ELSEVIER

www.elsevier.com/locate/carres

Carbohydrate Research 333 (2001) 165–171

CARBOHYDRATE RESEARCH

Note

Studies of acceptor site specificities for three members of UDP-GalNAc:*N*-acetylgalactosaminyltransferases by using a synthetic peptide mimicking the tandem repeat of MUC5AC

Daniel Tetaert,[a,*] Colette Richet,[a] Jean Gagnon,[b] Arnold Boersma,[a] Pierre Degand[a]

[a] *Unité Inserm no. 377, Place de Verdun, F-59045 Lille, France*

[b] *Institut de Biologie Structurale JP EBEL, CEA/CNRS/UJF, 41 rue Jules Horowitz, F-38027 Grenoble, France*

Received 16 January 2001; received in revised form 23 March 2001; accepted 3 May 2001

Abstract

The acceptor specificity of three major isoforms of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltranferases (murine recombinant proteins GaNTase-T1, -T2 and -T3) was investigated using the synthetic peptide (GTTP-SPVPTTSTTSAP) containing clusters of threonine residues mimicking the mucin tandem repeat unit of MUC5AC. The *O*-glycosylated products obtained after in vitro reactions were fractionated by capillary electrophoresis and the purified glycopeptides were characterized by MALDI mass spectrometry (number of *O*-GalNAc residues) and by Edman degradation (site location). A maximum of three GalNAc residues was transferred into the MUC5AC motif peptide and the preferential order of incorporation for each GaNTase isoform was determined. Our results suggest that clusters of threonine appear to be essential for site recognition of peptide backbone by the ubiquitous GaNTases and also support the notion that the different GaNTase isoforms with varying substrate specificities are involved in a hierarchical order of O-glycosylation processing of the mucin-type *O*-glycoproteins.

Keywords: Mucin; MUC5AC; O-Glycosylation; UDP-GalNAc; Polypeptide *N*-acetylgalactosaminyltranferase (GaNTase)

1. Introduction

Post-translational modification by O-glycosylation (i.e. mucin-type O-glycosylation) is an ubiquitous feature of secretory proteins and is also found in a selected array of nuclear and cytoplasmic proteins.[1,2] Glycosylation of proteins and lipids potentially provides a basis for tissue-specific interactions with the environment and the alterations in the O-glycosylation patterns may play a role in the pathogenesis of several diseases.[3,4] While the glycosyltransferases exhibit substrate specificity for underlying oligosaccharide chains, some of these enzymes also recognize specifically peptide and/or protein conformations. Concerning the initial step of O-glycosylation, it has been clearly established that a family of several UDP-GalNAc:*N*-acetylgalactosaminyltransferases (GaNTases) (E.C. 2.4.1.41)[5] con-

* Corresponding author. Tel.: +33-320-298857; fax: +33-320-538562.

E-mail address: tetaert@lille.inserm.fr (D. Tetaert).

0008-6215/01/$ - see front matter
PII: S0008-6215(01)00135-5

tributes to the process: at least ten distinct mammalian isoforms of this gene family have now been identified and functionally characterized.[6] Differences in kinetic properties, substrate specificities and expression patterns of isoforms could also explain differential regulation of *O*-glycan attachment sites and density on different peptide backbones.[7]

On the other hand, the different isoforms also appear distinct in sequence and exhibit an overall amino acid sequence identity of 45–50%, but several peptide stretches within the putative catalytic domains of the enzymes show sequence identities of more than 75–80%.[8] Moreover, homologous putative GaNTases (11 members) have been identified in *Caenorhabditis elegans* cDNA libraries and have exhibited 40–50% sequence similarity to each of the human genes. One may therefore assume that the family of GaNTases is old in evolutionary terms.[9] This large number of GaNTase members may reflect individual roles in tissue- and cell-type specific glycosylation of specific target glycoproteins. It is also possible that the various GaNTase members are located in different subcellular compartments including the ER and the Golgi: this scattered distribution through the cell argues also for diversity of GaNTase functions.[10,11]

