Carbohydrate Research, 175 (1988) 59–75
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

A COMPUTER-ASSISTED STRUCTURAL ANALYSIS OF REGULAR POLYSACCHARIDES ON THE BASIS OF ^{13}C-N.M.R. DATA

GRIGORY M. LIPKIND, ALEXANDER S. SHASHKOV, YURIY A. KNIREL*, EVGENY V. VINOGRADOV, AND NIKOLAY K. KOCHETKOV

N. D. Zelinsky Institute of Organic Chemistry, Academy of Sciences of the U.S.S.R., Moscow (U.S.S.R.)

(Received March 2nd, 1987; accepted for publication, August 26th, 1987)

ABSTRACT

A computerised approach to the structural analysis of unbranched regular polysaccharides is described, which is based on an evaluation of the ^{13}C-n.m.r. spectra for all possible primary structures within the additive scheme starting from the chemical shifts of the ^{13}C resonances of the constituent monosaccharides and the average values of the glycosylation effects. The analysis reveals a structure (or structures), the evaluated spectrum of which resembles most closely that observed. The approach has been verified by using a series of bacterial polysaccharides of known structure and, in combination with methylation analysis data, for the determination of the presently unknown structures of the O-specific polysaccharides from *Salmonella arizonae* O59 and O63, and *Proteus hauseri* O19.

INTRODUCTION

The formation of a glycosidic linkage leads to changes in the ^{13}C chemical shifts which depend on the type of linkage[1] and the stereochemistry of the monosaccharide residues connected[2]. The most significant changes are associated with the carbons directly involved in the glycosidic linkage and the adjacent carbons (α- and β-effects of glycosylation). Thus, oligo- and poly-saccharides of identical monosaccharide composition but different primary structures have different ^{13}C-n.m.r. spectra.

The glycosylation effects in the disaccharide fragments of unbranched oligosaccharides and polysaccharides are almost identical to those in the corresponding disaccharides, and, hence, they are transferable parameters. This similarity enables the evaluation, within the additive scheme, of the ^{13}C-n.m.r. spectra of carbohydrates of known structure starting from the chemical shift data for the constituent monosaccharides and the average values of the glycosylation effects, and, conversely, to establish the primary structure of carbohydrates on the basis of ^{13}C-

* Author for correspondence

0008-6215/88/$03.50

n.m.r. data. The possiblities of this system have been exemplified with some regular unbranched polysaccharides built up of simple hexoses[3]. We now report in detail the computerised approach to the structural analysis of carbohydrates and its application to unbranched bacterial polysaccharides of unknown structure built up of trisaccharide or tetrasaccharide repeating-units, containing variously Glc*p*, Man*p*, Gal*p*, Rha*p*, Glc*p*NAc, Gal*p*NAc, Man*p*NAc, and Fuc*p*NAc. These monosaccharides are widely distributed in natural carbohydrates and ^{13}C-n.m.r. data are available in the literature for a variety of oligosaccharides and polysaccharides containing them, thus allowing the glycosylation effects to be determined with confidence.

RESULTS AND DISCUSSION

General approach. — The proposed computer-assisted approach to the structural determination of regular polysaccharides involves three steps: (*1*) the generation of all possible unbranched structures of a polysaccharide of a given monosaccharide composition; (*2*) the evaluation of the ^{13}C-n.m.r. spectra for each of these structures starting from the chemical shift data for the constituent monosaccharides and the average values of the glycosylation effects; (*3*) searching for the structure (or structures) characterised by the calculated spectrum most similar to that observed for the polysaccharide.

In generating the possible primary structures of polysaccharides, account is taken of all permutations of the constituent monosaccharides in the repeating unit

TABLE I

^{13}C CHEMICAL SHIFT DATA FOR MONOSACCHARIDES [δ_o (*1*)] AND GLYCOSYLATION EFFECTS

Monosaccharide		δ_o *(1) (p.p.m.)*						B[a] *(p.p.m.)*	
		C-1	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-2'*	*C-5'*
Glc*p*	α	93.3	72.7	74.0	70.9	72.7	61.9	0.0	+0.5
	β	97.1	75.4	77.0	70.9	77.2	62.1	−1.0	0.0
Man*p*	α	95.3	72.0	71.5	68.2	73.7	62.3	−0.5	+0.5
	β	94.9	72.5	74.3	67.9	77.4	62.3	−1.0	0.0
Gal*p*	α	93.5	69.6	70.4	70.6	71.7	62.4	0.0	+0.5
	β	97.7	73.2	74.1	70.0	76.3	62.2	−1.5	0.0
Rha*p*	α	95.2	72.1	71.3	73.5	69.5	18.0	−0.5	+0.5
	β	94.7	72.6	74.0	73.1	73.2	18.0	−1.0	0.0
Glc*p*NAc	α	92.1	55.3	72.0	71.4	72.8	61.9	0.0	+0.5
	β	96.2	58.0	75.1	71.2	77.2	62.1	−1.0	0.0
Man*p*NAc	α	94.3	54.5	70.1	68.2	73.3	61.8	−0.5	+0.5
	β	94.2	55.3	73.3	67.9	77.6	61.8	−1.0	0.0
Gal*p*NAc	α	92.2	51.5	68.7	69.9	71.7	62.4	−0.5	+0.5
	β	96.6	55.0	72.4	69.1	76.4	62.2	−1.5	0.0
Fuc*p*NAc	α	92.1	51.2	68.7	72.4	67.5	16.7	−0.5	+0.5
	β	96.3	54.7	72.4	72.0	71.8	16.7	−1.0	0.0

[a]The effects for C-3',4',6' are taken as zero; the effects for C-1', which depend on the type of glycosidic linkage, are given in Table II.

