Published in final edited form as:

J Proteomics. 2016 September 02; 146: 90–98. doi:10.1016/j.jprot.2016.06.003.

Improvement of core-fucosylated glycoproteome coverage via alternating HCD and ETD fragmentation

Cheng Ma^{a,*}, Jingyao Qu^a, Xu Li^a, Xinyuan Zhao^b, Lei Li^a, Cong Xiao^a, Garrett Edmunds^a, Ebtesam Gashash^a, Jing Song^a, and Peng George Wang^{a,*}

^aDepartment of Chemistry, Center for Diagnostics & Therapeutics, Georgia State University, Atlanta, GA 30303, United States

^bNational Institute of Biological Sciences, Beijing 102206, People's Republic of China

Abstract

Core-fucosylation (CF) plays important roles in regulating biological processes in eukaryotes. Alterations of CF-glycosites or CF-glycans in bodily fluids correlate with cancer development. Therefore, global research of protein core-fucosylation with an emphasis on proteomics can explain pathogenic and metastasis mechanisms and aid in the discovery of new potential biomarkers for early clinical diagnosis. In this study, a precise and high throughput method was established to identify CF-glycosites from human plasma. We found that alternating HCD and ETD fragmentation (AHEF) can provide a complementary method to discover CF-glycosites. A total of 407 CF-glycosites among 267 CF-glycoproteins were identified in a mixed sample made from six normal human plasma samples. Among the 407 CF-glycosites, 10 are without the N-X-S/T/C consensus motif, representing 2.5% of the total number identified. All identified CF-glycopeptide results from HCD and ETD fragmentation were filtered with neutral loss peaks and characteristic ions of GlcNAc from HCD spectra, which assured the credibility of the results. This study provides an effective method for CF-glycosites identification and a valuable biomarker reference for clinical research.

Biological significance: CF-glycosylation plays an important role in regulating biological processes in eukaryotes. Alterations of the glycosites and attached CF-glycans are frequently observed in various types of cancers. Thus, it is crucial to develop a strategy for mapping human CF-glycosylation. Here, we developed a complementary method via alternating HCD and ETD fragmentation (AHEF) to analyze CF-glycoproteins. This strategy reveals an excellent complementarity of HCD and ETD in the analysis of CF-glycoproteins, and provides a valuable biomarker reference for clinical research.

*Corresponding authors. cma@gsu.edu (C. Ma), pwang11@gsu.edu (P.G. Wang).

Author contributions

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

Transparency document

The [Transparency document](#) associated with this article can be found, in online version.

Appendix A. Supplementary data

Two supplementary figures (Figs. S1–S2) and Table S1 that contain additional experimental data and supporting Information, which detail the result of identified CF-glycopeptides and CF-glycoproteins by chemical and enzymatic methods. Supplementary data associated with this article can be found in the online version, at <http://dx.doi.org/10.1016/j.jprot.2016.06.003>.

Keywords

Human plasma; Core-fucosylation; HCD; ETD

1. Introduction

Core-fucosylation (CF) is a type of N-linked glycosylation in which α 1,6 fucose is added to the innermost GlcNAc residue. It plays important roles in regulating biological processes in eukaryotes. Alterations of the glycosites and attached glycans are frequently observed in various types of cancers - including liver, ovarian, breast, prostate, and lung cancers - and are involved in almost every aspect of tumor progression [1–3]. Diagnostic biomarkers of clinical cancer are often glycoproteins, one of which, fucosylated α -fetoprotein (AFP-L3), has been approved by the Food and Drug Administration (FDA) as a diagnostic biomarker in hepatocellular carcinoma (HCC) [4,5] (the concentration of AFP is less than 20 ng/mL in healthy human serum). Moreover, Golgi membrane protein 1 (GOLM1), also known as GP73, is significantly increased in patients with HCC, even in those who had normal AFP levels [6]. Another two fucosylated glycoproteins, kininogen and alpha-1-antitrypsin (a1AT), have been identified as candidates of hepatic tumor markers [7, 8]. In addition, it has been proven that core-fucosylated haptoglobin is a much better biomarker in ovarian carcinoma than simply monitoring the expression of the haptoglobin [9]. All the examples show that, during pathological processes, CF-glycoprotein changes occur not only in protein abundance but also in the occupancy of CF-glycosylation. Alternatively, given the long known alterations in glycans associated with cancer, it is highly likely that monitoring of specific glycan sequence at specific-site of glycoproteins will have much higher sensitivity and specificity for early detection of cancer [10,11]. However, most current diagnostic tests only measure the expression of the proteins by ELISA and western blotting. Although these techniques are easily performed, the diagnostic results may be faulty. For example, alpha-fetal protein (AFP) is highly expressed in the patients of liver cirrhosis, which may provide an erroneous diagnosis of liver cancer [12].

