Extending the Mannose 6-Phosphate Glycoproteome by High Resolution/Accuracy Mass Spectrometry Analysis of Control and Acid Phosphatase 5-Deficient Mice*\[^{2}\]

David E. Sleat†‡§¶, Pengling Sun‡, Jennifer A. Wiseman‡, Ling Huang‡§, Mukarram El-Banna‡, Haiyan Zheng‡, Dirk F. Moore‖§, and Peter Lobel‡†‡¶§

In mammals, most newly synthesized lumenal lysosomal proteins are delivered to the lysosome by the mannose 6-phosphate (Man6P) targeting pathway. Man6P-containing proteins can be affinity-purified and characterized using proteomic approaches, and such studies have led to the discovery of new lysosomal proteins and associated human disease genes. One limitation to this approach is that in most cell types the Man6P modification is rapidly removed by acid phosphatase 5 (ACP5) after proteins are targeted to the lysosome, and thus, some lysosomal proteins may escape detection. In this study, we have extended the analysis of the lysosomal proteome using high resolution/accuracy mass spectrometry to identify and quantify proteins in a combined analysis of control and ACP5-deficient mice. To identify Man6P glycoproteins with limited tissue distribution, we analyzed multiple tissues and used statistical approaches to identify proteins that are purified with high specificity. In addition to 68 known Man6P glycoproteins, 165 other murine proteins were identified that may contain Man6P and may thus represent novel lysosomal residents. For four of these lysosomal candidates, (lactoperoxidase, phospholipase D family member 3, ribonuclease 6, and serum amyloid P component), we demonstrate lysosomal residence based on the colocalization of fluorescent fusion proteins with a lysosomal marker. Molecular & Cellular Proteomics 12: 10.1074/mcp.M112.026179, 1806–1817, 2013.

The lysosome is a key site within the cell for the digestion of macromolecules, including proteins, carbohydrates, lipids, and nucleic acids (1), and this catabolic function is enabled by the concerted action of numerous hydrolases that have evolved to function in the acidic environment of the lysosome. Most lysosomal hydrolases and their accessory proteins are soluble and located within the lumen of the lysosome, and to date, ~70 such proteins have been identified (2, 3). Mutations in the genes encoding more than 30 of these proteins result in lysosomal storage diseases in which unhydrolyzed substrates accumulate within lysosomes, disrupting cellular function and frequently resulting in cell death (32).

Most newly synthesized soluble, lumenal lysosomal proteins are directed to the lysosome via the mannose 6-phosphate (Man6P) pathway (4). Here, select N-linked carbohydrates receive the Man6P modification that is recognized by two Man6P receptors (MPRs) that direct the vesicular trafficking of the lysosomal proteins from the trans-Golgi to an acidified prelysosomal compartment. In most cell types, the Man6P modification is rapidly removed by acid phosphatase 5 (ACP5) (5) upon arrival within the lysosome.

The presence of Man6P provides a unique opportunity for the study of lysosomal proteins by allowing for the highly specific affinity purification of this class of proteins from complex mixtures using MPRs immobilized on a solid affinity support (6). Man6P-containing lysosomal proteins have been purified from a wide variety of biological sources, and their characterization by proteomic methods has provided valuable insights into the composition and function of the lysosome as well as the roles of lysosomal proteins in human disease (2, 7). In mammals, the tissue distribution of lysosomal proteins can be highly restricted, and thus one approach that has effectively increased proteome coverage has been to analyze multiple individual tissues (3).

\[^{1}\] The abbreviations used are: Man6P, mannose 6-phosphate; ACP5, acid phosphatase 5; LPO, lactoperoxidase; PLD3, phospholipase D family member 3; RNASE6, ribonuclease 6; APCS, serum amyloid P component; MPR, mannose 6-phosphate receptor; sCl-MPR, soluble cation-independent mannose 6-phosphate receptor; PBS-I, PBS with inhibitors; PBS-TI, PBS with inhibitors and Tween 20; Glc6P/Man, glucose 6-phosphate and mannose; GLB1L, β-galactosidase 1-like; ER, endoplasmic reticulum; SLRRP, small leucine-rich repeat proteoglycan.
One limitation to the MPR affinity purification approach is that in most cells and tissues, a relatively small fraction of lysosomal proteins contain Man6P (8) because of natural dephosphorylation by ACP5 (5), with the notable exception of brain (9). As a result, minor lysosomal components or those that are particularly susceptible to dephosphorylation, either in vivo or during the course of purification, may be overlooked in proteomic surveys. In addition, as the proportion of Man6P glycoproteins in the complex mixture of proteins of an initial tissue extract becomes smaller, the proportion of contaminants is likely to increase in the purified preparation. In practical terms, extensive dephosphorylation of lysosomal proteins can make purification from limiting tissue sources in whole organism surveys a difficult or impossible prospect.