(*i.e.* 2 and 6 for a trisaccharide and a tetrasaccharide, respectively), as well as the 8 types of linkage possible for each disaccharide fragment. The number of structures to be considered, which for a tetrasaccharide repeating-unit is 24,576, may be diminished by adopting some restrictions based on a preliminary analysis of the ^{13}C-n.m.r. spectra. This may be the presence or absence of (1→6) linkages, which can be inferred from the presence of the signals for C-6 at 60–63 p.p.m. for the unsubstituted, and at 65–71 p.p.m. for 6-substituted pyranose residues, as well as the number of α and β linkages, which can be established[4] easily on the basis of the $^{1}J_{C\text{-}1,H\text{-}1}$ values.

The ^{13}C chemical shift data for the monosaccharides were obtained for solutions in D_2O, using methanol as the internal standard. These data are given in Table I.

The effects of glycosylation were determined as the difference in the ^{13}C chemical shifts of the resonances for a disaccharide fragment P′ → P (P and P′ are the glycosylated and the glycosylating pyranose residues, respectively) and in the corresponding pyranoid forms of the free monosaccharides. The ^{13}C chemical shift data for various disaccharide fragments were taken mainly from published work[5,6], as well as from many other publications containing ^{13}C-n.m.r. data for oligosaccharides and polysaccharides. When different sources reported different data for a disaccharide fragment, the glycosylation effects were determined as average values, and, as a rule, deviations did not exceed 0.5 p.p.m.

The effects of glycosylation depend on the position and configuration of the glycosidic linkage, the general configuration of the residue P, and the relative absolute configuration of the residues P and P′. For (1→2) linkages, the glycosylation effects depend essentially on the anomeric configuration of the residue P, and, hence, they were determined separately for α and β anomers. For the other types of linkages, the average values of the effects were used for the two anomers unless otherwise stated in Table II. In addition, account was taken of the difference in the glycosylation effects, which depend on the residue P′ being Hex or HexNAc and whether it has the *manno* or *gluco (galacto)* configuration. When no ^{13}C-n.m.r. data for a disaccharide fragment were available in the literature, approximations were made based on general regularities in glycosylation effects[2].

The values of the glycosylation effects used for the calculations are listed in Tables I and II for each monosaccharide and each absolute configuration of the residues P and P′. In Table II, the columns correspond to the carbons and the lines correspond to the types of linkage between P and P′, which are denoted by the index k in the range 1–10. The matrix A and column B comprise the glycosylation effects for the residue P and C-1′ of the residue P′, respectively. The glycosylation effects for C-2′,3′,4′,5′ of the residue P′ do not depend on the types of glycosidic linkage and are listed in Table I. The values of the effects for C-3′,4′,6′ of the residue P′ are taken as zero.

The ^{13}C chemical shift data for each monosaccharide residue in the polysaccharide having a definite anomeric configuration and definite adjacent sugar

TABLE II

EFFECTS OF GLYCOSYLATION ON THE CHEMICAL SHIFTS OF THE ^{13}C RESONANCES FOR RESIDUE P AND C-1′ OF RESIDUE P′ OF THE DISACCHARIDE FRAGMENT P′→P, DEPENDING ON THE TYPE OF GLYCOSIDIC LINKAGE AND THE RELATIVE ABSOLUTE CONFIGURATION OF RESIDUES P AND P′ [D-D(L-L) OR D-L (L-D)][a]

[a]When P′ has the *manno* configuration or represents HexpNAc, the effects are given in parentheses or in square brackets, respectively, if they vary considerably from the average values. In the linkage type designations, the first Greek symbol means the anomeric configuration of P′, and the last one means the configuration of P. The superscripts α and β designate the anomeric configuration of P. The effects obtained by approximation are given in italic type.

(a) *P has the* gluco *configuration*

Linkage type	D-D(L-L)							D-L(L-D)						
	A						B	A						B
k	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-1′*	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-1′*
	Glucose													
α-(1→2)-α	−1.8	+4.0	−1.3	0	0	0	+3.4	*−0.5*	*+8.6*	*−0.9*	*0*	*0*	*0*	*+8.3*
α-(1→2)-β	+0.3	+4.7	−1.3	0	0	0	+5.6	*−1.3*	*+5.2* (+6.8)	*+0.7*	*0*	*0*	*0*	*+7.1*
α-(1→3)	0	−1.2	+6.8	−0.1 (+0.9)	0	0	+6.7	0	−1.2	+4.9 (+6.9)	−1.3	0	0	+7.1
α-(1→4)	0	0	+0.3	+7.4	−1.3	0	+7.6 (+7.1)	0	0	−1.3	+7.5 (+8.0)	−0.6	−0.6	+6.4
α-(1→6)	0	0	0	0	−1.8	+5.0	+5.6	0	0	0	0	−1.2	+6.5	+6.7
β-(1→2)-α	−0.7	+9.0	−1.3	0	0	0	+7.6	*−1.8*	*+6.0*	*−1.3*	*0*	*0*	*0*	*+4.0*
β-(1→2)-β	−1.6	+7.5	0	0	0	0	+7.0	*+0.3*	*+7.5*	*−1.3*	*0*	*0*	*0*	*+7.0*
β-(1→3)	0	−0.7	+9.5	−1.5	0	0	+6.7	*0*	*−1.2*	*+9.5*	*−0.1* (+0.9)	*0*	*0*	*+6.7*
β-(1→4)	0	0	−1.4	+9.0	−1.3	−0.8	+6.3	0	0	+0.1	+6.7	−1.4	0	+7.2
β-(1→6)	0	0	0	0	−1.0	+7.8 [+6.0]	+6.7	0	0	0	0	−1.0	+7.0	+6.1

	2-Acetamido-2-deoxyglucose[a]													
α-(1→3)								0	−0.2	+4.9 (+7.7)	−1.3	0	0	+7.0
β-(1→3)	0	−1.2	+9.6[α] +8.7[β] [+8.5][α] [+7.5][β]	−1.2	0	0	+6.7							

[a]The same effects as for glucose were used for other types of linkage.

