Carbohydrate Research, 188 (1989) 169–191
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

COMPUTER-ASSISTED STRUCTURAL ANALYSIS OF POLYSACCHARIDES WITH AN EXTENDED VERSION OF CASPER USING ^{1}H- AND ^{13}C-N.M.R. DATA*

PER-ERIK JANSSON, LENNART KENNE, AND GÖRAN WIDMALM
Department of Organic Chemistry, Arrhenius Laboratory, University of Stockholm, S-106 91 Stockholm (Sweden)

(Received November 7th, 1988; accepted for publication, January 4th, 1989)

ABSTRACT

The computer program CASPER, used in the structural analysis of polysaccharides composed of repeating units, has been extended. The extended version uses either unassigned ^{1}H- or ^{13}C-n.m.r. chemical shifts or the complete unassigned C,H-correlation spectrum, and can predict the structure of linear and branched oligo- and poly-saccharides. The number of possible structures, consistent with sugar and methylation analysis, can be decreased by the use of $^1J_{C,H}$ and $^3J_{H,H}$ values. The database, which contains ^{1}H- or ^{13}C-n.m.r. chemical shift data for monosaccharides and ^{1}H- or ^{13}C-glycosylation shifts for all types of glycosidic linkages obtained by combination of the monosaccharides, has been increased and now also contains correction values for sugar residues present in branch-point regions. The program has been tested on four polysaccharides of known structure but with different degrees of complexity. For three polysaccharides, the correct structure was suggested; for the fourth, two structures were consistent with the n.m.r. data, one of them being correct.

INTRODUCTION

In structural studies of carbohydrates, sequence analysis is still a difficult and time-consuming task. In earlier studies, the most common procedure was analysis of fragments obtained by chemical or enzymic degradations. This method has recently been partly replaced by n.m.r. spectroscopy, using information from nuclear Overhauser enhancement (n.O.e.)[1] and scalar couplings[2,3] over the glycosidic bond and from the chemical shifts of signals from structural reporter groups[4]. The structures of several oligo- and poly-saccharides have also been elucidated, mainly or exclusively by one- and two-dimensional n.m.r. spectroscopy. However,

*Presented at the XIVth International Carbohydrate Symposium, Stockholm, Sweden, August 14–19, 1988.

0008-6215/89/$ 03.50

these techniques require high-field n.m.r. spectrometers and it is often difficult and time-consuming to assign the signals in the rather complex spectra.

The ^{1}H- and ^{13}C-n.m.r. chemical shifts of the resonances from a monosaccharide residue within a larger saccharide depend mainly upon the structure of the monosaccharide and upon the nature of the flanking sugar residues. This fact has led to some computerised approaches to the structural analysis of carbohydrates, using ^{1}H- and ^{13}C-n.m.r. data. Computer programs have been developed for the analysis of oligosaccharide structures, derived from glycopeptides and glycolipids, using ^{1}H-n.m.r. chemical shifts[5-7]. As the structures of these oligosaccharides are closely related, the chemical shifts of signals from each monosaccharide residue can be compared to data for residues having the same surroundings in known structures.

Recently, we described[8] a computer program, CASPER, by which structural analysis of (*a*) linear polysaccharides having repeating units or (*b*) oligosaccharides could be performed. Sugar and methylation analysis data were used in combination with unassigned ^{13}C-n.m.r. chemical shift data. Similar approaches have been reported for the analysis of linear bacterial polysaccharides[9] and mannans[10], using ^{13}C-n.m.r. chemical shift data.

We now describe an improved and extended version of CASPER by which branched polysaccharides with one branch-point residue can also be analysed. The use of coupling constants for the anomeric carbons and protons ($^1J_{C\text{-}1,H\text{-}1}$ and $^3J_{H\text{-}1,H\text{-}2}$), the ^{1}H-n.m.r. chemical shifts from some or all signals, and data from 2-D C,H-COSY spectra have improved the analysis. Using ^{1}H and ^{13}C glycosylation shifts, the linkage position and the stereochemistry around the glycosidic bond are accounted for. By the introduction of C,H-COSY spectra, the influence on the chemical shifts of the monosaccharide residues is also considered. The basis for this is that the two-dimensional information present in a pair of chemical shifts is lost on going to one dimension, *i.e.*, the 1D spectrum. An example of this involves the chemical shifts for signals from H-3/C-3 and H-4/C-4 in β-D-Glc and in β-D-Gal, which, pair-wise, appear as 3.50/76.6 and 3.42/70.7 p.p.m. for β-D-Glc, and 3.59/73.8 and 3.89/69.7 p.p.m. for β-D-Gal. For the corresponding one-dimensional data, the identification is less obvious. C/H-Correlation spectra have not been used extensively earlier, because of the low sensitivity of the ^{13}C nucleus. Recently, a new technique, "inverse detection" HMQC, utilizing detection of the ^{1}H nucleus instead, has increased the sensitivity and made it possible to obtain C/H-correlation spectra of carbohydrates in considerably shorter times[11].

The use of the different possibilities in the program is demonstrated by analysis of some polysaccharides having known structures. A detailed account of the program will be published elsewhere.

RESULTS AND DISCUSSION

General overview. — The previous version of CASPER[8] consisted of three parts, namely, the database, the spectrum simulator, and the fitting procedure. The

program could handle linear oligosaccharides and polysaccharides composed of regular repeating units for which ^{13}C-n.m.r. spectra for each possible permutation could be simulated and compared to the experimental spectrum. The simulation is based on an additivity approach. Thus, the spectrum of a polysaccharide is assumed to be the sum of the chemical shifts of the monomeric residues and a number of induced chemical shift changes (glycosylation shifts). In the extended version of CASPER, the size of the database is increased, the program can handle branched structures with one branch-point residue, coupling constants for anomeric carbons and protons ($^1J_{C,H}$ and $^3J_{H,H}$) can be used to decrease the number of possible structures, and ^{1}H-n.m.r. data and C,H-correlation data obtained by 2D C,H-COSY experiments can be used. The program can also calculate the ^{1}H- and the ^{13}C-n.m.r. chemical shifts for a given structure.

A flow diagram for the extended version of CASPER is shown in Fig. 1. Input data are the identities of the constituent sugars in the repeating unit of the polysaccharide, or the oligosaccharide, and the linkage positions. Other data are the ^{1}H- or the ^{13}C-n.m.r. chemical shifts, or C,H-correlation data. For the C,H-spectra, the chemical shifts for each C/H signal are given to the program as pairs of chemical shifts in a descending order with respect to carbon. In addition, the $^3J_{H,H}$ values for signals from anomeric protons, if present, are given as the number of large (7–8 Hz), medium (3–4 Hz), and small (<2 Hz) couplings, referring to the normal values for the *β-gluco/galacto*, *α-gluco/galacto*, or *α/β-manno* configurations, respectively. Thus, the number 123 means that six sugar residues, one with the *β-gluco-*, two with the *α-gluco-*, and three with the *α-* or *β-manno* configuration, are the constituents of the repeating unit. The values for $^1J_{C,H}$, if available, are given as another number which refers to the size of each C-1,H-1 coupling

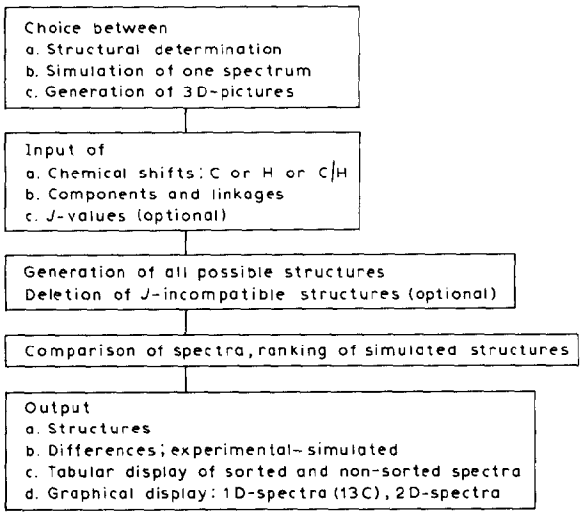


Fig. 1. Generalised flow diagram for the extended version of CASPER.

