Highly Efficient Identification of O-GalNAc Glycosylation by an Acid-Assisted Glycoform Simplification Approach

Xin You, Hongqiang Qin, Jiawei Mao, Yu Tian, Mingming Dong, Zhimou Guo, Xinmiao Liang, Liming Wang, Yan Jin, and Mingliang Ye*

Compared with N-linked glycosylation, the analysis of O-GalNAc glycosylation is extremely challenging due to the high structure diversity of glycans and lack of glycosidases to release O-GalNAc glycans. In this work, a glycoform simplification strategy by combining HILIC enrichment with chemical de-sialylation to characterize O-GalNAc glycosylation of human serum is presented. This method is first validated by using the bovine fetuin as the test sample. It is found that more than 90% of the sialic acid residues can be removed from bovine fetuin by the acid-assisted de-sialylation method, which significantly simplifies the glycan structure and improves identification sensitivity. Indeed, the number of identified peptide backbones increases nearly one fold when this strategy is used. This method is further applied to analyze the human serum sample, where 185 O-GalNAc modified peptide sequences corresponding to 94 proteins with high confidence (FDR (false detection rate) < 1%) are identified. This straightforward strategy can significantly reduce the variations of glycan structures, and is applicable to analysis of other biological samples with high complexity.

1. Introduction

As one important type of glycosylation, O-GalNAc participates in a variety of biological processes such as cell adhesion, cell signaling and immune response, etc.[1] The aberration of O-GalNAc glycosylation is closely related to many diseases, and the map of O-GalNAc glycosylation could help us to better interpret the significance of O-glycosylated proteins.[2] However, O-GalNAc glycosylation is much less explored compared to other types of glycosylation (N-linked and O-GlcNAc) because of its high complexity in glycoforms, as well as their low abundance.[3] There are up to eight core structures of O-glycosylation, and more than 20 distinct polypeptide GalNAc-transferases initiate and elongate the glycoforms, which is much more complex than that of GlcNAc-type O-glycosylation.[4,3] Additionally, much low abundance of O-GalNAc glycosylation makes the identification of O-GalNAc glycopeptides much more difficult.

In the analysis of N-glycoproteome, N-glycopeptides were typically subjected to PNGase F treatment to release glycans, which could effectively decrease the microheterogeneity of glycopeptides, as well as enhance the detection sensitivity of N-linked glycopeptides.[6] Yet, there was no such glycosidase that could release O-glycans with high efficiency. β-Elimination reaction could release O-glycans from glycoproteins/peptides, which was utilized to analyze O-glycans in glycomics analysis.[7] However, the identification of peptide sequences is difficult due to various side reactions occurring on the peptide backbones during the process of the β-elimination reaction, which limit its application in glycoproteomics analysis. Recently, a few methods have been reported on the analysis of GalNAc-type O-glycoproteome by simplifying the glycan structures. Darula et al. sequentially trimmed the enriched O-glycopeptides by using neuraminidase and galactosidase.[8] This partial deglycosylation approach leaves only the core N-acetylgalactosamine (GalNAc) residues on the peptides, which led to the identification of 124 O-glycosylation sites from 2 mL bovine serum. This method could decrease the complexity of O-glycan structures. Yet, the low efficiency of exoglycosidases, especially for some types of glycan structures, limited their applications in global analysis of O-GalNAc glycosylation. Further, the high cost of exoglycosidases is also a barrier for its applications at large scale. Instead of partial deglycosylation by using exoglycosidases, Steentoft et al. introduced zinc-finger nuclease to truncate the O-glycan elongation pathway in human cells, which could simplify all O-glycans to GalNAc and NeuAc-GalNAc.[9] By using this SimpleCell method, 1538 O-glycosites in 738 glycoproteins were identified from the cell lysates and secretomes in the Chinese hamster ovary cell line.[10] This method has great potential applications in the comprehensive analysis of glycosylation. However, it is only applicable to engineered cell lines, and it does not work for analysis of O-GalNAc glycosylation in tissues or humoral samples. In our previous work, O-GalNAc
glycopeptides were in silico deglycosylated for the interpretation of the spectra, and the peptide sequences could be identified without setting variable modifications.\(^{[11]}\) This method could be used for the analysis of tissues and humoral samples. Yet, low sensitivity and fragmentation of intact glycopeptides by MS limited its efficiency in analyzing O-GalNAc glycosylation.

