Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification

Authors
Tomislav Čaval, Jing Zhu, Weihua Tian, Sanne Remmelzwaal, Zhang Yang, Henrik Clausen, and Albert J.R. Heck

Correspondence
a.j.r.heck@uu.nl

In Brief
Here a Fe$^{3+}$-IMAC-based enrichment method is presented targeting mannose-6-phosphate (M6P) modified glycopeptides enabling the detailed profiling of these in both HeLa and CHO cells. A gene engineering strategy was used to remove the Acp2/5 phosphatases, which increased the abundance and coverage of the M6P-modified glycoproteome. We demonstrate that such a gene engineering strategy can be beneficial for the expression of alpha-galactoside and other enzymes used in the treatment of lysosomal storage diseases.

Graphical Abstract

Fe$^{3+}$-IMAC

Peptide mixture

HCD

EThcD

LVQAEYWHDPEDWYRHSLAD1NQER

Graphical Abstract

Highlights
- Fe$^{3+}$-IMAC enables co-enrichment of mannose-6-phosphate (M6P) modified glycopeptides.
- Detailed characterization of intact M6P modified glycopeptides from HeLa and CHO cell lines.
- EThcD results in a cleavage between 2 core GlcNAc residues enabling unambiguous glycoform identification.
- Double knock out of the phosphatases Acp2/5 results in a 20-fold increase in abundance of M6P modified glycopeptides.
- Recombinant expression of alpha-galactoside in an Acp2/5 KO CHO cell drastically increases its M6P content.
Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification*

Tomislav Čaval‡§**, Jing Zhu‡§**, Weihua Tian¶, Sanne Remmelzwaal‡§, Zhang Yang¶, Henrik Clausen¶, and Albert J.R. Heck‡§

Mannose-6-phosphate (M6P) is a distinctive post-translational modification critical for trafficking of lysosomal acid hydrolases into the lysosome. Improper trafficking into the lysosome, and/or lack of certain hydrolases, results in a toxic accumulation of their substrates within the lysosomes. To gain insight into the enzymes destined to the lysosome these glycoproteins can be distinctively enriched and studied using their unique M6P tag. Here we demonstrate, by adapting a protocol optimized for the enrichment of phosphopeptides using Fe³⁺-IMAC chromatography, that proteome-wide M6P glycopeptides can be selectively enriched and subsequently analyzed by mass spectrometry, taking advantage of exclusive phosphomannose oxonium fragment marker ions. As proof-of-concept of this protocol, applying it to HeLa cells, we identified hundreds of M6P-modified glycopeptides on 35 M6P-modified glycoproteins. We next targeted CHO cells, either wild-type or cells deficient in Acp2 and Acp5, which are acid phosphatases targeting M6P. In the KO CHO cells we observed a 20-fold increase of the abundance of the M6P-modification on endogenous CHO glycoproteins but also on the recombinantly over-expressed lysosomal human alpha-galactosidase. We conclude that our approach could thus be of general interest for characterization of M6P glycoproteomes as well as characterization of lysosomal enzymes used as treatment in enzyme replacement therapies targeting lysosomal storage diseases. Molecular & Cellular Proteomics 18: 16–27, 2019. DOI: 10.1074/mcp.RA118.000967.

Lysosomes and lysosome associated organelles are cytosolic acidic compartments responsible for the degradation and digestion of a variety of biological macromolecules (1). Its degradative activity originates from the presence of over 60 lysosomal acid hydrolases. Most of the lysosomal enzymes are incorporated into the organelle triggered by a specific post-translational modification involving a unique mannose-6-phosphate (M6P)¹ moiety. This sugar moiety can be recognized by two distinct mannose-6-phosphate receptors in the Golgi complex, that subsequently transport their cargo toward the lysosome (2, 3).

Biosynthesis of the M6P tag starts in the rough endoplasmic reticulum (ER), where co-translational modification results in the N-glycosidic linkage of a 14-sugar glycan entity on selected asparagine residues in targeted proteins (Fig. 1.1 to 1.4). Subsequent removal of three glucose molecules by glucosidases and a terminal mannose moiety by mannosidase, produces uniform glycoproteins prepared for export from the ER, to the Golgi apparatus (Fig. 1.5 to 1.8) (4). In the cis-Golgi network, selective phosphorylation of the proteins is achieved by the recognition of a lysine surface patch by the GlcNAc-phosphotransferase (5). The enzyme initiates a two-step reaction by catalyzing the addition of a GlcNAc-1-phosphate molecule to the outermost mannose residues (Fig. 1.9). The subsequent removal of the terminal GlcNAc group occurs by the GlcNAc-1-phosphodiester-N-acetylglucosaminidase ("uncovering enzyme" (UCE)), that finally produces the M6P that is recognized by the mannose-6-phosphate receptors. In the trans-Golgi network, the mannose-6-phosphate receptors recognize and bind to the M6P sugar moiety initiating the incorporation of newly synthesized lysosomal acid hydrolases in clathrin coated vesicles for their ultimate delivery to the lysosomes. Here, the mannose-6-phosphate receptor substrates dissociate from their receptor because of the substantial drop in pH, upon which many proteins undergo additional modifications to produce a fully functional enzyme. Some of such modifications are activated by the phosphatases Acp2 and Acp5 that remove phosphate groups from lysosomal acid

From the ‡Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science4Life, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands; §Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands; ¶Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, University of Copenhagen, Faculty of Health Sciences, Norre Alle 20, DK-2200 Copenhagen N, Denmark

Received July 12, 2018, and in revised form, September 20, 2018
Published, MCP Papers in Press, September 20, 2018, DOI 10.1074/mcp.RA118.000967

© 2019 Čaval et al. Published under exclusive license by The American Society for Biochemistry and Molecular Biology, Inc.
hydrolases, triggering them to execute their function as hydrolases, activating the degradation of macromolecules (6).