TABLE I

THE EFFECT OF INTRODUCING *J*-VALUES TO DECREASE THE NUMBER OF SUGGESTED STRUCTURES

Polysaccharide	*Number of possible structures*			
	Total[a]	$J_{H,H}$[b]	$J_{C,H}$[b]	$J_{H,H}+J_{C,H}$[b]
S. flexneri Y	96	48	24	6
S. pneumoniae 18A	1536	576	480	288
S. flexneri X	1536	768	240	96
E. coli O75	192	96	72	48

[a]Number of possible structures using information on components and linkages. [b]Number of structures compatible with information on $J_{H,H}$, *etc.*

constant, ~170 Hz for α, or ~160 Hz for β. A number of 12 thus means one α and two β configurations. These $^1J_{C,H}$ values also differentiate between α- and β-*manno* configurations, which have the same $^3J_{H,H}$ value (<2 Hz). Furthermore, the program requires information about the type of saccharide that is analysed, *i.e.*, a poly- or an oligo-saccharide. Using data on components and linkages, all possible repeating units or oligosaccharide structures can be generated by the computer by permutation of the constituent sugars and generation of all possible combinations of anomeric configurations. If $^3J_{H,H}$ and/or $^1J_{C,H}$ values are available, a comparison of the given numbers, representing the anomeric configurations, with the numbers from all suggested structures can be performed. This procedure drastically decreases the number of possible structures, as all suggested structures having a non-corresponding set of anomeric configurations will be omitted (Table I). Consequently, this saves calculation time for simulation of spectra and decreases the demand for computer memory.

The ^{1}H- and/or ^{13}C-n.m.r. spectra for all suggested structures are then calculated using chemical shifts, glycosylation shifts, and, for branched polysaccharides with vicinal disubstitution, correction values for sugars present in branching regions, using data from the database (see below). The fit according to peak-by-peak comparison with the experimental spectrum is calculated, and the simulated spectra are ranked according to their fit. In order to avoid a large number of calculations of C,H-spectra, which would be the case if the spectra of all possible structures were simulated and compared with the experimental spectrum, the ^{13}C-n.m.r. chemical shifts are simulated first and only structures with a good fit to these data are used in the further simulations. For the comparison of C,H-correlation spectra, a least-squares-fit procedure is used. This comparison is a time-consuming task, as there are many different combinations that must be tested to fit the spectra together. A simplification in the fitting procedure which significantly decreases the calculation time has been introduced by a division of the spectrum into groups of signals. This procedure is especially important for the signals from the ring atoms and is possible if there are gaps in the spectrum. Furthermore, as the spectral dis-

persion is different for ^{1}H- and ^{13}C-spectra, scaling factors have to be used. Consequently, C,H-differences are expressed in units instead of p.p.m. $\Delta\delta$-sums.

Databases. — In the previous version of CASPER, the database contained only chemical shifts for the ^{13}C resonances of both anomeric forms of six monosaccharides and ^{13}C glycosylation shifts observed for differently linked monosaccharides. In the extended version of CASPER, the database contains ^{1}H and ^{13}C chemical shifts for both anomeric forms of eleven monosaccharides, among which both amino sugars and uronic acids are represented (Tables II and III). The glycosylation shifts for all ^{1}H and ^{13}C resonances for all types of disaccharide elements having pyranosidic rings are given in the $\Delta\delta$-file. These glycosylation shifts are mainly obtained from studies of model disaccharides[12–16]. For some *Shigella* polysaccharides with branched structures, it was shown that, when the branched residue was not vicinally disubstituted, additivity of $\Delta\delta$-values holds; for vicinally disubstituted residues, deviations are found[17]. Therefore, the extended version of CASPER contains a file with correction values for sugar residues in the branch-point region. These values were obtained from model studies of "branched" trisaccharides[15,18].

The ^{1}H-n.m.r. $\Delta\delta$-file. — The ^{1}H $\Delta\delta$-file contains glycosylation shifts for di-

TABLE II

^{1}H-N.M.R. DATA[a] FOR GLYCOPYRANOSES

Sugar	*H-1*	*H-2*	*H-3*	*H-4*	*H-5*	*H-6*	*H-6*	*NAc*
α-D-Glc	5.23	3.54	3.72	3.42	3.84	3.76	3.84	
β-D-Glc	4.64	3.25	3.50	3.42	3.46	3.72	3.90	
α-D-Gal	5.22	3.78	3.81	3.95	4.03	3.69	3.69	
β-D-Gal	4.53	3.45	3.59	3.89	3.65	3.64	3.72	
α-D-Man	5.18	3.94	3.86	3.68	3.82	3.74	3.86	
β-D-Man	4.89	3.95	3.66	3.60	3.38	3.75	3.91	
α-L-Rha	5.12	3.92	3.81	3.45	3.86	1.28		
β-L-Rha	4.85	3.93	3.59	3.38	3.39	1.30		
α-L-Fuc	5.20	3.77	3.86	3.81	4.20	1.21		
β-L-Fuc	4.55	3.46	3.63	3.74	3.79	1.26		
α-D-GlcNAc	5.21	3.88	3.75	3.49	3.86	3.77	3.85	2.06
β-D-GlcNAc	4.72	3.65	3.56	~3.46	~3.46	3.75	3.91	2.06
α-D-GalNAc	5.28	4.19	3.95	4.05	4.13	3.79	3.79	2.06
β-D-GalNAc	4.68	3.90	3.77	3.98	3.72	3.82	3.84	2.06
α-D-ManNAc	5.13	4.31	4.07	3.63	3.86	~3.84	~3.84	2.10[b]
β-D-ManNAc	5.01	4.45	3.83	3.52	3.45	3.81	3.90	2.06
α-D-GlcA-Na	5.24	3.59	3.75	3.53	4.09			
β-D-GlcA-Na	4.65	3.30	3.52	3.54	3.72			
α-D-GalA-Na	5.30	3.83	3.92	4.29	4.39			
β-D-GalA-Na	4.56	3.51	3.69	4.23	4.03			
α-D-ManA-Na	5.22	3.90	3.87	3.83	4.04			
β-D-ManA-Na	4.89	3.93	3.66	3.72	3.63			

[a]Obtained at 70° for solutions in deuterium oxide (internal TSP; δ 0.00). [b]The assignment of signals from the *N*-acetyl groups in the respective anomeric forms was done using information on peak heights.

TABLE III

^{13}C-N.M.R. DATA[a] FOR GLYCOPYRANOSES

Sugar	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*NAc*	
α-D-Glc	92.99	72.47	73.78	70.71	72.37	61.84		
β-D-Glc	96.84	75.20	76.76	70.71	76.76	61.84		
α-D-Gal	93.18	69.35	70.13	70.28	71.30	62.04		
β-D-Gal	97.37	72.96	73.78	69.69	75.93	61.84		
α-D-Man	94.94	71.69	71.25	67.94	73.34	61.99		
β-D-Man	94.55	72.13	74.03	67.69	77.00	61.99		
α-L-Rha	94.84	71.81	71.00	73.19	69.12	17.67		
β-L-Rha	94.37	72.23	73.76	72.83	72.83	17.61		
α-L-Fuc	93.12	69.09	70.30	72.80	67.10	16.33		
β-L-Fuc	97.15	72.73	73.93	72.35	71.64	16.33		
α-D-GlcNAc	91.77	55.00	71.74	71.26	72.51	61.78	22.87[b]	175.13[b]
β-D-GlcNAc	95.85	57.86	74.81	71.06	76.82	61.85	23.10	175.49
α-D-GalNAc	91.95	51.16	68.40	69.56	71.36	62.11	22.91	175.43
β-D-GalNAc	96.29	54.80	72.01	68.85	75.98	61.89	23.12	175.78
α-D-ManNAc	93.96	54.14	69.75	67.93	73.00	61.54	22.84	175.43
β-D-ManNAc	93.91	54.94	73.00	67.65	77.25	61.54	22.98	176.39
α-D-GlcA-Na	92.96	72.26	73.52	72.91	72.47	177.42[b]		
β-D-GlcA-Na	96.77	75.00	76.53	72.69	76.93	176.47		
α-D-GalA-Na	93.07	69.03	70.26	71.64	72.30	176.43		
β-D-GalA-Na	96.89	72.62	73.79	71.18	76.44	175.59		
α-D-ManA-Na	94.80	71.42	71.10	69.95	73.68	177.58		
β-D-ManA-Na	94.48	71.93	73.86	69.58	76.95	176.82		

[a]Obtained at 70° for solutions in deuterium oxide (internal 1,4-dioxane; δ 67.40). [b]The assignment of signals from the *N*-acetyl groups and the carboxyl groups in the respective anomeric forms was done using information on peak heights.

saccharide elements obtained by combinations of the monosaccharides to all types of 1→2, 1→3, 1→4, and 1→6 linkages. These glycosylation shifts are mainly obtained from studies of model disaccharides[12–15] or by approximation of glycosylation shifts from these studies. Certain groups, *e.g.*, acetamido groups, cause shielding of neighbouring hydrogens on the adjacent ring. This effect varies between −0.1 and −0.2 p.p.m., depending upon substance, distance, and type of glycosidic linkage. For approximations from disaccharide elements lacking acetamido groups to elements with acetamido groups, such a shielding correction was introduced. An example is 3-linked 2-acetamido-2-deoxy-D-glucose derivatives.

