



Published in final edited form as:

J Proteome Res. 2008 July ; 7(7): 3010–3021. doi:10.1021/pr800135v.

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, NJ 08854

Abstract

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 towards 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 semi-quantitative 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.

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 transport¹. 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

* Corresponding authors. sleat@cabm.rutgers.edu; lobel@cabm.rutgers.edu.

† Center for Advanced Biotechnology and Medicine

‡ Department of Pharmacology

¥ Department of Biostatistics

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 information is available free of charge via the Internet at <http://pubs.acs.org>.

function result in lysosomal storage disorders² 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 in³). 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^{4, 5}. 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 centrifugation⁶⁻⁸. 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 subcellular 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 increasing⁸.

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 compartment⁹. 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 reagent¹⁰. This approach has been used to investigate the lysosomal proteomes from a number of sources including cultured cells and tissues¹⁰⁻²⁰. 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 cells²¹. 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 respects¹⁰. Similarly, expression profiling of soluble lysosomal proteins in 45 human tissues based upon the detection of their respective transcripts (Fig. 1, Panel A; Online Supplementary Material 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 (Fig. 1, Panel B). 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

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 to 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 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 ug/ml pepstatin A, 1 ug/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 hrs. 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 to Affigel 10 at a density of 5 mg/ml¹⁰. Columns were then flow-washed with 30 ml PBS-I containing 0.2 % Tween 20, then batch washed 4 times with 10 ml PBS-I containing 0.2 % Tween 20 and then 4 times with 10 ml PBS-I without Tween 20. Columns were then flow-washed overnight with 80ml PBS-I and sequentially batch eluted with: 1) PBS containing 10 mM mannose and 10 mM glucose 6-phosphate; 2) PBS containing 10 mM Man6-P, and; 3) 0.1 M glycine, pH 2.5. To perform batch elution, beads were resuspended in 4 ml of each respective elution buffer, incubated for 10 min and allowed to drain by 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 Centricon YM10 centrifugal concentrator and the protein concentration was determined²²

Tandem mass spectrometry

For each tissue, a sample of the specific (Man6-P) or nonspecific (mannose and glucose 6-phosphate) affinity purification eluates was heated for 10 minutes 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 described¹⁶. Samples were analyzed by LC-MS/MS using an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as described previously¹⁷. For the Man6-P eluates, typically a portion of each digest corresponding to 1 µg of starting material was analyzed by LC-MS. This was not possible for all samples and in some cases less was analyzed and the amount of each sample used for each LC-MS analysis is shown in Table 1. For each mock eluate, we analyzed the same proportion (v/v) of the total purified sample that we analyzed for the corresponding specific eluate. For example, if an LC-MS run was performed on 10% (v/v) of a total purified specific eluate, then we also analyzed 10% of the mock eluate regardless of the protein concentration of this sample. Two LC-MS/MS runs were conducted for each elution condition for each tissue source.

Generation of peak lists

Peak lists were generated from raw data using the TurboSEQUENT module of BioworksBrowser 3.1 SR1 (Thermo Fisher Scientific). Parameters were: peptide molecular weight range, 500–5000 Da; threshold intensity of 1000; a precursor mass tolerance of 1.4 m/z; minimum ion count of 50 and automatic charge state determination.

Database searching

The ENSEMBL rat protein database (see below) is incomplete as five known lysosomal proteins are absent even though these proteins are encoded by the rat genome. To help include such proteins in our analysis, we searched our data against both the rat and mouse databases, converting assignments made with the mouse database to the corresponding rat gene identifier where this was available. In cases where no rat gene identifier corresponding to the mouse assignment was available, the mouse gene identifier is instead used. Databases (rat, ENSEMBL, Feb. 2006, version 48.34m, which contains 18311 known genes; mouse, ENSEMBL April 2007 build of the NCBI m37 assembly, database version 48.37a, which contains 21928 known genes) were searched using a local implementation of GPM-XE Manager version 2.1.0 (Beavis Informatics Ltd, Winnipeg, Canada) which uses X! Tandem version 2007.07.1, to assign spectral data^{23, 24}. LTQ data was searched using the MudPit option to produce a merged output file which allows for a consistent assignment of spectra to similar or identical gene products. Parameters for searching were a precursor ion mass error of +4 and -0.5 Da and a fragment mass error of 0.4 Da. Errors in assignment of monoisotopic mass were not permitted. Cysteine carbamidomethylation was specified as a complete modification, and methionine oxidation was a permitted variable modification during development of the preliminary model with one missed cleavage site allowed. Methionine oxidation and deamidation at asparagine and glutamine residues were allowed during model refinement of those preliminary assignments achieving an expectation value of < 0.001. The threshold for protein assignment was a log GPM expectancy score (log(e)) of -10 based on the aggregate score from the merged data with a minimum of two peptides assigned per protein. When performing data analysis, sample information (e.g. source, eluate for affinity purification and spectral count data) was extracted from the merged output file. Data supporting protein assignments are given in Online Supplementary Material Table 2. Tentative subcellular locations for identified rat proteins were assigned from the human or mouse equivalents using the LOCATE subcellular localization database^{25, 26}.

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^{27, 28}. Statistical analysis was essentially as described previously²⁰. In brief, the method of Wilson²⁹ 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 computation and graphics (<http://www.r-project.org/>). Data for statistical analyses are presented in Online Supplementary Material Table 3.

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 Methods). The highest relative yield was obtained from placenta (16 µg/g). In other tissues, as predicted from earlier blotting experiments¹⁰, 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 (Online Supplementary Material 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 total^{16, 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 method^{27, 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 Methods). 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 Fig. 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 over-estimated and this is shown to be the case in Fig. 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 (Online Supplementary Material Table 3; Fig. 3, Panel A), 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 ($\log_2 > 1.5$) (Fig. 3, Panel A and Online Supplementary Material Table 3).

We can use relative enrichment to categorize proteins currently not classified as lysosomal (Table 3; Fig. 3, Panel B). 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 distributed whereas expression of others was more limited (Fig. 1, Panel A; Online Supplementary Material 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 (Fig. 4, Panel A). 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 to 16) of the sample types. Thus, the Man6-P forms of known lysosomal proteins appear to be quite

widely distributed in the tissues examined and the average number of tissues in which each protein was detected was ~15. For the proteins not assigned to the lysosome, the pattern of distribution was very different (Fig. 4, Panel B) with the majority of proteins found in three or less tissue samples.

Given that known lysosomal proteins tended to be relatively widely distributed, we considered the possibility that tissue distribution could help in the identification of candidates. In Fig. 5, we examined the tissue distribution of the individual groups of proteins that were categorized according to their enrichment in the specific eluate. For the proteins that were unclassified or categorized as not lysosomal, few were widely distributed, and each protein was found in an average of 1.3 and 2.4 tissues, respectively (Fig. 5, Panels A and B). More of the secondary candidates were widely distributed (each protein was found in an average of 4.4 tissues) but the majority were still found in few (≤ 3) tissues (Panel C). In contrast, for the primary lysosomal candidates, many more proteins were widely distributed and each was found in 7.9 tissues on average (Panel D).

