

Structure and Biosynthesis Gene Cluster of the O-Antigen of *Escherichia coli* O12

S. N. Senchenkova¹, Yuanyuan Zhang², A. V. Perepelov¹,
Xi Guo², A. S. Shashkov¹, Bin Liu², and Yu. A. Knirel^{1*}

¹Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russia;
fax: +7 (499) 135-5328; E-mail: yknirel@gmail.com

²Nankai University, TEDA Institute of Biological Sciences and Biotechnology,
300457 Tianjin, P. R. China; E-mail: striker198126@aliyun.com

Received October 5, 2015

Revision received November 21, 2015

Abstract—Two polysaccharides were isolated from *Escherichia coli* O12, the major being identified as the O12-antigen and the minor as the K5-antigen. The polysaccharides were studied by sugar analysis, Smith degradation, and one- and two-dimensional ¹H and ¹³C NMR spectroscopy. As a result, the following structure of the O12-polysaccharide was elucidated, which, to our knowledge, has not been hitherto found in bacterial carbohydrates: →2)-β-D-Glcp-(1→6)-α-D-GlcpNAc-(1→3)-α-L-FucpNAc-(1→3)-β-D-GlcpNAc-(1→. The →4)-β-D-GlcpA-(1→4)-α-D-GlcpNAc-(1→ structure established for the K5-polysaccharide (heparosan) is previously known. Functions of genes in the O-antigen biosynthesis gene cluster of *E. coli* O12 were assigned by comparison with sequences in the available databases and found to be consistent with the O12-polysaccharide structure.

DOI: 10.1134/S0006297916040106

Key words: *Escherichia coli*, O-polysaccharide structure, O-antigen gene cluster, glycosyltransferase

Escherichia coli bacteria are the predominant species in the human intestinal microflora. They include both pathogenic and commensal strains and are one of the most common causes of diarrheal diseases [1]. *Escherichia coli* isolates are classified into serotypes by a combination of their O-, H-, and sometimes K-antigens, which together define the serological specificity of the bacteria. The O-antigen represents a polysaccharide chain of the lipopolysaccharide (O-polysaccharide) and the K-antigen is a capsular polysaccharide. Antigen diversity helps survival of various bacterial clones in their specific niche. The surface antigens also are an important virulence factor, and their loss makes many pathogens serum sensitive or otherwise seriously impaired in virulence. Specific genes involved in their biosynthesis are generally localized in clusters on the chromosome [2, 3],

whose variations define the antigenic diversity of bacteria to a large extent. Currently, more than 180 *E. coli* O-serogroups are recognized [4].

This work was performed in the framework of a project devoted to elucidation of the structural and genetic diversity of *E. coli* O-antigens. Data on such a large number of various clones of one species are expected to shed light on the evolution history of the O-antigen diversification of enteric bacteria, to enable assignment of functions of genes involved in the O-antigen biosynthesis, including hundreds of glycosyltransferases, and improvement of classification scheme of *E. coli* isolates necessary for diagnostics of infectious diseases, epidemiological monitoring, and elaboration of vaccines against pathogenic strains. By now, O-antigen structures have been elucidated for most of the O-serogroups (<http://nevyn.org.au/su.se/ECODAB/>), but some remain unknown. Here we report data on the O-antigen of *E. coli* O12, which belonged to one of the O-serogroups that had not been investigated yet. The O12-polysaccharide structure was established, and the gene cluster for its biosynthesis was characterized. In addition to the O-polysaccharide, a minor polysaccharide was isolated from the bacterium and identified as the K5-antigen.

Abbreviations: ABC, ATP-binding cassette; COSY, correlation spectroscopy; FucNAc, 2-acetamido-2-deoxyfucose; GlcA, glucuronic acid; GlcNAc, 2-acetamido-2-deoxyglucose; HSQC, heteronuclear single-quantum coherence; ROESY, rotating-frame nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

* To whom correspondence should be addressed.

MATERIALS AND METHODS

RESULTS AND DISCUSSION

Bacterial strain and growth. *Escherichia coli* O12 type strain (laboratory stock number G1280) was obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia (IMVS). The bacterium was grown to late log phase in 8 liters of Luria–Bertani broth using a 10-liter fermenter (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 [5].

Isolation of the lipopolysaccharide and polysaccharides. The lipopolysaccharide was isolated in a yield 8.5% from bacterial cells by the phenol–water method [6], and to remove phenol, the crude extract was dialyzed without separation of the layers and freed from nucleic acids and proteins by treatment with 50% aqueous $\text{CCl}_3\text{CO}_2\text{H}$ to pH 2.0 at 4°C. The supernatant was dialyzed and lyophilized.

