

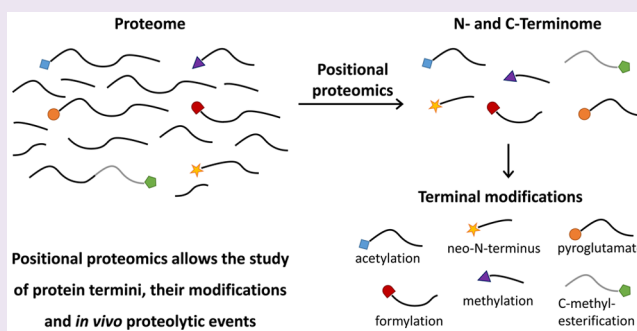
Protein Termini and Their Modifications Revealed by Positional Proteomics

Giada Marino,[†] Ulrich Eckhard,[†] and Christopher M. Overall^{*,†,‡}

[†]Centre for Blood Research, Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, British Columbia, Canada

[‡]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT: A myriad of co- and post-translational modifications occur at protein N- and C-termini, resulting in an extra layer of proteome complexity and an additional source of protein regulation. Here, we review N- and C-terminal modifications and the contemporary positional proteomics techniques used to isolate protein terminal peptides from complex protein mixtures and characterize their diversity and occurrence in biological systems. Furthermore, these degradomics strategies—often referred to as N- and C-terminomics—represent dedicated high-throughput techniques to study proteolysis in dynamic living systems. Over the past decade, terminomics studies have provided indispensable information on the functional states of individual proteins, cell types, tissues, and biological processes and delivered fundamental new data for the Human Proteome Project, including high confidence identifications of many so-called “missing proteins”, which had not been identified by traditional proteomics analyses.



As the number of expected protein-coding genes in humans has declined from over 100 000 to around 20 061 protein entities (www.neXtProt.org, release 2015-01-01) in the past 15 years,¹ proteins have stepped into the spotlight as key players to define the functional state of individual cells. It has been estimated that the human genome is translated into 1 billion protein isoforms, dramatically expanding the encoded flexibility of a living system.² These so-called proteoforms are defined as highly related protein molecules arising from a single gene by combinations of variation from all sources, including genetic variations, alternative splicing, and co- and post-translational modifications (PTMs; Figure 1).³ Attaching functional groups alters the physicochemical properties and overall structure of proteins, affecting their stability, localization, and function; the same holds true for proteolytic processing. Importantly, post-

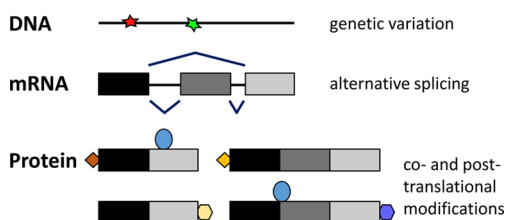


Figure 1. Proteoforms (protein isoforms) can arise from various biological processes, such as genetic mutations, alternative transcription and translation start sites, alternative splicing during transcription, and co- and post-translational modification at the protein level.

translational modifications provide a cell with an inherent flexibility to rapidly respond and adapt to internal and external stimuli. Even though proteins exist in a multitude of sizes and amino acid compositions, all possess a distinct start and end with specific biochemical properties and function, known as the N- and C-terminus, respectively, and collectively forming the terminome. Both terminal sites undergo general and distinct chemical modifications that are utilized and required in a variety of biological functions and processes such as protein sorting, membrane integration, protein transit, enzyme activity, and formation of protein complexes.^{4–6}

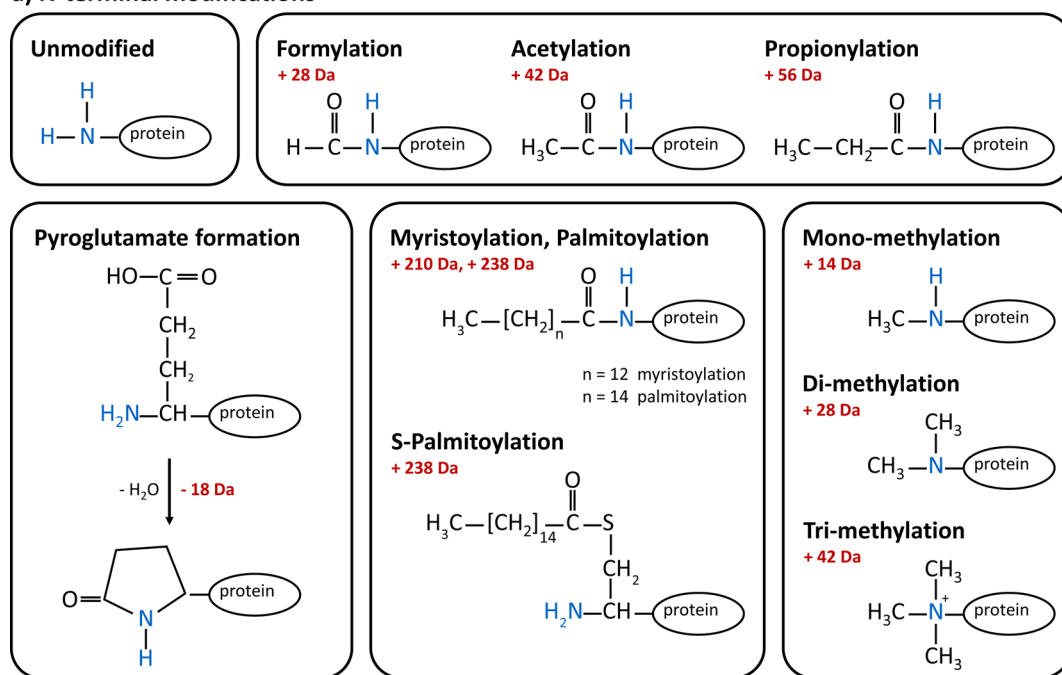
N-Terminal Modifications. N-terminal protein modifications occur either during protein synthesis with the polypeptide chain still bound to the mRNA-translating polyribosomes or after termination of protein synthesis. Hence, they are defined as co- and post-translational modifications. Co-translational modifications such as N-terminal methionine excision, N α -acetylation, N α -myristoylation, and N-terminal methionine deformylation of mitochondrial proteins in humans (and of bacterial and chloroplast proteins) are catalyzed by non-ribosomal proteins functioning as soon as the nascent polypeptide chain reaches the end of the ribosomal exit tunnel.⁷ For a detailed description of these processes, please refer to the excellent review by Giglione et al.⁸ After protein synthesis and proper folding, both N- and C-termini tend to be

Received: March 17, 2015

Accepted: June 4, 2015

Published: June 4, 2015

a) N-terminal modifications



b) C-terminal modifications

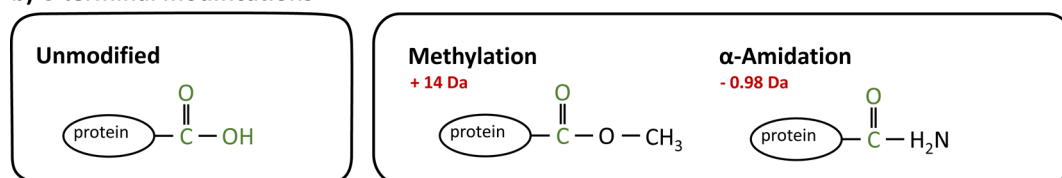


Figure 2. Schematic representation of selected co- and post-translational (a) N-terminal and (b) C-terminal modifications. (a) N-terminal α -amines can be either neutral or positively charged depending on the surrounding pH and chemical modifications. The most prominent protein terminal modification is N α -acetylation catalyzed by N-acetyl-transferases that also have the ability to formylate and propionylate protein N-termini. Pyroglutamate, commonly observed in antibodies and making them more resistant to aminopeptidases, forms through cyclization of N-terminal glutamine or glutamate, either spontaneously or enzymatically by glutaminyl cyclases. Palmitoylation and myristoylation can occur at free N-terminal glycines, and N-terminal cysteines can be palmitoylated (S-palmitoylation) at their side chain. Finally, N-terminal mono-, di- and tri-methylation modifications are important for protein-protein interaction. The monoisotopic mass shifts are according to UniMod:¹⁰¹ +27.9949 Da (formylation), +42.0106 Da (acetylation) and +56.0262 Da (propionylation), whereas pyroglutamate formation results in a mass loss of −18.0106 Da (from glutamate) or −17.0265 Da (from glutamine). For myristoylation and palmitoylation, +210.1984 Da and +238.2297 Da are observed, and for mono-, di, and tri-methylation: +14.0156 Da, +28.0313 Da, and +42.0470 Da. (b) C-terminal α -amidation neutralizes the negative charge of the carboxyl group at the C-terminus, improving stability and receptor binding ability of peptides and peptide-hormones. C-methyl-esterification is the most frequently annotated C-terminal modification as reported in TopFIND database. According to UniMod, the monoisotopic mass shifts are +14.0156 Da (methylation) and −0.9840 Da (α -amidation).

exposed in protein structure and are available for post-translational modification (Figure 2).⁹ Currently, the Termini-oriented protein Function INferred Database (TopFIND)^{10–12} together with UniMod (www.unimod.org) report 11 N-terminal modifications¹³ (number of instances reported in TopFIND are given in parentheses when available): mono-, di-, and trimethylation (5, 6, and 20, respectively), formylation (9), carbamylation, succinylation, cyclization (345), propionylation, palmitoylation, myristoylation, and the most prominent, N α -acetylation (9479).

