Published in final edited form as:

J Proteome Res. 2019 January 04; 18(1): 359–371. doi:10.1021/acs.jproteome.8b00654.

Differential Quantitative Determination of Site-Specific Intact *N*-Glycopeptides in Serum Haptoglobin between Hepatocellular Carcinoma and Cirrhosis Using LC-ET_hcD-MS/MS

Jianhui Zhu[†], Zhengwei Chen[‡], Jie Zhang[†], Mingrui An[†], Jing Wu[†], Qing Yu[§], St. John Skilton^{||}, Marshall Bern^{||}, K. Ilker Sen^{||}, Lingjun Li^{‡,§}, and David M. Lubman^{*,†}

[†]Department of Surgery, University of Michigan Medical Center, Ann Arbor, Michigan 48109, United States

[‡]Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States

[§]School of Pharmacy, University of Wisconsin—Madison, Madison, Wisconsin 53705, United States

^{||}Protein Metrics Incorporated, San Carlos, California 94070, United States

Abstract

Intact *N*-glycopeptide analysis remains challenging due to the complexity of glycopeptide structures, low abundance of glycopeptides in protein digests, and difficulties in data interpretation/quantitation. Herein, we developed a workflow that involved advanced methodologies, the ET_hcD-MS/MS fragmentation method and data interpretation software, for differential analysis of the microheterogeneity of site-specific intact *N*-glycopeptides of serum haptoglobin between early hepatocellular carcinoma (HCC) and liver cirrhosis. Haptoglobin was immunopurified from 20 μ L of serum in patients with early HCC, liver cirrhosis, and healthy controls, respectively, followed by trypsin/GluC digestion, glycopeptide enrichment, and LC-ET_hcD-MS/MS analysis. Identification and differential quantitation of site-specific *N*-

^{*}**Corresponding Author:** Tel: 734-647-8834. Fax: 734-615-2088. dmlubman@umich.edu.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00654.

SDS-PAGE analysis of the Hp eluent enriched from an HCC serum sample using the antibody-immobilized HPLC column; extracted ion chromatograms (XICs) of MS/MS spectra of the HexNAc oxonium ion within the *m/z* range 204.08–204.09 in early HCC, cirrhosis, and normal controls; MS1 spectrum of a low-abundant glycopeptide VVLHPN²⁴¹YSQyD with the glycan A4G4F2S3 showed the isotope distribution, with the S/N ratio greater than 26; Venn diagrams detail the overlap of site-specific *N*-glycopeptides and glycan compositions of serum haptoglobin among early HCC, cirrhosis, and normal controls; heat map of the altered *N*-glycopeptides observed in disease groups (early HCC and cirrhosis) but absent in healthy controls; top 10 abundant *N*-glycopeptides normalized against the total *N*-glycopeptides in serum Hp in early HCC, cirrhosis, and normal controls; overall distribution of bi-, tri-, and tetra-antennary *N*-glycoforms on sites N184, N207, and N241 of serum Hp in early HCC, cirrhosis, and normal controls, respectively; HPLC peak area of the eluted Hp fraction in each serum sample; relative abundance of site-specific intact *N*-glycopeptides identified in serum Hp from early HCC, cirrhosis, and normal controls ([PDF](#))

List of 279 *N*-glycopeptide spectral matches, including the glycosite, peptide sequence, modification summary, glycan composition, observed *m/z*, *z*, calculated *m/z*, ppm, Byonic score, and retention time ([XLSX](#))

Notes

The authors declare the following competing financial interest(s): Dr. Bern, Dr. Skilton, and Dr. Sen are from Protein Metrics Inc., and have a competing financial interest in the software package.

glycopeptides were performed using a combination of Byonic and Byologic software. In total, 93, 87, and 68 site-specific *N*-glycopeptides were identified in early HCC, liver cirrhosis, and healthy controls, respectively, with high confidence. The increased variety of *N*-glycopeptides in liver diseases compared to healthy controls was due to increased branching with hyper-fucosylation and sialylation. Differential quantitation analysis showed that 5 site-specific *N*-glycopeptides on sites N184 and N241 were significantly elevated in early HCC compared to cirrhosis ($p < 0.05$) and normal controls ($p = 0.001$). The result demonstrates that the workflow provides a strategy for detailed profiles of *N*-glycopeptides of patient samples as well as for relative quantitation to determine the level changes in site-specific *N*-glycopeptides between disease states.

Graphical Abstract

Keywords

site-specific glycopeptides; ETHcD-MS/MS; haptoglobin; HCC; Byonic; Byologic

INTRODUCTION

Glycosylation changes in serum proteins are highly associated with the progression of liver disease.^{1,2} Extensive efforts have been devoted to seeking specific alterations in cancer-related glycoforms for hepatocellular carcinoma (HCC).² Serum haptoglobin (Hp) has attracted particular attention due to its potential as a reporter molecule for aberrant glycosylations during the development of HCC.^{3,4} Hp has four potential glycosylation sites, i.e., N184, N207, N211, and N241, at its β chain.⁵ Increasing evidence has demonstrated that fucosylated/sialylated glycan structures of serum Hp have been significantly elevated in HCC compared to liver cirrhosis.^{4,6–8} Analysis of glycopeptides of serum Hp that simultaneously elucidates glycosylation sites and the attached glycan structures holds great promise to discover site-specific glyco-markers correlated with HCC.

Glycoproteomics has become an emerging field to the discovery of unique site-specific glycopeptides associated with cancer.⁹ Recently, site-specific glycoforms of Hp have been investigated by using either truncated *N*-glycopeptides of serum Hp with sialic acids or galactoses removed by exoglycosidases^{10–12} or a human haptoglobin standard whose glycoforms are much simpler than those in cancer or liver diseases.¹³ A recent study characterized the intact *N*-glycopeptides of serum Hp in gastric cancer (GC) patients vs healthy controls for GC biomarker discovery.¹⁴ However, comprehensive profiling and

quantitative analysis of intact glycopeptides of serum Hp in patients with HCC and liver cirrhosis has not been explored in detail.

Intact glycopeptide analysis remains challenging due to the complexities associated with the glycopeptide structure (i.e., peptide sequence, glycosylation sites, and glycan heterogeneity), low ionization efficiency and low abundance of glycopeptides compared to the peptides derived from the proteolytic digestion of biological samples, and challenges in data interpretation/quantitation. Due to the low abundance of glycopeptides in protein digests, enrichment strategies have been used to enrich glycopeptides via zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC),¹⁵ electrostatic repulsion liquid chromatography (ERLIC),¹⁶ graphitized carbon (PGC) solid-phase extraction,¹⁴ and cotton tips.¹⁷ In addition, CID/HCD,¹⁸ stepped HCD,¹⁹ CID/ETD,²⁰ HCD/ETD,²¹ and EThcD^{22–24} fragmentation techniques have been used to obtain detailed information on glycan fragments and peptide backbones.²⁵ Electron-transfer higher-energy collision dissociation (EThcD) MS/MS fragmentation in particular has recently been shown to have advantages in providing accurate structural information for intact glycopeptides.²²

Data interpretation of glycopeptide identification and quantitation remains a challenge in glycoproteomics.²⁶ Several search algorithms have been developed to interpret glycopeptide data, such as Byonic,²⁷ pGlyco 2.0,⁸ Integrated GlycoProteome Analyzer (I-GPA),²⁹ GP Finder,³⁰ gFinder,³¹ GlycoMaster DB,³² Glycopep grader (GPG),³³ GlycoPep Detector (GPD),³⁴ and GlycoPeptideSearch (GPS).¹³ Among these, most algorithms were developed to identify glycopeptides from mass spectral data produced with CID and/or HCD fragmentation. GlycoMaster DB³² and GlycoPep Detector³⁴ can analyze ETD tandem spectra, and Byonic can analyze all types of MS tandem spectra with any combination of CID, HCD, ETD, and/or EThcD fragmentation.²⁷ However, only limited software tools are available for relative quantitation of glycopeptides, such as pGlyco 2.0,²⁸ (I-GPA),²⁹ and LaCyTools,³⁵ which are applicable only for CID/HCD data. ETD or EThcD data interpreted by Byonic can be further quantified by Byologic (Protein Metrics Inc.), an automated quantitation software developed recently using both MS1 raw data and MS2 search results. This software performs label-free quantitation with extracted ion chromatograms (XICs) of glycopeptide precursor ions, where the peak area of a given glycopeptide is integrated and quantitated automatically at two levels: the total glycopeptide level and its corresponding glycosite level. The automated quantitation and the flexibility in quantifying glycopeptides among biological samples greatly facilitate glycopeptide quantitative analysis.

