The Mannose 6-Phosphate Glycoprotein Proteome

David E. Sleat,* †,‡ Maria Cecilia Della Valle, †,‡ Haiyan Zheng, † Dirk F. Moore, § and Peter Lobel* †,‡

Center for Advanced Biotechnology and Medicine and Departments of Pharmacology and Biostatistics, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Received February 18, 2008

Most luminal lysosomal proteins are synthesized as precursors containing mannose 6-phosphate (Man6-P) and a number of recent studies have conducted affinity purification of Man6-P containing proteins as a step toward defining the composition of the lysosome. Approximately 60 known lysosomal proteins have been found in such studies as well as many other Man-6-P glycoproteins, some of which represent new lysosomal proteins. The latter are of considerable interest from cell-biological and biomedical perspectives, but differentiating between them and other proteins remains a significant challenge. The aim of this study was to conduct a global analysis of the mammalian Man6-P glycoproteome, implementing technical and biostatistical methods to aid in the discovery and validation of lysosomal candidates. We purified Man6-P glycoproteins from 17 individual rat tissues. To distinguish nonspecific contaminants (i.e., abundant or “sticky” proteins that are not fully removed during purification) from specifically purified proteins, we conducted a semiquantitative mass spectrometric comparison of protein levels in nonspecific mock eluates versus specific affinity chromatography eluates to identify those proteins that are specifically purified. We identified 60 known lysosomal proteins, representing nearly all that are currently known to contain Man-6-P. We also find 136 other proteins that are specifically purified but which are not known to have lysosomal function. This approach provides a list of candidate lysosomal proteins and also provides insights into the relative distribution of Man6-P glycoproteins.

Keywords: lysosomal protein • mannose 6-phosphate • affinity purification • proteomic

Introduction

The lysosome is a eukaryotic organelle that plays a critical role in the degradation and recycling of cellular macromolecules including proteins, carbohydrates, nucleic acids and lipids. The catabolic function of the lysosome is conducted by the concerted action of soluble luminal hydrolases and their accessory proteins, as well as transmembrane proteins that function in vesicular transport, catalysis and molecular transport. To date, approximately 60 soluble lysosomal proteins have been described and this number continues to increase. The number of lysosomal transmembrane proteins has not been well-defined, although recent proteomic studies indicate that they appear to be numerous (see below).

Lysosomes and lysosomal proteins are of considerable biomedical importance as they are directly involved or have been implicated in numerous human diseases. Defects in lysosomal function result in lysosomal storage disorders which is a group of over 40 inherited diseases that are frequently progressive, neurodegenerative and which usually result in decreased life-span. In addition, alterations in lysosomal function have been implicated in cancer and metastasis, Alzheimer disease, immune system dysfunction and other widespread human diseases.

Given these links with human disease, there is considerable interest in defining the scope of cellular functions for the lysosome and one direction in which this has been recently explored is in the proteomic characterization of its constituent proteins (reviewed in ref 3). A particular emphasis of these studies has been in the identification of new lysosomal proteins to better understand the function of this organelle but also to identify candidates for the defective proteins in human lysosomal storage diseases of unknown etiology. Different approaches have been used in the proteomic characterization of lysosomal proteins and each has inherent advantages and disadvantages.

Proteomic surveys have been conducted on subcellular fractions enriched for lysosomes by gradient centrifugation. This approach allows for the identification of both soluble and transmembrane lysosomal proteins but, because lysosomes cannot be isolated to homogeneity due to an intrinsic overlap in the density of cellular organelles, enrichment for lysosomal proteins is relatively modest using such techniques (typically 50- to 100-fold). Thus, proteomic studies based upon subcel-
lular fractionation alone are prone to false positive errors in terms of assignment of lysosomal localization. However, as improvements in preparative methods and statistical analysis of data are implemented, the accuracy of lysosomal assignments from such studies appears to be increasing.8

An alternative approach that allows for much greater enrichment of the subset of proteins that reside within the lumen of the lysosome is affinity purification based upon the presence of a specific carbohydrate modification, mannose 6-phosphate (Man6-P). Man6-P is found on N-linked glycans of most newly synthesized soluble lysosomal proteins and is recognized by two Man6-P receptors (MPRs) that direct the vesicular trafficking of lysosomal proteins from the Golgi to an acidic prelysosomal compartment.9 While lysosomal proteins in transit contain the Man6-P modification, the total amount of any given lysosomal protein in the Man6-P glycoform is dependent on source, as it may be rapidly removed in some tissue or cell-types but may persist in others. Thus, depending on the type of sample analyzed, 1 to ∼50% of a given lysosomal protein may contain Man6-P and such glycoforms can be purified from complex mixtures using immobilized soluble forms of the MPRs as an affinity purification reagent.10 This approach has been used to investigate the lysosomal proteomes from a number of sources including cultured cells and tissues.10–20 This method allows for considerable purification factors (e.g., >106-fold when Man6-P glycoproteins were purified from human plasma17), but there are important limitations. First, while strongly suggestive, the presence of Man6-P does not always equate with lysosomal localization. Second, differentiating between true Man6-P glycoproteins and contaminants can represent a significant hurdle. For example, in any sample purified by affinity chromatography on immobilized MPR, in addition to Man6-P glycoproteins, there are also proteins that do not contain Man6-P but which instead bind and copurify with true Man6-P glycoproteins (i.e., specific contaminants) as well as highly abundant or “sticky” cellular proteins that are not completely removed by affinity chromatography (i.e., nonspecific contaminants).

While these different approaches to the purification of lysosomal proteins have their own particular merits, a general limitation of all of the studies conducted to date is that they have been performed on limited numbers of sources and this could potentially restrict the number of proteins found. Lysosomes are found in all nucleated cell types and many acid hydrolases appear to be present in all lysosomes, but levels of individual lysosomal proteins vary considerably according to cell type and tissue. In addition, some lysosomal proteins are only expressed in highly specialized tissues and cell types. For example, granzymes A and B are lysosomal proteins that play a role in immune function and which appear to be restricted to cytotoxic T lymphocytes and natural killer cells.21 Variations in the distribution of lysosomal proteins were clearly shown in an analysis of rat tissues demonstrating that the content of Man6-P glycoproteins varies considerably in both quantitative and qualitative respects.10 Similarly, expression profiling of soluble lysosomal proteins in 45 human tissues based upon the detection of their respective transcripts (Figure 1A; Supporting Information Table 1) demonstrates some lysosomal proteins to be quite widely distributed (e.g., present in as many as 44 tissues based on transcript analysis), whereas expression of others is more limited. In addition, the number of tissues in which transcripts corresponding to each lysosomal protein are found increase with the total number of ESTs assigned to each protein (Figure 1B). Tissue distribution may be particularly relevant in the search for new lysosomal proteins which could potentially have escaped classification as such because of a restricted expression pattern.

In this study, we have surveyed the mammalian Man6-P glycoproteome from 17 individual rat tissues using methods that allow the micropurification of these proteins from limiting amounts of sample. We estimated protein abundance in specific versus nonspecific mock affinity column eluates to help differentiate between Man6-P glycoproteins and nonspecific contaminants. The combination of a global purification approach with bioinformatic methods to eliminate nonspecific contaminants has allowed the generation of a database of mammalian proteins that are specifically purified by MPR affinity chromatography, many of which represent previously unrecognized candidate lysosomal proteins.

