Carbohydrate Research, 211 (1991) 183–190
Elsevier Science Publishers B.V., Amsterdam

Note

Structural studies of the *Escherichia coli* O1A O-polysaccharide, using the computer program CASPER

Herbert Baumann, Per-Erik Jansson*, Lennart Kenne, and Göran Widmalm

Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm (Sweden)

(Received June 16th, 1990; accepted for publication, August 31st, 1990)

It is possible, by using a computerised approach, to deduce the structures of oligo- or poly-saccharides[1-5]. The computer program CASPER[1,3] needs results from sugar and methylation analysis together with the chemical shifts from unassigned signals in 1D- and/or 2D-n.m.r. spectra. The program generates all possible structures that are compatible with the results from the chemical analysis and simulates their ^{1}H- and/or ^{13}C-n.m.r. spectra. The spectra are then ranked according to their fit to the experimental spectrum. In order to decrease the number of possibilities, the structures with anomeric configurations that are not compatible with the experimental $J_{H\text{-}1,H\text{-}2}$ and $J_{C\text{-}1,H\text{-}1}$ values can be omitted[3]. For several linear and branched polysaccharides with known and unknown structures, exclusive answers have been obtained. If the computer analysis is not decisive, then a limited number of selected chemical degradations may be attempted or other n.m.r. experiments may be performed, *e.g.*, measurements of the n.O.e. for inter-residue protons.

Escherichia coli serotype O1 is divided into three different subtypes based on their immunological behaviour. This difference is also reflected in the structures of their lipopolysaccharides[6]. We have now investigated the structure of the O-antigen from serotype O1A, using sugar and methylation analysis, n.m.r. spectroscopy, and subsequent analysis with the CASPER program.

The O-polysaccharide (PS) was released from the lipopolysaccharide by treatment with aqueous 1% acetic acid. A hydrolysate of the PS contained L-rhamnose, 2-amino-2-deoxy-D-glucose, and 2-amino-2-deoxy-D-mannose. The sugars were detected as their alditol acetates by g.l.c.–m.s. and the absolute configurations were determined by g.l.c. of the trimethylsilylated (–)-2-butyl glycosides as devised by Gerwig *et al.*[7]. A methylation analysis of PS that involved standard conditions for the methylation by treatment with sodium methylsulphinylmethanide for 16 h gave only 2,4-di-*O*-methyl-L-rhamnose and 4-*O*-methyl-L-rhamnose in the ratio 2:1. A ^{1}H-n.m.r. spectrum of an alkali-treated PS contained additional signals between 5.3 and 6.0 p.p.m. that

---

* Author for correspondence.

indicated elimination reactions and explained the loss of amino sugars in the methylation analysis. When the exposure of PS to sodium methylsulphinylmethanide was shortened to 5 min, 2-deoxy-3,4,6-tri-*O*-methyl-2-methylamino-D-mannose and 2-deoxy-3,6-di-*O*-methyl-2-methylamino-D-glucose were obtained in a near molar ratio. A ^{1}H-n.m.r. spectrum of the native PS contained signals for five anomeric protons (Fig. 1), two NAc groups, and three CH*Me* groups. Thus, the PS had a repeating unit that contained one 2,3-linked L-rhamnose, two 3-linked L-rhamnose, one 4-linked 2-acetamido-2-deoxy-D-glucose, and one terminal 2-acetamido-2-deoxy-D-mannosyl group.

