

Identification of Further Elongation and Branching of Dimeric Type 1 Chain on Lactosylceramides from Colonic Adenocarcinoma by Tandem Mass Spectrometry Sequencing Analyses^{*S}

Received for publication, August 30, 2007, and in revised form, February 20, 2008 Published, JBC Papers in Press, April 15, 2008, DOI 10.1074/jbc.M707274200

Yao-Yun Fan^{‡S}, Shin-Yi Yu^{‡S}, Hiromi Ito[¶], Akihiko Kameyama[¶], Takashi Sato[¶], Chi-Hung Lin^{§||}, Lung-Chih Yu^{‡S}, Hisashi Narimatsu[¶], and Kay-Hooi Khoo^{‡S1}

From the [‡]Institute of Biochemical Sciences, National Taiwan University, Taipei 106, Taiwan, [§]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan, [¶]Glycogene Function Team of Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, and ^{||}CBMB, Taiwan International Graduate Program, Academia Sinica, Taipei 115, Taiwan and Institute of Bioinformatics and Structural Biology, National Tsing-Hua University, Hsinchu 30013, Taiwan

Mammalian glycan chain elongation is mostly based on extending the type 2 chain, Gal β 1–4GlcNAc, whereas the corresponding type 1 chain, Gal β 1–3GlcNAc, is not normally extended. In a broader context of developing high sensitivity mass spectrometry methodologies for glycomic identification of Le^a versus Le^x and linear versus branched poly-*N*-acetyl-lactosamine (polyLacNAc), we have now shown that the dimeric type 1 glycan chain, as carried on the lactosylceramides of a human colonic adenocarcinoma cell line, Colo205, not only can be further extended linearly but can likewise be branched at C6 of 3-linked Gal in a manner similar to polyLacNAc. A combination of chemical and enzymatic derivatization coupled with advanced mass spectrometry analyses afforded unambiguous identification of a complex mixture of type 1 and 2 hybrids as well as those fucosylated variants founded exclusively on linear and branched trimeric type 1 chain. We further showed by *in vitro* enzymatic synthesis that extended type 1 and the hybrid chains can be branched by all three forms of the human I branching enzymes (IGnT) currently identified but with lower efficiency and stringency with respect to branching site preference. Importantly, it was found that a better substrate is one that carries a Gal site for branching that is extended at the non-reducing end by a type 2 and not a type 1 unit, whereas the IGnTs are less discriminative with respect to whether the targeted Gal site is itself β 3- or β 4-linked to GlcNAc at the reducing end.

Type 1 (Gal β 1–3GlcNAc) and type 2 (Gal β 1–4GlcNAc) glycan chains represent the two most common peripheral back-

bone sequences extending from the core structures of mammalian glycoconjugates to which linkage-specific additions of terminal fucose, sialic acid, and other glycosyl residues give rise to a variety of commonly found bioactive epitopes. Extension of type 2 chain with additional Gal β 1–4GlcNAc (LacNAc)² disaccharide units into polylactosaminoglycans (polyLacNAc) occurs readily with and without further branching (1, 2) and is often associated with onco-developmental activation (3, 4). In contrast, the type 1 disaccharide units are usually found attached directly to the glycan core or non-reducing termini of type 2 chains (5, 6) and not further elongated.

Among the first reported incidences of extended type 1 chains are those constituting the lacto series glycosphingolipids (GSL) of human meconium, which could be extended by one or two additional type 2 disaccharide units with varying degrees of fucosylation (7). A decade later, in the early 90s, fucosylated dimeric type 1 chains on lactosylceramides were identified as tumor-associated antigens in the form of Le^b-Le^a and Le^a-Le^a (8, 9). The colonic adenocarcinoma cell line, Colo205, was thought to exhibit such unusual glyco-phenotypes due to its abnormally high β 3-galactosyltransferase (β 3GalT) activities (10, 11) coupled with possibly a novel or up-regulated β 3-*N*-acetylglucosaminyltransferase (β 3GnT) activity (12). More recently, another rare occurrence of extended type 1 chain was reported for the GSLs isolated from the small intestine of an individual who is of blood group O, Le(a-b-) nonsecretor status (13). Non-fucosylated dimeric type 1 chain on lactosylceramide was found to carry an additional fucosylated type 2 branch. This suggests that although the fucosylated dimeric type 1 chain, Le^{b/a}-Le^a, may be truly tumor-associated, type 1 chain exten-

^{*} This work was supported by an Academia Sinica Program Project grant (to K.-H. K.) and was performed as a part of the R&D Project of the Industrial Science and Technology Frontier Program supported by the New Energy and Industrial Technology Development Organization (to H. N.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^S The on-line version of this article (available at <http://www.jbc.org>) contains a supplemental figure.

¹ To whom correspondence should be addressed: Institute of Biological Chemistry, Academia Sinica, 128, Academia Rd. Sec 2, Nankang, Taipei 11529, Taiwan. Fax: 886-2-27889759; E-mail: kkhoo@gate.sinica.edu.tw.

² The abbreviations used are: LacNAc, Gal β 1–4GlcNAc; polyLacNAc, poly-*N*-acetyl-lactosamine; β 3GnT, β 3-*N*-acetylglucosaminyltransferase; β 3GalT, β 3-galactosyltransferase; β 4GalT, β 4-galactosyltransferase; Fuc, fucose; GSL, glycosphingolipid; Hex, hexose; HexNAc, *N*-acetylhexosamine; IGnT, β 6-*N*-acetylglucosaminyltransferases; LacCer, Gal β 1–4-Glc (Lac) ceramide; NeuAc, *N*-acetylneuraminic acid; type 1, Gal β 1–3GlcNAc; type 2, Gal β 1–4GlcNAc; CID, collision induced dissociation; HPLC, high performance liquid chromatography; HPTLC, high performance thin layer chromatography; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; QIT, quadrupole ion trap; Q/TOF, quadrupole-time of flight; TOF/TOF, tandem time of flight; mAb, monoclonal antibody.

Extension and Branching of Dimeric Type 1 Chain

sion itself, either with another type 1 or type 2 unit, may have a wider occurrence on normal human tissues. It can be associated with a particular genetic background, developmental stage, or activation process whereby glycosyltransferases normally responsible for capping off type 1 chains are down-regulated coupled with elevated basal activities of $\beta 3\text{GnT}$ and $\beta 3\text{GalT}$.

A true evaluation of the occurrence and functions of extended type 1 chain is nonetheless difficult, largely hindered by a lack of sensitive detection methods either by analytical chemistry approach or biological probes such as monoclonal antibodies with strict specificity. However, recent advances in mass spectrometry (MS) analysis have opened up new possibilities in revisiting these issues. In particular, efficient MALDI-MS/MS utilizing both low and high energy collision induced dissociation (CID) is now possible at sufficiently high sensitivity to allow meaningful analysis of real biological samples (14). Using the well characterized GSLs from Colo205 and the glycans derived thereof, we aimed to develop a facile MS/MS sequencing method to critically map the various possible combination of extended and branched type 1 and type 2 glycan chains.

We now show that further extension or branching of dimeric type 1 chain carried on the lactosylceramides of Colo205 is possible. MS/MS fragmentation pattern was established from previous and this work, in which the signal assignment was validated by comparative analysis of authentic standards synthesized *in vitro* by enzymes. We further demonstrate that the human I branching $\beta 6\text{-N}$ -acetylglucosaminyltransferase (IGnT) is capable of transferring a GlcNAc onto the branch point of an extended type 1 chain albeit with significantly lower efficiency. Interestingly, by testing against all four possible combination of dimeric type 1 and type 2 chains as acceptors, we show that branching is significantly more efficient if a terminal type 2 and not type 1 unit is at the non-reducing terminus of the Gal and less dependent on whether the Gal itself is $\beta 3$ - or $\beta 4$ -linked to GlcNAc at the reducing end.

EXPERIMENTAL PROCEDURES

Extraction and Fractionation of GSLs—The Colo205 cells (ATCC no. CCL-222) were cultured in RPMI 1640 medium containing 10% fetal calf serum in a 5-liter spinner. GSLs were extracted from harvested cells by 15 volumes of 55:25:20 (v:v:v) isopropanol:hexane:water and then subjected to a Folch partition in 4:2:1 (v:v:v) chloroform:methanol:water (8). The Folch upper layer containing the GSLs were dried and re-dissolved in 30:60:8 (v:v:v) chloroform:methanol:water for loading onto a DEAE column. Neutral GSLs were washed through and collected in 3 volumes of 30:60:8 (v:v:v) chloroform:methanol:water. Size fractionation of the neutral GSLs was performed on a silica gel column (1.6 $\times$ 5 cm) packed with Iatrobeds 6RS-8060 (Iatron Laboratories Inc). The column was pre-equilibrated with 55:40:5 (v:v:v) isopropanol:hexane:water and stepwise eluted with decreasing ratios of 55:40:5 (v:v:v) to 55:20:25 (v:v:v) of isopropanol:hexane:water.

HPTLC of GSLs and Immunostaining—Neutral GSLs were spotted on HPTLC plate (silica gel 60 F254, Merck) and developed with a mobile phase containing chloroform:methanol:water in a ratio of 50:40:10 (v:v:v). GSLs were stained by spraying 0.2% orcinol (Sigma) in 10% H_2SO_4 and incubated for 10 min at

110 °C in oven. For immunostaining, the HPTLC plate was first fixed with 0.5% polyisobutylmethacrylate (GlycoTech, Gaithersburg, MD) in chloroform:hexane, 1:9 (v: v), for 30 s followed by blocking for 10 min in 3% bovine serum albumin/phosphate-buffered saline. The plates were then washed by phosphate-buffered saline and incubated with primary antibodies CF4C4 (anti-Le^a) and SH1 (anti-Le^x) (both kindly provided by GlycoNex Inc., Taiwan) at room temperature for 1 h followed by biotinylated secondary antibody at room temperature for 1 h. The plate was then further incubated with an avidin-biotin complex kit (Vector Laboratories, Burlingame, CA) at room temperature for 30 min to amplify signals from secondary antibody, followed by development with Immunostaining HRP-1000 kit (Konica Minolta).

Enzymatic Release, Digest, and Synthesis of Glycans—The glycan moiety of GSL was released from the ceramide core by digestion with 5 milliunits of ceramide glycanase (Calbiochem, catalog no. 219484) in 50 mM sodium acetate, pH 5.0, containing 0.1% sodium cholate at 37 °C overnight. The released glycans were separated from the ceramide, salt, and detergents by passing through a Supelclean Envi-carb SPE tube (Supelco) pre-conditioned in the sequential order of 1 N NaOH, water, 30% acetic acid, water, 75% acetonitrile + 0.1% trifluoroacetic acid, and water. After extensive washing with water, the glycan sample was eluted with 3 ml of 25% acetonitrile, 0.1% trifluoroacetic acid.