**Experimental Procedures**

**Purification of Man6-P Glycoproteins.** Rat tissues from adult Sprague–Dawley rats that were euthanized using hypobaric CO2 were obtained from Zivic Laboratories, Inc. (Pittsburgh, PA). Tissue samples were derived from 2–4 animals depending on the size of the particular tissue sample. Affinity purification of Man6-P glycoproteins was essentially as described10 with a number of modifications to allow a small-scale procedure for limiting amounts of tissue sample. All procedures were conducted at 4 °C. Tissues were homogenized using a Brinkmann...
Polytron homogenizer (Westbury, NY) with 20 mm generator in 100 mL of phosphate buffered saline (PBS) containing protease and phosphatase inhibitors (defined as “PBS-I” and comprising PBS containing 5 mM beta-glycerophosphate and 2.5 mM EDTA, 1 µg/mL pepstatin A, 1 µg/mL leupeptin and 0.25 mM Pefabloc). Tween-20 was added to a final concentration of 0.2% and the homogenate was centrifuged at 40 000 g for 2 h. The resulting supernatant was filtered through Whatman 3MM paper to remove insoluble lipids and other aggregates. Supernatants were loaded overnight onto 4 mL bed volume columns of sCI-MPR coupled with Affigel 10 at a gravity flow. Each elution was then repeated and pooled to give 8 mL per elution fraction. The volume of the different elutions were each reduced to ∼100 µL using a Centrifcon YM10 centrifugal concentrator and the protein concentration was determined.

**Tandem Mass Spectrometry.** For each tissue, a sample of the specific (Man6-P) or nonspecific (mannose and glucose 6-phosphate) affinity purification eluate was heated for 10 min at 60 °C in reducing, denaturing SDS-PAGE sample buffer, then fractionated on precast 10% polyacrylamide gels (Invitrogen, Carlsbad, CA) until the bromophenol blue dye-front had run ∼1 cm into the gel. Gel slices corresponding to each sample were excised and cut into small pieces, reduced, alkylated with iodoacetamide and digested with trypsin as described. Samples were analyzed by LC-MS/MS using an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as described previously. For the Man6-P eluates, typically a portion of each digest corresponding to 1 µg of starting material was analyzed by LC-MS. For the brain, we also analyzed 10% of the mock eluate regardless of the aggregate score from the merged data with a threshold for protein assignment was available, the mouse gene identifier was instead used. Databases (rat, ENSEMBL, February 2006, version 48.34m, which contains 18 311 known genes; mouse, ENSEMBL April 2007 build of the NCBI m37 assembly, database version 48.37a, which contains 21 928 known genes) were searched using a local implementation of GPM-XE Manager version 2.1.0 (Beavis Informatics Ltd., Winnipeg, Canada) which contains 21 928 known genes) were searched using the MudPit option to produce a merged output file which allows for a consistent assignment of spectra to the human or mouse equivalents using the LOCATE subcellular location database. Assignment of subcellular locations for identified rat proteins were assigned forming data analysis, sample information (e.g., source, eluate for affinity purification and spectral count data) was extracted from the merged output file which allows for a consistent assignment of spectra to the human or mouse equivalents using the LOCATE subcellular localization database.

### Table 1. Rat Tissue Man6-P Glycopeptide Purification*

<table>
<thead>
<tr>
<th>tissue</th>
<th>yield (µg/g)</th>
<th>load (ng)</th>
<th>LC-MS load (ng)</th>
<th>specific eluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>brain</td>
<td>3.7</td>
<td>0.2</td>
<td>2</td>
<td>11.1</td>
</tr>
<tr>
<td>cecum</td>
<td>1.0</td>
<td>0.4</td>
<td>2</td>
<td>10.0</td>
</tr>
<tr>
<td>duodenum</td>
<td>2.3</td>
<td>0.9</td>
<td>19</td>
<td>4.7</td>
</tr>
<tr>
<td>heart</td>
<td>5.0</td>
<td>0.3</td>
<td>5</td>
<td>6.5</td>
</tr>
<tr>
<td>kidney</td>
<td>5.0</td>
<td>0.4</td>
<td>10</td>
<td>4.0</td>
</tr>
<tr>
<td>liver</td>
<td>20.0</td>
<td>0.3</td>
<td>29</td>
<td>0.8</td>
</tr>
<tr>
<td>lung</td>
<td>5.0</td>
<td>0.1</td>
<td>3</td>
<td>4.9</td>
</tr>
<tr>
<td>mammary gland</td>
<td>4.5</td>
<td>0.9</td>
<td>12</td>
<td>7.4</td>
</tr>
<tr>
<td>pancreas</td>
<td>3.4</td>
<td>1.1</td>
<td>14</td>
<td>7.4</td>
</tr>
<tr>
<td>stomach</td>
<td>5.0</td>
<td>0.3</td>
<td>5</td>
<td>5.5</td>
</tr>
<tr>
<td>placentia</td>
<td>2.0</td>
<td>2.2</td>
<td>13</td>
<td>16.4</td>
</tr>
<tr>
<td>skeletal muscle</td>
<td>3.0</td>
<td>0.5</td>
<td>9</td>
<td>5.6</td>
</tr>
<tr>
<td>spleen</td>
<td>3.5</td>
<td>0.2</td>
<td>5</td>
<td>4.5</td>
</tr>
<tr>
<td>testis</td>
<td>4.0</td>
<td>2.0</td>
<td>16</td>
<td>11.9</td>
</tr>
<tr>
<td>thymus</td>
<td>2.6</td>
<td>0.6</td>
<td>9</td>
<td>1.5</td>
</tr>
<tr>
<td>vas deferens</td>
<td>0.5</td>
<td>0.2</td>
<td>&lt;1</td>
<td>3.4</td>
</tr>
<tr>
<td>uterus</td>
<td>0.5</td>
<td>0.2</td>
<td>&lt;1</td>
<td>3.4</td>
</tr>
</tbody>
</table>

*Yields are expressed as micrograms of protein recovered per gram of tissue wet weight. Load indicates nanograms of protein in a given eluate that was analyzed per duplicate LC-MS/MS run. *nd, not determined.

**Statistical Analysis.** Relative quantitation of protein abundance in different eluates was conducted by comparing the total number of spectra assigned to each protein in each sample. Statistical analysis was essentially as described previously. In brief, the method of Wilson was used to calculate the upper and lower limits of the 95% confidence interval for the ratio of spectral counts found in the specific compared to mock eluate. Analyses were conducted using R version 2.5.0, which is open source software for statistical
**Results and Discussion**

MPR-affinity purification was conducted on 17 different rat tissues which were chosen primarily on the basis of availability of sufficient material for predicted yields of Man6-P glycoproteins for multiple LC-MS/MS analyses (i.e., 10–40 μg). For the majority of tissues examined, yields were adequate, although, for several (skin, thymus, uterus and vas deferens), source material was limiting and the subsequent yields of purified protein were less than optimal (Table 1) although not sufficiently low to preclude analysis (see Experimental Procedures). The highest relative yield was obtained from placenta (16 μg/g). In other tissues, as predicted from earlier blotting experiments, relative yields of Man6-P glycoproteins were high from brain and testis (11–12 μg/g tissue).

In total, 793 proteins were assigned that met our criteria for significance (Supporting Information Table 2; summarized in Table 2) although 21 assignments were to proteins that are not of rodent origin and were eliminated from the analysis as contaminants. Of the remaining 772 assignments from rat tissues, 60 known soluble lysosomal proteins were identified and this number is comparable with that found in proteomic analysis of different human and mouse tissue sources (respectively, 60 and 56 proteins in total16,19,20). It is worth noting that, in rat, we found two highly similar (97%) yet genetically distinct rat CLN5 paralogs, encoded by genes on chromosomes 2 and 15. Given the similarity between these proteins, individual spectra cannot for the most part be assigned individually and we have thus considered these two proteins as a single entity in our analysis.

