



OPEN

## Analysis of site and structure specific core fucosylation in liver cirrhosis using exoglycosidase-assisted data-independent LC-MS/MS

Miloslav Sanda<sup>1,4</sup>, Jaeil Ahn<sup>3</sup>, Petr Kozlik<sup>5</sup> & Radoslav Goldman<sup>1,2,4</sup>

Carbohydrates form one of the major groups of biological macromolecules in living organisms. Many biological processes including protein folding, stability, immune response, and receptor activation are regulated by glycosylation. Fucosylation of proteins regulates such processes and is associated with various diseases including autoimmunity and cancer. Mass spectrometry efficiently identifies structures of fucosylated glycans or sites of core fucosylated N-glycopeptides but quantification of the glycopeptides remains less explored. We performed experiments that facilitate quantitative analysis of the core fucosylation of proteins with partial structural resolution of the glycans and we present results of the mass spectrometric SWATH-type DIA analysis of relative abundances of the core fucosylated glycoforms of 45 glycopeptides to their nonfucosylated glycoforms derived from 18 serum proteins in liver disease of different etiologies. Our results show that a combination of soft fragmentation with exoglycosidases is efficient at the assignment and quantification of the core fucosylated N-glycoforms at specific sites of protein attachment. In addition, our results show that disease-associated changes in core fucosylation are peptide-dependent and further differ by branching of the core fucosylated glycans. Further studies are needed to verify whether tri- and tetra-antennary core fucosylated glycopeptides could be used as markers of liver disease progression.

### Abbreviations

S	Serine
T	Threonine
CE	Collision energy
DIA	Data independent acquisition
GP-SWATH	Glycopeptide SWATH
HCV	Hepatitis C virus
HBV	Hepatitis B virus
ALD	Alcoholic
NASH	Non-alcoholic steatohepatitis

N-glycosylation is a frequent posttranslational modification of proteins with significant influence on physiological and pathological processes<sup>1–5</sup>. Fucose (deoxygalactose) can be attached directly to serines (S) and threonines (T) of some proteins<sup>6</sup> and is a common building block of glycans in the N- and O-glycoproteins<sup>7</sup>. Two distinct positional isomers are described in the context of fucosylated human N-glycoproteins. One set of fucosylated isomers, described as outer-arm or terminal fucosylation, is attached to the antennae of the complex type N-glycans by  $\alpha$ -(1–3) and  $\alpha$ -(1–4) linkage to the GlcNAc residues, by  $\alpha$ -(1–2) linkage to galactose, or by  $\alpha$ -(1–2) linkage

<sup>1</sup>Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 20057, USA. <sup>2</sup>Department of Biochemistry and Molecular and Cellular Biology, Georgetown University, Washington, DC 20057, USA. <sup>3</sup>Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington, DC 20057, USA. <sup>4</sup>Clinical and Translational Glycoscience Research Center, Georgetown University, Washington, DC 20057, USA. <sup>5</sup>Department of Analytical Chemistry, Faculty of Science, Charles University, Prague, Czech Republic. ✉email: ms2465@georgetown.edu

to GalNAc<sup>8,46</sup>. The second positional isomer called core fucose is attached by  $\alpha$ -(1–6) linkage to the GlcNAc bound to the protein's asparagine<sup>6</sup>. This positional isomer is formed in glycoproteins by the transfer of fucose from GDP-fucose to the innermost GlcNAc residue exclusively by the  $\alpha$ -(1–6)-fucosyltransferase (Fut8)<sup>9,10</sup>.

Core fucosylation is widespread and all classes of human N-glycans (paucimannose, high mannose, hybrid, complex) are core fucosylated to some extent<sup>11–13</sup>. It is also an impactful modification. Studies of Fut8(–/–) knock-out mice<sup>14</sup> or congenital disorders of glycosylation<sup>15</sup> show that lack of this essential modification leads to severe health consequences. Mechanisms of the pathophysiological processes are often unknown as core fucosylation affects many molecular pathways. Perhaps best defined function of the core fucose is modified affinity of the glycosylated IgG for Fc $\gamma$ RIII<sup>16</sup> which affects immune responses or the function of therapeutic antibodies<sup>17–19</sup>. A number of additional functions of core fucosylation was described in the studies of growth factor signaling<sup>14</sup> or T-cell receptors<sup>20,21</sup> but there is much to be discovered.

Alterations in these pathways were associated with various types of cancer<sup>22–25</sup>, including hepatocellular carcinoma<sup>23</sup>, as well as other diseases<sup>26,27</sup>. It is therefore not surprising that a number of mass spectrometric methods were introduced to study the extent of core fucosylation and its association with the disease processes<sup>28–33</sup>. In this study, we introduce a combination of exoglycosidase digests with data independent (DIA) LC-MS/MS using low collision energy (CE) fragmentation to achieve quantification of the core fucosylated glycopeptides with partially resolved N-glycan structures. We demonstrate applicability of the method to biomedical problems on the relative quantification of the changes in core fucosylation of 45 glycopeptides derived from 18 serum proteins of patients with liver cirrhosis of various etiologies.

## Experimental section

**Materials and reagents.** Acetonitrile (LC-MS grade), water (LC-MS grade), and dithiothreitol (DTT) were obtained from ThermoFisher Scientific (Waltham, MA). Iodoacetamide (IAA) was from MP Biomedicals (Santa Ana, CA). Trypsin Gold (V5280) was obtained from Promega (Madison, WI).  $\alpha$ 2-3,6,8,9 neuraminidase (P0722),  $\alpha$ 1-2 (P0724) and  $\alpha$ 1-3,4 (P0769) fucosidases with GlycoBuffer 1 (5 mM calcium chloride, 50 mM sodium acetate, pH 5.5) were purchased from New England BioLabs (Ipswich, MA). Hemopexin from human plasma was supplied by Athens Research and Technology (Athens, Georgia). Rapigest SF was from Waters (Milford, MA). All other chemicals were obtained from SigmaAldrich (Saint Louis, MO).

**Study population.** Applicability of the method was documented on serum samples of healthy volunteers and patients with cirrhosis of the liver of hepatitis C (HCV), hepatitis B (HBV), alcoholic (ALD) and non-alcoholic steatohepatitis (NASH) etiologies as described previously<sup>34</sup>. Briefly, participants were enrolled in collaboration with the Department of Hepatology and Liver.

Transplantation, Georgetown University Hospital, Washington, DC, under protocols approved by the the MedStar Health Research Institute IRB and carried out according to their guidelines and regulations; all participants provided written informed consent. Blood samples were collected using red-top serum vacutainer tubes (BD Diagnostics, Franklin Lakes, NJ) in line with manufacturer's recommendations and samples were processed within 6 h of the blood draw for storage at  $-80^{\circ}\text{C}$ . We created two pooled samples for each disease category. Each pool represents five participants and we analyzed duplicate samples of each of the pools (4 samples total per group). All the pools were age-matched and had a comparable degree of liver damage, as measured by the model for end-stage liver disease (MELD) score.

