Identification and Validation of Mannose 6-Phosphate Glycoproteins in Human Plasma Reveal a Wide Range of Lysosomal and Non-lysosomal Proteins*

David E. Sleat‡§, Yanhong Wang‡, Istvan Sohar‡, Henry Lackland‡, Yan Li‖, Hong Li‖, Haiyan Zheng‡, and Peter Lobel‡§**

Acid hydrolase activities are normally confined within the cell to the lysosome, a membrane-delimited cytoplasmic organelle primarily responsible for the degradation of macromolecules. However, lysosomal proteins are also present in human plasma, and a proportion of these retain mannose 6-phosphate (Man-6-P), a modification on N-linked glycans that is recognized by Man-6-P receptors (MPRs) that normally direct the targeting of these proteins to the lysosome. In this study, we purified the Man-6-P glycoforms of proteins from human plasma by affinity chromatography on immobilized MPRs and characterized this subproteome by two-dimensional gel electrophoresis and by tandem mass spectrometry. As expected, we identified many known and potential candidate lysosomal proteins. In addition, we also identified a number of abundant classical plasma proteins that were retained even after two consecutive rounds of affinity purification. Given their abundance in plasma, we initially considered these proteins to be likely contaminants, but a mass spectrometric study of Man-6-phosphorylation sites using MPR-purified glycopeptides revealed that some proportion of these classical plasma proteins contained the Man-6-P modification. We propose that these glycoproteins are phosphorylated at low levels by the lysosomal enzyme phosphotransferase, but their high abundance results in detection of Man-6-P glycoforms in plasma. These results may provide useful insights into the molecular processes underlying Man-6-phosphorylation and highlight circumstances under which the presence of Man-6-P may not be indicative of lysosomal function. In addition, characterization of the plasma Man-6-P glycoproteome should facilitate development of mass spectrometry-based tools for the diagnosis of lysosomal storage diseases and for investigating the involvement of Man-6-P-containing glycoproteins in more widespread human diseases and their potential utility as biomarkers. Molecular & Cellular Proteomics 5:1942–1956, 2006.

The lysosome is an acidic, membrane-delimited organelle that is responsible for the degradation and recycling of macromolecules, playing a role in endocytosis and autophagy (1). Most of the resident hydrolases (e.g. proteases, glycosidases, lipases, phosphatases, sulfatases, and nucleases) and accessory proteins (e.g. activator proteins) are transported to the lysosome by the mannose 6-phosphate (Man-6-P) targeting pathway. Like many other glycoproteins, soluble lysosomal proteins are synthesized in the endoplasmic reticulum and are cotranslationally glycosylated on select asparagine residues. As these proteins move through the secretory pathway, the lysosomal proteins are selectively recognized by a phosphotransferase that initiates a two-step reaction that results in the generation of the Man-6-P modification on specific N-linked oligosaccharides. The modified proteins are then recognized by two Man-6-P receptors (MPRs), the cation-dependent MPR and the cation-independent (CI-) MPR (2). These receptors bind the phosphorylated lysosomal proteins in the neutral environment of the trans-Golgi network and travel to an acidic prelysosomal compartment in which the low pH promotes dissociation of the receptors and ligands. The receptors then recycle back to the Golgi to repeat the process or to the plasma membrane. Here the CI-MPR can function in the endocytosis and lysosomal targeting of extracellular Man-6-P glycoproteins.

The primary location and site of function of lysosomal enzymes is by definition intracellular. However, lysosomal activities have also been identified in a variety of extracellular

---

1 The abbreviations used are: Man-6-P, mannose 6-phosphate; MPR, Man-6-P receptor; CI-MPR, cation-independent MPR; sCI-MPR, soluble cation-independent MPR; HS, Heremans-Schmid; bis-Tris, 2-[bis(2-hydroxyethyl) amino]-2-(hydroxymethyl) propane-1,3-diol; 1D, one-dimensional; 2D, two-dimensional; GPM, Global Proteome Machine; BTD, biotinidase; PTGDS, prostaglandin H2 d-isomerase; CP, ceruloplasmin; CPVL, carboxypeptidase vitelligenic-like; SMPDL3A, acid sphingomyelinase-like protein 3a; CREG1, cellular repressor of E1A-regulated gene transcription; EPDR1, mammalian ependymin-related protein.
environments including plasma (3). These proteins may be delivered to the plasma through leakage from dead or dying cells, through mobilization of secretory lysosome or granule contents, or by the release of lysosomal residual bodies by cell defecation. It is also possible that a portion of the newly synthesized lysosomal proteins escape the intracellular targeting pathway for transport to the lysosome and are secreted. Evidence for the latter is well documented; for example, in cultured mouse embryonic fibroblasts, wild-type cells secrete a small proportion of Man-6-P glycoproteins (~10% of that secreted in the absence of both MPRs, depending on the protein) (4–7), indicating that the sorting of newly synthesized lysosomal proteins is not absolutely efficient.

The biological significance of circulating lysosomal proteins is unclear, but there are several possibilities. First, some hydrolases may be synthesized and released by one cell type and delivered to another, and so circulating lysosomal proteins may represent intermediates in this process. Second, circulating lysosomal proteins might have a specific function in plasma. Third, the presence of lysosomal proteins in plasma may simply represent the steady state levels reflecting the balance between the rate of unwanted appearance in plasma (due to leakage and/or lack of absolute fidelity in the intracellular targeting system) versus the rate of uptake and clearance.

Regardless of biological function, circulating lysosomal hydrolases potentially represent valuable biomarkers for the study and diagnosis of human diseases. Mutations in genes encoding lysosomal proteins result in over 40 storage diseases (for a review, see Ref. 8). In addition, lysosomal activities have also been indirectly implicated in more widespread pathogenic processes (9–12) including tumor invasion and metastasis, Alzheimer disease, rheumatoid arthritis, and atherosclerosis as well as in normal processes such as aging (13) and immune system function (14). As a class, lysosomal proteins represent a relatively small subgroup of the plasma proteome that should be amenable to quantitative analysis.

In this study, we set out to characterize the human plasma proteome of mannose 6-phosphate glycoproteins to provide a reference dataset as a diagnostic tool for lysosomal storage diseases that could also facilitate further studies of the role of these proteins in other human conditions. We also viewed plasma as a promising source for the identification of potentially novel lysosomal proteins in our ongoing effort to identify components of the lysosomal proteome (15–17).

Given the complexity of the plasma protein population and the wide range of protein abundances, it is not currently feasible to directly detect lysosomal proteins in unfractionated plasma using proteomic methods due to their extremely low abundance compared with classical plasma proteins (18). We have therefore used an approach in which the Man-6-P-containing lysosomal proteins are purified by affinity chromatography on immobilized MPR (15–17, 19–22). As expected, we identified many lysosomal proteins, but we also purified many proteins that have not previously been assigned either Man-6-P or lysosomal function. Many of these proteins are highly abundant plasma glycoproteins that we initially suspected to represent likely contaminants that do not contain Man-6-P. Therefore, to distinguish between true Man-6-P glycoproteins and contaminants, we used a technique to directly identify Man-6-P-modified glycopeptides. Unexpectedly, we found that a small proportion of multiple abundant classical plasma proteins exist as Man-6-P-containing glycoforms.