Unpublished ^{1}H-n.m.r. data, obtained from model studies, which were used in this study are given in Table IV.

The ^{13}C-n.m.r. Δδ-file. — The ^{13}C Δδ-file contains glycosylation shifts for all ^{13}C resonances for the same disaccharide elements as the ^{1}H Δδ-file[12–16]. Approximations introduced to the glycosylation shifts for disaccharides that are not included in the model studies follow several rules. If similar stereochemistry is present around the glycosidic linkage, the same Δδ-set is used, *e.g.*, for β-D-Gal*p*-(1→3)-α-D-Man*p* and β-D-Glc*p*-(1→3)-α-D-Man*p*. For an axial hydroxyl sub-

TABLE IV

^{1}H-N.M.R. GLYCOSYLATION SHIFTS ($\Delta\delta$-VALUES) FROM UNPUBLISHED MATERIAL USED IN THE CALCULATION OF SIMULATED SPECTRA

Disaccharide element	*H-1*	*H-2*	*H-3*	*H-4*	*H-5*	*H-6*	*H-6'*	*CH_3*
β-D-Gal-(1→3)-β-D-GlcNAc	−0.11[a]	0.10	0.05	0.04	0.05	0.14	0.06	
	0.06	0.13	0.25	0.09	0.04	0.02	0.01	−0.01
β-D-GlcNAc-(1→3)-β-D-Gal	0.02	0.08	0.03	0.01	0.01	0.01	0.00	−0.02
	0.00	0.06	0.06	0.19	−0.01	0.00	0.00	

[a]The upper line refers to values of the glycosyl group.

TABLE V

^{13}C-N.M.R. GLYCOSYLATION SHIFTS ($\Delta\delta$-VALUES) FROM UNPUBLISHED MATERIAL USED IN THE CALCULATION OF SIMULATED SPECTRA

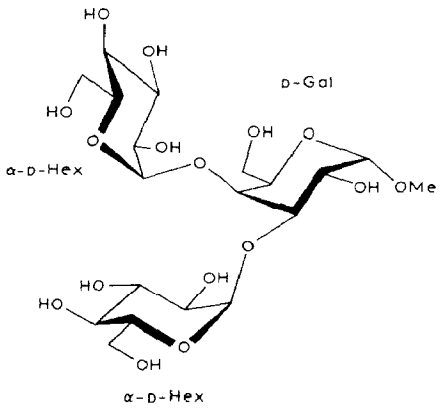
Disaccharide element	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*CO*	*CH_3*
β-D-Gal-(1→3)-β-D-GlcNAc	6.93[a]	−1.31	−0.26	−0.24	0.20	−0.03		
	−0.25	−1.07	8.76	−1.28	−0.45	−0.01	0.09	0.05
β-D-GlcNAc-(1→3)-β-D-Gal	7.39	−1.11	−0.19	−0.21	−0.22	−0.01	0.17	0
	0.00	−0.97	9.23	−0.39	−0.40	−0.06		

[a]The upper line refers to values of the glycosyl group.

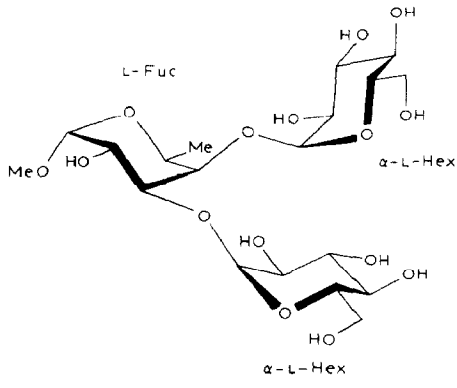
stituent in a 1,3-relation to the linkage, modified $\Delta\delta$-values are given, *e.g.*, for 2-*O*-substituted D-Gal*p*, 0.5 p.p.m. to the $\Delta\delta$-value of C-2, and for 4-*O*-substituted D-Man*p*, 1.0 p.p.m. to the $\Delta\delta$-value of C-4. If an α-glycosyl group has HO-2 axial, a correction of −0.5 p.p.m. for signals from C-1′ and C-2′ in the glycosyl group and for the linkage carbon in the substituted residue, *i.e.*, the aglycon, was introduced. For β-linked disaccharides with an axial substituent, only the linkage carbon will obtain an additional −0.5 p.p.m. for its $\Delta\delta$-value.

Unpublished ^{13}C-n.m.r. data, obtained from model studies, which were used in this study are given in Table V.

The branch-point correction $\Delta\delta$-file. — Files containing values for the correction of ^{1}H and ^{13}C $\Delta\delta$-values are implemented for some 2,3- and 3,4-branched residues, and a correction of the glycosylation shifts is introduced for the branch-point residue and for the two sugars substituting it. Reference data are derived from sixteen trisaccharides consisting of 3,4-diglycosyl-substituted α-D-galactopyranosyl residues that have recently been investigated[15,18]. Every possible combination of glycosyl groups with α-, β-, D-, and L-configurations has been investigated. For approximations of the glycosylation shifts for other vicinally substituted trisaccharide elements, only the $\Delta\delta$-values for signals from the linkage carbons and the corresponding protons are corrected. Thus, for example, for 3,4-branched L-fucosyl derivatives (**2**), the same correction values are used as those obtained for



1



2

the corresponding D-galactosyl derivative (**1**) in which the glycosyl groups have the opposite configuration. The main interactions between the different residues will then be identical, except that those to the hydroxymethyl group will now be to a methyl group.

In a similar manner, 2,3-disubstituted L-rhamnosyl residues are approximated from D-galactosyl trisaccharides. The 2-*O*-glycosyl group and the 3-*O*-glycosyl group in the L-rhamnosyl residue should have the same anomeric and absolute configuration as the 4-*O*-glycosyl group and the 3-*O*-glycosyl group, respectively, in the galactosyl derivative. For the two trisaccharides, the stereochemistry around the glycosyl bonds is the same and consequently the main interactions between the different residues are the same. The values for the 4-*O*-glycosyl group and C-4 in the D-galactosyl derivative are then used for the 2-*O*-glycosyl group and C-2 in the L-rhamnosyl derivative.

For 2,3-disubstituted D-mannosyl residues, values from D-galactosyl trisaccharides with reversed absolute configuration of the glycosyl groups are used to obtain the same stereochemistry at the glycosyl bonds. Thus, for a 3-*O*-α-L-glycosyl-4-*O*-β-D-glycosyl-substituted D-galactosyl residue, values for a 3-*O*-α-D-glycosyl-4-*O*-β-L-glycosyl-substituted D-mannosyl residue are obtained and the Δδ-value from the C-4 signal in the D-galactosyl derivative is given to the C-2 signal in the D-mannosyl derivative. For the latter two approximations, some deviations can be expected from the presence of the aglycon on C-1.

In the simulation of the n.m.r. spectra, the values from the monomer δ-file, the Δδ-file and, when applicable, the correction file are added.

Output of results. — The results from CASPER include a chosen number of suggested structures ranked according to their fit with the experimental spectrum. In addition to these, the Δδ-sum, the deviation/signal, the numbers for the coupling constants, and a check number are given for each suggested structure. The Δδ-sum is the total chemical shift difference between signals in the simulated and experimental spectrum when a signal-to-signal comparison is performed. The structures

are ranked according to this value. The check numbers describe the accuracy of the simulation; a low number is given when data from identical disaccharide elements or only minor approximations of these have been used. The calculated chemical shift data for each structure are given with all values in a descending order, or in a non-sorted form in which the chemical shifts for the signals from each monosaccharide residue are given in a Table. The simulated C,H-correlation spectra are given with the values for each C,H-pair in the order that they fit those in the experimental spectrum. In addition, the spectra can be shown and compared in the graphic mode.

