ELSEVIER

Carbohydrate Research 306 (1998) 11–17

CARBOHYDRATE RESEARCH

Computer-assisted structural analysis of oligo- and polysaccharides: An extension of CASPER to multibranched structures

Roland Stenutz [a], Per-Erik Jansson [b], Göran Widmalm [a,*]

[a] *Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden*

[b] *Clinical Research Centre, Analytical Unit, Karolinska Institute, Huddinge Hospital, NOVUM, S-141 86 Huddinge, Sweden*

Received 14 July 1997; accepted 8 September 1997

Abstract

The CASPER program which is used for determination of the primary structure of oligo- and polysaccharides has been extended. It can now handle a reduced number of experimental signals from an NMR spectrum in the comparison to the simulated spectra of structures that it generates, an improvement which is of practical importance since all signals in NMR spectra cannot always be identified. Furthermore, the program has been enhanced to simulate NMR spectra of multibranched oligo- and polysaccharides. The new developments were tested on four saccharides of known structure but of different complexity and were shown to predict the correct structures.

Keywords: CASPER; Structure; Computer; Saccharide; Glycopeptide; NMR; Lipopolysaccharide; Capsular polysaccharide

1. Introduction

Carbohydrate structure analysis is a prerequisite for further studies involving carbohydrate–protein interactions which are important, e.g., in protein trafficking, inflammation or bacterial infection processes. The analysis may be divided into the analysis of the primary structure and that of the three-dimensional structure. The former deals with the identity of the constituent monosaccharides, substituents if any, and the sequence of the monosaccharide residues. The latter includes conformation, flexibility and dynamics which are important properties to investigate in order to understand the function of carbohydrates in relation to other molecules, most often proteins. Determination of the primary structure of a saccharide may still be a time-consuming process and computerized approaches are useful in order to speed up the analysis.

We have previously reported a computerized approach [1], viz., CASPER which is an acronym for *C*omputer *A*ssisted *SP*ectrum *E*valuation of *R*egular polysaccharides (vide infra). A similar development which uses only ^{13}C NMR data has been reported by Lipkind et al. [2]. Computerized structure/data-base

* Corresponding author.

0008-6215/98/$19.00

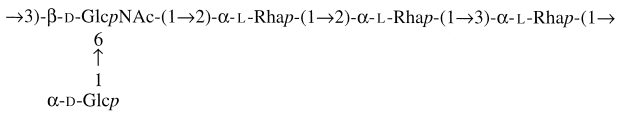
programs using ^{1}H NMR data have been reported both by van Kuik et al. [3] and Hounsell and Wright [4]. The neural network approach to solve structural problems has been put forward for carbohydrates by Meyer et al. [5]. Thus, a number of computerized approaches that complement each other have hitherto been developed in different laboratories.

The CASPER program was developed for the analysis of the primary structures of oligosaccharides and for polysaccharides with repeating units. The approach is based mainly on NMR spectroscopy and has been used in a number of structural determinations, e.g., the structures of the O-antigenic polysaccharides from *Escherichia coli* O 1 [6], *E. coli* O 3 [7], *E. coli* O 101 [8], and the capsular polysaccharide S-156 from a *Klebsiella pneumoniae* strain [9]. The program requires as input (i) sugar components, (ii) their linkage positions, and (iii) an NMR spectrum, preferably carbon-13. Often this is enough to suggest the correct structure. Further credibility to a suggested structure can be obtained by a ^{1}H,^{13}C correlated NMR spectrum. The approach is based on the fact that induced chemical shift differences are observed upon glycosylation of sugars. The NMR spectra are as a first approximation considered to be the result of their constituent disaccharide elements, i.e., the chemical shifts of the monosaccharides and the induced chemical shift displacements for the disaccharides. More complex spectra are observed for branched structures, but the analysis can be performed using additional correction sets for vicinally disubstituted sugar residues. Based on information on components and linkages, NMR spectra are simulated for all possible structures of an oligosaccharide or repeating unit in a polysaccharide. The experimental spectrum is subsequently compared to each of the simulated spectra and a ranking can be obtained based on the fit of spectral data. The ranking is based on a deltasum which is a summation of the absolute values of chemical shift differences between experiment and simulation and a measure of the quality of the fit. When the fit is very good and, at the same time, the structural suggestion ranked second place has a deltasum ≈ 50% higher for ^{13}C simulations, the first structural suggestion can be taken as the correct one. An excellent simulation is observed when the deviation between simulation and experiment is ≈ 0.1 ppm/signal based on ^{13}C NMR data [10]. In a number of cases, more than one structural suggestion show a good fit. These structures can then be differentiated by a judicious choice of an NMR experiment or a specific degradation.