The released glycans were first defucosylated by 48% aqueous hydrofluoric acid at 0 °C in the dark for 48 h before glycosidase digestions. β -1-3-Galactosidase (10 milliunits, from recombinant *Escherichia coli*, Calbiochem, catalog no. 345795) digestion was performed at 37 °C for 16 h in 50 mM sodium acetate, pH 4.5, and endo- β -galactosidase (5 milliunits, from *Escherichia freundii*, Seikagaku catalog no. 100455) digestion was performed at 37 °C for 16 h in 50 mM sodium acetate, pH 5.8. Recombinant α 1-3-fucosyltransferase from *Helicobacter pylori* strain NCTC11639 was kindly provided by Dr. Chun-Hung Lin (Institute of Biological Chemistry, Academia Sinica, Taiwan) (15). The fucosylation reaction was carried out in 50 mM Tris-HCl, pH 7.5, containing 1 mM MnCl_2 , and 1 mM GDP-fucose (50 nmol).

For *in vitro* enzymatic synthesis of glycans with hybrid type 1 and type 2 chains, 20 μM lacto-*N*-tetraose and lacto-*N*-neotetraose (Calbiochem) were separately incubated with recombinant human $\beta 3\text{-N}$ -acetylglucosaminyltransferase 2 ($\beta 3\text{GnT}2$) in 25 mM HEPES buffer, pH 7.0, containing 10 mM MnCl_2 and 2 mM UDP-GlcNAc at 37 °C for 24 h. The reactions were then terminated by boiling at 100 °C for 5 min. Subsequently, one half of each product was incubated with recombinant human $\beta 3$ -galactosyltransferase 5 ($\beta 3\text{GalT}5$), and the other half was incubated with recombinant human $\beta 4$ -galactosyltransferase 1 ($\beta 4\text{GalT}1$) along with 1 mM UDP-Gal for another 24 h at 37 °C. Glycans with linear extended type 1 or type 2 chains were synthesized by incubation with $\beta 3\text{GalT}5$ and $\beta 3\text{GnT}2$ or with $\beta 4\text{GalT}1$ and $\beta 3\text{GnT}2$, respectively, in 25 mM HEPES buffer, pH 7.0, containing MnCl_2 , UDP-GlcNAc, and UDP-Gal as described above. For the I-branching $\beta 6\text{-N}$ -acetylglucosaminyltransferase (IGnT) assay, the glycan chains were similarly incubated with recombinant human IGnT1-3 (16) in 25 mM HEPES containing 10 mM EDTA and 5 mM UDP-GlcNAc. All

recombinant human glycosyltransferases were produced in HEK293T cells and were used as described (17). On completion, the reaction mixtures were spun through an Ultrafree-MC filter unit (0.22 μ m, polyvinylidene difluoride, Millipore), and the synthesized free glycans along with any unreacted acceptor were recovered from the filtrate by passing through the Envirocarb SPE tube, as described above.

Fractionation and Purification of Trimeric Type 1 Chain—HPLC purification of the glycans released from GSLs was performed on a normal phase Type N PALPAK column (Takara, 250 $\times$ 4.6 mm). Buffer A was 75% acetonitrile in 200 mM triethylamine, and buffer B was 50% acetonitrile in 200 mM triethylamine. The sample was eluted at a flow rate of 0.8 ml/min with the program set at isocratic elution in 100% solvent A for 5 min followed by a stepwise linear gradient to 50% solvent B in 5 min, to 85% solvent B in 40 min, and finally to 100% solvent B in 10 min. Aliquots from every five fractions were combined and screened by MS to locate eluted glycans. (Gal-GlcNAc)₃Lac was eluted at 22–24 min by this program.

Quantification of *IGnT* Transcripts—Total RNA of the Colo205 colon cancer cell line was prepared using the RNeasy Mini kit (Qiagen GmbH, Hilden, Germany). The first-strand cDNAs were primed using oligo-dT primer and synthesized by SuperScript III reverse transcriptase (Invitrogen). The forward IAF16 (5'-GACACTCAACAGGATTCCCGGTGT-TCC-3', nucleotides +906 through +932 of exon 1A of the *IGnT* gene, translation initiation codon as nucleotides +1 ~ +3), IBF18 (5'-CCAGTGGTCCAAGGACACTTTCAGTC-CTGA-3', nucleotides +902 through +929 of exon 1B of the *IGnT* gene), and ICF20 (5'-CACTTAATAGGGTTTCAGG-TGTTCCCTGG-3', nucleotides +908 through +935 of exon 1C of the *IGnT* gene) primers were used to amplify *IGnTA*, *IGnTB*, and *IGnTC* cDNAs, respectively, by PCR. The reverse primer used, IRs (5'-CCAGCTGGGTTGTATCGCAG-3', complementary to the sequence locating in exon 3 of the *IGnT* gene), was common for the 3 *IGnT* cDNAs. PCR for the β -actin cDNA was performed using the ACTF (5'-GCACCAGGGCG-TGATGG-3') and ACTR (5'-GCCTCGGTCAGCAGCA-3') primers. The cDNA sample and 2 pmol of each forward and reverse primer were added to 20 μ l of the SYBR Green Master Mix (Applied Biosystems, Foster City, CA). Real-time PCR detection (GenAmp 7300 Sequence Detection System, Applied Biosystems) was used for quantification of the transcripts. Serial dilutions of the pGEM-T vector (Promega Co., Madison, WI) containing *IGnTA*, *IGnTB*, *IGnTC*, or β -actin cDNAs were used to generate standard curves for each transcript. The *IGnTA*, *IGnTB*, and *IGnTC* genes as described (18) encode for the *IGnT2*, *IGnT1*, and *IGnT3* enzymes (19), respectively, used in this work.

MS and MS/MS Analysis—All glycan samples were permethylated using the NaOH/dimethyl sulfoxide slurry method as described by Dell *et al.* (20). Briefly, NaOH pellets were grounded in DMSO until a slurry was formed. 0.5 ~ 1 ml of the DMSO/NaOH slurry was added to glycans dried down in screw-capped glass tubes followed by about 0.2 ml of methyl iodide (Merck). The reaction mixtures were vortexed and placed on the automatic shaker for 20 min at room temperature, after which the reaction was terminated by the a dropwise

addition of 1 ml of water. Permethylated sample was then extracted into 1 ml of chloroform and washed several times with water.

For MALDI-TOF MS profiling, permethylated samples were dissolved in acetonitrile and mixed 1:1 with 10 mg/ml 2,5-dihydroxybenzoic acid (Sigma) in acetonitrile for spotting onto the target plate. Data acquisition was performed on either a bench top MALDI LR system (Waters Micromass) or a 4700 Proteomics Analyzer (Applied Biosystems), both operated in the reflection mode.

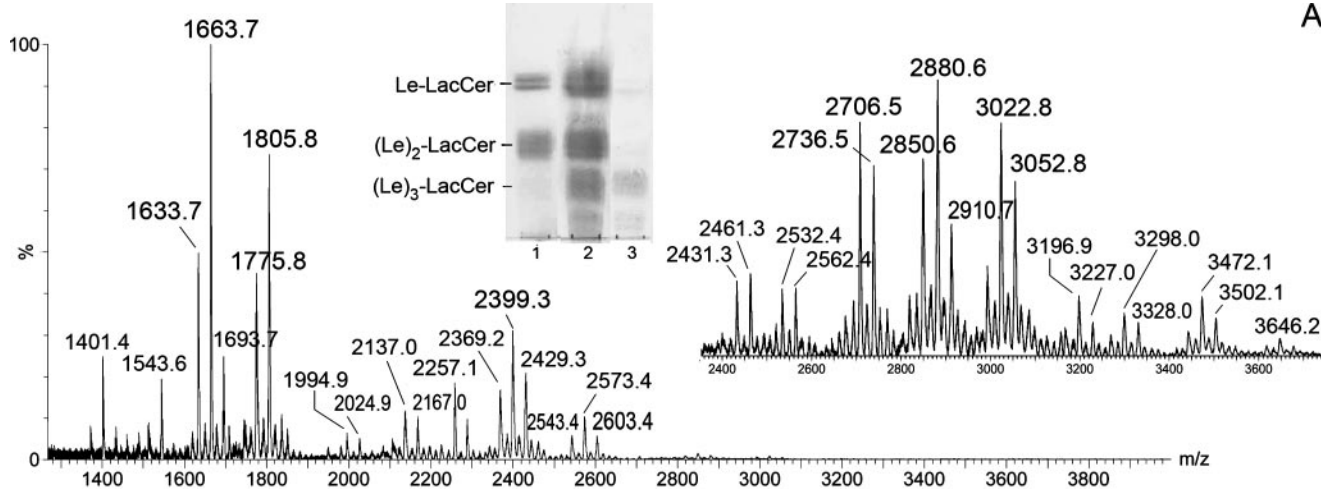
MALDI-MS/MS analyses of permethylated glycan and GSL samples at low collision energy were performed on a Q/TOF UltimaTM MALDI (Waters Micromass) using α -cyano-4-hydrocinnamic acid as matrix (5 mg/ml in 50% acetonitrile, 0.1% trifluoroacetic acid mixed 1:1 with sample dissolved in acetonitrile). Argon was used as the collision gas with a collision energy manually adjusted between 100 and 200 V to achieve an optimum degree of fragmentation. Alternatively, multistage low energy CID-MSⁿ of the permethylated glycans were acquired on an AXIMATM-QIT (Kratos Analytical, Shimadzu) using 2,5-dihydroxybenzoic acid as a matrix as described (21, 22). High energy CID MS/MS analyses were performed on the 4700 Proteomics Analyzer TOF-TOF system using 2,5-dihydroxybenzoic acid as matrix. The potential difference between the source acceleration voltage and the collision cell was set at 3 kV. The indicated collision cell pressure was increased from 3×10^{-8} (no gas) to around 2×10^{-6} torr by letting in argon as the collision gas. All CID MS/MS data were acquired and processed manually.

RESULTS

The Overall Profile and Characteristic Patterns of GSLs from Colo205—Multifucosylated variants of extended type 1 chain were first identified in the form of Le^b-Le^a and Le^a-Le^a on GSLs isolated from tumors in nude mice implanted with Colo205 (8, 9). Although it was evident from HPTLC immunostaining that smaller and larger Le^a-carrying GSLs were present (Fig. 1A, *inset*), a more definitive overall profile such as one that would be afforded by MS analysis was not presented. In this work, total and further fractionated GSLs similarly extracted from harvested Colo205 cell cultures were mapped by MALDI-MS before and after permethylation (Fig. 1A). Based on the mass shifts observed after derivatization and further MS/MS analyses, the GSLs detected can in general be defined as Fuc_n(Hex-HexNAc)_n-LacCer, where (Hex-HexNAc)_n refers to either type 1 or 2 units, and LacCer refers to Gal β 1-4-Glc (Lac) linked to the ceramide (Cer) core (Table 1).

