Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization

Lucy McDonald & Robert J Beynon

Proteomics and Functional Genomics Group, Faculty of Veterinary Science, University of Liverpool, Crown Street, Liverpool L69 7ZJ, UK. Correspondence should be addressed to R.J.B. (r.beynon@liv.ac.uk).

Published online 16 November 2006; doi:10.1038/nprot.2006.317

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

We describe a protocol for selective extraction of the amino (N)-terminal-most peptide of a protein or a mixture of proteins after proteolysis. The first stage of the protocol blocks the free amino groups α and ε (the latter being lysyl residues) on the intact proteins by acetylation. In the second stage, proteolysis of the acetylated proteins yields a mixture of N-terminally acetylated (true N-terminal) and non-acetylated (internal and carboxy-terminal) peptides. Affinity capture of peptides bearing free amino groups using an immobilized amine-reactive reagent removes internal peptides from the mixture. The unbound fraction is highly enriched in N-terminal peptides, which can be analyzed without further treatment. This method is compatible with a range of proteolytic enzymes and fragmentation methods, and should take 2 d to complete. The N-terminal peptides can then be analyzed by mass spectrometry. This low cost, rapid method is readily adopted using off the shelf reagents.

INTRODUCTION

Although other strategies have been proposed, most proteomics studies are based on 'bottom-up' approaches and employ mass spectrometry (MS) for the analysis of limit proteolytic peptides (i.e., the products of proteolytic digestion of proteins in which all cleavable bonds have been hydrolyzed) that are derived, usually by tryptic hydrolysis, from single proteins or protein mixtures[1]. In all proteomics studies, there is a decision to be made about when to leave 'protein space' and to move into 'peptide space', with the transfer usually being a tryptic digestion (**Fig. 1a**). Most proteomics studies aim to simplify a complex proteome, to the level of subproteomes (e.g., a subcellular fraction) or individual proteins (e.g., proteins separated by 2D gel electrophoresis) before proteolysis. After separation of the proteins by, for example, 2D gel electrophoresis, it is reasonably inferred that the connectivity of the entire set of peptides is such that they are all derived from the same parent protein. However, any mixture of proteins generates a correspondingly mixed set of limit peptides, and the connectivity between different peptides therefore cannot be assumed. This would usually mean that tandem MS (MS/MS) is required, as further information must be extracted from a single peptide. The extreme resolution of MS and of the information obtained by MS/MS can be used to obviate exhaustive protein separation in favor of more global 'shotgun' approaches[2], in which a mixture of proteins (sometimes an entire proteome) is initially proteolyzed after which the complex mixture of peptides is separated before mass-spectrometric analysis, often with direct

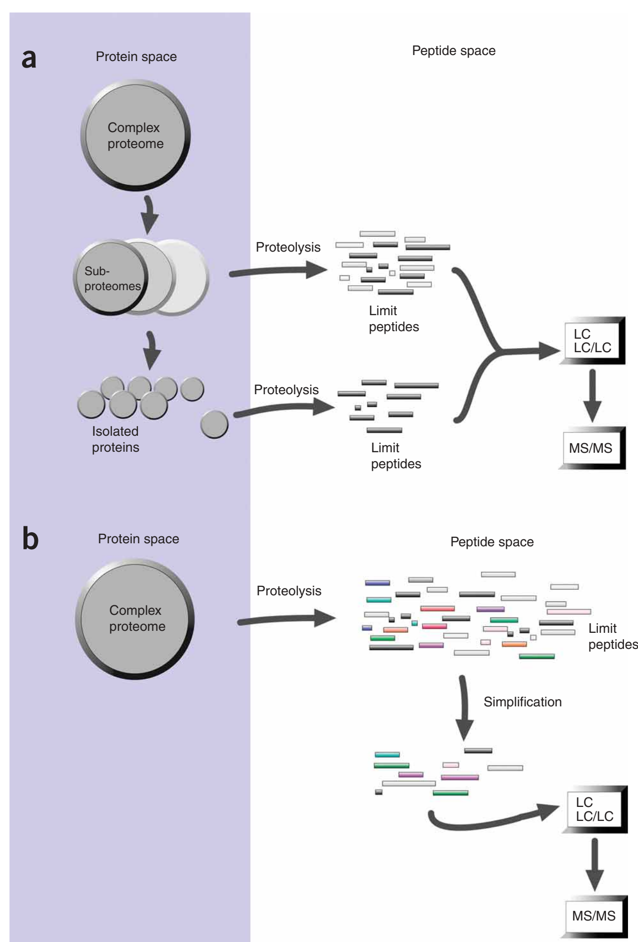

Figure 1 | Outline of a standard approach to protein identification. A complex proteome is simplified using a variety of separation techniques, either in protein space or, after proteolysis, in peptide space. The resulting subproteomes can be either proteolyzed directly or further subjected to separation into individual proteins before proteolysis. The resulting peptides are analyzed by mass spectrometry, with or without chromatographic separation, depending on the complexity of the final peptide mixture. (**a**) Strategy for selective/targeted peptide simplification. Proteolysis of a complete proteome creates a peptide mixture that is so complex that mass-spectrometric analysis is highly challenging. (**b**) By targeting specific structural regions on peptides, the mixture can be selectively purified in such a way that the majority of the proteome is discarded. LC, liquid chromatography.

coupling between the separation fluidics and the mass spectrometer. However, the complete repertoire of limit peptides generated from an entire proteome might contain many hundreds of thousands of peptides, which imposes a formidable analytical challenge. Even with the benefit of 2D chromatography (multidimensional protein identification technology or MUDPIT[3]), the mixture might deliver more peptides to the mass spectrometer than can feasibly be analyzed as the liquid stream flows through the electrospray source. Moreover, the data system is likely to direct MS analysis to more abundant proteins, therefore limiting dynamic range.