The advantage of the CASPER approach is that a probable structural suggestion can rapidly be obtained from unassigned NMR spectra. Previously, the program required all ^{13}C NMR signals (or a defined number of ^{1}H resonances, e.g., H-1 to H-4 for all sugar residues) to be given as input, something that may be difficult to obtain in heavily overlapped spectra. Furthermore, multibranched structures are often occurring in carbohydrates. The previous versions of CASPER could not handle an incomplete set of ^{13}C NMR data or multibranched structures and prompted further development of the approach.

2. Results and discussion

Simulation with a reduced number of experimental signals.—The structure of the O-polysaccharide of the lipopolysaccharide (LPS) from *Shigella flexneri* type 4a has been determined using chemical methods [11] and the ^{1}H and ^{13}C NMR spectra assigned in more recent studies [12,13]. The polysaccharide has pentasaccharide repeating units with structure.



The correct structure could be simulated by CASPER using the complete ^{13}C NMR spectrum or the two-dimensional ^{1}H, ^{13}C NMR spectrum as previously shown [14]. Some signals were then arbitrarily omitted from the experimental spectrum and a best fit between experimental and simulated data was performed. The fitting procedure for an experimental set with a reduced number of signals uses a recursive algorithm to find the best fit between the experimental spectrum and the spectra for each of the simulated structures. Omission of signals from the experimental ^{13}C NMR spectrum leads to a lower deltasum and to smaller differences between deltasums of simulated structures. A comparison of the deltasums from different structures is shown in Fig. 1 using: all signals, 2, 4 or 6 arbitrarily omitted signals. In these cases, the correct structure was always ranked as number one. When more resonances were omitted, the correct structure did not always turn out as number one, but was still among the top ranked. The approach can thus be considered reliable when a few chemical shifts cannot be identified in the experimental spectrum, a case which is often observed in practice, and

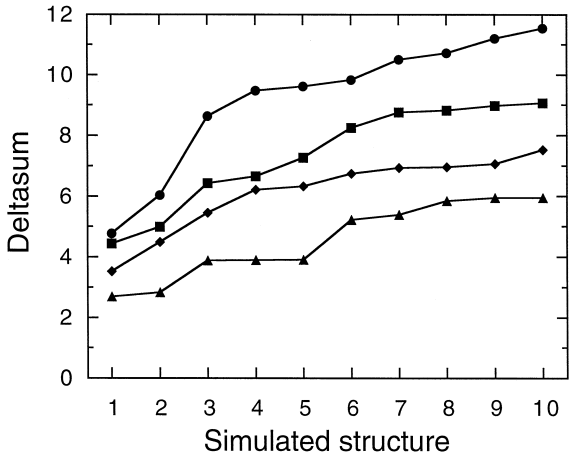


Fig. 1. Comparison of simulated structures vs. deltasum using all 32 (●), 30 (■), 28 (♦) or 26 (▲) resonances in the ^{13}C NMR spectrum of the O-antigenic polysaccharide from *S. flexneri* type 4a.

the procedure has recently been applied in the structural determination of the capsular polysaccharide from *Klebsiella* type 52 [15].

Simulation of a polysaccharide with a doubly-branched residue.—The structure of the polysaccharide of the LPS from an *Aeromonas cavie* strain was recently elucidated [16]. The polysaccharide is composed of pentasaccharide repeating units with structure

α-L-Rha*p*
1
↓
3
→3)-β-D-Gal*p*NAc-(1→6)-β-D-Man*p*NAc-(1→4)-β-D-Glc*p*A-(1→
4
↑
1
β-D-Gal*p*

in which one of the sugar residues, the 2-acetamido-2-deoxy-D-mannose residue, is doubly branched, and thus is 3-, 4- and 6-substituted. As input to CASPER, all ^{13}C NMR signals were used, besides information on sugars and their linkages. The output from the simulation is shown in Scheme 1.

