Quantitative Profile of Five Murine Core Proteomes Using Label-free Functional Proteomics*§

Pedro R. Cutillas‡§¶¶ and Bart Vanhaesebroeck‡¶¶**

Analysis of primary animal and human tissues is key in biological and biomedical research. Comparative proteomics analysis of primary biological material would benefit from uncomplicated experimental work flows capable of evaluating an unlimited number of samples. In this report we describe the application of label-free proteomics to the quantitative analysis of five mouse core proteomes. We developed a computer program and normalization procedures that allow exploitation of the quantitative data inherent in LC-MS/MS experiments for relative and absolute quantification of proteins in complex mixtures. Important features of this approach include (i) its ability to compare an unlimited number of samples, (ii) its applicability to primary tissues and cultured cells, (iii) its straightforward work flow without chemical reaction steps, and (iv) its usefulness not only for relative quantification but also for estimation of absolute protein abundance. We applied this approach to quantitatively characterize the most abundant proteins in murine brain, heart, kidney, liver, and lung. We matched 8,800 MS/MS peptide spectra to 1,500 proteins and generated 44,000 independent data points to profile the ~1,000 most abundant proteins in mouse tissues. This dataset provides a quantitative profile of the fundamental proteome of a mouse, identifies the major similarities and differences between organ-specific proteomes, and serves as a paradigm of how label-free quantitative MS can be used to characterize the phenotype of mammalian primary tissues at the molecular level. Molecular & Cellular Proteomics 6: 1560–1573, 2007.

Experiments on immortalized cell lines have resulted in the generation of a vast amount of information on the biological and biochemical processes that govern the function of cultured cells. However, discerning the mechanisms by which genes control mammalian physiology in vivo may only be achieved by investigations that involve the use of animal models of which the laboratory mouse (Mus musculus) offers many advantages (1). There is a wealth of resources related to the molecular biology of mouse cells that have benefited from genomics, transcriptomics, and, more recently, proteomics projects aimed at profiling the molecular composition of murine tissues (2–4). The generation of gene-targeted mice is particularly useful in advancing our understanding of how genes control fundamental processes of mammalian physiology (e.g. Refs. 5–9). Once created, finding the phenotype of a gene-targeted strain of mice is not a trivial task (1, 10). Several years are needed to fully characterize phenotypic alterations, and subtle phenotypes often go unnoticed. Robust and high throughput methods for profiling the proteomes of primary tissues in a quantitative fashion may expedite the search for phenotypic changes in gene-targeted and other animal models. Insights gained by powerful methods for the molecular characterization of primary tissues could also direct classical physiological experiments, reduce the number of experimental animals, and more comprehensively exploit the scientific knowledge that can be gained from animal models.

Thus, phenotypes could in principle be characterized systematically by comparing the proteomes of primary cells using unbiased proteomics approaches based on MS. Several analytical strategies exists that use MS for relative quantification of proteins and proteomes. However, current methods for quantitative proteomics have shortcomings for the analysis of primary tissues. First, metabolic labeling, the ideal approach for quantitative proteomics should not rely on chemical derivatization strategies if the method is to be useful for the comparison of an unlimited number of samples and to provide statistically sound results. Thus although powerful to quantify proteins from immortalized cell lines (11), stable isotope labeling by amino acids in cell culture (SILAC) and other metabolic labeling approaches for quantitative MS cannot be used to easily quantify proteins from primary tissues. Second, an ideal approach for quantitative proteomics should not rely on chemical derivatization strategies if the method is to be useful for the comparison of an unlimited number of samples and to provide statistically sound results. Thus although powerful in other contexts, strategies that rely on chemical labeling with compounds enriched in isotopes (e.g. isobaric tags for relative and absolute quantification (TROA) and ICAT approaches (12, 13)), heavy water (14), or fluorescence labels (e.g. the two-dimensional Dige approach (15)) are not ideal for the analysis of gene-targeted mice (or for clinical studies). This shortcom-
ing is also encountered when using metabolic labeling approaches. Third, the method should provide sufficient throughput, precision, and dynamic range, and the ideal technique for proteome profiling should therefore consider the trade-off between speed and depth of analysis. Approaches for extensive characterization of proteomes, which rely on complex cellular fractionation, are useful for characterizing the molecular architecture of cells (4, 16), but unfortunately these strategies do not offer the throughput required for comparison of proteomes.