A central aim of this study was to differentiate between true Man6-P glycoproteins and nonspecific contaminants. To this end, we used the spectral counting method27,28 to estimate the relative abundance of each protein in a given tissue sample that was released from the MPR affinity column using a glucose 6-phosphate/mannose (“mock”) or a Man6-P (“specific”) eluate. Our prediction was that true Man6-P glycoproteins (but possibly also specific contaminants associated with Man6-P glycoproteins, depending upon the strength of interaction) should be enriched in the Man6-P eluate relative to the mannose/glucose 6-phosphate eluate. In contrast, nonspecific contaminants (i.e., abundant or “sticky” proteins that leach from the column in a Man6-P independent manner) should be present at equal or greater levels in the mock compared to specific eluate. Given that the statistical power of spectral counting as a measure of protein abundance increases in proportion to the number of spectra counted, our approach was to compare the sum of spectra assigned to each protein from either the specific and mock eluates from all of the tissue samples combined together. The advantage of this approach is that it allows for confident conclusions to be drawn with respect to proteins that are present at low levels but in numerous samples. In these cases, the corresponding counts from individual samples would be insufficient to allow for useful conclusions.

We analyzed the same proportion of the total specific and mock eluates rather than equivalent amounts of protein (see Experimental Procedures). Thus, spectral counts measured in the two eluates are essentially normalized to unit weight of starting material. In terms of estimating enrichment in the specific Man6-P eluate, this represents a conservative approach as the total number of spectral counts is not directly proportional to the amount of protein analyzed due to sampling limitations during LC-MS. For example, with fewer peptides available for MS/MS analysis, each peptide may be measured more frequently when smaller amounts of protein are analyzed. In addition, when larger amounts of protein are analyzed, ion suppression by more abundant peptides may decrease the signal intensity and thus frequency of measurement of less abundant peptides. The relationship between amount of protein digest analyzed and number of spectra measured was determined experimentally and is shown in Figure 2A, where it is clear that the number of spectral counts plateaus with increasing amount of material analyzed. Thus, spectral counts measured in the mock eluate may be overestimated and this is shown to be the case in Figure 2B. Here, for each tissue, we plot the ratio of spectral counts for the specific versus mock eluates against the ratio of protein analyzed in the equivalent specific versus mock eluates. If spectral counts were directly proportional to amount of protein analyzed, then these two ratios would be expected to be the same, but this is not the case. Instead, the abundance of proteins in each of the mock eluates is overestimated, and thus, the stated enrichment factors are likely to be underestimates.

Most of the known soluble lysosomal proteins (59/60) were enriched in the Man6-P eluate (Supporting Information Table 3; Figure 3A), with GM2 activator protein being the only one that was depleted in the specific eluate. This may indicate that GM2 activator protein represents a low-affinity ligand for the immobilized MPR and that it readily dissociates during washing. Alternatively, some or all of the purified GM2 activator may be purified by virtue of association with other lysosomal proteins rather than by the presence of Man6-P. Interestingly, GM2 activator has been reported to traffic to the lysosome by both Man6-P-dependent and Man6-P-independent pathways,30,31 suggesting that this may be the case.

With the rationale that the specificity of purification for novel lysosomal candidates should be similar to known lysosomal proteins, we can use the enrichment observed for the latter to help in the identification of potential lysosomal candidates. We set the threshold for the lower 95% confidence interval of the specific/mock elution ratio to be >2.75 (log2 ≥ 1.5) (Figure 3A and Supporting Information Table 3).

We can use relative enrichment to categorize proteins currently not classified as lysosomal (Table 3; Figure 3B). For instance, 52 were found that were enriched to the same degree as the known lysosomal proteins (i.e., with an enrichment of >2.75-fold based on the lower limit of the 95% confidence interval) and we have categorized these as primary candidates for lysosomal residence (Table 4). We have also considered those proteins that are significantly enriched in the specific eluate but which are not enriched to the same degree as the known lysosomal proteins (lower limit of the 95% confidence interval for specific/mock > 1 but ≤2.75). These are categorized as secondary candidates. Proteins that are significantly depleted in the specific eluate (upper limit of the 95% confidence interval for spectral counts of specific/mock <1) are classified as not lysosomal. While this classification is arbitrary, we believe that it represents a useful approach to prioritizing candidates for further investigation.

**Relative Tissue Expression of Purified Proteins.** Expression profiling of soluble lysosomal proteins in human tissues demonstrated that some lysosomal proteins are quite widely...
### Table 2. Lysosomal Proteins in Rat Tissues

