Chemoenzymatic Approach for the Proteomics Analysis of Mucin-Type Core-1 O-Glycosylation in Human Serum

Xin You,†‡ Yating Yao,†‡ Jiawei Mao,†‡ Hongqiang Qin,†‡ Xinmiao Liang,†‡ Liming Wang,§ and Mingliang Ye*,†‡

†CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian, 116023, China
‡University of Chinese Academy of Sciences, Beijing, 100049, China
§Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, The Second Hospital of Dalian Medical University, Dalian, Liaoning 116023, China

Supporting Information

ABSTRACT: Human serum is a complex body fluid that contains various N-linked and O-linked glycoproteins. Compared with N-linked glycoproteins, the serum O-linked glycoproteins are not well-studied due to their high heterogeneity and their low abundance. Herein, we presented a novel chemoenzymatic method to analyze core-1 type of O-GalNAcylation in human serum. In this approach, the tryptic digest of serum was first subjected to PNGase F treatment to release the N-glycan and was then treated with strong acid to release sialic acid residues from mucin-type O-glycans. In this way, the internal Gal/GalNAc residues were exposed and were oxidized by the galactose oxidase to carry the aldehyde groups. The oxidized O-GalNAcylated peptides were then captured by hydrazide beads and eluted with methoxylamine for LC−MS/MS analysis. The de-N-deglycosylation decreased the abundance of N-glycopeptides, the desialylation simplified the O-glycans and the enzymatic oxidation conferred the enrichment specificity. We have demonstrated that this method was fitted to analyze O-GalNAcylated peptides with high confidence. This method was applied to analyze human serum, which resulted in the identification of 59 O-GalNAc modified peptide sequences corresponding to 38 glycoproteins from 50 μL of serum. This method is expected to have broad applications in the analysis of O-glycoproteome.

Received: July 3, 2018
Accepted: October 16, 2018
Published: October 16, 2018

© 2018 American Chemical Society

Supporting Information

Human serum has long been the focus of basic and clinical research for its ability to reflect an individual’s physical status.1 Nowadays, much attention has been paid to the proteomics analysis of human serum. However, the high dynamic range of human serum protein levels challenges the current analytical methods. Most of the high abundance N-glycoproteins are well-studied.2 While the low abundance mucin-type O-glycosylation in human serum is seldom explored due to the lack of effective methods. It is reported that approximately 90% of the total O-glycan pool were mucin-type core-1 structures.3 Thus, the analysis of mucin-type core-1 glycosylation in human serum is the key to obtain better understanding of O-glycosylation in human serum. As the aberrant O-GalNAcylation has close relationship to many diseases including tumor, the advance in the proteomics analysis of O-GalNAcylation should facilitate the identification of novel disease biomarkers. Only a limited number of studies were reported to analyze the O-GalNAcylated proteome (O-glycoproteome) of serum. Darula et al. have applied Jacalin affinity-chromatography combined with MS-analysis by higher-energy dissociation (HCD) and electron transfer dissociation (ETD) fragmentation for the characterization of the core-1 mucin-type glycoproteome of serum.4 To facilitate the identification of O-glycopeptides, the O-glycans attached on the peptides were simplified by enzymatic release of sialic acid residues and
galactose residues before MS analysis. In their initial study, they identified 32 O-glycosylation sites in 13 different proteins from bovine serum. To identify low-abundance glycoproteins, they used two different workflows by including different fractionation schemes after the first protein-level affinity enrichment. Totally, 124 O-glycosites in 51 proteins were identified from at least 2 mL of bovine serum. Recently, they also applied this extensive fractionation approach to analyze the human serum where only 27 glycosylation sites was identified by analyzing at least 30 fractions. They then developed a new workflow that used two lectins sequentially: wheat germ agglutinin and then Jacalin. The first lectin provides general glycopeptide enrichment, while the second one specifically enriches O-linked glycopeptides with Gal/β1-3GalNAcα structures. By using this method, they identified 52 unique glycosites in 20 proteins from 400 μL of human serum. They believed this was because of the poor specificity of the used lectins. Especially they thought the incomplete removal of sialic acids by neuraminidase would compromise the enrichment efficiency of Jacalin. In addition to lectins, hydrophilic interaction chromatography (HILIC) was also used to enrich O-glycopeptides for the analysis of O-glycoproteome. HILIC is a popular enrichment method for the enrichment of N-glycopeptides. We have demonstrated that HILIC can also be used effectively to enrich the O-glycopeptides when the N-linked glycans were removed by PNGase F. After extensive fractionation of the enriched intact O-glycopeptides by high pH reversed phase chromatography, 407 O-GalNAc glycopeptides (176 unique peptide backbone sequences) on 93 glycoproteins were identified from serum sample. The identification efficiency was still not high as only about 50 intact glycopeptides (about 34 unique backbone sequences) were identified in a 1D LC−MS/MS run. Recently, Yang et al. identified 68 unique intact O-glycopeptides corresponding to 13 unique peptide backbones from human serum by HILIC method. We found that chemical desialylation was helpful to improve identification efficiency and about 58 unique glycopeptide sequences identified in a 1D LC−MS/MS run. Compared with the lectin based method, a less amount of sample was consumed. For one 2D LC−MS analysis, typically only 200 μL of serum was enough for the HILIC based method. The strong anion exchange column can also be used to enrich and identify glycopeptide, which resulted in the identification of 31 peptide backbones. As different data processing methods were applied in these studies, the identification numbers are not comparable. However, the low identification numbers indicate that current methods are still not good enough for the profiling of O-GalNAcylated peptides in serum samples. Up until now, the only method enabling comprehensive analysis of O-glycoproteome is the ”Simple-Cell” strategy. In this strategy, zinc-finger nuclease was used to truncate the O-glycan elongation pathway in human cells, making all the mucin-type O-glycans in the engineered cells were GalNAc and NeuAc-GalNAc. Combining with ETD fragmentation, over 350 O-glycosites corresponding to over 100 proteins were identified from human cells. Later, the same group identified 1538 O-glycosites in 738 glycoproteins from the cell lysates and secretomes in the Chinese hamster ovary cell lines. However, this method is only applicable to the analysis of the engineered cell lines but not the humoral or the tissue samples.

