

System-wide proteomic identification of protease cleavage products by terminal amine isotopic labeling of substrates

Oded Kleifeld

Israel Institute of Technology, Faculty of Biology, Haifa 32000, Israel.

Alain Doucet

Department of Biochemistry and Molecular Biology, University of British Columbia.

Jayachandran N. Kizhakkedathu

Department of Pathology and Laboratory Medicine, University of British Columbia.

Christopher M. Overall

Department of Oral Biological and Medical Sciences, University of British Columbia

Method Article

Keywords: N-terminal labeling, degradomics, protease, quantitative proteomics, positional proteomics, polymer for proteomics, dimethylation, mass spectrometry, matrix metalloproteinase, MMP

Posted Date: March 7th, 2010

DOI: <https://doi.org/10.1038/nprot.2010.30>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Introduction

The sequence and nature of all the protein amino-termini (N-termini) within the proteome (the N-terminome) provides valuable functional annotation, since translation start sites, N-terminal isoforms, modifications and truncations determine the cellular localization, activity and fate of most proteins¹. As ~85% of eukaryotic proteins have an acetylated N-terminus² and all proteins undergo proteolysis³, these are not only two of the most ubiquitous, but also two of the most important post-translational modifications^{4,5}. The protein amino-terminus is susceptible to amino-terminal peptidase processing, modification of the alpha-amino group, and side-chain specific changes that can target a protein for ubiquitination and degradation or protect it from rapid turnover and so determines its half-life¹. In addition to constitutive proteolysis, regulated processing of protein amino termini can irreversibly change the protein activity or properties⁶⁻⁸ but the extent to which proteolysis sculpts the proteome is unknown⁴. Hence, it is important to determine the cleavage site within each protease substrate, since the biological activity of the cleavage products is commonly determined by the precise fragmentation pattern. With 569 members, proteases are the second largest enzyme class in man⁹ and are 5-10% of drug targets¹⁰. Crucial to linking a specific protease with a defined biological pathway and for drug development is determining the substrate repertoire, or substrate degradome¹¹, of a protease since this can generate hypotheses on its role and provide biomarkers of drug efficacy. Yet, for around half of the proteases in man no substrates are known and for the other half, the substrate degradome is incompletely annotated^{3,11}. Thus specific degradomics techniques are needed to rapidly identify and quantify the N-terminome in order to reveal the extent of proteolysis in a system, the functional state of key molecules, and to identify new substrates. Positional proteomics approaches that isolate only the N-terminal peptides of proteins, the N-terminome¹²⁻¹⁶, have been proposed for sample simplification before mass spectrometric (MS) analysis and for proteome annotation, but coverage is often limited^{12-15,17,18} and aside from combined fractional diagonal chromatography (COFRADIC) most of these approaches were not reported for global protease cleavage site analysis. The main obstacles towards this latter goal are the identification the neo-N-termini of the substrates generated by specific proteolysis and to distinguish these not only from the natural N-termini, but also from N-termini generated by background proteolysis of the proteins in a sample and by trypsin digestion (internal tryptic peptides) in proteomic workflows^{3,19}. Solving these problems requires innovative strategies to circumnavigate the very similar chemical properties of the primary amines of the lysine side chains and N-termini. Recent reports of different approaches to tackle this difficult task include lysine-specific blocking of intact proteins to expose only amino-termini for biotinylation followed by affinity capture²⁰, specific labeling followed by *in silico* selection of protease generated neo-N-terminal peptides²¹, specific enzyme-mediated biotinylation of unblocked alpha-amine groups with enrichment for the biotinylated N-terminal peptides after one reaction¹⁶. Although these approaches represent a welcome step forward for studying protease-generated neo-N-terminal peptides, they are still limited in different aspects such as quantification²⁰, coverage²¹ or are potentially biased¹⁶ and more importantly incapable of analysing naturally blocked N-terminal peptides. To date, COFRADIC is the only N-terminomics approach that provides, broad coverage and isotopic quantification and can be applied to study protease substrate degradomes as well as to completely annotate the N-terminome^{12,22-25}. However COFRADIC is an expensive and time-consuming procedure involving multiple and complicated enzymatic and chemical steps, multiple HPLC fractionations and up to 150 MS/MS analyses per experiment²⁵. To overcome these problems we developed a new positional proteomics approach: Terminal Amine Isotopic Labeling of Substrates (TAILS)²⁶. TAILS is a combined N-terminomics and protease substrate discovery degradomics platform for the simultaneous quantitative analysis of the N-terminome and proteolysis on a proteome-wide scale in one MS/MS analysis. By a three-day procedure with flexible labeling options, TAILS removes internal tryptic and C-terminal peptides to enrich for all forms of N-terminal peptides by negative selection. To overcome nonspecific peptide binding and the low capacity of derivatized chromatographic beads that in some techniques necessitates large sample amounts for analysis, we developed a novel class of dendritic polyglycerol aldehyde polymers optimized for efficient, high capacity tryptic peptide binding with virtually no non-specific interactions. Rather than deliberately excluding acetylated proteins^{13,16,20,21}, TAILS provides wide coverage of all forms of naturally blocked N-terminal peptides and allows for their quantification through isotopic labeling of lysine side-chains. In addition to annotating the proteome we utilize these peptides to form a statistical classifier for the TAILS experiments to determine statistically valid isotope ratio cutoffs.

