Carbohydrate Research, 246 (1993) 283–290
Elsevier Science Publishers B.V., Amsterdam

Structural elucidation of the capsular polysaccharide of *Escherichia coli* serotype K101 by high resolution NMR spectroscopy

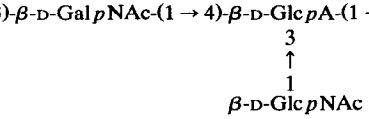
M. Ruth Grue, Haralambos Parolis * and Lesley A.S. Parolis

School of Pharmaceutical Sciences, Rhodes University, Grahamstown 6140 (South Africa)

(Received December 17th, 1992; accepted March 3rd, 1993)

ABSTRACT

The primary structure of the acidic capsular antigen of *Escherichia coli* O20:K101 was shown by glycose analysis, methylation analysis, and one- and two-dimensional ^{1}H and ^{13}C NMR spectroscopy to be composed of repeating branched-tetrasaccharide units having the structure:



INTRODUCTION

The great majority of the capsular (K) antigens of *Escherichia coli* are acidic polysaccharides and have been subdivided into two groups on the basis of their physical, chemical, and microbiological characteristics[1]. Group I polysaccharides are usually co-expressed with O groups 8, 9, 20, and 101. They are high molecular weight, acidic polysaccharides, are heat stable at pH 5–6, and may contain amino sugars. The capsular polysaccharide of *E. coli* O20:K101 belongs to the subgroup of I whose capsules do contain amino sugars.

RESULTS AND DISCUSSION

Isolation, composition, and linkage analysis of the capsular antigen.—*E. coli* K101 bacteria were grown on Mueller–Hinton agar and the acidic capsular polysaccharide (**PS**) was isolated and purified by precipitation with cetyltrimethylammonium

* Corresponding author.

bromide. A small sample was further purified for NMR spectroscopy by gel-permeation chromatography (GPC) on Sephacryl S500. The **PS** showed a broad distribution of molecular weights with an average M_r, at 1.5×10^5. Hydrolysis of **PS** followed by GLC–MS examination of the derived alditol acetates showed that Gal, GalN, GlcN, and Glc were present. Methanolysis and reduction of **PS** with $NaBH_4$, followed by hydrolysis, reduction, and acetylation, revealed no additional monosaccharides, but the proportion of Glc present increased, indicating the presence of GlcA. A further hydrolysis of the polysaccharide was performed, followed by reduction with $NaBD_4$, and the derived alditol acetates were examined by GLC–MS. The Glc component exhibited fragments two mass units higher than the corresponding Gal component, indicating that the Glc was in fact all derived from the reduction of GlcA and that there was no Glc present in the polysaccharide. Partial reduction of GlcA to Glc occurs via the formation of the glucurono-3,6-lactone during hydrolysis. The constituent sugars were shown to have the D configuration by GLC analysis of the derived acetylated (−)-2-octyl glycosides[2].

The ^1H NMR spectrum of **PS** in D_2O contained H-1 signals at δ 4.85 ($J_{1,2}$ 8.2 Hz), 4.75 ($J_{1,2}$ 6.1 Hz), 4.57 ($J_{1,2}$ 8.0 Hz), and 4.52 (broad doublet), and two signals for the methyl protons of two NAc groups at δ 2.03 and 2.06 (3 H each). The chemical shifts of the H-1 signals indicated that all linkages were β. The ^{13}C NMR data complemented the ^1H NMR results and confirmed the tetrasaccharide repeating unit for **PS**, with signals for C-1 at 104.77, 103.17, 102.16, and 102.69 ppm, and signals for carbonyl carbons at 175.25, 175.53, and 172.07 ppm. Signals at 56.92 and 51.86 ppm indicated the presence of two C–N bonds, and confirmed the presence of two NAc groups.

Methylation analysis of the polysaccharide gave 2-deoxy-4,6-di-*O*-methyl-2-methylacetamidogalactose, 2,4,6-tri-*O*-methylgalactose, 2-deoxy-3,4,6-tri-*O*-methyl-2-methylacetamidoglucose, and 2-*O*-methylglucose (after carboxyl reduction). These results indicated the presence of 3-linked Gal, 3-linked GlcNAc, terminal GlcNAc, and 3,4-linked GlcA in a branched-tetrasaccharide repeating unit with GlcA as the branch point.