(b) *P has the* manno *configuration*

Linkage type	D-D(L-L)							D-L(L-D)						
	A						B	A						B
k	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-1'*	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-1'*
	Mannose and rhamnose													
α-(1→2)-α	−2.5 (−1.4)	+9.7 (+8.2)	0	0	0	0	+9.2 (+8.1)	−2.5	+6.0 (+4.7)	0	0	0	0	+5.4 (+4.2)
α-(1→2)-β	−0.2	+7.1	+0.6	0	0	0	+7.4	*0*	*+7.1*	*0*	*0*	*0*	*0*	*+7.4*
α-(1→3)	*0*	−0.2	+7.5	−0.6	0	0	+8.3 [+7.8]	0	−3.3	+5.5 (+4.2)	−1.7	0	0	+3.4 (+2.4)
α-(1→4)	0	0	+0.2	+8.2 (+7.3)	−1.1	0	+7.7 (+7.0)	0	0	−1.3	+9.0	−1.0	0	+7.3
α-(1→6)	0	0	0	0	−1.8	+5.4 (+4.5)	+6.2 (+5.5)	*0*	*0*	*0*	*0*	*−1.2*	*+6.5*	*+6.7*
β-(1→2)-α	−2.2 [−3.5]	+7.4 [+5.6]	−0.1	0	0	0	+5.7 [+4.2]	−1.2	+8.4	−0.3	0	0	0	+8.4 [+7.3]
β-(1→2)-β	−0.1	+9.3	0	0	0	0	+7.7	−0.8	+9.8	+0.2	0	0	0	+7.4
β-(1→3)	0	−2.7	+7.1 [+7.9]	−1.7	0	0	+4.6 (+3.4)	0	−0.3	+9.5	−1.0	0	0	+7.7 (+7.3)
β-(1→4)	0	0	−1.0	+10.3 [+11.0]	−1.0	0	+6.7	0	0	0	+8.9 (+7.3)	−1.3	0	+7.2
β-(1→6)	0	0	0	0	−1.2	+7.8 [+6.0]	+6.7	*0*	*0*	*0*	*0*	*−1.0*	*+7.0*	*+6.1*

TABLE II *(continued)*

	2-Acetamido-2-deoxymannose[a]													
α-(1→3)								0	−3.4	+4.4 (+3.9)	−1.2 (−0.3)	0	0	+4.0 (+2.6)
β-(1→3)	0	−3.4	+5.9[α] +4.7[β]	−1.7	0	0	+3.2							

[a]The same effects as for mannose and rhamnose were used for other types of linkage.

(c) *P has the* galacto *configuration*

Linkage type	D-D(L-L)							D-L(L-D)						
	A						B	A						B
k	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-1'*	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*C-1'*
	Galactose													
α-(1→2)-α	−2.0	+5.0	−1.3	0	0	0	+4.0	−0.5	+8.6	−0.9	0	0	0	+8.3
α-(1→2)-β	−0.2	+5.1	−1.1	0	0	0	+5.8	−1.3	+5.2 (+6.8)	+0.7	0	0	0	+7.1
α-(1→3)	0	−1.5	+4.9 (+4.1) [+4.1]	−3.7	0	0	+3.3 (+2.4)	0	−0.7[α] −0.1[β]	+8.1	−0.4	0	0	+8.3
α-(1→4)	0	0	−0.5	+9.3[α] +8.3[β]	+0.2	−0.7	+8.0 [+6.8]	0	0	+0.3	+6.9	−0.1	0	+5.6
α-(1→6)	0	0	0	0	−2.1	+5.6	+5.7	0	0	0	0	−1.7	+6.0	+6.3
β-(1→2)-α	−0.2	+9.6	−1.0	0	0	0	+7.9	*−2.0*	*+7.0*	*−1.3*	*0*	*0*	*0*	*+4.0*
β-(1→2)-β	−1.4	+7.8	0	0	0	0	+6.9	*−0.2*	*+7.8*	*−1.1*	*0*	*0*	*0*	*+6.9*
β-(1→3)	0	−0.7	+9.8 (+9.1) [+9.1]	−0.3	0	0	+7.8	0	−1.2	+6.8	−2.8	0	0	+3.4
β-(1→4)	0	0	+0.3	+9.3[α] +8.6[β]	−0.7	0	+7.8	0	0	+0.6	+8.9	−1.8[α] −0.8[β]	−0.6	+6.8
β-(1→6)	0	0	0	0	−1.3	+7.8	+6.7 [+7.3]	0	0	0	0	−1.3	+7.1	+6.0

	2-Acetamido-2-deoxygalactose[a]												
α-(1→3)	0	−1.5	+4.9 (+4.1) [+4.1]	−3.7	0	0	+3.3 (+2.4)	0	−1.2	+9.2	−0.7	0	0	+7.9
α-(1→4)	0	0	−0.2 [−1.2]	+9.3$^{\alpha}$ +8.3$^{\beta}$ [+8.0]	+0.2	−0.7	+8.0 [+6.8]	0	0	−0.2	+6.4	−0.4	0	+6.6
β-(1→3)	0	−1.2	+9.8 (+9.1) [+9.1]	+0.1	0	0	+7.8							
β-(1→6)	0	0	0	0	−1.3	+7.8	+6.7 [+7.3]							
	2-Acetamido-2-deoxyfucose[a]												
α-(1→3)	0	−2.4	+4.9	−4.3	0	0	+2.7	0	−0.7	+8.2 [+6.1]	−0.4	0	0	+8.1
α-(1→4)	0	0	−0.2 [−1.2]	+9.3$^{\alpha}$ +8.3$^{\beta}$ [+8.0]	+0.4	−0.7	+8.0 [+6.8]	0	0	−0.2	+6.4	−0.4	0	+6.6
β-(1→3)	0	−1.8	+9.8 (+9.1) [+9.1]	−0.4 (−1.2)	0	0	+7.0	0	−1.5	+8.5	−2.5	0	0	+5.0 (+3.3)

[a]The same effects as for galactose were used for other types of linkage.

residues are calculated by the formula
$\delta(l) = \delta_o(l) + A(k,l) + B(k',l)$,
where l is the number of the carbon; $\delta_o(l)$ and $\delta(l)$ are the ^{13}C chemical shifts of that resonance for a monosaccharide in isolation and, with the same anomeric configuration, as a unit in the polysaccharide, respectively; $A(k,l)$ and $B(k',l)$ are the effects caused by the glycosylation of the unit (the index of linkage type k) and the formation of the linkage (the index of linkage type k'), respectively.