Nevertheless, several common characteristics of the mammalian GaNTase members are becoming apparent: (i) During evolution, the general structural features of GaNTases have been conserved: a catalytic unit with region of high similarity ($>80\%$) is present, whereas low identity is found in three N-terminal regions encoding the cytoplasmic tail, the putative signal anchor sequence and the stem regions;[8] (ii) discrepancies affect the genomic organization, the chromosomal localizations as well as the expression pattern of the enzymes in cells and human organs;[7,10,11] (iii) in vitro assays indicate that subtle differences in specificities of peptide substrate among the GaNTase activities exist, but there are overlaps.[12]

Table 1
Values of GalNAc incorporation rate ± SEM

GaNTase isoforms	Relative rates of GalNAc incorporation (nmol GalNAc/h)
GaNTase-T1	77.5 ± 1.2
GaNTase-T2	135.3 ± 3.6
GaNTase-T3	48.0 ± 0.6

The most widely expressed mammalian GaNTases (the murine T1, -T2 and -T3 isoforms) are analyzed here in order to gain further insight into the specificity of this family of enzymes towards the mucin MUC5AC motif peptide (GTTPSPVPTTSTTSAP). It has been reported previously that all threonine residues aligned in a cluster in the MUC5AC tandem repeat domain are fully glycosylated when a microsomal fraction issued from human gastric mucosa is used as enzyme source,[13] but the present investigations have been carried out to define whether the principal GaNTases (GaNTase-T1, -T2 and -T3) have a preferential site of recognition in terms of number and location of GalNAc residues along the peptide backbone.

2. Results and discussion

Acceptor substrate specificity of GaNTase-T1, -T2 and -T3 isoforms.—The oligopeptide G***TT******PSP***VPTTSTTSAP mimicking the mucin tandem repeat of the MUC5AC[14] motif consists of six threonine residues located as three pairs of hydroxy amino acids along the peptide backbone and, of a XTPXP sequence (position 2–6) reported as a signal favoring efficient glycosylation.[15,16] The three recombinant GaNTase isoforms and the peptide were incubated for various time periods (3–24 h) under our standard reaction conditions[17] and, the rate of GalNAc incorporation is presented in Table 1.

The capillary electrophoresis procedure provided adequate conditions for analyses of the newly synthesized glycosylated products (Fig. 1). The identification of the products was carried out by comparing the profiles of the peaks obtained in the absence or presence of UDP-GalNAc in the reaction mixture with and without the recombinant enzymes. The formation of new glycopeptide species was also confirmed by MALDI mass spectrometry analysis, which revealed the number of attached GalNAc residues (additional mass of