Providing lysosomes with the correct array of enzymes is of utmost importance in cellular function. Absence or improper activity of one of the lysosomal acid hydrolases results in diverging phenotypes ranging from improper antigen processing because of improper cathepsin function (7–9) all the way to lysosomal storage diseases. The most prominent phenotype of lysosomal storage diseases is the accumulation of glycoproteins, unprocessed lipids, mucopolysaccharides, and combinations thereof. Two well-known examples include Fabry’s and Gaucher disease, resulting from the absence of alpha-galactosidase A and glucocerebrosidase enzymes, respectively, causing sphingolipid accumulation inside the lysosome (10, 11). Currently, the only therapeutic option for such disorders is enzyme replacement therapy where a recombinant version of the enzyme is provided. One of the critical attributes of such replacement therapies in the presence of the M6P-moity on the enzyme, ensuring proper targeting to the lysosome where they can exert their therapeutic function.

To understand the origin of lysosomal storage diseases, it is essential to obtain an overview of the wide range of proteins that are contained within the lysosome. The current most

---

1 The abbreviations used are: M6P, Mannose-6-Phosphate; PSM, peptide to spectrum match; Man, Mannose; IMAC, immobilized metal affinity chromatography; ER, endoplasmic reticulum; GlcNAC, N-acetylglucosamine; UCE, uncovering enzyme; Acp2, acid phosphatase 2; Acp5, acid phosphatase 5; MS, mass spectrometry; HCD, higher energy c-trap dissociation; EThcD, electron-transfer higher-energy collision dissociation; KO, knock out.
widely used technique in the proteomic analysis of the lysosome is the affinity purification from tissue samples using columns with immobilized receptors (12-15). In these earlier used methods, M6P-modified proteins were extracted, deglycosylated with peptide N-glycosidase F and subsequently subjected to mass spectrometry (MS) analysis. However, these approaches can result in false positive identifications because of the applied removal of the M6P-tag. A few studies focused on intact M6P glycopeptides, although they were mainly focused on the analysis of a single lysosomal protein (16, 17).

Here we use a well-known enrichment method, for phosphopeptides, based on iron immobilized metal ion affinity chromatography (Fe$^{3+}$-IMAC) that we show enables the core-enrichment of M6P-modified glycopeptides from a complex cell lysate. To distinguish the M6P-modified glycopeptides from phosphopeptides we utilize a specific signature fragment ion of M6P (m/z 243.026) (18) observed during higher-energy c-Trap dissociation (HCD). Observation of this ion in HCD triggers electron-transfer/higher-energy collision dissociation (ETThCD), which results in a more confident characterization of the exact sites of the M6P-modified glycopeptides, and provides signature glycan fragments enabling confident glycoform identification. Initial analysis of a HeLa cell lysate with our modified enrichment and targeted MS approach resulted in the identification of 35 M6P-modified glycoproteins and 46 M6P glycosites from hundreds of detected M6P glycopeptides. Next, we applied our approach to wild-type CHO cells and CHO cells carrying a gene knockout (KO) of the acid phosphatases Acp2 and Acp5, which dephosphorylate lysosomal proteins. KO of these phosphatases resulted in >20 fold increase in abundance of M6P glycopeptides, enabling a deeper coverage of the CHO M6P-glycoproteome. Finally, we demonstrate that recombinant expression of lysosomal human alpha-galactosidase in WT CHO cells and CHO cells carrying the Acp2 and Acp5 KO results in a drastic decrease of the M6P-moiety on alpha-galactosidase in the KO CHO cell line. We conclude that Fe$^{3+}$-IMAC enables facile enrichment of intact M6P glycopeptides and could be used to extend biological insight gained from standard phosphoproteomic studies, additionally, when coupled with glycoengineering strategies it can be used for characterization of improved therapeutic enzymes used for the treatment of lysosomal storage diseases.

**EXPERIMENTAL PROCEDURES**

**Hela Cell Culture**—HeLa cells were cultured in Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum (Invitrogen, Landsmeer, the Netherlands) and 0.05 mg/ml penicillin/streptomycin (Invitrogen) at 37 °C in 5% CO$_2$ in 15 cm plate. Cells were collected at 80% confluency by centrifugation and washed three times with ice-cold PBS.

**CHO Cell Culture and Generation of Acp2/5 Knockout**—A CRISPR/Cas9 based approach was used for the gene knockout (KO) in CHO cells as described previously (19). CHOZN GS−/- cells with stable expression of alpha-galactosidase were used as the parental clone (as below: wild type) for the Acp2/S KO. Cells were maintained as suspension cultures in EX-CELL CHO CD Fusion serum-free media (Sigma-Aldrich, Brøndbyvester, Denmark) in 50 ml TPP TubeSpin® Bioreactors with 180 rpm shaking speed at 37 °C and 5% CO$_2$. Cells were seeded at 0.5 × 10$^6$ cells/ml in T25 flask (NUNC, Hvidovre, Denmark) 1 day prior to transfection. Electroporation was conducted with 2 × 10$^6$ cells with a DNA mixture of 1 μg of Cas9-GFP plasmid and 1 μg of gRNA plasmid (U6GRNA, Addgene Plasmid #68370) using an Amaxa kit V and program U24 with Amaxa Nucleofector 2B (Lonza, Copenhagen, Denmark). Forty-eight hours after nucleofection the 10−15% highest GFP expression pools of cells were enriched by FACS, and after 1 week cultured cells were single-cell sorted by FACS into 96-wells. KO clones were identified by Ident Detection by Ampliplex Analysis (IDAA) as described (20). Selected clones were further verified by Sanger sequencing. Knockout was then calculated.

