Journal of Biomolecular NMR, **14:** 33–45, 1999.
KLUWER/ESCOM

'Wave-type' structure of a synthetic hexaglycosylated decapeptide: A part of the extracellular domain of human glycophorin A

Oliver Schuster, Gunther Klich, Volker Sinnwell, Helge Kränz, Hans Paulsen & Bernd Meyer*
Institute for Organic Chemistry, Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, D-20146 Hamburg, Germany

Received 23 November 1998; Accepted 19 January 1999

Key words: glycophorin A, glycopeptide, MD simulation

Abstract

The three-dimensional structure of a glycopeptide, His-Thr*-Ser*-Thr*-Ser*-Ser*-Ser*-Val-Thr-Lys, with 2-acetamido-2-deoxy-α-D-galactose (GalNAc) residues linked to six adjacent amino acids from Thr-10 to Ser-15, was studied by NMR spectroscopy and molecular dynamics (MD) simulations. The hexaglycosylated decapeptide is part of the extracellular domain of human glycophorin A and shows an extended structure of the peptide backbone due to *O*-glycosylation. Furthermore, each GalNAc residue exhibits one and only one NOE contact from the NHAc proton to the backbone amide proton of the amino acid that the sugar is directly bound to. This indicates a strong preference for the orientation of all GalNAc residues towards the N-terminus. NOE build-up curves were used to determine 42 inter-proton distances that, in connection with ϕ angles of the peptide backbone obtained from ^3J-coupling constants, resulted in constraints for a MD simulation in water. The NMR data and the MD simulations show a preference for an extended backbone structure. The GalNAc residues are located alternatingly on opposite sides of the backbone and reduce the flexibility of the peptide backbone. The conformation of the molecule is relatively rigid and shows a 'wave-type' 3D structure of the peptide backbone within the glycosylation cluster. This new structural element is also supported by the unusual CD spectrum of the glycopeptide.

Introduction

Glycophorin A is a highly glycosylated glycoprotein (60% carbohydrate by mass) and is one of the major sialoglycoproteins located in the membranes of human erythrocytes (Furthmayr, 1978). Differences in the amino acid sequence are correlated with M or N blood group activity (Furthmayr, 1978; Marchesi, 1979) and various biological functions are known. It has been shown that glycophorin A protects K562 cells from natural killer cell attack (El Ouagari et al., 1995). The level of protection is correlated with the amount of glycophorin A embedded in the K562 cell membrane. The main effect is carried by the *N*-linked oligosaccharide. Another important function of glycophorin A is the preservation of the cellular shape of erythrocytes by the association of glycophorin and protein 4.1 (Anderson and Lovrien, 1984) and through the regulation of that association by polyphosphoinositide (Anderson and Marchesi, 1985).

The glycoprotein consists of 131 amino acids and has one trans-membrane domain. The extracellular part is highly glycosylated with 15 *O*-glycosylation sites and one *N*-glycosylation site. Two glycosylation clusters are found from Ser-2 to Thr-4 and from Thr-10 to Ser-15. The *O*-glycosylation sites carry a tetrasaccharide each, α-NeuNAc-(2-3)-β-Gal-(1-3)-[α-NeuNAc-(2-6)]-α-GalNAc-Ser/Thr. The *N*-glycan is a complex type saccharide and consists of 13 monosaccharides (Fukuda et al., 1987) (Figure 1b).

The role of glycosylation on the conformation of *O*-linked glycoproteins was shown by light-scattering studies of ovine submaxillary mucin (Shogren et al., 1989). The native and asialo mucin were described as highly extended random coils. Upon complete removal of the carbohydrates, the characteristic ex-

*To whom correspondence should be addressed. E-mail: bernd_meyer@sgi1.chemie.uni-hamburg.de.

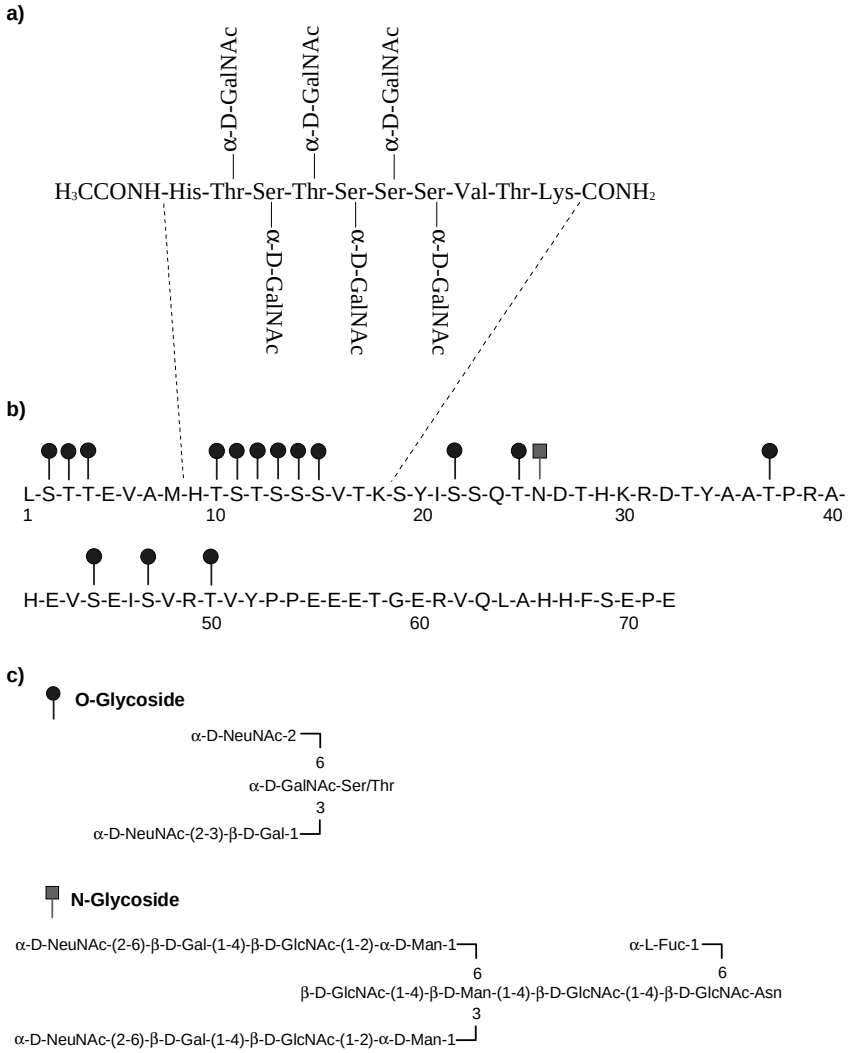

Figure 1. (a) Glycophorin A^N: Synthetic decapeptide hexaglycosylated with α-D-GalNAc. (b) Amino acid sequence in single letter code of the extracellular domain of glycophorin A^N with *O*- and *N*-glycosylation sites. The location of the synthetic glycopeptide relative to the native molecule is indicated by dashed lines from His-9 to Lys-18. (c) Oligosaccharide structures of the *O*- and *N*-type glycosylation sites.

tended structure of intact mucin collapses to dimensions typical of denatured globular proteins. Around the *O*-glycosylation site extended peptide conformations were found in eight *O*-glycopeptides sequences derived from the *N*-terminus of interleukin-2 containing two to ten amino acids, monoglycosylated at threonine 3 with a GalNAc residue (Paulsen et al., 1991).