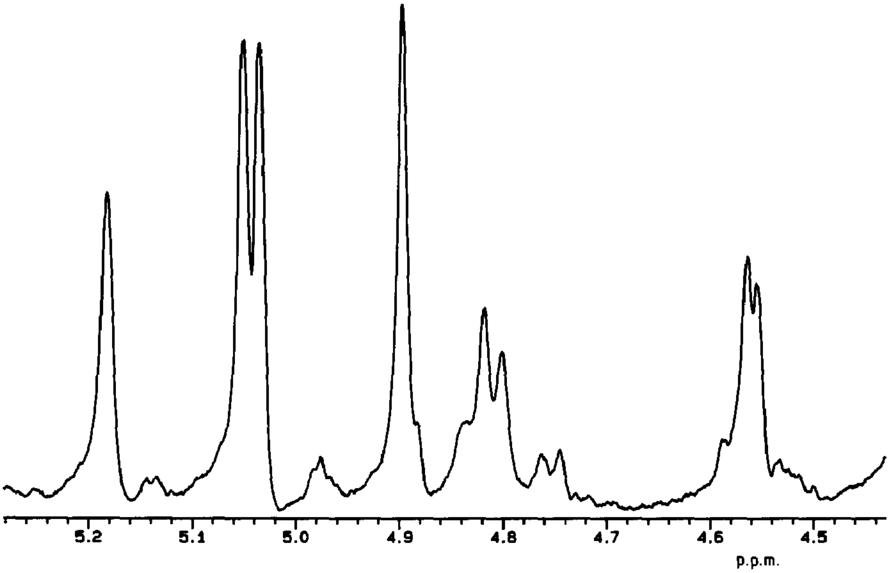
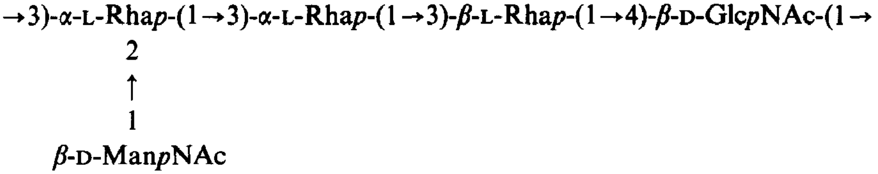


Fig. 1. The ^{1}H-n.m.r. spectrum of the *E. coli* O1A O-polysaccharide in the region for resonances of anomeric protons. Smaller extra signals probably originate from sugars residues in the core.

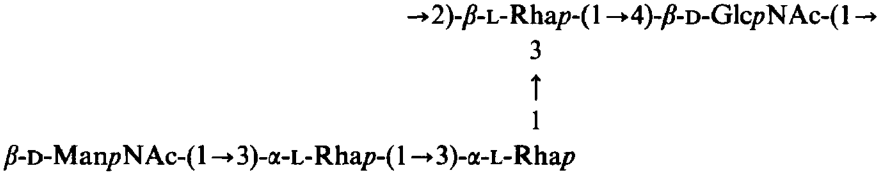
The ^{13}C-n.m.r. spectrum and the $J_{H-1,H-2}$ values were used for the computer analysis, the results of which are shown in Table I. All possible structures of the repeating unit containing the above sugar residues were formed by the CASPER program and the structures not compatible with the experimental $J_{H-1,H-2}$ values (four α- or *β-manno* and one *β-gluco*) were omitted. Since not all of the $J_{C-1,H-1}$ values could be obtained, due to overlapping signals, they were not used in the analysis. The ranking score is the difference in chemical shift between signals in the experimental and the simulated spectra derived from a peak-to-peak comparison. The structure with the lowest score is correct, as the evidence below shows. However, a few other spectra also have only minor deviations from the experimental spectrum. A CASPER run on these spectra (Table II), using the C,H-correlated spectrum, still gave structure **1** as the best fit, but altered the order of structures **2**–**4** compared to that of the first run. This finding strongly indicates that **1** is the correct structure.