Hydrazide chemistry is a popular method for analyzing N-glycoproteome. In this method, the cis-diols on N-glycans of N-glycopeptides are oxidized into aldehydes by sodium periodate and then are covalently captured by hydrazide beads. The peptide backbones are enzymatically released from the beads by PNGase F for LC−MS/MS analysis. Theoretically, hydrazide chemistry could also be utilized for analyzing O-glycosylation in biological samples. Due to lacking of effective endoglycosidase to cleave the total O-glycan, the captured O-glycopeptides can only be released by hydroxylamine and its homologues, leaving the modified glycans attached on the peptide backbones. Because the sodium periodate can oxidize any cis-diols on O-glycans and some diols may be partially oxidized, this makes the O-glycans highly heterogeneous. Due to the significantly expanded searching space by considering the complex structures of the modified O-glycans, the identification of the O-glycopeptides by this approach could be inefficient. In addition to sodium periodate, galactose oxidase was reported to specifically oxidize the terminal galactose, which enabled the characterization of Tn antigens in human cells. However, it is well-known that the O-glycans were predominantly sialylated and thus cannot be oxidized by galactose oxidase. The sialic acid residues could be removed from O-glycans by enzyme or chemical approach, which exposes the galactose residue and can be oxidized by the oxidase. The resulted O-glycans are more homogeneous and are fitted to the hydrazide chemistry for O-glycoproteome analysis. Based on this principle, we developed a chemoenzymatic approach for the analysis of O-glycosylation in human serum. In this approach, the tryptic digest of serum was first subjected to PNGase F treatment to release the N-glycan and was then treated with TFA to release sialic acid residues from mucin-type O-glycans. In this way, the internal Gal/GalNAc residues were exposed to be oxidized by the galactose oxidase to carry the aldehyde groups. The oxidized O-GalNAcylated peptides were captured by hydrazide beads and eluted with methoxylamine. The enriched O-GalNAcylated peptides were then analyzed by 1D LC−MS/MS in HCD (higher-energy collision dissociation) mode with stepped NCE (normalized collision energy), which resulted in the identification of 59 O-GalNAc modified peptide sequences corresponding to 38 glycoproteins from 50 μL of serum. For the HILIC based method, the short nonsialylated O-glycopeptides were hardly enriched due to their low hydrophilicity and false positive identifications were often occurred due to the presence of residual N-linked glycosylation. We found this chemoenzymatic approach can effectively circumvent the above problems.

**EXPERIMENTAL SECTION**

**Chemicals and Reagents.** Standard O-GalNAcylated peptide was purchased from Sussex Research (Ottawa, Canada). 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). Galactose oxidase (GAO) was purchased from Innovative Research (Michigan, USA). Chemical reagents were obtained from Sigma (St. Louis, MO). Formic acid (FA) was obtained from Fluka (Buchs, Germany). Acetonitrile and water for RPLC (HPLC grade) was purchased from Merck (Darmstadt, Germany). Ammonium bicarbonate and urea were obtained from Bio Basic Inc. (Ontario, Canada). Pure water for protein digestions and glycopeptide enrichment was purified with a Milli-Q system (Millipore, Milford, MA). 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 50 volunteers and pooled together with equal-volume. The samples were stored at −80 °C until usage.