For purposes of comparison, the signals of the evaluated and observed spectra are ordered according to their chemical shifts, then the squared deviations (Δ^2) for the chemical shifts of the signals with identical numbers, their sum ($\Sigma\Delta^2$), and the normalised value related to one monosaccharide residue ($S = \Sigma\Delta^2/n$, where n is the number of monosaccharide residues in the repeating unit of the polysaccharide) are determined. Since, as a rule, deviations of the glycosylation effects from the average values are not more than 0.5 p.p.m., the possible candidates for the real structures of polysaccharides are those for which the value S does not exceed or exceed insignificantly 1.5 ($0.5^2 \times 6$).

The above-mentioned dependence of the glycosylation effects on the absolute configuration of the connected monosaccharide residues[2] enables this approach to be employed also for the determination of the absolute configuration of one or several monosaccharides in polysaccharides, provided that the absolute configurations of the other constituent monosaccharides are known. For this purpose, the whole cycle of the calculation for each of the possible sets of the monosaccharides, taking into account different absolute configurations, should be re-iterated.

The same approach may also be employed for the structural analysis of linear oligosaccharides.

Verification. — The computerised method for the determination of primary structures has been verified on three linear O-specific polysaccharides from *Shigella flexneri* variant Y, *Shigella dysenteriae* type 10, and *Pseudomonas aeruginosa* O7, for which the structures **1**–**3**, respectively, have been established by conventional methods and the ^{13}C-n.m.r. data reported[7-12].

→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→
I II III

1[7,8]

→2)-β-D-Man*p*-(1→3)-α-D-Man*p*NAc-(1→3)-β-L-Rha*p*-(1→4)-α-D-Glc*p*NAc-(1→

2[9,10]

→3)-α-L-Fuc*p*NAc-(1→3)-β-D-Fuc*p*NAc-(1→2)-β-D-Glc*p*-(1→

3[11,12]

Taking into account that the ^{13}C resonance for C-6 of each Hex and HexNAc residue is in the region 61.5–62.5 p.p.m., the structures involving (1→6) linkages

TABLE III

CHEMICAL SHIFTS IN THE ^{13}C-N.M.R. SPECTRA OF THE *Shigella* AND *Pseudomonas aeruginosa* POLYSACCHARIDES (CALCULATED VALUES ARE IN PARENTHESES)

Constituent	*Chemical shifts (p.p.m.)*					
	C-1	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*
Shigella flexneri variant Y						
→2)-α-L-Rha*p*-(1→	102.1	79.9	71.1	73.7	70.4	17.9
I	(102.1)	(80.0)	(71.0)	(73.5)	(70.0)	(18.0)
→2)-α-L-Rha*p*-(1→	101.9	79.3	71.3	73.5	70.2	17.8
II	(102.1)	(79.8)	(71.3)	(73.5)	(70.0)	(18.0)
→3)-α-L-Rha*p*-(1→	102.2	71.8	78.6	72.8	70.2	17.6
III	(102.2)	(71.4)	(78.8)	(72.9)	(70.0)	(18.0)
→3)-β-D-Glc*p*NAc-(1→	103.3	56.7	82.7	69.8	77.1	62.1
	(103.5)	(56.8)	(82.8)	(69.9)	(77.2)	(62.1)
Shigella dysenteriae type 10						
→2)-β-D-Man*p*-(1→	97.7	78.7	74.9	68.0	77.7	62.2
	(97.9)	(78.6)	(74.9)	(67.9)	(77.4)	(62.3)
→3)-α-D-Man*p*NAc-(1→	96.7	50.4	76.2	66.3	73.4	61.8
	(96.7)	(50.6)	(76.0)	(66.5)	(73.8)	(61.8)
→3)-β-L-Rha*p*-(1→	101.9	68.0	78.1	71.6	73.4	17.9
	(101.9)	(68.3)	(78.2)	(71.4)	(73.2)	(18.0)
→4)-α-D-Glc*p*NAc-(1→	99.4	55.6	71.6	78.1	71.6	62.2
	(99.5)	(55.3)	(72.1)	(78.1)	(71.9)	(61.9)
Pseudomonas aeruginosa O7						
→3)-α-L-Fuc*p*NAc-(1→	100.3	49.2	77.1	69.3	67.4	16.6
	(100.2)	(49.2)	(77.2)	(69.9)	(68.0)	(16.7)
→3)-β-D-Fuc*p*NAc-(1→	104.0	52.9	78.1	71.5	72.1	16.6
	(103.3)	(53.0)	(78.5)	(71.6)	(71.8)	(16.7)
→2)-β-D-Glc*p*-(1→	100.3	82.5	77.1	71.3	77.1	62.1
	(100.5)	(81.9)	(77.0)	(70.9)	(77.2)	(62.1)

were not considered. Since the $^{1}J_{C\text{-}1,H\text{-}1}$ values for these polysaccharides have not been determined, no restrictions were placed on the number of α and β linkages. For the *S. flexneri* polysaccharide, there is only one possible sequence of the monosaccharide residues, whereas, for the other two polysaccharides, all possible permutations of the monosaccharides were considered. For the *P. aeruginosa* polysaccharide, it was assumed that the Fuc*p*NAc residues may have any absolute configuration (both different and identical); the other monosaccharides were assumed to have the absolute configurations shown in **1–3**.

