Published in final edited form as:

Anal Chem. 2019 September 17; 91(18): 11589–11597. doi:10.1021/acs.analchem.9b01698.

Dual-Functional Titanium (IV) Immobilized Metal Affinity Chromatography Enables Large-scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Reveals Its Predominant Substrates

Junfeng Huang¹, Jing Dong², Xudong Shi³, Zhengwei Chen⁴, Yusi Cui⁴, Xiaoyan Liu², Mingliang Ye^{2,*}, Lingjun Li^{1,4,5,*}

¹School of Pharmacy, University of Wisconsin, Madison, WI 53705, USA

²CAS Key Laboratory of Separation Sciences for Analytical Chemistry, National Chromatographic R&A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences (CAS), Dalian, China

³Department of Surgery, University of Wisconsin-Madison, Madison, WI 53705, USA

⁴Department of Chemistry, University of Wisconsin, Madison, WI 53705, USA

⁵School of Life Sciences, Tianjin University, Tianjin 300072, China

Abstract

Mannose-6-phosphate (M6P) glycosylation is an important post-translational modification (PTM) and plays a crucial role in transferring lysosomal hydrolases to lysosome and is involved in several other biological processes. Aberrant M6P modifications have been implicated in lysosomal storage diseases and numerous other disorders including Alzheimer's disease and cancer. Research on profiling of intact M6P glycopeptides remains to be challenging due to its extremely low stoichiometry. Here we propose a dual-mode affinity approach to enrich M6P glycopeptides by dual-functional titanium (IV) immobilized metal affinity chromatography (Ti(IV)-IMAC) materials. In combination with state-of-the-art mass spectrometry and database search engine, we profiled 237 intact M6P glycopeptides corresponding to 81 M6P glycoproteins in five types of tissues in mouse, representing the first large-scale profiling of M6P glycosylation in mouse samples. The analysis of M6P glycoforms revealed the predominant glycan substrates of this PTM. Gene ontology analysis showed that overrepresented M6P glycoproteins were lysosomal associated proteins. However, there were still substantial M6P glycoproteins that possessed different subcellular locations and molecular functions. Deep mining of their roles implicated in

*Corresponding author. M.Y. (mingliang@dicp.ac.cn) or L.L. (lingjun.li@wisc.edu).

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website: Sample preparation and protein digestion; LC-MS/MS analysis method; figure of investigation of influence of sample amount on identification results; figures of phosphopeptides and M6P glycopeptides in mouse liver, heart, kidney, and brain; HCD spectrum of Hspd1 M6P glycopeptide; molecular functions and biological processes analysis of M6P glycoproteins; data availability information; tables of identified M6P Glycopeptides and M6P Glycoproteins.

The authors declare no conflict of interest.

lysosomal and non-lysosomal function can provide new insights into functional roles of this important yet poorly studied modification.

Graphical Abstract

Keywords

titanium (IV) immobilized metal affinity chromatography; hydrophilic chromatography; dual-functional materials; Mannose-6-phosphate (M6P) glycopeptide; posttranslational modifications

Introduction

Glycosylation as a diverse protein post-translational modification (PTM) plays crucial roles in various biological processes of living organisms.^{1,2} It is well demonstrated that functions of proteins are frequently modulated by different types of glycans added to either asparagine (N-linked) or serine/threonine (O-linked) residues.³⁻⁵ Aberrant glycosylation is also involved in numerous diseases.⁶⁻⁸ Interestingly, glycan itself can also be modified by a variety of functional groups including phosphorylation, sulfation, acetylation and methylation at specific hydroxyl groups within the glycan chain, which may exert significant influences on the biology of glycoproteins possessing the glycans, while research on the modified glycans is largely underexplored due to numerous technical difficulties.⁹ Among them, phosphorylation of glycans occurring at the C-6 position of mannose in high-mannose N-glycan structures, which is called the mannose-6-phosphate (M6P) residues, is one of the most studied glycan modifications.^{10,11} M6P glycoproteins are mostly localized in lysosome and defined as the lysosomal hydrolase. They are biosynthesized in Golgi apparatus through a two-step enzymatic process.¹² Initially, a GlcNAc-1-phosphate group is transferred by UDP-acetylglucosamine 1-phosphotransferase to the selected terminal mannose residues in high-mannose type glycans from UDP-GlcNAc. The second step is performed by removal of the terminal GlcNAc group by an N-acetylglucosamidase. The newly generated M6P served as the recognition signal which allows the transportation of lysosomal hydrolases from the Golgi apparatus to the lysosome through binding with the M6P receptors. The whole process is called the M6P-dependent pathway, which is involved in cellular metabolic processes and its abnormality results in development of lysosomal storage diseases.^{13,14} Besides binding and transporting M6P enzymes to lysosomes, M6P receptors are also found to be anchored to the cell surface membrane, which has been implicated in the internalization of M6P bearing compounds and modulates the activity of a variety of non-lysosomal enzyme M6P bearing

glycoproteins including leukemia inhibitory factor (LIF), latent transforming growth factor- β (LTGF β) precursor, urokinase-type plasminogen activator receptor, glycoprotein D of the herpes virus, granzyme B, and proliferin.^{15,16} In addition, researchers have proposed that M6P glycoproteins may be associated with a variety of diseases, including tumorigenesis, neurodegenerative diseases, atherosclerosis, and arthritis¹⁷⁻²⁴. Significant efforts are underway to identify M6P modifications and link them with specific biological functions. Comprehensive and site-specific profiling of M6P modification at the proteome level would provide important insights into its biological function and could be translated into potential biomarkers or therapeutic targets for these diseases.^{13,23}