**EXPERIMENTAL PROCEDURES**

**Purification of Human Plasma Man-6-P Glycoproteins—** Refrozen lots of thawed human plasma were generously provided by V. I. Technologies, Inc. (Melville, NY). Plasma Man-6-P glycoproteins were purified using a modification of the affinity protocol described for human brain (17). For each purification, ~2 liters of frozen human plasma were rapidly thawed at 37 °C and then clarified by centrifugation at 13,000 × g for 90 min at 4 °C. The supernatant was filtered through six layers of cheesecloth and then added to an equal volume of ice-cold PBS containing 10 mM β-glycerophosphate, 10 mM EDTA, 2% Triton X-100, 0.4% Tween 20, and 0.2 mM Pefabloc (Pentafarm, Basel, Switzerland). The diluted plasma was then applied to a column of immobilized pentamannosyl phosphate-phosphate-aminooethyl agarose (100-m1 bed volume) to remove circulating soluble cation-independent MPR (sCI-MPR), and the flow-through was applied to a column of immobilized sCI-MPR (100-m1 bed volume at a coupling density of 3.3 mg/ml) at a flow rate of ~200 ml/h to capture circulating Man-6-P glycoproteins. The MPR resin was batch washed with 3 × 1 column volume of PBS containing 5 mM EDTA, 5 mM β-glycerophosphate, 1% Triton X-100 and then with 3 × 1 column volume of the same buffer without Triton X-100. The column was flow-washed with 20 column volumes of PBS/EDTA/β-glycerophosphate at 125 ml/h and then eluted with 2 column volumes of the same buffer containing 10 mM Man-6-P at a flow rate of 100 ml/h. Fractions containing Man-6-P glycoproteins were identified by protein and β-mannosidase assay, pooled, concentrated by ultrafiltration (Centriprep-10, Millipore, Billerica, MA), and buffer-exchanged to either 100 mM ammonium bicarbonate (single purification protocol) or PBS/EDTA/β-glycerophosphate (first elution from the double purification protocol). For repurification, the buffer-exchanged Man-6-P eluate was reapplied to the column of immobilized MPR and purified essentially as described above. The Man-6-P eluate was buffer-exchanged with 100 mM ammonium bicarbonate. Purified proteins were stored at ~80 °C, and volatile salts were removed by lyophilization prior to subsequent analyses.

**Two-dimensional Gel Electrophoresis—** First dimension isoelectric focusing was conducted using pH 3–7 or 3–10 non-linear Immobiline DryStrips (Amersham Biosciences), and second dimension gel electrophoresis was conducted using bis-Tris-HCl-buffered (pH 6.4) 10% Duracryl (Genomic Solutions, Ann Arbor, MI) gels run with MOPS running buffer. Transfer to nitrocellulose and detection of Man-6-P glycoproteins using iodinated sCI-MPR were as described previously (15, 22). Fractionated proteins were visualized using either a colloidal Coomassie Blue staining protocol (23) or SYPRO Ruby (Invitrogen).

**Tandem Mass Spectrometry—** Proteolytic digests of MPR-affinity purified plasma proteins were analyzed by nanospray LC-MS/MS using a ThermoElectron LTQ linear ion trap (ThermoElectron, San Jose, CA) and a Waters Micromass Q-TOF API US (Waters, Milford, MA) mass spectrometer. For both the singly and doubly purified human plasma samples, MPR-affinity purified mixtures were digested with trypsin in solution prior to LC-MS/MS. Forty micrograms of each sample were also fractionated by 1D SDS-PAGE (Novex 10% Nu-
Page gels with MOPS buffer (Invitrogen), and ~30 gel slices were digested with trypsin. For the doubly purified sample alone, 40 μg purified proteins were fractionated by two-dimensional (2D) gel electrophoresis, and ~60 spots were excised and digested with trypsin.

The LTQ was used to analyze digests of all the samples discussed above. Solution digests of each of the singly and doubly purified samples were analyzed in duplicate. Instrument, chromatographic conditions, and generation of peak lists were as described previously (24) except for the analysis of gel slices, which was conducted using a 50-min gradient. For the Q-TOF, tryptic digests of each MPR-affinity purified sample (5 μg) or tryptic digests of 1D gel slices were fractionated using a Micromass CapLC. Peptides were first desalted on a 300-μm × 1-mm PepMap C18 trap column with 0.1% formic acid in HPLC grade water at a flow rate of 20 μl/min. After desalting for 3 min, peptides were back-flushed onto an LC Packings 75-μm × 15-cm C18 nanocolumn (3 μm, 100 Å) at a flow rate of 200 nl/min. For solution digests of MPR-affinity purified proteins, peptides were eluted on a 170-min gradient of 3–42% acetonitrile in 0.1% formic acid. For digests of 1D gel slices, peptides were eluted on a 30-min gradient of 3–42% acetonitrile in 0.1% formic acid. The scan time for MS and MS/MS were 1.5 and 2.0 s, respectively. Mass ranges for the MS survey scan and MS/MS were m/z 300–1900 and m/z 50–1900, respectively. The top three multiply charged ions with an MS peak intensity greater than 30 counts/scan were chosen for MS/MS fragmentation with a precursor ion dynamic exclusion of 60 s. Instrument switching from MS/MS to MS mode occurred when either total MS/MS ion counts were over 5000 or when the total MS/MS scan time was over 6.0 s. Raw data were processed by ProteinLynx 2.0 to generate PKL files without peak threshold intensity limits (fast deisotoping) and with automatic charge state determination.

Database Searching—Peptide assignment and protein identification were conducted using a local implementation of X!Tandem Version 2005.12.01 (GPM-USB, Beavis Informatics Ltd., Winnipeg, Canada) to search the human proteome database (the February 2006 assembly of the National Center for Biotechnology Information (NCBI) 35, which contains a total of 33,869 entries representing 30,352 unique sequences) as described previously (24). Database searching using data generated with the Q-TOF was identical except that the parent ion mass error was ±0.5 Da. Methods for analysis of LC-MS/MS data from proteolytic digests of unfractionated mixtures or gel slices and for deglycosylated Man-6-P glycopeptides (see below) have been described previously (24). Data were exported into Microsoft Excel spreadsheets and filtered for threshold significance as described below. To avoid redundancy arising from proteins present in the human proteome database under multiple names and/or accession numbers, a list of ENSP numbers was compiled with all numbers corresponding to a single protein referenced to a single primary number. Protein assignments were compared with this database to ensure that multiple accession numbers representing the same protein were not listed.

Purification and Identification of Proteolytic Peptides Containing Mannose 6-Phosphate—The presence of Man-6-P on MPR-purified proteins was verified using a recently developed protocol for the purification and identification of glycopeptides containing Man-6-P from an unfractionated proteolytic digest of MPR-purified proteins (24).
The Human Plasma Mannose 6-Phosphate Glycoproteome

Unfractionated solution digests of MPR-affinity purified samples were analyzed in duplicate using the LTQ, and the highest confidence (i.e. lowest) log(e) score obtained in either of the two experiments is shown for each protein. For each set of duplicate experiments, we calculated the maximum log(f) for each individual protein and show here the differential between maximum log(f) scores obtained after one versus two cycles of affinity purification (i.e. \( \Delta \log(f) = \log(f)_{\text{ltq}} - \log(f)_{\text{q-tof}} \)). PPP, Plasma Proteome Project high confidence proteome database (25) (www.bioinformatics.med.umich.edu/hupo/ppp/).

