# A protease for 'middle-down' proteomics

Cong Wu[1,2], John C Tran[3], Leonid Zamdborg[2], Kenneth R Durbin[4], Mingxi Li[1], Dorothy R Ahlf[5], Bryan P Early[3], Paul M Thomas[3,4,6], Jonathan V Sweedler[2] & Neil L Kelleher[3–7]

**We developed a method for restricted enzymatic proteolysis using the outer membrane protease T (OmpT) to produce large peptides (>6.3 kDa on average) for mass spectrometry–based proteomics. Using this approach to analyze prefractionated high-mass HeLa proteins, we identified 3,697 unique peptides from 1,038 proteins. We demonstrated the ability of large OmpT peptides to differentiate closely related protein isoforms and to enable the detection of many post-translational modifications.**



The 'bottom-up' and 'top-down' approaches represent two strategies for proteomic studies using mass spectrometry. Bottom-up proteomics relies on enzymatic protein digestions before on-line liquid chromatography–coupled tandem mass spectrometry analysis (LC-MS/MS)[1]. Top-down proteomics omits proteolysis and focuses on complete characterization of intact proteins and their post-translational modifications (PTMs)[2]. Although both approaches continue to mature, they each have limitations[3]. The tryptic peptides used in the bottom-up approach are the primary unit of measurement, but their relatively small size (typically ~8–25 residues long) leads to problems such as sample complexity, difficulties in assigning peptides to specific gene products rather than protein groups[4], and loss of single and combinatorial PTM information. The top-down approach handles these issues by characterizing intact proteins, but its success declines in the high-mass region. Therefore, a hybrid approach based on 2–20 kDa peptides could unite positive aspects of both bottom-up and top-down proteomics.

We previously proposed a generic approach to 'middle-down' proteomics for interrogating high-mass proteomes, with two essential features: a size-dependent protein fractionation technique and a robust but restricted proteolysis method[5] (**Fig. 1a**). A continuous tube-gel electrophoresis technique can now provide the size-dependent fractionation of a complex proteome[6]. Previous efforts to explore restricted proteolysis options included using enzyme alternatives to trypsin (such as Lys-C[7] and Lys-N[8]) and chemical methods (such as microwave-assisted acid hydrolysis[9]). These methods, however, produced peptides only marginally longer than tryptic peptides in large-scale proteomic studies.

Here we present the protease OmpT to achieve a robust, yet restricted, proteolysis of a complex proteome (with its benefits described in the **Supplementary Note**). OmpT is known to cleave between two consecutive basic amino acid residues (Lys/Arg-Lys/Arg) and is reported to have favorable kinetics, with $k_{cat}/K_m$ in the $10^4$–$10^8$ $s^{-1}$ $M^{-1}$ range. Derived from the *Escherichia coli* K12 outer membrane, OmpT belongs to the novel omptin protease family[10]. In this study, we developed OmpT into an efficient reagent to generate >2-kDa peptides for middle-down proteomics.

We first performed *in silico* digestions of the human proteome using various enzymatic or chemical approaches to create histograms of their predicted peptide masses (**Supplementary Fig. 1**). Most traditional enzymatic approaches generated predominantly small peptides (<2 kDa). OmpT, which cleaves between less common dibasic sites, produced a distribution with a greater number of peptides >3 kDa.

We overexpressed and refolded OmpT to obtain active enzyme (see Online Methods), and then we optimized its digestion of four standard protein substrates. Optimal conditions for OmpT digestion were at pH 6.0 in 2–3 M urea at 22 °C (**Supplementary Fig. 2**). We used urea to reduce the higher-order structure present in large protein substrates—an important step for the cleavage efficiency of OmpT[11]. Characterization of digestion products from the 36-kDa glyceraldehyde-3-phosphate dehydrogenase (GAPDH) standard is shown as an example (**Fig. 1b**,**c** and **Supplementary Fig. 3**). In addition to the predicted dibasic cleavages, we also observed a Lys-Ala cleavage, corroborating previous reports that OmpT can still cleave with aliphatic amino acids in its P1′ position[12], especially under strongly denaturing conditions[13]. Although the GAPDH sequence contains a Lys-Lys site (**Supplementary Fig. 2g**), the cleaved product at this site (peptide 5 in **Fig. 1c**) was barely observable upon LC-MS/MS analysis (data not shown). This is likely because the flanking amino acids are two aspartic acids whose negative charges may prevent the binding of the nearby Lys-Lys site to the negatively charged OmpT active site[14]. We also characterized three other proteins digested with OmpT in detail (**Supplementary Fig. 2b**,**d**–**f**,**h**).