Initially thought to be only cotranslational, N α -acetylation is now also recognized as a widespread post-translational modification,^{13–19} playing an important role in modulating protein function, localization, and protein stability and turnover, as described by the destabilizing N-end rule^{20–22} and the stabilizing variant of the N-end rule of protein neo-N-termini.⁶⁷ The first amino acid of a cytosolic protein chain and

its modification define an N-terminal degradation signal called the N-degron, that determines the half-life of a protein *in vivo* by increasing or decreasing the likelihood of ubiquitination and degradation of the polypeptide by the proteasome.²³ Correct protein folding prevents recognition of a destabilizing residue and hence the N-end rule executes important protein quality control functions. Notable recent studies showed that N-acetyl-transferases also have the capacity to formylate and propionylate protein N-termini *in vitro* and *in vivo*, suggesting functional competition, although a biological implication for these PTMs has yet to be determined.^{24,25}

Even though N α -methylation has been documented for almost 40 years,²⁶ the first physiological role was described only in 2007 when Chen *et al.* revealed its importance in protein–DNA interactions.²⁷ Since many N α -methylated proteins are found in large multisubunit complexes, a role in the regulation of protein–protein interactions was also suggested.²⁸ Recent

experiments support an additional role in modulating protein stability, thus interplaying with N α -acetylation.²⁹ N-terminal cyclization, catalyzed *in vivo* by two glutaminyl cyclases, occurs at terminal glutamine and glutamate residues leading to the formation of pyroglutamic acid. This reaction also proceeds spontaneously at much lower rates,³⁰ but whether catalyzed or not, pyroglutamate formation is always faster for glutamine than glutamate.³¹ Cyclization protects many cytokines, antibodies, and structural proteins from protease degradation, e.g. fibrin, fibrinogen and collagen-like proteins by aminopeptidase trimming. Cyclization is essential for the function of many peptides such as the hormones gastrin and thyrotropin-releasing hormone (TRH), the neuropeptide neurotensin, and proteins such as chemokines.^{30,32} Of note, in several neurodegenerative disorders N-terminal pyroglutamate formation in neuronal peptides results in a higher tendency to aggregate and form plaques due to increased hydrophobicity and proteolytic stability, as shown for beta-amyloid (A β) deposition in Alzheimer's disease.³³ Protein N-termini can also be subject to lipid post-translational modification, where fatty acid chains (most frequently myristate or palmitate) are attached to free N-terminal glycines or to the side chain of N-terminal cysteines.^{34,35} Both myristylation and palmitylation are involved in mediating signal transduction and protein trafficking³⁶ and are implicated in the development and progression of several pathologies including cancer, neurodegenerative disease, and in viral and bacterial infection.^{34,37,38}

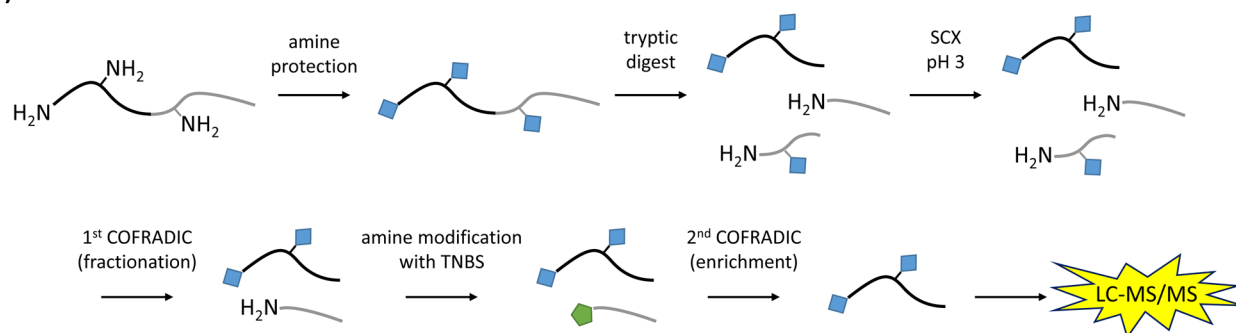
C-Terminal Modifications. Often neglected in analyses, the protein C-terminal residue can also be post-translationally modified (Figure 2b). However, due to the lower reactivity of the C-terminus in comparison to the N-terminus and therefore an inherent greater difficulty in identifying C-termini biochemically or by proteomics, these modifications are presumed infrequent, yet this may be a misnomer stemming from intractable analysis.^{39,40} Indeed, there are well-described examples in the literature of C-terminal PTMs, such as the addition of glycosylphosphatidylinositol (GPI) as a membrane anchor after proteolytic removal of a C-terminal propeptide, protein prenylation, or protein carboxyl methylation.^{41,42} With an isoprenyl group at the C-terminus, cytosolic proteins localize at the plasma membrane or organelle membranes. Extensive studies on oncogenic RAS proteins (NRAS, HRAS, and KRAS) have shown that farnesylation and geranylgeranylation (together referred to as prenylation) have a key role in tumor cell survival, growth, proliferation, migration, and metastasis.⁴³ Additionally, over half of all biologically active peptides and peptide hormones are C-terminally α -amidated, which is essential for their full biological activities.⁴⁴ The α -amidation neutralizes the negative charge of C-termini, resulting in more hydrophobic peptides and improved receptor binding for such proteins.⁴⁵ Studies on G-protein-coupled peptide receptors showed by NMR and receptor binding assays that the ligand's amide moiety is a key binding determinant.^{46,47} The most frequently annotated C-terminal modification reported in the knowledgebase TopFIND is C-methyl-esterification; however only 17 instances are reported in humans, six in mouse, and seven in yeast.¹³ One example is the methylation of the C-terminal leucine in the catalytic subunit of phosphoprotein phosphatase 2A (PP2A), which is required for the interaction with its regulatory subunit and is involved in modulating neural plasticity.^{48,49} It is obvious that much more investigation needs to be performed on the elusive C-termini. Such analyses will be facilitated by our recent introduction of a new protease for

proteomics sample preparation, LysargiNase, which mirrors trypsin selectivity and hence generates C-terminal peptides with a positive arginine or lysine residue at their N-terminus, thus greatly improving detection and identification of protein C-terminal peptides (see Enrichment of Protein C-Termini for more details).⁵⁰ With the advent of LysargiNase for shotgun analyses and positional proteomics techniques, such as C-TAILS (Carboxy-Terminal Amine-based Isotope Labeling of Substrates),³⁹ it is anticipated that the focus and understanding of protein C-termini will greatly increase in the near future.

Proteolytic Processing As a Function-Altering Post-Translational Modification. Post-translational modifications at protein termini can occur after specific proteolytic processing of a fully translated protein at internal sites.⁵¹ For instance, during protein maturation, removal of the signal peptide or the pro-domain by protease cleavage generates neo N-termini that can be subjected to N α -acetylation or cyclization.^{52,53} Exopeptidases typically trim the polypeptide chain by one residue or processively by up to three amino acids from the N- (aminopeptidases) or C-terminus (carboxypeptidases), thereby creating ragged termini, whereas endopeptidases precisely cleave within the protein, generating two new polypeptide chains, one possessing a neo-C-terminus, one a neo-N-terminus. As a post-translational modification, limited proteolysis, often referred to as proteolytic processing, is involved in modulating protein activity, function, and localization in several ways.⁵⁴ (i) In proteolytic cascades such as blood coagulation,⁵⁵ complement activation,⁵⁶ developmental signaling pathways,⁵⁷ and apoptosis,⁵⁸ proteases act as irreversible molecular switches, turning proteolytic activities either on or off; (ii) Protease cleavage can flip the activity of the target protein,⁵⁹ e.g., N-terminal truncation of the monocyte chemoattractant proteins CCL2 and CCL7 by matrix metalloproteases (MMPs) results in their conversion from receptor agonists to antagonists, and virtually all 54 human chemokines are now known to exhibit modified bioactivity including activation, inactivation, switching of receptor specificity and loss of glycosaminoglycan binding upon MMP cleavage;^{60,61} (iii) Specific proteolysis of proteins can release cryptic molecules or "neo-proteins" with a biological role completely different from the parental protein. Examples include the generation of angiogenesis inhibitors such as angiostatin from the plasmin precursor plasminogen, and the release of endostatin and vasostatin from the structural extracellular matrix protein type XVIII collagen and the calcium-binding chaperone protein calreticulin, respectively;⁵⁹ (iv) Shedding, the proteolytic release of extracellular domains (ectodomains) from cell membrane proteins such as cell adhesion molecules, growth factors, transmembrane cytokine precursors and receptors, and cell surface-anchored ectoenzymes, is facilitated by members of the ADAM (A Disintegrin And Metalloproteinase) and MMP protease families. Because of the broad variety of biological targets, shedding is involved in diverse processes such as signal transduction, cell adhesion, and protein relocalization.^{62,63}