Mild acid degradation of the lipopolysaccharide (150 mg) was performed with 2% aqueous HOAc at 100°C until precipitation of lipid (80 min). The precipitate was removed by centrifugation (13,000g, 20 min), and a mixture of polysaccharides (40 mg) was isolated by gel-permeation chromatography 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 using a differential refractometer (Knauer, Germany).

Chemical analyses. A sample of the polysaccharides (0.5 mg) was hydrolyzed with 2 M $\text{CF}_3\text{CO}_2\text{H}$ (120°C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on a Maestro (Agilent 7820) chromatograph (Interlab, Russia) equipped with an HP-5 column (0.32 mm × 30 m) and a temperature program of 160°C (1 min) to 290°C at rate 7°C/min. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides as described [7].

Smith degradation. A sample of the polysaccharides (20 mg) was oxidized with 0.1 M NaIO_4 (2 ml) in the dark for 48 h at 20°C, reduced with an excess of NaBH_4 , and desalted by dialysis. The products were hydrolyzed with 2% aqueous HOAc for 2 h at 100°C and fractionated by gel-permeation chromatography on a column (85 × 2.5 cm) of TSK HW-40 (S) in aqueous 1% HOAc to give oligosaccharide **1** (5 mg) and glycoside **2** (3.5 mg).

NMR spectroscopy. Samples were deuterium-exchanged by freeze-drying from 99.9% D_2O and then examined as solution in 99.95% D_2O . NMR spectra were recorded on an Avance II 600 MHz spectrometer (Bruker, Germany) at 40°C for the polysaccharides or 25°C for compounds **1** and **2** using sodium 3-trimethylsilylpropanoate-2,2,3,3- d_4 (δ_{H} 0, δ_{C} -1.6) as internal reference for calibration. Two-dimensional NMR spectra were obtained using standard Bruker software, and the Bruker TopSpin 2.1 program was used to acquire and process the NMR data. Mixing times of 100 and 150 ms were used in TOCSY and ROESY experiments, respectively.

Elucidation of polysaccharide structures. Lipopolysaccharide was obtained from *E. coli* O12 cells by phenol–water extraction and degraded with mild acid to give a mixture of polysaccharides isolated by Sephadex G-50 gel-permeation chromatography. Sugar analysis using GLC of the alditol acetates after full acid hydrolysis of the polysaccharides revealed Glc, GlcNAc, and FucNAc. GLC analysis of the acetylated (*S*)-2-octyl glycosides demonstrated the D-configuration of Glc and GlcNAc and the L-configuration of FucNAc.

The ^1H and ^{13}C NMR spectra showed two series of signals, belonging to major and minor polysaccharides consisting of tetrasaccharide and disaccharide repeating units, respectively. In the major series, there were signals for four anomeric atoms, one $\text{CH}_3\text{-C}$ group (C6 of FucNAc), three $\text{HOCH}_2\text{-C}$ groups (C6 of hexoses), and three nitrogen-bearing carbons (C2 of amino sugars). In the minor series, there were signals for two anomeric atoms, one $\text{HOCH}_2\text{-C}$ group, and one nitrogen-bearing carbon. The ^1H and ^{13}C NMR chemical shifts are given in the table.

The ^1H and ^{13}C NMR spectra of the polysaccharides were assigned using 2D $^1\text{H}, ^1\text{H}$ COSY, TOCSY, and $^1\text{H}, ^{13}\text{C}$ HSQC experiments, and spin systems for the major monosaccharides (β -Glc p NAc **A**, α -Fuc p NAc **B**, α -Glc p NAc **C**, β -Glc p **D**) and minor monosaccharides (α -Glc p NAc **E**, β -Glc p **F**) were recognized by characteristic chemical shifts [8, 9] and $^3J_{\text{H,H}}$ coupling constants. Significant downfield displacements of the signals for C2 of unit **D**, C3 of units **A** and **B**, and C6 of unit **C** in the major series as well as C4 of units **E** and **F** in the minor series, as compared with their positions in the spectra of the corresponding non-substituted monosaccharides [8, 9], were due to glycosylation and enabled determination of the positions of substitution of sugar residues in the repeating units.

A two-dimensional ROESY spectrum of the polysaccharides showed correlations between the anomeric protons and protons at the linkage carbons, which were assigned as follows: in the major series, for **D** H1/C H6, **C** H1/**B** H3, **B** H1/**A** H3, **A** H1/**D** H2; in the minor series, for **F** H1/**E** H4, **E** H1/**F** H4. These data are in agreement with positions of substitution and define the sequences of the monosaccharides in the repeating units.

