# Large-scale identification of *N*-glycan glycoproteins carrying Lewis x and site-specific *N*-glycan alterations in *Fut9* knockout mice

（α1,3-フコース転移酵素９ノックアウトマウスを用いた糖鎖および糖タンパク質の網羅的な解析）

２０１５

筑波大学大学院博士課程人間総合科学研究科

野呂　絵里花

筑　波　大　学

博 士 （医 学） 学 位 論 文

# CONTENTS

**ABSTRACT** …….. 2

**1. INTRODUCTION** …… 4

**2. EXPERIMENTAL SECTION** …… 7

2-1. Knockout Mice

2-2. Preparation of Mouse Kidney Glycopeptides

2-3. Collection of Glycopeptides Carrying Fucosylated Glycans by *Aleuria aurantia* lectin (AAL)-Affinity Chromatography

2-4. Deglycosylation and IGOT of *N*-Glycopeptides

2-5. Liquid Chromatography-Mass Spectrometry (LC–MS) Identification of AAL(+) Glycopeptides

2-6. MS Analysis of the *N*-Glycans Released from Glycopeptides

2-7. Site-Specific *N*-Glycome Analysis of AAL(+) Glycopeptides

***(1)*** *Assignment of MS Glycopeptide Signals*

***(2)*** *Prediction of the Core Peptide and Glycan Compositions for Each Glycopeptide Cluster*

***(3)*** *Confirmation of the Predicted Core peptide and the Site-Specific Glycome*

**3. RESULTS** …… 17

3-1. Profiling the *N*-glycan Structures in the Kidneys of WT and KO Mice

3-2. Identification of AAL(+) Glycopeptides and Their *N*-glycan Profiles

3-3. Identification of the Carrier Proteins of $Le^x$ by Site-Specific Glycome Analysis

**4. DISCUSSION** …….. 26

**5. ACKNOWLEDGEMENTS** ……..……...… 34

**6. REFERENCES** …….. 35

**7. TABLES** ..... 45

**8. FIGURE LEGENDS** ..... 47

**9. FIGURES** ..... 52

**ABSTRACT**

The Lewis x ($Le^x$) structure (Galβ1–4(Fucα1–3)GlcNAc-R) is a carbohydrate epitope comprising the stage-specific embryonic antigen-1 (SSEA-1) and CD15, and it is synthesized by α1,3-fucosyltransferase 9 (Fut9). *Fut9* is expressed specifically in the stomach, kidney, brain, and in leukocytes, suggesting a specific function in these tissues. In this study, the *N*-linked glycan mass spectrometry profile of wild-type mouse kidney glycoproteins revealed the presence of abundant terminal fucoses, which were lost following knockout of the *Fut9* gene; the terminal fucose was therefore concluded to be $Le^x$. These results suggested that $Le^x$ presence is widespread rather than being limited to specific proteins. We endeavored to comprehensively identify the $Le^x$ carriers in the mouse kidney. Glycopeptides carrying fucosylated glycans were collected by *Aleuria aurantia* lectin (AAL) affinity chromatography from kidney homogenates of wild-type and *Fut9* knockout mice. The site-specific *N*-glycomes on the glycopeptides were subsequently analyzed by adopting a new glycoproteomic technology composed of dissociation-independent assignment of glycopeptide signals and accurate mass-based prediction of the *N*-glycome on the glycopeptides. Our analyses demonstrated that

24/32 glycoproteins contained the $Le^x$ *N*-glycan structure in wild-type kidney; of these, $Le^x$ was lost from 21 in the knockout mice. This is the first report of large-scale identification of $Le^x$-carrying glycoproteins from a native sample based on the site-specific glycome analysis.

KEYWORDS: α1,3-fucosyltransferase 9 (Fut9), knockout mouse, kidney, Lewis x, glycoproteomics, IGOT-LC–MS, site-specific *N*-glycome analysis

## INTRODUCTION

Lewis x ($Le^x$) is an epitope of stage specific embryonic antigen-1 (SSEA-1) and is known as a neural stem cell marker.[1] Early embryos of mice from the eight-cell to the morula stage develop SSEA-1,[2] and it is thought that this antigen may be essential to the initial embryonic cell compaction in mice, as proposed by Solter. An *in vitro* fucosyltransferase assay suggested that the *Fut4* and *Fut9* gene products control biosynthesis of the $Le^x$ structure in mice.[3, 4] However, Fut9 has a higher ability to transfer fucose and synthesize $Le^x$ on a nonreducing end than does Fut4,[3] whereas Fut4 has strong activity for fucose transfer and $Le^x$ synthesis at a reducing end.[1] From various observations and reports, we estimated that $Le^x$ is synthesized primarily by Fut9 *in vivo*.[3–7]

To determine the function of Fut9 *in vivo*, Kudo et al. generated knockout mice (KO mice).[8] The KO mice had the ability to reproduce and grow normally. The $Le^x$ epitope of SSEA-1 was completely absent in early embryos and primordial germ cells in the KO mice, although the *Fut4* gene was expressed.[8] However, as the KO mice lacking the $Le^x$ epitope of SSEA-1 developed normally, the hypothesis with respect to its significance

during embryogenesis was incorrect.[8]

The *Fut9* gene is expressed in the stomach, kidney, brain, and certain other tissues in normal wild-type (WT) mice.[9–11] The Fut9 protein can produce a $Le^x$ epitope on both glycoproteins and glycolipids. For example, $Le^x$ structure is an epitope of the CD15 antigen on both glycoprotein and glycolipid. Neurobiologists have suggested that the CD15 antigen is important for synaptogenesis during brain development.[12–14] In concordance with this, it has recently been shown that *Fut9* is one of the causative genes of schizophrenia.[15] In addition, it has been predicted that CD15 might be involved in neutrophilic function, intracellular interactions, phagocytosis, and the stimulation of degranulation, as well as in respiratory bursts.[16, 17]

$Le^x$ and Fut9 protein are colocalized in neurons,[4] subglandular epithelial cells in the stomach[5], the proximal tubules in kidney,[18] and mature polymorphs.[19] This strongly indicates that Fut9 is responsible for $Le^x$ synthesis *in vivo*. The expression of Fut9 is closely regulated by both cell- and stage-specific mechanisms. However, the function of $Le^x$ in this process has not been determined, and little is known regarding the details of its biological function in other cells, as very few proteins modified by $Le^x$ have been

identified. To elucidate the function of $Le^x$, it is important to identify the core proteins carrying $Le^x$.

It is theoretically possible to use an antibody such as anti-SSEA-1 for capturing proteins carrying $Le^x$. We examined several anti-$Le^x$ antibodies; however, we could not find the antibody applicable to capturing $Le^x$-carrying glycopeptides comprehensively (data not shown). Thus, we designed a new approach to comprehensively identify $Le^x$-carrying glycoproteins. In this study, we identified glycopeptides carrying fucose(s) from normal mouse kidney using isotope-coded glycosylation site-specific tagging (IGOT),[20] and we specified $Le^x$-carrying glycopeptides among them by analyzing site-specific glycome using the newly developed *N*-glycoproteomics technology. Consequently, we revealed glycoproteins from which $Le^x$ was lost in *Fut9* KO mice compared to WT to determine the set of $Le^x$-carrying glycoproteins (Fut9 target proteins).

## EXPERIMENTAL SECTION

**Knockout Mice.** *Fut9* KO mice generated as previously described using standard gene-targeting techniques were backcrossed for five generations into the C57BL/6 genetic background.[8] Experimental animals were bred and housed in a specific pathogen-free animal facility and provided with sterilized chow and water ad libitum. The Institutional Animal Care and Use Committee approved the use of animals in this study.

**Preparation of Mouse Kidney Glycopeptides.** Male mice (60 or 64 weeks old) were anesthetized by 2-propanol inhalation. The blood was removed from an incision on the right atrium and was replaced by phosphate-buffered saline (PBS) into the left ventricle, followed by removal of the kidneys and storage at –80°C. Peptide preparation was carried out as previously described.[21, 22] Briefly, four kidneys from two mice (total weight: 1.06 g for WT mice and 1.02 g for KO mice) were homogenized using a polytron with a denaturing buffer (0.5 M Tris-HCl, pH 8.5, containing 7 M guanidine-HCl and 50 mM EDTA) and the supernatant was collected. The kidney homogenates were reduced with DTT and then S-alkylated with iodoacetamide,

followed by dialysis. An aliquot of the protein preparation was digested with lysyl endopeptidase (Wako) and *N*-tosylphenylalanyl chloromethyl ketone (TPCK)-treated trypsin (Thermo Scientific). An aliquot of the digested protein was subjected to hydrophilic interaction chromatography (HILIC) on an Amide-80 column to collect glycopeptides. [20, 21] The bound fraction was evaporated to remove MeCN.

**Collection of Glycopeptides Carrying Fucosylated Glycans by *Aleuria aurantia* lectin (AAL)-Affinity Chromatography.** An aliquot of the glycopeptide fraction was diluted with 10 mM Tris-HCl, pH 7.5 and applied to an AAL-immobilized column (5 mm i.d. × 50 mm, J-Oil Mills). After washing, the glycopeptides were eluted with 5 mM L-(–)-Fucose. To achieve maximal recovery, the flow-through fraction was reloaded onto the same column four times. The recovery rate was not increased during the fourth application. The bound fractions were combined and cleaned by reversed -phase chromatography on a Mightysil RP-18GP column (2 mm i.d. × 50 mm, Kanto Chemical Co., Inc.).

**Deglycosylation and IGOT of *N*-Glycopeptides.** AAL-bound (AAL(+)) *N*-glycopeptides were treated with peptide-*N*-glycosidase F (PNGase, Takara-Bio) in

$^{18}$O-labeled water to remove glycan moieties and to concomitantly label glycosylated Asn residues with the stable isotope as described previously.[20] This step is called IGOT.