| Gene  | Protein                                      | PPP     | Man-6-P glycopeptides | \( \Delta \text{Max log}(f) \) | Min log(e) |
|-------|---------------------------------------------|---------|-----------------------|---------------------------------|------------|
|       | Brain                                      | Plasma  |                      |                                 |            |
| ACP5  | Tartrate-resistant acid phosphatase         | +       | +                    | 6.06*                           | -28        |
| AGA   | Asparaginase                                | +       | +                    | 0.85                            | -173.4     |
| ARSA  | Arylsulfatase A                             | +       | +                    | 1.17                            | -66.8      |
| ASAH  | Acid ceramidase                             | +       | +                    | 0.45                            | -25.8      |
| CLN5  | Ceroid lipofuscinosis neuronal protein 5    | +       | +                    | -0.01                           | -17        |
| CPVL  | Probable serine carboxypeptidase CPVL       | +       | +                    | 1.4                             | -86.6      |
| CTBS  | Di-N-acetylhexitolase                       | +       | +                    | 0.8                             | -177.8     |
| CTSB  | Cathepsin B                                 | +       | +                    | 0.95                            | -102.3     |
| CTSB  | Cathepsin B                                 | +       | +                    | 1.05                            | -144.6     |
| CTSB  | Cathepsin D                                 | +       | +                    | 0.17                            | -116.7     |
| CTSB  | Cathepsin H                                 | +       | +                    | 1.05                            | -64.5      |
| CTSB  | Cathepsin L                                 | +       | +                    | 1.08                            | -292       |
| CTSB  | Cathepsin S                                 | +       | +                    | 1.03                            | -78.6      |
| CTSZ  | Cathepsin Z                                 | +       | +                    | 1.85                            | -100.9     |
| DNASE2| Deoxyribonuclease II                        | +       | +                    | 1.04                            | -74.7      |
| DPP7  | Dipeptidyl-peptidase II                     | +       | +                    | 0.71                            | -132.1     |
| FUGA1 | Tissue \( \alpha \)-fucosidase             | +       | +                    | 0.92                            | -71.7      |
| GAA   | Lysosomal \( \alpha \)-glucosidase         | +       | +                    | 0.96                            | -73.1      |
| GALNS | \( N \)-Acetylgalactosamine-6-sulfatase     | +       | +                    | 0.79                            | -80.1      |
| GGH   | \( \gamma \)-Glutamyl hydrolase            | +       | +                    | 0.58                            | -129.1     |
| GLB1  | \( \beta \)-Galactosidase                  | +       | +                    | 5.68*                           | -64.3      |
| GNS   | \( N \)-Acetylgalactosamine-6-sulfatase     | +       | +                    | 1.43                            | -67.6      |
| GUSB  | \( \beta \)-Glucuronidase                  | +       | +                    | 1.28                            | -126       |
| HEXA  | \( \beta \)-Hexosaminidase \( \alpha \) chain| +       | +                    | 1.1                             | -231.2     |
| HEXB  | \( \beta \)-Hexosaminidase \( \beta \) chain| +       | +                    | 0.82                            | -194.2     |
| HPSE  | Heparanase                                  | +       | +                    | 5.41*                           | -31        |
| IDS   | Iduronate 2-sulfatase                       | +       | +                    | 0.54                            | -174.1     |
| LGMN  | Legumain                                    | +       | +                    | 6.05*                           | -136.2     |
| LYPLA3| 1-O-Acylceramide synthase                   | +       | +                    | -0.76                           | -22.8      |
| MAN2B1| Lysosomal \( \alpha \)-mannosidase          | +       | +                    | 5.4*                            | -48.1      |
| MAN2B2| Epididymis-specific \( \alpha \)-mannosidase| +       | +                    | 0.58                            | -116.9     |
| MANBA | \( \beta \)-Mannosidase                     | +       | +                    | 0.64                            | -368.8     |
| MPO   | Myeloperoxidase                             | +       | +                    | 1.28                            | -147.5     |
| NAGA  | \( \alpha \)-N-Acetylgalactosaminidase      | +       | +                    | 6.01*                           | -87.1      |
| NAGLU | \( \alpha \)-N-Acetylgalactosaminidase      | +       | +                    | 0.68                            | -260.3     |
| NPC2  | NPC2 protein                                | +       | +                    | 1.15                            | -56.8      |
| PGCP  | Plasma glutamate carboxypeptidase           | +       | +                    | 0.34                            | -77.5      |
| PPG8  | Lysosomal protective protein                | +       | +                    | 0.87                            | -52.2      |
| PPT1  | Palmitoyl-protein thioesterase 1            | +       | +                    | 0.14                            | -26.6      |
| PRCP  | Lysosomal Pro-X carboxypeptidase            | +       | +                    | 1.09                            | -111.6     |
| PSAP  | Proactivator polypeptide                    | +       | +                    | 0.66                            | -43.4      |
| SGSH  | \( N \)-Sulfoglucosamine sulfohydrolase     | +       | +                    | -4.8b                           | -16.1      |
| SIAE  | Cytosolic sialic acid 9-O-acetyleresterase  | +       | +                    | 0.96                            | -151.6     |
| TPP1  | Tripeptidyl-peptidase l                     | +       | +                    | 0.12                            | -115       |

* Not identified in singly purified sample.

b Not identified in doubly purified sample.

ability that this represents a random match. When identifying proteins in the MPR affinity-purified protein mixtures, the score of highest confidence obtained at random when searching a reversed orientation human proteome was used as the threshold for acceptance of protein assignments. Method-specific thresholds were determined for unfraccionated digests (log(e) = 5.1 and \( \log(f) \), respectively) and for proteins fractionated by 1D PAGE when each gel slice was analyzed individually (log(e) = 4.3 and \( \log(f) \), respectively) for the LTQ and Q-TOF, respectively). Each threshold was used to filter the respective datasets obtained from searching the forward orientation human proteome. Samples were analyzed using different instruments (LTQ and Q-TOF), were singly and doubly purified, and were prepared for analysis using several different methods (solution digest and gel electrophoresis). Thus, as an additional level of stringency, proteins identified with one approach alone were rejected if the protein assignment scores failed to reach a threshold of log(e) less than 10. For analysis of purified Man-6-P glycopeptides, the threshold for acceptance of peptide assignments to proteins not previously classified as lysosomal was the score of lowest confidence of a deglycosylated peptide assigned to a known lysosomal species (pep-
The Human Plasma Mannose 6-Phosphate Glycoproteome

Non-lysosomal MPR-purified proteins from human plasma

Unfractionated solution digests of MPR-affinity purified samples were analyzed in duplicate using the LTQ, and the highest confidence (i.e. lowest) log(e) score obtained in either of the two experiments is shown for each protein. For each set of duplicate experiments, we calculated the maximum log(I) for each individual protein and show here the differential between maximum log(I) scores obtained after one versus two cycles of affinity purification (i.e. Δlog(I) = log(I)₂ - log(I)₁). PPP, Plasma Proteome Project high confidence proteome database (25) (www.bioinformatics.med.umich.edu/hupo/ppp/). In total, 29 immunoglobulin chains were identified in this analysis. Given the difficulties in unambiguous assignment of mass spectral data to individual genes encoding these proteins and the fact that these appear to represent nonspecific contaminants, these immunoglobulin chains are not included here. AMBP, α 1-microglobulin/bikunin precursor.

| Gene        | Protein                                      | PPP | Man-6-P Glycopeptides | ΔMax log(I) | Min log(e) |
|-------------|----------------------------------------------|-----|-----------------------|-------------|------------|
| LTF         | Lactotransferrin                             | +   | +                     | 5.94        | -145.2     |
| RNASE1      | Ribonuclease pancreatic                      | +   | +                     | 5.65        | -38.1      |
| HRNR        | Homerin                                      | +   | +                     | 5.62        | -123.4     |
| M6PR        | Cation-dependent mannose 6-phosphate receptor| +   | +                     | 5.62        | -12.3      |
| RNASE6      | Ribonuclease K6                              | +   | +                     | 5.46        | -27.8      |
| ICAM1       | Intercellular adhesion molecule 1            | +   | +                     | 5.41        | -50.2      |
| CREG1       | CREG1 protein                                | +   | +                     | 5.36        | -19.2      |
| NP.079105.3 | Laminin A FLJ22662                          | +   | +                     | 5.24        | -12.1      |
| HBA1        | Hemoglobin ζ chain                           | +   | +                     | 5.15        | -14.5      |
| CBLN4       | Cerebellin 4                                 | +   | +                     | 4.94        | -13.3      |
| DSP         | Desmopakin                                   | +   | +                     | 4.8         | -26.8      |
| FABP5       | Fatty acid-binding protein                   | +   | +                     | 4.76        | -13        |
| SUMF2       | Sulfatase-modifying factor 2                 | +   | +                     | 4.66        | -22.6      |
| ELA2        | Leukocyte elastase                           | +   | +                     | 4.53        | -10        |
| PLOD1       | Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1| +   | +                     | 4.38        | -19        |
| FTL         | Ferritin light chain                         | +   | +                     | 4.34        | -10.6      |
| FUCAL2      | Plasma α L-fucosidase                        | +   | +                     | 1.21        | -89.8      |
| SMPDL3A     | Acid sphingomyelinase-like phosphodiesterase 3A | +   | +                     | 1.09        | -143.6     |
| HBB         | Hemoglobin β chain                           | +   | +                     | 1.02        | -43.2      |
| NP.775813.1 | Laminin A LOC196463                         | +   | +                     | 0.81        | -70        |
| RNASET2     | Ribonuclease T2                              | +   | +                     | 0.73        | -148       |
| CECR1       | Cat eye syndrome critical region 1           | +   | +                     | 0.68        | -220.8     |
| LRG1        | Leucine-rich α 2-glycoprotein                | +   | +                     | 0.6         | -113.5     |
| SERPING1    | Antithrombin-III                             | +   | +                     | 0.53        | -191.3     |
| APCS        | Serum amyloid P-component                    | +   | +                     | 0.5         | -34.4      |
| SERPINA5    | Plasma serine protease inhibitor             | +   | +                     | 0.49        | -75.6      |
| HPR         | Haptoglobin-related protein                  | +   | +                     | 0.4         | -454.7     |
| C10A        | Complement C10 subcomponent, A chain         | +   | +                     | 0.35        | -29.6      |
| BCHE        | Cholinesterase                               | +   | +                     | 0.32        | -172.9     |
| AGT         | Angiotensinogen                              | +   | +                     | 0.25        | -204.3     |
| PTGDS       | Prostaglandin-H2 d-isomerase                 | +   | +                     | 0.25        | -47.3      |
| SERPING1    | Plasma protease C1 inhibitor                 | +   | +                     | 0.24        | -44.1      |
| EPDR1       | Mammalian ependymin-related protein          | +   | +                     | 0.21        | -46        |
| A1BG        | α 1-B-glycoprotein                           | +   | +                     | 0.18        | -145       |
| KLKB1       | Plasmin kallikrein                           | +   | +                     | 0.18        | -38.4      |
| HPX         | Hemopexin                                    | +   | +                     | 0.17        | -231.5     |
| AMBP        | AMBP protein                                 | +   | +                     | 0.12        | -86.4      |
| SERPINA1    | α 1-Antitrypsin                              | +   | +                     | 0.08        | -159.6     |
| SERPINF1    | Pigment epithelium-derived factor            | +   | +                     | 0.08        | -13.9      |
| A2M         | α 2-Macroglobulin                            | +   | +                     | 0.07        | -538.2     |
| SERPINA4    | Kallistatin                                  | +   | +                     | 0.05        | -94.4      |
| FCGBP       | Fc fragment of IgG-binding protein           | +   | +                     | 0.01        | -413.9     |
| SERPINA6    | Corticosteroid-binding globulin              | +   | +                     | 0.01        | -60.9      |
| AZGP1       | Zinc-α 2-glycoprotein                        | +   | +                     | -0.02       | -196.9     |
| AFM         | Afamin                                       | +   | +                     | -0.03       | -106       |
| TF          | Serotransferrin                              | +   | +                     | -0.03       | -226.4     |
| CP          | Ceruloplasmin                                | +   | +                     | -0.05       | -282.6     |
| BTD         | Biotinidase                                  | +   | +                     | -0.06       | -100.8     |
| C1QB        | Complement C1q subcomponent, B chain         | +   | +                     | -0.16       | -49.1      |
| C1OQ        | Complement C1q subcomponent, C chain         | +   | +                     | -0.16       | -40.7      |
| FCN3        | Ficolin-3                                    | +   | +                     | -0.24       | -73        |
| ALB         | Serum albumin                                | +   | +                     | -0.26       | -349.8     |
| ORM1        | α 1-Acid glycoprotein 2                      | +   | +                     | -0.26       | -98.2      |
| CDSL        | CD5 antigen-like                             | +   | +                     | -0.36       | -43.1      |
| ORM2        | α 1-Acid glycoprotein 2                      | +   | +                     | -0.36       | -44.2      |
| C2          | Complement C2 precursor                      | +   | +                     | -0.41       | -21.3      |

