Published in final edited form as:
J Proteome Res. 2012 March 2; 11(3): 1728–1740. doi:10.1021/pr201183w.

Semi-Automated Identification of N-Glycopeptides by Hydrophilic Interaction Chromatography, nano-Reverse-Phase LC-MS/MS, and Glycan Database Search

Petr Pompach,
Department of Oncology, Georgetown University, Washington, DC, USA, pp282@georgetown.edu

Kevin B. Chandler,
Department of Biochemistry & Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC, USA and Department of Oncology, Georgetown University, Washington, DC, USA, kbc@georgetown.edu

Renny Lan,
Department of Oncology, Georgetown University, Washington, DC, USA, rls25@georgetown.edu

Nathan Edwards, and
Department of Biochemistry & Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC, USA, nje5@georgetown.edu

Radoslav Goldman*
Department of Oncology, Georgetown University, Washington, DC, USA, rg26@georgetown.edu

Abstract

Glycoproteins fulfill many indispensable biological functions and changes in protein glycosylation have been observed in various diseases. Improved analytical methods are needed to allow a complete characterization of this complex and common posttranslational modification. In this study, we present a workflow for the analysis of the microheterogeneity of N-glycoproteins which couples hydrophilic interaction and nano-reverse-phase C18 chromatography to tandem QTOF mass spectrometric analysis. A glycan database search program, GlycoPeptideSearch, was developed to match N-glycopeptide MS/MS spectra with the glycopeptides comprised of a glycan drawn from the GlycomeDB glycan structure database and a peptide from a user-specified set of potentially glycosylated peptides. Application of the workflow to human haptoglobin and hemopexin, two microheterogeneous N-glycoproteins, identified a total of 57 distinct site-specific glycoforms in the case of haptoglobin and 14 site-specific glycoforms of hemopexin. Using glycan oxonium ions, peptide-characteristic glycopeptide fragment ions, and by collapsing topologically redundant glycans, the search software was able to make unique N-glycopeptide assignments for 51% of assigned spectra, with the remaining assignments primarily representing isobaric topological rearrangements. The optimized workflow, coupled with GlycoPeptideSearch, is expected to make high-throughput semi-automated glycopeptide identification feasible for a wide range of users.

CORRESPONDING AUTHOR FOOTNOTE: Radoslav Goldman, Department of Oncology, Georgetown University, LCCC Room S183, 3970 Reservoir Rd NW, Washington, DC 20057, Tel: 202-687-9868, Fax: 202-687-1988, rg26@georgetown.edu.
Petr Pompach, Lombardi Comprehensive Cancer Center, Georgetown University, 3970 Reservoir Rd NW, Washington, DC 20057-1465, USA.

Keywords

Mass spectrometry; HILIC chromatography; Q-TOF; Haptoglobin; Hemopexin; Glycopeptide search software; Detached glycans; Oxonium ions

INTRODUCTION

Glycosylation is one of the most ubiquitous and variable posttranslational modifications of proteins[1]. N-glycosylation occurs primarily on Asn residues of the NXS/T sequence motif where X represents any non-proline amino acid[2–4]. N-glycans are branched oligosaccharides synthesized by glycosyltransferases and glycosidases in a complex enzymatic pathway localized primarily in the endoplasmic reticulum and Golgi compartments[5]. It is estimated that two thirds of human protein sequences contain at least one N-glycosylation motif and that approximately half of all proteins are N-glycosylated[1].

Glycoproteins influence a broad spectrum of biological processes including cell-cell interaction, host-pathogen interaction, or protection of proteins against proteolytic degradation[6]. The severity of congenital disorders of glycosylation (CDGs), which result from mutations of enzymes involved in N-glycosylation, emphasizes the importance of N-glycosylation in pathophysiological processes[7–9]. Strong evidence links changes in protein glycosylation to diseases of etiology substantially more complex than congenital disorders, including infectious and autoimmune diseases, metabolic disorders, and cancer[10–13]. Qualitative and quantitative differences in cancer related N-glycans and glycoproteins have been described in a number of cancer-types. In addition, it has been observed by a number of authors that the majority of clinically used cancer markers are glycoproteins[12,14–18]. These observations have spurred interest in the characterization of N-glycans and glycopeptides in the context of disease progression; however, the complete characterization of glycoproteins remains a considerable analytical challenge[19]. Major problems represent the as yet unmapped number of N-glycan variants, isobaric species, unfavorable ionization and fragmentation of glycopeptides, limited tools for an efficient chromatographic resolution of glycopeptides, and lack of software tools for glycopeptide assignment.

The structure of glycopeptides has been studied by various analytical methods such as nuclear magnetic resonance, enzymatic digestion, or mass spectrometry. NMR has primarily been used as a tool for characterization of saccharide moiety of a glycopeptide or to observe conformation of glycan-peptide linkage[20,21]. Enzymatic digestion by exoglycosidases was used to determine the structure of released glycan from well-known glycoproteins such as fetuin, alpha 1-acid glycoprotein, or tissue plasminogen activator[22]. The mass spectrometric approaches for glycopeptide characterization have the advantage of sensitivity compatible with the analysis of biologically relevant glycoproteins; the MS methods have been well summarized in recent reviews[23–25].

A common problem in glycopeptide characterization in biological samples is the presence of a given peptide in multiple distinct glycoforms. The microheterogeneous species become hard to detect in a total protein digest even with state of the art mass spectrometry and multiple enrichment methods. We have only begun to uncover the complexity of glycoproteins' glycoforms in cancer and other diseases[26,27]. Enrichment of glycopeptides for mass spectrometric analysis is essential as the non-glycosylated peptides are more abundant than the glycopeptides. Moreover multiple glycoforms with different charge stages result in lower signal intensities. Affinity chromatography using lectins is a popular method to enrich glycoproteins and glycopeptides in complex samples[28,29]. Concanavalin A and wheat germ agglutinin are most frequently used for a broad enrichment of many N-

glycoforms; more specific lectins can be used to select subclasses of the structures[17,30]. However, the lectins' weak affinity for their glycan ligands and their cross-reactivity represent significant drawbacks; in addition, the lectin affinity separations cannot overcome the enormous dynamic range of secreted glycoproteins and are difficult to interface directly with mass spectrometry due to solvent incompatibility. Enrichment of glycoproteins on hydrazide beads following periodate-oxidation has been widely used in glycoproteomic research in the last decade[31,32]. This collapses the glycopeptide microheterogeneity onto a single modified peptide or a smaller set of glycoforms, boosting the abundance of measured glycosylated peptides, at the expense of observing the original glycopeptides.

Methods separating the glycan and peptide moieties, as with the analysis of detached N-glycans, simplify the analytical challenges somewhat[33–36]. N-glycans are detached using PNGaseF and the peptides and glycans are analyzed using separate analytical workflows. PNGaseF detachment of N-glycans boosts the abundance of glycopeptides' peptide substrate by aggregating the glycopeptides' glycoforms on a single peptide species. The enzymatic deglycosylation leads to a substitution of Asn for Asp (mass delta +0.98 Da), providing a mechanism for identifying a glycoproteins' N-linked glycosylation sites[37,38]. Analysis of detached N-glycans can provide important insights into the heterogeneity of protein glycosylation at the expense of the ability to characterize the glycans' protein or attachment site[36,39]. The analysis of detached glycans can be used to assemble a targeted glycan inventory but site specific characterization of protein glycoforms, the goal of our study, requires additional chromatographic and mass spectrometric steps[14,40,41].

Mass spectrometry has enabled important advances in the characterization of N-glycans and glycopeptides[42]. Sensitivity of MS detection and applicability of the MS methods to complex biological samples has been particularly helpful. State of the art mass spectrometry, in combination with multiple separation and enrichment methods, has begun to fill the gaps in our understanding of the glycoforms of proteins associated with disease[43,44]. These methods require a combination of specialized instrumentation, experimental methods, and skills that limits their successful application to a wide range of biological problems.

The aim of our study was to develop a workflow for qualitative glycopeptide analysis compatible with CID instruments of readily-available accuracy and resolution. To achieve this aim, we combine HILIC and nano-reverse-phase C18 chromatography, for an unbiased enrichment and separation of glycopeptides, with MS/MS characterization of the glycoforms on a quadrupole time of flight mass spectrometer (Q-TOF)[45]. The efficiency of the workflow is demonstrated here with haptoglobin and hemopexin, both of which have been previously demonstrated to have considerable glycoform microheterogeneity. The spectral interpretation bottleneck is addressed by a glycan database search tool that compares peptide-glycan pairs, representing putative glycopeptides, to tandem mass-spectra. The glycopeptide fragments observed in the MS/MS spectrum make it possible to avoid a brute-force enumeration of all peptide-glycan pairs whose mass is close to that of the precursor-ion, significantly boosting the quality of the glycopeptide assignments.

EXPERIMENTAL SECTION

Samples and Reagents

Haptoglobin (H3536), dithiothreitol (43815), iodoacetamide (A3221), trifluoroethanol (T63002), 2,5-dihydroxybenzoic acid (39319), sodium hydroxide (01209BH), trifluoroacetic acid (T6508), acetonitrile (34998), chloroform (C-2432), iodomethane (06416ME), sodium chloride (D-5545) and water (270733) were purchased from Sigma-Aldrich (St. Louis, MO). DMSO (327182500) was from Acros Organics (Pittsburgh, PA), hemopexin (PRO-560) was from ProSpec (East Brunswick, NJ). RapiGest SF Surfactant, HILIC XBridge column 5μm,

2.1 × 100mm (186004445), trap column 5μm Symmetry C18, 180μm × 20mm (186003514), analytical column 1.7μm BEH300 C18, 75μm × 150mm (186003813) and Sep-Pak Vac tC18 50mg cartridge (WAT054960) were obtained from Waters (Milford, MA). Charcoal solid-phase extraction column (744300) was from Harvard Apparatus (Hamden, CT). Trypsin-Gold (V5280) was obtained from Promega (Madison, WI) and endoproteinase Glu-C (10791156001) was from Roche (Indianapolis, IN). Peptide-N-Glycosidase F (P0705S) was from New England BioLabs (Ipswich, MA). Holder, manual macro trap kit (TH1/25111/02) and desalting peptide traps (TH1/25109/52) were from Michrom Bioresources (Auburn, CA). Water Optima LC/MS (W6-4), acetonitrile Optima LC/MS (A995-4), and formic acid Optima LC/MS (A117) were purchased from Fisher Scientific (Fair Lawn, NJ).