To date, direct analysis of M6P glycopeptides from a complex sample by the commonly used liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is challenged by their lower abundance and poor ionization efficiency compared to non-glycopeptides.²⁵⁻²⁷ Hydrophilic interaction chromatography (HILIC) is an excellent chromatographic technique that can efficiently enrich glycopeptides from biological samples by utilizing the hydrophilic/polar/hydrogen interactions between the glycan moieties in glycopeptides and the stationary phases, which would decrease the sample complexity and facilitate their identifications by LC-MS/MS.^{28,29} While other than the negatively charged phosphate that the M6P glycopeptides possess, they share similar physical and chemical properties with common glycopeptides, which means that analysis of low abundance M6P glycopeptides after HILIC enrichment is challenged by the interference from high abundance common glycopeptides. Anti-M6P antibody or M6P receptor immunoprecipitation methods have been developed to isolate M6P glycoproteins.^{30,31} However, non-specific binding could compromise the enrichment efficiency of the immunoprecipitation strategy. Moreover, the anti-M6P antibodies or M6P receptors currently are either not commercially available or are very expensive, which make them less accessible to most academic laboratories. Immobilized metal affinity chromatography (IMAC) has been developed as a technique for enrichment of phosphopeptides from the biological samples, which is based on the electrostatic interaction of negatively charged phosphate group on peptides with the positively charged transition metal ions (e.g., Fe³⁺, Al³⁺, Co²⁺, Ga³⁺, Ni²⁺, Zn²⁺, Zr³⁺, and Ti⁴⁺) immobilized on a resin beads or a column.^{32,33} This interaction makes enrichment of M6P glycopeptides from complex peptide samples possible as they also possess phosphate group and could be separated from common glycopeptides. Previously, our group demonstrated that the enrichment of phosphopeptides is accompanied with the detection of M6P glycopeptides with IMAC enrichment.³⁴ Later, Caval et al. used similar IMAC method to enrich M6P glycopeptides and then performed a targeted HCDpdETcD strategy that only the HCD acquisition with a M6P (243 Da) oxonium ion can trigger the ETcD fragmentation to improve the coverage of M6P peptides.³⁵ However, it should be emphasized that, in those mono-mode affinity enrichment, substantial phosphopeptides coeluted in the enrichment step can cause interference and suppression of the M6P glycopeptide signals in subsequent LC-MS/MS analysis. Therefore, development of new enrichment strategy to specifically isolate the M6P glycopeptides from complex biological samples and simultaneously eliminate the interference of common glycopeptides or phosphopeptides would be highly desirable and beneficial for the global mapping of this unique and important modification.

In view of this technological challenge, we develop a dual-mode affinity enrichment approach that employs a dual-functional Ti(IV)-IMAC material for specific enrichment of M6P glycopeptides. The Ti(IV)-IMAC material we developed possesses substantial phosphate chelated Ti(IV) ions, which can be used to enrich phosphopeptides through electrostatic interaction. Moreover, it also contains numerous hydroxyl, amine and phosphate groups on the material surface,³⁶⁻³⁸ which is highly hydrophilic and enable enrichment of glycopeptides through hydrophilic interactions.^{39,40} As M6P glycopeptides possess features of both phosphopeptides and glycopeptides, they could be captured by the dual-functional Ti(IV)-IMAC material through the synergistic electrostatic and hydrophilic interactions. Therefore, by performing HILIC-mode stepwise elution, M6P glycopeptides can be separated from phosphopeptides (Figure 1). We demonstrate the utility of this novel approach by conducting the first large-scale global profiling of M6P glycosylation in multiple mouse tissues and revealing its predominant substrates.

Experimental section

Typical Ti(IV)-IMAC enrichment and elution.

Sample preparation and protein digestion details are provided in Supporting Information. Ti(IV)-IMAC adsorbents were home-made and the enrichment protocol was adapted from our previous protocol.^{38,41,42} Briefly, the tryptic peptide mixture was added in the loading buffer (80% acetonitrile (ACN) (v/v) and 6% trifluoroacetic acid (TFA) (v/v)) with 1:1 ratio (v/v) and then they were incubated with Ti(IV)-IMAC adsorbents with a ratio of 1:10 (w/w) for 30 min. Then the non-specific adsorbed peptides on the surface of adsorbents were washed sequentially by washing buffer 1 (50% ACN (v/v) 6% TFA (v/v) and 200 mM NaCl) and washing buffer 2 (30% ACN (v/v) and 0.1% TFA (v/v)). Finally, the enriched phosphopeptides and M6P glycopeptides were eluted from the adsorbents with 10% NH₄OH (v/v). After centrifugation at 20,000 x g for 5 min, the supernatant was collected and immediately dried down in SpeedVac. Samples were stored at -20 °C and re-dissolved in 0.1% formic acid (FA) before LC-MS/MS analysis.

HILIC-mode Ti(IV)-IMAC elution.

For HILIC-mode Ti(IV)-IMAC elution, the sample enrichment is the same with typical Ti(IV)-IMAC enrichment described above. After enrichment, the elution was performed sequentially with the following six elution buffers: 80% ACN (v/v) and 5% NH₄OH (v/v), 60% ACN (v/v) and 10% NH₄OH (v/v), 40% ACN (v/v) and 10% NH₄OH (v/v), 20% ACN (v/v) and 10% NH₄OH (v/v), 10% ACN (v/v) and 10% NH₄OH (v/v), and 10% NH₄OH (v/v). After centrifugation at 20,000 x g for 5 min, each of the supernatant was collected and immediately dried down in SpeedVac. Samples were stored at -20 °C and re-dissolved in 0.1% formic acid (FA) before LC-MS/MS analysis (LC-MS/MS analysis details are provided in Supporting Information).

Data processing.