**Liquid Chromatography-Mass Spectrometry (LC–MS) Identification of AAL(+) Glycopeptides.** The $^{18}$O-labeled glycopeptides were analyzed by a nanoflow LC–MS system as described previously.[22] Briefly, the IGOT-peptides were separated by the LC system using a reverse phase tip column (150 μm i.d. × 75 mm, C18S, Nikkyo Technos), and the eluted peptides were electrosprayed directly into the LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific). The mass spectrometer was operated in the positive ion mode employing a data-dependent method. Full MS scans were acquired by the Orbitrap analyzer with a resolution of 30000 at m/z 400 in the mass range of m/z 440–1500. The 10 most intense ions were fragmented in a data-dependent mode by collision-induced dissociation (CID) with a normalized collision energy of 35, activation q = 0.25, activation time of 10 ms, and one microscan, and analyzed in the ion trap MS. The following conditions were used: spray voltage, 2.0 kV; ion transfer tube temperature, 250°C; ion selection threshold for MS/MS, 10000 counts; maximum ion accumulation times, 500 ms for full scans; and dynamic exclusion duration, 60 s (10

ppm window; maximum number of excluded peaks, 500).

Proteins were identified by Mascot database search as described previously.[23] Briefly, raw data files were converted to Mascot generic format files using Protein Discoverer software version 1.4 (Thermo Fisher Scientific) and then processed using the Mascot server version 2.4 (Matrix Science) to assign peptides by searching the NCBI mouse RefSeq protein sequence database (30427 entries, downloaded on December 13, 2011). The database-search was performed using the parameters described as follows: fixed modification, carbamoylmethylation (Cys); variable modifications, pyro-Glu (Gln at peptide *N*-terminus), oxidation (Met), and deamidation with $^{18}O$ (Asn, +2.98826 Da); enzyme, trypsin, and LEP; maximum number of missed cleavages, 2; peptide molecular weight tolerance, 7 ppm; peptide charges, +2 to +4; MS/MS tolerances, 0.8 Da; significance threshold, 0.05. False discovery rate, < 0.02. Peptide search results were exported as a CSV file and processed by Microsoft Excel. At first, we selected the peptides with rank 1 and expectation value within 0.05 as "identified peptides." If a prospective "identified peptide" contained one or more $^{18}O$-labeled aspartic acid residues in the consensus tripeptide sequence for *N*-glycosylation

(Asn-Xaa-[Ser/Thr/Cys], where Xaa is any amino acid except Pro), the peptide was accepted as an "*N*-glycopeptide."

**MS Analysis of the *N*-Glycans Released from Glycopeptides.** Two fractions of glycopeptides were collected to compare their *N*-glycomes by MS, that is, (1) whole glycopeptides collected with an Amide-80 column (Amide(+) glycopeptides) and (2) AAL(+) glycopeptides from WT and KO mice. Release and derivatization of *N*-glycans were performed according to the method as reported previously with slight modifications.[24] Briefly, the glycopeptides were treated with PNGase and applied to an OASIS HLB cartridge (Waters). The pass fractions were recovered, and permethylated with methyl iodide. The glycans were collected with a Sep-Pak C18 cartridge (Waters), eluted with 95% MeCN, and lyophilized.

A matrix (2,5-dihydroxybenzoic acid (DHB) (proteomics grade; Wako), 10 mg/mL DHB in 30% ethanol, 0.5 μL) was fixed onto a stainless sample plate for matrix assisted laser desorption/ionization (MALDI)-MS (MTP 384 target plate ground steel BC) (Bruker Daltonics). Then, the sample of permethylated *N*-glycans dissolved with MeCN was spotted onto the dried matrix (DHB). The mass spectrum was obtained by

MALDI-MS (Ultraflex TOF/TOF, Bruker Daltonics) in reflectron positive ion mode. The sample was ionized by nitrogen laser (337 nm, 7 Hz). The ions were accelerated at 20 kV with 400 ns delayed extraction. The spectrum was analyzed using flexAnalysis software (ver. 2.0, Bruker Daltonics). The CID-MS/MS spectra were obtained using MALDI-QIT-TOF MS (AXIMA-QIT, Shimadzu-Biotech) using Shimadzu Biotech Launchpad software (ver. 2.8.5, Shimadzu-Biotech). A nitrogen laser (337 nm, 5 Hz) was used for sample ionization. The DHB was used as a matrix, and it was mixed with the sample as described above. In CID-MS/MS analysis, the CID energy was controlled so that the signal of the precursor ion almost disappeared. Argon was used as a collision gas.

**Site-Specific *N*-Glycome Analysis of AAL(+) Glycopeptides.**

***(1) Assignment of MS Glycopeptide Signals.*** Accurate masses of AAL(+) glycopeptides were obtained by LC–MS analysis using an LTQ-Orbitrap mass spectrometer (Thermo Scientific). To maximize the mass accuracy, masses were measured by an Orbitrap analyzer (positive ion mode) at a resolution of 100000, and with mass locking at 445.12003 for polysiloxane. Simultaneously, MS/MS spectra were

acquired in a data-dependent manner for the two most intense signals by the high-energy collision-induced dissociation (HCD) method (collision energy: 50) using Orbitrap. All MS1 signals with intensities 50000 or higher were filtered from the complete raw data obtained; the list was composed of m/z, intensity, scan number, and retention time for each signal. Among these, the signal peaks were selected by the following data processing steps using an in-house software program: (1) the peaks of each signal that had the highest intensities, along with the scans (retention time), were selected within a specified mass range, e.g., 0.025; (2) the valence of the ion peaks was determined to assign a monoisotope mass value of each peak; and (3) the ion peaks of glycopeptides which had common peptides but different glycans, were selected as a cluster according to the following conditions. First, a pair of glycopeptide peaks must had a close retention time, e.g., within 0.4 min (24 s), and the mass difference between them needed to correspond to the mass of one of the specified monosaccharides, e.g., Hex (mannose, galactose, and glucose), HexNAc (GlcNAc and GalNAc), dHex (fucose), or optionally LacNAc, with an error of 10 ppm. Then, the pairing was extended to form a cluster of glycopeptides; that is, a pair of peaks A and B and a pair of peaks B and C

were connected as a cluster of peaks A, B, and C. To eliminate false assignment by accidental coincidence, the minimum number of members in a cluster was set to 4. At this stage, a number of glycopeptide clusters could be found even without use of MS/MS spectra.

***(2) Prediction of the Core Peptide and Glycan Compositions for Each Glycopeptide Cluster.*** Plausible combinations of a core peptide and glycan compositions of each glycopeptide cluster were predicted using the following simple equation and a set of three mass lists using in-house software: (1) the observed masses of glycopeptides, which were found by the above method as clusters; (2) the theoretical masses of core peptide candidates, which were calculated from the amino acid sequences and modifications of glycopeptides identified by IGOT-LC–MS analyses; and (3) the masses of the monosaccharides (Hex, HexNAc, dHex, and optionally LacNAc).

$$\text{Mglycopeptide}$$

$$= \text{Mpeptide} + \text{Mhex}*X + \text{Mhexnac}*Y + \text{Mdhex}*Z$$

$$= \text{Mglycopeptide} +/- \text{error}$$

An error of ± 10 ppm was allowed for matching. The ranges of numbers for Hex,

HexNAc, and dHex were set to 3–8, 2–8, and 0–4, respectively. Single or multiple combinations of specific peptides with various glycan compositions (*X*, *Y*, and *Z*) were listed for each glycopeptide cluster.

***(3) Confirmation of the Predicted Core peptide and the Site-Specific Glycome.*** To confirm the core predictions, first, the glycome assigned to a cluster was evaluated. The *N*-glycome profile of the whole sample of AAL(+) glycopeptides was obtained by MALDI-MS as described above. Major glycan compositions were listed as a table. If a predicted glycan composition coincided with any one of the listed glycomes, the predicted combination was marked in the prediction results table. The core peptides that got the most marks in a cluster were regarded as the primary candidates of that cluster. Then, the peptide sequences of the primary candidates were inspected manually for the HCD-MS/MS spectrum of the representative glycopeptide(s) of the cluster. If the partial sequence recognized was not coincident with that of the primary candidate, other candidate sequences in the same cluster were checked. If the sequence did not correspond to any one of the predicted candidate peptides, the sequence was searched against the protein sequence database using a BLAST search. After identification of

peptides, the glycome on the searched peptide was calculated. In the case in which the reassigned glycome was considerably different from the whole glycome, the assignment was discarded.

## RESULTS

Fut9 is strongly expressed in the kidney, brain, and stomach.[9] We used the kidneys to identify $Le^x$ carrier proteins. A schematic of the procedure used to identify glycoproteins carrying $Le^x$ is shown in Figure 1. The glycopeptides from trypsin-digested mouse kidneys were prepared by HILIC on an Amide80 column. An aliquot of this glycopeptide eluate was treated by PNGase and analyzed for isolated *N*-glycan structures by MALDI-MS. The remaining glycopeptides were applied to an AAL column to capture fucosylated glycopeptides. Then, the core peptides from the AAL(+) glycopeptide pool were identified by IGOT processing by PNGase digestion, and the isolated *N*-glycans were analyzed to profile the fucosylated glycans. To identify the glycoproteins which possessed Lewis type glycans, we analyzed the AAL(+) glycopeptides by LC–MS, made a list of glycopeptide clusters using site-specific glycome analysis, and by matching the information provided by the IGOT method, identified the glycan profile of individual glycopeptides and identified the carrier proteins of $Le^x$.

**Profiling the *N*-glycan Structures in the Kidneys of WT and KO Mice.** The

profiles of the *N*-glycans on mouse kidney glycoproteins were obtained by MS analysis (Figure 2, Supplementary Table 1a, b). In the figure, the assigned glycan compositions are indicated by digits, presented as XYZ, which represent the numbers of hexose (Hex), *N*-acetylhexosamine (HexNAc), and deoxyhexose (dHex) monosaccharides after eliminating those of the trimannosyl core. If sialic acids (Neu5Ac and Neu5Gc) were identified, these were added and indicated as Ac and Gc, respectively. For example, the glycan composition of the biggest signal in Figure 2a (signal 233) indicates Hex2HexNAc3dHex3. Incorporation of sialic acid can be seen, for example, in 221-Ac1, which is common in both Figure 2a, b.