TABLE II

Max log(I) score obtained in either of the two experiments is shown for each protein. For each set of duplicate experiments, we calculated the maximum log(I) for each individual protein and show here the differential between maximum log(I) scores obtained after one versus two cycles of affinity purification (i.e. Δlog(I) = log(I)₂ - log(I)₁). PPP, Plasma Proteome Project high confidence proteome database (25) (www.bioinformatics.med.umich.edu/hupo/ppp/). In total, 29 immunoglobulin chains were identified in this analysis. Given the difficulties in unambiguous assignment of mass spectral data to individual genes encoding these proteins and the fact that these appear to represent nonspecific contaminants, these immunoglobulin chains are not included here. AMBP, α 1-microglobulin/bikunin precursor.
### Purification of Plasma Man-6-P Glycoproteins

Man-6-P glycoproteins were initially purified from human plasma with a yield of ~500 µg/liter, which represents ~0.001% of the total starting protein content. Staining of the mixture after fractionation by two-dimensional gel electrophoresis using pH 3–10 isoelectric gradients revealed a complex pattern of purified proteins with >1500 spots resolved (Fig. 1A).

A significant problem encountered using an affinity purification approach to identify lysosomal proteins based on the presence of Man-6-P is the identification of false positives (24). Assuming correct protein assignment, false positives can represent abundant or “sticky” nonspecific contaminants (e.g. immunoglobulins or albumin) that bind nonspecifically and continuously leach off the column throughout the washing process and during the Man-6-P elution step. Alternatively non-Man-6-P proteins such as lectins or protease inhibitors can bind to and copurify with true Man-6-P glycoproteins, and these can be regarded as “specific contaminants” because they will elute specifically from the immobilized MPR with Man-6-P. A particular concern in this study, given that we expected lysosomal proteins to be very low abundance constituents of plasma, was the presence of nonspecific contaminants. To assess purity, we fractionated the mixture by two-dimensional gel electrophoresis, transferred the sample to nitrocellulose, and used a radiolabeled form of the sCI-MPR in a Western blot style assay (22) to visualize Man-6-P glycopeptides (Fig. 1B). When the resulting MPR blot was compared with the stained gel, it was clear that a considerable proportion of the purified proteins appeared not to contain Man-6-P. We therefore repeated the purification on our preparation and found that most of spots corresponding to the twice-purified proteins appeared to contain Man-6-P (Fig. 1, C and D). Given the stringency of this double purification, many of the spots that appeared not to contain Man-6-P probably represent either specific contaminants or proteolytically processed chains of lysosomal proteins that do not contain Man-6-P rather than nonspecific contaminants. Protein yields after two rounds of purification were approximately half of those obtained after a single round of purification (on average, 230 µg/liter compared with 480 µg/liter), reflecting the loss of contaminants.

### Comparison of Analytical Methods

A number of mass spectrometric approaches were used to characterize the singly and doubly purified mixtures of MPR-

---

**Table II—continued**

| Gene   | Protein                          | PPP | Man-6-P Glycopeptides | ΔMax log(l) | Min log(l) |
|--------|----------------------------------|-----|-----------------------|-------------|------------|
| ITIH4  | Inter-α1 -trypsin inhibitor heavy chain H4 | +   | Brain Plasma          | -0.48       | -52.3      |
| FCN2   | Ficolin-2                        | +   |                      | -0.53       | -133       |
| BF     | Complement factor B              | +   |                      | -0.73       | -146.9     |
| DBH    | Dopamine β1-hydroxylase          | +   |                      | -0.78       | -33.3      |
| FGB    | Fibrinogen β chain               | +   |                      | -4.19b      | -11.6      |
| SERPINA3 | α1-Antichymotrypsin           |     |                      | -4.31b      | -11.6      |
| ECM1   | Extracellular matrix protein 1   | +   |                      | -4.35b      | -18.2      |
| KNG1   | Kininogen-1                      | +   |                      | -4.49b      | -23.3      |
| LUM    | Lumican                          | +   |                      | -4.57b      | -11        |
| C4B    | Complement component 4B         |     |                      | -4.89b      | -53.1      |
| IGF2R  | Cation-independent mannose-6-phosphate receptor |     |                      | -4.95b      | -23.7      |
| ARTS-1 | Adipocyte-derived leucine aminopeptidase |     |                      | -5.09b      | -33.2      |
| FETUB  | Fetuin-B                         |     |                      | -5.1b       | -19.2      |
| NP_443122.2 | N-Acetylmuramoyl-L-alanine amidase |     |                      | -5.1b       | -35.1      |
| HRG    | Histidine-rich glycoprotein      | +   |                      | -5.13b      | -39.4      |
| VTN    | Vitronectin                      | +   |                      | -5.14b      | -26.5      |
| AHSG   | α2-HS-glycoprotein               | +   |                      | -5.16b      | -37.5      |
| ITIH2  | Inter-α1-trypsin inhibitor heavy chain 2 | +   |                      | -5.21b      | -36        |
| C3     | Complement C3 precursor          | +   |                      | -5.36b      | -88.2      |
| CLU    | Clusterin                        |     |                      | -5.38b      | -41.6      |
| C4BPA  | C4b-binding protein α chain      | +   |                      | -5.4b       | -62.5      |
| MASPI  | Mannan-binding lectin serine protease 1 | +   |                      | -5.49b      | -74.8      |
| TTR    | Transthyretin                    | +   |                      | -5.55b      | -53.3      |
| ITIH1  | Inter-α1 -trypsin inhibitor heavy chain 1 | +   |                      | -5.79b      | -135.7     |

---

*a* Not identified in singly purified sample.