We and others have shown that the data in LC-MS and LC-MS/MS experiments have inherent quantitative information such that it is possible to use this type of data to assess protein amounts in cells and tissues (17–19). Although label-free methods for quantitative proteomics based on spectral counts have been described (4, 20, 21), their level of precision only permits the use of spectral counts as an approximate indication of protein abundance (17). In contrast, quantitative MS methods based on the determination of peptide ion intensities may offer levels of precision close to those obtained by isotopic labeling strategies (17, 19). The aim of the current study was to evaluate the performance of a label-free quantitative proteomics approach (which we designed taking into account the considerations listed above) for the analysis of primary tissues. We report on the creation of a computer program and on the development of standardization procedures for downstream data processing that can be used to automate the quantitative analysis of label-free LC-MS/MS data such that this approach can be used for large scale comparative analysis of any protein mixture, including those in mammalian primary tissues. After assessing the performance of the created protocols, we used these tools in a proof-of-principle experiment aimed at obtaining a low resolution map of the major proteins in the mouse. We found that, in addition to cost and simplicity, an advantage of label-free methods is that they may be useful for providing relative and absolute values of protein amounts simultaneously. Our results describe five core proteomes of a mouse in quantitative terms, provide new insights into the major similarities and differences between the protein compositions of the main murine organs, and serve as an example of how this approach may be used to characterize mammalian organisms phenotypically at the molecular level.

**EXPERIMENTAL PROCEDURES**

**Tissue Extraction**

Cell culture reagents were from Invitrogen. The WEHI-231 B cell line was cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and 0.05 mM β-mercaptoethanol at 37 °C in 5% CO2. Mouse tissues were obtained from a typical 12-week-old C57BL/6 genetic background that was killed by cervical dislocation. Tissues were excised and processed without freezing steps.

Cells were lysed in Triton X-100 lysis buffer (150 mM NaCl, 1% (w/v) Triton X-100, 1 mM EDTA, 50 mM Tris-HCl, pH 7.4) supplemented with protein and phosphatase inhibitors. Fresh primary tissues were homogenized in Triton X-100 lysis buffer using a micropestle (Eppendorf). Protein concentrations in cell lysates and organ homogenates were determined using the Bradford assay.

**Immunoblotting**

Proteins for Western blotting were separated by 10% SDS-PAGE and transferred to PVDF membranes. Membranes were blocked with 5% skimmed milk in TBS-T buffer (20 mM Tris•HCl, pH 8.0, 150 mM NaCl, 0.1% Tween) and then probed with monoclonal antibodies against actin or tubulin (both from Sigma) followed by incubations with IRDye 800CW goat anti-mouse secondary antibody (LI-COR Biosciences, Cambridge, UK). Fluorescent immunoblot signals were quantified using an Odyssey imaging system (LI-COR Biosciences) and reported as -fold over the mean intensities of the samples to be compared.

**Preparation of Samples for MS**

Proteins were prepared for MS essentially as described previously (19). Briefly cell lysates or tissue homogenates were separated by 10% SDS-PAGE and visualized by colloidal Coomassie Blue staining. Gels were scanned in a Bio-Rad G-800 densitometer. The OD of each of the sections to be excised for downstream MS analysis was determined and expressed as a percentage of the OD of the total lane (this step was important for the estimation of absolute protein amounts; see below). Proteins in these gel sections were extracted by in-gel digestion using standard procedures with the exception that 2 pmol of a standard protein (fetuin or lysozyme) was added to each gel piece prior to in-gel digestion. These proteins served as internal standards to normalize intensity readings of endogenous proteins.

**Mass Spectrometry**

Protein-derived peptides were analyzed by LC-MS/MS in a Q-Tof-1 mass spectrometer (Micromass, Manchester, UK) connected on line with an Ultimate nanoflow HPLC system (LC-Packings/Dionex, Amsterdam, The Netherlands). Settings and conditions for this setup have been described (19, 22). Gradient elution was applied from 5% B to 40% B in 60 min followed by a ramp to 90% B over 5 min. Solvent B was 80% (v/v) acetonitrile, 0.1% (v/v) formic acid, and the balance solvent A was 0.1% (v/v) formic acid. To ensure that chromatographic peak shapes obtained from MS traces were compatible with quantification and that information such as peak heights were not lost during MS/MS analyses, we performed data-dependent acquisition (DDA) experiments with settings that switched from MS/MS to MS after just 1 s of MS/MS acquisition.