In this study, we have extended the proteomic analysis of mammalian Man6P glycoproteins with an analysis of multiple tissues from the mouse. We address the problem of dephosphorylation of lysosomal proteins with the inclusion of multiple tissues from ACP5-deficient mice (10), and we apply high resolution/accuracy mass spectrometry to confidently identify components of the mixture. Sixty eight known Man6P glycoproteins (65 lysosomal and 3 nonlysosomal) are identified, which is comparable with earlier studies. However, the modified approach has also resulted in the identification of 165 proteins (65 lysosomal and 3 nonlysosomal) are identified, which is comparable with earlier studies. However, the modified approach has also resulted in the identification of 165 proteins which are particularly susceptible to dephosphorylation, either in vivo or during the course of purification, may be overlooked in proteomic surveys. In addition, as the proportion of Man6P glycoproteins in the complex mixture of proteins of an initial tissue extract becomes smaller, the proportion of contaminants is likely to increase in the purified preparation. In practical terms, extensive dephosphorylation of lysosomal proteins can make purification from limiting tissue sources in whole organism surveys a difficult or impossible prospect.

In this study, we have extended the proteomic analysis of mammalian Man6P glycoproteins with an analysis of multiple tissues from the mouse. We address the problem of dephosphorylation of lysosomal proteins with the inclusion of multiple tissues from ACP5-deficient mice (10), and we apply high resolution/accuracy mass spectrometry to confidently identify components of the mixture. Sixty eight known Man6P glycoproteins (65 lysosomal and 3 nonlysosomal) are identified, which is comparable with earlier studies. However, the modified approach has also resulted in the identification of 165 other murine proteins that are specifically purified by MPR affinity chromatography and that may potentially represent lysosomal proteins. This is the most extensive list obtained to date for candidate mannose 6-phosphorylated proteins, and we demonstrate that four of these proteins localize to the lysosome using morphological methods.

**EXPERIMENTAL PROCEDURES**

**Animals**—Experiments and procedures involving live animals were conducted in compliance with protocols approved by the Robert Wood Johnson Medical School Institutional Animal Care and Use Committee. Acp5−/− mutant mice have been described previously (10) and were in an isogenic 129SvEv genetic background. Mice were deeply anesthetized with a mixture of sodium pentobarbital and sodium phenytoin (Euthasol, Delmarva Laboratories, Milton, DE) and killed by exsanguination following transcardiac perfusion with saline. Tissues were removed by dissection, frozen on dry ice, and stored at −80°C prior to use.

**Purification of Man6P Glycoproteins**—Proteins containing Man6P were isolated by affinity purification on immobilized bovine soluble cation-independent MPR (sCl-MPR) columns using a modification of a procedure originally developed for rat tissues (3). For Experiments 1–3 (supplemental Table 1), organs from multiple animals were dissected and pooled, and a maximum of 5 g of a given organ type were used for the purification of Man6P glycoproteins. Small organs for which the aggregate pool weighed 5 g or less were homogenized directly. Large organs (e.g. liver) were frozen, powdered with a Bessman tissue pulverizer (Spectrum Laboratories, Rancho Dominguez, CA), pooled together, and mixed, and 5 g were removed for homogenization. Tissues were homogenized on ice using a Polytron (Brinkmann Instruments) in 100 ml of PBS-TI (PBS containing 0.2% Tween 20, 2.5 mM EDTA, 5 mM β-glycerophosphate, 1 μg/ml pepstatin, 1 μg/ml leupeptin, and 0.5 mM Pefabloc). Carcasses were homogenized and analyzed individually. Each carcass (−20 g) was individually homogenized in 200 ml of buffer, and 50 ml of homogenate was removed and added to 50 ml of fresh buffer to maintain a volume/weight ratio of >20 as used for other tissues. Homogenates were centrifuged at 15,000 × g for 45 min at 4°C, filtered through Whatman No. 1 paper, and then loaded onto a 5-ml bed volume of immobilized sCl-MPR at a rate of 50 ml/h. Columns were washed with 30 ml of PBS-TI at 10 ml/h, gravity flow batch-washed four times with 10 ml of PBS-TI, then four times with 10 ml of PBS-I (i.e. PBS-TI without Tween 20), and then washed overnight with 80 ml PBS-I at 5 ml/hr. Columns were gravity flow-eluted two times with 5 ml of PBS containing 5 mM glucose 6-phosphate and 5 mM mannose (Glc6P/Mann) and then two times with 5 ml of PBS containing 5 mM Man6P. Eluates were concentrated using Amicon Ultra-4 Ultracel-10k filtration devices. For Experiment 4, Man6P glycoproteins were affinity-purified from individual wild-type mouse brains as described previously (12).