**Sample processing.** Affinity depletion of abundant serum proteins was carried out on a HP1290 HPLC system (Agilent, Santa Clara, CA) using the Multiple Affinity Removal Column Human 14 (MARS 14, Agilent) according to manufacturer's protocol. Briefly, human serum (20  $\mu\text{L}$ ) was mixed with buffer A (85  $\mu\text{L}$ ), remaining particulates were removed by centrifugation through a 0.22- $\mu\text{m}$  spin filter 1 min at 16,000  $\times g$ , and the clarified serum was injected on the column. Affinity buffer A and B were used as mobile phases, 100% A at a flow rate 0.25 ml/min at 0–9 min, 100% B at a flow rate 1 ml/min at 9–15 min, and 100% A at a flow rate 1 ml/min at 15–20 min. Collected flow-thru of each sample was precipitated using 80% acetone overnight at  $-20^{\circ}\text{C}$ , washed 2-times with 100% acetone, and dissolved in 50 mM ammonium bicarbonate buffer pH 8 for determination of the protein concentration by BCA assay. Aliquots of the samples (10  $\mu\text{g}$  of total protein) were reduced with 5 mM DTT for 60 min at  $60^{\circ}\text{C}$ , alkylated with 15 mM IAA for 30 min in the dark, residual IAA was quenched with 5 mM DTT, and digested with 2.5 ng/ $\mu\text{L}$  Trypsin Gold (Promega, Madison, WI) at  $37^{\circ}\text{C}$  in Barocycler NEP2320 (Pressure BioSciences, South Easton, MA) for 1 h. Trypsin was deactivated by heating to  $99^{\circ}\text{C}$  for 10 min, samples were evaporated using a vacuum concentrator (Labconco, Kansas City, MO) and digested with 2  $\mu\text{L}$  of  $\alpha$ 2-3,6,8,9 neuraminidase in GlycoBuffer 1 overnight at  $37^{\circ}\text{C}$  according to manufacturer's instructions. Neuraminidase was deactivated by heating at  $75^{\circ}\text{C}$  for 10 min and 2  $\mu\text{L}$  aliquots of each 1–3,4 fucosidase and 1–2 fucosidase were added to the samples for an overnight digestion at  $37^{\circ}\text{C}$ . The digests were evaporated using a vacuum concentrator (Labconco) and reconstituted at a final concentration 0.5  $\mu\text{g}/\mu\text{L}$  in a mobile phase A (0.1% formic acid in 2% acetonitrile) for mass spectrometric analysis.

**Glycopeptide identification by nano LC-MS/MS and selection of transitions.** Following 5 min trapping/washing step in solvent A (2% acetonitrile, 0.1% formic acid) at 10  $\mu\text{L}/\text{min}$ , peptide and glycopeptide separation was achieved at a flow rate of 0.3  $\mu\text{L}/\text{min}$  using the following gradient: 0–3 min 2% B (0.1% formic acid in ACN), 3–5 min 2–10% B; 5–60 min 10–45% B; 60–65 min 45–98% B; 65–70 min 98% B, 70–90 min equilibration at 2% B. We used an Orbitrap Fusion Lumos mass spectrometer to analyze the glycopeptides with the electrospray ionization voltage at 3 kV and the capillary temperature at  $275^{\circ}\text{C}$ . MS1 scans were performed over  $m/z$  400–1800 with the wide quadrupole isolation on a resolution of 120,000 ( $m/z$  200), RF Lens at 40%, intensity

threshold for MS2 set to 2.0e4, selected precursors for MS2 with charge state 3–8, and dynamic exclusion 30 s. Data-dependent HCD tandem mass spectra were collected with a resolution of 15,000 in the Orbitrap with fixed first mass 110 and normalized collision energy 10–30%. LC-MS datasets were processed by Protein Discoverer 2.3. (Thermo Fisher Scientific, Waltham, MA) with Byonic node (Protein Metrics, Cupertino, CA) followed by manual confirmation of the glycopeptides. We selected the glycopeptides based on the following criteria: pair of fucosylated and non-fucosylated glycoforms containing bi, tri and tetra antennary glycan, full tryptic digest, consistent detectability in control and disease samples; LLQ (S/N) greater than 3, fold change greater than 2.

**In-silico Database building.** We build database based on our previous results as well as analytes identified in DDA measured samples. All transition were build based on the expected charged loss of one antenna (366.1) using equation:  $m/z_{\text{quantification}} = (m_{\text{precursor}} - 366.1395)/(z_{\text{precursor}} - 1)$ . Transition were evaluated for quantification limit (LLOQ) based on the S/N ratio,  $S/N > 3$ .

**SWATH DIA quantification by nano LC-MS/MS.** Glycopeptide separation was achieved on a Nanoacquity LC (Waters, Milford, MA) using capillary trap, 180  $\mu\text{m} \times 0.5$  mm, and analytical 75  $\mu\text{m} \times 150$  Atlantis DB C18, 3  $\mu\text{m}$ , 300  $\text{\AA}$  columns (Water, Milford, MA) interfaced with 6600 TripleTOF (Sciex, Framingham, MA). A 3 min trapping step using 2% ACN, 0.1% formic acid at 15  $\mu\text{l}/\text{min}$  was followed by chromatographic separation at 0.4  $\mu\text{l}/\text{min}$  as follows: starting conditions 5% ACN, 0.1% formic acid; 1–55 min, 5–50% ACN, 0.1% formic acid; 55–60 min, 50–95% ACN, 0.1% formic acid; 60–70 min 95% ACN, 0.1% formic acid followed by equilibration to starting conditions for additional 20 min. For all runs, we have injected 1  $\mu\text{l}$  of tryptic digest on column. We have used Data Independent Acquisition (DIA) with one MS1 full scan (400–2000  $m/z$ ) and  $n$  MS/MS fragmentations (800–1600), dependent on the isolation window (15 Da), with rolling collision energy-optimized as described previously<sup>35</sup>. MS/MS spectra were recorded in the range 400–2000  $m/z$  with resolution 30,000 and mass accuracy less than 15 ppm using the following experimental parameters: declustering potential 80 V, curtain gas 30, ion spray voltage 2,300 V, ion source gas1 11, interface heater 180  $^{\circ}\text{C}$ , entrance potential 10 V, collision exit potential 11 V. Precursor and product ion masses and charge states together with the glycopeptide retention times (RT) used for quantification of the glycoforms are summarized in Supplemental Table 1. DDA were processed using Byonic software version 2.12. Data were searched against Uniprot database Human database 4/2017 with 70,709 entries. Protease was set to trypsin—full cleavage, precursor mass accuracy was set to 10 ppm and product mass accuracy to 20 ppm. Fixed modification was set to carbamidomethylation of cysteine, and as variable modification methionine oxidation and N-terminal acetylation.