*b* Not identified in doubly purified sample.
affinity purified proteins. All proteins that reached threshold significance (see “Experimental Procedures”) are listed in Supplemental Table 2 along with identification statistics achieved with each sample preparation method and mass spectrometric approach. These data are summarized in Supplemental Table 1. Both tryptic digests of the total unfractonated mixtures and gel slices of material fractionated by 1D gel electrophoresis were analyzed. Nanospray LC-MS/MS was conducted using both a linear ion trap (ThermoElectron LTQ) and a quadrupole time-of-flight instrument (Micromass Q-TOF). Results obtained using the different instruments and samples are summarized in Supplemental Table I. The doubly purified material was also further fractionated by 2D gel electrophoresis and analyzed using the LTQ (Supplemental Table 1). For the Q-TOF, we found that prefractionation of the singly purified sample by SDS-PAGE provided more overall unambiguous lysosomal protein identifications than analysis of the unfractonated solution digest. In contrast, prefractionation of the doubly purified sample by SDS-PAGE resulted in fewer lysosomal protein identifications with this instrument when compared with the unfractonated solution digest. This likely reflects the fine balance between sample losses due to 1D SDS-PAGE and inefficiencies in tryptic digestion of proteins in gel slices versus the reduction in sample complexity. The doubly purified sample, being less complex than the singly purified sample, will thus receive less benefit from prefractionation. With the LTQ, prefractionation of either sample by SDS-PAGE resulted in fewer protein identifications than the unfractonated mixture; this probably reflects the faster duty cycle of the LTQ, meaning that sample complexity is less likely to be a limiting factor. Prefractionation by 2D gel electrophoresis resulted in a significant decrease in identification in the number of proteins identified. In terms of LC-MS/MS approach, of a total of 148 unique protein assignments exceeding threshold expectation values, 147 were made with the LTQ compared with 94 with the Q-TOF (Supplemental Table 2).

Identification of Purified Plasma Proteins

Data obtained using the different analytical approaches were evaluated in terms of identification of known lysosomal proteins (Supplemental Tables 1 and 2) to determine which provided the greatest number of unambiguous assignments. The most effective approach was analysis of unfractonated solution digests using the LTQ, which resulted in the identification of 44 of a total of 45 lysosomal proteins that are found when results from all methods are combined. As a single, readily comparable dataset, we have therefore restricted our analyses here (Tables I–V) to data generated with unfractonated solution digests and the LTQ, although proteins assignments made using all approaches are provided in Supplemental Table 2.

Forty-four known lysosomal proteins were identified by LTQ analysis of the unfractonated solution digest of the mixture of Man-6-P glycoproteins purified from human plasma, and as would be expected, the majority (37 of 44) were identified in both the singly and doubly purified preparations. Six proteins achieved significance only in the doubly purified mixture and were presumably masked by contaminants in the more complex, singly purified mixture. Previously we examined human brain lysosomal proteins using MPR affinity chromatography (17, 24), and when combined with this study, a total of 53 lysosomal proteins were identified from these two human sources. The complement of lysosomal proteins identified from both sources was for the most part very similar with 41 proteins found in both brain and plasma. However, some proteins were found in brain but not in plasma (ARSB, CTSF, GALC, GLA, GM2A, IDUA, LIPA, SMPD1, and PPT2). In contrast, three lysosomal proteins (ACP5, CPVL, and heparanase) were found only in plasma.

The Plasma Proteome Project of the Human Proteome Organization (www.bioinformatics.med.umich.edu/hupo) has cataloged a high confidence (>95%) subset of the total plasma proteome database representing 889 plasma proteins (25). It is worth noting that of the known lysosomal proteins identified in this study, only one (γ-glutamyl hydrolase) is currently included in the high confidence dataset. This reflects the relatively low concentration of lysosomal proteins in plasma and the corresponding difficulties associated with their isolation and identification in a highly complex mixture.

Identification of Lysosomal Candidates

One of the aims of this and similar MPR affinity purification studies of the human proteome is to identify novel lysosomal candidates for further study. In our previous analysis of human brain (17), the number of proteins identified that were not classified as lysosomal was relatively small compared with the number of known lysosomal proteins; thus selection of candidates, based upon their respective biological and biochemical properties, was relatively straightforward. In contrast, in our analysis of the unfractonated solution digests of the mixture of proteins purified from plasma, we identified 80 different proteins (excluding contaminant immunoglobulin chains) that are not currently classified as lysosomal (Table II). Here selection of candidates of interest is not straightforward, especially given the significant challenges inherent in distinguishing a subset of plasma proteins with specific properties (e.g., the presence of carbohydrates containing Man-6-P) from highly abundant contaminants. We have therefore developed several experimental and bioinformatic approaches to help in the process of distinguishing the more probable lysosomal candidates from other purified proteins.

One approach is based on the hypothesis that true Man-6-P glycoproteins (and specific contaminants) will be enriched after two rounds of purification compared with a single purification as contaminants are removed. We used the X!Tandem log(ℓ) score (the sum log intensities for all matched MS/MS fragment ions for every peptide assigned to a given protein) from the LTQ analyses of unfractonated mixtures as a meas-
 Proteins not currently assigned lysosomal function

| Protein | Site | Score |
|---------|------|-------|
| Acid sphingomyelinase-like phosphodiesterase | 238NS | -7.7 |
| Afamin | 33NST | -3.2 |
| α1-Acid glycoprotein 2 | 33NAT | -3.3 |
| α1-Antitrypsin | 56NKS | -5.9 |
| α2-HS-glycoprotein | 72NKT | -3.5 |
| Cat eye syndrome critical region protein 1 | 349NAT | -10.3 |
| Ceruloplasmin | 138NTT | -7.8 |
| Complement C1q | 146NH9 | -6.8 |
| Corticosteroid-binding globulin | 330NTS | -9.3 |
| CREG1 protein | 216NT | -4.8 |
| Fc fragment of IgG-binding protein | 76NT | -5.1 |
| Haptoglobin | 123NLS | -6.0 |
| Hemopexin | 187NCS | -2.6 |
| Kallistatin | 157NTT | -5.9 |
| Leucine-rich α2-glycoprotein | 325NTT | -7.7 |
| Mammalian ependymin-related protein | 250NTS | -9.0 |
| Plasma serine protease inhibitor | 262NAT | -4.7 |
| Prostaglandin-H2 d-isomerase | 78NLT | -13.4 |
| Ribonuclease pancreatic | 104NNS | -7.4 |
| Ribonuclease T2 | 106NRS | -6.5 |
| Serotransferrin | 212NT | -4.9 |
| Stress 70 protein chaperone | 630NTV | -9.3 |
| Thymosine-binding globulin | 184NST | -6.3 |
| Zinc-α2-glycoprotein | 109NDS | -6.1 |
| 115NKS | -8.8 |

* Potential sites of Man-6-phosphorylation that were found on doubly glycosylated peptides for which singly glycosylated equivalents were not identified. These sites clearly contain endoglycosidase H-sensitive carbohydrates (indicated by the N*) modification, and at least one site must contain the Man-6-P modification. However, assignments of Man-6-phosphorylation to both sites should be regarded as tentative.

* Peptides containing phosphorylated glycosylation sites that were not identified in the mixture of MPR-affinity purified proteins from human brain.