Herein, we developed a workflow for differential site-specific intact *N*-glycopeptide analysis that can detect changes in glycoform microheterogeneity between disease states. The method involved EThcD-MS/MS fragmentation for determination of intact *N*-glycopeptide structures as well as for quantitation of alterations in site-specific glycoforms in serum haptoglobin among early HCC, liver cirrhosis, and healthy controls. This method was made possible using a combination of Byonic and Byologic softwares, where we were able to accomplish both identification and label-free quantitation of intact *N*-glycopeptides among different disease groups. On the basis of these advanced methodologies, we identified 279 *N*-glycopeptide spectral matches corresponding to 98 site-specific *N*-glycopeptides of serum haptoglobin with FDR < 1%. Five distinct glycopeptides at the sites N184 and N241

containing fully sialylated tri-or tetra-antennary *N*-glycans with either no fucose or two fucoses present were significantly increased in early HCC compared to cirrhosis. The differential quantitative LC-ET_hD-MS/MS strategy demonstrated the potential for detecting subtle changes in site-specific *N*-glycopeptides for discrimination of early HCC from cirrhosis. This workflow has provided a means of comparison in the identification and label-free relative quantitation of site-specific intact *N*-glycopeptides of patient samples between disease groups.

MATERIALS AND METHODS

Materials

Reagents were purchased from Sigma (St. Louis, MO) unless otherwise specified. Sequencing-grade trypsin and GluC were purchased from Promega (Madison, WI). The 7K MWCO Zeba Spin Desalting columns and the 4 mL YM-3 centrifugal devices were purchased from Thermo Scientific (Rockford, IL) and Millipore (Billerica, MA), respectively. Antihuman haptoglobin antibody was purchased from Abcam (Cambridge, MA, CatLog No. ab13429). The HILIC-packed TopTips were purchased from Glygen (Columbia, MD).

Serum Samples

Serum samples from patients with early-stage HCC ($n = 5$), liver cirrhosis ($n = 5$), or healthy subjects ($n = 5$) were provided by the University Hospital, Ann Arbor, MI, according to IRB approval. All HCC and cirrhosis patients in this study were hepatitis C virus (HCV) infected, which is the most prominent risk factor for cirrhosis and HCC in the United States.³⁶ The clinical features of patients are summarized in Table 1. Samples were aliquoted and stored at -80°C until further use.

Haptoglobin Purification

Haptoglobin (Hp) was purified from 20 μL of patient serum using an antibody-immobilized HPLC column which has a recovery of 40–50% of Hp from serum samples as described previously.³⁷ The immunopurification was performed on a Beckman Coulter HPLC system (Fullerton, CA) with a PEEK column (4.6 mm $\times$ 50 mm) packed in-house with the UltraLink hydrazide resin (Thermo Scientific, Rockford, IL) conjugated with an antihuman Hp antibody (Abcam, Cambridge, MA). The eluted haptoglobin fraction was desalted using a 4 mL YM-3 centrifugal device by buffer exchange 3 times with deionized water and then dried down in a SpeedVac concentrator (Thermo). The HPLC peak area of the eluted Hp fraction was measured and compared among patients (Supplemental Table S1), where SDS-PAGE gel analysis of the Hp eluent showed the purity of haptoglobin fraction (Supplemental Figure S1). One-tenth of the Hp eluent was loaded onto the gel, with 0.3 μg of a haptoglobin standard protein (Abcam) as a reference. On the basis of the gel result, the amount of haptoglobin used for the subsequent trypsin digestion was estimated to be around 3 μg .

Double Enzymatic Digestion

The purified serum Hp was dissolved in 50 mM NH_4HCO_3 , reduced with 20 mM DTT at 60°C for 45 min, and then alkylated with 50 mM iodoacetamide (IAA) in the dark at room

temperature for 30 min. The sample was then desalted using a 7K MWCO Zeba Spin Desalting column (Thermo Scientific, Rockford, IL), dried down, and resuspended in 50 mM NH_4HCO_3 . Trypsin (Promega, Madison, WI) was added to the sample and incubated at 37 °C overnight. The digestion was quenched at 95 °C for 10 min. A second digestion step with GluC (Promega, Madison, WI) was performed at 37 °C for 12 h. The sample was finally dried down in a SpeedVac (Thermo) for subsequent glycopeptide enrichment.

Glycopeptide Enrichment by HILIC TopTips

Glycopeptides were enriched using HILIC TopTips (Glygen, Columbia, MD) according to manufacturer's instructions. Briefly, the HILIC TopTips were activated with water 3 times, followed by equilibration with the binding buffer (85% acetonitrile containing 15 mM ammonium acetate, pH 3.5). The peptides were resuspended in binding buffer and loaded onto the HILIC TopTip, where the flow through was collected and reloaded 3 times to ensure complete binding. After glycopeptide capture, the HILIC TopTip was washed 3 times with 10 μL of 85% ACN containing 15 mM ammonium acetate. The bound glycopeptides were then eluted with 10 μL of water twice and dried under vacuum.

LC-MS/MS Analysis

The enriched glycopeptides were dissolved in 0.1% formic acid (FA) and analyzed with two injections on the Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo) coupled to a Dionex UPLC system. A binary solvent system was composed of H_2O containing 0.1% FA (A) and CH_3CN containing 0.1% FA (B) with a flow rate of 300 nL/min. Samples were separated on a 75 $\mu\text{m} \times 15$ cm column, packed in-house with 1.7 μm , 150 Å, BEH C18 material.²² The LC gradient for intact *N*-glycopeptides was set as follows, 3%–30% B (18–98 min), 30%–75% B (100–108 min), and 75%–95% B (108–118 min). The MS instrument was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. The MS1 scans (from *m/z* 400–1800) were acquired in the Orbitrap (120 000 resolution, 4e5 AGC, 100 ms injection time) followed by EThcD MS/MS acquisition of the precursors with the highest charge states in an order of intensity and detection in the Orbitrap (60 000 resolution, 3e5 AGC, 100 ms injection time). EThcD was performed with optimized user-defined charge-dependent reaction time (50 ms for 2+, 20 ms for 3+, 4+, and 5+; 9 ms for 6+, 7+, and 8+) supplemented by 33% HCD activation as reported previously.²²

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (<http://www.ebi.ac.uk/pride/archive/>) with the data set identifier PXD011239. Username: reviewer36928@ebi.ac.uk Password: AOvae09t

Data Analysis

All spectra were searched with Byonic (Protein Metrics, San Carlos, CA) incorporated in Proteome Discoverer 2.1 (Thermo) against the UniProt human haptoglobin database (P00738). A UniProt human protein database containing the complete human proteome with haptoglobin removed was also applied in data searching to evaluate the assignments of glycopeptides. A human *N*-glycan database containing 179 human *N*-glycans was employed for data searching, which combines a Byonic database of 164 mammalian *N*-glycans but no multiple fucoses and 15 *N*-glycans with multiple fucoses reported in the literature for serum

Hp in HCC and liver cirrhosis patients.^{7,11–13,37} Trypsin and GluC were selected as enzymes with a maximum of two missed cleavages allowed. The search was performed using the following parameters: (1) static modification, carbamidomethyl of C; (2) dynamic modifications, oxidation of M, deamidation of N and Q, and glycan modifications; (3) maximum missed cleavages, 2; (4) precursor ion mass tolerance 10 ppm; (5) fragment ion mass tolerance 0.01 Da. Results were filtered at 1% FDR and confidence threshold of Byonic score > 100, and further validation was performed manually. The manual check criteria include the retention time, the presence of oxonium ions, e.g., m/z 204.09 for HexNAc, 292.10 for NeuAc, 274.09 for NeuAc-H₂O, 366.13 for HexHexNAc, 512.20 for HexHexNAcFuc, and 657.23 for HexNAcHexNeuAc. The monoisotopic peak and diagnostic oxonium ion were further checked in the case of 2Fuc-1.02 = 1NeuAc to eliminate the false assignment of NeuAc to 2Fuc.

