MCP Papers in Press. Published on March 13, 2017 as Manuscript M117.067983

Structural diversity of human gastric mucin glycans

Chunsheng Jin[1], Diarmuid T Kenny[1], Emma C Skoog[1#], Medéa Padra[1], Barbara Adamczyk[1], Varvara Vitizeva[1], Anders Thorell[2], Vignesh Venkatakrishnan[1], Sara K. Lindén[1]*, and Niclas G. Karlsson[1]*

[1]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, University of Gothenburg, Box 440, Medicinaregatan 9A, 405 30 Gothenburg, Sweden

[2]Karolinska Institute, Department for Clinical Science and Department of Surgery, Ersta Hospital, Stockholm, Sweden

[#]Current address: Center for Comparative Medicine, University of California Davis, Davis, CA

*Corresponding authors: niclas.karlsson@medkem.gu.se (Tel: 46-31-7866528; Fax: 46-31-41608); sara.linden@gu.se (Tel: 46-31-7863057; Fax: 46-31-41608)

Abbreviations

CIM, clustered image map; deHex, deoxyhexose; ELISA, enzyme-linked immunosorbent assay; Hex, hexose; HexNAc, *N*-acetylhexosamine; LacdiNAc, *N*,*N*'-diacetyllactosamine; (oligo)-LacNAc, (oligo)-*N*-acetyllactosamine; LC-MS, liquid chromatography-mass spectrometry; $Le^{a/b/x/y}$, Lewis a/b/x/y antigen; MS/MS, tandem mass spectrometry; PLA, in situ proximity ligation assay; RT, retention time; $sLe^{a/x}$, sialyl Lewis a/x antigen; SNFG, the Symbol Nomenclature for Glycomics; Sul, sulfate

Summary

The mucin *O*-glycosylation of 10 individuals with and without gastric disease was examined in depth in order to generate a structural map of human gastric glycosylation. In the stomach, these mucins and their *O*-glycosylation protect the epithelial surface from the acidic gastric juice and provide the first point of interaction for pathogens such as *Helicobacter pylori,* reported to cause gastritis, gastric and duodenal ulcers and gastric cancer. The rational of the present study was to map the *O*-glycosylation that the pathogen may come in contact with. An enormous diversity in glycosylation was found, which varied both between individuals and within mucins from a single individual: mucin glycan chain length ranged from 2-13 residues, each individual carried 34-103 *O*-glycan structures and in total over 258 structures were identified. The majority of gastric *O*-glycans were neutral and fucosylated. Blood group I antigens, as well as terminal α1,4-GlcNAc-like and GalNAcβ1-4GlcNAc-like (LacdiNAc-like), were common modifications of human gastric *O*-glycans. Furthemore, each individual carried 1-14 glycan structures that were unique for that individual. The diversity and alterations in gastric *O*-glycosylation broaden our understanding of the human gastric *O*-glycome and its implications for gastric cancer research and emphasize that the high individual variation makes it difficult to identify gastric cancer specific structures. However, despite the low number of individuals, we could verify a higher level of sialylation and sulfation on gastric *O*-glycans from cancerous tissue than from healthy stomachs.

Introduction

Gastric cancer is the second most common cause of cancer-associated death and fourth most commonly diagnosed cancer worldwide (1). Annually, 0.7 million patients with gastric cancer die globally (2). The cancer is associated with glycosylation changes, but how alteration of gastric mucins relates to gastric cancer pathogenesis remains unknown. Despite the protection by mucins and the acidic gastric juice and proteolytic enzymes, the bacterium *Helicobacter pylori* manage to thrive in the gastric lining, infecting about half of the world's population (3). There is a direct correlation between infection and gastric cancer, where 0.1-3% of infected individuals develop gastric adenocarcinoma or mucosa-associated lymphoid tissue lymphoma and another 10-15% develop symptomatic gastritis or gastric and duodenal ulcers, whereas the majority show no symptoms (4).

In the stomach, MUC5AC and MUC6 are the major secreted mucins, while MUC1 is the dominant membrane-associated mucin. MUC5AC is produced by the surface epithelium, whereas MUC6 is secreted from the deep glands of the gastric mucosa (5, 6). Both MUC5AC and MUC6 are large oligomeric mucins that occur as distinct glycoforms (7). In gastric precancerous lesions and cancer, altered expression of MUC5AC, MUC6, MUC2 and MUC5B has been described, with MUC2 being a marker for intestinal metaplasia (8, 9). The gastric surface and foveolar epithelium are formed by a single layer of tall columnar mucin-producing cells that have a basal nucleus below an apical cup of mucin. These cells have a turnover rate of 3-6 days, but the mucus layer produced in these cells have an even shorter life span: the production rate from start of glycosylation until release at the apical side is about 6 hours (10), demonstrating that both the mucin repertoire and glycosylation theoretically can change rapidly. The

carbohydrate structures present on mucosal surfaces vary according to cell lineage, tissue location, and developmental stage (11). The massive *O*-glycosylation of the mucins protects them from proteolytic enzymes and induces a relatively extended conformation.

The dominating type of carbohydrate chains on mucins consist of extended oligosaccharides initiated with *N*-acetylgalactosamine (GalNAc) linked to a hydroxyl group on serine or threonine, elongated by the formation of the so-called core structures (core 1-8), and followed by the backbone region (type 1 or 2 chain). The chains are terminated by *e.g.* fucose (Fuc), galactose (Gal of type 1 or 2 chain), GalNAc or sialic acid (Neu5Ac) residues in the peripheral region, forming histo-blood group antigens such as A, B, and H, or Lewis type antigens such as Lewis a (Le^a), Le^b, Le^x, and Le^y, as well as sialyl-Le^a (sLe^a) and sLe^x structures. Immunohistochemical analysis has demonstrated that the Le^a and Le^b blood group antigens (Le type 1 structures) mainly appear in the surface epithelium, whereas the Le^x and Le^y antigens (Le type 2 structures) are expressed in mucous, chief and parietal cells of the glands (12). Thus, the Le type-1 structures co-localize with MUC5AC whereas Le type-2 structures co-localize with MUC6 (12), although this distribution is not always distinct (12-14). The carbohydrate structures present depend on the glycosyltransferases expressed in the cells, *i.e.* by the genotype of the individual. The terminal structures of mucin oligosaccharides are heterogeneous and vary between/within species and even with tissue location within a single individual (12, 15). Possibly, this structural diversity allows us to cope with diverse and rapidly changing pathogens, as reflected by the observation that susceptibility to specific pathogens differs between people with different histo-blood groups (16). Mucins appear to be the major carrier of aberrant glycosylation in carcinomas, and incomplete glycosylation, leading to expression of Tn and T antigens, and/or

sialylation/sulfation is common (15, 17). The sLe^x and sLe^a are frequently overexpressed in carcinomas, and expression of these antigens by epithelial carcinomas correlates with tumor progression, metastatic spread and poor prognosis (17).

Mucins from different individuals differ in their effect on *H. pylori* growth, adhesion and expression of virulence genes (18-20), and the Le^b and α1,4GlcNAc are two structural epitopes that have been shown to participate in regulation of *H. pylori* growth (21, 22). However, other, yet unknown, glycans may also affect *H. pylori*. The detailed characterization of *O*-glycosylation of a given tissue context is crucial for our understanding of its role during pathological and physiological conditions, such as *H. pylori* infection and gastric carcinogenesis. In addition, the alteration of *O*-glycosylation during gastric cancer progression, such as metastasis and cancer cell invasion, helps us to understand the control of *O*-glycosylation in gastric cancer. In this study, *O*-glycans from gastric adenocarcinoma tumors, normal mucosa of tumor-adjacent stomachs and normal mucosa are characterized. The diversity and alteration in gastric *O*-glycosylation broaden our understanding of the human gastric *O*-glycome and its implications for gastric cancer research.

Experimental procedures

Isolation of mucins

Gastric specimens were obtained after informed consent and approval of local ethics committees (Lund University Hospital, Lund, Sweden). Mucins were isolated from frozen gastric specimens as described previously (19). In brief, four of the specimens (P1T, P2T, P3T and P4T) were from gastric adenocarcinoma tumors (intestinal type) and another three (P5TA, P6TA and P7TA) were from macroscopically normal mucosa of tumor-adjacent stomachs (Table 1). Two of the tumors contained both soluble (S) and insoluble mucins (I, *e.g.* P1TS and P1TI, in which the insoluble MUC2 mucin was later solubilized by reduction and alkylation) whereas the insoluble fractions from the other tumors did not contain MUC2, MUC5B, MUC6 or MUC5AC (*i.e.* were considered negative for mucins). The specimens (approximately 1.5x1.5 cm) of normal mucosa isolated from tissues adjacent to gastric tumors (tumor-adjacent, TA) were separated into fundus (F) versus pyloric antrum (A), surface (S) versus gland material (G) according to tissue location, *e.g.* P5TA-AS and P6TA-FG. In addition, three specimens (P8H, P9H, and P10H) were from the junction between antrum and corpus of patients who underwent elective surgery for morbid obesity. Mucins were isolated by isopycnic density gradient centrifugation from these materials as previously described (23). Gradient fractions containing mucins were pooled together to obtain one sample for each gradient. The presence of MUC5AC, MUC6, MUC2 and MUC5B, as well as Le^b, sLe^a, sLe^x, and α1,4-GlcNAc, were evaluated in previous study (19).

In situ proximity ligation assay (PLA)

In situ proximity ligation assay (PLA) was performed with paraffin-embedded sections from human gastric tissues for the detection of proximity of blood group antigens (ABH) and MUC5AC. These samples were obtained after written informed consent (Ersta Diaktioni, Sweden) in conjunction with obesity surgery and they had a normal histology. The Duolink II kit (Olink Bioscience, Uppsala, Sweden) was used according to the manufacturer's instructions. The paraffin-embedded sections were dewaxed and rehydrated. Heat induced antigen retrieval was performed using 10 mM Tris, 1 mM EDTA and 0.05% Tween 20, pH 9.0. The sections were incubated with blocking solution (Olink Bioscience) for 1 hour at 37°C. Primary antibodies against blood type H (monoclonal mouse anti-human blood group H antigen, clone A70-A/A9, at a concentration of 2.5 μg/ml, ThermoFisher Scientific, Waltham, MA), A (monoclonal mouse anti-human blood group A, clone HE-193, dilution 1:80, ThermoFisher Scientific), B (monoclonal mouse anti-human blood group B, clone HEB-29, dilution 1:40, Abcam, Cambridge, UK), and MUC5AC (polyclonal rabbit anti-oligomeric mucus/gel-forming MUC5Ac N-term aa552-567, at a concentration of 5 μg/ml, antibodies-online GmbH, Aachen, Germany) were used and incubated at 4°C overnight. Antibodies conjugated with oligonucleotides were utilized to examine the proximity for 1 hour at 37°C (Olink Bioscience). Ligation and amplification were performed at 37°C for 30 min and 90 min, respectively. The cell nuclei were visualized by DAPI. Sections were examined under a Zeiss Imager Z1 Axio fluorescence microscope (Zeiss, Welwyn Garden City, UK). The proximity ligation resulted in bright red fluorescent dots. Images were acquired using a Zeiss Axio cam MRm and the AxioVision Rel 4.8 software.

Release of *O*-linked oligosaccharides for LC-MS

Isolated mucins were dot-blotted onto PVDF membranes (Immobillin P, Millipore), stained with direct blue 71 (Sigma-Adrich) and destained with a solution of 10% acetic acid in 40% ethanol. The *O*-glycans were released from PVDF membranes as described previously (24). Released *O*-glycans were analyzed by liquid-chromatography-mass spectrometry (LC-MS) using a 10 cm × 250 µm I.D. column, prepared in-house, containing 5 µm porous graphitized carbon (PGC) particles (Thermo Scientific, Waltham, MA). Glycans were eluted using a linear gradient from 0-40% acetonitrile in 10 mM ammonium bicarbonate over 40 min at a flow rate of approximately 10 µl/min. The eluted *O*-glycans were detected using an LTQ mass spectrometer (Thermo Scientific) in negative-ion mode with an electrospray voltage of 3.5 kV, capillary voltage of -33.0 V and capillary temperature of 300°C. Air was used as a sheath gas and mass ranges were defined depending on the specific structure to be analyzed. The data were processed using Xcalibur software (version 2.0.7, Thermo Scientific). Glycans were annotated from their MS/MS spectra manually and validated by available structures stored in UniCarb-DB database (2015-12 version) (25). The annotated structures were submitted to the UniCarb-DB database and they will be included in the next release.