Because of its important biological implications, the research of CF-glycoproteome with MS is of growing importance to biological and analytical scientists. However, several key problems restrict the large scale mapping of CF-glycoproteome. 1) CF-glycoproteins/glycopeptides capture. By now, the major approach of glycoprotein capture is lectin affinity. Some plant lectins, such as *Lens culinaris* agglutinin (LCA) and *Aleuria aurantia* lectin (AAL), have been successfully applied in CF-glycoproteins/glycopeptides enrichment. However, the affinity and specificity of lectins for glycoproteins is generally low, thereby causing the loss of glycoproteins [13] and non-specific capture of non-glycoproteins [14]. 2) Complex MS/MS fragment ions of intact CF-glycopeptides. Due to the complexity of MS/MS spectra and the lacking of commercial software, large scale mapping of intact N-glycopeptides is still very difficult. Endo-glycosidase F3 (Endo F3) recognizes core fucosylated N-glycans and cleave between the two GlcNAc. The residual Fuc α 1,6GlcNAc can be considered as a label for the CF-glycosites. Jia and his coworkers analyzed the simplified CF-glycopeptides using neutral loss-dependent MS³ to identify CF-glycosites. A

total of 107 CF-glycosites and 70 CF-glycoproteins were identified from healthy human plasma [15].

At present, several MS fragmentation techniques have been applied for proteomics research, such as collision-induced dissociation (CID) [16], electron capture dissociation (ECD) [17], electron-transfer dissociation (ETD) [18], and higher energy collisional dissociation (HCD) [19]. All of these fragmentation techniques are not only useful for peptides sequencing, but also critical for protein post-translational modification [20]. ETD cleaves randomly along the peptide backbone (generating c and z ions) while preserving side chains and modifications such as phosphorylation and glycosylation. Therefore, this technique is extremely useful in the research of PTM analysis. However, the limitations of the ETD technique are also obvious: 1) ETD only works well with precursor ions carrying more than 3 charges; 2) the reaction time of ETD is much longer than CID (100–200 ms vs. 5–10 ms). On the other hand, HCD breaks ions in a collision cell rather than an ion trap, and then transfers them back through the C-trap for high resolution analysis in an Orbitrap [21]. Compared with ion trap-based CID, HCD has high resolution ion detection, increased ion fragmentation and no low-mass cut-off, resulting in higher quality MS/MS spectra [22]. In our previous experiment, we developed CF-glycoproteins with low- and high-HCD fragmentation, and identified a total of 349 unique CF-glycosites and 209 CF-glycoproteins from human plasma [23] and 679 CF-glycosites and 490 CF-glycoproteins from mouse liver and brain [24]. Tan et al. applied a similar method to analyze CF-glycoproteins from pancreatic cancer patient plasma, identifying a maximum of 350 CF sites among 193 CF proteins using an optimized LCA enrichment method [25]. Although HCD fragmentation has been widely used in CF-proteome research, it still has the following problems: 1) a lot of simplified CF-spectra are not thoroughly fragmented with constant normalized collision energy (NCE). 2) The fragmented ions of simplified CF-glycopeptides by HCD remain complex because α -1,6 linked core-fucose are fragile and can fall off from GlcNAc easily.

Recently, combined fragmented techniques have been used for the analysis of intact glycoproteins [26,27]. Vakhrushev et al. used a combined strategy of HCD and ETD fragmentation to map human GalNAc-type O-glycoproteome with SimpleCells, and identified a total of 259 glycoproteins from three cell lines [28]. Go et al. analyzed N-glycosylation and O-glycosylation of HIV-1 gp120, derived from clade C transmitted/founder virus 1086.C expressed in Chinese hamster ovary (CHO) and human embryonic kidney containing T antigen (293T) cell lines. He applied a combined CID and ETD fragmentation techniques, and compared the difference in glycosylation structures of gp120 in these two culture system [29]. Singh et al. developed a method to analyze the N-glycoproteins via HCD Product Ion-Triggered ETD (HCD PI ETD) [30]. By using this technology, the glycosylation of ribonuclease B and immunoglobulin G was described. This foundational research gives us confidence to study the CF-glycoproteome through combined fragmentation.

Herein, an approach was shown to identify simplified CF-glycopeptides via alternating HCD and ETD fragmentation (AHEF). This approach incorporates the advantages of the two modes of fragmentation for protein posttranslational modification research. Our previous study demonstrated that glycosidic bonds are easier to break than peptide bonds. Fragmented

ions from GlcNAc can be observed in the MS/MS spectra of HCD fragmentation [31]. Thus, HCD spectra are used for the determination of CF-glycopeptides. Complementary to the strategy of HCD, ETD enables the elucidation of CF-glycosites by maintaining the glycan-peptide linkage. Thus, more CF-glycosites can be mapped with AHEF. In total, 407 CF-sites and 267 CF-glycoproteins from normal human plasma were identified with AHEF, 35% more than HCD fragmentation alone. In addition, ten CF-glycosites are identified without the N-X-S/T/C consensus motif, which occupied almost 2.5% of total identified CF-glycosites. This approach provides an effective method for large scale CF-glycosite identification. Furthermore, this technique can provide a valuable biomarker reference for clinical CF-glycoprotein research.

