ELSEVIER

Carbohydrate Research 257 (1994) 35–54

CARBOHYDRATE RESEARCH

Nuclear magnetic resonance and conformational investigations of the pentasaccharide of the Forssman antigen and overlapping di-, tri-, and tetra-saccharide sequences

Gunnar Grönberg [a], Ulf Nilsson [b], Klaus Bock [a,*], Göran Magnusson [b]

[a] *Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark*
[b] *Chemical Center, Organic Chemistry 2, P.O. Box 740, S-221 00 Lund, Sweden*

(Received August 6th, 1993, accepted in revised form October 26th, 1993)

Abstract

The ^{1}H and ^{13}C NMR parameters, i.e., chemical shifts and coupling constants, for the pentasaccharide of the Forssman antigen and overlapping di-, tri-, and tetra-saccharide sequences thereof have been measured and assigned completely using 1D and 2D techniques, and the oligosaccharide structures have thereby been confirmed. Nuclear Overhauser effect (NOE) experiments have been carried out at three different temperatures to assess the preferred conformations of the pentasaccharide and the component oligosaccharides.

The conformational preferences of the compounds mentioned above have subsequently been investigated by theoretical calculations. The flexibility and dynamics of the molecules have been studied by Metropolis Monte Carlo simulations using a modified HSEA force field, and ensemble average data have been generated and compared to data obtained experimentally.

1. Introduction

The Forssman antigen and the related glycolipids globotetraosylceramide (globoside) and globotriaosylceramide are the antigenic determinants of the P blood-group system [1,2]. They have been found to be important cell-surface receptor saccharides for a variety of biologically important proteins such as the

* Corresponding author.

0008-6215/94/$07.00 © 1994 Elsevier Science B.V. All rights reserved
SSDI 0008-6215(93)E0403-N

pilus-associated PapG adhesin protein [3] from *Escherichia coli*, verotoxin [4] (also from *E. coli)*, Shiga toxin [5,6] from *Shigella dysenteriae*, and an adhesin [7] from *Streptococcus suis*. Furthermore, glycolipids of the globo series are tumor-associated on Burkitt lymphoma cells [8] and are enriched in the body fluids of patients suffering from Fabry's disease [9].

Recently the syntheses of the Forssman pentasaccharide as well as its tetra-, tri-, and di-saccharide fragments were reported, together with conformational analysis based on oxygen-induced ^{1}H NMR chemical shifts and molecular mechanics [MM2(91)] calculations [10]. Conformational investigations have been performed on galabiose [α-D-Gal*p*-(1 → 4)-D-Gal*p*] [11,12], deoxy analogues of galabiose [11], the [-β-D-Gal*p*NAc-(1 → 3)-α-D-Gal-] sequence [13], globotriaosylceramide [14], globotetraosylceramide [15–17], globopentaosylceramide (Forssman ceramide) [18], and other glycosylceramides of the globo series [19]. For comparison, the oligosaccharides globotriose and Forssman pentasaccharide were also used in these studies [14,18]. NMR data and conformational analysis of the remaining fragments of the Forssman antigen have not yet been reported. Due to the low solubility of glycosyl ceramides in water, NMR investigations of the ceramides were performed in dimethyl sulfoxide with water added or in water with detergent added. In the present investigation, we have used synthetic, water-soluble glycosides, thus permitting the NMR experiments to be performed in water without additives. We report here the completely assigned ^{1}H and ^{13}C NMR spectral data for this series of compounds and describe their conformational preferences in aqueous solution, using state-of-the-art NMR techniques. The HSEA force field implemented in the GEGOP (geometry of glycoproteins) program [20], using a Metropolis Monte Carlo (MMC) [21] algorithm to search the conformational space of the oligosaccharides, was applied. The MMC approach is a stochastic way to generate conformations of the molecules, enabling simultaneous variation and optimization of the multiparameter system necessary to describe the conformational space of larger biomolecules. Comparisons were made with systematic grid searches for disaccharide fragments, using the same force field.

2. Results and discussion

NMR spectroscopy.—The ^{1}H NMR data, i.e., chemical shifts and coupling constants, of the 2-(trimethylsilyl)ethyl (TMSEt) or methyl glycosides **1–10** (Fig. 1), measured in D_2O at 500 MHz (27°C), are given in Tables 1 and 2, together with the data for methyl glycosides of the corresponding monosaccharides. The assignments were based on different variants of correlation spectroscopy (COSY) experiments to reveal the connectivities from well resolved anomeric signals to other protons in each residue. Nuclear Overhauser effect (NOE) monitoring experiments then served as a tool to verify the assignments made, and also to give information about linkage positions between the different residues. The assignments of shifts and coupling constants for H-5, H-6a, and H-6b were assisted by 1D inversion-recovery experiments [22], distinguishing CH_2 signals from CH signals according to their difference in T_1 relaxation time, thus simplifying the analysis.

1 β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→O)-TMSEt

2 α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→O)-TMSEt

3 β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→O)-Me

4 α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→O)-Me

5 α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→O)-TMSEt

6 β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→O)-TMSEt

7 α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→O)-Me

8 β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→O)-TMSEt

9 α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→O)-TMSEt

10 α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→3)-α-D-Gal*p*-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→O)-TMSEt

units: **e** **d** **c** **b** **a**

Fig. 1. Structures of compounds **1–10**. Letters **a–e** are used in the text and tables to indicate the monosaccharide units. TMSEt is 2-(trimethylsilyl)ethyl.

Table 1
^{1}H NMR chemical shift data for **1-10** [a]

Unit		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Reference monosaccharide (methyl glycoside)
a												
βGlc	H-1	4.50				4.50			4.50		4.50	4.38
	H-2	3.30				3.28			3.29		3.29	3.26
	H-3	3.63				3.62			3.63		3.62	3.48
	H-4	3.65				3.66			3.66		3.66	3.38
	H-5	3.57				3.57			3.57		3.57	3.46
	H-6a	3.97				3.98			3.98		3.99	3.92
	H-6b	3.80				3.83			3.83		3.83	3.72
b												
βGal	H-1	4.45	4.47			4.51	4.48		4.52	4.48	4.52	4.31
	H-2	3.54	3.52			3.58	3.53		3.59	3.53	3.59	3.50
	II-3	3.66	3.71			3.74	3.72		3.75	3.72	3.75	3.64
	H-4	3.92	4.03			4.04	4.03		4.04	4.03	4.04	3.92
	H-5	3.72	3.74			3.79	3.74		3.78	3.74	3.79	3.68
	H-6a	3.78	3.88			3.93	3.87		3.92	3.87	3.92	3.80
	H-6b	3.78	3.83			3.84	3.82		3.84	3.83	3.84	3.74
c												
αGal	H-1		4.97	4.80		4.95	4.94	4.81	4.92	4.94	4.93	4.84
	H-2		3.83	3.88		3.84	3.89	3.90	3.90	3.91	3.91	3.82
	H-3		3.91	3.88		3.90	3.96	3.91	3.96	3.99	3.99	3.78
	H-4		4.03	4.19		4.03	4.26	4.20	4.25	4.27	4.26	3.96
	H-5		4.36	3.90		4.35	4.40	3.91	4.38	4.40	4.39	3.88
	H-6a		3.70	3.73		3.70	3.69	3.73	3.69	3.71	3.69	3.78
	H-6b		3.70	3.73		3.70	3.69	3.73	3.69	3.69	3.68	3.72
d												
βGalNAc	H-1			4.63	4.46		4.63	4.71	4.64	4.71	4.72	4.40
	H-2			3.93	4.04		3.95	4.09	3.94	4.11	4.11	3.89
	H-3			3.74	3.79		3.75	3.81	3.75	3.82	3.82	3.72
	H-4			3.93	4.12		3.94	4.11	3.94	4.11	4.11	3.95
	H-5			3.66	3.65		3.67	3.63	3.68	3.64	3.64	3.63
	H-6a			3.79	3.83		3.80	3.80	3.80	3.81	3.81	3.82
	H-6b			3.75	3.76		3.76	3.74	3.76	3.75	3.75	3.77
	CH_3			2.03	2.05		2.04	2.05	2.04	2.06	2.06	
e												
αGalNAc	H-1				5.07			5.07		5.07	5.07	4.78
	H-2				4.21			4.21		4.22	4.22	4.16
	H-3				3.77			3.79		3.80	3.80	3.86
	H-4				4.01			4.01		4.01	4.01	4.02
	H-5				3.83			3.85		3.86	3.87	3.90
	H-6a				3.77			3.77		3.78	3.78	3.78
	H-6b				3.77			3.77		3.78	3.78	3.72
	CH_3				2.04			2.04		2.04	2.04	