The profile of WT mice showed *N*-glycans having multiple fucoses in addition to high-mannose glycans (Figure 2a). The glycan composition suggested that most of the *N*-glycans had bisecting GlcNAc and multiple fucoses (Figure 2). This spectrum agreed well with the data available at the Consortium for Functional Glycomics (CFG) homepage (http://www.functionalglycomics.org/fg/). The existence of multiple dHex molecules showed the existence of at least one fucose on the nonreducing end of the glycan, but it could not discriminate whether this structure was $Le^x$ or Lewis a ($Le^a$), or

another structure from the composition. However, because the mouse cannot synthesize $Le^a$, these fucoses were presumed to be $Le^x$. On the other hand, examination of the *N*-glycans of the KO mice showed that the *N*-glycans having multiple fucoses in WT resulted in those having a single fucose in KO mice. This change arising from KO of the *Fut9* gene suggested that one of the multiple fucoses of the WT was a core fucose, and that the others were terminal fucoses, which could be presumed to be $Le^x$. This is strongly supported by the data that mouse kidney does not express α1,2 fucosyltranseferases which are responsible for the synthesis of Fucα1,2 Gal epitope as reported in the BioGPS database, http://biogps.org/#goto=welcome, although Fut9 is abundantly expressed. From the CFG database, it is recognized that almost all complex type *N*-glycans have core fucose. The fact that most of the complexed glycans had multiple fucoses suggested that $Le^x$ existed on proteins globally rather than being limited to specific proteins. In contrast, in the KO mice, the disappearance of terminal fucoses was accompanied by an increase of sialic acids, which were primarily Neu5Gc.

**Identification of AAL(+) Glycopeptides and Their *N*-glycan Profiles.** In general, to identify glycoproteins carrying a specific glycan motif, methods to enrich such

glycopeptides using a specific lectin and antibodies are utilized. However, there is no known method to enrich glycopeptides carrying $Le^x$. Currently, there is no known lectin that specifically binds to $Le^x$. From the results of the structural analysis of kidney *N*-glycans in this study, core fucosylation was found to occur comprehensively on the complex type *N*-glycans. This result accords with the data shown in the CFG database. Therefore, we used AAL to comprehensively capture complex type *N*-glycopeptides, and we analyzed the site-specific glycome to identify $Le^x$-carriers. Because this analysis provides not only a list of $Le^x$-carriers but also the information on glycan alteration status, we selected AAL for comprehensive capturing rather than for specific capturing of $Le^x$-carriers. To attempt this, we enriched fucosylated glycopeptides using an AAL column and identified Lewis-carrying glycopeptides by confirming their glycan compositions. Accordingly, Amide(+) glycopeptides from the kidneys of WT and KO mice were applied to AAL columns and fucose-carrier glycopeptides were collected. A fraction of the AAL(+) glycopeptides was analyzed by IGOT and the core peptides were identified by LC–MS analysis (Table 1, Supplementary Table 2). Almost the same numbers of fucosylated glycopeptides were recovered from both WT and KO mice by

AAL capturing. For WT mice, 831 glycoproteins composed of 2311 glycopeptides were identified by the two analyses. Similarly, for KO mice, 856 glycoproteins composed of 2220 glycopeptides were identified. Of these, 1762 glycopeptides were identified as common between WT and KO mice, and 549 and 458 glycopeptides were identified only from WT or KO mice, respectively. In total, 2769 glycopeptides were identified. The bracketed numbers in Table 1 present the number of nonredundant *N*-glycosylated peptides. The information from these peptides was compiled as the basic data that was used in the subsequent site-specific glycome analysis.

*N*-glycans were isolated from the AAL(+) fraction by PNGase treatment and analyzed by MS after permethylation (Figure 3, Supplementary Table 3a, b). In WT mice, all *N*-glycans contained dHex (fucose), which indicated that the AAL capture was specific to fucosylated glycans (Figure 3a). The main *N*-glycan structures from WT kidneys were bisecting-GlcNAc bound to biantennary and triantennary glycans, and fucoses connected to chitobiose-GlcNAc and to a nonreducing end. The degree of fucosylated peptides was high overall, and a trace amount of sialic acids was also detected. On the other hand, only one fucose was found exclusively in kidneys from KO

mice (Figure 3b). Because all of the terminal fucoses were absent in the KO mice profile, we concluded that the terminal fucose detected in WT kidneys constituted $Le^x$. In addition, from the MS/MS analysis of the major *N*-glycans, it was confirmed that the remaining fucose in the KO mice was the core fucose (Figure 4a, b, Supplementary Table 4). Subsequently, we compared the *N*-glycans having the same glycan structures except for a fucose between the WT and KO mice. MS/MS analysis was applied to signals 2459.5 (WT) and 2284.9 (KO), which shared Hex1HexNAc3 in common (Figure 4a), and to signals 2838.0 (WT) and 2488.9 (KO), having Hex2HexNAc3 in common (Figure 4b). The enlarged picture (circled) in the lower left shows that signal 660.4, a Lewis-type glycan composed of three monosaccharides observed in WT mice, was not present in the KO profile. The remaining single fucose was linked to the innermost GlcNAc of chitobiose as a core fucose from these signals (Figure 4b).

**Identification of the Carrier Proteins of $Le^x$ by Site-Specific Glycome Analysis.**

CID-$MS^n$ , HCD-MS/MS, and electron transfer dissociation (ETD)/ electron capture dissociation (ECD)-MS/MS methods are available to analyze and identify both the peptide moieties and glycan compositions of glycopeptides.[25-28] However, analyses of

the fragment ions generated by the dissociations and the structural identification of both glycans and peptides depend upon complicated manual analysis and cause a low cover rate due to an elevation of the detection limit. Therefore, we devised a method to sort the glycopeptide signals as clusters based on the chromatographic behavior of the glycopeptides and the heterogeneity of glycans. Furthermore, we developed software that could estimate the glycan composition from the differences between the theoretical mass of the existing glycopeptide core and the accurate measured glycopeptide mass, and we utilized this to analyze the site-specific glycome. We note that this analysis was performed on glycopeptide samples in which sialic acid had been removed.

The desialylated AAL(+) glycopeptides of WT and KO mouse kidneys were analyzed by data-dependent LC–MS/MS method: MS was measured by Orbitrap analyzer. The peaks of the peptide signals were extracted from the survey MS1 data (Figure 5a, c). The glycopeptide clusters were searched against the extracted MS data using in-house software in which the number of minimum cluster members was set to 4 (Figure 5b, d). Accordingly, 103 and 141 glycopeptide clusters were assigned for the WT and KO samples, respectively (Figure 6a). The mass of the glycopeptides and the

mass of the core peptides identified previously by IGOT were compared comprehensively using the in-house software. Based on the predicted results, each core peptide and its glycome were confirmed by MS/MS analysis (Supplementary Figure 1). Finally, 72 and 99 clusters comprised of total 522 and 591 signals were identified for the WT and KO samples, respectively (Figure 6a, Supplementary Table 5). For the WT data, 72 clusters were assigned to 51 sites, and 42 out of 51 (82%) sites contained glycopeptides having multiple fucoses in a cluster (Figure 6b), and 24 glycoproteins were identified as $Le^x$ carriers (Figure 6c). In contrast, for the KOs, only two out of 99 clusters (one of 70 sites), i.e. one protein, carried the $Le^x$ structure, meaning that $Le^x$ modifications were generally absent in this sample.

In the WT sample, 24 of the 32 glycoproteins (75%) were identified as $Le^x$ carriers, suggesting that $Le^x$ presence is widespread rather than being limited to specific proteins. In the site-specific glycome data, 35 sites overlapped, and in the 28/35 sites, it was confirmed that the 28 sites contained $Le^x$, but their $Le^x$ were disappeared by Fut9 KO (Figure 6b). As shown in the representative data (Figure 7), the *N*-glycomes containing multiple fucoses in the WT sample were completely absent from the KOs, which agrees

with the result of the *N*-glycome isolated from the AAL(+) fraction, wherein no changes in the profile of glycan structures were observed except for a decrease of fucose (e.g., $Le^x$).

## DISCUSSION

$Le^x$ is a glycan motif (glycan structure) known as an epitope that comprises SSEA-1 and CD15. Fut9 is a key enzyme required for its synthesis that is thought to also bear on various life phenomena as it is expressed in a cell- and developmental stage-specific manner. To elucidate the function of the $Le^x$, it is important to identify the proteins that carry it. Conventional techniques using antibodies have been used to identify $Le^x$-carrier proteins. In the Western blotting of mouse kidney homogenates using the anti-SSEA-1 antibody, we could observe only a few bands, indicating that the SSEA-1 antibody bound to selected glycoproteins among many of the potential $Le^x$ carriers. However, most of the complex-type *N*-glycans contained multiple fucoses (Figure 2a). This suggested that the anti-SSEA-1 antibody might recognize not simply $Le^x$ but also a portion of the proteins carrying $Le^x$, or might have the potential for strong binding to $Le^x$ on only a specific subset of proteins. These results clearly suggest that $Le^x$ is an epitope indispensable for anti-SSEA-1 recognition, but it is not necessarily sufficient.

There have been some reports on successful identification of $Le^x$-carriers using

existing techniques. For example, Liu identified 36 $Le^x$-carrying glycoproteins from Namalwa cells (erythroid cells).[29] However, in that study, because Fut9 was artificially overexpressed, the identification might not reflect physiological conditions. From the physiological expression of $Le^x$ in the native tissue, Hashii reported 10 glycoproteins in the mouse kidney as $Le^x$ carriers, such as megalin (low-density lipoprotein receptor-related protein 2), cubilin, and γ-GTP1 and 18 $Le^x$-containing glycoforms. However, these glycoproteins were considered to be a small part of the glycoproteins having $Le^x$. In our study, the profile of *N*-glycans released from total glycoproteins demonstrated that most of the complex-type *N*-glycans contained multiple fucoses, and we therefore supposed that $Le^x$ is carried on widespread proteins (Figure 2a).

As the comprehensive identification of $Le^x$ carrying proteins is quite difficult by the conventional method using antibody capturing, we designed a new approach that consisted of site-specific glycome analysis of glycopeptides enriched by HILIC followed by AAL affinity column to select glycopeptides having fucoses. Conventionally, it had been difficult to analyze site-specific glycome of glycopeptides without separating the glycan moiety from peptides. The previously developed

IGOT-LC–MS method[20] is powerful and sensitive enough to identify deglycosylated peptides with information on the *N*-glycosylated site in a high-throughput manner. However, it cannot specify which glycan structure is attached to the glycosylated site. In this study, we developed a novel method for identifying the site-specific glycoforms on glycopeptides, on the basis of the information on peptide sequences (theoretical masses) provided by IGOT-LC–MS analysis (Figure 1). Software was developed to combine the results provided by the new method with the list of peptides provided by IGOT-LC‐MS. Another in-house software was also developed to select the glycopeptide signals as clusters.

