

Applications of the CASPER Program: Comparison between Experimental and Simulated Spectra and an Enhancement Procedure for the Database

Per-Erik Jansson, Lennart Kenne and Göran Widmalm

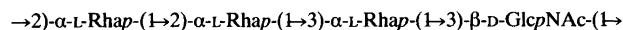
Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

Jansson, P.-E., Kenne, L. and Widmalm, G., 1991. Applications of the CASPER Program: Comparison between Experimental and Simulated Spectra and an Enhancement Procedure for the Database. *Acta Chem. Scand.* 45: 517–522.

CASPER, a computer program by which structural analysis of polysaccharides can be performed, has been tested for accuracy. This was done by simulation of the ^1H and ^{13}C NMR spectra of the *Shigella flexneri* type Y O-polysaccharide using data from the constituent disaccharides and then comparing these with experimental data. A close fit between experimental and calculated ^1H and ^{13}C NMR spectra was observed. Conformational and NMR studies on the disaccharides, as methyl glycosides, were also performed. Furthermore, an extraction procedure, aimed at improving the database in CASPER is demonstrated.

NMR spectroscopy is one of the most important techniques for structural and conformational studies of oligo- and poly-saccharides. By using the combination of NMR spectroscopy and the computer program CASPER^{1,2} structural studies of poly- and oligo-saccharides can be facilitated. The program includes, *inter alia*, a database consisting of chemical shift data arising from studies of mono-, di-, and tri-saccharides.³ In these studies simulation of spectra of homopolymers have successfully been made from data of the corresponding disaccharide elements.

In order to study the fit between simulated and experimental chemical shifts in a spectrum of a heteropolysaccharide using data of the corresponding disaccharide elements, *Shigella flexneri* type Y O-polysaccharide (SFY) was chosen as a model as its ^1H and ^{13}C NMR spectra have been assigned.^{4,5} The structure of the repeating unit of SFY⁶ is



The ^1H and ^{13}C NMR spectra of the four constituent disaccharides as methyl glycosides were first assigned. To investigate whether differences in conformation around the glycosidic bonds occur, the calculated dihedral angles, φ/ψ , of the disaccharides were compared with those calculated for the polysaccharide.

NMR studies on various oligosaccharides which are constituents of the repeating unit of SFY have previously been performed.^{7,8} Using NMR spectroscopy and HSEA energy calculations Bock *et al.* extensively studied the conformations of a number of oligosaccharide elements of SFY which were joined to a C_9 -linker arm. They concluded that the oligosaccharides and SFY had similar conformations.

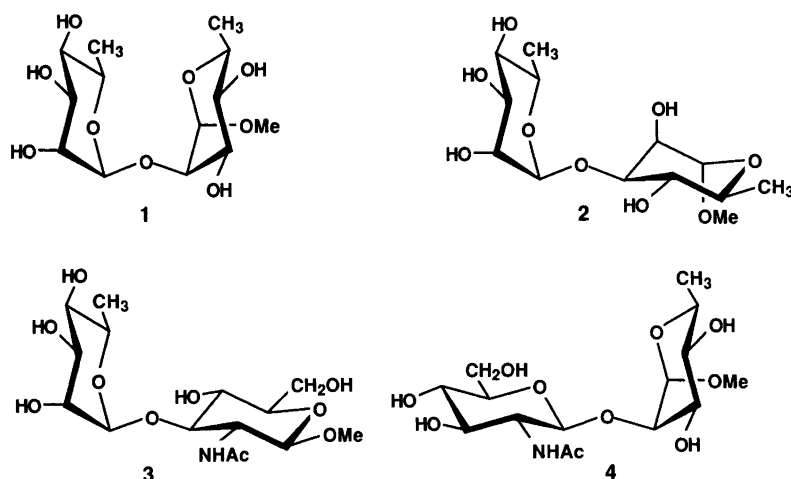
In NMR studies of polysaccharides, often most and sometimes all signals in the spectra are assigned.⁹ A simple method of acquiring NMR data from these spectra of elements not earlier studied would be of great value. Thus an extraction procedure to obtain such data from the spectra of oligo- or poly-saccharides is now described.