The Human Plasma Mannose 6-Phosphate Glycoproteome

| Peptide log(e) values are shown. AMBP, α1-microglobulin/bikunin precursor. | Table III |
|---|---|
| Lysosomal proteins | Protein Site Score |
| | AMBP protein 250NGT | -6.2 |
| Acid ceramidase | 47NED | -2.3 |
| β -N-Acetylgalactosaminidase | 124NFT | -2.8 |
| β -Gluconoridase | 272NGT | -2.3 |
| β -Hexosaminidase α chain | 157NKT | -3.5 |
| β -Hexosaminidase β chain | 84NST | -4.2 |
| Cathepsin D | 327NTT | -6.8 |
| Cathepsin F | 265NT | -3.0 |
| Cathepsin L | 367NFS | -3.9 |
| Deoxyribonuclease II | 376NDS | -6.0 |
| Di-N-acetylchitobiasin | 221NDS | -8.7 |
| Dipeptidyl-peptidase I | 266NCS | -6.9 |
| NCP2 protein | 290NST | -4.6 |
| γ -Glutamyl hydrolase | 280NIS | -2.8 |
| Iduronate 2-sulfatase | 246NIT | -7.6 |
| Legumain | 290NIS | -9.9 |
| Lysosomal α-glucosidase | 470NIT | -5.9 |
| Lysosomal protective protein | 149NDS | -9.3 |
| Lysosomal Pro-X carboxypeptidase | 47NYS | -6.5 |
| Myeloperoxidase | 336NYS | -6.4 |
| N-Acetylgalactosamine-6-sulfatase | 317NVS | -6.4 |
| N-Acetylglucosamine-6-sulfatase | 415NIS | -6.3 |
| N-Sulfoglucomucosamine sulfohydrolase | 322NIT | -5.5 |
| Palmitoyl-protein thioesterase 1 | 204NLT | -6.1 |
| Plasma glutamate carboxypeptidase | 189NYS | -6.1 |
| Proactivator polypeptide | 210NVN* | -8.1 |
| Probable serine carboxypeptidase CPVL | 279NSS | -5.9 |
| Sphingomyelin phosphodiesterase | 362NAT | -2.5 |
| Tissue α-L-fucosidase | 415NRT | -2.8 |
| Tripeptidyl-peptidase I | 197NHS | -7.3 |
| | 235NDS | -4.7 |
| | 179NYS | -6.2 |
| | 353NIS | -4.4 |
| | 80NAT | -9.3 |
| | 332NKT | -4.3 |
| | 426NST | -4.2 |
| | 81NKT | -2.9 |
| | 337NHS | -2.6 |
| | 377NTT | -2.6 |
| | 210NLT* | -8.4 |
| | 222NNS* | -8.4 |
| Proteins not currently assigned lysosomal function | 238NS | -7.7 |
| Acid sphingomyelinase-like phosphodiesterase | 33NSTT | -3.2 |
| | 33NAT | -3.3 |
| | 56NKS | -5.9 |
| | 72NKT | -3.5 |
| | 271NAT | -6.9 |
| | 156NDT | -3.1 |
after two rounds of purification was increased have known or
predicted enzymatic properties that might be consistent with
lysosomal function. These include several ribonucleases, a
fucosidase, a homolog of acid sphingomyelinase, and two
hypothetical proteins with some similarity to phospholipases.
However, there are also a number of proteins without reported
enzyme activity and several protease inhibitors that may have
been purified while bound to true Man-6-P glycoproteins (i.e.
specific contaminants).

Although a comparison of the intensity scores obtained for
individual proteins after a single versus double purification
may provide some clues toward the identification of true Man-6-P
glycoproteins, this is an indirect approach that cannot
differentiate between true Man-6-P glycoproteins and specific
contaminants. In addition, low abundance and/or sticky lysosom-
al proteins might be lost due to the increased sample
handling. As an alternative experimental classification ap-
proach, we have recently developed an affinity purification
protocol for the isolation of Man-6-P glycopeptides in a form
suitable for identification by LC-MS/MS, allowing direct veri-
fication of the presence of Man-6-P (24).

Forty-two Man-6-P glycopeptides were assigned to 28 of
44 (64%) of the known lysosomal proteins found in human
plasma (Table III and Fig. 2A). Most of these peptides have

| Table IV |
|-----------------------|
| **Known or predicted protease inhibitors purified from human plasma by MPR affinity chromatography** |
| “Man-6-P peptides” indicates whether Man-6-P glycopeptides corresponding to each protein were identified in the plasma samples. AMBP, α 1-microglobulin/bikunin precursor. |
| **Gene** | **Protein** | **Type of inhibitor** | **Domain** | **Man-6-P peptides** |
| A2M | α 2-Macroglobulin | Serine protease | | |
| AGT | Angiotensinogen | Serine protease | SERPIN | + |
| AHSG | α 2-HS-glycoprotein | Cysteine protease | Cystatin | + |
| AMBP | AMBP protein | Serine protease | IATI* | + |
| C3 | Complement C3 precursor | Endopeptidase | | |
| C4B | Complement component 4B | Endopeptidase | | |
| FCGBP | Fc fragment of IgG-binding protein | Serine protease | IATI | + |
| FETUB | Fettin-B | Cysteine protease | Cystatin | |
| HRG | Histidine-rich glycoprotein | Cysteine protease | Cystatin | |
| ITIH1 | Inter-α -trypsin inhibitor heavy chain 1 | Serine protease | | |
| ITIH2 | Inter-α -trypsin inhibitor heavy chain 2 | Serine protease | | |
| ITIH4 | Inter-α -trypsin inhibitor heavy chain 4 | Serine protease | | |
| KNG1 | Kininogen-1 | Cysteine protease | Cystatin | |
| SERPINA1 | α 1-Antitrypsin | Serine protease | SERPIN | + |
| SERPINA3 | α 1-Antichymotrypsin | Serine protease | SERPIN | |
| SERPINA4 | Kallistatin | Serine protease | SERPIN | + |
| SERPINA5 | Plasma serine protease inhibitor | Serine protease | SERPIN | + |
| SERPINA6 | Corticosteroid-binding globulin | Serine protease | SERPIN | |
| SERPIN C1 | Antithrombin-III | Serine protease | SERPIN | + |
| SERPIN F1 | Pigment epithelium-derived factor | Serine protease | SERPIN | |
| SERPING1 | Plasma protease C1 inhibitor | Serine protease | SERPIN | |

* IATI, inter-α -trypsin inhibitor.

| Table V |
|-----------------------|
| **Candidate lysosomal proteins identified in human plasma** |
| **Gene code** | **Protein name** | **Man-6-P glycopeptides?** |
| | | **Plasma** | **Brain** |
| SMPDL3A | Acid sphingomyelinase-like phosphodiesterase 3A | + | |
| RNASET2 | Ribonuclease T2 | + | + |
| RNASE6 | Ribonuclease K6 | + | |
| RNASE1 | Ribonuclease pancreatic | + | |
| NP_775813.1 | Laminin A LOC196463 | + | |
| NP_079105.3 | Laminin A FLJ22662 | + | |
| FUA2 | Plasma α-L-fucosidase | + | |
| EL2 | Leukocyte elastase | + | |
| CECR1 | Cat eye syndrome critical region 1 | + | |
been confirmed to contain Man-6-P in the analysis of lysosomal proteins from human brain (24), although a novel Man-6-P site was found here for di-N-acetylchitobiase, and a site was also identified for CPVL, a putative peptidase not identified in brain. In brain, we identified Man-6-P glycopeptides corresponding to 39 of 50 (78%) of the lysosomal proteins identified in the total mixture. This is a greater proportion than observed for plasma and may reflect the greater complexity of the Man-6-P glycopeptide mixture from this source.

Sites for Man-6-phosphorylation were also identified in 28 proteins that were not previously assigned lysosomal function (Table III). Similar to the set of known lysosomal proteins identified in plasma, these plasma proteins were enriched in the doubly purified sample. A number of these proteins (acid sphingomyelinase-like protein 3a (SMPDL3A), biotinidase (BTD), cellular repressor of E1A-regulated gene transcription (CREG1), mammalian ependymin-related protein (EPDR1), prostaglandin-H₂ d-isomerase (PTGDS), and RNASET2) were also previously found to contain Man-6-P in human brain (24). However, an intriguing observation was the identification of Man-6-P glycopeptides derived from a large number of classical plasma proteins. Some of these proteins, e.g. haptoglobin, transferrin, hemopexin, and α₂-HS-glycoproteins, are extremely abundant plasma species in the 0.5–3g/liter range that we originally considered to be likely nonspecific contaminants. Classical plasma glycoproteins for which we identified Man-6-P isoforms in this study can be classified into several broad categories.

Protease Inhibitors—One striking observation was the number of protease inhibitors that were purified from human plasma by MPR affinity chromatography (Table IV). In total, 21 protease inhibitors or proteins containing protease inhibitor domains were found, representing 26% of the proteins identified that have not been assigned lysosomal function, and of these, eight are members of the SERPIN family. Given that many lysosomal Man-6-P glycoproteins are proteases, we previously suggested that such protease inhibitors probably represent specific contaminants (17). This is clearly the case with some, e.g. cystatins, which cannot contain Man-6-P as their primary sequence lacks the consensus motif for N-linked glycosylation. However, the presence of Man-6-P is demonstrated in eight of the protease inhibitors found in the mixture of MPR-purified proteins (as well as in thyroxin-binding globulin (SERPINA7), which was not identified with threshold confidence in the analysis of the unfractionated solution digests), indicating that there are glycoforms of these proteins that do contain Man-6-P. Interestingly previous studies have demonstrated the presence of Man-6-P on several proteins containing the SERPIN domain that are secreted by the uterus of pregnant sheep and pigs including uterine milk proteins and uteroferrin-associated protein (26, 27). In addition, another member of the SERPIN family, C1 esterase inhibitor, may also contain Man-6-P under some circumstances (28).