1



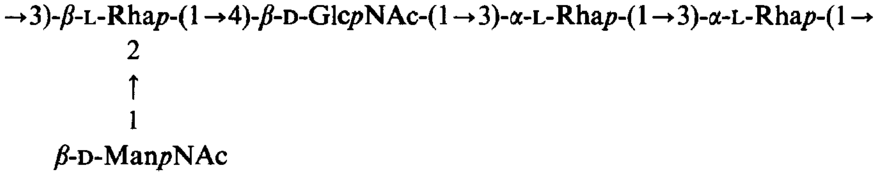
$$\begin{array}{cl}\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1 & \rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow \\ 2 & \\ \uparrow & \\ 1 & \\ \beta\text{-D-Man}p\text{NAc} & \end{array}$$

2



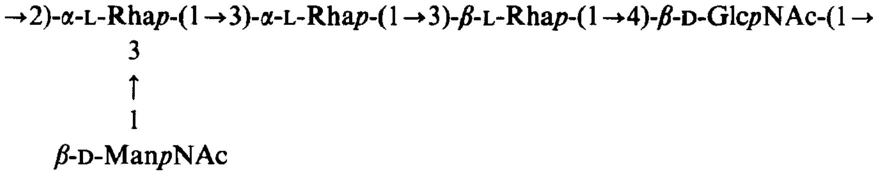
$$\begin{array}{rcl} & \rightarrow 2)\text{-}\beta\text{-L-Rha}p\text{-}(1 & \rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow \\ & 3 & \\ & \uparrow & \\ & 1 & \\ \beta\text{-D-Man}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-} & \alpha\text{-L-Rha}p & \end{array}$$

3



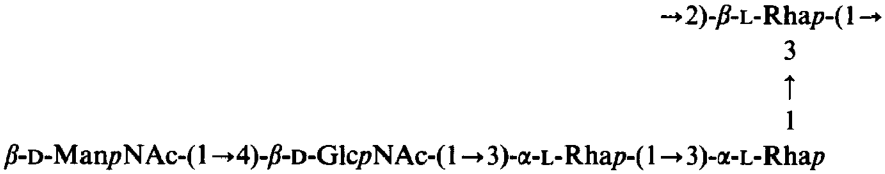
$$\begin{array}{cl}\rightarrow 3)\text{-}\beta\text{-L-Rha}p\text{-}(1 & \rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow \\ 2 & \\ \uparrow & \\ 1 & \\ \beta\text{-D-Man}p\text{NAc} & \end{array}$$

4



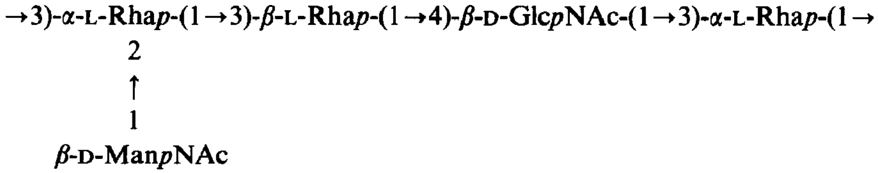
$$\begin{array}{cl}\rightarrow 2)\text{-}\alpha\text{-L-Rha}p\text{-}(1 & \rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow \\ 3 & \\ \uparrow & \\ 1 & \\ \beta\text{-D-Man}p\text{NAc} & \end{array}$$

5



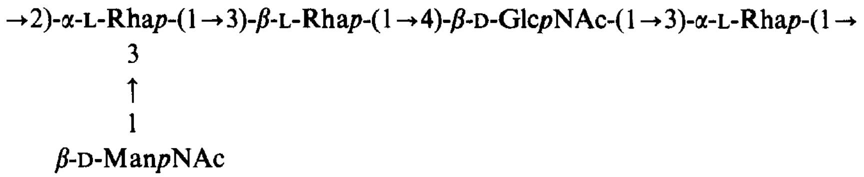
$$\begin{array}{rc} & \rightarrow 2)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow \\ & 3 \\ & \uparrow \\ & 1 \\ \beta\text{-D-Man}p\text{NAc-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-} & \alpha\text{-L-Rha}p \end{array}$$

