

Proteome-wide analysis of protein carboxy termini: C terminomics

Oliver Schilling^{1–3}, Olivier Barré^{1,2}, Pitter F Huesgen^{1,2} & Christopher M Overall^{1,2}

As proteome-wide C-terminal sequence analysis has been largely intractable, we developed a polymer-based enrichment approach to profile protein C-terminal peptides by mass spectrometry and identified hundreds of C-terminal peptides in the *Escherichia coli* proteome. We isotopically labeled GluC protease-digested and undigested samples and identified GluC substrates and their cleavage sites by quantification of neo-C-terminal peptides. Our method thus enables global annotation of protein C-terminal posttranslational modifications, including proteolytic truncations.

Determining the sequence and nature of N and C termini of proteins provides important functional annotation of proteomes. Proteolysis modifies proteins, generating new N and C termini, also called neo-N and neo-C termini, often altering protein function¹. The subfield of proteomics encompassing global techniques to identify proteases and their substrates, now known as degradomics¹, includes methods for proteome-wide analysis of protein N termini called N terminomics; this is an area of recent intense technique development^{2–7} that has provided new information on protein isoforms and N-terminal modifications such as acetylation and cyclization^{4,7}. Moreover, by data parsing for protease canonical cleavage sites^{5,6} and by peptide quantification for proteases with broad or unknown specificity^{4,7} such approaches can simultaneously identify protease cleavage sites and substrates. However, owing mainly to the lower chemical reactivity of carboxyl groups, similar advances to enrich C-terminal peptides and study C-terminal posttranslational modifications, including proteolytic processing of proteins, have not been made.

Important C-terminal proteolysis events⁸ include chemokine processing⁹, peptide hormone maturation and fibrinolysis¹⁰, but it is likely that many important C-terminal modifications remain unknown or poorly characterized owing to the lack of appropriate techniques for global analysis. Current approaches for C-terminal sequencing of single proteins use anhydrotrypsin capture, a combination of LysC protease cleavage and amine capture, carboxy-peptidase ladder sequencing or diagonal electrophoresis (reviewed

in ref. 11). These approaches, however, are not suitable for analyzing complex samples. Techniques for carboxyl-group modification in proteomes have been described¹² but have not been adapted for C-terminal sequencing. Alternatively, some protein C-terminal peptides can be separated via their low charge and reduced binding affinity to cation exchange resin by chromatography², but this is neither specific nor does it allow orthogonal validation of C-terminal peptide location in proteins. Thus, global C terminus-centric techniques have been missing in functional proteomics¹¹.

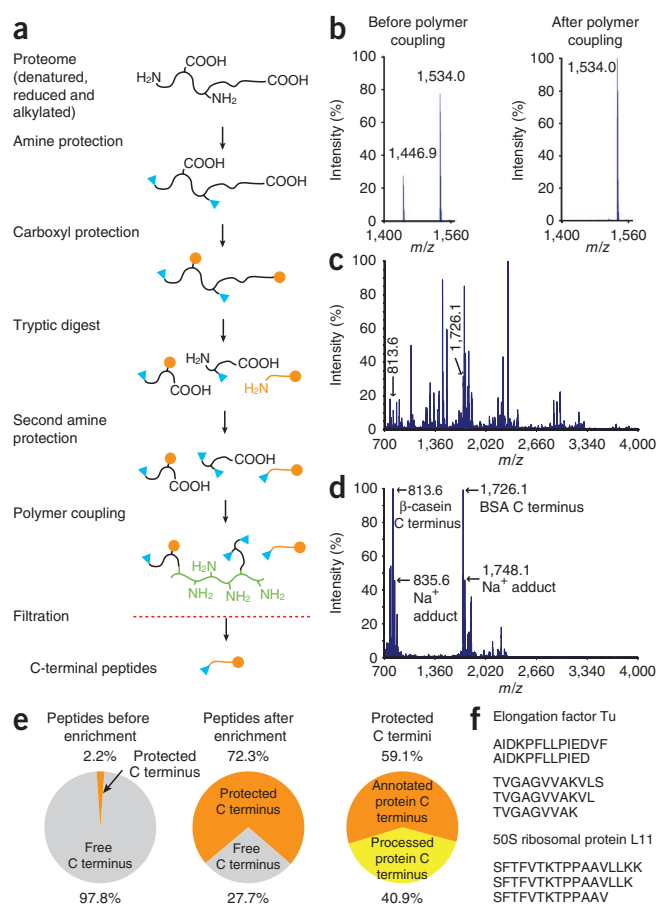
Here we report a strategy for the specific isolation and analysis of C-terminal peptides from complex proteomes, which we call C terminomics (Fig. 1a and Online Methods). In preparation for C-terminal enrichment, protein thiol groups are reduced and alkylated, and then protein N-terminal α -amines and lysine ϵ -amines are reductively methylated¹³. Carboxyl groups of C termini, aspartate and glutamate side chains are then protected by carbodiimide-mediated (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride; EDC) and *N*-hydroxy succinimide-assisted condensation with 1 M excess of ethanolamine. EDC-based carboxyl labeling proceeds under mild conditions, whereas an alternative we assessed, carboxyl esterification in methanolic HCl, was harsh, resulting in unacceptable side reactions, including deamidation of asparagine and glutamine (data not shown).