2D NMR studies of the E. coli *K101 polysaccharide.*—The sequence of the residues in the repeating unit was established by 2D NMR experiments, which also confirmed the glycosylation sites in the polysaccharide. The residues in the repeating unit were labelled **a**–**d** in order of decreasing chemical shift of their anomeric protons (Fig. 1). Most of the proton resonances were established from COSY[3], one- and two-step relay COSY[4], and 2D Homonuclear Hartmann–Hahn (HOHAHA)[5] experiments. The two-step relay COSY is shown in Fig. 1.

Residue **a** [*β*-D-*GlcNAc*]: The ^1H resonances for residue **a** were traced readily via their cross-peaks in the COSY and relay COSY spectra. Magnetism was relayed very efficiently through the spin system and all cross-peaks were clearly visible. The ^{13}C resonances for residue **a** were assigned by comparing the ^1H assignments with the ^1H–^{13}C correlation data obtained from an HMQC[6] experiment.

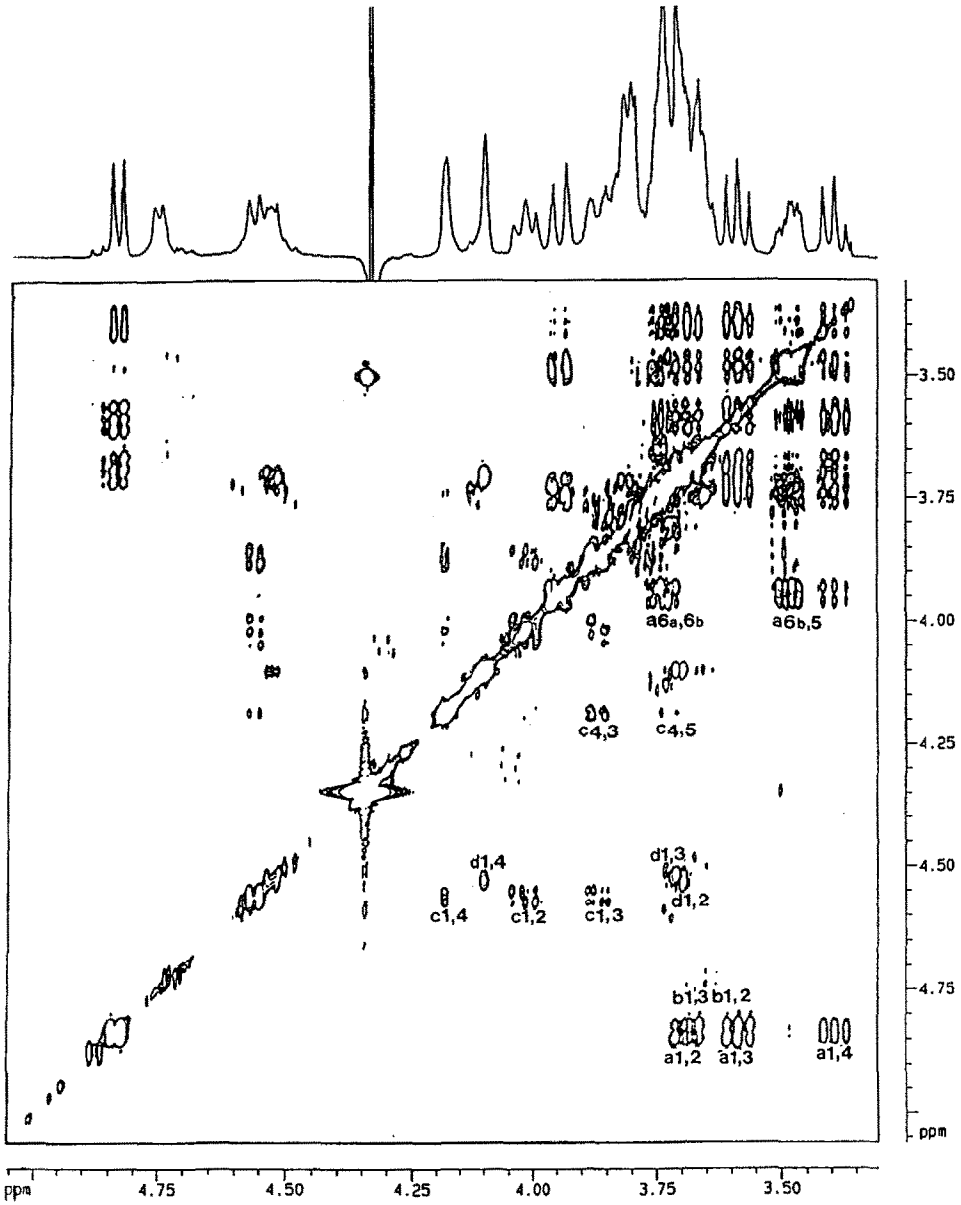


Fig. 1. Two-step RELAY COSY contour plot of **PS**: **a**1,2 connotes the cross-peak between H-1 and H-2 of residue **a**, etc. The ^{1}H NMR spectrum is projected along the f_2 axis. (See Table I for identification of **a**–**d**).

Residue **b** [→*3,4*]-*β*-D-*GlcA*]: The assignment of the ^{1}H resonances for this residue was complicated by the amount of signal overlap in the H-2 region of the COSY spectrum. The HOHAHA[5] spectrum showed that the signals for H-2, H-3, and H-5 were largely overlapped. As a result, it was not possible to distinguish