**Data Analysis**

Identification of Proteins by Database Searching—Deisotoped peak lists, obtained from MS/MS raw data using Distiller Version 1 (Matrix Science, London, UK), were used by Mascot Version 2.1.03 (Matrix Science) to interrogate the National Center for Biotechnology Information non-redundant (NCBI nr) July 6, 2005 protein database (containing 2,543,645 sequences) or the mouse International Protein Index (IPI) Information non-redundant (NCBInr) July 6, 2005 protein database (Matrix Science, London, UK), were used by Mascot Version 2.1.03 (Matrix Science) to interrogate the National Center for Biotechnology Information non-redundant (NCBI nr) July 6, 2005 protein database (containing 51,559 sequences). We used Integra (Matrix Science) for the automation of these database searches. This is a laboratory information management system that

---

1 The abbreviations used are: DDA, data-dependent acquisition; CV, coefficient of variation; GO, gene ontology; XIC, extracted ion chromatogram; Pescal, Peak Statistic Calculator; $t_{RP}$, retention time.
also allows for effective management of proteomics data. Settings were as follows: mass accuracy window for parent ion, 100 ppm; mass accuracy window for fragment ions, 150 millimass units; fixed modification, carbamidomethylation of cysteines; variable modifications, oxidation of methionine and acetylation of N termini. We accepted the returned protein identifications when Mascot scores were above the statistically significant threshold ($p < 0.05$) and at least two peptides matched the identified protein. Results from database searches were imported into Excel files.

**Equation 1.**

$$ RQ_{PEPT} = \frac{NI_{x}}{\frac{EI_{x}}{EI_{std}} \cdot \prod_{i=1}^{n} NI_{i}} \quad (Eq. 1) $$

**Equation 2.**

$$ RQ_{PROT} = \frac{1}{m} \cdot \sum_{i=1}^{m} RQ_{PEPT} \quad (Eq. 2) $$

For the estimation of variation, the standard deviation (S.D.) of the mean ($RQ_{PROT}$) was calculated taking $N$ as the number of peptides quantified per protein (Excel was used for these statistical analyses). The percentage coefficient of variance (CV%) was calculated by dividing the S.D. by $RQ_{PROT}$ times 100.

For absolute quantification of proteins ($AQ_{PROT}$), the intensities of peptides matching a given protein were added up. This aggregate intensity value for a given protein was then divided by the total ion current (TIC; which for the purpose of this manuscript was defined as the sum of the intensity values of all the identified peptides for all proteins in an LC-MS/MS experiment) and multiplied by the percentage of OD that this gel section ($OD_{BAND}$) had relative to the OD of the total gel lane ($OD_{LANE}$; Equation 4).

$$ AQ_{PROT} = \frac{\sum_{x=1}^{m} NI_{x} \cdot OD_{BAND}}{OD_{LANE}} \cdot 100 \quad (Eq. 4) $$

Hierarchical cluster analysis of quantitative LC-MS/MS data was performed using Cluster and Treeview (23) by complete linkage clustering using Pearson correlation (uncentered) similarity metric for both genes and arrays. Data were log$_{10}$-converted prior to cluster analysis.

**Gene Ontology (GO) Analyses**—GO annotations were retrieved using PIGOK (Protein Interrogation of Gene Ontology and KEGG databases) (24) and the IPI accession number of the identified protein entry.

**RESULTS AND DISCUSSION**

Numerous independent studies have shown that label-free approaches that use the inherent quantitative information in LC-MS/MS data are suitable for quantitative proteomics (17–19, 25). However, to be practical, this strategy requires informatics tools to automate the extraction of quantitative data for all the identified peptides in LC-MS experiments (18). We therefore wrote a script (which we named Pescal) to automate the generation of XICs and the calculation of their statistics (area under the curve and intensity at maximum peak height; see “Experimental Procedures” for more information). An important feature of Pescal is that it calculates peak areas and heights of generated XICs for peptide ions across samples even if these peptides are not always selected for LC/MS in DDA experiments due to the presence of peptides selected for MS/MS may vary across samples due to undersampling caused by the limited duty cycle of commercially available mass spectrometers.

Fig. 1 shows a scheme of the strategy we used to derive protein quantitative information. Note that XICs are generated from all the peptides identified from all data files to be compared irrespective of whether these peptides are present in all the samples. In cases where a peptide is not present in a sample, the returned intensity value (peak height or areas) is zero or very close to zero (background). Supplemental Tables 1 and 2 provide examples on the application of this analytical strategy. Pescal does not allow for manual correction of peak integration but permits calculation of peak intensities for thousands of peptide ions in a relatively short time (~5 peptide ions/s), making it a useful tool for comprehensive quantification of proteomes.