A quantitative estimation of the $\Delta\delta$-sum is dependent upon the type of simulation, *i.e.*, ^{13}C, ^{1}H, or C,H-correlation n.m.r. spectrum, and the quality of data used, *i.e.*, the check-number sum. As a rule of thumb, ^{13}C $\Delta\delta$-sums that are >1.5 times the value for the best structure are considered to be incompatible with the n.m.r. data. For C,H-correlation data, the corresponding value is 1.25. If a structure is ranked number one in both simulations, increased credit is given to that structure. This also requires that no coarse approximations are used. For some structures, the difference between the $\Delta\delta$-sums may be small, as similar or identical stereochemistry may occur for several of the linkages involved.

Structure determination with CASPER. — The procedure for the determination of polysaccharide structures will be exemplified by using four polysaccharides with known structures. Each of them will present a new type of complexity and illustrate how different data can be used in the calculations.

Shigella flexneri *type Y (a linear polysaccharide).* — The complete assignments of the ^{1}H and the ^{13}C resonances for the *S. flexneri* type Y O-polysaccharide were recently published[17,19]. The polysaccharide has structure **3**.

→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→
RhaI RhaII RhaIII GlcNAc

3

If information on components and linkages is given to the program, a total number of 96 suggested structures results after permutation of the components and generation of all possible combinations of anomeric configurations (Table I). Using the $^{3}J_{H,H}$ values (103 see above), the number is reduced to 48, and all structures containing an α-D-Glc*p*NAc residue are omitted as no anomeric proton signal has a $^{3}J_{H,H}$ value of 3–4 Hz. If also the $^{1}J_{C,H}$ values are used (13), the number of suggested structures is decreased to 6, as the structures containing β-L-Rha*p* residues are omitted (Table I). As two →2)-α-L-Rha*p*-(1→ residues are present in the repeating unit, the actual number of structures is three. By comparison of the simulated spectra with the experimental ^{13}C-n.m.r. spectrum (Fig. 2), a best-fitting structure is obtained. The order of the three possible structures is given in Scheme 1 and the correct structure has the best fit. The $\Delta\delta$-value for the second best suggestion is ~40% higher than that for the correct structure. A comparison between the non-sorted simulated spectrum and the assigned experimental spectrum[17] shows that

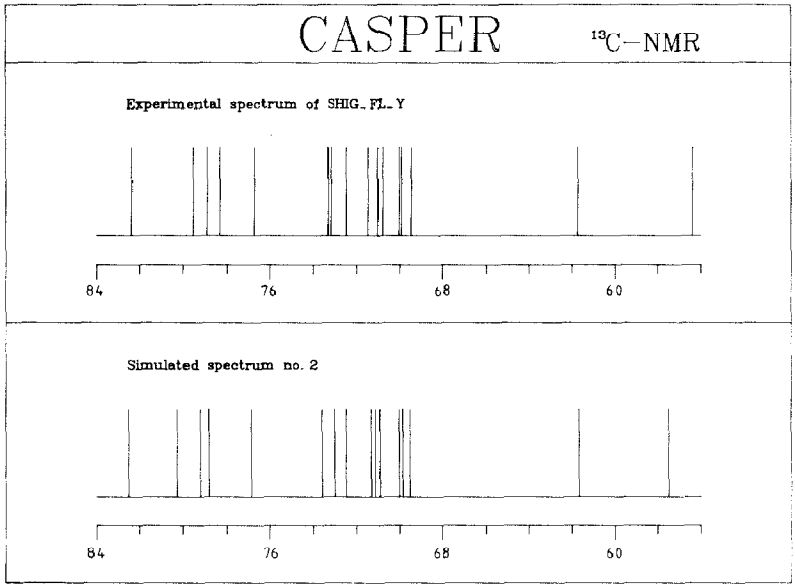


Fig. 2. Graphical output of the "ring carbon region" in the ^{13}C-n.m.r. spectrum of the *S. flexneri* type Y polysaccharide (upper) and the corresponding part of the simulated spectrum with the best fit (lower).

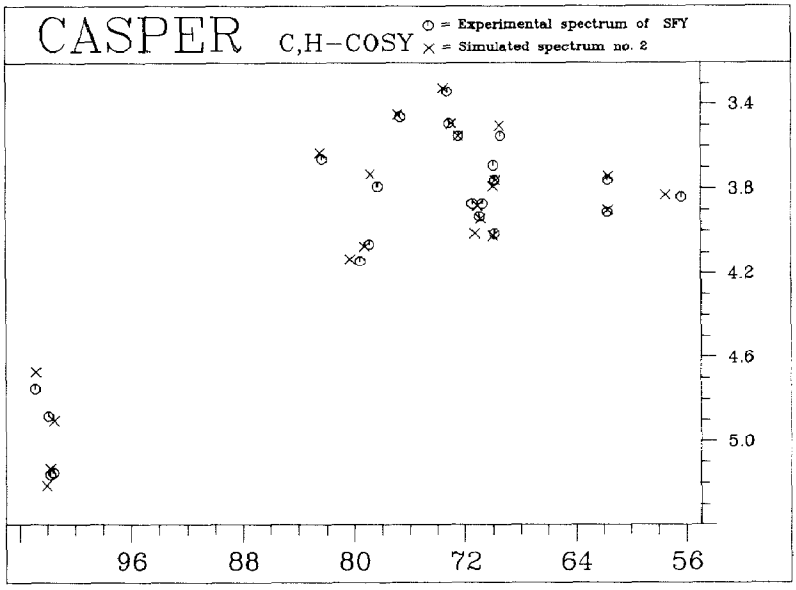


Fig. 3. Graphical output of a part of the C,H-COSY spectrum of the *S. flexneri* type Y polysaccharide (O) and the simulated spectrum with the best fit (X).

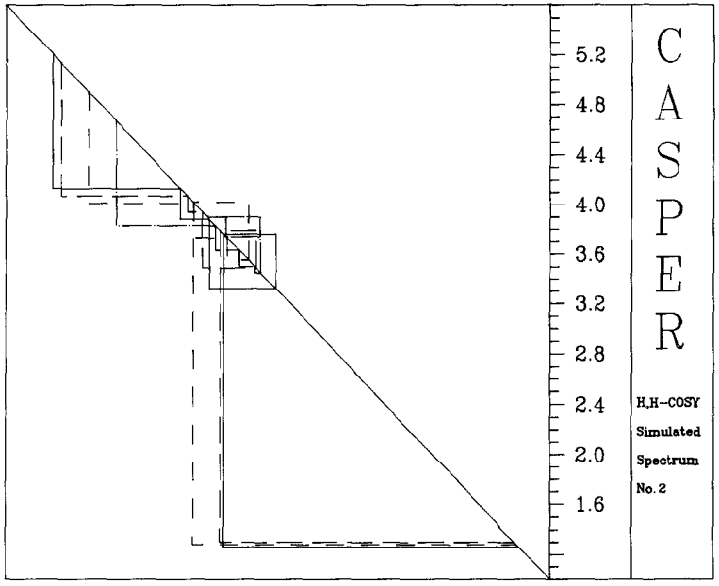


Fig. 4. Graphical output of the simulated H,H-COSY spectrum of the highest ranked *S. flexneri* type Y structure. The spin-coupling connectivity pathway for each spin system is depicted using different lines; alternatively, cross-peaks could be shown by markers. Only half of the cross-peaks were used in order to reduce overlap.

the largest deviations derive from C-2 and C=O in the β-D-GlcNAc residue and the linkage carbons in the flanking α-L-Rha*p* residues. This situation originates from the fact that glycosylation shifts obtained from disaccharides containing D-Glc*p* residues are used and that no corrections have been done for the acetamido group. In order to obtain additional evidence for the suggested structure, the complete C,H-COSY spectrum is given to CASPER. The fitting procedure gives a ranked list (Scheme 2) in which the order of the structures is identical to that from the run using only ^{13}C-n.m.r. data, but the fitting values are now different. A graphical output of the experimental spectrum and the simulated spectrum with the best fit is shown in Fig. 3. The correct structure is 1.7 units lower than the second best structure, and the third is another 1.5 units higher. It is not possible to compare ^{1}H-n.m.r. data in C,H correlation spectra directly, as the ^{13}C-chemical shifts may be interchanged and then the corresponding ^{1}H signals would be interchanged as well.

The capsular polysaccharide from Streptococcus pneumoniae *type 18A (a branched polysaccharide).* — The structure of this capsular polysaccharide was recently published[20]. It is composed of pentasaccharide repeating-units (**4**) and, in addition, contains a glycerol phosphate group. The structure was elucidated using mainly n.m.r. spectroscopy, and most of the signals in the spectra of the polysaccharide without substituents were assigned. In this study, the polysaccharide without the glycerol phosphate group was used.