For structural annotation, some assumptions were used in this study: monosaccharides in the reducing end were assumed as GalNAcol; GalNAc was used for HexNAc when identified in blood group A and LacdiNAc sequences, otherwise HexNAc was assumed to be GlcNAc; hexose was interpreted as Gal residues. The presence of core 1-4 has been reported in gastric tissue (15, 26, 27). In this study, reducing end with sequence of Hex-HexNAcol and retention time (RT) shorter than 8 min on PGC column was assumed be to core 1 disaccharide, Hex-(HexNAc-)HexNAcol as

core 2 trisaccharide, HexNAc-HexNAcol as core 3 and 5 disaccharides with core 3 having shorter RT on PGC column (28), and HexNAc-(HexNAc-)HexNAcol as core 4 trisaccharide. The discovery of core 5 structures (isomeric to core 3) were assumed to be only present as di- and tri-saccharides, and they were validated with RT compared to standards obtained from our previous studies (29, 30). *O*-glycans with linear cores (core 1, 3 and 5) were distinguished from branched cores (core 2 and 4) based on the presence of $[M - H]^-$-223 and $[M - H]^-$-$C_3H_8O_4$ (or 108) in MS/MS of structures with linear core (24, 28, 31).

Elongation was assumed to occur as *N*-acetyl-lactosamine units (Hex-HexNAc or Galβ1-4GlcNAcβ1-3). Terminal epitopes corresponding to blood group ABH, Lewis a/x, Lewis b/y and LacdiNAc were assumed based on the sequences detected in their MS/MS spectra (24, 28, 31). Terminal HexNAc was assumed to be αGlcNAc, since distal β1,3GlcNAc residues were usually capped with Gal residues as result of highly active galactosyltransferases. Validation of smaller structures (<7 residues) was made by RT comparison with standards (29, 30) and/or MS/MS spectral matching using Unicarb-DB database (25). Larger structures were identified by de-novo sequencing of MS/MS spectra, epitope specific fragmentation and biosynthetic pathways (core type and blood group ABH).

Proposed structures are depicted using the Symbol Nomenclature for Glycomics (SNFG) (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Data analysis

To identify the most closely related structural features (epitopes), we generated clustered image map (CIM) by using online software CIMminer available at

http://discover.nci.nih.gov/tools.jsp. Cluster analysis groups samples and glycan features with shared similar % abundance into trees whose branch lengths reflect the degree of similarity between the objects (34). Relative percentages of glycan in individual sample were used to represent the amount of each glyco feature or epitope (Lewis and ABO type) or modification (sialylation, fucosylation and sulfation). The relative amounts of the different *O*-glycans were given in percentage (%) of the total sum of integrated peak areas in the LC-MS chromatograms (Supplementary table). Due to high variability between samples, the rows were not clustered and we kept the order according to the subject groups (gastric adenocarcinoma, adjacent normal mucosa, and normal). The Manhattan distance algorithm was selected for distance measurements and average linkage was selected for clustering, which defined the distance as the average of all pairs from each cluster group. Color-coded CIM that determines the distance and linkage between clustered columns (calculated glycan structural features) is represented in Figure 8D.

Statistical analysis was performed using the GraphPad Prism 6.0 software package (GraphPad Software Inc., San Diego, CA). Results were expressed as mean±SD. The statistical differences were calculated by the two-tailed Student's t-test.

Results

The purpose of this study was to address the heterogeneity of mucin glycans present in the stomach. One question is if the blood group and secretor status of a patient (expressed on for instance MUC5AC, Figure 1) are the main factors contributing to the inter-individual variation of the gastric mucin glycans. Alternatively, there could be other differences that dominate differentiation of inter- and intra-individual mucin subpopulations. In order to address this, mucins from three types of tissues were included in this study: normal (P8H, P9H, and P10H), normal mucosa of tumor-adjacent tissue (P5TA, P6TA, and P7TA) and gastric adenocarcinoma tumor (P1T, P2T, P3T and P4T, Table 1). In total, 17 mucin samples were obtained from these 10 individuals according to the solubility (soluble/insoluble) and location (surface/gland) (Table 1). The presence of mucins (MUC5AC, MUC6, MUC2 and MUC5B), Lewis antigen (Le^b and sialyl $Le^{a/x}$) and α1,4-GlcNAc were evaluated by ELISA in our previous study (19). After reductive β-elimination, at least 258 *O*-glycans were identified by LC-MS/MS using on-line graphitized carbon column in negative-ion mode (Supplementary Table). As an overview, we found that human gastric *O*-glycans in most samples were dominated by neutral and fucosylated structures in all three tissue types (normal, tumor adjacent and tumor), although gastric *O*-glycans from cancerous tissue contained higher level of sialylation and sulfation than healthy tissue. The diversity of *O*-glycans were also reflecting the presence of blood type ABH epitopes and i/I-branches, as well as terminal α1,4-GlcNAc-like and GalNAcβ1-4GlcNAc (LacdiNAc)-like structures.

Core types of human gastric *O*-glycans

In the analyzed samples, structures with monosaccharide sequences corresponding to core 2 *O*-glycans were found to be the dominating core of the gastric *O*-glycans, followed by sequences interpreted as core 1, 3/4 and 5 (Figure 2, Table 1 and Supplementary table).

Core 1: 25 out of 258 (10% of the detected structures and 18.7±6.6% of the relative abundance) were identified as core 1 like. In addition to the ubiquitous structures interpreted as mono- and di-siaylated core 1, the vast majority of core 1 glycans were extended with what appeared to be oligo-*N*-acetyllactosamine (oligo-LacNAc) with or without terminal blood group like antigens (ABH). In addition, two extended core 1 like glycans (795-2 and 1203-3 in Supplementary table) were terminated with sequences interpreted as $Le^{a/x}$ and $Le^{b/y}$, respectively; three (587-2, 878-1, 952-3 in Supplementary table) were terminated HexNAc-indicative of α1,4-GlcNAc (*e.g.*, Figure 2B). MS/MS of extended core 1 *O*-glycans usually contained $^{2,4}A$ ions of 1,3-linked Hex (indicative of β1,3-linked Gal; *m/z* 789, $^{2,4}A_5$ in Figure 2A), B-ions from extended residues (*m/z* 729, B_4 in Figure 2A; *m/z* 567, B_3 in Figure 3B) and typical $^{0,2}A_{GlcNAc}$ and/or $^{0,2}A_{GlcNAc}$-H_2O fragment ions (*e.g.*, *m/z* 628, $^{0,2}A_4$-H_2O, Figure 2A; *m/z* 466, $^{0,2}A_4$-H_2O, Figure 3B). In comparison with branched core structures, linear core structures (core 1, 3 and 5) have typical diagnostic ions ($[M - H]^-$-223 and $[M - H]^-$- 108) in their MS/MS (24, 28, 31). As shown in MS/MS of two isomeric structures (Figure 2A-B), fragmentation ions at *m/z* 1006 ($[M - H]^-$-108) were used to assign a linear core (Figure 2A), which were absent in MS/MS of structures with branched core (*e.g.* Figure 2B).

Core 2: 206 out of 258 characterized *O*-glycans (80% of the detected structures and 76.0±10.2% of the relative abundance) contained sequences corresponding to core 2. Neutral core 2 like *O*-glycans usually showed dominant Z_{1i} ions in their MS/MS spectra (*e.g.*, *m/z* 934 in Figure 2B; *m/z* 877 and 1226 in Figure 4A-C) indicating the size of both C6 and C3 branch of core 2. In addition, the $^{4}A_{0}$ fragmentation (35) of the reducing end HexNAc alditol moiety (GalNAcol; *e.g.*, *m/z* 789, $^{4}A_{0\alpha}$ in Figure 2B; *m/z* 1081, $^{4}A_{0\alpha}$ in Figure 4C) and typical $^{0,2}A_{GlcNAc}$/$^{0,2}A_{GlcNAc}$-H_2O/$^{2,4}A_{GlcNAc}$ cross-ring cleavage on the C6 branch (*e.g.*, *m/z* 628, $^{0,2}A_{4\alpha}$-H_2O in Figure 2B; multiple A-ions in Figure 4A-C) were also key fragmentation ion in structural assignment of core 2 like *O*-glycans.

Core 4: Eleven out of 258 characterized glycans (4% of the detected structures and 0.2±0.4% of the relative abundance) were interpreted as core 4 *O*-glycans. Most core 4 like *O*-glycans were only extended with LacNAc like elongation (Hex-HexNAc) on one branch, while the other branch was terminated with blood group like antigens (ABH) or by sialylation. Similar to the core 2 like *O*-glycans, MS/MS spectra of typical core 4 like *O*-glycans dominated with Z_{1i} ions (*m/z 715*, Figure 2C). The $^{4}A_{0}$ fragmentation of the reducing end GalNAc alditol indicated the size of the C6 branch (*e.g.*, *m/z* 570, $^{4}A_{0\alpha}$ in Figure 2C).

Core 3/5: Both core 3 and core 5-like [13 (5.0%) and 2 (0.8%) out of 258 characterized glycans;2.2±2.5% and 0.2±0.4%, respectively] *O*-glycans were detected on the human gastric mucins. The presence of core 5 in addition to core 3 was based on that there were two isomers with composition of $HexNAc_2$ ($[M - H]^-$ ions of *m/z* 425) and $Neu5Ac_1HexNAc_2$ ($[M - H]^-$ ions of *m/z* 716), showing almost identical MS/MS spectra (*e.g.*, MS/MS spectra of ions of *m/z* 716 in Figure 2D). As shown in Figure 2D, the fragment ions at *m/z* 495 ($Z_{1\beta}$ or $^{0,2}X_{2\alpha}$) and 425 ($Y_{1\alpha}$)

indicated that Neu5Ac was linked to C6 of which was assumed to be a reducing end GalNAc alditol residue. Based on our previous study (28), the structure with short retention time (RT at 14.18 min) was assigned as sialylated core 3 like structure; while the one with longer RT at 17.03 min was assigned as sialylated core 5 (Figure 2D). The remaining 11 gastric *O*-glycans were most likely derived from extension of core 3 *O*-glycans although the possibility that they were also core 5 *O*-glycans cannot be ruled out.

Interpretation of i/I-branch like structures on human gastric *O*-glycan

Analysis of complex extensions of human gastric *O*-glycans showed that 36 out of 258 (14% of the detected structures) carried what was interpreted as I-branches, Galβ1-4GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ-. These were found to be attached to the C3 branches of core 2 like or extended core 1 like *O*-glycans (Figure 3A and Supplementary Table). These branches were further modified by ABH and/or Lewis blood group like antigens as well as sialylation and/or sulfation.

I-branched like *O*-glycans could also be distinguished from their linear i-like isomers in MS/MS spectra. As shown in Figure 3A, the fragmentation ions at *m/z* 975 suggested consecutive loss of three Hex residues from the parent ions. The discontinuous B-ions (B-ions at *m/z* 364 and 891) suggested that those ions were derived from a branched structure rather than a linear one. In addition, only cross-ring fragment ions from terminal LacNAc were observed (*e.g.*, A-ions at *m/z* 221, 263, 281 and 424 in Figure 3A) indicative of single LacNAc (branched) rather than linear oligo-LacNAc. Taken together, the annotated structure contained three terminal LacNAc like sequences with two of them forming the I-branch on the C3 branch of a core 2 like *O*-glycan.

Combining with other I-branch like containing *O*-glycans, the typical MS/MS spectra of I-branch containing *O*-glycans usually contained B-ions with the intact I-branch (*e.g.*, $B_{3\alpha}$ ions at *m/z* 891 in Figure 3A), A-ions from both C6 and C3 branch, and a lack of the continuous A- and B-ions present in the structures containing extended oligo-LacNAc like structures (*e.g.*, B_2, B_3 and B_4 ions at *m/z* 364, 567, and 729 in Figure 3B).