Reagents

****Cell lines**** Cell lines from an organism with a fully sequenced genome. ****Reagents**** These should be of the purest available: 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel plus loading and running buffers and silver stain solutions. Acetone - (Sigma-Aldrich). Acetonitrile - (Sigma-Aldrich). Dithiothreitol (DTT) - (Sigma-Aldrich). Ethylenediaminetetraacetic acid (EDTA) - (Sigma-Aldrich). ¹²C₂-formaldehyde (CH₂O) - 12.3 M for light labeling (37% (vol/vol), Sigma-Aldrich, cat. no. "252549":http://www.sigmaaldrich.com/catalog/ProductDetail.do?D7=0&N5=SEARCH_CONCAT_PNO%7C&N4=252549%7C&N25=0&QS=ON&F=SPEC). ¹³C₂-formaldehyde (¹³CD₂O) - 6.6 M for heavy labeling (20% in D₂O, 99% ¹³C, 98% D, Cambridge Isotopes, cat. no. "CDLM-4599-1":http://www.isotope.com/cil/products/displayproduct.cfm?prod_id=4208) formaldehyde vapors are toxic - prepare solutions in a fume hood. Formic acid (Sigma Aldrich). Guanidine hydrochloride (GuHCl) - (Sigma-Aldrich). 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) (Sigma-Aldrich), 1 M pH 7.0. High molecular weight-aldehyde derivatized (HPG-ALD) polymer at around 35 mg/ml. *_Note_*: "HPG-ALD":<http://www.flintbox.ca/technology.asp?page=3081> polymers for proteomics are available through Flintbox, The Global Intellectual Exchange and Innovation Network - "www.flintbox.ca":www.flintbox.ca, Flintbox Innovation Network Inc., 21 Water Street, 5th Floor, Vancouver BC, Canada, V6B 1A1. Iodoacetamide (IAA) - (Sigma-Aldrich), 0.5 M stock in water. *_Note_*: Iodoacetamide stock solution should be freshly prepared and kept at 4 °C. Methanol (Sigma-Aldrich). Sodium chloride (NaCl) - (Sigma-Aldrich). Sodium cyanoborohydride (NaBH₃CN) 1 M (ALD coupling solution, Sterogene, cat. no. "9704-01":http://www.sterogene.com/product_detail.php?ID=9704-01). *_Note_*: ALD solution should be kept at 4 °C and not stored past the expiry date to ensure good labeling efficiency. Phenylmethylsulphonyl fluoride (Sigma-Aldrich), 100 mM stock (100x). *_Caution_*: PMSF is carcinogenic. The stock solution in ethanol or acetonitrile should be freshly prepared. Phosphate buffered saline (PBS), sterile (138 mM NaCl, 2.7 mM KCl, 20 mM Na₂HPO₄, 1.5 mM

KH₂PO₄, pH 7.4). Protein concentration assay. Test protease for TAILS assay. Trifluoroacetic acid (TFA) (Sigma-Aldrich). Trypsin, mass spectrometry grade (Promega). Microcapillary tubes (Hamilton).

Equipment

1.5 mL microfuge tubes (e.g. Eppendorf or equivalent). **_Note_:** Polymers released from tubes and surfaces upon exposure to chemicals and solvents contaminate and interfere with mass spectrometric analysis. We have found that microtubes from Eppendorf show good chemical resistance and are suitable for the procedures described in this protocol. 15 mL polyethylene sterile tubes (e.g. "Nunc":<http://www.nuncbrand.com/us/page.aspx?ID=10558>, "Corning":<http://catalog2.corning.com/Lifesciences/en-US/Shopping/ProductDetails.aspx?productid=430052%28Lifesciences%29&categoryname=General+Labware+and+Equipment%28Lifesciences%29%7CCentrifuge+Tubes+%28Lifesciences%29%7C> or equivalent). **_Note_:** These tubes are going to be used for acetone precipitation, thus they require chemical resistance to acetone and methanol, and centrifugation at 15,000g. 50 mL tubes for collection of cell-conditioned medium. 50 mL 0.22 µm filtering devices such as Millipore "Steriflip":<http://www.millipore.com/catalogue/module/c3238> or equivalent. Argon gas. Centrifugal protein concentrator with 5-kDa molecular weight cut-off such as Millipore "Amicon-Ultra 15":<http://www.millipore.com/catalogue/module/c7715> concentrator. Centrifuge for volumes 50 ml tubes at up to 20,000g. Computer equipped with proteomic analysis software. Dialysis equipment with a 10-kDa molecular weight cut-off (the Pierce "Slide-A-Lyzer":<http://www.piercenet.com/Products/Browse.cfm?fldID=293902DB-5056-8A76-4E73-286D1FC1B2A5#minidialysisunits> works well). Standard laboratory equipment for cell culture, molecular biology, and protein chemistry. Reversed-phase solid phase extraction cartridges (e.g. Waters "Sep-Pak C18":http://www.waters.com/waters/partDetail.htm?locale=en_US&partNumber=WAT023501 light or equivalent). Millipore "Microcon":<http://www.millipore.com/catalogue/item/679> spin-filter device with a 30-kDa molecular weight cut-off. LC-MS/MS instrument. Liquid nitrogen. pH test strips range 6-9 (e.g. from "Merck":http://www.emd-chemicals.ca/ph-indicator-strips-non-bleeding/c_ViKb.s1L.CAAAAEWmOlfVhTI). Protein concentration determination assay. SDS-PAGE apparatus and power supply. Table-top centrifuge accommodating 2 mL reaction tubes. Vacuum evaporation system (commonly referred to as a "speed-vac").