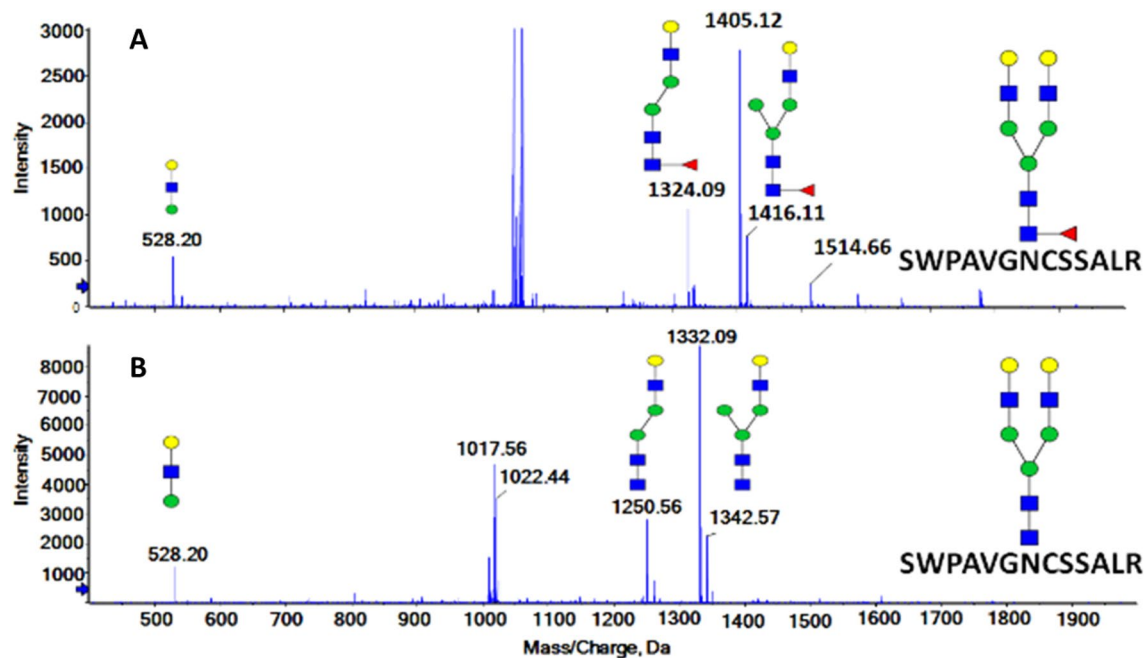
**Data analysis.** Y-ion isotope clusters with isolation window of 1.2 Da, extracted from the low CE SWATH-MS/MS with a 15 Da step window, were used for analysis of the glycopeptide intensities. MultiQuan 2.0 software was used for processing of the quantitative results. Processing methods were created for ion extraction from each SWATH window and for each glycoform. Areas of the fucosylated glycopeptides were normalized to the areas of the corresponding non-fucosylated analytes which removes the need to normalize to protein. Coefficients of variability (CV) were determined from the means and standard deviations of the measurements and were compared across the disease groups and glycoform categories using one-way analysis of variance (ANOVA). Normalized glycopeptide intensities (measured as peak areas) and their ratios were compared in the different disease groups by Kruskal–Wallis one-way ANOVA with statistical significance determined as a two-sided  $p < 0.05$ . The false discovery rate (FDR)  $< 0.05$  was considered statistically significant when multiple testing exists. Statistical analyses were performed using R version 3.40; further data processing and graphing were carried out in Microsoft Excel and Graphpad prism.

## Results

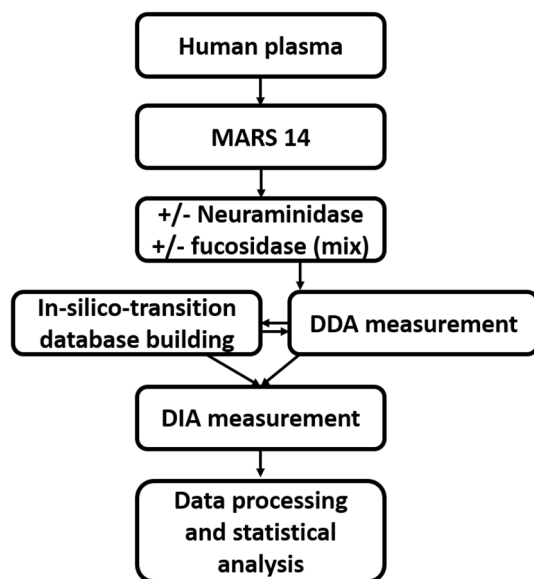
**Glycosidase treatment and identification of the N-glycopeptides in human serum.** Major goal of our study was to evaluate the feasibility of DIA quantification of core fucosylated glycopeptides (GP-SWATH) without an endoglycosidase digest. We do this to evaluate whether changes in the core fucosylation of liver secreted glycoproteins observed in fibrotic liver disease differ with branching of the glycoforms. To this end, we analyzed tryptic digests of MARS14 depleted human serum. The glycopeptides were further simplified by sequential treatment with exoglycosidases ( $\alpha$ 2-3,6,8,9 neuraminidase followed by  $\alpha$ 1-2 and  $\alpha$ 1-3,4 fucosidases) and were subsequently analyzed by LC-MS/MS without any further glycopeptide enrichment. We verified completeness of the exoglycosidase digests ( $> 98\%$ ) by nano HILIC separation of the hemopexin glycopeptide (Supplemental Fig. 1) as described previously<sup>36</sup>. This assures that we quantify core fucosylated glycoforms of each peptide but allows comparison of fucosylation on glycoforms with different degree of branching (see below).

We prepared pooled samples of healthy controls and patients with cirrhosis of the liver caused by HCV, HBV, ALD, and NASH etiologies. For each group, we prepared two pooled samples ( $n = 5$  participants each) which we analyzed first by LC-MS/MS with HCD fragmentation on an Orbitrap Fusion Lumos as described above. We selected for further analysis bi-, tri-, and tetra-antennary glycoforms of 45 N-glycopeptides of 18 proteins detectable in all the samples in their fucosylated and non-fucosylated forms (Supplemental Table 1). In addition to the 23 bi-antennary glycoform pairs (HexNAc4Hex5Fuc/HexNAc4Hex5), our analysis included 13 glycopeptides with tri-antennary (HexNAc5Hex6Fuc/HexNAc5Hex6) glycoforms, and 9 glycopeptides with tetra-antennary (HexNAc6Hex7Fuc/HexNAc6Hex7) glycoforms. In total, we quantified 90 glycopeptides by the DIA LC-MS/MS workflow.

**Core fucosylation and limited fragmentation.** To facilitate the DIA analysis of glycopeptides, we used lower CE settings (half of the CE used for peptides) to minimize interference of the peptides, to minimize rear-