Elucidating intact M6P glycopeptides is even more challenging than analyzing common glycopeptides due to their micro- (different glycan structures at one glycosylation site) and macro-heterogeneity (site occupancy) and additional difficulty that most of the current

glycopeptide identification software packages cannot recognize this unique modification because of the added phosphate group on M6P glycans. Here, MS/MS spectra of intact M6P glycopeptides were analyzed using Byonic software (Protein Metrics, San Carlos, CA), which allowed manually adding M6P glycans in the glycan database for searching. Raw files were searched against *Mus musculus* protein database of reviewed (Swiss-Prot) sequences downloaded from Uniprot. Precursor ion mass tolerance of 10 ppm and fragment ion mass tolerance of 0.01 Da were selected. The phosphorylation of serine (S), threonine (T) and tyrosine (Y) and the oxidation of methionine (M) were set as variable modifications. Meanwhile, the carbamidomethylation of cysteine (C) was set as fixed modification. N-linked glycan searching used a common mammalian N-glycome. In addition, M6P glycopeptides were searched by expanding the glycan database to include N-linked M6P glycans consisting of HexNAc (2–4)-Hex (5–9)-Phospho (1–2) modification. Moreover, to determine if low mannose type M6P glycopeptides are in existence, we also added HexNAc (2) Hex (3–4) Phospho (1) and HexNAc (3) Hex (3–4) Phospho (1) in the glycan database. Peptide identifications were filtered at 1% two-dimensional false discovery rate (2D FDR) calculated by Byonic and Byonic Score>50. Manual inspection of MS/MS spectra of Ti(IV)-IMAC enriched samples was performed to examine if Byonic identification results contain phosphorylated hexose diagnostic ions.³⁴ All searches allowed for a maximum total of two common modifications. The identified M6P proteins were sent to DAVID for Gene Ontology (GO) analysis. With the same raw files and database, phosphopeptides were searched against MaxQuant software (Version, 1.5.8.3). Precursor ion mass tolerance of 4.5 ppm and fragment ion mass tolerance of 0.05 Da were selected. The phosphorylation of serine (S), threonine (T) and tyrosine (Y) and the oxidation of methionine (M) were set as variable modifications. Meanwhile, the carbamidomethylation of cysteine (C) was set as fixed modification. Phosphopeptides with the false discovery rate <0.01 and minimum score of 40 were accepted as confident identifications.

Results and Discussion

Conventional mono-mode IMAC approach for M6P glycopeptide enrichment.

We initially employed conventional IMAC enrichment approach to investigate its performance for analysis of M6P glycopeptides.^{34,38} Tryptic digests from five types of mouse tissues were subjected to conventional IMAC enrichment where samples were loaded in 40% ACN/6% TFA and eluted by 10% NH₄OH. The resulting MS/MS spectra of intact M6P glycopeptides were analyzed by Byonic software, which allowed manually adding M6P glycans in the glycan database for searching. Peptide identifications were filtered at 1% false discovery rate and 50 Byonic score and were further inspected manually if they contain the phosphorylated hexose diagnostic ions (*m/z* 243.0269).³⁴ It was found, as shown in Figures 2A-B, the conventional IMAC enrichment protocol can enrich thousands of phosphopeptides in all of five mouse tissues; however, only 1, 33, 22, 6, and 29 M6P glycopeptides were identified from 25 µg of mouse liver, heart, lung, kidney, and brain protein digests, respectively. We also increased sample amount to see if M6P glycopeptide identification results can be improved. The results showed that neither identified phosphopeptides nor M6P glycopeptides showed significant improvements when using 50 or

100 µg mouse brain protein digests compared with the results obtained using 25 µg sample amount (Figure S1).

New dual-mode affinity approach for M6P glycopeptide enrichment.

To address limitations to the conventional mono-mode IMAC approach, we present a new dual-mode affinity enrichment approach by exploiting the dual-functional property of Ti(IV)-IMAC material (Figure 1). Ti(IV)-IMAC material possesses substantial phosphate chelated Ti(IV) ions, which is used to enrich phosphopeptides through electrostatic interaction. Moreover, it also contains lots of hydroxyl, amine and phosphate groups on the material surface, which is highly hydrophilic and can also enrich glycopeptides through hydrophilic interactions. As M6P glycopeptides possess high-mannose glycans that have similar hydrophilic characteristics compared with other glycopeptides, we hypothesize that dual-functional Ti(IV)-IMAC material may enable specific enrichment of M6P glycopeptides through synergistic electrostatic and hydrophilic interactions. Therefore, Ti(IV)-IMAC material captured M6P glycopeptides can be separated with phosphopeptides according to their hydrophilicity differences in HILIC-mode stepwise elution.

We performed dual-mode affinity enrichment approach in five mouse tissues (Figure 2C) and use mouse lung as an example sample to validate this new method (Figures 2D-E). Two indicators that are ratios of multi-phosphopeptides and numbers of M6P glycopeptides in individual fractions were used to show the peptide retention difference on Ti(IV)-IMAC material based on their hydrophilicity. After employing discontinuous organic solvent gradient in elution buffer, the phosphopeptides and the M6P glycopeptides were well separated. In the 80% ACN and 5% NH₄OH elution buffer, 737 phosphopeptides of which only 8.3% were multi-phosphopeptides were eluted. With reduction of acetonitrile ratio to 60%, 3791 phosphopeptides with only 13.5% of them being multi-phosphopeptides were eluted. Continually decreasing the acetonitrile ratio to 40%, totally 2484 phosphopeptides and only one M6P glycopeptide was identified, while the identified multi-phosphopeptides had increased sharply with the ratio rising to 38.1%. The rising of the multi-phosphopeptide ratio was consistent with characteristics that more phosphate groups on peptides enhanced their hydrophilicity and thus increased their retention. Meanwhile, only one M6P glycopeptide was identified in the 40% acetonitrile fraction, which indicated that most of the M6P glycopeptides were still retained in the aqueous layer due to their stronger hydrophilic interaction with the material. When further decreasing the acetonitrile content to be close to the aqueous elution buffer, which contained 20%, 10%, and 0% acetonitrile, 33, 81, and 77 M6P glycopeptides were identified, respectively, while the numbers of identified phosphopeptides decreased to 1365, 2090, and 1719, accordingly. The extracted ion chromatograms (XICs) for [Hex+phospho]⁺ at m/z 243.0269 ± 0.001 have been calculated with or without HILIC-mode step elution. The results in Table S1 showed that, in the non HILIC-mode step elution sample 901 (2.5%) of the 35974 spectra have the M6P diagnostic ions, while 2016 (6%) of the 33688 spectra, 2561 (7.6%) of the 33573 spectra, and 1507 (4.5%) of the 33322 spectra have M6P diagnostic ions in the 10% NH₄OH, 10% ACN/10% NH₄OH, and 20% ACN/10% NH₄OH elution, respectively. These results demonstrated that capturing of M6P glycopeptides on dual-functional Ti(IV)-IMAC material surface can be achieved through synergistic electrostatic and hydrophilic interactions, and by performing