In the main text, glycans were described as AxGxFxSx according to the Oxford glycan nomenclature, i.e., Ax, number (x) of antenna; Gx, number (x) of linked galactose on antenna; Fx, number (x) of fucose; Sx, number (x) of sialic acids. For example, A3G3F2S2 represents the triantennary trigalactosylated bifucosylated bisialylated glycan.

Automated quantitative analysis was performed using Byologic (Protein Metrics Inc.) which uses as input both MS1 raw data and Byonic search results. Byologic can compare the quantitative results among samples so all of the raw data and Byonic search results from 30 MS runs were uploaded into one Byologic project for comparison. Byologic incorporates the list of glycopeptides identified, the MS/MS spectrum, XIC, and precursor isotope plot in one single window, allowing one to gather complete information for quantitation analysis. The isotope plot of the precursor ion is one of the features of Protein Metrics software. Individual scans are used to find the isotopes within a given tolerance window (i.e., 18 ppm), and the isotope m/z apex is used to determine which m/z value to use for XIC creation. The use of a single isotope has previously been empirically justified by demonstrating more consistent XIC behavior.²⁷ The peak area of the XIC of a given glycopeptide was automatically integrated and normalized against that of total glycopeptides of Hp identified in each MS run. The abundance of a site-specific glycoform was represented by the sum of all glycopeptides containing the same glycan at the glycosite, where the peptide sequences may differ in the case of miscleavage or deamidation.

The relative quantitation of each site-specific glycopeptide was achieved at two levels: the total glycopeptide level and its corresponding glycosite level. Each sample was analyzed with two injections, and the average normalized abundance of a given *N*-glycopeptide from the two replicates was used to measure its relative levels between disease groups.

Statistical Analysis

The abundance of site-specific glycopeptides in early HCC, cirrhosis, and healthy controls was compared using GraphPad Prism 7 (La Jolla, CA). The reproducibility analysis of 4 replicates of a normal serum sample was performed with the Perseus software.³⁸ The heatmap was generated with HemI 1.0.³⁹ The significance of the difference in site-specific glycopeptides between groups was evaluated using Tukey's multiple comparisons test in ANOVA. The threshold $p < 0.05$ was used to determine the differentially expressed

glycopeptides. The scatter plots of the differentially expressed site-specific *N*-glycopeptides between early HCC and cirrhosis were generated with GraphPad Prism 7 (La Jolla, CA).

RESULTS AND DISCUSSION

In this study, we developed a workflow of differential quantitative determination of intact *N*-glycopeptides of serum Hp to characterize the alterations in site-specific glycoforms of Hp between early HCC and liver cirrhosis. Serum Hp has four potential glycosylation sites (N184, N207, N211, and N241) on its 40 kDa β chain,⁵ while tryptic digestion generates 3 glycosylated peptides, MVSHHN¹⁸⁴LTTGATLINEQWLLTTAK, NLFLN²⁰⁷HSEN²¹¹ATAK, and VVLHPN²⁴¹YSQVDIGLIK. The tryptic peptide NLFLN²⁰⁷HSEN²¹¹ATAK contains two glycosites, making it difficult to interpret the glycan compositions. GluC cleaves peptide bonds at the C-terminus of either aspartic (D) or glutamic acid (E) residues. Thus, GluC has been coupled with trypsin to further separate the two glycosites of N207 and N211.^{11–13} The trypsin/GluC digestion generates singly glycosylated peptides MVSHHN¹⁸⁴LTTGATLINE, nlfln²⁰⁷hse, n²¹¹atak, and VVLHPN²⁴¹YSQVD, which can facilitate the determination of glycoforms at the four sites separately.

The workflow used in this work is outlined in Figure 1. First, Hp was immunopurified from 20 μ L of patient sera in early HCC, cirrhosis, and healthy controls using an HPLC-based anti-Hp column. The abundance of the purified Hp in each patient was evaluated by HPLC peak area (Supplemental Table S1) as well as SDS-PAGE gel (Supplemental Figure S1) where the same amount of Hp from individual patients was used for MS analysis to eliminate variations in protein abundances between samples. Hp was digested with trypsin and GluC, and the glycopeptides were enriched using HILIC TopTips (Glygen), followed by *nano* LC-ET^hCD-MS/MS analysis in duplicate on the Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo). All spectra were searched with Byonic (Protein Metrics) against the UniProt human haptoglobin database (P00738) and a human *N*-glycan database containing 179 human *N*-glycans. Intact *N*-glycopeptides at each glycosylation site were quantitated using Byologic (Protein Metrics). Finally, statistical analysis was performed to evaluate the differentially expressed site-specific glycopeptides in early HCC compared to liver cirrhosis and normal controls.

Method Reproducibility

Four independent replicates of a normal serum sample that have been processed at four different times were applied to evaluate the reproducibility of the workflow. The Perseus software was employed for reproducibility analysis of 4 replicates. As shown in Figure 2a, the Pearson correlation coefficient R^2 values for the binary comparison of the 4 technical replicates were from 0.977 to 0.994, demonstrating good reproducibility of the method. A pie chart indicates the frequency of glycopeptide quantification, which were present in all 4 technical replicates, in 2–3, or only in 1 (Figure 2b). The result showed that over 90% of glycopeptides were identified in more than 2 replicates. The reproducibility of the AUC of five glycopeptides at site N241 which covered five orders of magnitude (10^6 – 10^{10}) has also been investigated across the 4 technical replicates (Figure 2c). The relative standard

deviation (RSD) of the 4 replicates for the most abundant glycopeptide at site 241 (VVLHPN²⁴¹YSQVD_A2G2S2) was 12.7%, followed by 9.1% for VVLHPN²⁴¹YSQVD_A3G3S3, and 14.7% for VVLHPN²⁴¹YSQVD_A2G2S1. The low-abundant glycopeptides of VVLHPN²⁴¹YSQVD_A3G3F1S3 and VVLHPN²⁴¹YSQVD_A3G3F1S2 had a RSD of 8.0% and 16.4%, respectively.

Identification of Intact *N*-Glycopeptides

Site-specific *N*-glycopeptides of haptoglobin enriched from individual patients were analyzed with two injections by EThcD-MS/MS. The representative extracted ion chromatograms (XICs) of MS/MS spectra of the HexNAc oxonium ion (HexNAc⁺) within the *m/z* range 204.08–204.09 in early HCC, cirrhosis, and normal controls are shown in Supplemental Figure S2. As shown in Figure S2, the major peaks at ~44, ~46, ~48, and ~68 min corresponded to the HexNAc oxonium ion derived from the glycopeptides of VVLHPN²⁴¹YSQVD, NLFLN²⁰⁷HSE, MVSHHN¹⁸⁴LTGATLINE, and VVLHPN²⁴¹YSQVDIGLIK, respectively. The oxonium ion derived from the glycopeptide of N²¹¹ATAK (~7 min) was difficult to detect, mainly due to the high hydrophilicity of the small peptide NATAK.^{11,12}

All *N*-glycopeptides were automatically identified by Byonic (Protein Metrics) with high confidence. Previous studies have addressed how to control FDR < 1% when searching glycopeptide data using Byonic.^{27,40} It has been reported that for glycopeptide profiling FDR is largely dependent on the precursor/product ion mass tolerances and the Byonic score threshold, where the peptide FDR was reduced as the scoring thresholds and the stringency of the mass tolerance were increased.⁴¹ In this study, strict mass tolerances (10 ppm/0.01 Da precursor/product ions) were applied while the assignments of glycopeptides with Byonic scores > 100 were confirmed with good MS/MS spectra. To control FDR well below 1%, we set the cut off of the Byonic score as 100.