For each of the three polysaccharides, the optimal structures revealed by calculation, which are characterised by the lowest values *S*, were identical to the structures **1–3** established previously. The calculated ^{13}C chemical shift data for these optimal structures and the tentative assignments of the signals in the observed ^{13}C-n.m.r. spectra, accomplished by using the calculated data, are listed in Table III.

Besides the optimal structure **1** for the *S. flexneri* polysaccharide with $S = 0.3$, other structures can be distinguished with relatively low S values (0.7–1.5; for all other possible structures, S is $\geqq 1.6$). This is a consequence of the fact that the ^{13}C chemical shift data for disaccharide fragments containing a Rha*p* residue glycosylated at position 2,3, or 4 are similar. In each of the alternative structures, three Rha*p* residues are α and the Glc*p*NAc residues are β. Hence, the analysis revealed unambiguously the configurations of the glycosidic linkages.

		S
→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→	**(1)**	0.3
→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→		0.7
→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→		0.8
→3)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→		1.0
→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→		1.1
→4)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→4)-β-D-Glc*p*NAc-(1→		1.1
→2)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→4)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→		1.2
→3)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→3)-β-D-Glc*p*NAc-(1→		1.5

The optimal structure **2** for the *S. dysenteriae* polysaccharide is characterised by an S value of 0.3. An alternative structure, with the much larger S value of 1.9, differs from the optimal structure by the configuration of the Rha*p* linkage, the sequence of the sugar residues, and the mode of substitution of the Rha*p* residue. It is inconsistent with the methylation analysis data[9] and, therefore, may be ruled out. Each of the other possible structures is characterised by an S value of >2.3 and, hence, they are inconsistent with the observed ^{13}C-n.m.r. spectrum of the polysaccharide.

		S
→2)-β-D-Man*p*-(1→3)-α-D-Man*p*NAc-(1→3)-β-L-Rha*p*-(1→4)-α-D-Glc*p*NAc-(1→	**(2)**	0.3
→2)-β-D-Man*p*-(1→3)-α-D-Man*p*NAc-(1→4)-α-D-Glc*p*NAc-(1→2)-α-L-Rha*p*-(1→		1.95

TABLE IV

MONOSACCHARIDE COMPOSITION OF THE POLYSACCHARIDES

Polysaccharide	*Constituents of the repeating unit*					
	Glc	*Gal*	*GlcN*	*GalN*	*FucN*	*Fuc3N*[a]
S. arizonae O59		1	1		1	
P. hauseri O19		1	1	1	1	
S. arizonae O63	1	1		2		1

[a]3-Amino-3,6-dideoxygalactose.

TABLE V

METHYLATION ANALYSIS DATA FOR THE POLYSACCHARIDES

Polysaccharide	*Partially methylated monosaccharides*[a]								
	2,3,6-Glc	*2,4,6-Gal*	*3,4,6-Gal*	*2,6-Gal*	*4,6-GlcN*	*3,6-GalN*	*4,6-GalN*	*4-FucN*	*2,4-Fuc3N*
S. arizonae O59			+		+			+	
P. hauseri O19		+			+	+		+	
S. arizonae O63	+			+		+	+		+
S. arizonae O63f21[b]	+	+				+	+		

[a]2,3,6-Glc represents 1,4,5-tri-*O*-acetyl-2,4,6-tri-*O*-methylglucitol, 4,6-GlcN represents 1,3,5-tri-*O*-acetyl-2-deoxy-4,6-di-*O*-methyl-2-(*N*-methylacetamido)glucitol, *etc.* [b]Data for the polysaccharide of structure **6**, which represents the backbone of the *S. arizonae* polysaccharide.

TABLE VI

CHEMICAL SHIFTS IN THE ^{13}C-N.M.R. SPECTRA OF *Salmonella arizonae* AND *Proteus hauseri* POLYSACCHARIDES (CALCULATED VALUES ARE IN PARENTHESES)

Constituents	*Chemical shifts (p.p.m.)*					
	C-1	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*
Salmonella arizonae O59						
→3)-α-L-Fuc*p*NAc-(1→	98.9	49.0	77.3	69.8	67.1	16.4
	(99.1)	(49.2)	(77.2)	(69.9)	(68.0)	(16.7)
→3)-β-D-Glc*p*NAc-(1→	103.1	56.7	79.7	69.8	77.3	62.3
	(103.1)	(56.8)	(80.0)	(69.9)	(77.2)	(62.1)
→2)-β-D-Gal*p*-(1→	101.1	79.2	74.7	70.2	76.3	62.3
	(101.3)	(79.5)	(74.1)	(70.0)	(76.3)	(62.2)
Proteus hauseri O19						
→3)-α-D-Gal*p*-(1→	101.9	68.8	79.4	70.0	72.2	62.0
	(101.5)	(68.9)	(79.5)	(70.3)	(72.2)	(62.4)
→4)-α-D-Gal*p*NAc-(1→	99.7	51.3	68.4	79.4	73.2	61.7
	(100.3)	(51.0)	(68.5)	(79.2)	(72.4)	(61.7)
→3)-α-L-Fuc*p*NAc-(1→	98.7	49.7	74.4	71.6	67.9	16.4
	(99.1)	(50.0)	(74.8)	(72.0)	(68.0)	(16.7)
→3)-β-D-Glc*p*NAc-(1→	104.0	56.9	80.3	69.6	76.8	61.7
	(104.0)	(56.8)	(80.0)	(69.9)	(77.2)	(62.1)
Salmonella arizonae O63[a]						
→4)-α-D-Gal*p*NAc-(1→	95.4	51.0	68.7	79.5	72.5	61.5
	(95.5)	(51.0)	(68.5)	(79.2)	(72.4)	(61.7)
→3)-β-D-Gal*p*NAc-(1→	104.0	52.1	76.5	65.3	76.0	62.2
	(104.4)	(52.0)	(76.5)	(65.4)	(76.4)	(62.2)
→3)-β-D-Gal*p*-(1→	103.8	71.3	82.9	69.7	76.0	62.1
	(104.0)	(71.0)	(83.2)	(69.7)	(76.3)	(62.2)
→4)-α-D-Glc*p*-(1→	101.3	72.8	72.7	79.8	72.2	61.1
	(101.3)	(72.7)	(72.6)	(79.9)	(71.9)	(61.1)
Salmonella arizonae O63						
→4)-α-D-Gal*p*NAc-(1→	95.3	51.0	68.8	79.2	72.4	61.5
→3)-β-D-Gal*p*NAc-(1→	103.8	52.2	76.8	65.5	76.0	62.3
→3,4)-β-D-Gal*p*-(1→	104.1	72.2	80.7	77.1	76.0	61.2
→4)-α-D-Glc*p*-(1→	101.3	72.8	72.8	79.8	72.3	61.2
α-D-Fuc*p*3NAc-(1→	99.7	67.7	51.8	71.8	68.1	16.6

