

## Protein Processing Characterized by a Gel-Free Proteomics Approach

Petra Van Damme, Francis Impens, Joël Vandekerckhove, and Kris Gevaert

### Summary

We describe a method for the specific isolation of representative N-terminal peptides of proteins and their proteolytic fragments. Their isolation is based on a gel-free, peptide-centric proteomics approach using the principle of diagonal chromatography. We will indicate that the introduction of an altered chemical property to internal peptides holding a free  $\alpha$ -N-terminus results in altered column retention of these peptides, thereby enabling the isolation and further characterization by mass spectrometry of N-terminal peptides. Besides pointing to changes in protein expression levels when performing such proteome surveys in a differential modus, protease specificity and substrate repertoires can be allocated since both are specified by neo-N-termini generated after a protease cleavage event. As such, our gel-free proteomics technology is widely applicable and amenable for a variety of proteome-driven protease degradomics research.

**Key Words:** Gel-free proteomics; N-terminal COFRADIC; protein processing; proteases; substrates.

### 1. Introduction

There are several advantages of gel-free proteomics following selection and identification of protein N-terminal peptides (*I*). First, the greatest reduction in sample complexity prior to mass spectrometry (MS)/MS analysis is achieved without any loss of information since every protein is represented only by its N-terminal peptide. Second, as many protein isoforms diverge mainly at their N-terminal extremities it is possible to distinguish them. As an example, so-called xenoproteomics experiments, i.e., simultaneous analysis of proteomes

From: *Methods in Molecular Biology*, vol. 484: *Functional Proteomics: Methods and Protocols*  
Edited by: J. D. Thompson et al., DOI: 10.1007/978-1-59745-398-1\_16, © Humana Press, Totowa, NJ

from different species as present in xenografts, have been performed successfully using N-terminal peptides (2). Third, newly generated N-termini are indicative for protein cleavage by proteases, allowing screening for their substrates in a differential proteomics setup (3). In the protocol to select N-terminal peptides by COmbined FRActional DIagonal Chromatography (COFRADIC) (4) outlined below, we focus on this latter application. The commercial rights for this and other COFRADIC applications belong to pronota ([www.pronota.com](http://www.pronota.com)).

Only a few techniques were reported for analyzing protein N-terminal sequences in a gel-free, high-throughput manner (5). Two methods somewhat related to N-terminal COFRADIC were recently reported: the use of protein sequence tags (6) and positional proteomics (7). However, both methods have not been applied for proteome-wide characterization of protein processing until now.

In our approach, following their extraction from cells or tissues, proteins are reduced, cysteines are alkylated, and free  $\alpha$ - and  $\epsilon$ -amines are blocked by trideuteroacetylation, making it possible to later characterize the *in vivo* nature (blocked or free, see below) of protein N-termini. Following protein cleavage this modification is an extra confirmation for the identification of newly formed N-termini since these should be trideuteroacetylated. As a consequence of this acetylation step, digestion by trypsin results in peptides ending on an arginine residue. The N-terminal COFRADIC procedure then serves to separate internal and C-terminal peptides from N-terminal ones. The modification reaction between the two sequential and identical chromatographic separation steps uses 2,4,6-trinitrobenzenesulfonic acid (TNBS). This bulky, hydrophobic reagent now reacts only with free  $\alpha$ -amines of internal peptides, hereby inducing a hydrophobic shift during the secondary separation. In this way, nonshifted N-terminal peptides (blocked by acetylation) are sorted for further MS/MS analysis.

To distinguish between different proteomes, stable isotope labeling is necessary, introducing known measurable peptide mass differences (**Subheading 3.3**). Using N-terminal COFRADIC in a differential way, the dynamics and status of N-terminal modifications on proteins are characterized. Furthermore, when screening for protease substrates, typically, samples with and without protease activity are compared. Peptides from newly generated N-termini will be present only in one proteome sample and will therefore be present as a peptide with a single isotopic envelope distribution in a mass spectrum (3). This, together with the trideuteroacetylation step mentioned above, makes proteome-wide identification and characterization of protease substrates very straightforward.

Just like other enzymatic systems proteases almost never work alone. They tend to work in networks in which one protease sequentially activates other

proteases (e.g., the caspase cascade and during blood clotting), or where several, different proteases become active at the same time (e.g., release of proteases by lysosomal membrane permeabilization). Together with unwanted protease activity induced by cell or tissue lysis, this often complicates the *in vivo* study of protease substrates. When used in a differential way, unwanted protein processing is evident following differential N-terminal COFRADIC since N-terminal peptides formed by this “unwanted activity” will be equally present in treated and control samples. Compensating for protease networking is more difficult and highly challenging, since often there is interest in categorizing the substrates of only one particular protease working in its normal *in vivo* environment or network. Therefore, we suggest performing two types of screens. First, we identify substrates by adding a purified or recombinant protease to a relevant lysate (further referred to as the *in vitro* screen) containing substrates in their native state. The generated list of substrates not only allows the assessment of cleavage site specificity, but can also be used to validate the results obtained from the second screen (further referred to as the *in vivo* screen) where the protease is active in its biological context. Based on the results of the *in vitro* screen those cleavage events in the *in vivo* screen that are due to activity of the protease of interest can be assigned.

## 2. Materials

### 2.1. Protein Extraction (Subheading 3.1)

1. Jurkat cell line (ATCC, Manassas, VA, #CRL-1658) and RPMI 1640 medium (Invitrogen, Carlsbad, CA, #61870-010) or adapted arginine-free RPMI medium (see **Subheading 3.3.2**).
2. Complete EDTA-free protease inhibitor cocktail tablet (Roche Diagnostics, Mannheim, Germany, #11873580001).
3. Complete protease inhibitor cocktail tablet (Roche Diagnostics, #11697498001).
4. Lysis buffer 1: 50 mM morpholinoethanesulfonic acid (MES), 50 mM sodium phosphate, pH 7.4, 150 mM NaCl, 1 mM dithiothreitol (DTT), 1 mM EDTA-free protease inhibitors (1 tablet per 100 mL of lysis buffer, see **Notes 1** and **2**).
5. Lysis buffer 2: 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.8% CHAPS, protease inhibitors (Roche, 1 tablet per 100 mL).
6. Bio-Rad DC Protein Assay Kit (Bio-Rad, München, Germany #500-0006).
7. Recombinant HIV-1 protease (ProteinOne, Bethesda, MD, #P5102).
8. Disposable desalting columns packed with Sephadex™ G-25 (GE Healthcare Bio-Sciences, Uppsala, Sweden, #17-0853-01, #17-0854-01, or #17-0851-01).

