WbuY, conducts ammonia to WbuX [12]. We suggest that Orf7 and Orf8 in *S. enterica* O47 have the same functions, which need experimental evidence.

Orf2 and Orf4 are the only two proteins with predicted transmembrane segments. Orf2 was predicted to have 12 well-proportioned transmembrane segments, which is a typical topology for O-antigen flippase (Wzx). Orf2 shares 53% similarity to Wzx of *Yersinia enterocolitica* O3, and belongs to a flippase protein family (PF01943, *E* value = $1.4{\cdot}e^{-34}$). Orf4 was found to have 10 predicted transmembrane segments. It shares 49 and 45% similarity to O-antigen polymerase (Wzy) of *E. coli* O104 and *Shigella boydii* type 11, respectively. Therefore, *orf2* and *orf4* were proposed to be *wzx* and *wzy* genes, respectively, which are O-antigen processing genes responsible for the translocation and polymerization of the O-unit.

In most *E. coli*, *Shigella*, and *S. enterica* strains, transfer of GlcNAc-1-P or GalNAc-1-P to an undecaprenol phosphate (UndP) carrier catalyzed by WecA initiates the O-unit synthesis. The *wecA* gene is located outside the O-antigen gene cluster [10, 15], and, therefore, only three transferase genes were expected in the O-antigen gene cluster of *S. enterica* O47.

Orf3 shares 47% similarity to a putative LicD-family phosphotransferase of *S. pneumoniae*. Orf12 is 85% identical to WbuB of *E. coli* O145, which is L-fucosamine

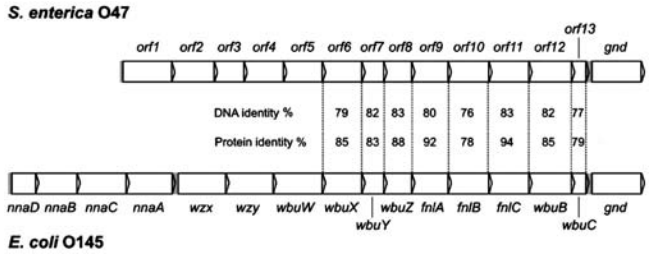

Fig. 2. Comparison of the O-antigen gene cluster of *S. enterica* O47 and *E. coli* O145. All genes are transcribed in the direction from *galF* to *gnd*.

transferase involved in the O-antigen synthesis. Orf5 shares 52% similarity to glycosyltransferase Cap33fG of *Streptococcus pneumoniae* and belongs to glycosyl transferases group 1 family (PF00534, E value = $3.5 \cdot e^{-26}$). Therefore, *orf3*, *orf5*, and *orf12* in *S. enterica* O47 were proposed to be putative transferase genes, which are responsible for sequential transfer of the activated derivatives of D-ribitol 5-phosphate, D-Gal, and L-FucNAm, respectively.

Orf13 share 79% identity to WbuC of *E. coli* O145, whose function has not been assigned.

The O-antigen gene cluster of *E. coli* O145 has been sequenced and analyzed by us earlier [12]. When the O-antigen gene clusters of *S. enterica* O47 and *E. coli* O145 were compared, it was found that their last 8 genes share DNA identity between 76 and 83% or protein identity between 78 and 94% (Fig. 2). These loci include genes for the synthesis and transfer of L-FucNAm, and their similarity is in agreement with the occurrence of the common →3)-α-L-Fuc*p*NAm-(1→3)-β-D-Glc*p*NAc-(1→ disaccharide fragment in the O-antigens of the two bacteria.

Escherichia coli and *S. enterica* diverged about 140 million years ago [17], and typical homologous proteins are, on average, 93% identical in the two species, ranging from 76.3 to 100% [18]. Three pairs of the identical O-antigens have been identified in *E. coli* and *S. enterica*, which are those of *E. coli* O111–*S. enterica* O35, *E. coli* O55–*S. enterica* O50, and *E. coli* O157–*S. enterica* O30. Their identical O-antigen gene clusters have been proposed to originate from a common ancestor [19]. Our data suggest that a close evolutionary relationship occurs between *S. enterica* O47 and *E. coli* O145, and a part of their O-antigen gene clusters may come from a common ancestor too.

This work was supported by the Russian Foundation for Basic Research (08-04-01205-a), RFFI-NSFC (08-04-92225), the Council on Grants at the President of the Russian Federation for Support of Young Russian Scientists (MK-157.2007.4), the National Natural Science Foundation of China (NSFC) Key Programs Grants 30530010 and 20536040, NSFC General Program Grant 30670038, and the Chinese National Science Fund for Distinguished Young Scholars (30788001).

REFERENCES

1. Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., and Tauxe, R. V. (1999) *Emerg. Infect. Dis.*, **5**, 607-625.
2. Kauffmann, F. (1966) *The Bacteriology of Enterobacteriaceae*, Williams and Wilkins, Baltimore.
3. 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.
4. Robbins, P. W., and Uchida, T. (1962) *Biochemistry*, **1**, 323-335.
5. Westphal, O., and Jann, K. (1965) *Meth. Carbohydr. Chem.*, **5**, 83-91.
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. Bock, K., and Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem.*, **41**, 27-66.
9. Shashkov, A. S., Lipkind, G. M., Knirel, Y. A., and Kochetkov, N. K. (1988) *Magn. Reson. Chem.*, **26**, 735-747.
10. Stenutz, R., Weintraub, A., and Widmalm, G. (2006) *FEMS Microbiol. Rev.*, **30**, 382-403.
11. Raetz, C. R. H., and Whitfield, C. (2002) *Annu. Rev. Biochem.*, **71**, 635-700.
12. Feng, L., Senchenkova, S. N., Tao, J., Shashkov, A. S., Liu, B., Shevelev, S. D., Reeves, P., Xu, J., Knirel, Y. A., and Wang, L. (2005) *J. Bacteriol.*, **187**, 758-764.
13. Samuel, G., and Reeves, P. (2003) *Carbohydr. Res.*, **338**, 2503-2519.
14. Mamat, U., Meredith, T. C., Aggarwal, P., Kuhl, A., Kirchhoff, P., Lindner, B., Hanuszkiewicz, A., Sun, J., Holst, O., and Woodard, R. W. (2008) *Mol. Microbiol.*, **67**, 633-648.
15. Kneidinger, B., O'Riordan, K., Li, J., Brisson, J. R., Lee, J. C., and Lam, J. S. (2003) *J. Biol. Chem.*, **278**, 3615-3627.
16. King, J. D., Mulrooney, E. F., Vinogradov, E., Kneidinger, B., Mead, K., and Lam, J. S. (2008) *J. Bacteriol.*, **190**, 1671-1679.
17. Ochman, H., and Wilson, A. C. (1987) *J. Mol. Evol.*, **26**, 74-86.
18. Sharp, P. M. (1991) *J. Mol. Evol.*, **33**, 23-33.
19. Samuel, G., Hogbin, J. P., Wang, L., and Reeves, P. R. (2004) *J. Bacteriol.*, **186**, 6536-6543.