Blood group like antigens on human gastric *O*-glycans

In total, 182 out of 258 *O*-glycans (71% of detected structures) were fucosylated, forming ABH and Lewis type histo-blood group like epitopes (Figure 4-5 and Supplementary table).

Blood group A, B, and H like epitopes) were determined by their diagnostic fragmentation ions in MS/MS spectra. As shown in Figure 4, core 2 like *O*-glycans were found to be the main carrier of all blood group ABH like epitopes. The dominant Z_i ions at *m/z* 877 (Figure 4A-B) and 1266 (Figure 4C) suggested the presence of blood group H (Figure 4A and 5C) or A (Figure 4B) on the C3 branch of these structures. The fragmentation $^{0,2}A_{3\alpha}$-H_2O ions at *m/z* 571 (Figure 4A-B) and 920 (Figure 4C) indicated the C6 branch was extended with type 2 like chain. According to the $^{2,4}A_{3\alpha}$ ions at *m/z* 529 (Figure 4A-B), they concluded the presence of a blood group B like epitope on the C6 branch of core 2 like *O*-glycan. Despite lacking of $Z_{3\alpha}/Z_{3\beta}$ and $Z_{3\alpha}/Z_{3\beta}$-CH_2O for blood group B (24), fragmentation ions at *m/z* 503 ($Z_{3\alpha}/Z_{3\beta}/Z_{1\gamma}$-$CH_2O$) caused by triple cleavage were observed, further confirming the presence of terminal blood group B like epitope (Figure 4A-B). As for the blood group H type 2 like epitopes, they were characterized by $^{0,2}A_{Gal}$-H_2O like fragments at *m/z* 409 (24). As shown in Figure 4C, a serial of cross-ring cleavages (A-ions at *m/z*

409/427, 570, 878, 920/938 and 1081) suggested all three fucose residues were linked to different Gal residues of α1,2-linkage.

In total, 77 out of 258 (30%) gastric *O*-glycans carried Lewis-like epitopes. Because *O*-glycans differing only in terminal Lewis epitope showed almost identical MS/MS spectra, we could not distinguish Le^a/Le^x from Le^b/Le^y in this study. The diagnostic ions of the Lewis-like epitopes were used to assign *O*-glycans from the gastric mucins (24). As shown in Figure 5, low molecular mass *O*-glycan structures showed Z_i/Z_i-CH_2O fragmentation (*m/z* 667 in Figure 5A-B), which were derived from double cleavages of what was interpreted as a C3 and C4 substituted GlcNAc in Lewis epitope configuration. Together with dominant Z ions at *m/z* 879 and 861 (Figure 5A-B), these two structures were assigned one with $Le^{a/x}$ and one with $Le^{b/y}$ on a core 2 like *O*-glycan C6 branch, respectively. For high-molecular-mass structures, the Z_i/Z_i-CH_2O fragmentation ions became less dominant (Figure 5C). However, other glycosidic and cross-ring cleavages provided enough structural information to annotate the structure. The dominant $Z_{1\gamma}$ ion at *m/z* 1064 together with $^4A_{0\alpha}$ ions at *m/z* 919 suggested a blood group A $Le^{b/y}$ like epitope on the C6 branch of core 2 *O*-glycan (Figure 5C). Up to four fucose residues were found in one structure (1698 in Supplementary table), which have both $Le^{a/x}$ and $Le^{b/y}$ like epitopes.

LacdiNAc-like and α1,4-GlcNAc-like epitopes on human gastric *O*-glycan

Among 258 *O*-glycans, 44 *O*-glycans (17% of the detected structures and 9.3±8.3% of the intensity) were assumed to be α1,4-GlcNAc–like structures, since distal β1,3GlcNAc residues were usually capped with Gal residues as result of highly active galactosyltransferases. Some of these structures were previously identified in porcine stomach (36, 37). Most of the α1,4-

GlcNAc-like structures in human tissue (91%) where found on core 2 like *O*-glycans (Figure 6 and Supplementary table). The terminal α1,4-GlcNAc-like structures were found on either of the branches of core 2 *O*-glycans (Figure 6B-C). In one case, α1,4-GlcNAc-like epitopes terminated both the branches (Figure 6 B). The dominant $Z_{1\beta}$ ions at *m/z* 772 and $^{4}A_{0\alpha}$ ions at *m/z* 627 suggested that this core 2 like *O*-glycan carried terminal HexNAc residues on both branches (Figure 6B). Cross-ring cleavages ions at *m/z* 466 ($^{0,2}A_{3\alpha}$-H_2O, Figure 6B) and *m/z* 424 ($^{2,4}A_{3\alpha}$, Figure 6B) concluded that a HexNAc-Gal-moiety was linked to C4 of a HexNAc (GlcNAc), *i.e.* a type 2 chain terminating with α1,4-GlcNAc-like epitope on the C-6 branch of the reducing end GalNAc moiety (Figure 4C), while the C-3 branch consisted only of a GlcNAcα1-4Galβ1-like disaccharide. *O*-glycans terminating with α1,4-GlcNAc-like structures were also detected on core 1 and 4 like structures. As shown in Figure 3B, an extended core 1 like glycan was found to be capped with an α1,4-GlcNAc-like epitope ($^{0,2}A_2$-H_2O ions at *m/z* 304).

The presence of LacdiNAc like structures on human gastric *O*-glycans was first described in our previous study based on *O*-glycan profiles from P10H and P5TA-AS (38). In this study, samples from additional 8 individuals were included. 9 (3% of the detected structures) *O*-glycans were detected to carry terminal HexNAc1-4HexNAc sequences interpreted as LacdiNAc-like epitopes including eight as core 2 like and one with core 3 like backbone. The presence of LacdiNAc-like epitopes was determined by the cross-ring cleavage of GlcNAc residues in a LacdiNAc motif (Figure 6A and 6C). The dominant $Y_{1\beta}/Z_{1\beta}$ ions at *m/z* 628/610 suggested the presence of HexNAc-HexNAc on the C6 branch of a core 2 *O*-glycan. The presence of cross-ring cleavages at *m/z* 322, 304 and 262 ($^{0,2}A_2$, $^{0,2}A_2$-H_2O, and $^{2,4}A_2$, respectively) indicated that the terminal HexNAc links to C4 of a HexNAc (GlcNAc). Taken together, we concluded that these glycans

contained terminal LacdiNAc-like structures. Some of these were previously identified as indeed carrying this epitope (38).

Acidic *O*-glycan on human gastric mucins

About one-fifth of *O*-glycans were sialylated (53 out of 258 detected structures). A sequence corresponding to sialyl Tn (Figure 7A), a tumor marker, was detected in all types of sample in varying amounts. Most sialylation (75%) were modifications on core 2 like structures. Unlike neutral core 2 like *O*-glycans, MS/MS spectra of sialylated core 2 like glycans were not dominated by Z_i ions. Instead, MS/MS spectra of sialylated core 2 like structures were dominated by Y_i ions without terminal Neu5Ac in singly charged ($[M - H]^-$) MS/MS spectra. As shown in Figure 7B, MS/MS spectra were dominated by two Y ions (*m/z* 1203 and 1041) indicating loss one Neu5Ac and $Neu5Ac_1Hex_1$ from the parent ions. The presence of $Z_{2\alpha}/Z_{2\beta}$ and $Z_{2\alpha}/Z_{2\beta}$-CH_2O ions at *m/z* 859 and 829 suggested this structure contained a terminal sialyl $Lewis^{a/x}$ ($sLe^{a/x}$) like epitope. Together with ions at *m/z* 692 and 674 ($Y_{1\alpha}/Z_{1\alpha}$), which indicated a blood group B like epitope linked to a GalNAc aditol, this structure was assigned as a core 2 like *O*-glycan consisting of $sLe^{a/x}$ on the C6 branch and blood group B like epitope on the C3 branch (Figure 7B). While singly charged MS/MS spectra only contained limited information about the location of NeuAc in the structure, the doubly charged $[M - 2H]^{2-}$ MS/MS spectra of sialylated core 2, however, contained more B_i/C_i ions consisting of terminal Neu5Ac. One example was shown in Figure 7C. The dominant ions at *m/z* 981 ($C_{4\alpha}$) suggested loss of terminal $Neu5Ac_1Hex_2HexNAc_1deHex_1$. Together with C ions at *m/z* 819 and 470 ($C_{3\alpha}$ and $C_{2\alpha}$) and ions at *m/z* 1032 ($Z_{3\alpha}/Z_{3\beta}$-CH_2O), it suggested it was a terminal $sLe^{a/x}$ like epitope. The fragment ions at *m/z* 1517 ($Z_{2\gamma}$) and 1186 ($Z_{1\gamma}$) suggested this structure also had a terminal $Le^{a/x}$ like epitope on

the other branch. Taken together, this structure was annotated as a core 2 *O*-glycan with one $Le^{a/x}$ on the C6 branch and one $sLe^{a/x}$ on the C3 branch (Figure 7C).

Approximately 10% of total *O*-glycans were mono-sulfated (29 out of 258 detected structures, and 4.9±9.5% of the relative abundance) with eight of them also sialylated, but the relative amounts were low (Supplementary table). Most sulfated glycans were found on core 2 like structure (25 out of 29), but sulfation was also found on core 1, 3 and 4-like structures. Three structures were detected to contain sulfo-(s)$Le^{a/x}$ like moieties (975-2, 1412, and 1557 in Supplementary table). One interesting sulfated *O*-glycan was shown in Figure 7D. The fragment ions at *m/z* 485 and 282 ($B_{2\alpha}$ and $Y_{2\alpha}/B_{2\alpha}$) indicated a terminal $HexNAc_2Sul_1$, likely a sulfated LacdiNAc. The fragment ions at *m/z* 870 and 667 ($Y_{2\beta}$ and $Y_{2\alpha}/Y_{2\beta}$) suggested this structure had one HexNAc and one Fuc as terminal. Together with cross-ring cleavage ions at *m/z* 356 ($^{3,5}A_{2\alpha}$), this structure was likely a sulfated LacdiNAc on a core 2 like glycan. Due to the migration of sulfate in MS/MS collision, the exact linkage and location of the sulfate could not be determined.

Diversity of *O*-glycans in health and diseased gastric tissues

The diversity of human gastric *O*-glycans reflected by chromatographic profiles was due to not only different peak numbers but also varied abundance of same peak in different samples (e.g., Figure 8A-C). In order to display the great variety of human gastric *O*-glycans a clustered image map with various glyco-epitopes was made (Figure 8D). The clustered groups revealed that the majority of structures was fucosylated core 2 *O*-glycans in human gastric mucins (Figure 8D). Indeed, the major fucosylation was attributed to the prevalence of blood type H like epitopes

(Fucα1-2Galβ-), indicating that blood group and secretor status were the main contributing factor for gastric blood group variation. Furthermore, despite that most sialylated structures detected were based on core 2 like structures, the sialylated core 1 like *O*-glycans, due to their less diversity, (675-1, 675-2 and 966 in Supplementary table) had a higher relative abundance. Thus, sialylation was clustered with core 1 like *O*-glycans in the map (Figure 8D). The sialyl Tn and sulfated glycans were often found in tumor tissue, though not exclusively (Table 1). Describing the overall glycosylation based on identified structural traits, normal (82.2±9.7%) and tumor-adjacent tissue (78.3±8.7%), tended to have less core 2 like *O*-glycans compared to tumor tissues (70.0±10.5%). On the contrary, tumor tissues tended to have higher content of core 1 *O*-glycans (22.9±3.1%) in comparison with that of normal tissues (14.6±4.6%) and tumor-adjacent tissue (17.2±1.4%, Table 1). However, these differences were not significant ($0.14>p>0.09$) and larger cohorts would be needed to investigate this. Interestingly, our data corroborates previous findings (39, 40) that the level of terminal α1,4-GlcNAc-like structures (20.0±9.4% of total relative abundance) where higher on mucins from gland mucous cells in comparison to mucins from the surface of tumor-adjacent tissue (3.1±2.8%, $p= 0.01$).