Given that known lysosomal proteins tend to be widely distributed, we examined the list of lysosomal candidates for proteins that are found in 13 or more of the 17 tissues analyzed (Fig. 6). We identify a number of proteins that are both enriched and widely distributed that are particularly promising candidates for lysosomal localization including several orthologs of known lysosomal proteins. However, while tissue distribution can help in identifying candidates, it cannot be used to exclude candidates as the population of previously discovered lysosomal proteins may be biased towards the most abundant lysosomal proteins with the widest tissue distribution. Some undiscovered lysosomal proteins (whose identification and classification is the goal of this study) may have escaped assignment to this organelle because they are rare or have very limited distribution.

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 sources¹⁰⁻²⁰ 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-DIRT procedure for identifying specific members of a protein complex that are isolated by the affinity tagging of one of its constituents³², 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^{11-14, 16-20}.

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 proposed³³. Retinoid inducible serine carboxypeptidase (RISC) is 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^{14, 17} protein that was recently demonstrated to be lysosomal³⁶. Based upon sequence homology, both FLJ22662 and LOC196463 may have phosphodiesterase activity.

However, for many of the enriched proteins, it is not easy to predict whether or not a lysosomal function is likely but frequently we find more than one representative of a particular class of protein. For example, while glycosyltransferase would not appear to be a classical lysosomal activity, we find enzymes of 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 microsomal-associated 6). However, it is also possible that these proteins are representatives of a hitherto unsuspected class of lysosomal protein. As noted previously¹⁷, 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 non-lysosomal 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 do contain Man6-P. Earlier studies that directly demonstrated sites of Man6-phosphorylation on a number of apparently non-lysosomal proteins tend to support this conclusion¹⁹.

It is worth noting that one property that is consistent with localization within the lumen of the lysosome is the presence of a signal sequence. In this study, we find that signal domains are predicted for the vast majority (46/52) of the identified proteins that are not assigned to the lysosome but which are enriched in the specific eluate to the same extent as known lysosomal proteins.

The demonstration here of numerous proteins that are not thought to reside within the lysosome that likely contain Man6-P raises two intriguing alternatives. First, the proteome of luminal resident lysosomal proteins could be considerably larger and more diverse than is currently thought. If this is the case, then the functional significance of the lysosomal residence of these “new” proteins would need to be carefully evaluated. Second, even if they are mannose 6-phosphorylated, it is possible that these proteins have no physiological role in the lysosome.

For instance, it is possible that the presence of Man6-P could simply indicate that these proteins represent low affinity substrates for the Man6-phosphotransferase and thus a proportion of a given protein may receive the Man6-P modification resulting in aberrant targeting to the lysosome that may be of little biological significance. Alternatively, there may be some non-lysosomal Man6-P glycoproteins that are not efficiently bound by the MPRs under physiological conditions but which are isolated when exposed to the high concentration of coupled MPR used in our affinity purification protocol.

In order to differentiate between these possibilities, methods for the accurate subcellular localization of the candidates identified here will be required. This can be achieved on a case-by-case basis (e.g., by generating appropriate antibody reagents and performing biochemical and morphological localization studies^{14, 36, 37}) but combining the resolution of subcellular fractionation with the sensitivity of mass spectrometry for protein identification together with the sort of technical and biostatistical approaches to validate conclusions described here and elsewhere⁸ currently appears to be the most promising route. In this approach, subcellular fractions are prepared that are enriched for lysosomal activities and the protein composition of these fractions are investigated using various mass spectrometric proteomic analyses. In principle, this could be an effective method for characterizing both the soluble component of the lysosome as well as the membrane proteins associated with this organelle. 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 towards 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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

References

1. Holtzman, E. Lysosomes. Plenum Press; New York: 1989. p. xvii. 439
2. Scriver, CR. The metabolic & molecular bases of inherited disease. Vol. 8th ed.. McGraw-Hill; New York: 2001. p. 4 v.p. xlvii. 6338p. I-140.
3. Sleat DE, Jadot M, Lobel P. Lysosomal proteomics and disease. *Proteomics - Clinical Applications*. 2007
4. Fan X, Zhang H, Zhang S, Bagshaw RD, Tropak MB, Callahan JW, Mahuran DJ. Identification of the gene encoding the enzyme deficient in mucopolysaccharidosis IIIC (Sanfilippo disease type C). *Am J Hum Genet* 2006;79(4):738–44. [PubMed: 16960811]
5. Naureckiene S, Sleat DE, Lackland H, Fensom A, Vanier MT, Wattiaux R, Jadot M, Lobel P. Identification of HE1 as the second gene of Niemann-Pick C disease. *Science* 2000;290(5500):2298–301. [PubMed: 11125141]
6. Bagshaw RD, Mahuran DJ, Callahan JW. A proteomic analysis of lysosomal integral membrane proteins reveals the diverse composition of the organelle. *Mol Cell Proteomics* 2005;4(2):133–43. [PubMed: 15579476]
7. Chataway TK, Whittle AM, Lewis MD, Bindloss CA, Davey RC, Moritz RL, Simpson RJ, Hopwood JJ, Meikle PJ. Two-dimensional mapping and microsequencing of lysosomal proteins from human placenta. *Placenta* 1998;19(8):643–54. [PubMed: 9859869]