the HILIC-mode stepwise elution of Ti(IV)-IMAC enriched samples, the phosphopeptides and M6P glycopeptides can be sufficiently separated. The results of other tissues are shown in Supplementary Figures S2-S5 and Tables S2-S5.

We then merged the results of all fractions in each of the five tissues. As shown in Figure 3, 5396, 4396, 5960, 5750, and 8454 phosphopeptides were identified in mouse liver, heart, lung, kidney, and brain tissues respectively, and in the same dataset, 10, 96, 109, 37, and 116 M6P glycopeptides corresponding 8, 26, 40, 20, and 38 M6P glycoproteins were profiled and site specifically characterized. With HILIC mode step elution of dual-functional Ti(IV)-IMAC materials, the identified M6P glycopeptides were increased by at least 3 fold in all of the five tissues. Overall, we identified 17128 phosphorylation sites corresponding to 4583 phosphoproteins in the five mouse tissues. Furthermore, we site-specifically identified 237 M6P glycopeptides corresponding to 81 M6P glycoproteins, which represents the first large-scale site-specific profiling of M6P glycosylation in mouse tissue samples (Tables S1-S2). Figure 4 shows a representative HCD spectrum of an identified M6P glycopeptide. In accordance with previously reported fragmentation spectra of phosphorylated glycans,³⁴ we also observed abundant ions at m/z 243.0269 corresponding to HexPhospho in the MS/MS spectra of M6P glycopeptides.

Of the 81 M6P glycoproteins, three (cathepsin L (Ctsl), N-acetylglucosamine-6-sulfatase (Gns), and alpha-glucosidase (Gaa)) were commonly identified and assigned in all of the five mouse tissues, four (tripeptidyl peptidase I (Tpp1), palmitoyl-protein thioesterase 1 (Ppt1), cathepsin D (Ctsd), and putative phospholipase B-like 2 (PLBD2)) were assigned in 4 of 5 mouse tissues except mouse liver. All of them are well-known M6P-modified lysosomal proteins, which demonstrated that M6P glycosylation as a lysosomal enzyme localization signal is highly conserved in majority of the mouse tissues. Meanwhile, 30 of the identified M6P glycoproteins were assigned in as least two mouse tissues, 51 of them were only detected in one mouse tissue, including 3, 6, 15, 5 and 22 M6P glycoproteins observed in the mouse liver, heart, lung, kidney, and brain respectively, which revealed distinct tissue-specificity of the M6P glycosylation and demonstrated that M6P glycosylation may play other important roles in different tissues such as cellular differentiation and tissue formation.

To evaluate and demonstrate the reproducibility of this new method, we used mouse brain as a test sample to conduct two additional replicate analyses. The phosphopeptide and M6P glycopeptide distributions in replicates 2 and 3 showed similar trends as that obtained in the first replicate (Figures S6-7). The comparison results in Figures S8A and S8B showed that 116, 104, and 109 M6P glycopeptides corresponding to 38, 47, and 45 M6P glycoproteins were profiled and site specifically characterized in these three replicates. Venn diagram in Figures S6C and S6D demonstrated the overlap of these three replicates, totally 60 M6P glycopeptides attributed to 20 M6P glycoproteins were commonly identified in all three replicates, which showed good reproducibility of the dual-functional Ti(IV)-IMAC material for M6P glycopeptide enrichment.

Gene ontology analysis of the identified M6P glycoproteins.

As this was the first global profiling of M6P glycoproteins in mouse tissues, gene ontology (GO) annotation analysis was performed in order to obtain an overview of subcellular compartments and better functional understanding of the M6P glycoproteins. The analysis was performed by using the bioinformatics tool DAVID (Version 6.8). Figure 3E and Figure S9 showed the annotation analysis results, where the bars from top to bottom are laid out according to their P values. Unlike global N-linked glycoproteome annotation analysis in which the leading annotated glycoproteins were plasma membrane or extracellular region proteins,⁴³ identified M6P glycoproteins distributed in lysosome were significantly overrepresented which is generally consistent with the reports that M6P glycoproteins are mainly lysosomal enzymes. The most significantly enriched GO categories for molecular function was hydrolase activity, and the most significantly enriched biological process was lysosome organization.

Global glycoform analysis of the identified M6P glycoproteins.

