ONLINE METHODS

Reagents. DNA restriction enzymes were purchased from Invitrogen and T4 DNA ligase was purchased from New England Biolabs. The *pET28a* vector and *E. coli* BL21(DE3) cells were obtained from EMD Biosciences. The SP-Sepharose media 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 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'-ATGCGGGCGAAACTTCTGGGAATAG-3' (forward) and 5'-TTAAAATGTGTACTTAAGACCAGCAGTAGTG-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'-ATTAATCCATGGCTTCTCGAGACTTTATCGTTTA-3' and 5'-ACTCGGGAATTCTTAAAAGTGTACTTAAGACCAG-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 *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 S.O.C. 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 liter H_2O) with 50 mg/l kanamycin at 37°C. The 5 ml starter culture was inoculated into 1 liter S.O.C medium with 50 mg/L 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 liter 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 pre-chilled lysis buffer was added quickly to introduce osmotic shock, followed by incubation on ice for another 30 min. The lysate was sonicated at 25 watts with a Sonic Dismembrator (Model 100, Fisher Scientific) every other minute 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, 1mM 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 pre-chilled 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*, filtered and the supernatant 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 which LPS-bound OmpT was found in two forms due to a single self-degradation site (Arg217-Lys218)[10], but > 80% of the enzyme still in its intact form (**Supplementary Fig. 3a, lane E2**). Based on SDS-PAGE analysis, fractions containing OmpT were pooled, aliquoted 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 stock solutions. Bovine serum albumin (BSA) was reduced in 5 mM dithiothreitol (DTT), alkylated with 10 mM iodoacetamide in the dark and precipitated with ice-cold acetone before resuspension 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, incubated with 100 mM iodoacetamide for 30 min in the dark, aliquoted 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 BCA assay and aliquoted 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 prior to resuspension in 8 M urea for OmpT digestion.

OmpT digestion and sample clean-up. 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), or mixed with high-mass HeLa GELFrEE samples at a substrate:enzyme ratio of 25:1 (final protein concentration of ~0.5 mg/ml 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 solubilizing at 100°C in sample loading buffer and were 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 in total, 16 fractions 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 nanoLC 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.), 150 μm inner diameter (i.d.) with a 3 cm media length, was used for sample loading, followed by a 10 cm long × 75 μm i.d. PLRP-S analytical column for sample separation. A linear gradient flowing at 300 nL/min 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. Collision-induced dissociation (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 5 method in separate CID or ETD runs, or top 2 or 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, complied 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 sub-sequence 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 ppm. In order to estimate the false discovery rate (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 = K, R; P1' = K, R, A, S, G, V, I, L) determined by biomarker search hits (see sequence logo in **Supplementary Fig. 8c**). 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 ppm for the first level search; an 81 Da precursor mass tolerance and ± 10 ppm 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 cut-off) 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. 9**). ProteinCenter software (Thermo Fisher Scientific) was used to group peptides and cluster protein identifications for unique protein counting (**Supplementary Table 1**).