8. Schroder B, Wrocklage C, Pan C, Jager R, Kusters B, Schafer H, Elsasser HP, Mann M, Hasilik A. Integral and associated lysosomal membrane proteins. *Traffic* 2007;8(12):1676–86. [PubMed: 17897319]
9. Ghosh P, Dahms NM, Kornfeld S. Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol* 2003;4(3):202–12. [PubMed: 12612639]
10. Sleat DE, Sohar I, Lackland H, Majercak J, Lobel P. Rat brain contains high levels of mannose-6-phosphorylated glycoproteins including lysosomal enzymes and palmitoyl-protein thioesterase, an enzyme implicated in infantile neuronal lipofuscinosis. *J Biol Chem* 1996;271(32):19191–8. [PubMed: 8702598]
11. Czupalla C, Mansukoski H, Riedl T, Thiel D, Krause E, Hoflack B. Proteomic analysis of lysosomal acid hydrolases secreted by osteoclasts: implications for lytic enzyme transport and bone metabolism. *Mol Cell Proteomics* 2006;5(1):134–43. [PubMed: 16215273]
12. Journet A, Chapel A, Kieffer S, Louwagie M, Luche S, Garin J. Towards a human repertoire of monocytic lysosomal proteins. *Electrophoresis* 2000;21(16):3411–9. [PubMed: 11079561]
13. Journet A, Chapel A, Kieffer S, Roux F, Garin J. Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. *Proteomics* 2002;2(8):1026–40. [PubMed: 12203898]
14. Kollmann K, Mutenda KE, Balleininger M, Ecker mann E, von Figura K, Schmidt B, Lubke T. Identification of novel lysosomal matrix proteins by proteome analysis. *Proteomics* 2005;5(15):3966–78. [PubMed: 16145712]
15. Sleat DE, Kraus SR, Sohar I, Lackland H, Lobel P. alpha-Glucosidase and N-acetylglucosamine-6-sulphatase are the major mannose-6-phosphate glycoproteins in human urine. *Biochem J* 1997;324(Pt 1):33–9. [PubMed: 9164838]
16. Sleat DE, Lackland H, Wang Y, Sohar I, Xiao G, Li H, Lobel P. The human brain mannose 6-phosphate glycoproteome: a complex mixture composed of multiple isoforms of many soluble lysosomal proteins. *Proteomics* 2005;5(6):1520–32. [PubMed: 15789345]
17. Sleat DE, Wang Y, Sohar I, Lackland H, Li Y, Li H, Zheng H, Lobel P. Identification and validation of mannose 6-phosphate glycoproteins in human plasma reveal a wide range of lysosomal and non-lysosomal proteins. *Mol Cell Proteomics* 2006;5(10):1942–56. [PubMed: 16709564]
18. Sleat DE, Zheng H, Lobel P. The human urine mannose 6-phosphate glycoproteome. *Biochim Biophys Acta*. 2006
19. Sleat DE, Zheng H, Qian M, Lobel P. Identification of sites of mannose 6-phosphorylation on lysosomal proteins. *Mol Cell Proteomics* 2006;5(4):686–701. [PubMed: 16399764]
20. Qian M, Sleat DE, Zheng H, Moore D, Lobel P. Proteomics analysis of serum from mutant mice reveals lysosomal proteins selectively transported by each of the two mannose 6-phosphate receptors. *Mol Cell Proteomics* 2008;7(1):58–70. [PubMed: 17848585]
21. Bots M, Medema JP. Granzymes at a glance. *J Cell Sci* 2006;119(Pt 24):5011–4. [PubMed: 17158907]
22. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;72:248–54. [PubMed: 942051]
23. Beavis RC. Using the global proteome machine for protein identification. *Methods Mol Biol* 2006;328:217–28. [PubMed: 16785652]
24. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* 2004;3(6):1234–42. [PubMed: 15595733]
25. Fink JL, Aturaliya RN, Davis MJ, Zhang F, Hanson K, Teasdale MS, Kai C, Kawai J, Carninci P, Hayashizaki Y, Teasdale RD. LOCATE: a mouse protein subcellular localization database. *Nucleic Acids Res* 2006;34(Database issue):D213–7. [PubMed: 16381849]
26. Sprenger J, Lynn Fink J, Karunaratne S, Hanson K, Hamilton NA, Teasdale RD. LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res* 2008;36(Database issue):D230–3. [PubMed: 17986452]
27. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76(14):4193–201. [PubMed: 15253663]
28. Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, Hettich RL, Samatova NF. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res* 2006;5(11):2909–18. [PubMed: 17081042]

29. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* 1927;22:209–212.
30. Glombitza GJ, Becker E, Kaiser HW, Sandhoff K. Biosynthesis, processing, and intracellular transport of GM2 activator protein in human epidermal keratinocytes. The lysosomal targeting of the GM2 activator is independent of a mannose-6-phosphate signal. *J Biol Chem* 1997;272(8):5199–207. [PubMed: 9030589]
31. Rigat B, Wang W, Leung A, Mahuran DJ. Two mechanisms for the recapture of extracellular GM2 activator protein: evidence for a major secretory form of the protein. *Biochemistry* 1997;36(27):8325–31. [PubMed: 9204879]
32. Tackett AJ, DeGrasse JA, Sekedat MD, Oeffinger M, Rout MP, Chait BT. I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J Proteome Res* 2005;4(5):1752–6. [PubMed: 16212429]
33. Wright KO, Messing EM, Reeder JE. Increased expression of the acid sphingomyelinase-like protein ASML3a in bladder tumors. *J Urol* 2002;168(6):2645–9. [PubMed: 12442002]
34. Chen J, Streb JW, Maltby KM, Kitchen CM, Miano JM. Cloning of a novel retinoid-inducible serine carboxypeptidase from vascular smooth muscle cells. *J Biol Chem* 2001;276(36):34175–81. [PubMed: 11447226]
35. Lee TH, Streb JW, Georger MA, Miano JM. Tissue expression of the novel serine carboxypeptidase Scep1. *J Histochem Cytochem* 2006;54(6):701–11. [PubMed: 16461364]
36. Jensen AG, Chemali M, Chapel A, Kieffer-Jaquinod S, Jadot M, Garin J, Journet A. Biochemical characterization and lysosomal localization of the mannose-6-phosphate protein p76 (hypothetical protein LOC196463). *Biochem J* 2007;402(3):449–58. [PubMed: 17105447]
37. Della Valle MC, Sleat DE, Sohar I, Wen T, Pintar JE, Jadot M, Lobel P. Demonstration of lysosomal localization for the mammalian ependymin-related protein using classical approaches combined with a novel density shift method. *J Biol Chem* 2006;281(46):35436–45. [PubMed: 16954209]