**Mass Spectrometry**—Samples of Man6P eluates and equivalent proportions of the total eluate of the corresponding Glc6P/Mann eluates were heated for 10 min at 60°C in reducing, denaturing SDS-PAGE sample buffer and then fractionated on precast 10% polyacrylamide gels (Invitrogen) until the bromphenol blue dye front had run ∼1 cm into the gel (3). Gel slices corresponding to each sample were excised and cut into small pieces, reduced, alkylated with iodoacetamide, and digested with trypsin (specificity, carboxyl side of lysine and arginine) using standard methods (13). Samples were analyzed in duplicate using an LTQ Orbitrap Velos tandem mass spectrometer coupled to a Dionex Ultimate 3000 RLSNANO System (Thermo Scientific, Somersett, NJ). Digests were solubilized in 0.1% TFA, and 0.5 μg were loaded on to a fused silica trap column of 100 μm × 2 cm packed with Magic C18 AQ (5 μm bead size, 200-Å pore size; Bruker-Michrom, Auburn, CA). The trap column was washed with 0.2% formic acid at a flow rate of 10 μl/min for 5 min. Retained peptides were separated on a fused silica column of 75 μm × 50 cm packed with Magic C18 AQ (3-μm bead size, 200-Å pore size; Bruker-Michrom) using a segmented linear gradient from 4 to 90% B (A: 0.1% formic acid, B: 0.08% formic acid, 80% ACN); 5 min, 4–10% B; 60 min, 10–40% B; 15 min, 40–55% B; 10 min, 55–90% B. For each cycle, one full MS was scanned in the Orbitrap with resolution of 60,000 from 300 to 2000 m/z, and the 20 most intense peaks were fragmented by CID using a normalized collision energy of 35%, and products were scanned in the ion trap. Data-dependent acquisition was set for a repeat count of 2 and exclusion of 60 s.

**Generation of Peak Lists**—Peak lists were generated using Proteome Discoverer 1.2 with minimum and maximum precursor masses of 350 and 6000 Da, respectively, minimum signal/noise of 1.5, and no constraints with respect to retention time, charge state, or peak count.

**Database Search**—A local implementation of the Global Proteome Machine (14, 15) Cyclone XE (Beavis Informatics Ltd., Winnipeg, Canada) with X!Tandem version CYCLONE (2011.12.01) was used to search mass spectrometry data. Data from murine samples analyzed using the Orbitrap Velos were searched against ENSEMBL release 64 of the NCBIM37 mouse genome assembly (21886 known genes) together with a contaminant database containing common contaminant proteins and the bovine cation-independent mannose-6-phosphate receptor precursor (gi27805933). The following search parameters were used: fragment mass error, 0.4 Da; parent mass error, 10 ppm; maximum charge, +4; minimum 15 peaks assigned; 1 missed cleavage allowed. Carbamidomethylation was a constant modification; methionine oxidation was a variable modification throughout the search; and tryptophan oxidation, asparagine deamidation, and glutamine deamidation were variable modifications during model refinement. MS data files were analyzed together in a MudPit analysis and individual data extracted. ENSEMBL protein identifiers were converted to associated gene names to collapse multiple protein identi-
levels of contaminants detected during MS/MS analysis. This dance and/or readily dephosphorylated. Second, the greater two ways. First, we may detect proteins that are of low abundance within the affinity column after washing via Man6-dependent interactions.

Examples include cystatins, some of which are found in tissues; thus, the analysis of Acp5−/− mice could potentially increase the sensitivity of Man6P glycoproteomics studies in two ways. First, we may detect proteins that are of low abundance and/or readily dephosphorylated. Second, the greater yield of Man6P glycoproteins would decrease the relative levels of contaminants detected during MS/MS analysis. This is important given that a major challenge is differentiating contaminants from proteins of interest.

Previously, we recognized that there are two classes of contaminant in Man6P glycoprotein purification schemes (3, 13). Specific contaminants elute from the affinity media in the presence of Man6P; these represent proteins that do not contain Man6P but interact with true Man6P glycoproteins. Examples include cystatins, some of which are found in preparations of purified Man6P glycoproteins (13) despite their lack of glycosylation and whose presence can be explained by copurification with lysosomal cysteine proteases. Nonspecific contaminants are proteins that are retained within the affinity column after washing via Man6P-independent interactions.

Our experiments are designed to distinguish between nonspecific contaminants and true Man6P glycoproteins and associated proteins. For each sample, after washing, we first elute the affinity column with a mixture of glucose 6-phosphate and mannose (Glc6P/Man). This mock elution should contain some nonspecifically bound proteins that are slowly leaching off the column as well as displacing some lectins. We then elute the column with Man6P. This specific elution releases proteins bound to the immobilized sCI-MPR via Man6P-dependent interactions. The relative amounts of Man6P glycoproteins and nonspecific contaminants should be increased in the Man6P eluate and decreased in the Glc6P/Man eluate. Thus, a given protein may be classified following mass spectrometric analysis by comparing its spectral counts in the two eluates, and we have developed statistical methods for this purpose to provide a point estimate and confidence interval for degree of enrichment (3) together with a probability (i.e. a $q$ value that is the $p$ value corrected for multiple measurements). In this way, we classify proteins as clear contaminants (higher in the Glc6P/Man than Man6P eluate, $q \leq 0.05$), potential Man6P glycoproteins (higher in the Man6P than Glc6P/Man eluate, $q \leq 0.05$), or ambiguous (e.g. $q > 0.05$). In the mass spectrometric analysis of the two eluates, we use the same proportion of the total Glc6P/Man and Man6P eluates rather than equal amounts of protein. It is worth noting that the amount of sample analyzed for the Man6P eluate likely lies outside the linear range in terms of spectral count response (data not shown). As a result, we may be underestimating total spectral counts obtained in the Man6P eluate and thus underestimating the degree of enrichment in this eluate compared with the Glc6P/Man eluate. In effect, this means that our results may be subject to conservative error (i.e. some proteins that are enriched in the Man6P eluate may be classified as ambiguous).