As expected, the Colo205 GSLs MS profile is clearly dominated by signal clusters corresponding to multiply fucosylated (Hex-HexNAc)_{1,2}-LacCer. The larger and those sialylated ones were visibly present but would become more apparent only after charge and size fractionation (Fig. 1, A and B, *upper spectra*). Within the signal clusters, mass differences of 14 (CH₂) or 30 (OCH₂) mass units indicated heterogeneity in the fatty acyl chain length and the additional presence of hydroxyl group, respectively, whereas 142 mass units is a common mass difference corresponding to a C16:0 to hydroxylated C24:0 substitution. Removal of the highly heterogeneous ceramide moiety by

A



B

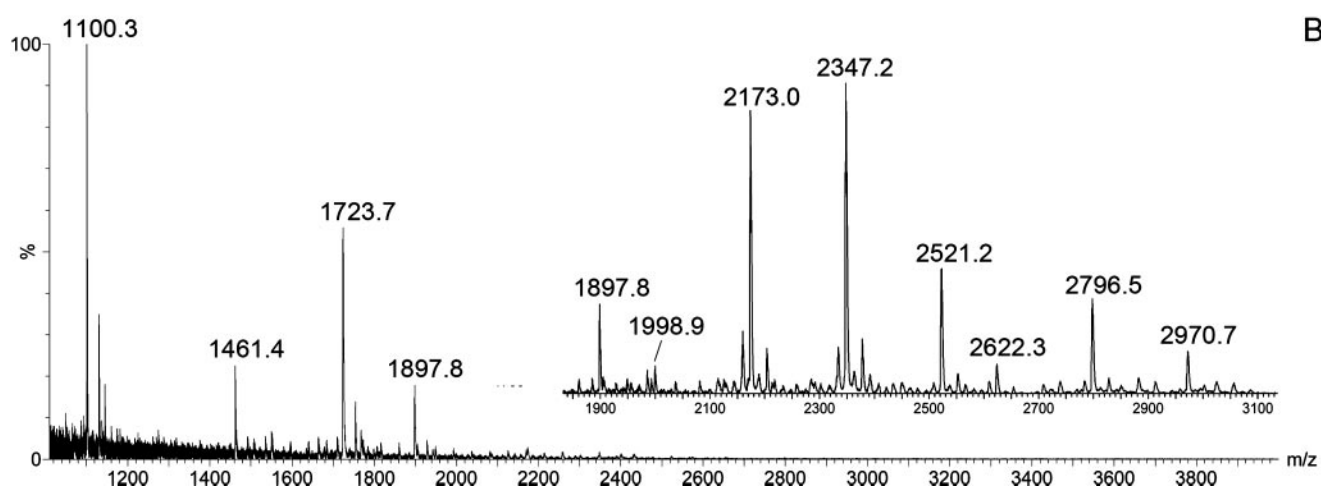


FIGURE 1. **Overall GSL profiles of Colo205.** MALDI-MS spectra of permethylated total GSLs isolated from Colo205 (A) and the glycans derived thereof by ceramidase digestion (B) are shown as the lower spectra in each panel. GSLs carrying trimeric Lewis units, corresponding to the HPTLC bands indicated as (Le)₃LacCer, were further purified from the total pools. The respective MS profiles of these are shown as the upper spectra. The molecular compositions of the major peaks are assigned as listed in Table 1. NeuAc-Lac was also detected as a major peak at m/z 838 but is not included in the mass range shown in B. The inset in A shows the HPTLC profiles of the neutral GSLs stained with orcinol/sulfuric acid spray (lane 1) or immunostained by anti-Le^a mAb CF4C4 (lane 2) and anti-Le^x mAb SH1 (lane 3), respectively. The bands corresponding to (Le)₃LacCer are not obvious by orcinol stain but are clearly visible by immunoblot against Le^a. (Le)₃LacCer also appeared as a faint band by staining against Le^x. Le, Le^a, Le^b, or Le^x.

ceramide glycanase yielded a much simplified MS profile with distinctive $[M+Na]^+$ molecular ion signals for the released glycan structures that share the backbone composition, (Hex-HexNAc)_n-Lac (Fig. 1B, Table 1). Thus, the major signal at m/z 1100 corresponds to permethylated Fuc₁(Hex-HexNAc)₁-Lac, and those at m/z 1723 and 1897 correspond to permethylated Fuc_{2,3}(Hex-HexNAc)₂-Lac, which are consistent with previous identification of Le^{b/a}-Le^a-LacCer as the characteristic GSLs of Colo205. Importantly, the data also unambiguously revealed the presence of Fuc₂₋₄(Hex-HexNAc)₃-Lac (m/z 2173, 2347, and 2521) and Fuc₂₋₄(Hex-HexNAc)₄-Lac (m/z 2622, 2796, and 2970) in the fraction highly enriched for the trimeric Lewis structures (Fig. 1, A and B, upper spectra). These were taken for MS/MS analyses along with the already well characterized Le^b-Le^a-Lac structure and a commercially available Le^b-Le^x-Lac standard, which serve to establish the fragmentation pattern that would allow discrimination of type 1 and 2 chains.

MALDI MS/MS Characteristics for Identification of Type 1 and 2 Chains—We have previously demonstrated that the low and high energy CID MALDI MS/MS fragmentation pattern of

complex glycans as acquired on the Q/TOF and TOF/TOF, respectively, are highly complementary. Although the overall sequence, branching pattern, core type, and terminal epitope substituents of an *N*- or *O*-glycan were found to be best determined via low energy CID MS/MS due to its simplicity, high energy CID MS/MS on a TOF/TOF can provide additional linkage information through a more complicated cleavage pattern (14). In general, a type 1 chain or Le^a epitope is distinguished from a type 2 chain or Le^x via the C3-linkage specific cleavage. Low energy CID MS/MS of permethylated glycans on a Q/TOF would readily afford a Z ion via elimination of glycosyl substituent at C3 position (6, 23). Thus, a Z ion corresponding to Δ GlcNAc-R would be afforded by -Gal-3GlcNAc-R and not -Gal-4GlcNAc-R. In the current context, the presence of an internal Le^a unit will, therefore, be easily identified by the corresponding Z ion, which would not be observed if it is a Le^x unit. This is schematically illustrated with the cleavage pattern afforded by low energy CID MS/MS on Le^b-Le^a-Lac versus Le^b-Le^x-Lac (Fig. 2 (i), MS/MS data are not shown). The former is distinguished by the additional presence of the

TABLE 1
Molecular composition of major GSLs from Colo205

[M+Na] ⁺ of released glycan ^a	Glycan composition	[M+Na] ⁺ of corresponding GSL ^a	Probable ceramide composition ^b
838.2	NeuAc-Lac	1401.4	d18:1 h16:0
		1543.6	t18:0 h24:1
1100.3	(Fuc) ₁ (LacNAc) ₁ -Lac	1633.7	d18:1 h16:0
		1663.7	d18:1 h16:0
		1693.7	t18:1 h16:0
		1775.8	d18:1 h24:0
		1805.8	t18:0 h24:1
1461.4	NeuAc-(Fuc) ₁ (LacNAc) ₁ -Lac	1994.9	d18:1 h16:0
		2024.9	d18:1 h16:0
		2137.0	d18:1 h24:0
		2167.0	t18:0 h24:1
1723.7	(Fuc) ₂ (LacNAc) ₂ -Lac	2257.1	d18:1 h16:0
		2287.1	d18:1 h16:0
		2369.2	d18:1 h24:0
		2399.2	d18:1 h24:0
		2429.3	t18:0 h24:1
1897.8	(Fuc) ₃ (LacNAc) ₂ -Lac	2431.3	d18:1 h16:0
		2461.3	d18:1 h16:0
		2573.4	d18:1 h24:0
		2603.4	t18:0 h24:1
1998.9	(Fuc) ₁ (LacNAc) ₃ -Lac	2532.4	d18:1 h16:0
		2562.4	d18:1 h16:0
2173.0	(Fuc) ₂ (LacNAc) ₃ -Lac	2706.5	d18:1 h16:0
		2736.5	d18:1 h16:0
		2850.6	d18:0 h24:0
2347.2	(Fuc) ₃ (LacNAc) ₃ -Lac	2880.6	d18:1 h16:0
		2910.7	d18:1 h16:0
		2992.8	d18:1 h24:0
		3022.8	d18:1 h24:0
		3052.8	t18:0 h24:1
2521.2	(Fuc) ₄ (LacNAc) ₃ -Lac	3196.9	d18:1 h24:0
		3227.0	t18:0 h24:1
2622.3	(Fuc) ₂ (LacNAc) ₄ -Lac	3298.0	d18:1 h24:0
		3328.0	t18:0 h24:1
2796.5	(Fuc) ₃ (LacNAc) ₄ -Lac	3472.1	d18:1 h24:0
		3502.1	t18:0 h24:1
2970.7	(Fuc) ₄ (LacNAc) ₄ -Lac	3646.2	d18:0 h24:1

^a The *m/z* values tabulated refer to the monoisotopic masses detected for each of the sodiated molecular ions afforded by the permethyl derivatives, as labeled on Fig. 1. *m/z* 828 was also detected as a major signal but not included in the mass range shown in Fig. 1B.

^b The ceramide moieties (sphingosine and fatty acid) for the major GSLs detected were mostly inferred from the molecular masses and Karlsson and Larson (7) and Stroud (8). The stereochemistry and the location of additional degrees of unsaturated bonds and hydroxylation were not defined in this work. For more abundant peaks, direct MS/MS on the intact GSLs were also performed to ascertain the major isomeric forms. For the larger GSLs, only those eluted into the fraction used for subsequent analysis are tabulated. d18:1, (4*E*)-sphingenine; t18:0, (4*D*)-hydroxysphinganine or phytosphingosine; h, hydroxyl form of the fatty acids.

C and Z ion pair at *m/z* 1056 and 864. A problem occurs, however, when both structural isomers are present, as often encountered in real samples not subjected to extensive purification, since there is no critical ion that would define the additional presence of an internal Le^x unit apart from elimination of Fuc.

Seeking corroborative evidence for better delineation of type 1/type 2 or Le^a/Le^x, it was found that an internal Le^x unit would readily afford a distinctive D ion (*m/z* 1269, Fig. 2*B*) under high energy CID MS/MS via elimination of the 3-linked Fuc in concert with glycosidic cleavage at the GlcNAc (Fig. 2*C*). This provides a critical ion indicative of an internal Le^x unit, the exact *m/z* value of which will be dependent on the retained substituents at C4. A Le^a unit would afford instead the corresponding D ion at *m/z* 442 irrespective of its C3 substituents. Although not informative with respect to its location, this abundant D ion (*m/z* 442) is very useful to diagnose the presence of Le^a unit at high sensitivity. Other ions detected and assigned for the high energy CID MS/MS analysis of the dimeric Lewis structures on MALDI-TOF/TOF (Fig. 2 (*ii*)) are consistent with a

previously established fragmentation pattern (14), including the linkage-specific satellite ions arising from concerted elimination of glycosyl substituents around the ring, coined as E and G ions (24), and the ^{1,5}X cross-ring cleavage ions formed at every glycosyl residues, which facilitate a complete sequencing from the non-reducing terminus. Another diagnostic ion identified is *m/z* 243, which corresponds to a C''/Y ion formed through a glycosidic cleavage at the non-reducing terminal Gal in concert with loss of the glycosyl substituent at C2 and, therefore, indicative of Fuc-2Gal epitope as found in Le^b or Le^y.

Based on these characteristic fragment ions, MALDI-MS/MS analysis of the larger Fuc₂₋₄(Hex-HexNAc)₃-Lac structures derived from Colo205 GSLs showed that in addition to a linear backbone chain, ions indicative of a branched Hex-HexNAc-(Hex-HexNAc)Hex-HexNAc-Lac backbone could be detected for the under-fucosylated structures, the inner Hex-HexNAc unit of which was mostly not fucosylated (data not shown). More significantly, apart from heterogeneity in the degree of fucosylation, the linkage-specific Z ions indicative of type 1 chain could be detected for each of the three Hex-HexNAc units (Fig. 3). However, D ions indicative of the type 2 chain could also be detected at least for the terminal and middle Hex-HexNAc units. Collectively, the MS/MS data were, therefore, suggestive of the occurrence of a trimeric type 1 chain but did not allow its unequivocal identification among the isomeric mixtures comprising different permutation of type 1 and 2 hybrids. We, thus, set out to determine specifically if a dimeric type 1 chain can be further extended linearly or branched with another type 1 or 2 unit.