Furthermore, manual verification of retention time and presence of diagnostic glycan fragments (specifically fucose and sialic acid) was employed to confirm the assignment. In the case of 2Fuc-1.02 = 1NeuAc, the glycopeptide assignment was confirmed with the isotopic evidence of the precursor ion extracted by Byologic (Protein Metrics), where the monoisotopic peak and diagnostic oxonium ion were checked to eliminate the false assignment of NeuAc to 2Fuc.

In total, there were 279 tandem mass spectral matches with *N*-glycopeptides, representing 98 site-specific *N*-glycopeptides containing bi-, tri-, and tetra-antennary sialylated and/or fucosylated *N*-glycans. Multiple charge states were observed for most of the *N*-glycopeptides. All of the *N*-glycopeptide spectral matches are listed in Supplemental Table S2, including the glycosite, peptide sequence, modification summary, glycan composition, observed *m/z*, charge (*z*), calculated *m/z*, ppm, Byonic score, and retention time.

The MS signal-to-noise ratio (S/N) was also evaluated. Supplemental Figure S3 shows the MS1 spectrum of a low-abundant glycopeptide VVLHPN²⁴¹YSQVD with the glycan A4G4F2S3 (*m/z* = 1197.978). The MS1 spectrum showed the isotope distribution. The intensity of noise was indicated with “N=”. As shown in Figure S3, the intensity of the

precursor peak was over 26-fold higher than that of noise. With the S/N ratio considerably greater than 3, these glycopeptides can be well quantified.

Evaluation of Hp Glycopeptide Assignments

Each glycopeptide assignment corresponded to a unique MS2 scan number which has been used as a reference to make sure it was not assigned to any other glycopeptides. To evaluate the assignments of glycopeptides of haptoglobin, a UniProt human protein database containing the complete human proteome with haptoglobin removed was also employed in data searching. As an example for an HCC sample, there were 46 matched glycopeptides that originated from 8 serum glycoproteins, including IGJ ($n = 4$), IGHG1 ($n = 5$), IGHG2 ($n = 8$), IGHM ($n = 8$), IGHA1 ($n = 10$), IGHA2 ($n = 7$), A1AG1 ($n = 3$), and HEMO ($n = 1$). The MS2 scan numbers of these glycopeptides did not overlap with those of the Hp glycopeptides. The result confirmed that all of the assignments of glycopeptides of haptoglobin in Table S2 were unique to haptoglobin.

It should be pointed out that haptoglobin was enriched from patient sera using an HPLC-based antibody column where the 1D gel analysis of the HPLC eluent of haptoglobin showed the purity of the haptoglobin fraction. A trace of IgGs, alpha-1 acid glycoprotein, and hemoglobin might coelute, though not detectable in the SDS-PAGE gel, but were detected in the Orbitrap Fusion Lumos due to LC separation and the high sensitivity of the mass spectrometer. However, the MS2 spectra intensity of those matched glycopeptides from glycoproteins other than haptoglobin was very low, confirming the low abundance in the enriched haptoglobin fraction.

ET_hcD-MS/MS of N-Glycopeptides

Representative MS/MS spectra of fingerprint N-glycopeptides of MVSHHN¹⁸⁴LTTGATLINE with glycan A3G3F2S2 and VVLHPN²⁴¹YSQVDIGLIK with glycan A3G3F3S1 and their m/z error plots are shown in Figure 3. The glycopeptide of VVLHPN²⁴¹YSQVDIGLIK was derived from incomplete cleavage by GluC.¹³ Both of the glycopeptides were quadruply charged ($z = 4$). Due to the complexity of the glycopeptide structure, it is difficult for a single fragmentation technique to produce unbiased glycan and peptide fragment ions in one MS/MS spectrum. However, ET_hcD-MS/MS has the advantage of incorporating both glycan and peptide fragment ions in a single MS/MS spectrum.²² Figure 3 represents the ET_hcD-MS/MS spectra with a variety of product ions, including glycan fragments (oxonium ions), b/y and c/z ions from peptide backbones, and glycosidic fragments with the sequential loss of monosaccharides from the parent glycopeptide. All glycopeptide spectrum matches were filtered with strict mass tolerances (10 ppm/0.01 Da precursor/product ions), resulting in over 30 unique fragments for sequencing a given glycopeptide (Figure 3).

As shown in Figure 3, the oxonium ions such as HexNAc (m/z 204.09), HexHexNAc (m/z 366.13), NeuAc (m/z 292.10), NeuAc-H₂O (m/z 274.09), and HexNAcHexNeuAc (m/z 657.23) were intensely present in the spectra. The b and y fragments from the peptide backbone were observed as well but at low intensity. Glycosidic fragments with the peptide remaining intact were well characterized in the ET_hcD MS/MS spectrum (Figure 3). The

distribution of these fragment ions was consistent with other reports in the literature.²² In a typical EThcD-MS/MS spectrum of glycopeptides, the most intense peaks are oxonium ions, followed by glycosidic fragments.²² The intensity of *b/y* and *c/z* ions from peptide backbones was less than that of oxonium ions and glycosidic fragments, but most of them were markedly above the noise level. By combining the presence of intense core ions, including Y0 (naked peptide), Y1 (peptide + HexNAc), Y2 (peptide +2HexNAc), Y3 (peptide +2HexNAcHex), Y4 (peptide +2HexNAc2Hex), Y5 (peptide +2HexNAc3Hex), and Y6 (peptide +3HexNAc3Hex), these fragments were sufficient to identify the glycopeptides with detailed spectral information for both glycans and peptide.

Notably, specific diagnostic fragment ions in EThcD spectra can further provide detailed structural information such as core and outer arm fucosylation. For example, the diagnostic fragment ion of HexHexNAcFuc (*m/z* 512.20, Figure 3A and 3B) confirmed the presence of the outer-arm fucose residue, while the diagnostic ions of peptide+2HexNAc+Fuc (*m/z* 1153.54 in Figure 3A and *m/z* 1174.12 in Figure 3B, marked by arrows) further confirmed the core fucose residue. The *m/z* error plots illustrated the accuracy of the fragment assignments where the observed mass of the oxonium ions (experimental quantity) matched exactly with the calculated mass (reference quantity), and in the case of glycosidic fragments with the intact peptide, the difference between the observed *m/z* and the calculated *m/z* was less than 0.004 Da.

Site-Specific *N*-Glycosylation Microheterogeneity in Serum Hp

Table S3 summarizes the 98 site-specific *N*-glycopeptides identified in this study, where the most abundant charge state of each *N*-glycopeptide, *m/z*, and relative abundance at each glycosite are listed. To simplify the annotation of glycan composition, a 4-digit nomenclature was employed in an order of HexNAc_Hex_Fuc_NeuAc (HexNAc = N-acetylhexos-amine; Hex = Hexose; Fuc = Fucose; NeuAc = sialic acid).

As shown in Table S3, there were 34 glycoforms identified at site N184, 22 at site N207, 4 at site N211, and 38 at site N241. The site N241 has the highest variety in *N*-glycosylation compared to other sites. A limited number of glycopeptides were identified at site N211 compared to the other three sites, mainly because of the high hydrophilicity of the small peptide N²¹¹ATAK which makes it difficult to detect.^{11,12} Although the resulting short glycopeptide N²¹¹ATAK was not quantifiable by C18-LC, GluC facilitated the identification of glycoforms at site N207.

Interestingly, a penta-antennary *N*-glycan A5G5S1 was identified on the site N184 for the first time. It should be noted that, in this study, more glycoforms were identified at sites N184, N207, and N241 than previous studies on intact¹⁴ or truncated¹¹ glycopeptide analysis of serum haptoglobin, which may benefit from the advanced EThcD MS/MS fragmentation technique and trypsin/GluC digestion for characterization of intact glycopeptides.

