ELSEVIER

Carbohydrate Research 258 (1994) 223–232

CARBOHYDRATE RESEARCH

The structure of the cyclic enterobacterial common antigen (ECA) from *Yersinia pestis*

Evgeny V. Vinogradov [a,*], Yuriy A. Knirel [b], Jane E. Thomas-Oates [c], Alexander S. Shashkov [b], Vyacheslav L. L'vov [a]

[a] *Institute of Immunology, Ministry of Health, Kashirskoe Shosse 24a, Moscow 115471, Russian Federation*
[b] *N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospect 47, Moscow 117913, Russian Federation*
[c] *Department of Mass Spectrometry, Bijvoet Center for Biomolecular Research, Universiteit Utrecht, P.O. Box 8.00.83, 3508 TB Utrecht, The Netherlands*

(Received June 1st, 1993; accepted December 6th, 1993)

Abstract

Two antigenic acidic polysaccharides related to enterobacterial common antigen (ECA) were isolated from a vaccine strain of a pathogenic microorganism *Yersinia pestis*. The low molecular weight polysaccharide (LMP) is composed of equal amounts of 2-acetamido-2-deoxy-D-mannuronic acid, 4-acetamido-4,6-dideoxy-D-galactose (Fuc4NAc), and 2-amino-2-deoxy-D-glucose which is partially *N*- and partially 6-*O*-acetylated. The structure of the trisaccharide repeating unit was established by analyses of LMP and the completely *N*-acetylated LMP (LMP-NAc) using ^{1}H and ^{13}C NMR spectroscopy, including 2D COSY and 1D NOE spectroscopy. Deamination of LMP with nitrous acid gave a set of oligomers terminated with 2,5-anhydromannose and ranging from tri- to dodeca-saccharides, thus indicating a random distribution of free amino groups. FABMS analyses of LMP and LMP-NAc showed that LMP consists mainly of the cyclic tetramer of the trisaccharide repeating unit together with a small amount of the cyclic trimer and a very small amount of the cyclic pentamer and has, thus, the following structure:

$$\text{cyclo-}[\rightarrow 4)\text{-}\beta\text{-D-Man}p\text{NAcA-}(1\rightarrow 4)\text{-}\alpha\text{-D-Glc}p\text{NR-}(1\rightarrow 3)\text{-}\alpha\text{-D-Fuc}p4\text{NAc-}(1\rightarrow]_n$$
$$\begin{array}{c} 6 \\ \uparrow \\ \text{R}' \end{array}$$

where R is Ac or H (~ 1:1), R′ is Ac or H (~ 1:4), and $n = 4$ (major), 3, 5 (minor). Small proportions of the linear trimer and the linear tetramer were also detected in the prepara-

* Corresponding author at: M.M. Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, Moscow 117871, Russian Federation.

SSDI 0008-6215(93)E0437-6

tions. The high molecular weight polysaccharide is linear and has the same (or a very similar) repeating unit as LMP.

1. Introduction

An enterobacterium *Yersinia pestis*, the causative factor of plague, is know to produce an R-type lipopolysaccharide [1]. We report now the isolation of two other surface polysaccharide antigens of *Y. pestis* related to ECA [2] and the detailed structure of one of them.

2. Results and discussion

Killed cells of *Y. pestis* strain EV were extracted [3] with cold trichloroacetic acid. The extract, which is used as a vaccine against *Y. pestis*, was fractionated by gel-permeation chromatography on Sephadex G-50 to give a high molecular weight polysaccharide (HMP), eluted immediately after the void volume of the column, and a low molecular weight polysaccharide (LMP) with apparent molecular weight of ~ 3000. Both polymers were found to be acidic and were purified by anion-exchange chromatography on TSK DEAE 650M. They reacted with polyclonal antibodies raised against the parent strain in an inhibition of precipitation test and induced a specific immune response.