**Cell Culture.** The Jurkat cells were cultured in RPMI 1640 medium, supplemented with 10% new born bovine serum, 100 U/mL of penicillin and 100 μg/mL of streptomycin in a humidified atmosphere containing 5% CO2 at 37 °C. The cells were harvested at about 75% density and were treated with 1 mM freshly prepared pervanadate for 15 min at 37 °C. After that, the cell pellets were homogenized in an ice-cold lysis buffer containing 50 mM Tris-HCl (pH 7.4), 8 M urea, phosphatase inhibitor (1 mM NaF and 1 mM Na3VO4), 2% protease inhibitor cocktail, and 1% Triton X-100 in ice bath, sonicated at 400 W for 10 min, and centrifuged at 25 000 g for 30 min. The supernatant was precipitated and purified as previously described. Then the obtained proteins were dissolved in buffer containing 50 mM Tris-HCl/8 M urea (pH 8.0), and the protein concentration was determined by BCA assay. After that, the redissolved proteins were reduced by 20 mM DTT at 37 °C for 2 h, alkylated by 40 mM IAA at room temperature for 40 min in the dark, then diluted to 1 M urea by 50 mM Tris-HCl buffer (pH 8.2), and finally digested by trypsin overnight at 37 °C with a ratio of enzyme-to-protein at 1:20 (w/w). The obtained digests were kept at −80 °C for further usage.

**Protein Digestion and De-N-Glycosylation.** The standard glycoprotein (bovine fetuin) and serum was denatured by dissolving in 50 mM HEPES/8 M urea buffer with 20 mM DTT. After incubation at 37 °C for 2 h, 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 = 7.8). The digestion process was lasted for about 20 h. Then the tryptic peptides were desalted by Oasis HLB column. After lyophilization, the peptides were redissolved in GlycoBuffer 2 (pH = 7.5) and PNGose 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.

**HILIC Enrichment.** The O-GalNAc glycopeptides were enriched by using HILIC tip materials according to our previous reports. The digestions of proteins were redissolved in the loading buffer (80% ACN/1% TFA) and loaded onto a HILIC tip with 5 mg of 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. The elution was combined and dried for MS analysis.

**Chemoenzymatic Enrichment.** The de-N-glycosylated peptides were desialylated by being treated with 1% TFA at 75 °C for 1 h. After lyophilization, the peptides were redissolved in GAQ buffer (25 mM Na2PO4 40 units/mL HRP, 25 units/mL GAQ, pH = 7.0) containing 10% DMSO. DMSO was used to scavenge the radicals generated during the reaction. The oxidation reaction was performed at 35 °C overnight in the dark. The reaction was neutralized by adding 50% acetic acid to adjust the pH to around 5.0, and the hydrazide beads was added to the oxidized peptide solution after prewashing with oxidation buffer (200 mM NH4Ac, 150 mM NaCl, pH = 5.0). The binding reaction was performed at 25 °C overnight. Then the nonspecific bound peptides were removed by washing the hydrazide beads with 50 mM HEPES/8 M urea, 80% ACN, and 1.5 M NaCl for three times, respectively. The bound O-GalNAcylated peptides were then eluted with elution buffer (0.2 M CH3ONa, 1.5 M NaCl, 0.1 M amine, and 0.1 M NaAc, pH = 4.5) for 12 h at 25 °C. The eluent were collected and desalted by Oasis HLB column. The desalted samples were lyophilized to dryness for LC–MS/MS analysis.

**MALDI TOF MS Analysis.** 5800 MALDI-TOF/TOF mass spectrometer (AB Sciex) equipped with a 355 nm pulsed Nd/YAG laser was used to conduct MALDI-TOF MS experiments. 2,5-Dihydroxybenzoic acid (DHB) was dissolved in ACN/H2O/H3PO4 = 50:49:1 to a final concentration 25 mg/ml as matrix. A 0.5 μL aliquot of the peptide sample and 0.5 μL of the DHB matrix were sequentially dropped into the MALDI plate for MALDI-TOF MS analysis.