**Cell Lysis and Protein Digestion**—HeLa, CHO wild type and CHO KO cell pellets were re-suspended in lysis buffer containing 100 mM Tris-HCl (pH 8.5) (Sigma-Aldrich), 7 M urea (Sigma-Aldrich), 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich), 30 mM chloroacetamide (CAA, Sigma-Aldrich), Triton X-100 (1%) (Sigma-Aldrich), 2 mM magnesium sulfate (Sigma-Aldrich), Benzonase (1%) (Merck Millipore, Darmstadt, Germany), phosphoSTOP (Roche) and complete mini EDTA free (Roche, Woerden, the Netherlands). Then, cells were disrupted by sonication for 45 min (20 s on, 40 s off) using a Bioruptor Plus (Diagenode, Seraing, Belgium). Cell debris were removed by centrifugation at 14,000 rpm for 1 h at 4 °C and the supernatant was kept. Impurities were removed by methanol/chloroform protein precipitation as follows: 1 ml of supernatant was mixed with 4 ml of methanol (Sigma-Aldrich), 1 ml chloroform (Sigma-Aldrich) and 3 ml ultrapure water with thorough vortexing after each addition. The mixture was then centrifuged for 10 min at 5000 rpm at room temperature (RT). The upper layer was discarded, and 3 ml of methanol was added. After sonication and centrifugation (5000 rpm, 10 min at RT), the solvent was removed, and the precipitate was allowed to air dry. The pellet was resuspended in digestion buffer containing 100 mM Tris-HCl (pH 8.5), 0.5% Rapigest (Water, Ettten-Leur, the Netherlands), 5 mM TCEP and 30 mM CAA. Trypsin (Sigma-Aldrich) and Lys-C (Wako, Neuss, Germany) were added to a 1:50 and 1:100 ratio (w/w), respectively. Digestion was performed overnight at 37 °C. After acidic precipitation of Rapigest with 0.5% trifluoroacetic acid (TFA), peptides were cleaned by Sep-Pak C18 1cc vac cartridge, dried and stored at −80 °C until Fe$^{3+}$-IMAC enrichment.

**Fe$^{3+}$-IMAC Enrichment**—Fe$^{3+}$-IMAC was performed in technical triplicates for each biological group as described previously (21). Tryptic peptides (2 mg) were re-suspended with ice-cold buffer A and the pH was adjusted to 2.3 using 10% TFA before injection into the Fe$^{3+}$-IMAC column (Propac IMAC-10 4 × 50 mm column, ThermoFisher Scientific, Landsmeer, the Netherlands). Mobile-phase solvent A consisted of 30% acetonitrile and 0.07% TFA, and mobile-phase solvent B consisted of 0.3% NH$_4$OH in water. Loading was performed at a flow rate of 0.1 ml/min for 7 min with 0% B and nonphosphorylated or nonphosphomannosylated peptides were washed out using a flow rate of 1 ml/min for 5 min with 0% B. Phosphopeptides and phosphomannose peptides were eluted at a flow rate of 1 ml/min with 50% B for 1.5 min, followed 0.5 ml/min for 2.5 min with 50% B, and
finally held at 1 ml/min of 0% B for 9 min. Collected phosphopeptides and phosphomannose peptides were dried down using a lyophilizer.

**LC-MS/MS Analysis**—Nanoflow LC-MS/MS was performed by coupling an Agilent 1290 (Agilent Technologies, Middelburg, Netherlands) to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific, Bremen, Germany) for the analysis of peptides enriched by Fe$^{3+}$-IMAC. After reconstitution in 0.1% TFA, peptides were separated by using a 100 μm inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 μm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 μm inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μm) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% FA in water, and mobile-phase solvent B consisted of 0.1% FA in ACN. Trapping was performed at a flow rate of 5 μl/min for 5 min with 0% B and peptides were eluted using a passively split flow of 300 nl/min for 85 min with 8% to 40% B over 75 min, 40% to 100% B over 3 min, 100% B for 1 min, 100% to 0% B over 1 min, and finally held at 0% B for 10 min. Peptides were ionized using 2.0 kV spray voltage and a capillary temperature of 320 °C. The mass spectrometer was set to acquire full-scan MS spectra (375–1700 m/z) for a maximum injection time of 50 ms at a mass resolution of 60,000 and an automated gain control (AGC) target value of 4e5. The dynamic exclusion was set to 20s at 50 ms at a mass resolution of 60,000 and an automated gain control (AGC) target value of 5e4 or a maximum injection time of 75 ms with a normalized resolution of 30,000, AGC target value of 2e5 or a maximum injection time of 100 ms with activation of ETD and supplemental activation with a great overlap in identified unique phosphopeptides. In the co-enrichment of M6P-modified glycopeptides, we seem-

