Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimie 85 (2003) 275–286
BIOCHIMIE
www.elsevier.com/locate/biochi

Enzymatic synthesis of the core-2 sialyl Lewis X O-glycan on the tumor-associated MUC1a' peptide

Ricardo Gutiérrez Gallego [a,1,3], Gregor Dudziak [b,3], Udo Kragl [b,2], Christian Wandrey [b], Johannis P. Kamerling [a], Johannes F.G. Vliegenthart [a,*]

[a] *Section of Glycoscience and Biocatalysis, Department of Bio-Organic Chemistry, Bijvoet Center, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands*
[b] *Forschungszentrum Jülich GmbH, Institut für Biotechnologie, 52425 Jülich, Germany*

Received 21 November 2002; accepted 3 February 2003

Abstract

Starting from a tumor-associated synthetic MUC1-derived peptide MUC1a' and using a completely enzymatic approach for the synthesis of the core-2 sialyl Lewis X glycopart, the following glycopeptide was synthesized: AHGV{Neu5Ac(α2-3)Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-6)[Gal(β1-3)]GalNAc(α1-O)}TSAPDTR. First, polypeptide *N*-acetylgalactosaminyltransferase 3 was used to site-specifically glycosylate MUC1a' to give MUC1a'-GalNAc. Then, in a one-pot reaction employing β-galactosidase and core-2 β6-*N*-acetylglucosaminyltransferase the core-2 O-glycan structure was prepared. The core-2 structure was then sequentially galactosylated, sialylated, and fucosylated by making use of β4-galactosyltransferase 1, α3-sialyltransferase 3, and α3-fucosyltransferase 3, respectively, resulting in the sialyl Lewis X glycopeptide. The overall yield of the final compound was 23% (3.2 mg, 1.4 μmol). During the synthesis three intermediate glycopeptides containing O-linked GalNAc, Gal(β1-4)GlcNAc(β1-6)[Gal(β1-3)]GalNAc, and Neu5Ac(α2-3)Gal(β1-4)GlcNAc(β1-6)[Gal(β1-3)]GalNAc, respectively, were isolated in mg quantities. All products were characterized by mass spectrometry and NMR spectroscopy.

Keywords: MUC1; O-glycopeptide; Enzymatic synthesis; Core-2 sialyl Lewis X

1. Introduction

Mucins are high-molecular-mass glycoproteins, characterized by the presence of a large extracellular domain made up of 30-100 tandem repeats of peptide sequences and a high degree of O-glycosylation (up to 90% of the glycoprotein by weight). Today, 14 human mucins have been cloned or partially cloned [1,2]. MUC1, the first one cloned, exhibits a ubiquitous organ expression, except skin epithelium and mesenchymal tissues [1]. Amongst the proposed functions of MUC1 are lubrication and cleaning of epithelial surfaces, protective adhesion to ingested pathogens [3], involvement in cell-signaling events [4], cell-adhesion [5], and tumor progression and metastasis [6]. The glycosylation of MUC1 in many organs is still unknown, but for breast cancer cells detailed data on the O-glycosylation have been reported [7–12]. Recently, the O-glycosylation profiles of a recombinant probe corresponding to six MUC1 tandem repeats expressed in the breast cancer cell lines T47D, MCF-7, MDA-MB231 and ZR75-1 have been published [12], and it was demonstrated that the latter three cell lines glycosylate the MUC1 repeat peptide preferentially with core 2-type glycans. Structural analysis of the O-glycans showed the presence of Gal(β1-4)GlcNAc(β1-6)[Gal(β1-3)]GalNAc for MCF-7, MDA-MB231 and ZR75-1 cells, of Neu5Ac(α2-3)Gal(β1-4)GlcNAc(β1-6)[Gal(β1-3)]GalNAc for MDA-MB231 and ZR75-1 cells, and of Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-6)[Gal(β1-3)]GalNAc for MCF-7 cells. Endogenous and recombinant MUC1 probes from T47D and MCF-7 cells revealed per cell type almost identical O-glycosylation patterns [12].

* Corresponding author. Tel.: +31-30-253-2281; fax: +31-30-254-0980.
E-mail address: j.f.g.vliegenthart@chem.uu.nl (J.F.G. Vliegenthart).

[1] Present address: Municipal Institute for Medical Research, C/Dr. Aiguader 80, E-08003 Barcelona, Spain.

[2] Present address: Fachbereich Chemie, Lehrstuhl für Technische Chemie, Universität Rostock, Albert-Einstein-Strasse 3a, 18059 Rostock, Germany.

[3] R. Gutiérrez Gallego and G. Dudziak contributed equally to this publication.

DOI: 10.1016/S0300-9084(03)00050-6

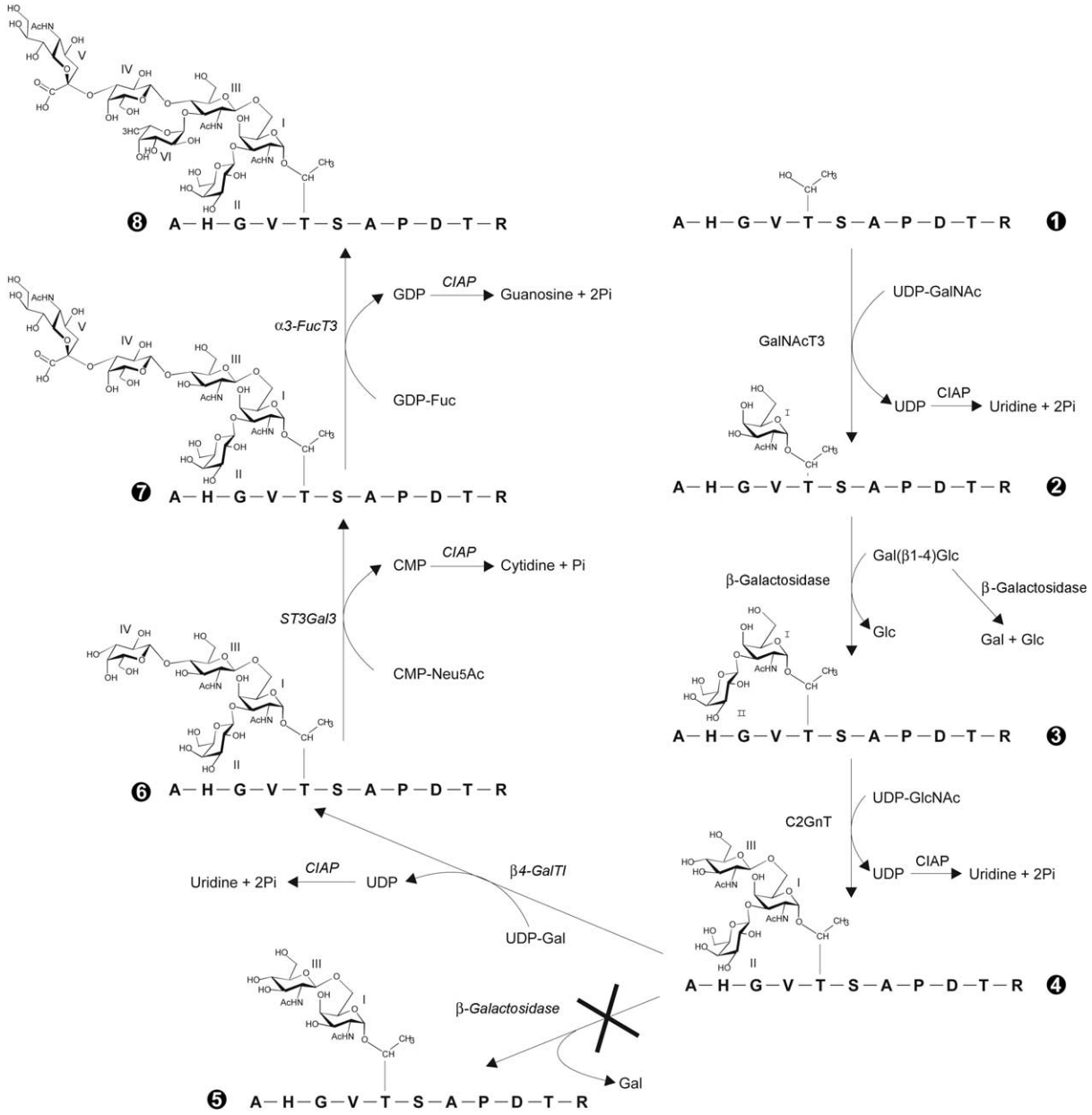


Fig. 1. Schematic representation of the sequential synthesis of core-2 sialyl Lewis X, O-glycosidically linked to the MUC1a' peptide. ✕ means that **4** is not converted into **5**.