Plasma Transporters—Vitamin E-binding protein afamin and metalloproteins ceruloplasmin (CP) and serotransferrin are all plasma transporters that appear to contain Man-6-P. Haptoglobin and hemopexin scavenge free hemoglobin from the bloodstream, and both were also found to contain Man-6-P. Other transport proteins found here include α₁-acid glycoprotein (ORM1 and ORM2) and fatty acid-binding protein 5.

Enzymes—Twenty-two of the proteins that are not classified as lysosomal have known enzymatic function or may be predicted to have catalytic function based on sequence homology with known enzymes. This category includes a number of abundant plasma glycoproteins such as CP, cholinesterase (BCHE), BTD, and PTGDS and complement factors (e.g. C2, C3, and BF) as well as a number of less abundant enzymes...
species that are not found in the Plasma Proteome Project high confidence dataset. The relatively high prevalence of proteins with enzymatic function probably reflects the inclusion in this category of a number of novel lysosomal hydrolases.

Circulating Glycoproteins of Unknown or Unclear Function—Man-6-P glycopeptides were assigned to a number of plasma proteins of unknown or unclear function including $\alpha_1$-acid glycoprotein 2 (ORM2), $\alpha_1$B-glycoprotein (A1BG), leucine-rich $\alpha_2$-glycoprotein (LRG1), and zinc-$\alpha_2$-glycoprotein (AZGP1).

DISCUSSION

In this study, we used an MPR affinity purification approach to isolate and characterize the proteome of circulating lysosomal proteins. In total, we identified 44 different lysosomal proteins that are associated with 23 lysosomal storage diseases. Lysosomal storage diseases are frequently difficult to diagnose using current clinical and biochemical approaches as presentation of genetically distinct diseases is often similar. In addition, different types of mutations within a single gene can result in a wide range of manifestations of disease (e.g. onset, severity, and symptom expression). One approach that may help definitive diagnoses in these diseases is the comparative proteomic profiling of disease and control samples to identify defective or deficient proteins. However, the complexity of the plasma proteome coupled with the low abundance of lysosomal proteins precludes a direct analysis of this subclass of plasma constituents. One of the aims of this study was therefore to determine whether MPR affinity purification of Man-6-P glycoproteins might form the basis for a comparative proteomic method for the global diagnosis of lysosomal storage diseases.

There are several approaches for proteomic profiling (29) that could be applicable to plasma Man-6-P glycoproteomes. One of the more frequently utilized methods is comparative two-dimensional gel electrophoresis, and in this study, we generated a 2D map of plasma Man-6-P glycoproteins (Supplemental Fig. 1 and Supplemental Table IV) as a potential clinical reference. Two problems were immediately apparent with this approach. First, LC-MS/MS analysis of even well resolved 2D gel spots (e.g. Supplemental Fig. 1 and Supplemental Table IV, gel spots 5–8) revealed a mixture of proteins; thus a deficiency in any given protein could potentially be difficult to identify. Second, limitations in the amount of clinical samples available, especially in young children undergoing diagnosis, would greatly limit use of the 2D approach. In terms of yield, we obtained ~200 $\mu$g of Man-6-P glycoproteins/liter after a double purification; thus in a clinical setting, yields of ~ 1 $\mu$g/ml of plasma could be achievable with appropriate method development. With such amounts, most of the individual spots on a stained 2D gel would be below the limits of detection. Sensitivity could be increased by visualization of Man-6-P glycoproteins using a labeled MPR probe (22), but this would result in a decrease in resolution (for example, in Fig. 1, compare A with B and C with D). Therefore, although the comparative profiling of plasma Man-6-P glycoproteins by 2D gel electrophoresis might have investigational applications, these issues combined with the technical expertise required suggest that this approach is probably unsuitable as a widely applicable method for the diagnosis of lysosomal storage diseases.

An alternative approach that may overcome the drawbacks of 2D gel electrophoresis is quantitative mass spectrometry for the comparison of Man-6-P glycoproteins from patients and normal reference controls (30). In this way, Man-6-P glycoproteins that are deficient as a consequence of a lysosomal storage disease may be readily identifiable, and a preliminary diagnosis may be made that can be verified using standard biochemical and/or DNA-based tests. This approach is more specific than comparative 2D gel electrophoresis and may be compatible with the projected yields of Man-6-P glycoproteins from clinical samples. Quite apart from its utility in the diagnosis of storage disorders, mass spectrometry-based clinical methods could also prove useful in investigating the role of lysosomal proteins in more widespread human diseases, including cancer, where alterations in lysosomal activities have been implicated. In addition, such approaches could also prove useful in evaluating the potential of lysosomal proteins as plasma biomarkers of prognostic or diagnostic value. Implementation of quantitative methods (e.g. isobaric peptide tagging (31)) combined with the enrichment methods outlined in this study may be useful in this respect.

Another aim was to identify Man-6-P glycoproteins for further study that could potentially represent new lysosomal proteins. A large number of proteins were identified in this study that have not been assigned lysosomal function so we have used several criteria to filter the data to identify the most promising potential candidates. The first criterion is that a candidate should be enriched after a second round of purification reflecting the loss of nonspecific contaminants. The second criterion for a candidate is that it should either be or be predicted to be a hydrolase like most soluble lysosomal proteins. Forty-one of 44 lysosomal proteins found in plasma here fall into this class of enzymes. The third criterion is that a candidate should not represent an abundant plasma protein. Lysosomal proteins are low abundance constituents of plasma, and this is reflected by the fact that only one of the known lysosomal proteins found here is currently represented in the Plasma Proteome Project high confidence dataset (25) of plasma proteins (Table I). Categorization of both plasma lysosomal proteins and proteins not assigned lysosomal function with respect to these three criteria is illustrated in Fig. 3. Of the known lysosomal proteins, most (38 of 44) are hydrolases that are absent from the Plasma Proteome Project and that are enriched by two rounds of affinity purification. Based upon the assumption that novel lysosomal proteins will have
properties similar to the known lysosomal proteins, we were able to identify nine highly promising candidates. These are summarized in Table V. Most of these proteins, i.e. the two phospholipase-like hypothetical proteins, SMPDL3A, plasma α-fucosidase (FUCA2), and several ribonucleases, have homology to known lysosomal proteins and are thus particularly promising. EPDR1 and CREG1 are also worth further investigation as they are low abundance, confirmed Man-6-P glycoproteins for which database searching reveals no similarity to other proteins that might provide clues to function. However, it should be emphasized that the objective of this and other large scale proteomic studies is to identify candidates of interest that will require further extensive characterization at the functional and compartmental levels before their actual cellular roles can be precisely ascertained.

An unexpected finding of this study was that a number of highly abundant classical plasma proteins were purified as Man-6-P-containing glycoforms (Table VI). Examples of such proteins include afamin, α1-acid glycoprotein, ceruloplasmin, and haptoglobin. It is possible that detection of Man-6-phosphorylated glycoforms of these proteins may arise from a lack of absolute fidelity that occurs in biological processes (see below) and may result in a very small proportion of the proteins receiving Man-6-P. Thus, it is worth considering the relative amount of a given protein that is in the phosphorylated form in plasma. We previously used enzyme activity measurements and MPR affinity chromatography to estimate that, for any given lysosomal protein, 5–50% of the total enzyme activity present in plasma was in the Man-6-phosphorylated form (16). Here after two rounds of purification, we obtained ~200 μg of Man-6-P glycoproteins/liter of plasma, representing 132 different proteins with an average yield of ~1.5 μg/liter for each individual protein, which is consistent with the predicted concentrations of lysosomal proteins containing Man-6-P in plasma. For example, total α-glucosidase is normally present at a median concentration of 17 μg/liter of plasma (32); thus purification of 1.5 μg/liter would correspond to 9% in the Man-6-P-containing form. In accordance, we previously determined by enzymatic assay that 28% of α-glucosidase in plasma actually contains Man-6-P (16). In contrast, we estimate that the proportion of the individual classi-