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

4

SHIG__FL__Y
No. Polysaccharide.

2 —2ALRHA —2ALRHA —3ALRHA —3BDGLCN—
4 —2ALRHA —3ALRHA —2ALRHA —3BDGLCN—
6 —2ALRHA —2ALRHA —3BDGLCN —3ALRHA—

No.	13C Deltasum	13C Sum /sig	13C Check #
2	5.4	0.21	0.22
4	7.7	0.30	0.22
6	8.1	0.31	1.12

J12 = 103 and JCH = 31 used to eliminate structures

13C Experimental spectrum

174.9	102.9	101.9	101.8	101.6	82.4	79.6	78.9	78.3	76.7
73.3	73.2	72.5	71.5	71.0	70.8	70.0	69.9	69.9	69.5
61.8	56.4	23.2	17.5	17.5	17.3				

Spectrum number 2.

175.5	102.8	102.0	101.8	101.5	82.5	80.3	79.3	78.8	76.9
73.6	73.0	72.5	71.3	71.1	70.9	70.0	70.0	69.9	69.5
61.7	57.5	23.1	17.6	17.4	17.4				

Spectrum number 4.

175.5	102.8	102.8	102.1	100.4	82.5	80.3	79.7	78.7	76.9
73.4	73.2	72.5	71.1	71.0	70.8	70.1	70.0	69.8	69.5
61.7	57.5	23.1	17.5	17.5	17.4				

Spectrum number 2.

102.0	80.3	71.1	73.6	69.9	17.4		
101.8	79.3	70.9	73.0	70.0	17.6		
101.5	71.3	78.8	72.5	70.0	17.4		
102.8	57.5	82.5	69.5	76.9	61.7	23.1	175.5

Scheme 1.

The branching at positions 2 and 4 is not vicinal and no corrections are introduced. The substitution at positions 1 and 2 is vicinal, but model studies of a limited number of substances indicate that smaller deviations from the glycosylation shifts in the respective disaccharides are observed for this type of substitution[21]. However, correction values for this type of substitution will be introduced.

The ^{13}C-n.m.r. data and information on the components and linkages are

SFY
No. Polysaccharide.

2 —2ALRHA —2ALRHA —3ALRHA —3BDGLCN—
4 —2ALRHA —3ALRHA —2ALRHA —3BDGLCN—
6 —2ALRHA —2ALRHA —3BDGLCN —3ALRHA—

No.	C,H Deltasum	13C Check#	1H Check#
2	2.6	0.22	0.22
4	4.3	0.22	0.22
6	5.8	1.12	0.22

J12 = 103 and JCH = 31 used to eliminate structures

Experimental C,H-correlation spectrum.

102.9	4.75	101.9	4.88	101.8	5.16	101.6	5.15	82.4	3.66
79.6	4.14	78.9	4.06	78.3	3.79	76.7	3.46	73.3	3.34
73.2	3.49	72.5	3.55	71.5	3.87	71.0	3.93	70.8	3.87
70.0	3.69	69.9	3.76	69.9	4.01	69.5	3.55	61.8	3.76
61.8	3.91	56.4	3.84	23.2	2.06	17.5	1.31	17.5	1.26
17.3	1.26								

Spectrum number 2.

102.8	4.67	101.5	4.90	102.0	5.21	101.8	5.13	82.5	3.63
80.3	4.13	79.3	4.07	78.8	3.73	76.9	3.45	73.6	3.32
73.0	3.49	72.5	3.55	71.1	3.88	71.3	4.01	70.9	3.94
69.9	3.76	70.0	3.79	70.0	4.02	69.5	3.50	61.7	3.74
61.7	3.90	57.5	3.83	23.1	2.06	17.6	1.30	17.4	1.28
17.4	1.26								

Spectrum number 4.

102.8	4.67	102.8	4.96	102.1	5.27	100.4	5.01	82.5	3.63
80.3	4.13	79.7	3.94	78.7	3.86	76.9	3.45	73.4	3.32
73.2	3.47	72.5	3.57	71.1	3.91	70.0	3.97	71.0	3.78
69.8	3.79	70.1	3.81	70.8	4.14	69.5	3.50	61.7	3.74
61.7	3.90	57.5	3.83	23.1	2.06	17.4	1.29	17.5	1.28
17.5	1.27								

Scheme 2.

given as input data. In order to reduce the number of suggested structures, both the $^3J_{H,H}$ and $^1J_{C,H}$ values are added to delete J-incompatible structures. The four suggested structures with the best fit are shown in Scheme 3, and the first suggested structure is correct. The second structure is similar to the first in most respects; the only difference is the position to which the β-L-Rha residue is linked, namely, to the β-D-Glc*p* residue in structure 1 and to the α-D-Glc*p* residue in structure 2, and no major differences in the simulated spectra should be present. The next two suggested structures have several disaccharide elements that are different from those in the first two structures.

The next part of the output shows that the correct structure has a $\Delta\delta$-sum that

```
     S18A
No. Polysaccharide.
1 —4BDGAL —4ADGLCN—3BLRHA —4BDGLC —
       2
     ADGLC
2 —2BDGAL —4ADGLCN—3BLRHA —4ADGLC —
       4
     BDGLC
3                          —2BDGAL —4ADGLCN—
                              4
     ADGLC —3BLRHA —4BDGLC
4                                    —4BDGAL —
                                        2
     BDGLC —4ADGLCN—3BLRHA —4ADGLC
```

No.	13C Deltasum	13C Sum /sig	13C Check#
1	9.2	0.29	0.50
2	11.5	0.36	0.50
3	11.9	0.37	0.50
4	12.3	0.39	0.50

J12 = 221 and JCH = 23 used to eliminate structures

13C Experimental spectrum.

175.1	104.7	102.3	101.3	98.8	94.7	79.4	78.4	77.6	77.4
76.6	76.4	75.5	75.2	74.7	73.8	73.0	72.8	72.6	72.3
71.5	71.2	70.4	70.1	68.2	61.9	61.7	61.4	61.0	54.2
22.8	17.6								

Spectrum number 1.

175.1	104.5	104.3	101.4	98.6	94.7	80.4	78.8	78.7	77.9
77.5	76.6	75.4	75.3	74.5	73.8	73.0	72.6	72.6	72.3
71.8	71.0	70.5	70.4	68.1	61.8	61.5	61.4	60.9	54.5
22.9	17.4								

Spectrum number 2.

175.1	104.6	104.3	101.4	98.5	94.7	80.4	78.8	78.7	77.8
77.5	76.8	76.8	75.3	74.7	73.5	73.0	72.6	72.2	71.8
71.2	71.0	70.7	70.4	68.1	61.8	61.4	61.4	60.9	54.5
22.9	17.4								

Spectrum number 1.

98.6	72.3	73.8	70.5	72.6	61.5		
104.3	77.5	72.6	78.7	75.3	61.4		
94.7	54.5	70.4	80.4	71.8	60.9	22.9	175.1
101.4	68.1	78.8	71.0	73.0	17.4		
104.5	74.5	76.6	77.9	75.4	61.8		

Scheme 3.

is ~2 p.p.m. lower than the second structure, and somewhat larger $\Delta\delta$-sums are observed for the third and fourth structures. A comparison with assigned experimental data[20] shows that the largest deviations are found for linkage carbons of the disaccharide element β-D-Gal*p*-(1→4)-α-D-Glc*p*NAc, 2.0 (C-1 of β-D-Gal*p*) and

```
    S18A__CH
No. Polysaccharide.
1 —4BDGAL —4ADGLCN—3BLRHA —4BDGLC —
        2
     ADGLC
2                                  —4BDGAL —
                                        2
     BDGLC —4ADGLCN—3BLRHA —4ADGLC
3                        —2BDGAL —4ADGLCN—
                              4
     ADGLC —3BLRHA —4BDGLC
4 —2BDGAL —4ADGLCN—3BLRHA —4ADGLC —
        4
     BDGLC
```

No.	H,C Deltasum	13C Check#	1H Check#
1	5.0	0.50	0.50
2	5.6	0.50	0.50
3	5.7	0.50	0.50
4	5.8	0.50	0.50

J12 = 221 and JCH = 23 used to eliminate structures

Experimental H,C-correlation spectrum

104.7	4.65	102.3	4.67	101.3	4.83	98.8	5.31	94.7	5.07
79.4	4.18	78.4	3.64	77.6	3.63	77.4	3.87	76.6	3.77
76.4	3.66	75.5	3.50	75.2	3.74	74.7	3.37	73.8	3.74
73.0	3.41	72.8	3.88	72.6	4.07	72.3	3.55	71.5	4.11
71.2	3.47	70.4	3.45	70.1	3.99	68.2	4.19	61.9	3.92
61.9	3.80	61.7	3.85	61.7	3.80	61.4	3.80	61.4	3.80
61.0	3.80	61.0	3.80	54.2	3.98	22.8	2.06	17.6	1.34

Spectrum number 1.