2. Materials and methods

2.1. Materials and chemicals

Endoglycosidase F3 (Endo F3) and formic acid (FA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Methyl- α -D-mannopyranoside was purchased from Fluka (St. Louis, MO). ZIC-HILIC media was acquired from Merck (Darmstadt, Germany). Sequencing grade porcine trypsin was purchased from Promega (Madison, WI); LCH-sepharose 4B was purchased from GE Healthcare (Little Chalfont, UK). Deionized water was produced by a Milli-Q A10 system from Millipore (Bedford, MA, USA). HPLC grade acetonitrile (ACN) was purchased from J. T. Baker Inc. Iodoacetamide, and DTT were purchased from ACROS ORGANIC. 3M Empore C8 disk was bought from 3M Bioanalytical Technologies (St. Paul, MN, USA). Filter YM-30 (30 kD) and zip-tip C18 were purchased from Millipore. Other materials, such as NaCl, CaCl₂, MgCl₂, and MnCl₂, were purchased from Sigma-Aldrich.

Normal human plasma samples were supplied by Cancer Institute & Hospital of Shanxi Province (Taiyuan, Shanxi, P.R. China). Previous institutional ethical approval was obtained and volunteers in the study gave written informed consent. The ethics committee approved the research protocol.

2.2. CF-glycoproteins capture

Six normal human plasma samples were pooled to create a mixed sample with equal quantity for subsequent analysis. In this study, the chosen lectin *Lens culinaris* agglutinin (LCA, LCH) showed high binding affinity to core-fucosylated glycans [32]. Approximately 10 μ L mixed human plasma sample (500 μ g protein) was subjected to CF-glycoprotein enrichment procedure as reported [33]: The plasma sample was lyophilized and resuspended with 300 μ L of binding buffer (20 mM Tris-HCl buffered saline, 0.5 M NaCl, 1 mM CaCl₂, 1 mM MnCl₂, pH 7.3). The sample was then mixed with 300 μ L LCH lectin-agarose beads slurry in a spin column and incubated at 20 °C for 1 h. After binding, the spin column was washed three times with 300 μ L of binding buffer. Separately, 400 μ L and two iterations of 300 μ L of 200 mM methyl α -D-mannopyranoside dissolved in binding buffer solution was added to plasma dilution, allowed to sit for 30 min, and then spun down. The plasma was eluted by 400 μ L of elution solution and 100 μ L deionized water then transferred to a 3 kD cut-off. The entire tube was centrifuged at 12,000 $\times g$ for 1 h at 4 °C. Then, 450 μ L of deionized water was added to the top portion of the Amicon filter twice and centrifuged at

8000 × *g* for 3 h at 4 °C. Lastly, 100 µL of 15% ACN was added three times to the Amicon filter and the enriched proteins were lyophilized.

2.3. Trypsin digestion

Approximately 60 µg enriched plasma protein was subjected to the filter aided sample preparation (FASP) procedure as reported [34]. In brief, approximately 60 µg enriched CF-glycoproteins were mixed with 30 µL of lysis buffer (20 mM Tris-HCl, 4% (v/v) SDS, 100 mM DTT, pH 7.6, and incubated for 5 min at 95 °C, then added into 200 µL of UA solution (8 M urea in 0.1 M Tris/HCl, pH 8.5), loaded into the 30 kD Microcon filtration devices, and centrifuged at 13,000*g* until the volume is less than 10 µL. The concentrates were washed in the devices with 200 µL of UA solution twice. After centrifugation, the concentrates were mixed with 100 µL of 50 mM iodoacetamide in UA solution and incubated in the dark at room temperature (RT) for 30 min followed by centrifugation for 20 min. Then, the protein concentrate was diluted with 200 µL of 8 M urea in 0.1 M Tris/HCl, pH 8.5, and concentrated again. This procedure was repeated twice. The samples were diluted twice with 100 µL of 40 mM NH₄HCO₃. Sequencing grade trypsin was dissolved in 100 µL 40 mM NH₄HCO₃ at an enzyme to protein ratio of 1:50 for digestion overnight at 37 °C. Subsequently, all trypsin digested peptides were collected by centrifugation of the filter units with 50 µL binding buffer for 20 min. This step was repeated 6 times with 50 µL binding buffer. The concentration of peptides was determined by Thermo Nanodrop UV-spectrometry, applying an extinction coefficient of 1.1 for 0.1% (g/L) solution at 214 nm.

2.4. CF-glycopeptides enrichment and simplification

Approximately 30 µg of digested peptides were added into 10 mg ZIC-HILIC media and enriched by Ma's procedure [35]. The strategy is as follows: a piece of C8 disk was put into 200 µL tip first; then about 10 mg ZIC-HILIC media was dissolved in 100 µL acetonitrile and injected into tips. In-solution digested peptides were re-dissolved in 80% ACN, 0.5% FA, and loaded into a 200 µL tip equilibrated with binding buffer (80% ACN, 0.5% FA). ZIC-HILIC tip was washed with 100 µL 80% ACN, 1% FA, and 19% H₂O six times and bounded peptides were eluted with 80 µL elution buffer (99% H₂O, 1% FA) three times. The concentration of enriched peptides by zic-HILIC was also determined by UV-spectrometry. Approximately 5 µg of enriched peptides were resuspended in 100 µL sodium acetate solution (100 mM, pH 4.5) and further incubated with endoglycosidase F3 overnight at 37 °C. The CF-glycopeptide treated with Endo F3 was named deglycosylated CF-glycopeptide. Zip-tip C18 was used for desalting partially deglycosylated CF-glycopeptides. Desalted peptides were stored at –80 °C awaiting analysis by LC-MS/MS.