Experimental

Disaccharides. Disaccharide **1** was provided by Dr. M. Szönyi and **2** by Mr. C. Hällgren (Stockholm University, Sweden). Disaccharides **3** and **4** were provided by Dr. D. R. Bundle (National Research Council, Ottawa, Canada).

NMR spectroscopy. NMR spectra were recorded for solutions in D_2O using JEOL GSX-270 or GX-400 instruments. For the assignment of signals in the spectra, different proton-proton and carbon-proton shift correlation experiments (COSY) were used. ^1H NMR chemical shifts of overlapping signals were obtained from the centre of the cross-peaks in the H,H-COSY spectra.

Energy calculations. To estimate minimum energy conformations, the GESA program¹⁰ was used which accounts for non-bonded interactions, as expressed by the Kitai-rodsky algorithm, together with a term for the exoanomic effect. The atoms in the glycosyl group are denoted by a prime. The torsional angles φ and ψ were defined for a disaccharide by $\text{H}(1')\text{-C}(1')\text{-O}(\text{X})\text{-C}(\text{X})$ and $\text{C}(1')\text{-O}(\text{X})\text{-C}(\text{X})\text{-H}(\text{X})$, respectively, where X is the linkage position. The bond angle τ [$\text{C}(1')\text{-O}(\text{X})\text{-C}(\text{X})$] was set as 117° . For the hydroxymethyl group of 2-acetamido-



2-deoxy- β -D-glucopyranose the *gauche-trans* conformer [O(5)-C(5)-C(6)-O(6) = 60°] was chosen for the energy minimisations. The co-ordinate set for α -L-rhamnopyranose,¹¹ was obtained from crystal data and that for 2-acetamido-2-deoxy- β -D-glucopyranose from disaccharide data.¹² The co-ordinates for the methyl glycosides were constructed with $\varphi = 50$ and 60° for methyl α -L-rhamnopyranoside and methyl 2-acetamido-2-deoxy- β -D-glucopyranoside, respectively. A tridecasaccharide is large enough to mimic the polysaccharide and was used in the calculation of the minimum energy conformation of the repeating unit. Rotational freedom was allowed at all gly-

cosidic linkages. The φ/ψ -angles of the minimum energy conformation of the disaccharides 1-4 were used as starting points in the minimisation and the tabulated values for the dihedral angles and inter-residue atomic distances were taken from the second repeat in the energy-minimised conformation of the tridecasaccharide.

Results and discussion

GESA calculations. NMR and conformational studies of the four disaccharides are analogous to previous studies.³

Table 1. Values for the φ and ψ angles, in degrees, together with inter-residue atomic distances ≤ 3.1 Å in the minimum-energy conformations of 1-4, as indicated by GESA calculations. Data from the energy-minimised conformation of the *Shigella flexneri* type O-polysaccharide are given in brackets.

Substance	φ/ψ angles	Inter-residue internuclear distances		
		H-1'	H-5'	O-5'
α -L-Rhap(1→2) α -L-RhapOMe (1)	48/14 [49/16]	2.68 (O-3) [2.61] 2.41 (H-2) [2.45]	2.28 (H-1) [2.30]	2.62 (H-2) [2.58]
α -L-Rhap(1→3) α -L-RhapOMe (2)	50/16 [50/11]	2.74 (O-4) [2.86] 2.47 (H-3) [2.42]	2.33 (H-2) [2.27]	2.57 (H-3) [2.60]
α -L-Rhap(1→3) β -D-GlcpNAcOMe (3)	42/15 [37/10]	2.66 (N-2) [2.83] 2.36 (H-3) [2.25] 2.48 (N-H) [2.60]	2.56 (O-4) [2.70]	2.70 (H-3) [2.81]
β -D-GlcpNAc(1→2) α -L-RhapOMe (4)	53/10 [59/-7]	2.77 (O-3) [3.19] 2.44 (H-2) [2.40]		3.10 (H-1) [2.65] 2.56 (H-2) [2.60]

Table 2. ¹H NMR chemical shifts of disaccharides and pertinent monosaccharides obtained at 70 °C relative to internal TSP (δ_H 0.00). Glycosylation shifts^a are given in parentheses.