Trypsin digestion then yields the original protein C-terminal peptides, now with protected carboxyl groups, which differ from the internal and N-terminal tryptic peptides that have free, newly generated C termini. In the analysis phase of our approach, peptides with a C-terminal label can be unequivocally validated as bona fide peptides derived from the C terminus of proteins. To prevent cross-reactivity, peptide concatamerization or cyclization, we chemically protected the newly generated N termini of tryptic and C-terminal peptides via a second reductive methylation step. Then we removed internal tryptic and N-terminal tryptic peptides by covalent coupling to a high-molecular-weight (~56 kDa) linear polyallylamine polymer by EDC-mediated condensation of the free carboxyl groups to the primary amines of the polymer. The superior efficiency of polymers over chromatography resin or magnetic beads for peptide enrichment has recently been shown for N-terminal peptide isolation⁴. The blocked original protein C-terminal peptides remain uncoupled and are readily separated from the polymer by ultrafiltration. Then we subjected the enriched C-terminal peptides to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

To validate our approach, we reductively methylated the N terminus of the synthetic peptide NH₂-AFQVWSDVTPLR-COOH, yielding a peptide with a molecular mass of 1445.7 Da. Then we labeled carboxyl groups of the aspartate residue and C terminus

¹Department of Biochemistry and Molecular Biology, University of British Columbia, Centre for Blood Research, Vancouver, British Columbia, Canada. ²Department of Oral Biological and Medical Sciences, University of British Columbia, Centre for Blood Research, Vancouver, British Columbia, Canada. ³Present address: University of Freiburg, Institute for Molecular Medicine and Cell Research, Freiburg, Germany. Correspondence should be addressed to C.M.O. (chris.overall@ubc.ca).

Figure 1 | C-terminomics workflow. (a) To enrich C-terminal peptides for mass spectrometry analysis, protein amine and carboxyl groups are first chemically protected. After trypsination, peptide N termini are also protected. N-terminal and internal tryptic peptides are coupled to polyallylamine via their free C terminus. The original protein C-terminal peptides remain unbound and after separation by ultrafiltration are analyzed by LC-MS/MS. (b) Polymer-based enrichment of the peptide AFQVWSDVTPLR with protected carboxyl groups (theoretical $[M+H]^+ = 1,533.8$, measured $[M+H]^+ = 1,534.0$) from a variant with free carboxyl groups (theoretical $[M+H]^+ = 1,446.7$ measured $[M+H]^+ = 1,446.9$) before and after polymer coupling. (c,d) Mass spectra of a sample of C-terminal peptides of β -casein (812.6 Da, $[M+H]^+ = 813.6$) and BSA (1,725.1 Da, $[M+H]^+ = 1,726.1$) before enrichment (c) and after enrichment (d; corresponding sodium-adduct ions (+22 Da) are indicated). The BSA C-terminal peptide originates from a chymotryptic cleavage, indicating the presence of chymotryptic impurities or partially autolysed trypsin¹⁸. (e) Summary of data averaged from two experiments for isolation of protein C termini from *E. coli*. (f) Examples of processive proteolysis of protein C termini detected after enrichment of *E. coli* protein C termini.



with ethanolamine, yielding a 1,532.8-Da species. We detected both peptide variants, when mixed, as single protonated ions ($[M+H]^+$) by matrix-assisted laser desorption ionization (MALDI)-time of flight (TOF) mass spectrometry (Fig. 1b and Online Methods). After polymer-based enrichment, however, we detected only the C terminus-protected peptide by MALDI-TOF (Fig. 1b).