Judging from the ninhydrin test, LMP contains free amino groups. The ^{1}H and ^{13}C NMR spectra indicated that LMP has an irregular structure probably due to the presence of a partially *N*-acetylated amino sugar [δ_C 23.1–23.3 (Me), 175.2–176.7 (CO)] and a partially *O*-acetylated monosaccharide [δ_C 21.7 (Me), 175.2–176.7 (CO)].

N-Acetylation of LMP resulted in a more homogeneous product (LMP-NAc). Its ^{1}H NMR spectrum (Fig. 1A) contained the signals for three anomeric protons at 4.90 (br s), 4.96 (d, $J_{1,2}$ 3.5 Hz), and 5.13 ppm (d, $J_{1,2}$ 3.5 Hz), one CH_3–C of a 6-deoxy sugar at 1.04 ppm (d, $J_{5,6}$ 6.5 Hz), and three NAc groups at 2.01, 2.06, 2.06 ppm (all s), and other signals in the region 3.7–4.6 ppm. In the ^{13}C NMR spectrum of LMP-NAc (Fig. 2), there were signals for three anomeric carbons at 94.2, 99.9, and 102.4 ppm; three carbons bearing nitrogen at 50.7, 53.8, and 54.2 ppm; one CH_3–C group (C-6 of a 6-deoxy sugar) at 16.6 ppm; one $HOCH_2$C–C group (C-6 of a hexose derivative) at 61.2 ppm; a COOH group (C-6 of a hexuronic acid) at 173.0 ppm; nine other sugar carbons at 67.9–79.6 ppm; as well as three NAc groups (CH_3 at 23.1, 23.2, and 23.3 ppm; CO at 175.2, 175.5, and 176.6 ppm). A series of minor, 5–6-fold less-intense signals was also present in the spectrum, those at 21.7 and 64.1 ppm belonging, most likely, to the CH_3 of an *O*-acetyl group and C-6 of a 6-*O*-acetylated hexose derivative, respectively (the displacement of the signals for C-6 from 61.2 to 64.1 ppm corresponds [4] to the α-effect of acetylation at O-6).

