Haidi Yin[1,2]
Mingrui An[3]
Pui-kin So[4]
Melody Yee-Man Wong[5]
David M. Lubman[3]
Zhongping Yao[2]

[1]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, P. R. China
[2]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[3]Department of Surgery, University of Michigan Medical Center, Ann Arbor, MI, USA
[4]University Research Facility in Life Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[5]University Research Facility in Chemical and Environmental Analysis, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Received November 13, 2017
Revised January 25, 2018
Accepted January 26, 2018

Research Article

The analysis of alpha-1-antitrypsin glycosylation with direct LC-MS/MS

A liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based methodology has been developed to differentiate core- and antennary-fucosylated glycosylation of glycopeptides. Both the glycosylation sites (heterogeneity) and multiple possible glycan occupancy at each site (microheterogeneity) can be resolved via intact glycopeptide analysis. The serum glycoprotein alpha-1-antitrypsin (A1AT) which contains both core- and antennary-fucosylated glycosites was used in this study. Sialidase was used to remove the sialic acids in order to simplify the glycosylation microheterogeneity and to enhance the MS signal of glycopeptides with similar glycan structures. β1-3,4 galactosidase was used to differentiate core- and antennary-fucosylation. In-source dissociation was found to severely affect the identification and quantification of glycopeptides with low abundance glycan modification. The settings of the mass spectrometer were therefore optimized to minimize the in-source dissociation. A three-step mass spectrometry fragmentation strategy was used for glycopeptide identification, facilitated by pGlyco software annotation and manual checking. The collision energy used for initial glycopeptide fragmentation was found to be crucial for improved detection of oxonium ions and better selection of Y1 ion (peptide+GlcNAc). Structural assignments revealed that all three glycosylation sites of A1AT glycopeptides contain complex N-glycan structures: site Asn70 contains biantennary glycans without fucosylation; site Asn107 contains bi-, tri- and tetra-antennary glycans with both core- and antennary-fucosylation; site Asn271 contains bi- and tri-antennary glycans with both core- and antennary-fucosylation. The relative intensity of core- and antennary-fucosylation on Asn107 was similar to that of the A1AT protein indicating that the glycosylation level of Asn107 is much larger than the other two sites.

Keywords:
Alpha-1-antitrypsin / Fucosylation / In-source collision-induced dissociation / Mass spectrometry
DOI 10.1002/elps.201700426

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Aberrant protein glycosylation especially fucosylation has been found to be associated with various diseases such as cancers [1]. The fucose that attaches to core N-acetyl glucosamine of N-glycans is core-fucosylation and those that attach to the antennary N-acetylglucosamine or galactose is antennary fucosylation. The change in core- or antennary-fucosylation of some proteins has been found to be indicative for various cancers. For example, the enhanced level of the core-fucosylation of alpha-fetoprotein (AFP-L3) in the serum was found to be associated with hepatocellular carcinoma [2]. AFP-L3 is detected using a *Lens culinaris* lectin (LCA) blot assay based on ivmmunoassay and the high affinity of LCA to core-fucosylated glycoproteins [3]. Another example is CA19-9, a type of antennary-fucosylation sialyl lewis A structure. The enhanced level of CA19-9 in the serum is the most widely used clinical marker for pancreatic cancer [4]. CA19-9 is monitored by immunoassay using a sialyl lewis A

Correspondence: Dr. Haidi Yin, The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, P. R. China; Department of Applied Biology and Chemical Technology, W704, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
E-mail: haidi.yin@polyu.edu.hk

Abbreviations: **A1AT**, alpha-1-antitrypsin; **CE**, collision energy; **CID**, collision-induced dissociation; **HCD**, higher-energy collisional dissociation; **IAA**, iodoacetamide; **LC-MS/MS**, liquid chromatography with tandem mass spectrometry; **TCEP**, tris (2-carboxyethyl) phosphine; **TEAB**, triethylammonium bicarbonate; **TFA**, trifluoroacetic acid; **XIC**, Extracting Ion Current

Color Online: See the article online to view Figs. 1–8 in color.

structure specific antibody [4]. This method relies on a specific antibody so it cannot be easily applied to other glycoproteins. In addition to the above immunoassay based method, another conventional approach for core- and antennary-fucosylation analysis involves a combination of various fucosidases and several cycles of HPLC separation [5] of glycans after cleaving glycans from glycoproteins. Although a recently developed immobilized PNGase F digestion procedure has enabled fast release of glycans from glycoproteins [6], the approach of fucosidase digestion is tedious. More importantly, most proteins have multiple fucosylation sites where the above analyses lose the site-specific information and thus cannot provide direct evidence for core- or antennary- fucosylation aberration of proteins, which is key for precise diagnosis.

Many studies have been exploring MS-based profiling of intact glycopeptides, such as increasing sensitivity, resolution and fragmentation of mass spectrometers and developing software for the data analysis of glycopeptides [7, 8]. Studies have been using CID, ECD, ETD, EThcD, low and high energy HCD fragmentation *et al*. or combinations of these fragmentation methods [9–11] to elucidate the structure of glycopeptides. Several different softwares for elucidating these spectra have been developed, among which Byonics [12] and GPQuest [13] are so far the most widely used. However, Byonics relies on peptide sequence-based scoring which underestimates the false positive discovery rate of glycopeptides [14] and GPQuest needs a sample-originated peptide library for matching of glycopepties, which makes the experiment more complicated [13]. Here we employed the newly developed pGlyco software from the groups of Yang PY and He SM to facilitate the MS analysis of glycopeptides. pGlyco uses HCD MS2 generated oxonium ions to filter glycopeptides, uses HCD MS3 on Y1 ions for peptide sequencing, and uses CID MS2 for glycan elucidation [15]. pGlyco2.0 is an updated version, which uses stepped HCD collision [14]. Although the involvement of MS3 in pGlyco makes the scan speed a bit slower, it enables manual check of both glycan structures and peptide sequences with much more complex fragments compared with pGlyco2.0. We therefore used pGlyco as the preferred method.

Using LC-MS/MS alone, it is often difficult to distinguish core- and antennary-fucosylation due to their similar retention time on a C18 column and the same m/z of the glycopeptides [16] and then there is also possible migration of fucose from antennary- to core-position during MS/MS fragmentation [17]. Glycan derivatization such as permethylation is able to solve the problem of fucose migration, but before derivatization glycans need to be released from glycopeptides so that the site-specific information is lost [17, 18]. We thus sought to develop a method to differentiate core- and antennary-fucosylation prior to LC-MS/MS and to use pGlyco facilitated mass spectrometry analysis to identify and semi-quantify core- and antennary-fucosylation. In this study, we applied sialidase and galactosidase double digestion to differentiate core- and antennary-fucosylation before mass spectrometry analysis. Sialidase was used to remove sialic acids to simplify the glycosylation microheterogeneity and to enhance the MS signal of glycopeptides. ß1-3,4 galactosidase (from bovine testis) was used to differentiate core- and antennary-fucosylation, where galactosidase is not able to cleave galactose from antennary fucosylated Lewis structures [19].