6



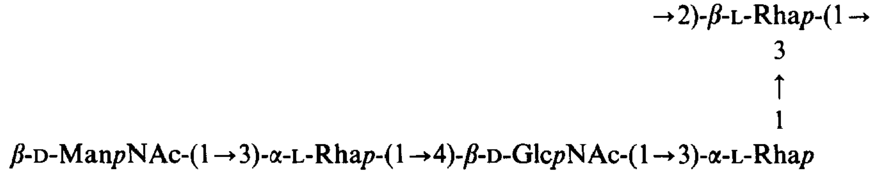
$$\begin{array}{cl}\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1 & \rightarrow 3)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow \\ 2 & \\ \uparrow & \\ 1 & \\ \beta\text{-D-Man}p\text{NAc} & \end{array}$$

7



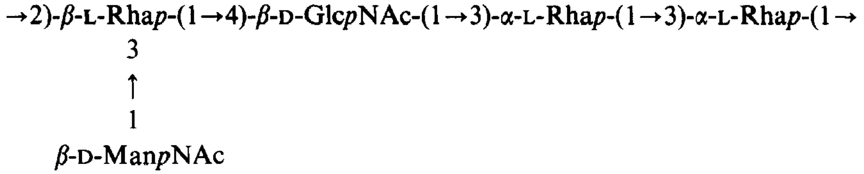
$$\begin{array}{l} \rightarrow 2)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow \\ \qquad\qquad\quad 3 \\ \qquad\qquad\quad \uparrow \\ \qquad\qquad\quad 1 \\ \qquad \beta\text{-D-Man}p\text{NAc} \end{array}$$

8



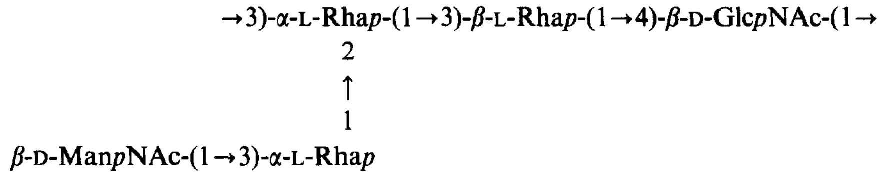
$$\begin{array}{r} \rightarrow 2)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow \\ 3\qquad\quad \\ \uparrow\qquad\quad \\ 1\qquad\quad \\ \beta\text{-D-Man}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p \end{array}$$

9



$$\begin{array}{l} \rightarrow 2)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow \\ \qquad\qquad\quad 3 \\ \qquad\qquad\quad \uparrow \\ \qquad\qquad\quad 1 \\ \qquad \beta\text{-D-Man}p\text{NAc} \end{array}$$

10



$$\begin{array}{l} \qquad\qquad\qquad \rightarrow 3)\text{-}\alpha\text{-L-Rha}p\text{-}(1\rightarrow 3)\text{-}\beta\text{-L-Rha}p\text{-}(1\rightarrow 4)\text{-}\beta\text{-D-Glc}p\text{NAc-}(1\rightarrow \\ \qquad\qquad\qquad\qquad\qquad 2 \\ \qquad\qquad\qquad\qquad\qquad \uparrow \\ \qquad\qquad\qquad\qquad\qquad 1 \\ \beta\text{-D-Man}p\text{NAc-}(1\rightarrow 3)\text{-}\alpha\text{-L-Rha}p \end{array}$$

In order to differentiate between the highest ranked structures, a Smith degradation of the PS was performed. In the methylation analysis of the product, the branched Rha*p* was replaced by 3-linked Rha*p*. This result leaves only structures **1**, **3**, and **6** for consideration; these have the same main chain, but with the 2-acetamido-2-deoxy-*β*-D-mannopyranosyl group linked to three different Rha*p* chain residues.