In addition to protein level analysis, global and site-specific M6P glycoform analysis was conducted. High mannose glycan *in vitro* phosphorylation process has demonstrated that N-glycans of Man6–8GlcNAc2 are optimal substrates for GlcNAc-1-phosphotransferase, due to the α -1,2-linked mannose residue terminating these glycans, but lacking in high mannose glycans with 5 or less mannose.^{44,45} However, the *in vivo* distribution of M6P glycoforms in different mammalian tissues has not been revealed previously. As we have conducted large-scale profiling of site-specific glycoforms of M6P glycosylation in five mouse tissues, we also performed compositional analyses of the M6P glycans. The results in Figure 5A indicated that the predominant N-glycan species for mono-phosphorylation was Man6GlcNAc2 and that for di-phosphorylation was Man7GlcNAc2. This is because the typical Man6GlcNAc2 glycan mainly possesses one α -1,2-linked mannose and the Man7GlcNAc2 glycan can possess two α -1,2-linked mannoses. These results are consistent with *in vitro* phosphorylation process that terminal α -1,2-linked mannose residue is more likely to be phosphorylated compared to mannose with other linkage.

Site-specific glycoform analysis of the commonly identified M6P glycoproteins.

Besides global analysis of distribution of M6P glycoforms, knowledge about the predominant N-glycan structures at each glycosylation site is also important. Here, three proteins that were commonly identified in the five mouse tissues were analyzed for their specific N-glycan compositions at each of the glycosylation sites. Glycosylation site analysis of Cathepsin L1 (Ctsl) showed Asn 221 as the common N-glycosylation site for this glycoprotein in all of the five tissues, which showed homogeneity of this protein in different tissues (Figure 5B) and is consistent with previous reports.⁴⁶⁻⁴⁸ However, the M6P glycans themselves displayed considerable microheterogeneity, with the same site being modified to 3 different glycoforms: Man6GlcNAc2Phospho1, Man7GlcNAc2Phospho1, and Man7GlcNAc2Phospho2. Moreover, the distribution of the 3 glycoforms in the five tissues were also different. By calculating the PSMs ratio of each glycoform, we found that the preferred M6P glycoform of Ctsl in mouse brain was Man6GlcNAc2Phospho1 and it possessed 55% of all glycopeptide PSMs, while the predominant M6P glycoform of Ctsl in

liver, heart, lung, and kidney was Man7GlcNAc2Phospho2, which possessed 100%, 47.8%, 40.9%, and 55.6% of all the glycopeptide PSMs, respectively (Figure 5E). To compare this result with glycoform distribution of non-M6P high mannose glycans, we examined a previous study on global glycoproteome analysis of these five mouse tissues.⁴⁷ The glycoform distribution (PSMs ratio) analysis of Ctsl was extracted from the glycoproteome data. The result showed that glycosylated Asn 221 of Ctsl was found in liver, lung and kidney. Moreover, the major glycoforms on Asn 221 of Ctsl in liver were Man4GlcNAc2 (46.2%), Man5GlcNAc2 (20.5%), Man8GlcNAc2(18.0%), and Man6GlcNAc2 (15.4%); in kidney were Man3GlcNAc2 (43.8%), Man6GlcNAc2 (25.0%), Man4GlcNAc2 (12.5%), and Man7GlcNAc2(12.5%); and in lung were Man6GlcNAc2 (100%). These results showed that the glycoform distribution of non-M6P high mannose glycans on Asn 221 of Ctsl were also tissue-specific, while the glycoform distributions of M6P and non-M6P were vastly different.

The identified M6P glycopeptides of lysosomal alpha-glucosidase (Gaa) were attributed to three N-glycosylation sites Asn140, Asn470, and Asn883 (Figure 5C). As comparison, Asn140, 390, 470, 883, and 926 of Gaa were reported N-glycosylation sites previously.^{47,48} M6P glycans occurring on these three sites were also tissue specific. Asn470 possessed M6P glycans in all five tissues; however, Asn140 and Asn 883 were modified with M6P glycans only in mouse heart and brain. On Asn470, M6P glycans also showed microheterogeneity, which contained 4 different glycoforms: Man5GlcNAc2Phospho1, Man6GlcNAc2Phospho1, Man7GlcNAc2Phospho1, and Man7GlcNAc2Phospho2. Man5GlcNAc2Phospho1 only occurred in heart and brain and was also the least representative M6P glycans in these two tissues, supporting that this type of high mannose glycan was difficult to be phosphorylated. Man6GlcNAc2Phospho1 was the predominant M6P glycan in heart and brain (45.8% and 37.5%, respectively). Man7GlcNAc2Phospho1 was the dominating one in kidney (66.7%). Man7GlcNAc2Phospho2 predominantly occurred in liver and heart (100% and 40%, respectively) (Figure 5F). The non-M6P high mannose glycan distribution on Asn470 of Gaa was also different. Glycosylated Asn470 was found in liver and brain and the Man5GlcNAc2 and Man6GlcNAc2 were the only two glycoforms found on this site. In brain, Man6GlcNAc2 was the major glycoform, which accounted for 75%. However, in liver, these two glycoforms shared similar occurrence frequency at 48.4% and 51.6%, respectively.⁴⁷

In N-acetylglucosamine-6-sulfatase (Gns), the M6P glycans were found at Asn271, 354, and 397 (Figure 5D). It had one glycosylation site Asn354 that possessed M6P glycans in all five tissues, while Asn271 was modified with M6P glycans in lung and kidney and Asn 397 in brain. On Asn354, Man7GlcNAc2Phospho1 was predominant M6P glycan in heart, lung, kidney and brain (57.1%, 30.8%, 37.5% and 28.6%), and Man7GlcNAc2Phospho2 was the dominating one in liver (100%) (Figure 5G). As comparison, Asn103, 109, 175, 271, 354, 379, and 397 of Gns were previously reported N-glycosylation sites^{47,48}. Asn354 of Gns possessed the non-M6P high mannose glycans in liver, lung, and kidney. In liver, the major glycoforms were Man7GlcNAc2 (40.0%), Man5GlcNAc2 (25.0%), and Man6GlcNAc2 (25.0%); in lung, they were Man5GlcNAc2 (34.4%), Man7GlcNAc2 (20.3%), and Man6GlcNAc2 (15.6%); in kidney, they were Man5GlcNAc2 (31.8%), Man4GlcNAc2 (22.7%), and Man7GlcNAc2 (13.6%).