2.5. LC-MS analysis

RP HPLC-MS experiments were performed on an LTQ-Orbitrap Elite ETD mass spectrometer (ThermoFisher) equipped with EASY-spray source and nano-LC UltiMate 3000 high performance liquid chromatography system (ThermoFisher). EASY-Spray PepMap C18 Columns (25 cm; particle size, 2 µm; pore size, 100 Å; ThermoFisher) were used for separation. Separation was achieved with a linear gradient from 3% to 40% buffer B for 80 min at a flow rate of 300 nL/min (mobile phase A: 2% ACN, 98% H₂O, 0.1% FA; mobile phase B: 80% ACN, 20% H₂O, and 0.1% FA). The LTQ-Orbitrap Elite mass

spectrometer was operated in data-dependent mode. A full-scan survey MS experiment (m/z range from 375 to 1600; automatic gain control target, 1,000,000 ions; resolution at 400 m/z , 60,000; maximum ion accumulation time, 50 ms) was acquired by the Orbitrap mass spectrometer, and ten MS/MS events for the 5 most intense ions were fragmented by alternating HCD and ETD. MS/MS spectra of HCD fragmentation were acquired in the Orbitrap analyzer with resolution of 15,000 at m/z 400 (automatic gain control target, 10,000 ions; maximum ion accumulation time, 200 ms). MS/MS spectra of ETD fragmentation were acquired in the ion-trap analyzer, and precursor ions with charge 2+ were eliminated. Activation time of ETD is set to 100 ms. MS/MS scanning model was set to the mode of centroid. The other conditions used were: temperature of 200 °C, S-lens RF level of approximately 60%, ion selection threshold of 50,000 counts for HCD.

2.6. Database searching and biological characteristics analysis

In this study, raw data was first converted to a mgf file by applying ProteomeDiscoverer 1.4. Potential CF-glycopeptide spectra were selected and respectively merged into HCD and ETD files with an in-house program. Then, the mgf files were generated to new mgf files after treated with the in-house program (the detailed description is shown in the Result and discussion section). Human protein database was downloaded from Uniprot_swissprot plus Uniprot_TrEMBL (Released on 2012-04, human, 65,493 entries), concatenated with reversed versions of all sequences. New mgf files were analyzed by pFind 2.1 software to search the human Uniprot_TrEMBL database and its reversed database [36]: The modification of HCD files was set as: static modification of carbamidomethyl (Cys); dynamic modification of GlcNAc (Asn, 203.079 Da), deamination (Asn), oxidation (Met), and acetylation (Lys). The modification of ETD files was set as similar as HCD, except dynamic modification of fucosyl GlcNAc (349.137 Da) is instead of GlcNAc (Asn, 203.079 Da) in ETD searching files. Trypsin was selected as the enzyme with two missed cleavages allowed. The mass tolerance of the precursor ion was set to 20 ppm and the fragmentation ions were set to 25 mmu. A false discovery rate (FDR) of 1% was estimated, and applied to all data sets at the total peptide level. Further, pBuild was used to remove redundant protein entries and to group related proteins into a single group entry. The motif was processed with an online software at <http://motif-x.med.harvard.edu/motif-x.html#>. Then the figure of CF-glycosylation consensus sequence was drawn by employing IceLogo [37]. The DAVID functional analysis GO classification tool (<http://david.abcc.ncifcrf.gov/home.jsp>), and commercial IPA pathway software were applied to derive pathways and protein location that are annotated in the KEGG database and associated with identified plasma CF-glycoproteins [38,39].

3. Result and discussion

3.1. Overview of AHEF strategy

In order to analyze CF-glycosites of human plasma with high-throughput, an alternating HCD and ETD fragmentation (AHEF) approach was developed in this study. The key procedure included the following steps: CF-glycoproteins were extracted with plant lectin LCH from normal human plasma. Then, enriched plasma CF-glycoproteins were denatured and then trypsin-digested with the FASP method. Next, all of trypsin-digested glycopeptides

were enriched with self-made ZIC-HILIC tips [35]. Further, Endo F3 was used to split CF-glycopeptides, and two glycans (core-fucose and GlcNAc) remain on the simplified CF-glycopeptides. Simplified CF-glycopeptides were analyzed afterwards using alternating HCD and ETD techniques with LTQ-Orbitrap Elite ETD system. Finally, HCD and ETD spectra were respectively searched in the database after selection and handling with our in-house software. Potential CF-glycopeptides from HCD spectra were selected according to 1) characteristic ions and 2) neutral loss peaks.