Although it is normally believed that chromatographic peak areas are best suited as the intensity readout for the quantitative analysis of LC-MS and HPLC data, peak heights are
sometimes preferred as the intensity value (26) especially when peaks are not well resolved as is often the case in the analysis of trace compounds by HPLC. Our initial experiments indicated that the quantitative data generated with our method was more precise when using peak heights rather than peak areas as the intensity readout (data not shown). This may be because of intrinsic integration errors associated with the calculation of peak areas of poorly resolved chromatographic peaks. This problem is analogous to that encountered in the analysis of trace compounds by HPLC (26), and it may be argued that quantification of peptides in complex mixtures by LC-MS presents a related analytical challenge. Furthermore chromatographic peak heights are more accurately calculated than peak areas when retention times shift across samples. Therefore, in the experiments described below, we used chromatographic peak heights as ion intensity readouts for correlation with protein abundance.

Validation of Chromatographic Peak Heights, as Obtained by Pescal, as the Intensity Readout for Relative Protein Quantification—We aimed at assessing the performance of this approach for protein quantification. To investigate whether the chosen retention time and mass accuracy windows were sufficiently narrow to ensure accurate calculations of protein amounts, we first performed experiments as in Ref. 19 in which serial dilutions of albumin (BSA) spiked in a constant amount of cell lysate from the murine WEHI-231 B cell line were separated by SDS-PAGE (Fig. 2A), and the proteins present in a gel section around 65 kDa were analyzed by LC-MS/MS and Pescal. Similar experiments were done on a constant amount of BSA spiked into a dilution series of WEHI-231 lysate before SDS-PAGE. Fig. 2B shows that, although proteins derived from WEHI-231 cells (and exogenously added trypsin) showed the same level of expression in replicate analyses, BSA signals correlated linearly with the amounts spiked in the WEHI-231 cell lysates (with the dynamic range being linear from 100 ng to at least 1000 ng on gel and 25 ng on column; this is close to the detection limit of the LC-MS/MS system used). Similarly the amounts of proteins in WEHI-231 cell lysates correlated linearly with the amounts of lysate loaded onto SDS-PAGE gels (Fig. 2C). In contrast, constant amounts of albumin and trypsin, added prior to and after electrophoretic separation, respectively, showed equal ion intensities in these samples (Fig. 2C).

Analysis of the variation indicated that this approach allows for protein quantification with reasonably good CV values (15–16% on average) with 70–80% of the data points showing CV values below 20% (Fig. 2, D and E). We have previously reported that by careful manual integration of peaks in XICs it is possible to quantify proteins with mean CV values of ~12% (19). We considered that the larger variation observed in our automated analysis (~16%) with Pescal (Fig. 2, D and E) is acceptable and that the loss of ~4 percentage points in precision may be the trade-off for being able to obtain quantity values for thousands of peptide ions in a time frame compatible with large scale analyses. Nevertheless this variation is still relatively low and adequate given that these data indicate that automated quantification of LC-MS/MS data using Pescal provides levels of precision close to those obtained using strategies
Validation of Chromatographic Peak Heights, as Obtained by Pescal, as the Intensity Readout for Estimation of Absolute Protein Amounts—Although the information provided by relative protein quantification is often sufficient to derive conclusions from certain biological experiments, the ability to obtain absolute quantification of proteins adds value to proteomics experiments. Thus absolute quantification is important to classify proteins within a sample according to abundance. This information may be particularly important for understanding the relative contribution of different isozymes to their biochemical pathway, a type of information that cannot be accessed by performing relative quantification experiments as in "classical" proteomics studies (28).
Targeted approaches for absolute quantification that use isotopically labeled peptides have been described (29–31), but this approach requires synthesis of an internal standard peptide for each of the proteins to be quantified. For large scale absolute quantification, it has been suggested that spectral counts (the number of peptides selected for MS/MS in DDA experiments) roughly correlate with protein abundance (21), providing an approximate estimation of absolute protein amounts. However, the accuracy of quantitative approaches based on spectral counts is very limited, and such approaches may only provide a semiquantitative estimation of absolute protein levels within a sample (17).

We tested the idea that large scale absolute quantification of proteins may be better achieved by adding up the ion intensities of all peptides derived from the same protein, thus combining the principle behind spectral counts with the added quantitative information stored in the ion intensities for each of these peptide ions. Testing this hypothesis was made possible because our approach makes it possible to obtain intensity values for all the peptides selected for MS/MS in LC-MS/MS experiments.