As an independent additional proof, a complete assignment of the n.m.r. chemical shift data *via* H,H- and C,H-COSY spectra was performed (Tables III and IV), together with establishment of inter-residue n.O.e. contacts. The information on sequence was gathered from 2D-n.O.e. spectra and n.O.e. difference spectra. The sets of chemical shifts were assigned to the respective sugar resonances by comparison with those of the appropriate *α* or *β* sugar[3] and relevant ^{1}H- and ^{13}C-n.m.r. glycosylation shifts for di- and tri-saccharides[8]. The H-1 and C-1 couplings indicated that the 2-acetamido-2-deoxy-D-glucose and 2-acetamido-2-deoxy-D-mannose residues were *β*, and that two of the three L-rhamnose residues were *α*, one of them being at the branch point.

TABLE I

Structures suggested by CASPER for the repeating unit of *E. coli* O1A O-polysaccharide, using ^{13}C-n.m.r. data

Structure	*^{13}C Chemical shift differences*	*^{13}C Diff./signal*
1	6.3	0.18
2	6.7	0.20
3	7.3	0.22
4	7.6	0.22
5	8.1	0.24
6	8.2	0.24
7	8.3	0.25
8	8.3	0.25
9	8.7	0.26
10	9.3	0.27

^{13}C Experimental spectrum

176.0 175.4 102.8 102.8 102.1 101.2 100.7 81.2 79.9 79.0
77.9 77.8 77.0 75.5 74.5 72.9 72.9 72.2 72.1 72.0
71.3 70.8 70.0 70.0 67.8 62.0 61.4 57.0 53.9 23.1
22.9 17.6 17.5 17.5

Spectrum No. 1

176.4 175.5 103.7 102.9 102.1 101.5 100.8 81.3 80.1 78.7
78.3 78.0 77.3 75.4 74.6 73.2 73.0 72.4 72.3 71.9
71.5 70.9 70.1 69.8 67.5 61.8 61.4 56.8 54.3 23.1
23.0 17.5 17.4 17.4

Spectrum No. 2

176.4 175.5 103.0 102.5 102.0 101.5 100.4 80.6 79.5 78.7
78.2 77.3 77.0 75.3 74.4 73.3 73.0 72.3 72.2 72.0
70.9 70.8 70.1 69.9 67.7 61.6 61.5 56.6 54.0 23.2
23.0 17.5 17.4 17.4

Assignments given by CASPER for spectrum No. 1

100.8 54.3 73.0 67.5 77.3 61.4 23.0 176.4
102.1 78.0 80.1 72.4 69.8 17.5
102.9 70.9 78.7 72.3 70.1 17.4
101.5 71.5 81.3 71.9 73.2 17.4
103.7 56.8 74.6 78.3 75.4 61.8 23.1 175.5

From the n.O.e. experiments, it was concluded that H-1 of the 2-acetamido-2-deoxy-β-D-glucose residue had an inter-residue n.O.e. contact to H-3 in the branch-point L-rhamnose residue. Also, there were contacts of H-1 of the terminal 2-acetamido-2-deoxy-β-D-mannosyl group to H-2 in the branch-point L-rhamnose residue, H-1 in the latter residue to H-3 in the 3-linked α-L-rhamnose residue, and its H-1 to H-3 in the 3-linked β-L-rhamnose residue. In the β-L-rhamnose residue, H-1 had an n.O.e. contact

TABLE II

Structures suggested by CASPER for the repeating unit of *E. coli* O1A O-polysaccharide, using C,H-COSY data

Structure	*C,H Score*
1	3.7
4	3.7
3	4.5
2	5.0

Experimental C,H-correlation spectrum

102.8 4.80 102.8 5.03 102.1 5.18 101.2 4.90 100.7 5.05
81.2 3.64 79.9 3.98 79.0 3.91 77.9 3.69 77.8 4.38
77.0 3.37 75.5 3.54 74.5 3.74 72.9 3.81 72.9 3.43
72.2 3.56 72.1 3.50 72.0 3.49 71.3 4.16 70.8 4.13
70.0 3.80 70.0 3.88 67.8 3.52 62.0 3.98 62.0 3.84
61.4 3.88 61.4 3.81 57.0 3.74 53.9 4.56 23.1 2.06
22.9 2.08 17.6 1.34 17.5 1.30 17.5 1.26