In a more complex setting, we selectively recovered the C-terminal peptides from bovine serum albumin (BSA) and β -casein and analyzed them by mass spectrometry (Fig. 1c,d). Before polymer enrichment, the C-terminal peptides were overshadowed by more intense tryptic-peptide ions and thus were in the lower third in terms of intensity (Fig. 1c). After polymer enrichment, the C-terminal peptides ranked as the two most intense ions in the mass spectrum (Fig. 1d) with Mascot scores of 128 for BSA and 30 for β -casein after MS/MS analysis (Supplementary Fig. 1).

We then applied our C-terminomics approach to *Escherichia coli* lysates. We determined that carboxyl protection was >95% complete by dataset searches for unblocked aspartate or glutamate residues after LC-MS/MS analysis. Before polymer enrichment, two independent experiments both yielded ~98% internal tryptic peptides with free C termini (Fig. 1e and Supplementary Tables 1 and 2). After polymer enrichment, we identified 460 peptides, of which >72% had a protected C terminus, from 196 proteins in two separate experiments by MS/MS analysis (Fig. 1e and Supplementary Tables 3 and 4). Overall, the original protein C termini represented ~60% of the C-terminally protected peptides identified, whereas ~40% of the peptides originated from internal positions owing to proteolytic processing in the cell generating stable cleavage products (Fig. 1e). The exact distinction between 'original' and 'proteolytically generated' protein C termini depends on functional proteome annotation, which for protein start and termination sites is incomplete¹⁴. The amount of C-terminal processing was surprisingly high, yet similar to the ~50% of N-terminally processed proteins (excluding initiator methionine removal) previously found in *E. coli* proteomes⁶.

Compared with previous C terminus-labeling approaches, such as aniline benzoic acid labeling analysis of *Lactococcus lactis*, which identified 188 proteins¹², we identified a greater number of C-terminal peptides owing to the selective removal of internal tryptic peptides. Using trypsin normally ensures that

peptides have a C-terminal basic residue that aids in LC-MS/MS detection and identification. However, semitryptic C-terminal peptides can have any terminal residue, leading to charge reduction as well as potential incompatibility with LC-MS/MS analysis and with analysis using search engines that are optimized for fully tryptic peptides.

In addition to routine matching of isolated C-terminal peptide sequences with those annotated in databases by Mascot, in our method the presence of the C-terminal label allows for orthogonal validation of original protein C-terminal peptides. For neo-C-terminal peptides with sequences from within the protein generated by proteolysis, it is crucial to have orthogonal validation to ensure that these are part of the proteolytic signature of the sample and are not due to cleavage during sample processing. The presence of the C-terminal label, incorporated in the intact proteins before digestion by trypsin, provides a reliable means of orthogonal validation that is not possible when relying on chromatographic separation alone. We identified cases of C-terminal truncations resulting from processive proteolysis, also known as ragging (Fig. 1f). Of the cleavage sites we identified, 62% were located in the C-terminal half of proteins. However, cleavage sites within 10 residues of either the original mature N or C terminus may not be identified by mass spectrometry if the neo-peptides are redundant as a result of being too short. This is lessened by a key feature of our approach whereby lysine residues are blocked, preventing trypsin cleavage. Hence, peptides are generated with ArgC-like specificity producing longer sequences, which considerably improves identification rates by mass spectrometry. Other cleavage sites occurred deep inside