**Precision of absolute quantification based on addition of peptide ion intensities**

The intensities of peptides matching a particular protein were added up and expressed normalized to total protein in the gel section. The results of five replicate experiments analyzing the expression of one gel band containing a total cell lysate fraction are shown. Raw data are shown in Supplemental Fig. 2. *n*, number of peptides used for quantification per protein.

<table>
<thead>
<tr>
<th>GI No.</th>
<th>Protein Name</th>
<th>n</th>
<th>Amount of protein in gel section (ng/µg of total protein) in replicate experiment number</th>
<th>Mean</th>
<th>S.D.</th>
<th>CV%</th>
</tr>
</thead>
<tbody>
<tr>
<td>31543113</td>
<td>Lymphocyte cytosolic protein 1</td>
<td>29</td>
<td>7.6 10.1 13.1 13.9 14.4</td>
<td>11.8</td>
<td>2.9</td>
<td>24.5</td>
</tr>
<tr>
<td>984636</td>
<td>65-kDa macrophage protein</td>
<td>26</td>
<td>6.7 9.0 11.8 12.3 12.7</td>
<td>10.5</td>
<td>2.5</td>
<td>24.2</td>
</tr>
<tr>
<td>11066098</td>
<td>Transketolase</td>
<td>16</td>
<td>6.2 9.9 10.9 12.3 12.1</td>
<td>10.3</td>
<td>2.5</td>
<td>24.2</td>
</tr>
<tr>
<td>52783203</td>
<td>Carboxykinase</td>
<td>7</td>
<td>1.3 1.9 2.4 2.4 2.5</td>
<td>2.1</td>
<td>0.5</td>
<td>23.5</td>
</tr>
<tr>
<td>226021</td>
<td>Growth-regulated nuclear 68</td>
<td>6</td>
<td>1.2 1.4 1.8 1.9 1.8</td>
<td>1.6</td>
<td>0.3</td>
<td>19.3</td>
</tr>
<tr>
<td>56385</td>
<td>SWAP-70</td>
<td>17</td>
<td>1.1 1.0 1.0 3.9 10.1</td>
<td>3.4</td>
<td>3.9</td>
<td>114.0</td>
</tr>
<tr>
<td>13242328</td>
<td>Lymphocyte cytosolic protein 1</td>
<td>5</td>
<td>1.1 1.3 1.6 1.7 1.7</td>
<td>1.5</td>
<td>0.3</td>
<td>18.1</td>
</tr>
<tr>
<td>293689</td>
<td>65-kDa macrophage protein</td>
<td>6</td>
<td>1.1 1.6 1.7 2.0 2.1</td>
<td>1.7</td>
<td>0.4</td>
<td>25.0</td>
</tr>
<tr>
<td>16359229</td>
<td>Transketolase</td>
<td>8</td>
<td>0.9 1.2 1.4 1.5 1.5</td>
<td>1.3</td>
<td>0.3</td>
<td>19.8</td>
</tr>
<tr>
<td>16741161</td>
<td>Carboxykinase</td>
<td>6</td>
<td>0.6 0.8 0.8 1.0 1.1</td>
<td>0.8</td>
<td>0.2</td>
<td>23.4</td>
</tr>
<tr>
<td>19882225</td>
<td>Growth-regulated nuclear 68</td>
<td>4</td>
<td>0.5 0.7 0.9 1.0 1.0</td>
<td>0.8</td>
<td>0.2</td>
<td>24.4</td>
</tr>
<tr>
<td>2137619</td>
<td>Lymphocyte cytosolic protein 1</td>
<td>2</td>
<td>0.3 0.4 0.5 0.5 0.6</td>
<td>0.5</td>
<td>0.1</td>
<td>24.0</td>
</tr>
<tr>
<td>2323410</td>
<td>65-kDa macrophage protein</td>
<td>3</td>
<td>0.2 0.3 0.2 0.3 0.3</td>
<td>0.3</td>
<td>0.0</td>
<td>12.0</td>
</tr>
</tbody>
</table>

*Fig. 3. Validation of Pescal for approximate absolute quantification of proteins by LC-MS/MS. A, graph showing that total ion currents were directly proportional to the amount of cell lysate (in protein weight) loaded in gels. B, the proportional amounts of BSA (in weight) were plotted against the proportional amount of ion intensities. The result showed a strong correlation between these two parameters. C, the sum of ion intensities of endogenous proteins in B cell lysates (from the experiment in Fig. 1C) produced responses that were linear (left panel) when plotted against the amount of protein loaded in gels with correlation coefficients ($R^2$) approximating unity in all instances (right panel).*
experiments. As shown in Fig. 3A, the added intensities of all peptide ions of a gel section centered at 65 kDa were directly proportional to the amount of cell lysate loaded in gels. In line with this observation, increasing amounts of BSA, when normalized to the total amount of all proteins in the sample, produced normalized ion intensities with good linearity (Fig. 3B).