Table 1 (continued)

Unit		1	2	3	4	5	6	7	8	9	10	Reference monosaccharide (methyl glycoside)
Aglycon												
	H-1a	4.04	4.04	3.41	3.53	4.03	4.04	3.41	4.03	4.04	4.04	
	H-1b	3.78	3.77			3.76	3.78		3.76	3.78	3.77	
	H-2a	1.08	1.07			1.07	1.08		1.08	1.07	1.08	
	H-2b	0.98	0.98			0.98	0.98		0.98	0.98	0.98	
	CH_3	0.03	0.03			0.03	0.03		0.03	0.03	0.03	

[a] Measured at 500.13 MHz in D_2O at 27°C, relative to internal acetone (2.225 ppm).

The intraring three-bond proton–proton coupling constants ($^3J_{HH}$) indicate that $^4C_1$ chair conformations are maintained in the monosaccharide units for all compounds (Table 2).

The effect on ^1H chemical shifts upon glycosylation is typically a deshielding of the protons across the glycosidic bond as well as of the protons at the two neighboring sites of the aglycon. The magnitude of deshielding, normally 0.4–0.1 ppm, depends on the type of monosaccharide, anomeric linkage, and conformation around the linkage. The main causes for this deshielding are the steric repulsion between hydrogens and fixation of oxygen lone pairs close in space to the hydrogens in question [23].

The most striking chemical shift effect in this series of oligosaccharides is found in the [-α-D-Gal*p*-(1 → 4)-β-D-Gal*p*-] (galabiose) sequence where H-5 of the α-D-Gal*p* unit is deshielded 0.47 ppm compared to H-5 of the reference monosaccharide and of structures **3** and **7**. This sequence has been thoroughly investigated by NMR and HSEA calculations on galabiose and several analogues, and the occurrence of a van der Waals contact between H-5 of the α-D-Gal*p* residue (**c**) and O-3 of the β-D-Gal*p* residue (**b**) was suggested [10,11,24].

The complete set of oligosaccharides available in this investigation should make it possible to monitor the influence on chemical shift not only for direct glycosylation but also for long-range effects between residues not directly connected. Analysis of the ^1H-chemical shift data (Table 1) reveals, however, that only very small shift changes can be related to long-range effects between residues. One effect is seen in the α-D-Gal*p* (**c**) unit where H-3 is shifted 0.03 ppm downfield in compounds **9** and **10** compared to **6** and **8** (without the terminal α-D-Gal*p*NAc). Another effect is observed for H-5 on the terminal α-D-Gal*p*NAc (**e**), shifting ca. 0.04 ppm, from 3.83 ppm in compound **4** to 3.87 ppm in compound **10**. These effects are probably due to the sterically crowded situation and loss of freedom for the *N*-acetyl group of the β-D-Gal*p*NAc being squeezed between the α-D-Gal*p* and α-D-Gal*p*NAc residues.

The ^{13}C NMR data (125.77 MHz) of compounds **1–10** are presented in Table 3 together with data for the methyl glycosides of the corresponding monosaccharide

Table 2
$^3J_{H,H}$ values (Hz) [a] for compounds **1**–**10**

Unit		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Reference monosaccharide (methyl glycoside)
a												
βGlc	$J_{1,2}$	8.0				8.0			8.0		8.0	8.2
	$J_{2,3}$	9.3				9.3			9.4		9.4	9.6
	$J_{3,4}$	nd [b]				nd			nd		10.0	9.6
	$J_{4,5}$	nd				nd			nd		11.0	9.6
	$J_{5,6a}$	2.3				2.3			2.2		2.5	2.4
	$J_{5,6b}$	5.2				5.0			5.1		5.0	5.8
	$J_{6a,6b}$	12.3				12.5			12.4		12.6	12.8
b												
βGal	$J_{1,2}$	7.8	7.8			7.8	7.8		7.8	7.8	7.8	8.0
	$J_{2,3}$	10.0	10.2			10.4	10.2		10.3	10.2	10.3	10.0
	$J_{3,4}$	3.4	3.4			3.3	3.2		3.0	3.3	3.2	3.8
	$J_{4,5}$	0.6	1.0			1.0	1.0		<1	<1	<1	0.8
	$J_{5,6a}$	4.5	6.6			4.6	5.5		4.5	nd	4.9	4.4
	$J_{5,6b}$	4.5	6.6			7.8	7.3		7.7	nd	7.5	7.6
	$J_{6a,6b}$	nd	11.8			11.4	nd		11.7	nd	11.9	11.2
c												
αGal	$J_{1,2}$		3.9	3.9		3.9	3.9	3.1	3.8	4.0	4.0	3.0
	$J_{2,3}$		10.5	10.5		10.5	10.5	nd	10.5	10.4	10.5	9.8
	$J_{3,4}$		3.3	3.1		3.4	3.1	2.7	3.1	3.2	3.2	2.3
	$J_{4,5}$		1.0	1.0		1.0	1.0	1.2	1.1	1.0	<1	1.0
	$J_{5,6a}$		6.1	6.1		6.1	6.8	6.6	6.5	6.4	6.2	4.6
	$J_{5,6b}$		6.9	7.1		7.1	7.3	6.6	6.5	nd	6.5	8.2
	$J_{6a,6b}$		nd	nd		nd	nd	nd	nd	nd	nd	12.2
d												
βGalNAc	$J_{1,2}$			8.6	8.5		8.6	8.5	8.4	8.4	8.6	8.4
	$J_{2,3}$			10.0	11.0		10.5	11.1	10.9	10.7	10.8	10.5
	$J_{3,4}$			3.3	3.4		3.4	3.3	3.2	3.3	3.2	3.2
	$J_{4,5}$			1.0	1.1		1.0	0.9	<1	1.0	<1	0.8
	$J_{5,6a}$			4.5	5.0		5.5	4.4	4.5	4.6	4.0	4.2
	$J_{5,6b}$			7.7	8.0		7.3	7.9	7.8	7.9	8.0	7.7
	$J_{6a,6b}$			11.8	11.0		11.7	nd	11.9	nd	nd	11.9
e												
αGalNAc	$J_{1,2}$				3.8			3.8		3.7	3.8	3.6
	$J_{2,3}$				11.1			11.1		11.0	11.0	10.8
	$J_{3,4}$				3.4			3.1		3.1	3.0	3.2
	$J_{4,5}$				<1			1.2		1.0	<1	0.8
	$J_{5,6a}$				7.0			6.2		6.0	7.0	6.0
	$J_{5,6b}$				7.0			7.0		7.0	7.0	6.0
	$J_{6a,6b}$				nd			nd		nd	nd	nd

[a] Measured at 500.13 MHz (D_2O, 27°C), first-order analysis with an accuracy of ±0.3 Hz. [b] nd, Not determined.

residues. Assignments were based on 2D heteronuclear correlation spectroscopy and by comparison with data for model compounds. The ^{13}C chemical shifts do not display any significant changes from expected values due to long range contacts or conformational effects and are additive throughout the series of oligosaccharides.