**RESULTS**

**Evaluation of the HeLa Cell Mannose-6-phosphate Modified Glycoproteome**—Somewhat inspired by studies (24–26) demonstrating the applicability of common phosphopeptide enrichment protocols to enrich also sialylated glycopeptides we investigated whether also M6P-modified glycopeptides could be enriched. To this end we first set out to reanalyze previously published data (22) in which three standard approaches for phosphopeptide enrichment were compared: Ti$^{4+}$-IMAC, Ti$^{4+}$-IMAC and TiO$_2$. C, Comparison of detected M6P glycopeptides in this and previous (22) study.
ingly did observe a bias, whereby the Fe$^{3+}$-IMAC based enrichment provided the highest number of identifications with 25 identified M6P-modified glycopeptides. Ti$^{4+}$ IMAC provided a slightly lower efficiency with 14 identified M6P glycopeptides, and more surprisingly, only 3 M6P glycopeptides were identified by using the TiO$_2$ based enrichment. Although in this experiment Fe$^{3+}$-IMAC and Ti$^{4+}$ IMAC show slightly different performance, we identified the same glycoforms, GlcNAc$_2$Man$_{6–7}$Phospho$_{1–2}$, albeit on a different peptide backbones indicating that the difference is likely because of stochastic nature of precursor ion selection during MS analysis. From these experiments, we decided to perform the rest of the experiments using Fe$^{3+}$-IMAC and we sought to further optimize the enrichment and analysis protocol to extend the coverage of M6P-modified glycoproteome. We hypothesized that there are two main reasons for the low numbers of M6P-modified glycopeptides. First, the MS fragmentation scheme based only on HCD is not optimal for glycopeptide identifications because of the preferential cleavage of the glycan moiety (27, 28). Second, combination of low starting material used in (22) and the inherent low abundance of M6P glycoproteins probably precluded their detection during MS analysis. To test our hypothesis, we performed an experiment where we increased the amount of starting material to 2 mg of protein and optimized the MS fragmentation scheme to combine HCD with EThcD fragmentation, which were triggered upon observation of the signature fragment ion in the HCD spectra for M6P-modified glycopeptides, namely the phosphomannose oxonium ion (243.026 m/z)(18). Using this approach, we could now identify almost a hundred of unique M6P-modified glycopeptides (supplemental Table S2) in HeLa cells corresponding to 35 M6P-modified glycoprotein and 46 M6P glycosites (Table I), which is significantly more than what we could identify from the published dataset (22) (Fig. 2C). We observed that the pool of M6P glycopeptides in frequency comprise around 2–3% of the total peptide (M6P glycopeptide and phosphopeptide) pool. Manual inspection of the spectra elucidated that the prime reason for the observed

### TABLE I

**Identified M6P-modified glycoproteins in HeLa cells**

| Uniprot ID | Protein name | M6P sites |
|------------|--------------|-----------|
| O00115     | Deoxyribonuclease-2-alpha | N$^{26,290}$ |
| O00469     | Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 | N$^{63}$ |
| O00584     | Ribonuclease T2 | N$^{106}$ |
| O00754     | Lysosomal alpha-mannosidase | N$^{367}$ |
| O14773     | Tripeptidyl-peptidase 1 | N$^{210,222}$ |
| O60911     | Cathepsin L2 | N$^{221}$ |
| O75503     | Ceroid-lipofuscinsosis neuronal protein 5 | N$^{178}$ |
| O75629     | Protein CREG1 | N$^{160,216}$ |
| P04062     | Glucosylceramidase | N$^{59}$ |
| P06280     | Alpha-galactosidase A | N$^{39,215}$ |
| P07339     | Cathepsin D | N$^{53}$ |
| P07602     | Prosaposin | N$^{32}$ |
| P07711     | Cathepsin L1 | N$^{37}$ |
| P10253     | Lysosomal alpha-glucosidase | N$^{145}$ |
| P10619     | Lysosomal protective protein | N$^{279,362,405}$ |
| P15586     | N-acetylgalactosamin-6-sulfatase | N$^{188}$ |
| P15848     | Arylsulfatase B | N$^{247,464}$ |
| P16278     | Beta-galactosidase | N$^{55}$ |
| P17405     | Sphingomyelin phosphodiesterase | N$^{151}$ |
| P20933     | N(4)-beta-N-acetylgalcosaminyl-L-asparaginase | N$^{157}$ |
| P06865     | Beta-hexosaminidase subunit alpha | N$^{145}$ |
| P42785     | Lysosomal Pro-X carboxypeptidase | N$^{197,212}$ |
| P50897     | Palmitoyl-protein thioesterase 1 | N$^{197}$ |
| P53634     | Dipeptidyl peptidase 1 | N$^{33,119,276}$ |
| P61916     | Epididymal secretory protein E1 | N$^{88}$ |
| Q02083     | N-acylethanolamine-hydrolyzing acid amidase | N$^{34,22}$ |
| Q02809     | Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 | N$^{17}$ |
| Q13510     | Acid ceramidase | N$^{157}$ |
| Q8NHFP8    | Putative phospholipase B-like 2 | N$^{197}$ |
| Q92484     | Acid sphingomyelinase-like phosphodiesterase 3a | N$^{238}$ |
| Q92820     | Gamma-glutamyl hydrolase | N$^{233}$ |
| Q99519     | Sialidase-1 | N$^{382}$ |
| Q99538     | Legumain | N$^{57}$ |
| Q9H3G5     | Probable serine carboxypeptidase CPVL | N$^{146}$ |
| Q9UBR2     | Cathepsin Z | N$^{184}$ |