Work on the impact of *N*-glycosylation on the peptide structure shows that the α-subunit of human chorionic gonadotropin exhibits interactions with the first carbohydrate residue (GlcNAc) (de Beer et al., 1996). Similar interactions were reported for the *N*-linked carbohydrate from the adhesion domain of human T lymphocyte glycoprotein CD2 that shows NOE contacts between the protein and the first GlcNAc of the glycan (Withka et al., 1993). Especially the first *N*-acetylglucosamine of the high mannose glycan is as rigid as the protein core (Wyss et al., 1997).

Structural changes were observed when most of the oligosaccharide chains of glycophorin $A^{M,N}$were removed, also lack of α-D-NeuNAc residues near the *N*-terminal amino acid induces a structural change (Batstone-Cunningham et al., 1983). Of interest are the effects of glycosylation near the *N*-terminus and their antigenic properties. Monoglycosylation of a

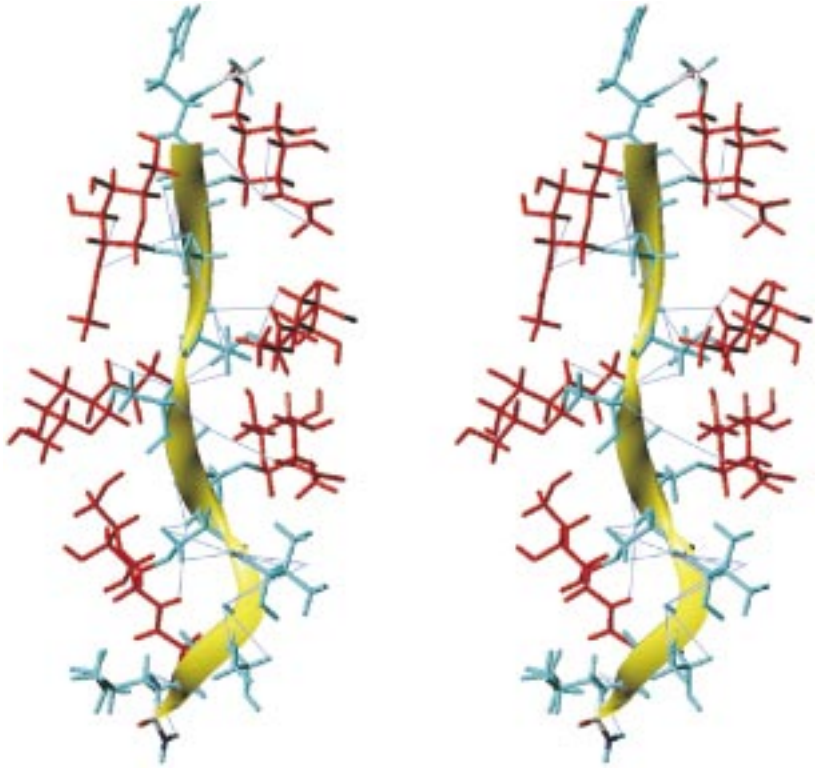

Figure 2. Stereo plot of an energy optimized average structure of the hexaglycosylated decapeptide obtained from the last 100 ps of the 1000 ps constrained MD simulation. α-D-GalNAc residues are coloured red, the peptide is coloured cyan. A yellow shaded ribbon shows the peptide backbone. The distance constraints are solid blue lines.

synthetic pentapeptide from the *N*-terminus of glycophorin A^M has an influence on the structure. T_1 relaxation times of the carbon atoms of the glycosylated pentapeptide are shorter than those of the unglycosylated peptide indicating that the glycosylation reduces conformational flexibility (Dill et al., 1985). It is known that the highly glycosylated T1 fragment of glycophorin A has a well defined saccharide and extended peptide backbone structure, because of peptide carbohydrate interactions at 13 glycosylation sites (Pieper et al., 1996). The *O*-glycosidic linkages of the first sugar residues to serine and threonine are of special interest. Nuclear Overhauser studies on model compounds, α-GalNAc-Thr, showed that the most stable conformation of the glycosidic bond in α-glycopeptides is that in which the anomeric proton and threonine proton CH_β are in close proximity (Pavia and Ferrari, 1983). This work also demonstrated an almost orthogonal orientation at the χ_1 angle of the glycosylated threonine.

Stochastic dynamics simulations of *O*-linked glycopeptides from ovine submaxillary mucin were used as simple models for the native and asialo form. The simulations show a significant decrease in end to end distance and radius of gyration (32% and 33%, respectively) upon complete removal of the mono- and disaccharide (α-NeuNAc-(2-6)-α-GalNAc and α-GalNAc) (Butenhof and Gerken, 1993).

Here we report the investigation of the impact of a highly clustered glycosylation with GalNAc residues on peptide conformation. It is of special interest whether the monosaccharide exerts the same effect as the tetrasaccharides present in the native glycophorin. The synthesis of the hexaglycosylated decapeptide (Figure 1a) has been described previously (Klich et al., 1997).

Materials and methods

NMR experiments

The spectra were recorded on Bruker DRX 500 spectrometer with a triple resonance 5 mm inverse probe head. The sample contained 11.3 mg of a glycopeptide