Spectrum No. 1

103.7 4.73 102.9 5.02 102.1 5.30 101.5 4.89 100.8 5.20
81.3 3.64 80.1 4.04 78.7 3.89 78.3 3.70 78.0 4.35
77.3 3.44 75.4 3.50 74.6 3.74 73.0 3.85 73.2 3.39
72.4 3.70 72.3 3.57 71.9 3.46 71.5 4.04 70.9 4.14
69.8 3.80 70.1 3.84 67.5 3.52 61.8 3.90 61.8 3.88
61.4 3.90 61.4 3.81 56.8 3.80 54.3 4.55 23.1 2.06
23.0 2.06 17.5 1.40 17.4 1.31 17.4 1.30

TABLE III

Chemical shifts in the ^{1}H-n.m.r. spectrum of the *E. coli* O1A O-polysaccharide

Sugar residue	*Chemical shift*[a] *(δ)*							
	H-1	*H-2*	*H-3*	*H-4*	*H-5*	*H-6a*	*H-6b*	*CH$_3$CO*
→3)-α-L-Rha*p*-(1→ 2 ↑	5.18 (n.r.)	4.38	3.98	3.50	3.80	1.26		
β-D-Man*p*NAc-(1→	5.05 (n.r.)	4.56	3.81	3.52	3.37	3.81	3.88	2.08[b]
→3)-α-L-Rha*p*-(1→	5.03 (n.r.)	4.13	3.91	3.56	3.88	1.30		
→3)-β-L-Rha*p*-(1→	4.90 (n.r.)	4.16	3.64	3.49	3.43	1.34		
→4)-β-D-Glc*p*NAc-(1→	4.80 (7.3)	3.74	~3.74	3.69	3.54	3.84	3.98	2.06[b]

[a] *J* values (Hz) in parentheses; n.r. = not resolved. [b] Assignments may be reversed.

TABLE IV

Chemical shifts in the ^{13}C-n.m.r. spectrum of the *E. coli* O1A O-polysaccharide

Sugar residue	*Chemical shift*^{a} (δ)							
	C-1	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*CH$_3$CO*	*CH$_3$CO*
→3)-α-L-Rha*p*-(1→	102.05 (174)	77.81	79.94	72.13^{b}	70.02	17.49^{b}		
2 ↑	[−]	[0.2]	[0.2]	[0.3]	[−0.2]	[−]		
β-D-Man*p*NAc-(1→	100.67 (163)	53.92	72.92^{b}	67.78	77.02	61.43	22.89^{b}	176.02^{b}
	[0.1]	[0.4]	[0.1]	[−0.3]	[0.3]	[−]	[0.1]	[0.4]
→3)-α-L-Rha*p*-(1→	102.78	70.75	79.02	~72.24	70.02	17.49^{b}		
	[0.1]	[0.1]	[−0.3]	[0.1]	[0.1]	[−0.1]		
→3)-β-L-Rha*p*-(1→	101.24 (160)	71.32	81.16	72.00^{b}	72.92^{b}	17.57^{b}		
	[0.3]	[0.2]	[0.1]	[−0.1]	[0.3]	[−0.2]		
→4)-β-D-Glc*p*NAc-(1→	102.78	57.05	74.48	77.94	75.46	61.97	23.11^{b}	175.37^{b}
	[0.9]	[−0.2]	[0.1]	[0.2]	[−0.1]	[−0.2]	[−]	[0.1]

^{a} $J_{\text{C-1,H-1}}$ values in Hz in parentheses, and chemical shift differences between experimental and calculated spectrum in brackets. ^{b} Tentative assignments.

to a proton giving a signal at δ 3.70. Signals from H-3 and H-4 in the 2-acetamido-2-deoxy-β-D-glucose residue each had this chemical shift but, as the amino sugar residue is 4-linked, the other possibility is excluded. The results give further evidence for structure **1**, which is the structure proposed by CASPER.