Our study represents the first global and site-specific, and tissue comparative M6P glycoform analysis in complex biological samples. We believe that the enrichment and identification strategies developed here and site-specific analysis of M6P glycoform distribution in different lysosome proteins and different tissues will provide a powerful tool for global M6P analysis. Furthermore, the knowledge gained through this analysis could also serve as instructions for process of optimizing recombinant enzyme to treat the lysosomal storage diseases (LSDs) through the enzyme replacement therapy (ERT).⁴⁹ The challenge of this therapy is related to the creation of a recombinant protein that has both therapeutic efficacy and appropriate targeting substrate for endocytosis into different tissues and transferring to the lysosome of patients. The differential glycosylation site occupancy and chain lengths of the mannose terminated N-glycans and the varying number of modified phosphate groups may alter the efficiency of the uptake of recombinant enzyme by different tissues and should be carefully optimized during their processing.⁵⁰

Conclusions

We developed a dual-mode affinity approach for the M6P glycopeptide enrichment. The synergistic electrostatic and hydrophilic interactions enabled separation and fractionation of M6P glycopeptides from complex biological samples. Overall, we site-specifically profiled 237 M6P glycopeptides corresponding to 81 M6P glycoproteins in five mouse tissues. By analyzing the M6P glycoforms, we identified the predominant glycan substrates of this PTM. GO analysis showed that overrepresented M6P glycoproteins were lysosomal associated proteins, while a substantial number of M6P glycoproteins possessed different subcellular locations, suggesting diverse functionalities of this important modification. Deep mining of their roles implicated in lysosomal and non-lysosomal functions will provide new insights into this important yet poorly studied modification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported, in part, by funds from the National Key R&D Program of China (2016YFA0501402, to MY), the National Natural Science Foundation of China (21535008, 21525524, to MY), the National Institutes of Health grants U01CA231081, RF1AG052324, and R01 DK071801 (to LL). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to LL) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. LL acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

- (1). Spiro RG Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 2002, 12, 43R–56R.
- (2). Lu H; Zhang Y; Yang P Advancements in mass spectrometry-based glycoproteomics and glycomics. *National Science Review* 2016, 3, 345–364.
- (3). Moremen KW; Tiemeyer M; Nairn AV Vertebrate protein glycosylation: diversity, synthesis and function. *Nature Reviews Molecular Cell Biology* 2012, 13, 448. [PubMed: 22722607]

- (4). Xiao H; Chen W; Smeekens JM; Wu R An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nature Communications* 2018, 9, 1692.
- (5). Chen Z; Huang J; Li L Recent advances in mass spectrometry (MS)-based glycoproteomics in complex biological samples. *TrAC, Trends Anal. Chem* 2018, 10.1016/j.trac.2018.1010.1009.
- (6). Ohtsubo K; Marth JD Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 2006, 126, 855–867. [PubMed: 16959566]
- (7). Sajic T; Liu Y; Arvaniti E; Surinova S; Williams EG; Schiess R; Hüttenhain R; Sethi A; Pan S; Brentnall TA; Chen R; Blattmann P; Friedrich B; Niméus E; Malander S; Omlin A; Gillesen S; Claassen M; Aebersold R Similarities and Differences of Blood N-Glycoproteins in Five Solid Carcinomas at Localized Clinical Stage Analyzed by SWATH-MS. *Cell Rep.* 2018, 23, 2819–2831. [PubMed: 29847809]
- (8). Chen Z; Yu Q; Hao L; Liu F; Johnson J; Tian Z; Kao WJ; Xu W; Li L Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (ETcD). *Analyst* 2018, 143, 2508–2519. [PubMed: 29687791]
- (9). Muthana SM; Campbell CT; Gildersleeve JC Modifications of Glycans: Biological Significance and Therapeutic Opportunities. *ACS Chem. Biol* 2012, 7, 31–43. [PubMed: 22195988]
- (10). Davis LG; Javadi JI; Brunngraber EG Identification of phosphoglycoproteins obtained from rat brain. *FEBS Lett.* 1976, 65, 30–34. [PubMed: 1278406]
- (11). Distler J; Hieber V; Sahagian G; Schmickel R; Jourdain GW Identification of mannose 6-phosphate in glycoproteins that inhibit the assimilation of β -galactosidase by fibroblasts. *Proc. Natl. Acad. Sci. U. S. A* 1979, 76, 4235–4239. [PubMed: 116230]
- (12). Coutinho MF; Prata MJ; Alves S Mannose-6-phosphate pathway: A review on its role in lysosomal function and dysfunction. *Mol. Genet. Metab* 2012, 105, 542–550. [PubMed: 22266136]
- (13). Tiels P; Baranova E; Piens K; De Visscher C; Pynaert G; Nerinckx W; Stout J; Fudalej F; Hulpiau P; Tännler S; Geysens S; Van Hecke A; Valevska A; Vervecken W; Remaut H; Callewaert N A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes. *Nat. Biotechnol* 2012, 30, 1225. [PubMed: 23159880]
- (14). Boustany R-MN Lysosomal storage diseases—the horizon expands. *Nature Reviews Neurology* 2013, 9, 583. [PubMed: 23938739]
- (15). Barnes J; Lim J-M; Godard A; Blanchard F; Wells L; Steet R Extensive Mannose Phosphorylation on Leukemia Inhibitory Factor (LIF) Controls Its Extracellular Levels by Multiple Mechanisms. *J. Biol. Chem* 2011, 286, 24855–24864. [PubMed: 21613225]
- (16). Agarwal V; Toshniwal P; Smith NE; Smith NM; Li B; Clemons TD; Byrne LT; Kakulas F; Wood FM; Fear M; Corry B; Swaminathan Iyer K Enhancing the efficacy of cation-independent mannose 6-phosphate receptor inhibitors by intracellular delivery. *Chem. Commun* 2016, 52, 327–330.
- (17). Du H; Grabowski GA Lysosomal acid lipase and atherosclerosis. *Curr. Opin. Lipidol* 2004, 15, 539–544. [PubMed: 15361789]
- (18). Reiner Ž; Guardamagna O; Nair D; Soran H; Hovingh K; Bertolini S; Jones S; Mori M; Calandra S; Hamilton J; Eagleton T; Ros E Lysosomal acid lipase deficiency – An under-recognized cause of dyslipidaemia and liver dysfunction. *Atherosclerosis* 2014, 235, 21–30. [PubMed: 24792990]
- (19). Dubland JA; Francis GA Lysosomal acid lipase: at the crossroads of normal and atherogenic cholesterol metabolism. *Front Cell Dev Biol* 2015, 3, 3. [PubMed: 25699256]
- (20). Wyss-Coray T Ageing, neurodegeneration and brain rejuvenation. *Nature* 2016, 539, 180. [PubMed: 27830812]
- (21). Wang Y; MacDonald RG; Thinakaran G; Kar S Insulin-Like Growth Factor-II/Cation-Independent Mannose 6-Phosphate Receptor in Neurodegenerative Diseases. *Mol. Neurobiol* 2017, 54, 2636–2658. [PubMed: 26993302]