The previously deduced structure is shown to be top ranked with a deviation of 0.34 ppm/signal and a systematic deviation of −0.07 ppm/signal. The differences to the following structures are fairly small, but this is one of the inherent limitations of the CASPER approach. A large systematic deviation (not present in this simulation) would indicate a discrepancy in experimental conditions or in referencing of the chemical shifts. Common for the first four structural suggestions is that the terminal α-L-Rha*p* residue substitutes the 3-position of the β-D-Man*p*NAc branch-point residue.

Simulation of a polysaccharide with two-branched residues.—The capsular polysaccharide from *Klebsiella* K8,52,59 [17,18] was shown to have two branching sugar residues out of the five in the repeating unit, the structure of which is given by

→3)-β-D-Glc*p*-(1→3)-β-D-Gal*p*-(1→3)-α-D-Gal*p*-(1→
4 2
↑ ↑
1 1
β-D-Glc*p*A α-D-Gal*p*

From the ^{13}C NMR spectrum all but two signals could be identified and these were given as input to the CASPER program. The result of the simulation is given in Scheme 2.

In an earlier study, we had simulated the ^{13}C NMR spectrum of the backbone [19]. The structure previously determined is now shown to be the first one after all structures have been ranked according to their respective deltasum. For the correct structure, a deviation of 0.34 ppm/signal and a systematic deviation of 0.16 ppm/signal were observed. It can be noted that in the first four structural suggestions the α-Gal residue is linked to the β-Glc residue and the permutation changes have occurred at the linkage positions of the branched residues.

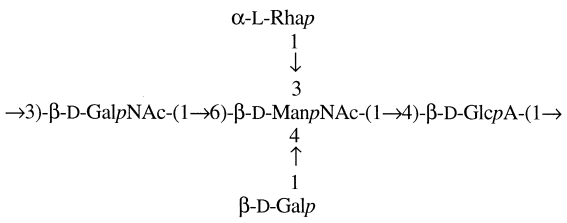
Simulation of a larger oligosaccharide having several branching points.—Oligosaccharides from glycoproteins are commonly multibranched and for such structures the number of permutations rapidly increase with increasing branching and size of the oligosaccharide. One such structure,

α-D-Man*p*
1
↓
6
α-D-Man*p*-(1→3)-α-D-Man*p*
1
↓
6
β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→NAsn
3
↑
1
α-D-Man*p*-(1→2)-α-D-Man*p*

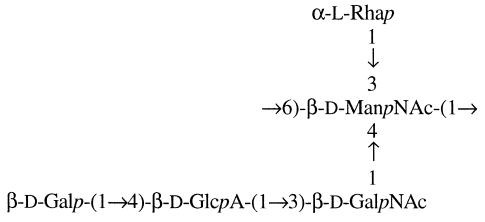
is a high mannose glycopeptide from hen ovalbumin [20]. As input to CASPER, all signals in the ^{13}C NMR spectrum originating from the carbohydrate part were used, together with information on sugars and their linkages. The output from the simulation is shown in Scheme 3.