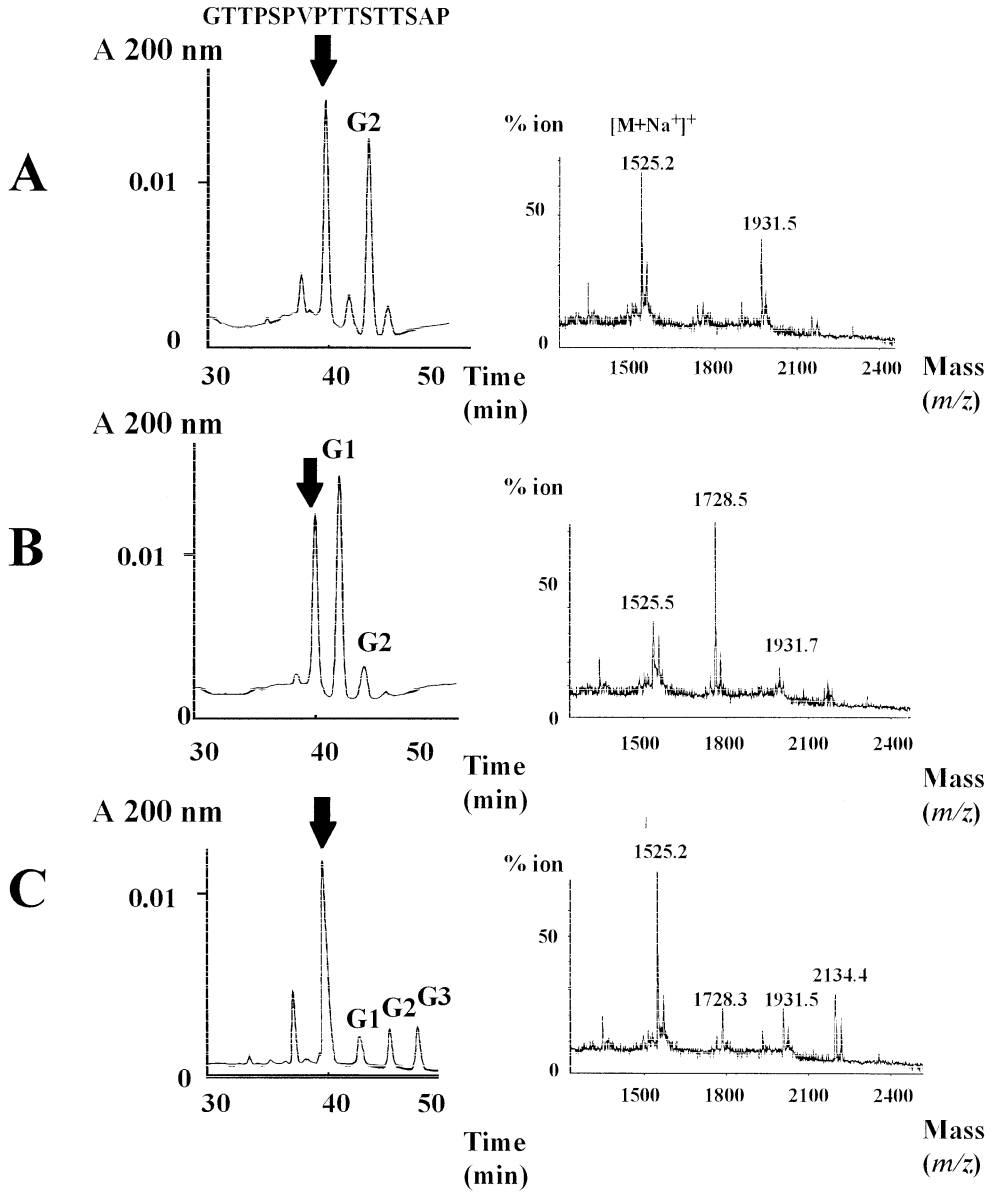

Fig. 1. Capillary electrophoresis (CE) analyses of in vitro O-glycosylation of the MUC5AC derived peptide by the different purified recombinant GaNTases. The corresponding MALDI mass spectra are associated to the CE profiles. (A) After GaNTase-T1 action; (B) after GaNTase-T2 action; (C) after GaNTase-T3 action. The G1, G2 and G3 peaks are the one-, two- and three-*O*-linked GalNAc glycopeptides. Arrows correspond to residual MUC5AC peptide.

203 for each). Under our standard conditions: (i) GaNTase-T1 transferred and was limited to 2 mol of GalNAc residues (m/z at 1931.5) even after a prolonged incubation of 24 h (Fig. 1(A)); (ii) with the same amount of GaNTase-T2, one GalNAc (m/z at 1728.5) was essentially transferred to the MUC5AC peptide after incubation of 24 h (Fig. 1(B)); (iii) GaNTase-T3 transferred three GalNAc to MUC5AC peptide and the glycopeptides G1, G2 and G3 (where the numeral represents the number of the O-linked GalNAc residues) (m/z at 1728.3, 1931.5 and 2134.4, respectively) appeared to be formed rapidly in equal amount (Fig. 1(C)). The number of *O*-GalNAc linkages is inversely proportional to the V_{max} of GalNAc incorporation for the corresponding GaNTases (number of O linkages: T2 < T1 < T3).

Our results are in agreement with those of Iida et al.,[18,19] since also on the MUC2 motif FM2-12 (PTTTPLK sequence), the number of

GalNAc linkages was, respectively, two for GaNTase-T1, one for GaNTase-T2 and three for GaNTase-T3. The recognition of attachment sites by the different GaNTase (T1, T2, T3) isoforms seems to be very precise and the resulting *O*-glycosylated products are not randomly generated, as previously thought.[20]