The variety of the *O*-glycans was reflected both on an individual level and between different types (tumor/normal/normal tumor-adjacent, Figure 9). The number of *O*-glycans from each sample ranged from 16 to 103 summing up to 258 oligosaccharides. 92 out of 258 (36% of detected structures) were detected in all three groups (normal, tumor-adjacent, and tumor tissue, Figure 9A). However, there were 32 (12%), 32 (12%) and 43 (17%) unique *O*-glycans present in normal, tumor-adjacent and tumor, respectively (Figure 9A), demonstrating that the variation between groups was similar to within groups. In addition, human gastric *O*-glycans

appeared individual-specific. More than one third (87 out of 258 characterized *O*-glycans) was found in only one individual including 29, 26, and 32 unique glycans isolated from normal, tumor-adjacent and tumor tissues (Figure 9B). Only 14 *O*-glycans (*i.e*. 5% of the structures were detected in all 10 individuals (Figure 9B). The structures present in 7 or more individuals, were present in all three tissue types (*i.e.* tumor, tumor-adjacent and normal), indicating that these are common structures that are independent of disease status. Mucins from normal, tumor-adjacent and tumor tissue had similar numbers of sialylated glycans (40, 32 and 34, respectively, Figure 9C). On the contrary, tumor tissue was the only tissue that had a large number of sulfated structures (27 out 29 *O*-glycan), and only 7 versus 3 sulfated structures were found in tumor-adjacent and normal tissue, respectively (Figure 9D). In addition, the relative amount of sulfated structures from normal tissue (0.5±0.6%) was lower in comparison with that of tumor (7.6±8.9%) and tumor-adjacent tissue (4.7±11.7%, Table 1). The data suggest that negatively charged glycans, especially sulfated glycans, in tumor-adjacent tissue may reflect the transition status from normal into tumor tissue.

Discussion

In the present study, the gastric *O*-glycosylation profile expanded to 258 *O*-glycans originating from normal, tumor-adjacent tissue and tumor tissue, demonstrating a great diversity of the human gastric *O*-glycan profile. In agreement with previous studies (38, 41), the core 2 *O*-glycans were the dominant structure in human gastric mucins. The inter-individual heterogeneity of gastric *O*-glycans was mainly due to ABH and Lewis blood group epitopes. The intra-individual heterogeneity of gastric *O*-glycans was, on the other hand, due to the properties and site of origin of the isolated mucins (glands versus surface; soluble versus insoluble). In addition to general modifications, we observed, for the first time, the presence of the I-branch on core 2 *O*-glycans in human gastric mucins, albeit in low amounts. The *O*-glycans with I-branch can serve as a scaffold, and was modified with blood group and Lewis like epitopes.

The majority of human gastric *O*-glycans were fucosylated (71%) including ABH (54%) and/or Lewis like blood group epitopes (30%). The high level of fucosylation and relatively low level of sialylation of human gastric *O*-glycans supported the hypothesis that there was an acidic gradient from stomach to colon (42), which can be speculated to regulate the regional distribution of bacterial species. Lewis epitopes, especially Le^b, are closely related to gastric pathology: attachment of pathogens such as *H. pylori*, to the mucous epithelial cells and the mucous layer lining the gastric epithelium is the critical step for the pathogenesis (11, 43). In addition to Le^b, there are also ALe^b, BLe^b and OLe^b blood group antigens at the non-reducing end. In our previous study, all samples except one normal (P8H) have shown expression of Le^b as determined by ELISA (19). However, the amount of $Le^{b/y}$ obtained by structural analysis of *O*-

glycans from human gastric samples (normal, tumor-adjacent and tumor tissue, Table 1) did not correlate with the Le^b signal obtained by ELISA (19). This discrepancy may be due to cross-reactivity of the antibody with the H1 structure, or that Le^b may be present on large heterogeneous structures below detection limit in the MS, or due to the low proportion of $Le^{b/y}$ in examined samples. The inter-individual heterogeneity of gastric *O*-glycans in this study was mainly due to ABH and Lewis like epitopes. This is strikingly different from the human colonic *O*-glycans, where MUC2 is the dominant mucin and its core 3-dominating *O*-glycans are largely lacking blood group antigens and almost identical inter-individually (44). Only 5% of the gastric glycan structures were found in all individuals in the current study, in comparison to that the corresponding number for common *O*-linked glycan structures in wild type salmon is 30% (28), indicating that in addition to the stomach being a region of high diversity in humans, the diversity in human glycosylation may be higher than for other species.

Several studies have implied that gastric *O*-glycans containing α1,4-GlcNAc inhibit *H. pylori* colonization and growth (21, 45, 46). And it is well known that *H. pylori* is a causative microbe for gastric cancer (47). In the present study, we detected high levels of α1,4-GlcNAc-like structures (9.3±8.0%, Table 1) in all tested samples except one. Thus, with our sample preparation, α1,4-GlcNAc was not a unique modification of mucins (such as MUC6) secreted from gastric gland mucous cells and Brunner's glands of the duodenal mucosa (39). The average level was higher than in another study where the level of expressed α1,4-GlcNAc was around 2.0±0.6% in 32 healthy individuals (48). The highest level was associated with mucins secreted from gland mucous cells of pyloric antrum and fundus (20.0±8.2%, Table 1). Higher prevalence of α1,4-GlcNAc-like structures in mucins from glands, where MUC6 dominate, is in agreement

with that glands has been shown to be responsible for its secretion (39, 40), and also with ELISA based results from the patients present in this study (19). Terminal α1,4-GlcNAc has also been indicated as a tumor suppressor for gastric adenocarcinoma in the study of *A4GNT*-deficient mice (49). However, the level of α1,4-GlcNAc in tumor tissue was similar to that from normal control (8.4 versus 5.2%, Table 1). However, it should be noted that the relative rather than absolute amount of selected glycans was used for comparison in this report. In our previous ELISA based study of these samples, the mucins from full gastric wall mucosa from tumor and healthy samples either had a low or no detectable level of α1,4-GlcNAc, whereas this structure was enriched in the mucins samples isolated from the glands only (19). Another study based on immunohistochemistry, also report tumor tissue to usually be negative for terminal α1,4-GlcNAc (40). The apparent discrepancy between the current study and these previous studies may be in that ELISA and immunohistochemistry usually are optimized to an antibody concentration where the highest signal of the sample set is set so it is in the linear range of the method, which may lead to those samples with low abundance fall below the detection limit.

The LacdiNAc-binding adhesin (LabA) from *H. pylori* binds to the LacdiNAc motif on MUC5AC (41). The expression of LacdiNAc was absent in cardiac gland, low in the surface of the fundic mucosa but more pronounced in pyloric glands (41). In this study, the *O*-glycans with LacdiNAc-like epitopes represented around 4% in normal tissue and tended to be lower in tumor and tumor-adjacent tissue. The value in normal tissue was similar to that of other studies where 3.4-7.0% relative abundance was reported, although a higher abundance has been found in intestinal metaplasia (8.5%) (41, 50). Sulfated LacdiNAc-like structure with sulfate linked to C6 of GlcNAc was different from reported sulfated LacdiNAc so far, where sulfate was linked to C4

position of GalNAc on both *N*- and *O*-glycans of human glycoprotein hormones as well as other glycoproteins (51, 52). This indicates that gastric LacdiNAc undergoes a different modification in comparison with that of brain and trachea (51, 52).

There was a trend that cancerous tissue tends to have higher level of sulfation and sialylation in this study. This is in agreement with the appearance of sulfomucins associated with the metaplastic process advances (50, 53). During the course of *H. pylori* infection, inflammation and cancer development, the *O*-glycosylation of mucins can change and display more sialylated and sulfated structures on the mucins (18, 54). We see presence of both sialylated structures including sTn and sulfated *O*-glycans in normal tissue as well, albeit to a low abundance. The detection of these structures by MS on healthy mucins, differ from another studies where no sulfomucin was detected in normal tissue (48). Interestingly, in A4GnT mutant mice, depleting terminal α1,4-GlcNAc lead to an increase of sialylation and fucosylation suggesting subtle remodeling of *O*-glycosylation in gastric mucosa (49). The increasing level of sialylation may also lead to the relative decline of core 2 in tumor tissue, leading to increased sialyl core 1 and sialyl Tn. Higher number and amount of sulfated *O*-glycans in tumor tissue in comparison to that of normal and tumor-adjacent tissue suggests that appearance of sulfated glycan may relate to tumorigenesis.

In conclusion, the gastric mucin *O*-glycosylation has a greater diversity than previously appreciated, and we identified some novel structures and linkages not described for this type of samples before. The diversity the gastric *O*-glycosylation broaden our understanding of the human gastric *O*-glycome and the structures presented in this study can function as a library for candidate structures important for pathogenesis, to be tested in biological assays.

References

1. Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 127, 2893-2917
2. Carter, D. (2014) New global survey shows an increasing cancer burden. *Am J Nurs* 114, 17
3. Cover, T. L., and Blaser, M. J. (2009) Helicobacter pylori in health and disease. *Gastroenterology* 136, 1863-1873
4. Suerbaum, S., and Michetti, P. (2002) Helicobacter pylori infection. *N Engl J Med* 347, 1175-1186
5. De Bolos, C., Garrido, M., and Real, F. X. (1995) MUC6 apomucin shows a distinct normal tissue distribution that correlates with Lewis antigen expression in the human stomach. *Gastroenterology* 109, 723-734.
6. Teixeira, A., David, L., Reis, C. A., Costa, J., and Sobrinho-Simoes, M. (2002) Expression of mucins (MUC1, MUC2, MUC5AC, and MUC6) and type 1 Lewis antigens in cases with and without Helicobacter pylori colonization in metaplastic glands of the human stomach. *J Pathol* 197, 37-43
7. Linden, S., Mahdavi, J., Hedenbro, J., Boren, T., and Carlstedt, I. (2004) Effects of pH on Helicobacter pylori binding to human gastric mucins: identification of binding to non-MUC5AC mucins. *Biochem J* 384, 263-270
8. Buisine, M. P., Devisme, L., Maunoury, V., Deschodt, E., Gosselin, B., Copin, M. C., Aubert, J. P., and Porchet, N. (2000) Developmental mucin gene expression in the gastroduodenal tract and accessory digestive glands. I. Stomach. A relationship to gastric carcinoma. *J Histochem Cytochem* 48, 1657-1666.
9. Ilhan, O., Han, U., Onal, B., and Celik, S. Y. (2010) Prognostic significance of MUC1, MUC2 and MUC5AC expressions in gastric carcinoma. *Turk J Gastroenterol* 21, 345-352
10. Navabi, N., Johansson, M. E., Raghavan, S., and Linden, S. K. (2013) Helicobacter pylori Infection Impairs the Mucin Production Rate and Turnover in the Murine Gastric Mucosa. *Infect Immun* 81, 829-837
11. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V., and McGuckin, M. A. (2008) Mucins in the mucosal barrier to infection. *Mucosal immunology* 1, 183-197
12. Linden, S., Semino-Mora, C., Liu, H., Rick, J., and Dubois, A. (2010) Role of mucin Lewis status in resistance to Helicobacter pylori infection in pediatric patients. *Helicobacter* 15, 251-258
13. Murata, K., Egami, H., Shibata, Y., Sakamoto, K., Misumi, A., and Ogawa, M. (1992) Expression of blood group-related antigens, ABH, Lewis(a), Lewis(b), Lewis(x), Lewis(y), CA19-9, and CSLEX1 in early cancer, intestinal metaplasia, and uninvolved mucosa of the stomach. *Am J Clin Pathol* 98, 67-75
14. Taylor, D. E., Rasko, D. A., Sherburne, R., Ho, C., and Jewell, L. D. (1998) Lack of correlation between Lewis antigen expression by Helicobacter pylori and gastric epithelial cells in infected patients. *Gastroenterology* 115, 1113-1122.