Identification of Linear and Branched Trimeric Type 1 Chain—Focusing on the type 1 and 2 backbone chains, the fraction highly enriched with Fuc₂₋₄(Hex-HexNAc)₃-LacCer was first treated with ceramide glycanase as before followed by defucosylation with hydrofluoric acid to yield a single major component corresponding to (Hex-HexNAc)₃-Lac, as defined by MALDI-MS of the permethylated product ([M+Na]⁺ molecular ion at *m/z* 1824). This was then isolated by normal-phase HPLC and subjected to endo-β-galactosidase digestion to degrade any chain with an internal type 2 unit. Under the same experimental conditions, the enzyme was independently shown not to act on an internal type 1 unit as tested against an enzymatically synthesized authentic standard of a trimeric type 1 linear chain (data not shown). A major degradation product was obtained that was identified by MS/MS analysis of the permethyl derivative ([M+Na]⁺ at *m/z* 1171) as comprising two structural isomers corresponding to endo-β-galactosidase cleavages at the respective two internal type 2 units. Thus, the first product, Hex-HexNAc-Hex-HexNAc-Hex, indicates that a type 2 chain can be extended by a type 1 unit (Fig. 4*A* (*J*)), which was resistant to digestion. More importantly, the second product, HexNAc-Hex-HexNAc-Lac, shows that a type 1 chain can be extended by a type 2 unit (Fig. 4*A* (*II*)). The internal type I unit is critically identified by the Z and C ion pair at *m/z* 690 and 504, which unlike other ions, cannot arise from other cleavages.

The resistant, non-digested product carrying internal type 1 unit was then isolated from the degraded products by nor-

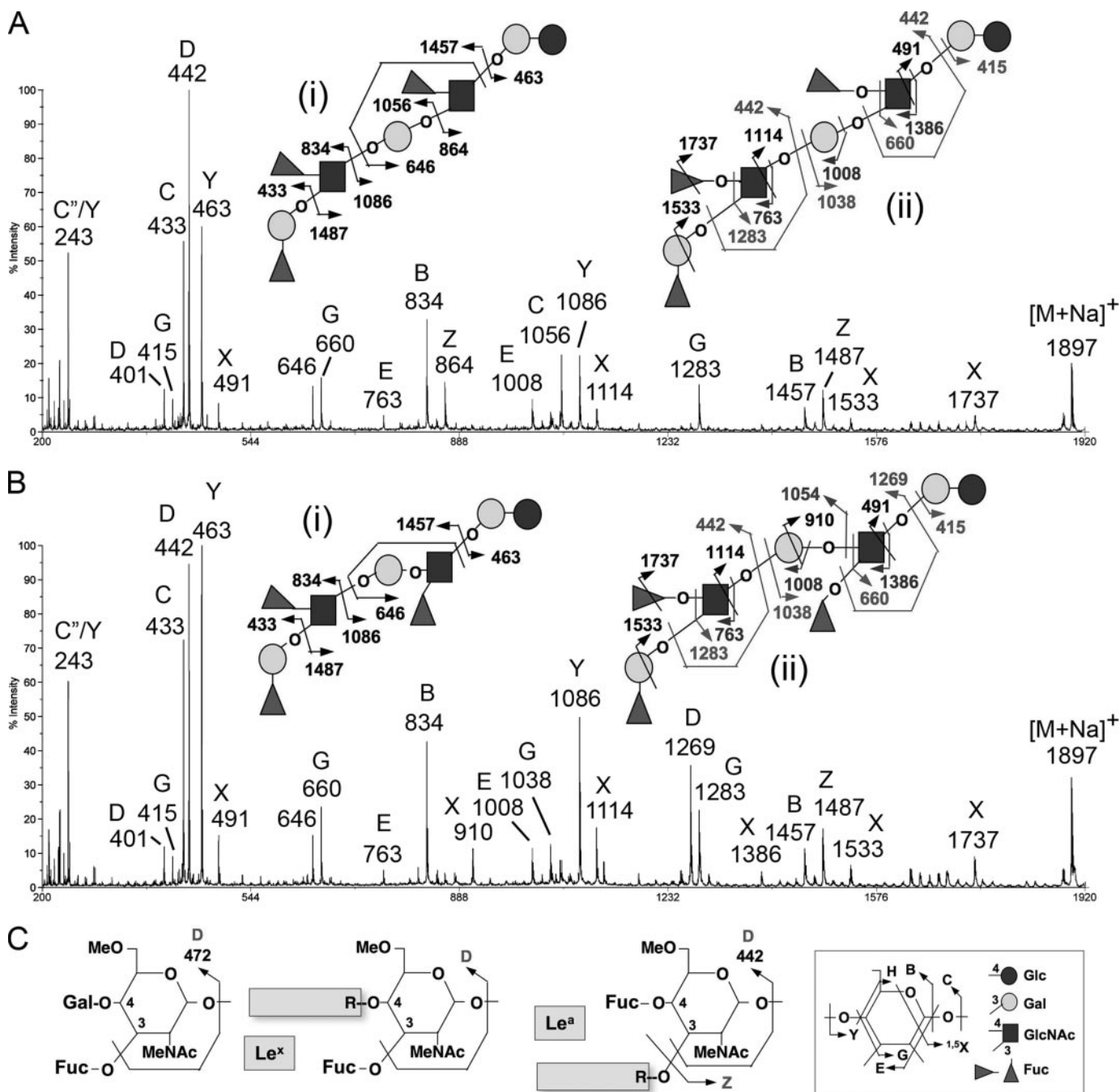


FIGURE 2. Characteristic high energy CID MS/MS fragmentation pattern of fucosylated types 1 and 2 glycan chains. MALDI-TOF/TOF MS/MS spectrum of permethylated $\text{Le}^b\text{-Le}^a\text{-Lac}$ isolated from Colo205 GSLs (A) is compared with that of an authentic $\text{Le}^b\text{-Le}^x\text{-Lac}$ standard (B). MALDI-Q/TOF analyses of the same samples afforded mainly the B, Y, C, and Z ions only (i), whereas MALDI-TOF/TOF analyses afforded additional cleavage ions (ii), as schematically illustrated on the respective figures. Symbol keys used and the nomenclature of the ions are shown in (C) together with schematic drawing illustrating the key D ions which allow discrimination of Le^a versus Le^x . The types of cleavage ions were annotated for each of the major fragment ions assigned, the origin of which can be referred to the schematic drawings (i) and (ii). For simplicity, $^{1,5}\text{X}$ ions were annotated as X ions without the superscript.

mal phase HPLC and incubated with $\alpha 3$ -fucosyltransferase from *H. pylori*. Isomers with a non-reducing terminal type 2 unit, which is not susceptible to endo- β -galactosidase, and/or internal type 2 unit that was incompletely digested would be $\alpha 3$ -fucosylated and, thus, shifted in molecular weight after the reaction. Indeed, major $[\text{M}+\text{Na}]^+$ molecular ion signals corresponding to mono- (m/z 1998) and difucosylated (m/z 2172) products were detected by MS analysis

of the permethyl derivatives (data not shown) along with the remaining unreacted $(\text{Hex-HexNAc})_3\text{-Lac}$ (m/z 1824), which is, thus, by definition, resistant to both endo- β -galactosidase and $\alpha 3$ -fucosyltransferase and corresponds to a structure with trimeric type 1 chain (Fig. 4B and the supplemental figure for a more detailed schematic elaboration of the overall workflow and rationale). This was further verified by high energy CID MALDI-MS/MS analysis, in comparison

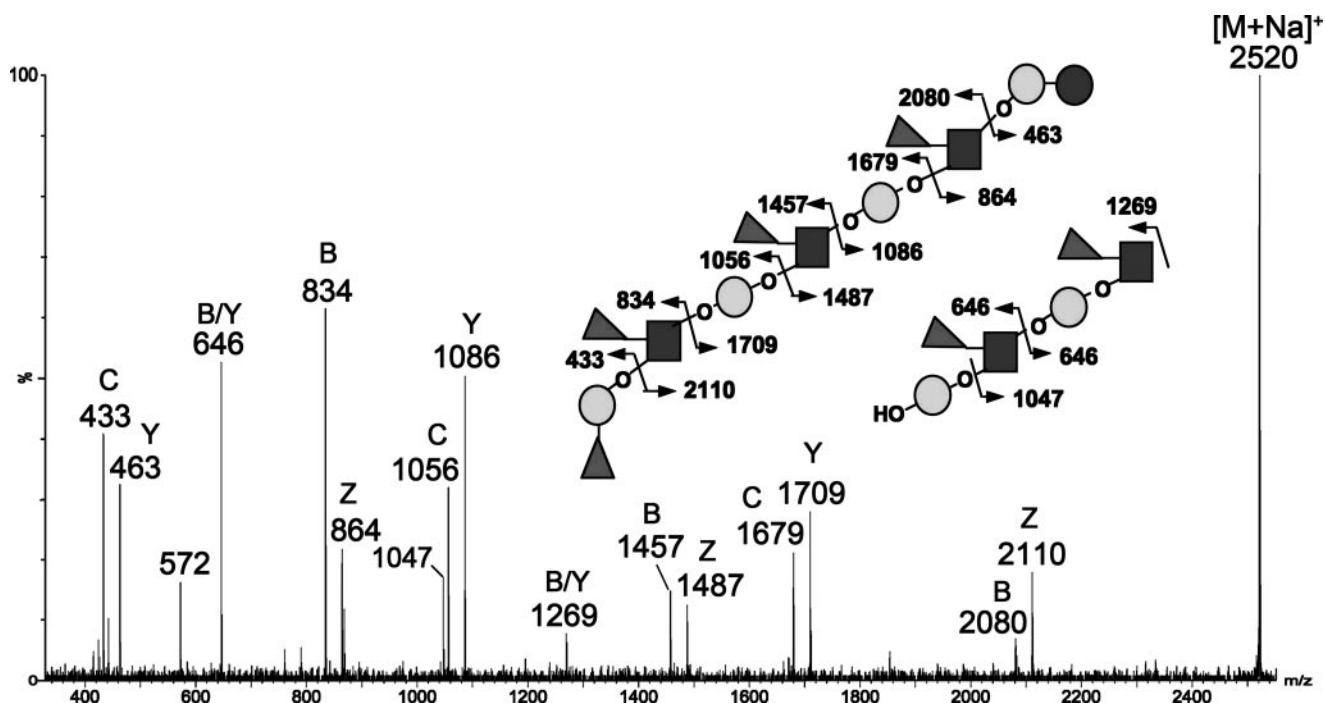


FIGURE 3. Identification of type 1 units in the fully fucosylated (Gal-GlcNAc)₃Lac derived from Colo205 GSLs. Assignments of the major fragment ions afforded by MALDI-Q/TOF MS/MS analysis of the permethylated glycan are schematically illustrated based on an extended linear type 1 chain structure to accommodate for the Z ion series. As noted under "Results", the data cannot resolve the co-existence of type 2 based Le^x units, which were in fact shown to be present by additional MALDI-TOF/TOF analysis (data not shown). Incomplete fucosylation introduced further heterogeneity, but a similar conclusion can be drawn with respect to the coexistence of both type 1 and 2 units.

with the original defucosylated sample not subjected to sequential enzymatic treatments, and an authentic trimeric type 1 linear chain was synthesized enzymatically *in vitro* (Fig. 5).