Localization of N-*acetylgalactosaminylated threonine residues.*—The *O*-glycosylated products obtained after action of GaNTase isoforms were purified by capillary electrophoresis (more than 20 analytical runs each for cumulative preparative purposes) and each fraction was subjected to amino acid sequencing. The PTH derivative of Thr-*O* GalNAc was identified as a pair of peaks with elution times near those of PTH-Ser and PTH-Thr and careful analysis indicated the decrease in repetitive yield at the expected Thr cycle if the residue was substituted, as illustrated in Fig. 2. All the results of the amino acid sequencing of the different glycopeptide fractions are summarized in Fig. 3. The preferred acceptor sites for the different GaNTase isoforms are different: GaNTase-T1 recognized simultaneously positions Thr # 13 and Thr # 3; GaNTase-T2 recognized position Thr # 9; GaNTase-T3 recognized position Thr # 3. The dominant activity of GaNTase-T3 isoform is secondarily directed towards another cluster of threonine residues, since compared

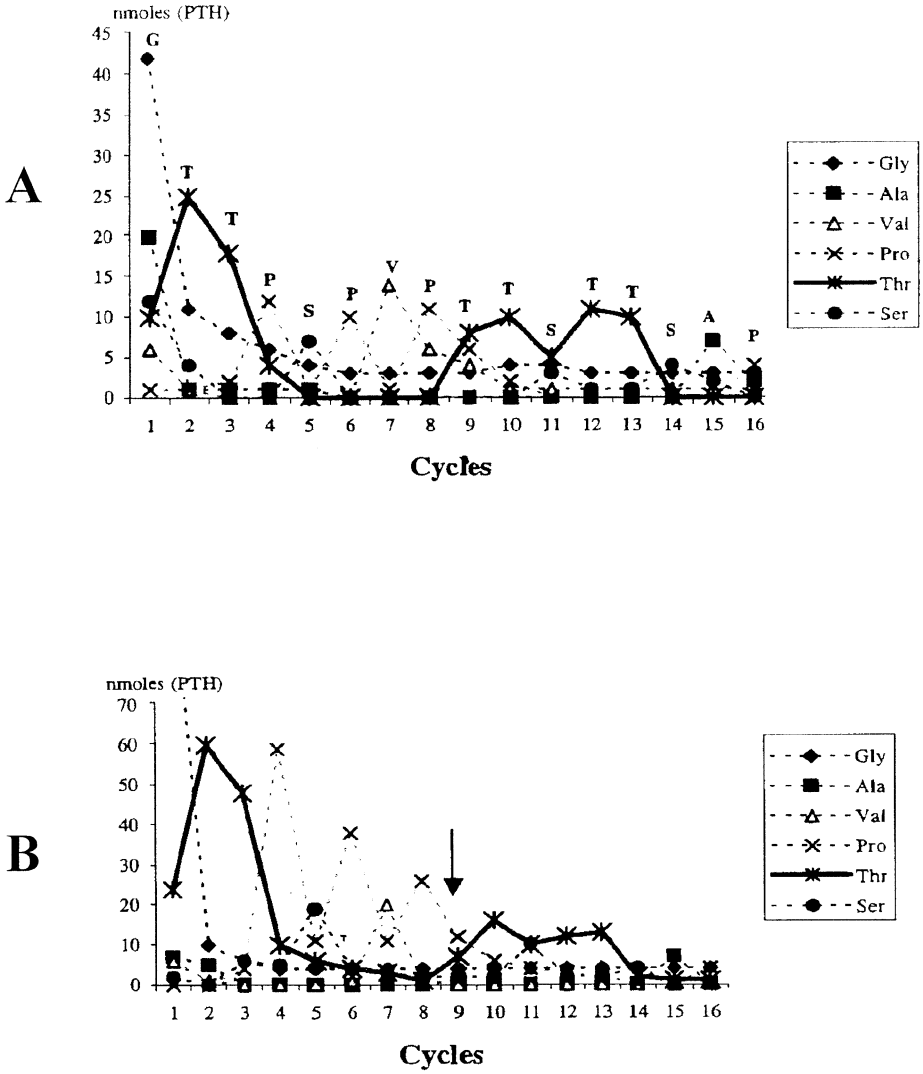

Fig. 2. Repetitive yields of Edman degradation for the parent peptide MUC5AC (A) and the purified glycopeptide G1 fraction obtained by action of GaNTase-T2 (B). The follow-up is mainly for PTH-Thr (—) the potential site(s) of O-glycosylation. Arrow indicates the decreased level of PTH-Thr at position # 9.