### 2.2. N-Terminal COFRADIC (Subheading 3.2)

1. Tris(2-carboxyethyl)phosphine (TCEP, Pierce, Rockford, IL, #20490).
2. Iodoacetamide (Fluka BioChemica, Buchs, Switzerland, #57670).

3. Sulfo-*N*-hydroxysuccinimide acetate (s-NHS-acetate, Pierce, #26777).
4. Trideutero-*N*-hydroxysuccinimide acetate (8).
5. Hydroxylamine (Fluka BioChemica, #55458).
6. Hydrogen peroxide (30% [w/w] in H<sub>2</sub>O, Sigma-Aldrich, St. Louis, MO, #H1009).
7. 2,4,6-Trinitrobenzenesulfonic acid (TNBS, Fluka BioChemika; 1 M solution in water, #92822).
8. Disposable desalting columns packed with Sephadex™ G-25 (GE Healthcare).
9. Sequencing grade modified trypsin (Promega, Madison, WI, #V5111).
10. Analytical reverse-phase high-performance liquid chromatography (RP-HPLC) column: 2.1 mm internal diameter (i.d.) × 150 mm (length) 300SB-C<sub>18</sub> column, Zorbax® (Agilent, Waldbronn, Germany).
11. Agilent 1100 Series HPLC system.
12. HPLC grade water (e.g., Baker HPLC analyzed, Mallinckrodt Baker B.V., Deventer, the Netherlands).
13. HPLC grade acetonitrile (e.g., Baker HPLC analyzed, Mallinckrodt Baker B.V.).
14. HPLC solvent A: 10 mM ammonium acetate (pH 5.5) or 0.1% trifluoroacetic acid (TFA) in water/acetonitrile, 98/2 (v/v) (see **Note 3**).
15. HPLC solvent B: 10 mM ammonium acetate (pH 5.5) or 0.1% TFA in water/acetonitrile, 30/70 (v/v) (see **Note 3**).
16. TFA (Rathburn, Walkerburn, UK).

### 2.3. Protein Isotopic Labeling (Subheading 3.3)

1. <sup>18</sup>O-rich water (93.7% H<sub>2</sub><sup>18</sup>O [w/w] pure, ARC Laboratories, Apeldoorn, The Netherlands, #OLM-240).
2. TCEP (Pierce, #OLM-240): prepare a 10 mM stock solution in water.
3. Iodoacetamide (Fluka BioChemica, #57670): prepare a 100 mM stock solution in water.
4. Guanidinium hydrochloride (Fluka BioChemica, #50939): prepare a 6 M stock solution in water.
5. Amino acids.
  - a. <sup>13</sup>C<sub>6</sub>-L-Arginine hydrochloride (Cambridge Isotope Laboratories, Andover, MA, #CLM-2265).
  - b. <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>4</sub>-L-Arginine hydrochloride (Cambridge Isotope Laboratories, #CNLM-539).
  - c. L-Arginine (Sigma-Aldrich, #A-8094).
6. Cell culture:
  - a. Dialyzed fetal bovine serum (Invitrogen, #26400-044).
  - b. Dulbecco's modified Eagle's medium (DMEM), F-12K or RPMI 1640 without L-arginine (Invitrogen). Note: the compositions of these media are available from Invitrogen as a custom service. The custom-synthesized media have exactly the same composition as the regular media (DMEM, #21885-108; RPMI 1640, #61870-010; F-12K, #21127-022 all from Invitrogen), except that they are deficient of the specified amino acid.

- c. Penicillin–streptomycin (10,000 U of penicillin and 10,000  $\mu\text{g}/\text{mL}$  of streptomycin) (Invitrogen, #15070-063).
  - d. HEK 293T cell line (ATCC, #CRL-11268).
  - e. Jurkat cell line (ATCC, #CRL-1658).
  - f. K-562 cell line (ATCC, #CCL-243).
  - g. A-549 cell line (ATCC, #CCL-185).
  - h. NK-92 cell line (ATCC, #CRL-2407).
  - i. NK-92MI cell line (ATCC, #CRL-2408).
  - j. SH-SY5Y cell line (ATCC, #CRL-2266).
7. Prepare concentrated stocks (400 mM) of  $^{13}\text{C}_6$ ,  $^{13}\text{C}_6^{15}\text{N}_4$ , and  $^{12}\text{C}_6$  L-arginine hydrochloride in phosphate-buffered saline (PBS) (f.c. [final concentration] for RPMI, #61870; 200  $\mu\text{g}/\text{mL}$  or 1.15 mM L-arginine or 1.15 mM, f.c.. for F-12K, #21127; 422  $\mu\text{g}/\text{mL}$  or 2 mM and f.c. for DMEM, #21885; 84  $\mu\text{g}/\text{mL}$  or 0.398 mM) to make complete RPMI 1640 (containing  $^{12}\text{C}_6$  L-arginine) and RPMI 1640 with  $^{13}\text{C}_6$  or  $^{13}\text{C}_6^{15}\text{N}_4$  L-arginine. Dissolve and divide in small aliquots to avoid multiple freeze–thaw cycles. Add the optimized amount of stock  $^{13}\text{C}_6$ ,  $^{13}\text{C}_6^{15}\text{N}_4$ , or  $^{12}\text{C}_6$  L-arginine hydrochloride to the reconstituted arginine-deficient RPMI 1640 media (containing 10% dialyzed fetal bovine serum [free of amino acids], 1% penicillin-streptomycin, and other components whenever required), as to prepare the heavy and light forms of the media. respectively. Subsequently, filter the medium through a 0.22- $\mu\text{m}$  filter and store it at 4°C until use.