All oligosaccharides **1–10** were analyzed by NOE-difference measurements at 500 MHz. The di- and tri-saccharides, which give positive NOE values under these conditions (500 MHz, 27°C), were analyzed by steady-state experiments and the results could then be compared with theoretical values as described below. The NOE values obtained from the disaccharides are given in Table 4. For the tetrasaccharides and the pentasaccharide, the correlation times gave NOE values close to zero, slightly positive for the tetrasaccharides and negative for the pentasaccharide. Therefore, truncated NOE-difference experiments with a mixing time of 200 ms were performed in order to avoid spin diffusion effects. With the pentasaccharide, experiments were performed at different temperatures (+8, +18, and +27°C) and these data are shown in Table 5. NOESY and ROESY 2D spectra were obtained for the pentasaccharide at 600 and 500 MHz, respectively, in order to compare with the 1D experiments. Finally, 2D NOESY data obtained at 600 MHz are shown in Table 6.

Force field calculations and conformational analysis.—Molecular modelling was performed with the GEGOP program [20], which utilizes a modified HSEA (hard sphere exo anomeric) force field [25,26] to calculate the energies of the conformers obtained. The conformations were generated from grid searches or MMC simulations. All calculations were performed on the corresponding methyl glycosides and the results are presented in Table 7.

In the GEGOP program, the conformations of the carbohydrate molecules are expressed in terms of torsional angles of the glycosidic linkage (ϕ,ψ) and the exocyclic groups, C-5–C-6 (ω) and C-2–N (see Experimental section for definitions), while the pyranose rings are kept rigid during the calculations. For disaccharides, a grid search systematically covering the whole ϕ,ψ-quadrant for the internal glycosidic linkage is a feasible method to generate the energy surface of the molecule, using the computational power available in our laboratory (INDIGO workstations). However, for a multiparameter system like the pentasaccharide, a systematic multidimensional grid search is not realistic. Therefore, an MMC simulation was used in which a stochastic method is employed to sample populations of simulated conformations of a molecule and calculate their energy according to the force field [27]. This methodology can be used to scan the conformational space of more complex molecules by varying all relevant parameters simultaneously and to calculate the energy of the resulting conformations. The MMC method scans the low energy areas only and is also able to overcome energy barriers to avoid the problem of being trapped in local minima. It is possible to monitor distances and dihedral angles from the conformations sampled during the simulation, and ensemble-averaged data can be obtained for calculation of, for example, theoretical NOE data, coupling constants, or average distances [28].

The energy surfaces of the four different disaccharides **1–4** were mapped in a grid search by varying the ϕ and ψ angles of the internal glycosidic bond in

Table 3
^{13}C NMR chemical shift data for **1–10** [a]

Unit		1	2	3	4	5	6	7	8	9	10	Reference monosaccharide (methyl glycoside)
a												
βGlc	C-1	102.2				102.2			102.2		102.2	104.0
	C-2	73.7				73.7			73.8		73.8	74.1
	C-3	75.4				75.4			75.5		75.5	76.0
	C-4	79.2				79,5			79.5		79.5	70.6
	C-5	75.6				75.6			75.6		75.6	76.8
	C-6	60.9				60.9			60.9		60.9	61.8
b												
βGal	C-1	103.7	103.0			104.1	103.0		104.1	103.0	104.1	104.5
	C-2	71.8	71.9			71.8	71.8		71.7	71.8	71.7	71.7
	C-3	73.4	73.4			73.0	73.3		72.9	73.3	73.0	73.8
	C-4	69.4	78.0			78.2	77.7		78.1	77.7	78.1	69.7
	C-5	76.1	75.8			76.2	75.8		76.3	75.7	76.2	76.0
	C-6	61.8	60.9			61.2	61.2		61.1	60.8	61.2	62.0
c												
αGal	C-1		101.1	100.3		101.1	101.1	100.4	101.2	101.1	101.2	100.1
	C-2		69.6	68.0		69.4	68.5	68.0	68.4	68.4	68.4	69.2
	C-3		70.0	80.0		70.0	80.0	80.1	79.5	80.1	79.7	70.5
	C-4		69.8	69.9		69.8	69.7	69.9	69.8	69.7	69.7	70.2
	C-5		71.7	71.1		71.7	71.0	71.1	71.1	71.1	71.1	71.6
	C-6		61.4	61.8		61.3	61.8	62.0	61.2	61.2	61.2	62.2
d												
βGalNAc	C-1			103.9	102.7		104.2	103.5	104.1	103.6	103.6	103.2
	C-2			53.6	51.3		53.4	51.7	53.4	51.7	51.7	53.1
	C-3			71.7	75.6		71.7	75.5	71.6	75.5	75.5	72.0
	C-4			68.6	64.4		69.7	64.4	68.6	64.4	64.4	68.7
	C-5			75.8	75.9		75.8	75.7	75.8	75.7	75.7	75.9
	C-6			60.7	61.8		60.7	61.8	61.8	61.8	61.9	61.8
	C=O			176.0	175.4		176.0	175.8	176.0	175.7	175.8	
	CH_3			23.1	23.0		23.1	23.0	23.1	23.1	23.1	
e												
αGalNAc	C-1				94.1			94.3		94.4	94.6	99.0
	C-2				50.1			50.2		50.2	50.2	50.7
	C-3				68.5			68.5		68.5	68.5	68.5
	C-4				69.1			69.1		69.1	69.1	69.3
	C-5				72.2			72.1		72.1	72.1	71.5
	C-6				61.7			61.7		61.7	61.7	62.0
	C=O				175.4			175.5		175.5	175.5	
	CH_3				22.8			22.8		22.8	22.8	
Aglycon	C-1	69.1	69.1	55.8	57.8	69.3	69.0	55.8	69.3	69.1	69.3	
	C-2	18.5	18.5			18.4	18.5		18.4	18.5	18.4	
	CH_3	−1.7	−1.7			−1.7	−1.7		−1.7	−1.7	−1.7	

[a] Measured at 125.77 MHz in D_2O at 27°C, relative to external dioxane (67.4 ppm).

Table 4
Steady-state NOE-difference data for compounds **1–4** at 27°C (500.13 MHz). Experimental steady-state NOE values and ensemble-averaged NOE values from MMC calculations (underlined)

Compound	Proton saturated	NOE observed (%)					
		Intraring			Interring		
			Exptl	Calcd		Exptl	Calcd
1	H-1b	H-2	3.6	5.5	H-2	nd	−0.8
		H-3	12.2 [a]	4.0	H-4	[a]	7.0
		H-4	−0.3	−0.6	H-6a	1.0	1.9
		H-5	5.8	6.1	H-6b	nd	−0.4
2	H-1c	H-2	14.8 [b]	14.1	H-2	nd	0.2
		H-5	1.6	0.4	H-3	nd	−0.3
					H-4	9.7	8.0
					H-5	nd	−0.2
					H-6a	[b]	1.2
					H-6b	0.7	0.9
3	H-1d	H-2	4.8 [c]	8.8	H-2	[e]	0.4
		H-3	8.4 [d]	7.0	H-3	13.0 [c]	11.4
		H-4	[c]	−1.2	H-4	1.1	0.8
		H-5	6.5	7.4	H-5	nd	−0.7
		H-6a	[d]	−0.1			
		H-6b	[d]	−0.4			
4	H-1e	H-2	12.7	14.3	H-2	nd	0.3
		H-3	nd [f]	0.3	H-3	4.0	3.6
		H-5	nd	0.3	H-4	7.0	6.2
					H-5	nd	−0.4

[a,b,c,d,e] Indicate overlapping signals; the sum of integrals is given for one resonance where indicated.
[f] nd, Not determined.

Table 5
NOE values (measured as NOE-difference experiments at 500.13 MHz) for compound **10** at different temperatures (mixing time 200 ms)

Residue	Proton saturated	NOE observed (%)							
		Intraring				Interring			
			+8°C	+18°C	+27°C		+8°C	+18°C	+27°C
c	H-1	H-2	−8.4	−4.4	−2.8	H-4	−6.4	−3.3	−1.9
						H-3	−0.8	−0.4	nd
						H-6a	−2.5	−1.1	−0.5
						H-6b			
d	H-1	H-2	−2.2	−1.2	−1.7	H-3	−7.0	−3.4	−2.2
		H-3	−3.6	−2.0	−1.0	H-2	−0.8	−0.5	nd
		H-5	−4.2	−2.2	−1.3	H-4	−1.2	−0.7	−0.4
e	H-1	H-2	−4.4	−2.6	−1.6	H-3	−3.0	−1.6	−1.1
		H-3	nd [a]	nd	nd	H-2	−0.4	nd	nd
						H-4	−4.4	−2.4	−1.1

[a] nd, Not determined.