Procedure

****Cell culture and protein collection:**** TAILS is based on the quantitative comparison of N-terminal peptides from protease-treated and control samples. The following protocol developed for studies of cell-conditioned medium proteins (secretome) can be easily adapted to cell lysates or samples derived from other sources. The introduction of the protease of interest, its inhibition or silencing can be done at the cellular level prior to proteome collection or in vitro after the proteome has been harvested. The latter requires collection under conditions to maintain the native structure of the constituent proteins. The minimal recommended protein amount is 100 µg for each sample (i.e. 100 µg for control and 100 µg for protease-treated), and can generally be achieved by collecting serum free condition medium from at least 6 cell culture flasks at around 70-80% confluence (175 cm², T175). 1) Grow cells in appropriate media up to 70% confluence. 2) Decant media and wash cells extensively (at least 3 times) with PBS to remove serum proteins. 3) Add serum-free media (i.e. the same medium used for growing the cells but without the addition of serum), usually 20 mL per T175 flask. 4) Grow cells overnight to synchronize the cells. 5) Decant media and wash cells at least 3 times with PBS. 6) Add fresh serum-free, phenol-free media. By using a lower amount of medium than for normal cell culture the secreted proteins will be more concentrated. The time of addition is set as the starting time. 7) Grow cells for the required time usually 24 h depending on the requirements of the experiment and tolerance of cells to serum free conditions. **_Note_:** After 24 h, serum starvation might occur. If cells are grown for shorter times larger number of flasks will be required to accrue sufficient quantities of protein in the medium. 8) Collect conditioned media in 50 mL tubes (i.e. 2 flasks per 50 mL tube). 9) Centrifuge conditioned media at 2,200g at 4 °C for 5 min to remove any cells. 10) Add protease inhibitors such as PMSF (1 mM final), EDTA (1 mM final), E64 according to the experimental question being addressed. For a complete list of protease inhibitors for all classes of proteases see reference 27. It is important to minimize any background proteolysis that inevitably occurs in all proteome samples after collection. **_Note_:** Excess and reversible protease inhibitors will be removed in the following steps by dialysis, however when the protease of interest is added **_in vitro_** after secretome collection, inhibitors of that protease should be avoided. 11) Filter supernatant using Millipore "Steriflip":<http://www.millipore.com/catalogue/module/c3238> or equivalent. **_Pause_:** At this point it is possible to freeze the samples in liquid nitrogen and store at -80 °C. 12) Apply the protein samples to protein concentration devices such as Millipore "Amicon-Ultra 15":<http://www.millipore.com/catalogue/module/c7715> concentrators. Concentrate condition medium proteins at 4 °C following the manufacture instructions to ≈ 1 mL volume. **_Note_:** To minimize the time for this and the following steps it is recommended to use several concentrators for treating each sample (i.e. one concentrator per 40 mL of collected conditioned medium proteins). 13) Add 14 mL of desired buffer to each concentrator. We recommend using 100 mM HEPES pH 7.0. **_Note_:** TAILS is based on the labeling of peptide primary amines and thus, other molecules with primary amines will interfere with the labeling step resulting in incomplete labeling of peptides. Thus ****primary amine containing buffers**** such as ammonium bicarbonate or Tris ****must not be used****. The purpose of the following buffer-exchange steps is to deplete the sample of free amino acids and other compounds with primary amines. If the protease of interest is added **_in vitro_** after secretome collection, the buffer of choice, pH and other additives should allow the optimal activity of the studied protease. It is also recommended to exclude any detergents as these can interfere with MS later. 14) Concentrate sample again to a volume of ≈ 1 mL or less. 15) Repeat steps 13 and 14 at least 3 times. 16) **_Optional_:** if several concentrators were used for each sample, pool all concentrates and concentrate them to final volume of ≈ 1 mL. Carefully recover as much protein as possible from each concentrator by gentle pipetting of the sample over the membrane before removal. 17) Measure protein concentration using your method of choice e.g. BCA or Bradford Assay. 18) Bring protein concentration to ≈ 1 mg/mL using buffer of choice (see section 13 above). 19) Keep a small aliquot of each sample (control and protease-treated) for quality control purposes designated "before labeling". **_Time taken_:** Cell culture: 1-7 days (depending on cells grown and required amounts). Media collection: 1 hour. Media concentration and buffer exchange: 3-6 hours (depending on sample volume and protein concentration). ****Optional:** Test protease cleavage of collected proteome ****** The following steps are required only if the proteome is exposed to the protease of interest in vitro. 1) Divide the proteome in two equal aliquots. **Optional:** prior to this, add a known substrate of the test protease that will serve as a positive control and allow for validation of the proteolytic activity and sensitivity of the TAILS procedure. If possible select a known substrate of different species than the test proteome (e.g. murine if the proteome is human), in