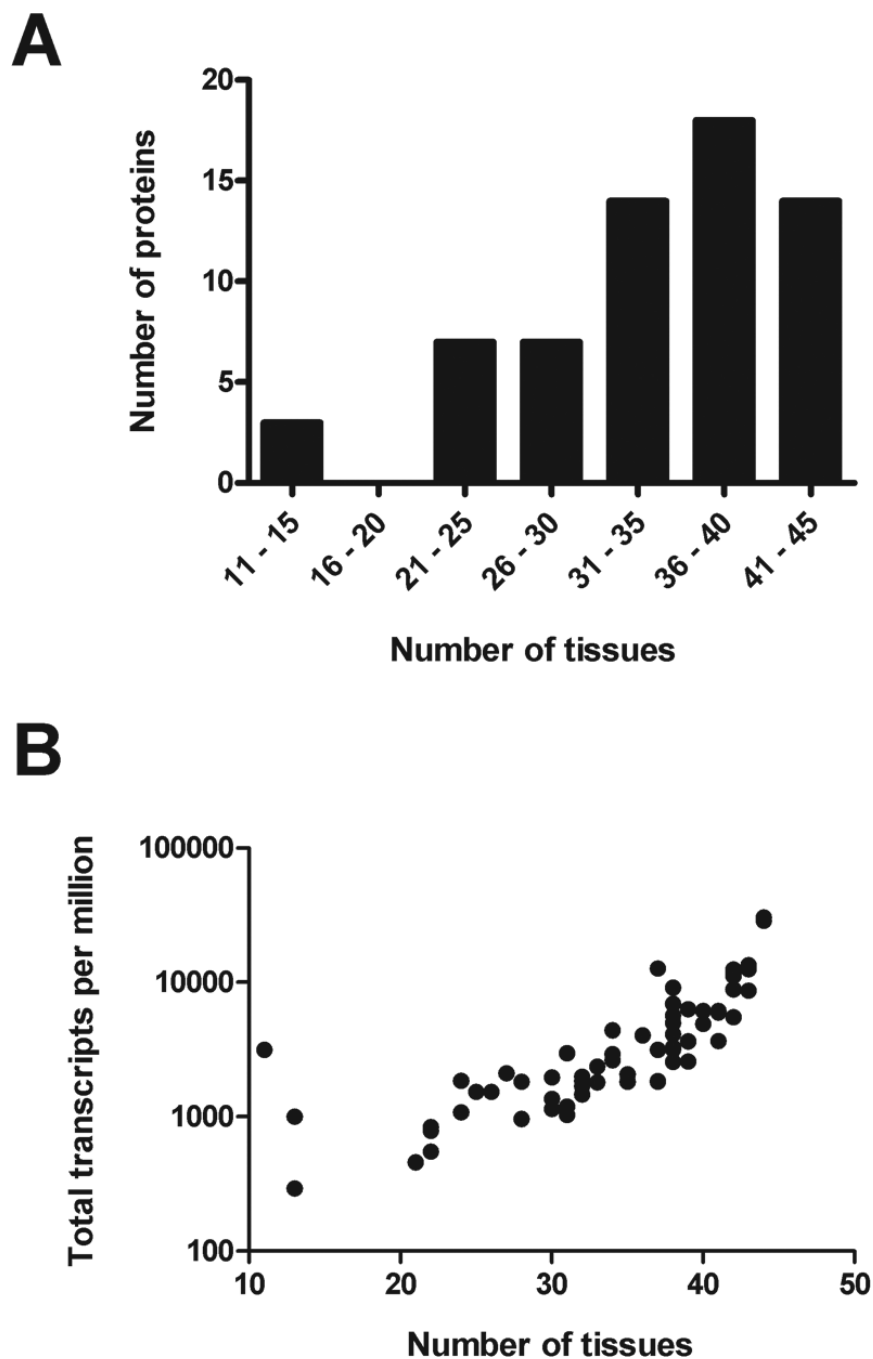


Figure 1. Distribution of known lysosomal proteins based upon expression profiling based upon number of EST counts. Panel A. Expression profiles of 63 known soluble lysosomal proteins were determined by the detection of transcripts in 45 different tissue samples in the human Unigene Build #247 database (<http://www.ncbi.nlm.nih.gov/UniGene/>). Transcript count data are shown in Supplemental Table 1. Panel B. Sum of transcripts per million for each lysosomal protein in all tissues as a function of the number of tissues in which ESTs corresponding to each protein were found.

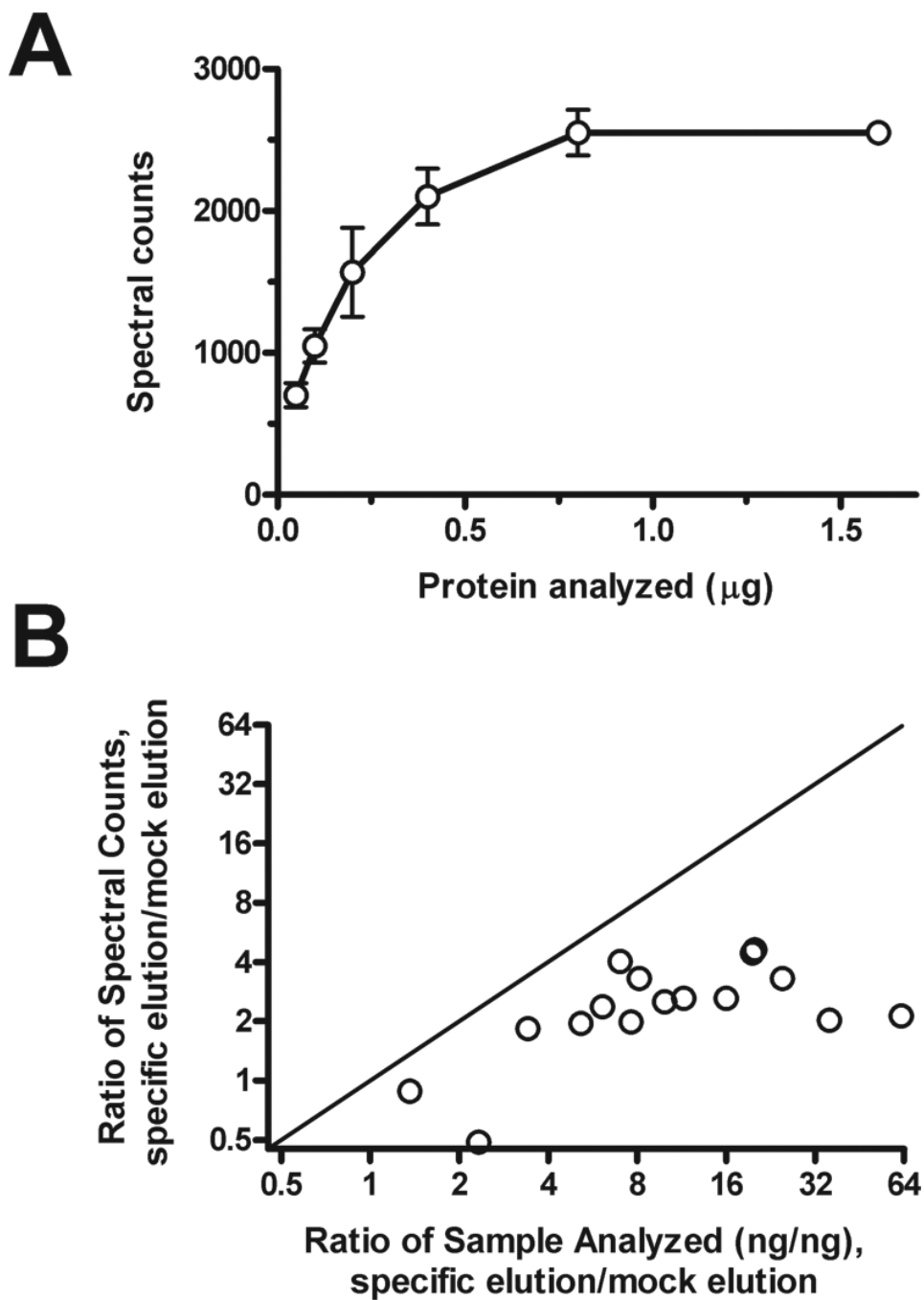


Figure 2. Relationship between amount of sample analyzed and number of spectral counts observed. Panel A. Increasing amounts of a tryptic digest of a rat liver preparation were analyzed in duplicate by LC-MS and the total number of spectral counts determined for each sample. Error bars represent range. Panel B. For each tissue sample, the ratio of spectral counts in specific versus mock eluates is plotted against the ratio of protein analyzed in the two samples.

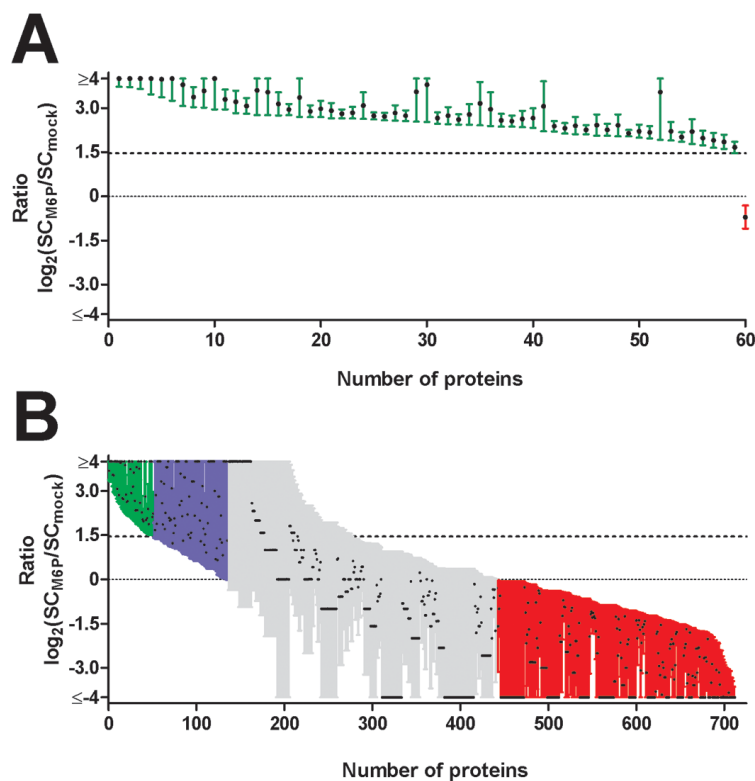


Figure 3.