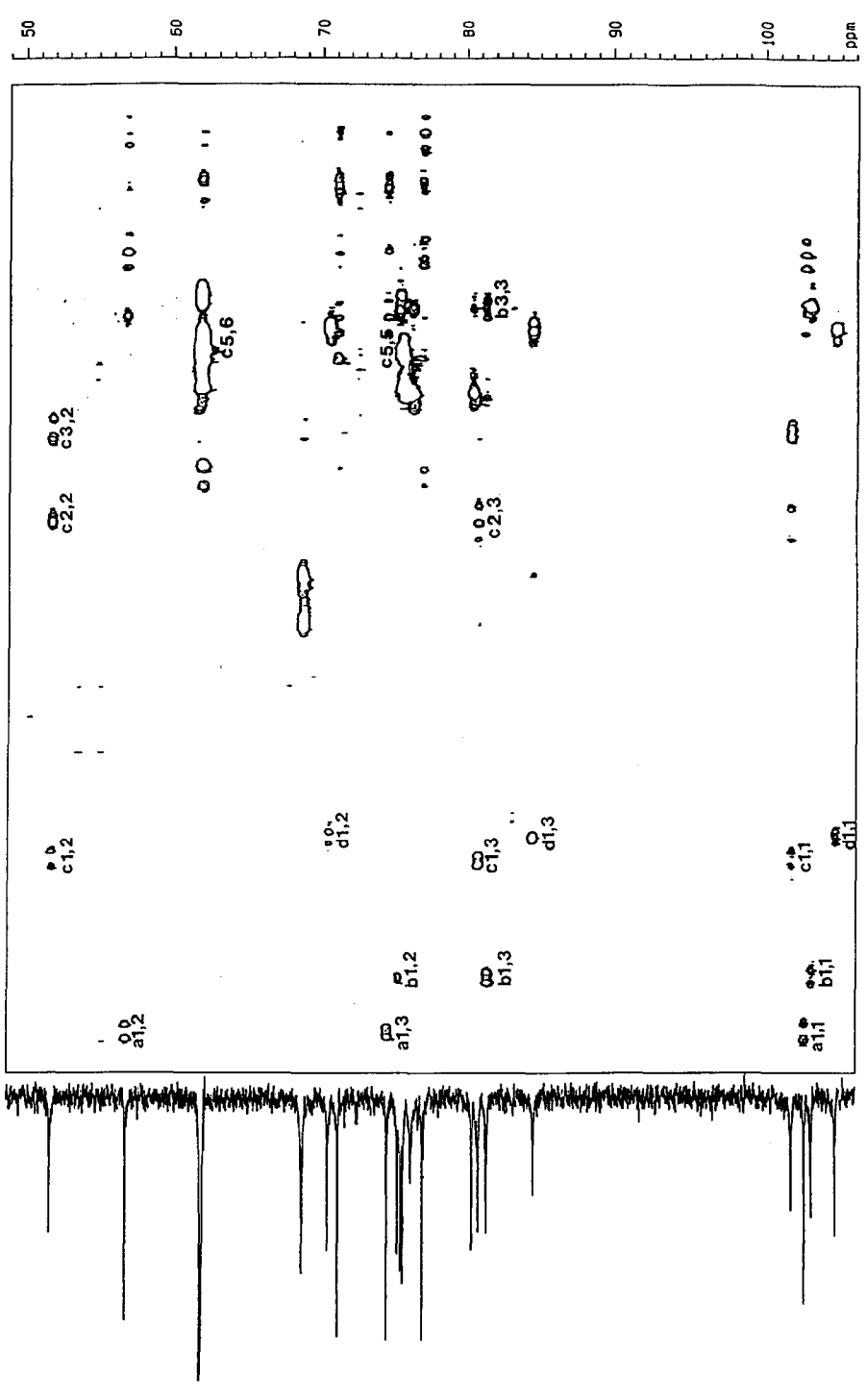


Fig. 2. HMQC-TOCSY contour plot of **PS**: **a**1,2 connotes the cross-peak between H-1 and C-2 of residue **a**, etc. The ^{13}C NMR spectrum is projected along the f_1 axis.

between the C-2 and C-3 resonances in the HMQC spectrum. Assignment of these carbon resonances was only possible using an HMQC-TOCSY[7] experiment, which clearly showed correlations between the H-1 and C-1, H-1 and C-2, and H-1 and C-3 resonances (Fig. 2).

Residue **c** [→ *3*)-β-D-*GalNAc*]: The ^{1}H resonances for H-1/5 of residue **c** were traced via their cross-peaks in the COSY and relay COSY spectra. The ^{13}C resonances for the corresponding C atoms were assigned by correlation from the HMQC spectrum. The C-6 signal was identified from the H-5–C-6 cross-peak in the HMQC-TOCSY spectrum (Fig. 2) and the H-6 signal was then confirmed from the $^{1}H-^{13}C$ correlation data.

Residue **d** [→ *3*)-β-D-*Gal*]: The ^{1}H resonances for H-1/4 were assigned from the COSY and relay COSY spectra. No further information could be obtained from the HOHAHA spectrum due to signal overlap and poor relay of magnetism in this spin system. The corresponding ^{13}C signals were assigned by comparison with the correlation data from the HMQC spectrum. The two remaining unassigned pairs of $^{1}H-^{13}C$ resonances from the $^{1}H-^{13}C$ correlation data could be assigned, by inspection, to H-5–C-5 and H-6–C-6 of residue **d**. The connectivity between these two pairs of resonances was confirmed from the HMQC-TOCSY spectrum.