The study was conducted as a series of four independent experiments. These are summarized in supplemental Table 1 together with identifiers for the 116 individual mass spectrometry runs and database searches. Experiment 1 was a direct comparison of proteins purified from wild-type and Acp5−/− liver. Experiment 2 was an initial analysis of total Acp5−/− mouse homogenate as well as six individual tissues (brain, heart, kidney, liver, small intestine, and spleen). Experiment 3 was an expanded analysis of Acp5−/− tissues (adrenal, brain, epididymis, heart, kidney, large intestine, liver, lung, ovary, small intestine, spleen, stomach, testes, and thymus) as well as the remaining carcass. Finally, Experiment 4 was an analysis of wild-type brain. We also conducted a reanalysis of a previous study of Man6P glycoproteins from multiple rat tissues (3) using more current rat protein databases as an additional validation of candidates arising from the analysis of mouse samples.

Analysis of Man6P Glycoproteins Purified from Control and Acp5−/− Liver—We examined the distribution of proteins in preparations of Man6P glycoproteins from liver from both...
wild-type and ACP5-deficient mice (supplemental Table 2). Proteins were classified as either “known Man6P” (2) or “other,” a category representing murine proteins that are not currently assigned to this organelle. Note our database of known Man6P category (supplemental Table 3) mainly includes characterized lysosomal proteins but also includes several proteins (myeloperoxidase, deoxyribonuclease 1, and renin) that are not considered primarily lysosomal. We classify granzymes A and B within the lysosomal category as they reside in specialized lysosome-related organelles. The other category includes contaminants as well as the novel lysosomal proteins whose identification represents the primary aim of this study. ACP5 is expressed at high levels in liver, and its loss results in greatly increased levels of Man6P glycoproteins within this organ (5). Although spectral counts assigned to known Man6P glycoproteins were similar in wild-type and mutant liver for both Man6P and Glc6P/Man eluates, the relative level of other proteins was greatly decreased in the absence of ACP5 (Fig. 1, Table 1). This demonstrates that, as predicted, reducing natural dephosphorylation increases the relative selectivity of purification of Man6P glycoproteins.

It is important to consider the effect of ACP5 on expression of the Man6P-containing forms of lysosomal proteins. In Fig. 2, we compared spectral counts assigned to known Man6P glycoproteins in the Man6P eluates from wild-type and ACP5-deficient mouse liver. We find that the levels of most known Man6P glycoproteins were elevated in the Acp5+/− mouse liver. However, in addition to ACP5 itself, several proteins were markedly decreased, including β-glucuronidase, prosaposin, cathepsins B, O, and S, and myeloperoxidase (supplemental Table 2). In view of this, our overall strategy to identify novel lysosomal candidates was modified to include liver and brain from wild-type animals in addition to the multiple organs from the Acp5+/− mice.

Analysis of Multiple Acp5+/− Mouse Tissues—Fig. 3 shows the relative levels based on spectral counting of lysosomal and other proteins present in the Glc6P/Man and Man6P eluates from the affinity chromatography of extracts of multiple organs from Acp5+/− mice as well as wild-type brain. Total spectral counts were lower in the Glc6P/Man eluates

| Genotype   | Protein class | Specificity of purification |
|------------|---------------|----------------------------|
|            |               | Enriched | Ambiguous | Depleted | Total |
| Acp5+/−    | Known Man6P   | 52       | 2         | 0        | 54    |
|            | other         | 39       | 25        | 17       | 88    |
| Acp5+/+    | Known Man6P   | 49       | 7         | 1        | 57    |
|            | other         | 41       | 68        | 106      | 222   |
FIG. 3. Spectral counts assigned in affinity chromatography eluates. Data show the mean of spectral count data from two replicate LC-MS runs of each purified sample with error bars representing the range. Contaminants of non-murine origin are excluded. All tissues are from Acp5−/− animals except for the indicated wild-type brains. Panel A, total spectral counts assigned in Man6P and Glc6P/Man eluates; panel B, spectral counts assigned to lysosomal and other proteins in Man6P eluates; panel C, spectral counts assigned to lysosomal and other proteins in Glc6P/Man eluates. Note that in some cases, in the Glc6P/Man eluate, levels of proteins not known to contain Man6P glycoproteins were significantly higher than other tissues (e.g. panel C, Brain Expt. 2 and Liver Expt. 3), possibly indicating that initial washing conditions were not sufficiently stringent to remove contaminants. However, in the Man6P eluate (panel B), these proteins were similar to those in other organs, indicating that the Glc6P/Man elution provided a final and effective wash.
compared with the Man6P eluate in most organs (Fig. 3, panel A). Analysis of the Man6P eluates revealed that most of the spectral counts were assigned to known Man6P glycoproteins (Fig. 3, panel B), the vast majority being lysosomal. In contrast, analysis of the Glc6P/Man eluates revealed that the majority of spectral counts were assigned to proteins not established to contain Man6P (Fig. 3, panel C). Levels of such proteins overall were relatively unchanged in the two eluates or depleted in the Man6P eluate, with the mean amount in the Man6P eluate being 0.8-fold that of the Glc6P/Man eluate (Fig. 3, compare panels A and C). For known Man6P glycoproteins, levels were on average ~5-fold higher in the Man6P eluate than in the Glc6P/Man eluate.