---

**Targeted Analysis of Lysosomal Directed Proteins**

---

**Molecular & Cellular Proteomics 18.1**

---

Downloaded from https://www.mcponline.org at Universiteitsbibliotheek Utrecht on February 10, 2020
increase in the number of M6P-modified glycopeptide identifications is because of the inclusion of EThcD as fragmentation technique. Additionally, we observed a cleavage between the first and second GlcNAc residues of M6P glycan, albeit only during EThcD fragmentation, which allows unambiguous assignment of the glycan form attached to the peptide backbone (Fig. 3B and supplemental Fig. S1). This combined with the observation that the M6P modification resulted in an increase in miss cleavage rates, resulting in larger peptides carrying more charges makes EThcD fragmentation attractive for characterization of M6P modified glycopeptides (29).

However, as demonstrated in Fig. 3A and supplemental Fig. S1, it is also possible to successfully identify M6P modified glycopeptides with HCD fragmentation alone, but greater care must be devoted to manual validation as a large fraction of the identifications is based solely on precursor mass.

Inspection of all the M6P-modified glycopeptide spectrum matches revealed that in HeLa cells each of the glycosites is predominantly modified with a GlcNAc$_2$Man$_n$Phospho$_2$ (Man7 PP) glycoform, but in total we did observe at least 7 different glycan compositions ranging from Man3 P to Man8 PP (Fig 4A). However, we did not observe any glycans corresponding to hybrid N-glycans harboring the M6P tag. Considering that hybrid N-glycans are the lowest abundant class of N-glycans it is most likely that glycopeptides harboring hybrid M6P glycans are simply too low abundant to be detected. We then further checked the specificity of our enrichment and counted the N-glycan signature oxonium ions (m/z 204). In total, we detected over 2000 scans containing both the 204 and 243 oxonium ions specific for N-glycans in general, and M6P modified glycans, respectively (supplemental Table S3). When compared with scans that contained only 204 ion and no 243 ion, around 200 scans, we calculated that over 90% of all glycopeptide scans are in fact originating from M6P-harboring glycopeptides. Our findings support in vitro studies, which revealed that the Man7 PP glycoform exhibits the highest binding affinity toward mannose-6-phosphate receptors (30, 31). We also further validated our results by performing a pathway enrichment analysis (32), which showed, as expected, significant enrichment for lysosomal proteins (Fig. 4B). Additionally, the Cell Atlas (33) was used to validate protein hits not previously reported to be lysosomal proteins, such as CREG1 protein, which verified that it can be a bone fide lysosomal protein.

Double KO of Acp2 and Acp5 Increases CHO M6P Glycoproteome Coverage—The Chinese hamster ovary cell line
(CHO) is often used in biological and medical research and commercially in the production of therapeutic proteins (34, 35), including enzymes. It has been used to produce recombinant lysosomal human acid alpha-galactosidase for the treatment of the lysosomal storage diseases: Fabry’s disease (36, 37). Therefore, we were interested to apply our method to reveal the M6P landscape in CHO cells. We used a CHO cell line with a stable expression of recombinant human acid alpha-galactosidase (referred to as CHO WT) and subjected it to the same protocol as described above for the HeLa cells. To our surprise, starting with similar amounts of input we observed 5-fold increase in scans containing the signature M6P ion (supplemental Table S3). We also identified a couple in these CHO cells.

To test this hypothesis and motivated by previous studies (15, 38) demonstrating increased selectivity of M6P affinity purification in Acp5 or Acp2/5 deficient mice, we created a CHO cell line with a double knock-out (KO) of the Acp2 and Acp5 phosphatases, responsible for the dephosphorylation of M6P-modified glycoproteins upon their entry into the lysosome. As expected, when using a same input and identical protocol, but now applied to the CHO Acp2/5 KO cell line, we were able to double the number of identified M6P-modified glycoproteins (28 in CHO KO versus 15 in CHO WT) and we observed 5-fold increase in scans containing the signature M6P ion (supplemental Table S3). We also identified a couple of hundreds of M6P-modified glycopeptide PSMs mapping to 160 unique M6P glycopeptides in the Acp2/5 KO cells, which represents a 4-fold increase in M6P glycopeptide PSMs corresponding to 42 unique M6P glycopeptides detected in CHO WT (Fig. 5A and supplemental Table S4). Furthermore, even for M6P-modified proteins observed in both WT and Acp2/5 KO CHO cells, such as Cathepsin Z (Fig. 5B), we observed an increase in glycoform coverage from 4 distinct glycoforms identified in CHO WT cell to 8 glycoforms identified in CHO Acp2/5 KO. To confirm that the increase of our identifications is because of the increase in M6P modifications itself and not the up-regulation of M6P modified glycoproteins we compared the intensities of unmodified peptides and M6P-modified peptides in both cell lines. We observed that the abundance of the unmodified tryptic peptides of Cathepsin Z remains alike in both CHO WT and CHO Acp2/5 samples (Fig. 5C and 5D), whereas the M6P modified peptides from the same protein (Fig. 5C and 5D) showed ~20-fold increase in abundance in the CHO Acp2/5 KO cells, demonstrating that Acp2/5 KO results in increased abundance of the M6P-modifications. A similar trend was also observed when comparing the entire ion chromatogram traces between CHO WT and CHO Acp2/5 experiments (supplemental Fig. S2). Interestingly, we also observed a GlcNac2Man3Phospho (Man3 P) modification on multiple proteins from the CHO Acp2/5 KO cell line. A representative spectrum is shown in Fig. 5E. Although we cannot pinpoint the exact mannose residue the phosphate is attached to, the ion signal observed at m/z 1051.45 (S/N > 5) indicates that it is likely attached to either an α-3 or α-6 mannose representing a previously undescribed M6P modification. However, this phosphate could also be located on an elongated mannose branch, which would be more in line with the known biosynthetic pathways. It remains to be answered whether these more processed high mannose M6P structures originate during biosynthesis or are products of mannosidase action inside the lysosomes.