<table>
<thead>
<tr>
<th>ENSG Identifier</th>
<th>ENSEMBL description</th>
<th>min log(e)</th>
<th>coverage</th>
<th>spectral counts (mock elute)</th>
<th>spectral counts (specific elute)</th>
<th>log_2 (specific/mock)</th>
<th>function/class</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSMUSG0000001348</td>
<td>Acid phosphatase 5, tartrate resistant</td>
<td>-44.6</td>
<td>17</td>
<td>1</td>
<td>72</td>
<td>6.17 (3.65 to 10.47)</td>
<td>e</td>
</tr>
<tr>
<td>ENSMUSG0000005043</td>
<td>N-sulfogalactosamine sulfohydrolase</td>
<td>-96.6</td>
<td>26</td>
<td>18</td>
<td>104</td>
<td>2.53 (1.82 to 3.25)</td>
<td>e</td>
</tr>
<tr>
<td>ENSMUSG0000016256</td>
<td>Cathepsin Z</td>
<td>-127</td>
<td>44</td>
<td>170</td>
<td>612</td>
<td>1.85 (1.6 to 2.09)</td>
<td>e</td>
</tr>
<tr>
<td>ENSMUSG0000025579</td>
<td>Glucosidase, alpha, acid</td>
<td>-223.1</td>
<td>21</td>
<td>14</td>
<td>263</td>
<td>4.23 (3.47 to 5)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOM0000000435</td>
<td>Lysosomal thioesterase PPT2</td>
<td>-92.5</td>
<td>40</td>
<td>14</td>
<td>163</td>
<td>3.54 (2.76 to 4.32)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOM000000571</td>
<td>Proaposin</td>
<td>-326.2</td>
<td>54</td>
<td>164</td>
<td>646</td>
<td>1.98 (1.73 to 2.22)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOM000001385</td>
<td>LAMA-like protein 2</td>
<td>-400.1</td>
<td>55</td>
<td>239</td>
<td>1843</td>
<td>2.95 (2.75 to 3.14)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOM000001465</td>
<td>Idurionate 2-sulfatase</td>
<td>-131.6</td>
<td>31</td>
<td>16</td>
<td>142</td>
<td>3.15 (2.41 to 3.89)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOM000002188</td>
<td>Heparanase</td>
<td>-106.5</td>
<td>35</td>
<td>8</td>
<td>94</td>
<td>3.55 (2.53 to 4.57)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000004199</td>
<td>N-acethylglucosamine hydrolase</td>
<td>-146.6</td>
<td>39</td>
<td>12</td>
<td>146</td>
<td>3.6 (2.77 to 4.44)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000003291</td>
<td>Cellular repressor of E1A-stimulated genes 1</td>
<td>-143.6</td>
<td>52</td>
<td>22</td>
<td>346</td>
<td>3.98 (3.36 to 4.59)</td>
<td>u</td>
</tr>
<tr>
<td>ENSRNOG000003759</td>
<td>Galactosylceramidase</td>
<td>-79.8</td>
<td>19</td>
<td>5</td>
<td>69</td>
<td>3.79 (2.52 to 5.05)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000004199</td>
<td>Glucosamine (N-acetyl)-6-sulfatase</td>
<td>-522</td>
<td>61</td>
<td>389</td>
<td>1867</td>
<td>2.26 (2.11 to 2.42)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000005526</td>
<td>Mannosidase 2, alpha B2</td>
<td>-353.2</td>
<td>35</td>
<td>29</td>
<td>555</td>
<td>4.26 (3.72 to 4.79)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000005931</td>
<td>Plasma glutamate carboxypeptidase</td>
<td>-313.3</td>
<td>60</td>
<td>282</td>
<td>890</td>
<td>1.66 (1.47 to 1.85)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000007089</td>
<td>Legumain</td>
<td>-358</td>
<td>60</td>
<td>204</td>
<td>946</td>
<td>2.21 (2.2 to 2.43)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000007351</td>
<td>Gamma-glutamyl hydrolase</td>
<td>-151.6</td>
<td>39</td>
<td>74</td>
<td>764</td>
<td>3.37 (3.02 to 3.71)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000008064</td>
<td>Alpha-N-acetylgalactosaminidase</td>
<td>-463.8</td>
<td>68</td>
<td>295</td>
<td>1798</td>
<td>2.61 (2.43 to 2.79)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000008310</td>
<td>Myeloperoxidase</td>
<td>-467.7</td>
<td>57</td>
<td>74</td>
<td>726</td>
<td>3.29 (2.95 to 3.64)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000009325</td>
<td>Tissue alpha-1-fucosidase</td>
<td>-187.3</td>
<td>41</td>
<td>99</td>
<td>459</td>
<td>2.21 (1.9 to 2.33)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000010034</td>
<td>N-acetylglucosamine amidohydrolase 1</td>
<td>-318.8</td>
<td>61</td>
<td>112</td>
<td>877</td>
<td>2.97 (2.69 to 3.25)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000010080/</td>
<td>CLN5 protein</td>
<td>-93.7</td>
<td>28</td>
<td>4</td>
<td>81</td>
<td>4.34 (2.95 to 5.73)</td>
<td>u</td>
</tr>
<tr>
<td>ENSRNOG000009759</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSRNOG000010196</td>
<td>Galactosidase, beta 1</td>
<td>-471</td>
<td>58</td>
<td>384</td>
<td>1552</td>
<td>2.01 (1.85 to 2.18)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000010252</td>
<td>Beta-hexosaminidase alpha chain</td>
<td>-533.3</td>
<td>62</td>
<td>282</td>
<td>1887</td>
<td>2.74 (2.56 to 2.92)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000010331</td>
<td>Cathepsin B</td>
<td>-515.2</td>
<td>82</td>
<td>298</td>
<td>1550</td>
<td>2.36 (2.2 to 2.56)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000010630</td>
<td>Prolylcarboxypeptidase</td>
<td>-304.4</td>
<td>51</td>
<td>60</td>
<td>556</td>
<td>3.21 (2.83 to 3.6)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000011150</td>
<td>Arylsulfatase B</td>
<td>-224.9</td>
<td>55</td>
<td>68</td>
<td>362</td>
<td>2.41 (2.04 to 2.79)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000011153</td>
<td>Galactosidase, alpha</td>
<td>-309.3</td>
<td>54</td>
<td>138</td>
<td>1040</td>
<td>2.91 (2.66 to 3.17)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000011164</td>
<td>GM2 ganglioside activator protein</td>
<td>-120.4</td>
<td>48</td>
<td>141</td>
<td>86</td>
<td>-0.71 (-1.1 to -0.33)</td>
<td>ap</td>
</tr>
<tr>
<td>ENSRNOG000012062</td>
<td>Epidermidal secretory protein E1</td>
<td>-191.2</td>
<td>63</td>
<td>108</td>
<td>571</td>
<td>2.4 (2.11 to 2.7)</td>
<td>ap</td>
</tr>
<tr>
<td>ENSRNOG000012616</td>
<td>Palmitoyl-protein thioesterase 1</td>
<td>-273.8</td>
<td>64</td>
<td>126</td>
<td>1056</td>
<td>3.07 (2.8 to 3.33)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000012640</td>
<td>Dipetidyl-peptidase 2</td>
<td>-461.6</td>
<td>56</td>
<td>260</td>
<td>1645</td>
<td>2.66 (2.47 to 2.85)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000012953</td>
<td>Arylsulfatase A</td>
<td>-248.3</td>
<td>59</td>
<td>86</td>
<td>544</td>
<td>2.66 (2.33 to 2.99)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000013190</td>
<td>Ribonuclease T2</td>
<td>-204.3</td>
<td>53</td>
<td>74</td>
<td>397</td>
<td>2.42 (2.07 to 2.78)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000013476</td>
<td>Beta-mannosidase</td>
<td>-489.8</td>
<td>49</td>
<td>226</td>
<td>1673</td>
<td>2.89 (2.69 to 3.09)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000014064</td>
<td>Cathepsin H</td>
<td>-316.8</td>
<td>62</td>
<td>292</td>
<td>1446</td>
<td>2.31 (2.13 to 2.49)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000014461</td>
<td>Galactosamine (N-acetyl)-6-sulfate sulfatase</td>
<td>-351.7</td>
<td>61</td>
<td>191</td>
<td>1146</td>
<td>2.58 (2.36 to 2.81)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000015573</td>
<td>Chitobiase, di-N-acetyl-</td>
<td>-238.8</td>
<td>51</td>
<td>107</td>
<td>716</td>
<td>2.74 (2.45 to 3.03)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000015857</td>
<td>Cathepsin A</td>
<td>-556.2</td>
<td>63</td>
<td>544</td>
<td>3568</td>
<td>2.71 (2.58 to 2.84)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000016496</td>
<td>Cathepsin C</td>
<td>-782.8</td>
<td>78</td>
<td>427</td>
<td>2988</td>
<td>2.83 (2.65 to 2.94)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000017977</td>
<td>Sphingomyelin phosphodiesterase 1</td>
<td>-176.3</td>
<td>37</td>
<td>20</td>
<td>205</td>
<td>3.36 (2.7 to 4.01)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRNOG000018566</td>
<td>Cathepsin L</td>
<td>-559.1</td>
<td>79</td>
<td>622</td>
<td>4129</td>
<td>2.73 (2.61 to 2.85)</td>
<td>e</td>
</tr>
</tbody>
</table>
distributed, whereas expression of others was more limited (Figure 1A; Supporting Information Table 1). Here, the number of tissues in which each individual protein was expressed was simply determined on the basis of presence or absence in the respective Man6-P eluates as determined by LC-MS/MS (Figure 4A). Thirty of the 60 known lysosomal proteins were found to be ubiquitously distributed and were present in all 17 tissue samples examined. An additional 26 proteins were present in most (12–16) of the sample types. Thus, the vast majority of Man6-P eluates obtained on the basis of protein expression were devoid of unknown proteins. Only a small number of additional proteins were present in a few tissues (1–4). No proteins were detected in the mock eluates (Figure 4B). Twenty-six proteins were present in all 17 tissue samples. In contrast, the majority of proteins were present in only a few tissues (1–4). No proteins were detected in the mock eluates. (A) Increasing amounts of a tryptic digest of a rat liver preparation were analyzed in duplicate by LC-MS, and the total number of spectral counts were determined. (B) For each tissue sample, the ratio of spectral counts in the specific eluate to the mock eluate was plotted against the ratio of protein analyzed in the two samples.