We established an OmpT-based middle-down platform to analyze complex mixtures pre-sorted by protein size (**Fig. 1a**). After integrating the data from the middle-down workflow that was applied to ~20–100-kDa proteins fractionated from the HeLa cell proteome, we identified 3,697 unique peptides (average size: 6.3 kDa) from

[1]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [2]The Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. [3]Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, USA. [4]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA. [5]Department of Chemistry, Northwestern University, Evanston, Illinois, USA. [6]Chemistry of Life Processes Institute, Northwestern University, Evanston, Illinois, USA. [7]Robert H. Lurie Comprehensive Cancer Center in the Feinberg School of Medicine, Northwestern University, Evanston, Illinois, USA. Correspondence should be addressed to N.L.K. (n-kelleher@northwestern.edu).

RECEIVED 5 JANUARY; ACCEPTED 10 MAY; PUBLISHED ONLINE 17 JUNE 2012; DOI:10.1038/NMETH.2074

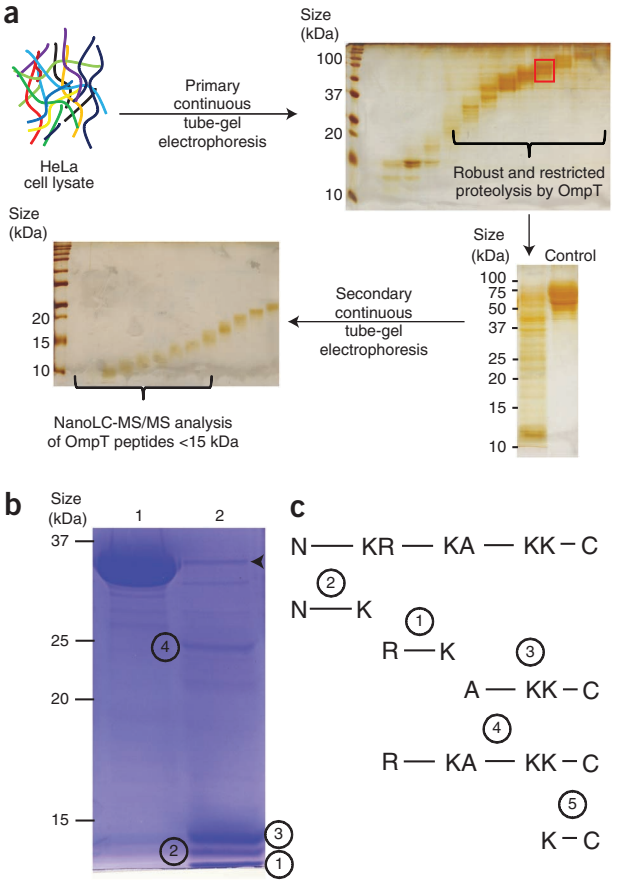


**Figure 1** | OmpT-based platform for middle-down proteomics and characterization of OmpT peptides from digestion of a standard protein. (**a**) The middle-down workflow was illustrated on proteins from a HeLa cell lysate sorted into narrow size ranges by molecular weight–based prefractionation (silver-stained gel, top right). A representative OmpT digestion of a fraction containing 50- to 75-kDa proteins (highlighted in the red box) was visualized by silver staining (left lane, bottom right) along with the control sample with no digestion (right lane). The digested samples were separated further, and fractions smaller than ~15 kDa were subjected to nanoLC-MS/MS analysis. (**b**) Peptide products from digestion of glyceraldehyde-3-phosphate dehydrogenase (GAPDH, 36 kDa) by OmpT were visualized on a Coomassie-stained SDS-PAGE gel. Lane 1, GAPDH incubated without OmpT. Lane 2, GAPDH after OmpT digestion. Major peptide products are numbered 1–4. Arrowhead indicates the intact OmpT enzyme. (**c**) Alignment of OmpT peptides identified by nanoLC-MS/MS, with the original GAPDH sequence on top. Peptide cleavage sites are illustrated; N and C represent the protein N and C termini.