Analysis of detached N-glycans

Analysis of permethylated detached N-glycans was carried out as described previously with minor modifications[13,46]. Briefly, samples of hemopexin or haptoglobin were reduced with dithiothreitol, alkylated with iodoacetamide, and N-glycans were detached with PNGaseF overnight at 37°C. Proteins were removed on a C18 cartridge, released N-glycans were further cleaned on an activated charcoal solid-phase extraction column, and the combined eluents were permethylated on sodium hydroxide beads in a solution consisting of 79.3% DMSO, 19.5% iodomethane, and 1.2% water. The permethylated samples were extracted in chloroform and the chloroform layer was further washed with three 1ml volumes of water. The dried permethylated samples were resuspended in a 50:50 methanol:water solution. Each sample (0.5 μL) was spotted directly on a MALDI plate and mixed with an equal volume of DHB matrix (10 mg/ml DHB in 50:50 methanol:sodium acetate, 2mM). The sample spots were dried under vacuum to achieve uniform crystallization. Spectra were acquired on an Applied Biosystems 4800 Mass Analyzer (AB Sciex, Framingham, MA) equipped with a Nd:YAG 355-nm laser. MALDI spectra were recorded in positive ion mode. Raw spectra were exported as text files for analysis (see below) and processed using an in house software described previously[39,47].

Protein digest

Haptoglobin and hemopexin were resuspended at 1 μg/μL of 50mM NH_4HCO_3 at pH7.8 with 0.05% RapiGest. The proteins were reduced for 30min at 60°C with 5mM DTT and alkylated for 30min in the dark at room temperature with 15mM iodoacetamide. Additional DTT was added to a final concentration of 50mM to stop the alkylation reaction. The proteins were digested with Trypsin Gold, 0.4 μg, at 37°C in Barocycler NEP2320 (Pressure BioSciences, South Easton, MA) for 60 cycles, each consisting of 50s period at 20kPsi and 10s period at low pressure, or with endoproteinase GluC, 1.2 μg, at 25°C, overnight. The digests were desalted on a Macro Trap peptide cartridge, and washed 5 times with 250uL of 0.1% aqueous TFA. The peptides were eluted with 150uL of 60% AcN with 0.1%TFA and the eluate was dried using a SpeedVac concentrator.

Separation of glycopeptides by HILIC chromatography

Twenty micrograms of tryptic-GluC/tryptic digests of haptoglobin and hemopexin were resuspended in 20 μL of 50%AcN in water and injected on a HILIC XBridge column (5μm, 2.1×100mm) using a loop of 900 μL volume. A smaller amount, five micrograms, of protein digest was initially tried, but resulted in insufficient UV signal for manual collection of low abundance fractions. The column was equilibrated in 90% of solvent B (0.01%TFA in acetonitrile). Separation was achieved using a 35 min linear gradient of 90% to 40% solvent B at 40°C (Solvent A: 0.01%TFA in water; Solvent B: 0.01%TFA in acetonitrile). Sixteen fractions were collected for haptoglobin and 24 fractions for hemopexin.

To confirm that glycopeptides are not eluted with the non-glycosylated peptides during HILIC separation, we conducted a separate experiment, in which we connected the HILIC XBridge column (5μm, 2.1×100mm) on line to a QTOF mass spectrometer using a Micro flow Ion spray, splitting the master flow in a 1:8 ratio (250μL/min were split, with 30μL/min used for MS analysis and the remaining volume going to waste). Chromatographic conditions were as above, with a 20μL sample loop. The extracted ions at m/z 204 and 366 corresponding to oxonium ions of N-Acetylhexosamine and Hexose-N-Acetylhexosamine were selected to monitor the retention of glycopeptides on HILIC column.

Nano LC – MS/MS

The collected HILIC fractions were analyzed by nanoLC-ES-MS/MS using a C18 column (1.7μm particles, 75 μm ID × 150mm) coupled to a QStar Elite mass spectrometer (Applied Biosystems, Foster City, CA). Each fraction was resuspended in 20μL of 2% acetonitrile, 0.1% formic acid. Two microliters of each fraction were injected and glycopeptides were separated at a flow rate of 0.4 μL/min, using the following gradient: 0min 99%A, 2min 90%A, 14min 55%A, 15min 1%A, 17min 1%A, 18min 99%A, 30min 99%A (Solvent A: 0.1% formic acid in 2% acetonitrile; Solvent B: 0.1% formic acid in 98% acetonitrile). The mass spectrometer was operated in data dependent mode; after the survey full-scan (m/z 200 to m/z 1800), the four most intense precursor ions were selected for collision induced dissociation. Collision energy and MS/MS accumulation time were set automatically and the MS/MS spectra were recorded from m/z 150 to m/z 2000. Dynamic exclusion was set at 20 seconds and 30 counts for two repeated precursors with a total cycle time of 3.8 seconds.

To demonstrate the effectiveness of two dimensional chromatography (HILIC + nano-reverse-phase C18) as compared to one dimensional separation (nano-reverse-phase C18 only) for enriching haptoglobin glycopeptides, we conducted a separate experiment in which 4pM of the tryptic-GluC digest of haptoglobin was injected directly into a nano-reverse-phase column: first under the same nano-reverse-phase chromatographic conditions described above (30 min gradient); and then with a longer gradient: 0min 99%A, 100min 65%A, 102min 10%A, 107min 10%A, 109min 99%A, 130min 99%A (Solvent A: 0.1% formic acid in 2% acetonitrile; Solvent B: 0.1% formic acid in 98% acetonitrile). Mass spectrometer settings were as described above.

Data interpretation of detached N-glycans

The MALDI-TOF spectra were calibrated on masses of a set of previously identified N-glycans and exported as text files for further processing[13,46]. We used in-house software implementing the following modifications to our previously published spectral processing methods[39,47]. We eliminated a binning step which was found to distort intensity-ratios of the N-glycan isotope clusters. Instead, the spectra were smoothed by the Savitsky-Golay algorithm, de-noised by Daubechies D20 wavelet transform, and the baseline was corrected by removing low frequency nodes using FFT convolution/deconvolution[48–50]. Similar to recent publications, the presence of previously identified N-glycans was determined by matching the theoretical distribution of their isotopic clusters to the observed spectra and resolving peak interval overlaps using an iterative prediction/correction procedure[40,47,51,52]. The detected N-glycans were subtracted from the spectrum and the remaining isotopic clusters, with intensity above a predefined cutoff and present in more than 20% of the analyzed spectra, were interpreted as unknown N-glycans. The identified peaks were normalized by scaling the total peak intensities to 100.

Data Interpretation of Glycopeptide Fragmentation Spectra

Version 1.0 of the GlycomeDB and associated structures were downloaded from www.glycome-db.org[53]. The glycan structures annotated as human were extracted in

GlycoCT condensed format[54]. The resulting set of 3,214 human glycan structures was later determined to contain 954 glycan structures with the N-linked glycan core motif.

Haptoglobin (HPT_HUMAN) and hemopexin (HEMO_HUMAN) protein sequences from UniProt underwent simultaneous *in silico* trypsin and GluC digestion (haptoglobin) or trypsin digestion (hemopexin). Peptides with the NXS/T N-linked glycosylation motif, residual mass less than 2000, and at most one missed cleavage were selected for further analysis. A total of 7 haptoglobin and 3 hemopexin peptides were considered (Table 1).

Prior to analysis, the MS/MS spectra were converted to mzXML format using msconvert from the ProteoWizard project[55]. The haptoglobin glycopeptide analysis resulted in 11 spectral data-files containing 3288 tandem mass-spectra, while the hemopexin analysis resulted in 19 spectral data-files containing 684 tandem mass-spectra.

After file conversion, spectra were submitted to GlycoPeptideSearch (GPS). GPS software searched each spectral data-file for characteristic glycopeptide oxonium ions at m/z 204 and 366 (used to automatically select likely glycopeptide fragmentation spectra, with at least one of these peaks required for further consideration). Spectra were then checked for glycopeptide fragment ions consistent with the intact mass of the peptides in Table 1. Peaks with m/z value corresponding to each peptide and the peptide with up to three monosaccharide residues (GlcNAc, GlcNAc-GlcNAc and GlcNAc-GlcNAc-Man) from the N-glycan core were selected. These intact peptide fragment ions where matched in charge states 1, 2, and 3. For spectra with at least two of the intact peptide fragment ions, the remaining precursor mass was determined to represent an N-linked glycan. Non-monoisotopic precursor selection was accounted for by correcting for up to 2 ^{13}C isotopes in the reported precursor mass. Putative glycan masses were matched to N-glycans from the human subset of GlycomeDB. Oxonium and intact peptide fragment m/z-match tolerance of 0.2 Da and glycan mass-match tolerance of 0.2 Da was used. Glycans with equivalent topology were collapsed to a single match. The individual runs were analyzed together using the GlycoPeptideSearch software. The results are presented in a single Excel spreadsheet. We adopt the N-glycan names from the NIBRT GlycoBase databases to be consistent with published literature[56].