15. Chik, J. H., Zhou, J., Moh, E. S., Christopherson, R., Clarke, S. J., Molloy, M. P., and Packer, N. H. (2014) Comprehensive glycomics comparison between colon cancer cell cultures and tumours: implications for biomarker studies. *J Proteomics* 108, 146-162

16. Marionneau, S., Cailleau-Thomas, A., Rocher, J., Le Moullac-Vaidye, B., Ruvoen, N., Clement, M., and Le Pendu, J. (2001) ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* 83, 565-573

17. Varki, A., Kannagi, R., and Toole, B. P. (2009) Glycosylation Changes in Cancer. In: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E., eds. *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, The Consortium of Glycobiology Editors, La Jolla, California, Cold Spring Harbor (NY)

18. Linden, S., Mahdavi, J., Semino-Mora, C., Olsen, C., Carlstedt, I., Boren, T., and Dubois, A. (2008) Role of ABO secretor status in mucosal innate immunity and H. pylori infection. *PLoS Pathog* 4, e2

19. Skoog, E. C., Sjoling, A., Navabi, N., Holgersson, J., Lundin, S. B., and Linden, S. K. (2012) Human gastric mucins differently regulate Helicobacter pylori proliferation, gene expression and interactions with host cells. *PLoS One* 7, e36378

20. Linden, S., Nordman, H., Hedenbro, J., Hurtig, M., Boren, T., and Carlstedt, I. (2002) Strain- and blood group-dependent binding of Helicobacter pylori to human gastric MUC5AC glycoforms. *Gastroenterology* 123, 1923-1930

21. Kawakubo, M., Ito, Y., Okimura, Y., Kobayashi, M., Sakura, K., Kasama, S., Fukuda, M. N., Fukuda, M., Katsuyama, T., and Nakayama, J. (2004) Natural antibiotic function of a human gastric mucin against Helicobacter pylori infection. *Science* 305, 1003-1006

22. Skoog, E. C., Padra, M., Aberg, A., Gideonsson, P., Obi, I., Quintana-Hayashi, M. P., Arnqvist, A., and Linden, S. K. (2017) BabA dependent binding of Helicobacter pylori to human gastric mucins cause aggregation that inhibits proliferation and is regulated via ArsS. *Sci Rep* 7, 40656

23. Nordman, H., Davies, J. R., Lindell, G., de Bolos, C., Real, F., and Carlstedt, I. (2002) Gastric MUC5AC and MUC6 are large oligomeric mucins that differ in size, glycosylation and tissue distribution. *Biochem J* 364, 191-200

24. Karlsson, N. G., Schulz, B. L., and Packer, N. H. (2004) Structural determination of neutral O-linked oligosaccharide alditols by negative ion LC-electrospray-MSn. *J Am Soc Mass Spectrom* 15, 659-672

25. Hayes, C. A., Karlsson, N. G., Struwe, W. B., Lisacek, F., Rudd, P. M., Packer, N. H., and Campbell, M. P. (2011) UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics* 27, 1343-1344

26. Hanisch, F. G., Chai, W., Rosankiewicz, J. R., Lawson, A. M., Stoll, M. S., and Feizi, T. (1993) Core-typing of O-linked glycans from human gastric mucins. Lack of evidence for the occurrence of the core sequence Gal1-6GalNAc. *Eur J Biochem* 217, 645-655

27. Brockhausen, I. (1999) Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta* 1473, 67-95

28. Jin, C., Padra, J. T., Sundell, K., Sundh, H., Karlsson, N. G., and Linden, S. K. (2015) Atlantic Salmon Carries a Range of Novel O-Glycan Structures Differentially Localized on Skin and Intestinal Mucins. *J Proteome Res* 14, 3239-3251

29. Liu, J., Jin, C., Cherian, R. M., Karlsson, N. G., and Holgersson, J. (2015) O-glycan repertoires on a mucin-type reporter protein expressed in CHO cell pools transiently transfected with O-glycan core enzyme cDNAs. *J Biotechnol* 199, 77-89

30. Cherian, R. M., Jin, C., Liu, J., Karlsson, N. G., and Holgersson, J. (2015) A Panel of Recombinant Mucins Carrying a Repertoire of Sialylated O-Glycans Based on Different Core Chains for Studies of Glycan Binding Proteins. *Biomolecules* 5, 1810-1831

31. Everest-Dass, A. V., Abrahams, J. L., Kolarich, D., Packer, N. H., and Campbell, M. P. (2013) Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom* 24, 895-906

32. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 25, 1323-1324

33. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj J* 5, 397-409

34. Weinstein, J. N., Myers, T. G., O'Connor, P. M., Friend, S. H., Fornace, A. J., Jr., Kohn, K. W., Fojo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sausville, E. A., Zaharevitz, D. W., Bunow, B., Viswanadhan, V. N., Johnson, G. S., Wittes, R. E., and Paull, K. D. (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* 275, 343-349

35. Karlsson, N. G., Karlsson, H., and Hansson, G. C. (1996) Sulphated mucin oligosaccharides from porcine small intestine analysed by four-sector tandem mass spectrometry. *J Mass Spectrom* 31, 560-572

36. Van Halbeek, H., Gerwig, G. J., Vliegenthart, J. F., Smits, H. L., Van Kerkhof, P. J., and Kramer, M. F. (1983) Terminal alpha (1 leads to 4)-linked N-acetylglucosamine: a characteristic constituent of duodenal-gland mucous glycoproteins in rat and pig. A high-resolution 1H-NMR study. *Biochim Biophys Acta* 747, 107-116

37. Ali, L., Kenny, D. T., Hayes, C. A., and Karlsson, N. G. (2012) Structural Identification of O-Linked Oligosaccharides Using Exoglycosidases and MSn Together with UniCarb-DB Fragment Spectra Comparison. *Metabolites* 2, 648-666

38. Kenny, D. T., Skoog, E. C., Linden, S. K., Struwe, W. B., Rudd, P. M., and Karlsson, N. G. (2012) Presence of terminal N-acetylgalactosaminebeta1-4N-acetylglucosamine residues on O-linked oligosaccharides from gastric MUC5AC: involvement in Helicobacter pylori colonization? *Glycobiology* 22, 1077-1085

39. Nakamura, N., Ota, H., Katsuyama, T., Akamatsu, T., Ishihara, K., Kurihara, M., and Hotta, K. (1998) Histochemical reactivity of normal, metaplastic, and neoplastic tissues to alpha-linked N-acetylglucosamine residue-specific monoclonal antibody HIK1083. *J Histochem Cytochem* 46, 793-801

40. Ferreira, B., Marcos, N. T., David, L., Nakayama, J., and Reis, C. A. (2006) Terminal alpha1,4-linked N-acetylglucosamine in Helicobacter pylori-associated intestinal metaplasia of the human stomach and gastric carcinoma cell lines. *J Histochem Cytochem* 54, 585-591

41. Rossez, Y., Gosset, P., Boneca, I. G., Magalhaes, A., Ecobichon, C., Reis, C. A., Cieniewski-Bernard, C., Joncquel Chevalier Curt, M., Leonard, R., Maes, E., Sperandio, B., Slomianny, C., Sansonetti, P. J., Michalski, J. C., and Robbe-Masselot, C. (2014) The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *J Infect Dis* 210, 1286-1295

42. Robbe, C., Capon, C., Maes, E., Rousset, M., Zweibaum, A., Zanetta, J. P., and Michalski, J. C. (2003) Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *J Biol Chem* 278, 46337-46348

43. McGuckin, M. A., Linden, S. K., Sutton, P., and Florin, T. H. (2011) Mucin dynamics and enteric pathogens. *Nature reviews. Microbiology* 9, 265-278

44. Larsson, J. M., Karlsson, H., Sjovall, H., and Hansson, G. C. (2009) A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* 19, 756-766

45. Lee, H., Wang, P., Hoshino, H., Ito, Y., Kobayashi, M., Nakayama, J., Seeberger, P. H., and Fukuda, M. (2008) Alpha1,4GlcNAc-capped mucin-type O-glycan inhibits cholesterol alpha-glucosyltransferase from Helicobacter pylori and suppresses H. pylori growth. *Glycobiology* 18, 549-558

46. Hidaka, E., Ota, H., Hidaka, H., Hayama, M., Matsuzawa, K., Akamatsu, T., Nakayama, J., and Katsuyama, T. (2001) Helicobacter pylori and two ultrastructurally distinct layers of gastric mucous cell mucins in the surface mucous gel layer. *Gut* 49, 474-480

47. Peek, R. M., Jr., and Blaser, M. J. (2002) Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* 2, 28-37

48. Rossez, Y., Maes, E., Lefebvre Darroman, T., Gosset, P., Ecobichon, C., Joncquel Chevalier Curt, M., Boneca, I. G., Michalski, J. C., and Robbe-Masselot, C. (2012) Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for Helicobacter pylori. *Glycobiology* 22, 1193-1206

49. Karasawa, F., Shiota, A., Goso, Y., Kobayashi, M., Sato, Y., Masumoto, J., Fujiwara, M., Yokosawa, S., Muraki, T., Miyagawa, S., Ueda, M., Fukuda, M. N., Fukuda, M., Ishihara, K., and Nakayama, J. (2012) Essential role of gastric gland mucin in preventing gastric cancer in mice. *J Clin Invest* 122, 923-934

50. Joncquel Chevalier Curt, M., Lecointe, K., Mihalache, A., Rossez, Y., Gosset, P., Leonard, R., and Robbe-Masselot, C. (2015) Alteration or adaptation, the two roads for human gastric mucin glycosylation infected by Helicobacter pylori. *Glycobiology* 25, 617-631

51. Toyoda, M., Kaji, H., Sawaki, H., Togayachi, A., Angata, T., Narimatsu, H., and Kameyama, A. (2016) Identification and characterization of sulfated glycoproteins from small cell lung carcinoma cells assisted by management of molecular charges. *Glycoconj J* 33, 917-926

52. Hiraoka, N., Misra, A., Belot, F., Hindsgaul, O., and Fukuda, M. (2001) Molecular cloning and expression of two distinct human N-acetylgalactosamine 4-O-sulfotransferases that transfer sulfate to GalNAc beta 1-->4GlcNAc beta 1-->R in both N- and O-glycans. *Glycobiology* 11, 495-504

53. Correa, P. (1988) A human model of gastric carcinogenesis. *Cancer Res* 48, 3554-3560

54. Sakamoto, S., Watanabe, T., Tokumaru, T., Takagi, H., Nakazato, H., and Lloyd, K. O. (1989) Expression of Lewisa, Lewisb, Lewisx, Lewisy, siayl-Lewisa, and sialyl-Lewisx blood group antigens in human gastric carcinoma and in normal gastric tissue. *Cancer Res* 49, 745-752

Footnotes

The work was supported by the Swedish Research Council (521-2011-2380 and 621-2013-5895), the Swedish Research Council Formas (223-2011-1073 and 221-2011-1036), The Swedish Cancer Society, Ruth and Richard Julins Foundation, the Jeanssons Foundations, the Th Nordströms Foundation, the Engkvist Foundation, the Erling-Persson Family Foundation, the European Union FP7 GastricGlycoExplorer ITN under grant agreement no.316929 and the Knut and Alice Wallenberg Foundation. The mass spectrometer (LTQ) was obtained by a grant (324-2004-4434) from the Swedish Research Council.

Figure Legends

Figure 1. *In situ* proximity ligation assay (PLA) showing human gastric MUC5AC decorated with ABH-blood group antigens. Primary antibodies recognizing MUC5AC and blood type A (**A**), B (**B**), and H (**C**) were used. The PLA signal was shown in red fluorescent spots indicating that MUC5AC and corresponding blood group antigens were in close proximity. The nuclei were stained blue with DAPI. The autofluoresence of tissue (green) delineated the tissue structure.