When we compared the characteristics of the M6P-modified glycoproteome of the HeLa and CHO cells we observed a very similar distributions of glycoforms, with possibly one notable exception being the absence of the GlcNac2 Man3Phospho (Man8 P) glycoform in CHO cells, for which
we have now no plausible explanation. Notably, we detected the protein CREG1 also in CHO cells as a M6P-modified glycoprotein, as we did in the HeLa cells, further providing credence that this is a "novel" genuine lysosomal protein.

**Expression of Alpha-galactosidase in CHO Acp2/5 KO for Improved Lysosomal Targeting**—Encouraged by the observed increase in frequency and abundance of M6P-modifications in the CHO Acp2/5 KO cells we hypothesized that the same could hold for a human protein recombinantly expressed in this CHO KO cell line. Therefore, we overexpressed human acid alpha-galactosidase, which is a 47 kDa N-glycosylated lysosomal hydrolase involved in sphingolipid metabolism. Deficiency of this alpha-galactosidase enzyme results in a lysosomal storage disorder known as Fabry disease and the current treatment option relies on enzyme replacement therapy with recombinant human alpha-galactosidase. The presence of the M6P modification on recombinant human alpha-galactosidase used for therapy is critical for its efficacy and targeting to the lysosomes.

Analysis of the recombinant human alpha-galactosidase enzyme expressed in CHO WT cell resulted in around 70 PSMs, while when it was expressed in the CHO Acp2/5 KO cell line the number of PSMs increased to over 300 PSMs (Fig. 6A). Next, the M6P glycoform analysis revealed a higher than expected heterogeneity, whereby we could observe 6 dominant glycoforms (GlcNAc2Man5– 8 Phospho1–2). This was in notable contrast with the glycoform distributions we observed in both the CHO and HeLa M6P glycoproteome analyses, which were dominated by the Man7 PP glycoform. Interestingly, we also noticed that in both sets of CHO cells we could observe the human alpha-galactosidase enzyme with a GlcNAc2Man8 Phospho1 (Man8 P) glycoform, which was absent from all other endogenous CHO M6P glycoproteins detected in CHO cells.

Finally, we also wanted to confirm that this increase is because of the increased abundance of the M6P modifications rather than up-regulation of the expressed protein. To this end we extracted ion chromatograms for both unmodified and M6P-modified tryptic peptides of human alpha-galactosidase (Fig. 6B). We observed that in the CHO Acp2/5 KO cells the abundance of the M6P-modified peptides was 20-fold higher, whereas the abundance of unmodified peptides showed very little variation. This suggest that glycoengineering efforts could represent a promising future avenue for

---

**Table II**

| Uniprot ID   | Protein name                  | M6P sites   |
|--------------|-------------------------------|-------------|
| A0A061I2K0   | Beta-glucuronidase            | N423,598    |
| A0A061I4G4   | Pro-cathepsin H-like protein   | N70,99      |
| A0A061I7D9   | Beta-galactosidase            | N262        |
| A0A061I8I4   | Cathepsin F                  | N350        |
| A0A061IU18   | Gamma-interferon-inducible lysosomal thiol reductase-like protein | N133 |
| G3GV64       | Mammalian epidymel-related protein 1a | N147,199   |
| G3GW3D3      | Alpha-mannosidase            | N190,629    |
| G3GZB2       | Acid ceramidase              | N342        |
| G3H2P3       | Beta-galactosidase (Fragment) | N229        |
| G3H8V5       | Carboxypeptidasea            | N140        |
| G3HCX3       | Deoxyribonuclease-2-alphaa    | N72,89      |
| G3HE67       | Protein CREG1                | N16         |
| G3HM6        | N(4)-(Beta-N-acetylglucosaminyl)-L-asparaginase | N39      |
| G3HK9        | Group XV phospholipase A2    | N197,212    |
| G3HN9        | Palmitoyl-protein thioesterase 1a | N159,319   |
| G3HQY6       | Lipase                       | N134,464,879|
| G3HTE5       | Lysosomal alpha-glucosidasea  | N284,293    |
| G3HZE3       | Sialidase-1                  | N268,277    |
| G3I1H5       | Legumain                     | N204        |
| G3I2I6       | N-acetyllactosaminide beta-1,3-N-acetylglycosaminyltransferasea | N60,110    |
| G3I4D4       | Ribonuclease T2a              | N134,259    |
| G3I4W7       | Cathepsin Da                 | N84         |
| G3I6T1       | Phospholipase B-like         | N201,284,327|
| G3I8P7       | N-acetylglicosamine-6-sulfatasea | N55      |
| G3I9G6       | CMP-N-acetylmuraminate-beta-galactosamidalpha-2, 3-sialyltransferaseb | N10      |
| G3IMH4       | Sphingomyelin phosphodiesterasea | N525,438   |
| G3IN6        | Dipetidy1-peptidase 2        | N22,221     |
| G3INC5       | Cathepsin L1a                | N139,192,215|
| P06280       | Alpha-galactosidase Aa        | N36         |
| Q8EEP7       | Cathepsin Z2a                | N156        |

aProteins also detected in CHO WT sample. bProteins detected only in CHO WT sample.