particular a substrate protein that has a different tryptic peptide spanning the cleavage site to avoid ambiguous identification if the source secretome also has the protein. Typically 0.5-1 µg of known substrate can be added to 200 µg proteome. 2) Add activated protease to the sample and an equivalent amount of buffer to the control sample. Typical protease to proteome ratios are 1:1000-1:50 (w/w) with 1:100 (w/w) being a useful ratio to be used for the first time. This will likely ensure that cleaved neo-N-terminal peptides can be identified. With experience the ratios of the protease to proteome can be reduced. 3) Incubate for 1-24 h at a temperature suitable for the protease under investigation. 4) Optional: heat inactivate the protease and control samples. Time taken - up to 24 hours depending on the selected assay conditions. ****Isotopic labeling of samples**** TAILS is based on isotopic labeling of the primary amines at protein N-termini and lysine side chains. Therefore any primary amine reactive isotopic reagent can be used. We chose to label the proteins by dimethylation using $^{12}\text{CH}_2$ -formaldehyde (light) and $^{13}\text{C}_2\text{H}_2$ -formaldehyde (heavy) and sodium cyanoborohydride (NaBH_3CN , ALD reagent) as the catalyst²⁸. This labeling approach is very fast, robust and efficient and utilizes relatively cheap reagents (≈\$1/labeling reaction). The labeling procedure must be carried out separately for the control and protease-treated sample. 1) Denature protein samples by adding 8 M GuHCl to a final concentration of 4 M GuHCl. Note: The labeling reaction can be carried out efficiently at lower denaturant concentrations but will require more time. Do not use urea which will modify amino acid residues in the sample and so reduce peptide identifications. 2) Check pH by pipetting 1 µL of sample onto a pH strip. Hamilton microcapillary tubes can also be used for 100 nL volumes. 3) Adjust pH to 7.0 by addition of small volumes of 1 N HCl or 1 N NaOH. 4) Reduce cysteine residues by adding 1.0 M DTT to a final concentration of 5 mM. 5) Incubate sample at 65 °C for 1 hour. 6) Cool samples to room temperature. Caution: Cooling down sample temperature prior to addition of iodoacetamide (IAA) is essential for preventing lysine modification by IAA²⁹. 7) Alkylate cysteines by adding 0.5 M IAA to final concentration of 10 mM. 8) Incubate sample at 25 °C in the dark for 30 minutes. 9) Quench excess IAA by adding 1.0 M DTT to a final concentration of 30 mM DTT. 10) Incubate for 30 minutes to achieve full quenching of IAA. Caution: Formaldehyde and sodium cyanoborohydride (ALD reagent) are extremely toxic and carcinogenic reagents. Therefore extra care should be taken while handling these. In addition, during the reductive dimethylation reaction lethal hydrogen cyanide gas is emitted. Therefore, all of the following steps should be performed in a fume hood. 11) Prepare 2.0 M working stocks of $^{13}\text{C}_2\text{H}_2$ -formaldehyde (heavy) and $^{12}\text{C}^1\text{H}_2$ -formaldehyde (light) in water. Note: The concentrations of the light and heavy formaldehyde stock solutions as supplied by the manufacturer are different: 37% or 12.3 M, and 20% or 6.6 M, respectively. 12) Add light formaldehyde to one sample (control) and heavy formaldehyde to the protease sample to a final concentration of 40 mM light/heavy formaldehyde. Note: If the experiment is repeated several times, labeling swaps are recommended for validation. By convention, heavy labels are used for the protease sample. 13) Add ALD reagent to each sample to a final concentration of 20 mM. 14) Vortex samples and adjust pH to 6-7 if required by adding 1.0 N HCl or 1.0 N NaOH (check pH as in step 2). 15) Incubate for at least 4 hours at 37 °C, but overnight is recommended. 16) Quench excess formaldehyde by adding 1.0 M ammonium bicarbonate to each sample up to a final concentration of 100 mM. 17) Vortex samples and check pH (as in step 2). If required adjust pH to 6-7 by adding small volumes of 1.0 N HCl or 1.0 N NaOH. 18) Incubate for at least 4 hours at 37 °C. 19) Keep a small aliquot (1-5%) of each sample for labeling validation (for troubleshooting purposes); label samples as "heavy" and "light". Suggestion: A fast, labeling test can be performed by analyzing the non-labeled (start) and light and heavy labeled samples by MALDI-TOF MS. Successful labeling is characterized by a complete shift by 26 m/z of the observed peaks in the labeled samples compared to the nonlabeled samples. Time taken - 9 to 24 hours. ****Proteolytic digestion of labeled samples**** The sample must be tryptic digested to prepare the sample for proteomics analysis. As will be described below and as in many other proteomics procedures, it is highly recommended to use trypsin for this purpose, although GluC or chymotrypsin can also be used. Following primary amine dimethylation, trypsin cannot cleave at the blocked lysines and so will cleave with ArgC specificity (cleaving C-terminal to arginine residues only). This generates longer peptides which significantly improves the likelihood of identifying neo-N-terminal peptides that have otherwise been shortened by proteolysis²⁶. If GluC or chymotrypsin are used this advantage is lost. Labeling reagents clean-up 1) Combine quenched heavy and light labeled samples in a 15 mL tube. 2) Keep small aliquot for quality control and label it "labeled samples before precipitation". 3) Add 8 sample volumes of ice-cold acetone and 1 sample volume of methanol to the labeled proteins. Caution: acetone and methanol should be stored in chemical resistant containers (i.e. glass bottles). The use of plastic-ware for storage can result in the extraction of contaminating plastic polymers into the solvent that will affect MS results. 4) Aliquot 1.2 mL sample into 1.5 mL microfuge tubes (unless a centrifuge capable of 15,000g for 15 mL tubes is available, in which case continue using the 15 mL tubes). 5) Precipitate labeled proteins for at least 4 hours at -80 °C, but overnight is recommended. 6) Centrifuge the samples at 14,000g in 4 °C for 10 min and carefully discard the supernatant. 7) Add 1 mL of ice-cold methanol to each tube (or 5 mL if 15 mL tube is used). Note: Washing the acetone pellet with methanol prevents unwanted acetylation of the N-termini of tryptic peptides in case of any NaCNBH_3 carry-over. 8) Centrifuge the samples at 14,000g in 4 °C for 10 min and carefully discard the supernatant. 9) Repeat steps 7 and 8. 10) Air-dry the sample. Note: Do not overdry the sample as it will be difficult to dissolve. 11) Resuspend samples 8 M GuHCl. For 1.5 mL it is recommended to start with 20 µL and increase volume if required. Use the minimal volume required to completely resuspend the sample. 12) Add 9 volumes of 50 mM HEPES pH 8.0 to each tube (i.e. 180 µL if 20 µL of 8 M GuHCl was used) and combine resuspensions from all tubes. Following this the final concentration of GuHCl should not exceed 0.75 M, which is suitable for trypsin digestion. If another protease is used, the final GuHCl should be adjusted accordingly. 13) Keep a small aliquot (1% of total or 1 µL if testing by MALDI-TOF) for quality control and label "labeled samples after precipitation". Time taken - 6 to 24 hours. Trypsin digest 1) Check pH and if required adjust to pH 8.0 by adding small volumes of 1.0 N HCl or 1.0 N NaOH. 2) Add mass spectrometry grade trypsin to a final ratio of 1:50 protease/protein (i.e. 4 µg trypsin per 200 µg sample) and gently pipette up and down to mix sample. 3) Incubate overnight (18 h) at 37 °C. 4) Optional: add additional trypsin as in step 2 and incubate for an additional 4 hours at 37 °C to ensure complete digestion. 5) Keep a small aliquot for quality control and label it "labeled samples after digestion". Optional: It is highly recommend to plan ahead and prepare sufficient starting material that will allow MS analysis of the sample prior to polymer negative selection. If the sample amounts permit such analysis, an aliquot should be stored for this purpose at this point. Time taken - 18 to 24 hours. Quality control for labeling and digestion To verify the successful completion of the above steps, run a 10% SDS-PAGE gel followed by silver staining of the aliquots that were stored in the previous steps: labeled samples before precipitation, labeled samples after precipitation, and labeled samples after digestion. Ensure similar protein bands and intensities appear before and after precipitation and that there is disappearance of all bands higher than 10 kDa after proteolytic digestion. Mismatching protein bands before and after precipitation indicates sample losses that will reduce the quality of the MS analysis. Protein bands after tryptic digestion may indicate incomplete digestion (e.g. due to a bad protease batch) and require repetition of the digestion step. Time taken - 4 hours. ****Negative selection of blocked peptides using HPG-ALD polymer and MS/MS readout**** This step enriches the naturally blocked as well as the dimethylated and labeled N-terminome peptides by negative selection.