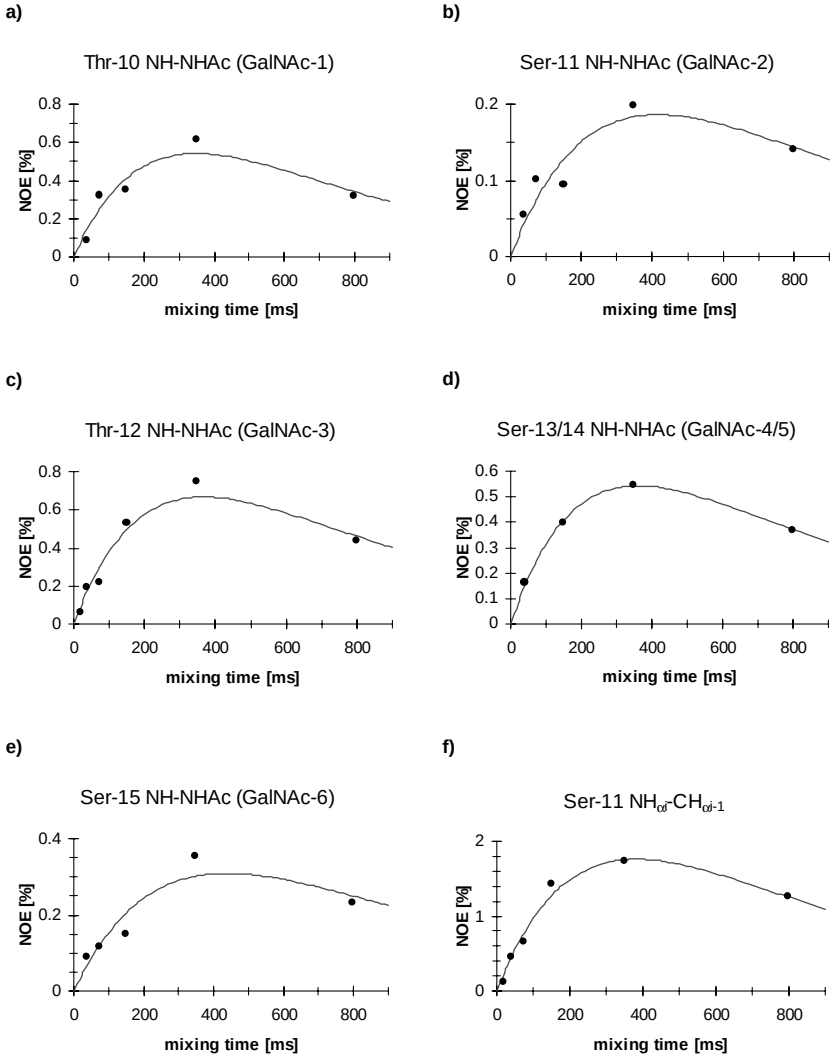

Figure 3. The experimental build-up curves a, b, c, d, e from the NH protons of the *O*-glycosylated amino acids (Thr-10, Ser-11, Thr-12, Ser-13, Ser-14, Ser-15) to the NHAc protons of the directly bound GalNAc residues. Curve f shows the build-up curve from the NH proton from Ser-14 to the $CH_{\alpha i-1}$ proton of Ser-13. The data was obtained from NOESY experiments with different mixing times (20 ms, 40 ms, 75 ms, 150 ms, 350 ms, 800 ms) (marked with a solid dot). The values were referenced to the diagonal signal of the amide proton from the C-terminus. The integral was extrapolated to a mixing time of 0 ms and set to 100%. The only cross peak, for the NH (amino acid)–NHAc (GalNAc) contacts, we could analyze at 20 ms was the one from Thr-12 to NHAc (GalNAc-3). At the mixing time of 75 ms the volumes of the signals from Ser-13/14 could not be determined. A double exponential fit was used to determine the buildup rate at $\tau_{mix} = 0$. The fitting function for the NOE contacts was $f(t) = p0 \cdot e^{-p2 \cdot t} \cdot (1 - e^{-p1 \cdot t})$ and is shown as a solid line.

(MW 2294.4 g/mol), in 550 µl H_2O/D_2O (90%/10%) (D_2O: 99.9%, Aldrich) which gives a concentration of 9.0 mmol/l. The pH value was adjusted to 3.5 (0.01% TFA, Merck). All spectra were measured at 300 K. All chemical shifts are referenced to the H_2O signal at 4.75 ppm for ^{1}H and to the TFA signal at 115.7 ppm for ^{13}C. Presaturation for water suppression was used in all experiments. Interpretation of the spectra was carried out with the XWI*N*-NMR (Bruker, v. 1.3) and the AURELIA program (Bruker, v. 2.0).

1D ^{1}H NMR spectra. The spectrum was recorded with 64 K data points, a spectral width of 15 ppm, and 1600 scans.

TOCSY spectrum. Sequence: $D_1(CW_{pres.})$-90°-t_1-P17-MLEV17-P17-t_2. Relaxation delay, $D_1 = 2$ s,

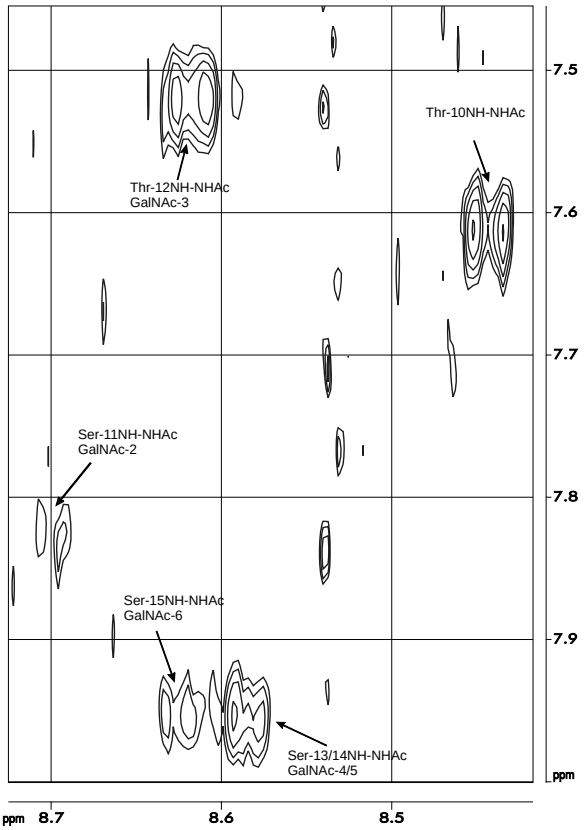

Figure 4. NH (amino acid)-NH (GalNAc) region of the 500 MHz NOESY spectrum (350 ms mixing time). Shown are the individual signals from the NH protons of the amino acids to the NH protons of the GalNAc residues. Importantly, each NH proton has only one cross peak.

mixing time = 99.4 ms, carrier frequency 2353 Hz, $\gamma B_2 = 5.2$ kHz, $SW_1 = SW_2 = 10$ ppm.

NOESY spectra. Sequence: $D_1(CW_{pres.})$-90°-t_1-90°-$\tau_{mix}(CW_{pres.})$-90°-t_2. Relaxation delay, $D_1 = 2$ s, mixing times $\tau_{mix} = 20$, 40, 75, 150, 350, and 800 ms, $SW_1 = SW_2 = 10$ ppm.

ROESY spectra. Sequence: $D_1(CW_{pres.})$-90°-t_1-P15-t_2. Relaxation delay, $D_1 = 2$ s, P15 pulse for spin lock $\tau_{mix} = 350$ ms, carrier frequency 2355 Hz, $\gamma B_2 = 1.2$ kHz, $SW_1 = SW_2 = 10$ ppm.

E.COSY spectra. Sequence: $D_1(CW_{pres.})$-90°-t_1-90°-90°-t_2. Relaxation delay, $D_1 = 2$ s. $f_1 = f_2 = 10$ ppm. (H_2O/D_2O, 9/1). In D_2O, $SW_1 = SW_2 = 6.3$.

Gradient selected HMQC. Sequence: D_1-90°(^{1}H)-D_2-90°(^{13}C)-t_1/2-G1-180°(^{1}H)-G2-t_1/2-90°(^{13}C)-G3-t_2 (GARP). Relaxation delays, $D_1 = 2$ s, $D_2 = 1/[2J(C,H)] = 3.45$ ms, calculated from $^1J(C,H) = 145$ Hz. G1, G2, G3 = sinusoidal shaped field gradients with 5% truncation, 2 ms duration. Gradient strength ratio: 5:3:4. $SW_1 = 10$ ppm and $SW_2 = 222$ ppm.

Gradient selected HMBC spectra. Sequence: D_1-90°(^{1}H)-D_2-90°(^{13}C)-D_6-90°(^{13}C)-t_1/2-G1-180°(^{1}H)-G2-t_1/2-90°(^{13}C)-G3-t_2. Relaxation delay, $D_1 = 2$ s, $D_2 = 1/[2J(C,H)] = 3.45$ ms, calculated from $^1J(C,H) = 145$ Hz, $D_6 = 1/[2J(C,H)] = 80$ ms, calculated from $^1J(C,H) = 6.25$ Hz. G1, G2, G3 = sinusoidal shaped field gradients with 5% truncation, 2 ms duration. Gradient strength ratio: 5:3:4. $SW_1 = 8$ ppm and $SW_2 = 222$ ppm.