1,038 unique proteins (26% average sequence coverage) at an estimated 1% false discovery rate (FDR)[2] (**Supplementary Table 1**). Two database search modes were used: biomarker and absolute mass (explained further below and in Online Methods). Both the forward and decoy databases for biomarker and absolute mass searches are available online using ProSightPTM 2.0 (http://prosightptm2.northwestern.edu/). ProSightPC users can also download these databases at ftp://prosightftp:gsX1gON@prosightpc.northwestern.edu/ and run both search modes locally. Results from an individual LC-MS/MS analysis of fractionated OmpT peptides from the middle-down workflow are provided as an example (**Supplementary Fig. 4**). We also performed a negative-control treatment of substrate proteins in the absence of OmpT, which showed no sample autodegradation (data not shown).

Proteotypic OmpT peptides can allow differentiation of specific protein isoforms. Detailed sequence alignments between protein isoforms revealed high sequence identity, whereas OmpT peptides, owing to their desirably large size, covered unique regions where isoform sequences differed (**Fig. 2a** and **Supplementary Fig. 5a**,**b**). Long peptides can also prove beneficial for detection and identification of modified peptides. In this study, ~25% of OmpT peptides were identified with PTMs (using annotated modifications from the UniProt database[2]), and several examples of multiply modified peptides were found (**Fig. 2b** and **Supplementary Fig. 5c**–**e**). An additional 8% of unique peptides with unexpected mass discrepancies were confidently identified in error-tolerant searching. Together, these data imply that the OmpT-based workflow can provide isoform-specific assignments, characterization of modified peptides and combinatorial PTM information complementary to traditional protease-based proteomic approaches.

The OmpT peptide size distribution was plotted in comparison with tryptic peptides (**Fig. 2c**). Because we analyzed only fractions below ~15 kDa, the average size of peptides identified here was 6.3 kDa, but OmpT peptides above 15 kDa were readily visible on gels (**Fig. 1a** and **Supplementary Fig. 4c**). We compared the performance between collision-induced dissociation (CID) and electron-transfer dissociation (ETD) using OmpT peptides (**Supplementary Fig. 6** and **Supplementary Table 2**). The low degree of overlap between the methods indicates that both CID and ETD can serve as highly complementary fragmentation approaches to identify and characterize OmpT peptides.

Although the substrate specificity of OmpT has been extensively studied, previous model substrates were mostly short peptides and unstructured protein linker regions[12,13,15]. This study helped to improve our understanding of OmpT's sequence preference under denaturing conditions (3 M urea) in which whole proteins were the substrates. We searched the entire data set in 'biomarker' mode against an intact protein database. A biomarker search assumes no specific proteolytic cleavage; rather, it queries every possible subsequence in the database within tolerance from an observed peptide mass. Confident biomarker peptide hits were then used to extract the P4 through P4 recognition sites of OmpT for the generation of an unbiased consensus sequence. From these data, we generated an iceLogo visualization that normalizes observed amino acid frequencies at each site to a reference set of proteomic amino acid frequencies (**Fig. 2d** and **Supplementary Fig. 7a**,**b**). For comparison, we also made a WebLogo, which illustrates amino acid frequencies at each site solely based on the input sequences without normalization (**Supplementary Fig. 7c**).

As shown in iceLogo and WebLogo representations, the P1 site was restricted almost exclusively to lysine and arginine, whereas the P1′ site was more permissive, allowing predominantly lysine and arginine but also alanine and serine. The relative promiscuity of OmpT at the P1′ position may be attributed to the location of the P1′-substrate binding site near the loops on top of the beta-barrel[14], which could have increased flexibility under denaturing conditions. Because of OmpT's broader specificity at the P1′ site, we defined the 'major cleavage sites' as Lys/Arg-Lys/Arg/Ala/Ser and performed another *in silico* digestion of human proteome at all these major sites assuming 0 and 2 missed cleavages (**Supplementary Fig. 1**). The resultant peptide size distributions strongly resembled the distributions assuming only Lys/Arg-Lys/Arg cleavages.