As expected, the $[M+Na]^+$ of the permethylated (Hex-HexNAc)₃-Lac (m/z 1824) from all three samples afforded a full complement of the ^{1,5}X and E ions at each residue together with the B and Y ions at HexNAc and the C and Z ions at Hex, which are fully supportive of a trimeric type 1 linear chain. More importantly, the D ions (m/z 921 and 1370) attributed to internal type 2 unit, the ^{3,5}A ion at m/z 329, and the H ions (m/z 1544, 1095, and 646) indicative of C4-linkage, which were present in the original samples (Fig. 5A), could no longer be detected in the treated sample (Fig. 5B) and standard (Fig. 5C). Furthermore, in comparison with the synthesized trimeric type 1 linear chain standard, the original and treated samples clearly gave two additional fragment ions at m/z 676 (D ion) and 574 (^{3,5}A ion), which are indicative of a co-existing branched structure, as schematically illustrated (Fig. 5(I)). Further digestion of the enzyme-treated product with β 1-3-galactosidase under conditions that preserved the enzyme specificity succeeded in complete removal of one and two terminal β 3-Gal residues, which gave more evidence for the linear and branched trimeric type 1 chain structures, respectively. MS/MS analysis of the branched structure afforded prominent fragment ions that are fully consistent with the deduced branching pattern, with two non-reducing terminal GlcNAcs located at the C3 and C6 of the 3,6-branched Gal of an internal type 1 unit (Fig. 5D (II) and supplemental Fig.).

Collectively, the data thus conclusively demonstrated that, in Colo205, the dimeric type 1 chain carried on its LacCer can be

further extended linearly or branched in a manner similar to polyLacNAc of type 2 chain. Although the dimeric structures were shown to be predominantly of extended type 1 chain, a significant proportion of the trimeric structures are based on a hybrid of type 1 and type 2 chains. Our MS-based analytical strategy additionally enabled an unambiguous identification of both linear and branched structures that consist exclusively of trimeric type 1 units among the isomeric mixtures. An accurate quantification of the relative abundance of each isomer detected is not possible by this approach. However, based on the normalized relative MS signal intensities of the recovered products, a rough approximation indicates that at least about 20% of the isolated trimeric structures was degradable by endo- β -galactosidase digestion, and another 20% of the resistant product can be further acted on by the α 3-fucosyltransferase (Fig. 4B).

In Vitro Synthesis of Branched Dimeric Type 1 and 2 Chains and Their Structure Characterization—The identification of branched structures on a dimeric type 1 chain implies that either the currently known I branching enzymes, β 6-N-acetylglucosaminyltransferases (IGnT, β 6GnT), which normally act on type 2 polyLacNAc chains, can act on one that is based on a hybrid or an extended type 1 chain or that a novel glycosyltransferase activity has yet to be identified. To this end, all combination of linear dimeric type 1 and 2 chains and their hybrids were enzymatically synthesized *in vitro* and used as acceptor substrates to assay for the branching activities of IGnTs (Fig. 6). Based on MS detection and MS/MS analyses of the permethyl derivatives of the reaction products, it was found that the human IGnT1, IGnT2, and IGnT3 could transfer one or more

Extension and Branching of Dimeric Type 1 Chain

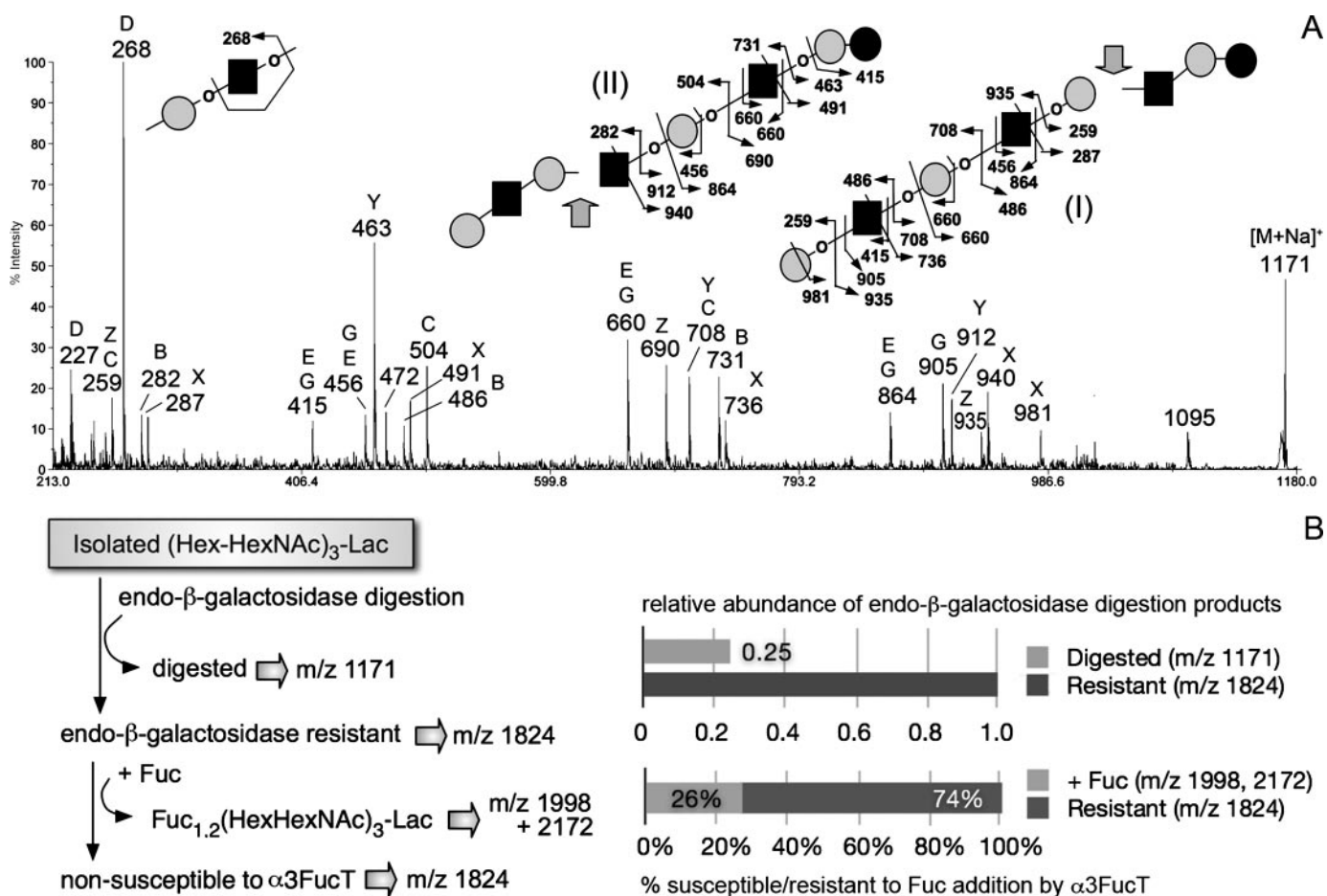


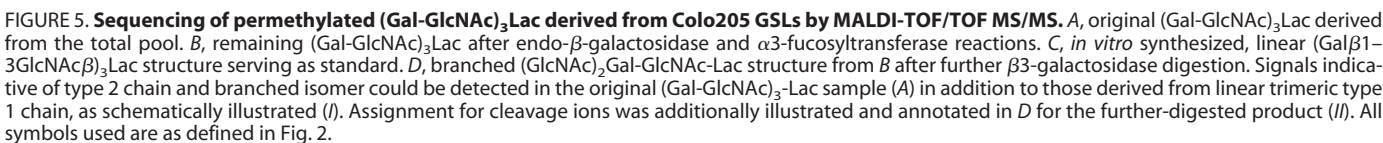
FIGURE 4. **Identification of internal and extended type 1 units in (Gal-GlcNAc)₃Lac derived from Colo205 GSLs.** A, MALDI-TOF/TOF MS/MS analysis of the major endo- β -galactosidase digested product, Hex₃HexNAc₂ ([M+Na]⁺ at m/z 1171). Assignments of fragment ions arising from either or both isomeric products (I) and (II) are schematically illustrated with all symbols used as defined in Fig. 2. The resistant internal type 1 unit is supported by the abundant D ion at m/z 268, although only in (II) can it be critically defined by the unique Z ion at m/z 690. For (I), the corresponding Z ion at m/z 486 can also be assigned as B ion. Ions that can be assigned to more than one origin are annotated as such. The signal at m/z 472 can be assigned as D ion of a terminal type 2 unit but can also arise from double cleavages (B/Y ion). The workflow from this endo- β -galactosidase digestion step to the next α 3-fucosylation step is schematically illustrated in B together with a plot of the relative abundance of the recovered products. Normalization was based on quantified peak areas for each of the isotopic signal cluster without correcting for differential response factors. Other products of endo- β -galactosidase digestion were not taken into consideration to arrive at an approximation of the relative amount of trimeric type 1 structure (see supplemental Fig.).

GlcNAc residues to all the synthetic substrates albeit with different efficiencies.

Strikingly, acceptors with a non-reducing terminal type 2 unit served as better acceptors than those with a non-reducing terminal type 1 unit. Under the same reaction conditions, the former were mostly converted to products with one to three GlcNAc residues added (Fig. 6, *A* and *C*, *m/z* 1620, 1865, and 2110), whereas a significant proportion of the latter remained unreacted (Fig. 6, *B* and *D*, *m/z* 1375). The relative susceptibility to a higher degree of GlcNAc branching as detected by MS monitoring provided a convenient indicator of reactivity from which it can be inferred that the type 2-type 2-Lac sequence served as the best acceptor substrate followed by type 2-type 1-Lac, type 1-type 2-Lac, and type 1-type 1-Lac in the order of decreasing reactivity (Fig. 6*E*). For a dimeric type 2 chain, the product with a single GlcNAc added on was shown by MS/MS analysis to be branched mostly at the Gal of the internal -3Gal-4GlcNAc unit as expected (data not shown). However, the Gal at the reducing end Lac unit and, to a lesser extent, the terminal

Gal at the non-reducing end, could also be acted on, thus, giving rise to a maximum of three GlcNAc residues added.