Here, we describe the completely enzymatic synthesis (Fig. 1) of a core-2 sialyl Lewis X O-glycan on the tumor-associated MUC1a' peptide (AHGVTSAPDTR), a compound directly related to the O-glycans mentioned above. By making use of a selective polypeptide *N*-acetylgalactosaminyltransferase (GalNAcT3) we could ensure a site-specific glycosylation. The absence of a recombinant β3-galactosyltransferase in the build-up of the core-2 structure was overcome by the use of a "one-pot" strategy employing a β-galactosidase in combination with a core-2 β6-*N*-acetylglucosaminyltransferase (C2GnT). The elongation of the GlcNAc residue with Gal using β4-galactosyltransferase 1 (β4-GalT1) (yielding an O-glycan from MCF-7, MDA-MB231 and ZR75-1 cells), Neu5Ac using α3-sialyltrans-

Abbreviations: BSA, bovine serum albumin; C2GnT, core-2 β6-*N*-acetylglucosaminyltransferase; CIAP, calf intestine alkaline phosphatase; CID, collision-induced dissociation; CMP, cytidine monophosphate; ESI, electrospray ionization; α3-FucT3, α3-fucosyltransferase 3; GalNAcT3, polypeptide *N*-acetylgalactosaminyltransferase 3; β4-GalT1, β4-galactosyltransferase 1; GDP, guanosine diphosphate; HPLC, high performance liquid chromatography; MALDI-TOF, matrix assisted laser desorption ionization time of flight; MES, 2-(*N*-morpholino)ethanesulfonic acid; MLEV, Malcolm Levitt; MS, mass spectrometry; MS-MS, tandem mass spectrometry; NMR, nuclear magnetic resonance; PMSF, phenylmethylsulfonyl fluoride; ROESY, rotating frame nuclear Overhauser enhancement spectroscopy; ST3Gal3, α3-sialyltransferase 3; TOCSY, total correlation spectroscopy; UDP, uridine diphosphate; WEFT, water eliminated Fourier transform.

Table 1
Summary of the enzymes used for the enzymatic synthesis of the core-2 sialyl Lewis X glycan on the MUC1a' peptide. Yields are given with reference to the acceptor substrate for the individual reaction steps and with reference to the first acceptor substrate (MUC1a' peptide) for the accumulated steps

Enzyme	Activity (mU/ml)	Yield (%)	Total yield (%)	Total yield (mg)
GalNAcT3	20	91 (**2**)	91	48
Galactosidase/C2GnT	1000/200	29 (**4**)	28	17
β4-GalT1	20	84 (**6**)	27	7.8
ST3Gal3	30	83 (**7**)	24	3.5
α3-FucT3	50	86 (**8**)	23	3.2

ferase 3 (ST3Gal3) (yielding an O-glycan from MDA-MB231 and ZR75-1 cells), and Fuc using α3-fucosyltransferase 3 (α3-FucT3) (compare with Lewis X O-glycan in MCF-7 cells) afforded the core-2 sialyl Lewis X epitope.

2. Materials and methods

2.1. Materials

MUC1a' peptide AHGVTSAPDTR was purchased from Neosystems (Strasbourg). Polypeptide *N*-acetylgalactosaminyltransferase 3 (GalNAcT3, EC 2.4.1.41) and core-2 β6-*N*-acetylglucosaminyltransferase (C2GnT, EC 2.4.1.102) were expressed and purified as described [13]. β-Galactosidase from bovine testes (EC 3.2.1.23) was obtained from Oxford Glycosciences, BSA, β4-galactosyltransferase 1 (β4-GalT1, EC 2.4.1.38), and calf intestine alkaline phosphatase (CIAP, EC 3.1.3.1) from Sigma, and α3-fucosyltransferase 3 (α3-FucT3, EC 2.4.1.65) and α3-sialyltransferase 3 (ST3Gal3, EC 2.4.99.5) from CalBiochem. All other chemicals were of the highest grade commercially available.

2.2. Mass spectrometry

Positive-ion mode MALDI-TOF mass spectrometric analysis of the products was performed on a Voyager-DE (PerSeptive Biosystems) instrument operating at an accelerating voltage of 22 kV (grid voltage, 92%; ion guide wire voltage, 0.1%) and equipped with a VSL-337ND-N_2 laser. The samples were dissolved in bidistilled water (1 μg/μl), mixed in the sample well with α-cyano-4-hydroxy-cinnamic acid (10 mg/ml in water:acetonitrile 1:1, v/v) at a ratio of 1:3, and allowed to crystallize at room temperature. Linear mass scans were recorded over 3000 Da using a pulse delay time of 90 ns. Recorded data were processed using GRAMS/386 software (v. 3.04, Galactic Industries Corporation).

Positive-ion mode ESI mass spectrometric analyses of compound **2** were performed on an Esquire-LC(TM) quadrupole ion-trap spectrometer (Bruker Daltonik GmbH). Samples were dissolved in 100 μl deionized water:acetonitrile:acetic acid (50:50:0.1, v/v/v), and were introduced by infusion at a flow rate of 2 μl/min. Ions were scanned between 50 and 1600 Da with a scan speed of 13 000 Da/s at unit resolution using resonance ejection at the hexapole resonance of one-third of the radio frequency (rf = 781 250 Hz). The calibration of the mass spectrometer was performed using ES tuning mix (Hewlett-Packard). CID MS-MS experiments were performed by using the quadrupole ion-trap to select the precursor ion for fragmentation. Helium was used as a collision gas, and the fragmentation energy applied on the endcaps varied between 0.5 and 1.8 V. Recorded data were processed using Esquire NT 3.1 software (Bruker) and interpreted using MS Bio Tools version 1.0 (Bruker).

2.3. NMR spectroscopy

For experiments performed in ^2H_2O, compounds were repeatedly exchanged in ^2H_2O (99.9 at.% ^2H, Isotec) with intermediate lyophilization, and finally dissolved in 450 μl ^2H_2O (99.96 at.% ^2H, Isotec). Resolution-enhanced ^1H 1D and 2D NMR spectra were recorded on Bruker DRX-500 or DRX-600 (Department of NMR Spectroscopy, Utrecht University and NSR Center, Nijmegen University) spectrometers at a probe temperature of 300 K. Chemical shifts (δ) were expressed in parts per million relative to internal acetate (δ 1.908; acetone δ 2.225). HO^2H signal suppression was achieved by applying a WEFT pulse sequence [14] in 1D ^1H experiments and by presaturation for 1 s in 2D experiments. 2D TOCSY spectra were recorded by using MLEV-17 mixing sequences [15] with effective spin-lock times between 20 and 100 ms. 2D ROESY [16] or 2D off-resonance ROESY [17] spectra were recorded with a mixing time of 250 ms. The spin-lock field strength corresponded to a 90° pulse of approximately 120 μs. ^1H 1D and 2D spectra were processed on Silicon Graphics IRIS work stations (Indigo 2 and O2) using XINSP2 software (Bijvoet Center, Department of Bio-Organic Chemistry).

3. Synthesis of MUC1a' glycopeptides

An overview of the sequential synthesis of the core-2 sialyl Lewis X MUC1a' O-glycopeptide (**8**) is depicted in Fig. 1, whereas the yields are given in Table 1. In the following section, details concerning the individual synthetic steps are summarized. The yields reported for the incubations are based on the total reaction mixture.