As a consequence of this complexity, there have been a number of attempts to simplify peptide mixtures. These approaches target specific chemical reactivities of the proteins or peptides within complex mixtures (**Fig. 1b**). The purpose of such methods is to eliminate the majority of the proteome-derived peptides but retain sufficient information necessary for analysis[4]. Lower abundance amino acids can be targeted as a means of extracting informative peptides. Methods such as isotope-coded affinity tagging (ICAT)[5] implicitly adopt this principle, using selective chemistry that recovers only those peptides that contain at least one cysteinyl residue and at the same time introducing tags for relative quantification. Immobilized metal-affinity chromatography (IMAC)[6], titanium dioxide-affinity chromatography[7], is used for the enrichment of phosphopeptides and lectin-affinity chromatography is used to target glycopeptides[8]. All of these simplification techniques have the potential to abstract more than one peptide for each protein and, in some instances, no peptides will be recovered. Furthermore, because the peptide could have been derived from any part of the parent protein sequence, no information on its location is obtained, which complicates the search strategy for identification proteomics. Finally, combined diagonal chromatography (COFRADIC) methods are able to enrich a subset of peptides according to their chemical composition (e.g., by the presence of methionyl residues) or enrich for amino (N)-terminal peptides[9–11].

An efficient strategy for proteome simplification would be the isolation of a single signature peptide from each protein in the proteome, optimizing the balance between analyte complexity and completeness of representation. As the disposition of protease cleavage sites within a protein is effectively random, the most obvious choice for a signature or proteotypic peptide would be the (amino N) or carboxy (C) terminus. Methods for recovery of C-terminal peptides have been reported, which are predominantly based on the capture of internal tryptic peptides using an anhydrotrypsin column[12,13]. Several strategies have been developed in order to selectively purify N-terminal peptides, including specific N-terminal sequencing by MS of gel-separated and blotted proteins[14], selective modification of N-terminal serine or threonine residues[15], and modification of the hydrophobicity of a peptide mixture to preferentially expose N-terminal peptides by diagonal chromatography[9].

We have previously described a strategy to selectively purify N-terminal peptides[16]. This approach required blocking of available amino groups on the intact proteins by acetylation, followed by proteolysis of the acetylated proteins. The resulting peptide mixture consisted of blocked N-terminal peptides mixed with internal peptides containing free amino groups on the proteolytically formed N-termini. Biotinylation of the peptide mixture using an N-hydroxysuccinimide (NHS) ester-derivative of biotin added a biotin moiety to the N-termini of the internal peptides, whereas true N-terminal peptides that had been previously blocked by acetylation would not be able to be biotinylated. The biotinylated peptides were removed by passing the mixture over streptavidin, and the unbound material containing the N-terminal peptides was analyzed without further treatment. As each protein yielded a single peptide, the resultant mixture had the same level of complexity as the initial proteome sample (**Box 1**).

The previously published protocol, although effective, required multiple peptide-purification steps to separate the peptides from the excess reagents that were used, which had the consequence of reducing the yield of material. Hence there was a need for an enhanced methodology that minimized the processing steps in order to maximize the yield. We have therefore developed an improved strategy, which is described here. The most significant enhancement to the method was the elimination of the

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

BOX 1 | REMOVAL OF INTERNAL PEPTIDES BY BIOTINYLATION AND STREPTAVIDIN PURIFICATION

1. Using a C18 ZipTip (following the manufacturer's instructions), desalt a small portion (10 µl) of the digested peptide mixture and elute into 10 µl elution solution.
2. Add 40 µl of 20 mM phosphate buffer (20 mM Na_2HPO_4 (pH 7.5)) to give a total volume of 50 µl.
▲ CRITICAL STEP Do not use an amine-containing buffer (e.g., Tris or ammonium bicarbonate). The free amines will quench the NHS-biotin reagent and prevent biotinylation of peptides.
3. Reconstitute 1 mg EZ-Link NHS biotin in 50 µl DMF.
▲ CRITICAL STEP EZ-Link NHS biotin should be prepared immediately before use, and the solution should not be stored and reused.
4. Add 1 µl biotin solution to the desalted peptide mixture and incubate overnight at 4 °C.
5. Using a C18 ZipTip, desalt the entire biotinylated peptide mixture and elute into 10 µl elution solution.
6. Dilute the biotinylated peptides in 10 µl binding buffer (20 mM Na_2HPO_4 and 0.15 M NaCl (pH 7.5)) to give a total volume of 20 µl.
7. Remove 20 µl streptavidin Sepharose from the stock preparation in ethanol and pipette into a 0.5 ml microcentrifuge tube.
8. Remove excess ethanol and wash the streptavidin Sepharose in 100 µl (five volumes) of binding buffer.
9. Centrifuge the Sepharose at 2,000*g* at room temperature for 20 s, remove the binding buffer and discard.
10. Add the desalted biotinylated peptide mixture, vortex and incubate with turning for 4 h at room temperature.
11. Centrifuge the Sepharose/peptide mixture at 2,000*g* at room temperature for 20 s, remove and retain supernatant (peptide mixture).
12. Proceed to mass-spectrometric analysis.

biotinylation step. An initial acetylation step is still required to block N-terminal peptides, but instead of targeting internal peptides by biotinylation and removing them with streptavidin we use a commercially available amine reactive immobilized reagent (NHS-activated Sepharose) to react and retain internal peptides in one step. NHS-Sepharose is efficient both in respect of amine binding and subsequent leakage of bound amines[17]. The peptide mixture is incubated with the NHS-activated Sepharose until coupling is complete, the Sepharose beads are then removed by brief centrifugation, and the unbound fraction is removed and analyzed without further treatment (**Fig. 2**).