The fucosylation level of serum protein alpha-1-antitrypsin (A1AT) has been identified as a potential biomarker for various cancers [20, 21] and inflammation [22]. In this study, sialidase and galactosidase double digestion of glycopeptide was followed by direct LC-MS/MS analysis without cleaving glycans from the glycopeptides. Both glycosylation site and multiple possible glycan occupancy at each site were resolved, with successful identification and semi-quantification of the glycopeptides of A1AT and with clear differentiation of core- and antennary-fucosylation.

2 Materials and methods

2.1 Trypsin digestion of protein into peptides

We added 10 μL of 50 mM ammonia bicarbonate to 10 μg alpha-1-antitrypsin (A1AT) and pipetted to dissolve the sample well. The dissolved A1AT was reduced with 10 mM tris (2-carboxyethyl) phosphine (TCEP) at 37°C for 30 min and alkylated with 20 mM iodoacetamide (IAA) at room temperature in the dark for 15 min. The sample solution is diluted for 3 times with 50 mM ammonia bicarbonate and incubated with 1 μL of 0.5 μg/μL trypsin (Promega, Madison, WI) at 37°C for 16 h. The trypsin is eventually deactivated at 95°C for 5 min and dried in a speedvac.

2.2 Enrichment and buffer-exchange of glycopeptides

3 K Ultra centrifugal filter-15 (Millipore Amicon) was used for glycopeptide enrichment and for buffer exchange. The buffer system was changed from the above system to 25 mM sodium acetate (pH5.5) for three times at 7500 × g for 1 h. Glycopeptides with modification were larger than 3 K so that only non-glycopeptides smaller than 3 K will pass through the 3 K membrane.

2.3 Sialidase/galactosidase double digestion

For the sialidase and galactosidase digestion, the glycopeptide mixture in 30 μL of 25 mM sodium acetate solution was incubated with 15 mU (3 μL) of non-specific α2-3,6,8,9 sialidase recombinant from *Arthrobacter ureafaciens* expressed in *E. coli* (Prozyme, Hayward, CA) and 75 mU (3 μL) of β1-3,4 galactosidase from bovine testis (Prozyme, Hayward, CA) at 37°C for 18 h to remove all sialic acid residues and galactose provided that no fucose is bound to the sub-terminal N-acetylglucosamine in an N-glycan. The glycosidases were deactivated at 95°C for 5 min.

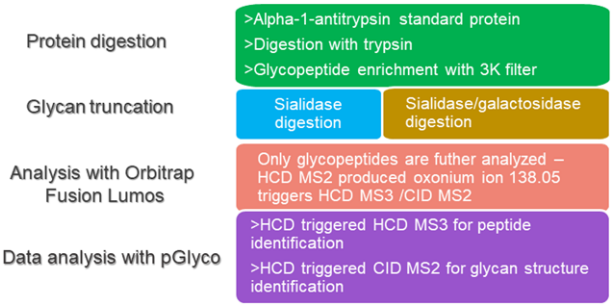

Figure 1. Work-flow of the experiment for determining glycosylation of A1AT. The glycosylated A1AT was first digested into peptides, followed by glycan truncation by sialidase/galactosidase double digestion. Glycopeptides were subjected to direct LC-MS/MS analysis without cleaving glycans.

2.4 C18 desalting

Trifluoroacetic acid (TFA) was added until the pH value reached 2. The C18 columns (Fisher Scientific, San Jose, CA) were activated with 200 μL 0.1% TFA in 50% acetonitrile for five times and equilibrated with 0.1% TFA in water for three times by centrifugation at 1500 × *g*/min for 1 min each time. The peptides were bound to the C18 beads for five times followed by three times washing with 0.1% TFA to remove non-specific binding by centrifugation as described above; 20 μL of 50% acetonitrile with 0.1% TFA was used for elution by centrifugation as described above. Elution was repeated once and the combined eluents were then dried in a speedvac.

2.5 LC-MS identification of glycopeptides

Nano LC-MS/MS conditions were as described in previous work [23]. A C18 capillary column (100 μm × 15 cm; 3 μm particles, 200 Å) (Thermo fisher Scientific, San Jose, CA) was used for LC separation, and gradient elution was performed using an Ultimate 3000 nanoLC system (Thermo fisher Scientific, San Jose, CA) with a flow rate of 350 nL/min. The mobile phase A was 2% acetonitrile with 0.1% formic acid in water and mobile phase B was 2% water with 0.1% formic acid in acetonitrile. The analytical gradient lasted for 100 min where after 10 min balancing time, the composition of solvent B rose from 3 to 7% in 2 min, from 7 to 14% in 8 min, from 14 to 25% in 55 min, followed by a washing and equilibration step where solvent B increased to 90% in 5 min and was held for 8 min, and then returned to 3% B in 0.1 min and was held for 17 min.

An Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) operated in positive ion mode was used for analysis. The ESI spray voltage and capillary voltage were set as described in the following part. Two runs of LC-MS were performed for each sample. Each run

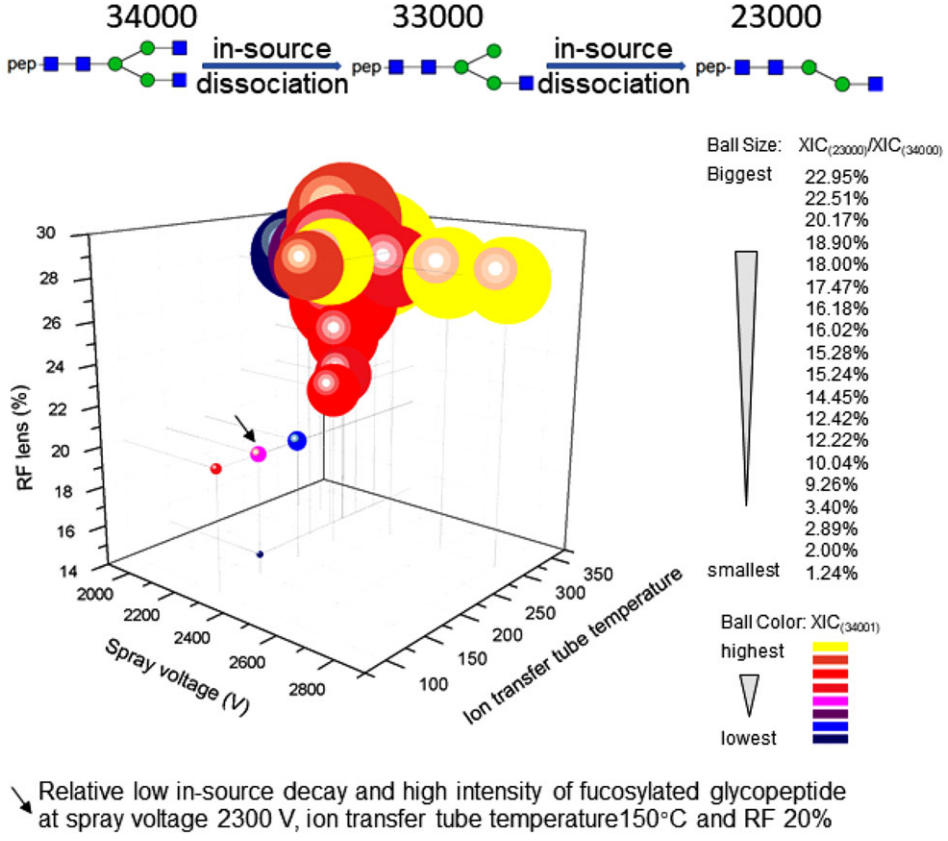

Figure 2. In-source dissociation of A1AT glycopeptides (Asn271) with A2 and FA2 glycan modification types under various settings of ion transfer tube temperature, spray voltage and RF%.