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_{M6P} and SC_{MOCK} , respectively) is plotted with bars representing the upper and lower 95% confidence indices for known lysosomal proteins (Panel A) or proteins not currently classified as lysosomal (Panel B). The lower confidence index 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 ($\log_2=0$)); grey error bars, proteins that cannot be classified (95% confidence interval includes $SC_{M6P}/SC_{MOCK}=1$); 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 than 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 are not of rodent origin are not shown.

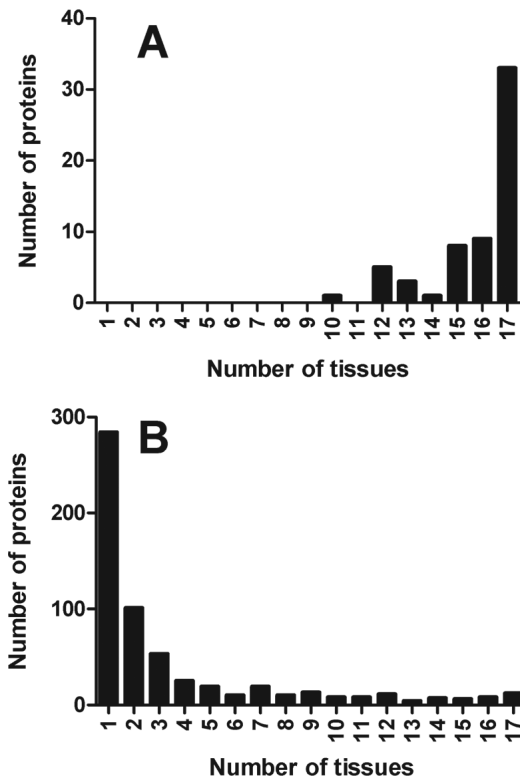


Figure 4.
Distribution of affinity purified proteins. The number of tissues is shown in which lysosomal proteins reported to contain Man6-P (Panel A) and the remaining proteins (Panel B) were identified in Man6-P eluates.

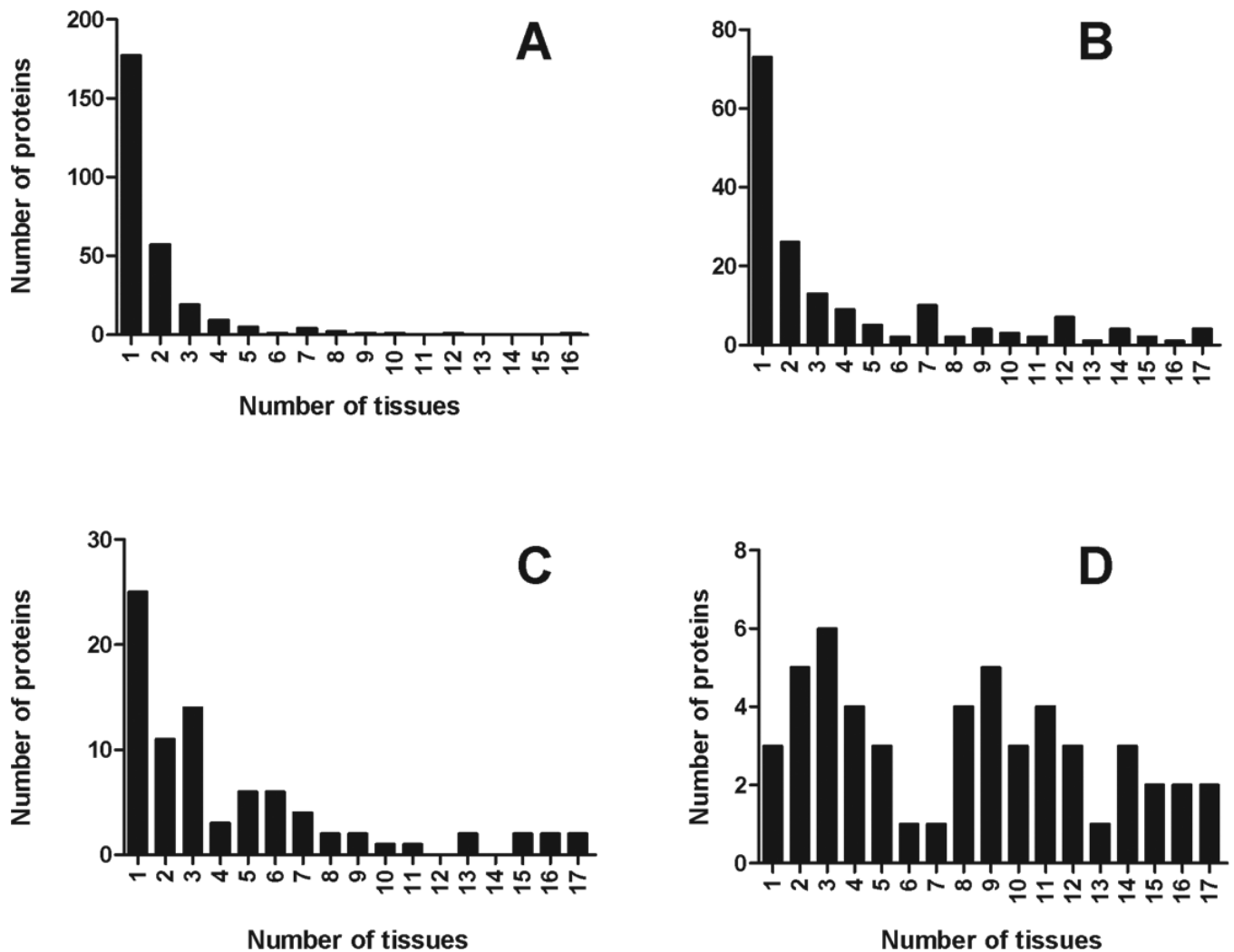


Figure 5. Tissue distribution as a function of enrichment in the specific eluate. The number of tissues is shown for which proteins categorized as unclassified (Panel A), non-lysosomal (Panel B), secondary lysosomal candidates (Panel C) or primary lysosomal candidates (Panel D) were identified in Man6-P eluates.