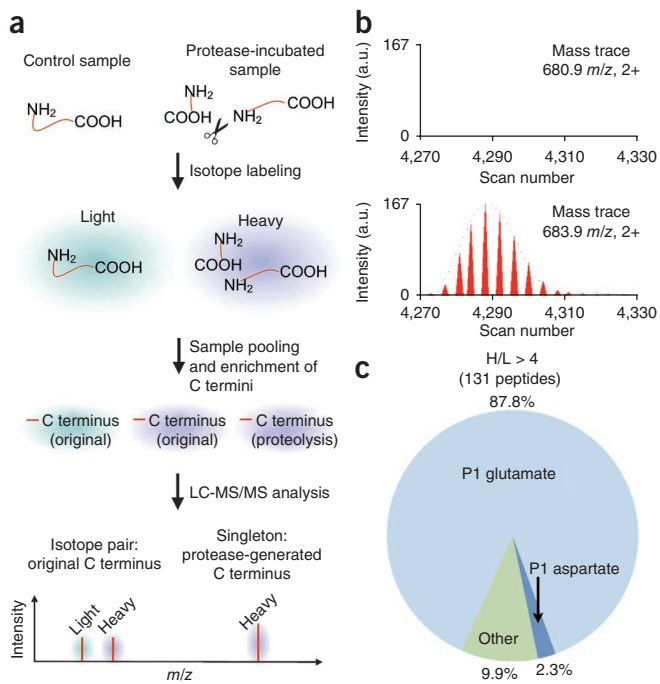


Figure 2 | C-TAILS profiling of protease cleavage sites. **(a)** The workflow is based on the general scheme for isolation of original protein C-terminal peptides. Protease cleavage generates neo-C termini. To distinguish induced proteolysis from background proteolysis, control and protease-treated samples are differentially labeled by stable isotope incorporation. After trypsin digestion and labeling, both samples are mixed, and C-terminal peptides are isolated as shown in **Figure 1a**. Neo-C-terminal peptides generated by the protease are predominantly found in the heavy isotopic form used for labeling of the protease-treated sample and lack the light isotopic counterpart of the control sample and thus are found as high-ratio peptides. **(b)** Identification of the peptide ALLNSMVIGVTE as a GluC-generated neo-C terminus. GluC-treated sample was labeled with $(d(2)C^{13})$ -formaldehyde (heavy) and the untreated control sample with $(d(0)C^{12})$ -formaldehyde (light). An extracted ion chromatogram¹⁹ for ALLNSMVIGVTE was only present in the heavy-labeled form (bottom; double charged (2+) peptide ion); this peptide was not detected in the light isotopic form (top); hence, this is a GluC-generated neo-C-terminal peptide. **(c)** Summary of data collected upon limited GluC proteolysis of *E. coli* proteome. H/L, heavy:light.

the protein; ~25% were cleaved within dibasic sequences with a lysine or arginine in both non-prime-side (P1) and prime-side (P1') substrate positions N- and C-terminal to protease cleavage sites, respectively, indicating cleavage by *E. coli* proteases such as omptin. Spiking standards for absolute quantification are necessary to determine the extent to which a protein is cleaved (R. Fahlman and C.M.O.; unpublished data).

We modified our C terminomics procedure to incorporate stable-isotope labeling of C-terminal peptides, which enables quantitative comparison of C termini from different samples to distinguish induced proteolysis (for example, by overexpression or addition of a protease) from background proteolysis (**Fig. 2a**). In this approach, which we call C-terminal amine-based isotope labeling of substrates (C-TAILS), we incorporated a heavy isotopic formaldehyde during both reductive methylation steps. This orthogonal labeling strategy targets lysine side chains as well as protein and peptide N termini. Direct isotope coding of carboxyl groups could be achieved through isotopic variants of ethanolamine, but ethanolamine is used in high concentrations (1.0 M) rendering this cost-ineffective. Alternatively, isobaric tag for relative and absolute quantitation (iTRAQ)¹⁵ or metabolic labeling could be used.

To test C-TAILS, we incubated native *E. coli* cell lysate proteins with *Staphylococcus aureus* protease V8 (GluC) (1:100) for 18 h at 25 °C, and we incubated a control sample without protease. GluC cleaves canonically at glutamate but also at aspartate¹⁶, thus enabling manual data parsing to validate the effectiveness of neo-C-terminal peptide substrate discrimination based on peptide isotope ratios⁴. We labeled the amino groups of the protease-treated sample with $(d(2)C^{13})$ formaldehyde and labeled the control sample with normal isotopic abundance formaldehyde ($(d(0)C^{12})$) (Online Methods). We pooled the samples and enriched the C termini as described above, followed by LC-MS/MS analysis (**Fig. 2b**). More than ~90% of the peptides identified had C-terminal residues corresponding to GluC cleavage sites (glutamic acid

and aspartic acid). We hypothesized that peptides with a high heavy:light ratio would represent GluC cleavage sites⁴. In total, we identified 241 heavy- or light-labeled, C-terminally protected peptides (**Supplementary Table 5**). Of these, 131 had a heavy:light ratio >4, with 87.8% of these having a P1 glutamate and 2.3% a P1 aspartate (**Fig. 2c**), in good agreement with a proteome-wide 'proteomic identification of protease cleavage site specificity' peptide assay that yielded 94.6% P1 glutamate or aspartate¹⁶. We also observed other C-terminal residues including 0.8% (each) of alanine, histidine and lysine, and 1.5% (each) of glycine, leucine, proline, arginine and threonine; the latter are likely generated by *E. coli* proteases after GluC cleavage and substrate unfolding.