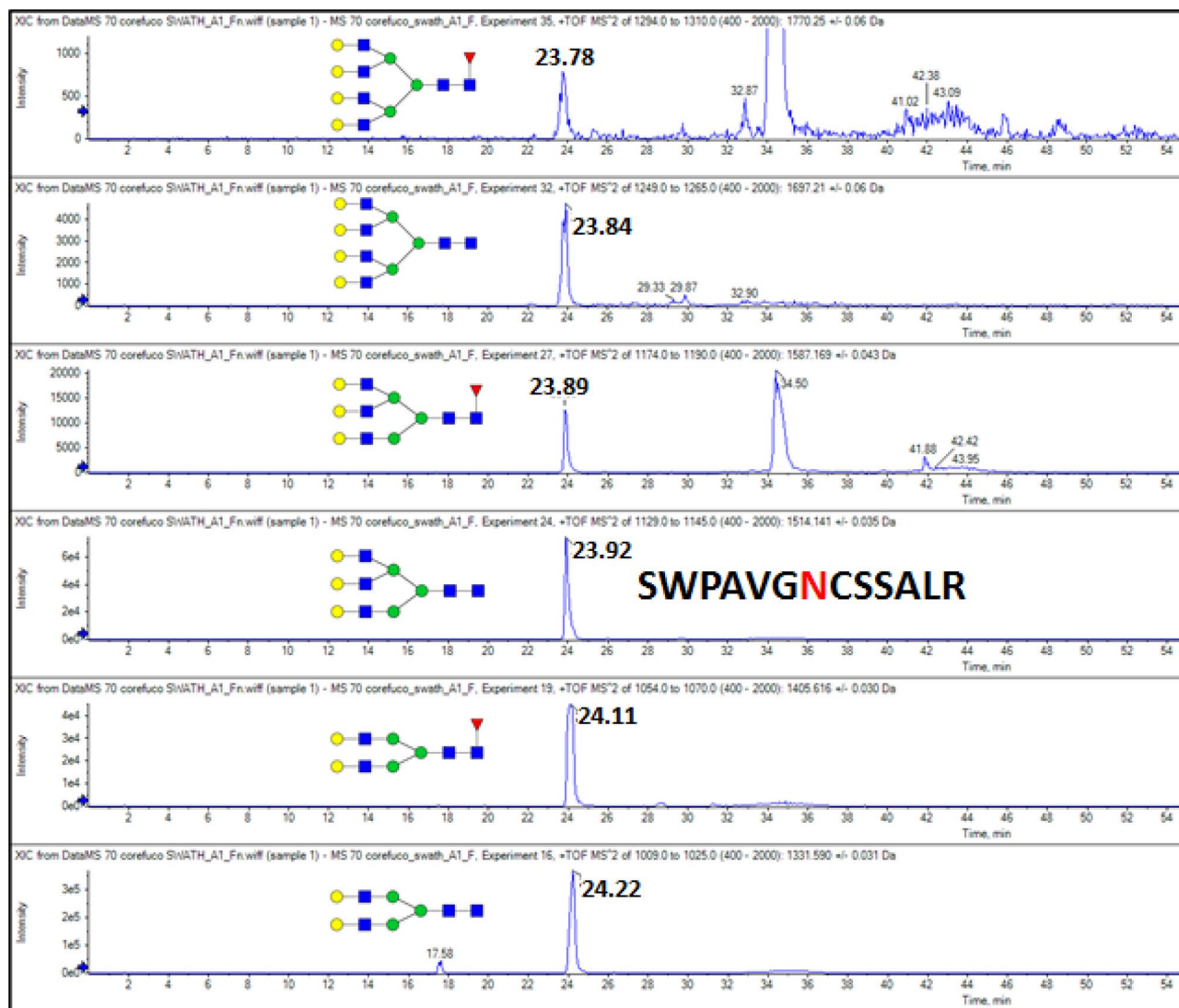


**Figure 1.** Low collision energy spectra from the GP-SWATH DIA of the following glycoforms of the SWPAVGNCSALR glycopeptides of hemopexin: (A) bi-antennary core fucosylated glycoform, precursor  $m/z = 1058.4$ , major quantification fragment  $m/z_{\text{mono}} = 1404.6$ ; and (B) bi-antennary glycoform without fucose, precursor  $m/z = 1009.8$ , major quantification fragment  $m/z_{\text{mono}} = 1331.6$ .



**Figure 2.** Scheme of workflow of exoglycosidase-assisted data-independent LC-MS/MS.

range of fucose and unwanted fragmentation of the glycans, and to maximize the yield of the quantification fragments as described previously<sup>35,37</sup>. Transitions for the quantification of the 90 glycopeptides use the Y-ions corresponding, typically, to the loss of an outer-arm of the glycan; this generates a Y-ion of  $[m/z - 1]$  compared to the precursor and the corresponding oxonium ion. The low CE fragmentation of the core fucosylated and non-fucosylated glycopeptide pairs is quite similar as documented on the spectra of the bi-antennary glycopeptide SWPAVGNCSALR of hemopexin (Fig. 1). The spectra of both the fucosylated (Fig. 1a) and non-fucosylated (Fig. 1b) glycoforms contains Y-ions ( $m/z$  1404.6 fucosylated and  $m/z$  1331.6 non-fucosylated) that represent the loss of an arm (singly charged HexNAc-Hex  $m/z$  366.1) and Y-ions ( $m/z$  1323.6 and  $m/z$  1250.5) from the loss of one arm with manose (singly charged HexNAc-Hex- $m/z$  528.2) visible in the low mass end. The  $m/z$  366.1 oxonium ions are not seen because we acquire the spectra from 400 Da due to optimized ion transmission for the Y-ions. Our complete workflow is described on Fig. 2.



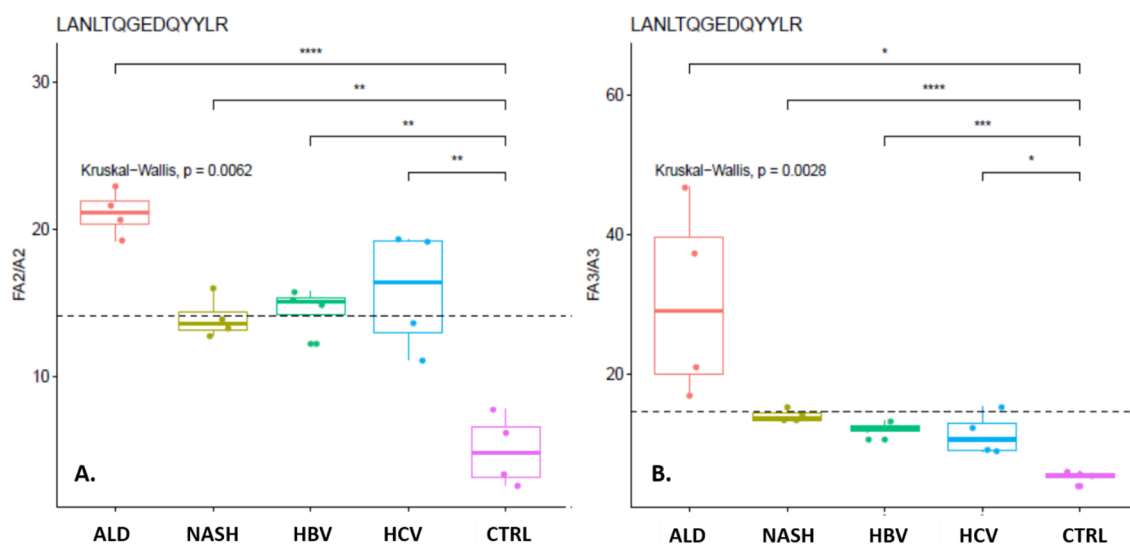
**Figure 3.** XIC chromatogram of six glycoforms of the SWPAVGNCSALR glycopeptide of hemopexin extracted from the GP-SWATH DIA data of a MARS14 depleted serum sample treated with  $\alpha$ 2-3,6,8,9 neuraminidase,  $\alpha$ 1-2 and  $\alpha$ 1-3,4 fucosidases.

The low CE fragmentation allows efficient extraction of the product ions for quantification of the glycoforms as documented on the SWPAVGNCSALR glycopeptide of hemopexin (Fig. 3). The XIC chromatograms show the expected trend of slight decrease in the retention time (RT) of the glycoforms with the addition of each neutral carbohydrate unit as described previously<sup>38</sup>. The RT alignment further confirms the identity of the quantified analytes. The low CE settings also minimize the potential rearrangement of fucose which keeps the error in quantification of the fucosylated glycoform below 15%<sup>37,39</sup> and maximizes the ion intensity sufficiently for reliable quantification of even the low-abundant tetra-antennary glycoforms.

This is documented by the evaluation of the coefficients of variation (CV) across all the 90 glycopeptides quantified; we summarize the results as average CVs in the respective disease groups and glycoform classes (Table 1). We measured every sample in duplicate and we compare the duplicate measurements to derive the technical CVs (Table 1A). The average technical CVs of all fucosylated or non-fucosylated glycopeptides are 16.9% and 18.2%, respectively, and they are not statistically different (one-way ANOVA  $p = 0.925$ ). Our analysis shows that the technical CVs do not differ between the disease groups but the CVs of tetra-antennary nonfucosylated glycoforms are higher compared to the tri- or bi-antennary analytes. Interestingly technical CVs of fucosylated tetraantennary glycoforms does not follow this trend even if they are more abundant than nonfucosylated and influence of noise could be lower. As expected, the biological replicates (two different samples per group) show similar CVs with average CV 23.3% for the fucosylated glycopeptides compared to the average CV of 22.0% for the non-fucosylated glycopeptides (one-way ANOVA  $p = 0.216$ ). We do not observe a significant differences in the biological CVs by disease group (one-way ANOVA  $p = 0.091$ ) or by the number of arms (one-way ANOVA  $p = 0.445$ ). Higher biological variability, observed especially in the disease groups, is expected and is minor compared to the fold-changes observed between the healthy controls and the cirrhotic disease groups (see below).