In the previous steps, protein original free N-termini and the protease-generated neo-N-termini were dimethylated, so together with naturally blocked N-termini (e.g. by acetylation and cyclization) of proteins, they all possess blocked N-termini. Trypsin digestion generated internal peptides with free N-termini. The HPG-ALD polymers developed for TAILS contain many aldehyde functional groups that readily react and bind the free N-terminal internal tryptic and C-terminal peptides when mixed with the digested sample in the presence of sodium cyanoborohydride. In contrast, the naturally blocked and isotopically-labeled mature N-terminal and neo-N-terminal peptides (and dimethylated lysines) are unreactive and will remain unbound for recovery by ultrafiltration. **Polymer preparation:** "HPG-ALD": <http://www.flintbox.ca/technology.asp?page=3081> polymer usually supplied at a concentration of ≈ 35 mg/ml. Although the polymer was dialyzed extensively it is recommended to dialyze again prior to use. 1) Dialyze 0.5 ml of HPG-ALDII polymer against 4 L of water overnight at room temperature with agitation. 2) Split HPG-ALDII stock into 20 μ L aliquots in microfuge tubes. 3) Flow argon gas on top of the liquid for 1 minute per each tube. **Caution:** do not use strong gas flow as it will cause the solution to splash out of the tube. 4) Close the microfuge tubes and freeze the polymer solution in liquid nitrogen. Store the polymer at -80 °C. These aliquots are ready to be used for experiments. **Note:** If the polymer solution is frozen other than by liquid nitrogen, a gel-like, opaque solution will be formed upon thawing that will require about one hour to form a clear, usable solution. **Polymer negative selection:** HPG-ALDII polymer has a binding capacity of 2.5 mg of peptide per mg of polymer. However, there are different versions of the HPG-ALD polymer with different binding capacity²⁶. Therefore, if a different version of HPG-ALD is used, the amount of polymer for the capture should be modified accordingly. Removal of the internal tryptic peptides results in a 90% decrease of the total peptide content, thus a maximum of ≈ 20 μ g of peptides can be recovered in the N-terminal enriched sample. The peptide content is in fact lower (around 10 μ g) due to sample loss through the different steps. 1) Add dialyzed HPG-ALDII to the trypsinized sample. We recommend capturing 100 μ g of peptides with 200 μ g of HPG-ALDII, representing a 5-fold excess of polymer. Therefore, if the polymer solution concentration is 35 mg/mL, 10 μ L of polymer stock should be added per 100 μ g of tryptic digest. 2) Add ALD reagent to final concentration of 20 mM. 3) Check pH and if required adjust to pH 6-7 range by adding small volumes of 1.0 N HCl or 1.0 N NaOH. 4) Incubate overnight at 37°C. 5) Add 1.0 M ammonium bicarbonate to 100 mM final concentration. **Note:** This step is used for blocking the excess functional aldehyde groups of the polymer which improves yield and reduces non-specific binding of peptides to the polymer. 6) Check pH and if required adjust to pH 6-7 by adding small volumes of 1.0 N HCl or 1.0 N NaOH. 7) Incubate at 37 °C for 30 minutes. **Time taken:** 6 to 12 hours. **Recovery of unbound blocked and labeled peptides:** 1) Pretreat a 10-kDa molecular cutoff Microcon spin-filter (Millipore) with 400 μ L of water as per manufacturer's instructions. Do not allow membrane to dry out before sample addition. 2) Load the tryptic digest/polymer reaction mixture. 3) Filter by centrifugation at 14,000 rpm for 15 min. 4) Monitor the sample volume above the filter and centrifuge until there remains just a few μ L on the filter. 5) Collect the filtrate, which contains the enriched N-terminal peptides. The internal tryptic peptides that are covalently bound to the polymer are retained by the filter. 6) Wash the filter by adding 200 μ L of 100 mM ammonium bicarbonate buffer and centrifuge again. 7) Collect the filtrate and combine it with the filtrate of Step 5. **Time taken:** up to 3 hours. **Desalting of blocked and labeled peptide solution:** This step is performed utilizing a C18 reverse-phase solid phase extraction cartridge. We used Waters Sep-Pak light, which accommodates a relatively small volume with a high binding capacity thus providing convenient sample concentration at this step. 1) Acidify the pooled filtrates (obtained at step 7 above) to pH 3 by adding formic acid and dilute to 3 mL 0.1% formic acid in water. 2) Condition a Sep-Pack light C18 cartridge by injecting 5 mL of 80% acetonitrile, 20% water, 0.5% formic acid with a syringe. 3) Discard the flow-through. **Caution:** Do not dry the cartridge by introducing air at the end of the injection. Always keep the cartridge wet. 4) Rinse the Sep-Pack light C18 cartridge with 5 mL of water with 0.1% formic acid and discard the flow-through. 5) Apply the sample to the cartridge at a maximum of 1 mL/min and collect the flow-through. Note: Measure the flow with a timer using the syringe volume marks. 6) Reapply the sample to the cartridge to improve peptide binding and recovery. 7) Wash the Sep-Pack light C18 cartridge twice with 5 mL 0.1% formic acid in water and discard the flow through. 8) Elute peptides with 1.5 mL of 80% acetonitrile, 20% water, 0.5% formic acid at a maximum of 1 mL/min. Collect the eluate into a microfuge tube. 9) Evaporate the eluate organic solvent under vacuum (using a speedvac). **Caution:** Do not dry completely. 10) Resuspend the peptides in 20 μ L of 3% acetonitrile, 97% water, 0.1% formic acid. Store the samples at -80 °C until mass spectrometry analysis. **Time taken:** 3 hours (strongly depend on speedvac speed). **Identification of N-terminal Peptides by Liquid Chromatography-Tandem Mass Spectrometry** TAILS-enriched N-terminal peptides have been analyzed on quadrupole-time of flight QSTAR (ABI) and an LTQ-Orbitrap (ThermoFisher) mass spectrometers, but can be analyzed on any tandem mass spectrometer. An LTQ-Orbitrap mass spectrometer is preferred because of its fast duty cycle time and high mass accuracy. Using the Orbitrap, TAILS data coverage has proven excellent without sample prefractionation due to the massive sample simplification achieved following removal of the internal tryptic peptides. However, higher coverage and potentially better quantification accuracy can be obtained using a 2D peptide separation system following TAILS, such as strong cation exchange chromatography to generate 10 fractions, with each then being loaded separately on the mass spectrometer. A description of the LC-MS/MS setup is not within the scope of this protocol so here we outline the conditions used for the LTQ-Orbitrap TAILS analysis²⁶ only briefly. These steps can be easily adapted to other mass spectrometers. 1) Load peptides onto a C18 reverse-phase (3 μ m ReproSil Pur C18 beads) capillary column (15 cm, 75 μ m inner diameter fused silica emitter with a 8 mm diameter opening) with a nanoflow HPLC in-line with the mass spectrometer as described. **Optional:** If required, prior to loading the column, desalt peptide sample using STop And Go Extraction (STAGE) tips³⁰. 2) Elute the peptides from the reverse-phase column with a gradient composed of Buffer A (0.5% acetic acid) and Buffer B (0.5% acetic acid and 80% acetonitrile) and inject it directly into the mass spectrometer by ion-spray ionization. The gradient is formed with 6 to 30% Buffer B in 60 min, then from 30 to 80% Buffer B in 10 min and held at 80% of Buffer B for 5 min. 3) Acquire MS1 scans between 350 and 1,500 m/z at a resolution of 60,000 and select the five most intense ions for fragmentation. Repeat this cycle for the period of the gradient. **Time taken:** 2-3 hours. **Data Analysis of the TAILS Tandem Mass Spectroscopy Spectra** Unlike most proteomics procedures, the successful outcome of TAILS negative selection is the generation and isolation of peptide "single hits" where a protein identification is based on a single peptide (i.e. only the original or neo-N-terminal peptide of each protein)³². To address this issue we set robust statistical and bioinformatics criteria. We collect 3 different biological samples for analysis that are treated and analyzed independently. High quality and accurate MS data are acquired using a Fourier-transform mass spectrometer. For the identification of protease generated neo-N-terminal peptides, database searches are performed utilizing 2 search engines. We use Mascot and X! Tandem. The search parameters include N-terminal and lysine dimethylation modifications. Data analysis depends on LC-MS/MS vendor-specific data formats and the choice and access to specific analysis software. We chose to analyze the data using the open source Trans-Proteomics Pipeline (TPP) software from the Systems Biology Institute in Seattle³¹, which allows input from different mass spectrometers and can incorporate MS/MS search results from different search engines including free and open-source engines such as X! Tandem and Omssa as well as the commonly used commercial programs Sequest and Mascot. For peptide identification we use an orthogonal validation strategy whereby the N-terminus must be dimethylated and the peptide must be within 5 p.p.m. of the expected m/z. Next, the peptides must pass

