Universal sample preparation method for proteome analysis

Jacek R Wiśniewski, Alexandre Zougman, Nagarjuna Nagaraj & Matthias Mann

We describe a method, filter-aided sample preparation (FASP), which combines the advantages of in-gel and in-solution digestion for mass spectrometry–based proteomics. We completely solubilized the proteome in sodium dodecyl sulfate, which we then exchanged by urea on a standard filtration device. Peptides eluted after digestion on the filter were pure, allowing single-run analyses of organelles and an unprecedented depth of proteome coverage.

There are two major strategies for converting proteins extracted from biological material to peptides suitable for mass spectrometry (MS)-based proteome analysis. The first involves solubilization of proteins with detergents, separation of proteins by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis and digestion of the gel-trapped proteins ('in-gel' digestion)[1]. The second method is detergent-free, comprising protein extraction with strong chaotropic reagents such as urea and thiourea, protein precipitation and digestion of proteins under denaturing conditions ('in-solution' digestion). This second approach is frequently followed by two-dimensional peptide separation, for example, in the 'MudPit' strategy[2]. Advantages of in-gel digestion include its robustness against impurities, which interfere with digestion, but the gel may prevent peptide recovery and the method cannot easily be automated. In-solution digestion is more readily automatable and minimizes sample handling, but the proteome may be incompletely solubilized, and digestion may be impeded by interfering substances.

SDS is the reagent of choice for total solubilization of cells and tissues, and is routinely used in biochemical studies. Unfortunately, detergents, even in small concentrations, can preclude enzymatic digestion and dominate mass spectra owing to their ready ionizability and their great abundance compared to individual peptides. Therefore, depletion of SDS is a prerequisite for efficient mass-spectrometric analysis in proteomics. Because in-solution removal of SDS has been thought to be impossible, various alternative approaches have been developed for analyzing membrane proteomes. Early attempts involved membrane solubilization with formic acid[2], organic solvents[3,4] or digestion of the protein chains protruding from the membrane bilayer of nonsolubilized membranes[5–8]. We had recently discovered that membrane proteins can be fully depleted from detergents by gel filtration in 8 M urea such that they can then be analyzed as efficiently as soluble proteins[9]. Using this observation as a starting point, we sought to develop a method that combines strong detergents for universal solubilization with a means to efficiently 'clean up' the proteome before digestion and obtain purified peptides after digestion while avoiding the disadvantages of the gel format.

We reasoned that a common ultrafiltration device could be used for detergent removal to enable subsequent proteome analysis. We describe a method, filter-aided sample preparation (FASP), in which the sample is solubilized in 4% SDS, then retained and concentrated into microliter volumes in an ultrafiltration device (Online Methods). The filter unit then acts as a 'proteomic reactor' for detergent removal, buffer exchange, chemical modification and protein digestion. The four critical steps of the FASP method are: (i) depletion of detrimental low-molecular-weight components in urea-containing buffer, (ii) carboamidomethylation of thiols, (iii) digestion of proteins and (iv) elution of peptides (**Fig. 1**). Notably, during peptide elution, the filter retains high-molecular-weight substances that would otherwise interfere with subsequent peptide separation.

As the key feature of the method is the ability of the filter membrane to retain high-molecular-weight substances (proteins and DNA) and to allow through low-molecular–weight substances (impurities and digested peptides), selecting a filter with the desired separation properties is essential. We tested filters with relative molecular mass (M_r) cut-offs of 3,000 (3k filter) and 10,000 (10k filter). Note that the manufacturer determined these cutoffs with folded rather than detergent-denatured proteins. We performed all MS analyses by electrospray liquid chromatography–tandem MS (LC-MS/MS) using a linear ion trap–orbitrap instrument (LTQ-Orbitrap) essentially as described previously[10,11] (Online Methods).

We first compared the distribution of molecular weights of the identified proteins using either a 3k or 10k filter. We observed no substantial differences in the number of proteins identified per molecular-weight interval down to the 5–10 kDa bin (**Supplementary Fig. 1a,c** online). Next, we compared the efficiency of peptide elution using either a 3k or 10k filter. The number of identified tryptic peptides with a molecular weight above about 1,500 Da was much reduced for the 3k filter compared to the 10k filter. The number of peptides larger than 1,500 Da decreased gradually with increasing size, and peptides with masses over 2,500 Da were almost completely retained by the 3k filter

Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, Martinsried, Germany. Correspondence should be addressed to J.R.W. (jwisniew@biochem.mpg.de) or M.M. (mmann@biochem.mpg.de).