Prior to the site-specific glycome analysis, we identified the profile of AAL(+) glycopeptides from WT and *Fut9* KO mouse kidneys. In each strain, approximately 2200 fucosylated-glycopeptides (800 fucosylated-glycoproteins) were identified, and 66% (1800 peptides) overlapped between the WT and KO mice. Most of the complex-type *N*-glycans were fucosylated with multiple fucoses in WT (Figure 2a), which was also observed in the profile of the AAL(+) glycopeptides (Figure 3a). In contrast, all of the fucosylated-glycopeptides from the KO mice captured by AAL had a

single fucose that was predicted to represent a core fucose. This suggested that the glycoproteins having multiple fucoses in the WT mouse kidney contained terminal fucoses, which form the $Le^x$ structure, on a widespread variety of proteins. From the disappearance of terminal fucose consequent to knockout of *Fut9* gene, it was confirmed that $Le^x$ is synthesized by Fut9. Furthermore, it could be speculated that almost all of the glycopeptides in AAL(+) fraction that overlapped between WT and KO mice carry $Le^x$ (66% of the total).

Thus, it became clear to us that $Le^x$ exists on a widespread spectrum of proteins rather than only on specific proteins in the mouse kidney. The glycopeptides contained in the AAL(+) fraction also include those carrying a core fucose alone without $Le^x$. Although there were few of these, it was difficult to identify them in this study. Therefore, after performing a glycome comparison using site-specific glycome analysis, comprehensive $Le^x$ addition was observed in 81% of WT proteins, which disappeared in KO mice (Figure 6). These results supported the hypothesis that $Le^x$ is attached comprehensively to kidney glycoproteins. This finding is a new discovery that has not been reported elsewhere. Furthermore, because the terminal fucose competes with the

addition of sialic acid and inhibits sialylation, sialylated *N*-glycans were observed at very low levels in WT mice (Figure 3a). Accordingly, addition of Neu5Gc was enhanced after the disappearance of the terminal fucose facilitated by knockout of the *Fut9* gene (Figure 3b), and there was no apparent alterations of the glycan structures other than sialylation. This strongly indicated that terminal fucosylation and sialylation *in vivo* are competitive for substrates. The terminal fucosylation, $Le^x$, might play a role in keeping kidney cells in a neutral state by protecting against sialylation. This distinctive feature of $Le^x$ glycans might correlate with an essential function of the kidney, but additional studies will be necessary to address this question.

In this study, we used homogenates of whole kidney as our starting material. It has been reported that the glomeruli, which probably occupy less than 10% of the volume of the whole kidney, are highly sialylated.[30-33] In WT kidney, glomeruli are strongly stained, whereas other areas are not stained by *Sambucus nigra* agglutinin (SNA) during lectin staining.[34, 35] As shown in Figure 2a in this study, the percentage of sialylated *N*-glycans was very low. This might correlate with the low occupancy of glomeruli in the kidney. Additional glycome and glycoproteomics analysis using isolated glomeruli

will be necessary in the future to resolve this issue.

Most of the major complex-type *N*-glycans from the kidney contained Lewis-type fucose. In addition, we concluded that the terminal fucose in WT kidney was $Le^x$ because the Lewis type motif disappeared in the KO mice. From the site-specific glycome analysis of the WT sample, we were able to identify 51 glycosylated sites in WT mouse, which included 42 $Le^x$-carrying glycosylation sites of 24 glycoproteins (Table 2). In our study, we reported 317 glycoforms from 24 $Le^x$ carrier proteins (Table 2). Furthermore, most of the proteins identified from the AAL(+) IGOT were also expected to be $Le^x$ carriers. This shows the effectiveness of our glycan heterogeneity-based glycopeptide signal assignment method. By using the novel glycoproteomics technology described in this study, we were able to identify a larger number of $Le^x$ glycoprotein carriers *in vivo* as well as to confirm a greater variety of glycan structures. This is the first report of a large-scale identification of $Le^x$-carrying glycoproteins.

In this study, we identified 2079 nonredundant AAL(+)-glycopeptides (Table 1, Supplementary Table 2) from both WT and KO mice by IGOT-LC–MS analyses as

candidates for the core peptides of glycopeptides; however, only 86 glycosylated sites were assigned by the site-specific glycome analysis by using new glycoproteomic technology (Supplementary Table 5). In general, the ionization efficiency of glycopeptides is significantly lower than that of nonglycosylated/deglycosylated peptides. Furthermore, the mass signals of glycopeptides having a common core peptide sequence but different glycans were dispersed due to the glycan heterogeneity. Therefore, the number of detectable glycopeptides was lower than anticipated. However, considering glycan variation on each glycopeptide, we could reveal 522 glycopeptide signals, 51 glycosylated sites in 24 glycoproteins (Table 2) in WT mice and 591 glycopeptide signals, 70 glycosylated sites in 48 glycoproteins. These numbers are supposed to be the largest that have been reported so far. In the future, improvement of the detection coverage is one of the issues to be addressed.

In conclusion, many proteins in the kidney except for those in the glomeruli are likely to carry $Le^x$. On the other hand, high sialylation has been observed in the glomerular glycan structures.[36] Proximal tubules showed strong positive staining with both anti-$Le^x$ antibody and AAL.[18] We plan to examine the tubular function in the kidney using the

*Fut9* KO mouse. Finally, the site-specific glycome analysis technology that was newly developed herein is applicable to other organs, cells, serum, and other systems. This technology might be a powerful tool for the elucidation of phenotypes of KO mice lacking specific carbohydrate epitopes, as well as for the discovery of new glyco-biomarkers of human disease.

**ACKNOWLEDGMENTS**

I would like to thank Dr. Hisashi Narimatsu and Dr. Hiroyuki Kaji for their assistance in carrying out this research.

And I am deeply grateful to Dr. Akira Togayachi and Dr. Takashi Sato for his assistance and invaluable advice.

Finally, acknowledged to all members of the Research Center for Medical Glycoscience of the National Institute of Advanced Industrial Science and Technology, AIST.

**REFERENCES**


1. Capela, A.; Temple, S. LeX/ssea-1 Is Expressed by Adult Mouse CNS Stem Cells, Identifying Them as Nonependymal. *Neuron* **2002,** *35*(5), 865–875.

2. Solter, D.; Knowles, B.B. Monoclonal Antibody Defining a Stage-Specific Mouse Embryonic Antigen (SSEA-1). *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*(11), 5565–5569.

3. Nishihara, S.; Iwasaki, H.; Kaneko, M.; Tawada, A.; Ito, M.; Narimatsu, H. Alpha1,3-Fucosyltransferase 9 (FUT9; Fuc-TIX) Preferentially Fucosylates the Distal GlcNAc Residue of Polylactosamine Chain While the Other Four Alpha1,3FUT Members Preferentially Fucosylate the Inner GlcNAc Residue. *FEBS Lett.* **1999**, *462*(3), 289–294.

4. Nishihara, S.; Iwasaki, H.; Nakajima, K.; Togayachi, A.; Ikehara, Y.; Kudo, T.; Kushi, Y.; Furuya, A.; Shitara, K.; Narimatsu, H. Alpha1,3-Fucosyltransferase IX (Fut9) Determines Lewis X Expression in Brain. *Glycobiology* **2003**, *13*(6), 445–455.

5. Kaneko, M.; Kudo, T.; Iwasaki, H.; Ikehara, Y.; Nishihara, S.; Nakagawa, S.;

Sasaki, K.; Shiina, T.; Inoko, H.; Saitou, N.; Narimatsu, H. Alpha1,3-Fucosyltransferase IX (Fuc-TIX) is Very Highly Conserved Between Human and Mouse; Molecular Cloning, Characterization and Tissue Distribution of Human Fuc-TIX. *FEBS Lett.* **1999**, *452*(3), 237–242.

6. Kudo, T.; Fujii, T.; Ikegami, S.; Inokuchi, K.; Takayama, Y.; Ikehara, Y.; Nishihara, S.; Togayachi, A.; Takahashi, S.; Tachibana, K.; Yuasa, S.; Narimatsu, H. Mice Lacking Alpha1,3-Fucosyltransferase IX Demonstrate Disappearance of Lewis X Structure in Brain and Increased Anxiety-Like Behaviors. *Glycobiology* **2007**, *17*(1), 1–9.

7. Shimoda, Y.; Tajima, Y.; Osanai, T.; Katsume, A.; Kohara, M.; Kudo, T.; Narimatsu, H.; Takashima, N.; Ishii, Y.; Nakamura, S.; Osumi, N.; Sanai, Y. Pax6 Controls the Expression of Lewis X Epitope in the Embryonic Forebrain by Regulating Alpha 1,3-Fucosyltransferase IX Expression. *J. Biol. Chem.* **2002**, *277*(3), 2033–2039.

8. Kudo, T.; Kaneko, M.; Iwasaki, H.; Togayachi, A.; Nishihara, S.; Abe, K.; Narimatsu, H. Normal Embryonic and Germ Cell Development in Mice Lacking

Alpha 1,3-Fucosyltransferase IX (Fut9) Which Show Disappearance of Stage-Specific Embryonic Antigen 1. *Mol. Cell. Biol.* **2004**, *24*(10), 4221–4228.

9. Kudo, T.; Ikehara, Y.; Togayachi, A.; Kaneko, M.; Hiraga, T.; Sasaki, K.; Narimatsu, H. Expression Cloning and Characterization of a Novel Murine Alpha1, 3-Fucosyltransferase, Mfuc-TIX, that Synthesizes the Lewis X (CD15) Epitope in Brain and Kidney. *J. Biol. Chem.* **1998**, *273*(41), 26729–26738.

10. Edwards, N.J.; Monteiro, M.A.; Faller, G.; Walsh, E.J.; Moran, A.P.; Roberts, I.S.; High, N.J. Lewis X Structures in the O Antigen Side-Chain Promote Adhesion of Helicobacter Pylori to the Gastric Epithelium. *Mol. Microbiol.* **2000**, *35*(6), 1530–1539.

11. Sheu, S.M.; Sheu, B.S.; Yang, H.B.; Lei, H.Y.; Wu, J.J. Anti-Lewis X Antibody Promotes *Helicobacter Pylori* Adhesion to Gastric Epithelial Cells. *Infect. Immun.* **2007**, *75*(6), 2661–2667.

