A seamless workflow for comprehensive analysis of the mucin-type O-linked glycoproteome

Kai Cheng[1]; Jun Zhu[1]; Jin Wenhai[2]; Fangjun Wang[1]; Mingming Dong[1]; Mingliang Ye[1]; Christie Hunter[3]; Hanfa Zou[1]

[1]Dalian Institute of Physical Chemistry, the CAS, Dalian, China; [2]AB SCIEX China, Shanghai, China; [3]AB SCIEX, Redwood Shore, CA, the United States

ABSTRACT

Mucin type O-glycosylation has drawn increasing attention recently as it has close relationship with diseases and cancers. However, previous studies only identified O-glycosylation sites, intact O-glycans especially those sialylated forms were difficult to be identified. Considering the important biological roles of O-glycans, it is necessary to profile its microheterogeneity. In this work, a novel strategy with enzymatic reaction assisted HILIC enrichment and Q-TOF-MS identification was established to analyze intact mucin-type O-glycopeptides. In short, O-glycopeptides from human serum were selectively captured and identified for not only the attachment sites but also the glycan structure. As a result, a total of 247 O-glycosylation sites corresponding to 89 different proteins were identified from only 120 μL human serum. Additionally, seven different O-glycoforms were characterized, including three types of sialylated core 1 O- and two types of core 2 O-glycoforms. This strategy could be further used to study the sialylation of O-glycoproteome in diseases and cancers.

.

INTRODUCTION

Protein glycosylation is a very heterogeneous group of post-translational modifications, but the analysis of protein O-glycosylation lags far behind that of N-glycosylation due to the lack of a universal endoglycosidase for removing of O-glycans, and the much lower abundance. Recently, three excellent strategies have been reported to identification of O-glycosylation sites from proteome samples. Larson et al. introduced a hydrazide chemistry based method where sialylated glycoproteins were selectively periodate-oxidized, captured on hydrazide beads, trypsinized and released by acid hydrolysis of sialic acid glycosidic bonds. The released glycopeptides were then analyzed by a hybrid linear quadrupole ion trap/FTICR mass spectrometer.(1). Zsuzsamma Darula *et al.* introduced a strategy by using jaclin affinity chromatography, higher-energy collision-dissociation (HCD)/electron transfer dissociation (ETD) and two fractionation methods (2). Catharina Steentoft *et al.* introduced zinc-finger nuclease to truncate the O-glycan elongation pathway in human cells which simplified all O-glycosylation to GalNAc (Tn) and NeuAc-GalNAc (STn) (3). After neuraminidase treatment and lectin chromatography enrichment, O-glycopeptides were identified using LC-MS/MS with ETD. Though many O-glycosylation sites were identified using above techniques, the attached glycans was not inadequately characterized mainly because the integrity of the glycan chains were broken before MS analysis. In the hydrazide chemistry method, the terminal sialic acids on the glycan chains were damaged and thus the O-linked glycoforms with different numbers and linkages (α2-3 or α2-6) of sialic acids could not be identified. In jaclin affinity chromatography method, to facilitate the identification of peptide sequences, i.e. to overcome the charge-density limits for successful ETD measurement, glycopeptides were treated with neuraminidase and -galactosidase, and the resulting glycopeptides retaining only the core GalNAc units were subjected to MS/MS analysis. Thus the inact glycan structure can not be obtained either. Up to now there is no single analytical approach that can readily identify both the glycosylation sites and the modifying sugar structures for a proteome sample. Even worse, there is no general enrichment approach for all types of O-glycopeptides.

In this work, a seamless workflow involving O-glycopeptide enrichment, MS analysis and interpretation was presented for the comprehensive analysis of O-glycosylation. An enzyme assisted HILIC was developed to selective isolation of O-glycopeptides, a high resolution triple TOF MS was applied to fragmentation of inact O-glycopeptides and a classification strategy was developed to decipher the acquired spectra. This workflow enables the identification of both the glycosylation sites and the modifying sugar structures in a high throughput way resulted in 360 O-glycosylation states including 247 O-glycosylation sites in 89 proteins from human serum. It was found that 70% of the identified sites contain only one glycosylation state and the rest 30% contain two and more glycosylation states.