Proteolytic processing is an irreversible process and occurs on almost all proteins, thus constituting a major, possibly the most universal, yet often overlooked PTM. Limited proteolysis plays key roles in most essential cellular processes, ranging from cell cycle division and proliferation to cell death.⁶⁴ It is also involved in the progression of many diseases including cancer, inflammation, and atherosclerosis, progressively highlighting proteases and their substrates as important diagnostic and therapeutic targets.^{65,66} Recent N-terminomics studies highlight

a) COFRADIC



b) TAILS

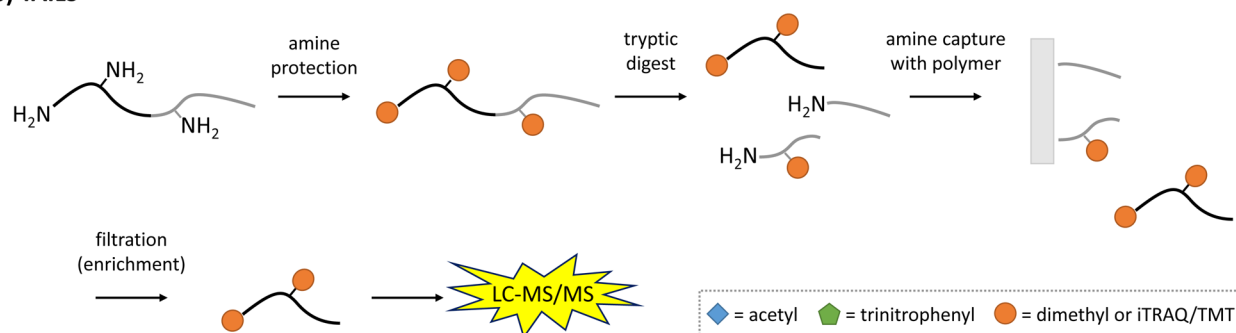


Figure 3. Schematic representation of N-terminomics COFRADIC and TAILS. (a) In COFRADIC, after proteome reduction and alkylation, all primary amines are blocked by acylation. Following proteolytic digestion, the sample is pre-enriched for N-terminal peptides by SCX chromatography. After a first fractionation by RP-HPLC, each fraction is treated with TNBS to chemically modify free α -amines of unwanted C-terminal and internal peptides. A second RP-HPLC run is used to separate the TNBS-modified peptides from N-terminal peptides, which are analyzed by LC-MS/MS. (b) In TAILS, after proteome reduction and alkylation, all primary amines are blocked by dimethylation or isobaric tags. Following proteolytic digestion, the internal tryptic peptides are captured with the HPG-ALD polymer and depleted by filtration. The blocked N-terminal peptides are collected and analyzed by LC-MS/MS. Refer to the main text for a detailed description of these strategies.

the pervasiveness of proteolytic processing *in vivo* by revealing that 61%, 64%, and 77% of all identified termini in murine skin,⁵³ human erythrocytes,⁶⁷ and human platelets,⁶⁸ respectively, differ from the genetically encoded termini and originate largely from proteolytic cleavage. The origin of N-termini in the human proteome has been recently analyzed by computational analysis on a global scale, finding that ~24 000 N-termini in the human proteome differ from the canonical encoded or methionine processed N-terminus.⁶⁹ These recent surprising analyses highlight the need to better understand the identity, nature, and functional implications of N- and C-terminal proteoforms in proteomes.

METHODS FOR CHARACTERIZING THE TERMINOME

Several positional proteomics techniques have emerged in the past decade for the identification and characterization of protein termini. These high-throughput mass spectrometry-based techniques have enabled experimental determination of protein N-termini and C-termini and their modifications in complex *in vivo* samples, as demonstrated for example by the published N-terminomes of skin,⁵³ blood plasma,⁷⁰ mitochondria,⁷¹ and red blood cells.⁶⁷ Additionally, quantitative N-terminal changes were investigated during apoptosis and upon platelet storage under blood banking conditions.^{68,72–74} Given the higher reactivity of primary amines over carboxyl groups, positional proteomics initially focused on identifying N-terminal peptides, with C-terminomics approaches slow to develop. Over the past decade, several laboratories have

developed various strategies to isolate and characterize protein termini by positive or negative enrichment of N- or C-termini from highly complex protein mixtures (for recent reviews, see refs 75 and 76). All of these strategies follow a general bottom-up proteomics approach where chemical labeling occurs at the protein level before the sample is proteolytically digested, thus unambiguously preserving terminal information. However, the negative selection procedures, TAILS^{77,78} and COFRADIC,^{52,78} enjoy the major advantage of also identifying naturally modified N-termini, that positive selection procedures (that rely upon chemical addition of handles to enrich) cannot capture, in addition to protease cleavage products.

Negative Enrichment of Protein N-Termini. The two most widely reported methods for enriching N-terminal peptides by negative selection are Combined FRActional Diagonal Chromatography (COFRADIC) and the more recent Terminal Amine Isotope Labeling of Substrates (TAILS; Figure 3). COFRADIC was the first successful method of negative enrichment of protein N-termini and was developed by the Gevaert laboratory.⁵² After reduction and alkylation of the proteome, all primary amines are blocked (typically by acylation), and proteins are subsequently digested with trypsin, or any other highly specific endoproteinase.⁷¹ Trypsin digestion thus results in two main classes of peptides: N-terminal peptides characterized by an acetylated α -amino group and internal peptides possessing a free N-terminus. At this point, the peptide mixture can be incubated with glutamine cyclotransferase and pyroglutamyl aminopeptidases, converting N-terminal glutamine into pyroglutamate, followed by

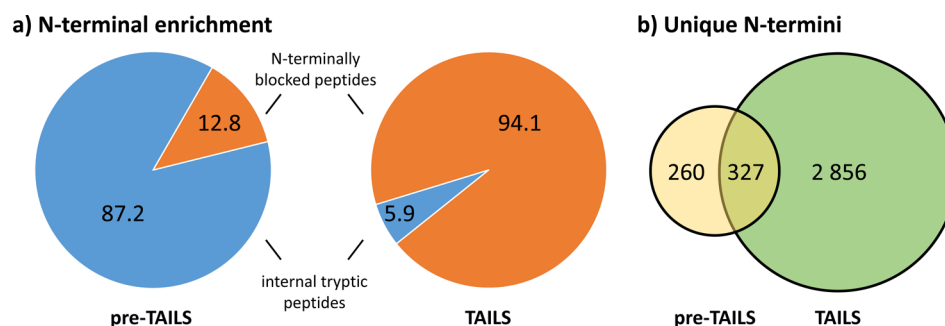


Figure 4. Illustration of N-terminal peptide enrichment by TAILS N-terminomics of human gingival tissue. (a) Percentage of N-terminally blocked and trypsin generated internal peptides identified before (pre-) and after TAILS enrichment. The 5.9% internal tryptic peptides can vary from 2 to 7% depending on the sample and generally represents peptide:peptide interactions that enable carry-over of such peptides to the unbound N-terminal peptide enriched fraction. As such, they are easily recognized based on the lack of blocked N-termini. (b) In total, 3443 unique N-terminal peptides were identified at a false discovery rate of 1%. Note: N-termini refers here to peptides possessing either a naturally (acetylated) or experimentally (dimethylated) blocked N-terminus, omitting peptides with a cyclized N-terminus. The mass spectrometry raw data associated with this figure are in preparation for submission and are available upon request now.

proteolytical removal, thus minimizing numbers of spontaneously formed pyroglutamyl peptides.⁷⁹ Afterward, the resulting peptide mixture is pre-enriched for N-terminal peptides by strong cation exchange (SCX) chromatography. At pH 3, internal peptides preferentially bind to the SCX resin, whereas the majority of N-terminal peptides with blocked α -amino groups and C-termini are recovered in the flow-through due to their higher hydrophobicity. After a first fractionation by reversed-phase liquid chromatography (RP-HPLC), each fraction is further subjected to treatment with 2,4,6-trinitrobenzenesulfonic acid (TNBS) to chemically modify the α -amines of unwanted C-terminal and internal peptides, dramatically increasing their hydrophobicity. Consequently, additional RP-HPLC runs are required to separate the TNBS-modified internal and C-terminal peptides from the now comparatively less hydrophobic N-termini, which are forwarded to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. One advantage of COFRADIC is that all the required materials are commercially available and relatively inexpensive. The extensive fractionation steps provide several opportunities for sample loss and make this technique with up to 100 MS/MS analyses per sample very instrument intensive. However, the latest version of the protocol reports a substantial improvement by introducing a sample pooling scheme to reduce the overall number of LC-MS/MS analyses to 36.⁸⁰ Nonetheless, compared to single or double shot analyses by TAILS, this is a disadvantage in terms of the more complex chemistry, workflow, and the increased amount of protein required for successful analysis. Nevertheless, it is a highly successful and elegant approach.