Molecular dynamics simulations

Constrained molecular dynamics (MD) simulations were carried out with the Sybyl program (v. 6.3, Tripos) on Silicon Graphics O2 (R10000) computers, using the Tripos force field. The starting structure was generated with fixed distance constraints obtained from the NOE distances. After energy optimization the resulting ϕ angles of the peptide backbone were compared to the possible solutions of the Karplus equation. The closest values were then used for torsional constraints. The amide bonds were assumed to be in the trans conformation. The structure was described by 42 distance and 10 torsion constraints (Table 2 and Figure 2). Constraints to pseudo atoms, generated by Sybyl, were used for non-stereospecifically assigned methylene groups and methyl groups. A harmonic potential was employed to the distance and torsion constraints. The force-field constants were set to 10 kcal/mol/Å^2 for the fixed distances. The constants for the torsional constraint of the ϕ angles of the peptide backbone were set to 0.01 kcal/mol/deg^2.

Additionally, the χ_1 values from the glycosylated amino acids were set to a conformation compatible with the ^3J coupling constants. The ϕ angles of the glycosidic linkage of α-D-glycosides, influenced by the *exo*-anomeric effect, were set close to −60° (definition of the ϕ angle: H1-C1-O1-Cβ). The values of the ψ angles (definition of the ψ angle: C1-O1-Cβ-Cα) of the GalNAc residues were in a range from −78° for GalNAc-4 to 173° for GalNAc-3.

This extended conformation was energy-optimized and put in a water box containing 1312 water molecules (WTR, Sybyl 6.3). The dimension of the result-

ing box was 39 Å × 33 Å × 33 Å. Before starting the MD simulation the box was energy-optimized over 1000 steps. The constrained simulation was performed at 300 K. The charges were calculated with the Gasteiger-Marsili method (Gasteiger and Marsili, 1980) and a dielectric constant of 1 was used. A cut-off radius of 8 Å was used for the *non*bonded interactions. The initial velocities for the atoms were taken from a Boltzman distribution at 300 K and the step size for the integration of Newton's equation was 1 fs. The coupling to the temperature bath was set to 100 fs and the *non*bonded interactions were updated each 25 fs. The MD simulation was run for 1000 ps at constant volume and temperature. The simulation was analyzed by Matlab scripts (Ver. 4.2c) to generate trajectories and population plots. The progress of the simulation was visualized in overlay plots of 10 structures (Figures 7 and 8), in 100 ps intervals. An energy-optimized average structure of the last 100 ps is shown in Figure 2.

CD spectra

CD experiments were carried out at 25 °C and a scan speed of 5 nm/min with a JASCO J-500 dichrograph. The data were digitally recorded and fed through the data processor for signal averaging and baseline subtraction. The pH value was adjusted to 5.5 (0.01% TFA, Merck) and the optical path length was 10 mm. The concentration of the hexaglycosylated decapeptide was 27.24 μmol/l. For the unglycosylated decapeptide we used a solution in water of 58.13 μmol/l. All spectra were recorded in a nitrogen atmosphere.

Results and discussion

NMR experiments

^{1}H NMR chemical shifts of the synthetic glycopeptide were assigned using TOCSY and COSY spectra (Table 1). The NH resonances of the GalNAc residues were located between 7.52 ppm and 7.96 ppm with four of the six being dispersed. The traces of the NH signals in the TOCSY were used to resolve the ^{1}H resonances of the *O*-linked GalNAc residues. The magnetization transfer within each GalNAc residue can be observed from NH (GalNAc) up to H-4. The resonance frequencies from H-5 and H-6a/b were obtained from the inverse ^{13}C-^{1}H correlation (HMQC). To complete and double check the assignment we used a long-range inverse ^{13}C-^{1}H correlation (HMBC). The sequential assignment of the amino acids was obtained from NOESY and ROESY spectra. A stereospecific assignment of the proton resonances of the CH_2groups of amino acid side chains and of the H6 resonances was not possible.

The dispersion of the NH proton signals of the GalNAc residues allows the assignment of the H-1, H-2, H-3, and H-4 cross peaks. The H-1 chemical shifts are located from 4.81 ppm to 4.84 ppm. The ranges of dispersion for the chemical shift of the H-2, H-3 and H-4 signals are 0.1, 0.02 and 0.03 ppm, respectively. Thus, the chemical shift dispersion of the NH resonances from the GalNAc residues was the largest, followed by that of H-2, H-1, H-3 and H-4, suggesting a proximity of the *N*-acetyl groups with the peptide. The H-5 signals could not be assigned explicitly to a specific GalNAc residue and the signals of all H-6a/b protons were overlapped at about 3.67 ppm.

The ^{1}H chemical shift of the NH protons of the amino acids range from 8.35 to 8.71 ppm. The CH_α protons of the glycosylated threonines are found significantly down-field from their normal position (δ = 4.55, 4.54). The CH_α protons of the serine residues are found from 4.49 to 4.81 ppm. The resonances of the CH_β protons of the glycosylated threonine residues were found at 4.27 and 4.28 ppm, the unglycosylated threonine at 4.06 ppm. The dispersion of the H_β protons of the serine residues show a dispersion from 0.1 ppm for Ser-13 to 0.19 ppm for Ser-11. The chemical dispersion of the H_α signals of the serine residues suggests a relatively well defined conformation of the peptide.

Coupling constants were used to determine dihedral angles ϕ and χ_1 of the peptide. The ^3J(NH,CH_α) (Table 2) coupling constant is associated to the torsion angle ϕ of the peptide backbone and the ^3J(CH_α,$CH_{\beta1,\beta2}$) coupling constants associated with the torsion angle χ_1 were determined from a 1D ^{1}H NMR spectrum and an E.COSY spectrum. The coupling constants were interpreted using a modified Karplus equation (Ludvigsen et al., 1991).

The ^3J(NH,CH_α) coupling constants range from 6.9 Hz (Lys-18) to 9 Hz (Thr-12) indicating a rather extended structure. The region from Thr-10 to Ser-13 shows the largest coupling constants between 7.9 and 9.0 Hz supporting the sterical demand of the threonine residues. The values correspond to dihedral angles of $\phi \approx -100°$ or $-145°$. The serine residues 14 and 15 as well as the unglycosylated amino acids show smaller coupling constants (cf. Table 2).

Table 1. ^{1}H NMR chemical shifts of the hexaglycosylated decapeptide in H_2O/D_2O 9/1 at 300 K

^{1}H [ppm]	NAc	His-9	Thr-10[a]	Ser-11[a]	Thr-12[a]	Ser-13[a]	Ser-14[a]	Ser-15[a]	Val-16	Thr-17	Lys-18	NH_2
NH		8.48	8.44	8.71	8.62	8.58	8.59	8.63	8.38	8.35	8.41	7.07
NH												7.59
CH_α		4.79	4.55	4.81	4.54	4.61	4.49	4.59	4.16	4.23	4.21	
$CH_{\beta'}$		3.22	4.28	3.96	4.27	3.85	3.84	3.83	2.00	4.06	1.76	
CH_β		3.08		3.77		3.75	3.71	3.68			1.70	
$CH_{\gamma'}$			1.21		1.21				0.86	1.14	1.38	
CH_γ									0.89		1.38	
CH_δ											1.62	
CH_ϵ											2.92	
H-2		8.55										
H-4		7.24										
Ac	1.92		1.97	1.95	1.98	1.99	1.99	1.99				
NH			7.62	7.83	7.52	7.96	7.96	7.96			7.48	
1-H			4.81	4.82	4.84	4.82	4.83	4.83				
2-H			4.00	4.08	4.03	4.10	4.10	4.08				
3-H			3.80	3.82	3.80	3.82	3.81	3.81				
4-H			3.90	3.89	3.92	3.90	3.90	3.90				
5-H			3.94[b]	3.93[b]	3.81[b]	3.81[b]	3.73[b]	3.73[b]				
6-Ha/b			3.67[c]	3.67[c]	3.67[c]	3.67[c]	3.67[c]	3.67[c]				

[a] Amino acids *O*-glycosylated with GalNAc. The resonances are obtained from a ^{1}H- and TOCSY-spectrum.
[b] Signals cannot be assigned to individual residues.
[c] Signals are overlapped.