PeptideProphet in the TPP with a false discovery rate of less than 1%. After high confidence peptide identification and protein identification the protease substrates are then identified by a process we term hierarchical substrate winnowing²⁶. The relative quantification (protease-treated vs control) of each peptide found in the database searches is analyzed using the XPRESS quantification tool included in the TPP. In this process high confidence cleavage sites must have a peptide abundance ratio (heavy/light, protease/control) >3. Such peptides are compiled for each of the three biological samples from the technical replicates and further validated through the TPP to generate a list of substrate candidates identified by single peptides in multiple biological samples. Only the following high ratio peptides are then selected as high confidence potential protease-generated peptides that are identified in more than one biological replicate sample, or in multiple charge states, or with and without oxidized methionine, or with and without arginine or glutamine deamidation. The rationale is that a peptide must be identified twice: either the same peptide in different samples or in different states in the same sample. It should be noted that a less stringent validation could be used when a protease with known and narrow cleavage specificity is being examined. The double identified, high confidence, high ratio neo-N-terminal peptides are then further examined to select the most biologically relevant candidate substrates for the tested protease. For proteases belonging to a family, candidate substrates can be winnowed from the candidates that are known to be cleaved by other proteases in the family, or are protein family members of a substrate cleaved by the protease. The identification of protein naturally blocked N-terminal peptides is performed by simply altering the database search parameters (i.e. including peptide N-terminal acetylation or cyclization instead of dimethylation) and by utilizing protein database annotations. We provide below examples based on data originating from a Thermo LTQ-Orbitrap™ instrument (".RAW" file format) analyzed by Mascot³³ and X! Tandem³⁴ for uninterpreted database searches. The steps are described briefly and detailed information regarding TPP usage can be found on the "TPP wiki":http://tools.proteomecenter.org/wiki/index.php?title=Main_Page, "tutorial":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tutorial, "TPP users discussion list":<http://groups.google.com/group/spctools-discuss?pli=1>, and in the links provided below. **_Time taken_** - hours to days. **_File conversions_**: 1) Make a directory for each analysis (Mascot protease cleavage, X! Tandem protease cleavage, Mascot N-terminal and so on). 2) Convert LTQ-Orbitrap RAW data to mzXML format in profile mode (not centroid) using "ReAdW":<http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW> tool in the TPP. **_Note_**: it is possible to convert to mzXML format, which is the new standard MS format set by HUPO - but the usage of this format for TAILS data analysis has not been tested yet. 3) Place 2 copies of the mzXML file in each directory and name differently for heavy and light. If required (see more details below on Mascot database search section). Note: The TPP requires that the mzXML and all other files generated during the analysis of a data set are kept in the same directory. If possible semi-links/shortcuts can be used to avoid storage of multiple copies of the mzXML files. 4) For Mascot searches convert the mzXML files to Mascot generic format (.mgf extension) using MzXML2Search tool in the TPP (use defaults). **_Mascot database search analysis_**: TPP quantitative analysis of "Mascot":http://tools.proteomecenter.org/wiki/index.php?title=TPP:Mascot_and_the_TPP search data for dimethylation requires running 2 separate searches, one for only heavy labeled peptides and one for only light labeled peptides. We recommend the use of decoy sequences in the searched database in order to improve peptide assignment validation in later steps of the analysis³⁵. For more information about decoy sequences, their generation and implantation in the database see "Mascot help":http://www.matrixscience.com/help/decoy_help.html. 1) Run a Mascot search for only light dimethylated peptides using the mgf file as input against an appropriate database using the following search parameters: Semi-ArgC cleavage specificity; up to 3 missed cleavages; precursor ion mass tolerance 10 ppm; fragment mass tolerance 0.8 Da; fixed modifications: cysteine carbamidomethylation (+57.021464), peptide N-terminal and lysine dimethylation (+28.031300); variable modification: methionine oxidation (+15.994915); scoring scheme ESI-TRAP. 2) Repeat the search for heavy dimethylated peptides by changing the settings of fixed modification of peptide N-terminal and lysine residue to heavy dimethylation (+34.063117). 3) Import the search result files (.dat extension) from the Mascot server to the analysis directory and name them according to the input file name (i.e. data1.dat for data1.mgf and so on). 4) Convert the .dat files to pepXML file using "Mascot2XML":<http://tools.proteomecenter.org/wiki/index.php?title=Software:Mascot2XML> tool. Do it for both the light and heavy labeled searches. Merge the heavy light search results, analyze and validate peptide MS/MS identifications and quantification using the "XInteract and XPRESS":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tutorial#Peptide_Level_Analysis, "PeptideProphet":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tutorial#Peptide_Level_Analysis tools of the TPP (respectively). The output of this step is an interact pepXML file that includes all the peptides and their related information (database search score, PeptideProphet score, relative abundance etc.). **_Note_**: All of these steps can be executed in a single step though the TPP "Petunia":http://tools.proteomecenter.org/wiki/index.php?title=TPP:Using_Petunia_interface (GUI). **_X! Tandem database search analysis_**: Quantitative analysis of "X! Tandem":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tandem_search searches for dimethylated peptides with the TPP does not require use of separate searches for heavy and light labeled peptides, but for simplification we will use the same search approach. In order to combine X! Tandem search results with Mascot results it is important to use the same database for both searches. 1) Perform an X! Tandem database search for light-labeled peptides directly from the "Petunia":http://tools.proteomecenter.org/wiki/index.php?title=TPP:Using_Petunia_interface of the TPP using the "k-score":http://tools.proteomecenter.org/wiki/index.php?title=TPP:X!Tandem_and_the_TPP option (X! Tandem searches are done with mzXML as input). Using the same parameters used for Mascot light search: Semi-ArgC cleavage specificity; up to 3 missed cleavages; precursor ion mass tolerance 10 ppm; fragment mass tolerance 0.8 Da; fixed modifications: cysteine carbamidomethylation (+57.021464), peptide N-terminal and lysine dimethylation (+28.031300); variable modification: methionine oxidation (+15.994915). The X! Tandem search will generate a .xml file with the search results data in the same directory of the mzXML file used for the search. **_Note_**: Running X! Tandem through the TPP is done by using an "input file":<http://www.thegpm.org/TANDEM/api/index.html> with the required parameters for the search. 2) Repeat the search for heavy dimethylated peptides by changing the settings of fixed modification of peptide N-terminal and lysine residue to heavy dimethylation (+34.063117). 3) Convert the .xml files to pepXML using "Tandem2XML":<http://tools.proteomecenter.org/wiki/index.php?title=Software:Tandem2XML> tool for both the light and heavy labeled searches. 4) Merge the heavy and light search results, analyze and validate peptide MS/MS identifications and quantification using the "XInteract":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tutorial#Peptide_Level_Analysis, "PeptideProphet and XPRESS":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tutorial#Peptide_Level_Analysis tools of the TPP (respectively). The output of this step is an interact pepXML file that includes all the peptides and their related information (database search score, PeptideProphet score relative abundance and so on). **_Note_**: All of these steps can be executed in a single step though the TPP "Petunia interface":http://tools.proteomecenter.org/wiki/index.php?title=TPP:Using_Petunia_interface (GUI). **_Data validation and selection of potential protease generated hits_**: 1) Combine the pepXML interact files of Mascot and X! Tandem using the "iProphet":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Demo2009#8_Further_peptide-level_validation_iProphet tool of the