Table 6
2D NOESY data for compound **10** at 27°C (600 MHz) and a mixing time of 200 ms. Experimental 2D NOESY values and ensemble-averaged NOE values from MMC calculation (underlined); in parenthesis is given the ensemble-averaged distance between the atoms considered

Residue	Proton saturated	NOE observed (%)						
		Intraring			Interring			
			Exptl	Calcd	Unit		Exptl	Calcd
a	H-1	H-2	−3.4	−0.6 (3.04)				
		H-3	−1.2	−1.2 (2.57)				
		H-5	−1.2	−0.8 (2.51)				
b	H-1	H-2	−1.6	−0.5 (3.06)	**a**	H-4	−0.6	−1.0 (2.41)
		H-3	−1.8 [a]	−0.8 (2.56)		H-6a		−0.2 (2.70)
		H-5	[a]	−1.0 (2.28)		H-6b	−0.3	−0.2 (2.77)
c	H-1	H-2	−2.0	−2.0 (2.40)	**b**	H-4	−1.2	−1.3 (2.15)
						H-6a	−0.4 [b]	−0.2 (2.77)
						H-6b	[b]	−0.2 (2.67)
	H-5	H-3	−0.7	−0.8 (2.49)		H-2	−0.6	−1.1 (2.69)
		H-4	−1.4	−1.0 (2.46)				
d	H-1	H-2	−1.4	−0.8 (3.06)	**c**	H-2	−0.1	−0.1 (3.96)
		H-3	−1.2	−0.5 (2.60)		H-3	−2.1	−1.1 (2.33)
		H-5	−0.8	−0.8 (2.41)		H-4	−0.2	−0.2 (3.01)
	CH_3				**e**	H-5		−0.3 (2.86)
e	H-1	H-2	−1.9	−2.1 (2.43)	**d**	H-2	−0.1	−0.1 (4.30)
		H-3	nd [c]	−0.1 (3.67)		H-3	−0.9	−0.7 (2.51)
						H-4	−1.1	−1.1 (2.33)

[a,b] Indicate overlapping signals, the sum of integrals is given where stated. [c] nd, Not determined

10-degree steps. The minimum energy conformations found in this way were used as starting points in a further minimization (Table 7). The disaccharides were also subjected to MMC simulations in order to correlate the result with the grid searches (Fig. 2) and to calculate ensemble-averaged theoretical NOEs and coupling constants, to be compared with the experimentally determined data (Table 4). MMC simulations were started from the minimum energy conformations. For

Fig. 2. ϕ/ψ Energy maps of the disaccharide subunits in the Forssman pentasaccharide. **A**, The lower maps shows the results of grid searches in 10° steps of the disaccharides **1**–**4**. The contour levels are set at 0.1, 1, 2, 3, 4, 5, and 10 kcal/mol relative to the low energy conformation. **B**, The middle row of ϕ/ψ energy maps was generated from conformation distributions sampled in MMC simulations (1 000 000 steps, acceptance rate 45%, at 600 K) of the disaccharides **1**–**4**. **C**, The upper row shows the results of an MMC simulation of the Forssman pentasaccharide **10** (1 000 000 steps, acceptance rate 43%, at 800 K). The MMC maps were contoured using a binsize of 6 degrees, and contour levels of 0.1, 1, 2, 3, 4, and 5 kcal/mol relative to the lowest energy conformation are shown.

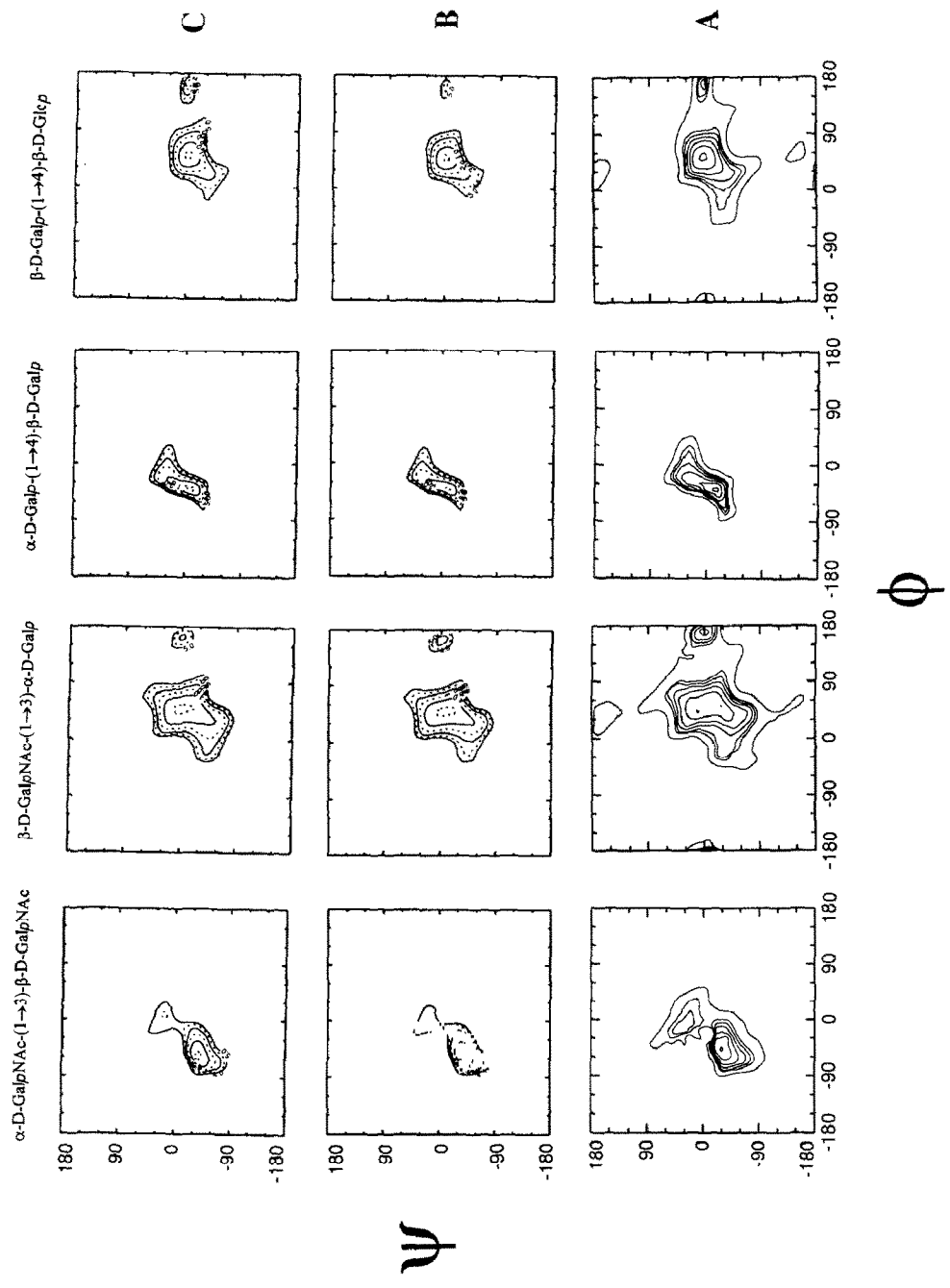
α-D-GalpNAc-(1→3)-β-D-GalpNAc
β-D-GalpNAc-(1→3)-α-D-Galp
α-D-Galp-(1→4)-β-D-Galp
β-D-Galp-(1→4)-β-D-Glcp
A
B
C
Ψ
Φ
180
90
0
-90
-180