Figure 2. Core type of human gastric *O*-glycans. (A) MS/MS spectra of a core 1 like *O*-glycan with a composition of $Hex_3HexNAc_3$ ($[M - H]^-$ of *m/z* 1114). (B) MS/MS spectra of a core 2 like *O*-glycan with a composition of $Hex_2HexNAc_2$ ($[M - H]^-$ of *m/z* 1114). (C) MS/MS spectra of a core 3 like *O*-glycan with a composition of $Hex_1HexNAc_3deHex_1$ ($[M - H]^-$ of *m/z* 936). (D) MS/MS spectra of two glycans with a composition of $Neu5Ac_1HexNAc_2$ ($[M - H]^-$ of *m/z* 716) containing either core 3 (upper) or core 5 like *O*-glycan (lower). The LC chromatograph of these two isomers showed different retention times on the PGC column. Symbol represents: yellow circle, Gal; yellow square, GalNAc; blue square, GlcNAc; empty square, HexNAc; purple diamond, Neu5Ac; S, sulfate; *, impurities. Structural annotation was based on the knowledge of glycan biosynthesis. Detailed assumptions related to linkage configuration and position, and validation of assigned structures can be found in materials and methods. Proposed structures are depicted using the Symbol Nomenclature for Glycomics (SNFG) (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Figure 3. I-branched and i-like extended structures on human gastric *O*-glycans. (A) MS/MS spectra of a glycan with the blood group I like antigen on the C3 branch of a core 2 like *O*-glycan ($Hex_4HexNAc_4$, $[M - 2H]^{2-}$ of *m/z* 739). (B) MS/MS spectra of i-like epitope extended core 1 like *O*-glycan with a composition of $Hex_2HexNAc_3$ ($[M - H]^-$ of *m/z* 952). Proposed structures are depicted using SNFG (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Figure 4. LC-MS/MS spectra of ABH histo blood group antigen containing *O*-glycans on human gastric mucins. (A) MS/MS spectra of a glycan with one blood group B like epitope on a core 2 like *O*-glycan ($Hex_3HexNAc_2deHex_2$, $[M - H]^-$ of *m/z* 1203). (B) MS/MS spectra of a glycan with one blood group A and B like epitope on core 2 like *O*-glycan ($Hex_3HexNAc_3deHex_2$, $[M - H]^-$ of *m/z* 1406). (C) MS/MS spectra of a fucosylated glycan with a composition of $Hex_3HexNAc_3deHex_3$ ($[M - 2H]^{2-}$ of *m/z* 775) containing internal H type 2 like epitope. Proposed structures are depicted using SNFG (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Figure 5. LC-MS/MS spectra of fucosylated *O*-glycans on human gastric mucins. (A) MS/MS spectra of a Lewis a/x containing glycan with a composition of $Hex_2HexNAc_2deHex_2$ ($[M - H]^-$ of *m/z* 1041). (B) MS/MS spectra of a Lewis b/y containing glycan with a composition of $Hex_2HexNAc_2deHex_3$ ($[M - H]^-$ of *m/z* 1187). (C) MS/MS spectra of a blood group A like Lewis b/y containing glycan with a composition of $Hex_2HexNAc_4deHex_3$ ($[M - 2H]^{2-}$ of *m/z* 976). Proposed structures are depicted using SNFG (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Figure 6. LC-MS/MS spectra of terminal α1,4-GlcNAc and LacdiNAc-like *O*-glycans in human gastric mucins. (A) MS/MS spectra of a LacdiNAc like glycan with a composition of $Hex_1HexNAc_3$ ($[M - H]^-$ of *m/z* 790). **(B)** MS/MS spectra of a terminal α1,4-GlcNAc-like glycan with a composition of $Hex_2HexNAc_4$ ($[M - H]^-$ of *m/z* 1155). **(C)** MS/MS spectra of a dual α1,4-GlcNAc-like and LacdiNAc-like glycan with a composition of $Hex_1HexNAc_4$ ($[M - H]^-$ of *m/z* 993). Proposed structures are depicted using SNFG (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Figure 7. LC-MS/MS spectra of acidic *O*-glycans on human gastric mucins. (A) MS/MS spectra of a glycan with a composition of $Neu5Ac_1HexNAc_1$ ($[M - H]^-$ of *m/z* 513). (B) MS/MS spectra of a sialyl Lewis a/x containing glycan with a composition of $Neu5Ac_1Hex_3HexNAc_2deHex_2$ ($[M - H]^-$ of *m/z* 1494). (C) MS/MS spectra of a sialyl Lewis a/x containing glycan with a composition of $Neu5Ac_1Hex_3HexNAc_3deHex_2$ ($[M - 2H]^{2-}$ of *m/z* 848). (D) MS/MS spectra of a sulfated LacdiNAc-like glycan with a composition of $Hex_1HexNAc_3deHex_1Sul_1$ ($[M - H]^-$ of *m/z* 1016). Proposed structures are depicted using SNFG (32) and nomenclature of fragments of carbohydrates as defined by Domon and Costello (33).

Figure 8. LC-MS chromatograms of *O*-glycans from mucins isolated from gastric tissue (A-C) and clustered image map of common glyco-epitopes based on their relative amount (D). (A-C) Spectra from mucins from tumor (A), normal tissue from a tumor affected stomach (B) and healthy (C) gastric tissue. The *O*-glycans were analyzed by LC-MS/MS in negative-ion mode. Major structures $[M - H]^-$ ions were depicted using CFG symbol nomenclature. (D) Clustered image map of the relative % abundance of the calculated structural glycan features. The structural features were plotted on the x-axis and subjects were plotted on the y-axis.

Calculated structural features were clustered based on Manhattan distance, with average linkage. 10 Samples were listed on the right-side of map including four (P1T-P4T) from gastric adenocarcinoma tumors, three (P6TA-P7TA) from normal mucosa of tumor-adjacent tissues, and three (P8H-10H) from normal tissue. For mucins from tumors, they were divided into soluble (S) and insoluble mucins (I). Mucins from tumor-adjacent tissues were separated into fundus (F) versus pyloric antrum (A), surface (S) versus gland material (G) according to tissue location.

Figure 9. Distribution of glycans between individuals and disease status. (A) Number of gastric *O*-glycans (in percentage) of the total 258 *O*-glycans, detected in stomachs of different disease status. (B) Number of glycans (of the total 258 *O*-glycans) detected in different individual(s): the bar above 1 on the x-axis represents the number of structures present in only one individual (structures unique to single individuals), whereas the bar above 10 represents structures present in all 10 individuals. (C-D) Distribution of the 53 sialylated *O*-glycans (C) and 29 sulfated *O*-glycans (D).

Tables

Table 1. Distribution of *O*-glycan features on mucins from gastric adenocarcinoma tumor (Tumor), normal mucosa of tumor-adjacet tissue (Norm.tum.adj) and normal tissue (Normal). The relative amounts of the different glycan features are given in percentage (%) in the relation to the total sum of integrated peak areas in the LC-MS chromatograms.

Samples[a]		Core 2[b]	Core 1	I antigen	Blood group	H type	A type	B type	AB type	$Le^{a/x}$	$sLe^{a/x}$	$Le^{b/y}$	Fucosylation	α1,4-GlcNAc	LacdiNAc	sTn	Sialylation	Sulfation
Tumor	P1TS	81.1	17.9	1.1	B	89.2	–	0.2	–	0.9	0.6	0.1	90.6	4.0	6.3	0.7	8.9	0.8
	P1TI	63.7	26.8	0.3	B	29.3	–	8.9	–	10.0	7.9	0.6	69.4	5.8	0.7	3.3	22.8	1.1
	P3TS	62.7	35.1	2.1	H	31.2	–	–	–	0.7	–	–	31.9	6.5	–	–	21.3	7.1
	P2TS	62.9	22.1	2.8	A	17.2	3.3	–	–	2.7	2.0	–	24.3	8.3	0.9	10.6	59.8	15.7
	P4TS	63.8	22.8	4.7	A	42.5	4.5	–	–	4.1	1.7	–	47.0	18.5	1.2	11.3	24.7	21.1
	P4TI	85.9	12.7	1.4	A	56.1	20.2	–	–	0.7	0.4	–	83.6	7.3	5.8	–	4.2	–
Norm.tum.adj	P6TA-FS	84.0	12.3	5.4	A	11.1	1.3	–	–	6.7	1.0	–	18.9	5.4	3.2	0.3	9.2	0.3
	P6TA-FG	77.1	18.8	9.1	A	50.6	18.9	–	–	5.7	–	0.2	73.7	16.5	1.3	0.5	4.7	0.3
	P6TA-AS	80.0	16.7	1.3	A	69.5	16.4	–	–	1.0	–	–	86.0	5.5	2.4	–	6.8	1.3
	P6TA-AG	85.3	14.7	22.5	A	33.3	6.4	–	–	18.9	0.8	–	58.7	32.1	–	–	14.1	–
	P5TA-AS	62.3	23.4	0.8	AB	19.1	0.9	4.2	0.3	26.0	23.7	0.5	45.5	1.6	0.2	12.5	67.7	1.4
	P5TA-AG	68.8	20.9	5.5	AB	22.5	18.9	8.7	3.5	–	–	–	53.6	21.6	0.8	–	–	–
	P7TA-AS	68.8	13.2	–	AB	71.7	–	–	–	–	–	–	61.0	–	–	–	21.5	33.6
	P7TA-AG	83.4	17.0	1.9	AB	34.9	50.9	2.3	1.3	14.8	–	6.1	93.3	9.6	3.6	0.3	1.0	0.4
Normal	P8H	78.0	19.1	4.3	A	50.4	7.9	–	–	6.0	0.3	0.9	62.2	3.6	2.1	0.9	9.5	0.3
	P9H	93.2	5.4	4.5	B	64.8	–	15.8	–	7.0	0.6	0.5	88.8	8.0	8.0	1.3	7.1	–
	P10H	75.4	19.1	1.1	AB	46.9	0.8	1.2	0.6	15.5	13.1	1.4	68.5	3.9	2.5	2.2	37.6	1.1

[a] 10 Samples included four specimens (P1-P4) from gastric adenocarcinoma tumors (T), three (P5-P7) from normal mucosa of tumor-adjacent tissues (TA), and three (P8-10) from normal tissue (H). For mucins from tumors, they were divided into soluble (S) and insoluble mucins (I). Mucins from tumor-adjacent tissues were separated into fundus (F) versus pyloric antrum (A), surface (S) versus gland material (G) according to tissue location.

[b] Identification of structural epitopes was based on knowledge of gastric glycan biosynthesis and assumptions made on linkage configuration and positions were summarized in material and methods.