104.5	4.62	104.3	4.56	101.4	4.93	98.6	5.28	94.7	5.09
78.7	4.16	80.4	3.73	77.9	3.63	78.8	3.69	77.5	3.73
76.6	3.70	75.4	3.48	75.3	3.72	74.5	3.37	73.8	3.75
73.0	3.34	72.6	3.83	72.6	4.00	72.3	3.54	71.8	4.07
71.0	3.50	70.5	3.44	70.4	3.95	68.1	4.19	61.8	3.91
61.8	3.85	60.9	3.92	61.5	3.83	61.5	3.77	61.4	3.83
61.4	3.66	60.9	3.86	54.5	3.98	22.9	2.04	17.4	1.30

Spectrum number 2.

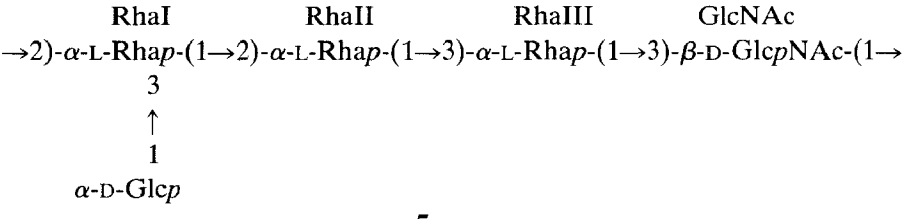
105.7	4.68	103.3	4.52	101.4	4.93	98.5	5.26	94.7	5.09
78.8	4.12	80.4	3.73	78.8	3.69	78.1	3.76	77.8	3.66
76.6	3.53	76.9	3.50	75.2	3.66	74.1	3.33	72.8	3.82
73.0	3.34	73.5	3.93	71.8	4.07	72.2	3.55	71.2	4.04
71.0	3.50	70.5	3.45	70.4	3.95	68.1	4.19	61.7	3.93
61.7	3.74	61.4	3.88	61.6	3.81	61.6	3.66	61.4	3.83
60.9	3.92	60.9	3.86	54.5	3.98	22.9	2.04	17.4	1.30

Scheme 4.

3.0 (C-4 of α-D-Glc*p*NAc) p.p.m., respectively. This may be the result of the 2-*O*-substituent. In order to provide additional evidence for the structure suggested by using ^{13}C-n.m.r. data, the complete C,H-COSY matrix is given as input. The result is shown in Scheme 4, and again the correct structure is found as number one.

Structures numbered four, three, and two from the preceding list now appear as suggested structures numbers two, three, and four. The difference values are relatively similar, however, but the combined evidence strongly favours the correct structure.

The O*-polysaccharide from* S. flexneri *type X (a branched polysaccharide with vicinal disubstitution).* — Both the ^{1}H- and the ^{13}C-n.m.r. spectrum for the *S. flexneri* type X polysaccharide have been assigned[17,19]. The polysaccharide has structure **5**.

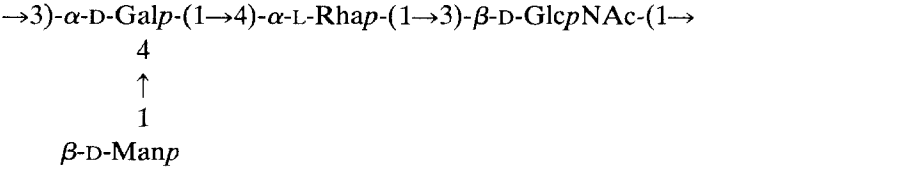


5

When using the ^{13}C-n.m.r. chemical shift data, and values for $^{3}J_{H,H}$ and $^{1}J_{C,H}$ as spectral input, the simulation results in the ranked list given in Scheme 5. The correct structure has the best fit and is closely followed by a structure that has RhaII and RhaIII interchanged. Structure number three has a considerably higher $\Delta\delta$-sum. No large differences between the experimental and the best simulated spectrum are observed.

If the C,H-correlation spectrum is given as additional input, the simulation yields the result shown in Scheme 6. Again, the correct structure is selected as suggested structure number one, followed by the same structure chosen when the ^{13}C-n.m.r. data were used. The differences from the experimental spectrum are ~4 and ~5 units, respectively. The combined evidence thus strongly indicates suggested structure number one to be correct.

The O*-polysaccharide from* Escherichia coli *O75 (a branched polysaccharide with vicinal disubstitution).* — The structure **6** was originally determined using mainly chemical degradations and information on ^{1}H-n.m.r. chemical shifts and $^{3}J_{H,H}$ values of signals from anomeric protons[22].



6

As a first attempt, the ^{13}C-n.m.r. chemical shift data and information on the components and linkages are given as input. The result of the CASPER simulation

SHIG_FL_X
No. Polysaccharide.

1 —2ALRHA —2ALRHA —3ALRHA —3BDGLCN—
3
ADGLC
2 —2ALRHA —3ALRHA —2ALRHA —3BDGLCN—
3
ADGLC
3 —2ALRHA —3ALRHA —3BDGLCN —2ALRHA —
3
ADGLC
4 —2ALRHA —3BDGLCN—2ALRHA —3ALRHA —
3
ADGLC

No.	13C Deltasum	13C Sum /sig	13C Check#	C-Branch Check#
1	9.5	0.30	0.32	1.00
2	10.6	0.33	0.32	1.00
3	14.1	0.44	0.32	1.00
4	14.3	0.45	0.32	1.00

J12 = 113 and JCH = 41 used to eliminate structures

13C Experimental spectrum.

174.6	102.4	102.0	101.9	101.4	95.6	82.5	79.2	78.2	76.9
75.1	74.9	74.1	73.2	72.5	72.4	72.3	71.8	71.5	70.9
70.6	70.3	69.9	69.9	69.5	61.8	61.4	56.2	23.4	17.6
17.6	17.3								

Spectrum number 1.

175.5	101.8	101.8	101.8	101.5	96.5	82.5	79.3	78.8	76.9
75.9	75.9	74.0	73.0	73.0	72.5	72.3	72.0	71.3	70.9
70.7	70.0	70.0	69.9	69.5	61.7	61.6	57.5	23.1	17.6
17.4	17.4								

Spectrum number 2.

175.5	102.8	101.9	101.8	100.4	96.5	82.5	79.7	78.7	76.9
76.0	75.9	74.0	73.2	73.0	72.5	72.3	71.8	71.0	70.8
70.7	70.1	70.0	69.8	69.5	61.7	61.6	57.5	23.1	17.5
17.4	17.4								

Spectrum number 1.

96.5	72.3	74.0	70.7	73.0	61.6		
101.8	75.9	75.9	72.0	69.9	17.4		
101.8	79.3	70.9	73.0	70.0	17.6		
101.5	71.3	78.8	72.5	70.0	17.4		
101.8	57.5	82.5	69.5	76.9	61.7	23.1	175.5

Scheme 5.

is given in Scheme 7 and only the four best suggested structures are shown. All the structures are compatible, since *J*-constraints were applied. The first two structures are similar, in that the branched residue is substituted by β-D-glycosyl residues in both cases, the difference being that the terminal group is linked to either position

SFX
No. Polysaccharide.

1 —2ALRHA —2ALRHA —3ALRHA —3BDGLCN—
3
ADGLC
2 —2ALRHA —3ALRHA —2ALRHA —3BDGLCN—
3
ADGLC
3 —2ALRHA —3BDGLCN—2ALRHA —3ALRHA —
3
ADGLC
4 —2ALRHA —3ALRHA —3BDGLCN—2ALRHA —
3
ADGLC

No.	C,H Deltasum	13C Check#	1H Check#	C-Branch Check#	H-Branch Check#
1	4.1	0.32	0.32	1.00	1.00
2	5.2	0.32	0.32	1.00	1.00
3	6.9	0.32	0.32	1.00	1.00
4	9.5	0.32	0.32	1.00	1.00

J12 = 113 and JCH = 41 used to eliminate structures

Experimental C,H-correlation spectrum.