RECEIVED 25 NOVEMBER 2008; ACCEPTED 11 MARCH 2009; PUBLISHED ONLINE 19 APRIL 2009; DOI:10.1038/NMETH.1322

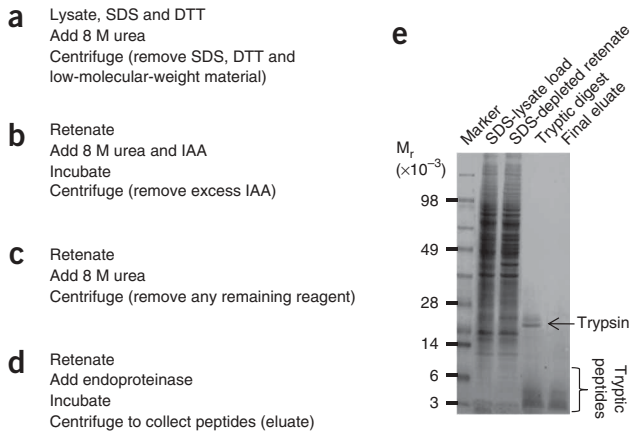

Figure 1 | Filter-aided sample preparation (FASP) for MS-based proteomic analysis. (**a**) Cell or tissue lysates can be prepared in the presence of high concentrations of detergent. Disulfide bridges are reduced with dithiothreitol (DTT). Detergent micelles and protein detergent complexes are dissociated in the presence of 8 M urea. The detergent, DTT and other low-molecular-weight components are removed by utrafiltration (Microcon units) facilitated by centrifugation. (**b**) Thiols are carboxyamidomethylated with iodoacetamide (IAA) and excess reagent is removed by ultrafiltration. (**c**) In repeated washes with 8 M urea any remaining detergent is depleted from the proteins. (**d**) The protein suspension is digested with endoproteinase, and the resulting peptides are collected as a filtrate. High-molecular-weight molecules including the endoprotease are retained on the filter. When nuclei or total cell lysates are processed in the units, DNA is retained on the filter. (**e**) SDS polyacrylamide gel electrophoresis analysis of total cell lysate, SDS-depleted and alkylated proteins, tryptic digest and eluted peptides. (Note that FASP does not involve any gel separation.)

(**Supplementary Fig. 1b**,**d**). As the 10k filter efficiently retained small proteins (5–10 kDa) and efficiently released peptides up to 5,000 Da (**Supplementary Fig. 1b**,**d**), we used it as the standard in the subsequent experiments.

To test the efficiency and range of applicability of the method, we processed and analyzed various amounts of bovine serum albumin (BSA) protein standard and total HeLa cell lysates. As judged by UV-light absorption and LC-MS/MS analysis of BSA peptides, we determined that FASP resulted in very high yield over at least three orders of magnitude of protein abundance (**Supplementary Fig. 2** online). Analysis of different numbers of HeLa cells, down to a few thousand cells, showed no substantial decrease in the number of identified peptides and proteins (**Supplementary Fig. 3** online).

In common with frequently used in-solution digestion protocols, FASP uses high concentrations of urea. In solution, a small fraction of urea decomposes to cyanic acid, which reacts with side chains of lysine and arginine and N-terminal amino groups to form carbamylated residues. As decomposition of urea is facilitated by high temperature, we performed all steps in FASP at room temperature (18–22 °C) and carried out the centrifugation steps at constant 20 °C. Under these conditions, less than 0.5% of identified peptides carried carbamylated arginine or lysine residues, a similar proportion as observed in our previous in-solution experiments (data not shown).