In this study, the critical step for mapping CF-glycosites is alternating HCD and ETD techniques. In our previous study, we have developed a strategy in large scale identification of CF-glycoproteins with low- and high-normalized collision energy (NCE). With low-NCE HCD, the neutral loss peaks and their precursor ions can be observed as the two highest peaks in MS/MS spectra. With high-NCE HCD, every peptide can be thoroughly fragmented into high-quality b+ and y+ fragment ions, and a number of diagnostic ions decomposed from the oxonium ion of GlcNAc ($m/z = 204.08$) can be easily observed. These evidences proved all the spectra are CF-glycopeptide spectra. With this strategy, a total of 347 CF-glycosylation sites from 209 CF-glycoproteins in human plasma were identified [40]. Herein, we adopted alternating high-HCD and ETD instead of low- and high-NCE strategy, and data-dependent MS method was applied to do MS/MS for the top 5 most intense peaks. In preliminary experiments, we can find the precursor ions and their neutral loss peaks of simplified CF-glycopeptides in MS/MS spectra with high-HCD (NCE = 27). In addition, ETD is a good method to analyze the glycosites by maintaining the glycan linkage. All of CF-glycopeptide spectra of ETD were verified with both neutral loss peaks and characteristic ions of GlcNAc from HCD spectra, which makes the result very credible. The AHEF strategy is shown in Fig. 1.

3.2. Analysis of HCD and ETD spectra of simplified plasma CF-glycopeptides

Although the combined fragmentation of HCD and ETD has been used in glycopeptides sequencing, research is lacking in the large scale mapping of CF-glycoproteome. In order to compare the two technologies, a total of 6 normal human plasma samples were equivalently pooled for investigation. Our initial idea was to respectively analyze plasma CF-glycopeptides via HCD and ETD fragmentation, then merge the identified results. However, our previous results showed that many “CF-glycosites” that did not match to the canonical sequence N-X-[S/T/C] (X is any amino acid except Proline) were identified via ETD fragmentation. We are suspicious that it might have false positive results of CF-glycopeptide via ETD fragmentation. To test these suspicions, a parallel experiment was carried out with a non-glycoprotein biological system, *Escherichia coli*, and then the sample was analyzed with MS by using ETD fragmentation only. As a result, 17 false “CF-glycopeptides” were identified from *E. coli* proteins within FDR < 1 (data not shown). This experiment indicated that the algorithm of database searching is defective for large scale PTM peptide mapping. In order to obtain credible CF-glycosites from ETD spectra, a precise strategy was designed to select potential CF-glycopeptides spectra from ETD fragmentation according to associated HCD fragmentation. With a certain normalized collision energy (NCE = 27) we found that CF-glycopeptides from HCD spectra can be recognized by diagnostic ions of GlcNAc and neutral loss peaks of precursor ions (Fig. 2). The diagnostic ions and neutral

loss peaks are strong evidences of CF-glycopeptides. Herein, we only selected ETD spectra according to their associated HCD spectra. According to this finding, a software package was developed in-house for the selection of partially deglycosylated CF-glycopeptides. As we know, database searching software uses different scoring algorithm to identify peptide sequences via comparison of theoretical MS/MS spectra with experimental data [41]. Matching the ratio between practical fragmentations of MS/MS spectra and theoretical MS/MS spectra will determine the confidence level of peptides. Some PTM, such as glycosylation and phosphorylation, generate complex MS/MS spectra due to the weaker chemical bond at the modification site compared to the peptide backbone amide bonds. In order to identify more CF-glycopeptides, we modified a previously reported procedure to improve potential CF-glycopeptides spectra scoring by deleting peaks that cannot be matched to any protein database, including diagnostic ions of GlcNAc, precursor ions and their isotope peaks [31]. The detailed selection process is explained as follows: 1) the mgf files were generated from original raw files, 2) HCD spectra containing the diagnostic ions and neutral loss peaks were extracted into a new file, with the mass of fucose subtracted from the precursors. 3) The ETD spectra which were relevant to HCD spectra were extracted into another file. 4) HCD files were further optimized by removing the diagnostic ions of GlcNAc and neutral loss ions from MS/MS spectra, and ETD files were optimized by removing the precursor ions from MS/MS files. Finally, the two files were searched against human databases, respectively.

By using this strategy, 267 CF-glycoproteins and 407 CF-glycosites were disclosed from triplicated parallel experiments (Supporting information Table 1). Among the results, 303 CF-glycosites from 190 CF-glycoproteins were identified with HCD fragmentation, and 232 CF-glycosites from 177 CF-glycoproteins were identified with ETD fragmentation. There are 102 newly identified CF-glycosites added by ETD fragmentation (Fig. 3a). This result indicates that HCD and ETD technique complement each other in the study of simplified CF-glycopeptides.

3.3. Analysis of CF-glycoproteins and sites in human plasma

We compared AHEF with LHNCE for CF-glycosites identification, and resulted in the assignment of 161 unique CF-glycosites from 105 CF-glycoproteins. Therefore, AHEF is a better option for global research of CF-glycoproteome than HCD or ETD fragmentation, separately. Previously, 301 CF-glycosites among 209 CF-glycoproteins were identified in normal human plasma [23]. In this study a total of 407 CF-glycosites from 267 CF-glycoproteins were disclosed. We compared the identified CF-glycosites with N-glycosylation information recorded at Uniprot- Prot. In all CF-glycosites, we found that 65.8% are known human N-glycosites, and we discovered additional 139 CF-glycosites on a diverse range of proteins. In this study, a number of low-abundance CF-glycoproteins were identified from healthy human plasma, such as P-selectin (0.95 ng/mL), interleukin-6 receptor subunit beta (1.3 ng/mL), sialic acid-binding Ig-like lectin 14 (2.4 ng/mL), lysosome-associated membrane glycoprotein 2 (7.3 ng/mL), and Golgi membrane protein 1 (19 ng/mL) (protein concentrations refer to [42]). Among these low-abundance plasma proteins, P-selectin (0.95 ng/mL) has the lowest reported plasma concentration in the list [43]. Golgi membrane protein 1 (GP73), a resident Golgi glycoprotein, is a novel plasma