Table 2. $^3J(HN,H_\alpha)$ coupling constants obtained from an E.COSY and a 1D ^{1}H spectrum and calculated dihedral ϕ angles

	His-9	Thr-10	Ser-11	Thr-12	Ser-13	Ser-14	Ser-15	Val-16	Thr-17	Lys-18
$^3J(HN,H_\alpha)$ [Hz]	7.7	8.5	7.9	9.0[a]	7.8[a]	7.0[a]	7.2[a]	7.9	7.4	6.9
$\phi_{exp.}$ [deg.]	−90, −150[b]	−98, −142[b]	−92, −148[b]	−105, −135[b]	−91, −149[b]	−85[b], −156	−86[b], −154	−92[b], −148	−88, −152[b]	−84, −156[b]

Experimental values of $^3J(HN,H_\alpha)$ coupling constants were determined from the 1D ^{1}H spectrum, or [a]from an E.COSY spectrum. Two values for the backbone angles ϕ were obtained from the Karplus equation (Ludvigsen et al., 1991).
[b] ϕ-angle used as constraint in the MD simulation.

Table 3. $^3J(CH_\alpha\text{-}CH_{\beta1,2}, CH_{\beta1}\text{-}CH_{\beta2})$ coupling constants obtained from an E.COSY and calculated staggered dihedral χ_1 angles of the side chains with the three state jump model[a]

		His-9	Thr-10	Ser-11	Thr-12	Ser-13	Ser-14	Ser-15	Val-16	Thr-17	Lys-18
$^3J_{CH\alpha-CH\beta1}$	[Hz]	5.3[1]	4.6	6.6	3.3	6.1	7.1	7.3	7.6	6.5	6.4
$^3J_{CH\alpha-CH\beta2}$	[Hz]	8.9[1]		5.4		5.1	5.4	5.6			8.2
$^3J_{CH\beta1-CH\beta2}$	[Hz]	15.6[1]		11.4		11.1	10.6	10.2			13.9
gauche	[%]		87		100				56	67	
trans	[%]		13		0				44	33	
$\chi_1 = 60°$	[%]	22		45		54	40	36			18
$\chi_1 = 180°$[2]	[%]	58		21		18	21	23			51
$\chi_1 = -60°$[2]	[%]	20		34		28	39	41			32

[a] The rotamer distribution was determined on the assumption that only the three staggered rotamers are populated ($\chi_1 = -60°$, 60° and 180°) and that the averaging proceeds between the *gauche* ^3J(H,H) coupling of 3.4 Hz and the *trans* ^3J(H,H) coupling of 12.9 Hz. The $^3J(CH_\alpha\text{-}CH_{\beta1,2}, CH_{\beta1}\text{-}CH_{\beta2})$ coupling constants were obtained from an E.COSY spectrum in D_2O or [1]in H_2O. Calculated populations of the staggered χ_1 angles (marked with [2]) can be exchanged.

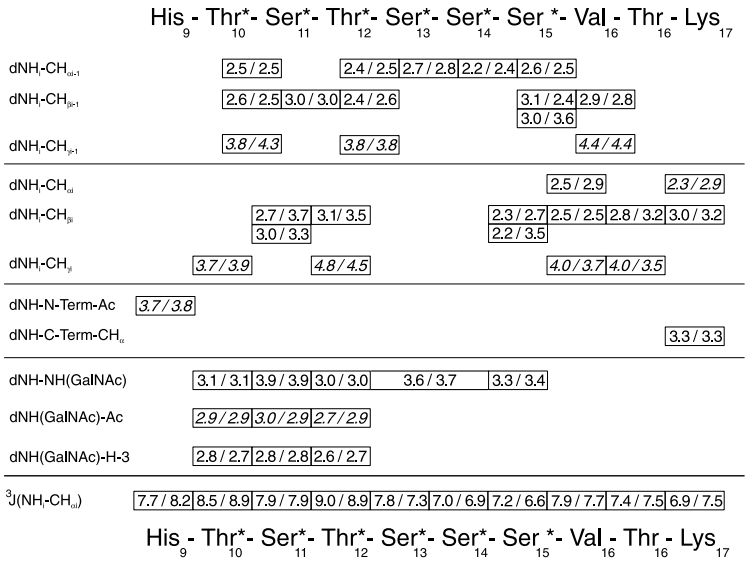

Figure 5. Sequence of the synthetic hexaglycosylated decapeptide with sequential, *non*-sequential, NH (amino acid) – NHAc (α-D-GalNAc), NHAc (α-D-GalNAc) – H-3 (α-D-GalNAc), NHAc (α-D-GalNAc) – Ac (α-D-GalNAc) distances [Å] and $^3J(\mathrm{NH_i,CH_{\alpha i-1}})$ coupling constants [Hz]. The first values in the boxes are from NOE build-up curves and the coupling constants are from an E.COSY spectrum, the second are average values from the constrained MD simulation. Amino acids, marked with an asterisk are monoglycosylated with α-D-GalNAc. Values written in italics are set to pseudo atoms. The NH signals (Ser-13, 14) – NHAc (α-D-GalNAc) for the distance calculation were integrated together and divided by two.

The populations of the side chain χ_1 rotamers were calculated with the three state jump model (Table 3). The glycosylated threonines show $^3J(\mathrm{CH_\alpha,CH_\beta})$ coupling constants of 4.6 Hz and 3.3 Hz, respectively, resulting in *gauche* orientations of the $\mathrm{CH_\alpha}$ protons and $\mathrm{CH_\beta}$ protons. In contrast, the unglycosylated threonine has a coupling constant of 6.5 Hz resulting in a 2:1 ratio of the *gauche* and the *trans* orientation. The glycosylated serine residues show a small population of about 20% of the *trans* conformer and 80% population of the *gauche* conformers. *Non*-glycosylated residues exhibit little preference for *gauche* or *trans* orientations.

Inter-proton distances were obtained by a set of NOESY spectra recorded with six different mixing times from 20 to 800 ms. Peak volumes were integrated on one side of the diagonal. The calculation is done by an iterative segmentation process which is implemented in the AURELIA program. A segmentation level of 0.5 (half intensity of the individual peak) was shown to be reasonable to separate overlapping signals. The calibration for the distance calculations was obtained from the integral of the cross peak between the amide protons at the C-terminus ($\delta = 7.07$ and 7.59). The reference distance was set to 1.73 Å. The NOE build-up curves were generated by a double exponential fit. The cross relaxation rate σ was calculated by the initial rate approximation and results in distance constraints for the MD calculation (Figure 5).