Identification of Candidate Man6P Glycoproteins—Our approach to the identification of novel lysosomal and other Man6P glycoproteins from mouse was to analyze data obtained for all experiments (supplemental Table 1), including multiple tissues from ACP5-deficient animals and liver and brain from wild-type controls. We used a MudPIT style, merged database search strategy and identified candidates by measuring the specificity of purification as described above (i.e. by comparing spectral counts obtained for each protein in the Glc6P/Man and Man6P eluates). In total, we identified 68 Man6P-containing proteins that included 65 known lysosomal proteins. In addition, we identified 1292 mouse proteins that are not currently thought to contain Man6P (protein and peptide assignments are shown in supplemental Tables 4 and 5, respectively). In Fig. 4, panel A, we compare the distribution of spectral counts assigned to each protein in the Glc6P/Man and Man6P eluates. Note that for the purposes of this analysis, data were filtered to accept only murine proteins with total spectral counts of ≥20, yielding 68 and 633 known Man6P and other proteins, respectively. (Identity and statistical analysis for these are listed in supplemental Table 6.)

We have evaluated the experiment to experiment variation of the relative levels of individual proteins in Man6P compared with the Glc6P/Man eluates, and these analyses are included in supplemental material. The supplemental Figs. 1–4 are heatmap diagrams that provide a global visual evaluation of experimental variation for each individual protein. The supplemental Figs. 5–8 are forest plots (19) depicting the proportion of individual proteins in Man6P compared with the total (Man6P plus Glc6P/Man) eluates, with bars indicating 95% percentile confidence, and plot box areas being proportional to this confidence. For both graphic depictions, protein assignments are categorized as known Man6P (supplemental Figs. 1 and 5), and proteins that are not known to contain Man6P but are either enriched in the Man6P eluate (supplemental Figs. 2 and 6), cannot be classified in terms of specificity of purification (supplemental Figs. 3 and 7), or are depleted in the Man6P eluate (supplemental Figs. 4 and 8).

All known Man6P glycoproteins were significantly enriched in the Man6P compared with the Glc6P/Man eluate with the exception of GM2 activator protein (GM2A), which was depleted. This may indicate that GM2A is a particularly low affinity ligand for the sIC-MPR or that it is purified in physical association with another Man6P glycoprotein, a possibility supported by the fact that its intracellular targeting is only partly dependent upon Man6P (20).

Six hundred and thirty three proteins that are not currently thought to contain Man6P were identified that met the threshold for total spectral counts. Of these, most (363) were significantly depleted in the Man6P eluate and are thus likely to represent contaminants. For 105 proteins, no significant conclusions could be drawn regarding distribution in the two eluates. However, 165 proteins were significantly elevated in the Man6P eluate, potentially indicating the presence of Man6P.

As an additional level of validation, we compared results obtained with mouse with a reanalysis of earlier data obtained from a study of multiple rat tissues (3) (protein and peptide assignments and statistical analyses are shown in supplemental Tables 7–9). The number (65) and degree of enrichment of known Man6P proteins was similar to that observed from mouse, with GM2A also being depleted in the specific eluate (Fig. 4, panel B). Two hundred and twenty five other rat proteins not known to contain Man6P that met the threshold for spectral counts (≥20) were detected, a number considerably less than the 633 found with mouse. This may reflect the number of LC-MS/MS analyses (116 for mouse and 67 for rat) as well as differences in instrumentation (Orbitrap Velos for mouse versus LTQ for rat). Nonetheless, we can use the overlapping datasets to further test the validity of assignments. In total, 209 proteins were identified in both species, including 64 known Man6P glycoproteins. In general, classification of proteins that were identified in both rat and mouse corresponded well, with 174/209 proteins classified similarly for the two species (Table II). This agreement is also reflected by the correlation between enrichment factors for proteins identified in preparations from mouse and rat (Fig. 4, panel C).

Identification and Validation of Lysosomal Candidates—Candidate Man6P glycoproteins identified from the aggregate analysis of all samples are broadly classified in Table III and listed in detail in supplemental Table 10. We selected four lysosomal candidates (serum amyloid P component (APCS), lactoperoxidase (LPO), phospholipase D family, member 3 (PLD3), and ribonuclease k6 (RNASE6)) to evaluate subcellular distribution. Our approach was to express fluorescent fusions between mCherry and the candidates of interest and compare cellular distribution with either a fluorescent-labeled NPC2 protein or a fluorescent-labeled dextran that are both endocytosed and delivered to the lysosome. These analyses were conducted in either stably transfected CHO cells (Fig. 5) or transiently transfected U2-OS and/or CHO cells (supplemental Figs. 8–12). Colocalization between the mCherry-can-
didate fusion and with fluorescent-labeled NPC2 or dextran lysosomal marker was quantified (supplemental Table 11). For APCS and PLD3, we observe a punctate cytoplasmic distribution of the fusion proteins with extensive colocalization with the lysosomal marker. When all experiments are considered together, we obtained a thresholded Mander’s colocalization coefficient between the lysosomal markers and APCS and PLD3 of 0.81 and 0.83, respectively. For RNASE6, we also find significant colocalization with the lysosomal marker (average thresholded Mander’s colocalization coefficient, 0.66), but there is additional cytoplasmic staining. For LPO, diffuse reticular cytoplasmic staining predominates, but careful inspection does reveal additional punctate cytoplasmic staining that colocalizes with the lysosomal marker (average thresh-
olded Mander’s colocalization coefficient, 0.48). In cells expressing low amounts of the LPO fusion construct, the mCherry signal clearly colocalizes with the lysosomal marker (for example, see Fig. 5, panel LPO, cell b). These results provide convincing evidence that intracellular APCS and PLD3 primarily reside within the lysosome and suggest that some of the intracellular LPO and RNASE6 also reside within this organelle.