**LC–MS/MS Analysis.** The analysis of enriched O-glycopeptides from fetuin and serum was performed by using Q-Exactive mass spectrometer equipped with a nanoprobe ion source and a U3000 RSLC nanosystem (Thermo, San Jose, CA, USA). The enriched glycopeptides were dissolved in 0.1% FA/water and subsequently loaded onto a trap column (4 cm length, 150 μm i.d.) packed with C18 AQ beads (1.9 μm, Dr. Maisch, Germany). 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 130 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. Data-dependent acquisition (DDA) mode was employed for the identification of glycosites, with the top 15 precursors subjected to MS/MS analysis with a resolution of 17 500.

**Database Searching.** For the analysis of O-GalNAc glycopeptides, the MS data of glycopeptides acquired by Q-Exactive MS was searched by using Sequest 1.4 against the Uniprot human database (70 709 sequences, 2017). Following parameters were used for searching: precursor and fragment ion tolerances were 10 ppm and 0.02 Da, respectively; for serum samples, enzyme specificity was set to KR/P with up to 2 missed cleavage; cysteine residue was set as variable modification of 57.0215 Da; Asn deamidation (+0.9840 Da) and methionine oxidation (+15.9949 Da) were set as variable modifications. For the core-1 type of O-GalNAc modifications, the mass tag of +230.0903 Da, +212.0797 Da (due to a water loss), +392.1431 Da, and +374.1325 Da (due to a water loss) on serine and threonine residues were set as variable modifications. Proteins and peptides with the FDR more than 1% were cutoff and the peptide spectral matches were filtered to 1% FDR.
RESULTS AND DISCUSSION

The overall workflow for this chemoenzymatic approach was shown in Figure 1A. It has three enzymatic steps and three chemical steps. The first enzymatic step is the digestion of serum proteins by trypsin, which is a general step for proteomics. Then the resulted digest is subjected to de-N-glycosylation using PNGase F to remove the N-glycans, the major interference for the detection of low abundance O-glycosylated peptides. Rather than using neuraminidase, a chemical approach was used for desialylation which exposed Gal or GalNAc residues on O-glycosites (Figure 1B), which was then allowed to be oxidized by galactose oxidase to carry the aldehyde groups (Figure 1C). The oxidized O-GalNAcylated peptides are covalently captured by hydrazide beads as the classic hydrazide chemistry method to enrich N-glycopeptides. Hydroxylamine and its homologues were previously employed to release the covalently captured (glyco)peptides from hydrazide beads due to their ability of breaking the hydrazone bonds between (glyco)peptides and hydrazide beads. In this study, we used methoxylamine to

Figure 1. Chemoenzymatic strategy for the enrichment of O-GalNAcylated peptides from protein digest (yellow ■, GalNAc; yellow ●, Gal; green ●, Man; blue ■, GlcNAc; purple ◆, NeuAc). (A) Workflow of the GAO-based strategy, (B) (S)Tn antigen and core-1 mucin-type O-GalNAcylated peptides before and after desialylation, and (C) covalently capture and release lead to addition of +27 Da tag.
release covalently captured O-glycopeptides from hydrazide beads, because the methoxylamine could leave the +27 Da mass tag on the terminal galactose residues (Figure 1C) which could be detected in MS/MS analysis. The released O-GalNAcylated peptides were then analyzed and identified by LC−MS/MS in higher-energy collision dissociation (HCD) mode with stepped normalized collision energy (NCE).