It is evident also that, with a limited amount of n.m.r. information, the list of possible alternatives can be significantly shortened. For instance, as the branch-point residue was demonstrated from ^{1}H- and ^{13}C-n.m.r. chemical shifts to be α, structures **2**, **3**, **5**, **8**, and **9** are eliminated.

Furthermore, in the structural elucidation of the structure of the *E.coli* O1A O-polysaccharide, the combined use of ^{13}C- and H,C-correlation n.m.r. spectra in the computer analysis with CASPER further indicates that **1** is the correct structure. The only major deviation ($>$0.4 p.p.m.) between the calculated and experimental chemical shifts is observed for the C-1 signal of the 4-linked 2-acetamido-2-deoxy-β-D-glucose residue. This is the signal that is influenced most by the interactions at the sterically crowded branch-point residue.

It is concluded also that ^{13}C-n.m.r. glycosylation shifts are reliable for the determination of linkage positions. However, the ^{1}H-n.m.r. glycosylation shifts are sometimes smaller at linkage positions than for neighbouring positions as, for example, with the 3-linked L-rhamnose residues, and appropriate reference substances are required.

EXPERIMENTAL

Sugar and methylation analyses were performed as described[9,10], but the treatment with base was for only 5 min before the addition of methyl iodide. Treatment of the polysaccharide with 0.25M sodium hydroxide at 70° for 16 h gave a degraded polysaccharide that was not investigated further. N.m.r. spectra of solutions in D_2O at 70° were recorded at 270 and 400 MHz (^1H, internal trimethylsilylpropanoate-d_4) and at 67.5 and 100 MHz (^{13}C, internal 1,4-dioxane, δ 67.4). The computer program CASPER was used as described[3]. The Smith degradation of the polysaccharide was performed essentially as described by Goldstein *et al.*[11].

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Natural Science Research Council, the Swedish Medical Research Council (03X-2522), and the Swedish National Board for Technical Development. Dr. K. Leontein (National Bacteriological Laboratory) is thanked for a gift of the *E. coli* O1A polysaccharide.

REFERENCES

1 P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 168 (1987) 67–77.
2 G. M. Lipkind, A. S. Shashkov, Y. A. Knirel, E. V. Vinogradov, and N. K. Kochetkov, *Carbohydr. Res.*, 175 (1988) 59–75.
3 P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 188 (1989) 169–191.
4 P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 193 (1989) 322–325.
5 G. M. Lipkind, A. S. Shashkov, and N. K. Kochetkov, *Carbohydr. Res.*, 198 (1990) 399–402.
6 A. Moll, B. Kusecec, G. Pluschke, G. Morelli, M. Kamke, B. Jann, K. Jann, and M. Achtman, *Infect. Immun.*, 53 (1986) 257–263.
7 G. J. Gerwig, J. P. Kamerling, and J. F. G. Vliegenthart, *Carbohydr. Res.*, 77 (1979) 1–7.
8 P.-E. Jansson and L. Kenne, *J. Chem. Soc., Perkin Trans. 1*, (1990) 2011–2018, and references therein.
9 T. A. Chowdhury, P.-E. Jansson, B. Lindberg, and U. Lindquist, *Carbohydr. Res.*, 190 (1989) 145–151.
10 T. J. Waeghe, A. G. Darvill, M. McNeil, and P. Albersheim, *Carbohydr. Res.*, 123 (1983) 281–304.
11 I. J. Goldstein, G. W. Hay, B. A. Lewis, and F. Smith, *Methods Carbohydr. Chem.*, 5 (1965) 361–370.