102.4	4.81	102.0	5.10	101.9	4.86	101.4	5.16	95.6	5.16
82.5	3.50	79.2	4.05	78.2	3.77	76.9	3.43	75.1	4.39
74.9	3.93	74.1	3.82	73.2	3.46	72.5	3.53	72.4	4.02
72.3	3.71	71.8	3.36	71.5	3.84	70.9	3.91	70.6	3.48
70.3	3.72	69.9	3.74	69.9	4.00	69.5	3.51	61.8	3.75
61.8	3.86	61.4	3.75	61.4	3.82	56.2	3.85	23.4	2.09
17.6	1.23	17.6	1.30	17.3	1.25				

Spectrum number 1.

101.8	4.80	101.8	5.13	101.5	4.90	101.8	5.20	96.5	5.13
82.5	3.63	79.3	4.07	78.8	3.73	76.9	3.45	75.9	4.30
75.9	3.90	74.0	3.79	73.0	3.49	72.5	3.55	71.3	4.01
72.3	3.59	72.0	3.44	73.0	3.91	70.9	3.94	70.7	3.43
70.0	3.79	69.9	3.78	70.0	4.02	69.5	3.50	61.7	3.74
61.7	3.90	61.6	3.76	61.6	3.84	57.5	3.83	23.1	2.06
17.4	1.27	17.6	1.30	17.4	1.28				

Spectrum number 2.

101.8	4.80	100.4	5.01	102.8	4.96	101.9	5.26	96.5	5.13
82.5	3.63	79.7	3.94	78.7	3.86	76.9	3.45	75.9	4.30
76.0	3.93	74.0	3.79	73.2	3.47	72.5	3.57	73.0	3.91
72.3	3.59	71.8	3.44	71.0	3.78	70.0	3.97	70.7	3.43
70.1	3.81	69.8	3.81	70.8	4.14	69.5	3.50	61.7	3.74
61.7	3.90	61.6	3.76	61.6	3.84	57.5	3.83	23.1	2.06
17.5	1.28	17.4	1.29	17.4	1.28				

Scheme 6.

ECO75
No. Polysaccharide.

```
1 —4ADGAL —4ALRHA —3BDGLCN—
        3
    BDMAN
2 —3ADGAL —4ALRHA —3BDGLCN—
        4
    BDMAN
3                        —4ADGAL—
                              3
    BDMAN —4ALRHA —3BDGLCN
4            —3ADGAL —4ALRHA —
                    4
    BDMAN —3BDGLCN
```

No.	13C Deltasum	13C Sum /sig	13C Check#	C-Branch Check#
1	5.6	0.21	0.40	0.01
2	6.0	0.23	0.40	0.01
3	8.1	0.31	0.40	0.01
4	8.9	0.34	0.40	0.01

J12 = 112 and JCH = 22 used to eliminate structures

13C Experimental spectrum.

174.9	102.4	101.8	101.3	100.4	82.6	81.6	78.2	77.0	77.0
76.8	74.1	71.8	71.4	71.1	70.1	69.9	69.1	69.0	68.1
62.2	62.0	61.3	56.6	23.1	17.8				

Spectrum number 1.

175.5	102.5	101.2	101.1	100.4	82.6	82.2	78.9	77.2	77.1
76.7	73.9	71.7	71.4	71.2	69.9	69.8	68.9	68.7	67.8
62.1	62.0	61.3	57.5	23.1	17.8				

Spectrum number 2.

175.7	102.4	101.2	101.0	100.5	82.3	82.2	78.8	76.9	76.7
76.6	74.1	71.7	71.4	71.1	69.9	69.8	68.9	68.7	67.8
62.1	61.9	61.3	57.2	23.1	17.8				

Spectrum number 1.

102.5	71.4	73.9	67.8	77.1	62.1		
100.4	68.7	78.9	77.2	71.2	61.3		
101.2	71.7	69.9	82.2	68.9	17.8		
101.1	57.5	82.6	69.8	76.7	62.0	23.1	175.5

Spectrum number 2.

101.0	71.4	74.1	67.8	76.9	62.1		
100.5	68.7	78.8	76.6	71.1	61.3		
101.2	71.7	69.9	82.2	68.9	17.8		
102.4	57.2	82.3	69.8	76.7	61.9	23.1	175.7

Scheme 7.

3 or 4. Structures three and four differ in the side chains. The correct structure is number two with a $\Delta\delta$-sum that is only 0.4 p.p.m. higher than that of structure number one. Structure number three has a value that is 2.1 p.p.m. higher and consequently of correspondingly lower probability. Thus, at this stage, the two first spectra are consistent with spectral data. That no significant difference for the $\Delta\delta$-sum for the two structures was obtained may be due to lack of adequate correction values for the branch-point region. In both cases, the same correction set was used, namely, that from a D-galactosyl residue substituted with two β-D-glucosyl groups. More adequate $\Delta\delta$-values for the disaccharide elements would significantly improve the analysis.

Extension of the input to the C,H-correlation spectrum gives a ranked list with the same difference values for the first two suggested structures (Scheme 8). A comparison between experimental data and simulated spectrum number two shows that ^{13}C-n.m.r. chemical shift differences up to 0.6 p.p.m. and proton chemical shift differences up to 0.13 p.p.m. are present. The differences are not confined to a few signals but spread over several. The identical difference values between the two first suggested structures show how difficult a prediction can be without proper values for glycosylation shifts and branch-point corrections. In this example, an n.O.e. experiment or a specific degradation should give the answer. It could for instance be *N*-deacetylation followed by acid hydrolysis, whereby the disaccharide β-D-Glc*p*N-(1→x)-D-Gal is formed, where x = 3 or 4. Knowledge of that particular linkage would exclude one of the top-ranked structures and thus completely determine the structure. In this study, it was inferred from a NOESY spectrum of the polysaccharide that structure two, identical to the published structure, was correct.

S. flexneri *type Y (the use of ^{1}H-n.m.r. data)*. — The possibility of using only ^{1}H-n.m.r. data is shown for the *S. flexneri* type Y polysaccharide (**3**). The input is thus limited to information on components, linkages, and a chosen number of proton signals. The same set of signals must be taken for all monosaccharide residues, *e.g.*, H-1 to H-4, as was done for this polysaccharide. Such data could, for example, be obtained from a partially assigned H,H-COSY spectrum. The result of the computer simulation is shown in Scheme 9. Only $^3J_{H,H}$ was used as a constraint and it is thus possible that structures containing β-L-Rha*p* residues will be suggested. The list which shows that the four structures with the best fit also contain two structures with β-L-Rha*p* residues, but still the correct structure has the best fit. The connectivity pathway for the simulation with the best fit is depicted in Fig. 4 and is an example of the graphical output that can be obtained in CASPER. The capability for differentiating the correct structure from the next is better than when only ^{13}C-n.m.r. data are used. Thus, the $\Delta\delta$-sum is ~60% higher for the second best structure than for the first structure. Examination of the experimental and the first simulated spectrum indicates that no differences larger than 0.06 p.p.m. are present. In the case when assigned signals are available, comparison with the non-sorted spectrum could easily assist the exclusion of erroneous alternatives. A better way to compare proton chemical shifts, than by decreasing order, should be by

ECO75_CH
No. Polysaccharide.

```
1 —4ADGAL —4ALRHA —3BDGLCN—
        3
    BDMAN
2 —3ADGAL —4ALRHA —3BDGLCN—
        4
    BDMAN
3                        —4ADGAL—
                             3
    BDMAN —4ALRHA —3BDGLCN
4           —3ADGAL —4ALRHA —
                    4
    BDMAN —3BDGLCN
```

No.	C,H Deltasum	13C Check#	1H Check#	C-Branch Check#	H-Branch Check#
1	3.6	0.40	0.40	0.01	0.01
2	3.6	0.40	0.40	0.01	0.01
3	4.6	0.40	0.40	0.01	0.01
4	5.0	0.40	0.40	0.01	0.01

J12 = 112 and JCH = 22 used to eliminate structures

Experimental C,H-correlation spectrum.

102.4	4.82	101.8	4.90	101.3	4.97	100.4	5.08	82.6	3.68
81.6	3.53	78.2	4.04	77.0	4.39	77.0	3.35	76.8	3.45
74.1	3.65	71.8	3.83	71.4	4.06	71.1	4.29	70.1	3.53
69.9	3.86	69.1	3.96	69.0	4.11	68.1	3.58	62.2	3.76
62.2	4.01	62.0	3.76	62.0	4.03	61.3	3.76	61.3	3.76
56.6	3.83	23.1	2.07	17.8	1.34				

Spectrum number 1.