Sialylation, often occurring at the terminal of glycans, is one of the main factors leading to rich diversity of O-glycan structures. If the sialic acid residues could be removed, the O-glycan structures could be significantly simplified as illustrated in Figure 1A. After the removal of sialic acid residues, the 17 glycans corresponding to core-1 and core-2 structures could be simplified into six glycan forms. The sensitivity of glycopeptide identification is expected to be enhanced: 1) the abundances of O-GalNAc glycopeptides could be increased as the simplification leads to the pooling of glycopeptides with the same core glycan structures; 2) the fragmentation efficiency of glycopeptides in MS could be improved due to the removal of sialic acid residues. Herein, we presented an acid-assisted approach to efficient desialylation of O-GalNAc glycopeptides (Figure 1B). This method exploits the feature that the glycosidic linkage of sialic acid residues is much more sensitive to acid hydrolysis than that of other monosaccharides.\(^{[12]}\) This method was first validated by using the bovine fetuin as the test sample. It was found more than 90% of the sialic acid residues could be removed from bovine fetuin by the acid-assisted de-sialylation method, which significantly simplified the glycan structure and improved identification sensitivity. This method was applied to analyze the human serum sample, where 185 O-GalNAc modified peptide sequences corresponding to 94 proteins with high confidence (FDR <1%) were identified. This straightforward strategy could significantly reduce the variations of glycan and is applicable to analysis of other biological samples with high complexity.

2. Experimental Section

Reagents and Materials: Standard proteins of bovine fetuin, trypsin and elastase were obtained from Sigma (St. Louis, MO). Glycosidases of PNGase F and neuraminidase were purchased from New England Biolabs (Ipswich, MA, USA). Chemical reagents of 1, 4-dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), iodacetamide (IAA) and trifluoroacetic acid (TFA) were obtained from Sigma (St. Louis, MO). Formic acid (FA) was obtained from Fluka (Buches, Germany). Acetonitrile and water for reverse phase liquid chromatography (RP LC) (HPLC grade) was purchased from Merck (Darmstadt, Germany). Ammonium bicarbonate and urea were obtained from BioBasic Inc. (Ontario, Canada). Pure water for protein digestions and glycopeptide enrichment was purified with a Milli-Q system (Millipore, Milford, MA). GE Loader tips (20 μL) were purchased from Eppendorf (Hamburg, Germany). The human serum used in the experiments was obtained from the Second Affiliated Hospital of Dalian Medical University (Dalian, China). All experiments were approved by the Ethics Committees of the Second Affiliated Hospital of Dalian Medical University in accordance with the ethical guidelines of the Declaration of Helsinki. The serum samples were collected from 100 volunteers, and pooled together with equal-volume. The samples were stored at -80 °C until usage. Click maltose-HILIC beads were prepared in house as reported.\(^{[13]}\)

Protein Digestion and Release of N-linked Glycans: The standard glycoprotein (bovine fetuin) and serum were denatured by dissolving in 50 mM HEPES/8 M urea buffer with 20 mM DTT. After incubation at 37 °C for 2 hours, the denatured proteins were alkylated by 40 mM IAA for 40 min in the dark at room temperature. Enzymes for digestion (elastase for bovine fetuin and
trypsin for serum) was then added at the ratio of 1:20 (enzyme: protein, w:w) into the reaction buffer after the concentration of urea was diluted to 1 M by adding 50 mM HEPES (pH = 7.4). The digestion process was last for about 20 hours. Then the tryptic peptides was desalted by Oasis HLB column. After lyophilization, the peptides were re-dissolved in GlycoBuffer 2 (pH = 7.5) and PNGase F was added to release the N-linked glycans at 37 °C overnight. The de-N-glycosylation peptides were desalted to remove N-linked glycans from samples. Finally, the digestion of proteins was collected by centrifugation and lyophilized to dryness for glycopeptide enrichment.