*3.1. AHGV[GalNAc(α1-O)]TSAPDTR (**2**)*

For the incubation, a Tris-HCl (50 mM, pH 7) buffer containing 10 mM $MnCl_2$, 0.5 mg/ml BSA, and 10 U/ml CIAP was used. To a solution of incubation buffer (4 ml) was added 10 mM MUC1a' (**1**) (45 mg), 12 mM UDP-GalNAc and 80 mU GalNAcT3. The mixture was incubated for 20 h at 30 °C. Then, the product was purified from the mixture by

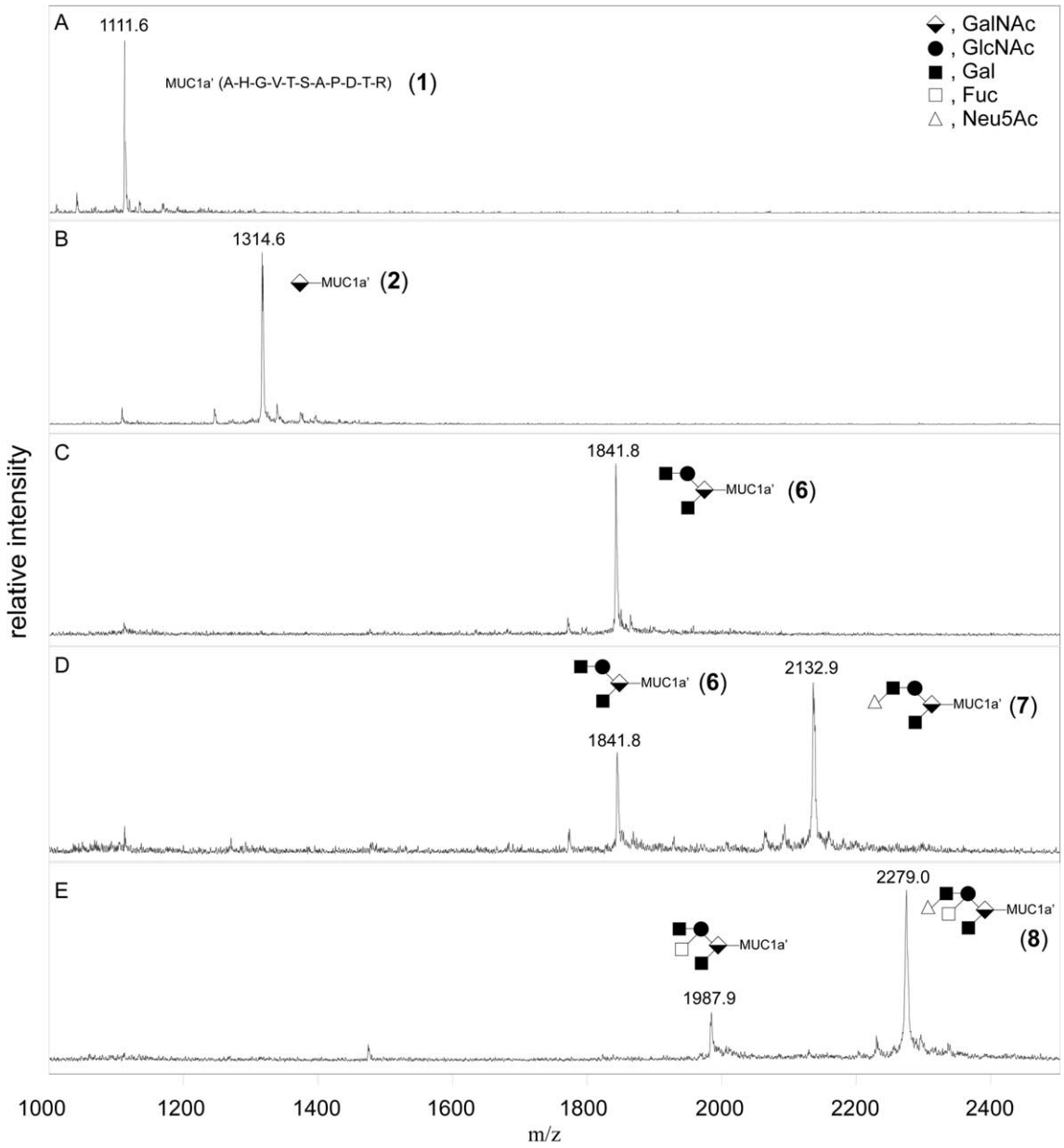


Fig. 2. Positive-ion mode MALDI-TOF mass spectra of MUC1a' peptide (A), intermediate glycopeptides (B-D), and the final product (E).

preparative HPLC, yielding 48 mg **2** (36.5 µmol, 91% with reference to the acceptor substrate). Compound **2** was characterized in detail, and used for further glycosylation (vide infra).

*3.2. AHGV{GlcNAc(β1-6)[Gal(β1-3)]GalNAc(α1-O)} TSAPDTR (**4**)*

For the incubation, a MES (50 mM, pH 6) buffer containing 10 mM $MgCl_2$, 0.5 mg/ml BSA, and 10 U/ml CIAP was used. To inhibit proteolytic digestion, PMSF (100 µg/ml), pepstatin (1 µg/ml), and leupeptin (0.5 µg/ml) were added. To a solution of incubation buffer (1.8 ml) was added 20 mM **2** (48 mg), and the pH was adjusted to 6. Then, 0.5 M lactose, 25 mM UDP-GlcNAc, 1.8 U β-galactosidase, and 360 mU C2GnT were added, and the solution was incubated for 40 h at 30 °C. Isolation of the product by preparative HPLC yielded 17 mg **4** (10.1 µmol, 29% with reference to the acceptor substrate), which was used for further glycosylation (vide infra).

*3.3. AHGV{Gal(β1-4)GlcNAc(β1-6)[Gal(β1-3)] GalNAc(α1-O)}TSAPDTR (**6**)*

For the incubation, a MES (50 mM, pH 6.5) buffer containing 10 mM $MgCl_2$, 0.5 mg/ml BSA, and 2 U/ml CIAP was used. To a solution of incubation buffer (12 ml) were added, 1 mM **4** (17 mg), 1.2 mM UDP-Gal, and 20 mU/ml β4-GalT1, and the solution was incubated for 8 h at 30 °C. Approximately half of the incubation mixture was used directly for a sialyltransferase experiment (vide infra), and the remaining mixture was purified by preparative HPLC, yield-