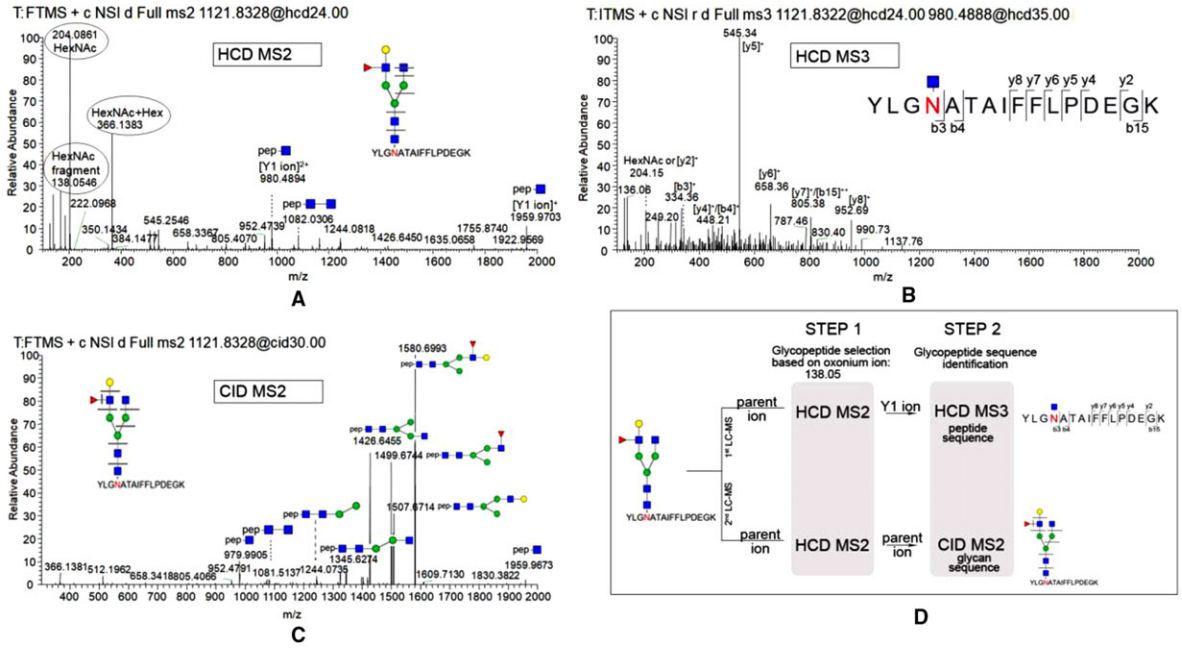

Figure 3. Spectra of A1AT glycopeptide (Asn271) with A2FG glycan modification type: (A) low energy HCD MS2 spectrum; (B) low energy HCD triggered high energy HCD MS3 spectrum; (C) low energy HCD triggered CID MS2; (D) Illustration of the fragmentation of the glycopeptide.

has two consecutive MS scan types. In the first run, glycopeptides were selected by the detection of oxonium ion 138.05 with low energy HCD MS2; consequently the Y1 ion (peptide+GlcNAc) from the glycopeptide fragment was subjected to high energy HCD MS3 for peptide sequencing. In the second run, after glycopeptide selection by low energy HCD MS2, the selected glycopeptide was subjected to CID MS2 for glycan structure analysis. The collision energy for each step of fragmentation was also optimized for better detection of oxonium ions, better selection of Y1 ion and better fragmentation of Y1 ion and glycans, as discussed in the results section. A full scan defines the mass range of m/z 600 to 1800, and MS/MS was performed with top speed mode.

2.6 Database search for glycopeptide identification

The search engines pGlyco and pFind developed by He SM's group were used for glycoprotein analysis. The raw data of two LC-MS runs were aligned first to make sure the retention time of the same precursor ion was the same in the two runs. pFind was used for Y1 peptide identification using MS3 spectra from the first LC-MS run: (i) fixed modification: cysteine carbamidomethylation (+57.021 Da); (ii) Dynamic modification: methionine oxidation (+15.995 Da) and NexHAc (+203.075 Da); (iii) One missed cleavage was allowed; (iv) Peptide ion tolerance: 15 ppm; (v) Fragment ion tolerance: 25 ppm. Identified Y1 peptides from the first LC-MS run and the raw data from the second LC-MS run were imported to pGlyco for glycopeptide matching and scoring. All identified glycopeptides were manually checked by GlycoWorkbench Software developed by the EUROCarbDB [24]. The nomenclature of glycans is used according to Essentials of Glycobiology [25] and the abbreviations are used according to the NIBRT GlycoBase.

3 Results and discussion

Both types of fucosylation structures of N-linked glycoproteins, core- and antennary-fucosylation, have been considered indicative in various cancers as biomarkers [1]. It is often difficult to distinguish the two structures. We thus sought to develop a method to distinguish core- and antennary-fucosylation at each glycosite of the target protein. A workflow of this study is shown in Fig. 1. Briefly, standard serum protein alph-1-antitrypsin (A1AT) was digested into peptides which were then treated by sialidase/galactosidase double digestion for glycan truncation. The truncated glycopeptides were semi-enriched and desalted by a 3K membrane and analyzed by direct LC-MS/MS. The core- and antennary-fucosylation of A1AT was thus successfully distinguished and quantified.