Substance	H-1' ^b	H-2'	H-3'	H-4'	H-5'	H-6a'	H-6b'	CH ₃ '	H-1	H-2	H-3	H-4	H-5	H-6a	H-6b	CH ₃	OMe
α-L-Rhap-(1→2)α-L-RhapOMe (1)	4.98 (-0.14)	4.07 (0.15)	3.81 (0.00)	3.47 (0.02)	3.77 (-0.09)	1.28 (0.00)			4.78 (0.09)	3.93 (0.00)	3.82 (0.10)	3.47 (0.02)	3.65 (-0.01)	1.31 (0.01)			3.41 (0.01)
α-L-Rhap-(1→3)α-L-RhapOMe (2)	5.04 (-0.08)	4.07 (0.15)	3.84 (0.03)	3.47 (0.02)	3.80 (-0.06)	1.29 (0.01)			4.67 (-0.02)	4.00 (0.07)	3.77 (0.05)	3.55 (0.10)	3.70 (0.04)	1.31 (0.01)			3.41 (0.01)
α-L-Rhap-(1→3)β-D-GlcpNAcOMe (3)	4.87 (-0.25)	3.80 (-0.12)	3.73 (-0.08)	3.44 (-0.01)	3.96 (0.10)	1.24 (-0.04)			4.49 (0.03)	3.78 (0.11)	3.63 (0.07)	3.51 (0.06)	3.47 (0.02)	3.76 (0.01)	3.94 (0.01)	2.05 (0.01)	3.51 (0.00)
β-D-GlcpNAc-(1→2)α-L-RhapOMe (4)	4.71 (-0.01)	3.70 (0.05)	3.57 (0.01)	3.45 (-0.01)	3.43 (-0.03)	3.75 (0.00)	3.91 (0.00)	2.05 (-0.01)	4.84 (0.15)	3.98 (0.05)	3.78 (0.06)	3.33 (-0.12)	3.63 (-0.03)	1.28 (-0.02)			3.39 (-0.01)
α-L-Rhap	5.12	3.92	3.81	3.45	3.86	1.28											
β-D-GlcpNAc	4.72	3.65	3.56	3.46 ^c	3.46 ^c	3.75	3.91	2.06									
α-L-RhapOMe									4.69	3.93	3.72	3.45	3.66	1.30			3.40
β-D-GlcpNAcOMe									4.46	3.67	3.56	3.45 ^c	3.45 ^c	3.75	3.93	2.04	3.51

^aGlycosylation shifts are calculated by subtraction of the chemical shifts of the corresponding hexose and methyl hexoside for the glycosyl part and the aglycone, respectively, and a positive difference indicates a downfield shift. ^bPrimed labels refer to the glycopyranosyl group and unprimed to the methyl glycoside residue. ^cApproximate values.

All φ/ψ-angles and inter-residue atomic distances ≤3.1 Å for the minimum energy conformations of **1–4** and the tridecasaccharide are given in Table 1.

Bock *et al.*⁸ used HSEA-calculations to estimate the conformation of a pentasaccharide as a model for SFY and we arrived at essentially the same values with our tridecasaccharide as they did. The tridecasaccharide was also chosen in order to include long-range effects and it was shown that such effects were present, influencing the conformation.