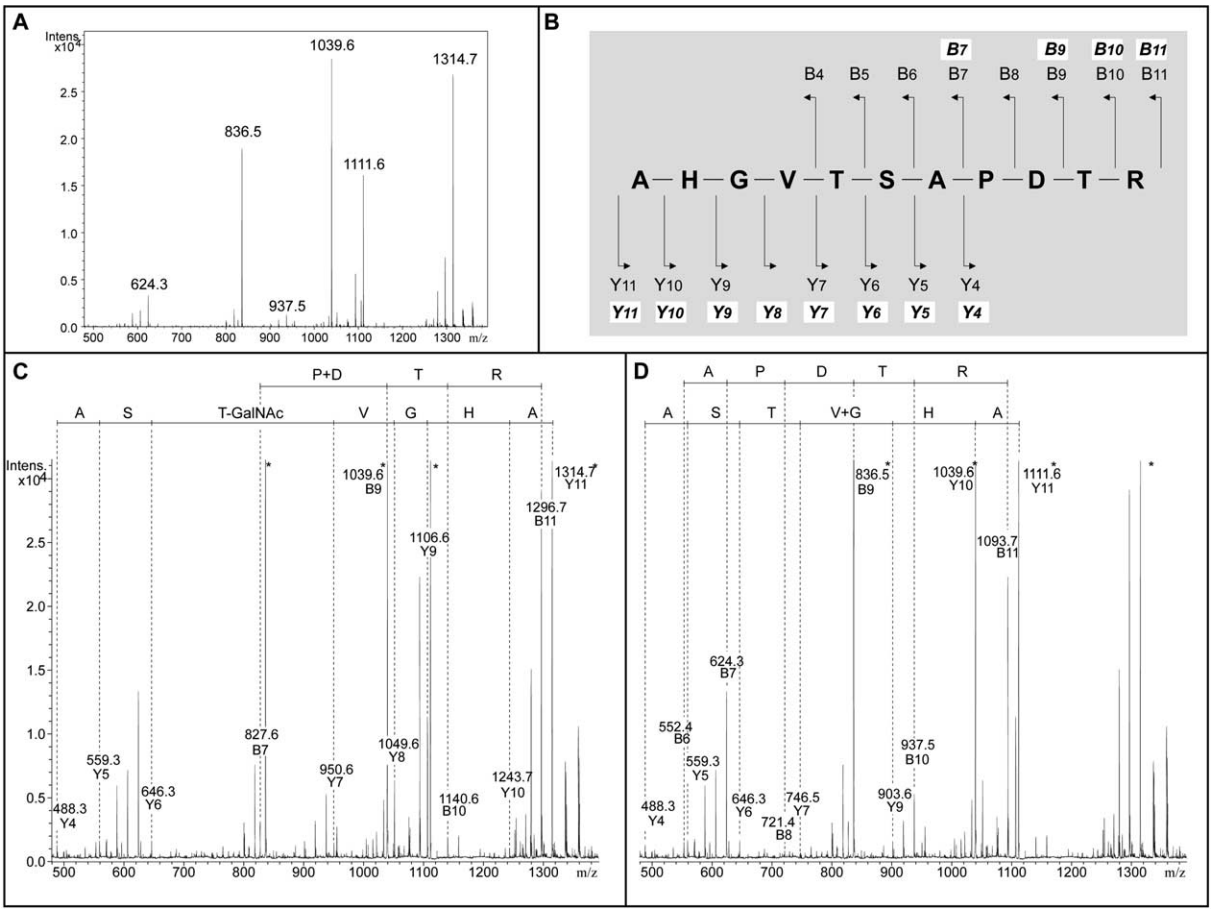


Fig. 3. (A) Positive-ion mode CID-MS-MS spectrum of the pseudo-molecular-ion (*m*/z 1314.7). (B) The fragments observed in C and D. The daughter ions marked in italics originate from the GalNAc-containing glycopeptide. The other daughter ions originate from the deglycosylated peptide. (C) Sequence ions of the glycopeptide. (D) Sequence ions after the loss of GalNAc (as 2-acetamidogalactal). Peaks marked with an asterix (*) are truncated at one third of the original intensity (see Fig. 3A).

ing 7.8 mg **6** (4.2 µmol, 84% with reference to the acceptor substrate), which was characterized in detail.

*3.4. AHGV{Neu5Ac(α2-3)Gal(β1-4)GlcNAc(β1-6)[Gal(β1-3)]GalNAc(α1-O)}TSAPDTR (**7**)*

For the incubation, 5 ml of the incubation mixture containing **6** (vide supra) was used. To the incubation buffer was added 2 mM CMP-Neu5Ac and 30 mU/ml ST3Gal3. The mixture was incubated for 24 h at 20 °C. Approximately one half of the solution was used directly for a α3-FucT3 experiment (vide infra), and the remaining mixture was purified by preparative HPLC, yielding 3.5 mg **7** (1.6 µmol, 83% with reference to the acceptor substrate), which was characterized in detail.

*3.5. AHGV{Neu5Ac(α2-3)Gal(β1-4)[Fuc(α1-3)]GlcNAc(β1-6)[Gal(β1-3)]GalNAc(α1-O)}TSAPDTR (**8**)*

For the incubation, 2.3 ml of the incubation mixture containing **7** (vide supra) was used. To the incubation buffer was added 2 mM GDP-Fuc and 50 mU/ml α3-FucT3. The reaction mixture was incubated for 8 h at 20 °C. HPLC purification yielded 3.2 mg **8** (1.4 µmol, 86% with reference to the acceptor substrate), which was characterized in detail.

3.6. Chromatographic purification

After each enzymatic incubation step, product was purified by preparative reversed phase HPLC (Pharmacia LKB) using a Sephasil Peptide C18 reversed phase column (4.6 × 250 mm, Pharmacia), equilibrated with solvent A (aqueous 95% acetonitrile in 0.1% trifluoroacetic acid), at a flow rate of 2.0 ml/min. Elutions were carried out with a linear gradient from 100% solvent A-0% solvent B (aqueous 5% acetonitrile in 0.1% trifluoroacetic acid) to 61% solvent A-39% solvent B in 30 min. The effluent was monitored by UV at 257 nm, and collected fractions were lyophilized.

4. Results

4.1. Synthesis of MUC1a' glycopeptides

In previous studies it was shown that both GalNAcT1 and GalNAcT3 efficiently glycosylate MUC1a' (**1**) [18,19], and in the present study GalNAcT3 was selected. Of the three potential glycosylation sites T5, S6 and T10, T5 was selectively occupied after 20 h of incubation (**2**; vide infra), as revealed by HPLC analysis (data not shown), in a yield of 91% (Table 1). UDP-induced deactivation of the enzyme was overcome by the addition of CIAP.

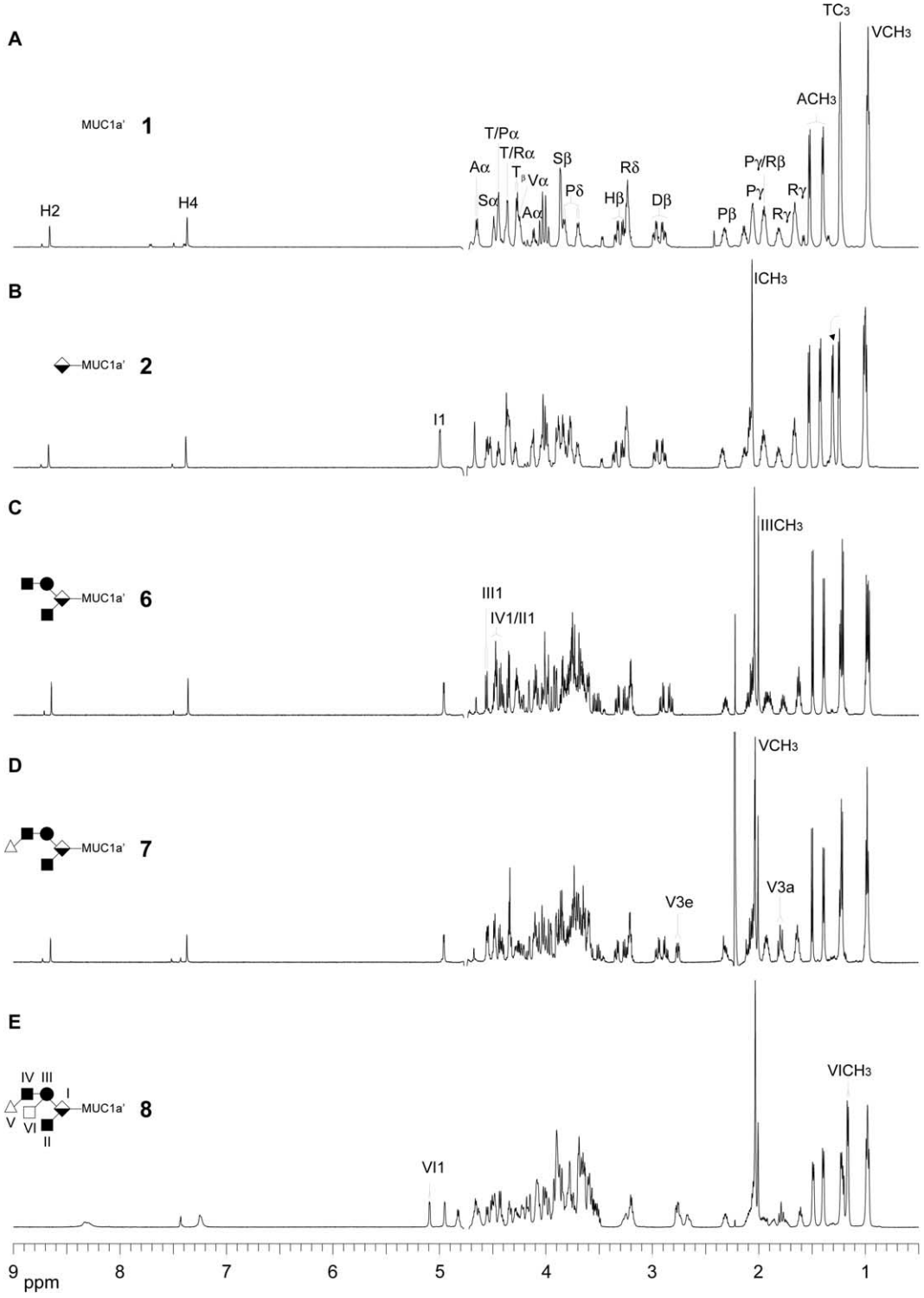


Fig. 4. 600 MHz ^{1}H NMR spectra of MUC1a' peptide (**1**, A), intermediate glycopeptides (**2**, B; **6**, C; **7**, D), and the final sialyl Lewis X containing MUC1a' glycopeptide (**8**, E). The assignments in A correspond to the amino acids; in B-E the incorporation of monosaccharides is indicated (**I**, GalNAc; **II**, Gal(β1-3); **III**, GlcNAc; **IV**, Gal(β1-4); **V**, Neu5Ac; **VI**, Fuc).