Table 7
ϕ/ψ Torsional angles in minimum energy conformation. The half-height widths obtained from 1D population plots derived from MMC simulations of the pentasaccharide and the disaccharides, respectively, are given in italics below each pair of ϕ/ψ angles. These values may serve as a measure of the flexibility the torsional angles experienced during the simulation

α-D-Gal*p*NAc-(1 → 3)-β-D-Gal*p*NAc-(1 → 3)-α-D-Gal*p*-(1 → 4)-β-D-Gal*p*-(1 → 4)-β-D-Glc*p*-(1 → O)Me				
e	**d**	**c**	**b**	**a**
Compound				
10	−43/−26 *(38/25*)* [a]	48/6 *(45/81)*	−39/−15 *(24/65)*	54/1 *(41*/32)*
1				53/4 *(36*/32)*
2			−40/−15 *(22/61*)*	
3		54/−11 *(40*/68)*		
4	−44/−24 *(34/32*)*			
Galabiose [b]			−39/−15	
1–4 [c]	−35/−32	27/−61	−41/−16	45/5

[a] * Indicates more than one distinct minimum for this angle in the 1D population plot. [b] Values from Bock et al. [11] (HSEA). [c] Values calculated for the disaccharide fragments **1–4** with MM2 91 [10]. *Note:* C-5–C-6 groups for units with the galactose configuration are found in *gauche–trans* (*gt*) conformation, while for the glucose residue the *gauche–gauche* (*gg*) conformation is preferred.

comparison, additional calculations were started at randomly chosen angles and the resulting population maps were in agreement.

Subsequently, the pentasaccharide **10** was subjected to conformational analysis. The minimum energy ϕ/ψ angles found for the corresponding disaccharides were used as a starting point for energy minimization. The minimum energy conformations of the glycosidic linkages in the pentasaccharide obtained in this way were similar to those found for the disaccharides as shown in Table 7. The only substantial deviation in the pentasaccharide compared to the individual disaccharide fragments is found for the linkage between units **d** and **c** [-β-D-Gal*p*NAc-(1 → 3)-α-D-Gal*p*-] where there is a change of +6° in ϕ-angle and −17° in ψ-angle (both these values are close to the limits of the accuracy expected for this method). Experimentally this is reflected in the deshielding of H-3 in unit **c** of the pentasaccharide **10** compared to disaccharide **3**, due to closer contact of H-3 with the ring oxygen in **d** for a positive value of the ψ-angle (2.67 and 2.76 Å for the minimum conformations of **10** and **3**, respectively).

To investigate the conformational space of the pentasaccharide, MMC simulations were initiated from the minimum energy conformation. From the sampled conformations, the iso-energy contour maps are calculated and compared to the maps obtained from grid searches and MMC calculations of the disaccharides **1–4** (Fig. 2). The sampled average distances and calculated NOE values together with experimental data are given in Table 6.

The shapes of the iso-energy ϕ/ψ-maps from the grid searches of the disaccharides are well reproduced by the MMC simulations of both the disaccharides and the pentasaccharide suggesting: (*1*) for the disaccharide fragments, the MMC simulations are able to reproduce the conformational space as determined in the grid searches. This result is important due to the fact that for larger oligosaccharides, such as the pentasaccharide **10**, the increasing number of parameters necessary to describe the molecule makes MMC simulations the only feasible way to perform the conformational analysis of the entire molecule, and at the same time obtain time-averaged data, such as distances or NOE values. (*2*) The shapes of the resulting MMC population (or energy) maps for the disaccharide fragments **1–4** are very similar to the ones obtained for the corresponding linkages in the pentasaccharide. Experimentally this was also indicated by the chemical shift consistency found for the whole series of oligosaccharides, indicating no measurable long-range effects upon addition of residues to form the pentasaccharide.

The projections of sampled populations of the ϕ- and ψ-angles are shown in Fig. 3. The width at half-height in Table 7 is presented as an indication of the flexibility experienced by each glycosidic torsion angle during the MMC simulation. Distance distributions between atoms, monitored during the MMC calculation of **10**, are displayed in Fig. 4.

The measured and calculated NOE values are shown in Table 4 for the disaccharides and in Table 5 for the pentasaccharide. The correlation times in the calculations were chosen to give NOE and T_1 values of the same order as the experimentally determined ones, so that experimentally derived and calculated relative NOE values for different hydrogen pairs can be compared. In general, the values for the experimental and calculated data are in good agreement, indicating that the ensemble-averaged NOE data calculated from the conformations generated during the MMC simulation reflect the validity of this approach to describe the conformational preferences of the pentasaccharide. An illustration of the flexibility of the pentasaccharide **10** along with the ϕ/ψ-maps of the individual linkages is found in Fig. 5, and stereoviews of the lowest energy conformer of the pentasaccharide are displayed in Fig. 6. In the following, a detailed discussion of the above analysis will be presented arranged according to the disaccharide fragments.

The α-D-*Gal*p*NAc-(1 → 3)-*β-D-*Gal*p*NAc segment* (**e–d**).—The sampled population distribution of the ϕ/ψ-angles shows an energy minimum around $-40°/-30°$ with a range at half height of (38°/25°). A less populated minimum is observed at $\sim 5°/30°$ at energy levels 3–4 kcal/mol higher than the minimum conformation.

A special feature of this linkage is the observation that an interresidue NOE from H-1**e** to H-4**d** is present, which is equal to or stronger than that observed from H-1**e** to H-3**d** (across the glycosidic bond). This feature was already observed for α-D-Gal*p* and α-D-Gal*p*NAc linked to the 3-position of galactose in blood-group active oligosaccharides [29] and has subsequently been noted to occur for α-D-Gal*p* and α-D-Glc*p* units linked to the 3-position of galactose. This may serve as a warning against the assumption that the strongest NOEs emanate from