The Zahedi laboratory has recently proposed a variation of the COFRADIC approach, termed Charged-based Fractional Diagonal Chromatography (ChaFRADIC).⁸¹ After proteome reduction and alkylation, all primary amines are blocked and isotopically labeled by dimethylation. The sample is then digested using trypsin, and the peptide mixture is separated by SCX at pH 2.7. Five fractions corresponding to charge states +1, +2, +3, +4, and >+4 are collected and treated with NHS trideutero-acetate, acetylating free N-termini of trypsin-generated internal peptides and thus reducing the peptide charge state. Each fraction is then subjected to a second round of SCX chromatography; peptides retaining the original charge states and thus corresponding to original and neo-protein N-termini are collected and used for subsequent LC-MS/MS

analysis. Thereby, this approach is less time-consuming and requires a smaller amount of starting material than COFRADIC.

TAILS N-terminomics is the other highly successful negative selection strategy introduced by our laboratory.^{77,78} In TAILS, after reduction and cysteine alkylation, the alpha and epsilon primary amines are blocked by dimethylation or isobaric tags like iTRAQ (isobaric Tags for Relative and Absolute Quantification,⁸² Sciex) or TMT (Tandem Mass Tags,⁸³ Thermo) at the whole protein level. This is an important difference from peptide level labeling performed after proteome digestion in shotgun experiments. As any free N-terminus is blocked and labeled at the protein level, their identification together with naturally blocked N-termini, by means of MS/MS, confirms the presence of these N-termini in the proteome sample. Following digestion with a highly specific endoprotease, such as trypsin or GluC, the peptide mixture is incubated with a commercially available (<http://www.flintbox.com/public/project/1948/>), water-soluble, high-molecular-weight, aldehyde-derivatized, hyperbranched polyglycerol (HPG-ALD) polymer. The HPG-ALD polymer covalently binds only the internal peptides generated postlabeling that carry a free α -amine at their N-terminus, whereas the natural and neo-N-termini have been naturally blocked or blocked by the isotopic labeling reagent and so are unreactive.^{77,78} Separation of internal peptides bound to the HPG-ALD polymer is achieved by ultrafiltration, where the flow-through fraction containing the unbound N-terminal peptides is collected and analyzed by LC-MS/MS. Advantages of TAILS include the small amount of starting material required (as low as 200 μ g) and the low number of MS/MS runs required (typically 1 or 2 for technical replicates). In a typical TAILS experiment, approximately 95% of the identified peptides possess a blocked N-terminus, either naturally acetylated, experimentally labeled, or cyclized, whereas samples without enrichment show around 12% blocked termini (Figure 4). The pharmaceutical company Roche modified the TAILS procedure by omitting the filtration step as a possible source of sample loss. They described applying the polymer-peptide mixture directly onto the trap column of their LC-MS/MS system.⁸⁴ Of note, we are currently developing a new polymer, containing paramagnetic particles anchored to HPG-ALD, to improve separation efficiency and speed (*unpublished*). An approach similar to TAILS is DICAS (Dimethyl Isotope-Coded Affinity Selection), which employs

an immobilized matrix (POROS-AL, Life Technologies) instead of a water-soluble polymer to capture and remove the nonterminal peptides.⁸⁵ Polymers for proteomics retain the distinct advantage of inherently zero nonspecific peptide binding and loss.

Recently, capitalizing on their extensive experience in phosphoproteomics, the Heck laboratory has also developed a modified form of the TAILS N-terminomics technique based on phospho tagging (PTAG). After blocking all primary amines by dimethylation, the complex protein sample is digested using trypsin. The trypsin-generated N-termini are tagged using glyceraldehyde-3-phosphate and depleted via titanium dioxide chromatography before MS/MS analyses.^{86,87}

Positive Enrichment of Protein N-Termini. Positive selection methods modify the α -amino group of the protein N-terminus with a tag that enables targeted enrichment of the N-terminal peptide from the majority of internal peptides and protein C-termini. Wells and co-workers have developed an engineered propriety subtilisin protease, subtiligase, to selectively label unblocked N-terminal α -amines with a biotinylated peptide ester containing a tobacco etch virus protease (TEV) cleavage sequence. After enzymatic labeling and tryptic digestion of the sample, the biotin-tagged N-terminal peptides are immobilized on avidin beads. Following TEV cleavage, the N-termini are released but retain a Ser-Tyr-dipeptide as a signature, simplifying identification by LC-MS/MS. However, large amounts (50 to 100 mg) of starting material are required, thus limiting potential applications such as cell culture secretome analyses.^{72,88} Another successful strategy based on positive enrichment is called N-CLAP (N-terminomics by Chemical Labeling of the α -Amine of Proteins). This method enables the selective labeling and capture of the α -amines present only on N-termini utilizing features of Edman chemistry; after blocking all primary amines using phenyl isothiocyanate (PITC), treatment with trifluoroacetic acid (TFA) specifically triggers cyclization of PITC-modified α -amines, resulting in peptide bond cleavage after the first amino acid, whereas PITC-modified lysines are left unaffected. The new α -amines are subsequently labeled with a thiol-cleavable amine-reactive biotinylation reagent. After digestion of the complex sample, biotinylated peptides are captured on avidin-based affinity resins, eluted by reduction, and analyzed by LC-MS/MS.⁷³ However, as already noted, positive selection is limited to protease cleavage product identification, whereas the negative selection procedures gather information on both the natural N- and neo-N-terminus. In this way, negative selection techniques identify more than 50% of proteins by two or more peptides, corresponding to the natural N-terminus and/or one or more neo-N-terminal peptides generated by proteolytic processing, thus increasing protein identification confidence.^{77,89}

Enrichment of Protein C-Termini. Due to the lower chemical reactivity of the α -carboxyl group and the lack of methods to selectively modify carboxyl groups in aqueous solution, C-terminal proteomics is a highly challenging task. One of the first attempts to isolate C-termini used SCX chromatography as reported by the Heck laboratory in 2007.¹⁸ However, this approach enriches also for N-acetylated peptides as both show only weak interactions with the resin. Nonetheless, it represents a straightforward approach to assess the C-terminome. Other more sophisticated methods to enrich for C-termini represent modified versions of TAILS and COFRADIC. Reported in back to back publications in 2010,^{39,40} they

have been used successfully for the high-throughput analysis of protein C-termini. In C-TAILS (Carboxy-Terminal Amine-based Isotope Labeling of Substrates), all primary amines are dimethylated, and the carboxyl groups are protected with ethanolamine using EDAC (1-ethyl-3-(3-(dimethylamino)-propyl) carbodiimide) chemistry. After trypsin digestion, the newly generated α -amines are blocked by a second round of dimethylation. The internal and N-terminal peptides are removed by covalent coupling to a high molecular weight polyallylamine polymer from Sigma via their unblocked C-termini. The C-TAILS approach was the first to incorporate isotope-encoded tags to provide quantitative data, which is essential to distinguish neo-termini derived from a specific protease versus cleavage products present in every sample *in vivo* or inadvertently generated in sample handling, though the latter is essentially eliminated by immediate use of universal protease inhibitors at sample collection and careful sample handling.³⁹ Similar to TAILS, the polymer with covalently bound internal peptides is removed by filtration, and the unbound ethanolamine blocked protein C-terminal peptides are analyzed by LC-MS/MS.³⁹

In C-terminal COFRADIC, N-terminal and C-terminal peptides are enriched simultaneously. After blocking all primary amines by acylation, the proteome is digested with trypsin. After SCX-enrichment, the TNBS used in the N-terminal COFRADIC workflow is replaced by an N-hydroxysuccinimide ester of butyric acid to chemically derivatize the free primary amines present on the C-terminal peptides, allowing segregation of N- and C-terminal peptides by RP-HPLC prior to LC-MS/MS analysis.⁴⁰ Besides carboxy-terminal COFRADIC and C-TAILS, several other strategies have been proposed, e.g. using TMPP tags,^{90,91} enzymatic labeling,⁹² or cyanogen bromide digestion.⁹³ For a detailed description and comparison of these approaches, please refer to the recent review by Tanco *et al.*⁴¹