We also observed that the signals of added peptide ion intensities derived from proteins identified in the experiment shown in Fig. 2C were linear (average correlation coefficient, $R^2 = 0.99$) with respect to the amount of cell lysate analyzed (Fig. 3C).

We further assessed the precision of quantification by performing replicate experiments (summarized in Table I). The data, obtained by comparing five replicate experiments as in Fig. 2C, demonstrated good reproducibility of quantification.

Fig. 4. Strategy for the identification and quantification of proteins extracted from mouse primary tissues and qualitative analysis of results. A, strategy for the identification and quantification of proteins from brain, heart, kidney, and lung murine tissues using Mascot for protein identification and Pescal for their quantification. B, distribution of number of peptides (left) and statistical scores (right) of returned protein hits. C, molecular weight of proteins identified in the 11 gel fractions analyzed (plotted values represent means ± S.E. of the mean). D, total ion counts per gel section. E, F, and G, distribution of peptides and proteins identified per gel section. H, overview of results tabulating some other statistics of the data obtained from these experiments.
when total ion intensities for a given protein were normalized to the sum of total intensities of the LC-MS/MS run. Taken together, these data indicate that the sum of the peptide intensities for a given protein correlates with the proportional amount of this protein in the protein mixture, and therefore this information can be used to estimate protein amounts in absolute units with reasonable precision. Thus we propose that the proportion of ion signal for a particular protein relative to the total ion intensity of an LC-MS/MS experiment may be translated to the proportional weight of this protein in the total protein mixture.

Although we are aware that the precision of this absolute quantification approach may not be as high as that afforded by isotope dilution MS, it permits cataloguing hundreds of proteins within a sample according to abundance with acceptable confidence. This is thus another useful strategy for maximizing the amount of information obtained from LC-MS/MS data.

Quantitative Profile of the Most Abundant Proteins in Mouse Organs—Having found that our protocols can be used to quantify proteins in complex mixtures, we designed experiments aimed at profiling the most prominent mouse proteins.
Fig. 4A shows a scheme of the approach used for these experiments. Proteins from brain, heart, kidney, liver, and lung homogenates were separated by SDS-PAGE, and lanes were cut into 11 sections. Proteins in these sections were digested with trypsin and analyzed by LC-MS/MS. Analyses were performed in duplicate. Mascot was used for searches against the mouse IPI database, and returned peptide hits were used by Pescal to derive peak intensity values for each of the identified peptide sequences. We obtained good data for 1,487 protein entries as evidenced by Mascot scores and the number of unique peptides matched per protein (Fig. 4B). Indeed ~75% of protein identifications matched to more than two peptide sequences, and >80% of these had Mascot scores >50. Another indication of the quality of the data was obtained from the molecular weight distribution of the identified proteins (Fig. 4C), which generally agreed with the predicted values; the exception was in the proteins found in fraction 1 probably because of precipitation occurring at the interface between the stacking and separating SDS-PAGE gels. Some other statistics of the data obtained with these experiments are shown in Fig. 4, D–H. No obvious bias was observed in the number of peptides of proteins identified per fraction (Fig. 4, E, F, and G), and the total ion counts per fraction (Fig. 4D) was consistent with Coomassie staining (Fig. 4A). After deleting duplicates (i.e. proteins that were present in more than one organ), we obtained a protein list consisting of 942 entries that matched 8,842 peptides. Ion intensity values for these peptides were obtained for the five different organs analyzed, leading to a total number of 44,210 independent quantitative data points and the quantification of 4,710 proteins (942 proteins per organ).

Relative Quantification of Proteins in Mouse Organs—Our data allowed comparison of the expression levels of the major
proteins present in distinct mouse tissues. We clustered the data using approaches originally designed for the analysis of cDNA microarray data (Fig. 5), allowing easy visualization of the protein patterns of these five mouse organs. The list of all the identified proteins together with their levels of expression is provided as Supplemental Tables 4 and 5.