A. Technical CV					
Alcoholic	HBV	HCV	NASH	Control	Glycan
16.39	15.92	17.95	13.01	14.24	FA2
17.46	21.84	19.60	14.34	20.40	FA3
16.71	17.93	17.44	14.27	16.90	FA4
15.23	13.95	13.27	11.75	13.97	A2
15.14	22.74	16.06	11.35	14.96	A3
24.00	30.16	22.55	24.85	23.33	A4
B. Biological CV					
Alcoholic	HBV	HCV	NASH	Control	Glycan
28.25	24.99	32.91	27.34	23.89	FA2
17.98	27.58	31.33	13.49	13.85	FA3
29.00	16.37	22.66	7.37	20.09	FA4
17.28	33.28	19.50	16.37	16.76	A2
17.51	30.78	24.25	14.00	15.40	A3
23.11	27.88	26.50	13.93	11.96	A4

**Table 1.** Distribution of the average coefficients of variability (CV) for the indicated glycopeptide classes and disease groups across the entire study set. Duplicate measurements were performed for both the (A) technical and (B) biological CV measures.



**Figure 4.** Changes in the relative abundance of the (A) bi-antennary and (B) tri-antennary core fucosylated glycoforms of the LANTQGEDQYYLR peptide of clusterin in healthy controls and patients with liver cirrhosis of the indicated etiologies. Statistical significance; \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$ , \*\*\* indicates  $p < 0.005$ , and \*\*\*\* indicates  $p < 0.001$ .

**Trends in the abundance of fucosylated and non-fucosylated glycopeptides.** The core fucosylated glycopeptides show a general trend to increased intensities in the cirrhotic groups compared to the healthy controls (Supplemental Table 1). When we compare peak areas of the same fucosylated glycoforms in the cirrhosis groups to the healthy controls, 23 of 45 glycopeptides have higher intensity in cirrhosis; 6 of the glycopeptides show significantly higher intensity in all four cirrhosis categories ( $p < 0.05$ ) compared to the healthy controls and all of them have intensity consistently more than twofold higher. Only 2 core fucosylated glycopeptides show consistently lower intensities in the four cirrhosis groups compared to the healthy controls. The highest difference in the peak areas of the fucosylated glycopeptide was observed for the tetra-antennary glycoform of the ELHHLQEQNVSNFLDK glycopeptide of ceruloplasmin which is on average 7.5 fold higher in the cirrhosis groups. One of the consistently and significantly changed core fucosylated glycopeptides is the biantennary and triantennary LANTQGEDQYYLR peptide of clusterin (Fig. 4) which was previously described in liver disease by our group<sup>33</sup> and by others (panel of analytes in patent US20090208926A1).

At the same time, we observe a trend to lower intensities of the non-fucosylated glycopeptides in the cirrhosis groups compared to the healthy controls (Supplemental Table 1). We observe a consistently lower intensity of 14 glycopeptides in the four cirrhosis groups ( $p < 0.05$ ) and one of them has intensities decreased more than twofold.

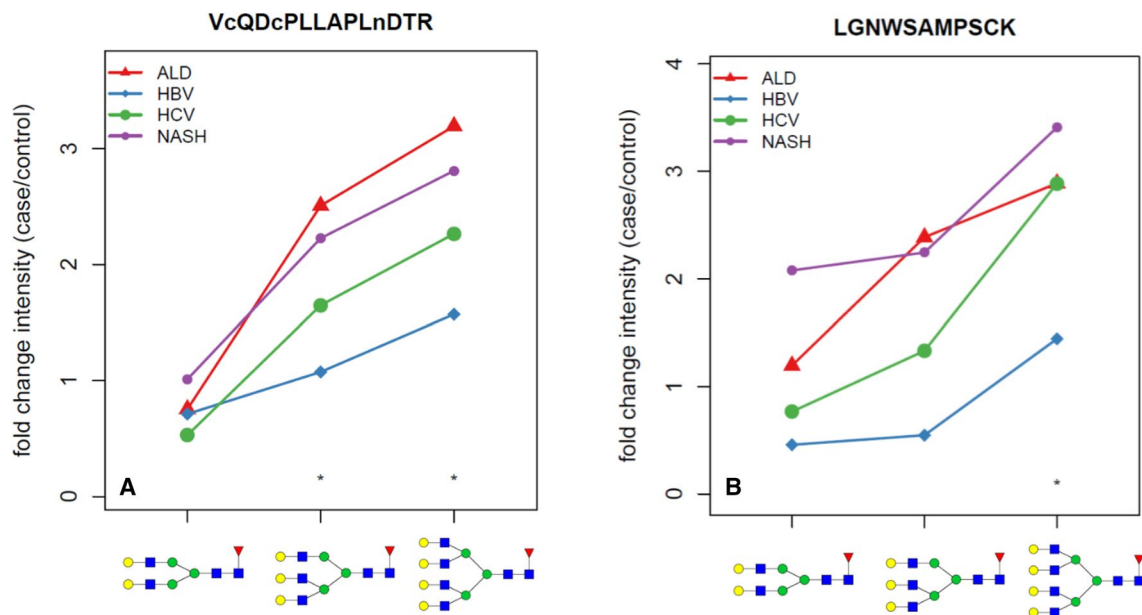
Peptide	Accession Number	Glycan ratio	Healthy	ALD	HBV	HCV	NASH	p-value	FDR
ELHHLQEQNVSN AFLDK	P00450	FA2/A2	0.738	3.52	2.456	2.526	2.479	0.026	0.034
ELHHLQEQNVSN AFLDK	P00450	FA3/A3	19.325	95.319	79.132	52.152	50.825	0.002	0.010
ELHHLQEQNVSN AFLDK	P00450	FA4/A4	33.133	152.706	190.84	106.261	101.359	0.003	0.010
LANLTQGEDQYYLR	P01042	FA2/A2	5.022	21.101	14.544	15.818	14.005	0.006	0.012
LANLTQGEDQYYLR	P01042	FA4/A4	4.428	21.234	12.066	8.596	8.51	0.004	0.01
LGNWSAMPSCCK	P02751	FA4/A4	26.469	118.536	259.891	40.991	37.053	0.022	0.03
LDAPTNLQFVNETD-STVLVR	P02751	FA2/A2	9.33	29	38.24	28.22	35.39	0.031	0.039
LGNWSAMPSCCK	P02751	FA3/A3	13.522	93.666	19.97	31.225	33.208	0.007	0.013
LPTQNITFQTESS-VAEQEAEFQSPK	Q14624	FA4/A4	35.121	155.44	395.277	107.626	72.535	0.002	0.01
QVFPGLNYCTSGAYS-NASSTDSASYPLTGD	P04114	FA2/A2	0.749	6.845	3.681	5.007	4.264	0.015	0.023
VDKDLQSQLEDIL-HQVEnK	P02671	FA2/A2	21.869	66.857	44.836	81.947	67.083	0.007	0.013

**Table 2.** Normalized intensities of the core fucosylated glycoforms in healthy controls and in patients with cirrhosis of the liver of alcoholic (ALD), hepatitis B viral (HBV), hepatitis C viral (HCV), and non-alcoholic steatohepatitis (NASH) etiologies for glycopeptides with ratio of the glycoforms increased more than threefold in at least two categories of the cirrhotic liver disease compared to the healthy controls. FA<sub>n</sub>/A<sub>n</sub>: ratio of the fucosylated glycopeptide to the nonfucosylated glycopeptide, n = number of antennae. FDR represents the false discovery rate.