Among the 98 site-specific glycopeptides were 93, 87, and 68 site-specific glycopeptides identified in early HCC, liver cirrhosis, and normal controls, respectively. A Venn diagram details the overlap of site-specific glycopeptides and glycans identified among disease

groups (Supplemental Figure S4). As shown in Supplemental Figure S4a, there are 62 intact *N*-glycopeptides commonly identified in the three groups, while 21 more site-specific glycopeptides overlapped between early HCC and cirrhosis but were absent in healthy controls. The increased number of glycopeptides in early HCC and cirrhosis is dominant due to increased fucosylation and sialylation, which can bear up to four (in cirrhosis) or five (in early HCC) fucoses and four sialic acids. A heatmap was employed to show the frequency of the altered glycopeptides identified in disease group across the 10 individual patients (Supplemental Figure S5). In the case of 5 early HCCs, 43.7% of the altered glycopeptides were identified in all 5 patients, 68.7% in at least 3 patients, and 12.5% in only 1 patient. Taking the glycopeptide N184_A3G3F2S3 as an example, it was identified in all 5 early HCCs but only in 3 cirrhosis patients with lower abundance.

When combining glycan compositions at the four sites, there were 41, 38, and 28 glycan compositions identified in early HCC, cirrhosis, and normal controls, respectively (Supplemental Figure S4b). Notably, 4 unique glycan compositions were exclusively identified in early HCC but absent in cirrhosis and normal controls, including A3G3F1, A4G4F1S4, A4G4F2S4, and A4G4F5S1. When compared to normal controls, there were 10 more glycan compositions identified in both early HCC and cirrhosis, i.e., A3G2S2, A3G3F2S3, A3G3F4S1, A4G4F1S1, A4G4F1S2, A4G4F1S3, A4G4F2S2, A4G4F3S1, A4G4F3S3, and A4G4F4S1, most of which are tri- and tetra-antennary glycans with increased fucose or sialic acid residues.

Quantitation of Site-Specific *N*-Glycopeptides at Total Glycopeptide Level

The abundance of site-specific glycopeptides was further quantitated using Byologic (Protein Metrics) and compared among disease groups. With the automated quantitation software, the peak area of the XIC of a given glycopeptide was integrated and normalized against the sum of abundance of all glycopeptides of Hp identified, providing a relative quantitation within each sample. The relative abundance of each glycopeptide was measured at two levels: the total glycopeptide level and individual glycosite level.

Supplemental Figure S6 shows the top 10 abundant *N*-glycopeptides identified in the three groups. The top 10 abundant glycopeptides were predominantly those with biantennary glycans, except in the cases of early HCC and cirrhosis, where the glycopeptides containing triantennary glycans with one or two fucoses and multiple sialic acids were included in the top 10 list. Overall, the top two abundant glycopeptides were those at sites N184 and N241 with the biantennary disialylated glycan A2G2S2, which account for $62.3 \pm 7.5\%$, $64.2 \pm 7.1\%$, and $67.1 \pm 6.0\%$ of total *N*-glycopeptides in early HCC, cirrhosis, and normal controls, respectively. The third most abundant glycopeptide was that at site N184 with biantennary monosialylated glycan (A2G2S1), which accounted for $4.7 \pm 1.3\%$, $4.8 \pm 1.5\%$, and $10.2 \pm 1.2\%$ of total *N*-glycopeptides in the three groups. The top 10 *N*-glycopeptides contributed $81.2 \pm 5.8\%$, $81.7 \pm 5.5\%$, and $89.8 \pm 1.6\%$ of *N*-glycopeptides in early HCC, cirrhosis, and normal controls, respectively.

Quantitation of Site-Specific *N*-Glycopeptides at Glycosite Level

We further determined the relative abundance of *N*-glycopeptides at each glycosite level. The mean value of relative abundance of *N*-glycopeptide at each glycosite is listed in Table S3, where the abundance of a glycopeptide was normalized against the sum of all glycopeptides of Hp at the same glycosite. Since the *N*-glycosylation profile at sites N184, N207, and N241 was well characterized, we focused on the quantitation at these three sites. Figure 4 shows the 10 most abundant *N*-glycopeptides at sites N184, N207, and N241 in early HCC as well as their abundances in cirrhosis and normal controls. The error bar shows the SD across the 5 individual patients in each group.

On the site N184, as shown in Figure 4A, the most abundant glycoform is the biantennary disialylated glycan A2G2S2, followed by the biantennary monosialylated glycan A2G2S1, in all samples. The levels of the A2G2S2 and A2G2S1 at site N184 are $61.59 \pm 5.37\%$ and $9.41 \pm 2.37\%$ in early HCC, $69.65 \pm 3.28\%$ and $9.59 \pm 1.09\%$ in cirrhosis, and $60.82 \pm 2.73\%$ and $22.16 \pm 2.13\%$ in normal controls, respectively. It should be pointed out that the abundance of these two glycoforms on site N184 was distinctly lower in early HCC ($71 \pm 4.90\%$) than that in normal controls ($82.98 \pm 1.83\%$) and cirrhosis ($79.24 \pm 2.45\%$). The rest of the abundant glycoforms at N184 in early HCC were bi- or triantennary *N*-glycans, including A2G2F2S1 (7.27%), A3G3F2S2 (3.43%), A3G3F1S3 (3.35%), A3G3S3 (3.34%), A2G2F1S2 (2.56%), A2G2F1S1 (1.22%), A3G3S2 (1.18%), and A3G3F1S2 (1.07%). The distribution of these 8 glycoforms in cirrhosis and normal controls was similar to that in early HCC. However, the tetra-antennary *N*-glycans with mono-, di-, or trisialic acids and mono-, di-, tri-, tetra-fucoses were also observed on the site N184 in early HCC patients but at a low level of $0.38 \pm 0.12\%$. The tetra-antennary glycans were distinctly decreased in cirrhosis ($0.19 \pm 0.04\%$) and normal controls ($0.04 \pm 0.02\%$). In addition, only one tetra-antennary structure, A4G4F2S1, was observed in normal subjects, and *N*-glycans with multiple ($n \geq 3$) fucoses or sialic acids were not detected in cirrhosis. The penta-antennary *N*-glycan A5G5S1 was identified on the site N184 for the first time, accounting for 0.317%, 0.075%, and 0.029% in early HCC, cirrhosis, and normal controls, respectively (Table S3).

As shown in Figure 4B, on the site N207, the biantennary di- or monosialylated *N*-glycans, A2G2S2 and A2G2S1, were also the top two glycoforms in all samples. However, the abundance of A2G2S1 increased when compared to the site N184, with a decrease in A2G2S2 level. The mean levels of the A2G2S2 and A2G2S1 on site N207 are 38.98% and 27.16% in early HCC, 38.91% and 28.41% in cirrhosis, and 36.09% and 40.27% in normal controls, respectively. The levels of triantennary *N*-glycans with 1, 2, or 3 sialic acids were increased on site N207 compared to site N184, with a contribution of $4.24 \pm 0.48\%$, $5.98 \pm 1.90\%$, and $4.80 \pm 0.56\%$, respectively, of all glycoforms on site N207 in early HCC. The tetra-antennary *N*-glycans were of low abundance in early HCC (1.22%) and cirrhosis (0.67%), while they were not present in normal controls.