In brief, samples are exchanged into a compatible (non-amine containing) buffer prior to acetylation — if the protein mixture can be generated directly in such a buffer this step is not required. After excess acetylation reagent is removed, the proteins are concentrated by acid precipitation prior to digestion with trypsin or another protease. The proteolyzed mixture is then exposed to activated NHS Sepharose, which removes all peptides with proteolytically exposed amino groups — in other words, all internal peptides (**Fig. 3**). To illustrate the method, we use soluble proteins from *Escherichia coli*, but this approach is amenable to any protein or protein mixture. Indeed, the soluble protein preparation from chicken skeletal muscle (obtained frozen from a local supermarket) has the advantage of a considerable bias in protein expression[18,19], and should generate an N-terminal preparation dominated by a few peptides from abundant proteins, which is readily assessed by matrix-assisted laser desorption ionization-time of flight (MALDI-ToF) MS.

We prefer to use acetylation as the amino-blocking reagent, although there are many other reagents that could be used. Many proteins are naturally acetylated at their N terminus, and by using the same chemical modification, we are effectively coalescing the N-terminal preparation into the same analytical space. It is straightforward to discriminate between naturally and chemically acetylated N-termini by the use of [$^3H_6$] acetic anhydride, which reacts with primary amines and introduces a mass shift of +45 Da instead of +42 Da. If modifications other than acetylation are used, it introduces the possibility of further resolving the N-terminal peptide mixture by, for example, affinity capture or a shift in hydrophobicity[11].

There is concern within the field of proteomics over the value of 'one hit wonders'[20], whereby a protein is identified using data derived from a single peptide. Part of this concern relates to the lack of information regarding the location of the peptide within the parent protein, and also to the search space required for identification (the entire database of candidate peptides must be given equal validity in the analysis). The strategy we have developed overcomes the stigma associated with 'one hit wonders', by anchoring the peptides at a precise location within the parent protein; it is therefore possible to limit the database search to a small subset of peptides.

The bias in the preparation of N-terminal peptides has consequences for the subsequent database search. Two strategies are used.

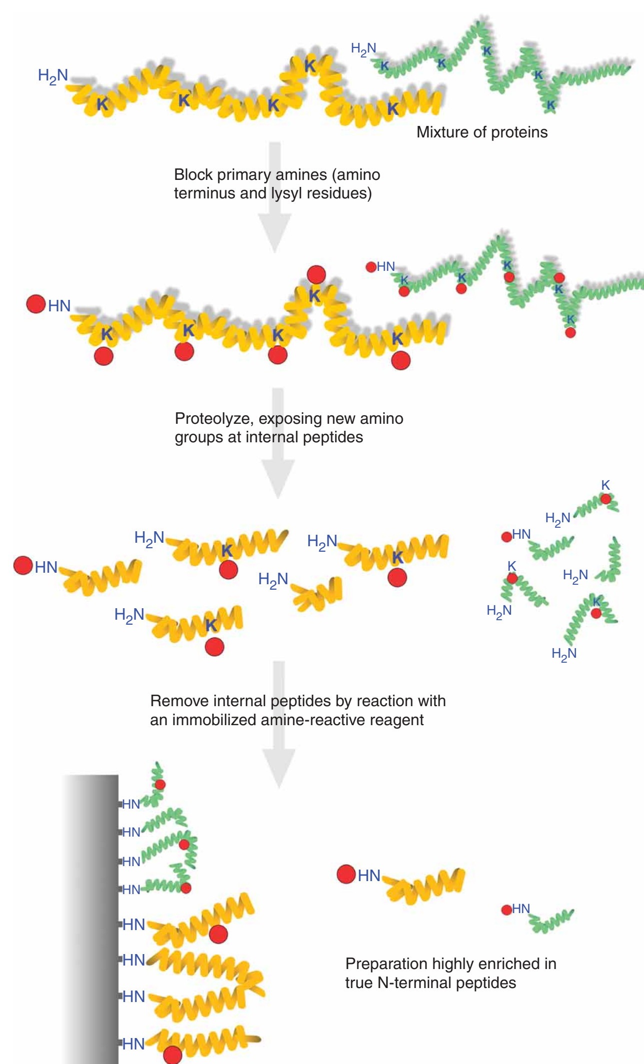

Figure 2 | Scheme showing the chemistry involved in N-terminal purification. Free amino groups (α and ε) are acetylated prior to proteolysis, which results in a mixture of N-terminally acetylated (true N-terminal) and non-acetylated (internal) peptides. Subsequent incubation of the peptide mixture with an immobilized amine-reactive reagent creates a preparation enriched in N-terminal peptides.

First, an entire database is searched, in which case the analyte peptides will be identified as high quality 'hits' that are also true N termini. However, this is predicated on the availability of knowledge of the true N terminus of each database entry. The complexity of post-translational processing and the fact that most proteomics databases are derived from cDNA or genomic data, creates a need for new search algorithms, yet to be developed, which capitalize on the drastically reduced search space and the positional bias. It is relevant to note that the preparation of N-terminal peptides can be employed to define the true N-termini of many proteins, and is therefore a useful tool in defining the true proteome.

MATERIALS

REAGENTS

- HPLC grade water
- HPLC grade acetonitrile (ACN) **! CAUTION** Flammable
- Luria broth (LB; Merck)
- Bugbuster protein extraction reagent (BB; Novagen-EMD Biosciences, cat. no. 70584) or any equivalent bacterial protein extraction reagent