MATERIALS AND METHODS

Modified filter-aided sample preparation (FASP) and centrifugation assisted click maltose-HILIC for O-glycopeptides enrichment. FASP procedure was performed according to Mann *et al.* The centrifugation assisted click maltose-HILIC enrichment was performed according to our previous report (17). The protein samples prepared by the modified FASP workflow were first redissolved in the enrichment loading buffer (1% TFA/80% ACN). Then, 20 μL solution equivalent to 20 μg standard glycoprotein digest or 20 μL human serum digest was pipetted into a HILIC tip. After centrifugation at 4000g for about 10 min, the HILIC tip was washed with 20 μL of loading buffer and 10 μL of 80% ACN. Finally, the enriched O-glycopeptides were eluted with 20 μL of 0.1% FA/H_2O and analysed by LC-MS/MS directly.

LC-MS/MS analysis. All the LC-MS/MS analyses were performed on TripleTOF® 5600+ (AB SCIEX, Redwood Shore, CA) with a Nano ACQUITY UPLC system (Waters, Milford, MA). The LC-MS/MS system includes a 3-cm C_{18} capillary trap column (200 μm i.d.)and a 12-cm C_{18} capillary analysis column (75 μm i.d.) with flow rate at 0.35 μL/min. The RP gradient was optimized to 90 min. A spray voltage of 2.3 kV was applied and the instrument was operated in IDA mode, with the top 40 precursors in each TOF MS scan (800 ms, 350-1250 m/z) subjected to MS/MS analysis (accumulation time 100 ms, 100-1400 m/z).

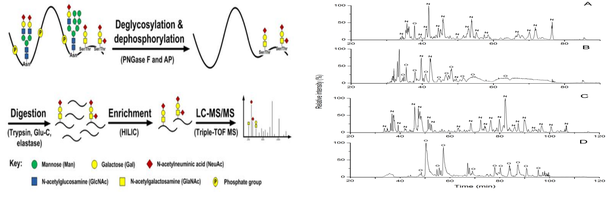

Fig1. Left, total flowchart for intact O-linked glycopeptide identification; Right, HIILIC enrichment of bovine fetuin (A,B) and human serum (C,D) before (A and C) or after (B and D) removing N-glycosylation. (Symbols: N means N-glycopeptide, O means O-glycopeptide.)

Database searching. All the WIFF files were converted to MGF files using PeakView® software v. 1.1.1 (AB SCIEX, Redwood Shore, CA). The MS/MS spectra containing peaks of 204.0873 Da or 274.0880 Da (mass accuracy: 0.01 Da) were extracted from the MGF files using a in-house software and the peak lists of glycan fragment ions (186.0766 Da, 204.0873 Da, 274.0880 Da, 292.1033 Da, 366.1401 Da, and 657.2355 Da) as well as peptide+glycan peaks (generated by neutral loss of different monosaccharide units from O-glycopeptide precursor) were all removed to generate a new MGF file. The new MGF files were searched with Mascot with variable modifications of HexNAc (203.0794 Da), HexNAc (365.1322 Da), HexHexNAc NeuAc (656.2276 Da) and HexHexNAcNeuAc$_2$ (947.3230 Da) of serine and threonine;

RESULTS AND DISCUSSION

Enzymatic reaction assisted HILIC for specific enrichment of O-linked glycopeptides. To evaluate the performance of the enrichment, two protein standards, kappa-casein and fetuin were used as the samples and the results are shown in Fig.1, Right Pane, indicting the enzymatic reaction assisted HILIC enrichment was suitable for O-linked glycopeptides.