biomarker in hepatocellular carcinoma [44]. The plasma GP73 level was significantly increased in liver cancer patients, even in HCC patients who had plasma AFP levels less than 20 ng/mL (the concentration of AFP is less than 20 ng/mL in normal human plasma). It has also been reported that the fucosylation of GP73 was increased in patients with HCC [44].

To investigate the reproducibility of AHEF, plasma samples were independently prepared in triplicate and measured by LC-MS/MS runs with 120 min acquisition time. From the result of three independent experiments, the overlap is around 83%. The Venn diagram highlighting the overlap among the three independent experiments of plasma samples is shown in Fig. 3b. Canonical N-linked glycosylation motifs has been known to accord with the sequence N-X-[S/T] and rarely the N-X-C (X is not Proline). In this study, the CF-glycosites that match with N-X-T (55%) occur more frequently than N-X-S (43%). N-X-C motif accounts for 1% (Fig. 3c). A total of ten irregular motif CF-glycosites were identified, and six sites were verified by both of HCD and ETD spectra (Supporting information Fig. S1). These interesting finding may give us a new cognition to analyze characteristics of CF-glycoproteins and CF-glycosylation.

The ratio results showed no significant differences between CF-glycosylation and N-glycosylation. Of the total CF-glycoproteins from the top confidence set, we found that approximately 72.2% carried a single N-glycosite (Fig. 3d), 13.2% carried two CF-glycosites, 7.9% carried three CF-glycosites, and 3.8% carried four CF-glycosites. Notably, there was a group of five CF-glycoproteins that contained six CF-glycosites and three CF-glycoproteins that contained six or more CF-glycosites. From the result, attractin is the protein with the heaviest CF-glycosylation. Attractin has been known to play multiple roles in regulating physiological processes that are involved in monocyte-T cell interaction, agouti-related hair pigmentation, and control of energy homeostasis.

3.4. Categorization and annotation of CF-glycoproteins in human plasma

DAVID functional analysis tool was applied to derive pathways that are annotated in the KEGG database and associated with identifying plasma CF-glycosylated proteins. From the identified results, we found that a total of 22 human plasma CF-glycoproteins are involved in complement and coagulation cascades. Further, we also wished to obtain an overview of the subcellular compartments and the cellular functions with which plasma CF-glycoproteins are preferentially associated. All 267 CF-glycoproteins identified in this work were categorized on the basis of their GO components and functions. For the GO molecular function categorization, 233 identifiers were mapped in the target data set, which included binding (40.7%), catalytic activity (15.4%), molecular function regulator (9.5%), molecular transducer activity (8.4%), enzyme regulator (7.4%), transporter activity (4.4%), and others (3.2%). For the cellular component categorization, 260 CF-glycoproteins were verified. Most plasma CF-glycosylated proteins were located in the extracellular region (32.7%), plasma membrane proteins (26.6%), or macromolecular complex (12.8%). Some CF-glycoproteins were located in membrane-enclosed lumen (9.7%), such as in the endoplasmic reticulum (ER) or Golgi apparatus, which are glycosylated protein synthesis and processing establishments. In addition, we found CF-glycoproteins associated with the cell junction

(2.0%) and collagen trimer (0.5%) (Fig. 4a). This result shows that 60% of the CF-glycoproteins are located on the cell membrane or extracellular region, which is similar with the mouse tissue N-glycoproteome [45]. Further, the IPA functional analysis tool was applied to analyze the network of identified CF-glycoproteins. We found that CF-glycoproteins concentrate in the network of neurological disease, inflammatory disease, and cancers. In this proteins' interaction network, we can match 22 proteins out of a total of 35 proteins (Fig. 4b).

4. Conclusion

In this study, a large scale approach of identifying CF-glycoproteins by altering HCD and ETD fragmentation was established. This strategy reveals the excellent complementarity nature of HCD and ETD in analysis of CF-proteins, and a total of 407 CF-glycosylation sites among 267 CF-glycoproteins were identified from normal human plasma. All of the CF-glycopeptide spectra were verified with both neutral loss peaks and characteristic ions of GlcNAc from HCD spectra, which make the result very credible. Compared with CF-glycoproteome analysis with HCD only, 77 CF-glycoproteins and 102 CF-glycosites were added. Meanwhile, a total of ten CF-glycosites were identified from human plasma without the N-X-S/T/C consensus motif, occupying almost 2.4% of total identified CF-glycosites. This approach supplies a more effective method for CF-glycosylation site identification, and provides a valuable biomarker reference for clinical research. Furthermore, this research strategy also provides an important approach for biomarker screening in cancers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We sincerely thank Georgia Research Alliance (GRA) and Georgia State University for purchasing the analytical instrument used in this research. Our work was partially supported by a funding from the Key Grant Project of Chinese Ministry of Education No. 313033.