### 3. Methods

#### 3.1. Extraction Procedures

Efficient protein extraction yielding soluble proteins after disruption of biological membranes is required prior to N-terminal COFRADIC. Since our main focus here is the identification of protease substrates, the major differences between the lysis methods described below depend on whether *in vitro* or *in vivo* substrate catalogues will be constructed. We describe three different protein extraction procedures preceding differential N-terminal COFRADIC approaches. **Subsection 3.1.1** outlines procedures for *in vitro* protease substrate screening whereas **Subsection 3.1.2** is recommended for protease-unrelated studies or studies in which postlytic *in vitro* enzymatic activity is unwanted. Both protocols use cells in culture. When starting from dissected animal tissue, **Subsection 3.1.3** must be applied.

For *in vitro* screens, as many potential substrates as possible must be extracted, preferably in their native form. In addition, the extraction conditions should be compatible with subsequent activity of the protease of interest. Therefore, we suggest extracting proteins by multiple freeze–thaw cycles on the cells of interest in a buffer optimal for protease activity or adaptable to achieve such conditions. Detergent-based cell lysis is to be avoided since most detergents are ineffi-

ciently removed and interfere with mass spectrometric analyses. Furthermore, detergents may lead to protein denaturation and thus to protease access to epitopes in irrelevant substrates. Also, detergents might influence protease activity.

As a major drawback, some proteins might be missed since their extraction needs detergents. To avoid contaminating downstream protease activity, broad-spectrum protease inhibitors against the three classes of proteases other than the one under investigation should be included, although many proteases are not well targeted by these inhibitors and behave as exceptions in their class. For reasons of general protein solubility the pH of the extraction buffer should be around 7. When studying proteases displaying an acid pH optimum, adjust the pH of the lysate after its extraction. Ionic strength, chelators, and other buffer components can best be optimized for each individual protease to reach its optimal activity.

Since the relevant “library” of possible substrates and specific conditions for activity differ considerably between proteases, in contrast to an *in vivo* screen, we cannot supply an optimal protocol well suited for every protease. As an example we describe the protein extraction steps and procedures to screen for substrates of the recombinant HIV-1 protease in a representative lysate of cultured human Jurkat T cells.

### 3.1.1. Protein Extraction from Cultured Cells for Subsequent Protease Incubation

1. a. In the case of metabolic labeling of proteins by  $^{13}\text{C}_6$ -Arg SILAC: culture Jurkat cells separately in adapted RPMI 1640 medium in the presence of  $^{12}\text{C}_6$  or  $^{13}\text{C}_6$ -arginine as described in **Subsection 3.3.2**. Harvest equal numbers of light and heavy labeled cells and wash them two times with PBS to remove residual media components.  
b. In the case of postmetabolic, enzymatic labeling of peptides by  $\text{H}_2^{18}\text{O}$  after protein extraction and digestion: harvest the cells cultured in normal RPMI 1640 medium and wash them two times with PBS. Divide the sample in two aliquots of equal cell numbers. Details with reference to the labeling procedure are outlined in **Subsection 3.3.1**.
2. Resuspend individual cell pellets in lysis buffer 1.
3. Freeze both samples by putting them on dry ice for 15 min followed by thawing on ice at  $4^\circ\text{C}$  for 15 min. Repeat this step three times.
4. Centrifuge the samples for 15 min at  $16,000 \times g$  ( $4^\circ\text{C}$ ) and recover the supernatant.
5. Measure the protein concentration using the Bio-Rad DC Protein Assay Kit according to the manufacturer’s instructions. Equalize small differences in protein concentration by diluting the most concentrated sample with the appropriate volume of lysis buffer 1.

6. Acidify both samples with 2 N HCl to pH 5.5 and increase the salt concentration to 300 mM NaCl using a 5 M stock since the HIV-1 protease has a slightly acid pH optimum and cleaves more efficiently at higher salt concentrations (9) (see **Note 4**).
7. Add to one sample the recombinant HIV-1 protease to a final concentration of 200 nM and incubate for 75 min at 37°C (treated sample, see **Note 5**). Add no protease or, alternatively, an inactive protease variant to the other sample and incubate under conditions identical to the treated sample (control sample, see **Note 6**).
8. After incubation, the protease activity can be blocked by adding an excess of a potent protease inhibitor to both samples; however, as is the case for the HIV-1 protease, such (often patented) inhibitors are not always available. In that case, immediately inhibit any remaining protease activity by adding chaotropes (e.g., guanidinium hydrochloride) in sufficiently high concentrations (4– 6 M) combined with cysteine alkylation (see below).
9. The pH of both samples is increased to 7.5 using 2 M NaOH and guanidinium hydrochloride is added dry to a final concentration of 4 M (see **Note 4**).
10. Proceed directly to *step 2* of **Subsection 3.2.1**. Mixing of both samples is discussed in **Subsection 3.4**.

In screens where extraction conditions do not need to be tuned for monitoring specific protease activity and the integrity of the three-dimensional structure of the substrate is unnecessary, postlysis effects due to remaining protease activity should be avoided during extraction. Below, we describe a general protocol for protein extraction for *in vivo* screens starting from cultured cells or dissected tissue.

### 3.1.2. Protein Extraction from Cultured Cells

1. In the case of metabolic labeling of proteins by  $^{13}\text{C}_6$ -Arg SILAC: culture cells separately in the appropriate medium and in the presence of  $^{12}\text{C}_6$  or  $^{13}\text{C}_6$ -arginine according to labeling conditions described in **Subsection 3.3.2**. Perform treatment of cells during culture (i.e., stimulate cells to evoke protease activity or use as control) and harvest numbers of light and heavy labeled cells such that equal amounts of proteins (see **Notes 6** and **7**) for treated and control cells are obtained. Wash the cells thoroughly with PBS.
2. In the case of postmetabolic, enzymatic labeling of peptides by  $\text{H}_2^{18}\text{O}$  after protein extraction and digestion: culture the cells in their normal medium, perform appropriate treatment of the cells during culture, and harvest numbers of light and heavy labeled cells to obtain equal amounts of protein (see **Notes 6** and **7**) for treated and control sample. Wash the cells thoroughly with PBS.
3. Resuspend each cell pellet in lysis buffer 2 and lyse the cells on ice for 15 min (see **Note 2**). More specific protease inhibitors can be added to this lysis buffer if required.
4. Centrifuge the samples for 15 min at  $16,000 \times g$  (4°C) and recover the supernatant.