Despite these advances, problems remain. After trypsin digestion, protein C-terminal peptides are rarely identified by LC-MS/MS, as they inherently lack basic residues, and thus recalcitrant to mass spectrometry identification.^{39,40} This year, our laboratory reported the discovery of a new protease for proteomics named LysargiNase, a thermostable metalloprotease from the Archaea *Methanosarcina acetivorans*.⁵⁰ LysargiNase improves protein C-termini identification by cleaving *before* lysine and arginine residues, the opposite to trypsin. In four independent experiments, Huesgen *et al.* showed a tremendous improvement of protein C-termini identification compared to trypsin.⁵⁰ Unlike trypsin, LysargiNase cleavage also occurs at methylated, dimethylated, and trimethylated lysines, so proving highly useful for improved phosphosite and epigenetic mark identification and overall proteome coverage.⁵⁰

3. APPLICATIONS

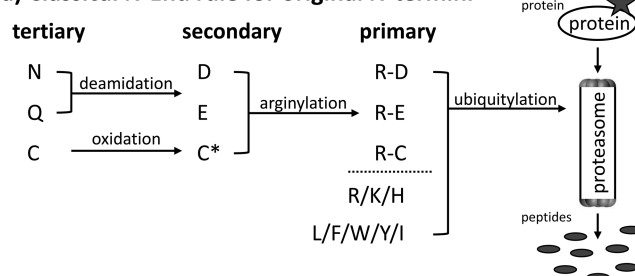
The choice of method determines the type of terminal modifications that are to be studied. By design, positive enrichment strategies exclude *in vivo* modified N-terminal peptides such as N α -acetylated peptides but solely focus on proteolytic cleavage products. In contrast, negative selection strategies additionally cover naturally blocked N-termini and allow for their quantification through isotopic labeling of all lysine side-chains. If metabolic SILAC labeling is used for MS1 quantification, even nonlysine containing terminal peptides will be quantifiable. However, peptide numbers are much greater for multiplex isobaric labeled peptides; e.g., recently we

identified more than 8500 high confidence N-terminal peptides using 10-plex TMT labels without prefractionation (Klein et al. submitted), which we had never approached using MS1 quantification methods. It is beyond the scope of this review to cover all the possible applications of the terminomics strategies described above. Hence, we focus here on the recent studies for characterization of protein N-termini.

N α -acetylation is the most abundant N-terminal modification, occurring on approximately 90 and 60% of all soluble proteins in humans and yeast, respectively.⁹⁴ Van Damme and colleagues recently compared the N-terminal acetylation profiles between human and yeast using COFRADIC.⁹⁵ They identified 1,345 (human) and 648 (yeast) unique acetylated N-termini and described the N-terminal acetyltransferase NatF as one of the key drivers for the significant shift in N-terminal acetylation occurrence from lower to higher eukaryotes. The presence of NatF in higher eukaryotes boosts their capacity to acetylate Met-Lys-, Met-Leu-, and other Met-starting N-termini, and introduction of NatF into yeast increased N-terminal acetylation levels significantly.⁹⁵ In another study, the same group identified 180 human and 110 yeast substrates of NatB and established the requirement of NatB activity for maintaining the structure and function of actomyosin fibers and proper cellular migration in humans.¹⁷ N-terminal acetylation mainly occurs on the initiator methionine or on the first residue after methionine removal by methionine aminopeptidases. Recently, our lab analyzed the amino acid distribution of N-terminal acetylated peptides and compared it with the specificity preference of methionine aminopeptidases.^{53,67,68} N-terminal methionine excision is preferred when the second amino acid is either alanine, glycine, or serine, and the same residues are favored targets for N-terminal acetylation, indicating concerted cotranslational action. Of note, the presence of proline at the primary or secondary position of the protein N-terminus prevents acetylation and thus ensures a free N-terminus.^{53,96} In contrast, the presence of glutamate, aspartate, or phenylalanine at position 2 leads to acetylation of the initiator Met, in perfect agreement with previous *in vitro* studies.^{53,68} Notably, different sequence dependencies were detected for acetylated peptides starting at positions other than 1 and 2 and not preceded by methionine, suggesting another set of enzymes for post-translational N α -acetylation. These different sequence dependencies may allow the identification of alternative translation sites based on their N-terminal acetylation profile.^{53,68}

Recently, we found that post-translational N α -acetylation is an efficient mechanism to either stabilize or destabilize intracellular proteins after proteolytic processing and allow functional diversification of the proteome under conditions of limited *de novo* protein synthesis, as occurs in mature erythrocytes or platelets.^{67,68} Based on the observed amino acid frequency and N α -acetylation status of internal protein N-termini in the TAILS analysis of human erythrocytes, a modified N-end rule was developed, namely, “the stabilizing N-end rule of processed protein termini.” According to the original N-end rule, amino-terminal amino acids can be grouped as primary, secondary, and tertiary destabilizing residues (Figure 5a).⁹⁷ Primary destabilizing residues are directly recognized by components of the ubiquitin system (N-recognins) and are targeted for degradation by the proteasome. These residues are further subdivided into types 1–2 based on the dominant N-recognin responsible for their targeting. Secondary destabilizing residues require arginylation

a) Classical N-End rule for original N-termini



b) Stabilizing N-End rule for processed N-termini

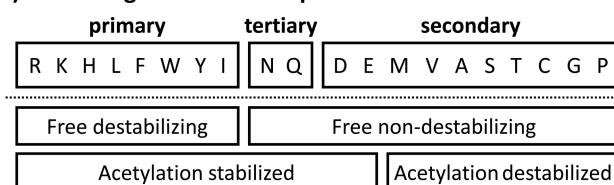


Figure 5. (a) The classical N-end rule pathway in mammals. Primary, secondary, and tertiary denote mechanistically distinct subsets of destabilizing N-terminal residues. The tertiary destabilizing residues Asn and Gln are enzymatically deamidated into the secondary destabilizing residues Asp and Glu. The arginylation of N-terminal Asp, Glu, and oxidized Cys (C*) leads to the formation of primary destabilizing residues. Primary destabilizing residues are recognized directly by components of the ubiquitin system and targeted for proteasome-mediated degradation. (b) Stabilizing N-end rule for processed protein N-termini. Proposed terminus stability classification categories based on an updated N-end rule and N-terminal acetylation status determined by TAILS analysis of human erythrocytes.⁶⁷ Unmodified primary residues are grouped as “free destabilizing,” whereas secondary and tertiary residues are classified as “free non-destabilizing.” After N α -acetylation, most secondary residues become destabilizing, whereas Asp and Glu together with primary and tertiary residues become stabilizing. Amino acid residues are indicated by single-letter abbreviations.

prior to recognition by the ubiquitin system, whereas tertiary destabilizing residues must be deamidated (asparagine, glutamine) or oxidized (cysteine) prior to arginylation. Importantly, most proteins are translated with an N-terminal degradation signal, but this is immediately masked by folding or protein–protein interactions, resulting in protected and stabilized proteins: Misfolded or misassembled proteins display their N-terminus and thus are promptly targeted by the ubiquitin proteasome pathway. For proteolytically processed protein N-termini in human erythrocytes, we observed, following an updated amino acid grouping, significantly higher levels of N α -acetylation among primary compared to secondary destabilizing residues. Based on these observations, we grouped N-terminal amino acids into four categories: (i) free destabilizing termini, including unmodified types 1 and 2 primary destabilizing residues; (ii) free nondestabilizing termini, comprising unmodified secondary and tertiary destabilizing residues; (iii) acetylation-destabilized termini, including acetylated secondary destabilizing residues; and (iv) acetylation-stabilized termini, including aspartic and glutamic acid, and primary destabilizing residues and tertiary destabilizing residues for which acetylation prevents deamidation and arginylation (Figure 5b).⁶⁷

Searching a customized protein database, Van Damme and colleagues showed that more than 20% of all identified protein N-termini in higher eukaryotes result from alternative translation initiation sites indicating incorrectly annotated or

overlooked protein starts and/or alternative splicing.⁹⁸ They combined information from the publicly available protein sequence database UniProtKB/Swiss-Prot with ribosome profiling and translation initiation mapping data to identify novel protein reading frames. This approach allowed the authors to identify more than 1700 potential alternative protein N-termini originating from in-frame, downstream translation initiation sites in human and murine cellular proteomes. Additionally, they identified 34 N-terminally extended protein products and several instances of N-terminally truncated proteins and overlapping upstream open reading frames resulting from translation initiation at near-cognate start codons, i.e. using GUG, CUG, or ACG for the initiator methionine instead of their canonical amino acids, valine, leucine, and threonine.