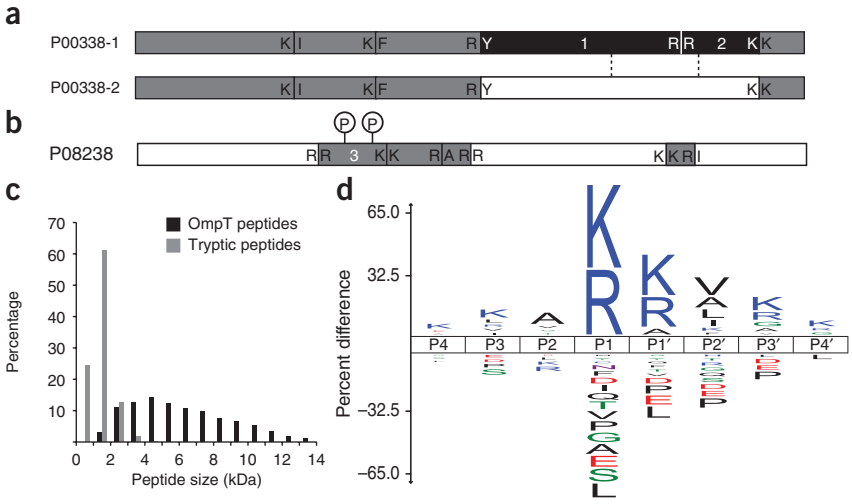


**Figure 2** | Proteotypic OmpT peptides, peptide size distribution and iceLogo of OmpT recognition site. (**a**) Peptides 1 and 2 (10.8 kDa and 5.4 kDa, respectively) cover a proteotypic sequence region of 37-kDa L-lactate dehydrogenase A chain isoform 1 (UniProt accession code P00338-1, 87% identity to isoform 2). Cleavage sites for OmpT peptides are shown. The schematic isoform alignment (detailed sequence alignment in **Supplementary Fig. 5a**) marks the region where the two isoform sequences differ between dashed lines. Peptides covering the distinct part of a certain isoform are shaded in black; peptides covering the common regions of all isoforms are in gray. (**b**) The 84-kDa heat shock protein Hsp90β (UniProt accession code P08238) identified by peptides in gray; phosphorylation sites (P) in peptide 3 (8.9 kDa) are indicated (survey spectrum of the singly and doubly modified species in **Supplementary Fig. 5c**). (**c**) Mass distribution of identified OmpT peptides (below ~15 kDa) in comparison with that of tryptic peptides[16]. (**d**) IceLogo of OmpT recognition sequences from P4 through P4′ sites. OmpT cleaves between P1 and P1′. The *y* axis displays the percent difference of amino acid frequencies between the experimental set and the reference set at each position.

In addition to selectivities at the P1 and P1′ sites, the P2′ site also had a slight preference for aliphatic amino acids. Overall, OmpT favored positively charged residues across its recognition sites (with the exception of P2) and resisted negatively charged and proline residues. Selectivities outside P1 and P1′ have been previously reported[12,15] and might explain the average number of observed missed cleavages (0.99 ± 1.29) at the major sites. In spite of these preferences, OmpT is still a stringent protease with well-defined substrate specificities, which will be better understood with future experimentation and data mining.

The stable beta-barrel structure of this membrane endopeptidase endows it with a remarkable resistance to both denaturants and surfactants, allowing extensive denaturation of large protein substrates under strongly solubilizing conditions for robust proteolysis. The analysis of widely distributed OmpT peptides across a broad mass range will necessitate adjustments in separation protocols and LC-MS/MS methods accordingly. New and next-generation instruments will further increase the routine size range accessible for sequencing, and many proteomics search engines will require modification to identify these larger peptides. With a demonstrated capacity for robust and restricted proteolysis, OmpT is an attractive option for mass spectrometry–based interrogation of protein primary structure.

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Supplementary information is available in the online version of the paper.*


### ACKNOWLEDGMENTS

We sincerely thank current and former members of the Kelleher and Sweedler research groups, especially B. Evans, P. Chu, P. Compton, A. Catherman, S. Sweet, I. Ntai, J. Lee, A. Vellaichamy and K. Catherman, for technical support and insightful suggestions, P. Yau and B. Imai from the Protein Sciences Facility at the University of Illinois for the synthesis of the fluorogenic substrate and M. Burke for the access to a fluorimeter. The project was supported by the US National Institutes of Health through awards R01 GM067193, P30 DA018310 and F30 DA026672, by the National Science Foundation through award DMS 0800631 and by the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust.