These quantitative data were first validated by comparison of the expression levels of proteins known to be enriched in specific organs. For example, Fig. 6A shows that synaptotagmin-1 and excitatory amino acid transporter 1, two proteins involved in neurotransmission (32, 33), were found to be more abundant in brain. Tissue-specific isoforms of fatty acid-binding proteins were enriched in the respective organs (the heart isoform in heart and the liver isoform in liver; Fig. 6A). Similarly and as expected, megalin (the multifunctional protein receptor responsible for reabsorbing polypeptides from the glomerular filtrate) and meprin (a glomerular structural protein) were expressed at high levels in the kidney. Also in line with expectations was the high expression of myoglobin in heart and alcohol dehydrogenase-1 in liver and kidney (Fig. 6A) consistent with the known expression pattern of this metabolic enzyme (34, 35). Fig. 6A also shows enriched expression of the lung isoform of carbonyl reductase and endothelial filamin A in the lung; the latter is consistent with the abundance of endothelial cells in this tissue. Other proteins were detected with similar levels of expression across different tissues. As an example, \( \beta \)-tubulin and actin-1 by Western blot using specific antibodies and by Pescal analysis of LC-MS/MS data. Fig. 6B shows that expression patterns of \( \beta \)-tubulin and
actin-1 were similar when analyzed by these two different methods, a finding that was confirmed by the high degree of correlation between the two sets of quantitative data (Fig. 6C). Further analysis of replicates indicated good reproducibility of the quantitative data (Supplemental Fig. 1 and Supplemental Table 3).

**Absolute Quantification of Mouse Organ Proteins** — As discussed above, our approach can also provide an estimation of the absolute abundance of proteins in complex mixtures. Thus we obtained estimates of absolute quantities of all identified proteins as a percentage of total ion intensities, which, as discussed above, we propose may be translated to percentage of total weight (or mg of protein/100 mg of total tissue protein). Fig. 7A shows that the quantitative readings spanned at least 3 orders of magnitude, whereas Fig. 7B indicates that metabolic enzymes (e.g. α-enolase) and structural proteins (e.g. actin-1) are among the most abundant proteins in these organs. Also highly present are proteins such as albumin due to the presence of blood in the organs. Nevertheless some of these abundant proteins are also known to have organ-specific roles. Thus, Na⁺/K⁺-ATPases subunits were found to represent a high percentage of the brain and kidney proteomes (Fig. 7B) when compared with those of heart, lung, and liver. This is consistent with the high requirement of neurons and renal epithelial cells for the electrical and concentration gradients that are generated by the Na⁺/K⁺-ATPase. These gradients are used by nerve cells for the transmission of information in the form of postsynaptic potentials and by renal epithelial cells for the reabsorption of solutes in the glomerular filtrate. It is also interesting to note that enzymes involved in the generation of ATP (ATP synthase α and β) constitute a larger percentage of heart and kidney proteomes than in other organs (Fig. 7B). This finding is consistent with the high energy requirements of these organs to maintain constant heart muscle contraction and for the large rate of vacuolar ATPase-dependent endocytosis taking place in renal epithelial cells (37).

**Gene Ontology Analysis Identifies the Most Prevalent Cellular Functions** — We next performed a bioinformatics analysis of our dataset to provide insight into the most common cellular functions in the different organs analyzed (Fig. 8). For this we used a program termed PIGOK (24) to query GO annotations on the IPI protein database (38). About 40 and 60% of all the identified proteins returned GO annotations on biological and molecular function process, respectively. We then compared the total number of proteins matching a particular GO annotation with the total absolute amounts of proteins (calculated by LC-MS and Pescal) within each GO description.
We noted that the number of proteins matching a particular GO description did not always correlate with the combined ion intensities (Fig. 8). As an example, Fig. 8A shows that the number of proteins matching structural, carrier, transporter, regulator, and motor ontologies was lower than their combined intensities. In other words, the number of genes matching a particular GO group may not always reflect the abundance of this activity or process in cells and argues for the use of absolute protein amounts for refined functional analysis of protein expression data.

Metabolism was the most enriched biological process in all tissues with about 16% of the total ion intensity corresponding to metabolic proteins (Fig. 8A). Other abundant biological processes in all tissues were transport (14%), cell death (9%), and development (7%) (Fig. 8A).

As for their molecular function, proteins with oxidoreductase, hydrolase, and transferase activities produced the largest combined intensity values in all tissues (Fig. 8A), indicating that these are the three most abundant biochemical functions in primary murine cells. Overall these prevalent molecular functions were about 2 orders of magnitude more abundant (as judged by their combined intensities) than proteins matching signal transduction and antioxidant ontologies and 1 order of magnitude more abundant than receptors, regulatory proteins, and proteins with kinase activity. These values may be an underestimate because only about 60% of all proteins identified had GO annotations. However, we do not expect the ratios to change significantly after all entries have been annotated.