In depth TAILS analysis of murine skin,⁵³ human erythrocytes,⁶⁷ and platelets⁶⁸ revealed that more than 50% of all identified protein termini represent neo-N-termini derived from proteolytic processing. These findings reinforce the concept of limited proteolysis as an important and pervasive post-translational modification that cannot be neglected in interpreting any biological process.^{11,12} To integrate data on protein termini, their chemical and proteolytic modifications, and to provide functional insights into the biological processes involved, the knowledgebase TopFIND is most useful.^{10–12} This knowledgebase combines relevant information from experimental terminomics data, the protein resource UniProt, and the peptidase database MEROPS. In the latest release, TopFIND 3.0,¹² all biological processes leading to the formation of alternative protein termini, such as alternative splicing and translation initiation, were integrated by including data from Ensembl and TISdb¹² and analyzed recently.⁶⁹

Finally, targeted analysis of protein termini can be used as an alternative strategy for reducing sample complexity and hence facilitates identification of rare or low abundance proteins in the human proteome.^{67,68} The main goal of the Chromosome-centric Human Proteome Project (C-HPP) is to experimentally detect all protein species, ideally including their post-translational modifications, expressed from the human genome throughout different developmental stages.⁹⁹ With TAILS positional proteomics studies, we established a general workflow suitable for the in-depth and high-throughput analysis of position and nature of protein N-termini in different tissues and disease states, as envisioned by the HPP. Importantly, due to the altered *m/z*, ionization, and fragmentation properties of semitryptic N-terminal peptides in comparison to their fully tryptic counterparts originating from conventional shotgun proteomics,⁶⁷ we are using TAILS to identify hundreds of additional missing proteins.^{67,68}

CONCLUSIONS

The large amount of data generated in the past decade by using TAILS and COFRADIC strategies has positioned terminomics as a compelling new field of proteomics in the broader realm of degradomics.⁵⁹ By reducing sample complexity >95% by depletion of the internal peptides, these techniques have tremendously improved both the coverage and dynamic range of proteome analyses and provided mechanistically informative data on proteins by quantitative analysis of terminal peptides. Still, rare terminal modifications are often disregarded to limit the search space during peptide-spectrum matching or excluded by design, as in the case of pyroglutamate when using N-terminal COFRADIC. However, given their functional

relevance in physiological and pathological processes, development of specific enrichment strategies for these rare modifications is important. The integration of quantitative labeling strategies and constant improvement of mass spectrometers has led to an increasing interest in positional proteomics as the dynamics of proteolytic processing can now be studied. Multiplexing with isotopic variants of formaldehyde dimethylation or isobaric tags such as iTRAQ or TMT is the current gold standard, allowing the comparison of up to 10 different samples or conditions in a single experiment. Recent studies include N-terminome changes during skin inflammation⁵³ and platelet storage,⁶⁸ and profiling studies comparing terminal changes in health and disease and during medical treatment and for the discovery of novel protease substrates are underway in multiple laboratories. These studies will lead to new biological insights into physiological and pathological processes at a systems level, probably establishing proteolytic signatures as candidate biomarkers of disease.¹⁰⁰ The limitation that most currently used multiplexing strategies target primary amines, which renders termini lacking a free N-terminus (e.g., due to *N*-acetylation) and internal lysines (ϵ -amines) unquantifiable, will be overcome by ongoing developments in the near future. Another exciting direction is the emerging use of C-terminomics, particularly in nonenriched shotgun analyses using the archaeal endopeptidase LysargiNase.⁵⁰ Whereas C-terminal peptides generated by trypsin inherently lack basic residues, LysargiNase mirrors trypsin specificity resulting in protein C-terminal peptides with an N-terminal lysine or arginine residue, thus tremendously increasing the likelihood of identifying protein C-termini by shotgun mass spectrometry. The end game is ahead of us, in which we will witness the next transformation of terminomics from a dark horse in proteomics to an indispensable component of clinical diagnosis and personalized medicine.

AUTHOR INFORMATION

Corresponding Author

*E-mail: chris.overall@ubc.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.M.O. holds a Canada Research Chair in Proteinase Proteomics and Systems Biology. G.M. and U.E. were supported by a CBR Internal Collaborative Training Award and a postdoctoral fellowship from the Michael Smith Foundation for Health Research (MSFHR), respectively. The authors thank G. Butler for outstanding editing assistance, and all members of the Overall lab for valuable discussions, feedback, and support.

NOMENCLATURE

BOTTOM-UP PROTEOMICS, all fields of proteomics in which a biological sample is subjected to proteolytic cleavage into peptides (e.g., by trypsin, GluC, LysargiNase) prior to its analysis by mass spectrometry, whereas intact proteins and protein complexes are analyzed in top-down proteomics; COFRADIC, COMbined FRActional Diagonal Chromatography is a positional proteomics technique based on a series of orthogonal liquid chromatographic separations to negatively select for protein termini prior to MS/MS analysis; DEGRADOMICS, the system-wide study of proteases and

substrates, by employing high-throughput technologies such as genomics, transcriptomics, and proteomics; N-TAILS, Terminal Amine Isotopic Labeling of Substrates is a positional N-terminal proteomics technique harnessing whole protein amine labeling that inherently labels the natural and cleaved neo-termini, but not tryptic peptide N-termini, which are generated after labeling and digestion, labeled N-termini are separated from unlabeled tryptic peptide N-termini by aldehyde-derivatized polymers and thus negatively enrich for protein termini prior to MS/MS analysis; PROTEOLYTIC PROCESSING, specific proteolytic cleavage of target proteins at one or two sites in a living system, irreversibly modulating protein biological function, as opposed to unspecific protein degradation responsible for protein turnover; PROTEOFORMS, all molecular variants of a protein product stemming from one single gene, e.g. by genetic variations, alternative splicing, and co- and post-translational modifications, and combinations thereof, giving rise to squillions of protein forms from ~20 000 genes in humans; PROTEOLYTIC SIGNATURES, ensembles of characteristic protein termini originating from altered proteolytic networks in the protease web, which provide qualitative and quantitative information in physiological and pathological processes; TERMINAL MODIFICATIONS, N- or C-terminal protein modifications occurring during protein translation (cotranslational modifications) or after protein synthesis (post-translational modifications), including chemical modifications (e.g., N α -acetylation, N α -methylation, cyclization) and limited proteolysis (e.g., N-terminal methionine excision, signal peptide removal, zymogen maturation, signal molecule processing); TERMINOME, ensemble of all protein N- and C-termini present in a proteome, comprising original protein termini and neo-termini generated by proteolytic processing; TERMINOMICS or positional proteomics, a subfield of bottom-up proteomics dedicated to the study of the terminome by high-throughput technologies; peptides carrying protein N- or C-termini are enriched from complex whole proteome digests and are identified by mass spectrometry