CONCLUSIONS

This study exploits our earlier discovery that the Man6P lysosomal recognition signal is removed within the lysosome by the lysosomal phosphatase, ACP5. We demonstrate that preparations of affinity-purified Man6P glycoproteins from mice lacking ACP5 have significantly lower levels of non-lysosomal contaminants than preparations from wild-type mice. However, we find that the presence of the additional contaminants associated with the sample purified from wild-type animals has no adverse effect on the identification of known Man6P glycoproteins and presumably lysosomal candidates. This is likely because the purified samples from both mice genotypes are of relatively low complexity, and thus analytical capacity is not a limiting factor with the specific instrumentation and workflow employed to achieve the goals of this study.

Spectral counts of most known Man6P glycoproteins were relatively increased in preparations of Man6P glycoproteins from Acp5−/− liver compared with wild type. Among the exceptions were β-glucuronidase, prosaposin, and several cathepsins (Fig. 3 and supplemental Table 2). It is not clear why the Man6P forms of these proteins are not elevated in liver, but it is possible that they are preferential substrates for an alternative ACP5-independent pathway for the removal of Man6P. For example, there is evidence that lysosomal acid phosphatase (ACP2) may play a role in the dephosphorylation of lysosomal proteins (21) or alternatively, Man6P could be lost as a consequence of deglycosylation. The fact that the loss of ACP5 does not result in the complete retention of Man6P in liver suggests that such a pathway exists. It is worth noting that in an earlier study (5), we found that the loss of ACP5 did not result in the retention of Man6P on β-glucuronidase to the same extent as it does for β-galactosidase and β-hexosaminidase, providing support for the possibility that it may be partially dephosphorylated by an ACP5-independent pathway.

Candidate lysosomal proteins arising from this study are shown in Table III and supplemental Table 10, and they fall into several broad classes. We identified numerous protease inhibitors, and it is likely that many of these represent false
Mannose 6-Phosphate Glycoproteome

positives that are purified in association with lysosomal proteases. We also identified a number of endoplasmic reticulum (ER) glycosyltransferases. It is possible that these may be purified with lysosomal glycoproteins if they possess lectin-like properties. However, it is worth noting that for some (e.g. B3GNT1 and POFUT2), we previously demonstrated that they do contain Man6P (22). This suggests that some or all of this class of specifically purified proteins may actually represent true Man6P glycoproteins. The biological significance of the Man6P lysosomal targeting modification on ER glycosyltransferases is unclear. It may be that these proteins are low affinity substrates for the phosphotransferase responsible for the addition of Man6P, and they receive the modification as a result of prolonged exposure to the pool of phosphotransferase residing in the ER (23).

Six small leucine-rich repeat proteoglycans (SLRRPs) were specifically purified. SLRRPs are a class of proteins containing repetitive leucine sequences and various classes of glycosaminoglycan side chains, including keratin sulfate, chondroitin sulfate, dermatan sulfate, and heparin sulfate chains. Originally recognized for their structural properties, there is increasing evidence that proteoglycans such as SLRRPs play a role in cellular signaling (24). Again, it is not clear whether their specific isolation here reflects the presence of Man6P or interaction with true Man6P glycoproteins (as outlined above). There is evidence that proteoglycans of the serglycin type may interact with the CI-MPR in competition with Man6P-containing ligands, suggesting that this may be a direct interaction (25). In addition, SLRRPs contain O-linked glycans, and there is evidence that another protein (α-dystroglycan) contains similar structures with phosphorylated O-mannosyl residues that may possibly interact with the CI-MPR (26).

The most numerous of the specifically isolated proteins were catabolic hydrolases, and given that the majority of lysosomal Man6P glycoproteins also fall within this class, we have examined these proteins in more detail (Table IV). The majority of the enriched hydrolases were proteases/peptidases with a large number of serine proteases identified, including eight members of the family of kallikrein-related peptidases. Many of the enriched hydrolases have a relatively limited cellular distribution; for example, we find a number of mast cell-specific proteases and eosinophil-specific ribonucleases, suggesting that some cell types may have unique lysosomal proteomes reflecting specialized function.