The differences between the φ/ψ-values in SFY and those in the corresponding disaccharides were ≤6°, except for ψ in disaccharide **4**, β-D-GlcpNAc (1→2)-α-L-Rhap, for which the value of the angle was 17° smaller in the polymer. Decreased distances, >0.1 Å, between atoms in the tridecasaccharide compared with those in the corresponding disaccharides were found between H-1' and H-3 in **3**, and between O-5' and H-1 in **4**. An increase in the distances, >0.1 Å, was found for the pairs H-1'---O-4 in **2**, H-1'---N-2, H-1'---N-H, H-5'---O-4 and O-5'---H-3 in **3** and H-1'---O-3 in **4**.

¹H NMR Glycosylation shifts of **1–4**. The ¹H NMR chemical shifts and the glycosylation shifts (Δδ, chemical shift differences relative to the chemical shifts of the respective monomers) are given in Table 2. Chemical shifts of signals which are not of first order are approximate only. The coupling constants were all of the expected order.

The NMR spectra of other disaccharide alkyl glycosides, analogous to **1–4**, in which the alkyl group is a C₉-carbox-

ylic acid, have been reported.⁸ The differences in chemical shift between those and the ones for our disaccharides were <0.05 ppm with a few values of ca. 0.1 ppm, possibly due to different recording temperatures and different aglycones.

¹³C NMR Glycosylation shifts. The ¹³C NMR chemical shifts for compounds **1–4** and relevant monomers and the glycosylation shifts (Δδ), are given in Table 3.

Fair agreement with data for the C₉-alkyl glycosides is observed with the exception of signals for C-1, C-3, and C-5 which are displaced up to 1.5 ppm, probably owing to the different aglycone.

Temperature dependence of the ¹³C NMR chemical shifts. Changes in chemical shift upon heating is often characteristic of different types of glycosidic linkage, and they can, in favourable cases, be used for assignment of signals.³

The differences in chemical shift for the signals on changing the temperature from 30 to 70 °C are given in Table 3. The values are relative to the signal for internal dioxane, which has been given the same chemical shift, δ 67.40, at both temperatures.

On raising the temperature, most signals were shifted to lower field. No differences were larger than 0.33 ppm, a relatively low value. It should be noted that for three of the disaccharides a negative value for the C-1' signal is observed and for **1** a negative value is also observed for signals from the linkage carbon.

Table 3. ^{13}C NMR chemical shifts of disaccharides and pertinent monosaccharides obtained at 70 °C relative to internal dioxane (δ_{C} 67.40). Glycosylation shifts^a are given in parentheses and chemical shift differences in ppm from variation in temperature^b are given in brackets.

Substance	C-1' ^c	C-2'	C-3'	C-4'	C-5'	C-6'	CH ₂ '	C=O'	C-1	C-2	C-3	C-4	C-5	C-6	CH ₂	C=O	OMe
L-Rhap-(1→2) α -L-RhapOMe (1)	102.91 (8.07) [-0.11]	71.02 (-0.79) [0.12]	71.15 (0.15) [0.25]	73.20 (0.01) [0.22]	69.90 (0.78) [-0.04]	17.45 (-0.22) [-0.01]			100.56 (-1.18) [0.14]	79.12 (8.18) [-0.19]	71.06 (-0.24) [0.21]	73.00 (-0.01) [0.15]	69.40 (0.17) [0.00]	17.55 (0.09) [0.06]			55.74 (0.20) [0.02]
L-Rhap-(1→3) α -L-RhapOMe (2)	102.97 (8.13) [-0.18]	71.08 (-0.73) [0.15]	71.16 (0.16) [0.23]	73.01 (-0.18) [0.17]	69.86 (0.74) [-0.02]	17.46 (-0.21) [0.03]			101.66 (-0.08) [-0.02]	70.73 (-0.21) [0.09]	78.88 (7.58) [0.07]	72.31 (-0.70) [0.08]	69.46 (0.23) [0.06]	17.43 (-0.03) [0.04]			55.58 (0.04) [0.01]
L-Rhap-(1→3) β -D-GlcpNAcOMe (3)	102.13 (7.29) [0.06]	71.60 (-0.21) [0.04]	71.18 (0.18) [0.14]	72.87 (-0.32) [0.17]	69.79 (0.67) [0.07]	17.40 (-0.27) [0.08]			102.18 (-0.58) [-0.14]	55.96 (-0.43) [-0.05]	82.51 (7.61) [0.03]	69.74 (-1.32) [0.33]	76.82 (0.03) [0.00]	61.81 (-0.01) [-0.19]	23.04 (-0.02) [0.07]	174.86 (-0.49) [-0.19]	57.69 (-0.05) [-0.13]
D-GlcNAcp-(1→2) α -L-RhapOMe (4)	103.29 (7.44) [-0.12]	56.79 (-1.07) [0.15]	74.66 (-0.15) [0.21]	70.91 (-0.15) [0.25]	76.70 (-0.12) [0.10]	61.70 (-0.15) [0.23]	23.17 (0.07) [0.04]	175.46 (-0.03) [-0.15]	100.64 (-1.10) [0.08]	79.41 (8.47) [0.00]	70.98 (-0.32) [0.22]	73.33 (0.32) [0.19]	69.31 (0.08) [-0.01]	17.47 (0.01) [0.03]			55.62 (0.08) [0.02]
L-Rhap	94.84	71.81	71.00	73.19	69.12	17.67											
D-GlcpNAc	95.85	57.86	74.81	71.06	76.82	61.85	23.10	175.49									
L-RhapOMe									101.74	70.94	71.30	73.01	69.23	17.46			55.54
D-GlcpNAcOMe									102.76	56.39	74.90	71.06	76.79	61.82	23.06	175.35	57.74