Enrichment and Simplification of O-GalNAc Glycopeptides: The O-GalNAc glycopeptides were enriched by using HILIC tip materials according to our previous reports.[16] The digestions of proteins were re-dissolved in the loading buffer (80% ACN/1% TFA). Then, 60 μL solution equivalent to 100 μg bovine fetuin digest or equivalent to 50 μL serum digest was pipetted into a HILIC tip with 5 mg click-maltose materials. After centrifugation, the HILIC tip was washed with 60 μL of loading buffer for one time and 20 μL of loading buffer for two times. Finally, the enriched O-glycopeptides were eluted with 100 μL of 30% ACN/1% TFA for twice. And the elution was combined and dried for MS analysis.

For the simplification of glycans by using acid-assisted method, the O-GalNAc glycopeptides from about 20 μg fetuin or 5 μL human serum were dissolved in 100 μL of 1% TFA, and the acidic solution was incubated at 75 °C for 60 min. Then the digests were dried for further analysis. For the simplification of glycans by using neuraminidases, the O-GalNAc glycopeptides from the same amount of proteins as above were dissolved in 100 μL of 50 mM PBS buffer (pH = 7.5), and about 200 U of the neuraminidase was added. The solution was vortexed and incubated at 37 °C for 20 h, and dried for MS analysis.

LC-MS/MS Analysis of O-GalNAc Glycopeptides: The analysis of enriched O-glycopeptides from fetuin and serum was performed by using Q-Exactive mass spectrometer equipped with a nanospray ion source and a U3000 RSLC nano-system (Thermo, San Jose, CA, USA). The simplified glycopeptides and native glycopeptides were dissolved in 0.1% FA/water, and subsequently loaded onto a trap column (4 cm length, 200 μm i.d.) packed with C18 AQ beads (5 μm, Daison, Osaka, Japan). The peptides were separated by a capillary analysis column (15 cm length, 150 μm i.d.) packed with C18 AQ beads (1.9 μm, Dr. Maisch, Germany). For the analysis of glycopeptides from fetuin, a 60 min gradient method was employed while 120 min gradient method was utilized for analysis of glycopeptides enriched from human serum: loading at 3% Buffer B (80%ACN/20%H₂O/0.1%FA) for 10 min, from 3% to 7% Buffer B for 5 min, from 7% to 45% Buffer B for 88 min, from 45 to 90% Buffer B in 2 min. The separation system was equilibrated by Buffer A (98%H₂O/2%ACN/0.1%FA) for 15 min. The full mass scan was acquired from 400 to 2000 with a resolution of 70,000. And information-dependent acquisition (IDA) mode was employed for the identification of glycosites, with the top 15 precursors subjected to MS/MS analysis with a resolution of 17,000. The HCD mode was used for MS² fragmentation with stepped NCE (normalized collision energy) of 23, 25 and 27.

For expanded analysis of glycosylation in human serum sample, fractionation before O-glycopeptide enrichment was performed. The O-glycopeptides enriched from serum digest equivalent to 66 μL serum was fractionated by using a PLRP-S column (150 × 2.1 mm i.d., 3 μm, Agilent Technologies). Mobile phases A (98% H₂O, adjusted pH to 10.0 using NH₄H₂O) and B (98% acetonitrile, adjusted pH to 10.0 using NH₄H₂O) were used to develop a gradient. The solvent gradient was set as follows: 0–2% B, 1 min; 2–10 B, 1 min; 10–15% B, 4 min; 15–40% B, 19 min; 40–70% B, 4 min; 70–85% B, 2 min. The column was then equilibrated by phase A. Fractions were collected in every 1 minute from 1 to 31 min, and the 30 fractions were mixed into ten fractions as in previous report.[14] The samples were dried under vacuum for subsequent MS analysis.