We first validated if the galactose oxidase can oxidize the desialylated O-glycopeptides and allowed them to be enriched from complex peptide mixture. It is reported that approximately 90% of the total O-glycan pool in serum were mucin-type core-1 structures. Among them, the majority are sialylated, including NeuAc α2-3 Gal β1-3 GalNAc (~60%), NeuAc α2-3 Gal β1-3 (NeuAc β2-6) GalNAc (~20%), Gal β1-3 GalNAc (~10%), and NeuAc β2-6 GalNAc (~0.1%). When they are desialylated, the core-1 mucin-type O-GalNAc glycans could be simplified into only two types of glycans, i.e., GalNAc and Gal β1-3 GalNAc (Figure 1B). Since the galactose oxidase had been proved to be effective in oxidizing the GalNAc modified peptides, we further validate it can also oxidize the Gal β1-3 GalNAc modified peptides. A synthetic peptide carrying Gal β1-3 GalNAc was treated with galactose oxidase and detected with MALDI-TOF-MS. As shown in Figure S1, a 2 Da mass shift was observed which indicated the oxidation. The origin glycopeptide was barely observed in the spectra meant the oxidation was quite complete. The 2 Da mass shift also indicated that only one hydroxyl on the terminal galactose residue was oxidized by galactose oxidase (as shown in Figure 1C). In a serum sample, the O-GalNAcylated peptides coexists with huge amounts of nonglycopeptides and N-linked glycopeptides. A peptide mixture containing the synthetic peptide carrying Gal β1-3 GalNAc, IgG digest, and BSA digest with the mole ratio of 1:2:10 was used to mimic the real complex sample with both high abundant nonglycoproteins and N-glycoproteins and low abundant O-glycoproteins. This chemoenzymatic strategy was compared with the HILIC for the enrichment of the O-glycopeptide from this peptide mixture. An aliquot of the peptide mixture was treated with PNGase F to remove the N-glycan. Both samples, without and with de-N-glycosylated, were subjected to the enrichment by the HILIC and the chemoenzymatic method. The HILIC method can effectively enrich the N-glycopeptides on the IgG from the mixture due to the high hydrophilicity of the large N-glycans (Figure 2A). However, the peak of the synthetic O-glycopeptide was not observed. While for the chemoenzymatic method, an intensive peak with m/z of 1240.79 Da corresponding to the synthetic peptide was observed (Figure 2B). This indicated that this O-glycopeptide was effectively enriched. However, there were some intensive peaks with m/z over 2400 Da. These peaks vanished when the de-N-glycosylated sample was used (Figure 2D). This is because the N-glycopeptides carrying glycan with terminal galactose residues can also be oxidized and so can be captured by the hydrazide beads. The vanishing of these peaks indicates the interference of high abundant N-glycosylated peptides could be effectively eliminated by the deglycosylation using PNGase F. For HILIC enrichment, the synthetic O-glycopeptidase was not observed even when the de-N-glycosylated sample was used (Figure 2C). This means HILIC is unable to enrich this type of O-glycopeptide. Furthermore, HILIC is likely to enrich hydrophilic non-glycopeptides mainly appeared in the range with m/z <2000 (Figure 2A,C). While much less such peaks were observed in the chemoenzymatic method (Figure 2B,D), indicating high specificity of this method. A peptide mixture without any glycopeptide other than the synthetic O-glycopeptidase was also subjected to the enrichment by the HILIC and chemoenzymatic method. Some abundant peaks were observed in the HILIC method (Figure 2E) while only a dominant peak of the O-glycopeptide was observed (Figure 2F). This further
indicated that the HILIC will bring the interference of hydrophilic nonglycopeptides. Bring together, above results indicated that this method can specifically enrich Gal β1-3 GalNAc modified peptides.

We further prove that this chemoenzymatic strategy is more suitable for enriching relatively short O-glycan modified peptides by using total cell lysate of Jurkat cell line as the samples. Due to the defect in C1GALT1C (core 1 β3-galactosyltransferase specific molecular chaperone [Cosmc]) in Jurkat cells, the mucin-type O-glycans are mostly Tn antigens, i.e., GalNAc. Therefore, the GalNAc modified peptides in the de-N-glycosylated protein digest could be oxidized by galactose oxidase and captured by hydrazide beads without the desialylation step. For comparison, the same sample was also treated with HILIC method. As shown in Table S1 and Figure S2, the chemoenzymatic method showed higher identification efficiency than that of HILIC method in both O-glycopeptide sequence level (23 vs 13) and O-glycoprotein level (21 vs 10). We found the overlapped identifications between these two methods are very low (Figure S2B). We suspect that most of the O-glycopeptides identified by HILIC method are not correct. The O-glycopeptide backbones should rarely contain N-glycosylation motif (NXS/T, X≠P). However, there were 9 out of 13 identified peptide sequences contained N-glycosylation motif for the HILIC method while only 1 out of 23 contained the motif in the chemoenzymatic method. The high frequency of the N-glycosylation motif hinted the high false positive identification in the HILIC method. The high false positive identification could also be evidenced by the cellular location of the identified proteins. The O-glycoproteins identified by chemoenzymatic method were mostly on membranes while the O-glycoproteins identified by HILIC method were almost nucleus proteins (Figure S2C). For example, the UBQL2_HUMAN and NUF2_P2_HUMAN, annotated as nucleus proteins, were identified from HILIC method as O-GalNAcylated proteins and covered by many PSMs (peptide spectrum matches). Apparently, these two proteins were false positive identification because the O-GalNAcylated and N-glycosylation do not occur on nucleus proteins. The different performance of the two enrichment methods were mainly attributed to the properties of mucin-type O-glycans in Jurkat cell lines. As mentioned above, the majority of mucin-type O-glycans contains no other monosaccharides but only one GalNAc residue in Jurkat cell lines. O-Glycopeptides modified with this kind of short O-glycans were hardly enriched by HILIC method due to its low hydrophilicity as we observed above. As a result, the “glycopeptides” obtained from HILIC method were basically the nonspecifically captured hydrophilic non-Glycopeptides peptides and residue amount of N-glycopeptides that were not completed deglycosylated. These peptides could be false positive identified as O-glycopeptides during the database searching (Figure S3). While the chemoenzymatic method can enrich the glycopeptides oxidized by galactose oxidase without any discrimination on the hydrophilicity of the modified glycans. Therefore, for the analysis of O-GalNAcylated with short glycans, this chemoenzymatic method is more reliable.