Glycosylation shifts are calculated by subtraction of the chemical shifts of the corresponding hexose and methyl hexoside for the glycosyl part and the aglycone, respectively, and a positive difference indicates a downfield shift. ^b $\Delta\delta = \delta(70^\circ\text{C}) - \delta(30^\circ\text{C})$. Dioxane was taken as δ 67.40 ppm for all temperatures. ^cPrimed labels refer to the glycopyranosyl group and unprimed to the methyl glycoside residue.

When predicting NMR spectra of polysaccharides, problems often occur especially for polysaccharides with vicinally di-substituted branch point residues. For these, as well as for the substituting sugar residues, the simple additivity approach with values only from disaccharide elements does not hold. Thus for these three residues additional glycosylation shifts, called branch point corrections are added when calculating NMR spectra using CASPER.

These values have been obtained from studies of several trisaccharides containing a 3,4-disubstituted galactose residue.^{14,15} In the database, correction values for 2,3- and 3,4-vicinally disubstituted sugar residues are included at the moment. The branch point residue and the substituting sugars may each have the α -D-, α -L-, β -D- or β -L-configuration, but whether the sugars are hexoses, hexosamines, or hexuronic acids has not been taken into consideration.

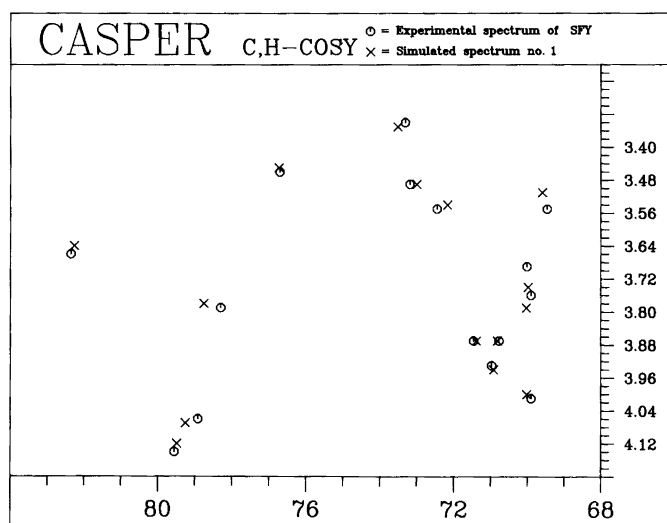


Fig. 1. Graphical output of the experimental C,H-correlation spectrum of SFY (O) and of the simulated spectrum (X).