Data Analysis of O-GalNAc Glycopeptides: For the analysis of O-GalNAc glycopeptides before and after sialic acid release, the MS data of glycopeptides acquired by Q-Exactive MS was searched by using Byonic™ (PROTEIN METRICS, v2.13.2) against UniProt human database (70,709 sequences, 2017). Following parameters were used for searching: precursor and fragment tolerances were 10 and 20 ppm respectively; for serum samples, enzyme specificity was set to KR/P with up to two missed sites; cysteine residue was set as a static modification of 57.0215 Da; Asn deamidation (+0.9840 Da) and methionine oxidation (+15.9949 Da) were set as variable modifications. For the O-linked glycan modification, the six most common glycans (list in Figure 1, Supporting Information) were set as variable modifications on Ser/Thr residues. The protein FDR cutoff was 1% and peptides score <200 was cut off. To guarantee the confidence of the identifications, the hits with score between 200 and 300 was checked manually according to the existence of both oxonium ions and sufficient b and y-ions. After manually checking, the number of identified O-GalNAc modified proteins/peptide sequences slightly reduced (from 83/54 to 65/44 for 1D analysis; from 199/100 to 185/94 for 2D analysis).

3. Results and Discussion

Bovine fetuin-A, a major component in bovine fetuin, is known to bear both N-glycans and O-GalNAc glycans. And most of the O-GalNAc glycans attached on the bovine fetuin were reported to contain at least one sialic acid residue.[15] The theoretic tryptic peptide with four known O-glycosites (S271, T280, S282 and S296) of fetuin-A has the molecular weight of 5247.93 Da, which is too big to be identified by LC-MS/MS. To generate smaller peptides, elastase was used to digest the proteins. The digest was subjected to PNGase F treatment to release the N-linked glycans. And HILIC enrichment was performed for the isolation of O-GalNAc glycopeptides. Then the enriched O-glycopeptides were dissolved in 1% TFA and incubated at 75 °C for 60 min to desialylate the O-GalNAc glycans. The resulting peptide sample was finally submitted to LC-MS/MS analysis. For comparison, the sample without acid treatment was also analyzed by LC-MS/MS. To investigate the efficiency of de-sialylation, the acquired mass spectra for both cases were searched by setting variable modifications of the six most common O-linked glycoforms (Figure 1, Supporting Information). The identified peptides, O-GalNAc modified peptides, peptide sequences and O-glycans were listed in Table 1, Supporting Information. The spectra counts roughly indicate the abundance of the intact glycopeptides. It was found that the O-glycopeptides with sialic acid residues
Table 1. The number of O-glycopeptide and O-glycans identified from bovine fetuin without sialic acid (SA) removal, and removed by enzymatic (neuraminidase) and chemical (acid-assisted) methods.

<table>
<thead>
<tr>
<th>O-glycopeptides</th>
<th>O-glycans</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSMs</td>
<td>Unique Sequences</td>
</tr>
<tr>
<td>Native</td>
<td>259</td>
</tr>
<tr>
<td>Neuraminidase</td>
<td>258</td>
</tr>
<tr>
<td>Acid-assisted</td>
<td>342</td>
</tr>
</tbody>
</table>