RESULTS AND DISCUSION

Both haptoglobin and hemopexin are heavily glycosylated proteins secreted by the liver. The main function of haptoglobin, which contains four N-glycosylation sites, is to bind free hemoglobin released by intravascular hemolysis; haptoglobin prevents the loss of iron and subsequent kidney damage[57]. Hemopexin is a heme-binding glycoprotein containing five N-glycosylation sites and one O-glycosylation site. Its major biological role is the defense against hemoglobin-mediated oxidative damage during intravascular hemolysis[58]. It is known that the glycosylation of haptoglobin and hemopexin is affected by disease processes of liver cirrhosis or cancer[15,17,59]. The study of glycosylation changes in the disease context could elucidate pathophysiology of the processes or provide putative markers for disease.

We have developed a workflow to address the current bottleneck in the analysis of glycopeptides to move toward a more comprehensive understanding of glycoprotein microheterogeneity in biological contexts. Figure 1 presents a schema for the experimental and data interpretation workflows applied in this study, including (1) proteolytic digestion of glycoproteins by trypsin and/or trypsin-GluC proteases, (2) separation of glycopeptides via HILIC chromatography, (3) online nano-reverse-phase LC CID MS/MS, and (4) interpretation of MS/MS spectra using GlycoPeptideSearch software to aid in the matching of glycopeptide spectra with database-derived glycan structures.

HILIC Separation of Glycopeptides

We have used the XBridge HILIC column with underivatized silica stationary phase and a simple gradient of water and acetonitrile with 0.01%TFA to enrich the glycopeptides for the MS analysis. One of the major benefits of using HILIC chromatography is the efficient enrichment of glycopeptides, which cannot be achieved with RP chromatography, as peptides and glycopeptides co-elute under reverse phase conditions. Glycopeptide enrichment via HILIC chromatography, followed by analysis of glycopeptide-enriched fractions using RP-LC MS/MS increases the likelihood that glycopeptides are selected for fragmentation by removing non-glycopeptides that compete for ionization. The resulting spectra contain a useful number of fragment ion peaks with good signal-to-noise characteristics, facilitating data interpretation[66].

A number of different HILIC surfaces and solvents have been used for the analysis of N-glycans or glycopeptides[66–68]. The HILIC columns can be divided into two common groups: 1. the underivatized silica stationary phases containing siloxanes or silanols; and 2. the derivatized silica, such as PolycatA, PolyWax or zwitterionic HILIC[25]. Zwitterionic (ZIC)-HILIC stationary phase is one of the most popular and is also frequently used for solid phase extraction (SPE) of glycopeptides[67]. In our experience, the underivatized silica is an excellent sorbent. The retention time of glycopeptides on the HILIC column is primarily dependent on the glycan moiety. Non-glycosylated peptides are significantly less-well retained and elute, under our conditions, close to the injection head volume while the glycopeptides are well retained and separated from the non-glycosylated peptides and from each other[45].

To evaluate the extent to which non-glycosylated peptides co-elute with the glycopeptides, each HILIC fraction was subject to peptide-identification analysis with the ProteinPilot software, in order to identify spectra matching the (non-glycosylated) human haptoglobin protein sequence. The results reveal the presence of some non-glycosylated peptides amongst the glycopeptides, but they represent a very small number of low intensity precursors.

To confirm the separation and enrichment of glycopeptides from non-glycosylated peptides, we coupled HILIC chromatography directly to a micro spray source on a QTOF mass-spectrometer. Despite splitting the primary flow and the reduced sensitivity of the micro flow sprayer, the extracted ion chromatograms correlate well with the UV trace. After alignment, the UV trace and the extracted ion chromatograms for the two most intense glycopeptide oxonium ions (m/z 204.0 corresponding to N-acetylhexosamine and m/z 366.1 corresponding to N-acetylhexosamine-hexose) in the fragmentation spectra demonstrate the enrichment and separation of glycopeptides from non-glycosylated peptides (Supplemental figure 1).

To assess the benefit of using HILIC chromatography for enrichment of glycopeptides for glycopeptide discovery we made two attempts to analyze the same tryptic-GluC digest of haptoglobin without first fractionating the peptides on the HILIC column.

In the first experiment, haptoglobin peptides were injected directly onto the nano-reverse-phase column and a short chromatographic separation (30 min), consistent with the nano-reverse-phase gradient for HILIC fractions, was performed. To avoid overloading the column, four picomoles of the protein digest were analyzed. This analysis yielded 515 MS/MS spectra; just 92 (18%) of these spectra contained oxonium ion peaks (*m/z* 204 or 366) with relative ion intensities greater than 10%. In the second experiment, a long nano-reverse-phase separation (130 min) was performed using the same amount of starting

material. This workflow yielded 2465 MS/MS spectra, 572 (23%) of which contained oxonium ion peaks (*m/z* 204 or 366) with relative intensity greater than 10%.

In comparison with these more direct analyses, the spectra from the two-dimensional separation demonstrate the value of HILIC separation for enriching for glycopeptides from the haptoglobin digest. Manual fraction collection from the HILIC column followed by injection of the fractions onto the nano-reverse-phase column resulted in the acquisition of 3288 MS/MS spectra, with 2930 (89%) spectra containing oxonium ion peaks (*m/z* 204 or 366) with relative intensity greater than 10%. The HILIC separation of peptides allowed tenfold greater relative load of glycopeptides compared to one dimensional chromatography. The significant boost in the number of putative glycopeptides fragmentation spectra can be directly attributed to the HILIC fractionation.

The HILIC column separates glycopeptides based on glycan content, glycan size, and the presence of negatively charged sialic acids[69]. We observe this in our analysis, with the less hydrophilic glycopeptides observed in early fractions, and more hydrophilic glycopeptides observed in later fractions. To observe this, we count the spectra matching haptoglobin glycopeptide MVSHHNLTTGATLINE with non-fucosylated tri-antennary structures A3G3S1, A3G3S2 and A3G3S3. For this peptide, the mono-sialylated MVSHHNLTTGATLINE-A3G3S1 isoform is observed in the early HILIC fractions, elution from the HILIC column first. The other two glycopeptides are observed in progressively later fractions (Figure 2A).

This effect, while influenced by the peptide, is understood to be predominantly due to the glycan moiety[45,69]. To show this, we have plotted the number of spectra matching haptoglobin NLFLNHSE, MVSHHNLTTGATLINE, and VVLHPNYSQVDIGLIK glycopeptides with glycans A3G3S1, A3G3S2 and A3G3S3 against the HILIC fraction of the associated spectrum. (Figure 2B). This graph confirms the predicted elution behavior of glycopeptides containing the A3G3S1 glycans before the glycopeptides containing A3G3S2 and A3G3S3 glycans. This concordance between the expected physico-chemical behavior of glycopeptides and the peptide-glycan pairs proposed by the fraction-oblivious GPS tool increases our confidence that the GPS hits are more than random mass-matches.

Protein Digestion to Facilitate Glycopeptide Identification

Our analysis of glycopeptides was carried out on a commercially available QTOF instrument; no novel fragmentation technologies, high-accuracy precursor, or fragment ion measurements were required for reliable identification of the glycopeptides. We should point out, however, that we do not resolve attachment position or linkage of isobaric glycopeptides, though the position of an N-glycan on a peptide with a single N-glycosylation motif may be inferred. Our strategy for the analysis of haptoglobin and hemopexin was identical except for the inclusion of GluC in the case of haptoglobin. One of the tryptic peptides of haptoglobin, containing two N-linked glycosylation consensus motif sites, can be cleaved into two glycopeptides by the GluC digest. Hemopexin also contains a tryptic peptide with two N-linked glycosylation motif sites, but for this peptide, we have not identified a protease that is able to cleave between the glycosites. At this stage the GPS tool is not optimized for the interpretation of doubly glycosylated peptides and these spectra were assigned under the assumption of a single attached glycan. A manual analysis was able to match spectra for some doubly glycosylated peptides.

In some cases, observed glycopeptides contained internal missed cleavage sites or unexpected termini. Where these peptide sequences can be predicted or manually derived from the spectra, these peptides can be added to the GPS peptide search list. The haptoglobin peptide KVVLHPNYSQVD, with one internal missed cleavage and unexpected

C-terminus was observed in spite of a high excess of both proteases, digestion in a barocycler in case of trypsin, and overnight digestion in case of the GluC digestion. GluC protease is expected to cleave peptide bonds C-terminal to glutamic acid; however a lower-efficiency cleavage has been observed at aspartic acid residues, consistent with the Asp at the C-terminus of this peptide. Detergents or devices facilitating digestion in glycoproteomic research should be further explored[62].

High-Throughput Glycopeptide Search

The lack of effective data interpretation software complicates the characterization of glycopeptides from large LC-MS/MS datasets. Traditional software for peptide identification from tandem mass-spectra support only relatively simple post-translation modifications of known, relatively small, mass offset; thus the size and diversity of N-glycans represents a challenge. Furthermore, the relatively high mass and charge-state of glycopeptides and unfavorable peptide fragmentation in the presence of the glycan moiety make the automated interpretation of spectra a task that awaits a complete solution. We developed a new software tool in house to automate the identification of N-glycopeptides. The software takes advantage of abundant oxonium ions and commonly observed fragments, which indicate the mass of the peptide substrate, and can be reliably observed in CID fragmentation spectra of N-glycopeptides. These ions, including N-acetyl-hexosamine at m/z 204, N-acetyl-hexosamine-hexose at m/z 366, N-acetyl-neuraminic acid at m/z 292 and N-acetyl-hexosamine-hexose-acetyl-neuraminic acid at m/z 657 as markers of N-glycosylation[23]. The software, called GlycoPeptideSearch (GPS), analyzes LC-MS/MS spectral datasets in open data formats (MGF, mzXML, mzData) and outputs its results as an Excel spreadsheet. We expect that the use of open-data formats will make GPS and associated tools useful for researchers on a variety of different CID platforms.