The overall fucosylation level of serum proteins is quite low [26]. With the routine mass spectrometry settings for peptide analysis, in-source collision-induced dissociation (also called nozzle-skimmer dissociation, abbreviated as in-source dissociation herein) of glycopeptides would occur. This is a process where an ion dissociates as a result of collisional excitation during ion transfer from an atmospheric pressure ion source to the vacuum chamber of the mass spectrometer [27]. In our experiment in-source dissociation of

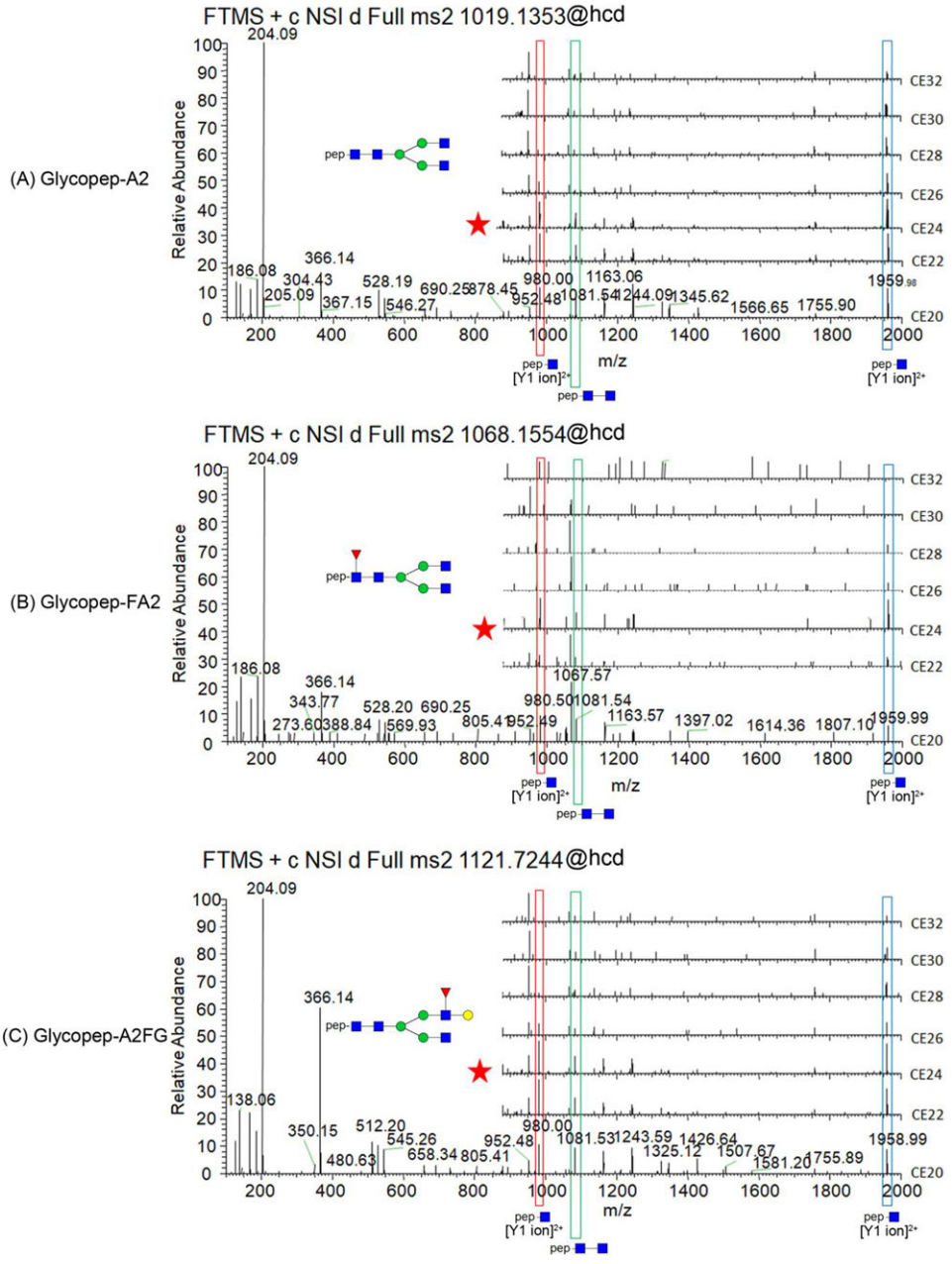

Figure 4. MS2 fragmentation patterns of glycopeptides (Asn 271) with A2, FA2 and A2FG glycan modification types under various low energy HCD collision energies, indicating that low energy HCD with CE24 provides the strongest Y1 ion (peptide+GlcNAc).

glycopeptides was found to severely affect the identification and semi-quantification of the low abundance fucosylated peptides. However, there is so far no detailed analysis of this problem. Taking the most abundant glycan modification type A2 on one glycopeptide of A1AT for example, A2 should have three of Hex (mannose), four of HexNAc (GlcNAc), zero of NeuAc, zero of NeuGlc and zero of dHex after sialidase/galactosidase double digestion, abbreviated as 34 000 herein. As shown in Fig. 2, more than 20% of 34 000 glycans were decayed into 23 000 (calculated as XIC_{23000}/XIC_{34000}) with the "universal" method settings for peptide analysis on Thermo Scientific Orbitrap Fusion mass spectrometers developed by Thermo Scientific (ion transfer tube temperature = 300°C, RF = 30%) [28] or harsher settings. A series of MS settings for glycopeptides analysis were thus optimized. We found that with lower temperature and lower RF (150°C, 20%), the in-source dissociation of glycopeptides was reduced to less than 3%, whereas the signal of core-fucosylated peptide with 34001 glycan did not reduce significantly. We also found that higher spray voltage (spray voltage > 2300 V) provided better signal but also increased in-source

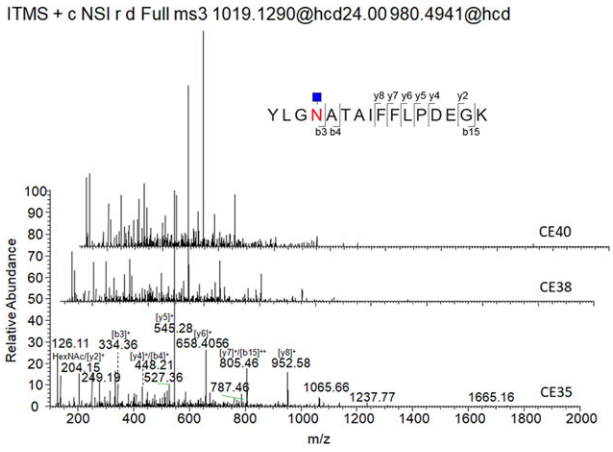

Figure 5. MS3 fragmentation patterns of Y1 ion (peptide+GlcNAc) of glycopeptides (Asn271) under various high energy HCD collision energies, indicating that high energy HCD with CE35 provides the best fragmentation profile.

dissociation. Therefore the lowest spray voltage 2300 V for a stable spray was used. This optimized setting was used further to identify site-specific glycosylation and to semi-quantify core- and antennary-fucosylation of A1AT glycopeptides.