The polysaccharide structures were confirmed by Smith degradation, which resulted in destruction of the Glc and α -GlcNAc residues in the major polymer and the GlcA residue in the minor polymer. As a result, oligosaccharide **1** and glycoside **2** were obtained from the major and minor polysaccharides, respectively, and were separated by gel chromatography. Their structures shown in Fig. 1 were established by one- and two-dimensional NMR spectroscopy as described above and were consistent with the structures of the initial polysaccharides.

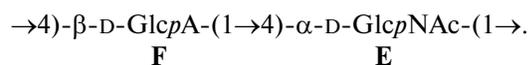
^1H and ^{13}C NMR chemical shifts (δ , ppm)

Sugar residue	Nucleus	1	2	3	4	5	6
O12-polysaccharide							
$\rightarrow 3$)- β -D-GlcpNAc-(1 \rightarrow	^1H	4.79	3.88	3.69	3.50	3.44	3.80, 3.94
A	^{13}C	102.6	57.1	79.3	69.9	77.1	62.3
$\rightarrow 3$)- α -L-FucpNAc-(1 \rightarrow	^1H	4.99	4.28	3.97	3.88	4.41	1.16
B	^{13}C	98.7	49.7	75.0	72.1	67.8	16.5
$\rightarrow 6$)- α -D-GlcpNAc-(1 \rightarrow	^1H	5.01	3.96	3.76	3.77	3.84	3.97, 4.08
C	^{13}C	99.9	54.7	72.1	72.0	72.8	70.0
$\rightarrow 2$)- β -D-Glcp-(1 \rightarrow	^1H	4.51	3.49	3.49	3.38	3.40	3.71, 3.89
D	^{13}C	103.3	80.8	77.9	71.0	76.9	61.9
K5-polysaccharide							
$\rightarrow 4$)- α -D-GlcpNAc-(1 \rightarrow	^1H	5.36	3.90	3.81	3.70	3.78	3.79, 3.87
E	^{13}C	98.4	54.5	70.1	79.6	70.3	60.6
$\rightarrow 4$)- β -D-GlcpA-(1 \rightarrow	^1H	4.55	3.37	3.72	3.80	4.02	
F	^{13}C	103.5	74.5	77.1	77.5	75.9	not found
Oligosaccharide 1							
α -L-FucpNAc-(1 \rightarrow	^1H	5.01	4.11	3.90	3.77	4.42	1.19
B	^{13}C	99.0	50.7	69.0	72.4	68.1	16.6
$\rightarrow 3$)- β -D-GlcpNAc-(1 \rightarrow	^1H	4.64	3.85	3.68	3.53	3.47	3.78, 3.94
A	^{13}C	102.2	56.9	79.8	69.8	77.1	62.1
R	^1H	4.93	3.91	3.70, 3.82	3.66, 3.92	3.86	3.60, 3.68
D'	^{13}C	103.8	81.0	61.5	70.2	81.0	63.8
Glycoside 2							
α -D-GlcpNAc-(1 \rightarrow	^1H	5.11	3.94	3.78	3.50	3.89	3.79, 3.86
E	^{13}C	99.0	55.2	72.3	71.3	73.5	61.8
$\rightarrow 2$)-erythronic acid	^1H		3.97	4.25	3.72, 3.76		
F'	^{13}C	not found	82.3	74.5	61.7		

Note: Chemical shifts for the N-acetyl groups are for the K5- and O12-polysaccharides: δ_{H} 1.97-2.04, δ_{C} 23.2-23.6 (Me) and 174.9-175.2 (CO); for oligosaccharide **1**: δ_{H} 1.98 and 2.04, δ_{C} 23.4, 23.5 (both Me), 175.4 and 175.7 (both CO); for glycoside **2**: δ_{H} 2.04, δ_{C} 23.2 (Me) and 175.7 (CO).

Based on the data obtained, it was concluded that the major polysaccharide from *E. coli* O12 has the structure shown in Fig. 2. Genetic analysis indicated that this polysaccharide represents the O-antigen of *E. coli* O12 (see below).

The minor polysaccharide from *E. coli* O12 has the structure of heparosan:



It is known as the capsular polysaccharide of *E. coli* (K5-antigen) [3, 10] and *Pasteurella multocida* type D [11]. Heparosan is a biosynthetic precursor of biologically important mammalian polysaccharides heparin and heparan sulfate [12].