For the generation of **4**, rather than using an enriched β3-GalT preparation [20] which resulted in relatively low overall yield, an earlier reported "one-pot" strategy was preferred, employing both β-galactosidase and C2GnT [13,21]. In a pilot experiment using lactose as donor substrate and β-galactosidase from bovine testes, **2** was tested as acceptor substrate for galactosylation. Protease activities, present in the galactosidase preparation were efficiently suppressed using a protease-inhibitor cocktail. Both HPLC and capillary electrophoresis analysis (data not shown) revealed the formation of one new product. The low yield (~10%) was attributed to the hydrolysis of the generated product. Subsequently, the

Table 2
600 MHz ^{1}H NMR chemical shifts of the amino acids in five (glyco)peptides (chemical shifts are relative to internal acetone, δ 2.225) measured in HO^{2}H at 300 K. n.d., not determined

Proton	**1**	**2**	**6**	**7**	**8**
A1α	4.113	4.105	4.093	4.096	4.093
A1β	1.521	1.518	1.497	1.498	1.493
H2α	4.743	4.740	4.717	4.731	4.668
H2β	3.338, 3.266	3.330, 3.263	3.331, 3.253	3.252, 3.335	3.199, 3.262
H22	8.662	8.653	8.646	8.651	8.314
H24	7.370	7.363	7.363	7.371	7.245
G3α	4.020	n.d.	n.d.	n.d.	n.d.
V4α	4.275	4.340	4.361	4.348	4.336
V4β	2.137	2.119	2.105	2.096	2.093
V4γCH$_3$	0.974	0.991	0.979	0.981	0.978
T5α	4.443	4.658	4.659	4.674	4.673
T5β	4.272	4.331	4.286	4.283	4.282
T5γCH$_3$	1.235	1.297	1.235	1.235	1.231
S6α	4.488	4.506	4.471	4.484	4.478
S6β	3.863	3.817, 3.879	3.857, 3.789	3.856, 3.789	3.841, 3.871
A7α	4.648	4.541	4.486	4.489	4.489
A7β	1.398	1.414	1.391	1.390	1.400
P8α	4.447	4.435	4.411	4.409	4.418
P8β	2.325	2.332	2.317	2.318	2.317
P8γCH$_2$	2.059, 1.955	2.077, 1.946	2.058, 1.930	2.061, 1.931	2.057, 1.941
P8δCH$_2$	3.826, 3.698	3.844, 3.691	3.821, 3.668	3.821, 3.669	3.804, 3.660
D9α	4.765	4.751	4.711	4.734	4.634
D9β	2.892, 2.977	2.883, 2.961	2.829, 2.913	2.872, 2.950	2.747, 2.664
T10α	4.361	4.358	4.348	4.342	4.352
T10β	4.250	4.276	4.269	4.258	4.259
T10γCH$_3$	1.235	1.236	1.213	1.218	1.205
R11α	4.377	4.339	4.269	4.334	4.182
R11β	1.811, 1.949	1.806, 1.932	1.770, 1.897	1.777, 1.927	1.752, 1.869
R11γCH$_2$	1.660	1.656	1.625	1.637	1.611
R11δCH$_2$	3.233	3.232	3.205	3.210	3.201

incubation of **2** was repeated including C2GnT and UDP-GlcNAc. After HPLC purification, **4** was obtained in a yield of 29%. No formation of the core-6 glycopeptide (**5**, vide infra) could be observed, in contrast to previous studies [13] whereby, using GalNAc(α1-OBn) as acceptor substrate, both GlcNAc(β1-6)GalNAc(α1-OBn) (yield 56%) and GlcNAc (β1-6)[Gal(β1-3)]GalNAc(α1-OBn) (yield 34%) were formed. Apparently, **2** is not such a good acceptor substrate, and after incorporation of GlcNAc the glycopeptide (**4**) ceases to be a substrate for β-galactosidase. It should be noted that after 1 d the formation of **4** had come to an end (data not shown). UDP, UMP and uridine concentrations were stable indicating that no UDP-GlcNAc was being metabolized after 1 d. It is suggested that the protease-inhibitor cocktail influences the activity of one or both enzymes since this restricted conversion was not observed in an identical experimental set-up, without the protease inhibitors, using GalNAc(α1-OBn) as acceptor substrate [13].

In order to prevent substrate induced inhibition of β4-GalT1 [22] the galactosylation was performed using only 1 mM of acceptor substrate and a 20% excess of donor substrate. Using these conditions the acceptor substrate could no longer be observed (HPLC analysis, data not shown) after 8 h. After purification of half of the mixture by HPLC, **6** was isolated in a yield of 84%.

The remaining 50% was directly incubated with recombinant ST3Gal3 and a 20% excess of CMP-Neu5Ac. As expected from the substrate specificity of the enzyme only one product, **7** (vide infra), was observed after 18 h. Half of the mixture was purified by HPLC yielding **7** in an 83% conversion.

To the remaining reaction mixture α3-FucT3 and an excess of GDP-Fuc were added. After 8 h the reaction was stopped and the mixture purified by HPLC. Compound **8** (vide infra) was obtained in a yield of 86%. The total overall yield of the five steps was calculated to be 23% representing approximately 11.1 mg (the actual yield was 3.2 mg since only 50% of the yields of steps 3 and 4 were used for further incubations). It is clear that the limiting step in the procedure is the synthesis of the core-2 structure (Table 1).

4.2. Mass spectrometry

Positive-ion mode MALDI-TOF mass spectrometry was performed on (glyco)peptides **1**, **2**, **6**, **7**, and **8** (Fig. 1). Starting from the non-glycosylated peptide **1** (Fig. 2A; *m*/*z*

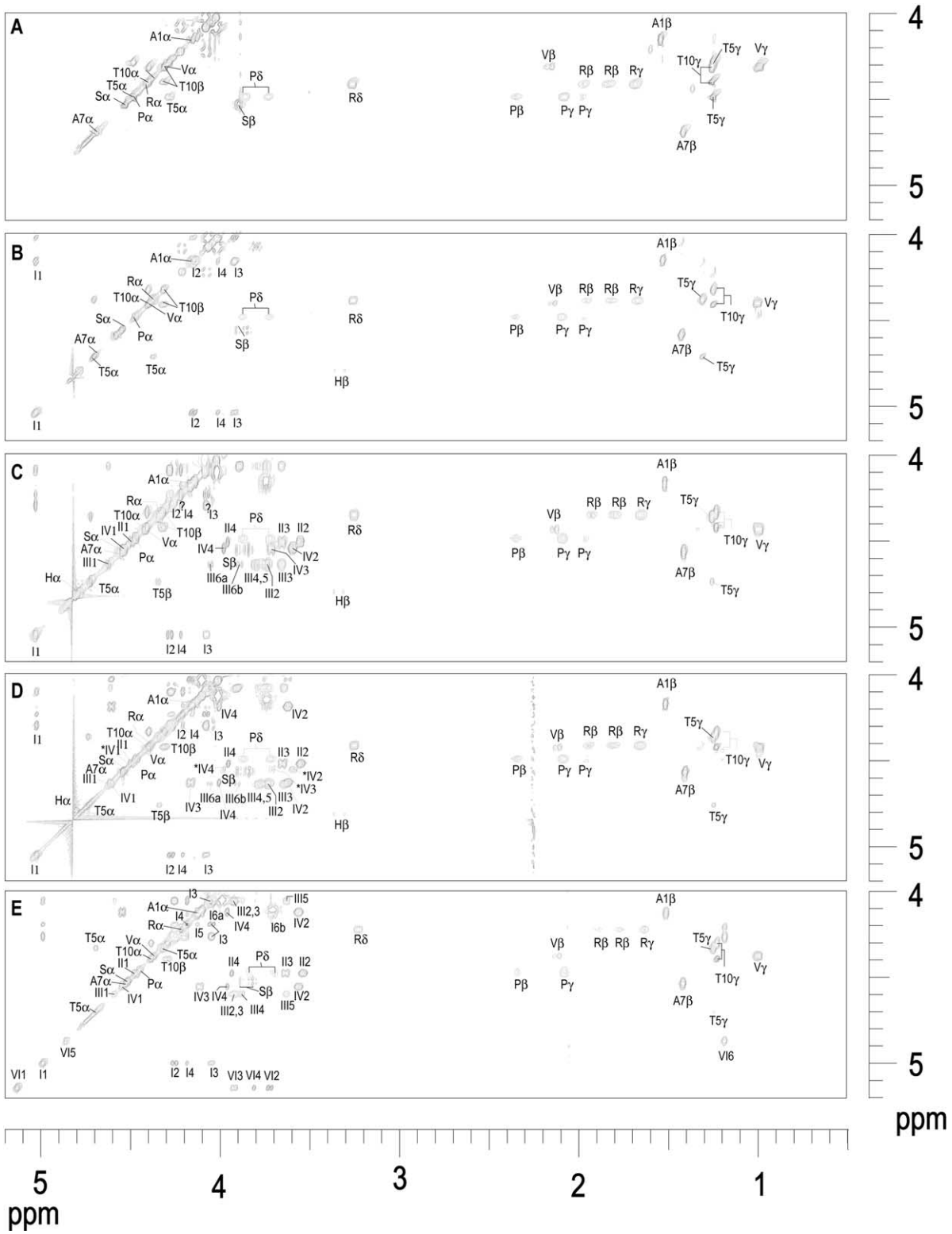
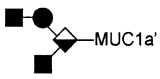