The direct LC-MS/MS strategy is shown in Fig. 3, where first glycopeptides were selected by the detection of oxonium ion 138.05 with low energy HCD MS2 (Fig. 3A); then the Y1 ion from the glycopeptide fragment was subjected to high energy HCD MS3 (Fig. 3B) for peptide sequencing; while the selected glycopeptide was subjected to CID MS2 (Fig. 3C) for glycan structure analysis; and the entire procedure is summarized in Fig. 3D. The HCD collision energy (CE) of the first step was found to be crucial for fragmentation of glycopeptides. It was optimized for improved detection of oxonium ions and improved selection of the Y1 ion. As shown in Fig. 4, either non-fucosylated or core-fucosylated or antennary-fucosylated biantennary glycopeptides with Asn271, low energy HCD MS2 with CE24 (among series of HCD from CE20 to CE32) provided the strongest Y1 ion fragment. This optimal CE seems to be irrelevant with glycan structures or peptide sequences. As shown for glycopeptides (with Asn271), either non-fucosylated or core-fucosylated or antennary-fucosylated structures have the same optimal CE (Figure 4), where the other two glycopeptides (with Asn107 or Asn70) of A1AT also have the same optimal CE (Supporting Information Fig. 1).

In contrast, the HCD CE for peptide annotation of the second step and the CID CE for glycan annotation of the third step were found not to be that sensitive. HCD MS3 with either CE35, 38 or 40 showed similar fragmentation patterns (shown in Fig. 5), where HCD MS3 CE35 resulted in a somewhat stronger signal for higher m/z fragments. Both CID CE30 and CE35 provided similar fragmentation of the glycan structures of glycopeptides (Fig. 6). Thus in the following experiment, low energy HCD CE24, CID CE30 and high energy HCD CE 35 were used respectively.

Sialidase releases α2-3,6,8,9 N-acetylneuraminic acid leaving galactose as the terminal of the N-glycan. Subsequently, β-galactosidase cleaves β1–3,4 galactose on condition that no fucose is bound to the subterminal N-acetyl glucosamine in an N-glycan, thus providing a means to distinguish core-fucosylation and antennary fucosylation [19]. The core-fucosylated and the antennary-fucosylated glycopeptides have the same *m/z* in a sialidase digested sample (Fig. 7B), thus the spectrum of the sialidase digested sample is a mixture of core- and antennary-fucosylated peptides (Fig. 7B1). By contrast, core- and antennary-fucosylated peptides in the sialidase/galactosidase double digested sample have different *m/z* (Fig. 7A). Therefore with sialidase/galactosidase double digestion on glycopeptides the two types of fucosylation are distinguished without further MS/MS analysis or extensive sequential exoglycosidase digestion, similar to previous work at the glycan level [19]. Also, with sialidase/galactosidase digestion, the retention time of an antennary-fucosylated glycopeptide is earlier than its corresponding core-fucosylated glycopeptides (Fig. 7A), indicating that the addition of a galactose enhanced its hydrophilicity. We found that not only the elution time but also the fragmentation patterns of core- and antennary-fucosylated glycopeptides were different. In the CID MS/MS spectrum of core-fucosylated glycopeptides, several core-fucosylated glycopeptide fragments were observed in a cluster (Fig. 7A1); whereas in the CID MS/MS spectrum of an antennary-fucosylated glycopeptide, pep-43001, pep-43000 and pep-33001 always appear as the three strongest fragments (Fig. 7A2). This difference of fragmentation patterns of core- and antennary-fucosylation was found in other glycopeptides as well (Supporting Information Fig. 2). Therefore, using sialidase/galactosidase double digestion, a solid differentiation can be made between core- and antennary-fucosylated peptides using direct LC-MS/MS analysis. Traditional exoglycosidase with fucosidase α1-2,3,4,6 and fucosidase α1-3,4 were further applied on the sialidase/galactosidase double digested glycopeptides, showing the efficacy of this strategy (Supporting Information Fig. 3).

In our previous study, we analyzed the glycans cleaved from the glycoprotein A1AT and found that A1AT has 12 different glycan structures after sialidase/galactosidase double digestion [19]. In this study, we identified ten of these structures on specific glycosylation sites. Their retention times on a C18 column and their relative intensities are shown in Table 1. Low energy HCD MS2, high energy MS3 and CID MS2 spectra of all identified glycopeptides are shown in Supporting Information Fig. 2. Glycopeptide1 ADTHDEILEGLNFnLTEIPEAQIHEGFQELLR with Asn107 (modified amino acid is shown in lower case) has the most various glycan modification types including A2, FA2, A2FG, A3, FA3, A3FG, A4, FA4, A4FG, A3F2G2, whereas the other two sites have fewer glycan modification types (glycopeptide2 YLGnATAIFFLPDEGK with Asn271: A2, FA2, A2FG, A3, A3FG and glycopeptide3 QLAHQSnSTNIFFSPVSIATAFAMLSLGTK with Asn70: A2). The extent to which each site is glycosylated