accounted for over 60% of total spectra counts of O-glycopeptide identifications for the sample without acid treatment, while this percentage decreased to 6% when the enriched O-glycopeptides were treated with acid (Table 1). The de-sialylation efficiency of neuraminidase, a type of exogenous glycosidase which is widely used to release sialic acid residues form N-linked and O-linked glycans, was also tested by using bovine fetuin as the sample. It was found that the neuraminidase could catalyze the removal of sialic acid residues to some extent, but efficiency (Table 1, the sialylated glycopeptides accounts for approximately 29% of total spectra counts) is not as good as the acid-assisted strategy. We further compared the number of identified unique glycopeptide sequences for the three strategies. It was found that nearly two times of unique O-GalNAc modified peptide sequences were identified by acid-assisted desialylation strategy compared with the method without treatment and the neuraminidase strategy (Table 1). Especially, the number of peptide sequences identified covering at least one of the four known O-glycosites (S271, T280, S282 and S296) of fetuin-A increased from 38 to 86 after acid treatment. One important reason for the improvement of peptide sequence identification efficiency could be the removal of sialic acid residues improves the fragmentation of the peptide backbones. In the MS2 spectra of sialylated glycopeptides, the m/z 274 ion is often very intense which indicating NeuAc is protonated very well. Thus less charge is left on the rest of the peptide sequence which backbones would be more efficiently fragmented when the NeuAc was removed before MS analysis. Above results demonstrated that the acid-assisted desialylation strategy significantly enhanced the sensitivity for the identification of O-GalNAc modified peptide sequences.

Human serum is a complex body fluid that contains various glycoproteins ranging in concentration over at least 9 orders of magnitude. N-linked glycoproteome of human serum has been extensively investigated by using the established approaches. However, due to lack of effective methods, the O-GalNAc glycoproteome of human serum is seldom explored. By taking the good performance of glycoform simplification strategy, this method was utilized to analyze the O-GalNAc glycosylation in human serum. The procedure for the analysis is basically the same except trypsin was used for digestion. Due to the severe neutral loss of glycans during the HCD fragmentation, the O-glycosities could not be identified unambiguously if the peptides contain multiple Ser/Thr residues. Thus only the peptide sequences, glycans on the peptides and proteins were identified and summarized in Table S2. First, we evaluated the efficiency of de-sialylation for the complex human serum digest. The spectra contained 274.09213 Da (oxonium ion of dehydrated NeuAc residues) and 204.08665 Da (oxonium ion of GalNAc residues) were extracted from the raw data before and after acid treatment. We believed that the intensity of XIC of 274 could reveal the level of sialylation while that of 204 could reveal the level of O-GalNAc modification. As shown in Figure 2, the intensity of XIC of m/z 204 changed little before and after acid hydrolysis while the intensity of XIC of m/z 274 decreased two magnitudes of orders after the acid treatment. This result indicated that the acid treatment released the sialic acid residues with high efficiency while leave the O-GalNAc modification attached on the peptide backbones. Due to the high diversity of native serum O-glycans, much more types of O-glycopeptides should be presented in the sample without acid treatment compared with that with acid treatment where the glycans were simplified. On the contrary, slightly more intact glycopeptides (233 vs. 201) were identified by the acid treated sample. This is attributed to the improved identification sensitivity due to the simplified glycans. More importantly, significant increase in the sensitivity for the identification of glycopeptide sequences and glycoproteins were observed (Figure 3A). Only 49 glycopeptide sequences corresponding to 28 glycoproteins were identified when the sialic acid residues were not removed. However, there were 57.1% more O-glycopeptides (44) identified by using acid-assisted glycoform simplification strategy. And the number of unique peptide sequences in glycopeptides also increased from 49 to 65, which clearly indicated the significantly improved analysis coverage. For example, the insulin-like growth factor-binding protein 6 (IBP6_HUMAN, P24592) was reported to carry (HexNAc)2(Hex)2(NeuAc)2 glycoform on the T126. However, the corresponding tryptic peptide carrying this site was not identified for the analysis of the serum digest without removal of sialic acid residues. While the tryptic peptide with the simplified glycoform (HexNAc)2(Hex)2 was identified after the sample was treated with acid (Figure S2A). Extracellular matrix protein 1 (ECM1_HUMAN, Q16610) was another example which was reported to be modified with O-linked glycans on the range from T325 to S335. The peptide containing this sequence was identified in the acid-assisted glycoform simplification strategy while not in the direct analysis (Figure S2B). It should be mentioned that the de-sialylation is in favor of HCD fragmentation due to the removal of negative-ion sialic acid residues. In Figure 3B and 3C, the MS2 spectrum of the same sequence bearing different glycoforms ((HexNAc)2(Hex)2 and (Hex)2(NeuAc)2) were shown. Apparently, fragment ions of de-sialylated O-glycopeptides were more efficient than the ones with native O-glycan. In a word, the acid-assisted glycoform simplification strategy offers better chance for the identification of relatively low abundant O-glycoproteins. The simplification of the glycans could be illustrated by an example shown in Figure 3, Supporting Information. Inter-alpha-trypsin inhibitor heavy chain H1 (ITIH1_HUMAN, P19827) was a well-studied O-GalNAc modified glycoprotein with the T653 O-glycosylated. Native ITIH1 identified with tryptic peptides with three types of O-glycans which were (HexNAc)2(Hex)2 (23%), (HexNAc)2(Hex)2(NeuAc)1;
Figure 2. The extracted ion chromatography (XIC) of m/z = 204.09213 (A and C) and m/z = 274.08665 (B and D) before (A and B) and after (C and D) acid treatment.