In this work, we have applied the software to the analysis of glycopeptides derived from previously characterized glycoproteins haptoglobin and hemopexin. CID of glycopeptides results in very limited fragmentation of the peptide backbone[64]. Prior knowledge of the N-glycosylated proteolytic peptide sequences (containing the N-glycosylation motif NXS/T) facilitates the analysis of the CID spectra. First, MS/MS spectra are scanned for evidence of glycan fragmentation by searching for oxonium ions at m/z 204 and 366, and their corresponding neutral water-loss ions. These peaks correspond to protonated N-acetyl-hexosamine and hexose-N-acetyl-hexosamine, respectively. In some cases, especially for glycopeptides of very low abundance, oxonium ions are the only abundant ions observed. Regardless of whether the spectra are ultimately interpretable, the oxonium ions indicate the presence of a glycan moiety. Our search tool uses oxonium ions to focus time and effort on the glycopeptide fragmentation spectra in LC-MS/MS spectral datasets containing fragmentation spectra of non-glycosylated peptides.

In addition to oxonium ions, glycopeptide CID MS/MS spectra also contain glycopeptide fragments corresponding to a specific peptide. We have observed that ions representing the intact peptide substrate, with or without a small number of the N-linked glycan core monosaccharaides still attached, are often observed in CID spectra of glycopeptides (see, for example, Figure 3). In our tool, we assess the relative intensity of the first four of these fragment ions in multiple charge states, representing the peptide, the peptide with one or two N-acetylglucosamines attached, and the peptide with two N-acetylglucosamines and one mannose attached. To boost the discriminating power of these matches, we require at least two of these four intact-peptide fragment ions be matched. Together, the presence of two or more of these peptide peaks gives powerful evidence that the spectrum is the result of N-glycopeptide CID fragmentation, and suggests the mass of the peptide substrate. A single glycoprotein has a limited number of protease-specific peptides containing the NXS/T motif. The masses of these peptides, coupled with the specificity of two or more peptide ion

matches (peptide-GlcNAc, peptide-GlcNAc-GlcNAc, peptide-GlcNAc-GlcNAc-Man), provide strong evidence that a peptide matches the spectrum. Even with very low thresholds for testing the presence of the intact-peptide fragment ions, it is rare for this approach to permit more than one peptide to be associated with an N-glycopeptide spectrum.

Once matched to an N-glycopeptide fragmentation spectrum, the mass of the putative peptide is subtracted from the implied precursor mass; the remaining mass is presumed to account for the mass of a single attached glycan, and is used to select N-linked glycans from the human subset of the GlycomeDB glycan database. All glycans within 0.2 Da of the remaining mass are listed, along with the putative peptide substrate, as potential matches to the spectrum.

A given MS/MS spectrum may be matched to one or more peptides paired with one or more glycan structures, all of which match the precursor m/z. Since this approach cannot always discriminate glycopeptides differing only in monosaccharide linkage, we post-process the matching glycopeptides to collapse glycopeptides with identical glycan topology to a single hit. The set of matches from the entire dataset of multiple spectral data-files are output as a single Excel spreadsheet for subsequent curation and analysis. In Table 2 we show the site specific glycoforms of haptoglobin and hemopexin identified by this workflow.

This process would be of little value if hundreds of glycopeptides were proposed for each spectrum, but we have found that our conservative filtering approach typically results in only a few spectra with unclear assignments. First, the oxonium ion and intact peptide fragment filters ensure that only good quality N-glycopeptide fragmentation spectra with good evidence for the mass of the peptide substrate are considered. We find relatively few spectra match to glycopeptides representing more than one peptide. Second, the N-linked human GlycomeDB glycan database is larger than we would consider manually, but each entry is derived (indirectly, from the source glycan databases) from literature, so the set of considered glycans is much more focused than a pure enumerative strategy. In addition, the relative mass-sparsity of glycan compositions and proposed peptides in this general mass range (up to, say, 5,000Da) often results in a single glycan composition, even with the relatively modest 0.2Da mass accuracy used. Third, once topological equivalents are collapsed to a single hit, the glycopeptide hits reduce to a few glycopeptide structures, generally three or less. Usually one of the few remaining hits represents a commonly observed class of glycans and the identification can be made.

Several additional checks can be performed on the glycopeptide matches using information in the MS/MS spectrum. The Excel output of the GPS tool captures additional information to assist in the curation task. For example, spectra with strong oxonium ion peaks at m/z 292 likely contain N-acetylneuraminic acid (NeuAc); therefore, glycopeptide hits without NeuAc may be disregarded. Similar logic may be useful in ascertaining the presence and location of fucose (core-versus outer-arm) was assigned based on oxonium ion fragments at m/z 512.2 (Hex-HexNAc-Fuc) and at m/z 803.3 (NeuAc-Hex-HexNAc-Fuc). Recently published work on the rearrangement of fucose and hexose residues in protonated glycans during fragmentation suggests that interpretation of glycan fragments must be done with care[63]. Also provided in the GPS output are various features which describe the precursor ion, including the charge-state and number of ^{13}C isotopes the proposed glycopeptide ion contains, the mass delta, and the intensity of the precursor's isotope cluster peaks at its proposed monoisotopic m/z value. Ultimately, we hope that these features might be useful for ranking the glycopeptide hits – at this time, we do not attempt to rank the glycopeptides hits matching each spectrum.

Several overlapping peptides with the same glycosylation site were detected in the haptoglobin CID spectra (VVLHPNYSQVDIGLIK, VVLHPNYSQVD, KVVLHPNYSQVD) using GPS. These overlapping peptides often capture multiple independent observations of the same glycosylation event. The overlapping peptides VVLHPNYSQVDIGLIK and VVLHPNYSQVD, with mass-delta of approximately 540 Da, are matched to an almost identical set of glycan-peptide pairs. This offers further evidence that our fragment-directed mass-based matching of glycan-peptide pairs with glycopeptides spectra is a robust method for site-specific glycopeptide identification, ready to be expanded to the analysis of more complex protein mixtures.

The most abundant glycoform, observed with nearly all glycopeptides, was the bi-antennary glycan terminated with two sialic acids (A2G2S2). Other abundant glycoforms include A2G2S1, A3G3S2, and A3G3S3. Collision induced dissociation (CID) of haptoglobin tryptic-GluC peptide NLFNHSE and hemopexin C-terminal tryptic peptide ALPQNVTSLLGCTH, both carrying the A2G2S2 glycan, are shown in Figure 3. The glycopeptides are readily identified by the high intensity glycan oxonium ions, glycopeptide fragments, and even peptide backbone fragments (y_3, y_4, y_5, b_3, b_6 ions in the case of the haptoglobin peptide; y_1, y_3, y_4, y_5, y_6, y_7, y_8, y_{14}, b_{14} ions in case of the hemopexin peptide). Oxonium ions at m/z 657, 292 and 274 confirm the presence of sialic acid and the abundant glycopeptides fragments allow unequivocal assignment of the glycan composition.

Identification of low abundance glycoforms can be more challenging. The fragmentation spectra typically contain only glycan oxonium ions and some glycopeptide fragments, without peptide b- or y-ion fragments, consistent with previously published accounts[23,64]. Identification of low abundance glycoforms, especially of large glycopeptides which tend to ionize and fragment with lower efficiency generally requires a more careful manual analysis.

Many of the glycopeptide matches have significant structural similarity, with smaller glycans matching larger glycans after the removal of one or more terminal monosaccharides. While these matches may well represent the rich microheterogeneity of haptoglobin and hemopexin glycan structures, we cannot rule out the possibility that some of these matches may result from in-source fragmentation of the glycan moiety. To explore this possibility, we varied the QTOF declustering potential (DP) from 20 to 80 volts in a series of glycopeptide analyses, and checked for changes in the MS glycopeptide profile. The absence of significant changes in the MS profile for different DP values suggests that in-source fragmentation is not a common occurrence in our analysis. In our primary analysis above, we use DP 60V.

In the GlycoPeptideSearch software settings, we require identification of several glycopeptide fragments in addition to the parent ion and oxonium ions for identification as described above; in addition, all assignments were manually verified. Limited retention on the reverse phase column may be responsible for the lack of small glycopeptides in our results; in particular, we did not observe the tryptic-GluC glycopeptide NATAK in our C18 trap column based workflow.

Site-Specific Glycoforms of Haptoglobin and Hemopexin

The site-specific glycoforms of haptoglobin have been characterized by several groups[18,32,43]. Wang and coworkers separated tryptic glycopeptides on a polystyrene-divinyl-benzene porous layer tubular column connected to LTQ-CID/ETD mass spectrometer[43]. This ultra-sensitive method allowed an efficient glycopeptide characterization with only 10fmol of sample injected. By manual data annotation, up to 29 site specific glycoforms were identified including two-glycosylation forms of one of the tryptic glycopeptides. The results of Wang et al correlate well with the results of Nakano et

al., who observed some additional tetra-antennary glycopeptides with up to two fucoses, which is likely related to the use of haptoglobin isolated from patients with pancreatic cancer[18]. One study showed a correlation between pancreatic cancer and increased levels of haptoglobin N-glycan fucosylation[61].

In our analysis, we have not observed the tetra-antennary fucosylated glycans; however, we have observed some small glycoforms such as A1G1S1, A2G1 and A2G2, also observed in our detached glycan analysis. The A2G2 glycan was also found by Park et al.[2]. Given the differences in instrument setup and amounts of starting material available, the results of the analyses are quite consistent and appear to represent a thorough description of haptoglobin glycoform microhetrogeneity.

Hemopexin is a 60kDa plasma glycoprotein which contains five N-glycosylation sites and one O-glycosylation site at the N-terminus[65]. Nilsson et al. detected 13 glycoforms in cerebrospinal fluid but the use of hydrazine beads to capture the glycoproteins prevented resolution of the sialylated glycoforms[32]. Our analysis, using semi-automated glycopeptide assignment, identified 14 glycoforms. The results correlate well with the findings of Nilsson et al.; neither study found tetra-antennary glycoforms while our study identified the FA3G3, A3G3 and A4G3 glycoforms also observed by Nilsson et al.