A different distribution of the glycoforms was observed on site N241 than the other two sites. Site N241 has the most diverse *N*-glycosylation due to the complexity of tetra-antennary and hyper-fucosylation and sialylation, where in this study 38 *N*-glycans were identified on this site containing up to four sialic acids and five fucoses in early HCC patients (Table S3). As shown in Figure 4C, though the most abundant glycan was also the

biantennary disialylated *N*-glycan A2G2S2, which contributed $69.42 \pm 2.25\%$ in early HCC, $65.65 \pm 3.92\%$ in cirrhosis, and $82.76 \pm 1.73\%$ in normal controls, the level of A2G2S1 in each group was only $2.0 \pm 0.15\%$, $3.22 \pm 0.33\%$, and $2.84 \pm 0.2\%$, respectively. The second most abundant glycan was the triantennary disialylated structure A3G3S2 (4.42%), followed by A2G2F2S1 (3.64%), A3G3S3 (3.55%), and A3G3F1S3 (2.02%) in early HCC. Interestingly, among the top 10 most abundant glycoforms on site N241 was a triantennary *N*-glycan with four fucoses A3G3F4S1, which had a contribution of $1.24 \pm 0.35\%$ in early HCC and $1.0 \pm 0.40\%$ in cirrhosis but was absent in normal controls. Notably, the overall level of tetra-antennary *N*-glycans was higher in early HCC (4.49%) than in cirrhosis (3.75%) and normal controls (0.91%). The relative quantitation provides a reasonable measure in electrospray mass spectrometry which is a concentration-dependent technique. It was not possible to measure variation in ionization efficiency of the wide variety of peptides, and it is reasonable to expect that normalization helps to neutralize some of this effect.⁴²

Supplemental Figure S7 summarizes the overall distribution of bi-, tri-, and tetra-antennary *N*-glycoforms on sites N184, N207, and N241 of serum Hp in early HCC, cirrhosis, and normal controls, respectively. In general, the biantennary *N*-glycoforms were the most abundant, which had a >70% contribution in all cases. The triantennary glycoforms were medium abundant, with a 10–28% contribution on the individual sites. The tetra-antennary glycoforms were at low abundance, accounting for 0.05–5% on each site. Interestingly, the tetraantennary glycoforms were significantly elevated on the site N241 in early HCC and cirrhosis patients compared to the other two sites, where this site has the most complexity of hyper-fucosylation and sialylation in liver disease as reported previously.¹²

A recent study identified and quantitated 96 glycopeptides in Hp derived from patient serum between gastric cancer versus healthy controls with a multispecific protease, Pronase, to generate short peptides at the 4 glycosties.¹⁴ Among the 96 glycopeptides were 20 at N184, 25 at N207, 26 at N211, and 25 at N241, where the glycan moieties were primarily bi- and triantennary glycans with one or two sialic acid residues or/and one fucose.¹⁴ Herein, we identified more tetra-antennary glycoforms at sites N184 and N241 with multiple fucose and sialic acid residues in serum Hp from HCC and cirrhosis patients, probably due to the different types of diseases. The Goldman group also revealed increased fucosylation of serum Hp in HCC and liver cirrhosis with multiple fucoses found in the intact glycopeptides of serum Hp.¹¹ After treatment with neuraminidase to remove sialic acids, up to 6 fucoses were observed in site-specific glycopeptides in HCC.¹¹

Differentially Expressed *N*-Glycopetides in Early HCC Compared to Cirrhosis

Since there were clear differences in site-specific *N*-glycopeptides between liver disease and healthy controls, we further focused on the determination of the differentially expressed *N*-glycopeptides between early HCC and cirrhosis. Five site-specific *N*-glycopeptides were found to be significantly elevated in early HCC compared to cirrhosis (Table 2), with at least a 2-fold change and $p < 0.05$. Among them were one glycoform on site N184 (A3G3F2S3) and four glycoforms on site N241 (A3G3F2S3, A4G4S4, A4G4F2S3, and A4G4F2S4), which are mainly fully sialylated tri- or tetra-antennary *N*-glycans with either no fucose or two fucoses present. The ratio and p value of these site-specific glycopeptides between early

HCC and normal are also listed in Table 2 as a reference. The result showed that the glycopeptide on site N241 with A4G4F2S4 was exclusively expressed in early HCC but was absent in cirrhosis and normal controls ($p < 0.001$). Moreover, the glycopeptides with A3G3F2S3 on sites N184 and N241 were both absent in normal controls, while the glycopeptides on site N241 with A4G4S4 and A4G4F2S3 were significantly decreased in normal controls compared to early HCC ($p = 0.001$).

To evaluate the reproducibility of these changes in early HCCs, we further repeated the analysis for 3 early-stage HCC samples (i.e., HCC1, HCC3, HCC4), starting with a new 20 μL aliquot for each sample. Taking the *N*-glycopeptide of N241_A3G3F2S3 as an example, the relative standard deviation (RSD) of its abundance in the two technical replicates was 12.6%, 13.5%, and 10.1% for HCC1, HCC3, and HCC4, respectively. The result demonstrated that this *N*-glycopeptide, though at a low abundance, was consistently present in these early-stage HCC samples.

The scatter plots of these differentially expressed glycopeptides between early HCC and cirrhosis are shown in Figure 5. The result indicated that the site-specific sialylation is of importance for determination of early HCC from cirrhosis, where the site-specific *N*-glycopeptides with fully sialylated tri- or tetra-antennary glycans were significantly associated with early HCC ($p < 0.01$). This result further illustrated that the bifucosylated tri- and tetra-antennary glycoforms can discriminate early HCC from cirrhosis, which was consistent with our glycomics studies on serum haptoglobin in early HCC.^{6,8} The increased tetra-antennary glycoforms with hyper-fucosylation (up to 5 fucoses) on site N241 were observed in a portion of early HCC patients of this sample set but did not significantly distinguish early HCC from cirrhosis.

CONCLUSIONS

In this study, we developed a method for differential analysis of site-specific intact *N*-glycopeptide structures between disease states. This involved a workflow that incorporated the LC-ET_hCD MS/MS method and data quantitation softwares into characterizing the alterations in site-specific *N*-glycopeptides of haptoglobin in patient serum among early HCC, liver cirrhosis, and healthy controls. With the advantage of ET_hCD-MS/MS fragmentation²² and strict MS1/MS2 tolerances and Byonic score, a total of 98 site-specific *N*-glycopeptides were identified in serum haptoglobin with high confidence (FDR < 1%). A significantly increased number of site-specific *N*-glycopeptides was identified in early HCC (93 *N*-glycopeptides) and liver cirrhosis (87 *N*-glycopeptides) relative to healthy controls (68 *N*-glycopeptides). The increased complexity of *N*-glycopeptides in liver diseases compared to healthy controls was mainly due to glycopeptides containing tetra-antennary *N*-glycans with multiple fucoses and sialic acids.

Automated quantitative analysis using Byologic software, with the flexibility in quantitation at both total glycopeptide level and glycosite level, greatly facilitated the label-free glycopeptide quantitation between disease groups. Site-specific quantitation at sites N184, N207, and N241 of serum Hp showed that the biantennary glycopeptides were the most abundant at each site, while the triantennary glycopeptides were medium abundant, with a

total contribution of 10–28%. Although the tetraantennary glycopeptides have a low contribution of <5% at each site, they were found distinctly increased in early HCC and cirrhosis patients compared to healthy controls. Also, five site-specific *N*-glycopeptides on sites N184 and N241 were significantly elevated in early HCC compared to cirrhosis ($p < 0.05$), where these were fully sialylated tri- or tetra-antennary *N*-glycans with either no fucose or two fucoses present. The increased tetra-antennary glycoforms with hyper-fucosylation (up to 5 fucoses) were observed in a portion of early HCC patients of this sample set but did not significantly distinguish early HCC from cirrhosis.

This work demonstrated that EThcD together with the powerful Byonic/Byologic software has provided a means to study the detailed glycan structures with quantitation information for site-specific glycopeptides and to perform differential quantitative analysis between these structures. We have done this on an initial test set to demonstrate its capabilities for identifying often minor but important structures, which will be useful for future biomarker studies on site-specific *N*-glycopeptides for discrimination of cancer from cirrhosis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We acknowledge the support of this work from the National Cancer Institution under grants 1R01 CA160254 (DML), U01 CA225753 (DML) and R50 CA221808 (JZ) and from the National Institutes of Health through grant R01 GM 49500 (DML). MB acknowledges the support from the National Institute of General Medical Science through the grant R21 GM122634 (MB). LL acknowledges funding support from the National Institutes of Health through grants R01AG052324 (LL) and R21AG055377 (LL). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to LL) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison.