Figures

Figure 1

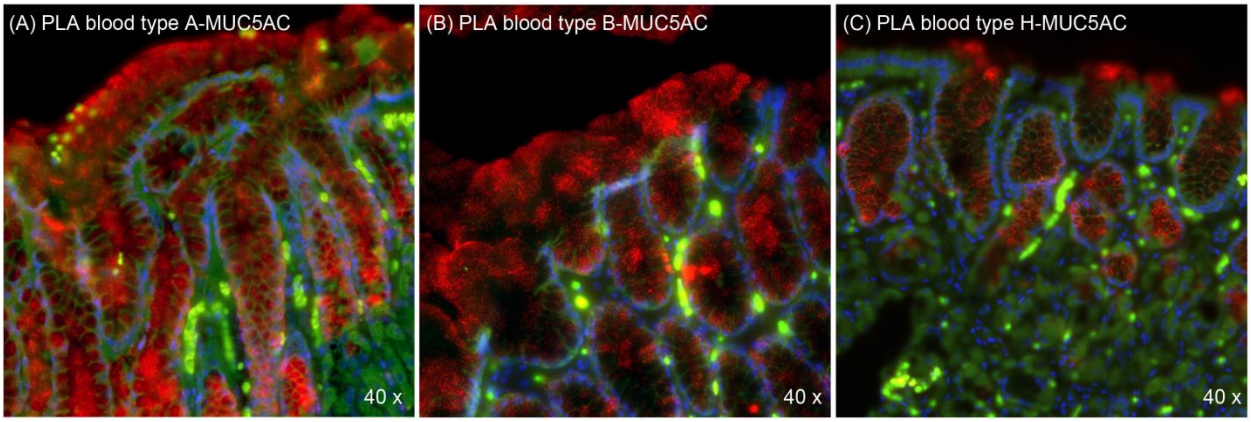

Figure 2

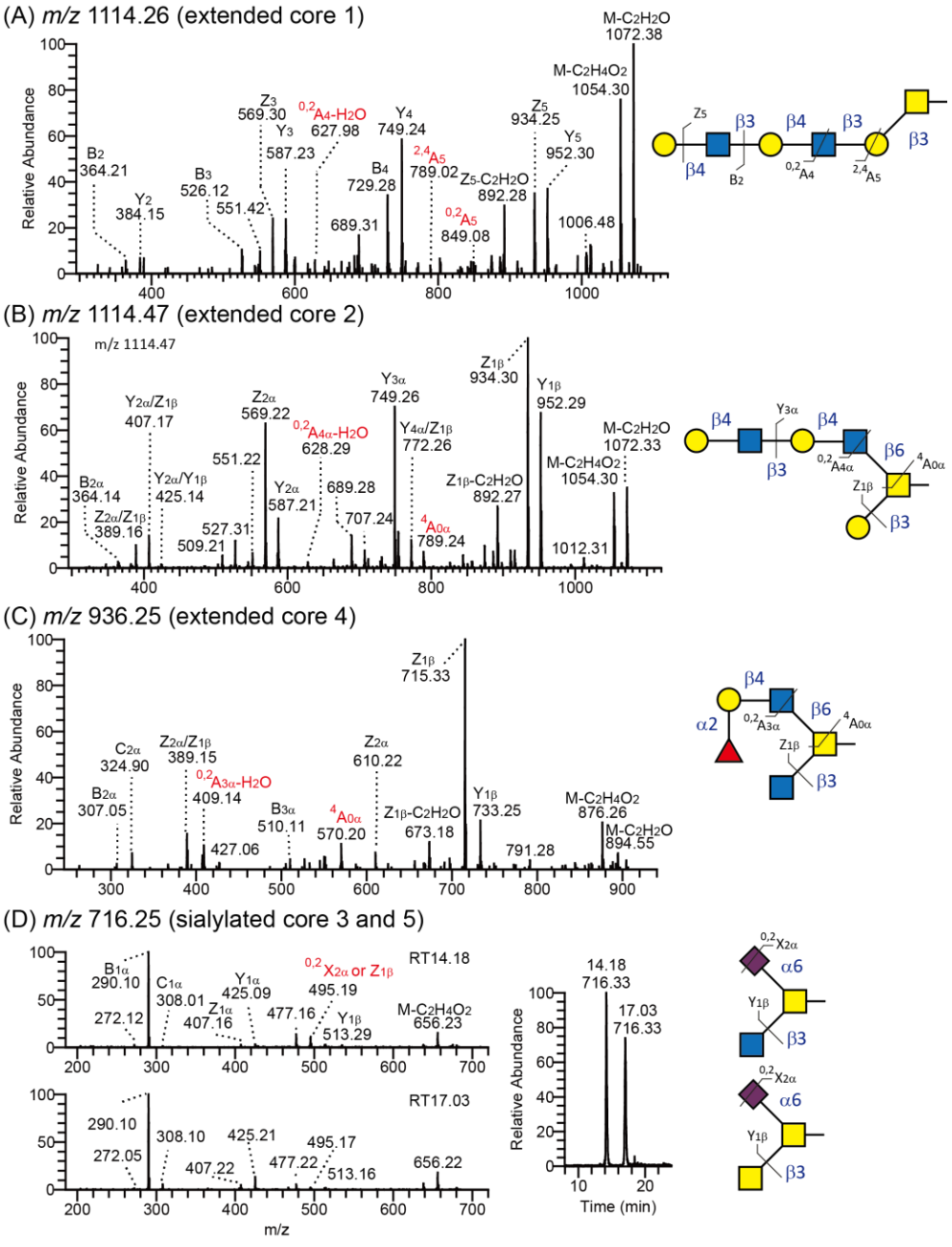

Figure 3

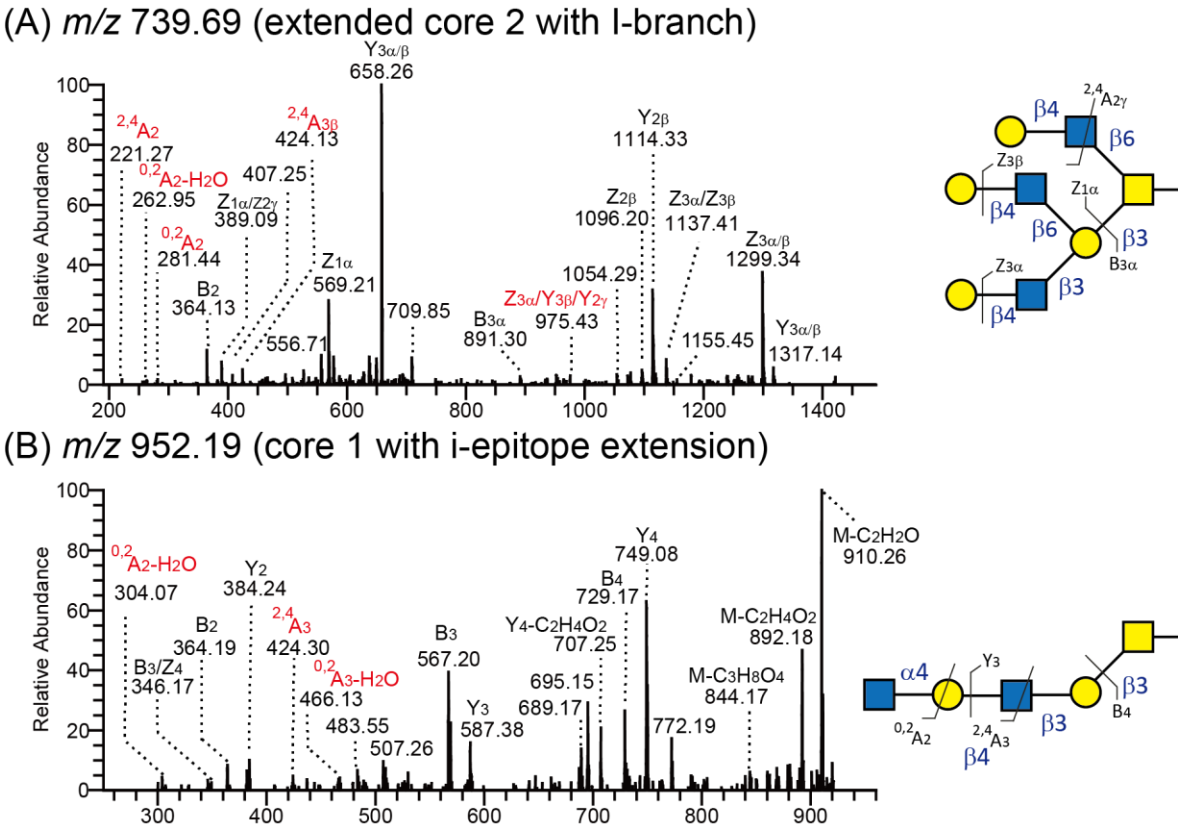

Figure 4

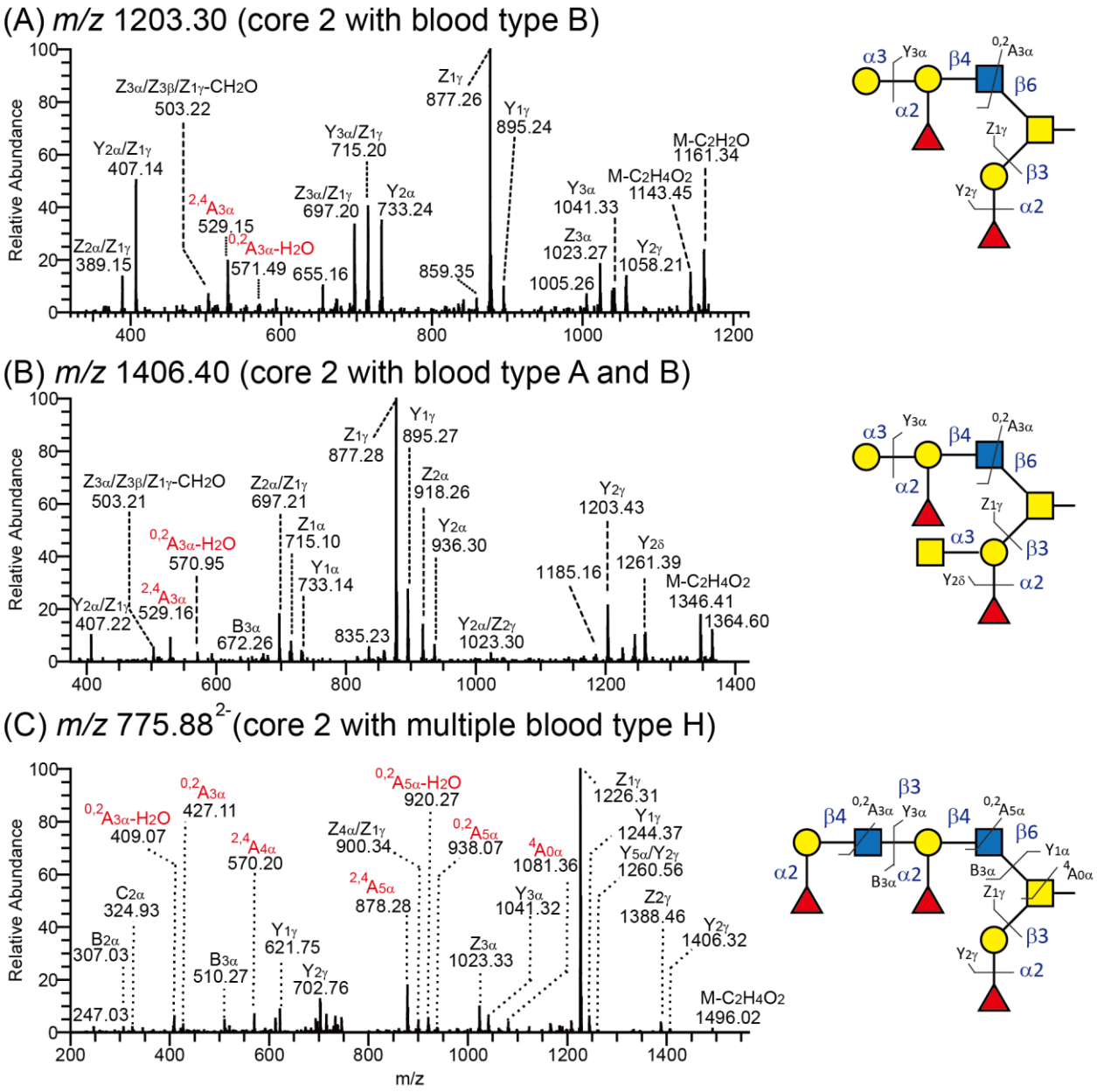