Comparison of the chemical shift data for residues **a**–**d** with those reported for methyl glycosides[8–10] permitted identification of residue **a** as terminal β-GlcNAc, residue **b** as 3,4-linked β-GlcA, residue **c** as 3-linked β-GalNAc, and residue **d** as 3-linked β-Gal. These data are in agreement with the methylation results for **PS**.

The sequence of the residues **a**–**d** in the repeating unit was established by a heteronuclear multiple bond correlation (HMBC) experiment[11], which measures through-bond connectivity between C and H atoms two and three bonds distant. Correlations between H-1 of GlcNAc and C-3 of GlcA, between H-1 of GlcA and C-3 of Gal, between H-1 of GalNAc and C-4 of GlcA, and between H-1 of Gal and

TABLE I

NMR data[a] for *E. coli* K101 polysaccharide

Residue		Proton or carbon						
		1	2	3	4	5	6a	6b
β-D-GlcNAc (**a**)	H	4.85	3.68	3.59	3.39	3.48	3.74	3.94
	C	102.69	56.92	74.62	71.21	77.06	62.01	
→ 3,4)-β-D-GlcA (**b**)	H	4.75	3.66	3.67	3.83	3.72		
	C	103.17	75.36	81.35	80.47	75.58	172.07	
→ 3)-β-D-GalNAc (**c**)	H	4.57	4.02	3.88	4.18	3.73	3.82	3.82
	C	102.16	51.86	80.85	68.69	75.74	61.92	
→ 3)-β-D-Gal (**d**)	H	4.52	3.69	3.71	4.11	3.64	3.73	3.73
	C	104.77	70.49	84.59	68.69	75.58	61.88	

[a] Chemical shifts in ppm with acetone as internal standard, δ 2.23 and 31.07 ppm for ^{1}H and ^{13}C, respectively.