### AUTHOR CONTRIBUTIONS

C.W. designed research, conducted experiments, analyzed data and wrote the paper; J.C.T., L.Z., K.R.D. and B.P.E. analyzed data; M.L., K.R.D. and D.R.A. conducted cell culture; P.M.T. analyzed data and wrote the paper; J.V.S. interpreted data and wrote the paper; N.L.K. designed research, interpreted data and wrote the paper.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

**Published online at http://www.nature.com/doifinder/10.1038/nmeth.2074. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.**


1. de Godoy, L.M. *et al. Nature* **455**, 1251–1254 (2008).
2. Tran, J.C. *et al. Nature* **480**, 254–258 (2011).
3. Chait, B.T. *Science* **314**, 65–66 (2006).
4. Nesvizhskii, A.I. & Aebersold, R. *Mol. Cell. Proteomics* **4**, 1419–1440 (2005).
5. Forbes, A.J., Mazur, M.T., Patel, H.M., Walsh, C.T. & Kelleher, N.L. *Proteomics* **1**, 927–933 (2001).
6. Tran, J.C. & Doucette, A.A. *Anal. Chem.* **80**, 1568–1573 (2008).
7. Wu, S.L., Kim, J., Hancock, W.S. & Karger, B. *J. Proteome Res.* **4**, 1155–1170 (2005).
8. Taouatas, N., Drugan, M.M., Heck, A.J. & Mohammed, S. *Nat. Methods* **5**, 405–407 (2008).
9. Cannon, J. *et al. J. Proteome Res.* **9**, 3886–3890 (2010).
10. Kramer, R.A., Zandwijken, D., Egmond, M.R. & Dekker, N. *Eur. J. Biochem.* **267**, 885–893 (2000).
11. White, C.B., Chen, Q., Kenyon, G.L. & Babbitt, P.C. *J. Biol. Chem.* **270**, 12990–12994 (1995).
12. Dekker, N., Cox, R.C., Kramer, R.A. & Egmond, M.R. *Biochemistry* **40**, 1694–1701 (2001).
13. Okuno, K. *et al. Biosci. Biotechnol. Biochem.* **66**, 127–134 (2002).
14. Vandeputte-Rutten, L. *et al. EMBO J.* **20**, 5033–5039 (2001).
15. McCarter, J.D. *et al. J. Bacteriol.* **186**, 5919–5925 (2004).
16. Olsen, J.V. *et al. Mol. Cell. Proteomics* **8**, 2759–2769 (2009).

## ONLINE METHODS

**Reagents.** DNA restriction enzymes were purchased from Invitrogen and T4 DNA ligase from New England Biolabs. The pET28a vector and *E. coli* BL21(DE3) cells were obtained from EMD Biosciences. The SP-Sepharose medium and the K16/20 cation exchange column were bought from GE Healthcare Life Sciences. Isopropyl-β-D-thiogalactopyranoside (IPTG) was from Roche; all other chemicals were purchased from either Thermo Fisher Scientific or Sigma-Aldrich unless otherwise noted. The fluorogenic substrate Abz-Ala-Arg-Arg-Ala-Tyr($NO_2$)-$NH_2$ (Abz, *o*-aminobenzoyl; Tyr($NO_2$), 3-nitrotyrosine) was synthesized by the Protein Sciences Facility at the University of Illinois[10].

**Cloning of the OmpT gene and construction of the expression plasmid.** All PCR used Phusion Hot Start Polymerase (Finnzymes) and PCR-grade dNTPs (Invitrogen). PCR products and restriction-digested DNA were purified with the QIAquick gel extraction and PCR cleanup kits (Qiagen). The OmpT gene was amplified from the genomic DNA of *E. coli* K12 DH5α. The primer sequences used for cloning OmpT were 5′-ATGCGGGCG AAACTTCTGGGAATAG-3′ (forward) and 5′-TTAAAATGTGT ACTTAAGACCAGCAGTAGTG-3′ (reverse) from IDT. After the OmpT gene was cloned, another pair of primers containing restriction sites was used to amplify the gene without the N-terminal signal peptide with the sequences 5′-ATTAATCCATGGCTTC TCGAGACTTTATCGTTTA-3′ and 5′-ACTCGGGAATTCTT AAAAGTGTACTTAAGACCAG-3′. The amplified OmpT gene contains an NcoI restriction site at the 5′ end and an EcoRI site at the 3′ end (underlined). Both the pET28a vector and OmpT were doubly digested with NcoI and EcoRI (Invitrogen) and ligated to produce the new vector pNK1009, which was used to transform *E. coli* BL21(DE3) for protein expression after sequence confirmation by the University of Illinois Core DNA Sequencing Facility.