Characterization of the O-antigen biosynthesis gene cluster. Most heteropolysaccharide O-antigens of *E. coli* are synthesized by the flippase Wzx/polymerase Wzy-dependent pathway, including the assembly of the O-unit

accession number ACD37144.1). Orf4 shares 30% identity and 50% similarity to predicted glycosyltransferase WfbF of *E. coli* O123 [17] (GenPept accession number ABG81791.1). The polysaccharides of O164 [15, 18] and O123 [19] have the β -D-Glcp-(1 \rightarrow 6)- α -D-Glcp and α -D-GlcpNAc-(1 \rightarrow 3)- α -L-QuipNAc disaccharide fragments, which are similar to the β -D-Glcp-(1 \rightarrow 6)- α -D-GlcpNAc and α -D-GlcpNAc-(1 \rightarrow 3)- α -L-FucpNAc disaccharide fragments, respectively, of the O12-polysaccharide. Therefore, Orf1 is evidently responsible for the transfer of D-Glcp from UDP-D-Glcp to GlcpNAc and Orf4 for the transfer of D-GlcpNAc from UDP-D-GlcpNAc to L-FucpNAc (Fig. 2).

As in many other O-serogroups [20], GlcNAc is likely the first sugar of the O-unit of *E. coli* O12. The *wecA* gene initiates the O-antigen biosynthesis by transfer of GlcpNAc 1-phosphate from UDP-D-GlcpNAc to undecaprenyl phosphate (UndP) to give UndPP- α -D-GlcpNAc. The *wecA* gene is located in the enterobacterial common antigen gene cluster [21] and is not duplicated in the O-antigen gene cluster.

The K5-polysaccharide (K5-antigen) belongs to group 2 capsules, which are commonly expressed in pathogenic extraintestinal *E. coli* and closely resemble the capsules of *Neisseria meningitidis* and *Haemophilus influenzae* [22]. The biosynthesis of the K5 polysaccharide in *E. coli* is well studied both biochemically and genetically [3, 22]. Briefly, it occurs by the ABC transporter-dependent pathway through the sequential addition of GlcA and GlcNAc residues to the nonreducing end of the growing polysaccharide chain catalyzed by two glycosyltransferases, KfiA and KfiC. The mature polymer is exported across the inner membrane by the ABC transporter. The proteins involved are encoded by *kfiABCD* genes in the *kps* locus located near the *serA* gene [23].

Therefore, two polysaccharides were found in bacteria *E. coli* O12. The major polysaccharide has a structure that, to our knowledge, has not been hitherto identified in bacterial carbohydrates. The genetic content of the O-antigen gene cluster of *E. coli* O12 is fully consistent with the established structure of this polysaccharide, and hence it represents the O12-antigen. Functions of genes and the corresponding proteins involved in biosynthesis of the O12-polysaccharides were determined. Particularly, each predicted glycosyltransferase was assigned to one of three glycosidic linkages that form the tetrasaccharide repeating unit. Our data extend the notion of the genetic basis of the structural diversity of the O-antigens of enteric bacteria and the database of carbohydrate biosynthesis enzymes with predicted functions. The minor polysaccharide was identified as the K5-antigen (heparosan), whose structure and genetics of biosynthesis have been elucidated earlier.

This work was supported by the Russian Science Foundation (project No. 14-14-01042).

Xi Guo was supported by the International Science and Technology Cooperation Program of China (2012DFG31680 and 2013DFR30640), the National Key Program for Infectious Diseases of China (2013ZX10004216-001-001 and 2013ZX10004221-003), the National Natural Science Foundation of China (NSFC) Program (31371259 and 81471904), and the Research Project of Chinese Ministry of Education (No. 113015A).