TOF-CID for fragmentation of intact O-linked glycopeptides. we compared CID, HCD, and ETD in LTQ-Orbitrap MS with CID in TripleTOF 5600 MS (TOF-CID) (Fig. 2). It was found that CID and HCD in LTQ-Orbitrap only break glycosidic bonds between monosaccharide units of the attached glycans, so little information on peptide backbone could be obtained by the two fragmentation methods (Fig. 2A and Fig. 2B). ETD could provide some information on peptide sequence and O-glycosylation site, however, the low fragmentation efficiency and mass accuracy make it difficult to interpret the ETD spectrum. Moreover, when sialic acid is not removed, the fragmentation will be even worse due to the low charge-density of sialylated peptides and the extensive neutral loss of sialic acid. Therefore, the Mascot score after database searching was only 19 (Fig. 2C) In contrast, TOF-CID is a versatile fragmentation method for O-linked glycopeptides which can simultaneously provide sufficient information for peptide sequence and the structure for attached O-glycans

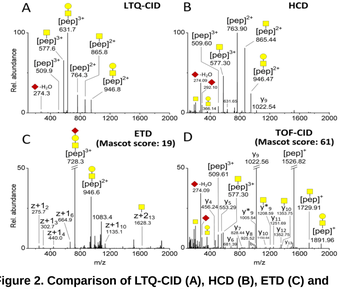

Figure 2. Comparison of LTQ-CID (A), HCD (B), ETD (C) and TOF-CID (D) for identification of O-linked glycopeptides. The four spectra were all generated from the same O-glycopeptide, SSTTKPPFKPHGS(GalNAcGalNeuAc)R (M.W. 2182.0277 Da).

By using TOF-CID, a continuous y_4 to y_{13} ions series for the peptide backbone was clearly obtained (Fig. 2D). The resulting Mascot score was 61 which is much better than the ETD spectrum. The carbohydrate oxonium ions at m/z 204.09, 274.09, 292.10, and 366.14 were used as reporting ions for recognition of the MS/MS spectra generated from glycopeptide precursors. The peptide+glycan ions, i.e. [peptide+H]$^+$, [peptide+HexNAc+H]$^+$, and [peptide+HexNAcHex+H]$^+$ could be used for the elucidation of glycan composition. The B+glycan or Y+glycan ions, such as [y^*_9+HexNAc+H]$^+$ and [y_{10}+HexNAc+H]$^+$, could be used for O-glycosylation site identification. By using TOF-CID, peptide sequence, O-linked glycosylation site and the attached O-glycan were all successfully profiled, therefore TOF-CID is more suitable for the identification of intact O-glycopeptides compared with the other three fragmentation methods.

Fig.3 Amino acid distribution around the O-glycosylation sites determined.

Characterization for human serum O-glycoproteome. The developed workflow was applied to comprehensive analysis of O-glycoproteome of human serum. To improve the coverage of O-glycoproteome, three digestion strategies by using different enzymes were used. All together, we identified 247 O-glycosylation sites in 89 different glycoproteins from only 120 μL human serum; 177 (72%) of these sites and 48 of the 89 proteins are novel. Moreover, by using TOF-CID, 154 (62%) of these sites are identified unambiguously by manual validation, which is comparative with Adnan Halim *et al.* reported by using CID and ECD (40 in 57 O-sites were exactly identified in their work) (4). These results demonstrated that TOF-CID could be used as a alternative technique for ETD in O-glycosylation site identification. Although no universal consensus motif were exists for O-linked glycosylation, the frequency of amino acids around the glycosylation sites was said to have some regular patterns, as shown in Fig. 3; It is interesting that Pro residue was the most frequent amino acid in the sequences, especially in -1 and +3 positions of the glycosylation sites. Statistically, 34% and 35% of the total sequences contain Pro residues at the -1 and +3 positions.

Determination of the O-glycan microheterogeneity. By using our strategy, seven different O-glycoforms attached to Ser/Thr were successfully discovered in human serum, including five sialylated O-glycoforms and core 2-like glycoforms with five or six monosaccharide units (Hex$_2$HexNAc$_2$NeuAc, Hex$_4$HexNAc$_2$NeuAc$_2$), which are all very difficult to be identified in ETD and MS^n strategies (2,4, 5). As different glycoforms were always found at the same O-glycosylation site, we termed one glycoform in one specific site as a "glycosylation state", therefore, about 360 glycosylation states were identified in this work. According to the statistical result, 70% of the identified sites contain only one glycosylation state and the rest 30% contain two and more glycosylation states (Fig. 4A). The distribution of glycoforms was also calculated (Fig. 4B). The sialylated glycoforms (HexHexNAcNeuAc and HexHexNAcNeuAc$_2$) contribute to more than 90% of all glycoforms. These results demonstrate that our strategy is well suited for O-glycan heterogeneity analysis.