12. Streit, A.; Yuen, C.T.; Loveless, R.W.; Lawson, A.M.; Finne, J.; Schmitz, B.; Feizi, T.; Stern, C.D. The Le(X) Carbohydrate Sequence is Recognized by Antibody to L5, a Functional Antigen in Early Neural Development. *J.*

*Neurochem.* **1996**, *66*(2), 834–844.

13. Allendoerfer, K.L.; Magnani, J.L.; Patterson, P.H. FORSE-1, an Antibody that Labels Regionally Restricted Subpopulations of Progenitor Cells in the Embryonic Central Nervous System, Recognizes the Le(X) Carbohydrate on a Proteoglycan and Two Glycolipid Antigens. *Mol. Cell Neurosci.* **1995**, *6*(4), 381–395.

14. Allendoerfer, K.L.; Durairaj, A.; Matthews, G.A.; Patterson, P.H. Morphological Domains of Lewis-X/FORSE-1 Immunolabeling in the Embryonic Neural Tube are Due to Developmental Regulation of Cell Surface Carbohydrate Expression. *Dev. Biol.* **1999**, *211*(2), 208–219.

15. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological Insights from 108 Schizophrenia-Associated Genetic Loci. *Nature* **2014**, *511*(7510), 421–427.

16. Melnick, D.A.; Meshulam, T.; Manto, A.; Malech, H.L. Activation of Human Neutrophils by Monoclonal Antibody PMN7C3: Cell Movement and Adhesion can be Triggered Independently from the Respiratory Burst. *Blood* **1986**, *67*(5),

1388–1394.

17. Warren, H.S.; Altin, J.G.; Waldron, J.C.; Kinnear, B.F.; Parish, C.R. A Carbohydrate Structure Associated with CD15 (Lewis X) on Myeloid Cells is a Novel Ligand for Human CD2. *J. Immunol.* **1996**, *156*(8), 2866–2873.

18. Comelli, E.M.; Comelli, E.M.; Head, S.R.; Gilmartin, T.; Whisenant, T.; Haslam, S.M.; North, S.J.; Wong, N.K.; Kudo, T.; Narimatsu, H.; Esko, J.D.; Drickamer, K.; Dell, A.; Paulson, J.C. A Focused Microarray Approach to Functional Glycomics: Transcriptional Regulation of the Glycome. *Glycobiology* **2006**, *16*(2), 117–131.

19. Nakayama, F.; Nishihara, S.; Iwasaki, H.; Kudo, T.; Okubo, R.; Kaneko, M.; Nakamura, M.; Karube, M.; Sasaki, K.; Narimatsu, H. CD15 Expression in Mature Granulocytes is Determined by Alpha 1,3-Fucosyltransferase IX, but in Promyelocytes and Monocytes by Alpha 1,3-Fucosyltransferase IV. *J. Biol. Chem.* **2001**, *276*(19), 16100–16106.

20. Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin Affinity Capture, Isotope-Coded

Tagging and Mass Spectrometry to Identify N-Linked Glycoproteins. *Nat. Biotechnol.* **2003**, *21*(6), 667–672.

21. Kaji, H.; Yamauchi, Y.; Takahashi, N.; Isobe, T. Mass Spectrometric Identification of N-Linked Glycopeptides Using Lectin-Mediated Affinity Capture and Glycosylation Site-Specific Stable Isotope Tagging. *Nat. Protoc.* **2006**, *1*(6), 3019–3027.

22. Kaji, H.; Shikanai, T.; Sasaki-Sawa, A.; Wen, H.; Fujita, M.; Suzuki, Y.; Sugahara, D.; Sawaki, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Takahashi, N.; Isobe, T.; Narimatsu, H. Large-Scale Identification of N-Glycosylated Proteins of Mouse Tissues and Construction of a Glycoprotein Database, Glycoprotdb. *J. Proteome Res.* **2012**, *11*(9), 4553–4566.

23. Sugahara, D.; Kaji, H.; Sugihara, K.; Asano, M., Narimatsu, H. Large-Scale Identification of Target Proteins of a Glycosyltransferase Isozyme by Lectin-IGOT-LC/MS, an LC/MS-Based Glycoproteomic Approach. *Sci. Rep.* **2012**, *2*, 680.

24. Ito, H.; Kuno, A.; Sawaki, H.; Sogabe, M.; Ozaki, H.; Tanaka, Y.; Mizokami,

M.; Shoda, J.; Angata, T.; Sato, T.; Hirabayashi, J.; Ikehara, Y.; Narimatsu, H. Strategy for Glycoproteomics: Identification of Glyco-Alteration Using Multiple Glycan Profiling Tools. *J. Proteome Res.* **2009**, *8*(3), 1358–1367.

25. Hashii, N.; Kawasaki, N.; Itoh, S.; Nakajima, Y.; Harazono, A.; Kawanishi, T.; Yamaguchi, T.; Identification of Glycoproteins Carrying a Target Glycan-Motif by Liquid Chromatography/Multiple-Stage Mass Spectrometry: Identification of Lewis X-Conjugated Glycoproteins in Mouse Kidney. *J. Proteome Res.* **2009**, *8*(7), 3415–3429.

26. Scott, N.E.; Parker, B.L.; Connolly, A.M.; Paulech, J.; Edwards, A.V.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S.P.; Højrup, P.; Packer, N.H.; Larsen, M.R.; Cordwell, S.J. Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of *Campylobacter Jejuni*. *Mol. Cell Proteomics* **2011**, *10*(2), M000031–MCP201.

27. Wu, S.W.; Pu T.H.; Viner, R.; Khoo, K.H. Novel LC-MS(2) Product Dependent

Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides. *Anal. Chem.* **2014**, *86*(11, 5478–5486.

28. Håkansson, K.; Cooper, H.J.; Emmett, M.R.; Costello, C.E.; Marshall, A.G.; Nilsson, C.L. Electron Capture Dissociation and Infrared Multiphoton Dissociation MS/MS of an N-Glycosylated Tryptic Peptic to Yield Complementary Sequence Information. *Anal. Chem.* **2001**, *73*(18), 4530–4536.

29. Liu, T.W.; Kaji, H.; Togayachi, A.; Ito, H.; Sato, T.; Narimatsu, H. A Chemoenzymatic Approach Toward the Identification of Fucosylated Glycoproteins and Mapping of N-Glycan Sites. *Glycobiology* **2012**, *22*(5), 630–637.

30. Mohos, S.C.; Skoza, L. Glomerular Sialoprotein. *Science* **1969**, *164*(3887), 1519–1521.

31. Dekan, G.; Gabel, C.; Farquhar, M.G. Sulfate Contributes to the Negative Charge of Podocalyxin, the Major Sialoglycoprotein of the Glomerular Filtration Slits. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*(12), 5398–5402.

32. Gelberg, H.; Healy, L.; Whiteley, H.; Miller, L.A.; Vimr, E. *In vivo* Enzymatic Removal of Alpha 2-->6-Linked Sialic Acid from the Glomerular Filtration Barrier Results in Podocyte Charge Alteration and Glomerular Injury. *Lab. Invest.* **1996**, *74*(5), 907–920.

33. Quaggin, S.E. Sizing Up Sialic Acid in Glomerular Disease. *J. Clin. Invest.* **2007**, *117*(6), 1480–1483.

34. Kakani, S.; Kakani, S.; Yardeni, T.; Poling, J.; Ciccone, C.; Niethamer, T.; Klootwijk, ED.; Manoli, I.; Darvish, D.; Hoogstraten-Miller, S.; Zerfas, P.; Tian, E.; Ten Hagen, K.G.; Kopp, J.B.; Gahl, W.A.; Huizing, M. The Gne M712T Mouse as a Model for Human Glomerulopathy. *Am. J. Pathol.* **2012**, *180*(4), 1431–1440.

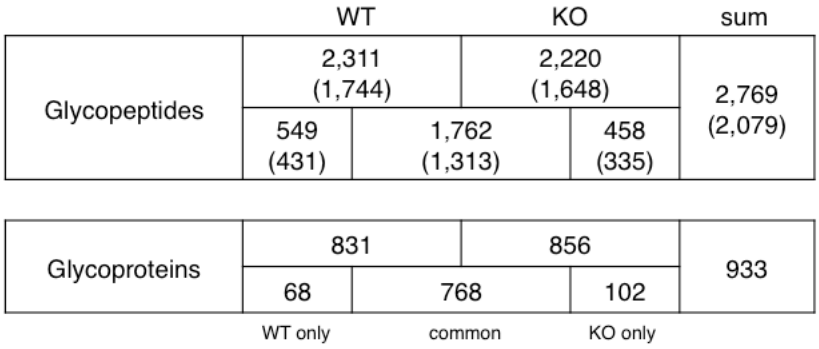
35. Macauley, M.S.; Arlian, B.M.; Rillahan, C.D.; Pang, P.C.; Bortell, N.; Marcondes, M.C.; Haslam, S.M.; Dell, A.; Paulson, J.C. Systemic Blockade of Sialylation in Mice With a Global Inhibitor of Sialyltransferases. *J. Biol. Chem.* **2014**, *289*(51), 35149–35158.

36. Mohos, S.C.; Skoza, L. Glomerular Sialoprotein. *Science*. **1969**, *164* (3887),

1519–1521.

**TABLES**

**Table 1.** Identification of AAL(+) glycopeptides by the IGOT-LC‐MS method.




Glycopeptides carrying fucosylated glycans were identified by the IGOT-LC–MS method followed by a Mascot search. The numbers of identified glycopeptides and glycoproteins of WT and KO mice are summarized. The numbers in parentheses are the number of nonredundant glycopeptide sequences.