REFERENCES

1. Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004) Pathogenic *Escherichia coli*, *Nat. Rev. Microbiol.*, **2**, 123-140.
2. Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., Hayashi, T., and Thomson, N. R. (2014) A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster, *DNA Res.*, **22**, 105-107.
3. Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*, *Annu. Rev. Biochem.*, **75**, 39-68.
4. Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015) Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data, *J. Clin. Microbiol.*, **53**, 2410-2426.
5. Robbins, P. W., and Uchida, T. (1962) Studies on the chemical basis of the phage conversion of O-antigens in the E-group *Salmonellae*, *Biochemistry*, **1**, 323-335.
6. Westphal, O., and Jann, K. (1965) Bacterial lipopolysaccharides. Extraction with phenol-water and further applications of the procedure, *Methods Carbohydr. Chem.*, **5**, 83-91.
7. Leontein, K., and Lönngren, J. (1993) Determination of the absolute configuration of sugars by gas-liquid chromatography of their acetylated 2-octylglycosides, *Methods Carbohydr. Chem.*, **9**, 87-89.
8. Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V., and Kochetkov, N. K. (1988) A computer-assisted structural analysis of regular polysaccharides on the basis of ^{13}C -n.m.r. data, *Carbohydr. Res.*, **175**, 59-75.
9. Jansson, P.-E., Kenne, L., and Widmalm, G. (1989) Computer-assisted structural analysis of polysaccharides with an extended version of CASPER using ^1H - and ^{13}C -n.m.r. data, *Carbohydr. Res.*, **188**, 169-191.
10. Vann, W. F., Schmidt, M. A., Jann, B., and Jann, K. (1981) The structure of the capsular polysaccharide (K5 antigen) of urinary-tract-infective *Escherichia coli* O10:K5:H4, *Eur. J. Biochem.*, **116**, 359-364.
11. DeAngelis, P. L., Gunay, N. S., Toida, T., Mao, W. J., and Linhardt, R. J. (2002) Identification of the capsular polysaccharides of type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively, *Carbohydr. Res.*, **337**, 1547-1552.
12. Sugahara, K., and Kitagawa, H. (2002) Heparin and heparan sulfate biosynthesis, *IUBMB Life*, **54**, 163-175.
13. Reeves, P. R., and Wang, L. (2002) Genomic organization of LPS-specific loci, *Curr. Top. Microbiol. Immunol.*, **264**, 109-135.
14. Kneidinger, B., O'Riordan, K., Li, J., Brisson, J. R., Lee, J. C., and Lam, J. S. (2003) Three highly conserved proteins

- catalyze the conversion of UDP-*N*-acetyl-D-glucosamine to precursors for the biosynthesis of O antigen in *Pseudomonas aeruginosa* O11 and capsule in *Staphylococcus aureus* type 5. Implications for the UDP-*N*-acetyl-L-fucosamine biosynthetic pathway, *J. Biol. Chem.*, **78**, 3615-3627.
15. Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R., and Wang, L. (2008) Structure and genetics of *Shigella* O antigens, *FEMS Microbiol. Rev.*, **32**, 627-653.
 16. D'Souza, J. M., Samuel, G. N., and Reeves, P. R. (2005) Evolutionary origins and sequence of the *Escherichia coli* O4 O-antigen gene cluster, *FEMS Microbiol. Lett.*, **244**, 27-32.
 17. Beutin, L., Wang, Q., Naumann, D., Han, W., Krause, G., Leomil, L., Wang, L., and Feng, L. (2007) Relationship between O-antigen subtypes, bacterial surface structures and O-antigen gene clusters in *Escherichia coli* O123 strains carrying genes for Shiga toxins and intimin, *J. Med. Microbiol.*, **56**, 177-184.
 18. Linnerborg, M., Weintraub, A., and Widmalm, G. (1999) Structural studies of the O-antigen polysaccharide from the enteroinvasive *Escherichia coli* O164 cross-reacting with *Shigella dysenteriae* type 3, *Eur. J. Biochem.*, **266**, 460-466.
 19. Perepelov, A. V., Liu, B., Shevelev, S. D., Senchenkova, S. N., Shashkov, A. S., Feng, L., Knirel, Y. A., and Wang, L. (2010) Relatedness of the O-polysaccharide structures of *Escherichia coli* O123 and *Salmonella enterica* O58, both containing 4,6-dideoxy-4-{*N*-(*S*)-3-hydroxybutanoyl]-D-alanyl]amino-D-glucose; revision of the *E. coli* O123 O-polysaccharide structure, *Carbohydr. Res.*, **345**, 825-829.
 20. Rojas-Macias, M. A., Stahle, J., Lutteke, T., and Widmalm, G. (2015) Development of the ECODAB into a relational database for *Escherichia coli* O-antigens and other bacterial polysaccharides, *Glycobiology*, **25**, 341-347.
 21. Lehrer, J., Vigeant, K. A., Tatar, L. D., and Valvano, M. A. (2007) Functional characterization and membrane topology of *Escherichia coli* WecA, a glycosyltransferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide, *J. Bacteriol.*, **189**, 2618-2628.
 22. Whitfield, C., and Roberts, I. S. (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli*, *Mol. Microbiol.*, **31**, 1307-1319.
 23. Rigg, G. P., Barrett, B., and Roberts, I. S. (1998) The localization of KpsC, S and T, and KfiA, C and D proteins involved in the biosynthesis of the *Escherichia coli* K5 capsular polysaccharide: evidence for a membrane-bound complex, *Microbiology*, **144**, 2905-2914.