Aeromonas cavie

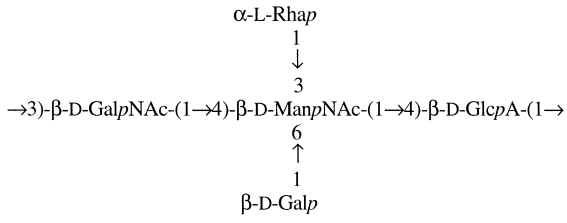
Structure 1



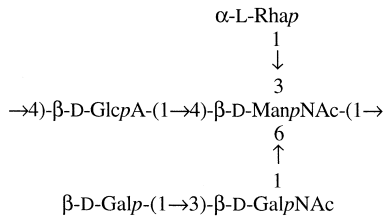
Structure 2



Structure 3



Structure 4



Structure	13C-Deltasum
1	11.6
2	12.7
3	12.9
4	13.1

^{13}C and ^{1}H chemical shifts of simulated structure 1

1	2	3	4	5	6	Me	CO	Residue
96.97	70.95	70.96	73.32	69.21	17.38			α-L-Rhap-(1→
5.15	3.98	3.97	3.48	4.42	1.27			
100.13	51.56	72.19	71.25	75.86	68.21	22.98	176.39	→3,4,6)-β-D-ManpNAc-(1→
4.94	4.79	4.08	4.13	3.81	4.13,4.29	2.06		
102.97	52.72	81.89	68.51	75.66	61.83	23.12	175.78	→3)-β-D-GalpNAc-(1→
4.57	4.17	3.93	4.20	3.75	3.85,3.87	2.06		
104.09	73.89	74.88	81.19	75.53	175.67			→4)-β-D-GlcpA-(1→
4.70	3.49	3.65	3.77	3.86				
102.58	72.14	73.52	69.25	75.73	62.20			β-D-Galp-(1→
4.44	3.49	3.65	3.94	3.59	3.76,3.76			

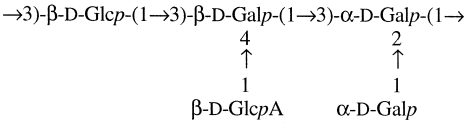
Scheme 1. Result from simulation based on the ^{13}C NMR spectrum from the O-antigen polysaccharide from an *Aeromonas cavie* strain.

The correct structure with a deviation of 0.27 ppm/signal was observed to be top ranked. In this simulation, *J* values were not used in order to limit some possibilities and already for the second structure a significant change has occurred in the core region of the glycopeptide and the structure is no longer in accord with *N*-linked glycans.

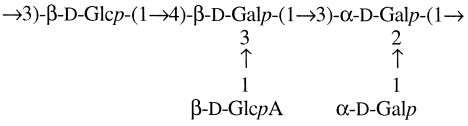
Conclusions.—The CASPER program has been developed to find the best fit to the simulated spectra from an experimental NMR spectrum with a reduced

Klebsiella K8,52,59

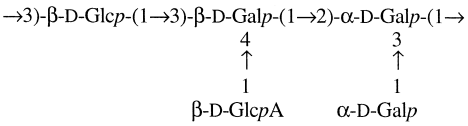
Structure 1



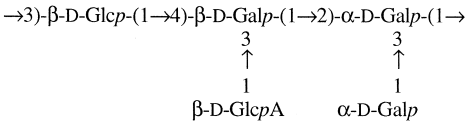
Structure 2



Structure 3



Structure 4



Structure	13C-Deltasum
1	9.5
2	9.7
3	10.2
4	10.3

Experimental ^{13}C spectrum

```
176.50, 104.97, 104.64, 102.67, 96.78, 95.45, 83.51, 82.83, 77.94, 76.64,
 76.51,  75.94,  75.35,  74.62, 74.24, 73.27, 72.70, 71.43, 71.10, 70.94,
 70.84,  70.37,  70.08,  69.16, 62.11, 61.73, 61.54, 60.70
```

^{13}C and ^{1}H chemical shifts of simulated structure 1

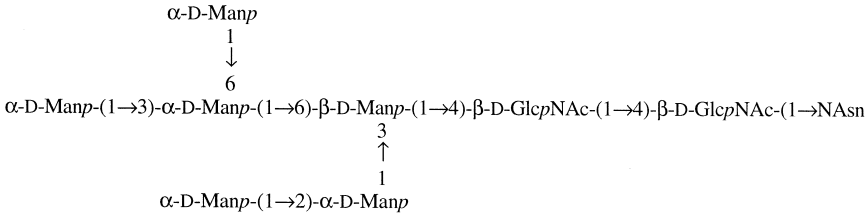
1	2	3	4	5	6	Residue
104.92	71.47	82.94	76.56	75.31	61.29	→3,4)-β-D-Gal*p*-(1→
4.59	3.87	3.83	4.35	3.71	3.62,3.75	
103.22	74.32	76.64	72.82	76.82	176.58	β-D-Glc*p*A-(1→
4.85	3.39	3.51	3.54	3.66		
97.56	73.72	78.62	69.93	71.52	61.69	→2,3)-α-D-Gal*p*-(1→
5.48	4.11	4.09	4.24	4.21	3.71,3.71	
97.76	69.14	70.11	70.11	71.78	61.71	α-D-Gal*p*-(1→
5.06	3.80	3.87	3.98	4.11	3.70,3.70	
104.83	73.32	83.75	71.03	76.53	61.77	→3)-β-D-Glc*p*-(1→
4.69	3.45	3.61	3.61	3.48	3.69,3.91	

Scheme 2. Result from simulation based on the ^{13}C NMR spectrum from the capsular polysaccharide from *Klebsiella* K8,52,59.