5. Measure the protein concentration using the Bio-Rad DC Protein Assay Kit according to the manufacturer's instructions. Equalize small differences in concentration by diluting with an appropriate volume of lysis buffer 2.
6. Desalt the protein mixture using disposable desalting columns according to the manufacturer's instructions with the appropriate volume of guanidinium hydrochloride in sodium phosphate (pH 7.5). The final concentration of guanidinium hydrochloride should be 4 M after drying down the protein mixture to its original starting volume.
7. Proceed directly to *step 2* of **Subsection 3.2.1**. Mixing of both samples is discussed in **Subsection 3.4**.

### 3.1.3. Protein Extraction from Dissected Animal Tissue

1. During dissection, wash the tissue samples several times thoroughly with PBS and remove residual body fluid components as completely as possible. Snap-freeze the samples in liquid nitrogen and store at  $-80^{\circ}\text{C}$  until further processing.
2. Subject the frozen tissue to mechanical dissociation by a pestle in a liquid nitrogen-cooled mortar.
3. Suspend the powder in 4 M guanidinium hydrochloride and 50 mM sodium phosphate buffer at pH 7.5 (see **Note 2**).
4. Extract proteins by incubating this suspension on an orbital shaker for 1 h at  $4^{\circ}\text{C}$ .
5. Centrifuge the protein sample for 60 min at  $90,000 \times g$  and at  $4^{\circ}\text{C}$  and recover the supernatant.
6. Measure the protein concentration using the Bio-Rad DC Protein Assay Kit according to the manufacturer's instructions. Equalize small differences in concentration by adding lysis buffer.
7. Proceed directly to *step 2* of **Subsection 3.2.1**. Mixing of both samples is discussed in **Subsection 3.4**.

## 3.2. N-Terminal COFRADIC

### 3.2.1. Sorting of N-Terminal Peptides

1. Prepare proteomes from treated and control samples as described in **Subheading 3.1**.
2. Desalt the protein mixtures on a disposable desalting column according to the manufacturer's instructions with the appropriate amount of guanidinium hydrochloride in sodium phosphate (pH 7.5) to generate a final concentration of 4 M guanidinium hydrochloride in 50 mM sodium phosphate (pH 7.5) after vacuum drying the desalted protein mixtures to their original volume.
3. Add freshly prepared TCEP-HCl (1 mM f.c.) and iodoacetamide (2 mM f.c.) solutions. Let the reduction/alkylation reaction proceed in the dark for 1 h at  $37^{\circ}\text{C}$ .
4. Desalt the protein mixtures on a desalting column in 2 M guanidinium hydrochloride in 50 mM sodium phosphate (pH 8.0) after drying down to its original volume.

5. Add freshly prepared 5 mM sulfo-*N*-hydroxysuccinimide acetate or 10 mM trideutero-*N*-hydroxysuccinimide acetate (prepare a fresh 500 mM stock in 1% DMSO). Incubate for 90 min at 30°C.
6. Revert partial acetylation of hydroxyl groups by adding 2  $\mu$ L of hydroxylamine and incubate for an additional 15 min at 30°C.
7. Desalt the mixtures of modified proteins in 20 mM  $\text{NH}_4\text{HCO}_3$  (pH 7.6).
8. Reduce the overall volume of each sample to 1 mL by vacuum drying.
9. Boil the protein mixtures for 10 min at 95°C and then transfer for 10 min to an ice bath.
10. Add sequence grade modified trypsin (the enzyme/substrate ratio should be about 1/50) and incubate overnight at 37°C.
11. Proceed to *step 2* of **Subsection 3.3.1** when using differential  $^{18}\text{O}$  labeling.
12. Acidify the modified primary fractions by adding 2  $\mu$ L of TFA or 4  $\mu$ L of 100% acetic acid (*see Note 8*) and centrifuge the peptide mixtures for 10 min at  $10,000 \times g$  to remove insoluble material. Transfer the supernatant to an HPLC sample vial.
13. Add the appropriate volume of 30% (w/v)  $\text{H}_2\text{O}_2$  solution to reach a final concentration of 0.5% and incubate for 30 min at 30°C (*see Note 9*).
14. Load the sample on the reverse-phase column (*see Subheading 3.2.2*) for the primary COFRADIC separation and fractionate in 12–15 consecutive fractions of 4 min each starting 20 min following sample injection (about 7% of acetonitrile concentration), as very few peptides elute earlier in the gradient.
15. Dry these primary fractions to complete dryness and redissolve each primary fraction in 50  $\mu$ L sodium borate buffer (pH 9.5).
16. Add 10  $\mu$ L of a 15 mM TNBS solution and incubate for 1 h at 37°C.
17. Repeat the previous step three times to ensure near quantitative TNBS modification of free  $\alpha$ -amino groups.
18. Load the TNBS-treated fraction onto the reverse-phase column, starting with the most hydrophobic primary fraction, and subsequently fractionate using the same solvent gradient as during the primary run. Collect the N-terminal peptides (*see Note 10*) in 16 equal-volume secondary fractions in an 8-min-long time interval starting 2 min prior to and ending 2 min after the primary collection interval (*see Note 11*). An example of COFRADIC sorting N-terminal peptides is depicted in **Fig. 1**.
19. Dry the collected N-terminal peptides and store at  $-20^\circ\text{C}$  until further LC-MS/MS analysis (*see Note 12*).

### 3.2.2. Setting Up the Reverse-Phase Diagonal Chromatographic System for Sorting N-Terminal Peptides

1. Apply the following binary solvent gradient for separating the peptide mixture:
  - a. Following injection of the sample onto the column, apply a 10 min isocratic run with 100% of solvent A at a constant flow rate of 80  $\mu\text{L}/\text{min}$  (*see Note 13*).