Only 3 non-fucosylated glycopeptides show significantly higher intensity in at least 2 of the cirrhosis groups. This trend reflects to some degree the reported trends in the serum concentration of the proteins in liver cirrhosis. We observe lower intensities in the non-fucosylated glycopeptides of proteins (hemopexin, clusterin, ITIH4, fibrinogen) reported to decrease in cirrhosis while the non-fucosylated glycopeptide of alpha-2 macroglobulin has higher intensity in cirrhosis in line with its increased serum concentration<sup>40–42</sup>. However, additional factors affect the abundance of the non-fucosylated glycopeptides because some glycoforms of the same peptide show inconsistent trends. For example, the tri- and tetra-antennary glycoforms of some glycopeptides have higher intensity in cirrhosis while their bi-antennary glycoforms are lower. This trend may be potentially explained by increased branching in the cirrhosis groups. This is likely the case of the ELHHLQEQNVSN AFLDK glycopeptide of ceruloplasmin where the tetra-antennary non-fucosylated glycoform has higher intensity in the cirrhosis groups while the bi-antennary glycoform is slightly lower in cirrhosis than in the healthy controls.

These trends suggest that the abundance of the core fucosylated glycoforms of some proteins goes up in cirrhosis in spite of the fact that the protein's concentration may decrease. It is also a reason why we normalize the fucosylated glycoforms by their non-fucosylated counterparts and consider ratio as the two glycoforms as our final measure of interest.

**Structure specific changes of core fucosylation in liver cirrhosis.** We find multiple core fucosylated glycoproteins with significantly higher intensities of fucosylated glycopeptides in the serum of some category of the cirrhotic disease. Core fucosylated glycoforms (normalized intensity) of 11 glycopeptides of these proteins increase significantly (FDR < 0.05) threefold or greater in at least two categories of liver cirrhosis compared to the healthy controls (Table 2). Here we compare the normalized fucosylated intensities represented by the ratios of the same fucosylated/non-fucosylated glycoforms of a glycopeptide (i.e. bi-antennary fucosylated/bi-antennary non-fucosylated etc.). Overall, we observe the highest (5- to ninefold) increase in the normalized fucosylated intensity for the bi-antennary glycoform of the QVFPGLNYCTSGAYSNASSTDSASYPLTGD glycopeptide of apolipoprotein B-100; the increase is significant in all the cirrhosis groups compared to the healthy control group. However, we also observe decrease in the normalized fucosylated intensities for some glycopeptides. For example, the bi-antennary FGCEIENNR glycopeptide of Zn-alpha2-glycoprotein has significantly lower normalized fucosylated intensities in the cirrhosis groups (Supplemental Table 1) in line with our previous reports of site- and structure-specific changes in the fucosylation of serum proteins in the context of liver cirrhosis<sup>35,43</sup>. In addition, we observe an interesting trend towards higher intensities of branched core fucosylated glycoforms for several peptides. For example, the fucosylated intensities of the tri-antennary glycoform of the SWPAVGNCSSALR peptide of hemopexin increase in cirrhosis consistently 2- to fivefold while the intensities of the bi-antennary glycoforms increase 1- to threefold. Equally interesting are the trends of the bi-, tri- and tetraantennary glycoforms of the VCQDCPLLAPLNDTR peptide of Alpha-2-HS-glycoprotein and the LGNWAMPSCCK glycopeptide of beta-2-glycoprotein (Fig. 5). For all three glycopeptides, the increase in the intensity of the fucosylated analyte goes up with branching. We do not know the reason for this observation; it could be related to different rates of synthesis or to the post-synthetic selection of the fucosylated glycoforms. Nonetheless, the results show clearly that the presented DIA LC-MS/MS workflow enables quantification of the partially resolved



**Figure 5.** Average ratio of the increase in the intensity of the following core fucosylated glycopeptides in the liver cirrhosis groups of four etiologies compared to the healthy controls: A. bi-, tri- and tetra-antennary glycoforms of the VCQDCPLLAPLNDR peptide of Alpha-2-HS-glycoprotein; B. bi-, tri-, and tetra-antennary glycoforms of the LGNWSAMPSCCK peptide of beta-2-glycoprotein.

fucosylated structures which would not be achieved with the commonly used truncation of the glycoforms with endoglycosidases.

## Discussion

Knowledge of the extent of core fucosylation of human proteins substantially expanded in recent years<sup>28–33</sup>. The reported studies use a variety of enrichment strategies and mass spectrometric workflows to improve coverage of the site-specific core fucosylated proteins but quantification of the core fucosylated analytes remains challenging<sup>28–33</sup>. This is due to some extent to the complex nature of quantification of the glycoforms at specific sites of attachment to proteins<sup>44</sup> but also due to the fact that synthetic isotopically labeled standards are not available for the quantification of glycopeptides. The studies therefore focus on relative quantification using iTRAQ or TMT labeling<sup>30–32</sup> which is often combined with truncation of the N-glycans with endoglycosidases<sup>29,31</sup> to improve identification and relative quantification of the truncated (peptide with GlcNAc with/out fucose) analytes. This is, however, by no means straightforward as the efficiency of cleavage of the iTRAQ or TMT tags is affected by the presence of complex glycans and because the rate of cleavage of some types of complex glycans (fucosylated, branched) by endoglycosidases differs substantially (some are not appreciably cleaved). As a consequence, we know of several thousand sites modified to some extent by core fucosylated N-glycans but we know their relative intensities to a limited extent<sup>28–33</sup>.

In this study, we introduce a workflow that eliminates the endoglycosidase digest and extends the glycopeptide SWATH (GP-SWATH) analyses, that we introduced recently<sup>35</sup>, to relative quantification of the core fucosylated glycopeptides. We combine exoglycosidase digests with DIA LC-MS/MS using low CE fragmentation which enables quantification of the core fucosylated glycopeptides with partially resolved N-glycan structures, namely branching into bi-, tri-, and tetra-antennary glycoforms. We focus on the optimization of a DIA LC-MS/MS workflow that quantifies pairs of glycopeptides with/out core fucose and removes the need to normalize the glycopeptides to protein concentration. We use the previously described example of altered core fucosylation in fibrotic liver disease<sup>33–35</sup> to document the applicability of our workflow to a relevant biomedical problem. This is important because the knowledge of the core fucose distribution on glycoproteins by their branching is limited.

We use exoglycosidases ( $\alpha$ 2-3,6,8,9 neuraminidase, 1–3,4 fucosidase and 1–2 fucosidase) because we focus on the core fucose. The outer-arm modifications dilute the MS signal into multiple low abundant analytes which may have overlapping isotope clusters<sup>45</sup>; removal of the outer-arm fucose also simplifies data interpretation. Low CE fragmentation is an important component of the workflow because it minimizes interferences of peptides (which remain to a large extent unfragmented), eliminates excessive fragmentation of the glycans (Fig. 1) and potential rearrangement of fucose as described recently<sup>37,39</sup>. This set of optimizations makes the glycopeptide Y-ions sufficiently selective<sup>34,35</sup> for reliable quantification of the branched (bi- to tetra-antennary) core fucosylated glycopeptides by the GP-SWATH DIA LC-MS/MS. Our approach eliminates the need for endoglycosidase H (F) treatment and issues related to the limited cleavage of certain types of N-glycans, especially the branched multi-antennary glycans, by the endoglycosidases. We selected a set of analytes, including representative glycopeptides with branched multi-antennary structures, for our DIA-LC-MS/MS quantification of the site and structure specific core fucosylated glycoforms. We focused on the optimization of the quantitative DIA workflow



and the breadth of coverage was not our goal. Our results document relative quantification of 45 glycopeptide pairs (with/out fucose) with average technical CV around 17% (Table 1) and demonstrate meaningful biological assessment of differences in the core fucosylation of serum glycoproteins in liver cirrhosis of various etiologies.