■ REFERENCES

- (1) Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M. L. (2014) Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* 23, 5866–5878.
- (2) Kelleher, N. L. (2012) A cell-based approach to the human proteome project. *J. Am. Soc. Mass Spectrom.* 23, 1617–1624.
- (3) Smith, L. M., and Kelleher, N. L. (2013) Proteoform: a single term describing protein complexity. *Nat. Methods* 10, 186–187.
- (4) Deribe, Y. L., Pawson, T., and Dikic, I. (2010) Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* 17, 666–672.
- (5) Woodsmith, J., Kamburov, A., and Stelzl, U. (2013) Dual Coordination of Post Translational Modifications in Human Protein Networks. *PLoS Comput. Biol.* 9, e1002933.
- (6) Mann, M., and Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255–261.
- (7) Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* 16, 589–597.
- (8) Giglione, C., Fioulaine, S., Meinel, T. N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie*. In press.
- (9) Jacob, E., and Unger, R. (2007) A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics* 23, e225–e230.
- (10) Lange, P. F., and Overall, C. M. (2011) TopFIND, a knowledgebase linking protein termini with function. *Nat. Methods* 8, 703–704.
- (11) Lange, P. F., Huesgen, P. F., and Overall, C. M. (2012) TopFIND 2.0—linking protein termini with proteolytic processing and modifications altering protein function. *Nucleic Acids Res.* 40, D351–D361.
- (12) Fortelny, N., Yang, S., Pavlidis, P., Lange, P. F., and Overall, C. M. (2015) Proteome TopFIND 3.0 with TopFINDER and PathFINDER: database and analysis tools for the association of protein termini to pre- and post-translational events. *Nucleic Acids Res.* 43, D290–D297.
- (13) Lange, P. F., and Overall, C. M. (2013) Protein TAILS: when termini tell tales of proteolysis and function. *Curr. Opin. Chem. Biol.* 17, 73–82.
- (14) Van Damme, P., Arnesen, T., and Gevaert, K. (2011) Protein alpha-N-acetylation studied by N-terminomics. *FEBS J.* 278, 3822–3834.
- (15) Van Damme, P., Evjenth, R., Foyn, H., Demeyer, K., De Bock, P.-J., Lillehaug, J. R., Vandekerckhove, J., Arnesen, T., and Gevaert, K. (2011) Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. *Mol. Cell. Proteomics*, DOI: 10.1074/mcp.M110.004580.
- (16) Helsens, K., Van Damme, P., Degroove, S., Martens, L., Arnesen, T., Vandekerckhove, J., and Gevaert, K. (2011) Bioinformatics analysis of a *Saccharomyces cerevisiae* N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation. *J. Proteome Res.* 10, 3578–3589.
- (17) Van Damme, P., Lasa, M., Polevoda, B., Gazquez, C., Elosegui-Artola, A., Kim, D. S., De Juan-Pardo, E., Demeyer, K., Hole, K., Larrea, E., Timmerman, E., Prieto, J., Arnesen, T., Sherman, F., Gevaert, K., and Aldabe, R. (2012) N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *Proc. Natl. Acad. Sci. U. S. A.* 109, 12449–12454.
- (18) Dormeyer, W., Mohammed, S., Breukelen, B. van, Krijgsveld, J., and Heck, A. J. R. (2007) Targeted Analysis of Protein Termini. *J. Proteome Res.* 6, 4634–4645.
- (19) Helbig, A. O., Rosati, S., Pijnappel, P. W., Breukelen, B. van, Timmers, M. H., Mohammed, S., Slijper, M., and Heck, A. J. (2010) Perturbation of the yeast N-acetyltransferase NatB induces elevation of protein phosphorylation levels. *BMC Genomics* 11, 685.
- (20) Hollebeke, J., Van Damme, P., and Gevaert, K. (2012) N-terminal acetylation and other functions of N α -acetyltransferases. *Biol. Chem.* 393, 291–298.
- (21) Kim, H.-K., Kim, R.-R., Oh, J.-H., Cho, H., Varshavsky, A., and Hwang, C.-S. (2014) The N-Terminal Methionine of Cellular Proteins as a Degradation Signal. *Cell* 156, 158–169.
- (22) Shemorry, A., Hwang, C.-S., and Varshavsky, A. (2013) Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Mol. Cell* 50, 540–551.
- (23) Hwang, C.-S., Shemorry, A., and Varshavsky, A. (2010) N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science* 327, 973–977.
- (24) Foyn, H., Damme, P. V., Støve, S. I., Glomnes, N., Evjenth, R., Gevaert, K., and Arnesen, T. (2013) Protein N-terminal Acetyltransferases Act as N-terminal Propionyltransferases In Vitro and In Vivo. *Mol. Cell. Proteomics* 12, 42–54.
- (25) Zhang, X., Ye, J., and Højrup, P. (2009) A proteomics approach to study in vivo protein N α -modifications. *J. Proteomics* 73, 240–251.
- (26) Brosius, J., and Chen, R. (1976) The primary structure of protein L16 located at the peptidyltransferase center of *Escherichia coli* ribosomes. *FEBS Lett.* 68, 105–109.
- (27) Chen, T., Muratore, T. L., Schaner-Tooley, C. E., Shabanowitz, J., Hunt, D. F., and Macara, I. G. (2007) N-terminal α -methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* 9, 596–603.

- (28) Stock, A., Clarke, S., Clarke, C., and Stock, J. (1987) N-terminal methylation of proteins: structure, function and specificity. *FEBS Lett.* 220, 8–14.
- (29) Petkowski, J. J., Schaner Tooley, C. E., Anderson, L. C., Shumilin, I. A., Balsbaugh, J. L., Shabanowitz, J., Hunt, D. F., Minor, W., and Macara, I. G. (2012) Substrate Specificity of Mammalian N-Terminal α -Amino Methyltransferase NRMT. *Biochemistry (Mosc.)* 51, 5942–5950.
- (30) Kumar, A., and Bachhawat, A. K. (2012) Pyroglutamic acid: throwing light on a lightly studied metabolite. *Curr. Sci. India* 102, 288–297.
- (31) Seifert, F., Schulz, K., Koch, B., Manhart, S., Demuth, H.-U., and Schilling, S. (2009) Glutaminyl Cyclases Display Significant Catalytic Proficiency for Glutamyl Substrates. *Biochemistry (Mosc.)* 48, 11831–11833.
- (32) Shih, Y.-P., Chou, C.-C., Chen, Y.-L., Huang, K.-F., and Wang, A. H.-J. (2014) Linked Production of Pyroglutamate-Modified Proteins via Self-Cleavage of Fusion Tags with TEV Protease and Autonomous N-Terminal Cyclization with Glutaminyl Cyclase In Vivo. *PLoS One* 9, e94812.
- (33) Becker, A., Kohlmann, S., Alexandru, A., Jagla, W., Canneva, F., Bäuscher, C., Cynis, H., Sedlmeier, R., Graubner, S., Schilling, S., Demuth, H.-U., and von Hörsten, S. (2013) Glutaminyl cyclase-mediated toxicity of pyroglutamate-beta amyloid induces striatal neurodegeneration. *BMC Neurosci.* 14, 108.
- (34) Thinon, E., Serwa, R. A., Broncel, M., Brannigan, J. A., Brassat, U., Wright, M. H., Heal, W. P., Wilkinson, A. J., Mann, D. J., and Tate, E. W. (2014) Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* 5, No. 4919.
- (35) Linder, M. E., and Deschenes, R. J. (2007) Palmitoylation: policing protein stability and traffic. *Nat. Rev. Mol. Cell Biol.* 8, 74–84.
- (36) Resh, M. D. (2006) Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat. Chem. Biol.* 2, 584–590.
- (37) Young, F. B., Butland, S. L., Sanders, S. S., Sutton, L. M., and Hayden, M. R. (2012) Putting proteins in their place: Palmitoylation in Huntington disease and other neuropsychiatric diseases. *Prog. Neurobiol.* 97, 220–238.
- (38) Wright, M. H., Heal, W. P., Mann, D. J., and Tate, E. W. (2009) Protein myristoylation in health and disease. *J. Chem. Biol.* 3, 19–35.
- (39) Schilling, O., Barré, O., Huesgen, P. F., and Overall, C. M. (2010) Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat. Methods* 7, 508–511.
- (40) Van Damme, P., Staes, A., Bronsoms, S., Helsens, K., Colaert, N., Timmerman, E., Aviles, F. X., Vandekerckhove, J., and Gevaert, K. (2010) Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nat. Methods* 7, 512–515.
- (41) Tanco, S., Gevaert, K., and Van Damme, P. (2015) C-terminomics: Targeted analysis of natural and posttranslationally modified protein and peptide C-termini. *Proteomics* 15, 903–914.
- (42) Chung, J.-J., Shikano, S., Hanyu, Y., and Li, M. (2002) Functional diversity of protein C-termini: more than zipcoding? *Trends Cell Biol.* 12, 146–150.
- (43) Berndt, N., Hamilton, A. D., and Sebt, S. M. (2011) Targeting protein prenylation for cancer therapy. *Nat. Rev. Cancer* 11, 775–791.
- (44) Kim, K.-H., and Seong, B. L. (2001) Peptide amidation: Production of peptide hormones in vivo and in vitro. *Biotechnol. Bioprocess Eng.* 6, 244–251.
- (45) Prigge, S. T., Mains, R. E., Eipper, B. A., and Amzel, L. M. (2000) New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. *Cell. Mol. Life Sci.* 57, 1236–1259.
- (46) Gigoux, V., Escrieut, C., Fehrentz, J. A., Poirot, S., Maigret, B., Moroder, L., Gully, D., Martinez, J., Vaysse, N., and Fourmy, D. (1999) Arginine 336 and asparagine 333 of the human cholecystokinin-A receptor binding site interact with the penultimate aspartic acid and the C-terminal amide of cholecystokinin. *J. Biol. Chem.* 274, 20457–20464.
- (47) Edison, A. S., Espinoza, E., and Zachariah, C. (1999) Conformational ensembles: the role of neuropeptide structures in receptor binding. *J. Neurosci. Off. J. Soc. Neurosci.* 19, 6318–6326.
- (48) Sontag, J.-M., and Sontag, E. (2014) Protein phosphatase 2A dysfunction in Alzheimer's disease. *Front. Mol. Neurosci.* 7, 16.
- (49) Stock, J. B., and Zhang, S. (2013) The biochemistry of memory. *Curr. Biol.* 23, R741–R745.
- (50) Huesgen, P. F., Lange, P. F., Rogers, L. D., Solis, N., Eckhard, U., Kleifeld, O., Goulas, T., Gomis-Rüth, F. X., and Overall, C. M. (2015) LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification. *Nat. Methods* 12, 55–58.
- (51) Meinel, T., and Giglione, C. (2008) Tools for analyzing and predicting N-terminal protein modifications. *Proteomics* 8, 626–649.
- (52) Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* 21, 566–569.
- (53) Auf dem Keller, U., Prudova, A., Eckhard, U., Fingleton, B., and Overall, C. M. (2013) Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation. *Sci. Signal.* 6, rs2.
- (54) Doucet, A., Butler, G. S., Rodríguez, D., Prudova, A., and Overall, C. M. (2008) Metadegradomics Toward in Vivo Quantitative Degradomics of Proteolytic Post-translational Modifications of the Cancer Proteome. *Mol. Cell. Proteomics* 7, 1925–1951.
- (55) Adams, R. L. C., and Bird, R. J. (2009) Review article: Coagulation cascade and therapeutics update: Relevance to nephrology. Part 1: Overview of coagulation, thrombophilias and history of anticoagulants. *Nephrology* 14, 462–470.
- (56) Holers, V. M. (2014) Complement and Its Receptors: New Insights into Human Disease. *Annu. Rev. Immunol.* 32, 433–459.
- (57) LeMosy, E. K., Hong, C. C., and Hashimoto, C. (1999) Signal transduction by a protease cascade. *Trends Cell Biol.* 9, 102–107.
- (58) Cryns, V., and Yuan, J. (1998) Proteases to die for. *Genes Dev.* 12, 1551–1570.
- (59) López-Otin, C., and Overall, C. M. (2002) Protease degradomics: A new challenge for proteomics. *Nat. Rev. Mol. Cell Biol.* 3, 509–519.
- (60) McQuibban, G. A., Gong, J.-H., Wong, J. P., Wallace, J. L., Clark-Lewis, I., and Overall, C. M. (2002) Matrix metalloproteinase processing of monocyte chemoattractant proteins generates CC chemokine receptor antagonists with anti-inflammatory properties in vivo. *Blood* 100, 1160–1167.
- (61) Cox, J. H., and Overall, C. M. (2008) Cytokine Substrates: MMP Regulation of Inflammatory Signaling Molecules, in *The Cancer Degradome: Proteases and Cancer Biology*, pp 517–538. Springer Science & Business Media.
- (62) Hayashida, K., Bartlett, A. H., Chen, Y., and Park, P. W. (2010) Molecular and Cellular Mechanisms of Ectodomain Shedding. *Anat. Rec., Adv. Integr. Anat. Evol. Biol.* 293, 925–937.
- (63) Clark, P. (2014) Protease-mediated ectodomain shedding. *Thorax* 69, 682–684.
- (64) Puente, X. S., Sánchez, L. M., Overall, C. M., and López-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* 4, 544–558.
- (65) Drag, M., and Salvesen, G. S. (2010) Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discovery* 9, 690–701.
- (66) Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discovery* 5, 785–799.
- (67) Lange, P. F., Huesgen, P. F., Nguyen, K., and Overall, C. M. (2014) Annotating N termini for the human proteome project: N termini and N α -acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome. *J. Proteome Res.* 13, 2028–2044.
- (68) Prudova, A., Serrano, K., Eckhard, U., Fortelny, N., Devine, D. V., and Overall, C. M. (2014) TAILS N-terminomics of human platelets reveals pervasive metalloproteinase-dependent proteolytic processing in storage. *Blood* 124, e49–e60.