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

- Coomassie Plus® protein assay (Perbio Science)
- Acetylation reagent: sulfo-NHS acetate (Pierce, cat. no. 26777)
- Acetylation buffer: 20 mM sodium carbonate (Na_2CO_3), pH 8.5, or other non-amine-containing buffers, such as phosphate or HEPES (pH 7–9) ▲ CRITICAL Do not use an amine-containing buffer (e.g., Tris or ammonium bicarbonate) as the free amines will quench the reagent and prevent protein acetylation
- Quenching reagent: Tris(2-aminoethyl)amine, polymer bound (Sigma, cat. no. 472107)
- 1 mM HCl
- Trichloroacetic acid (TCA) ! CAUTION Causes severe burns
- Trifluoroacetic acid (TFA) ! CAUTION Causes severe burns
- Diethyl ether ! CAUTION Highly flammable
- Digestion buffer: 20 mM Na_2HPO_4, pH 7.5
- Trypsin, sequencing grade (Roche, cat. no. 11 478 475 001) or any other proteolytic enzyme
- EZ-Link NHS-biotin (Pierce, cat. no. 20217)
- Dimethylformamide (DMF) ! CAUTION Harmful by inhalation, ingestion or skin contact
- Streptavidin Sepharose™, High Performance (GE Healthcare, cat. no. 17-5113-01)
- NHS-activated Sepharose™ 4 Fast Flow, stored in propanol (GE Healthcare, cat. no. 17-0906-01)
- Binding buffer: 20 mM Na_2HPO_4 and 0.15 M NaCl (pH 7.5)
- MALDI matrix: α-cyano-4-hydroxycinnamic acid (CHCA; Sigma, cat. no. C2020) ! CAUTION Irritating to eyes, respiratory system and skin
- Reverse-phase running buffer (A) 0.1% formic acid
- Reverse-phase eluting buffer (B) 90% ACN:0.1% formic acid
- Bacterial strains: *E. coli* BL21 ëDE3 (any other commercially available laboratory strain is suitable); frozen competent cells can be obtained from many suppliers, including Stratagene, Promega and Genlanatis ! CAUTION Use good microbiological practice in handling and disposing of this *E. coli* laboratory strain
- Chicken muscle (we used skeletal muscle tissue from *Gallus gallus* obtained frozen from a local supermarket)

EQUIPMENT

- Slide-A-Lyzer® dialysis cassettes, 500 µl to 3 ml, 10,000 molecular weight cut-off (Pierce, cat. no. 66425)
- ZipTip C18 pipette tips (Millipore, cat. no. ZTC18S008)
- Standard spectrophotometer for absorbance readings in the visible range, including 600 nm
- 1.5 and 0.5 ml plain microcentrifuge tubes
- Homogenizer
- Reverse-phase column: C18 3 µm particle size (100), 75 µm diameter × 150 mm long (Dionex)
- MALDI-ToF mass spectrometer (Waters MALDI-R, Shimadzu Axima TOF2 or equivalent)
- Electrospray-ionization tandem mass spectrometer coupled to a high-resolution nanoflow chromatography system (Dionex 3000 coupled to a Thermo Finnigan LTQ or other tandem mass spectrometer)

REAGENT SETUP

LB Dissolve 25 g LB powder in 1 l distilled water. The pH should be 7.0 ±0.2 at 25 °C; if not, adjust with HCl or NaOH as appropriate. Sterilize by autoclaving for 15 min at 121 °C.

Matrix for MALDI-ToF MS Prepare 50 ml of 50% (vol/vol) acetonitrile-0.1% (vol/vol) TFA; store at room temperature (20–25 °C). Prepare a fresh saturated solution of ~10mg CHCA in 1ml 50% (vol/vol) ACN-0.1% (vol/vol) TFA.

NHS-activated Sepharose Centrifuge the NHS-Sepharose slurry at 2,000*g* at room temperature for 20 s and remove excess propanol from the beads Wash the beads in five volumes of cold 1 mM HCl, vortex and remove HCl by centrifugation (as before). Wash in two volumes of binding buffer, remove by centrifugation. ▲ CRITICAL In order to retain maximum binding capacity of the pre-activated medium prior to the coupling step, use cold (0–4 °C) solutions. The time interval for all washing steps must be minimized. Prepare all required solutions prior to coupling ligand.

Chicken muscle Homogenize 0.5 g chicken skeletal muscle in 5 ml acetylation buffer. Centrifuge for 45 min at 13,000*g* at 4 °C, remove the supernatant fraction and use immediately or store at –20 °C. Determine the protein concentration of the soluble fraction using the Coomassie Plus® protein assay.

***E. coli* cell lysate** Using a single colony of *E. coli*, inoculate 10 ml LB and incubate overnight at 37 °C with shaking. Transfer 500 µl of the overnight culture to 50 ml prewarmed (to 37 °C) fresh LB media (1:100 dilution) and incubate the culture with shaking. Remove 1 ml samples at hourly intervals and determine the absorbance at 600 nm. Monitor growth rate until the early stationary phase is reached. Transfer the culture to a pre-weighed 50 ml centrifuge tube and centrifuge at 1,200*g* for 10 min at 4 °C. Decant the supernatant and weigh the tube again to determine the wet weight of the cell pellet. For ≤1 g of wet cell pellet, add 2.5 ml BB and, to ensure good resuspension, place cells on a rocker platform at room temperature for 15 min. Centrifuge the cells at 16,000*g* at room temperature for 20 min. Set the braking speed at low to give a gentle rotor deceleration. Remove the supernatant (soluble fraction) and use immediately or store at –20 °C. Determine the protein concentration of the soluble fraction using the Coomassie Plus® Protein assay.

EQUIPMENT SETUP

Slide-A-Lyzer® cassettes Before using the cassettes check for leaks by injecting the maximum amount of distilled water (3 ml) into each. Inject water through one of the four valves located in each corner of the cassette. Each valve can be used only once, so mark the cassette using a pen when a valve has been used. Attach a float to the top of the cassette and place into 1 litre acetylation buffer for 20 s in order to wet the membrane. Dry off the cassette by tapping it lightly onto a paper towel.

HPLC reverse-phase gradient The reversed phase chromatography was conducted at a flow rate of 0.3 µl min^{-1} over 3 h. The three phases of the gradients were as follows: 0–140 min, 0–50% buffer B (linear); 140–160 min, 50% buffer to 80% buffer B (linear); and 160–180 min, 80% buffer B to 0% buffer B.

PROCEDURE

Dialysis of *E. coli* proteins into acetylation buffer ● TIMING 4 h

1| Dialyze *E. coli* cell lysate into acetylation buffer using Slide-A-Lyzer® cassettes (see EQUIPMENT SETUP). Inject 3 ml *E. coli* cell lysate into the cassette through an unused valve. Attach the float and place into 1 l acetylation buffer. Leave to dialyze for 4 h at room temperature with stirring.
■ PAUSE POINT Dialyzed samples can be used immediately or stored for several months at –20 °C.