In contrast, branching site preference was less apparent for the single GlcNAc residue added onto an extended dimeric type 1 chain, MS/MS analysis of which afforded critical ions that are indicative of GlcNAc addition at all three possible Gal sites (Fig. 7). In particular, GlcNAc added to the Gal of internal type 1 unit (I) is established by the double cleavage ions at m/z 717, 944, and 490 in both the MS^2 and MS^3 spectra. Additional GlcNAc at the Gal of Lac (II) is defined by the presence of B and Y (or C and Z) ion pairs at m/z 935 and 708, respectively. GlcNAc addition at the non-reducing end Gal (III) is identified by the C_2 and B_3 ions at m/z 504 and 731, respectively. The relative low intensities of these ions in both MS^2 and MS^3 spectra and the lack of the corresponding Z ions (m/z 1139 and 699) indicate that this isomeric product is likely to be least abundant, consistent with it being the least preferred site of GlcNAc addition. Furthermore, none of these site-specific fragment ions indicative of GlcNAc addition at the non-reducing end Gal or Gal of Lac was observed in the corresponding products with single GlcNAc



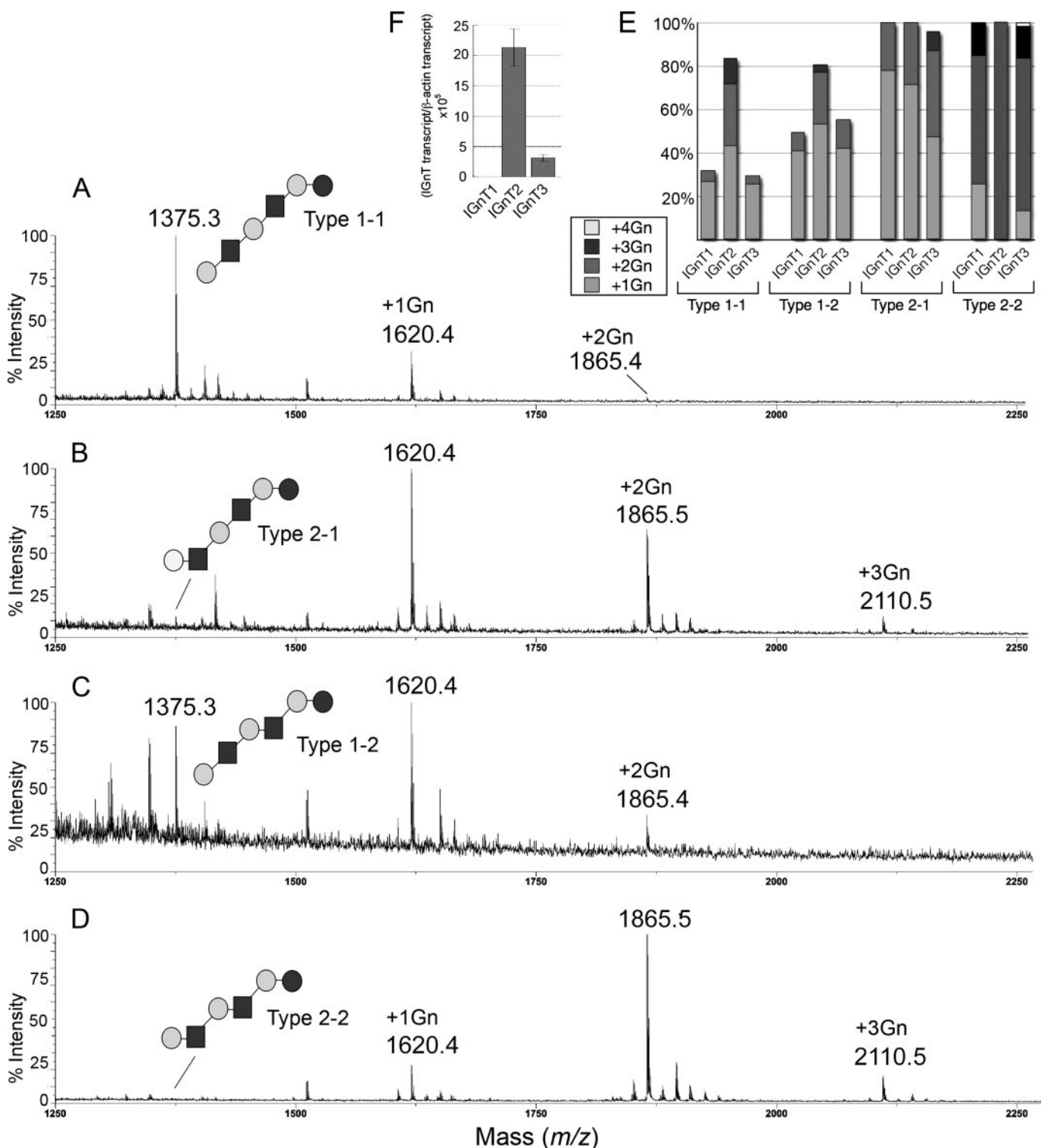


FIGURE 6. Branching activities of IGnT against (Gal-GlcNAc)₂Lac substrates *in vitro*. Equal amounts of the four different synthetic acceptors representing type 1-type 1 (A), type 1-type 2 (B), type 2-type 1 (C), and type 2-type 2 (D) chains were treated with the same amount of the branching enzymes (IGnT1–3) under the same experimental conditions. The reaction products were permethylated after sample clean-up and profiled by MALDI-MS. Only the dataset corresponding to the activities of IGnT3 is shown. Relative abundance of each of the products with one or more GlcNAc added (annotated as +*n*Gn) based on peak areas are normalized to generate a representation of the respective activity profile of IGnT1–3 against the four substrates in terms of % conversion of the initial substrate to the branched products (E). The relative abundance of each of the IGnT mRNA transcript/ β -actin transcript expressed in Colo205 cells, as detected by reverse transcription-PCR (IGnT1, not detected; IGnT2, $2.13 \times 10^{-4} \pm 3.04 \times 10^{-5}$; IGnT3, $3.16 \times 10^{-5} \pm 5.61 \times 10^{-6}$), is shown in F.

addition from acceptors where the non-reducing terminal unit is a type 2 chain. The results, therefore, support the notion that a type 1 chain at the non-reducing end of the Gal makes it a less

reactive acceptor substrate than that extended by a type 2 chain and, consequently, reduces the overall branching site preference.

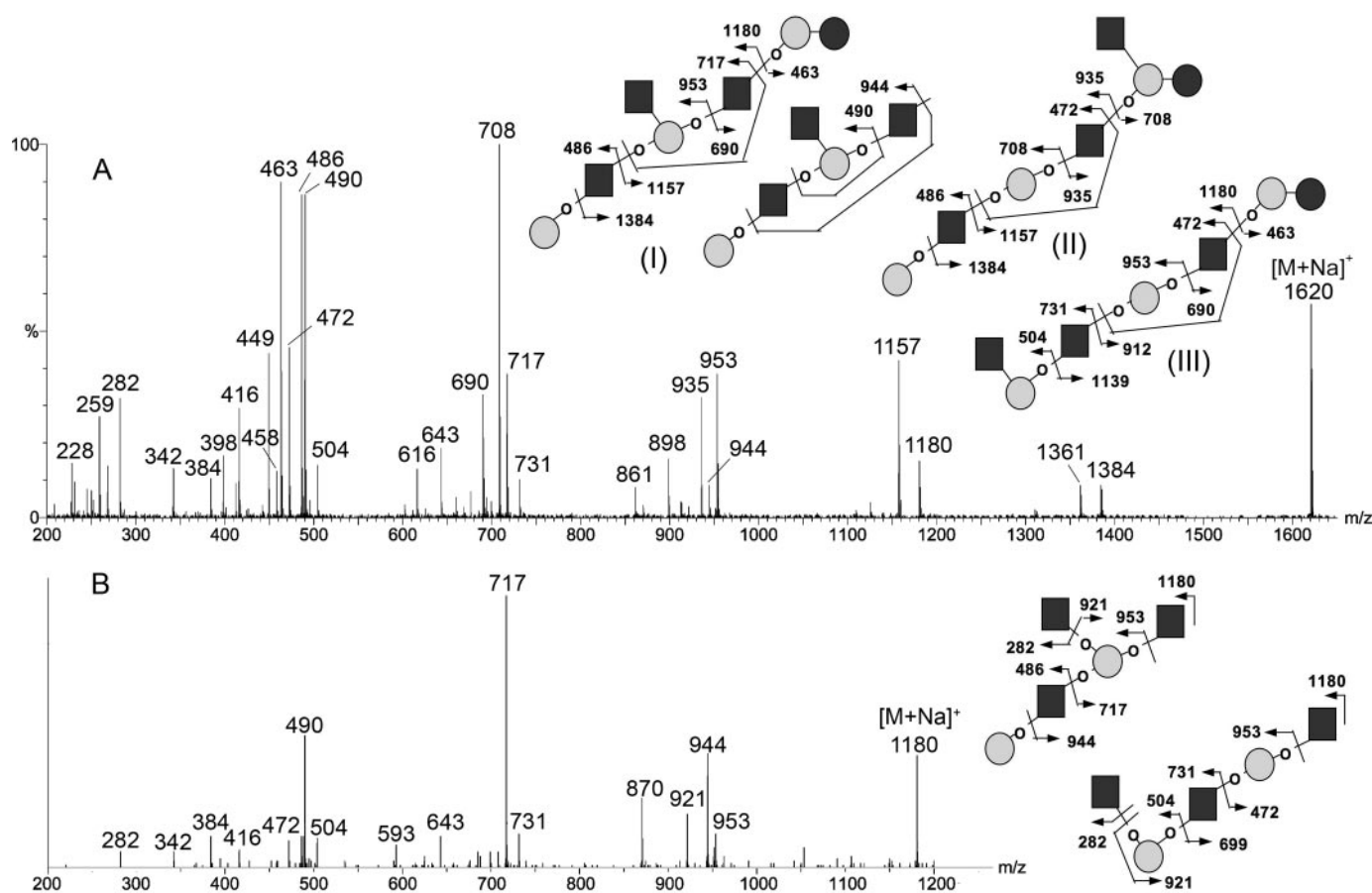


FIGURE 7. The branching sites of the extra GlcNAc added *in vitro* by IGnT onto the linear extended type 1 chain. The MALDI-QIT/TOF MS² fragment ion at *m/z* 1180 produced by the sodiated parent ion *m/z* 1620 of the permethyl derivatives (A) was further selected for MS³ analysis (B). A fragmentation pattern afforded by QIT/TOF was found to be similar to that of Q/TOF, namely dominated by B, Y, C, and Z ions, as illustrated schematically in (I), (II), and (III), corresponding, respectively, to GlcNAc addition on the three probable Gal sites. Signals at *m/z* 1361, 898, and 449 correspond to loss of terminal GlcNAc from the molecular and primary fragment ions at *m/z* 1620, 1157, and 708, respectively. Additional non-assigned ions correspond mostly to a further loss of 74 mass units from those assigned, the origin of which is not clear. Symbols used are as defined in Fig. 2.

All three human IGnTs (IGnT1–3) (16) demonstrated similar substrate and branching site preference, as can be concluded from the MS/MS analyses, with the IGnT2 enzyme preparation marginally more reactive against the type 1-type 1- and type 1-type 2-Lac substrates (Fig. 6E). True specific activities were, however, difficult to define given the different combinations of products which can only be distinguished by MS and MS/MS analyses. *In vivo*, only the transcripts of *IGnT2* and *IGnT3* can be detected by reverse transcription-PCR, with the former expressed at 6.7-fold higher (Fig. 6F) and, thus, most likely account for the observed branching activities. In a separate set of experiments, a mixture of the *in vitro* synthesized dimeric, trimeric, and tetrameric linear type 2 chains were found to mostly have 2, 3, and 4 GlcNAc residues added by IGnT2, respectively (Fig. 8A). In each case, MS/MS analysis of the major products localized the extra site of branching to Gal of the Lac unit (data not shown). Under the same reaction conditions, a majority of the corresponding dimeric and trimeric linear type 1 chains synthesized remained unreacted, with a small portion receiving only a single GlcNAc residue (Fig. 8B).