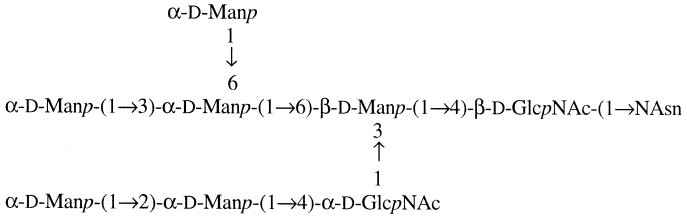
number of signals. Furthermore, the program can simulate polysaccharides and oligosaccharides that are multibranched, e.g., those occurring in glycoproteins. In the cases shown, the correct structures have consistently been selected to be the highest ranked structure. However, the discrimination be-

Glycopeptide

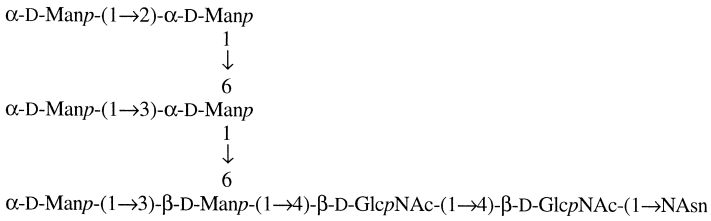
Structure 1



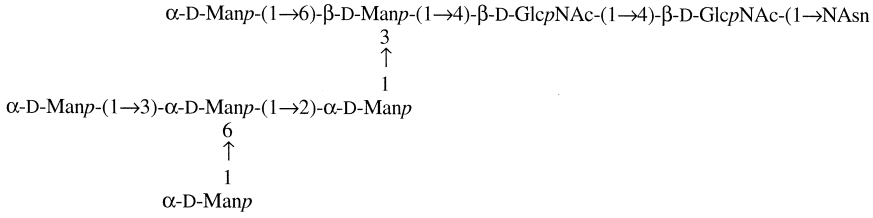
Structure 2



Structure 3



Structure 4



Structure	13C-Deltasum
1	14.3
2	15.4
3	15.4
4	15.9

^{13}C chemical shifts of simulated structure 1

1	2	3	4	5	6	Me	CO	Residue
102.99	70.87	71.33	67.87	74.13	61.95			α-D-Manp-(1→2)
102.97	70.97	71.37	67.80	74.14	61.81			α-D-Manp-(1→3)
101.30	71.65	71.58	67.90	73.74	61.80			α-D-Manp-(1→6)
102.56	80.28	71.14	68.09	74.11	61.95			→2)-α-D-Manp-(1→
101.28	71.18	79.38	67.09	72.40	67.51			→3,6)-α-D-Manp-(1→
100.94	71.08	82.94	66.53	75.66	67.23			→3,6)-β-D-Manp-(1→
102.03	56.54	73.12	79.56	75.58	61.06	23.10	175.49	→4)-β-D-GlcpNAc-(1→
78.88	54.86	73.76	79.68	77.25	60.91	23.07	173.60	→4)-β-D-GlcpNAc-(1→NAsn

Scheme 3. Result from simulation based on the ^{13}C NMR spectrum from a glycopeptide of hen ovalbumin.

tween these multibranched saccharides has not been large and, from the data shown, it would not have been possible to deduce unambiguously the structure, which would need to be corroborated by some additional data. The present developments make the program usable for most structural problems in carbohydrate chemistry that can be addressed by NMR spectroscopy.

3. Methods

The ^{13}C NMR spectrum of the capsular polysaccharide from *Klebsiella* K8,52,59 was recorded at 70 °C on a JEOL 270 MHz NMR spectrometer and referenced relative to internal acetone (δ 31.00). Only readily identifiable resonances were used as input to CASPER. The chemical shifts of the glycopeptide were adjusted and referenced to dioxane (δ 67.40). The chemical shifts of β-D-Glc*p*NAc-(1 → NAsn [21] were adjusted and referenced to acetone.