TPP. This will generate a pepXML file with a combined list of identified and quantified peptides. 2) Open the resulting iProphet pepXML file (using pepXML viewer). 3) Determine PeptideProbability score that corresponds to a false discovery rate of 1% by using the "calculate stats" option under "other options" tab. 4) Select peptides with a PeptideProbability score above the score found in step 2, corresponding to a false discovery rate of 1%. 5) Manually verify the quantification data (extracted ion chromatograms from "XPRESS":http://tools.proteomecenter.org/wiki/index.php?title=TPP_Tutorial#XPRESS_Results) of peptides with undefined ratios and heavy and light singletons and correct if required. 6) Export the final list of peptides to a Microsoft Excel sheet using the "export spreadsheet" option under "other options" tab. 7) Select peptides with high (>3) and low ratios (<0.33), which are considered as potential substrates of the protease of interest. 8) Perform these steps separately for all biological repeats analyzed by TAILS. 9) Compare the high and low peptide lists obtained in steps 7 and 8 and select only the peptides appearing in two biological samples or that were identified by two different tandem mass spectra (different charge states, different labels, oxidized and non-oxidized methionine). 10) Optional: as extra validation select only high and low peptides with precursor mass error <5 ppm (experimental v.s. theoretical). Analysis of natural N-termini of proteins: Analysis of natural N-terminal peptides of proteins using TAILS requires only changing the database search parameters. For analysis of acetylated peptides the search parameters listed above should be changed by replacing the fixed modification on peptide N-termini from dimethylation (+28.031300 or +34.063117) to acetylation (+42.010565). The rest of the parameters should remain the same. It was suggested that for general N-terminome mapping aiming at enrichment and identification of proteins natural N-terminal peptides, the validation of "single hits" could also rely on the actual position of the identified peptide within the protein sequence^{13,15}. Obtaining peptide positional information is dependent on the quality of the database annotation. To assist in obtaining the positional information of identified peptide we recommend the usage of a documented Perl script developed in our lab that can be found at ["www.clip.ubc.ca/resources/index.html"](http://www.clip.ubc.ca/resources/index.html):<http://www.clip.ubc.ca/resources/index.html>, under CLIP-PICS³⁶.

Timing

Highly variable depending on cell lines and data analysis time. About 5 days including a 24 h cell culture time in serum-free media.

Critical Steps

Do not use buffers or reagents with primary amines prior to completion of labeling and the last step of polymer negative selection. Monitor and adjust pH when indicated. Cool samples to room temperature before alkylating cysteine with IAA. Do not use urea for sample resolubilization. Quench completely excess IAA with DTT. Quench completely excess labeling reagent with ammonium bicarbonate prior to mixing heavy and light labeled samples and tryptic digest. Block polymer functional groups with ammonium bicarbonate after completion of binding of the internal tryptic and C-terminal peptides to the polymer in the negative selection step of blocked peptides. Perform all labeling steps in fume hood.

Troubleshooting

TAILS is a highly optimized procedure. Many of the steps are performed to be mass spectrometer compatible at later stages. Do not vary the parameters and conditions until satisfactory results are obtained. If very few peptides are identified by the database search: Verify the quality of the MS data (mzXML) using "Pep3D":<http://tools.proteomecenter.org/wiki/index.php?title=Software:Pep3D>. Change database search parameters to include potential modifications. Validate the tested protease activity (as suggested in "Test protease cleavage of collected proteome"). Increase the protease:proteome ratio or incubation time. Verify labeling by analyzing initial labeled samples (subsection 19 in "isotopic labeling"). Use protease of known specificity such as GluC or caspase to set the analysis conditions before using the test protease.

Anticipated Results

Anywhere from 100 to 1000s of cleavage sites for the tested protease and ~ 100-1000 original free and blocked N-terminal will be identified. This depends on the size and complexity of the tested proteome and activity and presence of substrates of the test protease.

References

1. Meinnel, T., Serero, A. & Giglione, C. Impact of the N-terminal amino acid on targeted protein degradation. *Biol. Chem.* **387**, 839-851 (2006).
2. Brown, J.L. & Roberts, W.K. Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are N-alpha-acetylated. *J. Biol. Chem.* **251**, 1009-1014 (1976).
3. Overall, C.M. & Blobel, C.P. In search of partners: linking extracellular proteases to substrates. *Nat. Rev. Mol. Cell Biol.* **8**, 245-257 (2007).
4. Doucet, A., Butler, G.S., Rodriguez, D., Prudova, A. & Overall, C.M. Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol. Cell. Proteomics* **7**, 1925-1951 (2008).
5. Plevoda, B. & Sherman, F. N. alpha-terminal acetylation of eukaryotic proteins. *J. Biol. Chem.* **275**, 36479-36482 (2000).
6. Overall, C.M. Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol. Biotechnol.* **22**, 51-86 (2002).
7. McQuibban, G.A. et al. Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3. *Science* **289**, 1202-1206 (2000).
8. Vergote, D. et al. Proteolytic processing of SDF-1alpha reveals a change in receptor specificity mediating HIV-associated neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19182-19187 (2006).
9. Puente, X.S., Sanchez, L.M., Overall, C.M. & Lopez-Otin, C. Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4**, 544-558 (2003).
10. Turk, B. Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* **5**, 785-799 (2006).
11. Lopez-Otin, C. & Overall, C.M. Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell Biol.* **3**, 509-519 (2002).
12. Gevaert, K. et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* **21**, 566-569 (2003).
13. McDonald, L., Robertson, D.H., Hurst, J.L. & Beynon, R.J. Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat. Methods* **2**, 955-957 (2005).
14. Kuhn, K. et al. Isolation of N-terminal protein sequence tags from cyanogen bromide cleaved