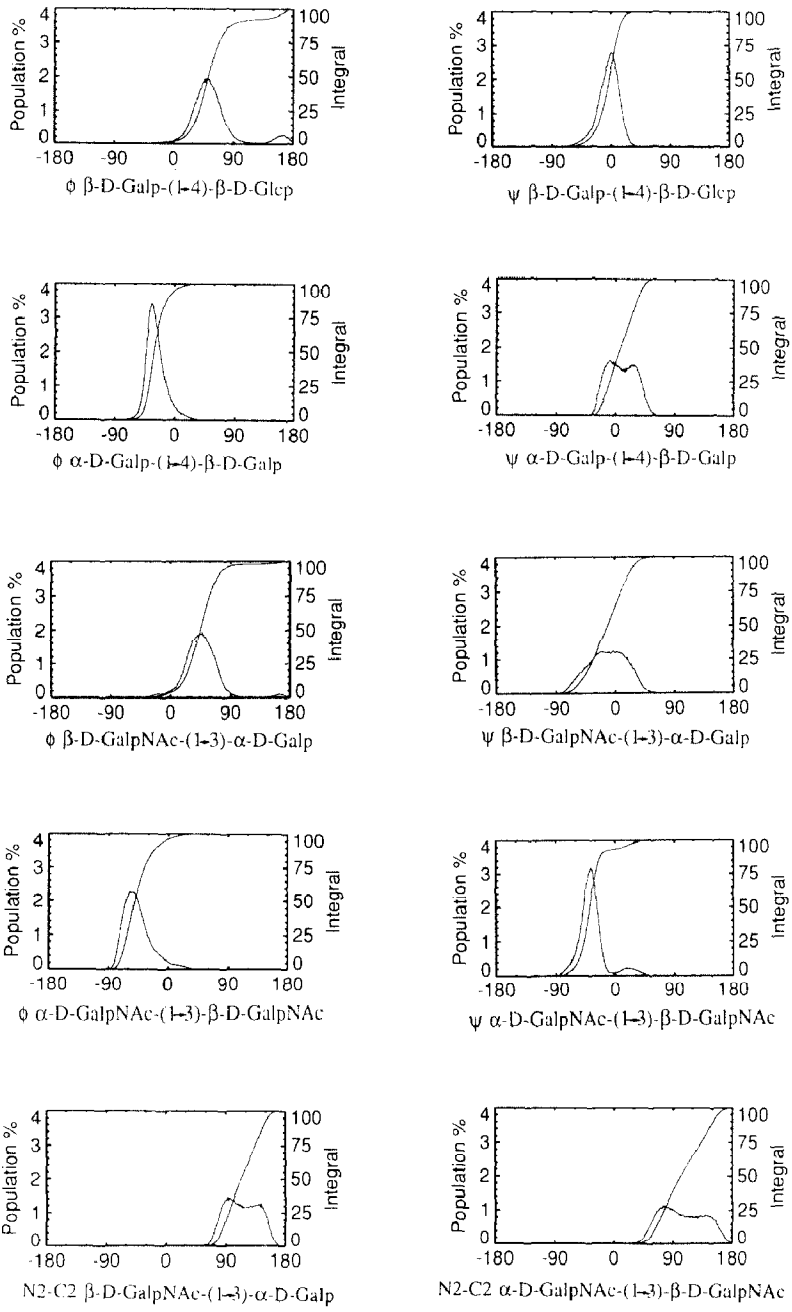

Fig. 3. One-dimensional population plots of ϕ and ψ angles generated from the MMC simulation of the pentasaccharide **10**.

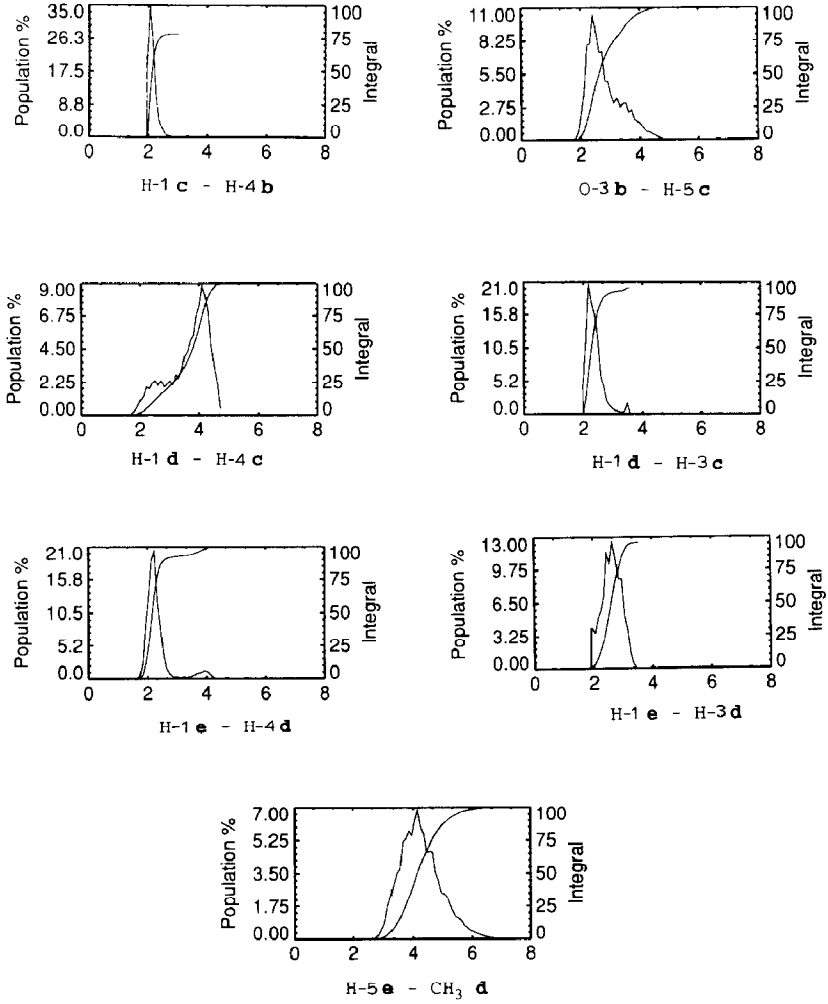

Fig. 4. Distribution of selected interatomic distances (Å) monitored during the MMC simulation of **10**.

protons across the glycosidic bond when deducing structures based on NOE results.

The distribution of distances between H-1**e** and H-4**d** is narrow with values between 1.8–2.8 Å and a maximum at 2.15 Å (Fig. 4), while the distance distribution between H-1**e** and H-3**d** is broader (2–3.7 Å). The populations with ϕ-values near zero and positive ψ-angles are seen as a small shoulder with a H-1**e**–H-4**d** distance of 3.9 Å. The *N*-acetyl group occupies conformations with N-H trans to H-2, with a torsional angle between 40° and 180°, energy minima at 65° and 150°.

The calculated and experimental NOE data are all in good agreement for this linkage. In the NOESY spectrum, a weak NOE between the *N*-acetyl CH_3-group of unit **d** and H-5**e** was also observed, which agrees with the distance between

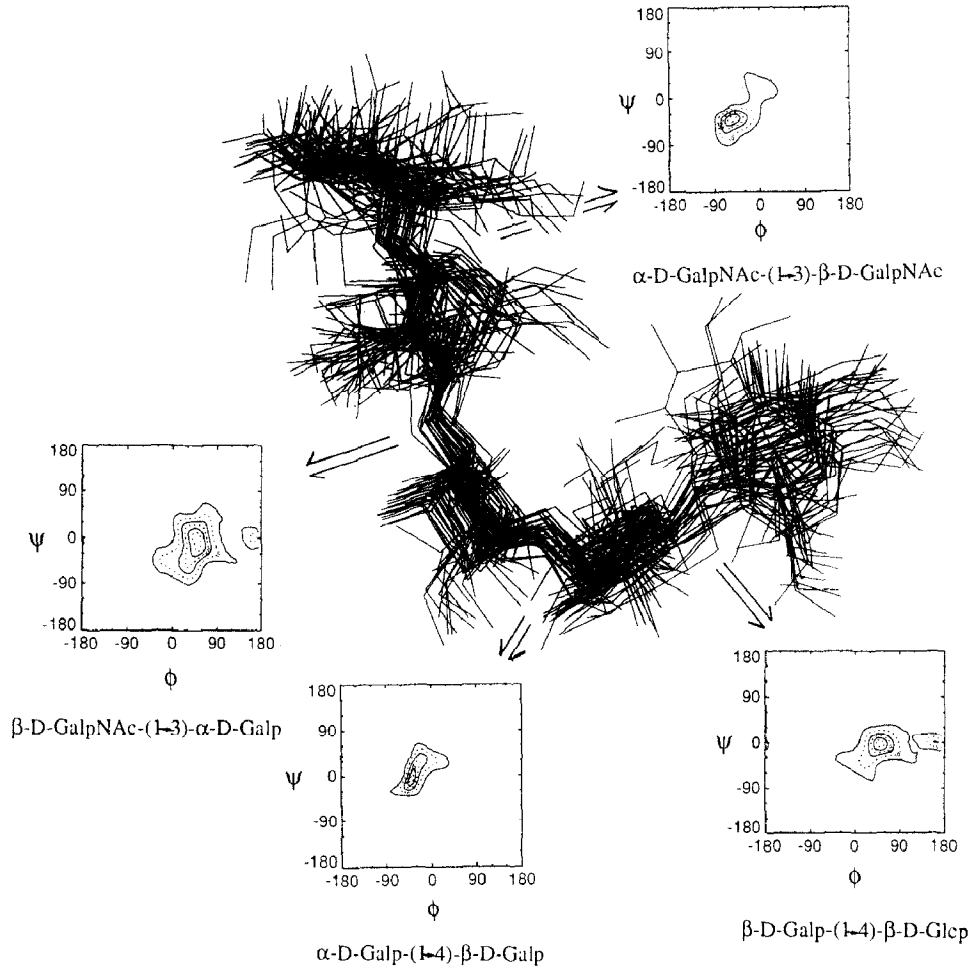

Fig. 5. Conformations obtained from the MMC simulation of the pentasaccharide **10**, by superimposition of 40 conformations selected randomly in the energy interval 0 to 10 kcal relative to the low energy conformation. The ϕ/ψ population maps are based on all conformations sampled during the MMC simulations.

these protons varying over a large range (3–6 Å) during the simulation as seen in Fig. 4.