[a]Data for the polysaccharide of structure **6**, which represents the backbone of the *S. arizonae* O63 polysaccharide.

The optimal structure **3** for the *P. aeruginosa* polysaccharide is characterised by an *S* value of 0.7. An alternative structure with the nearest *S* value (3.5) and all other possible structures with *S* values of >4 are ruled out, since they are inconsistent with the observed ^{13}C-n.m.r. spectrum. Therefore, the calculation allowed the unambiguous establishment of the structure of the polysaccharide without the use of any additional data, as well as the determination of the absolute configurations of both Fuc*p*NAc residues.

Application. — The computerised approach was applied to the O-specific polysaccharides from *Proteus hauseri* O19, and *Salmonella arizonae* O59 and O63 (*Arizona* O19 and O8, respectively), the structures of which had not been established. The monosaccharide compositions of these polysaccharides, determined after acid hydrolysis, are given in Table IV. The ^{13}C-n.m.r. spectra of these polysaccharides (Table VI) showed that they are built up of tri- (*S. arizonae* O59), tetra- (*P. hauseri* O19), and penta-saccharide (*S. arizonae* O63) repeating-units, which accords with the results of sugar analysis. The spectra also indicated that all the amino sugar residues are *N*-acetylated (signals for CH_3 and CO of the acetamido groups are present at 23.2–23.5 and 174–175.7 p.p.m., respectively).

The methylation analysis data (Table V) revealed the *S. arizonae* O59 and *P. hauseri* O19 polysaccharides to be linear and the *S. arizonae* O63 polysaccharide to be branched with a Gal residue at the branching point and 3-acetamido-3,6-dideoxygalactose (Fuc3NAc) as the terminal residue of the side chain.

In applying the computerised approach to the two linear polysaccharides, structures with (1→6) linkages were rejected since the signals for C-6 of all the Hex and HexNAc residues were in the region 61–62.5 p.p.m. (Table VI). Glc, Gal, GlcNAc, and GalNAc were assumed to be D as is usual for bacterial polysaccharides, whereas both D and L configurations were assumed for the FucNAc residues.

The analysis of the *S. arizonae* O59 polysaccharide led to the optimal structure **4** with $S = 0.6$; for all other possible structures, the S values were >3.0, and, hence, none was consistent with the observed ^{13}C-n.m.r. spectrum of this polysaccharide.

→3)-α-L-Fuc*p*NAc-(1→3)-β-D-Glc*p*NAc-(1→2)-β-D-Gal*p*-(1→

4 (*S. arizonae* O59)

Thus, the analysis enabled unambiguous establishment of the structure of the polysaccharide, including the linkage positions, the sequence and configuration of the monosaccharide residues, and the L configuration for the Fuc*p*NAc residue. The structure **4** is confirmed by the methylation analysis data of the polysaccharide (Table V) and the $^1J_{C\text{-}1,H\text{-}1}$ values of 173.3, 161.7, and 161.1 Hz, which indicate[4] that one of the sugar units is α and that the other two are β.

In the analysis of the *P. hauseri* O19 polysaccharide, an additional restriction on the number of α and β linkages was adopted (three α and one β, as indicated by the $^1J_{C\text{-}1,H\text{-}1}$ values of 173, 169, and 168 Hz for the signals at 98.7, 99.7, and 101.9 p.p.m., respectively, which are typical[4] of the α anomers, and from the location of the fourth signal* at 104.0 p.p.m. consistent[1] with a β anomer).

The analysis of this polysaccharide afforded an optimal structure with $S = 0.7$ and four other structures with $S = 1.0$–1.9. Only the optimal structure **5** is consis-

* It was not possible to determine the corresponding $^1J_{C\text{-}1,H\text{-}1}$ value with sufficient accuracy because of the marked broadening of the signal

tent with the results of methylation analysis (Table V). All other possible structures had S values of >2.3 and were inconsistent with the observed ^{13}C-n.m.r. spectrum.

		S
→3)-α-D-Gal*p*-(1→4)-α-D-Gal*p*NAc-(1→3)-α-L-Fuc*p*NAc-(1→3)-β-D-Glc*p*NAc-(1→	**(5)**	0.7
→2)-α-D-Gal*p*-(1→4)-α-D-Gal*p*NAc-(1→3)-α-L-Fuc*p*NAc-(1→3)-β-D-Glc*p*NAc-(1→		1.0
→4)-α-D-Gal*p*-(1→4)-α-D-Gal*p*NAc-(1→3)-α-L-Fuc*p*NAc-(1→3)-β-D-Glc*p*NAc-(1→		1.1
→4)-α-D-Gal*p*-(1→4)-α-D-Gal*p*NAc-(1→3)-α-L-Fuc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→		1.6
→3)-α-D-Gal*p*-(1→4)-α-D-Gal*p*NAc-(1→3)-α-L-Fuc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→		1.9

The calculations also allowed tentative assignment of the ^{13}C signals in the spectra of the *S. arizonae* O59 and *P. hauseri* O19 polysaccharides. The assignment together with the chemical shift data, calculated for the structures **4** and **5**, are given in Table VI.