Figure 3. A) Distributions of serum O-linked glycopeptides, glycopeptide sequences and glycoproteins identified without (in blue circle) and with (in yellow circle) the glycoform simplification strategy. The MS² spectrum of peptide sequence TVVPSVGAAAGPPC carrying different O-linked glycans of (HexNAc)₁(Hex)₁(NeuAc)₁ (B) and (HexNAc)₁(Hex)₁ (C).
(23%) and \((\text{HexNAc})_{1}(\text{Hex})_{1}(\text{NeuAc})_{2}\) (54%). While after de-sialylated, only one type of glycan \((\text{HexNAc})_{1}(\text{Hex})_{1}\) (100%) was observed.

To further map the O-GalNAc glycosylation in human serum, high pH reversed phase liquid chromatography was employed to the fractionation of enriched glycopeptides. After LC-MS/MS analysis, 185 unique peptide sequences corresponding to 94 proteins were identified to be O-glycosylated with FDR < 1% and score > 200. It was found over 60% of these proteins were reported as O-GalNAcylated before\([18,19,21]\) which indicated the high confidence of this identification. Though the glycans were simplified, this proteomics analysis still revealed that the high heterogeneity of O-GalNAc glycosylation. For example, plasma protease C1 inhibitor (\text{IC1_HUMAN}, P05155) was identified to be O-GalNAcylated on the peptide sequence \((\text{K})_{44}\text{VATTVISK}_{52}\). As shown in Figure S5, several glycan compositions, majority of them were sialic acid free, were identified to bind on the peptide sequence. We also investigated the distribution of glycoproteins on the number of identified unique O-glycopeptides and their spectra counts. As shown in Figure 4, the identified O-glycoproteins with high unique O-glycopeptide counts (or peptide spectrum match, PSMs) were almost reported as O-GalNAc modified proteins before. The high ratio of known O-glycoproteins in the datasets of high unique O-glycopeptide counts (> 20) and PSMs of O-glycopeptides (> 100) revealed the reliability of this method for the identification of high-abundant O-glycoproteins. While the ability for discovering low-abundant O-glycoproteins of this method could be reflected in the high ratio of newly discovered O-glycoproteins from the dataset of low unique O-glycopeptide counts (PSMs of O-glycopeptides). Immunoglobulin heavy constant alpha 1 (\text{IGA1_HUMAN}), which is a high abundant human blood protein with O-GalNAc modifications on the range of \((\text{K})_{89}\text{HYTNPSQDVTPCPVPSTPSTPSTPSTPSTPSTPSTPCCHPR}_{126}\),\([22]\) was identified most times in this method. This peptide sequence is rich in proline residues, making its O-GalNAc...