*The β-*D*-Gal*p*NAc-(1 → 3)-α-*D*-Gal*p *segment* (**d–c**).—The energy map for the sampled conformations of this linkage covers a large area with small energy differences ($\leqslant 1$ kcal/mol). For the ϕ-angle, the energy minimum value is found around $+50°$ with a flexibility of approximately $\pm 15°$. An inverted conformation ($\phi \approx 170°$) is also indicated, comprising 2–3% of the total population for the pentasaccharide and somewhat more for the disaccharide **3**. Large flexibility around the ψ-angle is evident, with populations sampled between $-50°$ and $+30°$. For the pentasaccharide, the minimum energy conformers seem to be somewhat more restricted than for the disaccharide.

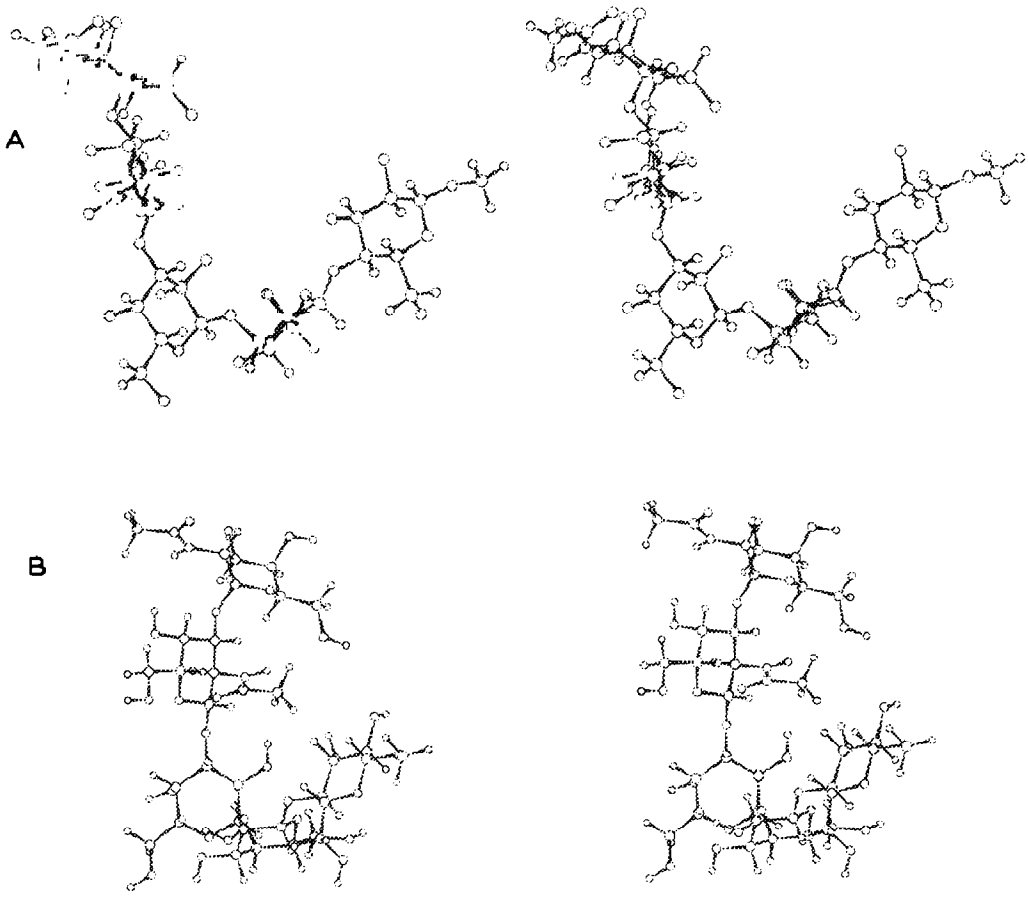

Fig. 6. A, Ball and stick stereoview representation of the lowest energy conformation of the Forssman pentasaccharide **10**. B, The lower pair is obtained by a rotation of 90° along the *z*-axis.

The *N*-acetyl group has approximately the same conformational space available as in the α-D-Gal*p*NAc residue **e**, although within a narrower range (70–170°, energy minima at 90° and 145°).

Investigations of globoside (cf. compound **8**, Fig. 1) in Me_2SO [17], based on hydrogen bonds detected for different conformers, indicated a large variation of the ψ-angle while the φ-angle was more rigid.

The α-D-*Gal*p-*(1 → 4)*-β-D-*Gal*p *segment* (**c**–**b**).—The φ-angle is relatively rigid with values centered around −39°, while the ψ-angle covers −30° to +50° showing two almost equally populated energy minima at −15° and +30° with a slight preference for the negative value. However, for the disaccharide fragment, the preference for the −15° conformation is more pronounced, indicating more flexibility in the pentasaccharide than in the disaccharide, although the simulated overall population maxima turn out to be the same in both compounds. This shows the importance of considering the overall shape of the conformational space

without putting too much emphasis on the exact values of the angles. Poppe et al. [14] found a positive ψ-value in the low energy conformation for the globotriose saccharide and a long-range NOE contact between H-5c and H-2b. This contact was also predicted in the present MMC simulation and verified experimentally.

The β-D-*Gal*p-(*1* → *4*)-β-D-*Glc*p *segment* (**b–a**).—The linkage has a rather well defined population maximum around ϕ/ψ-angles 50°/0° and with populations sampled for inverted structures ($\phi \approx 170°$). The experimental data are in good agreement with these observations.

3. Conclusions

These investigations have shown that the energy surfaces obtained from a grid search approach were well reproduced in the MMC simulations in the case of the disaccharides corresponding to the Forssman pentasaccharide. The next step, to use a molecule as large as the Forssman pentasaccharide in an MMC simulation, showed no large energy differences for the individual linkages, as compared to the disaccharide components. This result was strongly supported by the NMR data of all overlapping di-, tri-, and tetra-saccharides. A thorough analysis clearly showed that long-range chemical shift effects and unexpected NOEs were absent.

We thus conclude that the low energy conformations of all the different Forssman fragments are very similar to those of the pentasaccharide and, conversely, the data for the small fragments can be added together to give reasonable conformations for larger linear fragments, including the complete Forssman pentasaccharide. This might be applicable also to other large saccharides.

4. Experimental

Compounds **1–10** were available from previous work [10]. NMR data of monosaccharides used as references in Tables 1, 2, and 3 were obtained from the literature [30].

NMR spectroscopy.—Solutions of 4–5 mg of saccharide in 0.5 mL of D_2O were used. Spectra were recorded at 27°C in 5-mm tubes at 500.13 MHz for ^1H NMR and 125.77 MHz for ^{13}C NMR with a Bruker AM 500 spectrometer, and at 600 MHz for ^1H NMR with a Bruker AMX-600 instrument. The ^1H resonances were measured relative to internal acetone (2.225 ppm) and coupling constants were determined on a first-order basis (±0.3 Hz). The ^{13}C resonances are reported indirectly via the spectrometer lock frequency relative to internal dioxane (67.4 ppm). 2D NMR spectroscopy: COSY, relayed COSY, double relayed COSY, long-range coupling optimized COSY, double-quantum filtered phase-sensitive COSY, C–H correlation experiments, and ROESY experiments were performed as published previously [31]. The 2D NOESY spectrum at 600 MHz was measured with a mixing time of 200 ms in phase-sensitive mode according to States et al. [32]. 1024 FIDs were sampled in 4K real points and with a sweep width of 6250 Hz. The

data were processed in 4K × 2K complex points and shifted sine-bell windows were applied in both dimensions before Fourier transformation. To simplify spectra and to measure coupling constants for hydroxymethyl groups, partially relaxed 1D spectra [22] and triple quantum filtered COSY (TQFCOSY) experiments [33] were performed. Assignments of carbonyl carbons were accomplished using 1D selective decoupling and 2D COLOC experiments.