Fig. 5. Cross sections of 600 MHz ^{1}H NMR TOCSY spectra of MUC1a' peptide (**1**, A), intermediate glycopeptides (**2**, B; **6**, C; **7**, D), and the final sialyl Lewis X containing MUC1a' glycopeptide (**8**, E). The letter codes correspond to the amino acids and the roman numbers to the individual monosaccharides (**I**, GalNAc; **II**, Gal(β1-3); **III**, GlcNAc; **IV**, Gal(β1-4); **V**, Neu5Ac; **VI**, Fuc).

1111.6 $[M + H]^+$), a mass difference of 203.0 a.m.u. between **2** (Fig. 2B; *m*/*z* 1314.6 $[M + H]^+$) and **1** was observed, indicating the incorporation of only one *N*-acetylhexosamine residue. To exactly determine the glycosylation site, **2** was analyzed by CID-MS-MS. The first-stage mass spectrum (not shown) revealed essentially four peaks at *m*/*z* 1314.7, 657.9, 439.0, and 556.3, respectively. The first three peaks were assigned to the singly, doubly, and triply charged pro-

Table 3
600 MHz ^{1}H NMR chemical shifts of the carbohydrate residues in four glycopeptides (chemical shifts are relative to internal acetone, δ 2.225) measured in HO^2H at 300 K. Coupling constants are given between brackets. n.d., not determined

Residue	Reporter group	**2**	**6** 	**7** 	**8** 
GalNAc **I**	H-1	4.976 (3.5)	4.960 (3.6)	4.957 (3.6)	4.954 (s)
	H-2	4.119	4.224	4.221	4.223
	H-3	3.890	4.024	4.028	4.024
	H-4	3.981	4.160	4.156	4.159
	H-5	4.030	4.111	4.099	4.093
	H-6a	n.d.	3.694	3.692	3.892
	H-6b	n.d.	n.d.	3.956	n.d.
	CH_3	2.045	2.043	2.038	2.034
Gal **II**	H-1		4.432 (7.8)	4.432 (7.8)	4.433(7.8)
	H-2		3.505	3.506	3.509
	H-3		3.607	3.605	3.603
	H-4		3.905	3.905	3.904
	H-5		n.d.	3.646	3.634
	H-6a		n.d.	n.d.	n.d.
	H-6b		n.d.	n.d.	n.d.
GlcNAc **III**	H-1		4.560 (8.4)	4.557 (8.4)	4.552 (8.3)
	H-2		3.750	3.734	3.890
	H-3		3.607	3.589	3.850
	H-4		3.690	3.685	3.774
	H-5		3.690	3.685	3.600
	H-6a		3.996	4.007	4.018
	H-6b		3.836	3.856	n.d.
	CH_3		2.006	2.006	2.007
Gal **IV**	H-1		4.468 (7.8)	4.544 (7.8)	4.504 (7.6)
	H-2		3.545	3.576	3.533
	H-3		3.665	4.108	4.081
	H-4		3.927	3.955	3.928
	H-5		n.d.	n.d.	3.655
	H-6a		n.d.	n.d.	n.d.
	H-6b		n.d.	n.d.	n.d.
Neu5Ac **V**	H-3a			1.797	1.790
	H-3e			2.761	2.765
	H-4			3.699	3.682
	H-5			3.848	3.851
	H-6			3.641	3.660
	H-7			3.596	n.d.
	H-8			3.889	n.d.
	H-9a			n.d.	n.d.
	H-9b			n.d.	n.d.
	CH_3			2.033	2.034
Fuc **VI**	H-1				5.092 (3.3)
	H-2				3.693
	H-3				3.778
	H-4				3.890
	H-5				4.823 (5.9)
	CH_3				1.164 (5.9)

tonated glycopeptide, respectively. The peak at *m*/*z* 556.3 was assigned to the doubly charged protonated peptide, formed after elimination of GalNAc (as 2-acetamidogalactal; 203 a.m.u.) from the pseudo-molecular ion. The peak at *m*/*z* 1314.7 was selected in the ion-trap, and after fragmentation the mass spectrum (Fig. 3A) revealed several higher and lower intensity daughter-ions. Analysis of the data using MS Bio Tools revealed two main sets of fragment ions (Fig. 3C,D). The first set, originating from the parent ion, included the B_7 and $B_{9\text{-}11}$ ions revealing that T10 was not glycosylated (Fig. 3B). An almost complete set of ions belonging to the Y-series (Y_4 to Y_{11}) was also identified. While the mass difference between Y_6 and Y_5 corresponded to non-substituted S6, the difference between Y_7 and Y_6 was 304.3 a.m.u. representing a HexNAc-substituted T5. The second main set of fragment ions (Fig. 3D) could be assigned