To verify the calculated proton distances we compared distances for proton pairs with a relatively fixed distance with their experimentally determined values calculated from NOE buildup rates. Intra-residue distances between the amide protons of the GalNAc residues 1, 2, and 3 and the methyl protons of their own acetyl group were used. The deviations between experiment and fixed distance were less than $\pm$ 0.15 Å with the exact distance being positioned in the middle of the experimental values validating the choice of the reference atom pair. The NOEs between H-3 and the NH within a monosaccharide were used as an additional test for the distances obtained from analysis of the build-up rates. The distance obtained from the model is 2.62 Å and the distances obtained from build-up rates is between 2.6 Å and 2.8 Å showing a maximum deviation of 0.18 Å.

The build-up curves for the inter residue NOEs from the NH signals of the GalNAc residues to the NH signals of the peptide backbone are shown in Figure 3. NOEs are only observed between the sugar and the NH of the amino acid they are directly attached to (Figure 4). For Ser-13 and Ser-14 we found one overlapped cross peak from the NH protons of the amino acids to the related NHAc protons of the

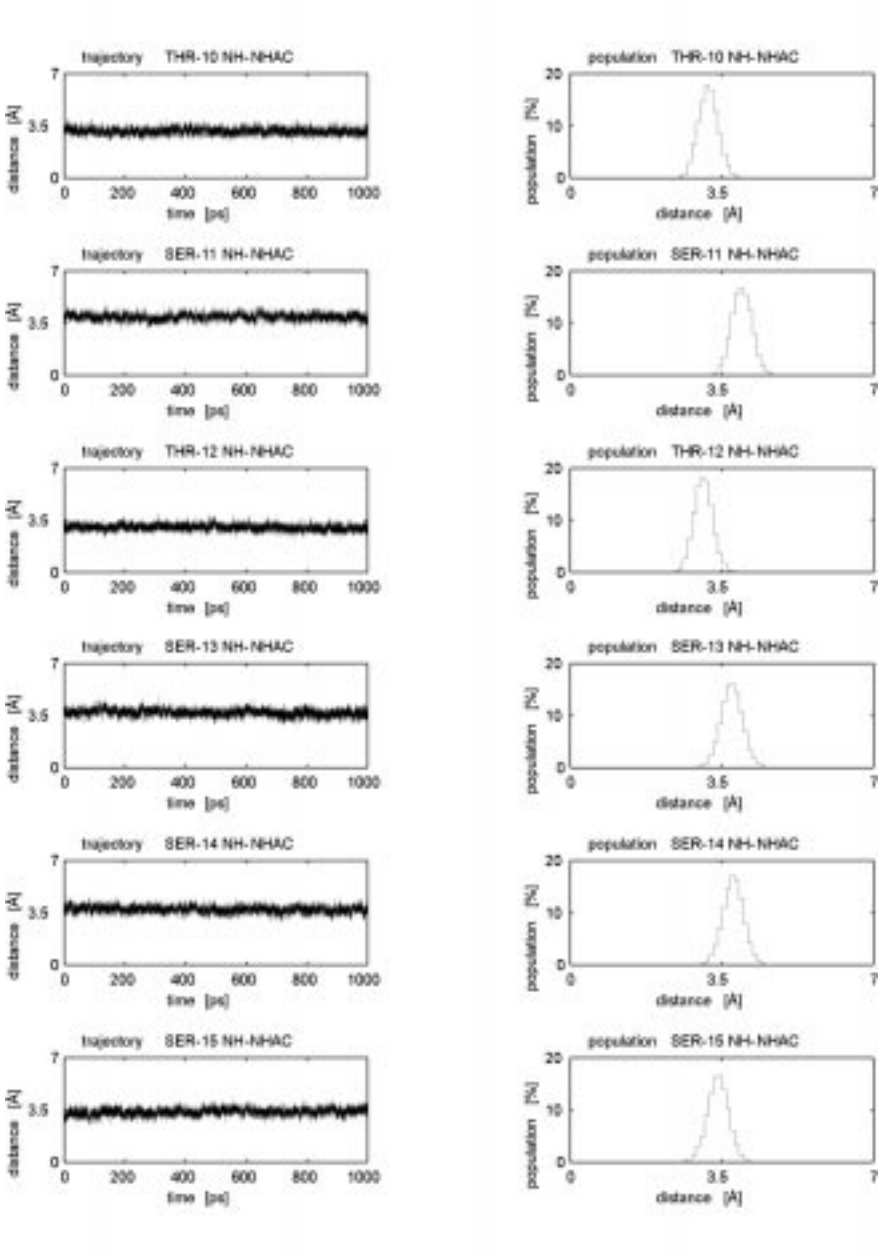

Figure 6. Trajectories and population plots of the amino acid NH proton – NHAc (GalNAc) proton distances, obtained from the 1000 ps constrained MD simulation.

corresponding GalNAc residue. These signals were integrated together and divided by two to obtain estimates for the inter proton distances. The shortest distances were observed for the threonines, 3.0 Å for Thr-12 NH to NHAc of GalNAc-3 and 3.1 Å for Thr-10 NH to NHAc of GalNAc-1. The corresponding distances of serine NH atoms to the connected GalNAc residues were significantly longer, between 3.3 Å and 3.9 Å (Figure 5). The distances found for the threonine-GalNAc contacts are very close to the minimum distance of 2.6 Å that is possible by considering the ϕ and χ_1 angle constraints. A ROESY spectrum showed that no significant chemical exchange took place and that the NOEs reported here are largely caused by dipolar coupling.

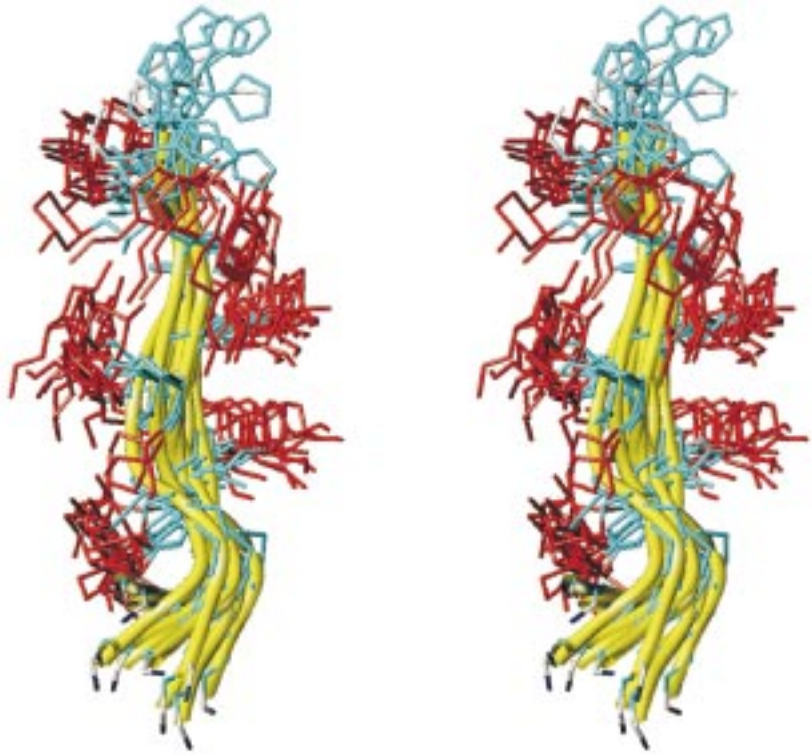

Figure 7. Stereo plot of a superposition from 10 structures, obtained from a 1000 ps constrained MD simulation. Each structure is a snapshot every 100 ps. The sugar residues are colored red, the peptide cyan. The peptide backbone conformation is highlighted by a yellow tube.