- (22). Song E; Mayampurath A; Yu C-Y; Tang H; Mechref Y Glycoproteomics: Identifying the Glycosylation of Prostate Specific Antigen at Normal and High Isoelectric Points by LC-MS/MS. *J. Proteome Res* 2014, 13, 5570–5580. [PubMed: 25327667]
- (23). Vaillant O; Cheikh KE; Warther D; Brevet D; Maynadier M; Bouffard E; Salgues F; Jeanjean A; Puche P; Mazerolles C; Maillard P; Mongin O; Blanchard-Desce M; Raehm L; Rébillard X; Durand J-O; Gary-Bobo M; Morère A; Garcia M Mannose-6-Phosphate Receptor: A Target for Theranostics of Prostate Cancer. *Angew. Chem. Int. Ed* 2015, 54, 5952–5956.
- (24). Olszewska-Slonina D; Jung S; Matewski D; Olszewski KJ; Krzyzyska-Malinowska E; Braszkiewicz A; Kowaliszyn B Lysosomal enzymes in serum and synovial fluid in patients with osteoarthritis. *Scand. J. Clin. Lab. Invest* 2015, 75, 145–151. [PubMed: 25594799]
- (25). Stadlmann J; Taubenschmid J; Wenzel D; Gattlinger A; Dürnberger G; Dusberger F; Elling U; Mach L; Mechtler K; Penninger JM Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* 2017, 549, 538. [PubMed: 28959962]
- (26). Ruhaak LR; Xu G; Li Q; Goonatilleke E; Lebrilla CB Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem. Rev* 2018, 118, 7886–7930. [PubMed: 29553244]
- (27). Yu Q; Wang B; Chen Z; Urabe G; Glover MS; Shi X; Guo L-W; Kent KC; Li L Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass. Spectrom* 2017, 28, 1751–1764. [PubMed: 28695533]
- (28). Zhu J; Sun Z; Cheng K; Chen R; Ye M; Xu B; Sun D; Wang L; Liu J; Wang F; Zou H Comprehensive Mapping of Protein N-Glycosylation in Human Liver by Combining Hydrophilic Interaction Chromatography and Hydrazide Chemistry. *J. Proteome Res* 2014, 13, 1713–1721. [PubMed: 24495048]
- (29). Li X; Jiang J; Zhao X; Wang J; Han H; Zhao Y; Peng B; Zhong R; Ying W; Qian X N-glycoproteome analysis of the secretome of human metastatic hepatocellular carcinoma cell lines combining hydrazide chemistry, HILIC enrichment and mass spectrometry. *PLoS One* 2013, 8, e81921. [PubMed: 24324730]
- (30). Sleat DE; Wang Y; Sohar I; Lackland H; Li Y; Li H; Zheng H; Lobel P Identification and Validation of Mannose 6-Phosphate Glycoproteins in Human Plasma Reveal a Wide Range of Lysosomal and Non-lysosomal Proteins. *Mol. Cell. Proteomics* 2006, 5, 1942–1956. [PubMed: 16709564]
- (31). Sleat DE; Zheng H; Qian M; Lobel P Identification of Sites of Mannose 6-Phosphorylation on Lysosomal Proteins. *Mol. Cell. Proteomics* 2006, 5, 686–701. [PubMed: 16399764]
- (32). Huang J; Wang F; Ye M; Zou H Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications. *J. Chromatogr. A* 2014, 1372, 1–17.
- (33). Thingholm TE; Larsen MR: Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography In *Phospho-Proteomics: Methods and Protocols*; von Stechow L, Ed.; Springer New York: New York, NY, 2016; pp 123–133.
- (34). Glover MS; Yu Q; Chen Z; Shi X; Kent KC; Li L Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass spectrom* 2018, 427, 35–42.
- (35). Caval T; Zhu J; Tian W; Remmelzwaal S; Yang Z; Clausen H; Heck AJR Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification. *Mol. Cell. Proteomics* 2019, 18, 16–27. [PubMed: 30237200]
- (36). Zhou H; Ye M; Dong J; Han G; Jiang X; Wu R; Zou H Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J. Proteome Res* 2008, 7, 3957–3967. [PubMed: 18630941]
- (37). Yu Z; Han G; Sun S; Jiang X; Chen R; Wang F; Wu R; Ye M; Zou H Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. *Anal. Chim. Acta* 2009, 636, 34–41. [PubMed: 19231353]