- (69) Fortelny, N., Pavlidis, P., and Overall, C. M. (2015) The path of no return—Truncated protein N-termini and current ignorance of their genesis. *Proteomics*, DOI: 10.1002/pmic.201500043.
- (70) Wildes, D., and Wells, J. A. (2010) Sampling the N-terminal proteome of human blood. *Proc. Natl. Acad. Sci. U. S. A.* 107, 4561–4566.
- (71) Vögtle, F.-N., Wortelkamp, S., Zahedi, R. P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N., and Meisinger, C. (2009) Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* 139, 428–439.
- (72) Mahrus, S., Trinidad, J. C., Barkan, D. T., Sali, A., Burlingame, A. L., and Wells, J. A. (2008) Global Sequencing of Proteolytic Cleavage Sites in Apoptosis by Specific Labeling of Protein N Termini. *Cell* 134, 866–876.
- (73) Xu, G., Shin, S. B. Y., and Jaffrey, S. R. (2009) Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19310–19315.
- (74) Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005) Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nat. Methods* 2, 771–777.
- (75) Rogers, L. D., and Overall, C. M. (2013) Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol. Cell. Proteomics* 12, 3532–3542.
- (76) Huesgen, P. F., and Overall, C. M. (2012) N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. *Physiol. Plant.* 145, 5–17.
- (77) Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N., and Overall, C. M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* 28, 281–288.
- (78) Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J. N., and Overall, C. M. (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* 6, 1578–1611.
- (79) Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* 8, 1362–1370.
- (80) Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011) Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat. Protoc.* 6, 1130–1141.
- (81) Venne, A. S., Vögtle, F.-N., Meisinger, C., Sickmann, A., and Zahedi, R. P. (2013) Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55. *J. Proteome Res.* 12, 3823–3830.
- (82) Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169.
- (83) Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003) Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904.
- (84) Guryčá, V., Lamerz, J., Ducret, A., and Cutler, P. (2012) Qualitative improvement and quantitative assessment of N-terminomics. *Proteomics* 12, 1207–1216.
- (85) Shen, P.-T., Hsu, J.-L., and Chen, S.-H. (2007) Dimethyl isotope-coded affinity selection for the analysis of free and blocked N-termini of proteins using LC-MS/MS. *Anal. Chem.* 79, 9520–9530.
- (86) Mommen, G. P. M., van de Waterbeemd, B., Meiring, H. D., Kersten, G., Heck, A. J. R., and de Jong, A. P. J. M. (2012) Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO(2)-based depletion. *Mol. Cell. Proteomics* 11, 832–842.
- (87) Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. R. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* 76, 3935–3943.
- (88) Yoshihara, H. A. I., Mahrus, S., and Wells, J. A. (2008) Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* 18, 6000–6003.
- (89) Auf dem Keller, U., Prudova, A., Gioia, M., Butler, G. S., and Overall, C. M. (2010) A Statistics-based Platform for Quantitative N-terminome Analysis and Identification of Protease Cleavage Products. *Mol. Cell. Proteomics* 9, 912–927.
- (90) Kuyama, H., Shima, K., Sonomura, K., Yamaguchi, M., Ando, E., Nishimura, O., and Tsunasawa, S. (2008) A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry. *Proteomics* 8, 1539–1550.
- (91) Nakajima, C., Kuyama, H., and Tanaka, K. (2012) Mass spectrometry-based sequencing of protein C-terminal peptide using α -carboxyl group-specific derivatization and COOH capturing. *Anal. Biochem.* 428, 167–172.
- (92) Xu, G., Shin, S. B. Y., and Jaffrey, S. R. (2011) Chemoenzymatic Labeling of Protein C-Termini for Positive Selection of C-Terminal Peptides. *ACS Chem. Biol.* 6, 1015–1020.
- (93) Nika, H., Hawke, D. H., and Angeletti, R. H. (2014) C-terminal protein characterization by mass spectrometry: isolation of C-terminal fragments from cyanogen bromide-cleaved protein. *J. Biomol. Technol.* 25, 1–18.
- (94) Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J. E., Vandekerckhove, J., Lillehaug, J. R., Sherman, F., and Gevaert, K. (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8157–8162.
- (95) Van Damme, P., Hole, K., Pimenta-Marques, A., Helsens, K., Vandekerckhove, J., Martinho, R. G., Gevaert, K., and Arnesen, T. (2011) NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet.* 7, e1002169.
- (96) Goetze, S., Qeli, E., Mosimann, C., Staes, A., Gerrits, B., Roschitzki, B., Mohanty, S., Niederer, E. M., Laczkó, E., Timmerman, E., Lange, V., Hafen, E., Aebersold, R., Vandekerckhove, J., Basler, K., Ahrens, C. H., Gevaert, K., and Brunner, E. (2009) Identification and Functional Characterization of N-Terminally Acetylated Proteins in *Drosophila melanogaster*. *PLoS Biol.* 7, e1000236.
- (97) Gonda, D. K., Bachmair, A., Wünnig, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.* 264, 16700–16712.
- (98) Van Damme, P., Gawron, D., Van Criekeing, W., and Menschaert, G. (2014) N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol. Cell. Proteomics* 13, 1245–1261.
- (99) Paik, Y.-K., Omenn, G. S., Uhlen, M., Hanash, S., Marko-Varga, G., Aebersold, R., Bairoch, A., Yamamoto, T., Legrain, P., Lee, H.-J., Na, K., Jeong, S.-K., He, F., Binz, P.-A., Nishimura, T., Keown, P., Baker, M. S., Yoo, J. S., Garin, J., Archakov, A., Bergeron, J., Salekdeh, G. H., and Hancock, W. S. (2012) Standard Guidelines for the Chromosome-Centric Human Proteome Project. *J. Proteome Res.* 11, 2005–2013.
- (100) Huesgen, P. F., Lange, P. F., and Overall, C. M. (2014) Ensembles of protein termini and specific proteolytic signatures as candidate biomarkers of disease. *Proteomics: Clin. Appl.* 8, 338–350.
- (101) Creasy, D. M., and Cottrell, J. S. (2004) Unimod: Protein modifications for mass spectrometry. *Proteomics* 4, 1534–1536.