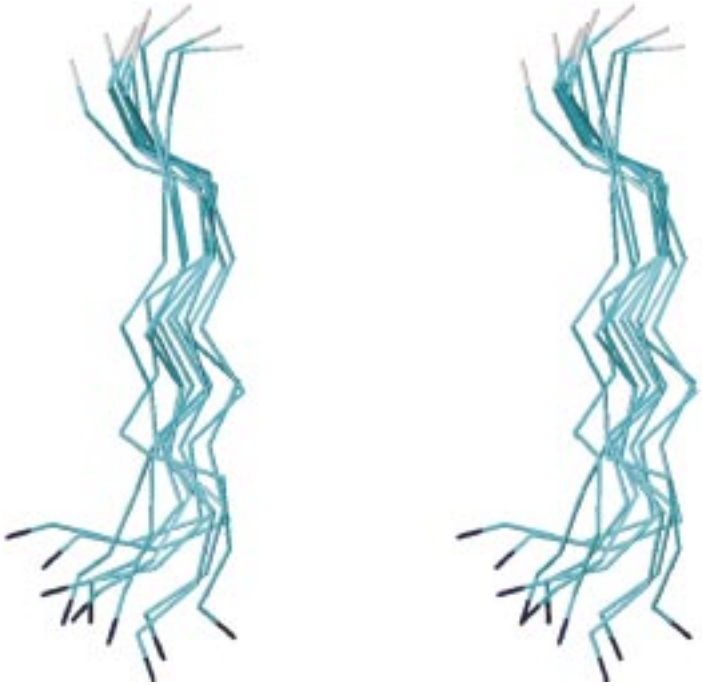

Figure 8. Stereo plot of a superposition from 10 structures, obtained from a 1000 ps constrained MD simulation. Each structure is a snapshot every 100 ps. Only the C_α atoms are shown. Note the 'wave-type', structure at each glycosylation site.

Altogether, we found 16 sequential NOEs and 14 intra-residue NOEs of the amino acids, 5 NOEs between protons of the GalNAc residues and those of amino acids and 6 intra-residue NOEs within GalNAc residues (Figure 5). All were used to calculate distances from build-up curves. We found no NH_i-NH_{i-1} contacts, characteristic for helical conformation, no mid-range and no long-range NOEs.

As expected, NOE interactions from H-1_{GalNAc} to CH_α protons and CH_β protons of serine residues and to CH_α, CH_β and CH_γ protons of threonine residues were observed. Due to overlap of the H-1 resonances with the water signal these NOEs could not be quantified. The interactions suggest a conformational preference for the *O*-glycosylation linkages of the synthetic glycopeptide.

The standard *trans* orientation of the NAc groups of the GalNAc residues was confirmed by strong NOEs from NH_{GalNAc} to H-3 and H-2. NOEs from the NH signals to the H-1 could not be identified due to overlap with the water resonance.

Molecular dynamics simulations

To investigate the dynamic behaviour and the conformation of the synthetic glycopeptide we ran a constrained MD simulation in water. The constraints originated from NOEs and coupling constants described above. First, an energy minimization was done starting from a β-strand conformation of the peptide backbone to obtain a collision-free structure. Attempting the same procedure from an α-helical start conformation gave no low energy structure that could be used for a constrained MD simulation. The β-strand structure was then surrounded by 1312 water molecules resulting in a 10% solution (w/w). The water box was energy-minimized. Subsequently, an MD simulation was run over 1000 ps.

During the MD simulation no constraints were violated, i.e. the average values obtained from the MD simulation deviated by less than 0.9 Å from those derived directly from the build-up curves of the NOEs. Most deviations were in the range of 0.2 Å. The backbone of the peptide remained in an extended structure over the simulation time. The flexibility of the backbone increases towards the C-terminus. The least flexibility is found in the glycosylated Thr_{10}-Ser_{11}-Thr_{12} cluster. The sugars showed some flexibility around the glycosidic linkages which is higher at Thr-10 and Ser-15 than at the linkages located between these terminal glycosylation sites. In general, GalNAc residues on

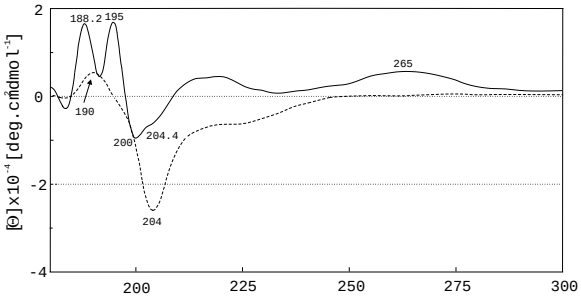

Figure 9. CD spectra of the hexaglycosylated decapeptide (solid line) and of the unglycosylated decapeptide (dashed line). The unglycosylated peptide shows a CD spectrum that is characteristic for a random coil structure. The CD spectrum of the hexaglycosylated decapeptide is characterized by two maxima at 188 nm and 195 nm, respectively and a strong negative band at 200 nm with a shoulder at 204 nm. Additionally, the CD band at 265 nm which originates from the aromatic region of the His-9 suggests that this amino acid is highly ordered in the glycopeptide compared to the unglycosylated reference molecule.

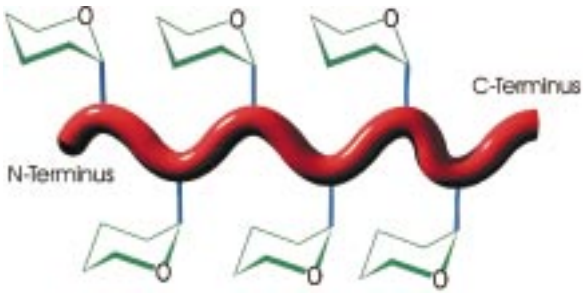

Figure 10. Glycosylation with a monosaccharide in a cluster seems to introduce a structural motif in the backbone that is characterized by alternating positions of the carbohydrates along the peptide backbone accompanied with a 'wave-type' conformation of the peptide backbone. The orientation of all GalNAc residues is directed towards the *N*-terminus.

threonine residues seem to have a lower mobility than those on serine residues.

All except one standard deviation for the interproton distance constraints were in the range of 0.1 Å to 0.25 Å. The distance distributions have an almost symmetrical shape. For the distances of the amide protons to the $CH_{\alpha i-1}$ the minimum and maximum distances were about 0.5 Å away from the mean value. For the distances of the amide protons of the amino acids to the amide protons of the GalNAc residues the minimum and maximum distance were 0.9 Å away from the mean (Figure 6). These values indicate an increase in flexibility of the carbohydrates relative to the peptide backbone.

^3J (NH,CH_α) coupling constants were calculated using all conformations generated over the total simulation time and differ a maximum of 0.6 Hz from

the experimental values. The standard deviations of each of the $\phi_{peptide}$ dihedral angles are between 6.3° and 8.0°. In Figure 6 trajectories and population plots are shown for the distances from the amide proton of the amino acids to the amide proton of the GalNAc residues.