### Table 2. Continued

<table>
<thead>
<tr>
<th>ENSG identifier</th>
<th>ENSEMBL description</th>
<th>min log(e)</th>
<th>coverage</th>
<th>spectral counts (mock eluate)</th>
<th>spectral counts (specific eluate)</th>
<th>log2 (specific/mock)</th>
<th>function/class</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSRN0G00000018989</td>
<td>Ependymin related protein 1 (zebrafish)</td>
<td>-181.4</td>
<td>55</td>
<td>255</td>
<td>1490</td>
<td>2.55 (2.36 to 2.74)</td>
<td>u</td>
</tr>
<tr>
<td>ENSRN0G00000019077</td>
<td>Lysozyme A</td>
<td>-136.9</td>
<td>47</td>
<td>44</td>
<td>372</td>
<td>3.08 (2.63 to 3.53)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G0000001912</td>
<td>Lysozyme A</td>
<td>-382.3</td>
<td>55</td>
<td>218</td>
<td>979</td>
<td>2.17 (1.96 to 2.38)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000019837</td>
<td>Interferon gamma inducible protein 30</td>
<td>-150.7</td>
<td>53</td>
<td>26</td>
<td>203</td>
<td>2.96 (2.38 to 3.55)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G000000199708</td>
<td>Cathepsin F</td>
<td>-278</td>
<td>60</td>
<td>75</td>
<td>514</td>
<td>2.78 (2.43 to 3.13)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000019859</td>
<td>Lysozyme A</td>
<td>-155.3</td>
<td>31</td>
<td>16</td>
<td>221</td>
<td>3.79 (3.06 to 4.51)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000020006</td>
<td>Cathepsin D</td>
<td>-551.7</td>
<td>72</td>
<td>867</td>
<td>3822</td>
<td>2.14 (2.03 to 2.25)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G000000211155</td>
<td>Cathepsin K</td>
<td>-178.4</td>
<td>38</td>
<td>12</td>
<td>100</td>
<td>3.06 (2.21 to 3.91)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G000000211157</td>
<td>Cathepsin S</td>
<td>-140.5</td>
<td>37</td>
<td>56</td>
<td>257</td>
<td>2.2 (1.78 to 2.61)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000023830</td>
<td>Deoxyribonuclease-2-alpha</td>
<td>-186.5</td>
<td>58</td>
<td>26</td>
<td>312</td>
<td>3.58 (3.01 to 4.16)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000023910</td>
<td>Mannosidase 2, alpha B1</td>
<td>-807.4</td>
<td>61</td>
<td>220</td>
<td>1575</td>
<td>2.84 (2.64 to 3.04)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000025274</td>
<td>Beta-hexosaminidase beta chain</td>
<td>-364.1</td>
<td>52</td>
<td>246</td>
<td>1178</td>
<td>2.26 (2.06 to 2.46)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000031266</td>
<td>Sialic acid acetylerase</td>
<td>-231.6</td>
<td>52</td>
<td>58</td>
<td>156</td>
<td>3.14 (2.75 to 3.54)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000032381</td>
<td>Alpha-N-acetylgalactosaminidase</td>
<td>-370.8</td>
<td>39</td>
<td>46</td>
<td>807</td>
<td>4.13 (3.71 to 4.56)</td>
<td>e</td>
</tr>
<tr>
<td>ENSRN0G00000032942</td>
<td>Neuraminidase 1</td>
<td>-222.4</td>
<td>45</td>
<td>157</td>
<td>586</td>
<td>1.9 (1.65 to 2.15)</td>
<td>e</td>
</tr>
</tbody>
</table>

For each protein, fold enrichment (the number of spectral counts in the specific eluate divided by the number in the mock eluate) is shown together with the upper and lower 95% confidence indices. “Function” is a tentative assignment based upon known or predicted biological properties. Abbreviations are sc, spectral counts; e, enzyme; ap, accessory protein; u, unknown. “Min log(e)” represents the highest confidence expectation score assigned by X! Tandem and “coverage” indicates the total percentage of sequence within the assigned peptides to each protein.
Figure 3. Enrichment factors in specific versus mock eluates obtained for proteins identified by affinity chromatography on immobilized MPRs. The log$_2$ of the ratio of spectral counts in the specific and mock eluates (SC$_{MPR}$ and SC$_{MOCK}$, respectively) is plotted with bars representing the upper and lower 95% confidence indices for known lysosomal proteins (A) or proteins not currently classified as lysosomal (B). The lower confidence index for all but one of the known lysosomal proteins is greater than 2.75 (log$_2$ = 1.5) and this threshold is plotted as a dotted line. Plots represent: green error bars, proteins that achieve this threshold; blue error bars, proteins that do not achieve this threshold but which are significantly enriched in the specific eluate (lower 95% confidence interval is greater than 1 (log$_2$ = 0)); gray error bars, proteins that cannot be classified (95% confidence interval includes SC$_{MPR}$/SC$_{MOCK}$ = 1); and red error bars, proteins that are significantly depleted in the specific eluate (upper 95% confidence interval is less than 1 (log$_2$ = 0)). For graphical representation, fold-enrichments that are greater than 16 or less that 1/16 are arbitrarily assigned to be 16 (a log$_2$ value of 4) or 1/16 (a log$_2$ value of −4), respectively. Proteins that not are of rodent origin are not shown.

Table 3. Classification of Affinity Purified Proteins Based upon Relative Abundance in Specific and Mock Eluates$

<table>
<thead>
<tr>
<th>category</th>
<th>number of proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>lysosomal</td>
<td>60</td>
</tr>
<tr>
<td>not lysosomal</td>
<td>272</td>
</tr>
<tr>
<td>primary candidate</td>
<td>52</td>
</tr>
<tr>
<td>secondary candidate</td>
<td>84</td>
</tr>
<tr>
<td>unclassified</td>
<td>304</td>
</tr>
<tr>
<td>Total</td>
<td>772</td>
</tr>
</tbody>
</table>

$^a$ Categories are lysosomal, known lysosomal proteins; primary lysosomal candidates, proteins that are enriched in the Man6-P eluate to the same degree as known lysosomal proteins (i.e. log$_2$ of lower 95% confidence index >1.5); secondary lysosomal candidates, proteins that are significantly enriched in the specific eluate (i.e. log$_2$ of lower 95% confidence index >0 but ≤1.5); not lysosomal, proteins that are significantly depleted in the specific eluate (i.e. log$_2$ of upper 95% confidence index <0); and unclassified, proteins for which statistically meaningful conclusions cannot be drawn. Non-rodent contaminants (21 proteins) are excluded from this analysis.

Concluding Remarks

It is becoming increasingly apparent from recent studies that the soluble proteome of the lysosome is more expansive than previously imagined. While over 60 Man6-P containing proteins are established as residing within the lumen of this organelle, analyses of proteins isolated by MPR-affinity chromatography from a variety of mammalian sources have revealed a significant number of additional proteins that may have lysosomal function. In this study, we have surveyed the proteome of MPR-affinity purified proteins from a broad selection of rat tissues. We have used mass spectrometric and biostatistical methods to distinguish specifically purified proteins from nonspecific contaminants by filtering the extensive list of identified proteins with parameters based upon the relative abundance of known lysosomal proteins in specific versus mock affinity column eluates. In concept, this approach is not dissimilar to the I-DIERT procedure for identifying specific members of a protein complex that are isolated by the affinity tagging of one of its constituents, with the main difference being that we have relied upon spectral counting for protein abundance measurement rather than isotopic labeling.