We observe typically higher core fucosylation in liver cirrhosis compared to the healthy controls; a typical increase is up to twofold but we observed up to eightfold increased intensity of the fucosylated glycoform in case of the tetra-antennary ELHHLQEQNVSN AFLDK glycopeptide of ceruloplasmin, depending on the etiology of the cirrhosis. The increase in core fucosylation is more pronounced in the groups of cirrhosis of ALD (average: 1.79) and NASH (average 1.88) etiologies (15 and 21 of the 45 glycopeptides with significantly higher core fucosylation compared to the healthy controls, respectively) as opposed to HBV and HCV etiology (8 and 12 glycopeptides). At the same time, we observe a trend towards lower intensity of the non-fucosylated glycoforms in cirrhosis which is likely related, to some degree, to the decrease in concentration of some of the proteins in the cirrhotic patients. It is well known that synthesis of proteins in the liver is affected by the progressing fibrotic remodeling. We therefore normalize the intensity of the fucosylated glycoforms and use the ratio of fucosylated to the non-fucosylated forms as the quantified readout (Table 2).

Some bi-antennary glycopeptides decrease in cirrhosis compared to healthy controls but increase in cirrhosis in the tri- or tetra-antennary forms of the same peptide. This cannot be explained by lower protein concentration but could perhaps be attributed to increased branching; further studies will be needed to define whether this originates at synthesis or post-synthetic selection. The VC(cam)QDC(cam)PLLAPLNDTR glycopeptide of Alpha-2-HS-glycoprotein is a good example with higher contribution of tetra-antennary glycoforms in the cirrhosis groups while the bi-antennary and triantennary glycoforms remain the same. Relative contribution of the bi-, tri and tetra-antennary structures in the healthy group for this peptide is 18%, 78% and 4%, respectively and the ratio of the glycoforms in the cirrhotic groups compared to the healthy group range from 0.71 to 1.04 for the bi-antennary glycopeptides, 0.98 to 1.05 (for tri-antennary, and 1.20 to 1.40 for the tetra-antennary glycopeptides. This is true for several glycopeptides (for example LGNWSAMPSCK), and supports the usefulness of our analytical approach where we use increase of core fucosylation and increase of glycan branching to simultaneously. The greater than fourfold increases of the normalized intensities of some of the glycopeptides in cirrhosis suggests that the core fucosylated glycopeptides might be useful in serologic monitoring of liver fibrosis; this hypothesis is, however, by no means confirmed by our preliminary results and needs to be further evaluated.

Our study has several weaknesses. We rely on the quantification of relative intensities using a label-free DIA workflow because we do not have synthetic isotopic standards of the glycopeptides. Nonetheless, the workflow has an average technical CV of 17.5% and shows consistent trends which suggests good performance. In addition, we do not have yet an automated pipeline for the analysis of the GP-SWATH DIA datasets which will be needed to improve the throughput of the studies; this is an active topic of our ongoing studies.

In summary, we introduce an exoglycosidase-assisted GP-SWATH analysis of core fucosylated glycopeptides with resolved branching of the N-glycans. The main advantage of our workflow is higher sensitivity compared to a DDA workflow, especially in the complex matrix. In addition, we resolve structure-specific quantities of the core fucosylated glycopeptides, including the multi-antennary glycoforms. We show that the DIA LC-MS/MS workflow identifies meaningful differences in the relative quantities of site-specific core fucosylated glycopeptides in a preliminary study of liver cirrhosis. The data suggest that the degree of core fucosylation and its disease-associated alterations differ by branching of the N-glycans as documented on the case of ceruloplasmin or hemopexin. We expect that the workflow will enable quantitative studies of core fucosylation in biological model systems and in liquid biopsies such as serum, urine, or cerebrospinal fluid (CSF).

## Data availability

The mass spectrometry data have been deposited to the jPOST repository JPST000928 and <https://repository.jpostdb.org/preview/12562715376144958ca7b7d> Access key: 3189.

Received: 19 April 2021; Accepted: 3 November 2021

Published online: 02 December 2021

## References

- Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**(5512), 2370–2376 (2001).
- Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**(5512), 2364–2369 (2001).
- Cummings, R. D. & Pierce, J. M. The challenge and promise of glycomics. *Chem. Biol.* **21**(1), 1–15 (2014).
- Hudak, J. E. & Bertozzi, C. R. Glycotherapy: new advances inspire a reemergence of glycans in medicine. *Chem. Biol.* **21**(1), 16–37 (2014).
- Varki, A., & Gagneux, P. Biological Functions of Glycans. **2015**, 77–88.
- Schneider, M., Al-Shareffi, E. & Haltiwanger, R. S. Biological functions of fucose in mammals. *Glycobiology* **27**(7), 601–618 (2017).
- Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **15**(6), 346–366 (2019).
- Li, J., Hsu, H. C., Mountz, J. D. & Allen, J. G. Unmasking fucosylation: From cell adhesion to immune system regulation and diseases. *Cell Chem. Biol.* **25**(5), 499–512 (2018).
- Miyoshi, E. *et al.* The alpha1-6-fucosyltransferase gene and its biological significance. *Biochim. Biophys. Acta* **1473**(1), 9–20 (1999).
- Breton, C., Oriol, R. & Imberty, A. Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. *Glycobiology* **8**(1), 87–94 (1998).
- Balog, C. I. *et al.* N-glycosylation of colorectal cancer tissues: a liquid chromatography and mass spectrometry-based investigation. *Mol. Cell Proteomics.* **11**(9), 571–585 (2012).
- Chandler, K. & Goldman, R. Glycoprotein disease markers and single protein-omics. *Mol. Cell Proteomics.* **12**(4), 836–845 (2013).
- Tjondro, H. C., Loke, I., Chatterjee, S. & Thaysen-Andersen, M. Human protein paucimannosylation: Cues from the eukaryotic kingdoms. *Biol. Rev. Camb. Philos. Soc.* **94**(6), 2068–2100 (2019).