The CASPER program, version 3.0, was written in C and run on Silicon Graphics workstations. In the comparison with incomplete experimental data, all possible combinations of removal of simulated peaks were generated whereafter a peak-to-peak fit to experimental data was performed. The systematic deviation (ppm/signal) is the average chemical shift difference between signals in a simulated spectrum and those in the experimental spectrum and may thus be positive or negative. A large systematic deviation indicates differences in referencing. The database has been modified to include cross references between glycosylation shifts for different residues. This implementation drastically reduces the size of the database and leads to a facile addition of new monomers. Possible anomeric configurations were not eliminated by use of coupling constants for anomeric protons except for β-D-Glc*p*NAc-(1 → NAsn which was implemented as a single residue.

Acknowledgements

This work was supported by grants from the Swedish Natural Science Research Council.

References

[1] P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 188 (1989) 169–181.
[2] G.M. Lipkind, A.S. Shashkov, Y.A. Knirel, E.V. Vinogradov, and N.K. Kochetkov, *Carbohydr. Res.*, 175 (1988) 59–75.
[3] J.A. van Kuik, K. Hård, and J.F.G. Vliegenthart, *Carbohydr. Res.*, 235 (1992) 53–68.
[4] E.F. Hounsell and D.J. Wright, *Carbohydr. Res.*, 205 (1990) 19–29.
[5] B. Meyer, T. Hansen, D. Nute, P. Albersheim, A. Darvill, W. York, and J. Sellers, *Science*, 251 (1991) 542–544.
[6] H. Baumann, P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 211 (1991) 183–190.
[7] E.C. Medina, G. Widmalm, A. Weintraub, P.A. Vial, M.M. Levine, and A.A. Lindberg, *Eur. J. Biochem.*, 224 (1994) 191–196.
[8] M. Staaf, F. Urbina, A. Weintraub, and G. Widmalm, *Carbohydr. Res.*, 297 (1997) 297–299.
[9] A. Johansson, P.-E. Jansson, and G. Widmalm, *Carbohydr. Res.*, 253 (1994) 317–322.
[10] P.-E. Jansson, L. Kenne, and G. Widmalm, *Acta Chem. Scand.*, 45 (1991) 517–522.
[11] L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen, and E. Romanowska, *Eur. J. Biochem.*, 91 (1978) 279–284.
[12] P.-E. Jansson, L. Kenne, and T. Wehler, *Carbohydr. Res.*, 166 (1987) 271–282.
[13] P.-E. Jansson, L. Kenne, and T. Wehler, *Carbohydr. Res.*, 179 (1988) 359–368.
[14] P.-E. Jansson, L. Kenne, and G. Widmalm, *J. Chem. Inf. Comput. Sci.*, 31 (1991) 508–516.
[15] R. Stenutz, B. Erbing, G. Widmalm, P.-E. Jansson, and W. Nimmich, *Carbohydr. Res.*, 302 (1997) 79–84.
[16] M. Linnerborg, G. Widmalm, M.M. Rahman, P.-E. Jansson, T. Holme, F. Qadri, and M.J. Albert, *Carbohydr. Res.*, 291 (1996) 165–174.
[17] P.-E. Jansson, B. Lindberg, G. Widmalm, G.G.S. Dutton, A.V.S. Lim, and I.W. Sutherland, *Carbohydr. Res.*, 175 (1988) 103–109.
[18] B. Erbing, P.-E. Jansson, G. Widmalm, and W. Nimmich, *Carbohydr. Res.*, 273 (1995) 197–205.
[19] P.-E. Jansson, L. Kenne, and G. Widmalm, *Carbohydr. Res.*, 168 (1987) 67–77.
[20] A. Allerhand and E. Berman, *J. Am. Chem. Soc.*, 106 (1984) 2412–2420.
[21] B.W. Dijkstra, J.F.G. Vliegenthart, G. Strecker, and J. Montreuil, *Eur. J. Biochem.*, 130 (1983) 111–115.