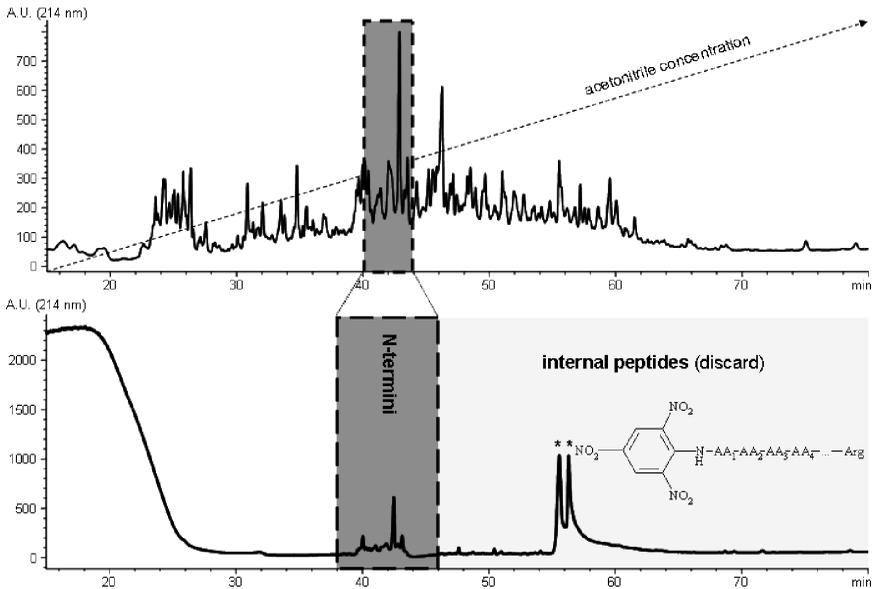


Fig. 1. Sorting of N-terminal peptides. Cultured human Jurkat cells were subjected to three freeze–thaw cycles to extract proteins and subsequently processed as indicated under the method in **Subsection 3.1.1, step 4**. The upper panel shows the RP-HPLC chromatogram (UV absorbance measured at 214 nm) of the separation of the tryptic digest of this protein mixture (i.e., the primary COFRADIC run). This peptide mixture was fractionated into 13 primary fractions of 4 min each (from 20 to 72 min). Shown in the lower panel is the RP-HPLC chromatogram of secondary fraction 6 after treatment of the peptide mixture with TNBS (i.e., the secondary COFRADIC run). Unaltered N-terminal peptides are collected in 16 equal-volume secondary fractions in an 8-min-wide time window starting 2 min prior to the original, primary elution interval of fraction 6 (indicated in a gray background with a dashed line). TNBS-modified peptides (i.e., internal peptides that carried a free  $\alpha$ -amino group) now obtained a hydrophobic trinitrophenyl group and are thus shifted to later elution times. Note that background peaks due to impurities in TNBS are indicated with an asterisk.

- b. Apply a linear, binary gradient over 100 min to 100% of solvent B.
  - c. Apply a 10 min isocratic wash with 100% of solvent B, followed by a linear gradient over 5 min to 0% of solvent B (100% of solvent A).
  - d. Reequilibrate the column for another 20 min with 100% of solvent A before injection of another sample.
2. Depending upon the type of peptide isolated and thus the preceding protein preparation steps we observed that peptides typically elute between 20 and 100 min of gradient time, corresponding to acetonitrile concentrations of 7% and 63%, respectively. Collect the primary fractions as indicated in *step 13* of **Subsection 3.2.1**.

### 3.3. Differential Quantitative Proteomic Labeling Approaches Exploited for N-Terminal COFRADIC

When performing large-scale, differential proteomics surveys, labeling methods incorporating stable, heavy isotopes into proteins or peptides are typically used. By determining the ratio of the intensities originating from the isotopically “light” and “heavy” ion signals of a peptide in a mass spectrum, the relative abundance of the peptide (and protein) in the two represented varieties can be assessed. Isotope labeling can be done on two different levels: either through physiological incorporation (metabolic labeling) or by introduction of a specific enzymatic or chemical derivatization step on the peptide or protein level (postmetabolic labeling) (**10,12,17**). Here, we focus on the strategies that we routinely follow to introduce stable heavy isotopic label(s) when performing N-terminal COFRADIC, the selection of which mainly depends on the sample’s origin.

We recently introduced an acetylation step on the protein level introducing a trideutero-acetyl group (**8**) on every free  $\alpha$ - and  $\epsilon$ -amino group. As mentioned above, cleavage event(s) will now appear as single trideuteroacetylated neo-N-termini (*see Note 14*).

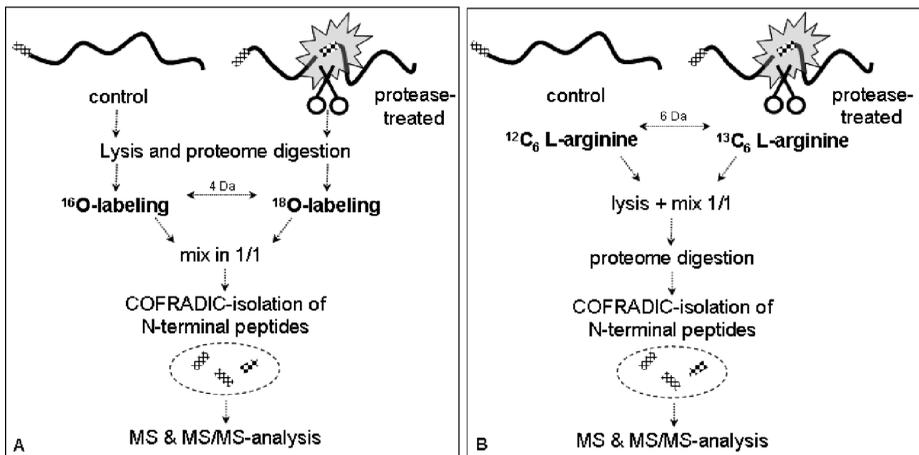
Representative for postmetabolic peptide labeling is proteolytic  $^{18}\text{O}$  labeling by trypsin (**13**). Trypsin catalyzes the exchange of oxygen atoms at the C-terminal carboxyl groups of tryptic peptides and produces in this way labelled peptides that carry two oxygen-18 isotopes at their C-termini. This labeling is introduced following proteome digestion and before chromatographic and mass spectrometric analyses to identify and quantify (relatively) peptides. The primary advantage of this labeling approach is that it is applicable to every proteolytic digest independent of its origin of sampling, whether tissue extractions, body fluids, or cell culture lysates.