In order to complement the database with such branch point corrections, the extraction procedure was applied to the ^{13}C NMR spectrum of SFY400 to obtain values for a 3,4-disubstituted residue with the *D*-gluco-configuration. Thus the following values were extracted.

(c)	(b)							
$\alpha\text{-L-Rhap-(1}\rightarrow\text{3)-}\beta\text{-D-GlcpNAc-(1}\rightarrow$								
	4							
	↑							
	1							
	$\alpha\text{-D-Glcp}$	(a)						
	C-1	C-2	C-3	C-4	C-5	C-6	Me	CO
(a)	-4.2	-1.5	-0.2	0.1	0.0	0.0		
(b)	0.9	-0.3	-0.5	-6.7	1.2	0.6	0.3	0.1
(c)	-1.7	0.2	-1.3	0.5	0.6	0.3		

These values are the deviation from the disaccharide additivity approach. Large values are often observed for signals from linkage carbons, almost always negative, i.e. the real glycosylation shift is smaller than with the simple additivity approach. In accordance with this, signals from C-1 in residues (a) and (c), and C-4 in residue (b) are shifted -4.2 , -1.7 and -6.7 ppm, respectively. Furthermore, three values larger than 1 ppm were observed, two of them for signals from carbons next to a linkage.

Extraction of glycosylation shifts is also recommended when the data used in the simulation are approximated. By repeated extraction and averaging from different NMR spectra, the quality of the data will improve. The extraction of chemical shift differences and automatic modification of the database is underway.

In the simulation of NMR spectra of polysaccharides using data from disaccharide elements the accuracy of future studies can be expected to have a mean deviation of ca. 0.1 and 0.01 ppm/signal for ^{13}C and ^1H NMR chemical

shifts, respectively. The extraction procedure developed will constitute an alternative method of obtaining glycosylation shifts to studies using synthetic oligosaccharides.

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

References

- Jansson, P.-E., Kenne, L. and Widmalm, G. *Carbohydr. Res.* 168 (1987) 67.
- Jansson, P.-E., Kenne, L. and Widmalm, G. *Carbohydr. Res.* 188 (1989) 69.
- Jansson, P.-E., Kenne, L. and Ottosson, H. *J. Chem. Soc., Perkin Trans. 1* (1990) 2011 and references therein.
- Jansson, P.-E., Kenne, L. and Wehler, T. *Carbohydr. Res.* 166 (1987) 271.
- Jansson, P.-E., Kenne, L. and Wehler, T. *Carbohydr. Res.* 179 (1988) 359.
- Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E. and Romanowska, E. *Carbohydr. Res.* 56 (1977) 363.
- Thøgersen, H., Lemieux, R. U., Bock, K. and Meyer, B. *Can. J. Chem.* 60 (1982) 44.
- Bock, K., Josephson, S. and Bundle, D. R. *J. Chem. Soc., Perkin Trans 2* (1982) 59.
- Jones, D. N. M. and Sanders, J. K. M. *J. Chem. Soc., Chem. Commun.* (1989) 167.
- Paulsen, H., Peters, T., Sinnwell, V., Lebuhn, R. and Meyer, B. *Liebigs Ann. Chem.* (1985) 489.
- Takagi, S. and Jeffrey, G. A. *Acta Crystallogr., Sect. B* 34 (1978) 2551.
- Mo, F. *Acta Chem. Scand., Ser. B* 33 (1979) 207.
- Wehler, T. and Carlin, N. I. A. *Eur. J. Biochem.* 176 (1988) 471.
- Baumann, H., Erbing, B., Jansson, P.-E. and Kenne, L. *J. Chem. Soc., Perkin Trans 1* (1989) 2153.
- Baumann, H., Erbing, B., Jansson, P.-E. and Kenne, L. *J. Chem. Soc., Perkin Trans 1* (1989) 2167.

Received November 13, 1990.