Hemopexin contains the tryptic peptide NGTGHGNSTHHGPEYMR, which has two N-glycosylation sites. Table 2 includes only partially occupied glycoforms with a single attached glycan because the GPS tool does not currently attempt to associate multiple glycans with the peptide substrate. Manual interpretation revealed combinations of A2G2S2+A2G2S2, FA2G2S2+A2G2S2 and A3G3S3+A2G2S2 glycans on this peptide. We have also found O-glycosylation on the N terminal peptide TPLPPTSAHGNVAEGETKPDPDVTER by manual interpretation. This peptide is associated at least 5 O-linked glycoforms. These glycopeptides are well separated by HILIC chromatography with the GalNAc+Hex glycoform eluting in fraction 14, GalNAc+Hex +Sialic acid in fraction 16, GalNAc+Hex+2Sialic acid in fraction 20, 2GalNAc+2Hex +2Sialic acid in fraction 23, and 2GalNAc+2Hex+3Sialic acid in fraction 25. Ultimately, we hope to extend the GPS software tool for the analysis of O-glycans and multiply glycosylated peptides.

The glycopeptide analytical workflow, which pairs HILIC separation of glycopeptides with reverse-phase nanoLC-MS/MS analysis, achieves good coverage of the glycoforms observed in the analysis of detached N-glycans (Figure 4) and those reported by other groups[15,18,32,43,61]. While we have observed very few glycopeptides with tetra-antennary glycans in our workflow, we have observed a number of glycans not observed in the detached N-glycan analysis, including glycans with composition A1G1, A3G3, M3 and A4G3. The complementary nature of the glycopeptide and detached glycan analyses is not unexpected, with a variety of factors contributing to the observation of differing glycan lists. A partial list of these possible factors includes: 1) the analysis of glycans in detached and glycopeptide forms has different sensitivity related to varying recovery, ionization efficiency, or fragmentation behavior; 2) some glycans have very low relative intensities compared to the abundant A2G2S2 glycoform and are below detection limit of the workflow; and 3) the starting material contains low amounts of an additional protein impurity that cannot be detected in glycopeptide analysis.

Sometimes, glycopeptides observed in the MS spectrum cannot be readily identified from their glycopeptide fragmentation spectrum, but in other cases, precursor selection captures low-abundance ions and provides a high-quality tandem mass-spectrum that yields a matching peptide-glycan pair. The haptoglobin glycopeptide VVLHPNYSQVD with

associated A4G4S3 glycan was observed (by manual inspection) in the MS data in charge stages 4+ and 5+, but while the fragmentation spectrum of this glycopeptide was acquired, its quality was so low that it did not pass the peptide-glycan filter. This method may aid in identifying multiple site-specific glycoforms from CID MS/MS data, however, as the previous example demonstrates, there are instances in which we are not able to identify all site-specific glycoforms.

Detached N-Linked Glycan Analysis

To provide a general overview of the N-glycosylation of haptoglobin and hemopexin and to provide a basis for comparison with the glycopeptide workflow, we also analyzed N-glycans detached by PNGaseF using MALDI-TOF mass-spectrometry (Figure 4). This is a well-established approach for surveying detached N-glycans[39,60]. The observed peaks in the haptoglobin spectra were matched to hybrid and complex glycans; while the hemopexin spectra were less diverse and represented complex glycans only (Figure 4). Our results are generally consistent with the results of other studies of glycans and site specific glycoforms of haptoglobin[43,61].

Less information is available on the glycosylation of hemopexin. Several glycans were found in the study of hemopexin purified from serum of healthy volunteers; however, the desialylation of the glycans prior to analysis eliminated the resolution of the sialylated glycoforms[15]. We did not find other papers specifically examining N-glycosylation of hemopexin.

The N-glycans detached from haptoglobin and hemopexin provide a limited set of glycans for manual interpretation of glycopeptide fragmentation spectra in our study. These manual glycopeptide assignments provided a gold-standard set of identifications, facilitating the development of the automated search tool, and the expansion of glycan matching to the human glycans contained in GlycomeDB.

CONCLUSION

Our results show that nano scale reverse phase chromatography combined with QTOF mass spectrometry of glycopeptides enriched by HILIC chromatography allows an efficient identification of the site specific glycoforms of haptoglobin and hemopexin. The identified glycopeptides are in a good agreement with the current literature; in addition, we have identified some glycoforms, especially in the case of hemopexin, that have not been previously reported. The identification of the glycoforms was achieved on a commercially available QTOF instrument with CID fragmentation and without high resolution or high accuracy precursor or fragment ions. The semi-automated assignment of the glycopeptide glycoforms was facilitated by a newly developed GlycoPeptideSearch tool for semi-automatic glycopeptide assignment, which can be downloaded from the Edwards Lab web-site (http://edwardslab.bmcb.georgetown.edu/GPS). We believe that this combination of analytical methods will facilitate the analysis of glycopeptides by a wide community of researchers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by RO1 CA115625 and CA135069 awarded to RG and the CCSG grant NIH P30 CA51008 to the Lombardi Comprehensive Cancer Center supporting the Proteomics and Metabolomics Shared

Resource. Kevin B Chandler was supported by the Graduate Research Fellowship from the National Science Foundation. Petr Pompach was supported in part by the Ministry of Education, Youth and Sports of the Czech Republic (LC545).

REFERENCES

1. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta. 1999; 1473(1):4–8. [PubMed: 10580125]
2. Park SY, Lee SH, Kawasaki N, Itoh S, Kang K, Ryu SH, Nashii N, Kim JM, Kim JY, Kim JH. a1–3/4 fucosylation at Asn 241 of β-haptoglobin is a novel marker for colon cancer: a combinatorial approach for development of glycan biomarkers. International Journal of Cancer. 2011 [Epub ahead of print].
3. Chi YH, Koo YD, Dai SY, Ahn JE, Yun DJ, Lee SY, Zhu-Salzman K. N-glycosylation at non-canonical Asn-X-Cys sequence of an insect recombinant cathepsin B-like counter-defense protein. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology. 2010; 156(1):40–47.
4. Valliere-Douglass JF, Eakin CM, Wallace A, Ketchem RR, Wang W, Treuheit MJ, Balland A. Glutamine-linked and Non-consensus Asparagine-linked Oligosaccharides Present in Human Recombinant Antibodies Define Novel Protein Glycosylation Motifs. Journal of Biological Chemistry. 2010; 285(21):16012–16022. [PubMed: 20233717]
5. Varki, A.; Esko, J.; Colley, JK. Cellular Organization of Glycosylation. In: Varki, A., editor. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. p. 37-46.
6. Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 1993; 3(2):97–130. [PubMed: 8490246]
7. Chantret I, Dupre T, Delenda C, Bucher S, Dancourt J, Barnier A, Charollais A, Heron D, Bader-Meunier B, Danos O, Seta N, Durand G, Oriol R, Codogno P, Moore SE. Congenital disorders of glycosylation type Ig is defined by a deficiency in dolichyl-P-mannose:Man7GlcNAc2-PP-dolichyl mannosyltransferase. J Biol Chem. 2002; 277(28):25815–25822. [PubMed: 11983712]
8. Di Rocco F, Nonaka Y, Hamada H, Yoshino M, Nakazaki H, Oi S. Endoscopic biopsy interpretation difficulties in a congenital diffuse intracranial teratoma. Childs Nerv Syst. 2006; 22(1):84–89. [PubMed: 15789215]
9. Eklund EA, Merbouh N, Ichikawa M, Nishikawa A, Clima JM, Dorman JA, Norberg T, Freeze HH. Hydrophobic Man-1-P derivatives correct abnormal glycosylation in Type I congenital disorder of glycosylation fibroblasts. Glycobiology. 2005; 15(11):1084–1093. [PubMed: 16079417]
10. Mkhikian K, Grigorian A, Li CF, Chen HL, Newton B, Zhou RW, Beeton C, Torossioan S, Tatarian GG, Lee SU, Lau K, Walker E, Siminovitch KA, Chandy KG, Yu Z, Dennis JW, Demetriou M. Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. Nat Communication. 2011; 2(334) Epub.
11. Dennis JW, Nabi IR, Demetriou M. Metabolism, Cell Surface Organization, and Disease. Cell. 2009; 139(7):1229–1241. [PubMed: 20064370]
12. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat. Rev. Cancer. 2005; 5(11):845–856. [PubMed: 16239904]
13. Goldman R, Ressom HW, Varghese RS, Goldman L, Bascug G, Loffredo CA, Abdel-Hamid M, Gouda I, Ezzat S, Kyselova Z, Mechref Y, Novotny MV. Detection of hepatocellular carcinoma using glycomic analysis. Clin. Cancer Res. 2009; 15(5):1808–1813. [PubMed: 19223512]
14. Saldova R, Royle L, Radcliffe CM, Abd Hamid UM, Evans R, Arnold JN, Banks RE, Hutson R, Harvey DJ, Antrobus R, Petrescu SM, Dwek RA, Rudd PM. Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. Glycobiology. 2007; 17(12):1344–1356. [PubMed: 17884841]
15. Debruyne EN, Vanderschaeghe D, Van VH, Vanhecke A, Callewaert N, Delanghe JR. Diagnostic value of the hemopexin N-glycan profile in hepatocellular carcinoma patients. Physiol. Rev. 2010; 56(5):823–831.

16. Kyselova Z, Mechref Y, Al Bataineh MM, Dobrolecki LE, Hickey RJ, Vinson J, Sweeney CJ, Novotny MV. Alterations in the serum glycome due to metastatic prostate cancer. J Proteome. Res. 2007; 6(5):1822–1832. [PubMed: 17432893]

17. Comunale MA, Wang M, Hafner J, Krakover J, Rodemich L, Kopenhaver B, Long RE, Junaidi O, Bisceglie AM, Block TM, Mehta AS. Identification and Development of Fucosylated Glycoproteins as Biomarkers of Primary Hepatocellular Carcinoma. J. Proteome. Res. 2009; 8(2): 595–602. [PubMed: 19099421]

18. Nakano M, Nakagawa T, Ito T, Kitada T, Hijioka T, Kasahara A, Tajiri M, Wada Y, Taniguchi N, Miyoshi E. Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: a novel approach for the development of tumor markers. Int. J Cancer. 2008; 122(10): 2301–2309. [PubMed: 18214858]

19. Medzihradszky KF. Characterization of protein N-glycosylation. Mass Spectrometry: Modified Proteins and Glycoconjugates. 2005; 405:116–138.

20. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA. Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. Chemical Reviews. 2002; 102(2):371–386. [PubMed: 11841247]

21. Ye LB, Li JR, Zhang JS, Pan YJ. NMR characterization for polysaccharide moiety of a glycopeptide. Fitoterapia. 2010; 81(2):93–96. [PubMed: 19686815]

22. Sutton CW, Oneill JA, Cottrell JS. Site-Specific Characterization of Glycoprotein Carbohydrates by Exoglycosidase Digestion and Laser-Desorption Mass-Spectrometry. Analytical Biochemistry. 1994; 218(1):34–46. [PubMed: 8053566]

23. Wuhrer M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics based on tandem mass spectrometry of glycopeptides. J Chromatogr. B Analyt. Technol. Biomed. Life Sci. 2007; 849(1–2):115–128.

24. Wuhrer M, de Boer AR, Deelder AM. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. Mass Spectrom. Rev. 2009; 28(2):192–206. [PubMed: 18979527]

25. Boersema PJ, Mohammed S, Heck AJ. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal. Bioanal. Chem. 2008; 391(1):151–159. [PubMed: 18264818]

26. Hua S, Nwosu CC, Strum JS, Seipert RR, An HJ, Zivkovic AM, German JB, Lebrilla CB. Site-specific protein glycosylation analysis with glycan isomer differentiation. Anal Bioanal. Chem. 2011 Jun 8. [Epub ahead of print].

27. Mayampurath AM, Wu Y, Segu ZM, Mechref Y, Tang HX. Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity. Rapid Communications in Mass Spectrometry. 2011; 25(14):2007–2019. [PubMed: 21698683]

28. Yang Z, Harris LE, Palmer-Toy DE, Hancock WS. Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. Clin. Chem. 2006; 52(10):1897–1905. [PubMed: 16916992]

29. Madera M, Mechref Y, Klouckova I, Novotny MV. High-sensitivity profiling of glycoproteins from human blood serum through multiple-lectin affinity chromatography and liquid chromatography/tandem mass spectrometry. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 2007; 845(1):121–137.

30. Zhang W, Wang H, Zhang L, Yao J, Yang PY. Large-scale assignment of N-glycosylation sites using complementary enzymatic deglycosylation. Talanta. 2011; 85(1):499–505. [PubMed: 21645732]

31. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. 2003; 21(6):660–666. [PubMed: 12754519]

32. Nilsson J, Ruetschi U, Halim A, Hesse C, Carlsohn E, Brinkmalm G, Larson G. Enrichment of glycopeptides for glycan structure and attachment site identification. Nat. Methods. 2009; 6(11): 809–811. [PubMed: 19838169]

33. Arnold JN, Saldova R, Hamid UM, Rudd PM. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. Proteomics. 2008; 8(16):3284–3293. [PubMed: 18646009]

34. Mechref Y, Novotny MV. Structural investigations of glycoconjugates at high sensitivity. Chem. Rev. 2002; 102(2):321–369. [PubMed: 11841246]
35. Chu CS, Ninonuevo MR, Clowers BH, Perkins PD, An HJ, Yin H, Killeen K, Miyamoto S, Grimm R, Lebrilla CB. Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. Proteomics. 2009; 9(7):1939–1951. [PubMed: 19288519]
36. Ashline D, Singh S, Hanneman A, Reinhold V. Congruent strategies for carbohydrate sequencing. 1. Mining structural details by MSn. Anal. Chem. 2005; 77(19):6250–6262. [PubMed: 16194086]
37. Jiang H, Wu SL, Karger BL, Hancock WS. Characterization of the glycosylation occupancy and the active site in the follow-on protein therapeutic: TNK-tissue plasminogen activator. Anal. Chem. 2010; 82(14):6154–6162. [PubMed: 20552988]
38. Ueda K, Takami S, Saichi N, Daigo Y, Ishikawa N, Kohno N, Katsumata M, Yamane A, Ota M, Sato TA, Nakamura Y, Nakagawa H. Development of Serum Glycoproteomic Profiling Technique; Simultaneous Identification of Glycosylation Sites and Site-Specific Quantification of Glycan Structure Changes. Molecular & Cellular Proteomics. 2010; 9(9):1819–1828. [PubMed: 20811073]
39. Ressom HW, Varghese RS, Goldman L, An Y, Loffredo CA, Abdel-Hamid M, Kyselova Z, Mechref Y, Novotny M, Drake SK, Goldman R. Analysis of MALDI-TOF mass spectrometry data for discovery of peptide and glycan biomarkers of hepatocellular carcinoma. J Proteome. Res. 2008; 7(2):603–610. [PubMed: 18189345]
40. Goldberg D, Bern M, Parry S, Sutton-Smith M, Panico M, Morris HR, Dell A. Automated N-glycopeptide identification using a combination of single- and tandem-MS. J Proteome Res. 2007; 6(10):3995–4005. [PubMed: 17727280]
41. Isailovic D, Kurulugama RT, Plasencia MD, Stokes ST, Kyselova Z, Goldman R, Mechref Y, Novotny MV, Clemmer DE. Profiling of human serum glycans associated with liver cancer and cirrhosis by IMS-MS. J Proteome. Res. 2008; 7(3):1109–1117. [PubMed: 18237112]
42. Novotny MV, Mechref Y. New hyphenated methodologies in high-sensitivity glycoprotein analysis. J Sep. Sci. 2005; 28(15):1956–1968. [PubMed: 16276785]
43. 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 mum 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]
44. Alley WR Jr, Mechref Y, Novotny MV. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. Rapid Commun. Mass Spectrom. 2009; 23(1):161–170. [PubMed: 19065542]
45. Gilar M, Yu YQ, Ahn J, Xie HW, Han HH, Ying WT, Qian XH. Characterization of glycoprotein digests with hydrophilic interaction chromatography and mass spectrometry. Analytical Biochemistry. 2011; 417(1):80–88. [PubMed: 21689629]
46. Kang P, Mechref Y, Novotny MV. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. Rapid Commun. Mass Spectrom. 2008; 22(5):721–734. [PubMed: 18265433]
47. Tang Z, Varghese RS, Bekesova S, Loffredo CA, Hamid MA, Kyselova Z, Mechref Y, Novotny MV, Goldman R, Ressom HW. Identification of N-Glycan Serum Markers Associated with Hepatocellular Carcinoma from Mass Spectrometry Data. J Proteome. Res. 2010; 9:104–112. [PubMed: 19764807]
48. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 1964; 36:1627–1639.
49. Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung MC, Kuerer HM. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. Proteomics. 2005; 5(16):4107–4117. [PubMed: 16254928]
50. Hegland M. A Self-Sorting In-Place Fast Fourier-Transform Algorithm Suitable for Vector and Parallel-Processing. Numerische Mathematik. 1994; 68(4):507–547.

51. Tang H, Mechref Y, Novotny MV. Automated interpretation of MS/MS spectra of oligosaccharides. Bioinformatics. 2005; 21 Suppl 1:431–439.
52. Goldberg D, Sutton-Smith M, Paulson J, Dell A. Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. Proteomics. 2005; 5(4):865–875. [PubMed: 15693066]
53. Ranzinger R, Herget S, Wetter T, der Lieth CW. GlycomeDB - integration of open-access carbohydrate structure databases. BMC Bioinformatics. 2008; 9:384. [PubMed: 18803830]
54. Herget S, Ranzinger R, Maass K, Lieth CW. GlycoCT-a unifying sequence format for carbohydrates. Carbohydr. Res. 2008; 343(12):2162–2171. [PubMed: 18436199]
55. Kessner D, Chambers M, Burke R, Agusand D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008; 24(21):2534–2536. [PubMed: 18606607]
56. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim YG, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA. HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. Anal Biochem. 2008; 376(1):1–12. [PubMed: 18194658]
57. Dobryszycka W. Biological functions of haptoglobin--new pieces to an old puzzle. Eur J Clin Chem Clin Biochem. 1997; 35(9):647–654. [PubMed: 9352226]
58. Delanghe JR, Langlois MR. Hemopexin: a review of biological aspects and the role in laboratory medicine. Clin Chim. Acta. 2001; 312(1–2):13–23. [PubMed: 11580905]
59. Zhang S, Shu H, Luo KX, Kang XN, Zhang Y, Lu HJ, Liu YK. N-linked glycan changes of serum haptoglobin beta chain in liver disease patients. Molecular BioSystems. 2011; 7(5):1621–1628. [PubMed: 21380457]
60. Wada Y, Azadi P, Costello CE, Dell A, Dwek RA, Geyer H, Geyer R, Kakehi K, Karlsson NG, Kato K, Kawasaki N, Khoo KH, Kim S, Kondo A, Lattova E, Mechref Y, Miyoshi E, Nakamura K, Narimatsu H, Novotny MV, Packer NH, Perreault H, Peter-Katalinic J, Pohlentz G, Reinhold VN, Rudd PM, Suzuki A, Taniguchi N. Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. Glycobiology. 2007; 17(4):411–422. [PubMed: 17223647]
61. Lin ZX, Simeone DM, Anderson MA, Brand RE, Xie XL, Shedden KA, Ruffin MT, Lubman DM. Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer. Journal of proteome research. 2011; 10(5):2602–2611. [PubMed: 21417406]
62. Drapeau GR, Houmard J, Boily Y. Purification and Properties of An Extracellular Protease of Staphylococcus-Aureus. Journal of Biological Chemistry. 1972; 247(20) 6720-&.
63. Wuhrer M, Koeleman CA, Hokke CH, Deelder AM. Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. Rapid Commun Mass Spectrom. 2006; 20(11):1747–1754. [PubMed: 16676317]
64. Huddleston MJ, Bean MF, Carr SA. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. Anal Chem. 1993; 65(7):877–884. [PubMed: 8470819]
65. Takahashi N, Takahashi Y, Putnam FW. Complete Amino-Acid Sequence of Human Hemopexin, the Hemebinding Protein of Serum. Proceedings of the National Academy of Sciences of the United States of America. 1985; 82(1):73–77. [PubMed: 3855550]
66. Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J Proteome Res. 2004; 3(3):556–566. [PubMed: 15253437]
67. Mysling 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):5598–5609. [PubMed: 20536156]
68. Zhang J, Wang DI. Quantitative analysis and process monitoring of site-specific glycosylation microheterogeneity in recombinant human interferon-gamma from Chinese hamster ovary cell culture by hydrophilic interaction chromatography. J Chromatogr. B Biomed. Sci Appl. 1998; 712(1–2):73–82. [PubMed: 9698230]

69. Gilar M, Jaworski A. Retention behavior of peptides in hydrophilic-interaction chromatography. J. Chromatogr. A. 2011 [Epub ahead of print].