Next, we tested FASP on samples including mouse liver and brain tissues in addition to cultured cells. Preparation of tissue lysates was extremely simple, consisting only of tissue homogenization in the presence of SDS and subsequent application of an aliquot of this homogenate to the membrane reactor, taking less than 10 min. Notably, the presence of SDS efficiently inactivated detrimental enzymatic functions such as protease and phosphatase activity. In single-run analyses with 4-h gradients, we identified 1,800–2,200 proteins, with 99% confidence and at least two identified peptides per protein using the MaxQuant algorithms[12]. When we added proteins identified with one peptide, the number of identified proteins increased to 2,200–2,700 proteins (**Fig. 2a** and **Supplementary Table 1** online). In comparison, in our recent characterization of the liver proteome, using extensive cytosolic and membrane fractionation with analysis of 20 in-gel slices, we identified 2,210 proteins in total[13]. In the FASP datasets, 75–80% of fragmentation events resulted in the identification of the peptide in a database. Such high identification rates had been previously only observed for stable isotope labeling with amino acids in culture (SILAC) pairs[12], suggesting that the high purity of eluted FASP peptides minimized fragmentation events associated with chemical noise, which cannot lead to peptide identifications. Gene Ontology analysis revealed 42% (HeLa total cell lysate) and 52% (brain tissue) proteins matching to the membrane category (**Supplementary Table 1** and **Supplementary Fig. 4** online). This high percentage of membrane proteins indicated the absence of bias against hydrophobic proteins compared to soluble proteins. We also observed better sequence coverage for membrane proteins via the FASP preparation method than by the standard in-solution digestion method (**Supplementary Fig. 5** online).

We had previously reported the identification of 22,905 peptides and 3,979 proteins from HeLa cells by combining peptide isoelectric focusing in the 'OFFGEL' fractionator (Agilent) with 12 peptide fractions and conventional in-solution digestion[14]. In comparison, using FASP we identified 40,582 unique peptides corresponding to 7,093 proteins from HeLa cells (**Supplementary Table 1**). To our knowledge, this is the largest reported proteome in any single experiment. The measurements took only 2 d, showing that deep proteome coverage is possible within a reasonable analysis time. In comparison, comparable in-depth measurement of the embryonic stem cell proteome took more than three weeks of measurement time[15].

We next used Gene Ontology analysis to investigate whether the FASP-prepared proteome was biased for proteins from any compartments or protein classes. As a reference set for the expressed genome in HeLa cells, we used the Affymetrix GeneChip Human Genome U133 Plus 2.0 dataset, which detected 23,348 probes that have at least two positive absorbance calls in three experiments corresponding to 10,937 genes[16]. For all genes whose messages were detected on the chip, we determined the proportion of Gene Ontology cellular compartments as well as biological functions. We then compared them to the same categories in the FASP-based proteome measurements. None of the categories were considerably different, demonstrating that the FASP preparation method is universal in that it does not lead to preferential extraction of proteins from specific cellular compartments or with specific functions (**Fig. 2b–d**).

As expected, compared to the protein coverage of the previous HeLa cell experiment, low-abundance protein classes were represented more extensively using the FASP method. For example, the percentage of proteins that were Gene Ontology–annotated for transcription, signal transduction and receptor activity increased by 20–30%. This was paralleled by a corresponding decrease in the percentage of proteins annotated for metabolic and catalytic