- (38). Zhou H; Ye M; Dong J; Corradini E; Cristobal A; Heck AJR; Zou H; Mohammed S Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc* 2013, 8, 461–480. [PubMed: 23391890]
- (39). Boersema PJ; Mohammed S; Heck AJR Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal. Bioanal. Chem* 2008, 391, 151–159. [PubMed: 18264818]
- (40). Bensaddek D; Nicolas A; Lamond AI Evaluating the use of HILIC in large-scale, multi dimensional proteomics: Horses for courses? *Int. J. Mass spectrom* 2015, 391, 105–114. [PubMed: 26869852]
- (41). Huang J; Qin H; Dong J; Song C; Bian Y; Dong M; Cheng K; Wang F; Sun D; Wang L; Ye M; Zou H In Situ Sample Processing Approach (iSPA) for Comprehensive Quantitative Phosphoproteome Analysis. *J. Proteome Res* 2014, 13, 3896–3904. [PubMed: 25072903]
- (42). Liu Z; Wang F; Chen J; Zhou Y; Zou H Modulating the selectivity of affinity absorbents to multi-phosphopeptides by a competitive substitution strategy. *J. Chromatogr. A* 2016, 1461, 35–41. [PubMed: 27470094]
- (43). Zielinska DF; Gnad F; Wi niewski JR; Mann M Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 2010, 141, 897–907. [PubMed: 20510933]
- (44). Couso R; Lang L; Roberts RM; Kornfeld S Phosphorylation of the oligosaccharide of uteroferrin by UDP-GlcNAc:glycoprotein N-acetylglucosamine-1-phosphotransferases from rat liver, *Acanthamoeba castellanii*, and *Dictyostelium discoideum* requires alpha 1,2-linked mannose residues. *J. Biol. Chem* 1986, 261, 6326–6331. [PubMed: 2939075]
- (45). Ketcham CM; Kornfeld S Characterization of UDP-N-acetylglucosamine:glycoprotein N-acetylglucosamine-1-phosphotransferase from *Acanthamoeba castellanii*. *J. Biol. Chem* 1992, 267, 11654–11659. [PubMed: 1317874]
- (46). Uematsu R; Furukawa J.-i.; Nakagawa H; Shinohara Y; Deguchi K; Monde K; Nishimura S-I High Throughput Quantitative Glycomics and Glycoform-focused Proteomics of Murine Dermis and Epidermis. *Molecular & Cellular Proteomics* 2005, 4, 1977–1989. [PubMed: 16170054]
- (47). Liu MQ; Zeng WF; Fang P; Cao WQ; Liu C; Yan GQ; Zhang Y; Peng C; Wu JQ; Zhang XJ; Tu HJ; Chi H; Sun RX; Cao Y; Dong MQ; Jiang BY; Huang JM; Shen HL; Wong CCL; He SM; Yang PY pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat. Commun* 2017, 8, 438. [PubMed: 28874712]
- (48). Huang JF; Qin HQ; Sun Z; Huang G; Mao JW; Cheng K; Zhang Z; Wan H; Yao YT; Dong J; Zhu J; Wang FJ; Ye ML; Zou HF A peptide N-terminal protection strategy for comprehensive glycoproteome analysis using hydrazide chemistry based method. *Sci. Rep* 2015, 5, 10164. [PubMed: 25959593]
- (49). Liu L; Lee W-S; Doray B; Kornfeld S Engineering of GlcNAc-1-Phosphotransferase for Production of Highly Phosphorylated Lysosomal Enzymes for Enzyme Replacement Therapy. *Molecular Therapy - Methods & Clinical Development* 2017, 5, 59–65. [PubMed: 28480305]
- (50). Pierce OM; McNair GR; He X; Kajiura H; Fujiyama K; Kermode AR N-glycan structures and downstream mannose-phosphorylation of plant recombinant human alpha-L-iduronidase: toward development of enzyme replacement therapy for mucopolysaccharidosis I. *Plant Mol. Biol* 2017, 95, 593–606. [PubMed: 29119347]

Figure 1.
Scheme of IMAC and HILIC dual-mode affinity enrichment approach for the enrichment of M6P glycopeptides.

Figure 2.

Conventional mono-mode IMAC enrichment yields considerable phosphopeptide coverage (A), while less coverage of M6P glycopeptides (B); Dual-mode affinity enrichment approach enables the separation of M6P glycopeptides from phosphopeptides: (C) Workflow of large-scale profile of phosphopeptides and M6P glycopeptides in five mouse tissues; (D) Phosphopeptide identification results from mouse lung; (E) M6P glycopeptide identification results from mouse lung.

Figure 3.

The dual-mode affinity approach significantly improves the analysis coverage of both phosphopeptides (A) and M6P glycopeptides (B); Venn diagram analysis reveals tissue-specific differential expression of M6P glycopeptides (C) and M6P glycoproteins (D) in different mouse tissues; (E) Cellular component analysis reveals diverse localizations of the M6P glycoproteins, according to DAVID analysis.

Figure 4.
Representative HCD spectrum of an annotated M6P glycopeptide corresponding to
Cathepsin L identified in mouse lung.

Figure 5.

Global and site-specific M6P glycoform analysis. (A) Global analyses reveal the predominant glycan substrates for this unique PTM; The site-specific glycoforms reveal subtle differential M6P microheterogeneity among different tissues: (B, C, D) site-specific M6P glycoform annotation of the three commonly identified proteins in five mouse tissues (the number in brackets behind each glycan is the PSMs number); (E, F, G) M6P glycoform distribution (PSMs ratio) on the commonly identified glycosylation site of each protein in five mouse tissues. Glycan composition abbreviation: H = Hex/Man, N = HexNAc/GlcNAc, P = phosphate group.