proteins as a novel approach to investigate hydrophobic proteins. *J. Proteome Res.* **2**, 598-609 (2003). 15. McDonald, L. & Beynon, R.J. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat. Protoc.* **1**, 1790-1798 (2006). 16. Mahrus, S. et al. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **134**, 866-876 (2008). 17. Ji, C., Guo, N. & Li, L. Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. *J. Proteome Res.* **4**, 2099-2108 (2005). 18. Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J. & Heck, A.J. Targeted analysis of protein termini. *J. Proteome Res.* **6**, 4634-4645 (2007). 19. Schilling, O. & Overall, C.M. Proteomic discovery of protease substrates. *Curr. Opin. Chem. Biol.* **11**, 36-45 (2007). 20. Timmer, J.C. et al. Profiling constitutive proteolytic events in vivo. *Biochem. J.* **407**, 41-48 (2007). 21. Enoksson, M. et al. Identification of proteolytic cleavage sites by quantitative proteomics. *J. Proteome Res.* **6**, 2850-2858 (2007). 22. Van Damme, P. et al. Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nat. Methods* **2**, 771-777 (2005). 23. Van Damme, P. et al. Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. *Mol. Cell. Proteomics* **8**, 258-272 (2008). 24. Vande Walle, L. et al. Proteome-wide Identification of HtrA2/Omi Substrates. *J. Proteome Res.* **6**, 1006-1015 (2007). 25. Staes, A. et al. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* **8**, 1362-1370 (2008). 26. Kleifeld, O. et al. Proteomic Identification of Protease Cleavage Products from N-Terminome Analysis by Terminal Amine Isotopic Labeling of Substrates. *Nat. Biotechnol.* (2010). 27. Overall, C.M. & Kleifeld, O. Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* **6**, 227-239 (2006). 28. Hsu, J.L., Huang, S.Y., Chow, N.H. & Chen, S.H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843-6852 (2003). 29. Nielsen, M.L. et al. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat. Meth.* **5**, 459-460 (2008). 30. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663-670 (2003). 31. Keller, A., Eng, J., Zhang, N., Li, X.J. & Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **1**, 2005 0017 (2005). 32. Higdon, R. & Kolker, E. A predictive model for identifying proteins by a single peptide match. *Bioinformatics* **23**, 277-280 (2007). 33. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999). 34. Craig, R. & Beavis, R.C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466-1467 (2004). 35. Choi, H. & Nesvizhskii, A.I. Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *J. Proteome Res.* **7**, 254-265 (2008). 36. Schilling, O. & Overall, C.M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.* **26**, 685-694 (2008).

Acknowledgements

O.K. was supported by the Centre for Blood Research (CBR) (University of British Columbia), Canadian Institute for Health Research/Heart and Stroke Foundation of Canada (CIHR/HSFC) Strategic Training Program in Transfusion Science research fellowship. A.D. acknowledges the Fonds Quebecois de la Recherche sur la Nature et les Technologies and the Michael Smith Foundation for Health Research (MSFHR) for research fellowships. J.N.K. is the recipient of a Canadian Blood Services (CBS)/CIHR new investigator award in transfusion science. C.M.O is supported by a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology. This work was supported by grants from the CIHR and from a program project 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 the Canada Foundation for Innovation (CFI) and MSHFR. We thank Dr Wei Chen and the CBR mass spectrometry core facility for proteomics analysis and Dr Georgina Butler for critical reading of the manuscript.

Figures

Figure 1

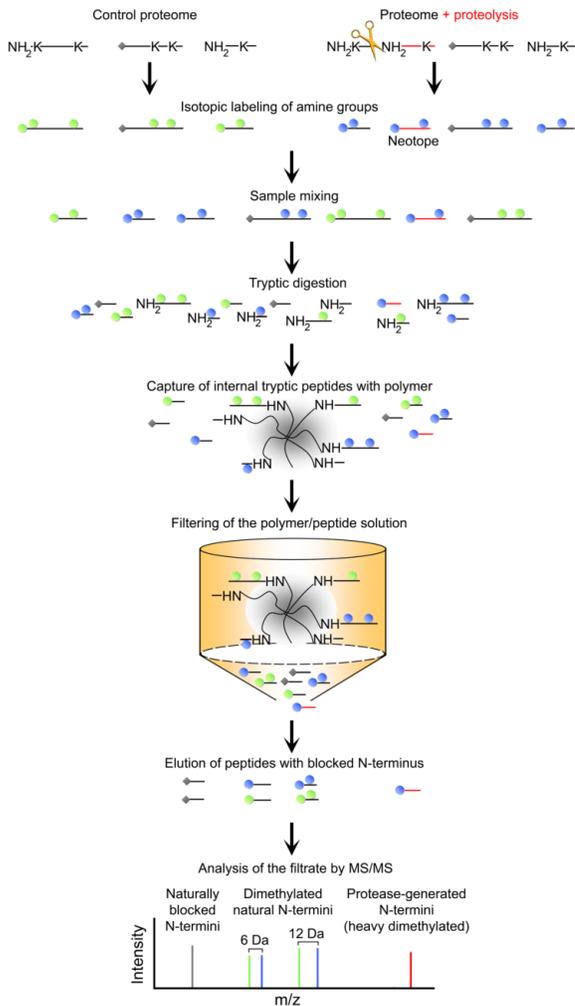


Figure 1

Schematic representation of TAILS workflow. Protein derived from control and protease treated samples are labeled on the primary amines of the peptides N-termini (NH₂) and lysines (K) via dimethylation using (d(0)C¹²)-formaldehyde (L) (green spheres) or (d(2)C¹³)-formaldehyde (H) (blue spheres), respectively. Protease substrate is cleaved (represented by scissors) to generate a prime site fragment (carboxyl) shown in red and a non-prime cleaved fragment. After mixing of protease-treated and control proteins, the sample is digested with trypsin. The newly formed internal tryptic peptides are removed by reaction of the tryptic peptide amino-termini with the amine reactive polymer. Isotopically labeled N-terminal peptides remain in the unbound fraction and are analyzed and quantified by high accuracy LC-MS/MS. Naturally acetylated or cyclized N-terminal peptides are also recovered in the filtrate after separation from the polymer. High protease/control ratio or singletons correspond to protease-generated neo-amino termini (neotope). These can be easily distinguished from background proteolysis products that occur in both samples and from natural N-terminal peptides, which have an isotope ratio centered on 1.0.