References

1. Kufe DW. Mucins in cancer: function, prognosis and therapy. *Nat. Rev. Cancer.* 2009; 9:874–885. [PubMed: 19935676]
2. Perego P, Gatti L, Beretta GL. The ABC of glycosylation. *Nat. Rev. Cancer.* 2010; 10:523.
3. Slawson C, Hart GW. O-GlcNAc signalling: implications for cancer cell biology. *Nat. Rev. Cancer.* 2011; 11:678–684. [PubMed: 21850036]
4. Shimizu K, Katoh H, Yamashita F, Tanaka M, Tanikawa K, Taketa K, et al. Comparison of carbohydrate structures of serum alpha-fetoprotein by sequential glycosidase digestion and lectin affinity electrophoresis. *Clin. Chim. Acta.* 1996; 254:23–40. [PubMed: 8894307]
5. Aoyagi Y, Isemura M, Yosizawa Z, Suzuki Y, Sekine C, Ono T, et al. Fucosylation of serum alpha-fetoprotein in patients with primary hepatocellular carcinoma. *Biochim. Biophys. Acta.* 1985; 830:217–223. [PubMed: 2411292]
6. Marrero JA, Romano PR, Nikolaeva O, Steel L, Mehta A, Fimmel CJ, et al. GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. *J. Hepatol.* 2005; 43:1007–1012. [PubMed: 16137783]

7. Comunale MA, Rodemich-Betesh L, Hafner J, Wang M, Norton P, Di Bisceglie AM, et al. Linkage specific fucosylation of alpha-1-antitrypsin in liver cirrhosis and cancer patients: implications for a biomarker of hepatocellular carcinoma. *PLoS One*. 2010; 5:e12419. [PubMed: 20811639]
8. Miyoshi E, Noda K, Ko JH, Ekuni A, Kitada T, Uozumi N, et al. Overexpression of alpha1-6 fucosyltransferase in hepatoma cells suppresses intrahepatic metastasis after splenic injection in athymic mice. *Cancer Res*. 1999; 59:2237–2243. [PubMed: 10232614]
9. Garibay-Cerdenares OL, Hernandez-Ramirez VI, Osorio-Trujillo JC, Hernandez-Ortiz M, Gallardo-Rincon D, Cantu de Leon D, et al. Proteomic identification of fucosylated haptoglobin alpha isoforms in ascitic fluids and its localization in ovarian carcinoma tissues from Mexican patients. *J. Ovarian Res*. 2014; 7:27. [PubMed: 24576319]
10. Packer NH, von der Lieth CW, Aoki-Kinoshita KF, Lebrilla CB, Paulson JC, Raman R, et al. Frontiers in glycomics: bioinformatics and biomarkers in disease. An NIH white paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11–13, 2006). *Proteomics*. 2008; 8:8–20. [PubMed: 18095367]
11. Taniguchi N. Human disease glycomics/proteome initiative (HGPI). *Mol. Cell. Proteomics*. 2008; 7:626–627. [PubMed: 18322298]
12. Moriwaki K, Miyoshi E. Fucosylation and gastrointestinal cancer. *World J. Hepatol*. 2010; 2:151–161. [PubMed: 21160988]
13. Lee A, Nakano M, Hincapie M, Kolarich D, Baker MS, Hancock WS, et al. The lectin riddle: glycoproteins fractionated from complex mixtures have similar glycomic profiles. *OMICS*. 2010; 14:487–499. [PubMed: 20726804]
14. Lee A, Kolarich D, Haynes PA, Jensen PH, Baker MS, Packer NH. Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *J. Proteome Res*. 2009; 8:770–781. [PubMed: 19125615]
15. Jia W, Lu Z, Fu Y, Wang HP, Wang LH, Chi H, et al. A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol. Cell. Proteomics*. 2009; 8:913–923. [PubMed: 19139490]
16. Wells JM, McLuckey SA. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol*. 2005; 402:148–185. [PubMed: 16401509]
17. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem*. 2000; 72:563–573. [PubMed: 10695143]
18. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A*. 2004; 101:9528–9533. [PubMed: 15210983]
19. Frese CK, Altelea AF, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, et al. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res*. 2011; 10:2377–2388. [PubMed: 21413819]
20. Nagaraj N, D'Souza RC, Cox J, Olsen JV, Mann M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J. Proteome Res*. 2010; 9:6786–6794. [PubMed: 20873877]
21. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods*. 2007; 4:709–712. [PubMed: 17721543]
22. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol. Cell. Proteomics*. 2011; 10 (M111.009910).
23. Ma C, Zhang Q, Qu J, Zhao X, Li X, Liu Y, et al. A precise approach in large scale core-fucosylated glycoprotein identification with low- and high-normalized collision energy. *J. Proteome*. 2015; 114:61–70.
24. Cao Q, Zhao X, Zhao Q, Lv X, Ma C, Li X, 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*. 2014; 86:6804–6811. [PubMed: 24914453]