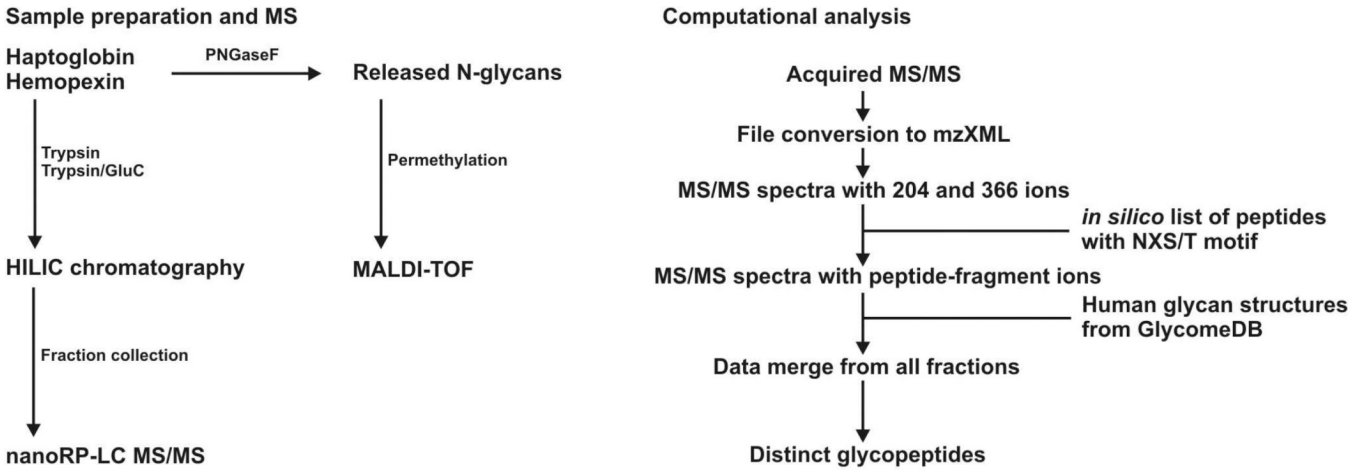

Figure 1.
Flowchart demonstrating experimental and data interpretation workflows.

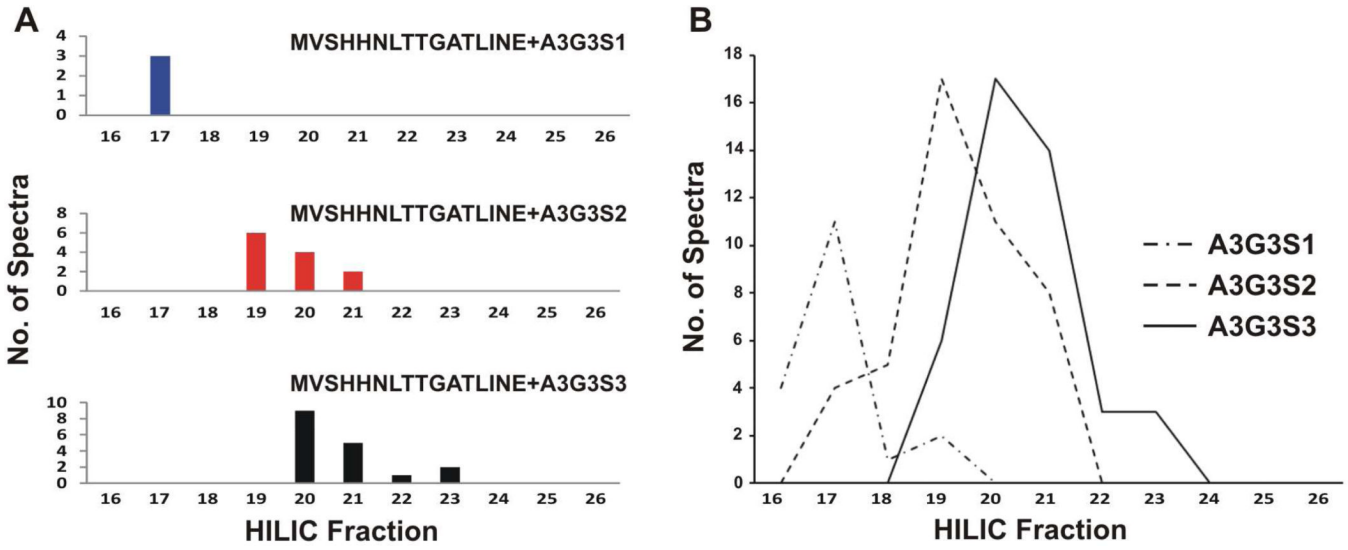

Figure 2.

A) Plots of one glycopeptide (MVSHHNLTTGATLINE) with three non-fucosylated tri-antennary N-glycans with increasing number of N-acetylneuraminic acid residues demonstrating the influence of sialylation on HILIC retention and B) plot merging three haptoglobin glycopeptides (MVSHHNLTTGATLINE, NLFLNHSE and VVLHPNYSQVD) bearing non-fucosylated tri-antennary N-glycans with increasing number of N-acetylneuraminic acid residues.

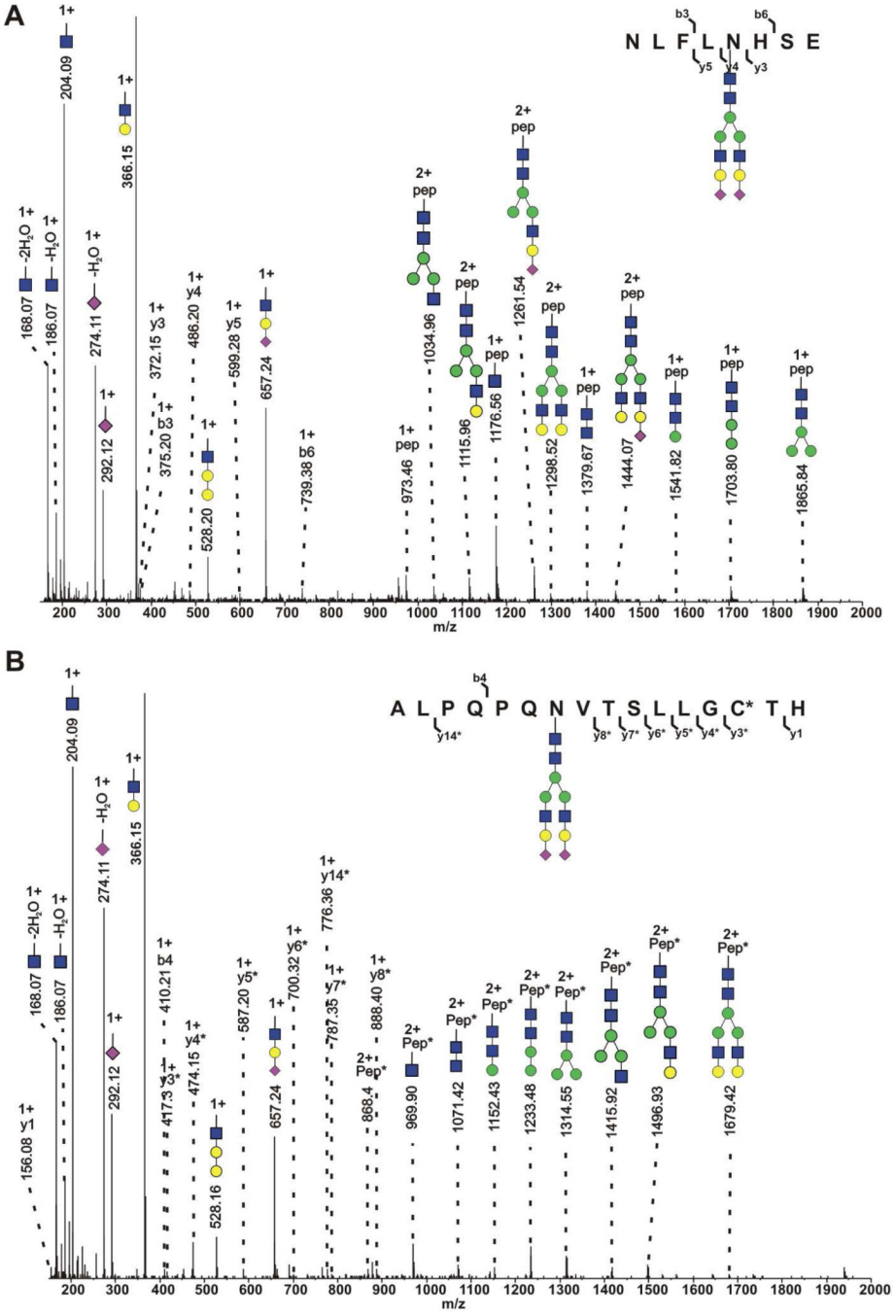

Figure 3.
CID spectrum of A) tryptic-GluC glycopeptide from haptoglobin and B) tryptic glycopeptide from hemopexin. Fragment ions modified by iodoacetamide are marked with asterisk.