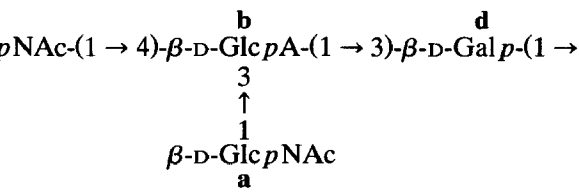
TABLE II

Two- and three-bond $^{1}H-^{13}C$ correlations for **PS**

Residue	Proton	Correlation to
β-D-GlcNAc (**a**)	H-1	56.92 (**a**, C-2), 81.35 (**b**, C-3)
	H-2	102.69 (**a**, C-1), 74.62 (**a**, C-3)
	H-3	56.92 (**a**, C-2), 71.21 (**a**, C-4)
	H-4	74.62 (**a**, C-2), 77.06 (**a**, C-5), 61.88 (**a**, C-6)
	H-6a,b	71.21 (**a**, C-4), 77.06 (**a**, C-5)
→ 3,4)-β-D-GlcA (**b**)	H-1	84.59 (**d**, C-3)
	H-2	103.17 (**b**, C-1), 81.35 (**b**, C-3), 75.58 (**b**, C-5)
	H-4	81.35 (**b**, C-3), 75.58 (**b**, C-5)
→ 3)-β-D-GalNAc (**c**)	H-1	80.47 (**b**, C-4)
	H-2	102.16 (**c**, C-1), 80.85 (**c**, C-3)
	H-3	51.86 (**c**, C-2)
	H-4	51.86 (**c**, C-2), 80.85 (**c**, C-3)
	H-5	68.69 (**c**, C-4), 61.88 (**c**, C-6)
→ 3)-β-D-Gal (**d**)	H-1	80.85 (**c**, C-3)
	H-2	104.77 (**d**, C-1), 84.59 (**d**, C-3)
	H-4	70.49 (**d**, C-2), 84.59 (**d**, C-3)
	H-5	68.69 (**d**, C-4), 61.88 (**d**, C-6)

C-3 of GalNAc were clearly visible (Table II). Other intraresidue connectivities served to confirm the assignments made for the residues as listed in Table I.

The combined chemical and NMR data permit the structure of the tetrasaccharide repeating unit of the K101 capsular polysaccharide to be written as:



This is the first *E. coli* capsular polysaccharide found to contain a terminal D-GlcNAc residue and joins a group of 13 other K-antigens whose constituent monosaccharides are all β-linked.

EXPERIMENTAL

General methods.—Analytical GLC was performed on a Hewlett–Packard 5890A gas chromatograph, fitted with flame-ionization detectors and a 3392A recording integrator, with He as carrier gas. A J&W Scientific fused-silica DB-17 bonded-phase capillary column (30 m × 0.25 mm; film thickness, 0.25 μm) was used for separating partially methylated alditol acetates (programme I), alditol acetates (programme II) and acetylated octyl glycosides (programme II). A J&W Scientific fused-silica DB-Wax bonded-phase capillary column (30 m × 0.25 mm; film thickness, 0.15 μm) was used for separating alditol acetates of ManNAc, GlcNAc, and

GalNAc (240°C isothermal). A J&W Scientific OV-225 bonded-phase capillary column (30 m × 0.25 mm; film thickness, 0.25 μm) was also used for separating acetylated octyl glycosides (240° isothermal). The temperature programmes used were: I, 180°C for 2 min, then 3°C min^{-1} to 240°C; II, 180°C for 2 min, then 2°C min^{-1} to 240°C. The identities of all derivatives were determined by comparison with authentic standards and confirmed by GLC–MS on a Hewlett–Packard 5988A instrument, using the appropriate column. Spectra were recorded at 70 eV and an ion-source temperature of 200°C.