**Table 2.** List of $Le^x$-carrying proteins by site-specific glycome analysis.

| $Le^x$ (+) protein | Protein accession | Protein_description | Expression and localization |
|---|---|---|---|
| 1 | gi\|238637279 | 4F2 cell-surface antigen heavy chain isoform b | The basolateral membrane of kidney proximal tubules and small intestine epithelia. |
| 1 | gi\|160333226 | alkaline phosphatase, tissue-nonspecific isozyme precursor | Expressed in liver/bone/kidney (tissue-nonspecific). |
| 1 | gi\|46559389 | angiotensin-converting enzyme isoform 1 precursor | Expressed in apical brush borders of the proximal tubular epithelial cells. |
| | gi\|33468873 | angiotensin-converting enzyme isoform 2 precursor | |
| 1 | gi\|210147589 | beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase | Expressed in proximal tubular cytoplasmical cells. |
| 1 | gi\|6680900 | cadherin-16 isofrom 1 precursor | Expressed in basolateral membranes of renal tubular epithelial cells . |
| | gi\|357933652 | cadherin-16 isoform 2 precursor | |
| | gi\|357933654 | cadherin-16 isoform 3 precursor | |
| 1 | gi\|6753484 | collagen alpha-1(VI) chain precursor | Basement membrane. |
| 1 | gi\|124487348 | cubilin precursor | Expressed in proximal tubule epithelial cells. |
| | gi\|124487348 | cubilin precursor | |
| 1 | gi\|161016833 | dipeptidase 1 precursor | Expressed in apical cell membrane (renal cortex > renal medulla >> renal medulla). |
| 1 | gi\|150378458 | disintegrin and metalloproteinase domain-containing protein 10 precursor | One of a sheddase (membrane-bound enzymes) and expressed in fetal brain, liver, heart, kidney and lung, and in lymphoid tissues. |
| 1 | gi\|31982845 | folate receptor alpha precursor | Expressed in distal renal tubule. |
| 1 | gi\|6679995 | gamma-glutamyltranspeptidase 1 precursor | Located on the outer surface of the cell membrane and widely distributed in mammalian tissues (highest expression in the proximal tubule of the kidney). |
| 1 | gi\|224967126 | granulins precursor | Candidate growth factors (highest levels in kidney). |
| 1 | gi\|45504394 | integrin beta-1 precursor | Cell adhesion molecule. |
| 1 | gi\|124487372 | low-density lipoprotein receptor-related protein 2 precursor | The apical membrane domain of epithelial cells, including proximal tubules. |
| 1 | gi\|31982199 | meprin A subunit alpha | Expressed in renal proximal tubules. |
| 1 | gi\|147901863 | meprin A subunit beta precursor | |
| 1 | gi\|6680552 | napsin-A precursor | An aspartic protease present in the epithelial cells of the lung and kidney. |
| 1 | gi\|31543255 | neprilysin | Expressed in podocytes, renal proximal tubular epithelium. |
| 1 | gi\|82998931 | PREDICTED: beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase-like | Expressed in proximal tubular cytoplasmical cells. |
| 1 | gi\|6753138 | sodium/potassium-transporting ATPase subunit beta-1 | The active enzyme, which catalyzes the hydrolysis of ATP coupled with the exchange of Na(+) and K(+) ions across the plasma membrane. |
| 1 | gi\|119120879 | solute carrier family 3, member 1 | Expressed in cytoplasmic and membrane in tubules. |
| 1 | gi\|225735649 | sulfated glycoprotein 1 isoform A preproprotein | Localized mainly to the epithelial cells of the distal renal tubules or to the parietal epithelial cells of glomeruli. |
| | gi\|225735645 | sulfated glycoprotein 1 isoform B preproprotein | |
| | gi\|225735651 | sulfated glycoprotein 1 isoform C preproprotein | |
| | gi\|225735653 | sulfated glycoprotein 1 isoform D preproprotein | |
| | gi\|225735655 | sulfated glycoprotein 1 isoform E preproprotein | |
| | gi\|225735657 | sulfated glycoprotein 1 isoform F preproprotein | |
| 1 | gi\|251823824 | sushi domain-containing protein 2 isoform 1 precursor | Expressed in cytoplasmic and luminal membrane in tubules. |
| | gi\|251823826 | sushi domain-containing protein 2 isoform 2 precursor | |
| 1 | gi\|227499499 | tubulointerstitial nephritis antigen precursor | Expressed in proximal tubule and interstitium. |

## FIGURE LEGENDS

**Figure 1.** Procedure to identify glycoproteins carrying $Le^x$ in mouse kidney. To identify glycoproteins carrying $Le^x$ structures, first, glycopeptides having fucosylated glycans were enriched by AAL affinity chromatography from kidney homogenates of both WT and KO mice. Then (1) their core peptide sequences were identified by the IGOT-LC–MS method and the glycan portions released from the glycopeptides (glycome) were analyzed by MALDI-MS; (2) To assign $Le^x$ carriers, the glycopeptide fractions were analyzed by LC–MS. Finally, using a set of three mass lists, the site-specific glycome was predicted using an in-house program. The cores of one or more members of each glycopeptide cluster were assigned by manual inspection of their respective MS/MS spectra. Lysyl endopeptidase (Lys C)

**Figure 2.** Mass spectra of permethylated *N*-glycans released from whole (Amide(+)) glycopeptides of WT and KO mice. Mass spectra of the *N*-glycans released from Amide(+) glycopeptides of (a) WT and (b) KO mice are shown. The assigned glycan

compositions are presented by digits (XYZ) indicating the numbers of Hex, HexNAc, and dHex monosaccharides, respectively, after elimination of the trimannosyl core (Hex3HexNAc2). In addition, two kinds of sialic acids, Neu5Ac and Neu5Gc, are indicated as Ac and Gc, respectively. On the basis of their compositions, most *N*-glycans were assigned as being high-mannose (gray) or fucosylated (black boxed) type.

**Figure 3.** Mass spectra of permethylated *N*-glycans released from AAL(+) glycopeptides of WT and KO mice. Mass spectra of the *N*-glycans released from AAL(+) glycopeptides of (a) WT and (b) KO mice are shown. The assigned glycan compositions are presented by digits (XYZ), as indicated in Figure 2. Glycans having multiple fucoses suggesting the presence of $Le^x$ structures are indicated with black boxes.

**Figure 4.** Comparisons of the MS/MS spectra of representative *N*-glycans. *N*-glycans released from the AAL(+) glycopeptides of WT and KO mice were permethylated and

analyzed by MALDI-MS/MS. (a) Comparison of the MS/MS spectra between glycan precursors sharing Hex1HexNAc3 of WT (upper panel, signal 132 of Figure 3a) and KO (lower panel, inverted, signal 131 of Figure 3b) mice. (b) Comparison of the MS/MS spectra between glycan precursors sharing Hex2HexNAc3 of WT (upper panel, signal 233 of Figure 3a) and KO (lower panel, inverted, signal 231 of Figure 3b) mice. The masses of each precursor are shaded in gray. The hypothetical structures of precursors and fragment ions are shown around the signals. The fragments suggesting the presence of $Le^x$ are boxed. The glycans listed in this figure were assigned using the GlycoWorkBench software.

**Figure 5.** Selection of glycopeptide signals as glycopeptide clusters using an in-house program. The AAL(+) glycopeptides were analyzed by a data-dependent LC–MS/MS method: MS was measured by Orbitrap analyzer. All MS signals (intensity >50000) are plotted (with gray squares) (WT: (a), KO: (c)) along with the retention time. Among them, the highest monoisotopic signals of each ion are shown in red. (b), (d) Glycopeptide signals were selected as clusters based on the conditions described in the

Methods by in-house program. Each cluster is shown by verticallyconnected symbol.

**Figure 6.** Detection and core assignment of glycopeptide clusters and identification of glycosylation sites and glycoproteins carrying $Le^x$. (a)Using the in-house program, 103 and 141 glycopeptide clusters were detected for WT and KO mice, respectively. Among them, for 72 (WT) and 99 (KO) clusters, their core peptides were confirmed by MS/MS spectrum inspection. (b) Identification of $Le^x$-carrying glycopeptides by site-specific glycome analysis. For WT, 72 clusters correspond to 51 glycosylated sites, and 42 (=14+28) sites included $Le^x$-containing glycopeptides; for KO, 99 clusters correspond to 70 sites and all but one sites did not contain $Le^x$-containing glycopeptide. (c) Identification of $Le^x$-carrying glycoproteins. For WT, 51 sites correspond to 32 glycoproteins, and 24 (= 3 + 21) were $Le^x$-carrying proteins.

**Figure 7.** Comparison of the site-specific *N*-glycomes of a representative glycosylated site (Asn-489 in cadherin-16) found in both WT and KO mice. The number of monosaccharides (glycan compositions) after eliminating the trimannosyl core is listed.

The presence or absence of the glycan structures at each site in WT and KO mice are indicated with a “+” or “-“, respectively. The symbol (+) including multiple fucoses in WT mice are indicated in gray. Presumed glycan structures having maximum fucosylation in the same glycan skeleton, e.g., Hex1HexNAc3 (dHex1 or 3), are shown in the right columns.