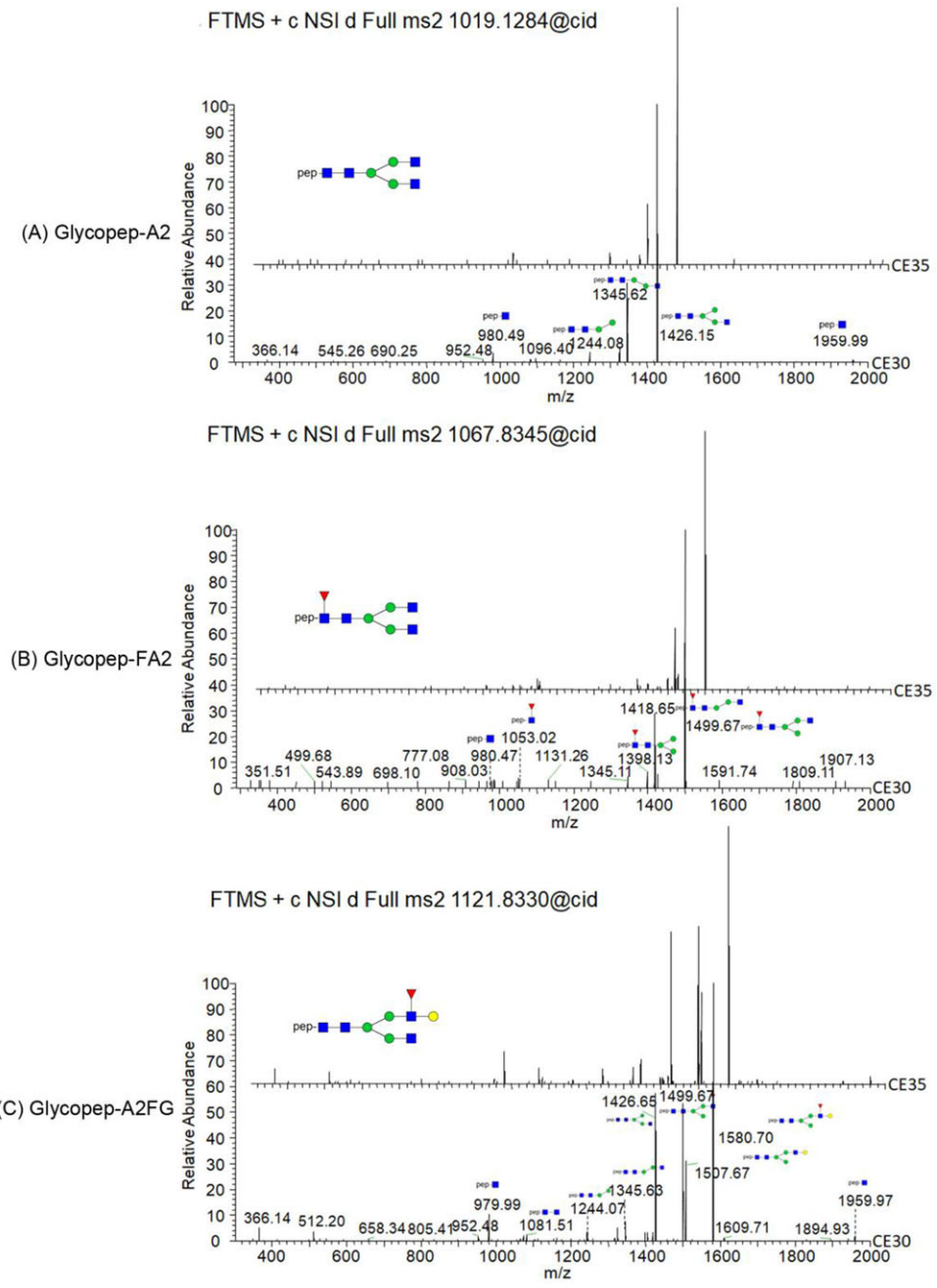

Figure 6. MS2 fragmentation patterns of glycopeptides (Asn271) with A2, FA2 and A2FG glycan modification types under various CID collision energies, indicating that either CE30 or CE35 provides similar fragmentation profile of these glycopeptides.

may possibly depend on the protein structure or proximity of the site to certain amino acids or to the N/C terminus [22]. As shown in Fig. 8 (crystal structure from [29]), all three sites are located at the protein surface and in loops, among which Asn107 is almost in the center of a big loop and may be more accessible by various glycosyltransferases while Asn271 and Asn70 are closer to the alpha helix or beta sheet structures and have smaller spaces. This may partially explain why Asn107 has the most various glycosylation modification.

As expected, after sialidase/galactosidase double digestion, the extra fucose of core-fucosylated glycopeptides made the glycopeptide more hydrophilic, thus its elution from the C18 column was earlier than its corresponding non-fucosylated glycopeptides for bi- and tri-antennary glycan modifications. An extra galactose made the antennary-fucosylated bi-antennary glycopeptide even more hydrophilic compared to its corresponding core-fucosylated case. However, a further galactose and/or fucose did not make the tri- or tetra- antennary glycopeptides significantly

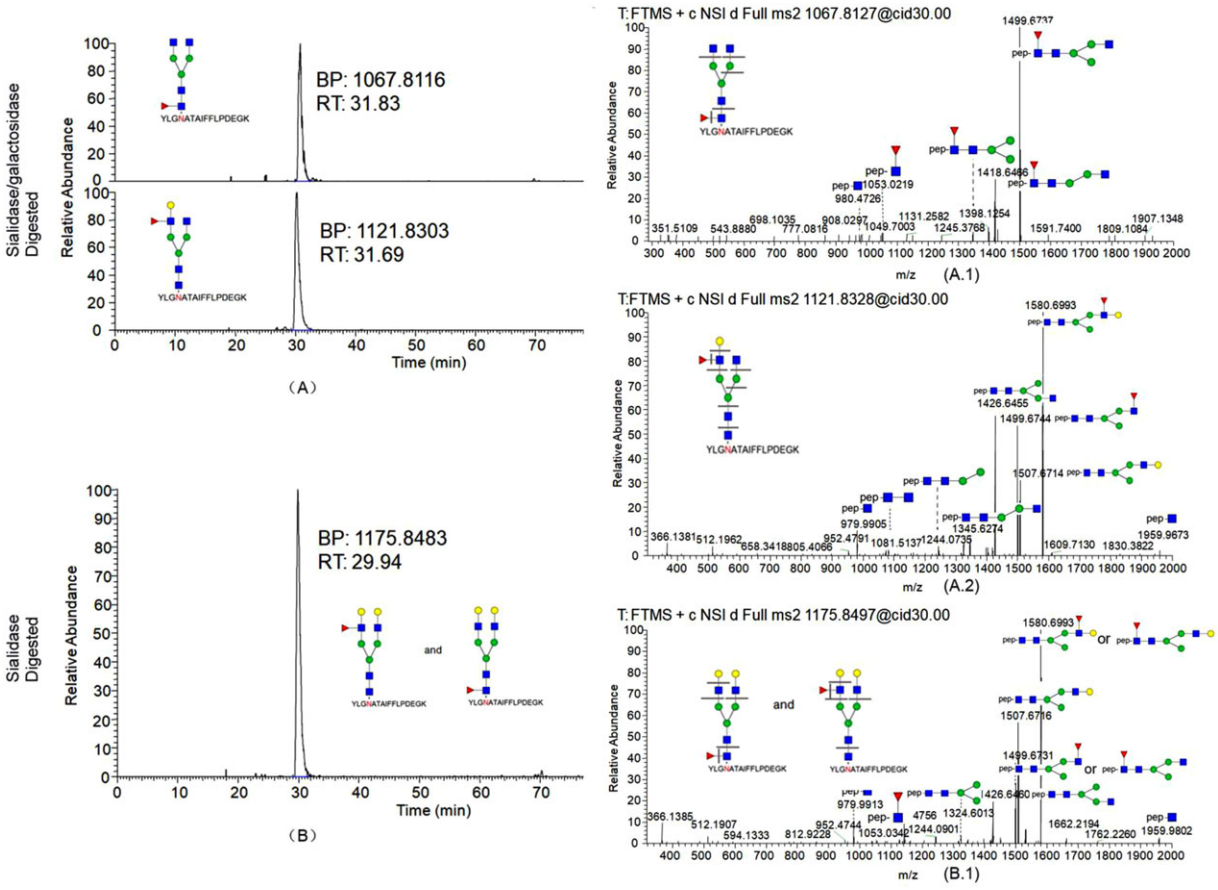

Figure 7. Differentiation of core- and antennary-fucosylation of glycopeptides (Asn271) by sialidase/galactosidase digestion (A): retention time; A.1: spectrum of FA2; A.2: spectrum of A2FG. As a comparison, no differentiation was observed for sialidase-digested case (B): retention time; B.1 spectrum of A2G2(F).

more hydrophilic and the elution times of all fucosylated tri-antennary glycopeptides or all tetra-antennary glycopeptides were similar on a very slow elution gradient (0.2% ACN/min). It can be concluded that after sialidase/galactosidase double digestion, the hydrophobicity of glycopeptides is mainly determined by its peptide backbone and only a slight hydrophobicity change was found in bi-antennary glycans. This change may be due to the fact that bi-antennary glycans have fewer sugar units and one or two extra sugars may contribute overall more hydrophilicity to the glycopeptides compared to tri- and tetra- antennary glycans.