Figure 5

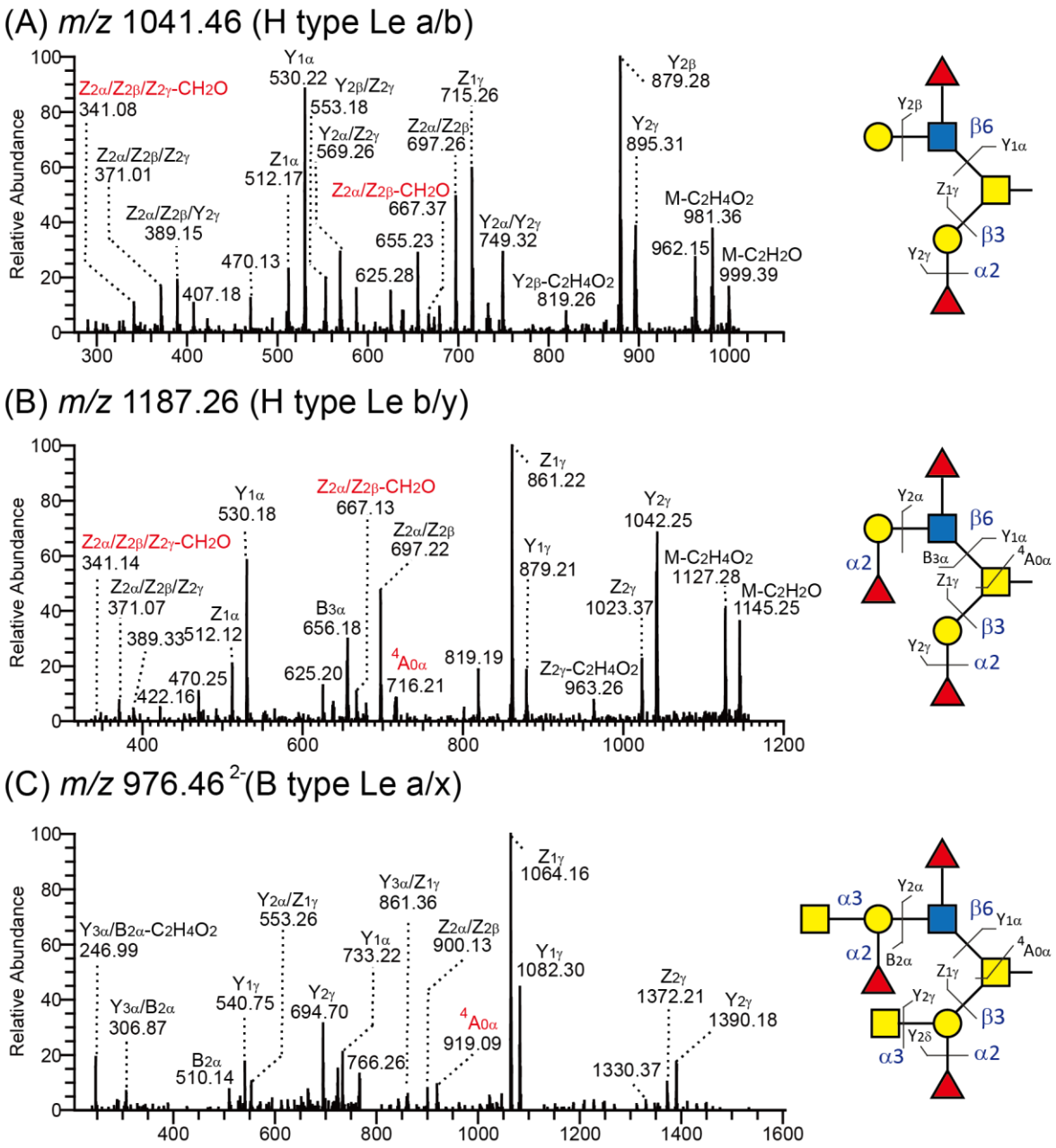

Figure 6

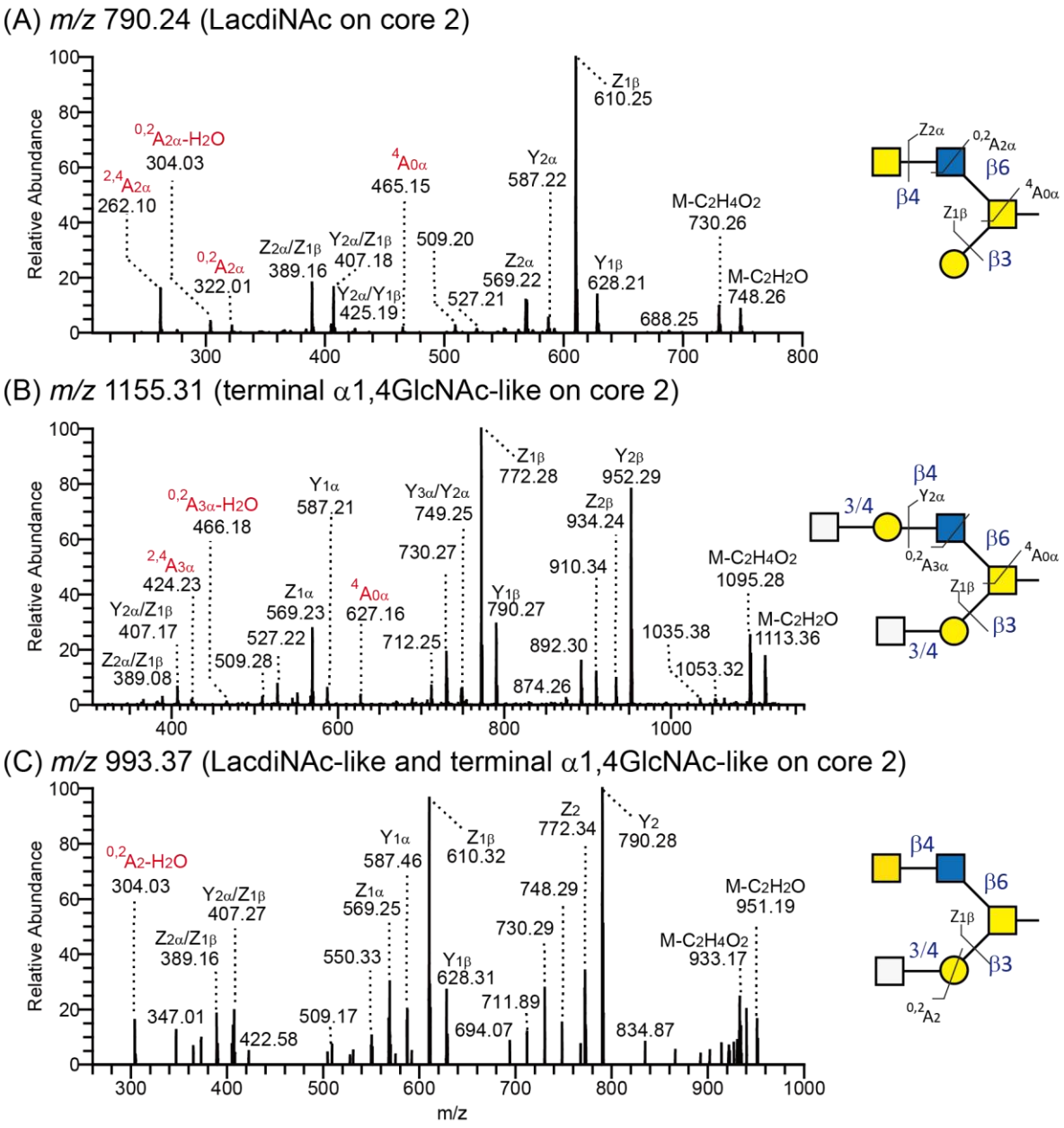

Figure 7

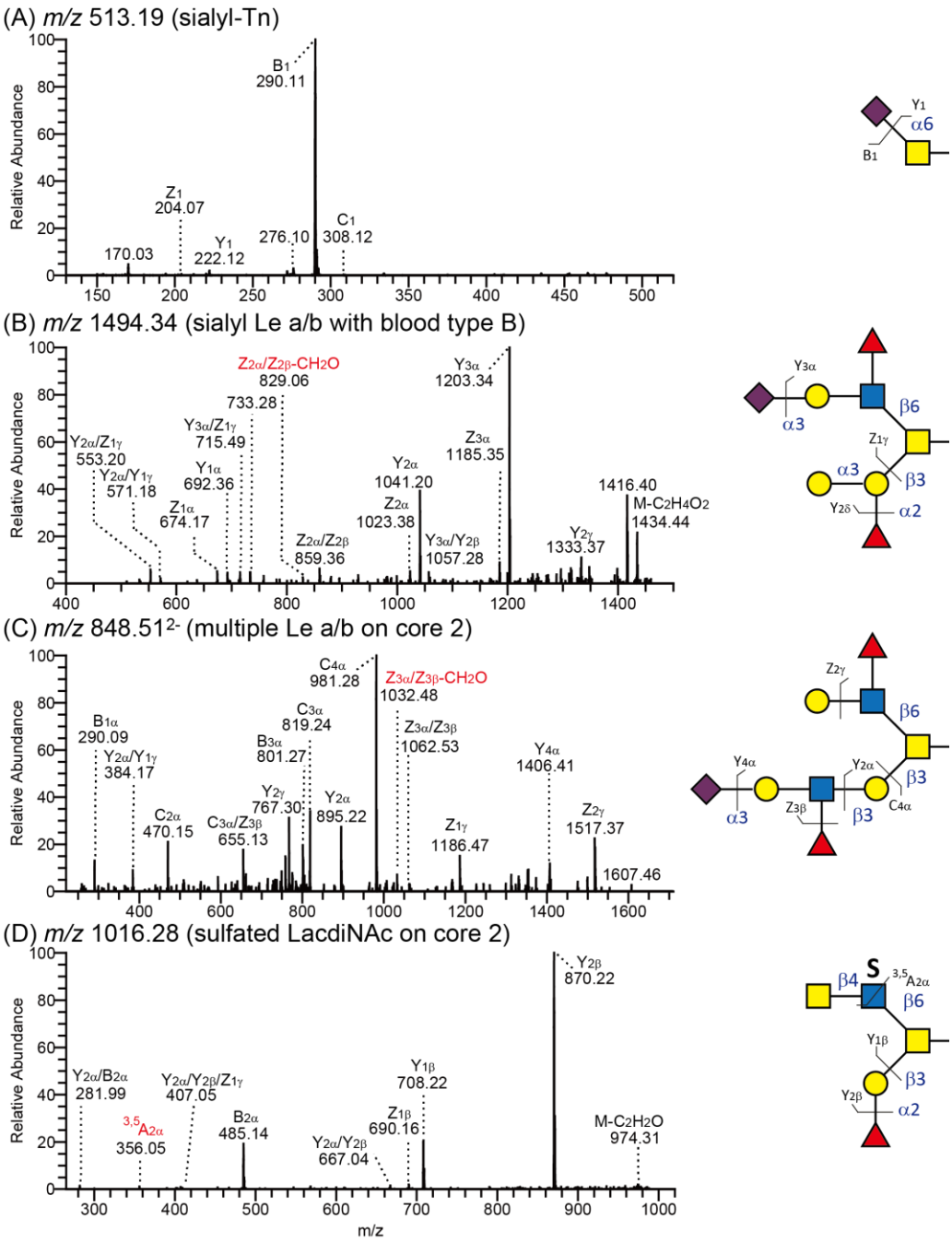

Figure 8

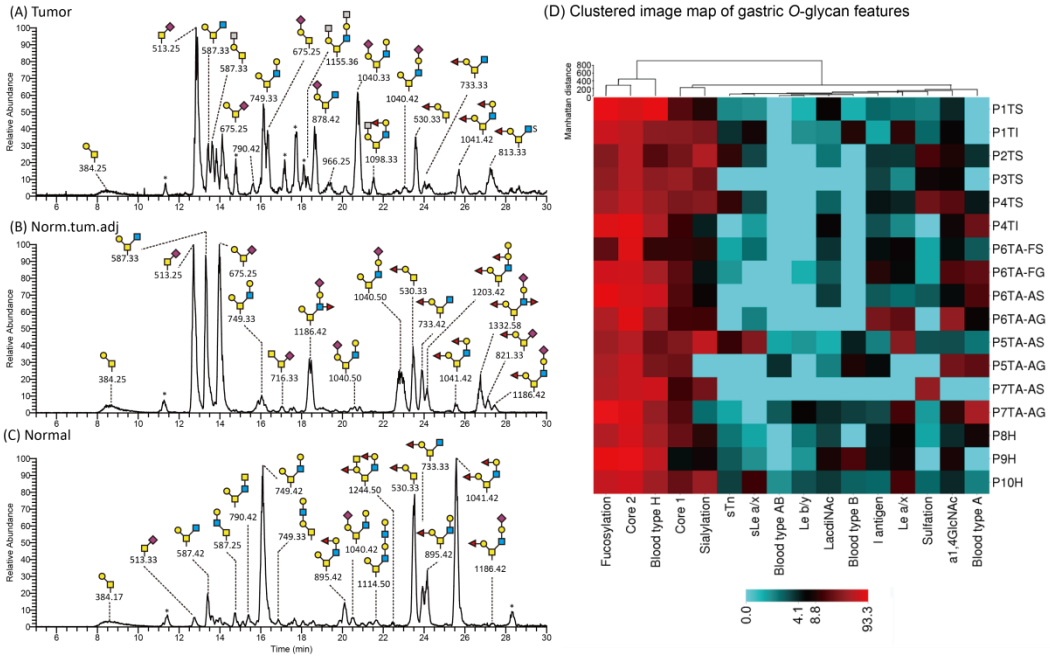

Figure 9