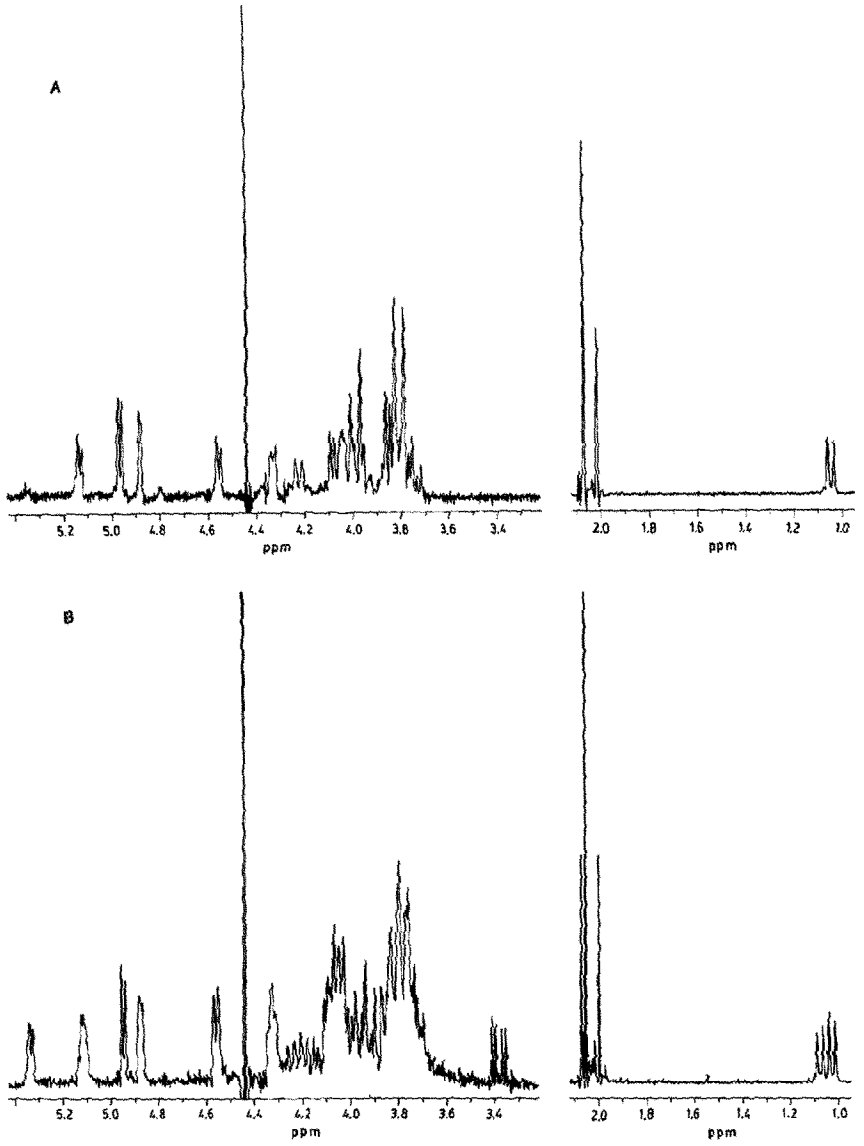

Fig. 1. 250-MHz ^{1}H NMR spectra of LMP-NAc (A) and LMP (B). The high-field region is given with 2-fold decreased intensity.

Completely *N*-acetylated LMP was also isolated, by HPLC on *reversed-phase* C_{18}, from the original LMP preparation where it is present *as a minor species*. It displayed a ^{1}H NMR spectrum almost identical to that of LMP-NAc, except that the minor signals due to *O*-acetylation were not detected.

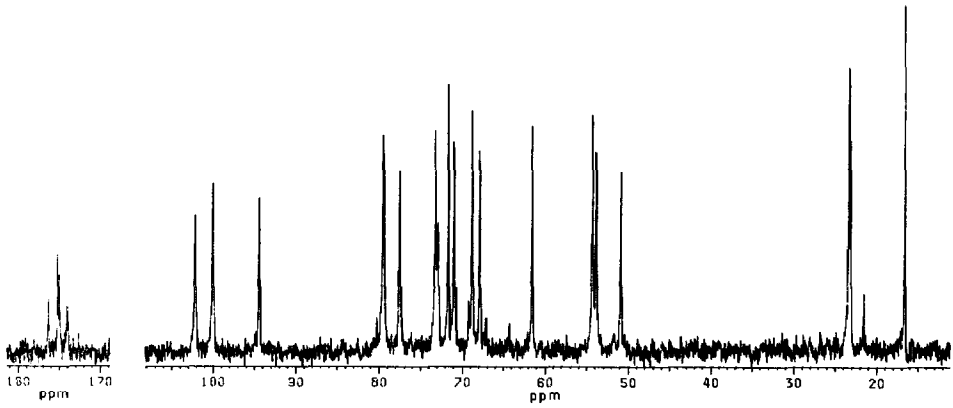

Fig. 2. 75-MHz ^{13}C NMR spectrum of LMP-NAc.