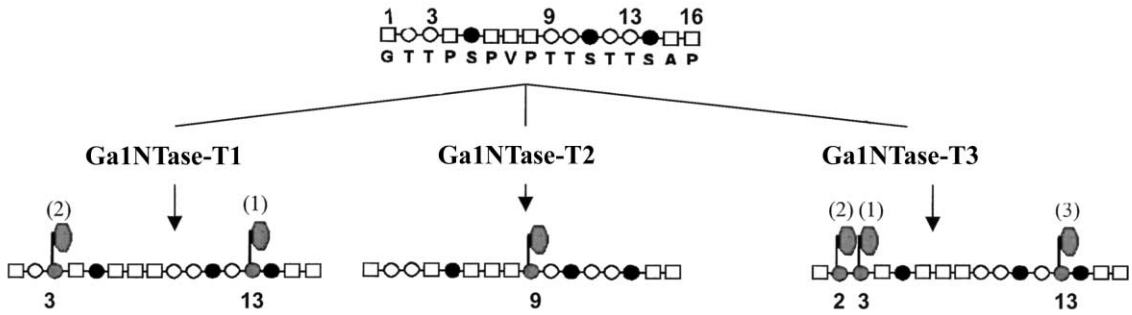

Fig. 3. Schematic depiction of the three GaNTase activities towards the different positions (2, 3, 9 and 13) of the MUC5AC motif peptide. The number (1) indicates the privileged sites for each GaNTase whereas (2) and (3) are secondary and tertiary sites for GaNTase-T1 and -T3. Circles correspond to potential *O*-glycosylated sites: open circles are Thr residues, closed circles are Ser residues; dotted circles correspond to glycosylated positions; Dotted octagons indicate GalNAc residues.

to the simultaneous action of GaNTase-T1 towards the positions Thr # 13/Thr # 3 (G2), GaNTase-T3 acted at positions Thr # 2 and Thr # 13 respectively (G3), after Thr # 3 was substituted.

To sum up, the O-glycosylation mechanism requires the coordinated action of a large number of enzymes involving sialyltransferases, fucosyltransferases, galactosyltransferases and UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltranferase (GaNTases). The latter are particularly important because these enzymes govern the number as well as the length of the glycan chains. Therefore, the data reported here suggest that the initial O-glycosylation reaction in terms of GalNAc site occupancy is less random than previously expected[20] and processing appears to occur in a highly hierarchical manner. Indeed, the recognition of attachment site by GaNTase-T1, -T2 and -T3 towards three clusters of threonine residues in the peptide mimicking the tandem repeat sequence of MUC5AC is very specific and the preference for a particular Thr residue appears to depend on the environment: GaNTase-T3 is influenced by the N-terminal region; GaNTase-T2 by the P × P × P β-turn segment (corresponding to the efficient signal XTPXP sequence[16]); GaNTase-T1 by the C-terminal region. Subsequent GalNAc attachment is rapidly completed by GaNTase-T1 with simultaneous O linkage at Thr in position # 3 whereas the GaNTase-T3, after recognition of the Thr # 3, places GalNAc residues at Thr in position # 2 followed by Thr in position # 13. Apart from the first occupancy, GaNTase-T3 seemed to perform the glycosylation in a reverse order compared to GaNTase-T1. These three enzymes share a similar acceptor specificity towards the tandem repeat sequence of MUC1, with the three Thr residues separated within its tandem repeat unit.[12] GaNTase-T3 is thought to play an atypical role by its action towards specific peptide substrates such as the HIV V3 loop[21] or by the multigalactosaminylation of MUC2 motif peptide.[18] All these results are of biological importance since the positioning and spacing of oligosaccharide chains in mucins is known to modulate interactions properties (i.e., lectins and clusters of glycan epitopes).[22,23]

3. Experimental

Purification of secreted isoforms.—Recombinant enzymes GaNTase-T1, -T2 and -T3 were kindly provided by KG Ten Hagen and LA Tabak (Center for Oral Biology, Aab Institute for Biomedical Sciences, University of Rochester, Rochester, NY, USA). The transferases were purified from the culture media of transfected cells (COS 7) using anti-FLAG™ M2 affinity gel (Eastman Kodak Co.) as described previously.[24] Levels of partially purified recombinant enzymes were analyzed by Tricine SDS-PAGE after labeling with γ^{32}P-rATP using heart muscle kinase (HMK) as described previously. Gels were dried under vacuum and exposed to film (XAR-Kodak) for quantitation on a PhosphorImager (Molecular Dynamics).