**FIGURES**

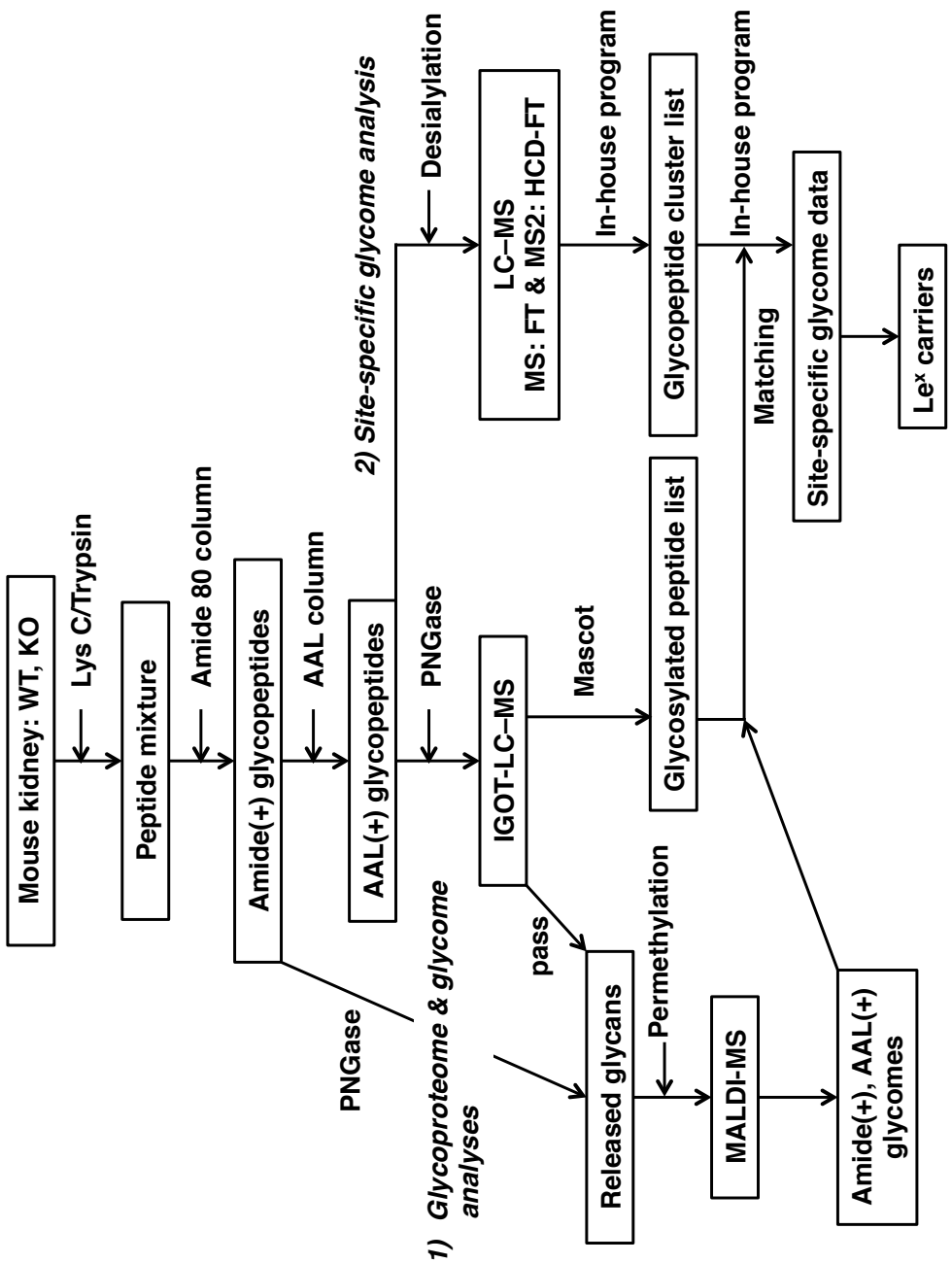



Figure 1. Noro et al.