Routinely, we also use SILAC (stable isotopic labeling of amino acids in cell cultures; *see Note 15*). SILAC was developed as a simple and accurate approach for MS-based quantitative proteomics (**14**) and relies on the incorporation of essential amino acids with substituted stable isotopic nuclei (D,  $^{13}\text{C}$ , and  $^{15}\text{N}$ ). During the N-terminal COFRADIC protocol, except for the majority of the C-terminal peptides, all peptides end on arginine. Accordingly, heavy form(s) of arginine are the SILAC amino acids to be used since these will introduce (at least) one label per peptide. Interestingly, there are at least three benefits when using  $^{13}\text{C}_6$  or  $^{13}\text{C}_6^{15}\text{N}_4$  L-arginine. First, the spacing between the light and heavy isotopes is increased (6 to 10 Da) as compared to oxygen-16/18 labeling making the determination of abundance ratios straightforward, since peaks are more easily declustered. Second, SILAC labels are very stable during COFRADIC and MS experiments in contrast to the oxygen-16/18 labeling where back-exchange can occur in acidic environments. Finally, triplex experiments may be performed

since  $^{12}\text{C}_6$ ,  $^{13}\text{C}_6$ , or  $^{13}\text{C}_6^{15}\text{N}_4$  arginine forms can be used. The flow path for both labeling strategies is illustrated in **Fig. 2**. One possible flaw is the arginine-to-proline conversion, which can occur in mammalian cells. This results in label dilution in two different peptide forms both representing the heavy form of the peptide (see **Fig. 3**). Thus far, in our hands, in all cell lines tested (including primary cell lines), proline conversion occurs but can be reduced to background levels by reducing the L-arginine concentration to 5–20% of the concentration suggested by manufacturers of cell media, and this without notably affecting cell growth and morphological appearances (see **Note 16**).

### 3.3.1. Peptide Labeling with Oxygen-18 Atoms

1. Step 2 of this protocol is preceded by step 10 of **Subsection 3.2.1**.
2. Following digestion in 10 mM ammonium bicarbonate (pH 7.6), vacuum dry peptide mixtures.
3. Redissolve the peptides in 25  $\mu\text{L}$  of 0.1  $\text{MKH}_2\text{PO}_4$  (pH 4.5) and redry.
4. Add 100  $\mu\text{L}$  of  $^{18}\text{O}$ -rich water (“heavy peptides”) or 100  $\mu\text{L}$  of natural water (“light peptides”) and incubate overnight at 37°C.
5. Transfer 10  $\mu\text{L}$  of the 10 mM TCEP solution to an Eppendorf tube and vacuum dry. Add 10  $\mu\text{L}$  of the 100 mM iodoacetamide solution to 75  $\mu\text{L}$  of 6 M guanidinium



**Fig. 2.** Schematic strategic experimental outline when making use of diverse quantitative proteomic labeling approaches. As outlined, the flow path for sample processing differs when making use of either oxygen-18 (**A**) or SILAC labeling (**B**). When using SILAC, samples can be processed simultaneously, ruling out potential artifacts introduced by parallel processing of samples with postmetabolic oxygen-18 labeling. For oxygen-18 labeling, samples are mixed at the peptide level.

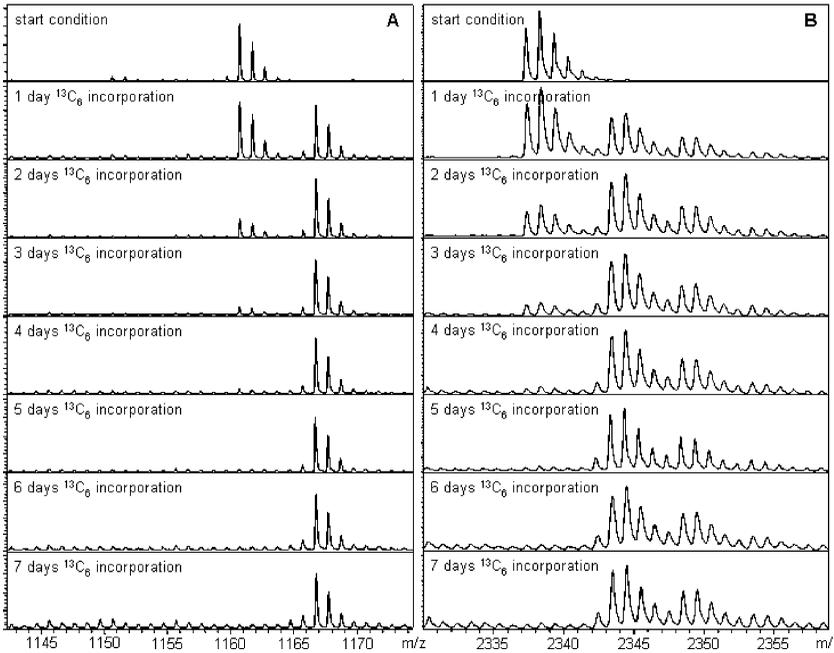


Fig. 3. SILAC labeling strategy in combination with N-terminal COFRADIC. **(A)** The SILAC labeling with  $^{13}\text{C}_6$  L-arginine at various points in time. Jurkat cells were switched to  $^{13}\text{C}_6$  L-arginine-containing RPMI medium on day 0 and samples were obtained on days 0, 1, 2, 3, 4, 5, 6, and 7 during the labeling process. After acetylation, lysates were digested with sequencing-grade modified trypsin and separated on an RP-HPLC. Corresponding fractions in time in the different setups were analyzed by MALDI-MS. The panels show the extent of incorporation of  $^{13}\text{C}_6$  L-arginine into the peptide at the indicated time points. Complete incorporation of  $^{13}\text{C}_6$  L-arginine into proteins was observed in digests obtained from cell lysates harvested on day 5. **(B)** Jurkat cells readily convert  $^{13}\text{C}_6$  L-arginine to  $^{13}\text{C}_5$  L-proline. This results in the formation of two clusters of heavy peptides differing by 5 Da for all proline-containing peptides. The correct weight of the heavy peptides is thus the sum of the  $^{13}\text{C}_6$  L-arginine and the  $^{13}\text{C}_5$  L-proline peak. By reducing the amount of  $^{13}\text{C}_6$  L-arginine, proline conversion was no longer observed.

hydrochloride (this is for 100  $\mu\text{L}$  of  $^{18}\text{O}$ -rich water or natural water to achieve an f.c. for guanidinium hydrochloride of 4 M) in a second Eppendorf tube and dry.