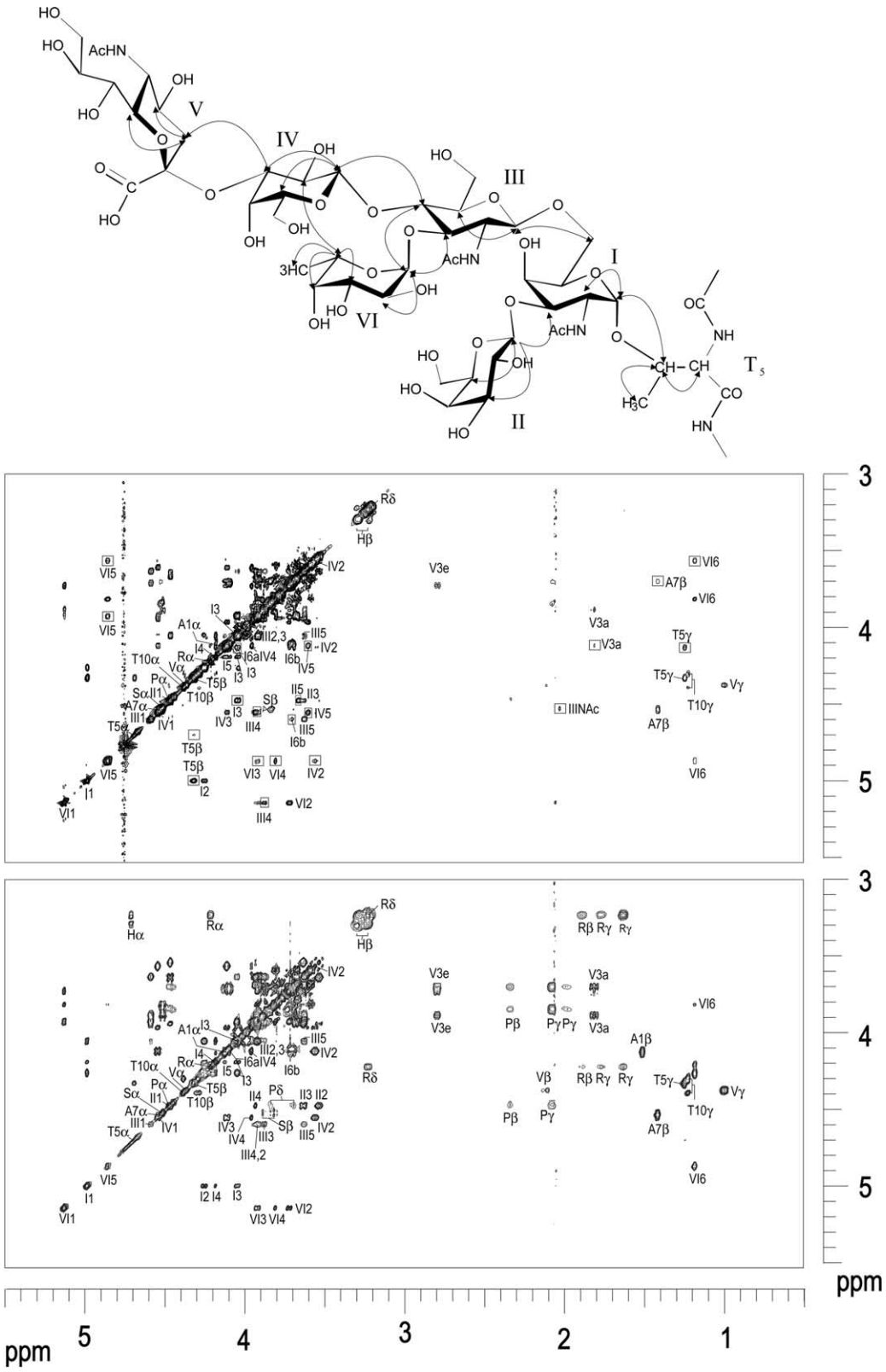


Fig. 6. 600 MHz ^{1}H NMR TOCSY (lower panel) and off-resonance ROESY (upper panel) spectra of the sialyl Lewis X containing MUC1a' glycopeptide **8**. Intra- and interresidual contacts observed for the core-2 sialyl Lewis X structure in the ROESY spectrum are indicated in the structure and in the spectrum.

to the Y- (Y_4 to Y_{11}, except Y_8) and the B-series (B_4 to B_{11}) of the deglycosylated peptide. The confirmation of the T5-glycosylation came from the $Y_{7\text{-}11}$ ions, which were 203 a.m.u. lower than in the first series. The site-specific glycosylation of threonine at position 5 (T5) was in agreement with previous experiments, in which the substrate specifici-

ties of GalNAcT1, GalNAcT2, and GalNAcT3 were investigated [18].

The protonated pseudo-molecular ion at *m*/*z* 1841.8 in the MALDI-TOF mass spectrum of **6** (Fig. 2C) revealed the incorporation of two hexose and one *N*-acetylhexosamine residues with respect to **2**. It should be noted that the one-pot reaction, using β-galactosidase from bovine testes and the core-2 β 6-*N*-acetylglucosaminyltransferase yielded only the core-2 structure.

Analysis of the MALDI-TOF mass spectrum of purified **7** (Fig. 2D) showed the presence of two pseudo-molecular ions at *m*/*z* 2132.9 ($M + H^+$) and 1841.8 ($M + H^+$), respectively. The highest value corresponded with **7**, whereas the lower value originated from desialylation of **7** (resulting in **6**). It has been observed before that desialylation can occur during the ionization process in MALDI-TOF mass spectrometry [23], particularly when α-cyano-4-hydroxy-cinnamic acid is used as a matrix. However, the presence of both the sialylated and non-sialylated compound was also observed in the ^{1}H NMR spectrum of purified **7** (vide infra). Here, the slightly acidic pH (~6) of the solution during the NMR measurements combined with the prolonged measuring time at 300 K may account partly for the desialylation observed. As the MS analysis was carried out on the NMR sample, the peak at *m*/*z* 1841.8 probably reflects both mechanisms.

The MALDI-TOF mass spectrum of **8** (Fig. 2E) revealed two protonated pseudo-molecular ions at *m*/*z* 1987.9 and 2279.0, corresponding to the MUC1a' glycopeptides bearing a core-2 Lewis X and sialyl Lewis X structure, respectively. The ^{1}H NMR spectrum of **8** excluded the presence of the non-sialylated structure, indicating that in this case the desialylated compound was obviously formed in the mass spectrometer due to the earlier mentioned ionization phenomenon [23].

4.3. NMR spectroscopy

1D and 2D ^{1}H NMR spectroscopy was performed on (glyco)peptides **1**, **2**, **6**, **7**, and **8** (Fig. 1). The amino acid signals in the 1D spectrum (Fig. 4A) of the non-glycosylated peptide (**1**) could be assigned (Table 2) by comparing the values to literature data [24], and by the characteristic correlation tracks in the 2D TOCSY (100 ms) spectrum (Fig. 5A, the tracks of H and D are not shown). The observation of an interresidual cross peak in the 2D off-resonance ROESY spectrum (not shown) between P8δ and A7α/A7β allowed the discrimination between both A residues present. The spin systems of both T residues could only be correctly assigned after the incorporation of the GalNAc residue (vide infra).

In the 1D ^{1}H NMR spectrum of **2** (Fig. 4B) the incorporation of an α-linked *N*-acetylhexosamine residue (**I**) was concluded from the observation of the *N*-acetyl singlet at δ 2.045 and the anomeric doublet at δ 4.976 ($^3J_{1,2}$ = 3.5 Hz). The anomeric track in the 2D TOCSY spectrum (100 ms) (Fig. 5B) identified **I** as *N*-acetylgalactosamine based on the characteristic Gal*p* NAc spin system (**I** H2,3,4). **I** H5 could be observed on the **I** H4-track (Table 3). The incorporation of **I** in peptide **2** resulted in the downfield shifts of the α, β, and γ protons of one T residue only, confirming the above mentioned site-specific glycosylation. The glycosylation of T5 was further evidenced by the downfield shift of the α protons of the neighboring S6 and V4 residues. Such a feature was not observed for D9 or R11 flanking T10. In addition, the S6β protons were observed as a double doublet indicating the modified chemical environment with respect to **1** in which S6β appeared as a doublet. The ROESY spectrum (not shown) showed two interresidual sets of cross peaks between P8δ and A7α, A7β, and between **I** H1 and T5β providing further evidence for the site of glycosylation.

The 1D ^{1}H NMR spectrum of **6** (Fig. 4C) showed the presence of additional anomeric signals at δ 4.432 ($^3J_{1,2}$ = 7.8 Hz, **II**), 4.560 ($^3J_{1,2}$ = 8.4 Hz, **III**), and 4.468 ($^3J_{1,2}$ = 7.8 Hz, **IV**), and a second *N*-acetyl singlet at δ 2.006, when compared with the carbohydrate NMR data of **2**. The anomeric track of **III** in the 2D TOCSY (100 ms) spectrum (Fig. 5C) showed five cross peaks, representing six protons, which could be assigned (Table 3) by making use of a TOCSY experiment with a short mixing time (not shown). This allowed the identification of **III** as *N*-acetylglucosamine. The anomeric tracks of **II** and **IV** revealed both a typical set of H-2,3,4 galactose cross peaks, which were easily assigned on the basis of the peak shape and corroborated in the 10 ms TOCSY spectrum. The presence of interresidual cross peaks in the off-resonance ROESY spectrum (not shown) between **II** H1 and **I** H3 and between **IV** H1 and **III** H4 permitted the assignment of **II** as galactose β-1,3-linked to GalNAc (**I**) and **IV** as galactose β-1,4-linked to GlcNAc (**III**). The β-1,6-linkage between **III** and **I** was ascertained via the observation of interresidual connectivities between **III** H1 and **I** H6a.

The presence of a Neu5Ac residue (**V**) in **7** was indicated in the 1D ^{1}H NMR spectrum (Fig. 4D) by the presence of an additional *N*-acetyl signal at δ 2.033 and the **V** H3a and **V** H3e resonances at δ 1.797 and δ 2.761, respectively, when comparing the spectrum to that of **6**. Although the off-resonance ROESY spectrum (not shown) did not show cross peaks between **V** H3 and **IV** H3 as reported in the literature [25], the α-2,3-linkage between **V** and **IV** became evident from the TOCSY (100 ms) spectrum (Fig. 5D). The incorporation of **V** to HO-3 of **IV** resulted in the downfield shift of all **IV** resonances. The most pronounced downfield shift was observed for **IV** H3 (Δδ 0.443 ppm), in agreement with earlier observations [20]. The presence of **6** as a contaminant of **7** (see 4.2.) was evidenced by the presence of the anomeric track of **IV** from **6** in the TOCSY spectrum of **7** (Fig. 5D), marked with an asterix (*).