**Protease expression and purification.** OmpT was expressed in inclusion bodies in BL21(DE3) as previously described with some modifications[10]. Briefly, BL21(DE3) cells containing pNK1009 were grown overnight in 5 ml SOC medium (20 g Bacto-Tryptone, 5 g Bacto Yeast Extract, 0.5 g NaCl, 2.5 ml of 1 M KCl, 20 ml of 1 M glucose in 1 l $H_2O$) with 50 mg $l^{-1}$ kanamycin at 37 °C. The 5-ml starter culture was inoculated into 1 l SOC medium with 50 mg $l^{-1}$ kanamycin and grown to an $OD_{600}$ between 1.0 and 1.5. The expression of OmpT inclusion bodies was induced by the addition of 1 M IPTG to a final concentration of 0.4 mM, followed by further incubation at 37 °C for 6–9 h.

For OmpT purification, inclusion bodies were first isolated from the cell pellet as described before with some modifications[17]. Briefly, the cell pellet from a 1 l culture was resuspended in 12 ml lysis buffer (50 mM Tris-HCl, 40 mM EDTA, pH 8.0) and incubated with 3 mg lysozyme on ice for 30 min, and another 12 ml prechilled lysis buffer was added quickly to introduce osmotic shock, after which the mixture was incubated on ice for another 30 min. The lysate was sonicated at 25 W with a Sonic Dismembrator (Model 100, Fisher Scientific) every other minute for five cycles until the lysate was no longer viscous. Inclusion bodies were collected by centrifugation at 4,500*g* for 30 min. The pellet was washed once with 30 ml of wash buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and extracted with 4 ml of dissolving buffer (8 M urea, 50 mM glycine, pH 8.3) on ice for 30 min. To this solution, 16 ml of prechilled 31.25 mM *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate ($DodMe_2NPrSO_3$) was added to initiate OmpT refolding. Then the pH of the refolding mixture was adjusted to 4.0 using 10% acetic acid. The solution was centrifuged at 20,450*g*, and the supernatant was then filtered and loaded onto a 10-ml Fast Flow SP-Sepharose column (16 mm in diameter, 5 cm in length) equilibrated with buffer A (10 mM $DodMe_2NPrSO_3$, 20 mM sodium acetate, pH 4.0). The column was washed with 5 column volumes of buffer A, and proteins were eluted off with a linear gradient of NaCl to 1 M in 300 ml of buffer A. After cation exchange, OmpT was activated with lipopolysaccharide (LPS)[10,14] and dialyzed against enzymatic buffer to remove high-concentration salt; after this, LPS-bound OmpT was found in two forms due to a single self-degradation site (between between Lys217 and Arg218)[10], but >80% of the enzyme was still in its intact form (**Supplementary Fig. 2a**, lane E2). Based on SDS-PAGE analysis, fractions containing OmpT were pooled, divided into aliquots and frozen at −80 °C for storage after the OmpT activity was confirmed using the synthetic fluorogenic substrate Abz-Ala-Arg-Arg-Ala-Tyr($NO_2$)-$NH_2$[10].