These data showed that LMP-NAc has a trisaccharide repeating unit containing three amino sugars, one of which is a 6-deoxy sugar, the second a hexuronic acid, and the third a hexose derivative; up to 20% of the hexose derivative is *O*-acetylated at position 6. The ^{1}H NMR spectrum of LMP-NAc was assigned using 1D sequential, selective spin decoupling, 2D homonuclear shift-correlated spectroscopy (COSY), and relayed coherence transfer spectroscopy (COSYRCT) (Table 1). The coupling constants $^{3}J_{H,H}$, determined from the spectrum, indicated that the uronic acid has the *β-manno* configuration (unit **A**), the hexose derivative has the *α-gluco* configuration (unit **B**), and the 6-deoxy sugar has the *α-galacto* configuration (unit **C**). On sequential pre-irradiation of H-1 of units **A**, **B**, and **C**, NOEs were observed on H-4 of unit **B**, H-3 and H-4 of unit **C**, and H-4 of unit **A**, respectively. These data revealed the sequence of the sugar residues and the substitution of units **A** and **B** at position 4. As for unit **C**, it is evidently substituted at position 3 since no NOE on H-3 would appear in the case of 4-substitution, whereas in α-(1 → 3)-linked disaccharides with the *galacto* configuration of the glycosylated pyranose, an NOE on H-4 is characteristic when the constituent monosaccharides have the same absolute configuration [5]. The ^{13}C NMR spectrum of LMP-NAc was assigned by using 2D heteronuclear $^{13}C/^{1}H$ COSY (Table 2). It was shown that the carbon atoms bearing nitrogen were C-4 of unit **C** and C-2 of units **A** and **B** (δ_C 50.7, 53.8, and 54.2, respectively). Hence, unit **A** is 2-acetamido-2-deoxymannuronic acid, unit **B** is 2-acetamido-2-deoxyglucose, and unit **C** is 4-acetamido-4,6-dideoxygalactose (Fuc4NAc). The D configuration of GlcNAc was determined by the method of Gerwig et al. [6]. The D configuration of Fuc4NAc followed from the NOE data (see above) and a very small glycosylation effect on C-1 of α-D-GlcNAc (+2.1 ppm) which is characteristic of the same absolute configuration of the sugar constituents in α-(1 → 3)-linked disaccharides with the glycosylated pyranose having the *galacto* configuration [7]. The relatively large β-effect of glycosylation on C-3 of GlcNAc (−1.4 ppm) indicated [7] that the glycosylating β-ManNAcA residue had the same D configuration as the glycosylated D-GlcNAc.

Table 1
^{1}H NMR data (δ in ppm, J in Hz)

	H-1 $J_{1,2}$	H-2 $J_{2,3}$	H-3 $J_{3,4}$	H-4 $J_{4,5}$	H-5 $J_{5,6}$	H-6
LMP-NAc						
β-D-ManNAcA (unit **A**)						
δ	4.90	4.57	4.08	3.84	4.07	
J	<1	4	9	9		
α-D-GlcNAc (unit **B**)						
δ	4.96	3.98	3.82	3.76	3.76	
J	3.5	10	10			
α-D-Fuc4NAc (unit **C**)						
δ	5.13	3.82	4.06	4.34	4.21	1.04
J	3.5	11	4	1.5	6	
LMP [a]						
β-D-ManNAcA (unit **A**)						
δ	4.90	4.58	4.10	3.86		
J	<1	4	9	9		
α-D-GlcN (unit **B**)						
δ	5.35	3.40	3.95	3.77		
J	3.5	10	9.5	9.5		
α-D-Fuc4NAc (unit **C**)						
δ	5.12	3.84	4.12	4.35	4.21	1.09
J	3.5	11	4	1.5	6	
Trisaccharide **3**						
α-D-Fuc4NAc						
δ	5.32	3.64	3.95	4.17	4.17	1.04
J	4	10.5	4		6	
β-D-ManNAcA						
δ	4.89	4.45	4.07	3.83	3.94	
J	<1	4	9	9		
D-2,5anhMan						
δ	4.98	3.73	4.34	4.12	4.02	3.67
J	6	4	4	5		

[a] The data for the trisaccharide unit containing GlcN with the free amino group.