When data obtained from all 17 tissues are considered together, we found that no significant conclusions could be drawn for 304/772 of the identified rodent proteins. In most cases, this could be attributed to low spectral counts for both the specific and mock eluates resulting in an extremely wide 95% confidence interval for the ratio. However, about a third (272/772) of all of the identified proteins could be confidently excluded from further analysis because they were significantly depleted in the specific compared to the mock eluate. One hundred and ninety-six proteins were significantly enriched in the Man6-P compared to mock eluate. Of these, 60 are known soluble lysosomal proteins and the rest are proteins that are not currently thought to have lysosomal function. Of the latter, 52 proteins were enriched to levels comparable to the known lysosomal proteins (Table 4). Some of these proteins (21/52) were also identified in previous proteomic studies of purified Man6-P glycoproteins.

The enriched proteins that are not assigned to the lysosome fall into numerous functional categories. Many are known or predicted to be hydrolases or other enzymes, and as such, they represent promising lysosomal candidates, especially those that resemble known lysosomal proteins and which have a widespread tissue distribution. Several proteins fall into this category. Acid sphingomyelinase-like 3A (SMPDL3A) has been identified in many studies of purified Man6-P glycoproteins and is a paralog of the lysosomal hydrolases, acid sphingomyelinase. Increased expression of SMPDL3A has been observed in bladder cancer and a role in tumorigenesis has been proposed. Retinoid inducible serine carboxypeptidase (RISC)
<table>
<thead>
<tr>
<th>ENSRNOG00000011287</th>
<th>Multiple inositol polyphosphate phosphatase 1</th>
<th>−52</th>
<th>15</th>
<th>24</th>
<th>0</th>
<th>∞ (2.64 to ∞)</th>
<th>no</th>
<th>9</th>
<th>e</th>
<th>2</th>
<th>yes</th>
<th>secreted</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSRNOG00000015941</td>
<td>Similar to 65 kDa FK506-binding protein</td>
<td>−141</td>
<td>30</td>
<td>85</td>
<td>18</td>
<td>2.24 (1.51 to 2.97)</td>
<td>no</td>
<td>12</td>
<td>e</td>
<td>7</td>
<td>yes</td>
<td>secreted</td>
<td>ER</td>
</tr>
<tr>
<td>ENSRNOG0000001935</td>
<td>Carboxypeptidase B2 Amiloride binding protein 1</td>
<td>−86</td>
<td>33</td>
<td>44</td>
<td>3</td>
<td>3.87 (2.27 to 5.48)</td>
<td>no</td>
<td>8</td>
<td>e</td>
<td>6</td>
<td>yes</td>
<td>secreted</td>
<td>extracellular</td>
</tr>
<tr>
<td>ENSRNOG0000001913</td>
<td>Ceruloplasmin, Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1</td>
<td>−330.8</td>
<td>35</td>
<td>145</td>
<td>30</td>
<td>2.27 (1.71 to 2.84)</td>
<td>yes</td>
<td>9</td>
<td>e</td>
<td>6</td>
<td>yes</td>
<td>secreted</td>
<td>extracellular</td>
</tr>
<tr>
<td>ENSRNOG0000001823</td>
<td>Sialyltransferase 1 Acid sphingomyelinase-like phosphodiesterase 3a</td>
<td>−120.8</td>
<td>33</td>
<td>62</td>
<td>7</td>
<td>3.15 (2.05 to 4.25)</td>
<td>no</td>
<td>3</td>
<td>e</td>
<td>3</td>
<td>yes</td>
<td>secreted</td>
<td>Golgi</td>
</tr>
<tr>
<td>ENSRNOG00000019014</td>
<td>N- heparan sulfate sulfotransferase 1 Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1</td>
<td>−38.5</td>
<td>6.1</td>
<td>14</td>
<td>0</td>
<td>∞ (1.87 to ∞)</td>
<td>no</td>
<td>5</td>
<td>e</td>
<td>4</td>
<td>yes</td>
<td>secreted</td>
<td>type II membrane</td>
</tr>
<tr>
<td>ENSRNOG0000007763</td>
<td>Cathepsin O Acid sphingomyelinase-like phosphodiesterase 3a</td>
<td>−21</td>
<td>11</td>
<td>16</td>
<td>0</td>
<td>∞ (2.06 to ∞)</td>
<td>yes</td>
<td>9</td>
<td>e</td>
<td>4</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000002358</td>
<td>Retinoid-inducible serine carboxypeptidase Lactoperoxidase</td>
<td>−211.6</td>
<td>41</td>
<td>673</td>
<td>80</td>
<td>3.07 (2.74 to 3.41)</td>
<td>yes</td>
<td>16</td>
<td>e</td>
<td>5</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000008422</td>
<td>LAMA-like protein 1 Cathepsin O</td>
<td>−79.9</td>
<td>19</td>
<td>23</td>
<td>0</td>
<td>∞ (2.58 to ∞)</td>
<td>no</td>
<td>3</td>
<td>e</td>
<td>6</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000008933</td>
<td>Carboxypeptidase A3 Carboxypeptidase A3, alpha-1, 2, plasma</td>
<td>−55.2</td>
<td>16</td>
<td>27</td>
<td>2</td>
<td>3.75 (1.83 to 5.68)</td>
<td>no</td>
<td>6</td>
<td>e</td>
<td>4</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000011181</td>
<td>Fucosidase, alpha-L, Pancreatic lipase-related protein 1</td>
<td>−221.4</td>
<td>50</td>
<td>322</td>
<td>52</td>
<td>2.63 (2.21 to 3.05)</td>
<td>yes</td>
<td>15</td>
<td>e</td>
<td>5</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000015551</td>
<td>Pancreatic lipase-related protein 1</td>
<td>−214.1</td>
<td>43</td>
<td>117</td>
<td>27</td>
<td>2.12 (1.52 to 2.71)</td>
<td>no</td>
<td>3</td>
<td>e</td>
<td>3</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000017908</td>
<td>Chymosin family Tryptase alpha/beta 1</td>
<td>−253.7</td>
<td>66</td>
<td>323</td>
<td>28</td>
<td>3.53 (2.97 to 4.08)</td>
<td>no</td>
<td>12</td>
<td>e</td>
<td>1</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000024181</td>
<td>Tryptase alpha/beta 1 Arylsulfatase K</td>
<td>−79.9</td>
<td>33</td>
<td>54</td>
<td>2</td>
<td>4.75 (2.86 to 6.65)</td>
<td>no</td>
<td>12</td>
<td>e</td>
<td>2</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000026937</td>
<td>Arylsulfatase K Carboxypeptidase A2</td>
<td>−144.5</td>
<td>27</td>
<td>57</td>
<td>1</td>
<td>5.83 (3.31 to 10.14)</td>
<td>yes</td>
<td>8</td>
<td>e</td>
<td>7</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000028092</td>
<td>Carboxypeptidase A2 Mast cell protease 8</td>
<td>−100.7</td>
<td>34</td>
<td>37</td>
<td>3</td>
<td>3.62 (2.01 to 5.24)</td>
<td>no</td>
<td>2</td>
<td>e</td>
<td>0</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000029604</td>
<td>Mast cell protease 8 Mast cell protease 9</td>
<td>−235.7</td>
<td>49</td>
<td>241</td>
<td>38</td>
<td>2.66 (2.17 to 3.16)</td>
<td>no</td>
<td>9</td>
<td>e</td>
<td>4</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000030899</td>
<td>Mast cell protease 9 Granzyme-like protein 2</td>
<td>−445.3</td>
<td>76</td>
<td>1121</td>
<td>87</td>
<td>3.69 (3.37 to 4)</td>
<td>no</td>
<td>11</td>
<td>e</td>
<td>2</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000032717</td>
<td>Granzyme-like protein 2</td>
<td>−398.4</td>
<td>56</td>
<td>439</td>
<td>65</td>
<td>2.76 (2.38 to 3.13)</td>
<td>no</td>
<td>10</td>
<td>e</td>
<td>3</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSRNOG0000039865</td>
<td>Vanin 3</td>
<td>−167</td>
<td>39</td>
<td>121</td>
<td>5</td>
<td>4.6 (3.35 to 5.85)</td>
<td>no</td>
<td>8</td>
<td>e</td>
<td>3</td>
<td>yes</td>
<td>secreted</td>
<td>secreted</td>