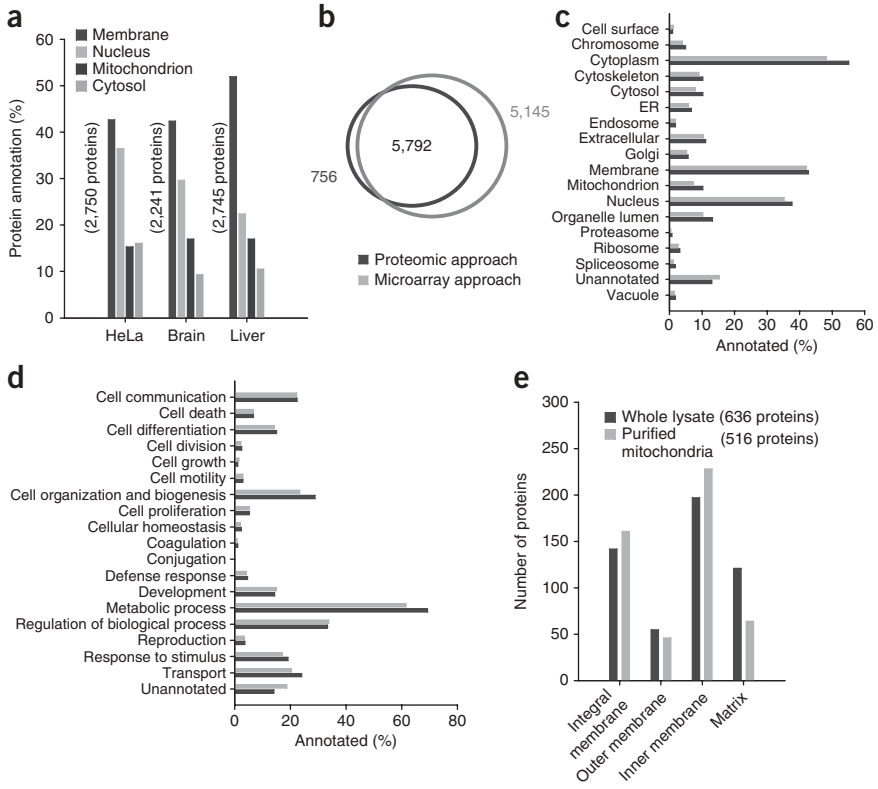

Figure 2 | FASP-based proteomic analysis of SDS lysates. (**a**) Single-run analysis of total lysates of HeLa cells, mouse brain and mouse liver processed in 10k filter units using two-step LysC and trypsin digestion. Bars show the percentage of proteins with the indicated Gene Ontology annotations. Total numbers of proteins identified per run are indicated in parentheses. (**b**) Venn diagram shows the overlap of genes identified by the FASP-based proteomic and microarray approaches. Note that a subset of identified genes cannot be matched to Affymetrix identifiers. (**c**,**d**) Comparison of Gene Ontology annotations for cell component (**c**) and biological processes (**d**) show that proteome and mRNA data are in concordance. (**e**) Single-run analysis of mouse liver mitochondria compared to mitochondrial proteins identified in 12 isoelectric focusing fractions of 'whole lysate' of HeLa cells. Bars show the number of proteins with the indicated Gene Ontology annotations. Total numbers of mitochondrial proteins identified in each experiment are indicated in parentheses.

function (**Supplementary Fig. 4**). We identified more than 90% of the proteins involved in the oxidative phosphorylation pathway, assembly of the ribosome, RNA polymerase and the polymerase II transcriptional machinery (**Supplementary Figs. 6–8** online). Considering that some of these proteins were cell type–specific and cell stage–specific and therefore were not expressed in all conditions, our data had very high coverage.

Before establishing the FASP protocol, we often separated proteomes into soluble and pellet fractions to achieve uniform representation of the proteome. These pellet fractions led to particularly poorly focused peptides in isoelectric focusing, with many peptides in three or more fractions, presumably because of contamination by nucleic acids, which are highly charged. With FASP, there is only a single proteome fraction, and focusing of all peptides was improved considerably (we detected 82% of peptides only in a single well, and 14% were focused into two wells; **Supplementary Fig. 9** online).

One major advantage of the FASP over the 'in-gel' and 'in-solution' approaches is its ability to accommodate a wide range of digestion conditions. We observed specific digestion for five different endoproteases (**Supplementary Table 2** and **Supplementary Fig. 10** online) and efficient digestion even without urea (**Supplementary Table 3** online).

Last, we investigated whether FASP would allow extensive proteomic analysis of isolated organelles without any fractionation. Indeed, with two single runs of a mouse liver fraction enriched in mitochondria, we identified 516 proteins annotated to this organelle (**Fig. 2e**) with a very high coverage of the core machinery of oxidative phosphorylation (**Supplementary Fig. 11** online). For this experiment, we used a filter unit with a relative molecular mass cut-off of 30,000 (30k filter) as the FASP reactor. This device shortened preparation time by a factor of three (2 h) and did not prevent identification of very small proteins.

The FASP method allowed processing of total SDS lysates of essentially any class of protein from biological material of any origin, thus solving the long-standing problem of efficient and unbiased solubilization of all cellular proteins irrespective of their subcellular location. In particular, FASP enables digestion of membrane proteins under conditions previously applied only to soluble proteins. With larger volume filter units, FASP also allowed handling of milligram amounts of protein.