On the basis of the data obtained, it was concluded that the repeating unit of LMP-NAc has the structure **1**. Analysis of the ^{1}H NMR spectrum of LMP (Fig. 1B, Table I), assigned as described above for LMP-NAc, revealed the presence of two main series of signals having nearly equal integral intensities, showing that LMP contains trisaccharide units of two types. One of them, as in LMP-NAc, has the structure **1**, whereas the other differs by the absence of one of the NAc groups, namely that of GlcNAc. In fact, the chemical shift of the signal for H-2 of unit **B** differs significantly in the two series (δ_H 3.98 for GlcN and 3.40 for GlcNAc), while those for H-2 of ManNAcA and H-4 of Fuc4NAc were similar; another

Table 2
^{13}C NMR chemical shifts (δ in ppm)

	C-1	C-2	C-3	C-4	C-5	C-6
β-D-ManNAcA (unit A)						
LMP-NAc	102.4	53.8	72.7 [a]	79.6	77.0	173.0
3	99.7	54.5	73.6	75.6	77.0	[b]
4 [c]	100.2	54.6	73.3	75.8	77.2	[b]
	99.7	54.5	73.6	75.4	77.1	[b]
α-D-GlcNAc (unit **B**) or D-2,5anhMan						
LMP-NAc	94.2	54.2	70.6	79.0	71.4	61.2
3	90.4	87.1	77.8	86.3	84.1	62.3
4 [c]	95.2	54.3	70.6	80.2	71.6	61.2
	90.4	87.1	77.8	86.3	84.1	62.3
α-D-Fuc4NAc (unit **C**)						
LMP-NAc	99.9	68.5	73.2 [a]	50.7	67.9	16.6
3	100.2	69.7	69.6	55.1	67.1	16.7
4 [c]	100.2	69.8	69.6	55.1	67.1	16.7
	100.3	68.3	74.3	51.2	67.1	16.7

[a] Assignment could be interchanged.
[b] Not found.
[c] The data for the sugar residue close to the nonreducing end are listed first.
Additional signals for LMP-NAc: NAc 23.1–23.3 ppm (Me), 175.2–176.6 (CO).

signal which is markedly shifted in the spectrum is that for H-1 of unit **B** (δ_H 5.35 for GlcN and 4.96 for GlcNAc). Therefore, the structure of the repeating unit of LMP may be depicted as **2**.

A **B** **C**

→ 4)-β-D-Man*p*NAcA-(1 → 4)-α-D-Glc*p*NR-(1 → 3)-α-D-Fuc*p*4NAc-(1 →

6
↑
R′

1 R is Ac, R′ is Ac or H (1:4).

2 R is Ac or H (1:1), R′ is Ac or H (1:4).

Deamination of LMP with nitrous acid led to a mixture of products which was separated by gel-permeation chromatography into three fractions. Two of them, eluted second and third, were further purified by anion-exchange chromatography and analysed in detail with the help of ^1H and ^{13}C NMR spectroscopy as described above for LMP-NAc (Tables 1 and 2). The lower molecular weight fraction was found to be a trisaccharide **3** which represents a modified repeating unit of LMP ('monomer') with 2,5-anhydromannose (2,5-anhMan) at the reducing end (this sugar is formed by deamination of GlcN [8]). The middle fraction is a hexasaccharide **4** ('dimer') composed of an unmodified trisaccharide repeating unit attached to trisaccharide **3**. The higher molecular weight fraction displayed complex ^1H and