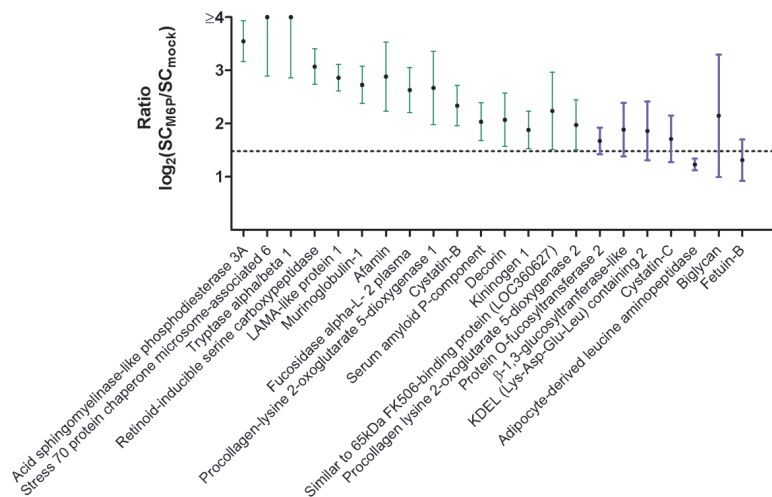


Figure 6. Widely-distributed proteins that are significantly enriched in the specific eluate. These proteins were identified in Man6-P eluates from >12 different tissues. The \log_2 of the ratio of spectral counts in the specific and mock eluates (SC_{M6P} and SC_{MOCK} , respectively) is plotted with bars representing the upper and lower 95% confidence indices. Plots in green represent primary lysosomal candidates (i.e., where the lower 95% confidence index is greater than 1.5, a threshold that is plotted as a dotted line) and plots in blue indicate secondary lysosomal candidates (i.e., proteins that are significantly enriched in the specific eluate).

Table 1

Rat tissue Man6-P glycoprotein purification

Yields are expressed as micrograms of protein recovered per gram tissue wet weight. Load indicates nanograms of protein in a given eluate that was analyzed per duplicate LC-MS/MS run.

Tissue	Starting material		Mock eluate		Specific eluate	
	wet wt (g)	Yield (µg/g)	LC-MS load (ng)	Yield (µg/g)	LC-MS load (ng)	Yield (µg/g)
brain	3.7	0.2	2	11.1	1000	1000
cecum	1.0	0.4	2	10.0	520	520
duodenum	2.3	0.9	19	4.7	1000	1000
heart	5.0	0.3	5	6.5	1000	1000
kidney	5.0	0.4	10	4.0	1000	1000
liver	20.0	0.3	29	0.8	1000	1000
lung	5.0	0.1	3	4.9	1000	1000
mammary gland	4.5	0.9	12	7.4	1000	1000
pancreas and stomach	3.4	1.1	14	7.4	1000	1000
placenta	2.0	2.2	13	16.4	1000	1000
skeletal muscle	3.3	0.5	9	5.6	1000	1000
skin	1.2	2.1	15	2.8	200	200
spleen	3.5	0.2	5	4.5	1000	1000
testis	4.0	2.0	16	11.9	1000	1000
thymus	2.4	0.6	9	1.5	200	200
uterus	0.5	0.2	<1	3.4	80	80
vas deferens	0.2	nd*	nd*	nd*	nd*	nd*

* nd, not determined.

Table 2

Lysosomal proteins in rat tissues

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 expectation score assigned by X! Tandem and "Coverage" indicates the total percentage of sequence within the assigned peptides to each protein.

ENSG Identifier	ENSEMBL Description	Min log(e)	Coverage	Spectral counts (mock eluate)	Spectral counts (specific eluate)	Log ₂ (specific/mock)	Function/class
ENSMUSG000000001348	Acid phosphatase 5, tartrate resistant	-44.6	17	1	72	6.17 (3.65 to 10.47)	e
ENSMUSG000000005043	N-sulfoglucosamine sulfohydrolase	-96.6	26	18	104	2.53 (1.82 to 3.25)	e
ENSMUSG000000016256	Cathepsin Z	-127	44	170	612	1.85 (1.6 to 2.09)	e
ENSMUSG000000025579	Glucosidase, alpha, acid	-223.1	21	14	263	4.23 (3.47 to 5)	e
ENSRNOG000000000043	Iduronidase, alpha-L-	-145	28	6	127	4.4 (3.25 to 5.55)	e
ENSRNOG000000000108	Aspartylglucosaminidase	-321.5	61	138	983	2.83 (2.58 to 3.09)	e
ENSRNOG000000000435	Lysosomal thioesterase PPT2	-92.5	40	14	163	3.54 (2.76 to 4.32)	e
ENSRNOG000000000571	Prosaposin	-326.2	54	164	646	1.98 (1.73 to 2.22)	ap
ENSRNOG000000000913	Beta-glucuronidase	-309	45	118	728	2.63 (2.34 to 2.91)	e
ENSRNOG000000001385	LAMA-like protein 2	-400.1	55	239	1843	2.95 (2.75 to 3.14)	e
ENSRNOG000000001465	Iduronate 2-sulfatase	-131.6	31	16	142	3.15 (2.41 to 3.89)	e
ENSRNOG000000002188	Heparanase	-106.5	35	8	94	3.55 (2.53 to 4.57)	e
ENSRNOG000000002273	N-acyl ethanolamine-hydrolyzing acid amidase	-146.6	39	12	146	3.6 (2.77 to 4.44)	e
ENSRNOG000000003291	Cellular repressor of E1A-stimulated genes 1	-143.6	52	22	346	3.98 (3.36 to 4.59)	u
ENSRNOG000000003759	Galactosylceramidase	-79.8	19	5	69	3.79 (2.52 to 5.05)	e
ENSRNOG000000004919	Glucosamine (N-acetyl)-6-sulfatase	-522	61	389	1867	2.26 (2.11 to 2.42)	e
ENSRNOG000000005526	Mannosidase 2, alpha B2	-353.2	35	29	555	4.26 (3.72 to 4.79)	e
ENSRNOG000000005931	Plasma glutamate carboxypeptidase	-313.3	60	282	890	1.66 (1.47 to 1.85)	e
ENSRNOG000000007089	Legumain	-358	60	204	946	2.21 (2 to 2.43)	e
ENSRNOG000000007351	Gamma-glutamyl hydrolase	-151.6	39	74	764	3.37 (3.02 to 3.71)	e
ENSRNOG000000008064	Alpha-N-acetylgalactosaminidase	-463.8	68	295	1798	2.61 (2.43 to 2.79)	e
ENSRNOG000000008310	Myeloperoxidase	-467.7	57	74	726	3.29 (2.95 to 3.64)	e
ENSRNOG000000009325	Tissue alpha-L-fucosidase	-187.3	41	99	459	2.21 (1.9 to 2.53)	e
ENSRNOG000000010034	N-acylsphingosine amidohydrolase 1	-318.8	61	112	877	2.97 (2.69 to 3.25)	e
ENSRNOG000000010080 / ENSRNOG000000009759	CLN5 protein	-93.7	28	4	81	4.34 (2.95 to 5.73)	u
ENSRNOG000000010196	Galactosidase, beta 1	-471	58	384	1552	2.01 (1.85 to 2.18)	e
ENSRNOG000000010252	Beta-hexosaminidase alpha chain	-533.3	62	282	1887	2.74 (2.56 to 2.92)	e
ENSRNOG000000010331	Cathepsin B	-515.2	82	298	1550	2.38 (2.2 to 2.56)	e