When compared with the carbohydrate NMR data of **7**, the 1D ^{1}H NMR spectrum of **8** (Fig. 4E) showed the presence of an additional anomeric resonance at δ 5.092 ($^3J_{1,2}$ = 3.3 Hz) and a doublet at δ 1.164 ($^3J_{6,5}$ = 5.9 Hz) originating from the fucose methyl-protons (**VI**). By means of 2D TOCSY (Fig. 5E) and ROESY (Fig. 6) experiments all chemical shifts stemming from **VI**, and almost all other chemical shifts

could be assigned (Table 3). The cross peak between **VI** H1 and **III** H3 in the off-resonance ROESY spectrum demonstrated the **VI** (α1-3) **III** linkage. This connectivity, and all other intra- and interresidual connectivities observed in the ROESY spectrum (Fig. 6) are visualized in Fig. 6 providing unambiguous evidence that the O-glycan synthesized on the MUC1a'-peptide is core-2 sialyl Lewis X.

5. Conclusions

In this study we have demonstrated a rapid and easy access to a complex hexasaccharide, containing sialyl Lewis X. It was synthesized regio- and stereoselectively on a biologically interesting peptide backbone, in mg amounts. The effective synthesis time, using optimized reaction conditions was less than 1 week, including intermediate purification steps which allowed the isolation of essentially pure compounds. In total four glycopeptides were generated, representing intermediates in glycosylation, naturally encountered as glycoforms on a single O-glycosylation site.

Acknowledgements

We thank Dr. Y.E.M. van der Burgt and Dr. A. Schneider (Bruker Daltonic, Bremen) for allowing the use of the Esquire-LC mass spectrometer and their help with the measurements. This work was financially supported by the EU, project BIO CT 95-0138 "Engineering O-Glycosylation for the Production of Receptor Blockers".

References

[1] F.-G. Hanisch, S. Müller, MUC1: the polymorphic appearance of a human mucin, Glycobiology 10 (2000) 439–449.
[2] B.W. Yin, K.O. Lloyd, Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16, J. Biol. Chem. 276 (2001) 27371–27375.
[3] R.H. Yolken, J.A. Peterson, S.L. Vonderfecht, E.T. Fouts, K. Midthun, D.S. Newburg, Human milk mucin inhibits rotavirus replication and prevents experimental gastroenteritis, J. Clin. Invest. 90 (1992) 1984–1991.
[4] P. Pandey, S. Kharbanda, D. Kufe, Association of the DF3/MUC1 breast cancer antigen with Grb2 and the Sos/Ras exchange protein, Cancer Res. 55 (1995) 4000–4003.
[5] M. Yamamoto, A. Bharti, Y. Li, D. Kufe, Interaction of the DF3/MUC1 breast carcinoma-associated antigen and β-catenin in cell adhesion, J. Biol. Chem. 272 (1997) 12492–12494.
[6] S.P. Treon, J.A. Mollick, M. Urashima, G. Teoh, D. Chauhan, A. Ogata, N. Raje, J.H. Hilgers, L. Nadler, A.R. Belch, L.M. Pilarski, K.C. Anderson, Muc-1 core protein is expressed on multiple myeloma cells and is induced by dexamethasone, Blood 93 (1999) 1287–1298.
[7] K.O. Lloyd, J. Burchell, V. Kudryashov, B.W.T. Yin, J. Taylor-Papadimitriou, Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines; demonstration of simpler and fewer glycan chains in tumor cells, J. Biol. Chem. 271 (1996) 33325–33334.
[8] I. Brockhausen, J.M. Yang, J. Burchell, C. Whitehouse, J. Taylor-Papadimitriou, Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells, Eur. J. Biochem. 233 (1995) 607–617.
[9] I. Brockhausen, Pathways of O-glycan biosynthesis in cancer cells, Biochim. Biophys. Acta 1473 (1999) 67–95.
[10] S. Müller, K. Alving, J. Peter-Katalinic, N. Zachara, A.A. Gooley, F.-G. Hanisch, High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells, J. Biol. Chem. 274 (1999) 18165–18172.
[11] J.M. Burchell, A. Mungul, J. Taylor-Papadimitriou, O-Linked glycosylation in the mammary gland: changes that occur during malignancy, J. Mammary Gland Biol. Neoplasia 6 (2001) 355–364.
[12] S. Müller, F.-G. Hanisch, Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogeneous breast cancer mucin; high density and prevalent core 2-based glycosylation, J. Biol. Chem. 277 (2002) 26103–26112.
[13] G. Dudziak, S. Zeng, E.G. Berger, R. Gutiérrez Gallego, J.P. Kamerling, U. Kragl, C. Wandrey, In situ generated O-glycan core 1 structure as substrate for Gal(β1-3)GalNAc β-1,6-GlcNAc transferase, Bioorg. Med. Chem. Lett. 8 (1998) 2595–2598.
[14] K. Hård, G. van Zadelhoff, P. Moonen, J.P. Kamerling, J.F.G. Vliegenthart, The Asn-linked carbohydrate chains of human Tamm-Horsfall glycoprotein of one male. Novel sulfated and novel *N*-acetylgalactosamine-containing N-linked carbohydrate chains, Eur. J. Biochem. 209 (1992) 895–915.
[15] A. Bax, J. Davis, MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy, J. Magn. Reson. 65 (1985) 355–360.
[16] A. Bax, J. Davis, Practical aspects of two-dimensional transverse NOE spectroscopy, J. Magn. Reson. 63 (1985) 207–213.
[17] H. Desvaux, P. Berthault, N. Birlirakis, M. Goldman, M. Piotto, Improved versions of off-resonance ROESY, J. Magn. Reson. 113 (1995) 47–52.
[18] H.H. Wandall, H. Hassan, E. Mirgorodskaya, A.K. Kristensen, P. Roepstorff, E.P. Bennett, P.A. Nielsen, M.A. Hollingsworth, J. Burchell, J. Taylor-Papadimitriou, H. Clausen, Substrate specificities of three members of the human UDP-*N*-acetyl-α-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase family, GalNAc- T1,-T2, and -T3, J. Biol. Chem. 272 (1997) 23503–23514.
[19] E.P. Bennett, H. Hassan, U. Mandel, E. Mirgorodskaya, P. Roepstorff, J. Burchell, J. Taylor-Papadimitriou, M.A. Hollingsworth, G. Merkx, A.G. van Kessel, H. Eiberg, R. Steffensen, H. Clausen, Cloning of a human UDP-*N*-acetyl-α-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat, J. Biol. Chem. 273 (1998) 30472–30481.
[20] S. Zeng, R. Gutiérrez Gallego, A. Dinter, M. Malissard, J.P. Kamerling, J.F.G. Vliegenthart, E.G. Berger, Complete enzymic synthesis of the mucin-type sialyl Lewis x epitope, involved in the interaction between PSGL-1 and P-selectin, Glycoconjugate J. 16 (1999) 487–497.
[21] C. Hoh, G. Dudziak, A. Liese, Optimization of the enzymatic synthesis of O-glycan core 2 structure by use of a genetic algorithm, Bioorg. Med. Chem. Lett. 12 (2002) 1031–1034.
[22] M. Ujita, J. McAuliffe, T. Schwientek, R. Almeida, O. Hindsgaul, H. Clausen, M. Fukuda, Synthesis of poly-*N*-acetyllactosamine in core 2 branched O-glycans. The requirement of novel β-1,4-galactosyltransferase IV and β-1,3-*N*-acetylglucosaminyltransferase, J. Biol. Chem. 273 (1998) 34843–34849.
[23] A.K. Powell, D.J. Harvey, Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry, Rapid Commun. Mass Spectrom. 10 (1996) 1027–1032.
[24] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley (1986) 14–17.
[25] G.E. Ball, R.A. O'Neill, J.E. Schultz, J.B. Lowe, B.W. Weston, J.O. Magy, E.G. Brown, C.J. Hobbs, M.D. Bednarski, Synthesis and structural characterization using 2D NMR of sialyl Lewis X (SLe^x) and Lewis X (Le^x) oligosaccharides: ligands related to E-selectin (ELAM-1) binding, J. Am. Chem. Soc. 114 (1992) 5449–5451.