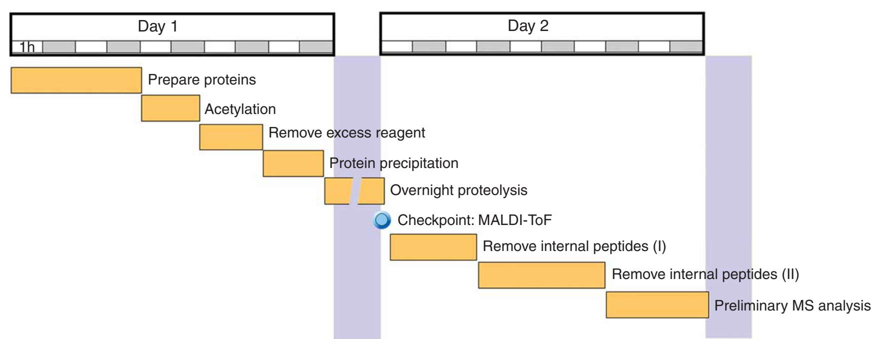

Figure 3 | Gantt chart for a typical N-terminal peptide-purification experiment. The entire procedure takes <2 d.

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

2| Determine the protein concentration of the dialyzed soluble fraction using the Coomassie Plus® protein assay. The protein concentration should be in the range of 1 to 5 μg μl^{-1}.

Acetylation of intact proteins ● TIMING 2 h

3| Reconstitute 1 mg sulfo-NHS acetate into 50 μl acetylation buffer.

▲ CRITICAL STEP Reconstitute sulfo-NHS acetate immediately before use. The NHS-ester readily hydrolyses and becomes unreactive.

4| Add 50 μl (1 mg) of the reconstituted acetylation reagent to 50 μg of the protein recovered from the dialyzed *E. coli* cell lysate (or protein mixture of choice). Incubate at room temperature for 2 h.

Removal of excess acetylation reagent ● TIMING 1 h

5| Add ~5 mg quenching reagent (Tris(2-aminoethyl)amine, polymer bound), vortex for 1 min and incubate with gentle agitation for 1 h.

▲ CRITICAL STEP This treatment has a major influence on the overall success of the process, and obviates the addition of free amines to inactivate excess reagent, which would then have to be removed before proceeding.

6| Remove amine-scavenging beads by filtration or centrifugation.

Precipitation of acetylated proteins ● TIMING ~1.5 h

7| Add 600 μl (five volumes) of cold 30% TCA to the protein mixture, vortex and incubate on ice for 1 h.

8| Centrifuge at 13,000*g* at room temperature for 2 min to pellet the protein.

9| Carefully remove the TCA supernatant fraction from the pellet and discard.

10| Add 200 μl diethyl ether to the pellet and agitate using a pipette tip.

! CAUTION Use a fume hood when pipetting ether.

11| Centrifuge for 10 s at 13,000*g* at room temperature.

12| Repeat Steps 10 and 11 twice more (three ether washes in total).

13| Remove diethyl ether and place tube at 37 °C for 5 min with the lid open to evaporate the excess.

Proteolysis ● TIMING overnight

14| Resuspend the diethyl ether-washed protein pellet in 50 μl digestion buffer. Digest overnight at 37 °C with 1 μg trypsin (or any other proteolytic enzyme; 1:50 enzyme:substrate).

■ PAUSE POINT Once digested, acetylated peptides can be stored for a few months at −20 °C.