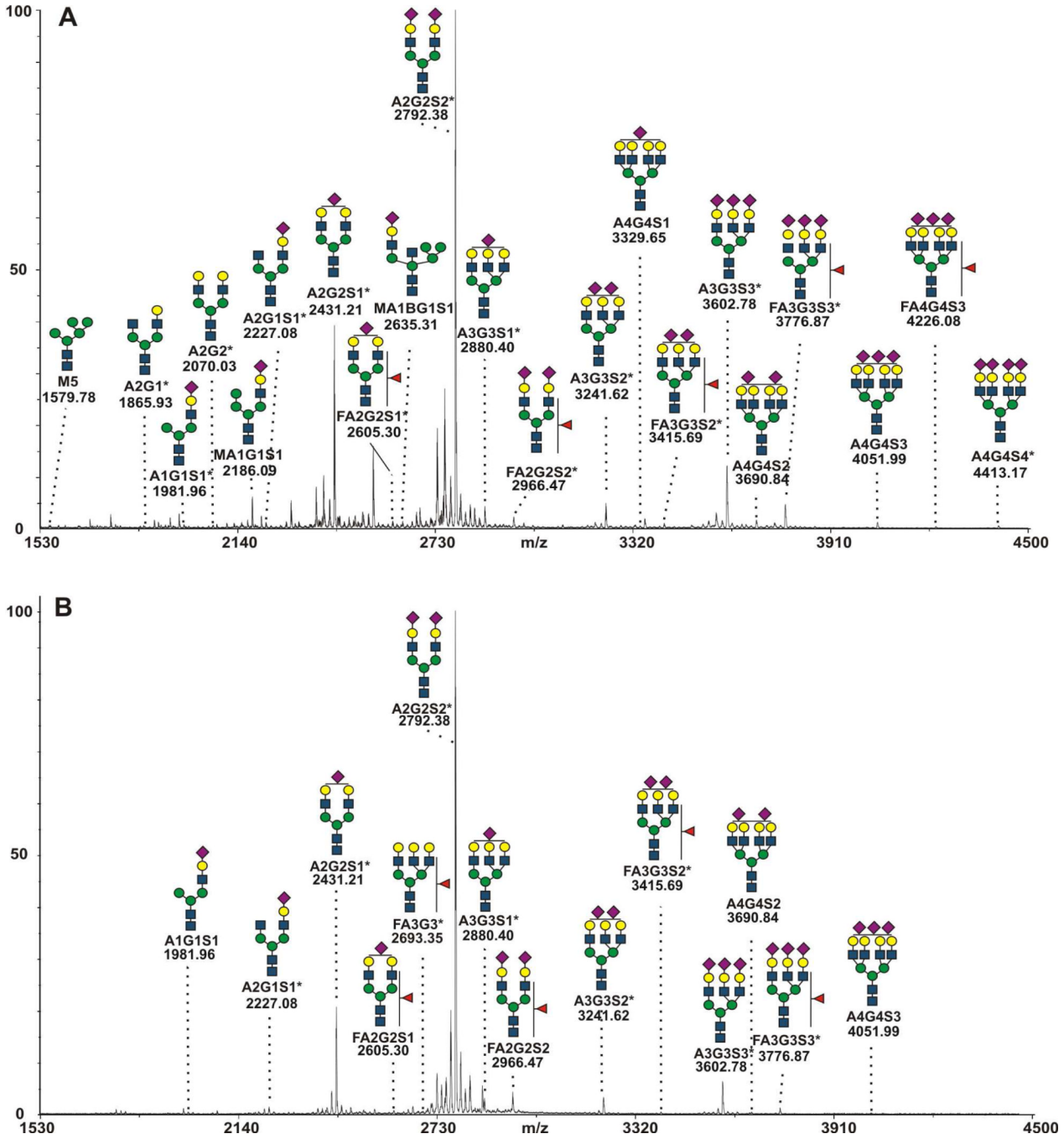

Figure 4.

MALDI-TOF spectra of detached permethylated glycans from A) haptoglobin and B) hemopexin. Glycans observed as site specific forms are marked with an asterisk.

Table 1

Haptoglobin and hemopexin peptides with N-linked glycosylation sequence motif considered in the manual and automated analysis.

Protein / Protease(s)	Amino Acid Sequence	$[M+H^+]^+$
Haptoglobin		
Trypsin/GluC	^{211}*N*ATAK	504.28
Trypsin/GluC	NLFL^{207}*N*HSE	973.47
Trypsin/GluC	VVLHP^{241}*N*YSQVD	1270.64
Trypsin/GluC	KVVLHP^{241}*N*YSQVD	1398.74
Trypsin	NLFL^{207}*N*HSE^{211}*N*ATAK	1458.73
Trypsin/GluC	MVSHH^{184}*N*LTTGATLINE	1737.86
Trypsin	VVLHP^{241}*N*YSQVDIGLIK	1795.01
Hemopexin		
Trypsin	SWPAVG^{164}*N*CSSALR	1347.64
Trypsin	ALPQPQ^{430}*N*VTSLLGCTH	1678.85
Trypsin	^{217}*N*GTGHG^{223}*N*STHHGPEYMR	1851.79

Table 2

Site specific glycoforms of haptoglobin and hemopexin identified by the GlycoPeptideSearch software. N-glycan notation was adopted from NIBRT GlycoBase database. The charge represents the most abundant or unique charge state of the glycopeptide. M/z column shows the deconvoluted glycopeptide masses calculated from observed ions.

Protein / Glycopeptide	Glycan Composition	Charge	$[M+H^+]^+$	Error (ppm)
Haptoglobin				
KVVLHPNYSQVD	A1G1	3+	2656.26	8
	A1G1S1*	3+	2947.30	3
	FA1G1S1	4+	3296.39	6
	A2G2S1*	3+	3312.46	12
	A2G2S2*	4+	3603.53	6
	A3G3S1*	4+	3677.56	5
MVSHHNLTTGATLINE	M3	2+	2630.17	0
	A1G1	3+	2995.34	13
	A2G1*	3+	3198.39	6
	A1G1S1*	3+	3286.44	9
	A2G2*	3+	3360.48	15
	A2G2S1*	3+	3651.54	0
	A3G3	3+	3725.53	13
	FA2G2S1*	4+	3797.59	5
	A2G2S2*	3+	3942.62	3
	A3G3S1*	4+	4016.70	8
	FA2G2S2*	4+	4088.70	5
	A3G3S2*	3+	4308.73	7
	FA3G3S2*	4+	4453.79	4
	A3G3S3*	4+	4598.84	4
	FA3G3S3*	4+	4744.91	2
NLFLNHSE	A1	2+	2068.87	0
	A1G1	2+	2230.93	4
	A1G1S1*	2+	2522.02	4
	A2G2*	3+	2596.10	19
	A2G2S1*	3+	2887.19	10
	A2G2S2*	3+	3178.28	9
	A3G3S1*	3+	3252.30	3
	FA2G2S2*	3+	3324.30	0
	FA3G3S2*	4+	3689.54	8

Protein / Glycopeptide	Glycan Composition	Charge	$[M+H^+]^+$	Error (ppm)
	A3G3S3*	3+	3834.49	3
	FA3G3S3*	4+	3980.61	16
VVLHPNYSQVD	M3	2+	2162.96	0
	A1G1	2+	2528.12	12
	A1G1S1*	3+	2819.17	11
	A2G2*	3+	2893.24	7
	A2G1S1*	3+	3022.27	3
	A2G2S1*	3+	3184.33	0
	A2G2S2*	3+	3475.46	12
	FA2G2S1*	3+	3330.38	6
	A3G3S1*	4+	3549.45	3
	FA2G2S2*	3+	3621.42	14
	A3G3S2*	4+	3840.53	5
	FA3G3S2*	4+	3986.67	10
	A3G3S3*	4+	4131.59	15
	FA3G3S3*	3+	4277.75	9
VVLHPNYSQVDIGLIK	M3	2+	2687.33	4
	A1G1	3+	3052.51	20
	A1G1S1*	3+	3343.58	6
	A2G2*	3+	3417.58	0
	A2G1S1*	3+	3546.73	31
	A2G2S1*	4+	3708.69	0
	A2G2S2*	4+	3999.72	15
	A3G3S1*	5+	4073.92	24
	FA2G2S2*	4+	4145.89	14
	A3G3S2*	4+	4364.99	18
	FA3G3S2*	4+	4511.13	2
	A3G3S3*	4+	4656.07	13
	FA3G3S3*	4+	4802.15	17
	A4G4S4	5+	5311.08	28
	FA4G4S4*	5+	5457.39	20
Hemopexin				
SWPAVGNC#SSALR	A2G1S1*	3+	3156.22	28
	A3G3	4+	3392.36	9
	A2G2S2*	3+	3609.28	47

Protein / Glycopeptide	Glycan Composition	Charge	$[M+H^+]^+$	Error (ppm)
	A3G3S2*	4+	3974.44	38
	FA3G3S2*	4+	4120.48	38
	A3G3S3*	4+	4265.52	38
	FA3G3S3*	4+	4411.57	38
SWPAVGNCSSALR	A3G3S1*	4+	3626.30	44
ALPQPQNVTSLLGC#TH	A2G2S1*	4+	3649.47	27
	A2G2S2*	4+	3940.62	13
	A3G3S3*	4+	4596.74	33
ALPQPQNVTSLLGCTH	FA3G3S3*	4+	4685.77	32
NGTGHGNSTHHGPEYMR	FA3G3*	4+	3985.42	30
	A4G3	4+	4042.51	20

Carbamidomethylated cysteines are indicated by the pound symbol (#) and glycopeptides observed as detached glycans are indicated by an asterisk (*).