GPC of the K101 polysaccharide was performed on a dextran-calibrated column (1.6 × 65 cm) of Sephacryl S500, using 0.1 M sodium acetate buffer (pH 5.00) as eluent.

Polysaccharide samples were hydrolysed with 4 M CF_3CO_2H for 1 h at 125°C. Alditol acetates were prepared by reduction of the products in aqueous solutions of hydrolysates with $NaBH_4$ for 1 h followed by acetylation with 2:1 Ac_2O–pyridine for 1 h at 100°C. Samples were methanolysed by refluxing with methanolic 3% HCl for 16 h. Native and methylated polysaccharides were carboxyl-reduced with $NaBH_4$ in dry MeOH after methanolysis. Methylations were carried out on the acid form of the polysaccharide, using potassium dimsyl[12] and MeI in Me_2SO.

Preparation of the K101 polysaccharide.—An authentic culture of *E. coli* O20:K101 was obtained from Dr. I. Ørskov (Copenhagen) and propagated on Mueller–Hinton agar. The capsular polysaccharide was extracted with aq 1% phenol, separated from the cells by ultracentrifugation, and purified by precipitation with cetyltrimethylammonium bromide.

NMR spectroscopy.—Samples were deuterium-exchanged by freeze-drying several times from D_2O and then examined as solutions in 99.99% D_2O containing a trace of acetone as internal standard (δ 2.23 for ^1H and 31.07 ppm for ^{13}C). Spectra were recorded at 67°C on a Bruker AMX-400 spectrometer equipped with an X32 computer. The parameters used for 2D experiments were as follows: COSY, two-step relay COSY, and HOHAHA [256 × 2048 data matrix, zero-filled to 1024 data points in t_1; 96 or 128 scans per t_1 value; spectral width 1362.4 Hz; reycle delay 1.0 s; fixed delays of 0.036 s (relay COSY); unshifted sine-bell filtering in t_1 and t_2 (COSY and relay COSY); and shifted sine-squared filtering in t_1 and t_2 (HOHAHA)]. HMQC and HMQC-TOCSY [512 × 4096 data matrix, zero-filled to 1024 data points in t_1; spectral width 14022 Hz in t_1 and 1362.4 Hz in t_2; 52 and 48 scans, respectively, per t_1 value; recycle delay 1.0 s; fixed delay 3.45 ms; shifted sine-squared filter]. HMBC [256 × 2048 data matrix, zero-filled to 1024 data points in t_1; spectral width 20733 Hz in t_1 and 1362.4 Hz in t_2; 96 scans per t_1 value; recycle delay 1.0 s; fixed delay 3.45 ms; shifted sine-squared filter].

ACKNOWLEDGMENTS

We thank Dr. I. Ørskov (Copenhagen) for the test strain of *E. coli* O20:K101, and the Foundation for Research Development (Pretoria) for a doctoral bursary (to M.R.G.) and for financial support (to H.P.).

REFERENCES

1 K. Jann and B. Jann, *Rev. Infect. Dis.*, (1987) s517–s526.
2 K. Leontein, B. Lindberg, and J. Lönngren, *Carbohydr. Res.*, 62 (1978) 359–362.
3 A. Bax and R. Freeman, *J. Magn. Reson.*, 46 (1982) 535–539.
4 A. Bax and G. Drobny, *J. Magn. Reson.*, 61 (1985) 306–320.
5 A. Bax and D.G. Davis, *J. Magn. Reson.*, 65 (1985) 355–360.
6 A. Bax and S. Subramanian, *J. Magn. Reson.*, 67 (1986) 565–569.
7 L. Lerner and A. Bax, *J. Magn. Reson.*, 69 (1986) 365–380.
8 K. Bock and H. Thörgersen, *Annu. Rep. NMR Spectrosc.*, 13 (1982) 1–57.
9 K. Izumi, *Carbohydr. Res.*, 170 (1987) 19–25.
10 P.A.J. Gorin and M. Mazurek, *Can. J. Chem.*, 53 (1975) 1212–1223.
11 A. Bax and M.F. Summers, *J. Am. Chem. Soc.*, 108 (1986) 2092–2094.
12 L.R. Phillips and B.A. Fraser, *Carbohydr. Res.*, 90 (1981) 149–152.