As for core-1 O-GalNAcylation, the terminal galactose residues are usually modified with sialic acid residues, which prevents the oxidation of Gal or GalNAc by galactose oxidase. While once the sialic acid residues are removed, almost all the core-1 O-glycans were terminated with Gal or GalNAc residues. This means that basically all the core-1 type of O-GalNAcylated peptides could be enriched by this chemoenzymatic method after desialylation. Also, desialylation step would decrease the heterogeneity of O-glycans (Figure 1B), which facilitate the identification of core-1 mucin-type O-glycosylation. Therefore, the removal of sialic acid residues is critical in our enrichment workflow. Enzymatic removal of sialic acid residues was reported to be not efficient with some glycan structures. Instead, we use a chemical approach in this study. As the bonds between sialic acid and other saccharides is more sensitive to acid, TFA was used to release sialic acid residues from core-1 O-glycans due to its compatibility with proteomic analysis. Bovine fetuin is a well-studied O-glycoprotein standard, the O-glycoforms of which are almost sialylated core-1 O-glycans. We employed bovine fetuin to test the efficiency of desialylation. Due to the absence of tryptic cleavage sites around known O-glycosylated sites, elastase was used for digestion. Bovine fetuin digest with and without TFA treatment was enriched by this chemoenzymatic method and analyzed with LC−MS/MS. As shown in Table S2 and Figure 3A, the PSMs and modified peptide sequences identified from desialylation group is significantly more than that of native group. This result indicated that the removal of sialic acid residues especially for heavily sialylated O-GalNAcylated

Figure 3. (A) The number of PSMs, unique O-glycopeptides and peptide sequences identified from bovine fetuin-A by chemoenzymatic method before and after desialylation by TFA. (B) The number of unique O-GalNAc modified peptide sequences and proteins identified from human serum by chemoenzymatic method under different NCEs.
Fragmentation of glycopeptides in mass spectrometry is another key step for analyzing O-GalNAcylated peptides. ETD was reported to be suitable for the MS/MS fragmentation of O-glycopeptides due to its ability to generate b, y ions with the sugar chains attached. However, the limited fragmentation efficiency of ETD inhibited its widely applications in analysis of O-GalNAcylated peptides with low abundance. Also, ETD is usually equipped by expensive mass spectrometry such as Fusion Series of ThermoFisher. These kinds of mass spectrometry are not available in our lab, so we choose HCD for MS/MS fragmentation in this work. We found in our experiments that neutral loss of glycans was easily occurred on the O-GalNAcylated peptides in HCD fragmentation mode. The neutral loss would absorb some HCD energy resulting in the incomplete fragmentation of peptide backbones due to the absence of enough fragmentation energy. Therefore, a stepped NCE technique was used for the fragmentation of the enriched O-GalNAc modified peptides. The stepped NCE technique referred to the fragmentation of the same ions under at most three different collision energies in the HCD cell sequentially generating one spectrum containing all the fragment ions from different collision energy. This stepped NCE technique guaranteed adequate collision energy for the fragmentation of peptide backbones and so more peptide fragment ions were observed. To optimize the NCE of O-GalNAcylated peptides, four groups of stepped NCE were used to analyze the samples from human serum (NCE 25 represented stepped NCE: 23, 25, and 27; NCE 28 represented stepped NCE: 25, 28, and 31; NCE 31 represented stepped NCE: 28, 31, and 34; NCE 34 represented stepped NCE: 31, 34, and 37). Each group of NCE was performed three parallel experiments with ~50 μL of human serum in total. As shown in Figure 3B and Tables S3–S6, the number of identified O-glycopeptide sequences and proteins with the NCE 31 (stepped NCE: 28, 31, 34) is significantly higher than that of other three kinds of NCE. There were 146% more O-glycopeptide sequences (59 vs 24) and 81% more O-glycoproteins (38 vs 21) were identified with the NCE 31 compared with NCE 25. These results indicated the importance of collision energy optimization in the analysis of O-GalNAcylation.