The only obstacle to the evaluation of the ^{13}C-n.m.r. spectra of branched polysaccharides is the lack of a good data base derived from branched model oligosaccharides. The glycosylation effects derived from unbranched models (Table II) cannot be used for this purpose since the chemical shifts calculated for the resonances of monosaccharide residues involved in branching often differ[13] (up to 5 p.p.m.) from the observed values. This difficulty may be avoided by employing a two-step analysis involving first, determination of the structure of the backbone, and second, the locations of the side chains. However, this approach requires that the monosaccharide composition of the backbone be known and one way to solve the problem is to split off the side chains. This approach has been used in establishing the structure of a branched polysaccharide from *S. arizonae* O63.

Solvolysis of the polysaccharide with anhydrous hydrogen fluoride under mild conditions selectively cleaved the 3-acetamido-3,6-dideoxygalactose residues. Acid hydrolysis and ^{13}C-n.m.r. data (Table VI) showed the resulting polysaccharide to be built up of tetrasaccharide repeating-units containing Glc, Gal, and GalNAc in the ratios 1:1:2. Methylation analysis (Table V) revealed the polysaccharide to be unbranched, and, hence, to represent the backbone of the original polysaccharide.

In the computerised analysis of this modified polysaccharide, the (1→6) linkages were not considered and all the constituent monosaccharides were assumed to be D. The calculation led to the structure **6**, having the lowest S value (0.3), in which the modes of substitution of the constituent monosaccharides are consistent with the methylation analysis data (Table V). In four alternative structures with S values of 1.4–1.7, the constituent sugars have the anomeric configurations as in **6** but different linkage positions and, hence, are inconsistent with the methylation data. For

the other possible structures, the *S* values are >2.3. Thus, the structure **6** is the only one consistent both with the observed ^{13}C-n.m.r. spectrum (Table VI) and the methylation analysis data.

		S
→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-α-D-Glc*p*-(1→	(**6**)	0.3
→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→2)-β-D-Gal*p*-(1→2)-α-D-Glc*p*-(1→		1.4
→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→2)-β-D-Gal*p*-(1→3)-α-D-Glc*p*-(1→		1.5
→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→4)-β-D-Gal*p*-(1→3)-α-D-Glc*p*-(1→		1.5
→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→		1.7

In considering the structure of the original branched polysaccharide, the methylation analysis data (Table V) showed that the 3-acetamido-3,6-dideoxygalactose residue was attached to Gal at position 4. Three of the five constituent monosaccharides in this polysaccharide are α and two are β, as follows[4] from the $^1J_{C\text{-}1,H\text{-}1}$ values of 170.9, 170.9, 169.7, 162.3, and 162.3 Hz for the signals at 95.3, 99.7, 101.3, 103.8, and 104.1 p.p.m., respectively. In the modified polysaccharide **6**, two constituent monosaccharides are α and two are β; hence, the 3-acetamido-3,6-dideoxygalactose residues in the original polysaccharide are α. The $[\alpha]_D$ value of the 3-amino-3,6-dideoxygalactose hydrochloride (Fuc3NAc), isolated from the hydrolysate of the polysaccharide, indicated it to be D, so that the polysaccharide has the structure **7**.

→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-α-D-Glc*p*-(1→
4
↑
1
α-D-Fuc*p*3NAc

7 (*S. arizonae* O63)

The ^{13}C-n.m.r. data for this polysaccharide (Table VI) are consistent with the structure proposed. The tentative assignment of the signals in Table VI was accomplished by comparison with the data for the modified polysaccharide **6** and methyl 3-acetamido-3,6-dideoxy-α-L-galactopyranoside[14].

The computerised approach to structural analysis allowed the structures of three new polysaccharides, including a branched polysaccharide with a pentasaccharide repeating-unit, to be established without the preparation of oligosaccharide fragments, thus saving time and material. Knowledge of the monosaccharide composition and the ^{13}C-n.m.r. data may sometimes be sufficient for the unambiguous determination of the structure of the repeating unit (*e.g.*, the *P. aeruginosa* O7 and *S. arizonae* O59 polysaccharides). Otherwise, the structure obtained in the

computerised approach should be confirmed by methylation analysis, which may also allow a choice of several structures with similar calculated ^{13}C-n.m.r. spectra. It is possible that the data provided by both of these methods will not indicate unambiguously the structure of a polysaccharide, but the calculations should reduce markedly the number of possible structures.

EXPERIMENTAL

The programme was developed in the non-dialogue mode, using a BESM-6 computer (U.S.S.R.) and the programming language ALGOL-60. The input data were as follows: (*1*) the codes of constituent sugars (glucose, 1; galactose, 2; *etc.*) and the number of each sugar residue in the repeating unit, (*2*) the ^{13}C chemical shift data of the free monosaccharides (Table I), (*3*) the matrices *A* (1:10,1:6) and *B* (1:10,1:6) comprising the glycosylation effects for the constituent monosaccharides (Tables I and II), and (*4*) the chemical shifts in decreasing order from the observed ^{13}C-n.m.r. spectrum of the polysaccharide. The data output for each generated structure characterised by an *S* value of <4.0 are (*1*) the sequence of the constituent sugar codes, which determines the sequence of the monosaccharide residues in the repeating unit; (*2*) the sequence of indices *k* (see Table II), which determines the types of intersugar linkages; (*3*) the value of *S*; (*4*) the assigned calculated ^{13}C chemical shifts; and (*5*) the calculated ^{13}C chemical shifts in decreasing order.