6. Transfer the peptide mixture to the "TCEP vial," mix thoroughly, and incubate at 37°C for 1 h.
7. Transfer the reduced peptide mixture to the "iodoacetamide + Gu.HCl vial" and incubate again for 1 h at 37°C in the dark. At this time point, samples can be stored at -20°C (see **Note 17**).

8. Mix both samples in a 1/1 ratio.
9. Continue with *step 12* of **Subsection 3.2.1**.

### 3.3.2. SILAC Labeling with Heavy Arginine

1. Label the cell population with  $^{13}\text{C}_6$ ,  $^{13}\text{C}_6^{15}\text{N}_4$ , or  $^{12}\text{C}_6$  L-arginine hydrochloride during cell culture at 37°C, 5%  $\text{CO}_2$  for at least five population doublings (usually complete incorporation is achieved after six doublings).
2. Harvest cells from each population and extract proteins as outlined in **Subsection 3.1**.

## 3.4. Sample Mixing

For differential proteomics, mixing of peptide samples in a near 1/1 ratio is favored for meaningful quantification information. As this ratio is based on the total protein amount present in both samples it is important to start with equal sample amounts. Also, small differences in total protein concentration after lysis can be accounted for as indicated in the protocols of **Subsection 3.1** to obtain unavoidable but similar losses of protein material in the following desalting steps.

For  $^{18}\text{O}$ -labeled peptides, the point of sample mixing is fixed in the procedure (after labeling and before the primary COFRADIC run) as described in **Subsection 3.3.1**. As for SILAC-labeled proteins, samples can basically be mixed as early as possible in the protocol (directly after lysis or even before) guaranteeing like treatment of samples. However, to avoid postlysis effects when studying protease substrates it is beneficial to mix samples at a later time point in the procedure when high chaotrope concentration, alkylation (e.g., in the case of cysteine proteases or proteases depending on disulfide bridges for their activity), and acetylation blocked any protease activity. Since most proteases will have lost their activity after one of the modification reactions, samples can then be mixed rather safely before subsequent desalting steps.

In any case, a precise measurement of protein concentration should precede the mixing step. By mixing different sample volumes it is possible to adjust for small differences in protein concentration. However, for both  $^{18}\text{O}$ -labeled peptides and SILAC peptides it is preferable to first mix a small part of the samples in a 1/1 ratio and to use this mixture for a “preprimary” COFRADIC run. Collected primary fractions of this separation are then measured in MS mode. Based on the observed average ratio of the peptide peaks, the mixing volumes of the rest of the samples can be adjusted to obtain a 1/1 ratio.

#### 4. Notes

1. Besides EDTA, these tablets also lack pepstatin A, a generally used inhibitor for aspartic proteases.
2. Since the total protein amount after extraction depends on the cell type and the number of lysed cells or amount of tissue, it is necessary to determine the amount of protein material harvested. A total protein amount of at least 2 mg when using a total extract should be obtained. Correct the buffer volume to get a total protein concentration between 2 and 4 mg/mL taking into account the volume of sample that needs to be loaded on desalting columns. Several types of columns are available, differing in the sample volume applied. The same type of columns should be used during the whole procedure.
3. The elution profile of a peptide depends on the ion pairing agent of the HPLC solvents. In ammonium acetate systems peptides tend to elute at lower concentrations of organic solvent than in TFA systems. For cataloguing proteomes we suggest using TFA as this ion-pairing agent produces extremely sharp peaks and as such a high resolution can be obtained when sorting amino terminal peptides.
4. High concentrated stock solutions are used to avoid large decreases in total protein concentration by volume increase of the sample. The volume of stock solutions added to a given buffer solution should be tested in advance.
5. The conditions of concentration and time for protease incubation should be optimized using alternative techniques (e.g., Western analysis of known substrates to follow their processing in function of time/protease concentration). During optimization and for the final analysis a constant protein (substrate) concentration should be respected (*see also Note 2*).
6. The analysis should be repeated with label swapping between samples. Besides accounting for an extra validation of substrates from a single experiment, repeating the analysis will partly overcome the undersampling problem, which is an intrinsic drawback of mass spectrometers working in automated MS/MS mode due to random selection of peptide ions for fragmentation.
7. Cell treatment can influence both the amount and nature of proteins extracted (e.g., lysis of cells in different phases of cell death). Therefore, it is necessary to determine the extracted protein amount upon stimulation and correct for differences between treated and control samples.
8. When peptides are labeled with oxygen-18, TFA cannot be used as an ion pairing ion in HPLC solvents as this may lead to acid-catalyzed exchange of oxygen atoms in carboxyl groups (*13*). Typically, acetic acid is first added to lower the pH to 5 before injecting peptides onto the RP column, which is run in an ammonium acetate system.
9. The use of hydrogen peroxide to uniformly oxidize methionines to their sulfoxide form is recommended since this prevents accidental hydrophilic shifts of methionyl peptides between chromatographic runs. When performing methionine oxidation prior to the primary RP-HPLC separation (*step 9 of Subsection 3.2.1*) it is important to respect the oxidation time (30 min) and temperature (30°C) since prolonged incubation leads to unwanted and uncon-

trolled oxidation of methionine to methionine sulfone, and the side chain of other amino acids such as cysteine and tryptophan is also oxidized. This implies that following the oxidation step it is necessary to proceed immediately with the RP-HPLC separation of the peptide mixture.