**Figure 2. Noro et al.**

**Figure 3. Noro et al.**

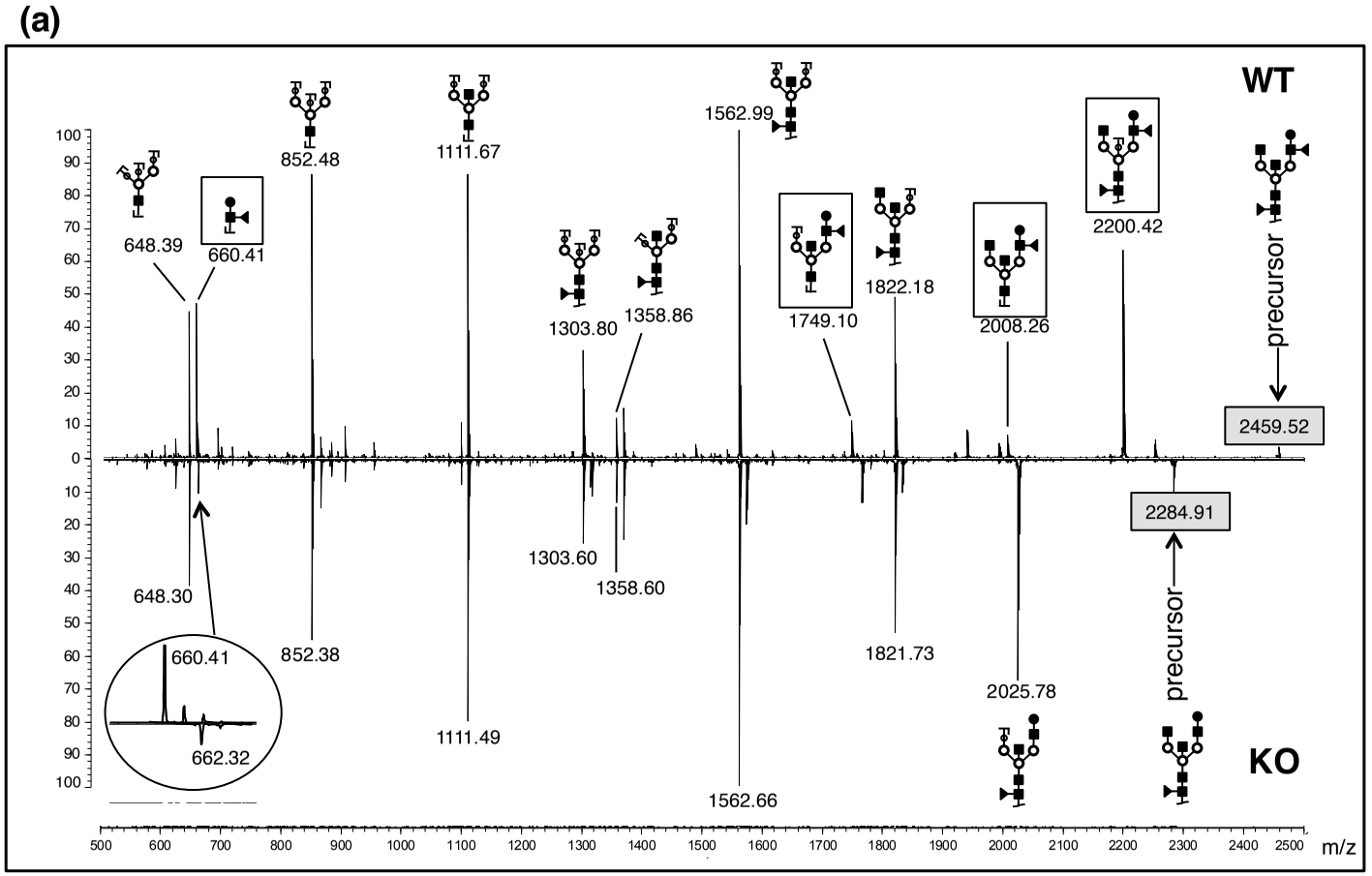



**Figure 4a. Noro et al.**

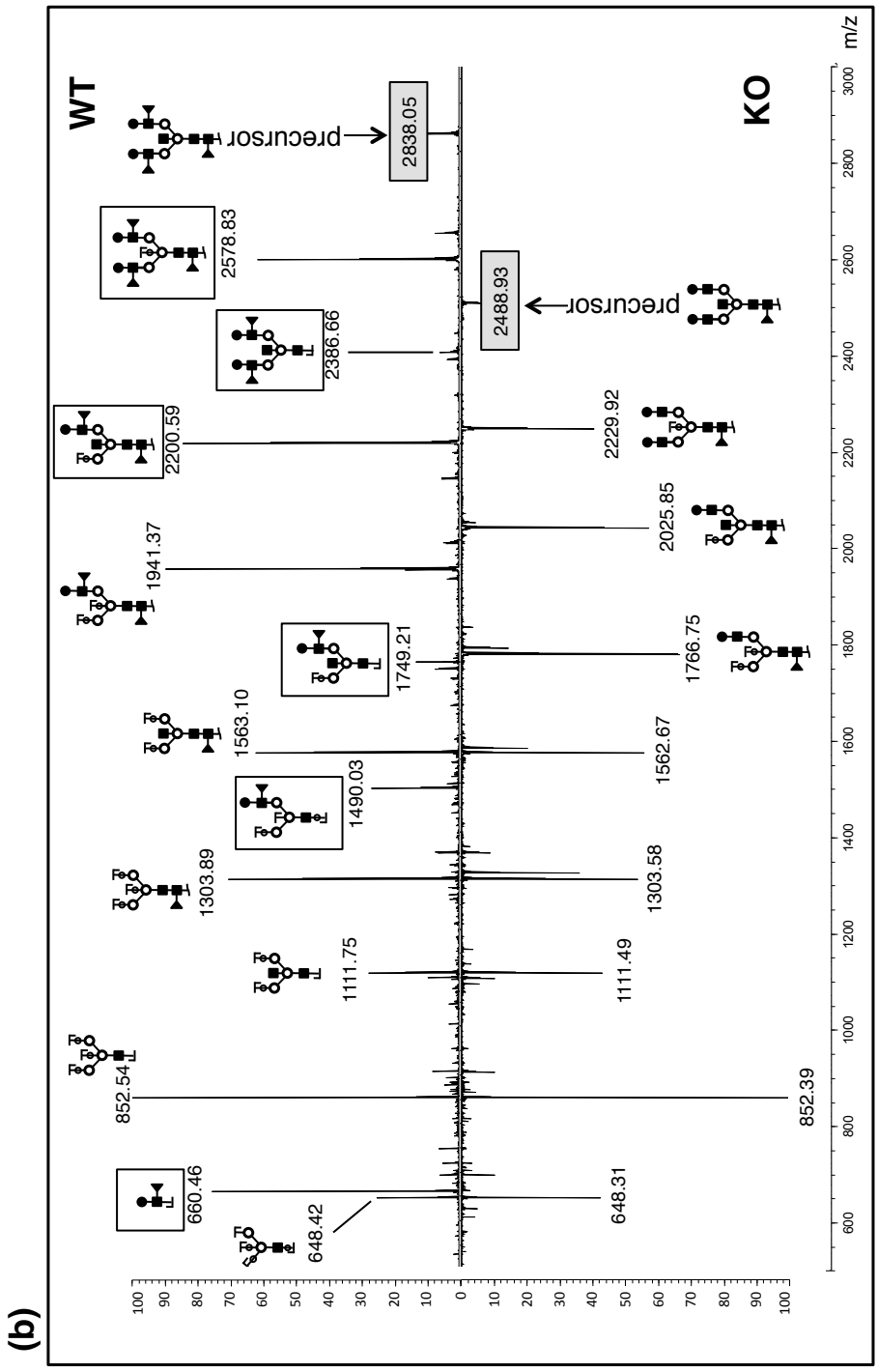



**Figure 4b. Noro et al.**

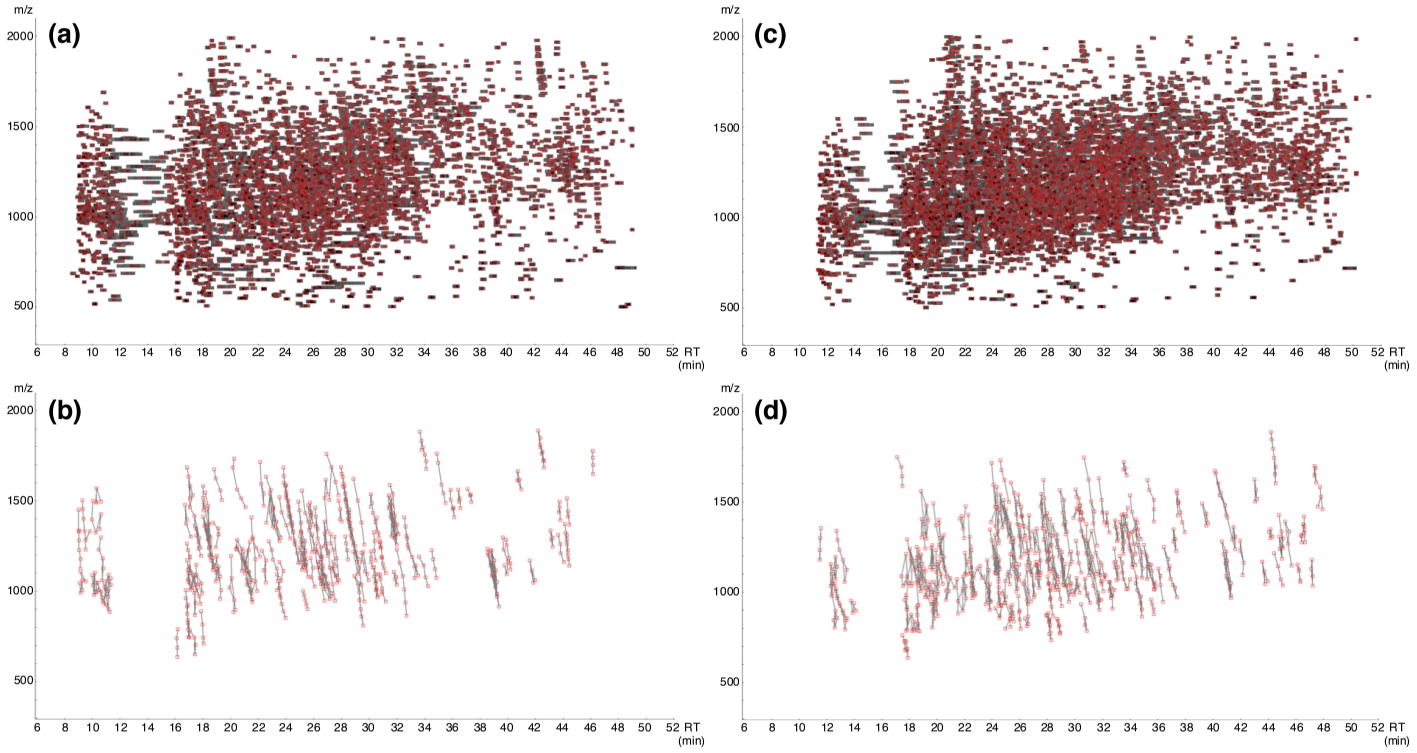



**Figure 5. Noro et al.**

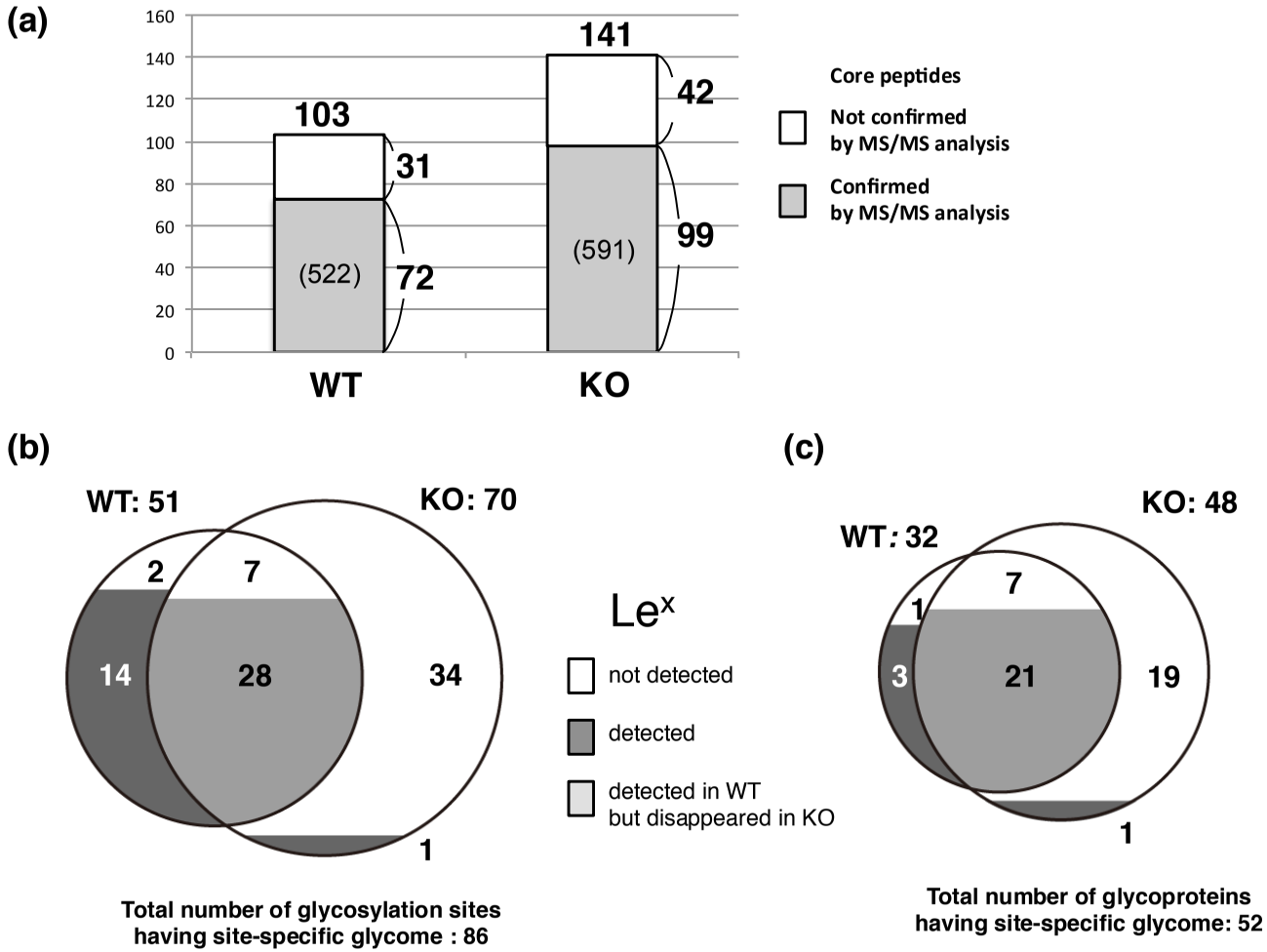
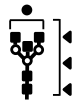
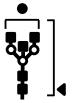
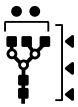
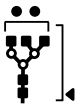



**Figure 6. Noro et al.**

| Protein description | Glycosylated site |
|---|---|
| cadherin-16 isoform 2 precursor | 489(NLS) |

| Glycan composition | | | mouse | | Typical predicted glycome | |
|---|---|---|---|---|---|---|
| Hex [-3] | HexNAc [-2] | dHex | WT | KO | WT | KO |
| 0 | 2 | 1 | + | + | | |
| 1 | 2 | 1 | + | + | | |
| 2 | 2 | 1 | + | + |  |  |
| 2 | 2 | 2 | + | - | | |
| 2 | 2 | 3 | + | - | | |
| 0 | 3 | 1 | - | + | | |
| 1 | 3 | 1 | + | + |  |  |
| 1 | 3 | 2 | + | - | | |
| 1 | 3 | 3 | + | - | | |
| 2 | 3 | 0 | - | + |  |  |
| 2 | 3 | 1 | + | + | | |
| 2 | 3 | 2 | + | - |  |  |
| 2 | 3 | 3 | + | - | | |
| 3 | 3 | 0 | - | + |  |  |
| 3 | 3 | 1 | + | + | | |
| 3 | 3 | 4 | + | - | | |

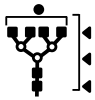
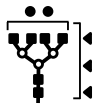
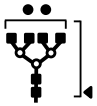
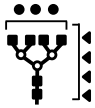
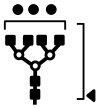
| Glycan composition | | | mouse | | Typical predicted glycome | |
|---|---|---|---|---|---|---|
| Hex [-3] | HexNAc [-2] | dHex | WT | KO | WT | KO |
| 0 | 4 | 1 | - | + |  |  |
| 0 | 4 | 3 | + | - | | |
| 1 | 4 | 1 | - | + |  |  |
| 1 | 4 | 3 | + | - | | |
| 2 | 4 | 1 | - | + |  |  |
| 2 | 4 | 2 | + | - | | |
| 2 | 4 | 3 | + | - | | |
| 3 | 4 | 1 | - | + |  |  |
| 3 | 4 | 2 | + | - | | |
| 3 | 4 | 3 | + | - | | |
| 3 | 4 | 4 | + | - | | |
| 4 | 4 | 1 | - | + | | |
| 3 | 5 | 1 | - | + | | |
| 3 | 5 | 3 | + | - | | |
| 4 | 5 | 1 | - | + |  |  |
| 4 | 5 | 3 | + | - | | |

**Figure 7. Noro et al.**

**Supporting Information**

The Supporting Information is attached to CD-R.

List of masses (*m/z*) of glycopeptides captured by Amide 80 column; list of AAL(+) glycopeptides identified by Mascot search; list of masses (*m/z*) of glycopeptides captured by AAL column; list of mass values of representative fragment ions of glycans; list of $Le^x$-carrying proteins by site-specific glycome analysis; representative MS/MS spectra of glycopeptides to confirm their core peptides and glycan composition (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

* Hisashi Narimatsu, Research Center for Medical Glycoscience (RCMG), National Institute of Advanced Industrial Science and Technology (AIST), Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568, Japan

Tel: +81-29-861-3200, Fax: +81-29-861-3201, E-mail: h.narimatsu@aist.go.jp

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Notes**

The authors declare no competing financial interest.