Figure 4. The distribution of the identified glycoproteins on A) the number of identified unique O-glycopeptides and B) their spectra counts. The percentages on the top of columns represent the ratio of known O-glycoproteins among the covered range. (The number of known O-glycoproteins were represented in blue and the newly identified in orange.)
modified forms quite hydrophilic, which is in favor of the HILIC enrichment method in our strategy. The high abundance and hydrophilicity are probably the reasons why IGA1 were identified most times in this method. Furthermore, this glycan simplification method could reduce the molecular weight of O-glycopeptides to some extent. The reduced molecular weight would help the O-glycopeptides, especially the peptide sequences with multiple amino acid residues and potential O-glycosites just as the the sequence of IGA1, to be better identified in shotgun proteomics analysis. Though the O-glycosites were not identified unambiguously on specific residue due to the serious neutral loss of glycans in HCD, the glycopeptide sequences were confidently identified which could locate the regions where O-glycosylation occured. Therefore, the distribution of amino acid residues around the potential glycosites were also investigated (Figure S6). It can be seen that sequences around the O-GalNAc modified sites are rich in proline residues, which is consistent with the reported rules of O-GalNAc glycosylation occurrence.\[6\] Thus above results indicate that O-GalNAc is highly heterogeneous and the simplification strategy is fitted for global characterization of O-GalNAc glycosylation at proteome level.

The serum O-GalNAc glycoproteome is not well explored. In our previous work\[11\], the intact O-GalNAc glycopeptides derived from human serum samples were directly analyzed by MS, which has the strength of revealing the high heterogeneity of O-glycosylation. For 1D LC–MS/MS analysis, there were 47 and 54 intact glycopeptides identified from normal human serum in two runs, respectively. While the number of unique glycopeptide sequences identified by 1D LC–MS/MS in this study were 56 and 50 in two runs (Table S4). Considering the high diversity of the O-glycans, the number of identified unique glycopeptide sequences in previous study should be much less than that obtained in this study. Clearly the glycoform simplification approach has higher sensitivity. Multiple 2D LC–MS/MS runs were applied to analyze two different serum samples in the previous study, which resulted in the identification of 407 intact O-GalNAc glycopeptides on 176 unique O-glycopeptide backbone sequences. Slightly more peptide sequences and glycoproteins were identified in this study (Figure S4). It should be mentioned that the number of identifications between these two studies were not comparable due to the total different workflows, different mass spectrometers and different data processing schemes used. However, we found the identifications at the levels of both peptide sequences and proteins are highly complimentary and additional 48 O-glycoproteins were identified in this study. The improved sensitivity by simplifying the glycoform should contribute to the identification of the additional O-glycoproteins in this study.

4. Concluding Remarks

In summary, a strategy based on acid-assisted glycoform simplifications was presented for the analysis of O-GalNAc glycosylation. We have demonstrated that the identification sensitivity was improved when the O-GalNAc glycan structures were simplified. Though this method is not as sensitive as the SimpleCell approach, it is applicable to analyse any type of biological samples. Thus this approach could be an alternative, cost-effective approach for the analysis of O-GalNAc proteome.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgement

X.Y. and H.Q. contributed equally to this work. The authors acknowledge funds from the China State Key Basic Research Program Grants (2016YFA0501402, 2017YFA0505003), the National Natural Science Foundation of China (21535008, 81430072, 21775146) and the innovation program (DICP TMR201601) of science and research from the DICP, CAS. M.Y. is a recipient of the National Science Fund of China for Distinguished Young Scholars (21525524).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

O-GalNAc, HILIC enrichment, N-acetyl-neuraminic acid