^{13}C-N.m.r. spectra were recorded with a Bruker AM-300 instrument for solutions in D_2O at 60° (internal methanol; δ_C 50.15). Optical rotations were measured with an EPO-1 polarimeter for aqueous solutions at 20°. G.l.c.–m.s. was performed on a Varian MAT Gnom 111 instrument, using a column packed with 3% of OV-1 on Diatomite CQ (100–200 mesh). Gel filtration was performed on a colunn (80 × 1.6 cm) of TSK HW 40 by elution with water, and on a column (70 × 3.5 cm) of Sephadex G-50 by elution with pyridine acetate buffer (pH 4.5). Ion-exchange chromatography of amino sugars was carried out on a column (25 × 0.6 cm) of Chromex UA-8 by elution with 0.3M hydrochloric acid.

S. arizonae lipopolysaccharides were obtained[15] from acetone-dried cells and treated[16] with aqueous 1% acetic acid followed by gel filtration on Sephadex G-50 to give the polysaccharides.

Acid hydrolysis was carried out with 4M hydrochloric acid (3 h, 100°) and the resulting monosaccharides were identified by using sugar and amino acid analysers[17]. Ion-exchange chromatography of the hydrolysate of the *S. arizonae* O63 polysaccharide gave 3-amino-3,6-dideoxy-D-galactose hydrochloride, $[\alpha]_D$ +55° (*c* 0.2), *cf.* $[\alpha]_D$ +114° (water) for the *N*-acetyl derivative[18]. Methylation analysis was carried out by the reported method[19], and the partially methylated monosaccharides were identified as their alditol acetates, using published data[20,21]. Solvolysis with anhydrous hydrogen fluoride was carried out at −40° (40 min), and the products were isolated by precipitation with ether[22] followed by gel filtration on TSK HW 40.

ACKNOWLEDGMENTS

We thank Professor E. S. Stanislavsky (I. I. Mechnikov Institute of Vaccines and Sera, Moscow) for providing the *S. arizonae* cells, Dr. W. Kaca (University of Lodz, Institute of Microbiology) for a preparation of *P. hauseri* lipopolysaccharide, and Dr. V. N. Shibaev (N. D. Zelinsky Institute of Organic Chemistry) for helpful discussion.

REFERENCES

1 P. A. J. Gorin, *Adv. Carbohydr. Chem. Biochem.*, 38 (1980) 13–104.
2 N. K. Kochetkov, O. S. Chizhov, and A. S. Shashkov, *Carbohydr. Res.*, 133 (1984) 173–185.
3 G. M. Lipkind, A. S. Shashkov, and N. K. Kochetkov, *Bioorg. Khim.*, 13 (1987) 833–841.
4 K. Bock and C. Pedersen, *J. Chem. Soc., Perkin Trans. 2*, (1974) 293–297.
5 J. H. Bradbury and G. A. Jenkins, *Carbohydr. Res.*, 126 (1984) 125–156.
6 K. Bock, C. Pedersen, and H. Pedersen, *Adv. Carbohydr. Chem. Biochem.*, 42 (1984) 193–225.
7 K. Bock, S. Josephson, and D. R. Bundle, *J. Chem. Soc., Perkin Trans. 2*, (1982) 59–70.
8 N. K. Kochetkov, N. E. Byramova, Y. E. Tsvetkov, and L. V. Backinowsky, *Tetrahedron*, 41 (1985) 3363–3375.
9 B. A. Dmitriev, Y. A. Knirel, O. K. Sheremet, N. K. Kochetkov, and I. L. Hofman, *Bioorg. Khim.*, 3 (1977) 1219–1225.
10 A. S. Shashkov, B. A. Dmitriev, Y. A. Knirel, O. K. Sheremet, and N. K. Kochetkov, *Bioorg. Khim.*, 5 (1979) 583–587.
11 B. A. Dmitriev, Y. A. Knirel, N. A. Kocharova, N. K. Kochetkov, E. S. Stanislavsky, and G. M. Mashilova, *Eur. J. Biochem.*, 106 (1980) 643–651.
12 A. S. Shashkov, Y. A. Knirel, N. A. Kocharova, B. A. Dmitriev, and N. K. Kochetkov, *Bioorg. Khim.*, 6 (1980) 1332–1337.
13 R. U. Lemieux, K. Bock, L. T. J. Delbaere, S. Koto, and V. S. Rao, *Can. J. Chem.*, 58 (1980) 631–653.
14 V. L. L'vov, N. V. Tochtamysheva, A. S. Shashkov, B. A. Dmitriev, and K. Čapek, *Carbohydr. Res.*, 112 (1983) 233–239.
15 O. Westphal and K. Jann, *Methods Carbohydr. Chem.*, 5 (1965) 83–91.
16 E. Müller-Seitz, B. Jann, and K. Jann, *FEBS Lett.*, 1 (1968) 311–314.
17 B. A. Dmitriev, Y. A. Knirel, N. K. Kochetkov, and I. L. Hofman, *Eur. J. Biochem.*, 66 (1976) 559–566.
18 G. Ashwell and W. A. Volk, *J. Biol. Chem.*, 240 (1965) 4549–4555.
19 H. E. Conrad, *Methods Carbohydr. Chem.*, 6 (1972) 361–364.
20 P.-E. Jansson, L. Kenne, H. Liedgren, B. Lindberg, and J. Lönngren, *Chem. Commun. Univ. Stockholm*, (1976) 1–15.
21 K. Stellner, H. Saite, and S. Hakomori, *Arch. Biochem. Biophys.*, 155 (1973) 464–472.
22 A. J. Mort, *Carbohydr. Res.*, 122 (1983) 315–321.