25. Tan Z, Yin H, Nie S, Lin Z, Zhu J, Ruffin MT, et al. Large-scale identification of core-fucosylated glycopeptide sites in pancreatic cancer serum using mass spectrometry. *J. Proteome Res.* 2015; 14:1968–1978. [PubMed: 25732060]
26. Wang D, Hincapie M, Rejtar T, Karger BL. Ultrasensitive characterization of site-specific glycosylation of affinity-purified haptoglobin from lung cancer patient plasma using 10 μ m i.d. porous layer open tubular liquid chromatography-linear ion trap collision-induced dissociation/electron transfer dissociation mass spectrometry. *Anal. Chem.* 2011; 83:2029–2037. [PubMed: 21338062]
27. Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AV, Crossett B, et al. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol. Cell. Proteomics.* 2011; 10 (M000031-mcp201).
28. Vakhrushev SY, Steentoft C, Vester-Christensen MB, Bennett EP, Clausen H, Levery SB. Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with Simple Cells. *Mol. Cell. Proteomics.* 2013; 12:932–944. [PubMed: 23399548]
29. Go EP, Liao HX, Alam SM, Hua D, Haynes BF, Desaire H. Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *J. Proteome Res.* 2013; 12:1223–1234. [PubMed: 23339644]
30. 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; 11:4517–4525. [PubMed: 22800195]
31. Ma C, Qu J, Meisner J, Zhao X, Li X, Wu Z, et al. Convenient and precise strategy for mapping N-glycosylation sites using microwave-assisted acid hydrolysis and characteristic ions recognition. *Anal. Chem.* 2015; 87:7833–7839. [PubMed: 26161579]
32. Tateno H, Nakamura-Tsuruta S, Hirabayashi J. Comparative analysis of core-fucose-binding lectins from *Lens culinaris* and *Pisum sativum* using frontal affinity chromatography. *Glycobiology.* 2009; 19:527–536. [PubMed: 19218400]
33. Zhao Y, Jia W, Wang J, Ying W, Zhang Y, Qian X. Fragmentation and site-specific quantification of core fucosylated glycoprotein by multiple reaction monitoring-mass spectrometry. *Anal. Chem.* 2011; 83:8802–8809. [PubMed: 21970473]
34. Wisniewski JR, Zougman A, Mann M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* 2009; 8:5674–5678. [PubMed: 19848406]
35. Ma C, Zhao X, Han H, Tong W, Zhang Q, Qin P, et al. N-linked glycoproteome profiling of human serum using tandem enrichment and multiple fraction concatenation. *Electrophoresis.* 2013; 34:2440–2450. [PubMed: 23712678]
36. Wang LH, Li DQ, Fu Y, Wang HP, Zhang JF, Yuan ZF, et al. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 2007; 21:2985–2991. [PubMed: 17702057]
37. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods.* 2009; 6:786–787. [PubMed: 19876014]
38. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009; 4:44–57. [PubMed: 19131956]
39. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37:1–13. [PubMed: 19033363]
40. Ma C, Zhang Q, Qu J, Zhao X, Li X, Liu Y, et al. A precise approach in large scale core-fucosylated glycoprotein identification with low- and high-normalized collision energy. *J. Proteome.* 2014; 114c:61–70.
41. Fu Y, Yang Q, Sun R, Li D, Zeng R, Ling CX, et al. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics (Oxford, England).* 2004; 20:1948–1954.

42. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics*. 2011; 10 (M110.006353).
43. Scalia R, Armstead VE, Minchenko AG, Lefer AM. Essential role of P-selectin in the initiation of the inflammatory response induced by hemorrhage and reinfusion. *J. Exp. Med.* 1999; 189:931–938. [PubMed: 10075976]
44. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, et al. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:779–784. [PubMed: 15642945]
45. Zielinska DF, Gnad F, Wisniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*. 2010; 141:897–907. [PubMed: 20510933]

Fig. 1.

AHEF strategy for mapping CF-glycosites. In this study, a data dependent scan mode with top 5 precursor ions were selected for alternating HCD and ETD fragmentation. All of the HCD and ETD spectra were selected by diagnostic ions of GlcNAc and neutral loss peaks.

Fig. 2.

HCD and ETD spectra of ISVQVHnATCTVR. (a) From the HCD spectrum, the fragment ions are lack because of the strong neutral loss peak of precursor ion and diagnostic ions of GlcNAc. (b) ETD enabled the elucidation of CF-glycosites by maintaining the glycan-peptide linkage.

Fig. 3. Large scale mapping of human plasma CF-glycoproteome via HCD and ETD fragmentation. (a) Overlap of CF-glycosites and CF-glycoproteins via HCD and ETD fragmentation. (b) Repeatability analysis of identified CF-glycosites via AHEF. (c) N-glycosylation consensus sequence was used to create relative frequency plots with Motif-X and generated with the IceLogo tool. The significantly enriched motifs are the canonical serine or threonine containing motifs <http://iomics.ugent.be/icelogoserver/>. (d) Distribution of singly and multiply glycosylated proteins.

Fig. 4. (a) Gene Ontology (GO) analysis of plasma CF-glycoproteins. (b) Interaction pathway analysis of plasma CF-glycoproteins by using IPA software and DAVID pathway analysis.