</tr>
<tr>
<td>ENSG identifier</td>
<td>ENSEMBL description</td>
<td>min log(e)</td>
<td>spectral counts (specific eluate)</td>
<td>spectral counts (mock eluate)</td>
<td>log, (specific/mock)</td>
<td>previously identified number of tissues found</td>
<td>function</td>
<td>number of potential N-linked glycosylation sites</td>
<td>signal domain</td>
<td>protein class</td>
<td>subcellular location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------</td>
<td>------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td>-----------------------------------------------</td>
<td>----------</td>
<td>-----------------------------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000039971</td>
<td>Cytotoxic T lymphocyte-associated protein 2 beta-like</td>
<td>−55.5</td>
<td>57</td>
<td>38</td>
<td>0</td>
<td>∞ (3.31 to ∞)</td>
<td>no</td>
<td>6</td>
<td>e</td>
<td>0</td>
<td>no</td>
<td>soluble, nonsecreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000009086</td>
<td>Serum amyloid A-component</td>
<td>−144.6</td>
<td>58</td>
<td>320</td>
<td>78</td>
<td>2.04 (1.68 to 2.39)</td>
<td>yes</td>
<td>15</td>
<td>l</td>
<td>2</td>
<td>yes</td>
<td>secreted extracellular</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG0000009217</td>
<td>F-box only protein 6</td>
<td>−71.9</td>
<td>91</td>
<td>6</td>
<td>4</td>
<td>3.95 (2.55 to 5.36)</td>
<td>yes</td>
<td>11</td>
<td>l</td>
<td>1</td>
<td>no</td>
<td>soluble, nonsecreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000032834</td>
<td>Stress 70 protein chaperone microsome-associated 6</td>
<td>−138.4</td>
<td>34</td>
<td>67</td>
<td>3</td>
<td>4.48 (2.89 to 6.07)</td>
<td>yes</td>
<td>14</td>
<td>o</td>
<td>5</td>
<td>yes</td>
<td>secreted ER</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000020129</td>
<td>P-cadherin</td>
<td>−41.4</td>
<td>9.1</td>
<td>11</td>
<td>0</td>
<td>∞ (1.52 to ∞)</td>
<td>no</td>
<td>1</td>
<td>o</td>
<td>4</td>
<td>yes</td>
<td>type I membrane</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000008383</td>
<td>Gastrokine 1</td>
<td>−66.8</td>
<td>34</td>
<td>91</td>
<td>15</td>
<td>2.6 (1.82 to 3.38)</td>
<td>no</td>
<td>5</td>
<td>o</td>
<td>1</td>
<td>yes</td>
<td>secreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000008716</td>
<td>Neurofilament heavy polypeptide</td>
<td>−138.5</td>
<td>14</td>
<td>28</td>
<td>2</td>
<td>3.81 (1.88 to 5.73)</td>
<td>no</td>
<td>4</td>
<td>o</td>
<td>1</td>
<td>no</td>
<td>soluble, nonsecreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000018400</td>
<td>Golgi membrane protein 1</td>
<td>−26.7</td>
<td>7.4</td>
<td>11</td>
<td>0</td>
<td>∞ (1.52 to ∞)</td>
<td>no</td>
<td>4</td>
<td>o</td>
<td>4</td>
<td>yes</td>
<td>type II membrane</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG0000013572</td>
<td>Latexin</td>
<td>−42.2</td>
<td>22</td>
<td>21</td>
<td>0</td>
<td>∞ (2.45 to ∞)</td>
<td>no</td>
<td>4</td>
<td>pi</td>
<td>3</td>
<td>no</td>
<td>soluble, nonsecreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000030387</td>
<td>Kininogen 1</td>
<td>−253.1</td>
<td>56</td>
<td>302</td>
<td>82</td>
<td>1.88 (1.53 to 2.23)</td>
<td>yes</td>
<td>13</td>
<td>pi</td>
<td>5</td>
<td>yes</td>
<td>secreted extracellular</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000005599</td>
<td>Alpha-1-inhibitor 3</td>
<td>−606.3</td>
<td>13</td>
<td>99</td>
<td>13</td>
<td>2.93 (2.11 to 3.75)</td>
<td>yes</td>
<td>11</td>
<td>pi</td>
<td>12</td>
<td>yes</td>
<td>secreted extracellular</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000037188</td>
<td>Murinoglobulin-1</td>
<td>−365.5</td>
<td>44</td>
<td>497</td>
<td>75</td>
<td>2.73 (2.38 to 3.08)</td>
<td>yes</td>
<td>15</td>
<td>pi</td>
<td>1</td>
<td>yes</td>
<td>secreted lysosomes</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000010527</td>
<td>Cystatin-B</td>
<td>−96.7</td>
<td>72</td>
<td>329</td>
<td>65</td>
<td>2.34 (1.96 to 2.72)</td>
<td>yes</td>
<td>17</td>
<td>pi</td>
<td>0</td>
<td>no</td>
<td>soluble, nonsecreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000009855</td>
<td>Serine (or cysteine) peptidase inhibitor, clade A,</td>
<td>−38</td>
<td>21</td>
<td>13</td>
<td>0</td>
<td>∞ (1.76 to ∞)</td>
<td>yes</td>
<td>1</td>
<td>pi</td>
<td>1</td>
<td>yes</td>
<td>secreted cytoplasm</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000020455</td>
<td>Cystatin E/M</td>
<td>−62.4</td>
<td>51</td>
<td>44</td>
<td>1</td>
<td>5.46 (2.93 to 9.78)</td>
<td>no</td>
<td>9</td>
<td>pi</td>
<td>0</td>
<td>yes</td>
<td>secreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000033245</td>
<td>Murinoglobulin family</td>
<td>−340.7</td>
<td>3</td>
<td>12</td>
<td>0</td>
<td>∞ (1.64 to ∞)</td>
<td>no</td>
<td>5</td>
<td>pi</td>
<td>14</td>
<td>yes</td>
<td>secreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000010527</td>
<td>Serine (or cysteine) peptidase inhibitor, clade A,</td>
<td>−227.7</td>
<td>57</td>
<td>209</td>
<td>41</td>
<td>2.35 (1.87 to 2.83)</td>
<td>no</td>
<td>10</td>
<td>pi</td>
<td>3</td>
<td>yes</td>
<td>type I membrane</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG0000004067</td>
<td>Neuronal cell adhesion molecule</td>
<td>−51.7</td>
<td>6.7</td>
<td>11</td>
<td>0</td>
<td>∞ (1.52 to ∞)</td>
<td>yes</td>
<td>1</td>
<td>s</td>
<td>23</td>
<td>yes</td>
<td>type I membrane</td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000024109</td>
<td>Neurexin I</td>
<td>−64.6</td>
<td>9.3</td>
<td>17</td>
<td>0</td>
<td>∞ (2.15 to ∞)</td>
<td>yes</td>
<td>2</td>
<td>s</td>
<td>2</td>
<td>no</td>
<td>membrane</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG00000004610</td>
<td>Lumican</td>
<td>−52.8</td>
<td>26</td>
<td>40</td>
<td>3</td>
<td>3.74 (2.13 to 5.35)</td>
<td>yes</td>
<td>11</td>
<td>s</td>
<td>4</td>
<td>yes</td>
<td>secreted</td>
<td></td>
</tr>
<tr>
<td>ENSRNOG0000001554</td>
<td>Decorin</td>
<td>−134</td>
<td>31</td>
<td>164</td>
<td>39</td>
<td>2.07 (1.57 to 2.57)</td>
<td>no</td>
<td>12</td>
<td>s</td>
<td>4</td>
<td>yes</td>
<td>secreted extracellular</td>
<td></td>
</tr>
</tbody>
</table>
A widely distributed protease that colocalizes with lysosome-associated membrane protein 2 and is probably a lysosomal protein.\(^{34,35}\) FLJ22662 is a paralog of LOC196463, a previously discovered protein that was recently demonstrated to be lysosomal.\(^{36}\) On the basis of sequence homology, both FLJ22662 and LOC196463 may have phosphodiesterase activity. However, for many of the enriched proteins, it is not easy to predict whether a lysosomal function is likely, but frequently, we find more than one representative of a particular class of proteins. For example, while glycosyltransferase would not appear to be a classical lysosomal activity, we find enzymes of this class among the enriched proteins.