Our previous study of glycans showed that the nonfucosylated bi- and tri-antennary glycans are the top two most abundant glycan structures of A1AT, comprising 35.8 and 25.2%, respectively [19]. In this study of glycopeptide1 and glycopeptide2, we found the bi- and tri-antennary glycan modification on Asn107 comprised 48 and 34%, respectively whereas those on Asn271 comprised 97 and 1%, respectively (Table 1). The analysis of glycopeptide3 showed that there was only A2 glycan modification on Asn70; thus we consider the glycan modification on Asn70 does not contribute much to the overall glycosylation of A1AT protein. A chi-square test is used for comparison of the relative peak intensity of major glycan modification types between glycan types cleaved from A1AT protein (data from previous result [19]) and glycan modification types of glycopeptide1 (with Asn107) or glycan modification types of glycopeptide2 (with Asn271) (Table 2). The glycan types cleaved from A1AT protein are significantly different from the glycan modification types on Asn271 (p value < 0.01), but not different from those on Asn107 (p value $= 0.97$), indicating that the glycosylation level of Asn107 overwhelms the other two sites and contributes more to the glycan structure of the A1AT protein. The core- and antennary-fucosylation level of the two sites also varied significantly, where the antennary-fucosylated tri-antennary glycan was the most abundant fucosylation type on Asn107 but it was negligible on Asn271.

One study with classical lectin blot assay found that the up-regulation of core-fucosylated but not antennary-fucosylated A1AT could be indicative for hepatocellular cancer diagnosis [20] while antennary-fucosylation of A1AT indicates inflammation especially in HBV-infected patients [22]. Our previous glycan study has indicated that bi-antennary

Table 1. Summary of glycopeptides of A1AT with truncated glycan and % relative peak area of core- and antennary-fucosylation of each site

Glycopeptide	Experimental m/z	Charge (z)	Theoretical m/z	delta ppm	Retention time (min)	Percent relative peak area
GLYCOPEP1 with Asn107	*ADTHDEILEGLNFnLTEIPEAQIHEGFQELLR*					
A2	1248.8282	4+	1248.8334	4.2	49.59	48.4
FA2	1285.3411	4+	1285.3479	5.3	49.51	2.1
A2FG	1325.8546	4+	1325.8611	4.9	49.37	1.1
A3	1299.5972	4+	1299.6033	4.7	49.51	33.8
FA3	1069.0868	5+	1069.0957	8.4	49.37	0.3
A3FG	1376.6236	4+	1376.6309	5.3	49.37	11.3
A3F2G2	1453.6519	4+	1453.6586	4.6	49.37	0.3
A4	1350.3662	4+	1350.3731	5.1	49.37	2.3
FA4	1386.8769	4+	1386.8876	7.7	49.37	0.2
A4FG	1142.1147	5+	1142.1222	6.5	49.37	0.2
GLYCOPEP2 with Asn271	*YLGnATAIFFLPDEGK*					
A2	1019.4596	3+	1019.4657	6.0	32.21	97.2
FA2	1068.1455	3+	1068.1517	5.8	31.83	0.2
A2FG	1122.1628	3+	1122.1693	5.8	31.69	1.4
A3	1087.1553	3+	1087.1589	3.3	31.86	1.0
A3FG	1189.8561	3+	1189.8624	5.3	31.36	0.2
GLYCOPEP3 with Asn70	*QLAHQSnSTNIFFSPVSIATAFAMLSLGTK*					
A2	1125.2794	4+	1125.2850	4.9	53.6	null

Table 2. A chi-square test is used for comparison of the relative peak intensity of major glycan modification types between glycan types cleaved from A1AT protein (data from previous result [19]) and glycan modification types on glycopeptide1 (with Asn107) or glycopeptide2 (with Asn271)

	Percent relative peak area					Chi-square test *p* value[a)]
	A2	FA2	A2F	A3	A3FG	
GLYCAN	35.8	1.8	0.5	25.2	11.3	–
GLYCOPEP1 with Asn107	48.4	2.1	1.1	33.8	11.3	0.97
GLYCOPEP2 with Asn271	97.2	0.2	1.4	1	0.2	<0.01

a) p value < 0.05 is considered significant.

core-fucosylation was the most abundant core-fucosylation type of A1AT protein. From Table 1 we may deduce that if there were alteration in A1AT core-fucosylation types in patients, the bi-antennary core-fucosylation on Asn107 is most likely the possible target that can be precisely monitored and quantified by mass spectrometry. Another classical lectin blot-based study showed that the overall increase of A1AT fucosylation level was able to distinguish lung adenocarcinoma from benign diseases or other lung cancer subtypes [21]. The strategy developed in this study would enable the identification and quantification of core- and antennary-fucosylation on specific sites of A1AT. In future work, this methodology will be used to study changes in serum A1AT glycosylation during the progression of various cancers. The more precise fucosylation analysis with site-specific information should provide improved diagnostic value. Also, this strategy can be applied to the study of other key glycoproteins during the progression of various diseases.

4 Concluding remarks

We have developed a pipeline to study the glycosylation of A1AT to identify the presence of core- versus antennary-fucosylation without separating glycans and peptides from glycopeptides. This was performed on a standard protein A1AT which was digested by trypsin followed by sialidase/galactosidase digestion. Galactosidase removes terminal galactose residues in an N-glycan except when the subterminal N-acetylglucosamine is modified by fucosylation, thus providing a means to distinguish core-fucosylation and antennary-fucosylation. The sites and structures of glycans could be determined simultaneously by this procedure. In total, we identified 1 glycan structure (A2) on Asn70 of glycopeptide QLAHQSnSTNIFFSPVSIATA, 10 glycan structures (A2, FA2, A2FG, A3, FA3, A3FG, A4, FA4, A4FG and A3F2G2) on Asn107 of glycopeptide ADTHDEILEGLNFnLTEIPEAQIHEGFQELLR, and 5 glycan structures (A2, FA2, A2FG,

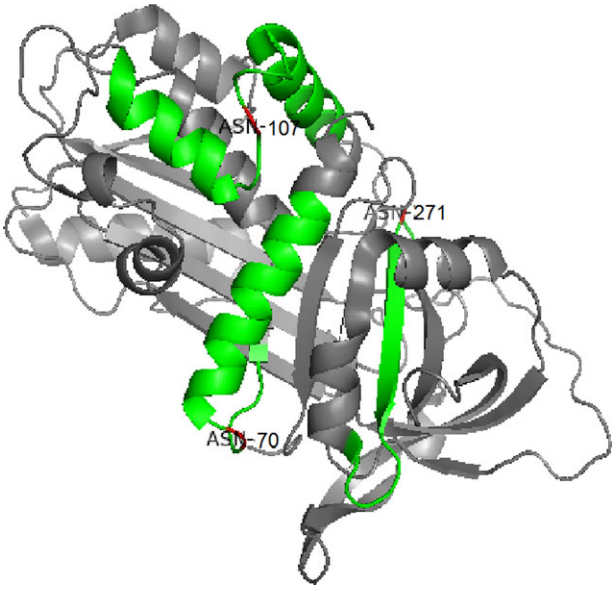

Figure 8. The three glycosylated sites Asn70, Asn107 and Asn271 of A1AT are labeled red and the detected peptides by mass spectrometry are labeled green in the 3D structure.