Functional expression assays of secreted recombinant GaNTases.—The GaNTase activities were measured against the peptide substrate MUC5AC (GTTPSPVPTTSTTSAP) expressed predominantly in the tracheobronchial and gastric tissues.[14] Subsequent assays used 5 μL of culture media from COS7 cells transfected with murine GaNTase-T1, -T2, -T3. Equivalent amounts (units) of each enzyme within the media (1 μg of enzyme as determined by SDS PAGE gel quantitation) were used in each assay. Assays were performed with MUC5AC peptide in a total volume of 40 μL at the following final concentrations: 1 mM peptide substrate; 125 mM MES buffer (pH 7.0) containing 0.2% (v/v) Triton X-100; 12.5 mM $MnCl_2$ and 1.25 mM AMP. The reactions were allowed to begin with the addition of 1 mM UDP-*N*-acetylgalactosamine plus 1.25 nM UDP-[^{3}H]GalNAc (150,000 DPM; 0.6 Gbq. $mmol^{-1}$; 25.1 kBq/40 μL). Reactions proceeded for 3 and 24 h at 37 °C and the products were passed through AG1-X8 resin, eluted with 3 mL water and incorporation rates calculated by scintillation counting. Background values obtained from controls incubated without peptide substrate were substracted from each experimental value.

Analytical.—Reaction products from the aforementioned enzyme assays were also analyzed by mass spectrometry and/or capillary electrophoresis. To desalt samples prior to capillary electrophoresis and/or mass spectrometry, Sep-Pak C18 reverse-phase cartridges were used as described previously.[25] Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) was performed using a Vision 2000 time-of-flight instrument (Finnigan MAT, Bremen, Germany) equipped with a 337 nm UV laser. The mass spectra were acquired in reflectron mode under 6 kV acceleration voltage and positive detection. The samples were prepared by mixing directly onto the target 1 μL of analyzed solution (typically 50 pmol) and 1 μL of a 2,5-dihydroxybenzoic acid (DHB) matrix solution (12 mg/mL in 7:3 MeOH–water), then allowed to crytallize at rt. External calibration was performed using the MUC5AC peptide (MW 1502.7). Shots (10–30) were accumulated for the mass spectrum.

Capillary electrophoresis was performed on a P/ACE system model 5000 (Beckman, Fullerton, CA, USA) under conditions previously described. For the separation of the hexadecapeptides, 2 N formic acid buffer was made 2.5% with respect to polyvinyl alcohol (MW 15,000) (v/v).[26] To determine O linkage sites, a preparative scale procedure was performed as described by Bielher and Schwartz[27] and the recovered fractions were then analyzed by Edman degradation using an Applied Biosystems gas-phase Sequencer, model 477A as described previously.[25]

Acknowledgements

The authors thank Drs KG Ten Hagen and LA Tabak (Center for Oral Biology, Aab Institute for Biomedical Sciences, University of Rochester, Rochester, NY, USA) for providing generously the recombinant GaNTase enzymes. We are indebted to Dr Guy Ricart USTL Villeneuve d'Ascq for help in mass spectrometry and Dominique Demeyer, Nadine Parsy, Delphine Cools for their skilful help in performing the experiments.

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97–130.
2. Hart, G. W. *Ann. Rev. Biochem.* **1997**, *58*, 841–874.
3. Dennis, J. W.; Granovsky, M.; Warren, C. E. *Biochim. Biophys. Acta* **1999**, *1473*, 21–34.
4. Brockhausen, I. *Biochim. Biophys. Acta* **1999**, *1473*, 67–95.
5. Clausen, H.; Bennett, E. P. *Glycobiology* **1996**, *6*, 635–646.
6. Ten Hagen, K. G.; Bedi, G. S.; Tetaert, D.; Kingsley, P.; Hagen, F. K.; Balys, M. M.; Beres, T. M.; VanWuyckhuyse, B.; Degand, P.; Tabak, L. A. *J. Biol. Chem.* **2001**, *276*, in press.
7. Mandel, U.; Hassan, H.; Therkildsen, M. H.; Rygaard, J.; Jakobsen, M. H.; Juhl, B. R.; Dabelsteen, E.; Clausen, H. *Glycobiology* **1999**, *9*, 43–52.
8. Hagen, F. K.; Hazes, B.; Raffo, R.; deSa, D.; Tabak, L. A. *J. Biol. Chem.* **1999**, *274*, 6797–6803.
9. Hagen, F. K.; Nehrke, K. *J. Biol. Chem.* **1998**, *273*, 8268–827752.
10. Wandall, H. H.; Hassan, H.; Mirgorodskaya, E.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. *J. Biol. Chem.* **1997**, *272*, 23503–23514.