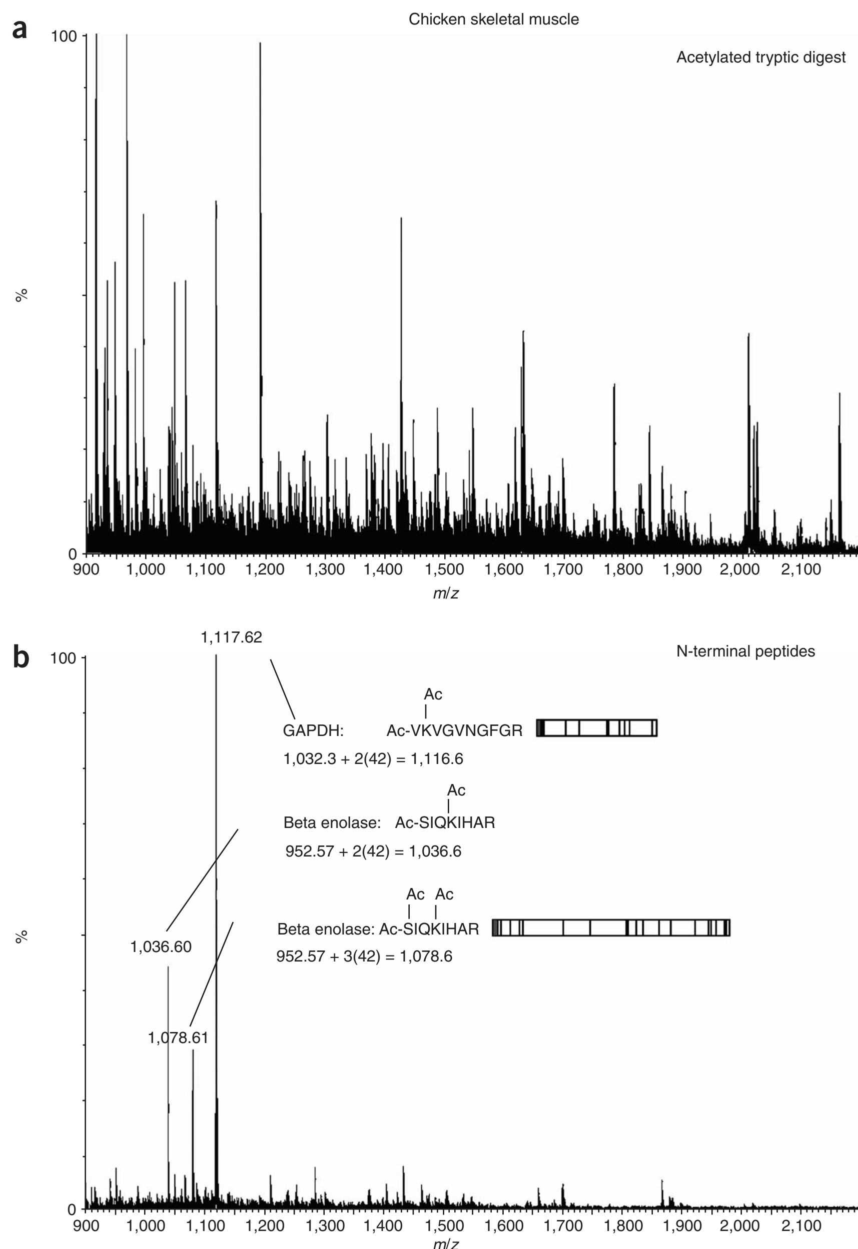

Figure 4 | Isolation of N-terminal peptides from chicken skeletal muscle soluble fraction. The starting material is a complex mixture of proteins, which simplifies to a relatively straightforward N-terminal peptide mix. (**a**) The entire tryptic digest of acetylated proteins. (**b**) Following application of the simplification protocol, the major ions labeled correspond to the N-terminal peptides from some of the most abundant soluble proteins in skeletal muscle.

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

MALDI-ToF analysis ● TIMING ~1 h

15| At this stage in the protocol, it is good practice to monitor acetylation and proteolysis of the digested peptides by MALDI-ToF using a Waters MALDI-R or any equivalent MALDI-ToF mass spectrometer. Prepare samples as follows: (i) dilute sample 1:20 in matrix (2 μl sample + 38 μl matrix); (ii) pipette 1 μl onto a clean MALDI target and allow to air dry; (iii) acquire data over the range of 900 to 3,500 *m/z*. At this stage of the protocol, a complex spectrum should be observed (corresponding to peptides derived from the entire protein mixture). The level of acetylation is difficult to determine; however, the observation of ArgC (as opposed to tryptic) peptides is a good indication that a sufficient degree of modification has occurred. Internal peptides can now be removed using NHS-activated Sepharose (continue to Step 16) or by biotinylation and streptavidin purification (**Box 1**).

Coupling of internal peptides to NHS-activated Sepharose ● TIMING ~24 h

16| Dilute the digested peptides in 50 μl binding buffer (20 mM Na_2HPO_4 and 0.15 M NaCl (pH 7.5)).

17| Remove 100 μl NHS-activated Sepharose from the stock preparation in propanol, and pipette into a 1.5-ml microcentrifuge tube. Wash as described in the REAGENT SETUP.

18| Add acetylated peptides (50 μg) to NHS-Sepharose, vortex and incubate with turning for 4 h at room temperature.

19| Centrifuge the Sepharose/peptide mixture at 2,000*g* at room temperature for 20 s, then remove and retain the supernatant (peptide mixture).

20| Prepare a second aliquot (100 μl) of NHS-activated Sepharose (repeat Step 17) and add the peptide mixture.

21| Incubate overnight at 4 °C with turning.

▲ **CRITICAL STEP** A second incubation with NHS Sepharose is necessary for complete coupling of peptides to Sepharose and to minimize leakiness of the procedure, wherein internal peptides can appear in the N-terminal peptide preparation.

22| Remove the peptide mixture from NHS-Sepharose (as before), divide into 10 μl aliquots and proceed to mass-spectrometric analysis.

■ **PAUSE POINT** The peptide mixture can be stored at −20 °C prior to MS.

MALDI-ToF analysis ● TIMING ~1 h

23| Prepare samples for MALDI-ToF analysis as described in Step 15.