The global analysis of O-GalNAcylation in human serum was performed by using ~50 μL of human serum. The human serum sample was digest by trypsin and de-N-glycosylated by PNGase F before being treated with GAO. The GAO-treated sample was then enriched by hydrazide beads and analyzed with LC−MS/MS under stepped NCE mode. The results from three parallel experiments were listed and summarized in Table S5. Totally, 59 core-1 type of O-GalNAc modified peptide sequences corresponding to 38 O-glycoproteins were identified by our chemoenzymatic strategy with FDR less than 1% at PSM level (Figure 4A). After the FDR filtering, the lowest value of Xcorr of PSMs was around 1. These identifications were still quite confident as sufficient b, y-ions and oxonium ions were observed (Figure S4). We believed that the high enrichment specificity was the probable reason why the PSMs with low value of Xcorr were confidently identified. Among all identified O-glycoproteins, 65.8% were reported as O-GalNAcylated proteins before. The high percentage of known O-glycoproteins also indicated the high confidence of the identifications. It is well-known that O-glycopeptides suffer from severe neutral loss during the HCD fragmentation. So virtually no b/y-ions carrying saccharide presents in the MS/MS spectra, which makes the assignment of the glycan(s) to specific residue(s) impossible. So in this work, the O-GalNAc modified peptides were identified without the accurate modified sites. The identified O-glycoproteins were annotated and characterized with keywords by DAVID tool (Figure 4B). The result showed that the most apparent characters of identified proteins were secreted and glycoproteins, which proved the confidence of our identification. The character “Signal” was consistent with one of the main functions of

![Figure 4](https://example.com/image.png)

**Figure 4.** (A) Overlap of identified O-glycopeptide sequences and O-glycoproteins from three experimental runs by chemoenzymatic method. (B) Keywords annotation of identified O-glycoproteins by David tools (p < 0.01, FDR < 1%).
According to Zheng’s reports, the GAO-treated Tn antigen modified peptides would generate report ions during the MS/MS fragmentation (m/z = 171.0766 and 195.0763 Da). The same report ions were also found from the Tn antigen modified peptides spectrum in our experiments (Figure S5). Similarly, we discovered specific report oxonium ions (m/z = 204.0866 and 393.1497 Da) from the spectrum of Gal β1-3 GalNAc modified peptides (Figure 5). The 393.1497 is the oxonium of modified Gal β1-3 GalNAc and the 204.0866 is the oxonium of GalNAc, which was produced from HCD fragmentation of modified Gal β1-3 GalNAc. The coexistence of these two report ions was an obvious character for the recognition of spectrum from Gal β1-3 GalNAc modified peptides. All the PSMs were manually checked to make sure the existence of corresponding report ions and sufficient b, y-ions. Due to the microheterogeneity of O-GalNAc modified peptides, the intensity of report ions varied from PSM to PSM, but all the PSMs contained the corresponding report ions. As shown in Figure S6A, the peptide TTPPTATPPIR was identified to carry GalNAc and Gal1-3GalNAc simultaneously. Correspondingly, all the four reporter oxonium ions were found in the low mass range (Figure S6B). IGHG3_HUMAN, as the constant region of immunoglobulin heavy chains, was previously found to be O-GalNAcylated on the range of TPLGDTHHTCPPR\textsubscript{115}–\textsubscript{1272}. While in our strategy, the peptide range of SCDTPPPCPR was also identified to be modified with O-GalNAcyl. Since this sequence of amino acids repeats three times among the range from S118 to R158, we inferred that the region from T102 to R158 in IGHG3_HUMAN could be O-GalNAcylated (Figure S7).

The analysis of serum O-glycopeptidome was also achieved by enrichment of O-glycopeptides with HILIC in our previous work.\textsuperscript{7} The multiple 2D separations of multiple serum samples led to the identification of 407 intact O-GalNAc glycopeptides from 93 glycoproteins. However, for 1D analysis of a serum sample, only 50 intact O-glycopeptides corresponding to about 34 peptide sequences were identified. The identified unique O-GalNAcylated peptide sequences account for 57.6% of that achieved in this study where only 1D separation was performed. The overlap of this work and our previous work with 1D analysis at peptide sequence and protein levels was shown in Figure S8. Very low overlap at both levels, 22.4% at peptide level and 29.8% at protein level, indicated that these two methods were highly complementary. The HILIC method is known to have poor enrichment specificity and is likely to generate higher false positive identification as we mentioned before. While this chemoenzymatic approach has exclusive specificity for the O-glycopeptides containing terminal galactose residues, this makes the identification more confident. There are some drawbacks for this approach. For example, the long sample preparation workflow makes this method time-consuming and the desialylation step makes this method unable to reveal the heterogeneity of sialylation. However, the ability of enriching O-glycopeptides with relatively low hydrophilicity makes this chemoenzymatic method an excellent complementary method of the HILIC-based methods. It should be mentioned that this chemoenzymatic method is not specific to core-1 glycosylation. Theoretically all the O-glycoforms with terminal galactose could be oxidized by galactose oxidase and enriched by hydrazide beads. While for serum, over 90% O-glycosylation was core-1 type as mentioned above. Thus, other kinds of O-glycosylation are of low abundance and could barely affect the enrichment specificity of our method in the analysis of core-1 O-glycosylation in human serum.