The identification of more than 2,000 proteins in single runs using only 1–2 μg of material opens up interesting applications for proteomics, especially as the entire sample workflow is very streamlined. In organelle analysis, for example, this depth of analysis may already be sufficient: it is at least an order of magnitude greater in sensitivity and number of identified proteins than widely used proteome techniques such as two-dimensional gel analysis. For in-depth analysis of complex, mammalian proteomes FASP could be a crucial enabling sample preparation technology.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank S. Ren for bioinformatic support. This work was supported by the Max Planck Society for the Advancement of Science, High-throughput Epigenetic Regulatory Organization In Chromatin (HEROIC), Role of Ubiquitin and Ubiquitin-like Modifiers in Cellular Regulation (RUBICON) 6th Framework grants of the European Commission and the Munich Center for Integrated Protein Science (CIPSM).

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. *Anal. Chem.* **68**, 850–858 (1996).
2. Washburn, M.P., Wolters, D. & Yates, J.R., III *Nat. Biotechnol.* **19**, 242–247 (2001).

3. Blonder, J., Xiao, Z. & Veenstra, T.D. *Expert Rev. Proteomics* **3**, 483–496 (2006).
4. Blonder, J., Conrads, T.P. & Veenstra, T.D. *Expert Rev. Proteomics* **1**, 153–163 (2004).
5. Wu, C.C., MacCoss, M.J., Howell, K.E. & Yates, J.R., III *Nat. Biotechnol.* **21**, 532–538 (2003).
6. Olsen, J.V. *et al. Mol. Cell. Proteomics* **3**, 82–92 (2004).
7. Nielsen, P.A. *et al. Mol. Cell. Proteomics* **4**, 402–408 (2005).
8. Le Bihan, T. *et al. J. Proteome Res.* **5**, 2701–2710 (2006).
9. Nagaraj, N., Lu, A., Mann, M. & Wisniewski, J.R. *J. Proteome Res.* **7**, 5028–5032 (2008).
10. Olsen, J.V. *et al. Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
11. Rappsilber, J., Ishihama, Y. & Mann, M. *Anal. Chem.* **75**, 663–670 (2003).
12. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
13. Shi, R. *et al. J. Proteome Res.* **6**, 2963–2972 (2007).
14. Hubner, N.C., Ren, S. & Mann, M. *Proteomics* **8**, 4862–4872 (2008).
15. Graumann, J. *et al. Mol. Cell. Proteomics* **7**, 672–683 (2008).
16. Mense, S.M. *et al. Physiol. Genomics* **25**, 435–449 (2006).

ONLINE METHODS

Tissue and cell solubilization. We homogenized 50-mg pieces of frozen mouse liver or brain in 0.40 ml of 0.1 M Tris-HCl, pH 7.6 using an Ultra Turbax blender (IKA) at maximum speed (approximately 25,000 r.p.m.) at 4 °C for 30 s. Then, 50 µl aliquots of 20% SDS and 1 M DTT were added to the homogenate and the mixture was incubated for 3 min at 95 °C. Frozen aliquots of 5×10^7 HeLa cells were lysed in 0.5 ml of 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at room temperature and briefly sonicated to reduce viscosity of the lysate. BSA was denatured and reduced in 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at 95 °C for 3 min. Membrane and cytosolic fractions were prepared from HeLa cells and mouse brains as described previously[9].

Filter-aided sample preparation (FASP). Aliquots of lysates corresponding to 1 mg wet tissue (0.1 mg protein) or 2×10^5 HeLa cells (0.13 mg of protein) were mixed with 200 µl of 8 M urea in Microcon devices YM-10 or YM-3 (Millipore). The device was centrifuged at 14,000*g* at 20 °C for 40 min. All following centrifugation steps were performed applying the same conditions allowing maximal concentration. The concentrate was diluted with 200 µl of 8 M urea in 0.1 M Tris-HCl, pH 8.5 and the device was centrifuged. Subsequently, 100 µl of 0.05 M iodoacetamide in 8 M urea in 0.1 M Tris-HCl, pH 8.5 were added to the concentrate followed by centrifugation. The resulting concentrate was diluted with 100 µl 8 M urea in 0.1 M Tris-HCl, pH 7.9 and concentrated again. This step was repeated 2 times, and the concentrate was subjected to proteolytic digestion (**Supplementary Table 4** online) The digests were collected by centrifugation, and the filter device was rinsed with 50 µl 0.5 M NaCl and centrifuged. Detailed instructions for performing FASP are described in the **Supplementary Protocol** online. The combined filtrates were desalted on MILI-SPE Extraction disk cartridge (C_{18}-SD); 7 mm per 3 ml (Millipore).