102.5	4.92	101.1	4.88	101.2	4.93	100.4	5.10	82.6	3.60
82.2	3.52	78.9	4.02	77.2	4.43	77.1	3.39	76.7	3.40
73.9	3.68	71.7	3.96	71.4	4.08	71.2	4.24	69.8	3.50
69.9	3.80	68.7	4.10	68.9	4.14	67.8	3.60	62.1	3.77
62.1	3.90	62.0	3.76	62.0	3.91	61.3	3.77	61.3	3.72
57.5	3.82	23.1	2.06	17.8	1.36				

Spectrum number 2.

102.4	4.70	101.2	4.93	101.0	5.09	100.5	5.09	82.3	3.66
82.2	3.52	78.8	3.93	76.6	4.41	76.9	3.32	76.7	3.48
74.1	3.65	71.7	3.96	71.4	4.04	71.1	4.20	69.8	3.49
69.9	3.80	68.7	3.98	68.9	4.14	67.8	3.60	62.1	3.77
61.9	3.90	61.9	3.75	62.1	3.92	61.3	3.76	61.3	3.71
57.2	3.80	23.1	2.04	17.8	1.36				

Scheme 8.

comparing experimental spin systems with those simulated for each structure, but this procedure has not yet been implemented.

Conclusions and future aspects. — The extension of CASPER to the use of more n.m.r. parameters than ^{13}C-n.m.r. chemical shift data gives a larger possibility

SH__FL__Y-H
No. Polysaccharide.

2 —2ALRHA —2ALRHA —3ALRHA —3BDGLCN—
4 —2ALRHA —2ALRHA —3BLRHA —3BDGLCN—
6 —2ALRHA —3ALRHA —2ALRHA —3BDGLCN—
8 —2ALRHA —2ALRHA —3BDGLCN—3BLRHA —

No.	1H Deltasum	1H Sum /sig	JCH	1H Check#
2	0.37	0.02	31	0.22
4	0.61	0.04	22	0.31
6	0.65	0.04	31	0.22
8	0.90	0.06	22	0.31

J12 = 103 used to eliminate structures

1H Experimental spectrum.

5.16	5.15	4.88	4.75	4.14	4.06	3.93	3.87	3.87	3.84
3.79	3.66	3.55	3.55	3.49	3.34				

Spectrum number 2.

5.21	5.13	4.90	4.72	4.13	4.07	3.94	3.93	3.92	3.88
3.78	3.68	3.54	3.54	3.49	3.32				

Spectrum number 4.

5.21	5.13	4.88	4.76	4.13	4.11	4.07	3.94	3.88	3.84
3.75	3.69	3.66	3.50	3.49	3.32				

Spectrum number 2.

5.21	4.13	3.88	3.32	3.76	1.26		
5.13	4.07	3.94	3.49	3.79	1.30		
4.90	3.93	3.78	3.54	4.02	1.29		
4.72	3.92	3.68	3.54	3.46	3.74	3.91	2.06

Scheme 9.

for selecting the correct structure. For compounds for which only ^{1}H-n.m.r. data can be obtained, it is now possible to run a computer analysis. The inclusion of branched structures makes it possible to investigate a large number of carbohydrate structures. For some polysaccharides, the database will not be accurate enough, or the $\Delta\delta$-sum differences will be too small, to differentiate between some structures. Additional chemistry or n.m.r. experiments then have to be undertaken. These can normally be selective, as most of the disaccharide elements in the structure are established through the computer analysis.

Further investigations on model substances will enhance the quality of the database. There is also a large amount of information obtained from known polysaccharide structures which are examined by CASPER. The integration of such information automatically into the database is a challenging next step in the development. This will probably be performed best with an artificial intelligence. Manual input of such data is already underway.

EXPERIMENTAL

^{1}H-N.m.r. (400 MHz) and ^{13}C-n.m.r. spectra of polysaccharides for solutions in D_2O were obtained at 70° with a JEOL GX-400 instrument. Chemical shifts of the ^{13}C resonances were obtained from a ^{13}C-n.m.r. spectrum referenced relative to either internal 1,4-dioxane (δ 67.40), methanol (δ 49.78), or acetone (δ 31.00).

Chemical shifts of overlapping ^{1}H resonances were obtained from the centre of the cross-peaks in the proton–proton correlated spectra (H,H-COSY) and referenced relative to internal sodium 3-trimethylsilylpropanoate-d_4 (TSP). CASPER is written in VAX-11 Pascal. GPGS-F, developed by the Norwegian Association for Computer Graphics at the Computing Centre, University of Trondheim, was used for the graphics, and the pictures were obtained by an HP-7550 plotter. The program was run on a VAX 11/750 computer.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Natural Science Research Council and the Swedish National Board for Technical Development.

REFERENCES

1 J. H. PRESTEGARD, T. A. W. KOERNER, JR., P. C. DEMOU, AND R. K. YU, *J. Am. Chem. Soc.*, 104 (1982) 4993–4995.
2 G. BATTA AND A. LIPTAK, *J. Am. Chem. Soc.*, 106 (1984) 248–250.
3 M. J. GIDLEY AND S. M. BOCIEK, *J. Chem. Soc., Chem. Commun.*, (1985) 220–222.
4 J. F. G. VLIEGENTHART, L. DORLAND, AND H. VAN HALBEEK, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209–374.
5 E. F. HOUNSELL, D. J. WRIGHT, A. S. R. DONALD, AND J. FEENEY, *Biochem. J.*, 223 (1984) 129–143.
6 D. R. ANDERSON AND W. J. GRIMES, *Anal. Biochem.*, 146 (1985) 13–22.
7 D. S. M. BOT, P. CLEIJ, H. A. VAN'T KLOOSTER, H. VAN HALBEEK, G. A. VELDINK, AND J. F. G. VLIEGENTHART, *J. Chemometr.*, 2 (1988) 11–27.
8 P.-E. JANSSON, L. KENNE, AND G. WIDMALM, *Carbohydr. Res.*, 168 (1987) 67–77.
9 G. M. LIPKIND, A. S. SHASHKOV, Y. A. KNIREL, E. V. VINOGRADOV, AND N. K. KOCHETKOV, *Carbohydr. Res.*, 175 (1988) 59–75.
10 D. A. CUMMING, C. G. HELLERQVIST, AND O. TOUSTER, *Carbohydr. Res.*, 179 (1988) 369–380.
11 A. BYRD, W. EGAN, M. F. SUMMERS, AND A. BAX, *Carbohydr. Res.*, 166 (1987) 47–58.
12 H. BAUMANN, P.-E. JANSSON, AND L. KENNE, *J. Chem. Soc., Perkin Trans. 1*, (1988) 209–217.
13 I. BACKMAN, B. ERBING, P.-E. JANSSON, AND L. KENNE, *J. Chem. Soc., Perkin Trans. 1*, (1988) 889–898.
14 P.-E. JANSSON, L. KENNE, AND E. SCHWEDA, *J. Chem. Soc., Perkin Trans. 1*, (1988) 2729–2736.
15 H. BAUMANN, B. ERBING, P.-E. JANSSON, AND L. KENNE, *J. Chem. Soc., Perkin Trans. 1*, submitted.
16 M. FORSGREN, P.-E. JANSSON, AND L. KENNE, *J. Chem. Soc., Perkin Trans. 1*, (1985) 2383–2388.
17 P.-E. JANSSON, L. KENNE, AND T. WEHLER, *Carbohydr. Res.*, 179 (1988) 359–368.
18 H. BAUMANN, B. ERBING, P.-E. JANSSON, AND L. KENNE, *J. Chem. Soc., Perkin Trans. 1*, submitted.
19 P.-E. JANSSON, L. KENNE, AND T. WEHLER, *Carbohydr. Res.*, 166 (1987) 271–282.
20 P.-E. JANSSON, N. S. KUMAR, B. LINDBERG, G. WIDMALM, AND J. HENRICHSEN, *Carbohydr. Res.*, 173 (1987) 217–233.
21 A. ADEYEYE, P.-E. JANSSON, L. KENNE, AND G. WIDMALM, unpublished data.
22 C. ERBING, L. KENNE, B. LINDBERG, AND S. HAMMARSTRÖM, *Carbohydr. Res.*, 60 (1978) 400–403.