REFERENCES

- (1). Naitoh A; Aoyagi Y; Asakura H Highly enhanced fucosylation of serum glycoproteins in patients with hepatocellular carcinoma. *J. Gastroenterol. Hepatol* 1999, 14 (5), 436–45. [PubMed: 10355508]
- (2). de Oliveira RM; Ornelas Ricart CA; Araujo Martins AM Use of Mass Spectrometry to Screen Glycan Early Markers in Hepatocellular Carcinoma. *Front. Oncol* 2018, 7, 328. [PubMed: 29379771]
- (3). Clerc F; Reiding KR; Jansen BC; Kammeijer GS; Bondt A; Wuhler M Human plasma protein N-glycosylation. *Glycoconjugate J* 2016, 33 (3), 309–43.
- (4). Ang IL; Poon TC; Lai PB; Chan AT; Ngai SM; Hui AY; Johnson PJ; Sung JJ Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: a glycoproteomic approach. *J. Proteome Res* 2006, 5 (10), 2691–700. [PubMed: 17022640]
- (5). Kurosky A; Barnett DR; Lee TH; Touchstone B; Hay RE; Arnott MS; Bowman BH; Fitch WM Covalent structure of human haptoglobin: a serine protease homolog. *Proc. Natl. Acad. Sci. U. S. A* 1980, 77 (6), 3388–92. [PubMed: 6997877]
- (6). Zhu J; Lin Z; Wu J; Yin H; Dai J; Feng Z; Marrero J; Lubman DM Analysis of serum haptoglobin fucosylation in hepatocellular carcinoma and liver cirrhosis of different etiologies. *J. Proteome Res* 2014, 13 (6), 2986–97. [PubMed: 24807840]
- (7). Huang Y; Zhou S; Zhu J; Lubman DM; Mechref Y LC-MS/MS isomeric profiling of permethylated N-glycans derived from serum haptoglobin of hepatocellular carcinoma (HCC) and cirrhotic patients. *Electrophoresis* 2017, 38 (17), 2160–2167. [PubMed: 28543513]

- (8). Zhang Y; Zhu J; Yin H; Marrero J; Zhang XX; Lubman DM ESI-LC-MS Method for Haptoglobin Fucosylation Analysis in Hepatocellular Carcinoma and Liver Cirrhosis. *J. Proteome Res* 2015, 14 (12), 5388–95. [PubMed: 26503433]
- (9). Frost DC; Li L Recent advances in mass spectrometry-based glycoproteomics. *Adv. Protein Chem. Struct. Biol* 2014, 95, 71–123. [PubMed: 24985770]
- (10). Takahashi S; Sugiyama T; Shimomura M; Kamada Y; Fujita K; Nonomura N; Miyoshi E; Nakano M Site-specific and linkage analyses of fucosylated N-glycans on haptoglobin in sera of patients with various types of cancer: possible implication for the differential diagnosis of cancer. *Glycoconjugate J* 2016, 33 (3), 47182.
- (11). Pompach P; Brnakova Z; Sanda M; Wu J; Edwards N; Goldman R Site-specific glycoforms of haptoglobin in liver cirrhosis and hepatocellular carcinoma. *Mol. Cell. Proteomics* 2013, 12 (5), 1281–93. [PubMed: 23389049]
- (12). Sanda M; Pompach P; Brnakova Z; Wu J; Makambi K; Goldman R Quantitative liquid chromatography-mass spectrometry-multiple reaction monitoring (LC-MS-MRM) analysis of site-specific glycoforms of haptoglobin in liver disease. *Mol. Cell. Proteomics* 2013, 12 (5), 1294–305. [PubMed: 23389048]
- (13). Chandler KB; Pompach P; Goldman R; Edwards N Exploring site-specific N-glycosylation microheterogeneity of haptoglobin using glycopeptide CID tandem mass spectra and glycan database search. *J. Proteome Res* 2013, 12 (8), 3652–66. [PubMed: 23829323]
- (14). Lee J; Hua S; Lee SH; Oh MJ; Yun J; Kim JY; Kim JH; Kim JH; An HJ Designation of fingerprint glycopeptides for targeted glycoproteomic analysis of serum haptoglobin: insights into gastric cancer biomarker discovery. *Anal. Bioanal Chem* 2018, 410 (6), 1617–1629. [PubMed: 29285644]
- (15). Myslinska S; Palmisano G; Hojrup P; Thaysen-Andersen M Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem* 2010, 82 (13), 5598–609. [PubMed: 20536156]
- (16). Zacharias LG; Hartmann AK; Song E; Zhao J; Zhu R; Mirzaei P; Mechref Y HILIC and ERLIC Enrichment of Glycopeptides Derived from Breast and Brain Cancer Cells. *J. Proteome Res* 2016, 15 (10), 3624–3634. [PubMed: 27533485]
- (17). Selman MH; Hemayatkar M; Deelder AM; Wührer M Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal. Chem* 2011, 83 (7), 2492–9. [PubMed: 21366235]
- (18). Lee JY; Lee HK; Park GW; Hwang H; Jeong HK; Yun KN; Ji ES; Kim KH; Kim JS; Kim JW; Yun SH; Choi CW; Kim SI; Lim JS; Jeong SK; Paik YK; Lee SY; Park J; Kim SY; Choi YJ; Kim YI; Seo J; Cho JY; Oh MJ; Seo N; An HJ; Kim JY; Yoo JS Characterization of Site-Specific N-Glycopeptide Isoforms of alpha-1-Acid Glycoprotein from an Interlaboratory Study Using LC-MS/MS. *J. Proteome Res* 2016, 15 (12), 4146–4164. [PubMed: 27760464]
- (19). Yin H; An M; So PK; Wong MY; Lubman DM; Yao Z The analysis of alpha-1-antitrypsin glycosylation with direct LC-MS/MS. *Electrophoresis* 2018, 39 (18), 2351–2361. [PubMed: 29405331]
- (20). Mechref Y Use of CID/ETD mass spectrometry to analyze glycopeptides. *Current Protocols in Protein Science*; John Wiley & Sons, Inc: Hoboken, NJ, 2012; Chapter 12, Unit 12 11.1–11.
- (21). Singh C; Zampronio CG; Creese AJ; Cooper HJ Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J. Proteome Res* 2012, 11 (9), 4517–25. [PubMed: 22800195]
- (22). Yu Q; Wang B; Chen Z; Urabe G; Glover MS; Shi X; Guo LW; Kent KC; Li L Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glyco-proteome Characterization. *J. Am. Soc. Mass Spectrom* 2017, 28 (9), 1751–1764. [PubMed: 28695533]
- (23). Chen Z; Yu Q; Hao L; Liu F; Johnson J; Tian Z; Kao WJ; Xu W; Li L Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD). *Analyst* 2018, 143 (11), 2508–2519. [PubMed: 29687791]
- (24). Glover MS; Yu Q; Chen Z; Shi X; Kent KC; Li L Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD

fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass Spectrom* 2018, 427, 35–42.