In summary, we present an isotope-encoded quantitative C terminomics strategy for proteome-wide C termini enrichment and LC-MS/MS analysis. C-TAILS can be used to reliably identify neo-C-terminal sequences and protease substrates to complement N-terminal cleavage site analysis by N-TAILS⁴, extending substrate coverage. The method uses commercially available and cost-effective reagents; additional development of optimized carboxyl-labeling reagents and carboxyl-capturing polymers could potentially increase the rate of peptide identification. Overall, the method should be useful for functional proteome annotation and can provide crucial information on protein modification and proteolysis in complex systems. With a method for global analysis of the C terminome, many possibilities are now open for improved understanding of the role of the C terminus in protein function and cell biological processes.

METHODS

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

Accession codes. Proteomics Identifications database¹⁷ (PRIDE): 11684, 11685, 11686, 11687 and 11688 (LC-MS/MS data).

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

ACKNOWLEDGMENTS

O.S. acknowledges support from the Deutsche Forschungsgemeinschaft (grants SCHI 871/1-1 and 871/2-1) and the Michael Smith Foundation for Health Research (MSFHR). O.B. acknowledges support from the Swiss National Science

Foundation and Canadian Institutes of Health Research. P.F.H. is supported by Postdoctoral Fellowships of the German Academic Exchange Service (DAAD) and the MSFHR. C.M.O. is supported by a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology. This work was supported by a grant from Canadian Institutes of Health Research and from a program grant in Breast Cancer Metastases from the Canadian Breast Cancer Research Alliance with funds from the Canadian Breast Cancer Foundation and the Cancer Research Society as well as with an Infrastructure grant from MSHFR. We thank W. Chen, S. Perry and S. He for LC-MS/MS measurements. The University of British Columbia Centre for Blood Research Mass Spectrometry Suite is supported by the Canada Foundation for Innovation and the MSFHR.

AUTHOR CONTRIBUTIONS

O.S. and C.M.O. conceived the approach. All authors contributed to experimental design. O.S. performed all experiments, except O.B. and P.F.H. performed the BSA and β -casein peptide enrichment studies, and O.B. contributed to the proteome experiments. O.S. and C.M.O. wrote the paper, and all authors edited the paper. O.S. and P.F.H. prepared the figures. C.M.O. supervised the project and provided grant support.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturemethods/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- López-Otín, C. & Overall, C.M. *Nat. Rev. Mol. Cell Biol.* **3**, 509–519 (2002).
- Dormeyer, W., Mohammed, S., Breukelen, B. Krijgsveld, J. & Heck, A.J.R. *J. Proteome Res.* **6**, 4634–4645 (2007).
- Gevaert, K. *et al. Nat. Biotechnol.* **21**, 566–569 (2003).
- Kleifeld, O. *et al. Nat. Biotechnol.* **28**, 281–288 (2010).
- Mahrus, S. *et al. Cell* **134**, 866–876 (2008).
- Timmer, J.C. *et al. Biochem. J.* **407**, 41–48 (2007).
- Van Damme, P. *et al. Nat. Methods* **2**, 771–777 (2005).
- Gomis-Rüth, F.X. *Crit. Rev. Biochem. Mol. Biol.* **43**, 319–345 (2008).
- Cox, J.H., Dean, R.A., Roberts, C.R. & Overall, C.M. *J. Biol. Chem.* **283**, 19389–19399 (2008).
- Reznik, S.E. & Fricker, L.D. *Cell. Mol. Life Sci.* **58**, 1790–1804 (2001).
- Impens, F. *et al. Proteomics* **10**, 1284–1296 (2010).
- Panchaud, A. *et al. Mol. Cell. Proteomics* **7**, 800–812 (2008).
- Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A.J.R. *Nat. Protocols* **4**, 484–494 (2009).
- Ansong, C., Purvine, S.O., Adkins, J.N., Lipton, M.S. & Smith, R.D. *Brief. Funct. Genomics Proteomics* **7**, 50–62 (2008).
- Prudova, A., Auf dem Keller, U., Butler, G.S. & Overall, C.M. *Mol. Cell. Proteomics* **9**, 865–882 (2010).
- Schilling, O. & Overall, C.M. *Nat. Biotechnol.* **26**, 685–694 (2008).
- Vizcaino, J.A. *et al. Proteomics* **9**, 4276–4283 (2009).
- Keil-Dlouha, V., Zylber, N., Imhoff, J.-M., Tong, N.-T. & Keil, B. *FEBS Lett.* **16**, 291–295 (1971).
- Han, D.K., Eng, J., Zhou, H. & Aebersold, R. *Nat. Biotechnol.* **19**, 946–951 (2001).