11. Kingsley, P. D; TenHagen, K. G.; Maltby, K. M.; Zara, J.; Tabak, L. A. *Glycobiology* **2000**, *10*, 1317–1323.
12. Bennett, E. P.; Weghuis, D. O.; Merkx, G.; Van Kessel, A. G.; Eiberg, H.; Clausen, H. *Glycobiology* **1998**, *8*, 547–555.
13. Soudan, B.; Hennebicq, S.; Tetaert, D.; Boersma, A.; Richet, C.; Demeyer, D.; Briand, G.; Degand, P. *J. Chromatogr. B.* **1999**, *729*, 65–74.
14. Guyonnet-Dupérat, V.; Audié, J. P.; Debailleul, V.; Laine, A.; Buisine, M. P.; Galiegue-Zouitina, S.; Pigny, P.; Degand, P.; Aubert, J. P.; Porchet, N. *J. Biochem.* **1995**, *305*, 211–219.
15. Yoshida, A.; Suzuki, M.; Ikenaga, H.; Takeuchi, M. *J. Biol. Chem.* **1997**, *272*, 16884–16888.
16. Asada, M.; Orikasa, N.; Yoneda, A.; Ota, Y.; Ota, K.; Imamura, T. *J. Glycoconjugate* **1999**, *16*, 321–326.
17. Ten Hagen, K. G.; Tetaert, D.; Hagen, F. K.; Richet, C.; Beres, T. M.; Gagnon, J.; Balys, M. M.; Van-Wuyckhuyse, B.; Degand, P.; Tabak, L. A. *J. Biol. Chem.* **1999**, *274*, 27867–27874.
18. Iida, S.-I.; Takeuchi, H.; Hassan, H.; Clausen, H.; Irimura, T. *FEBS Lett.* **1999**, *449*, 230–234.
19. Iida, S.-I.; Takeuchi, H.; Kato, K.; Yamamoto, K.; Irimura, T. *J. Biochem.* **2000**, *347*, 535–542.
20. Wang, Y.; Abernethy, J. L.; Eckhardt, A. E.; Hill, R. L. *J. Biol. Chem.* **1992**, *267*, 12709–12716.
21. Bennett, E. P.; Hassan, H.; Clausen, H. *J. Biol. Chem.* **1996**, *271*, 17006–17012.
22. Kim, Y. J.; Borsig, L.; Han, H.-L.; Varki, N. M.; Varki, A. *Am. J. Pathol.* **1999**, *155*, 461–472.
23. Iida, S.-I.; Yamamoto, K.; Irimura, T. *J. Biol. Chem.* **1999**, *274*, 10697–10705.
24. Hagen, F. K.; Ten Hagen, K. G.; Beres, T. M.; Balys, M. M.; Van-Wuyckhuyse, B. C.; Tabak, L. A. *J. Biol. Chem.* **1997**, *272*, 13843–13848.
25. Hennebicq, S.; Tetaert, D.; Soudan, B.; Boersma, A.; Briand, G.; Richet, C.; Gagnon, J.; Degand, P. *J. Glycoconjugate* **1998**, *15*, 275–282.
26. Boulis, Y.; Richet, C.; Haupt, K.; Hennebicq, S.; Soudan, B.; Tetaert, D.; Degand, P.; Vijayalakshmi, M. A. *J. Chromatogr. A.* **1998**, *805*, 285–293.
27. Bielher, R.; Schwartz, H. E. *Tech. Bull. TIBC (Palo Alto, Calif Beckman Instr. Spinco Division)* **1995**, *105*, 1–5.