### Table 4. Continued

<table>
<thead>
<tr>
<th>ENSG identifier</th>
<th>ENSEMBL description</th>
<th>min log(e)</th>
<th>coverage</th>
<th>spectral counts (specific eluate)</th>
<th>spectral counts (mock eluate)</th>
<th>log(_2) (specific/mock)</th>
<th>previously identified</th>
<th>number of tissues found</th>
<th>function</th>
<th>number of potential N-linked glycosylation sites</th>
<th>signal domain</th>
<th>protein class</th>
<th>subcellular location</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSRNOG00000015410</td>
<td>Similar to asporin</td>
<td>−75.1</td>
<td>28</td>
<td>42</td>
<td>6</td>
<td>2.81 (1.61 to 4.01)</td>
<td>no</td>
<td>11</td>
<td>s</td>
<td>2</td>
<td>yes</td>
<td>secreted</td>
<td>extracellular</td>
</tr>
<tr>
<td>ENSRNOG0000002878</td>
<td>Afamin</td>
<td>−206.6</td>
<td>36</td>
<td>155</td>
<td>21</td>
<td>2.88 (2.23 to 3.54)</td>
<td>yes</td>
<td>12</td>
<td>tp</td>
<td>6</td>
<td>yes</td>
<td>secreted</td>
<td>extracellular</td>
</tr>
<tr>
<td>ENSRNOG00000021001</td>
<td>Gastric intrinsic factor</td>
<td>−32</td>
<td>11</td>
<td>17</td>
<td>0</td>
<td>∞ (2.15 to ∞)</td>
<td>yes</td>
<td>2</td>
<td>tp</td>
<td>5</td>
<td>yes</td>
<td>secreted</td>
<td>extracellular</td>
</tr>
<tr>
<td>ENSRNOG00000020148</td>
<td>Interleukin-4 induced gene-1</td>
<td>−135.2</td>
<td>25</td>
<td>68</td>
<td>13</td>
<td>2.39 (1.54 to 2.23)</td>
<td>no</td>
<td>5</td>
<td>o</td>
<td>3</td>
<td>yes</td>
<td>multipass</td>
<td>nucleus</td>
</tr>
<tr>
<td>ENSRNOG00000021258</td>
<td>Endogenous retroviral sequence 3</td>
<td>−174.9</td>
<td>64</td>
<td>226</td>
<td>56</td>
<td>2.01 (1.59 to 2.43)</td>
<td>no</td>
<td>4</td>
<td>o</td>
<td>3</td>
<td>yes</td>
<td>secreted</td>
<td>extracellular</td>
</tr>
<tr>
<td>ENSRNOG00000037782</td>
<td>Gastrokine family</td>
<td>−27</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>∞ (2.31 to ∞)</td>
<td>no</td>
<td>3</td>
<td>o</td>
<td>1</td>
<td>yes</td>
<td>type II</td>
<td>membrane</td>
</tr>
</tbody>
</table>

For each protein, fold enrichment (the number of spectral counts in the specific eluate divided by the number in the mock eluate) is shown together with the upper and lower 95% confidence indices. Only those proteins with a lower confidence limit that exceeds the threshold determined from known lysosomal proteins (Table 2) are shown. Function/class is a tentative assignment based upon known or predicted biological properties. Abbreviations: Sc, spectral counts. For "function": e, enzyme; tp, transport protein; u, unknown; pi, protease inhibitor; s, structural; o, other. Protein class and tentative subcellular location were assigned using the LOCATE subcellular localization database from the corresponding human and/or mouse proteins. "Min log(e)" represents the highest confidence expectation score assigned by X! Tandem (see Experimental Procedures) and "coverage" indicates the total percentage of sequence within the assigned peptides to each protein. "Previously identified" indicates whether a given protein was assigned in one or more of the earlier analyses\(^{11–14,16–20}\) of purified Man6-P glycoproteins.
this type (including GDP-fucose protein O-fucosyltransferase 2 (POFUT2), beta 3-glycosyltransferase-like and sialyltransferase 1) that are enriched in the Man6-P eluate. It is possible that they represent ER proteins of which some proportion may be aberrantly decorated with Man6-P, especially as we purify other proteins that may have ER localization (e.g., procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1 and 2, KDEL containing protein 2 and stress 70 protein chaperone microsome-associated). However, it is also possible that these proteins are representatives of a hitherto unsuspected class of lysosomal protein. As noted previously, we also identified a number of protease inhibitors that appear to contain Man6-P. In this study, we also find a significant number of small leucine-rich proteoglycans.

Enrichment in the specific eluate during affinity purification is consistent with a lysosomal function, but it is not indicative of such. For example, a protein that is enriched in the Man6-P eluate could represent a specific contaminant or a nonlysosomal Man6-P glycoprotein rather than a bona fide lysosomal resident. While some of the enriched proteins are unquestionably purified in association with true Man6-P glycoproteins (e.g., cystatins, that lack N-linked glycosylation sites), for many or most of the purified proteins there seems little biological basis to suspect such an interaction; thus, they most probably represent ER proteins of which some proportion may be aberrantly decorated with Man6-P, especially as we purify other proteins that may have ER localization (e.g., procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1 and 2, KDEL containing protein 2 and stress 70 protein chaperone microsome-associated). However, lysosomal subcellular fractions are highly complex samples, with significant contamination by other organelles and this poses technical hurdles for a global, data-independent mass spectrometric approach toward cellular localization. As an alternative, the database of lysosomal candidates identified in this study should provide an excellent resource for targeted MS studies that address a subpopulation of candidate proteins within the complex lysosomal fractions.

Acknowledgment. This work was supported by NIH grants DK054317 and S10RR017992 (P.L.). We thank Caifeng Zhao for her excellent assistance with the mass spectrometry.

Supporting Information Available: Raw mass spectrometry data files are available upon request. Supplementary data in the form of an Excel workbook is provided which details lysosomal protein tissue distribution in terms of transcript and spectral counts as well as Supporting Information for protein assignment and statistical analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

References
The Man6-P Glycoproteome


(24) Craig, R.; Cortens, J. P.; Beavis, R. C. Open source system for analyzing, validating, and storing protein identification data. J. Proteome Res. 2004, 3 (6), 1234–42.


PR800135V