ONLINE METHODS

Polymer-based enrichment of peptide AFQVWSDVTPLR. For amine labeling, the peptide AFQVWSDVTPLR was incubated with 20 mM formaldehyde (Sigma) in the presence of 20 mM of sodium cyanoborohydride (NaBH_3CN) (Sterogene) for 16 h at 25 °C in 100 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) (pH 7.5). The dimethylated peptide was then desalted by Sep-Pak C18 cartridges (Waters). Carboxyl labeling was performed in 200 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) (pH 5), 1.0 M ethanolamine, 5 mM *N*-hydroxy succinimide (NHS), 50 mM EDC for 2 h at 25 °C. Another 50 mM EDC was added and the reaction was completed for 16 h at 25 °C. The peptide was then desalted by Sep-Pak C18. Both peptide variants (protected amines and free carboxyls as well as protected amines and protected carboxyls) were mixed at equal abundance, yielding a 0.1 mM peptide mix in 200 mM MES (pH 5) (sample 'before enrichment'). This sample was then incubated with 1 mM polyallylamine (molecular weight, ~56,000; Sigma), 50 mM EDC, 5 mM NHS for 18 h at 25 °C, yielding the sample 'after enrichment'. For mass spectrometry analysis, 1 μl of every sample was mixed with 1 μl of a saturated α -cyano-4-hydroxycinnamic acid (CHCA) solution in 50% (vol/vol) acetonitrile containing 1% (vol/vol) trifluoroacetic acid and spotted on a steel target. Samples were analyzed by an Applied Biosystems 4700 matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF-MS) in reflector mode.

Enrichment of C-terminal peptides of bovine β -casein and serum albumin. We dissolved 20 nmol bovine β -casein (>90% pure; Sigma) and 20 nmol bovine serum albumin fraction V (BSA, > 90% pure; Gibco) in 2.0 M guanidine hydrochloride and 100 mM HEPES (pH 7.5). Disulfide bonds were reduced by incubation with 10 mM dithiothreitol (DTT) for 60 min at 37 °C and free thiols were protected by incubation with 40 mM iodoacetamide for 2 h at 25 °C in the dark, followed by quenching of the remaining iodoacetamide by addition of further 20 mM DTT. Primary amines were protected by reductive dimethylation with 40 mM formaldehyde and 20 mM NaCNBH_3 for 16 h at 37 °C. Proteins were precipitated with chloroform-methanol as described elsewhere²⁰. Carboxyl protection, trypsin digestion and chemical protection of new N termini were performed as described below. The β -casein and BSA digest was purified with a Sep-Pak C18 cartridge using 10 mM HCl for washes and 60% acetonitrile for elution, followed by concentration with a vacuum concentrator (Savant). Polymer coupling of internal peptides and ultrafiltration were performed as described below, with the exception that 0.5 mM polyallylamine and 10 mM sulfo-NHS were used. For MALDI-TOF-MS/MS, aliquots of the two-protein digest before and after polymer coupling were desalted with Omix C18 pipette tips (Varian) according to manufacturer's instructions. Samples were then mixed with 3 mg ml^{-1} CHCA solution in 80% (vol/vol) acetonitrile containing 0.1% (vol/vol) trifluoroacetic acid and spotted on a steel target. Samples were analyzed by an Applied Biosystems 4700 MALDI-TOF-MS, including MS/MS measurements. MS/MS spectra were analyzed by Mascot²¹ searches against the SwissProt²² database using the following parameters: fixed modifications were carbamidomethylation of cysteine (+57.02 Da), dimethylation at lysine and peptide N termini (+28.03 Da), ethanolamine at aspartic acid and

glutamic acid (+43.04 Da); variable modifications were ethanolamine labeling at peptide C termini (+57.02 Da).