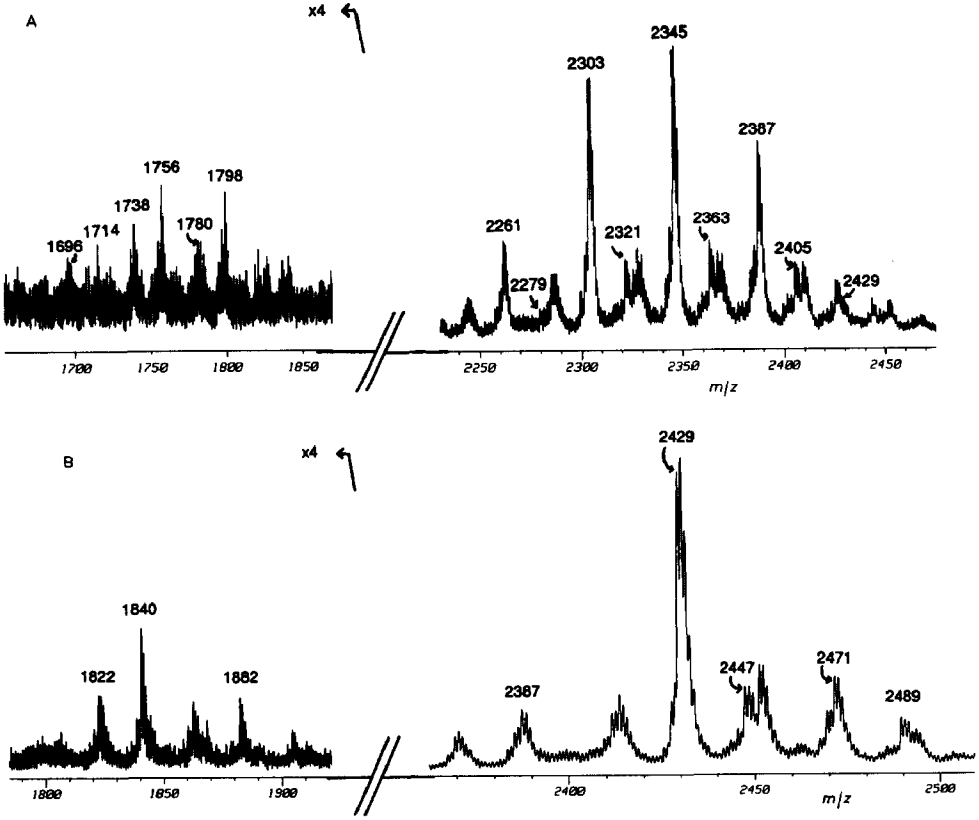

Fig. 3. Positive-ion FABMS of LMP (A) and LMP-NAc (B).

^{13}C NMR spectra and seems to be a mixture of linear 'trimer' **5**, 'tetramer' **6**, and unsplit LMP-NAc present in LMP.

$$\alpha\text{-D-Fuc}p4\text{NAc-(1-[}\rightarrow 4)\text{-}\beta\text{-D-Man}p\text{NAcA-(1}\rightarrow 4)\text{-}\alpha\text{-D-Glc}p\text{NR-(1}\rightarrow$$
$$3)\text{-}\alpha\text{-D-Fuc}p4\text{NAc-(1-]}_n\rightarrow 4)\text{-}\beta\text{-D-Man}p\text{NAcA-(1}\rightarrow 4)\text{-D-2,5-anhMan}$$

3 $n = 0$; **4** $n = 1$; **5** $n = 2$; **6** $n = 3$

The formation of tri- to dodeca-saccharides on deamination, i.e., all the theoretically possible products, indicated random distribution of free amino groups over the different GlcN residues in LMP.

Despite the low molecular masses of LMP and LMP-NAc, their ^{1}H and ^{13}C NMR spectra contained no signals for terminal sugar residues. This feature suggested a cyclic structure for these polysaccharides and prompted us to examine them by FABMS. Positive-ion FABMS analysis of original LMP revealed that the sample is heterogeneous. The spectrum (Fig. 3A) contained one minor and three major $[M + H]^+$ pseudomolecular ions at m/z 2261, 2303, 2345, and 2387, corresponding to the cyclic species which contain four ManNAcA–GlcN–Fuc4NAc

trisaccharide repeating units and which bear from 0 to 3 *N*- and/or *O*-acetyl groups. A very small trace of the species with four acetyl groups corresponding to the naturally occurring LMP-NAc tetramer was detected at *m/z* 2429. The less intense $[M + H]^+$ pseudomolecular ions observed at *m/z* 1696, 1738, and 1780 represent the corresponding cyclic species having three trisaccharide units and from 0 to 2 acetyl groups, while very minor ions were observed at *m/z* 2910 and 2952 (data not shown) for the analogous species containing five repeating units and 2 and 3 acetyl groups. Additionally there is evidence for small amounts of the linear analogues containing three (trimer) and four (tetramer) trisaccharide repeating units: the ions at *m/z* 1714, 1756, and 1798 correspond to the species with three units and from 0 to 2 acetyl groups, while the ions at *m/z* 2279, 2321, 2363, and 2405 arise from the linear species composed of four units and from 0 to 3 acetyl moleties.