---

**TABLE VI**

| Protein Name                                           | Concentration | PPP entry | Ref. |
|--------------------------------------------------------|---------------|-----------|------|
| Afamin precursor (α-albumin)                           | 59.8 mg/liter | +         | 38   |
| α1-Acid glycoprotein 2                                 | 0.5–1.2 g/liter | + | 39   |
| α1-Antitrypsin precurso                                   | 0.85–1.85 g/liter | + | 39   |
| α2-HS-glycoprotein                                      | 400–900 mg/liter | + | 40   |
| AMBP protein precursor                                   | ~54 mg/liter  | +         | 41   |
| Angiotensinogen precursor                               | ~1.7 mg/liter | +         | 42   |
| Antithrombin-III                                        | ~290 mg/liter | +         | 43, 44 |
| Biotinidase                                             | 2 mg/liter    | +         | 45   |
| Ceruloplasmin                                           | 250–550 mg/liter | + | 39   |
| Complement C1q                                           | 65 mg/liter   | +         | 46   |
| Corticosteroid-binding globulin                         | 24–44 mg/liter | + | 47   |
| Fc fragment of IgG-binding protein                       | +             |           |      |
| Haptoglobin                                              | 1.5–2.1 g/liter | + | 48   |
| Hemopexin                                               | ~720 mg/liter | +         | 49   |
| Kallistatin                                             | 22 mg/liter   | +         | 50   |
| Leucine-rich α2-glycoprotein                            | 21 mg/liter   | +         | 51   |
| Plasma serine protease inhibitor (SERPINA5)             | >200 mg/liter | +         | 52   |
| Prostaglandin-H2 o-isomerase                            | 0.5 mg/liter  | +         | 53   |
| Serotransferrin                                          | 1.5–3 g/liter | +         | 39   |
| Thyroxine-binding globulin                              | ~20 mg/liter  | +         | 54   |
| Zinc-α2-glycoprotein                                    | ~100 mg/liter | +         | 55   |
cal plasma proteins that contains Man-6-P is much lower. The average concentration of the classical plasma proteins purified here in Man-6-P glycoforms is ~500 mg/liter (Table VI); thus the average proportion of each that contains Man-6-P would be $1.5 \times 10^{-4}\%$ of each respective total; this is many orders of magnitude lower than that observed for known lysosomal proteins.

Although only a very small proportion of the classical plasma proteins appear to receive the Man-6-P modification, these glycoforms are probably detected here simply because these proteins are so abundant. These glycoproteins may represent substrates for the GlcNAc-phosphotransferase that are recognized with low affinity compared with lysosomal proteins bearing similar glycans in the same way that the non-lysosomal glycoprotein ribonuclease B can be Man-6-phosphorylated but with lower efficiency than the lysosomal protein uteroferrin (33). Thus, we expect that the GlcNAc-phosphotransferase will have a spectrum of $K_m$ values for different glycoproteins with bona fide lysosomal proteins exhibiting high substrate specificity constants ($k_{cat}/K_m$) for the phosphotransferase and thus being Man-6-phosphorylated preferentially. Other glycoproteins will exhibit a range of lower substrate specificity constants and, depending on the upper value of this range, may be phosphorylated at a low level. This may reflect an intrinsic balance between the requirement for the phosphotransferase to specifically recognize many different types of lysosomal proteins and its ability to distinguish non-lysosomal and lysosomal proteins. Should the concentration of low affinity non-lysosomal proteins be extremely high, then this would allow them to compete with lysosomal proteins for the phosphotransferase and receive the Man-6-P modification. Given the three-dimensional nature of the determinant for Man-6-phosphorylation, it is likely that the ability of the GlcNAc-phosphotransferase to distinguish between lysosomal and non-lysosomal proteins is less stringent if the latter bear some structural similarity to the former. An example is the non-lysosomal protein renin that receives the Man-6-P modification and that is structurally similar to cathepsin D (34).

It is possible that the SERPIN domain bears some structural similarity to the lysosomal determinant for Man-6-phosphorylation; this could explain why we found that a number of proteins of this family contain Man-6-P. Alternatively it is possible that such proteins are associated with bona fide lysosomal proteins in transit through the Golgi and become phosphorylated due to a proximity effect.

In addition to receiving the Man-6-P modification, isofoms of non-lysosomal proteins must also escape the cellular pathway for targeting to the lysosome to allow secretion or must be delivered into the plasma by some other mechanism (see the Introduction). Given that they may be Man-6-phosphorylated aberrantly, these proteins could also represent low affinity ligands for the MPRs that are not efficiently recognized and targeted to the lysosome but that are instead secreted. However, the high concentration of MPR used in the affinity chromatography would allow such proteins to be purified. For example, cathepsin L is a protein that binds to the MPR with relatively low affinity and, as a result, is largely secreted in cell culture (35, 36), but it is still purified by MPR chromatography (15, 17, 21).

Irrespective of biological significance and source, the observation of Man-6-P glycoforms of abundant classical plasma proteins has important implications for the increasing number of studies directed toward the characterization of lysosomal proteomes using MPR affinity chromatography (15, 17, 19, 20, 37). Although the Man-6-P modification is suggestive of lysosomal localization and can provide a valuable clue toward novel candidates, its detection does not obviate the necessity for detailed characterization of subcellular localization and functional analyses. Regardless the ability to detect and quantify a small and distinct subset of plasma proteins with both established and proposed biomedical importance is likely to have valuable future application in the discovery of useful biomarkers for human disease.

Acknowledgment—We thank Dr. Zygmund Roth (V. I. Technologies, Inc.) for generously providing plasma samples.

REFERENCES

1. Holtzman, E. (1989) Lysosomes, Plenum Press, New York
2. Ghosh, P., Dahms, N. M., and Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twists in the tale. Nat. Rev. Mol. Cell. Biol. 4, 202–212
3. Lentner, C., Lentner, C., and Wink, A. (1984) Geigy Scientific Tables, pp. 166–199, Ciba-Geigy, West Caldwell, NJ
4. Ludwig, T., Munier-Lehmann, H., Bauer, U., Hollinshead, M., Ovitt, C., Lobel, P., and Hoffack, B. (1994) Differential sorting of lysosomal enzymes in mannos 6-phosphate receptor-deficient fibroblasts. EMBO J. 13, 3430–3437
5. Pohlmann, R., Boeker, M. W., and von Figura, K. (1995) The two mannos 6-phosphate receptors transport distinct complements of lysosomal proteins. J. Biol. Chem. 270, 27311–27318
6. Kasper, D., Dittmer, F., von Figura, K., and Pohlmann, R. (1996) Neither type of mannos 6-phosphate receptor is sufficient for targeting of lysosomal enzymes along intracellular routes. J. Cell Biol. 134, 615–623
7. Munier-Lehmann, H., Maxion, F., Bauer, U., Lobel, P., and Hoffack, B. (1996) Re-expression of the mannos 6-phosphate receptors in receptor-deficient fibroblasts. Complementary function of the two mannos 6-phosphate receptors in lysosomal enzyme targeting. J. Biol. Chem. 271, 15166–15174
8. Scriver, C. R., Beaudet, A. L., Sly, W. S., Valle, D., Childs, B., Kinzler, K. W., and Vogelstein, B. (eds) (2001) The Metabolic & Molecular Bases of Inherited Disease, 8th Ed, Vol. III, pp. 134–3984, McGraw-Hill, New York
(1992) Protein C inhibitor in human body fluids. Seminal plasma is rich in inhibitor antigen deriving from cells throughout the male reproductive system. J. Clin. Investig. 89, 1094–1101
53. Tumani, H., Reiber, H., Nau, R., Prange, H. W., Kauffmann, K., Mader, M., and Felgenhauer, K. (1998) Beta-trace protein concentration in cerebrospinal fluid is decreased in patients with bacterial meningitis. Neurosci. Lett. 242, 5–8
54. Cavalieri, R. R., McMahon, F. A., and Castle, J. N. (1975) Preparation of 125I-labeled human thyroxine-binding a globulin and its turnover in normal and hypothyroid subjects. J. Clin. Investig. 56, 79–87
55. Jirka, M., Blanicky, P., and Cerna, M. (1974) The Zn-α2-glycoprotein level in human serum during ontogenesis. Clin. Chim. Acta 56, 31–36