The differences in molecular function and biological process (as determined by GO analysis) between the different organs were not pronounced. Fig. 8B shows that proteins involved in metabolic functions produced a similar proportion of added ion intensities in the different organs. Nevertheless we observed several interesting differences in the functional classification of the proteomes analyzed. For example, proteins with ligase and transferase ontologies produced more abundant ion signals in liver than in brain, heart, and lung (Fig. 8B). Ligase and transferase activities are involved in biosynthetic pathways, and their enrichment in the liver may reflect a high rate of anabolic metabolism, a known liver function, in this organ (39, 40).

Interestingly proteins with signal transduction and protein transport activity were 3–5-fold more abundant in brain than in the other organs, whereas other functions were diluted in this organ, including proteins with electron transport activity (needed for the generation of metabolic energy in mitochondria) whose added intensities were ~3.5-fold less abundant in brain than in heart. This finding is consistent with the ~3-fold higher number of mitochondria in rodent heart than in brain (41, 42).

**Comparison between Relative and Absolute Protein Quantification**—Our dataset also allowed us to directly compare the merits and utilities of relative and absolute methods for protein quantification. As an example, we assessed relative and absolute expression of the isoforms of the enolase enzyme in different organs. In agreement with published data (43), the expression of γ-enolase was found to be enriched in brain (Fig. 9A, top panel); expression of α-enolase and β-enolase was more equally distributed in the different tissues. However, absolute quantification (Fig. 9A, bottom panel) revealed that, although γ-enolase expression was more pronounced in brain relative to other organs, the most abundant enolase isozyme (in absolute terms) was the α isozyme in all organs analyzed, including the brain.

The expression patterns of the 17 distinct small GTPases that we identified provide another example of the extra information provided by absolute protein quantification. Relative quantification revealed that most Rab isoforms were ≈2 times more abundant in brain than in other organs (Fig. 9B, upper
Quantitative Profile of a Mouse Core Proteome

**panel, graphs 1–12**. In contrast, R-Ras expression seems to be diluted in brain and more abundant in the lung (Fig. 9B, graph 17). These data, however, did not allow us to assess which of the identified small GTPases are more abundant within a defined tissue.

Absolute quantification analysis revealed that the most abundant Rab isoforms in brain are Rab-1A, Rab-3A, and Rab-3D (Fig. 9B, bottom panel), which are approximately 1 order of magnitude more abundant in this tissue than other members of the Rab family of small GTPases such as Rab-11B and Rab-15. It is also interesting to note that although relative quantification suggests that R-Ras may be more abundant in lung than in any other organ (Fig. 9B, upper panel, graph 17), assessment of absolute amounts indicates that R-Ras is expressed at low amounts in all tissues analyzed, including the lung (Fig. 9B, bottom panel).

**Conclusion**—Using a first generation Q-TOF mass spectrometer, the approach described here allowed the quantification of the ~1,000 most abundant proteins in different murine organs, thus demonstrating the power of this technique for quantitative proteomics. This approach can be used for profiling the proteomes of cell lines and primary tissues and comparing an unlimited number of samples is facilitated by the fact that it is not restricted to the number of available isotopic labels. These are two important requirements for phenotypic analysis of primary clinical samples and cells and tissues from model organisms. The bioinformatics tools and standardization procedures described here (in combination with new generation mass spectrometers) will provide us with the opportunity for maximizing the biochemical information that may be obtained from our gene targeting efforts (5, 44–46), which are ultimately aimed at understanding the mechanisms of signal transduction in physiologically relevant systems.

**Acknowledgments**—We thank M. D. Waterfield for allowing access to analytical instrumentation, M. Graupera for animal dissection, R. Jacobs for advice using Integra, and John Timms for feedback on the manuscript.

* Work in the laboratory of B. V. was supported by the Ludwig Institute for Cancer Research. This work was also supported by additional funds from the Association for International Cancer Research (to P. R. C. and B. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Available online** The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

**Present address:** Analytical Signalling Laboratory, Centre for Cell Signalling, Inst. of Cancer, Barts and the London Medical School, 3rd fl. John Vane Science Bldg., Charterhouse Square, London EC1M 6BQ, UK.

**To whom correspondence should be addressed:** Cell Signalling in Cancer Laboratory, Ludwig Inst. for Cancer Research, 91 Riding House St., London W1T 7BS, UK. Tel.: 44-020-7882-8264; E-mail: p.cutillas@qmul.ac.uk.

**REFERENCES**