CONCLUSION

In conclusion, a chemoenzymatic method, by combining the step of de-N-glycosylation, chemical assisted desialylation, terminal galactose oxidation, covalently capturing by hydrazide beads, was presented for the profiling the core-1 type of O-GalNAcyl in human serum. We demonstrated that this method could offer excellent enrichment specificity and reliable O-glycopeptide identifications. By using this method, 59 core-1 O-GalNAcylated peptide sequences corresponding to 38 O-glycopeptides were identified from 50 μL of human serum. This method is an excellent complementary approach for the analysis of O-glycopeptidome and is expected to have broad applications.
Figure S1, MALDI spectra of the standard O-GalNAcylated peptide; Figure S2, number of unique O-GalNAcylated peptide sequences and proteins from Jurkat cell lines digest identified by HILIC enrichment and chemoenzymatic enrichment, overlap of identified O-GalNAc modified peptide sequences and proteins by chemoenzymatic and HILIC methods, and cellular components analysis of identified O-glycoproteins by chemoenzymatic and HILIC methods; Figure S3, MS/MS spectrum of false positive identified O-glycopeptidies in HILIC enrichment method; Figure S4, MS/MS spectra of four O-glycopeptides identified with low Xcorr; Figure S5. MS/MS spectrum of (VQAAVGT-SAAPVPDSNH + GalNAc); Figure S6, MS/MS spectrum of O-glycopeptide (TTPPTTATPIR + GalNAc + Gal/J-3GalNAc); Figure S7, O-GalNAcylated range (marked with red column) on IGHG3_HUMAN identified by our method; and Figure S8, overlap between this work and our previous work of 1D analysis at O-GalNAc modified O-glycosylation in bovine fetuin-A and chemoenzymatic enrichment, overlap of identified O-GalNAcylated peptide sequences and proteins from Jurkat cell lines digest identified by HILIC enrichment and chemoenzymatic enrichment, overlap of identified O-GalNAc modified peptide sequences and proteins by chemoenzymatic and HILIC methods, and cellular components analysis of identified O-glycoproteins by chemoenzymatic and HILIC methods; Figure S3, MS/MS spectrum of false positive identified O-glycopeptidies in HILIC enrichment method; Figure S4, MS/MS spectra of four O-glycopeptides identified with low Xcorr; Figure S5. MS/MS spectrum of (VQAAVGT-SAAPVPDSNH + GalNAc); Figure S6, MS/MS spectrum of O-glycopeptide (TTPPTTATPIR + GalNAc + Gal/J-3GalNAc); Figure S7, O-GalNAcylated range (marked with red column) on IGHG3_HUMAN identified by our method; and Figure S8, overlap between this work and our previous work of 1D analysis at O-GalNAc modified O-glycosylation in Jurkat cell by HILIC and GAO (XLSX)

Table S1, identified O-glycosylation in Jurkat cell by HILIC and GAO (XLSX)

Table S2, identified O-glycosylation in bovine fetuin-A native and deSA (XLSX)

Table S3, identified O-glycosylation in human serum NCE25 (XLSX)

Table S4, identified O-glycosylation in human serum NCE28 (XLSX)

Table S5, identified O-glycosylation in human serum NCE31 (XLSX)

Table S6, identified O-glycosylation in human serum NCE34 (XLSX)

**AUTHOR INFORMATION**

**Corresponding Author**

E-mail: mingliang@dicp.ac.cn.

**ORCID**

Hongqiang Qin: 0000-0002-7508-0872

Xinniao Liang: 0000-0001-5802-1961

Mingliang Ye: 0000-0002-5872-9326

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The authors acknowledge funds from the China State Key Basic Research Program Grants (Grants 2018YFC0910302, 2016YFA0501402, 2017YFA0505003), the National Natural Science Foundation of China (Grants 21535008, 81430072, 21775146, 81272368, 81471755) and the Innovation Program (Grant DICP TMSR201601) of Science and Research from the DICP, CAS. M.Y. is a recipient of the National Science Fund of China for Distinguished Young Scholars (Grant 21525524).