***E. coli* growth and lysis.** *E. coli* strain MG1655 cultures were grown in Luria-Bertani rich medium and lysed in 100 mM HEPES (pH 7.4) by ultrasonication in the presence of the following protease inhibitors: 1 mM trans-epoxysuccinyl-L-leucylamido (4-guanidino)butane (E-64), 10 mM phenylmethanesulfonyl fluoride (PMSF) and 1 mM ortho-phenanthroline. Cell debris and insoluble proteins were removed by centrifugation at 12,000g for 10 min. DNA and RNA were degraded by addition of 10 $\mu\text{g ml}^{-1}$ of DNase I and RNase A. Small-molecule components were removed by size exclusion chromatography using Sephadex G-10 (GE-Healthcare) pre-equilibrated in 100 mM HEPES (pH 7.4) and 150 mM NaCl.

Protection of sulfhydryls and primary amines. Proteins were denatured in 2.0 M guanidine hydrochloride. Disulfide bonds were reduced by incubation with 20 mM DTT for 30 min at 70 °C. Subsequently, free sulfhydryls were protected with 50 mM iodoacetamide (3 h at 25 °C) followed by quenching of unreacted iodoacetamide with 5 mM DTT. Primary amines were protected by dimethylation using 20 mM formaldehyde and 20 mM NaCNBH_3 for 16 h at 25 °C. Proteins were precipitated by addition of 15% (vol/vol) trichloroacetic acid (TCA), incubation on ice for 1 h and centrifugation at 12,000g for 10 min. The pellet was washed three times with ice-cold methanol to remove traces of TCA.

Protection of carboxyls. Proteins were resolubilized in 200 mM MES, pH 5.0, 2.0 M guanidine hydrochloride, 1.0 M ethanolamine and 5 mM NHS. The reaction was started by addition of 20 mM EDC with an incubation for 3 h at 25 °C. The EDC coupling step was repeated twice with the last incubation lasting 16 h. EDC coupling is a potentially problematic step owing to rapid decomposition of EDC in water and fluctuations of the pH value, requiring frequent pH control and adjustment. Proteins were precipitated by addition of 15% TCA, incubation on ice for 1 h and centrifugation at 12,000g for 10 min. The pellet was washed three times with ice-cold methanol to remove traces of TCA and EDC to reduce the potential for carry-over and inadvertent labeling of tryptic peptides.

Trypsin digestion and chemical protection of new N termini. Proteins were resuspended in a minimal volume of 20 mM HEPES (pH 7.5) and 2.0 M guanidine hydrochloride. The guanidine hydrochloride concentration was then reduced to 0.5 M by addition of 20 mM HEPES (pH 7.5). Proteins were digested for 18 h at 37 °C with trypsin (1-chloro-3-tosylamido-4-phenyl-2-butanone-treated; Sigma) at a ratio of 1:100. After digestion, primary amines of internal peptides were protected by reductive dimethylation with 20 mM formaldehyde and 20 mM NaCNBH_3 and incubation for 6–16 h at 37 °C.

Polymer preparation. A 2 mM stock solution of polyallylamine was prepared in 200 mM MES (pH 5.0), 2 M guanidine hydrochloride and 20% (vol/vol) acetonitrile.

Polymer coupling of internal peptides and ultrafiltration. Tryptic digests were incubated with an equal volume of stock polyallylamine polymer giving a final 1 mM polyallylamine

(representing a primary amine concentration of ~1.0 M), 5 mM sulfo-NHS in 200 mM MES (pH 5). Polymer coupling of internal and N-terminal peptides was started by addition of 50 mM EDC with an incubation for 3 h at 25 °C. The EDC step was repeated twice with the last incubation lasting 16 h with regular monitoring and adjusting of pH. Separation of original protein C-terminal blocked peptides (unbound) from polymer and polymer-coupled internal peptides was achieved by filtration with a 10-kDa cut-off Microcon filter device (Millipore) followed by two 500 µl washes with water. Elution and wash samples were pooled together and desalted using Sep-Pak C18.