DISCUSSION

It has been observed repeatedly that malignant transformed cells and metastatic tumor cells often make and pres-

ent more polyLacNAc chains than the parent cells or low metastatic counterparts (25). Because a type 1 chain is not normally extended, most descriptions pertaining to the functioning of polyLacNAc chain, either linear or branched, are referred to type 2 chains. It should, however, be pointed out that whereas detection of polyLacNAc chains may not be too daunting, detailed structural delineation to ascertain the absence of co-existing or co-polymerized type 1 units is not an easy task. In the absence of data suggesting otherwise, the expression of extended type 1 chain appears to be rather restricted, which to a certain extent may simply be due to its being under-investigated systematically and specifically. This is not helped by lack of specific mAbs for a rapid screening of tissue expression. All reported anti-fucosylated extended type 1 chain mAbs currently available (8, 26) demonstrate a certain degree of cross-reactivities against fucosylated type 2 units.

An important consideration for the polyLacNAc chain is that it presents various fucosylated and/or sialylated antigens, not only at the non-reducing termini but also internal ones at high valency. In this context, a monomeric *versus* dimeric or oligomeric and a homotypic *versus* a hybrid of $\pm$ sialyl $\pm$ sulfo Le^x and Le^a on a linear *versus* branched backbone chains are unlikely to be functionally equivalent. Although the subtle differences may

Extension and Branching of Dimeric Type 1 Chain

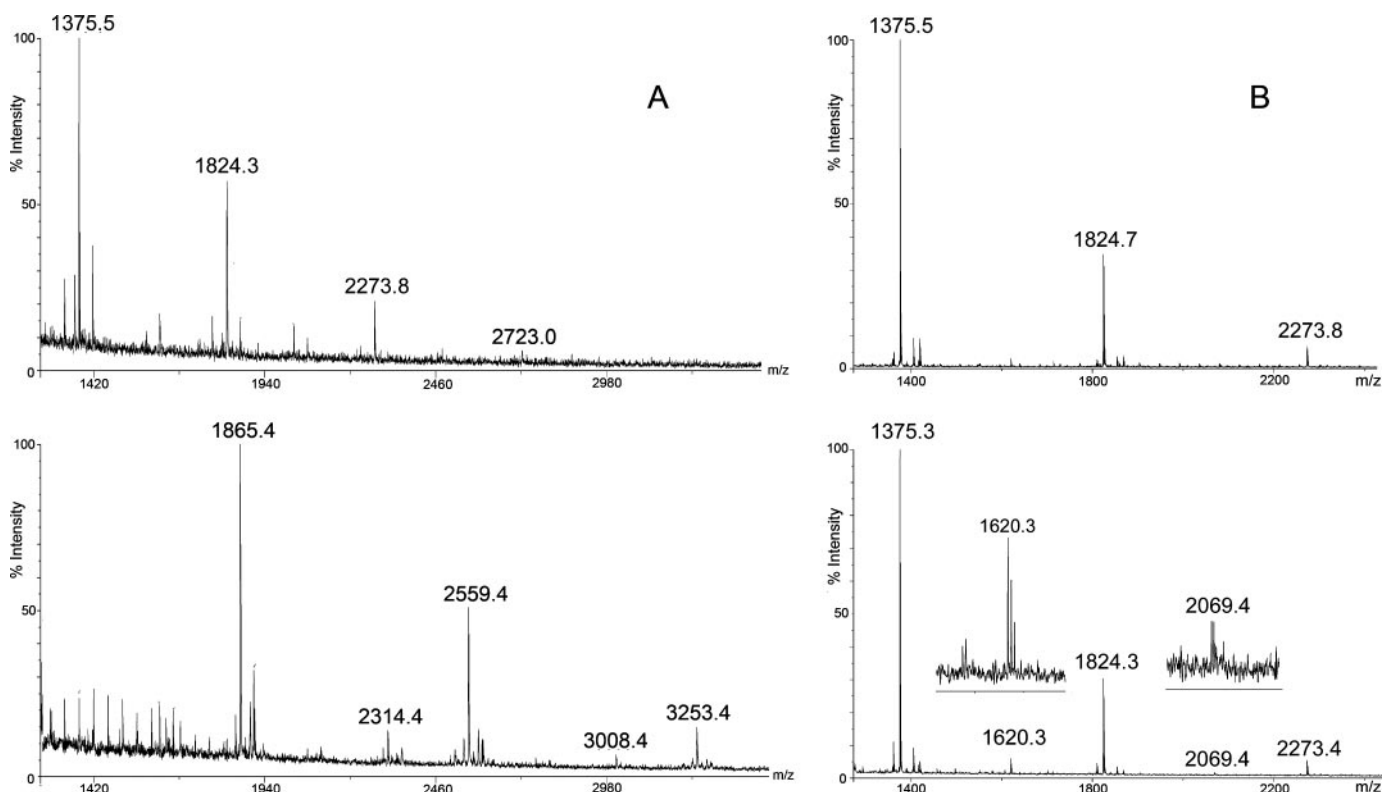


FIGURE 8. *In vitro* enzymatic extension and branching of extended type 2 and type 1 chains. Linear extension of type 2 (A) and type 1 (B) chains were performed by incubating the respective acceptors lacto-*N*-neotetraose and lacto-*N*-tetraose with β 3GnT2/ β 4GalT1 and β 3GnT2/ β 3GalT5, respectively, and aliquots of the resulting products were permethylated after clean-up for MALDI-MS analysis (upper panels). The major signals afforded correspond to $[M + Na]^+$ molecular ions of the starting material and products extended by 1 and 2 Gal-GlcNAc units at m/z 1375, 1824 and 2273, respectively. With lacto-*N*-neotetraose, a small amount of tetrameric type 2 units was also detected at m/z 2723. The synthesized products were then further treated with IGnT2 under the same conditions and similarly analyzed to assay for the respective branching activities (lower panel). For extended type 2 chains, the detected products correspond to GlcNAc₂(Gal-GlcNAc)₂-Lac (m/z 1865), GlcNAc₃(Gal-GlcNAc)₃-Lac (m/z 2314), GlcNAc₃(Gal-GlcNAc)₃-Lac (m/z 2559), GlcNAc₃(Gal-GlcNAc)₄-Lac (m/z 3008), and GlcNAc₄(Gal-GlcNAc)₄-Lac (m/z 3253). For extended type 1 chains, only small amounts of products corresponding to GlcNAc₁(Gal-GlcNAc)₂-Lac (m/z 1620) and GlcNAc₂(Gal-GlcNAc)₃-Lac (m/z 2069) were detected along with unreacted substrates.

not be apparent when assayed with simplified model systems, there are now indications that our natural immune system is capable of surveying this difference and translates it into proper immuno-reactive consequences. In one notable example, the mannan-binding protein, a serum lectin associated with innate immunity, was shown to recognize specifically tandem repeats of terminal Le^b and Le^a, and not Le^{x/y}, carried on multiantennary complex type *N*-glycans of SW116 cells, another human colorectal carcinoma (6). This mannan-binding protein and Le^{b/a}-dependent cell-mediated cytotoxicity has also been reported for other colonic carcinomas including DLD-1 and Colo205 (27), although the implicated ligand and its carriers were poorly characterized.

Technically, it is a challenging task for any glycomic analysis to rival the resolving power of biological systems in distinguishing the occurrence and positioning of type 1 and 2 chains, which underpins the need in developing the enabling MS methodologies required for definitive structural studies at high sensitivity. In this work we have capitalized on previously established fragmentation pattern afforded by both low and high energy CID MS/MS (14) to identify ions that are indicative of type 1 and 2 linkages. A plethora of fragment ions can normally be detected that can unambiguously establish the chain sequence and branching position. However, only a handful of these ions are truly linkage-specific. A fundamental problem relates to situa-

tions when the characteristic ion signals for both type 1 and 2 units can be detected for each of the Gal-GlcNAc unit along the chain. It is not possible to resolve if a strictly type 1 or type 2 chain exists among the many permutation of hybrids. Because the overall aim in the current work was to ascertain if an extended dimeric type 1 chain can be further extended or branched, we have relied on a strategy that removed the isomers with internal type 2 units. A strictly extended type 1 chain would then be positively identified by the respective type 1-specific ions (C and Z ions by Q/TOF and also TOF/TOF) in conjunction with not detecting the type 2-specific ions (D ions by TOF/TOF).

Using this analytical strategy, we provided convincing structural evidence for the occurrence of extended linear and branched trimeric type 1 chains as well as the hybrids. Although it is not unusual for a single type 1 unit to be found extending from type 2 LacNAc or polyLacNAc chains, giving rise to terminal Le^{a/b} or sialyl Le^a carried on a type 2 chain (5, 6, 28), the reverse is rather uncommon (7). Similar hybrid structures were also found among the LacCer with tetrameric GalGlcNAc units using similar analytical approach. In contrast to the presence of only a small quantity (5–10%) of type 2 chain in the more abundant dimeric species (8), our current data indicated that the proportion of the type 2 chain in those larger GSLs is significantly higher (up to 40% or more), most

of which constitutes hybrids of type 1 and 2 chains in either direction.

A priori, elevated expression of type 1 chain would necessitate enhanced or up-regulated activity of $\beta 3\text{GalT}$ because this is the enzyme that is responsible for making a type 1 unit by adding a β -galactosidase to the 3-position of a β -GlcNAc, in competition against the more ubiquitous $\beta 4\text{GalT}$. This has indeed been elegantly demonstrated by Holmes *et al.* for Colo205 (10). The $\beta 3\text{GalT}$ activity was detectable in only a few cell lines, such as Colo205 and SW403 (10, 11), whereas the $\beta 4\text{GalT}$ activity was found to be highly expressed in all tumor cell lines as well as normal colonic epithelial HCMC cells. More recently, Colo205 had further been shown to express high amount of $\beta 3\text{GalT5}$ transcript relative to $\beta 4\text{GalT1}$ (11). Thus, the synthesis of type 1 unit itself does not appear to be rate-limiting or responsible for the occurrence of extended type 1 chains in cell lines such as Colo205, which is endowed with abnormally high $\beta 3\text{GalT}$ activity. For a type 1 unit to be further extended, its terminal Gal needs to be equally susceptible to the addition of a β -GlcNAc by a $\beta 3\text{GlcNAcT}$ ($\beta 3\text{GnT}$) activity that would normally extend a type 2 chain. To date, there are no data investigating and distinguishing this scenario from one in which there is an elevated expression of a particular $\beta 3\text{GnT}$ which would preferentially elongate a type 1 unit.