The positive-ion FAB mass spectrum (Fig. 3B) of LMP-NAc contained one major $[M + H]^+$ pseudomolecular ion at *m/z* 2429 which corresponds to the fully *N*-acetylated cyclic tetramer of the trisaccharide repeating unit. Two additional, less intense ions were observed (*m/z* 2387 and 2471) corresponding to related species containing one fewer and one extra acetyl group, respectively. The former is probably the result of incomplete *N*-acetylation, while the latter probably corresponds to the mono-*O*-acetylated species. Again, a small amount of the cyclic trimer was observed (*m/z* 1822 for the fully *N*-acetylated species) and a very small amount of the fully *N*-acetylated pentamer was detectable at *m/z* 3036 (data not shown). The linear species were once again present: the ions at *m/z* 1840 and 1882 correspond to the fully *N*-acetylated and mono-*O*-acetylated linear trimers, while *m/z* 2447 and 2489 correspond to the fully *N*-acetylated and mono-*O*-acetylated linear tetramers, respectively.

On the basis of the data obtained, it was concluded that LMP is heterogeneous with respect to the number of *N*- and *O*-acetyl groups, the chain length, and the type of the structure (cyclic or linear). Most of the molecular species represent the cyclic tetramer of **2** (**7**, $n = 4$), in which the GlcN residues are partially *N*- and/or partially 6-*O*-acetylated, and the minor species correspond to the cyclic trimer (**7**, $n = 3$), the linear tetramer, and the linear trimer, together with a negligible amount of the cyclic pentamer (**7**, $n = 5$). Each of the species contains no more than one *O*-acetyl group per molecule and the GlcN residues may carry in total from 0 to 4 acetyl groups.

cyclo-[→4)-β-D-Man*p*NAcA-(1→4)-α-D-Glc*p*NR-(1→3)-α-D-Fuc*p*4NAc-(1→]$_n$
6
↑
R′

7 R and R′ = Ac or H, $n = 3$–5

The minor species, both cyclic and linear, revealed by FABMS are below the limits of NMR detection.

LMP from *Y. pestis* looks similar to the cyclic ECA from *Shigella sonnei* phase I [9] but is on average smaller in size (the cyclic tetramer predominates in *Y. pestis*

LMP) and more homogeneous with respect to the cycle size. Another difference between LMP and ECA from *S. sonnei* is the reduced degree of *O*-acetylation (20 compared to 70%) and *N*-acetylation of LMP. A cyclic ECA with four trisaccharide repeats has been found also in *Salmonella montevideo* [10].

It is noteworthy that the position of the signal of H-1 of Fuc4NAc (δ_H 5.13) differs significantly in the ^{1}H NMR spectrum of LMP-NAc as compared with those of **3** and **4** (δ_H 5.32). This is also the case for the signal of C-1 of ManNAcA, which resonates at 102.4 ppm in the cyclic LMP-NAc, but at 99.7 in **3** and **4** and 100.0 ppm in the linear ECA from *Plesiomonas shigelloides* [11]. These chemical shifts are thus indicative of the cyclic versus open-chain structure of ECA-type polysaccharides. Analysis of the ^{1}H and ^{13}C NMR spectra of HMP showed that it has the same, or a very similar, repeating unit to that of LMP and is linear. The NMR spectra also contained signals of different intensities which seem to belong to other unidentified sugar and lipid components. The presence of a lipid (phospholipid) moiety is characteristic of linear forms of ECA from different sources [2,11]. The role of ECA in the immunogenicity of the vaccine against *Y. pestis* is being studied.