14. Wang, X., Gu, J., Miyoshi, E., Honke, K. & Taniguchi, N. Phenotype changes of Fut8 knockout mouse: Core fucosylation is crucial for the function of growth factor receptor(s). *Methods Enzymol.* **417**, 11–22 (2006).
15. Ng, B. G. *et al.* Biallelic mutations in FUT8 cause a congenital disorder of glycosylation with defective fucosylation. *Am. J. Hum. Genet.* **102**(1), 188–195 (2018).
16. Ferrara, C., Stuart, F., Sondermann, P., Brunker, P. & Umana, P. The carbohydrate at FcγRIIIa Asn-162. An element required for high affinity binding to non-fucosylated IgG glycoforms. *J. Biol. Chem.* **281** (8), 5032–5036 (2006).
17. Majewska, N. L., Tejada, M. L., Betenbaugh, M. J., & Agarwal, N. N-glycosylation of IgG and IgG-like recombinant therapeutic proteins: Why is it important and how can we control it? *Annu. Rev. Chem. Biomol. Eng.* (2020).
18. Dekkers, G., Treffers, L., Plomp, R., Bentlage, A. E. H., de, B. M., Koeleman, C. A. M., Lissenberg-Thunnissen, S. N., Visser, R., Brouwer, M., Mok, J. Y., Matlung, H., van den Berg, T. K., van Esch, W. J. E., Kuijpers, T. W., Wouters, D., Rispen, T., Wuhrer, M., & Vidarsson, G. Decoding the human immunoglobulin G-glycan repertoire reveals a spectrum of Fc-receptor- and complement-mediated-effector activities. *Front. Immunol.* **2017**, 8, 877.
19. Okazaki, A. *et al.* Fucose depletion from human IgG1 oligosaccharide enhances binding enthalpy and association rate between IgG1 and FcγRIIIa. *J. Mol. Biol.* **336**(5), 1239–1249 (2004).
20. Liang, W. *et al.* Core fucosylation of the T cell receptor is required for T cell activation. *Front Immunol.* **9**, 78 (2018).
21. Okada, M. *et al.* Blockage of core fucosylation reduces cell-surface expression of PD-1 and promotes anti-tumor immune responses of T cells. *Cell Rep.* **20**(5), 1017–1028 (2017).
22. Agrawal, P. *et al.* A systems biology approach identifies FUT8 as a driver of melanoma metastasis. *Cancer Cell* **31**(6), 804–819 (2017).
23. Wang, Y. *et al.* Loss of alpha1,6-fucosyltransferase inhibits chemical-induced hepatocellular carcinoma and tumorigenesis by down-regulating several cell signaling pathways. *FASEB J.* **29**(8), 3217–3227 (2015).
24. Liu, Y. C., Yen, H. Y., Chen, C. Y., Chen, C. H., Cheng, P. F., Juan, Y. H., Chen, C. H., Khoo, K. H., Yu, C. J., Yang, P. C., Hsu, T. L., & Wong, C. H. Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proc. Natl. Acad. Sci. USA* **108** (28), 11332–11337 (2011).
25. Honma, R. *et al.* Expression of fucosyltransferase 8 is associated with an unfavorable clinical outcome in non-small cell lung cancers. *Oncology* **88**(5), 298–308 (2015).
26. Fujii, H. *et al.* Core fucosylation on T cells, required for activation of T-cell receptor signaling and induction of colitis in mice, is increased in patients with inflammatory bowel disease. *Gastroenterology* **150**(7), 1620–1632 (2016).
27. Wang, T. T. *et al.* IgG antibodies to dengue enhanced for FcγRIIIA binding determine disease severity. *Science* **355**(6323), 395–398 (2017).
28. Jia, W. *et al.* A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol. Cell Proteomics.* **8**(5), 913–923 (2009).
29. Cao, Q. *et al.* Strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry. *Anal. Chem.* **86**(14), 6804–6811 (2014).
30. Yin, H. *et al.* Mass-selected site-specific core-fucosylation of serum proteins in hepatocellular carcinoma. *J. Proteome. Res.* **14**(11), 4876–4884 (2015).
31. Ma, C. *et al.* A precise approach in large scale core-fucosylated glycoprotein identification with low- and high-normalized collision energy. *J. Proteomics.* **114**, 61–70 (2015).
32. Zhou, J. *et al.* Site-specific fucosylation analysis identifying glycoproteins associated with aggressive prostate cancer cell lines using tandem affinity enrichments of intact glycopeptides followed by mass spectrometry. *Anal. Chem.* **89**(14), 7623–7630 (2017).
33. Ma, J., Sanda, M., Wei, R., Zhang, L. & Goldman, R. Quantitative analysis of core fucosylation of serum proteins in liver diseases by LC-MS-MRM. *J. Proteomics.* **189**, 67–74 (2018).
34. Yuan, W., Benicky, J., Wei, R., Goldman, R. & Sanda, M. Quantitative analysis of sex-hormone-binding globulin glycosylation in liver diseases by liquid chromatography-mass spectrometry parallel reaction monitoring. *J. Proteome. Res.* **17**(8), 2755–2766 (2018).
35. Sanda, M., Zhang, L., Edwards, N. J. & Goldman, R. Site-specific analysis of changes in the glycosylation of proteins in liver cirrhosis using data-independent workflow with soft fragmentation. *Anal. Bioanal. Chem.* **409**(2), 619–627 (2016).
36. Kozlik, P., Goldman, R. & Sanda, M. Hydrophilic interaction liquid chromatography in the separation of glycopeptides and their isomers. *Anal. Bioanal. Chem.* **410**(20), 5001–5008 (2018).
37. Sanda, M. & Goldman, R. Data independent analysis of IgG glycoforms in samples of unfractionated human plasma. *Anal. Chem.* **88**(20), 10118–10125 (2016).
38. Kozlik, P., Goldman, R. & Sanda, M. Study of structure-dependent chromatographic behavior of glycopeptides using reversed phase nanoLC. *Electrophoresis* **38**(17), 2193–2199 (2017).
39. Acs, A., Ozohanics, O., Vekey, K., Drahos, L. & Turiak, L. Distinguishing core and antenna fucosylated glycopeptides based on low-energy tandem mass spectra. *Anal. Chem.* **90**(21), 12776–12782 (2018).
40. Grieco, A. *et al.* Plasma levels of fibronectin in patients with chronic viral and alcoholic liver disease. *Hepatogastroenterology* **45**(23), 1731–1736 (1998).
41. Niu, L., Geyer, P. E., Wewer Albrechtsen, N. J., Glud, L. L., Santos, A., Doll, S., Treit, P. V., Holst, J. J., Knop, F. K., Vilsboll, T., Junker, A., Sachs, S., Stemmer, K., Muller, T. D., Tschop, M. H., Hofmann, S. M., & Mann, M. Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol. Syst. Biol.* **15** (3), e8793 (2019).
42. Gangadharan, B., Antrobus, R., Dwek, R. A. & Zitzmann, N. Novel serum biomarker candidates for liver fibrosis in hepatitis C patients. *Physiol. Rev* **53**(10), 1792–1799 (2007).
43. Benicky, J., Sanda, M., Pompach, P., Wu, J. & Goldman, R. Quantification of fucosylated hemopexin and complement factor h in plasma of patients with liver disease. *Anal. Chem.* **86**(21), 10716–10723 (2014).
44. Goldman, R. & Sanda, M. Targeted methods for quantitative analysis of protein glycosylation. *Proteomics. Clin. Appl.* **9**(1–2), 17–32 (2015).
45. Sanda, M. *et al.* Quantitative liquid chromatography-mass spectrometry-multiple reaction monitoring (LC-MS-MRM) analysis of site-specific glycoforms of haptoglobin in liver disease. *Mol. Cell Proteomics.* **12**(5), 1294–1305 (2013).
46. Brockhausen, I., Möller, G., Merz, G., Adermann, K. & Paulsen, H. Control of mucin synthesis: The peptide portion of synthetic O-glycopeptide substrates influences the activity of O-glycan Core 1 UDPgalactose:N-acetyl-d-galactosaminyl-R β3-galactosaminyltransferase. *Biochemistry* **29**, 10206–10212 (1990).

## Acknowledgements

This work was supported by National Institutes of Health Grants U01CA230692, R01CA135069, and R01CA238455 (to RG); further support was provided by the Office Of The Director, National Institutes of Health under Award Number S10OD023557 supporting the operation of the Clinical and Translational Glycoscience Research Center, Georgetown University, CCSG Grant P30 CA51008 (to Lombardi Comprehensive Cancer Center) supporting the Proteomics and Metabolomics Shared Resource, and the Czech Science Foundation,

Grant No 19-18005Y (to PK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### Author contributions

Conceptualization, M.S.; methodology, M.S.; validation, M.S and J.A; formal analysis, M.S. and P.K; investigation, M.S.; resources, R.G.; data curation, M.S.; writing—original draft preparation, M.S and R.G.; writing—review and editing, M.S and R.G., visualization, M.S and J.A; All authors have read and agreed to the published version of the manuscript.

### Funding

National Institutes of Health, U01CA230692, U01CA230692, Czech Science Foundation, No 19-18005Y.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-02838-3>.

**Correspondence** and requests for materials should be addressed to M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021