The specificity and regulated expression of $\beta 3\text{GnTs}$ is a complicated issue beyond the scope of current studies. Of those investigated, $\beta 3\text{GnT2}$, -3, -4, and -5 were shown to be expressed in Colo205 (19), and $\beta 3\text{GnT2}$, $\beta 3\text{GnT3}$, and $\beta 3\text{GnT5}$ all showed a higher activity against type 2 chain on lacto-*N*-neotetraose than type I chain on lacto-*N*-tetraose (29). In this work we have additionally demonstrated that $\beta 3\text{GnT2}$ in the presence of $\beta 3\text{GalT5}$ can act in concert to elongate the type 1 chain acceptor substrate, lacto-*N*-tetraose, to a series of extended type 1 chains of varying lengths. This *in vitro* synthesis using recombinant enzymes did not resolve the issue of which endogenous $\beta 3\text{GnT(s)}$ is responsible for *in vivo* synthesis using the natural Lc_4Cer ($\text{Gal}\beta 1-3\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}\beta 1-1'\text{Cer}$) substrates. It simply demonstrated that if a type 1 unit could be readily extended to dimeric type 1 chain, it could be further oligomerized to give the trimeric extended type 1 chain identified in this study. Importantly, this provided a pure, authentic standard for us to validate our MS/MS fragmentation pattern as well as a substrate for FucT to make polyLe^a antigens for use in defining the mAb binding specificity.

The identification of a branched trimeric type 1 structure synthesized by a specific cell line further raised an interesting issue not investigated previously. Because all polyLacNAc branching has in the past been studied in the context of type 2 chains, it is currently not known if any of the three cloned human IGnTs (16, 18) or $\beta 6\text{GnTs}$ that can branch a type 2 chain can similarly branch an extended type 1 chain. Our studies based on *in vitro* synthesis employing the three recombinant IGnTs have shown the highest activity of IGnT2 against the dimeric extended type 1 substrate. IGnT1 appears to be least active or not active, whereas the activity of IGnT3 is somewhat intermediate. Both demonstrated a higher activity against the natural substrate, *i.e.* type 2 polyLacNAc. Interestingly, when considered in conjunction with results obtained on the type 2

and 1 hybrids, it would appear that a better substrate is one that carries a Gal site for branching which is extended at the non-reducing end by a type 2 and not a type 1 unit. The IGnTs are less discriminative against the aglycon of the Gal itself whether it is in a type 2 or type 1 unit. Thus, $\text{Gal}\beta 1-4\text{GlcNAc}\beta 1-3\text{Gal}\beta 1\text{-R}$, as in a dimeric type 2 or a type 2-type 1 hybrid chain, constitutes a better acceptor substrate than $\text{Gal}\beta 1-3\text{GlcNAc}\beta 1-3\text{Gal}\beta 1\text{-R}$, as in a dimeric type 1 or a type 1-type 2 hybrid chain.

An important conclusion from the work presented here is that IGnT2 can readily branch a dimeric extended type 1 chain substrate *in vitro*, adding a GlcNAc to the expected Gal. In fact, when the enzyme activity is rate-limiting under the experimental conditions employed, our MS/MS sequencing data on the products revealed that a GlcNAc could also be added to both the Gal of the reducing end Lac unit and the non-reducing end terminal Gal. However, this apparent lack of preference is only evident against a dimeric type 1 chain. Against a dimeric type 2 chain, a single GlcNAc is clearly preferentially added to the expected middle Gal site, *i.e.* $\text{Gal}\beta 1-4\text{GlcNAc}\beta 1-3\text{Gal}\beta 1\text{-R}$. Thus, it could be concluded that although the dimeric extended type 1 chain did not serve as an optimum substrate for the IGnTs, the occurrence of branched structures is supported by the expression of a significant level of IGnT2 mRNA transcript in Colo205. In other words, the expression of IGnT2 is capable of converting the natural type 2 polyLacNAc substrate and possibly the aberrant glycosylation products, extended type 1 or the hybrid chains, into the branched structures identified in this work.

In summary, our data indicate that once a yet to be identified oncogenic event perturbs or subverts the normal constraints imposed on extending the type 1 chain, it can be further elongated and subsequently branched using the existing arrays of βGalTs and βGnTs despite presenting a somewhat less than ideal substrates. The high activity of $\beta 3\text{GalT}$ itself relative to $\beta 4\text{GalT}$ is a requisite for enhanced level of type 1 unit and with it Le^a and sialyl Le^a commonly associated with colonic carcinoma cells (30, 31). However, it may or may not be the determining factor to tilt the balance in promoting the synthesis of extended type 1 chains, which is a relatively rare and unique feature of Colo205. The key to this aberrant glycosylation may additionally be dependent on the regulated activities of the $\beta 3\text{GnTs}$ expressed. The Colo205 and other colonic adenocarcinoma cell lines, therefore, provide a good model system to further investigate the relative specificities and activities of $\beta 3\text{GalTs}$, $\beta 4\text{GalTs}$, $\beta 3\text{GnTs}$, and $\beta 6\text{GnT}$, in the biosynthesis of extended linear and branched glycan chains.

Acknowledgments—MALDI-Q/TOF and TOF/TOF data were acquired at the National Core Facilities for Proteomics (NSC Grant 95-3112-B-001-014), located at the Institute of Biological Chemistry, Academia Sinica. We are grateful to GlycoNex Inc. for the generous gift of monoclonal antibodies used in HPTLC staining and Mei-Chun Yang (GlycoNex Inc. Taiwan) for technical support and advice in GSL sample preparation.

REFERENCES

1. Ujita, M., McAuliffe, J., Suzuki, M., Hindsgaul, O., Clausen, H., Fukuda, M. N., and Fukuda, M. (1999) *J. Biol. Chem.* **274**, 9296–9304
2. Ujita, M., Misra, A. K., McAuliffe, J., Hindsgaul, O., and Fukuda, M. (2000) *J. Biol. Chem.* **275**, 15868–15875
3. Fukuda, M., Fukuda, M. N., and Hakomori, S. (1979) *J. Biol. Chem.* **254**, 3700–3703
4. Miyake, M., Kohno, N., Nudelman, E. D., and Hakomori, S. (1989) *Cancer Res.* **49**, 5689–5695
5. Kannagi, R., Levery, S. B., and Hakomori, S. (1985) *J. Biol. Chem.* **260**, 6410–6415
6. Terada, M., Khoo, K. H., Inoue, R., Chen, C. I., Yamada, K., Sakaguchi, H., Kadowaki, N., Ma, B. Y., Oka, S., Kawasaki, T., and Kawasaki, N. (2005) *J. Biol. Chem.* **280**, 10897–10913
7. Karlsson, K. A., and Larson, G. (1981) *J. Biol. Chem.* **256**, 3512–3524
8. Stroud, M. R., Levery, S. B., Nudelman, E. D., Salyan, M. E., Towell, J. A., Roberts, C. E., Watanabe, M., and Hakomori, S. (1991) *J. Biol. Chem.* **266**, 8439–8446
9. Stroud, M. R., Levery, S. B., Salyan, M. E., Roberts, C. E., and Hakomori, S. (1992) *Eur. J. Biochem.* **203**, 577–586
10. Holmes, E. H. (1989) *Arch. Biochem. Biophys.* **270**, 630–646
11. Isshiki, S., Togayachi, A., Kudo, T., Nishihara, S., Watanabe, M., Kubota, T., Kitajima, M., Shiraishi, N., Sasaki, K., Andoh, T., and Narimatsu, H. (1999) *J. Biol. Chem.* **274**, 12499–12507
12. Holmes, E. H., Hakomori, S., and Ostrander, G. K. (1987) *J. Biol. Chem.* **262**, 15649–15658
13. Angstrom, J., Larsson, T., Hansson, G. C., Karlsson, K. A., and Henry, S. (2004) *Glycobiology* **14**, 1–12
14. Yu, S. Y., Wu, S. W., and Khoo, K. H. (2006) *Glycoconj. J.* **23**, 355–369
15. Lin, S. W., Yuan, T. M., Li, J. R., and Lin, C. H. (2006) *Biochemistry* **45**, 8108–8116
16. Inaba, N., Hiruma, T., Togayachi, A., Iwasaki, H., Wang, X. H., Furukawa, Y., Sumi, R., Kudo, T., Fujimura, K., Iwai, T., Gotoh, M., Nakamura, M., and Narimatsu, H. (2003) *Blood* **101**, 2870–2876
17. Togayachi, A., Sato, T., and Narimatsu, H. (2006) *Methods Enzymol.* **416**, 91–102
18. Yu, L. C., Twu, Y. C., Chou, M. L., Reid, M. E., Gray, A. R., Moulds, J. M., Chang, C. Y., and Lin, M. (2003) *Blood* **101**, 2081–2088
19. Shiraishi, N., Natsume, A., Togayachi, A., Endo, T., Akashima, T., Yamada, Y., Imai, N., Nakagawa, S., Koizumi, S., Sekine, S., Narimatsu, H., and Sasaki, K. (2001) *J. Biol. Chem.* **276**, 3498–3507
20. Dell, A., Reason, A. J., Khoo, K. H., Panico, M., McDowell, R. A., and Morris, H. R. (1994) *Methods Enzymol.* **230**, 108–132
21. Koy, C., Mikkat, S., Raptakis, E., Sutton, C., Resch, M., Tanaka, K., and Glocker, M. O. (2003) *Proteomics* **3**, 851–858
22. Kameyama, A., Kikuchi, N., Nakaya, S., Ito, H., Sato, T., Shikanai, T., Takahashi, Y., Takahashi, K., and Narimatsu, H. (2005) *Anal. Chem.* **77**, 4719–4725
23. Dell, A., Morris, H. R., Egge, H., von Nicolai, H., and Strecker, G. (1983) *Carbohydr. Res.* **115**, 41–52
24. Spina, E., Sturiale, L., Romeo, D., Impallomeni, G., Garozzo, D., Waidelich, D., and Glueckmann, M. (2004) *Rapid Commun. Mass Spectrom.* **18**, 392–398
25. Dennis, J. W., Granovsky, M., and Warren, C. E. (1999) *Biochim. Biophys. Acta* **1473**, 21–34
26. Ito, H., Tashiro, K., Stroud, M. R., Orntoft, T. F., Meldgaard, P., Singhal, A. K., and Hakomori, S. (1992) *Cancer Res.* **52**, 3739–3745
27. Muto, S., Sakuma, K., Taniguchi, A., and Matsumoto, K. (1999) *Biol. Pharm. Bull.* **22**, 347–352
28. Martensson, S., Due, C., Pahlsson, P., Nilsson, B., Eriksson, H., Zopf, D., Olsson, L., and Lundblad, A. (1988) *Cancer Res.* **48**, 2125–2131
29. Togayachi, A., Akashima, T., Ookubo, R., Kudo, T., Nishihara, S., Iwasaki, H., Natsume, A., Mio, H., Inokuchi, J., Irimura, T., Sasaki, K., and Narimatsu, H. (2001) *J. Biol. Chem.* **276**, 22032–22040
30. Itzkowitz, S. H., Yuan, M., Fukushi, Y., Lee, H., Shi, Z. R., Zurawski, V., Jr., Hakomori, S., and Kim, Y. S. (1988) *Cancer Res.* **48**, 3834–3842
31. Nakamori, S., Kameyama, M., Imaoka, S., Furukawa, H., Ishikawa, O., Sasaki, Y., Kabuto, T., Iwanaga, T., Matsushita, Y., and Irimura, T. (1993) *Cancer Res.* **53**, 3632–3637