3. Experimental

General methods.—^{1}H NMR spectra were recorded with a Bruker WM-250 spectrometer for solutions in D_2O at 70°C, and ^{13}C NMR spectra with a Bruker AM-300 instrument for solutions in D_2O at 70°C for polysaccharides and 30°C for oligosaccharides; acetone was used as an internal standard (δ_H 2.23, δ_C 31.45). Positive-ion FAB mass spectra were obtained using MS 1 of a Jeol JMX-SX/SX 102A tandem mass spectrometer operated at 6 kV accelerating voltage, as described [12].

GPC was performed on a column (70 × 3 cm) of Sephadex G-50 in a pyridine–acetate buffer (pH 5.5) or on a column (80 × 1.6 cm) of TSK HW-40 (S) in water, and monitored with a Knauer differential refractometer. Reversed-phase HPLC was performed on a DuPont Zorbax C_{18} column (25 × 0.9 cm), using a linear gradient of 0 → 20% MeCN in aq 0.05% CF_3CO_2H, and monitored with a Kratos Spectroflow UV monitor at 220 nm. Anion-exchange chromatography was performed on a column (20 × 2 cm) of TSK DEAE 650M, using a linear gradient of 0 → 0.3 M aq NaCl, and monitored as for HPLC.

Determination of the absolute configuration of GlcN was achieved using GLC analysis of acetylated (*S*)-2-butyl glycosides according to the published [6] and modified [12] method.

N-*Acetylation of LMP.*—LMP (20 mg) was dissolved in 2 mL of satd aq $NaHCO_3$ (2 mL), to which Ac_2O (50 μL) was added, and the solution was stirred for 20 min, deionized with Dowex 50W-X4 (H^+-form) resin, and freeze-dried to give LMP-NAc (18 mg).

Deamination of LMP.—LMP (70 mg) was dissolved in aq 15% AcOH (5 mL), aq 5% $NaNO_2$ (3 mL) was added, and the mixture was kept for 1 h at 20°C and then applied to a column of TSK HW 40(S) to afford trisaccharide **3**, hexasaccharide **4**,

and a mixture of higher oligosaccharides; **3** and **4** were further purified by anion-exchange chromatography.

References

[1] L. Kenne and B. Lindberg, in G.O. Aspinall (Ed.), *The Polysaccharides*, Vol. 2, Academic, New York, 1983, pp 287–363.
[2] H.-M. Kuhn, U. Meier-Dieter, and H. Mayer, *FEMS Microbiol. Rev.*, 54 (1988) 195–222.
[3] A.-M. Staub, *Methods Carbohydr. Chem.*, 5 (1965) 92–93.
[4] P.-E. Jansson, L. Kenne, and E. Schweda, *J. Chem. Soc., Perkin Trans. 1*, (1987) 377–383.
[5] G.M. Lipkind, A.S. Shashkov, S.S. Mamyan, and N.K. Kochetkov, *Carbohydr. Res.*, 181 (1988) 1–12.
[6] G.J. Gerwig, J.P. Kamerling, and J.F.G. Vliegenthart, *Carbohydr. Res.*, 77 (1979) 1–7.
[7] A.S. Shashkov, G.M. Lipkind, Y.A. Knirel, and N.K. Kochetkov, *Magn. Reson. Chem.*, 26 (1988) 735–747.
[8] J.M. Williams, *Adv. Carbohydr. Chem. Biochem.*, 31 (1975) 9–79.
[9] A. Dell, J. Oates, C. Lugowski, E. Romanowska, L. Kenne, and B. Lindberg, *Carbohydr. Res.*, 133 (1984) 95–104.
[10] A. Dell, personal communication, cited in H.-M. Kuhn, U. Meier-Dieter, and H. Mayer, *FEMS Microbiol. Rev.*, 54 (1988) 195–222.
[11] S. Basu, H.-M. Kuhn, A. Neszmelyi, K. Himmelspach, and H. Mayer, *Eur. J. Biochem.*, 162 (1987) 75–81.
[12] E.V. Vinogradov, O. Holst, J.E. Thomas-Oates, K. Broady, and H. Brade, *Eur. J. Biochem.*, 210 (1992) 491–498.