A number of candidates were identified with primary sequence and/or functional similarities to known lysosomal proteins. Most of these candidates (e.g. SMPD3A, PLBD1, ARSK, and RNASE6) have been discussed previously (2), but one notable lysosomal candidate is identified in this study. β-Galactosidase 1-like (GLB1L) was present at relatively low levels based on total spectral counts but was highly enriched in the Man6P eluate. This protein has significant sequence similarity to β-galactosidase 1 (human GLB1 and GLB1L are 55% identical and 67% similar), and although it is likely to be a glycosylhydrolase, its precise catalytic function remains to be determined. It is interesting to note that of the tissues analyzed in the Acp5−/− mice, expression of GLB1L was almost completely restricted to the reproductive system, with 34 of the 42 total spectral counts derived from testis and epididymis and four derived from ovary. Examination of the EST profile (www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist = Hs.181173) for the human ortholog also shows highest relative expression levels in testis. These results suggest that this protein may play a specialized role in reproduction.

We investigated the subcellular localization of four candidates arising from this study by expressing fluorescently tagged fusion constructs and found evidence for lysosomal localization for each. PLD3 (phospholipase D family, member 3) was previously identified as a glycosylated, endoplasmic reticulum-resident transmembrane protein (27). A recent study showing that PLD3 colocalizes with lysosomal protein LAMP1, and that its expression is mediated by a transcription factor that regulates lysosomal biogenesis and function (28), is consistent with a lysosomal distribution. LPO is found in secretory fluid such as milk and saliva and plays an important role in host defense. LPO is a member of the heme family of peroxidases, which includes myeloperoxidase, a component of neutrophil secretory granules that contains Man6P (29) that may be important for segregation and intracellular transport. Immunohistochemical analysis from the Human Protein Atlas Project (30) indicates some cytoplasmic staining for LPO that may be consistent with lysosomal localization. However, we also observed diffuse cytoplasmic staining for LPO that may reflect ER localization. Ribonuclease k6 (RNASE6) is a ubiquitously expressed ribonuclease that has also been postulated to play a role in host defense (31). To date, the only ribonuclease that has been localized to the lysosome is ribonuclease T2 (32). Our results suggest that RNASE6 represents an additional lysosomal, Man6P-containing ribonuclease, and the first of the RNase A family to be shown to reside within this organelle.

A lysosomal localization for PLD3, LPO, and RNASE6 is consistent with their predicted enzymatic activities and the presence of similar enzymes within the organelle. However, one intriguing result of this study was the lysosomal localization of serum APCS, a circulatory glycoprotein that binds β-amyloid peptide (33) and that is a component of human amyloid deposits (34). Given the intriguing connections between the lysosomal system and Alzheimer disease (35), the observation that APCS can reside within the lysosome may have significant implications for disease.

Analysis of Man6P glycoproteins in this study and elsewhere has resulted in a significant number (~50–100) of new proteins that may potentially reside within the lumen of the lysosome, and systematic validation of their localization will
be the next major step in defining this part of the lysosomal proteome. In this study, we have used fluorescent-tagged expressed proteins to demonstrate colocalization with two lysosomal markers (fluorescent NPC2 and dextran) in two different cell types (CHO and U2-OS), and this approach does demonstrate that the respective candidates can be delivered to this organelle. Two of the candidate fusions (PLD3 and APCS) appear to predominantly localize to the lysosome, but for LPO and to a lesser extent, RNASE6, we also observe nonlysosomal cytoplasmic staining. The physiological relevance of nonlysosomal staining in the latter cases is not clear. It can be argued that the approach of expression of fluorescent fusion proteins, although a mainstay of cell biology, is perturbing, and thus results may not necessarily reflect the distribution of the native endogenous protein. For example, the C-terminal mCherry fusion could interfere with normal protein folding or glycosylation, and this may in turn disrupt normal intracellular targeting. Alternatively, overexpression may interfere with normal targeting pathways, resulting in abnormal subcellular location, e.g. an accumulation within the ER. However, it is also possible that LPO and RNASE6 normally have multiple cellular localizations. An example of such a protein is superoxide dismutase 1 (36). Thus, in future studies it will be important to determine the proportion of a