- (25). Chen Z; Glover MS; Li L Recent advances in ion mobility-mass spectrometry for improved structural characterization of glycans and glycoconjugates. *Curr. Opin. Chem. Biol* 2018, 42, 1–8. [PubMed: 29080446]
- (26). Tsai PL; Chen SF A Brief Review of Bioinformatics Tools for Glycosylation Analysis by Mass Spectrometry. *Mass Spectrom* 2017, 6, S0064.
- (27). Bern M; Kil YJ; Becker C Byonic: advanced peptide and protein identification software. *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc: Hoboken, NJ, 2012; Chapter 13, Unit 13.20.
- (28). Liu MQ; Zeng WF; Fang P; Cao WQ; Liu C; Yan GQ; Zhang Y; Peng C; Wu JQ; Zhang XJ; Tu HJ; Chi H; Sun RX; Cao Y; Dong MQ; Jiang BY; Huang JM; Shen HL; Wong CCL; He SM; Yang PY pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat. Commun* 2017, 8 (1), 438. [PubMed: 28874712]
- (29). Park GW; Kim JY; Hwang H; Lee JY; Ahn YH; Lee HK; Ji ES; Kim KH; Jeong HK; Yun KN; Kim YS; Ko JH; An HJ; Kim JH; Paik YK; Yoo JS Integrated GlycoProteome Analyzer (I-GPA) for Automated Identification and Quantitation of Site-Specific N-Glycosylation. *Sci. Rep* 2016, 6, 21175. [PubMed: 26883985]
- (30). Strum JS; Nwosu CC; Hua S; Kronewitter SR; Seipert RR; Bachelor RJ; An HJ; Lebrilla CB Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. *Anal. Chem* 2013, 85 (12), 5666–75. [PubMed: 23662732]
- (31). Kim JW; Hwang H; Lim JS; Lee HJ; Jeong SK; Yoo JS; Paik YK gFinder: A Web-Based Bioinformatics Tool for the Analysis of N-Glycopeptides. *J. Proteome Res* 2016, 15 (11), 41164125.
- (32). He L; Xin L; Shan B; Lajoie GA; Ma B GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J. Proteome Res* 2014, 13 (9), 3881–95. [PubMed: 25113421]
- (33). Woodin CL; Hua D; Maxon M; Rebecchi KR; Go EP; Desaire H GlycoPep grader: a web-based utility for assigning the composition of N-linked glycopeptides. *Anal. Chem* 2012, 84 (11), 4821–9. [PubMed: 22540370]
- (34). Zhu Z; Hua D; Clark DF; Go EP; Desaire H GlycoPep Detector: a tool for assigning mass spectrometry data of N-linked glycopeptides on the basis of their electron transfer dissociation spectra. *Anal. Chem* 2013, 85 (10), 5023–32. [PubMed: 23510108]
- (35). Jansen BC; Falck D; de Haan N; Hipgrave Ederveen AL; Razdorov G; Lauc G; Wührer M LaCyTools: A Targeted Liquid Chromatography-Mass Spectrometry Data Processing Package for Relative Quantitation of Glycopeptides. *J. Proteome Res* 2016, 15 (7), 2198–210. [PubMed: 27267458]
- (36). Rosenberg ES; Hall EW; Sullivan PS; Sanchez TH; Workowski KA; Ward JW; Holtzman D Estimation of State-Level Prevalence of Hepatitis C Virus Infection, US States and District of Columbia, 2010. *Clin. Infect. Dis* 2017, 64 (11), 15731581.
- (37). Zhu J; Wu J; Yin H; Marrero J; Lubman DM Mass Spectrometric N-Glycan Analysis of Haptoglobin from Patient Serum Samples Using a 96-Well Plate Format. *J. Proteome Res* 2015, 14 (11), 4932–9. [PubMed: 26448449]
- (38). Tyanova S; Temu T; Sinitcyn P; Carlson A; Hein MY; Geiger T; Mann M; Cox J The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 2016, 13 (9), 731–40. [PubMed: 27348712]
- (39). Deng W; Wang Y; Liu Z; Cheng H; Xue Y HemI: a toolkit for illustrating heatmaps. *PLoS One* 2014, 9 (11), e111988. [PubMed: 25372567]
- (40). Bern MW; Kil YJ Two-dimensional target decoy strategy for shotgun proteomics. *J. Proteome Res* 2011, 10 (12), 5296–301. [PubMed: 22010998]
- (41). Lee LY; Moh ES; Parker BL; Bern M; Packer NH; Thaysen-Andersen M Toward Automated N-Glycopeptide Identification in Glycoproteomics. *J. Proteome Res* 2016, 15 (10), 39043915.

- (42). Song E; Pyreddy S; Mechref Y Quantification of glycopeptides by multiple reaction monitoring liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom* 2012, 26 (17), 1941–54. [PubMed: 22847692]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Workflow of quantitative LC-ETHeD-MS/MS analysis of intact *N*-glycopeptides derived from serum haptoglobin in patients between early hepatocellular carcinoma (HCC), liver cirrhosis, and normal subjects.

Figure 2.

Reproducibility analysis of 4 technical replicates of a normal serum sample that have been processed at four different times. (a) Pearson correlation coefficient R^2 values for the binary comparison of the 4 technical replicates. R^2 values range from 0.977 to 0.994, demonstrating good reproducibility of the method. (b) Pie chart of the frequency of glycopeptide quantitation, which were present in all 4 technical replicates, in 2–3, or only in 1. (c) Reproducibility of the AUC of five glycopeptides at site N241 which covered five orders of magnitude (10^6 – 10^{10}) for the 4 technical replicates. Relative standard deviation (RSD) of each glycopeptide for the 4 replicates was provided.

Figure 3. Representative MS/MS spectra of fingerprint *N*-glycopeptides of (A) MVSHHN¹⁸⁴LTTGATLINE with glycan A3G3F2S2 and (B) VVLHPN²⁴¹YSQyDIGLIK with glycan A3G3F3S1 and their *m/z* error plots. Both of the glycopeptides were quadruply charged (*z* = 4). Oxonium ions such as HexNAc (*m/z* 204.09), HexHexNAc (*m/z* 366.13), NeuAc (*m/z* 292.10), NeuAc-3/4O (*m/z* 274.09), and HexNAcHexNeuAc (*m/z* 657.23) were present at high intensity in the spectra. *b* and *y* fragments from the peptide backbone were observed as well but at low intensity. Glycosidic fragments with the loss of monosaccharides from the parent glycopeptide were well characterized in the EThcD MS/MS spectra. Specific diagnostic fragment ions further provided detailed structural information such as core and outer arm fucosylation (marked by red arrows). For example, the diagnostic fragment ion of HexHexNAcFuc (*m/z* 512.20) confirmed the presence of the outer-arm fucose residue, while the ions of peptide+2HexNAc+Fuc (*m/z* 1153.54 in Figure 3A and *m/z* 1174.12 in Figure 3B) confirmed the core fucose residue. *m/z* error plots illustrated the

accuracy of the fragment assignments where the mass of oxonium ions matched exactly with the calculated value, and in the case of glycopeptide fragments, the difference between the observed m/z and the calculated m/z was within the range of 0.004 Da.

Figure 4. *N*-Glycopeptide quantitation at glycosite level where the 10 most abundant glycopeptides on sites N184 (A), N207 (B), and N241 (C) in early HCC (red bars) were present, with a comparison of their abundances in cirrhosis (green bars) and normal controls (blue bars), respectively. Error bar shows the standard deviation across the 5 individual patients in each disease group.

Figure 5. Scatter plots of the 5 differentially expressed site-specific *N*-glycopeptides between early HCC and cirrhosis. (A and B) *N*-Glycopeptides containing the glycan A3G3F2S3 on sites N184 and N241, respectively. (C, D, and E) *N*-glycopeptides on site N241 containing glycans A4G4S4, A4G4F2S3, and A4G4F2S4, respectively.

Table 1.

Clinical Characteristics of Patients Involved in This Study

disease diagnosis	early HCC	cirrhosis	normal
no. ^a	5	5	5
gender % (M/F)	80/20	60/40	80/20
age (mean $\pm$ SD)	63.6 $\pm$ 10	58.6 $\pm$ 9	62.8 $\pm$ 15
etiology	HCV ^b	HCV	NA
AFP lever ^c (median, ng/mL)	5.3	7.3	NA
MELD ^d score	6 $\pm$ 3.4	8 $\pm$ 2.2	NA
tumor size (cm)	2.3 $\pm$ 0.2	NA	NA
TNM stage % (I/II/III/IV)	60/40/0/0	NA	NA

^a Samples and their AFP levels were provided by Division of Gastroenterology, University of Michigan.

^b HCV: Hepatitis C virus.

^c Samples and their AFP levels were provided by Division of Gastroenterology, University of Michigan.

^d MELD: Model for end stage liver disease.

Site-Specific N-Glycopeptides Significantly Elevated in Early HCC Compared to Cirrhosis and Normal Controls

Table 2.

peptide sequence	glycan composition	early HCC/Cirr	p value	early HCC/N	p value
MVSHHN ¹⁸ LTTGATLINE	A3G3F2S3	3.21	0.0028	NA ^a	<0.0001
WLHPN ¹¹ YSQyD	A3G3F2S3	4.16	0.044	NA	0.0006
	A4G4S4	3.37	0.0038	2.67	0.001
	A4G4F2S3	2.06	0.034	43.93	0.0004
	A4G4F2S4	NA	0.0004	NA	0.0004

^aNA represents that the glycopeptide was uniquely present in early HCC but was absent in cirrhosis or normal control samples.