The population maxima of the ϕ angles of GalNAc residues show values between $-50°$ and $-63°$. Ψ angles of the threonine linked GalNAc residues have their population maximum at about 170°. GalNAc residues linked to serine residues 11 and 14 have only one population maximum at about 180° whereas those linked to serine 13 and 15 have two population maxima at 180° and 70°, populated. Threonine-linked GalNAc residues have a lower flexibility during the MD simulation than those attached to serine residues.

The overlay of 10 structures from the constrained MD simulation shows an inflexible area within the glycosylation cluster (cf. Figures 7 and 8). The flexibility increases towards the C-terminus. The structure of the peptide and the orientation of the sugar residues to the *N*-terminus and alternating on opposite sides of the peptide are conserved over the time of the MD simulation.

During the MD simulation the sugar residues were largely oriented towards the *N*-terminus as identified through the $NH_{peptide}$ backbone to NHAc NOEs. Depending on the length of the experimentally determined distance the sugar residues have a lower or higher mobility.

The conformations adopted during the MD simulation are in full agreement with the NMR data.

Less than 1.2 kcal/mol in constraint violation energy is obtained for distances within the backbone during the MD simulation with average violation energies of ≈ 0.2 kcal/mol. Therefore, the constraints do not impose significantly on the 3D structure of the peptide. The violation of the $NH_{peptide}$-NH_{sugar} constraints are less than 2.5 kcal/mol during the MD simulation with average violation energies of ≈ 0.3 kcal/mol.

An earlier MD simulation of a mucin type glycopeptide (Butenhof and Gerken, 1993) showed that the sugar residues should be oriented towards the C-terminus with a H-bond to the backbone carbonyl group. This is in contrast with our experimental and theoretical findings reported here.

CD spectra

The CD spectrum of the synthetic glycopeptide differs significantly from the spectrum of the corresponding unglycosylated decapeptide (cf. Figure 9). The peptide shows a CD spectrum that is characteristic of a random coil structure with a negative CD band at 204 nm and a positive at 190 nm. In contrast we find very unusual patterns for the CD spectrum of the hexaglycosylated decapeptide, characterized by two maxima at 188 nm and 195 nm, respectively, and a strong negative band at 200 nm with a shoulder at 204 nm. The changes of the CD patterns from the peptide to the glycopeptide are not caused by the α-D-GalNAc residues. The CD spectrum of GalNAc was reported (Coduti et al., 1977) to have a negative CD at 209 nm and a positive slope up to 190 nm. In contrast, we observe about zero intensity for the glycopeptide at 209 nm. The CD bands of sugar and peptide portion of *N*-type glycopeptides were shown to be additive (Watanabe et al. 1981). Therefore, the dual positive bands at 188 nm and 195 nm observed here and the reduction of the intensity of the negative CD at 204 nm of the unglycosylated peptide by a factor of about four cannot be caused by a mere addition of the GalNAc residues. Thus, we assume that the CD pattern of the glycopeptide is characteristic for the conformation of the peptide backbone. Further support comes from the CD bands at about 265 nm which originate from the aromatic region of the His-9 residue suggesting that this amino acid is highly ordered in the glycopeptide compared to the unglycosylated reference molecule.

Conclusions

Glycosylation of peptides with monosaccharides arranged in a cluster seems to introduce a structural motif in the backbone that is characterized by alternating positions of the carbohydrates along the peptide backbone accompanied with a 'wave-type' conformation (Figure 10) of the peptide backbone where the carbohydrate attachment sites alternate between the tops and bottoms of the wave. The plane of the wave can tilt somewhat between glycosylation sites. Overall, this results in an extended conformation of the peptide backbone. The data on the 3D structure of the synthetic glycopeptide reported here further refine earlier data obtained on the T1 fragment isolated from human glycophorin A^N (Pieper et al., 1996). The major effect of the glycosylation on the 3D structure of the glycopeptide in terms of extending the peptide backbone is already obtained by glycosylation with a monosaccharide. As pointed out earlier glycosylation with a tetrasaccharide on each glycosylation site

should amplify the effect of extending the peptide backbone. This is in agreement with data obtained from light scattering analysis of mucins (Shogren et al., 1989). The extension of the peptide backbone can aid in presenting the neuramic acids on the outside of the erythrocytes and, thus, effect the interaction with antibodies and cellular receptors.

Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft through Resource Center Grant SFB 470, by the Fonds der Chemischen Industrie and by the State of Hamburg is acknowledged.

References

Anderson, R.A. and Lovrien, R.E. (1984) *Nature*, **307**, 655–658.
Anderson, R.A. and Marchesi, V.T. (1985) *Nature*, **318**, 295–298.
Batstone-Cunningham, R.L., Hardy, R.E., Daman, M.E. and Dill, K. (1983) *Biochim. Biophys. Acta*, **746**, 1–7.
Butenhof, K.J. and Gerken, T.A. (1993) *Biochemistry*, **10**, 2650–2663.
Coduti, P.L., Gordon, E.C. and Bush, C.A. (1977) *Anal. Biochem.*, **78**, 9–20.
de Beer, T., van Zuylen, C.W., Leeflang, B.R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P. and Vliegenthart, J.F. (1996) *Eur. J. Biochem*., **241**, 229–242.
Dill, K., Carter, R.D., Lacombe, J.M. and Pavia, A.A. (1985) *Carbohydr. Res.*, **152**, 217–228.
El Ouagari, K., Teissie, J. and Benoist, H. (1995) *J. Biol. Chem.*, **45**, 26970–26975.
Fukuda, M., Lauffenburger, M., Sasaki, H., Rogers, M.E. and Dell, A.J. (1987) *J. Biol. Chem.*, **262**, 11952–11957.
Furthmayr, H. (1978) *Supramol. Struct.*, **9**, 195–211.
Furthmayr, H. (1978) *Nature*, **271**, 519–524.
Gasteiger, J. and Marsili, M. (1980) *Tetrahedron*, **36**, 3219–3228.
Klich, G., Paulsen, H., Meyer, B., Meldal, M. and Bock, K. (1997) *Carbohydr. Res.*, **299**, 33–48.
Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) *J. Mol. Biol.*, **217**, 731–736.
Marchesi, V.T. (1979) *Semin. Hematol.*, **16**, 3–20.
Paulsen, H., Pollex-Kruger, A. and Sinnwell, V. (1991) *Carbohydr. Res.*, **214**, 199–226.
Pavia, A.A. and Ferrari, B. (1983) *Int. J. Pept. Protein Res.*, **22**, 539–548.
Pieper, J., Ott, K.H. and Meyer, B. (1996) *Nat. Struct. Biol.*, **3**, 228–232.
Shogren, R., Gerken, T.A. and Jentoft, N. (1989) *Biochemistry*, **28**, 5525–5536.
Sybyl 6.3 (1996), *Force Field Manual*, Tripos, St. Louis, MO, pp. 359–371.
Watanabe, K., Matsuda, T. and Sato, Y. (1981) *Biochim. Biophys. Acta*, **667**, 242–250.
Withka, J.M., Wyss, D.F., Wagner, G., Arulanandam, A.R., Reinherz, E.L. and Recny, M.A. (1993) *Structure*, **1**, 69–81.
Wyss, D.F., Dayie, K.T. and Wagner, G. (1997) *Protein Sci.*, **6**, 534–542.