ENSG Identifier	ENSEMBL Description	Min log(e)	Coverage	Spectral counts (mock eluate)	Spectral counts (specific eluate)	Log ₂ (specific/mock)	Function/class
ENSRNOG00000010630	Prolylcarboxypeptidase	-304.4	51	60	556	3.21 (2.83 to 3.6)	e
ENSRNOG00000011150	Arylsulfatase B	-224.9	55	68	362	2.41 (2.04 to 2.79)	e
ENSRNOG00000011513	Galactosidase, alpha	-309.3	54	138	1040	2.91 (2.66 to 3.17)	e
ENSRNOG00000011864	GM2 ganglioside activator protein	-120.4	48	141	86	-0.71 (-1.1 to -0.33)	ap
ENSRNOG00000012062	Epididymal secretory protein E1	-191.2	63	108	571	2.4 (2.11 to 2.7)	ap
ENSRNOG00000012616	Palmitoyl-protein thioesterase 1	-278.3	64	126	1056	3.07 (2.8 to 3.33)	e
ENSRNOG00000012640	Dipeptidyl-peptidase 2	-461.6	56	260	1645	2.66 (2.47 to 2.85)	e
ENSRNOG00000012953	Arylsulfatase A	-248.3	59	86	544	2.66 (2.33 to 2.99)	e
ENSRNOG00000013190	Ribonuclease T2	-204.3	53	74	397	2.42 (2.07 to 2.78)	e
ENSRNOG00000013476	Beta-mannosidase	-489.8	49	226	1673	2.89 (2.69 to 3.09)	e
ENSRNOG00000014064	Cathepsin H	-316.8	62	292	1446	2.31 (2.13 to 2.49)	e
ENSRNOG00000014461	Galactosamine (N-acetyl)-6-sulfate sulfatase	-351.7	61	191	1146	2.58 (2.36 to 2.81)	e
ENSRNOG00000015573	Chitinase, di-N-acetyl-	-238.8	51	107	716	2.74 (2.45 to 3.03)	e
ENSRNOG00000015857	Cathepsin A	-556.2	63	544	3568	2.71 (2.58 to 2.84)	e
ENSRNOG00000016496	Cathepsin C	-782.8	78	427	2968	2.8 (2.65 to 2.94)	e
ENSRNOG00000017977	Sphingomyelin phosphodiesterase 1	-176.3	37	20	205	3.36 (2.7 to 4.01)	e
ENSRNOG00000018566	Cathepsin L	-559.1	79	622	4129	2.73 (2.61 to 2.85)	e
ENSRNOG00000018989	Ependymin related protein 1 (zebrafish)	-181.4	55	255	1490	2.55 (2.36 to 2.74)	u
ENSRNOG00000019077	Lysosomal acid lipase	-136.9	47	44	372	3.08 (2.63 to 3.53)	e
ENSRNOG00000019212	Tripeptidyl peptidase I	-382.3	55	218	979	2.17 (1.96 to 2.38)	e
ENSRNOG00000019387	Interferon gamma inducible protein 30	-150.7	53	26	203	2.96 (2.38 to 3.55)	e
ENSRNOG00000019708	Cathepsin F	-278	60	75	514	2.78 (2.43 to 3.13)	e
ENSRNOG00000019859	Lysophospholipase 3	-155.3	31	16	221	3.79 (3.06 to 4.51)	e
ENSRNOG00000020206	Cathepsin D	-551.7	72	867	3822	2.14 (2.03 to 2.25)	e
ENSRNOG00000021155	Cathepsin K	-178.4	38	12	100	3.06 (2.21 to 3.91)	e
ENSRNOG00000021157	Cathepsin S	-140.5	37	56	257	2.2 (1.78 to 2.61)	e
ENSRNOG00000023830	Deoxyribonuclease-2-alpha	-186.5	58	26	312	3.58 (3.01 to 4.16)	e
ENSRNOG00000023910	Mannosidase 2, alpha B1	-807.4	61	220	1575	2.84 (2.64 to 3.04)	e
ENSRNOG00000025274	Beta-hexosaminidase beta chain	-364.1	52	246	1178	2.26 (2.06 to 2.46)	e
ENSRNOG00000031266	Sialic acid acetyltransferase	-231.6	52	58	513	3.14 (2.75 to 3.54)	e
ENSRNOG00000032381	Alpha-N-acetylglucosaminidase	-370.8	39	46	807	4.13 (3.71 to 4.56)	e
ENSRNOG00000032942	Neuraminidase I	-222.4	45	157	586	1.9 (1.65 to 2.15)	e

Table 3**Classification of affinity purified proteins based upon relative abundance in specific and mock eluates**

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.

Category	Number of proteins
lysosomal	60
not lysosomal	272
primary candidate	52
secondary candidate	84
unclassified	304
Total	772

Table 4

Primary lysosomal candidates that are enriched in the Man6-P eluate to the same extent as known lysosomal proteins

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; 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 expectation confidence score assigned by X! Tandem (see “Experimental”) 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.

ENSG Identifier	ENSEMBL Description	Min log(e)	Coverage	Spectral counts		Log2(specific/mock)	Previously identified	Number of tissues found	Function	Number of potential N-linked glycosylation sites	Signal domain	Protein Class	Subcellular location
				(specific eluate)	(mock eluate)								
ENSRNOG00000011287	Multiple inositol polyphosphate phosphatase 1	-52	15	24	0	∞ (2.64 to ∞)	no	9	e	2	yes	secreted	ER
ENSRNOG00000015941	Similar to 65kDa FK506-binding protein	-141	30	85	18	2.24 (1.51 to 2.97)	no	12	e	7	yes	secreted	ER
ENSRNOG00000010935	Carboxypeptidase B2	-86	33	44	3	3.87 (2.27 to 5.48)	no	8	e	6	yes	secreted	extracellular
ENSRNOG00000008575	Amitofide binding protein 1	-449.5	55	435	111	1.97 (1.67 to 2.27)	no	6	e	4	yes	secreted	extracellular
ENSRNOG00000011913	Ceruloplasmin	-330.8	35	145	30	2.27 (1.71 to 2.84)	yes	9	e	6	yes	secreted	extracellular
ENSRNOG000000030183	Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2	-278.3	42	177	45	1.98 (1.51 to 2.45)	yes	13	e	6	yes	secreted	extracellular, ER
ENSRNOG00000001823	Sialyltransferase 1	-120.8	33	62	7	3.15 (2.05 to 4.25)	no	3	e	3	yes	soluble, non-secreted	Golgi
ENSRNOG00000019014	N-heparan sulfate sulfotransferase 1	-38.5	6.1	14	0	∞ (1.87 to ∞)	no	5	e	4	yes	type II membrane	lysosomes
ENSRNOG00000007763	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	-196.3	35	121	19	2.67 (1.98 to 3.36)	yes	15	e	4	yes	soluble, non-secreted	mitochondria, ER
ENSMUSG000000028015	Cathepsin O	-21	11	16	0	∞ (2.06 to ∞)	yes	9	e	4	yes	secreted	
ENSRNOG000000000815	Acid sphingomyelinase-like phosphodiesterase 3a	-341.8	55	679	58	3.55 (3.16 to 3.93)	yes	17	e	6	yes	secreted	
ENSRNOG000000002358	Retinoid-inducible serine carboxypeptidase	-211.6	41	673	80	3.07 (2.74 to 3.41)	yes	16	e	5	yes	secreted	
ENSRNOG000000008422	Lactoperoxidase	-79.9	19	23	0	∞ (2.58 to ∞)	no	3	e	6	yes	secreted	
ENSRNOG000000008933	LAMA-like protein 1	-409.8	52	1062	146	2.86 (2.61 to 3.11)	yes	16	e	5	yes	secreted	
ENSRNOG00000011181	Carboxypeptidase A3	-55.2	16	27	2	3.75 (1.83 to 5.68)	no	6	e	4	yes	secreted	
ENSRNOG00000015551	Fucosidase, alpha-L-2, plasma	-221.4	50	322	52	2.63 (2.21 to 3.05)	yes	15	e	5	yes	secreted	
ENSRNOG00000017908	Pancreatic lipase-related protein 1	-214.1	43	117	27	2.12 (1.52 to 2.71)	no	3	e	3	yes	secreted	
ENSRNOG00000018236	Chymosin family	-253.7	66	323	28	3.53 (2.97 to 4.08)	no	5	e	1	yes	secreted	
ENSRNOG000000024181	Tryptase alpha/beta 1	-59.2	33	54	2	4.75 (2.86 to 6.65)	no	12	e	2	yes	secreted	
ENSRNOG000000026937	Arylsulfatase K	-144.5	27	57	1	5.83 (3.31 to 10.14)	yes	8	e	7	yes	secreted	