A3 and A3FG) on Asn271 of glycopeptide YLGnATAIFFLPDEGK. We believe that this methodology will be widely used to identify and quantify core- and antennary-fucosylation on A1AT or other key glycoproteins during the progression of various diseases.

This work was funded by the National Natural Science Foundation of China (81601828). Thanks for the software support from Dr. Zeng WF from the pGlyco group in the Institute of Computing Technology, CAS, Beijing. Also we thank the support of the University Research Facility in Chemical and Environmental Analysis and the University Research Facility in Life Sciences of Hong Kong Polytechnic University. Partial support was also provided by the National Cancer Institute under grant no. R01 CA160254 (D.M.L.).

The authors have declared no conflicts of interest.

5 References

[1] Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., Packer, N. H., *Proteomics* 2014, *14*, 525–546.

[2] Pompach, P., Brnakova, Z., Sanda, M., Wu, J., Edwards, N., Goldman, R., *Mol. Cell. Proteomics* 2013, *12*, 1281–1293.

[3] Marrero, J. A., Feng, Z. D., Wang, Y. H., Nguyen, M. H., Befeler, A. S., Roberts, L. R., Reddy, K. R., Harnois, D., Llovet, J. M., Normolle, D., Dalhgren, J., Chia, D., Lok, A. S., Wagner, P. D., Srivastava, S., Schwartz, M., *Gastroenterology* 2009, *137*, 110–118.

[4] Goonetilleke, K. S., Siriwardena, A. K., *European J. Surg. Oncol.* 2007, *33*, 266–270.

[5] Guttman, A., *Electrophoresis* 1997, *18*, 1136–1141.

[6] Szigeti, M., Bondar, J., Gjerde, D., Keresztessy, Z., Szekrenyes, A., Guttman, A., *J. Chromatogr.* 2016, *1032*, 139–143.

[7] Desaire, H., *Mol. Cell. Proteomics* 2013, *12*, 893–901.

[8] Chandler, K. B., Costello, C. E., *Electrophoresis* 2016, *37*, 1407–1419.

[9] Wuhrer, M., Catalina, M. I., Deelder, A. M., Hokke, C. H., *J. Chromatogr.* 2007, *849*, 115–128.

[10] Ma, C., Zhang, Q., Qu, J., Zhao, X., Li, X., Liu, Y., Wang, P. G., *J Proteomics* 2015, *114*, 61–70.

[11] Yu, Q., Wang, B., Chen, Z., Urabe, G., Glover, M. S., Shi, X., Guo, L. W., Kent, K. C., Li, L., *J. Am. Soc. Mass Spectrom.* 2017.

[12] Bern, M., Kil, Y. J., Becker, C., *Current Protocols in Bioinformatics*, *Chapter 13*, Unit13 20, 2012.

[13] Sun, S., Shah, P., Eshghi, S. T., Yang, W., Trikannad, N., Yang, S., Chen, L., Aiyetan, P., Hoti, N., Zhang, Z., Chan, D. W., Zhang, H., *Nat. Biotechnol.* 2016, *34*, 84–88.

[14] Liu, M. Q., Zeng, W. F., Fang, P., Cao, W. Q., Liu, C., Yan, G. Q., Zhang, Y., Peng, C., Wu, J. Q., Zhang, X. J., Tu, H. J., Chi, H., Sun, R. X., Cao, Y., Dong, M. Q., Jiang, B. Y., Huang, J. M., Shen, H. L., Wong, C. C. L., He, S. M., Yang, P. Y., *Nat. Communications* 2017, *8*, 438.

[15] Zeng, W. F., Liu, M. Q., Zhang, Y., Wu, J. Q., Fang, P., Peng, C., Nie, A., Yan, G., Cao, W., Liu, C., Chi, H., Sun, R. X., Wong, C. C., He, S. M., Yang, P., *Scientific Reports* 2016, *6*, 25102.

[16] Pompach, P., Ashline, D. J., Brnakova, Z., Benicky, J., Sanda, M., Goldman, R., *J Proteome Res* 2014, *13*, 5561–5569.

[17] Wuhrer, M., Koeleman, C. A., Hokke, C. H., Deelder, A. M., *Rapid Communications Mass Spectrometr.* 2006, *20*, 1747–1754.

[18] Veillon, L., Huang, Y., Peng, W., Dong, X., Cho, B. G., Mechref, Y., *Electrophoresis* 2017, *38*, 2100–2114.

[19] Yin, H., Zhu, J., Wu, J., Tan, Z., An, M., Zhou, S., Mechref, Y., Lubman, D. M., *Electrophoresis* 2016, *37*, 2624–2632.

[20] Comunale, M. A., Rodemich-Betesh, L., Hafner, J., Wang, M., Norton, P., Di Bisceglie, A. M., Block, T., Mehta, A., *PloS One* 2010, *5*, e12419.

[21] Liang, Y., Ma, T., Thakur, A., Yu, H., Gao, L., Shi, P., Li, X., Ren, H., Jia, L., Zhang, S., Li, Z., Chen, M., *Glycobiology* 2015, *25*, 331–340.

[22] McCarthy, C., Saldova, R., Wormald, M. R., Rudd, P. M., McElvaney, N. G., Reeves, E. P., *J Proteome Res* 2014, *13*, 3131–3143.

[23] Yin, H., Tan, Z., Wu, J., Zhu, J., Shedden, K. A., Marrero, J., Lubman, D. M., *J Proteome Res* 2015, *14*, 4876–4884.

[24] Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., Haslam, S. M., *J Proteome Res* 2008, *7*, 1650–1659.

[25] Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E.,

Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, New York 2008.

[26] Bones, J., Mittermayr, S., O'Donoghue, N., Guttman, A., Rudd, P. M., *Anal Chem* 2010, *82*, 10208–10215.

[27] Murray, K., Boyd, R., Eberlin, M., Langley, G., Li, L. and Naito, Y., *Pure Applied Chem.* 2013, *85*, 1515–1609.

[28] Eliuk, S., Soltero, N., Remes, P., Senko, M., Zabrouskov, V., 13 Human Proteome Organization World Congress, Madrid, 2014, pp. 791–P.

[29] Patschull, A. O., Segu, L., Nyon, M. P., Lomas, D. A., Nobeli, I., Barrett, T. E., Gooptu, B., *Acta Crystallographica Section F, Structural Biology Crystallization Communications* 2011, *67*, 1492–1497.