**Preparation of standard proteins and high-mass proteome samples.** The standard proteins carbonic anhydrase (bovine), GAPDH (rabbit) and phosphorylase b (rabbit) were directly dissolved in 8 M urea to make 2–5 mg $ml^{-1}$ stock solutions. Bovine serum albumin was reduced in 5 mM dithiothreitol (DTT), alkylated with 10 mM iodoacetamide in the dark and precipitated with ice-cold acetone before being resuspensed in 8 M urea for OmpT digestion. For the human proteome sample, HeLa S3 cells were obtained from the American Type Culture Collection and grown as previously described[18]. Cells were lysed by boiling in cell lysis buffer (4% SDS, 100 mM Tris-HCl, 10 mM DTT, pH 7.5) for 10 min; they were then incubated with 100 mM iodoacetamide for 30 min in the dark, divided into aliquots and frozen at −80 °C for future use. To fractionate the whole proteome into molecular-mass bins, a continuous tube-gel electrophoresis technology, Gel-eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE), was applied for primary separation[19]. Specifically, an eight-channel, multiplexed commercial continuous tube-gel electrophoresis device (GELFREE 8100 fractionation system, Protein Discovery Inc.) was used with 8% or 10% gel cartridges (Protein Discovery) to prepare the high-mass HeLa proteome. The HEPES-SDS buffer system, pH 7.8, was used as recommended by the vendor. To load samples onto the GELFrEE devices, protein concentrations were measured using a BCA assay and aliquots of HeLa lysates corresponding to 1–2 mg of total protein were thawed on ice, precipitated by cold acetone at −20 °C for 30 min and air-dried before resuspension with sample loading buffer, and then heated at 50 °C for the commercial GELFrEE. After sample loading, the commercial GELFrEE device was operated as described in the manufacturer's instructions. Each fraction contained 1.2 ml of sample volume (150 µl for each channel; samples from eight channels were pooled together for the same fraction) and fractions corresponding to the high-mass proteome (20–100 kDa) were cleaned up by cold acetone precipitation and air-dried before resuspension in 8-M urea for OmpT digestion.

**OmpT digestion and sample cleanup.** To obtain active enzyme, aliquoted OmpT solution was thawed on ice, activated with 0.1 mM

LPS overnight[10,14,20] and dialyzed against enzymatic buffer (10 mM Bis-Tris-HCl, 2 mM EDTA, pH 6.0). Immediately after dialysis, OmpT (liganded to LPS) was mixed with resuspended standard proteins at a substrate:enzyme ratio of up to 75:1, (final protein concentration of 0.3–0.75 mg $ml^{-1}$) or with high-mass HeLa GELFrEE samples at a substrate:enzyme ratio of 25:1 (final protein concentration of ~0.5 mg $ml^{-1}$ and final urea concentration of 3.2 M). The mixtures were incubated at 22 °C overnight. Digested standard proteins or GELFrEE samples were cleaned up by methanol-chloroform precipitation[18,21] before they were solubilized at 100 °C in sample loading buffer and loaded onto a single-channel custom GELFrEE device for high-resolution secondary continuous tube-gel electrophoresis separation[6]. The buffer system of this custom device was Tris-glycine (25 mM Tris, 0.2 M glycine, 0.1% SDS). Tube gels with Tris-glycine were cast at 15% T in this secondary continuous tube-gel electrophoresis for resolving digested peptides. The custom GELFrEE device was operated at 180 V, and 16 fractions total were collected containing peptides up to 30 kDa over 100 min. SDS was removed from collected fractions by methanol-chloroform precipitation. The resultant protein pellets from either standard protein digestions or GELFrEE digestions by OmpT were recovered by buffer A (95% $H_2O$, 5% acetonitrile, 0.2% formic acid) solubilization and injected onto a nanocapillary liquid chromatography column coupled to a mass spectrometer for on-line characterization as described below.

**Nanocapillary liquid chromatography-mass spectrometry (nanoLC-MS/MS).** A PLRP-S trap column (New Objective, Inc.), with a 150-µm inner diameter and a 3-cm medium length, was used for sample loading, and a 10-cm-long, 75-µm inner diameter PLRP-S analytical column was used for sample separation. A linear gradient flowing at 300 nl $min^{-1}$ from an Eksigent 2D system started from 95% buffer A and 5% buffer B (5% $H_2O$, 95% acetonitrile, 0.2% formic acid), ramped to 40% B in 55 min and finally 85% B in 15 min. Samples eluted from the nanoLC were electrosprayed into a custom hybrid linear-ion-trap Fourier-transform ion cyclotron resonance mass spectrometer (11-Tesla LTQ-FT-Ultra mass spectrometer, Thermo Fisher Scientific). Samples were analyzed using a data-dependent top-2 or top-3 method. CID was applied with a 10–15 *m/z* isolation window and normalized collision energy of 41%; for MS1, 1–6 microscans at 160,000 resolving power at 400 *m/z* were used with a target value of 1 million and scan range of *m/z* 450–1,800 in the Fourier-transform ion cyclotron resonance cell (FT-ICR); for MS2, 2–6 microscans at 80,000 resolving power were used with a target value of 1–1.5 million in the FT-ICR. For CID and ETD comparison analysis, a Velos Orbitrap Elite system was used (Thermo Fisher Scientific). Samples were analyzed either using a data-dependent top-3 or top-5 method in separate CID or ETD runs, or top-2 or top-3 method in alternating CID and ETD runs. Both CID and ETD were applied with a 15 *m/z* isolation window; normalized collision energy for CID was set at 41%, and reaction time for ETD was 5–25 ms. For MS1, 2–4 microscans at 120,000 resolving power at 400 *m/z* were used with a target value of 1 million and scan range of *m/z* 400–1,500 in the Orbitrap; for MS2, 3–6 microscans at 60,000 resolving power were used with a target value of 1 million in the orbitrap. All the raw data files collected are available at Tranche (https://proteomecommons.org/group-data.jsp?i=360).