The peptide content was estimated by UV light spectral density at 280 nm using an extinctions coefficient of 1.1 of 0.1% (g l^{-1}) solution that was calculated on the basis of the frequency of tryptophan and tyrosine (the main UV light–absorbing amino acids at 280 nm) in vertebrate proteins[17].

Isoelectric focusing of peptides. We separated 0.05 mg of peptides into 12 fractions on the 3100 OFFGEL Fractionator (Agilent Technologies) as described previously. The immobilized pH gradient strips (IPG strips) from GE Healthcare (Immobiline DryStrip pH 3–10, 13 cm) were rehydrated with 20 µl per well of isoelectric focusing buffer containing 5% glycerol and 50-fold diluted IPG buffer pH 3–10 (GE Healthcare) for 20 min. Peptides were dissolved in 1.68 ml of the isoelectric focusing buffer and 0.14 ml of the solution were loaded into to each well. Mineral oil was added to the ends to prevent the drying of the filter wicks wetted with the buffer. Focusing was performed at 20 °C with maximum values of 4,500 V and 200 mW. The limiting maximum current was set to 50 µA. Focusing was carried out for a target of 20 kVh. The focused peptides were acidified by adding 20 µl of acidic mixture (0.5% acetic acid, 1% TFA and 2% acetonitrile) before desalting and LC-MS/MS analysis.

Preparation and FASP of mitochondria. Frozen mouse liver was homogenized in a motor-driven glass-Teflon Potter-Elvehjem homogenizer at a 1:10 ratio of tissue to homogenization buffer (0.3 M sucrose, 10 mM MOPS-NaOH, 1 mM EDTA). The cell debris and nuclei were removed by centrifugation at 1,000*g* for 10 min. Then, the supernatant was collected and centrifuged at 16,000*g* for 15 min. The mitochondrial pellet was washed once by resuspending in the homogenization buffer and pelleting at 16,000*g* for 15 min. Mitochondrial pellet was lysed in 0.5 ml of 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at room temperature. The lysates were processed with the FASP method as described above but using 30k filtration units (Microcon; Millipore).

Mass spectrometric analysis. The digests were purified and stored in C_{18} StageTips as described[11]. Usually up to 10 µg peptide mixture was loaded on a StageTip containing two membrane plugs. Approximately a half of the sample was applied to the high-performance liquid chromatography column in each experiment. Peptide mixtures were analyzed by online capillary LC-MS/MS. The LC-MS/MS setup was similar to that described before[18]. Briefly, samples were separated on an in-house made 15 cm reversed-phase capillary emitter column (inner diameter 75 µm, 3 µm ReproSil-Pur C18-AQ medium; Dr. Maisch GmbH) using 240 min (cell and tissue lysates) or 60 min (BSA standard) gradients and analyzed using the LTQ-Orbitrap instrument (Thermo Fisher Scientific). Survey MS scans were acquired in the orbitrap with 60,000 resolution. For accurate mass measurements, the lock-mass option was used[10]. Up to 10 most intense ions in each full MS scan were fragmented and analyzed in the LTQ.

Peak list generation, database searching and validation. Raw MS files were processed with MaxQuant, an in-house developed software suite[12]. Peak list files were searched against decoy International Protein Index mouse database version 3.46 containing both forward and reverse protein sequences by the MASCOT search engine[19]. Initial parent and fragment ion maximum mass deviation[20] were set to 7 p.p.m. and 0.5 Da, respectively. The search included variable modifications of oxidation of methionine and protein N-terminal acetylation. Peptides with at least six amino acids were considered for identification. The false discovery rate for both peptides and proteins were set at 0.01. All peptides and proteins identified in this study are listed with posterior error probability values in **Supplementary Data 1–7** online.

Bioinformatics analysis. Gene ontology analysis of the identified proteins was performed using the Protein Center platform (Proxeon Biosystems).

17. Zhuang, Y., Ma, F., Li-Ling, J., Xu, X. & Li, Y. *Mol. Biol. Evol.* **20**, 1978–1985 (2003).
18. Olsen, J.V. & Mann, M. *Proc. Natl. Acad. Sci. USA* **101**, 13417–13422 (2004).
19. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. *Electrophoresis* **20**, 3551–3567 (1999).
20. Zubarev, R. & Mann, M. *Mol. Cell. Proteomics* **6**, 377–381 (2007).