| Gene symbol | Protein description | Specificity of purification |
|-------------|---------------------|-----------------------------|
| 2210010C04Rik | Trypsin-3 | Depleted Elevated |
| Adam28 | A disintegrin and metallopeptidase domain 28 | Elevated Elevated |
| Amy2a1 | Pancreatic α-amylase | Elevated Elevated |
| Arsk | Arylsulfatase K | Elevated Elevated |
| Bch | Butyrylcholinesterase | Elevated Elevated |
| Btd | Biotinidase | Elevated Elevated |
| Ces1c | Carboxylesterase 1C | Elevated Elevated |
| Ces1e | Carboxylesterase 1E | Elevated Elevated |
| Ces2g | Carboxylesterase 2G | Elevated Elevated |
| Ces5a | Carboxylesterase 5A | Elevated Elevated |
| Cma1 | Chymase 1, mast cell | Elevated Elevated |
| Cp | Ceruloplasmin | Elevated Elevated |
| Cpa2 | Carboxypeptidase A2 | Elevated Elevated |
| Cpa3 | Carboxypeptidase A3 mast cell | Elevated Elevated |
| Cpb2 | Carboxypeptidase B2 (plasma) | Elevated Elevated |
| Cym | Embryonic pepsinogen | Elevated Elevated |
| Ear1 | Eosinophil-associated ribonuclease A family | Elevated Elevated |
| Ear2 | Eosinophil-associated ribonuclease A family | Elevated Elevated |
| Ear6 | Eosinophil-associated ribonuclease A family | Elevated Elevated |
| Earp1 | Endoplasmic reticulum aminopeptidase 1 | Elevated Elevated |
| Fuc2a | Fucosidase α-L-2 | Elevated Elevated |
| Gb1I | Galactosidase β1-like | Elevated Elevated |
| Klk1b1 | Kallikrein 1-related peptidase b1 | Elevated Elevated |
| Klk1b11 | Kallikrein 1-related peptidase b11 | Elevated Elevated |
| Klk1b16 | Kallikrein 1-related peptidase b16 | Elevated Elevated |
| Klk1b27 | Kallikrein 1-related peptidase b27 | Elevated Elevated |
| Klk1b3 | Kallikrein 1-related peptidase b3 | Elevated Elevated |
| Klk1b4 | Kallikrein 1-related peptidase b4 | Elevated Elevated |
| Klk1b5 | Kallikrein 1-related peptidase b5 | Elevated Elevated |
| Klk1b9 | Kallikrein 1-related peptidase b9 | Elevated Elevated |
| Mcpt1 | Mast cell protease 1 | Elevated Elevated |
| Mcpt2 | Mast cell protease 2 | Elevated Elevated |
| Mcpt8 | Mast cell protease 8 | Elevated Elevated |
| Mcpt8l3 | Mast cell protease 8-like 3 | Elevated Elevated |
| Mmp1 | Multiple inositol polyphosphate histidine phosphatase | Elevated Elevated |
| Ngly1 | N-Glycanase 1 | Elevated Elevated |
| Pibd1 | Phospholipase B domain containing 1 | Elevated Elevated |
| Pid3 | Phospholipase D family member 3 | Elevated Elevated |
| Philp1 | Pancreatic lipase-related protein 1 | Elevated Elevated |
| Prtn3 | Proteinase 3 | Elevated Elevated |
| Rnase6 | Ribonuclease A family | Elevated Elevated |
| Smad3 | Sphingomyelin phosphodiesterase acid-like 3A | Elevated Elevated |
| Tpsab1 | Trypsin α/β1 | Elevated Elevated |
| Vnn3 | Vanin 3 | Elevated Elevated |
given protein that localizes to the lysosome, and this will be key in interpreting lysosomal residence. For example, for a given protein that is predominantly secreted with a small proportion retained intracellularly within the lysosome, interpretation of the lysosomal residence demonstrated by morphological methods is not as straightforward as that of a classical lysosomal protein that is primarily targeted to the organelle. Similar considerations will apply to proteins that have multiple intracellular locations, including the lysosome and nonlysosomal proteins undergoing lysosomal degradation. Future analyses applicable to lysosomal candidates, including analysis of cellular retention and targeted mass spectrometry to investigate subcellular distribution in differential centrifugation fractions (36), will provide useful information in such cases.

Supporting Information—Raw mass spectrometry data files will be made available in the Tranche public repository. Supporting information in the form of an Excel workbook is provided with information for protein assignment and statistical analyses.

Acknowledgments—We thank Cai Feng Zhao and MeiQian Qian for their excellent assistance with the mass spectrometry. We would also like to thank Dr. Zhiping Pang and the Child Health Institute of New Jersey (supported by the Robert Wood Johnson Foundation) for their excellent assistance with the mass spectrometry. We would also like to thank Dr. Zhipei Pang and the Child Health Institute of New Jersey (supported by the Robert Wood Johnson Foundation) for their excellent assistance with the mass spectrometry.
31. Rosenberg, H. F., and Dyer, K. D. (1996) Molecular cloning and characterization of a novel human ribonuclease (RNase k6): increasing diversity in the enlarging ribonuclease gene family. *Nucleic Acids Res.* **24**, 3507–3513

32. Campomenosi, P., Salis, S., Lindqvist, C., Mariani, D., Nordström, T., Acquati, F., and Taramelli, R. (2006) Characterization of RNASET2, the first human member of the Rh/T2/S family of glycoproteins. *Arch. Biochem. Biophys.* **449**, 17–26

33. Hamazaki, H. (1995) Ca²⁺-dependent binding of human serum amyloid P component to Alzheimer’s β-amyloid peptide. *J. Biol. Chem.* **270**, 10392–10394

34. Cathcart, E. S., Wollheim, F. A., and Cohen, A. S. (1967) Plasma protein constituents of amyloid fibrils. *J. Immunol.* **99**, 376–385

35. Nixon, R. A., and Cataldo, A. M. (2006) Lysosomal system pathways: genes to neurodegeneration in Alzheimer’s disease. *J. Alzheimers Dis.* **9**, 277–289

36. Della Valle, M. C., Sleat, D. E., Zheng, H., Moore, D. F., Jadot, M., and Lobel, P. (2011) Classification of subcellular location by comparative proteomic analysis of native and density-shifted lysosomes. *Mol. Cell. Proteomics* **10**, M110.006403