---

**Targeted Analysis of Lysosomal Directed Proteins**

Molecular & Cellular Proteomics 18.1
creation of improved biotherapeutics for enzyme replacement therapy.

**DISCUSSION**

In this study we demonstrated that Fe\(^{3+}\)-IMAC enrichment typically used for phosphopeptides, can also be used to enrich in a single step M6P-modified glycopeptides. This enrichment combined with optimized LC MS/MS fragmentation schemes combining HCD and EThcD, triggered by specific M6P-glycopeptide oxonium fragment ions, enabled the identification of hundreds of M6P-modified glycopeptides in HeLa and CHO cells. When compared with other studies aimed at profiling the M6P glycoproteome (14, 15, 39) that included a deglycosylation step and consequently lost all information on the actual glycan compositions that bear the M6P tag, our method provides deeper insights into the complexity of the glycan forms attached to the M6P glycopeptides. However, deeper insight comes with a price of lower protein IDs, as we detected 30–50% less M6P glycoproteins than described in above mentioned studies. A plausible cause of this observed decrease in M6P glycoprotein identifications may be the low ionization efficiency of glycosylated peptides (40) that is further exacerbated by the presence of multiple glycoforms on a single glycosite splitting the glycopeptide.
signal into multiple lower abundant signals. This was mainly best demonstrated in our comparison of CHO WT versus CHO Acp2/5 KO where we could detect twice as many M6P glycoproteins and observe a 700% increase in PSMs in the CHO Acp2/5 cells, demonstrating that the main challenge in identification of intact M6P glycopeptides is their low abundance (supplemental Fig. S2). However, we feel that this tradeoff in M6P glycoprotein identifications is well compensated by the additional information obtained on glycan structures attached to each modified site that will enable further studies of glycoprotein functions. Another challenge in intact glycopeptide profiling of complex mixtures is the high proportion of false positive identifications stemming from the inability to control the FDR for both the peptide and glycan moieties simultaneously (41). To address this issue, we relied on the known biosynthetic pathways that describes possible M6P glycan compositions (Fig. 1), presence of signature oxonium ions, improved sequence coverage and unambiguous glycoform identification obtained with EThcD method, and known subcellular location of M6P modified proteins, which all together enabled us to conclude that we achieve here confident identification of M6P modified glycopeptides.

Additionally, we demonstrated a new gene editing strategy involving the knock-out of Acp2 and Acp5 phosphatases in CHO cells that results in an ~20-fold increase in the abundance of M6P-modified peptides enabling us to gain a deeper insight into the lysosomal M6P glycoproteome. Finally, we also established that the CHO Acp2/5 KO cells can be used to produce improved glycoengineered lysosomal enzymes, carrying a higher degree of required M6P modifications, which could be potentially beneficial for lysosomal enzyme replacement therapies. However, it must be noted that in this work we have focused on characterization of lysosomal human alpha-galactosidase, whereas the enzyme used for the treatment Fabry’s disease is secreted human alpha-galactosidase.

The main difference is that secreted human alpha-galactosidase does not traffic to the lysosome where membrane bound Acp2 and soluble Acp5 phosphatases reside (38) and exhibit their dephosphorylative function, but it is known that Acp5 can be actively secreted (42). Further studies will be necessary to elucidate the effects of its’ KO on M6P levels of secreted human alpha-galactosidase.

In conclusion, the straightforward method described in this paper will enable further characterization of M6P glycoproteomes and provide further biological insight in the areas where lysosomal proteins have critical functions such as modulation of immune response during infection (43). Additionally, this method is also applicable for the characterization and quality control of biological therapeutics wherein M6P is a critical quality attribute of the product, and in combination with genetic glycoengineering strategies could pave the way for better biological products.

DATA AVAILABILITY

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE (23) partner repository with the data set identifier PXD010333 and 10.6019/PXD010333. For review purposes, review account has been made available. As PRIDE does not fully support glycopeptide data Byonic results, in the form of freely available Byonic Preview files, containing all identified and assigned spectra are available at the following link: https://figshare.com/s/ce1596a84a2fe0927b1e.

* The Netherlands Organization for Scientific Research (NWO) supported this research through funding of the large-scale proteomics facility Proteins@Work (project 184.032.201) embedded in the Netherlands Proteomics Centre. T. C and A.J.R.H are further supported by the NWO Gravitation program Institute for Chemical Immunology. A.J.R.H. acknowledge further support by the NWO TOP-Punt Grant 718.015.003. This work was further supported by The Lundbeck
Targeted Analysis of Lysosomal Directed Proteins

Foundation, The Novo Nordisk Foundation, and The Danish National Research Foundation (DNRF107). We acknowledge additional funding through the European Union’s Horizon 2020 research and innovation program under grant agreement 686547 (MSMed). J.Z. acknowledges support from the Chinese Scholarship Council (CSC).

This article contains supplemental material.

To whom correspondence should be addressed: Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science4Life, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands. E-mail: a.j.r.heck@uu.nl.