10. Besides N-terminal peptides, other types of peptides are unavoidably cosorted by COFRADIC. Peptides carrying (or acquiring) a blocked, nonacetylated N-terminal amino acid such as a pyrrolidone carboxylic acid or a cyclic S-carbamoylmethylcysteine are cosorted since they do not react with TNBS. Although they appear to “pollute” the mixture of sorted peptides, for differential proteomics purposes their presence can be beneficial as several peptides per protein can be quantified, thus increasing the accuracy of the abundance ratio of their proteins.
11. In theory, N-terminal peptides should elute in the same time frame during the primary and secondary runs. In practice, given the fact that HPLC is not absolutely reproducible, the elution window tends to enlarge and especially abundant N-terminal peptides tend to smear over larger intervals. Therefore, peptides are collected both before (2 min) and after (2 min) their primary collection interval. Since the number of peptides collected in these intervals is much lower than those collected in the expected elution window, such a secondary fraction may be pooled reducing the number of LC-MS/MS analyses.
12. To link MS/MS spectra of COFRADIC-sorted peptide ions efficiently to peptide/protein sequences in databases, search engines such as Mascot (**15**) need to consider the (potential) presence of several modifications on the analyzed peptides. An overview of both the fixed modifications (due to the protein preparation method) and potential (variable) modifications (modifications that are likely to be present in [a part of] the sorted peptides) is presented in **Table 1**. Furthermore, the sequence of a sorted peptide indicating irreversible protein processing is often not exactly predicted by search engines as they do not consider *in vivo* “processing and ragging” of protein (termini). Hence, identification of such peptides may be missed. To overcome such flaws, we constructed DBToolkit (freely available via <http://www.proteomics.be>), an algorithm that uses protein databases as input, imitates protein processing, and creates FASTA-formatted, peptide databases (**16**). Using such peptide-centric databases, we noted an increase of at least 30% of identified MS/MS spectra of N-terminal peptides using Mascot (**3**).
13. In the overall COFRADIC setup the reproducibility of peptide separation is critical. Adequate HPLC instrumentation is now available creating highly reproducible solvent gradients and thus equally reproducible peptide separations. We use Agilent’s electronic flow controller for maintaining a constant solvent flow through the column independent of the backpressure and we thermo-control as many parts of the system as possible (e.g., the column compartment as well as the tubing delivering the solvent to the column and the fraction collector). Taking care of these issues, we generally observe a standard deviation of only a few seconds on the retention time of peptides in a complex peptide mixture over a gradient of nearly 2 h.

**Table 1**  
**Recommended Parameters for Searching Databases with MS/MS Spectra of Peptides Sorted by N-Terminal COFRADIC <sup>a</sup>**

Fixed modifications	Variable modifications
Trideutero-acetylation (K)	Acetylation (N-terminus)
Carbamidomethyl (C)	Trideutero-acetylation (N-terminus)
Oxidation (M)	Deamidation (NQ)
	Oxidation (M)
	Pyrocarbamidomethyl cysteine (C)
	Pyroglutamic acid (N-terminal Q)
Optional fixed modifications	Optional variable modifications
<sup>18</sup> O Labeling	
<sup>18</sup> O C-term (double)	
SILAC labeling	
<sup>13</sup> C <sub>6</sub> L-arginine	<sup>13</sup> C <sub>5</sub> proline*

<sup>a</sup> Since the COFRADIC sorting chemistries lead to additional modifications on sorted peptides we here provide an overview of recommended and essential settings of amino acid modifications when searching databases with engines such as Mascot or SEQUEST.

\*Only when proline conversion occurs.

14. Protease substrates are often characterized by only one identified MS/MS spectrum (“single-hits”). The presence of a trideutero-acetyl group at the  $\alpha$ -amino group of peptides, being present in single isotopic forms, the searched peptide/protein database indicative of the cleavage specificity of the protease of interest, the internal start position and manual validation of identified MS/MS spectra that strictly met the criterium of being ranked one, and scoring above Mascot’s 95% confidence interval score are all making the identification more confident.
15. SILAC cannot be applied for labeling harvested tissue samples, although metabolic labeling of intact species (**17,18**) has been performed.
16. As for some cell lines, the propagation in media containing dialyzed serum (devoid of all substances less than about 10 kDa) may require some optimization, meaning supplementing extra growth factors to the serum.
17. To obtain complete trypsin inactivation the combined action of reductive alkylation under strong denaturing conditions is required.

## Acknowledgments

F.I. is a Research Assistant of the Fund for Scientific Research–Flanders (Belgium) F.W.O.–Vlaanderen).

## References

1. Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003) *Nat. Biotechnol.* **21**, 566–569.
2. Meuleman, P., Libbrecht, L., De Vos, R., de Hemptinne, B., Gevaert, K., Vandekerckhove, J., Roskams, T., and Leroux-Roels, G. (2005) *Hepatology* **41**, 847–856.
3. Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005) *Nat Methods* **2**, 771–777.
4. Gevaert, K., Van Damme, J., Goethals, M., Thomas, G. R., Hoorelbeke, B., Demol, H., Martens, L., Puype, M., Staes, A., and Vandekerckhove, J. (2002) *Mol. Cell. Proteomics* **1**, 896–903.
5. Gevaert, K., Van Damme, P., Ghesquiere, B., and Vandekerckhove, J. (2006) *Biochim. Biophys. Acta* **1764**, 1801–1810.
6. Kuhn, K., Thompson, A., Prinz, T., Muller, J., Baumann, C., Schmidt, G., Neumann, T., and Hamon, C. (2003) *J. Proteome Res.* **2**, 598–609.
7. McDonald, L., Robertson, D. H., Hurst, J. L., and Beynon, R. J. (2005) *Nat. Methods* **2**, 955–957.
8. Ji, J., Chakraborty, A., Geng, M., Zhang, X., Amini, A., Bina, M., and Regnier, F. (2000) *J. Chromatogr. B Biomed. Sci. Appl.* **745**, 197–210.
9. Szeltner, Z. and Polgar, L. (1996) *J. Biol. Chem.* **271**, 5458–5463.
10. Beynon, R. J. and Pratt, J. M. (2005) *Mol. Cell. Proteomics* **4**, 857–872.
11. Mann, M. (2006) *Nat. Rev. Mol. Cell. Biol.* **7**, 952–958.
12. Miyagi, M. and Rao, K. C. (2007) *Mass Spectrom. Rev.* **26**, 121–136.
13. Staes, A., Demol, H., Van Damme, J., Martens, L., Vandekerckhove, J., and Gevaert, K. (2004) *J. Proteome Res.* **3**, 786–791.
14. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) *Mol. Cell. Proteomics* **1**, 376–386.
15. Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C. P., Plasterk, R. H., and Heck, A. J. (2003) *Nat. Biotechnol.* **21**, 927–931.
16. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., and Yates, J. R., 3rd (2004) *Anal. Chem.* **76**, 4951–4959.
17. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) *Electrophoresis* **20**, 3551–3567.
18. Martens, L., Vandekerckhove, J., and Gevaert, K. (2005) *Bioinformatics* **21**, 3584–3585.