ENSG Identifier	ENSEMBL Description	Min log(e)	Coverage	Spectral counts (specific eluate)	Spectral counts (mock eluate)	Log2(specific/mock)	Previously identified	Number of tissues found	Function	Number of potential N-linked glycosylation sites	Signal domain	Protein Class	Subcellular location
ENSRNOG000000028092	Carboxypeptidase A2	-100.7	34	37	3	3.62 (2.01 to 5.24)	no	2	e	0	yes	secreted	
ENSRNOG000000029604	Mast cell protease 8	-235.7	49	241	38	2.66 (2.17 to 3.16)	no	9	e	4	yes	secreted	
ENSRNOG000000030909	Mast cell protease 9	-445.3	76	1121	87	3.69 (3.37 to 4)	no	11	e	2	yes	secreted	
ENSRNOG000000032717	Granzyme-like protein 2	-398.4	56	439	65	2.76 (2.38 to 3.13)	no	10	e	3	yes	secreted	
ENSRNOG000000039865	Vannin 3	-167	39	121	5	4.6 (3.35 to 5.85)	no	8	e	3	yes	soluble, non-secreted	
ENSRNOG000000039971	Cytotoxic T lymphocyte-associated protein 2 beta -like	-55.5	57	38	0	∞ (3.31 to ∞)	no	6	e	0	no	soluble, non-secreted	
ENSRNOG00000009086	Serum amyloid P-component	-144.6	58	320	78	2.04 (1.68 to 2.39)	yes	15	1	2	yes	secreted	extracellular
ENSRNOG00000009217	F-box only protein 6	-71.9	31	62	4	3.95 (2.55 to 5.36)	yes	11	1	1	no	soluble, non-secreted	
ENSRNOG000000032834	Stress 70 protein chaperone microsome-associated 6	-138.4	34	67	3	4.48 (2.89 to 6.07)	yes	14	0	5	yes	secreted	ER
ENSRNOG000000020129	P-cadherin	-41.4	9.1	11	0	∞ (1.52 to ∞)	no	1	0	4	yes	type I membrane	plasma Membrane
ENSRNOG000000008838	Gastrokine 1	-66.8	34	91	15	2.6 (1.82 to 3.38)	no	5	0	1	yes	secreted	secretory granule
ENSRNOG000000008716	Neurofilament heavy polypeptide	-138.5	14	28	2	3.81 (1.88 to 5.73)	no	4	0	1	no	soluble, non-secreted	
ENSRNOG000000018400	Golgi membrane protein 1	-26.7	7.4	11	0	∞ (1.52 to ∞)	no	4	0	4	yes	type II membrane	cytoplasm
ENSRNOG000000013572	Latexin	-42.2	26	21	0	∞ (2.45 to ∞)	no	4	pi	3	no	soluble, non-secreted	
ENSRNOG000000030387	Kininogen 1	-253.1	56	302	82	1.88 (1.53 to 2.23)	yes	13	pi	5	yes	secreted	extracellular
ENSRNOG000000005599	Alpha-1-inhibitor 3	-606.3	13	99	13	2.93 (2.11 to 3.75)	yes	11	pi	12	yes	secreted	extracellular
ENSRNOG0000000037188	Murine globulin-1	-635.5	44	497	75	2.73 (2.38 to 3.08)	yes	15	pi	1	yes	secreted	lysosomes
ENSRNOG000000001201	Cystatin-B	-96.7	72	329	65	2.34 (1.96 to 2.72)	yes	17	pi	0	no	soluble, non-secreted	nucleus, Cytoplasm
ENSRNOG000000009855	Serine (or cysteine) peptidase inhibitor, clade A,	-38	21	13	0	∞ (1.76 to ∞)	yes	1	pi	1	yes	secreted	
ENSRNOG000000020455	Cystatin E/M	-62.4	51	44	1	5.46 (2.93 to 9.78)	no	9	pi	0	yes	secreted	
ENSRNOG000000033245	Murine globulin family	-340.7	3	12	0	∞ (1.64 to ∞)	no	5	pi	14	yes	secreted	
ENSRNOG000000010527	Serine (or cysteine) peptidase inhibitor, clade A,	-227.7	57	209	41	2.35 (1.87 to 2.83)	no	10	pi	3	yes	type I membrane	
ENSRNOG000000004067	Neuronal cell adhesion molecule	-51.7	6.7	11	0	∞ (1.52 to ∞)	yes	1	s	23	yes	type I membrane	plasma Membrane
ENSMUSG000000024109	Neurexin I	-64.6	9.3	17	0	∞ (2.15 to ∞)	yes	2	s	2	no	type II membrane	plasma Membrane
ENSRNOG000000004610	Lumican	-52.8	26	40	3	3.74 (2.13 to 5.35)	yes	11	s	4	yes	secreted	extracellular
ENSRNOG000000004554	Decorin	-134	31	164	39	2.07 (1.57 to 2.57)	no	12	s	4	yes	secreted	extracellular
ENSRNOG000000015410	Similar to asporin	-75.1	28	42	6	2.81 (1.61 to 4.01)	no	11	s	2	yes	secreted	extracellular
ENSRNOG000000002878	Afamin	-206.6	36	155	21	2.88 (2.23 to 3.54)	yes	12	tp	6	yes	secreted	extracellular

ENSG Identifier	ENSEMBL Description	Min log(e)	Coverage	Spectral counts (specific eluate)	Spectral counts (mock eluate)	Log2(specific/mock)	Previously identified	Number of tissues found	Function	Number of potential N-linked glycosylation sites	Signal domain	Protein Class	Subcellular location
ENSRNOG000000021001	Gastric intrinsic factor	-32	11	17	0	∞ (2.15 to ∞)	no	2	tp	5	yes	secreted	
ENSRNOG000000020148	Interleukin-4 induced gene-1	-135.2	25	68	13	2.39 (1.54 to 2.23)	no	5	o	3	yes	multipass membrane	nucleus
ENSRNOG000000021258	Endogenous retroviral sequence 3	-174.9	64	226	56	2.01 (1.59 to 2.43)	no	4	o	3	yes	type I membrane	
ENSRNOG000000037782	Gastrokine family	-27	19	19	0	∞ (2.31 to ∞)	no	3	o	1	yes	type II membrane	