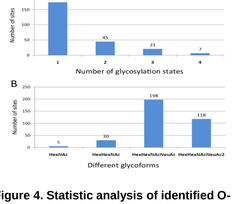

Figure 4. Statistic analysis of identified O-glycosylation states (A) and O-glycoforms (B) in human serum.

In conventional ETD and MS^n strategies, the terminal sialic acids were inevitably damaged or removed before LC-MS/MS analysis, thus no information about the attached sialic acids was obtained. Fortunately, in our strategy, intact mucin type O-glycopeptides including those sialylated forms could be directly analyzed. Therefore, the microheterogeneity of O-glycans could be well studied. For example, two O-glycosylation sites of apolipoprotein E (APOE) was identified, involving eight glycosylation states (Fig. 5). And five different O-glycoforms were successfully identified with Hex$_2$HexNAc$_2$NeuAc$_2$ (core 2 structure) being the most complex glycan. Similar results were also reported by Youra Lee *et al.* (6). Another example from the standard protein (fetuin) was also provided (Fig. 6). In this case, we identified six glycosylation states with five different glycoforms, including four sialylated glycoforms. May mention the advantage of this enrichment method.

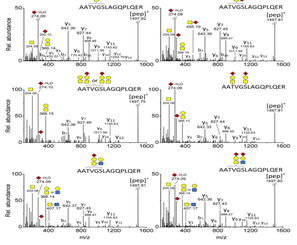

Figure 5. Diversity of O-glycoforms identified from Apolipoprotein E in human serum.

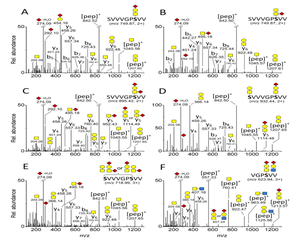

Figure 6. Diversity of O-glycoforms identified from bovine alpha-2-HS-glycoprotein.

CONCLUSIONS

In summary, we developed a novel strategy by the combination of modified FASP, HILIC enrichment and TOF-CID identification for mucin type O-glycopeptides analysis. By using this strategy, intact O-glycopeptides attached with different glycoforms from human serum were well enriched and identified. Unlike ETD and MS^n strategies, the O-glycopeptides together with their attachment sites and attached O-glycans were successfully identified from the same TOF-MS/MS spectrum. What's more, the microheterogeneity of the attached O-glycans were well studied and many sialylated O-glycans including some core 2 glycans were profiled

REFERENCES

1Halim A, *et al.* (2011) Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein / amyloid β-peptides in human cerebrospinal fluid. *Proc. Natl. Acad. Sci. U.S.A.* 108(29):11848-11853.

2Darula Z, Sherman J, Medzihradszky KF, Francisco S, & Medzihradszky KF (2012) How to dig deeper ? Improved enrichment methods for mucin core-1 type. *Mol Cell Proteomics*:doi: 10.1074/mcp.O1111.016774

3Steentoft C, *et al.* (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8(11):977-9825

4Halim A. NJ, Rüetschi U., Hesse C. and Larson G. (2012) Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. Mol Cell Proteomics 11(4):doi: 10.1074/mcp.M1111.013649

5Darula Z & Medzihradszky KF (2009) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. Mol Cell Proteomics 8(11):2515-2526.

6Lee Y, *et al.* (2010) Glycosylation and sialylation of macrophage-derived human apolipoprotein E analyzed by SDS-PAGE and mass spectrometry: evidence for a novel site of glycosylation on Ser290. *Mol Cell Proteomics* 9:1968-1981.

TRADEMARKS/LICENSING