Liquid chromatography-tandem mass spectrometry. LC-MS/MS was performed on a Packings capillary LC system (Dionex) coupled to a quadrupole time-of-flight mass spectrometer (QSTAR Pulsar; Applied Biosystems, operated by members of the Michael Smith Laboratory and Laboratory for Molecular Biophysics Proteomics Core Facility; or QSTAR XL, Applied Biosystems, operated by members of the University of British Columbia Centre for Blood Research Mass Spectrometry Suite). For the QSTAR Pulsar instrument, samples were resuspended in 2% (vol/vol) acetonitrile, 0.1% formic acid and loaded onto a column packed with PepMap C18 resin (Dionex). Peptides were eluted using a 5–40% gradient of organic phase (buffer B) over 90 min. Buffer A was 2% acetonitrile and 0.1% formic acid, whereas buffer B was 85% acetonitrile and 0.1% formic acid. For the QSTAR XL instrument, samples were resuspended in 5% (vol/vol) acetonitrile, 3% formic acid and loaded onto a column packed with Magic C18 resin (Michrom Bioresources). Peptides were eluted using a 7–40% gradient of organic phase over 95 min. Buffer A was 2% acetonitrile with 0.1% formic acid and buffer B was 98% acetonitrile, 0.1% formic acid. For both instruments, mass spectrometry data were acquired automatically using Analyst QS software (Applied Biosystems) for information-dependent acquisition based on a 1 s mass spectrometry survey scan followed by up to three (QSTAR Pulsar) or two (QSTAR XL) MS/MS scans of 3 s each.

Tandem mass spectrometry data analysis. Peptides were identified by X!Tandem²³ using a decoy search strategy in conjunction with PeptideProphet²⁴ at 95% confidence level. Mass tolerance for QSTAR XL data was 100 p.p.m. for parent ions and 0.1 Da

for fragment ions. Mass tolerance for QSTAR Pulsar data was 200 p.p.m. for parent ions and 0.2 Da for fragment ions. As trypsin does not cleave at dimethylated lysines, cleavage specificity was defined as C-terminal to arginine (ArgC-like specificity). Semistyle cleavage searches were applied with up to three missed cleavages. Static modifications included carboxyamidomethylation of cysteine residues (+57.02 Da), dimethylation of lysines and amino termini (+28.03 Da for normal isotope abundance; 34.06 Da for d(2)C¹³ formaldehyde), ethanolamine-modification (+43.04 Da) for aspartate and glutamate residues. In searches for peptides with free C termini, the C termini were defined as unmodified; in searches for peptides with protected C termini, the C termini were defined as being modified with ethanolamine. Peptides of pre-enrichment samples possessed free N termini. X!Tandem refinement parameters included methionine oxidation as well as glutamine and asparagine deamidation.

Identification of GluC cleavage products. To demonstrate applicability of our approach for protease substrate discovery, *E. coli* lysate was first incubated for 4 h at 25 °C to break down the serine protease inhibitor PMSF, which had been added during *E. coli* lysis with the other protease inhibitors listed above to block endogenous protease activity. The lysate was then incubated with the serine protease GluC (1:100) for 18 h at 25 °C. A control sample was kept for 18 h at 25 °C without added protease. After incubation, sulfhydryls, primary amines and carboxyls were protected as described above with the following modification: primary amines of the GluC-treated sample were labeled with isotopically heavy formaldehyde, (d(2)C¹³) formaldehyde (Cambridge Isotopes) both before and after trypsin digestion. The control sample was labeled with light formaldehyde (d(0)C¹²). After trypsin digestion and primary amine protection samples were mixed at equal abundance. C termini were isolated and analyzed by LC-MS/MS as described above. Heavy:light ratios were determined by XPRESS¹⁹.

A step-by-step protocol is available in the *Nature Protocols* Network <http://dx.doi.org/10.1038/nprot.2010.100>.

- Wessel, D. & Flügge, U.I. *Anal. Biochem.* **138**, 141–143 (1984).
- Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. *Electrophoresis* **20**, 3551–3567 (1999).
- Boeckmann, B. *et al. Nucleic Acids Res.* **31**, 365–370 (2003).
- Craig, R. & Beavis, R.C. *Bioinformatics* **20**, 1466–1467 (2004).
- Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. *Anal. Chem.* **74**, 5383–5392 (2002).