** These authors contributed equally to this work.

Author contributions: T.C. and A.J.R.H. designed research; T.C., J.Z., and S.R. performed research; T.C. and J.Z. analyzed data; T.C., J.Z., H.C., and A.J.R.H. wrote the paper; W.T., Z.Y., and H.C. contributed new reagents/analytic tools.

REFERENCES

1. Salti, P., and Klumperman, J. (2009) Lysosome biogenesis and lysosomal membrane proteins: Trafficking meets function. Nat. Rev. Mol. Cell Biol. 10, 623–635

2. Dahms, N. M., Lobel, P., and Kornfeld, S. (1989) Mannose 6-phosphate receptors and lysosomal enzyme targeting. J. Biol. Chem. 264, 12115–12118

3. Ghosh, P., Dahms, N. M., and Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twist in the tale. Nat. Rev. Mol. Cell Biol. 4, 202–212

4. Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Sadler, J. E. (1995) Molecular Cell Biology. W. H. Freeman, New York.

5. Dahms, N. M., Olson, L. J., and Kim, J.-J. P. (2008) Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. Glycobiology 18, 664–678

6. Makrypidi, G., Damme, M., Mülter-Loennies, S., Trusch, M., Schmidt, B., Schlueter, H., Heeren, J., Lübke, T., Salti, P., and Braulke, T. (2012) Mannose 6 dephosphorylation of lysosomal proteins mediated by acid phosphatases Acp2 and Acp3. Mol. Cell. Biol. 32, 774–82

7. Blum, J. S., Wearsch, P. A., and Cresswell, P. (2013) Pathways of Antigen Processing. Annu. Rev. Immunol. 31, 443–473

8. Turk, V., Turk, B., Günçar, G., Turk, D., and Kos, J. (2002) Lysosomal cathepsins: Structure, role in antigen processing and presentation, and cancer. Adv. Enzyme Regul. 42, 285–303

9. Manoury, B. (2013) Proteases: Essential actors in processing antigens and intracellular toll-like receptors. Front. Immunol. 4, 1–5

10. Boustany, R. M. N. (2013) Lysosomal storage diseases – The horizon expands. Nat. Rev. Neuro. 9, 583–598

11. Platt, F. M. (2014) Sphingolipid lysosomal storage disorders. Mol. Cell. Proteomics 13, 311–328

12. Thaysen-Andersen, M., Packer, N. H., and Schulz, B. L. (2016) Maturing glycoproteomics technologies provide unique structural insights into the N -glycoproteome and its regulation in health and disease. Mol. Cell. Proteomics 2, mcp.O115.057638

13. Frese, C. K., Altelaar, A. F. M., Van Den Toorn, H., Nolting, D., Griep, M., and Van Der Goes, T. (2015) A review of glycobiology and higher-energy collision dissociation tandem mass spectrometry. Anal. Chem. 84, 9688–9673

14. Song, X., Lasanaijak, Y., Olson, L. J., Boonen, M., Dahms, N. M., Kornfeld, S., Cummings, R. D., and Smith, D. F. (2009) Glycan microarray analysis of P-type lectins reveals distinct phosphomannose glycan recognition. J. Biol. Chem. 284, 35201–35214

15. Blackler, R. J., Evans, D. W., Smith, D. F., Cummings, R. D., Brooks, C. L., Braulke, T., Liu, X., Evans, S. V., and Müller-Loennies, S. (2015) Single-chain antibody-fragment M6P-1 possesses a mannose 6-phosphate monosaccharide-specific binding pocket that distinguishes N-glycan phosphorylation in a branch-specific manner. Glycobiology 26, 181–192

16. Chen, J., Bardes, E. E., Aronow, B. J., and Jegga, A. G. (2009) TopGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37, 305–311

17. Thu, P. J., Åkesson, K., Gernmandel, M., Jonsdottir, D., Geladaki, A., Attil Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnan, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulvey, C. M., Nilsson, P., Oksvold, P., Rockberg, J., Schwenk, J. M., Sivertsson Å, Sjöstedt, E., Skogs, M., Stadler, C., Sullivan, D. P.,
Mannose 6 dephosphorylation of lysosomal proteins mediated by acid phosphatases Acp2 and Acp5. Mol. Cell. Biol. 32, 774–782
39. Sleat, D. E., Zheng, H., Qian, M., and Lobel, P. (2006) Identification of sites of mannose 6-phosphorylation on lysosomal proteins. Mol. Cell. Proteomics 5, 686–701
40. Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Silva, D. V., Fuchser, J., Kaspar, S., Rapp, E., Seeberger, P. H., and Kolarich, D. (2013) Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: An evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. J. Mass Spectrom. 48, 627–639
41. Lee, L. Y., Moh, E. S. X., Parker, B. L., Bern, M., Packer, N. H., and Thaysen-Andersen, M. (2016) Toward automated N-glycopeptide identification in glycoproteomics. J. Proteome Res. 15, 3904–3915
42. Baumbach G. A, Saunders, P. T., Bazer, F. W., and Roberts, R. M. (1984) Uteroferrin has N-asparagine-linked high-mannose-type oligosaccharides that contain mannose 6-phosphate. Proc. Natl. Acad. Sci. U.S.A. 81, 2985–9
43. Gao, Y., Chen, Y., Zhan, S., Zhang, W., Xiong, F., and Ge, W. (2017) Comprehensive proteome analysis of lysosomes reveals the diverse function of macrophages in immune responses. Oncotarget 8, 7420–7440