HSEA calculations.—The torsion angles ϕ and ψ are defined as H-1–C-1–O-1–C-X and C-1–O-1–C-X–H-X, respectively; ω as O-6–C-6–C-5–O-5; and for *N*-acetyl groups, a torsion angle is defined as as C-1–C-2–N-2–C-7. The glycosidic bond angle, τ, is defined as C-1–O-1–C-X. Coordinates for α- and β-D-galactopyranose [34], methyl β-D-galactopyranoside, and 2-acetamido-2-deoxy-α- and -β-D-galactopyranose [35] were taken from neutron diffraction data, and for β-D-glucopyranose [36] from X-ray studies of gentiobiose. The protons are attached as described previously [25]. Grid searches were performed in 10-degree steps for the ϕ and ψ angles while keeping exocyclic groups at values found for the minimum energy optimized conformation.

Monte Carlo simulations were performed for the disaccharides at a temperature of 600 K. The number of performed MMC steps was 1 000 000 with maximum step lengths of 24.0° (ϕ and ψ), 27.0° (ω, and *N*-acetyl groups), 33.0° (dihedral angles of hydroxyl groups) and 1.0° (τ), resulting in a total number of accepted conformers between 40 and 48%. For the pentasaccharide the MMC simulation was performed at 800 K, with 1 000 000 steps and the corresponding maximum step lengths of 16.0°, 19.0°, 21.0°, and 1.0°, giving a total acceptance rate of 45%. Simulations were also performed at temperatures of 300, 600, and 700 K.

Plots were generated with the PV-wave plot program (Precision Visuals Inc) and all calculations were performed on Silicon Graphics INDIGO workstations.

5. Acknowledgment

We thank Dr. Rainer Stuike-Prill for introduction to the GEGOP program and valuable discussions.

6. References

[1] D.M. Marcus, M.A. Naiki, and S.K. Kundu, *Proc. Natl. Acad. Sci. U.S.A.*, 73 (1976) 3263–3267.
[2] D.M. Marcus and S.K. Kundu, *Semin. Hematol*, 18 (1981) 63–71.
[3] S. Normark, M. Båga, M. Göransson, F.P. Lindberg, B. Lund, M. Norgren, and B.-E. Uhlin, *Microbial Lectins and Agglutins*, in D. Mirelman (Ed.), *Ecological and Applied Microbiology*, Wiley, New York, 1986, pp 113–143.
[4] C.A. Lingwood, H. Law, S. Richardsson, M. Petric, J.L. Brunton, S. De Grandis, and M. Karmali, *J. Biol. Chem.*, 262 (1987) 8834–8839.
[5] M. Jacewics, H. Clausen, E. Nudelman, A. Donohue-Rolfe, and J.E. Keusch, *J. Exp. Med.*, 163 (1986) 1391–1401.

[6] A.A. Lindberg, J.E. Brown, N. Strömberg, M. Westling-Ryd, J.E. Schultz, and K.A. Karlsson, *J. Biol. Chem.*, 262 (1987) 1779.
[7] S. Haataja, K. Tikkanen, J. Liukkonen, and J. Finne, *Internat. Carbohydr. Symp., XVIth, Paris, 1992*, p B-089.
[8] E. Nudelman, R. Kannagi, S. Hakomori, M. Parsons, M. Lipinski, J. Weils, M. Fellous, and T. Tursz, *Science*, 220 (1983) 509–511.
[9] C.C. Sweely and B. Klionski, *J. Biol. Chem.*, 238 (1963) 3148–3150.
[10] G. Magnusson, U. Nilsson, A.K. Ray, and G. Taylor, *ACS Symp. Ser.*, 519 (1993) 92–110.
[11] K. Bock, T. Frejd, J. Kilhberg, and G. Magnusson, *Carbohydr. Res.*, 176 (1988) 253–270.
[12] D.E. Portlock, G.S. Lubey, and B. Borah, *J. Org. Chem.*, 54 (1989) 2327–2330.
[13] J.S. Yadav and P. Luger, *Int. J. Quant. Chem.*, 23 (1983) 1433–1439.
[14] L. Poppe, J. Dabrowski, C.W.V.D. Lieth, K. Koike, and T. Ogawa, *Eur. J. Biochem.*, 189 (1990) 313–325.
[15] J.N. Scarsdale, R.K. Yu, and J.H. Prestegard, *J. Am. Chem. Soc.*, 108 (1986) 6778–6784.
[16] J.N. Scarsdale, P. Ram, J.H. Prestegard, and R.K. Yu, *J. Comput. Chem.*, 9 (1988) 133–147.
[17] L. Poppe, C.-W. von der Lieth, and J. Dabrowski, *J. Am. Chem. Soc.*, 112 (1990) 7762–7771.
[18] L. Poppe, J. Dabrowski, and C.W.V.D. Lieth, *Biochem. Biophys. Res. Commun.*, 174 (1991) 1169–1175.
[19] K. Bock, M. Breimer, A. Brignole, G.C Hansson, K.A. Karlsson, G. Larson, H. Leffler, B.E. Samuelsson, C. Svanborg-Ede'n, and J. Thurin, *J. Biol. Chem.*, 260 (1985) 8545–8551.
[20] R. Stuike-Prill and B. Meyer, *Eur. J. Biochem.*, 194 (1990) 903–919.
[21] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.*, 21 (1953) 1087–1092.
[22] L.D. Hall, *Adv. Carbohydr. Chem. Biochem.*, 29 (1974) 11–40.
[23] A. De Bruyn, *J. Carbohydr. Chem.*, 10 (1991) 159–180.
[24] S. Haataja, K. Tikkanen, J. Liukkonen, C. Francois-Gerard, and J. Finne, *J. Biol. Chem.*, 268 (1993) 4311–4317.
[25] K. Bock, *Pure Appl. Chem.*, 55 (1983) 605–622.
[26] L. Poppe, R. Stuike-Prill, B. Meyer, and H. van Halbeek, *J. Biomol. NMR*, 2 (1992) 109–136.
[27] T. Peters, B. Meyer, R. Stuike-Prill, R. Somorjai, and J.-R. Brisson, *Carbohydr. Res.*, 238 (1993) 49–73.
[28] B. Meyer, M. Zsiska, and R. Stuike-Prill, *Monte Carlo Simulations in Glycoproteins*, Springer Proceedings, in D.P. Landau, K.K. Mon, and H.-B. Schlütter (Eds.), *Physics Vol. 72, Computer Simulation Studies in Condensed Matter Physics IV*, 1993, pp 90–112.
[29] R.U. Lemieux, K. Bock, L.T.J. Delbaere, S. Koto, and V.S. Rao, *Can. J. Chem.*, 58 (1980) 631–653.
[30] K. Bock and H. Thøgersen, *Annu. Rep. NMR. Spectrosc.*, 13 (1982) 1–57.
[31] K. Bock, J.Ø. Duus, B. Norman, and S. Pedersen, *Carbohydr. Res.*, 211 (1991) 219–233.
[32] D.J. States, R.A. Haberkorn, and D.J. Ruben, *J. Magn. Reson.*, 48 (1982) 286–292.
[33] J. Boyd, C.M. Dobson, and C. Redfield, *FEBS Lett.*, 186 (1985) 35–40.
[34] S. Takagi and G.A. Jeffrey, *Acta Crystallogr., Sect. B*, 35 (1979) 902–906.
[35] W.T. Winter, S. Arnott, D.H. Isaac, and E.D.T Atkins, *J. Mol. Biol.*, 125 (1978) 1–19.
[36] G.A. Jeffrey, R.K. McMullan, and S. Takagi, *Acta Crystallogr., Sect. B*, 33 (1977) 728–733.