**Data reduction and database searching.** Each LC-MS/MS run was collected as a .raw file and processed with ProSightPC 2.0 SP1 software (Thermo Fisher Scientific). Briefly, monoisotopic neutral precursor and fragment masses were determined using the Xtract algorithm, compiled into a .puf file (ProSight Upload Format) and searched on a 168-core cluster in two different search modes (absolute mass and biomarker) against two shotgun-annotated human proteome databases.

Biomarker search mode does not assume any hypothetical cleavages in the database and queries every possible subsequence of any protein in the intact protein database (UniProt release 2011_10) for a match within the defined mass tolerance window. In this mode, the precursor mass tolerance window was set to 1.1 Da and the fragment mass tolerance was set to ±10 p.p.m. To estimate the FDR in biomarker search mode, a *q*-value evaluation approach was applied as previously described[2]. A decoy database was built by scrambling the protein sequences from the forward intact database[22]. All data were searched against both the forward and decoy databases separately using identical search parameters. All search hits were scored using a Poisson-based model[23] (*p* score), and a posterior probability-based *q* value was calculated for each hit to estimate the FDR for each identification event[24,25].

For the absolute mass search, a custom peptide database was constructed using the OmpT cleavage propensities (P1 = Lys, Arg; P1′ = Lys, Arg, Ala, Ser, Gly, Val, Ile, Leu) determined by biomarker search hits (see sequence logo in **Supplementary Fig. 7c**). Eight missed cleavages were considered in constructing this middle-down database, which contained 20 million peptide forms (including signal peptides, alternative splice variants and PTMs). To search data in absolute mass mode, ProSightPC's iterative searching was used, with the precursor mass tolerance window set to 2.2 Da and the fragment tolerance to ±10 p.p.m. for the first level search; an 81-Da precursor mass tolerance and ±10-p.p.m. fragment tolerance were used for the second level search. FDR estimation was performed as described above.

Peptide hits with a *q* value lower than 0.01 (1% FDR cutoff) from both the biomarker and absolute mass search modes were reported and used for further analysis in this study. A brief comparison was drawn between biomarker hits and absolute mass hits (**Supplementary Fig. 8**). ProteinCenter software (Thermo Fisher Scientific) was used to group peptides and cluster protein identifications for unique protein counting (**Supplementary Table 1**).


17. Dekker, N., Merck, K., Tommassen, J. & Verheij, H.M. *Eur. J. Biochem.* **232**, 214–219 (1995).
18. Lee, J.E. *et al. J. Am. Soc. Mass Spectrom.* **20**, 2183–2191 (2009).
19. Tran, J.C. & Doucette, A.A. *Anal. Chem.* **81**, 6201–6209 (2009).
20. Kramer, R.A. *et al. Eur. J. Biochem.* **269**, 1746–1752 (2002).
21. Wessel, D. & Flugge, U.I. *Anal. Biochem.* **138**, 141–143 (1984).
22. Elias, J.E. & Gygi, S.P. *Nat. Methods* **4**, 207–214 (2007).
23. Meng, F. *et al. Nat. Biotechnol.* **19**, 952–957 (2001).
24. Benjamini, Y. & Hochberg, Y. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
25. Storey, J.D. & Tibshirani, R. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).






doi:10.1038/nmeth.2074