Tandem mass-spectrometric analysis ● TIMING ~3 h for one sample

24| Separate the samples using a microcapillary reverse-phase column

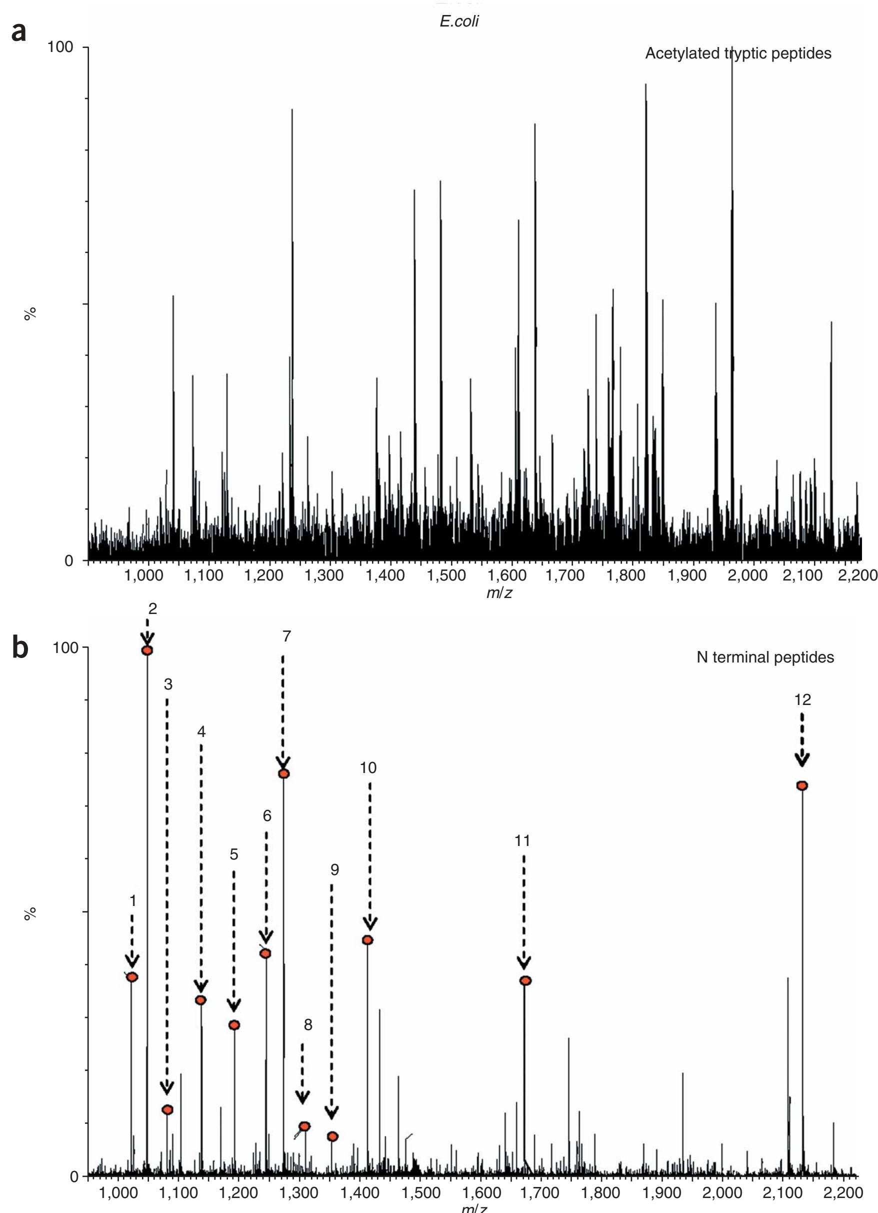

Figure 5 | Isolation of N-terminal peptides from *E. coli* cell lysate. (**a**) The entire tryptic digest of acetylated proteins. (**b**) Following application of the simplification protocol, the major ions labeled correspond to N-terminal peptides from some of the most abundant proteins in the sample.

(using reverse phase running buffer (A) and reverse phase eluting buffer (B)) in line with an eletrospray-ionization ion-trap tandem mass spectrometer.

Data analysis

25| Search the tandem mass spectra against Swiss-Prot using MASCOT or the TurboSequest program (ThermoElectron). Database search parameters include the fixed modifications of N-terminal acetylation and lysine acetylation, and the variable modification of O-acetylated serine (this is rare, however, so searches should be conducted both with and without this modification).

● TIMING

Dialysis of *E. coli* proteins into acetylation buffer: 4 h
Acetylation of intact proteins: 2 h
Removal of excess acetylation reagent: 1 h
Precipitation of acetylated proteins: ~1.5 h
Proteolysis: overnight
MALDI-ToF analysis: ~1 h
Coupling of internal peptides to NHS-activated Sepharose: ~24 h
MALDI-ToF analysis: ~1 h
Tandem mass-spectrometric analysis: ~3 h for one sample

? TROUBLESHOOTING

See **Table 1**.

TABLE 1 | Troubleshooting for the preparation of N-terminal peptides.

Problem	Possible cause	Solutions
Incomplete acetylation	Instability of acetylation regent (sulfo-NHS acetate and acetic anhydride)	Prepare sulfo-NHS acetate freshly and use immediately
	Acidification of mixture by excess acetic anhydride lowers pH	Use an acetylation buffer that is able to maintain the pH at ~8.5 for the duration of the reaction
	Accessibility of some primary amino groups is impaired by the 3D structure of the protein	Use buffers that denature or partially unfold the protein, or increase the time for acetylation (e.g., chaotropic buffers and/or reducing agents)
	Amines present in sample buffer compete with proteins for acetylation	Ensure that the buffer used does not contain amines
Incomplete proteolysis	Precipitated proteins not washed adequately to remove all residual trichloroacetic acid, which will acidify the digestion buffer and shift the pH down from the optimal value for trypsin action	Wash protein pellets carefully with ether to remove all TCA, or precipitate proteins with acetone; TCA will denature the protein mixture more effectively and increase the efficiency of proteolysis
Incomplete removal of peptides, identified by database searching.	Insufficient NHS-activated Sepharose used in the coupling step	Use a larger excess of NHS-activated Sepharose; ensure the NHS-Sepharose is freshly prepared
	Not enough time allowed for coupling	Use longer coupling times; repeat the reaction with a second batch of freshly washed NHS-Sepharose
	pH of coupling reactions not optimal	Ensure the system is buffered effectively at pH 8.0 ± 0.5; ensure that no amine-containing buffers, including Tris, have been added to the mixture; make sure that the internal peptide is not a true intracellular N terminus generated by endogenous proteolysis or by ectopic digestion after tissue breakage
Poor MALDI spectra	Phosphate buffer/PBS affecting ionization	Dilute the sample further or desalt using a C18 column; change the proteolysis buffer to a different system, such as HEPES
	Sample too complex for meaningful interpretation	This will be normal for a complex protein mixture in which no proteins predominate; separate the sample using HPLC (as in Step 24) and analyze fractions by MALDI, or move directly to LC-MS/MS or LC2-MS/MS (MUDPIT)

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

TABLE 2 | *E. coli* N-terminal peptides from MALDI spectrum.

Spot	Protein	Mass (Da)	Sequence
1	UPF0304 protein yfbU (P0A8W8)	1,021.43	MEMTNAQR
2	Elongation factor Tu (EF-Tu) (P0A6N1)	1,048.52	SKEKFER
3	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (P62707)	1,081.69	AVTKLVLVR
4	Enolase (P0A6P9)	1,138.71	SKIVKIIGR
5	10 kDa chaperonin (P0A6G1)	1,192.61	MNIRPLHDR
6	Glyceraldehyde-3-phosphate dehydrogenase A (P0A9B4)	1,244.69	TIKVGINGFGR
7	Copper-resistance protein D (Q47455)	1,274.67	MNDLIMIVIR
8	D,D-heptose 1,7-bisphosphate phosphatase (Q8FKZ1)	1,299.72	AKSVPAIFLDR
9	Putative HTH-type transcriptional regulator yeaT (P76250)	1,353.71	MNNLPLLNDLR
10	Elongation factor Ts (EF-Ts) (P0A6P1)	1,412.79	AEITASLVKELR
11	Phosphoglycerate kinase (P0A799)	1,671.89	SVIKMTDLDLAGKR
12	β-lactamase (P62593)	2,132.30	HPETLVKVKDAEDQLQR

These ions should be visible even in a MALDI-ToF spectrum of an entire N-terminal preparation of *E.coli* soluble proteins.

ANTICIPATED RESULTS

MALDI spectra of a digest of acetylated proteins should yield a complex mass spectrum, which represents the most abundant peptides in the sample. This analysis serves as a 'check point' in which the extent of acetylation can be monitored, provided that a few key peptides can be recognized. Due to the complexity of the peptide mixture, it is difficult to assign peptides to individual peaks at this stage in the protocol. Following abstraction of internal peptides to NHS-activated Sepharose, the N-terminally enriched supernatant should produce a notably simpler mass spectrum. At this stage, it might be possible to assign the highest intensity signals to true N-terminal peptides.

Skeletal muscle soluble protein preparations are dominated by ~10–20 major proteins that are predominantly glycolytic enzymes[19,20]. This preparation has the advantage of generating relatively simple spectra, and is a valuable test system with which to practice the method. **Figures 4a** and **5a** show the entire tryptic digests for the acetylated chicken muscle and *E. coli* cell lysate samples. As expected the spectra are complex and it is not possible to identify any N-terminal peptides. However, **Figures 4b** and **5b** represent the unbound fraction for the two samples, which should be substantially N-terminally enriched. These elicit much simpler mass spectra, and it is possible to assign identities to the most intense signals (*E. coli* MALDI-ToF peak assignments are listed in **Table 2**). Due to the dynamic range of the skeletal muscle proteome, the N-terminal spectrum for chicken muscle is substantially less complex than the *E. coli* sample. The high abundance of glycolytic proteins found in skeletal muscle means that the lower-abundance proteins are not visible at this stage of analysis.

Using a 3-h HPLC gradient, tandem mass-spectrometric analysis should provide data on hundreds of N-terminal peptides (depending on sample complexity). When analyzed in this way, the *E. coli* N-terminally enriched preparation yielded >300 protein identifications. All identifications were from a search of the entire SwissProt database of *E. coli* proteins, and relatively few peptides were identified as internal sequences.

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

COMPETING INTERESTS STATEMENT The authors declare that they have no competing financial interests.

ACKNOWLEDGMENTS This work was supported by the Engineering and Physical Sciences Research Council.

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

1. Bogdanov, B. & Smith, R.D. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* **24**, 168–200 (2005).
2. Wu, C.C. & MacCoss, M.J. Shotgun proteomics: tools for the analysis of complex biological systems. *Curr. Opin. Mol. Ther.* **4**, 242–250 (2002).
3. Liu, H., Lin, D. & Yates, J.R. Multidimensional separations for protein/peptide analysis in the post-genomic era. *Biotechniques* **4**, 898–902 (2002).
4. Mirzaei, H. & Regnier, F. Structure specific chromatographic selection in targeted proteomics. *J. Chromatogr. B* **817**, 23–34 (2005).
5. Smolka, M.B., Zhou, H., Purkayastha, S. & Aebersold, R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem.* **297**, 25–31 (2001).
6. Raggiaschi, R., Gotta, S. & Terstappen, G. Phosphoproteome analysis. *Biosci. Rep.* **25**, 33–44 (2005).
7. Kweon, H.K. & Hakansson, K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* **78**, 1743–1749 (2006).
8. West, I. & Goldring, O. Lectin affinity chromatography. *Methods Mol. Biol.* **59**, 177–185 (1996).
9. Gevaert, K., Van Damme, P., Martens, L. & Vandekerckhove, J. Diagonal reverse-phase chromatography applications in peptide-centric proteomics: ahead of catalogue-omics? *Anal. Biochem.* **345**, 18–29 (2005).
10. Martens, L. *et al.* The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile. *Proteomics* **5**, 3193–3204 (2005).
11. 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).
12. Kasai, K.-I. Trypsin and affinity chromatography. *J Chromatogr. A* **597**, 3–18 (1992).
13. Sechi, S. & Chait, B.T. A method to define the carboxyl terminal of proteins. *Anal. Chem.* **72**, 3374–3378 (2000).
14. Yamaguchi, M. *et al.* High-throughput method for N-terminal sequencing of proteins by MALDI mass spectrometry. *Anal. Chem.* **77**, 645–651 (2005).
15. Chelius, D. & Shaler, T.A. Capture of peptides with N-terminal serine and threonine: a sequence-specific chemical method for peptide mixture simplification. *Bioconjugate Chem.* **14**, 205–211 (2003).

16. 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).
17. Van Sommeren, A.P.G., Machielsen, P.A.G.M. & Gribnau,, T.C.J. Comparison of three activated agaroses for use in affinity chromatography: effects on coupling performance and ligand leakage. *J. Chromatogr. A* **639**, 23–31 (1993).
18. Hayter, J.R., Robertson, D.H.L., Gaskell, S.J. & Beynon, R.J. Proteome analysis of intact proteins in complex mixtures. *Mol. Cell Proteomics* **2**, 85–95 (2003).
19. Doherty, M.K. *et al.* The proteome of chicken skeletal muscle: changes in soluble protein expression during growth in a layer strain. *Proteomics* **4**, 2082–2093 (2004).
20. Veenstra, T.D., Conrads, T.P. & Issaq, H.J. What to do with "one-hit wonders"? *Electrophoresis* **25**, 1278–1279 (2004).

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols